scottish institute for research in economics

SIRE DISCUSSION PAPER

SIRE-DP-2010-13

Multi-agent contracting with countervailing incentives and limited liability

Daniel Danau

University of St Andrews

Annalisa Vinella

University of Bari. Italy

www.sire.ac.uk

Multi-agent contracting with countervailing incentives and limited liability*

Daniel Danau[†] Annalisa Vinella[‡]

Hereinafter entitled:

Contract design with countervailing incentives, correlated information and limited liability

Abstract

We consider a principal who deals with a privately informed agent protected by limited liability in a correlated information setting. The agent's technology is such that the fixed cost declines with the marginal cost (the type), so that countervailing incentives may arise. We show that, with high liability, the first-best outcome can be effected for any type if (1) the fixed cost is non-concave in type, under the contract that yields the smallest feasible loss to the agent; (2) the fixed cost is not very concave in type, under the contract that yields the maximum sustainable loss to the agent. We further show that, with low liability, the first-best outcome is still implemented for a non-degenerate range of types if the fixed cost is less concave in type than some given threshold, which tightens as the liability reduces. The optimal contract entails pooling otherwise.

Keywords: Countervailing incentives; Limited liability; Correlation; Pooling
J.E.L. Classification Numbers: D82

Revised draft date: February 15, 2010

*We gratefully acknowledge useful comments from participants at the first NERI Meeting (Pescara), the XXI Siep Meeting (Pavia), the 2009 ASSET Meeting (Istanbul) as well as seminar participants at the University of St Andrews, the Max-Planck Institute for Research on Collective Goods (Bonn) and the SIRE FORUM (Edinburgh). The usual disclaimer applies.

†University of St Andrews, School of Economics & Finance, Castlecliffe, The Scores, Fife, KY16 9AL (UK). E-mail: Daniel.Danau@st-andrews.ac.uk

‡University of Bari, Department of Economics and Mathematics, Via C. Rosalba 53, 70124 Bari (Italy). E-mail: a.vinella@dse.uniba.it

Contract design with countervailing incentives, correlated information and limited liability

1 Introduction

In some agency problems, the agent displays countervailing incentives, *i.e.* the temptation either to overstate or to understate his private information (the type) in the report to the principal, depending upon its specific realization. For instance, Lewis and Sappington [6] show that countervailing incentives arise in regulator-firm hierarchies when the firm/agent's production technology is such that the fixed cost depends negatively on the marginal cost, which is unknown to the regulator/principal. As Maggi and Rodriguez-Clare [8] point out, this is a plausible case because, in regulated sectors, low marginal costs are likely associated with high overhead costs.

The literature on contract design in the presence of countervailing incentives has neglected the possibility that some additional piece of information, correlated with the agent's type, become publicly available after the contract is drawn up. This approach appears to be restrictive with regards to various real-world contexts in which countervailing incentives may arise. For instance, a national regulator dealing with some local monopoly (say, a public utility) under *ex ante* asymmetric information is generally able to acquire some more knowledge over time, say, by observing the performance of regulated monopolies in neighboring countries, who share stochastically some information with the home country monopoly. This can affect the contractual strategies at hand and, hence, the achievements of the national regulator.

In this article, we characterize the optimal contract between a principal and an agent who may have countervailing incentives, focusing on situations where the agent's type is correlated with some external signal that is observed after the contract has been signed and the agent has chosen his production for the principal. To make the analysis truly realistic, we consider the case in which the agent is limitedly liable, that is also typical of regulatory settings, for instance. We thus extend the research work about countervailing incentives to correlated information settings with limited liability on the agent's side[1].

From the literature we learn that information correlation enables the principal to retain surplus from the agent (Crémer and McLean [2], Riordan and Sappington [9]). However, limited liability can reduce the benefits from correlation and make rent extraction more difficult (Demougin and Garvie [3], Gary-Bobo and Spiegel [4]). The literature further shows that the presence of countervailing incentives tends to relax informational constraints (Lewis and Sappington [6], Maggi and Rodriguez-Clare [8] etc.). This suggests

[1]In certain environments, the perspective of new correlated information appearing at later stage may induce the principal to delay contracting. However, this possibility may not apply to the settings we have in mind mainly. For instance, regulators may not be in a position to delay contracting with regulated firms that provide services of general interest. Anyway, it would be beyond the scope of the present paper to tackle issues related to strategic waiting.

that also countervailing incentives may help the principal in the relationship with the agent. That is, they may operate in the same direction as correlation. One may thus wonder how the principal can still exploit correlation and what she can achieve in so doing in environments where the agent displays countervailing incentives but is protected by limited liability. Besides, from the literature we know that, under some circumstances, countervailing incentives induce pooling in the optimal contract (Lewis and Sappington [6], Maggi and Rodriguez-Clare [8] etc.). Therefore, another open question is how information correlation (which is beneficial to the principal) and limited liability (which is detrimental) affect the screening ability of the principal in the presence of countervailing incentives. Our study is meant to address these issues.

At this aim, we construct a model that brings together countervailing incentives, information correlation and limited liability as core ingredients. In the model, countervailing incentives may arise because the agent's production technology is taken to include a fixed cost that declines with the privately known marginal cost (the type), a representation inspired to that of Lewis and Sappington [6][2]. Besides, the agent's type is known to be correlated with a random signal observable *ex post*, a feature that has only appeared in standard models with systematic incentives to misrepresent type so far (Riordan and Sappington [9], Demougin and Garvie [3], Gary-Bobo and Spiegel [4])[3]. Finally, the agent can only sustain bounded financial losses *ex post*, in the same vein as in Demougin and Garvie [3] and Gary-Bobo and Spiegel [4], among others.

To begin with, our analysis evidences how the incentive scheme that yields the *M*inimum *F*easible *L*oss to the agent (the MFL scheme), and is thus most likely to implement the first-best outcome under limited liability, is specifically affected by the existence of countervailing incentives. In a setting with systematic incentives to overstate type, Gary-Bobo and Spiegel [4] show that, under a scheme of this kind, all agent's types are rewarded if some critical signal is observed and incur an equal deficit whenever it is not, the critical signal being chosen such that the deficit is minimized. A similar result is obtained in a framework where countervailing incentives may arise, with the difference that the signal is no longer the same for all possible types. To understand this difference, one can interpret the finding of Gary-Bobo and Spiegel [4] as follows. As long as the agent exhibits systematic incentives to overstate type, the principal picks the signal that is most likely to

[2] The modelling device we use for the technology, which fits particularly well the regulatory settings mentioned in the text, is actually flexible enough to sketch circumstances other than technological features *stricto sensu*. As Lewis and Sappington [6] evidence, it further captures the possibility that valuable managerial skills be associated with profitable outside options. Countervailing incentives also appear when the agent's reservation utility is decreasing with his productivity in the activity performed for the principal (compare, for instance, Lewis and Sappington [7], Maggi and Rodriguez-Clare [8], Brainard and Martimort [1]; Jullien [5]). As an application, Lewis and Sappington [7] describe the relationship between a landowner who endows her farmer with some capital grant at the outset of their agreement, which limits the farmer's incentive to exaggerate production costs. Jullien [5] provides further applications concerning linear and nonlinear pricing. These are all situations in which it is very likely that the principal can exploit some external piece of information correlated with the agent's type.

[3] Actually, in our model, new information is conveyed by a purely informational signal as in Riordan and Sappington [9] and Demougin and Garvie [3]. In Gary-Bobo and Spiegel [4], the agent's type is correlated with a signal that is, in fact, a shock affecting the agent's cost itself.

be drawn by *higher* types. Indeed, for some given type, the probability of reward is lower the higher the likelihood of reward for higher types. Thus, a lower loss suffices to remove that type's incentives to over-report. The more likely the type is to incur a deficit, the smaller the deficit the principal needs to impose to prevent mimicking. In a framework with countervailing incentives, the result of Gary-Bobo and Spiegel [4] carries over for a low-marginal-cost agent, who has an incentive to over-report, but it needs to be amended for a high-marginal-cost agent, who has an incentive to under-report. Actually, for the latter, minimizing the deficit requires that the reward be assigned when the signal that is most likely to be drawn by *lower* types is observed. Overall, in the presence of countervailing incentives, whether the agent is rewarded or bears a loss depends not only on the signal realization but also on his type realization.

Beside characterizing the MFL scheme in our framework, we identify a sufficient condition on the agent's cost function for it to implement first best. We find that the MFL contract entails full efficiency if the cost function (or, more precisely, the fixed cost) is non-concave in type, provided the conditional likelihood function of the signal is concave in the agent's type. This means that, as far as FB is to be effected by the MFL scheme, not only the presence of countervailing incentives adds requirements in terms of relevant signals, it also tightens the condition on costs. This can be seen by comparing our finding with those of Riordan and Sappington [9]. In a model similar to that of Gary-Bobo and Spiegel [4] but without limited liability concerns, these authors assess that one simple way to implement first best is to use the same payoff scheme as in Gary-Bobo and Spiegel [4], except that the relevant signal is not necessarily chosen to minimize the loss. They show that such a scheme does effect first best if the agent's cost function is less concave in type than so is the conditional likelihood function of the relevant signal, a condition that is surely satisfied in the model of Gary-Bobo and Spiegel [4] where the agent's cost function is taken to be strictly convex. Our result differs from that of Riordan and Sappington [9] for the following reason. In a setting where countervailing incentives arise, there exists some intermediary type that displays no incentive to cheat because it produces at highest *total* costs. From the principal's perspective, this is the least efficient type. Under the MFL contract, this type is assigned a payoff equal to zero, whatever the signal. That is, information correlation plays no role in the payoff profile designed for the least efficient type. Because of this, the principal would be unable to extract surplus from other types that were to mimic this particular type. It turns out that, with a concave cost function, the MFL scheme leaves all other types with an incentive to actually mimic this particular type.

The next contribution of our research rests on the observation that, under the MFL contract, first best can only arise when the penalty it yields is smaller than the maximum sustainable deficit for all types (or, at the limit, it equals that deficit for some types), but this penalty results independently of the agent's actual liability. Because of this, the principal takes less advantage of type correlation than she would if any larger sustainable deficit were assigned to the agent. This aspect is especially relevant in environments with countervailing incentives. Indeed, in the latter, intermediary types are assigned particularly low losses and this contractual offer may become attractive for other types. To circumvent this difficulty, we propose an alternative scheme under which the minimum feasible deficit is replaced with the Maximum Sustainable Loss (from which MSL scheme) for all types, including the one that produces at highest total costs. We show that the MSL scheme is the best possible contractual option for the principal as it allows to exploit type correlation at maximum, given the agent's liability. The benefit is that first best is effected under milder conditions. While under the MFL contract first best arises if the cost function is non-concave in type, under the MSL scheme it arises if the cost function is "not excessively" concave, a condition that is relaxed the more liable the agent is.

When the minimum feasible loss exceeds the maximum deficit the agent can sustain, first best cannot be implemented (at least for some types). Our subsequent contribution is to characterize the optimal second-best contract for this case. We find that, once again, the contractual features heavily depend both on the nature of the agent's incentives and on the cost characteristics that determine their intensity. Two relevant situations can arise, depending on the curvature of the cost function with respect to type.

If the cost function is less concave in type than some relevant threshold (the first possible situation), then there exist some types for which first best is still effected. This outcome follows from the possibility to exploit type correlation by inflicting (bounded) losses in the presence of countervailing incentives. It does not appear in correlated information environments with systematic incentives to over-report, as represented by Gary-Bobo and Spiegel [4], in which tight limited liability prevents first-best implementation for *any* type. In our model, first best survives for a continuum range of intermediate types neighboring the least efficient one. This is explained by considering that, because such types display weak incentives to cheat, as they are turned between the desire to over-report and that to under-report, the principal does not need to assign large losses (and rewards) to induce them to truthtell. Therefore, as far as intermediate types are concerned, the limits on liability remain irrelevant in the contractual design. The first-best outcome is beyond reach for all remaining types, instead. Moving away from the intermediate types, lower and higher types exhibit increasingly stronger incentives to over and under-report respectively. This involves that, for the types immediately below and above the intermediate ones, the quantity is distorted just enough to retain all surplus and, at the same time, to solicit information release and satisfy the limits on liability. On the other hand, an information rent is conceded to very low and very high types. As usual, this rent is contained by distorting the quantity till the ensuing loss exactly compensates the surplus extraction gain (the familiar efficiency/rent-extraction trade-off). The second-best contract in our framework compares with that in Gary-Bobo and Spiegel [4] with sole regards to this last case. More precisely, the similarity concerns the contract designed for low types with intense incentives to over-report. This is so because the countervailing effect is weak for such types, so that the principal faces a (nearly) standard adverse selection problem.

If the cost function is more concave in type than the relevant threshold aforementioned (the second possible situation), then also this cost characteristics, and not only the limits on liability, has an impact on contractual performance. The incentive problem is exacerbated to the point that information is not released unless the principal induces pooling in the contract, *i.e.* an inflexible rule for some given bunch of types. Under this rule, the quantity that is efficient for the type that has no incentive to cheat (the least efficient type) is assigned to all types in its neighborhood. As that type is also the sole from which all surplus is retained *ex ante*, it is the sole for which the first-best outcome is still enforced. Comparing with Lewis and Sappington [6] - [7], it emerges that this incentive scheme is similar in structure to the contract that is optimal in uncorrelated information settings with countervailing incentives when the fixed cost is concave in type.

Further comparing the whole bulk of our results with those obtained in uncorrelated information contexts, we are able to shed light on how the presence of correlation and liability affects the "knife-edge" situation between pooling and separating equilibria. This is the last contribution of our study. Maggi and Rodriguez-Clare [8] show that, in uncorrelated information settings, the knife-edge situation is represented by the case of linear fixed cost (linear reservation utility, in their model), in which the optimal contract entails pooling and no rent for a range of intermediate types. Pooling is removed as soon as the fixed cost becomes convex. It persists with all types but one getting a rent as soon as the fixed cost becomes concave. According to our results, in correlated information environments, the linear case would still be the knife-edge situation if the agent were unable to sustain any deficit *ex post*. With the agent (limitedly) liable, it is rather given by the concavity threshold we mentioned to distinguish the two situations that can be realized with tight limited liability. Importantly, we find that this threshold relaxes as the maximum loss the agent can bear raises. We thus conclude that the possibility to take advantage of correlation by inflicting penalties to the agent removes pooling in a class of situations in which it would otherwise arise, *i.e.* in contexts where technologies are such that the fixed cost is concave but not too concave in type, and that this class enlarges as the agent's pocket becomes deeper.

The reminder of the article is organized as follows. In section 2, we present the model. Section 3 focuses on implementation of the first-best outcome. In section 4, we characterize the optimal contract for the case of tight limited liability. Section 5 concludes. Mathematical details are relegated to the Appendix.

2 The model

A risk-neutral principal P contracts with a risk-neutral agent for the production of q units of some good. Production costs are given by

$$C(q, c) = cq + K(c), \tag{1}$$

where c is the marginal cost and $K(c)$ the fixed cost. Similarly to Lewis and Sappington [6], we take the fixed cost to decrease with the marginal cost ($K'(c) < 0$). However, unlike these authors who suppose concavity, we make no assumption about the curvature of $K(\cdot)$.

At the contracting stage, the agent is privately informed about c (the type). It is commonly known that c is drawn from the continuous support $[\underline{c}, \overline{c}]$ with density function $f(c)$ and cumulative distribution function $F(c)$. Moreover, the marginal cost is correlated with a random signal s that is drawn from the discrete support $N \equiv \{1, ..., n\}$ and is publicly observable *ex post.* We denote $p(c, s) = \Pr ob(s \mid c)$ the probability of observing the signal s conditional on type c. The larger the value of s, the more likely c is to be large in turn. We take $p(c, s)$ to be differentiable everywhere with respect to c.

2.1 The principal's programme

As usual, the Revelation Principle applies and attention can be restricted to direct revelation mechanisms in which the agent reports his true type. A mechanism designed for an agent of type c and some signal s is an allocation $\{q(c), t(c, s)\}$, with $q(c)$ the quantity to be produced and $t(c, s)$ the transfer to be paid. Under truthful reporting, the agent's *ex post* and *interim* profit are respectively given by

$$\pi(c, s) = t(c, s) - [cq(c) + K(c)] \tag{2a}$$

$$E_s[\pi(c, s)] \equiv \sum_{s=1}^{n} \{t(c, s) - [cq(c) + K(c)]\} p(c, s). \tag{2b}$$

Truthful reporting in a Bayesian setting is induced by satisfying the following incentive constraints

$$E_s[\pi(c, s)] \geq \sum_{s=1}^{n} \{t(r, s) - [cq(r) + K(c)]\} p(c, s), \ \forall c \in [\underline{c}, \overline{c}]. \tag{IC}$$

Besides, P needs to satisfy the participation constraints

$$E_s[\pi(c, s)] \geq 0, \ \forall c \in [\underline{c}, \overline{c}], \tag{PC}$$

and the limited liability constraints

$$\pi(c, s) \geq -L, \ \forall c \in [\underline{c}, \overline{c}], \ \forall s \in N, \tag{LL}$$

for some given $L \geq 0$.

Let $S(q(c))$ the gross utility P obtains when $q(c)$ units of the good are provided, with $S(0) = 0$, $S' > 0$, $S'' < 0$, $S'(0) = +\infty$ and $S'(+\infty) = 0$. P's objective is to achieve the highest attainable level of utility. The latter is taken to be a weighed sum of gross utility net of transfer, namely $V(q(c)) = S(q(c)) - t(c, s)$, and the agent's profit. Formally, P's programme is written:

$$\begin{aligned} \underset{\{q(c);\pi(c,s)\}}{Max}\ \widetilde{W} \quad &\equiv \quad \int_{\underline{c}}^{\overline{c}} \sum_{s=1}^{n} \left[V\left(q\left(c\right)\right) + \alpha\pi\left(c,s\right)\right] p\left(c,s\right) f\left(c\right) dc \\ &\text{subject to} \\ &\text{(IC), (PC) and (LL),} \end{aligned} \qquad (\Gamma)$$

with $\alpha \in [0,1]$[4].

3 First-best implementation

The first-best outcome (FB hereafter) entails whenever P can design transfers $t^{fb}(c,s)$ for the production of $q^{fb}(c)$ units of the good, such that $S'(q^{fb}(c)) = c$, and retain all surplus *ex ante* from the agent, *i.e.* $E_s[\pi^{fb}(c,s)] = 0$. In this section, we explore in which ways and under which conditions this is feasible, indeed.

To make (LL) most likely satisfied, a natural strategy for P is to offer the mechanism that minimizes the loss to be assigned to the agent for all possible types. Gary-Bobo and Spiegel [4] present this strategy in a framework in which the agent has a systematic incentive to over-report type. We hereafter adopt the same method to show how the mechanism that yields the minimum feasible loss is to be designed in the presence of countervailing incentives. We highlight that, in the context we consider, such a mechanism fails to be incentive compatible for a wide class of technologies. To circumvent this difficulty, we subsequently characterize an alternative incentive scheme that implements FB under milder conditions.

3.1 The Minimum-feasible-loss scheme

Before moving to the analysis, we observe that, under FB implementation, (IC) is conveniently replaced by the pair of conditions

$$q^{fb}(c) + K'(c) + \sum_{s=1}^{n} \pi^{fb}(c,s) \frac{dp(c,s)}{dc} = 0 \qquad \text{(LIC)}$$

$$\sum_{s=1}^{n} \left\{ t^{fb}(r,s) - cq^{fb}(r) - K(c) \right\} p(c,s) \leq 0, \qquad \text{(GIC)}$$

where $t^{fb}(r,s)$ is the FB transfer that P makes to the agent when the latter reports r and s is observed. (LIC) requires that the agent has no incentive to report $r \neq c$ in a neighborhood of his true type c (*local* incentive compatibility)[5]. (GIC) ensures that

[4] As standard in the literature, in a regulation context, transfers to agents (the regulated firms) can be thought of as made out of the public budget if products are public goods. In the case of private goods, $S(\cdot)$ can be interpreted as the gross consumer surplus for each product (*i.e.* the integral of the inverse demand function) and transfers to agents as including both the usage fees and the fixed fees paid by consumers (or, as an alternative to the latter, a subsidy made out of the public budget). In a procurement context, a natural choice would be to set $\alpha = 0$.

[5] (LIC) is obtained from standard calculations, that are reported in Appendix A.1.

the agent has no interest in reporting any $r \neq c$ within the feasible set (*global* incentive compatibility).

Consider now the reduced programme in which (LL) and (GIC) are neglected. It is defined as follows:

$$\begin{array}{c} \underset{\{q(c);\pi(c,s)\}}{Max} \widetilde{W} \\ s.t. \text{ (LIC) and (PC)}. \end{array} \qquad (\Gamma')$$

Let any solution to (Γ') the vector of profits $\Pi^{fb}(c) \equiv \left\{\pi^{fb}(c,1),...,\pi^{fb}(c,n)\right\}$ and the quantity $q^{fb}(c)$ for each given c. As there are more combinations of profits $\pi^{fb}(c,s)$ that solve (Γ') for each given c, define Ω the set of all vectors $\Pi^{fb}(c)$. Furthermore, define Φ the set that contains the lowest element of each vector $\Pi^{fb}(c)$. Finally, let $\pi^*(c)$ the largest element of Φ. This is the minimum feasible loss under which FB is implemented for any given c. Once $\pi^*(c)$ is identified, it is possible to identify also the specific set of profits, among all those in Ω, to which it belongs. Denote such a set $\Pi^*(c)$. This is the set of FB profits under which (LL) is least likely to be binding in the original programme (Γ). From now on, we refer to the incentive scheme that implements FB with profits in $\Pi^*(c)$ as to the "Minimum-feasible-loss" (MFL) scheme.

Lemma 1 *(Gary-Bobo and Spiegel [4]) Under the MFL scheme, for any* $c \in [\underline{c},\overline{c}]$, *the agent is rewarded whenever* s *takes some value* $s = \widetilde{s}(c)$ *and bears the smallest feasible loss whenever* $s \neq \widetilde{s}(c)$, *the loss being equal in size for all* $s \neq \widetilde{s}(c)$.

By this lemma, under the MFL scheme, the set $\Pi^*(c)$ reduces to only two values for each type c. As Gary-Bobo and Spiegel [4] explain, spreading punishments over as many realizations of s as possible (*i.e.*, all feasible realizations but one) allows P to minimize the highest possible loss for each type of agent. This requires that the largest reward-loss wedge that can be realized over all possible realizations of s be minimized.

Based on Lemma 1, for all $c \in [\underline{c},\overline{c}]$, the MFL pair of profits is found to be

$$\pi^{fb}(c,s)\Big|_{s=\widetilde{s}(c)} = \left[q^{fb}(c)+K'(c)\right]\frac{1-p(c,\widetilde{s}(c))}{dp(c,\widetilde{s}(c))/dc} \equiv \overline{\pi}^{fb}(c,\widetilde{s}(c)) \qquad (3)$$

$$\pi^{fb}(c,s)\Big|_{s\neq\widetilde{s}(c)} = \left[q^{fb}(c)+K'(c)\right]\frac{-p(c,\widetilde{s}(c))}{dp(c,\widetilde{s}(c))/dc} \equiv \underline{\pi}^{fb}(c,\widetilde{s}(c)), \qquad (4)$$

the difference $\left[\overline{\pi}^{fb}(c,\widetilde{s}(c)) - \underline{\pi}^{fb}(c,\widetilde{s}(c))\right]$ being the lowest feasible wedge when FB is implemented under (LIC) and (PC).

It is interesting to illustrate how $\widetilde{s}(c)$ should be selected in our framework, in which the agent does not exhibit a systematic incentive to misrepresent type. The choice of $\widetilde{s}(c)$ depends on the sign of the sum $q^{fb}(c)+K'(c)$, which may not be the same for all $c \in [\underline{c},\overline{c}]$ as $K'(c) < 0$. When this sum is positive, the situation is similar to that Gary-Bobo and Spiegel [4] consider (*i.e.* $K'(c) = 0$) and the same result obtains. For (4) to be a loss $(\underline{\pi}^{fb}(c,\widetilde{s}(c)) < 0)$, $\widetilde{s}(c)$ must be such that the probability of its realization raises with c, *i.e.* $dp(c, \widetilde{s}(c))/dc > 0$. Moreover, for (4) to be the smallest feasible loss, $r(c)$ must be such that the ratio $\frac{p(c,\widetilde{s}(c))}{dp(c,\widetilde{s}(c))/dc}$ is minimized. This is tantamount to requiring that the ratio $\frac{dp(c,\widetilde{s}(c))/dc}{p(c,\widetilde{s}(c))}$ be maximized. One can interpret this result (and hence that of Gary-Bobo and Spiegel [4]) by observing that $\frac{dp(c,\widetilde{s}(c))/dc}{p(c,\widetilde{s}(c))}$ is the rate of increase of the conditional likelihood that signal $\widetilde{s}(c)$ be drawn as c increases. This means that, for any given c, $\widetilde{s}(c)$ is to be the signal for which *higher* types are most likely to be drawn. Intuitively, because any type c that has an incentive to over-report is more likely to incur a deficit than higher types are, a smaller deficit suffices to remove that type's incentive to mimic. Similar reasoning applies, *mutatis mutandis*, when $q^{fb}(c) + K'(c) < 0$. In that case, $\widetilde{s}(c)$ must be such that its conditional likelihood decreases with c, *i.e.* $dp(c, \widetilde{s}(c))/dc < 0$, and the ratio $\frac{dp(c,\widetilde{s}(c))/dc}{p(c,\widetilde{s}(c))}$ is minimized. That is, for any given c, the agent is to be rewarded when the signal that is most likely to be drawn by *lower* possible types does materialize. By doing so, a smaller deficit can be imposed to remove incentives to under-report. The sole situation in which the value of $\widetilde{s}(c)$ is irrelevant arises when $q^{fb}(c) + K'(c) = 0$, in which case both (3) and (4) reduce to zero.

To identify the signal $\widetilde{s}(c)$ for each feasible c it is useful to make the following assumptions.

Assumption 1 $K''(c) < -\frac{dq^{fb}(c)}{dc}, \forall c \in [\underline{c}, \overline{c}]$.

Assumption 2 *The conditional likelihood function is such that:*

$$\frac{dp(c,n)}{dc} > 0 \quad and \quad \frac{d^2p(c,n)}{dc^2} < 0, \ \forall c \in [\underline{c}, \overline{c}] \tag{5}$$

$$\frac{dp(c,1)}{dc} < 0 \quad and \quad \frac{d^2p(c,1)}{dc^2} < 0, \ \forall c \in [\underline{c}, \overline{c}]. \tag{6}$$

Assumption 3 *The conditional likelihood function satisfies the following properties:*

$$\frac{d}{dc}\left(\frac{p(c,s)}{p(c,n)}\right) \leq 0, \ \forall c \in [\underline{c}, \overline{c}], \ \forall s \in N \tag{7}$$

$$\frac{d}{dc}\left(\frac{p(c,s)}{p(c,1)}\right) \geq 0, \ \forall c \in [\underline{c}, \overline{c}], \ \forall s \in N. \tag{8}$$

Assumption 1 is equivalent to saying that the sum $q^{fb}(c) + K'(c)$ decreases with c, so that it is positive for low types and possibly negative for high types. Negative values appear if and only if there exists some $\widehat{c} \in [\underline{c}, \overline{c}]$ at which $q^{fb}(c) + K'(c) = 0$. (5) (resp. (6)) in Assumption 2 tells that the probability of drawing $\widetilde{s}(c) = n$ (resp. $\widetilde{s}(c) = 1$) increases (resp. decreases) with type c at a decreasing rate. Assumption 2 is thus a requirement on the behaviour of $p(c, s)$ at two values of s, which we take to be n and 1. (7) (resp. (8)) in Assumption 3 requires that the rate of increase (resp. decrease) of the conditional likelihood that $s = n$ (resp. $s = 1$) be drawn is higher (resp. lower) than that of any other signal.

An assumption similar to (5) is found in both Riordan and Sappington [9] and Gary-Bobo and Spiegel [4]. The former explore a single-agent framework in which the agent's types are correlated with an *ex post* observable signal, the space of which is smaller than that of the possible types. With regard to this framework, they show that P can implement FB using information about a *unique* signal, provided the conditional likelihood function has analogous features to those described in (5) at that sole signal. Additionally, Gary-Bobo and Spiegel [4] identify the relevant signal (a shock affecting the cost, in their case) to be the highest feasible one and make an assumption analogous to (7) to warrant that this is indeed the state in which the agent should be rewarded for the loss to be minimized. Unlike in Riordan and Sappington [9] and Gary-Bobo and Spiegel [4], in our environment with countervailing incentives, it is necessary to refer to *two* signals, which explains the introduction of (6) and (8) on top of (5) and (7).

Our previous assumptions allow us to determine the values of $\widetilde{s}(c)$ in (3) and (4), as stated in the lemma hereafter.

Lemma 2 *Suppose there exists* $\widehat{c} \in [\underline{c}, \overline{c}]$ *such that* $q^{fb}(\widehat{c}) + K'(\widehat{c}) = 0$. *Then, under Assumption 1 - 3, in the MFL scheme, the agent is rewarded in two states of nature, namely* $\widetilde{s}(c) = n$ *if* $c \in [\underline{c}, \widehat{c})$ *and* $\widetilde{s}(c) = 1$ *if* $c \in (\widehat{c}, \overline{c}]$.

In what follows, we maintain the hypothesis that $\widehat{c}$ does exist, unless differently specified. Then, the MFL scheme is actually "region-specific", *i.e.* it is specifically characterized over different cost ranges, whereas this is not the case in Gary-Bobo and Spiegel [4][6].

Proposition 1 *Suppose*

$$L \geq \left[q^{fb}(c) + K'(c)\right] \frac{p(c, \widetilde{s}(c))}{dp(c, \widetilde{s}(c))/dc}, \ \forall c \in [\underline{c}, \overline{c}], \tag{9}$$

with $\widetilde{s}(c) = n$ *for* $c < \widehat{c}$ *and* $\widetilde{s}(c) = 1$ *for* $c < \widehat{c}$. *Then, under Assumption 1 - 3, the first-best outcome is implemented with ex post profits (3) and (4) if*

$$K"(c) \geq 0, \ \forall c \in [\underline{c}, \overline{c}]. \tag{10}$$

The conditions reported in Proposition 1 are explained as follows. Condition (9) follows from the circumstance that the agent cannot bear unbounded losses. The solution to (Γ) that is picked by the MFL scheme does not implement FB unless (9) is satisfied. This condition is similar to that in Proposition 2 of Gary-Bobo and Spiegel [4], although it specifies differently according to whether $c < \widehat{c}$ or $c > \widehat{c}$. Condition (10) suffices for the MFL payoff profile to be globally incentive compatible in (Γ). It requires that the fixed cost function be (weakly) convex in c.

Let us illustrate the intuition behind (10). The transfer an agent of type c receives when he reports r and $\widetilde{s}(c)$ is observed is given by

$$t(r, s) = rq^{fb}(r) + K(r) + \pi(r, s).$$

[6] The particular choice of n and 1 as the signals that trigger a reward is without loss of generality in the model. The properties of the likelihood function in Assumption 2 and 3, which ensure that n and 1 are the optimal reward signals indeed, could refer to any other pair of cost values.

This transfer is composed of two elements. The first element, namely $rq^{fb}(r)+K(r)$, is a fixed payment equal to the total cost the agent would bear if he were of type r. The second element, namely $\pi(r,s)$, is an uncertain payment whose value depends on the signal realization. Because this realization is unknown to the agent, he faces a lottery with expected value

$$\begin{aligned}\sum_{s=1}^{n}\pi(r,s)\,p(c,s) &= -\left[\overline{\pi}^{fb}(r,\widetilde{s}(c))-\underline{\pi}^{fb}(r,\widetilde{s}(c))\right]\left[p(r,\widetilde{s}(r))-p(c,\widetilde{s}(r))\right] \\ &= -\frac{q^{fb}(r)+K(r)}{dp(r,\widetilde{s}(r))/dr}\left[p(r,\widetilde{s}(r))-p(c,\widetilde{s}(r))\right].\end{aligned}$$

The introduction of this lottery is meant to offset the benefit the agent might obtain with a convenient report as a difference between the fixed payment and his true cost. For this to occur, the lottery should yield sufficiently high expected costs for mimicking types. This requires that the wedge between the reward and the loss designed for type r, as expressed by the ratio $\frac{q^{fb}(r)+K'(r)}{dp(r,\widetilde{s}(r))/dr}$, be large enough. Indeed, this allows P to exploit the correlation between types, as represented by the difference $[p(r,\widetilde{s}(r))-p(c,\widetilde{s}(r))]$, to extract surplus. Recall however that, under the MFL scheme, the wedge $\overline{\pi}^{fb}(r,\widetilde{s}(c))-\underline{\pi}^{fb}(r,\widetilde{s}(c))$ is set at the minimum feasible level for each r and, in particular, it equals zero for $r=\widehat{c}$. Thus, whenever $\widehat{c}$ is reported, the lottery disappears. Under this circumstance, type $c\neq\widehat{c}$ is discouraged from reporting $\widehat{c}$ if and only if

$$\int_{c}^{\widehat{c}}\left[K'(x)-K'(\widehat{c})\right]dx\leq 0,\ \forall c\in[\underline{c},\overline{c}],$$

which explains (10).

The literature has shown that some restriction on the properties of the cost function is required for FB implementation also in the absence of countervailing incentives. From Riordan and Sappington [9], we learn that, when the signal space is smaller than the type space, together with the conditions on the likelihood function of the relevant signal (the counterpart of (5) in our model), FB enforcement calls for restrictions on the shape of the agent's cost function. It is thus not surprising that a lower bound on the concavity of K appears also in our setting. However, the restriction imposed by (10) on the fixed cost function is tighter than the condition identified by Riordan and Sappington [9]. The latter only requires that the agent's cost function be less concave in type than the conditional likelihood function at the relevant signal. As stated in the corollary below, a similar result would entail in our model if the agent were to display a systematic incentive either to overstate or to understate type, whatever the cost realization.

Corollary 1 *If there exists no $\widehat{c}\in[\underline{c},\overline{c}]$ such that $q^{fb}(\widehat{c})+K'(\widehat{c})=0$, then, $\forall c\in[\underline{c},\overline{c}]$, (10) is replaced by*

$$K\text{"}(c)\geq\left[q^{fb}(c)+K'(c)\right]\frac{d^2p(c,\widetilde{s}(c))/dc^2}{dp(c,\widetilde{s}(c))/dc}, \tag{11}$$

with $\widetilde{s}(c) = n$ *when* $q^{fb}(c) + K'(c) > 0$ *and* $\widetilde{s}(c) = 1$ *when* $q^{fb}(c) + K'(c) < 0$.

The corollary above emphasizes that the presence of countervailing incentives exacerbates the requirement on the properties of the cost function.

It should by now be clear that the concavity restriction appears because P adopts the payoff profile that allows her to minimize the deficit for each agent's type. As already illustrated, when this scheme is designed for an agent displaying countervailing incentives, the lottery tends to vanish as r approaches $\widehat{c}$. To circumvent this problem, P should design a different mechanism under which (1) all agent's types do face an effective lottery whatever the report and (2) (LL) is still satisfied. We hereafter describe how a scheme with these characteristics can be constructed.

3.2 The Maximum-sustainable-loss scheme

Suppose (10) holds for all possible types, meaning that limited liability does not (necessarily) compromise FB implementation. As already explained, to induce information release at no agency cost, the expected value of the lottery is to be low enough.

Lemma 3 *The ex post profits that minimize the expected value of the lottery and, at the same time, satisfy (LIC), (PC) and (LL) are such that, for each type* $c \in [\underline{c}, \overline{c}]$, *the agent is rewarded for one sole signal* $s \in N$ *and incurs the highest admissible loss* $(-L)$ *for all the other signals.*

The scheme presented in the lemma is similar to the MFL scheme in that it includes only one reward and equal losses. Yet, losses are here fixed at the largest feasible level so as to minimize the incentive to misreport type for any given c. For this reason, we refer to it as to the Maximum-sustainable-loss (MSL) scheme. Taken together, Lemma 1 and 3 evidence that spreading losses over as many realizations of s as possible is beneficial to P in two different ways. First, when the MFL scheme is adopted, spreading losses and minimizing the reward-loss wedge for the state s in which this wedge is maximum enables P to minimize the deficit that the agent could be required to incur. Second, under the MSL scheme, spreading losses and maximizing the reward-loss wedge for each possible realization of s allows P to minimize the expected value of the lottery that the agent is called to face.

Assumption 3 ensures that the signal s for which the agent is rewarded under the MSL scheme remains the same as under the MFL scheme, namely $\widetilde{s}(c) = n$ if $c < \widehat{c}$ and $\widetilde{s}(c) = 1$ if $c > \widehat{c}$. For any $c \in [\underline{c}, \overline{c}]$, the payoff profile is given by

$$\pi^{fb}(c,s)\Big|_{s=\widetilde{s}(c)} = \frac{1 - p(c, \widetilde{s}(c))}{p(c, \widetilde{s}(c))} L \equiv \overline{\pi}^*(c, \widetilde{s}(c), L) \quad (12)$$

$$\pi^{fb}(c,s)\Big|_{s \neq \widetilde{s}(c)} = -L \equiv \underline{\pi}^*(c, \widetilde{s}(c), L), \quad (13)$$

so that the expected value of the lottery is written

$$\sum_{s=1}^{n} \pi(r,s)\, p(c,s) = -L\left[1 - \frac{p(c,\widetilde{s}(r))}{p(r,\widetilde{s}(r))}\right].$$

This lottery is actually more effective at extracting surplus from the agent, as compared to the one associated with the MFL scheme, because it allows P to take better advantage of the correlation between type and signal. As a result, FB is enforced under milder conditions.

Proposition 2 *Suppose condition (9) holds. Then, under Assumption 1 - 3, the first-best outcome is implemented with ex post payoffs (12) and (13) if,* $\forall c \in [\underline{c}, \overline{c}]$,

$$\begin{aligned} K''(c) \quad \geq \quad & \left[q^{fb}(c) + K'(c) - \frac{dp(c,\widetilde{s}(c))/dc}{p(c,\widetilde{s}(c))} L\right] \\ & + \frac{d^2 p(c,\widetilde{s}(c))/dc^2}{p(c,\widetilde{s}(c))} L, \end{aligned} \tag{14}$$

with $\widetilde{s}(c) = n$ *for* $c < \widehat{c}$ *and* $\widetilde{s}(c) = 1$ *for* $c > \widehat{c}$.

This proposition states that, whenever the lowest loss that is compatible with FB implementation is smaller than the largest deficit the agent can sustain, FB is enforced with payoffs (12) and (13) provided that K'' does not fall below the lower bound imposed by (14). Condition (9) and Assumption 2 ensure that this bound is negative, showing that the requirement on the curvature of K is now relaxed as compared to (10).

A clear message ensues from our analysis. FB is at hand also when K is concave, provided that P is available to abandon the MFL scheme and opt for a mechanism that possibly inflicts a more important (though still feasible) penalty to the agent. Gary-Bobo and Spiegel [4] emphasize that resorting to the MFL scheme, rather than offering a payoff profile that entails a larger deficit, can be especially convenient for a principal. They point out that regulators prefer to avoid financial difficulties for the regulated firms both to avoid activity interruptions and because this can be embarrassing for themselves. Our investigation evidences that, when the MFL scheme is adopted in environments with countervailing incentives, the loss it yields might result excessively low for efficiency to be achieved. This is actually the case when types are very intensely turned between the desire to over-report and that to under-report, *i.e.* when condition (10) is not met.

In the sequel of the analysis, we take (14) to be satisfied. We thus neglect the possibility that FB does not attain because fixed costs are too concave. We rather focus on the more interesting case in which FB implementation is beyond reach because the limits on liability are particularly stringent.

4 The optimal contract with tight limited liability

In this section, we explore the situation in which (LL) is so tight that condition (9) in Proposition 1 fails to hold. Under this circumstance, P cannot find a profile of transfers (and thus of profits) that implement FB. She thus designs a second-best (SB hereafter) contract, which is to be characterized in the sequel of the analysis.

To begin with, notice that, in fact, (9) is not violated for all feasible values of c.

Lemma 4 *Under Assumption 1 - 3, for any* $L \geq 0$, *at the solution to* (Γ), *there exists a unique range of types* $[c_2, c_3] \subseteq [\underline{c}, \overline{c}]$, *such that* $\widehat{c} \in [c_2, c_3]$, *for which the first-best outcome is implemented.*

First of all, limited liability is not an issue as far as type $\widehat{c}$ is concerned. Indeed, for this type, (9) is surely satisfied as $q^{fb}(\widehat{c}) + K'(\widehat{c}) = 0$. Furthermore, (9) holds for the types that lie in a neighborhood of $\widehat{c}$, *i.e.* for all values of c for which the absolute value of $q^{fb}(c) + K'(c)$ is sufficiently low. (LL) is more and more likely to be binding as c diverges from $\widehat{c}$.

Lemma 5 *Under Assumption 1 - 3, there exists (at most) one cost value* $c_1 \in (\underline{c}, c_2)$ *(resp.* $c_4 \in (c_3, \overline{c})$*) such that, at the solution to* (Γ), (PC) *is slack* $\forall c \in [\underline{c}, c_1)$ *(resp.* $\forall c \in (c_4, \overline{c}]$*) and binding* $\forall c \in [c_1, c_2]$, *(resp.* $\forall c \in [c_3, c_4]$*). When no such a cost value exists,* (PC) *is binding* $\forall c \in [\underline{c}, c_2)$ *(resp.* $\forall c \in (c_3, \overline{c}]$*).*

At the solution to (Γ) under tight limited liability, not only P enforces FB for all types in $[c_2, c_3]$. She is also able to extract all surplus from some types below c_2 and some types above c_3[7]. From the proof of Lemma 5 in Appendix, one deduces that the range of types below c_2 (resp. above c_3) from which surplus is fully retained spans to the whole set $[\underline{c}, c_2]$ (resp. $[c_3, \overline{c}]$) if (i) the rate of increase (resp. decrease) of the conditional likelihood that signal n (resp. 1) be drawn as c raises, namely $\frac{dp(c,n)/dc}{p(c,n)}$ (resp. $\frac{dp(c,1)/dc}{p(c,1)}$), is sufficiently large (resp. small) and/or (ii) the fixed cost function is not very concave in type. Intuitively, and in line with the insights from the FB analysis, this means that P is more likely to induce truthtelling at zero rent when the two signals are especially informative about type and/or when the agent's incentive to cheat is not particularly intense[8]. Otherwise, surplus extraction becomes unfeasible for very low and very high types, whose incentives to misreport are strongest. These types are then assigned a positive *interim* payoff.

The possibility that efficient types obtain an information rent depending on the conditional likelihood of signal n raises a similarity with the findings of Gary-Bobo and Spiegel [4]. Indeed, in their model, the agent's participation constraint holds strictly for all types but the least efficient one if the derivative of the conditional likelihood function at n is

[7] Although surplus is retained, FB is not implemented for these types because quantities are distorted away from the efficient level, as will become clear shortly.

[8] The conclusion that convexity of the cost function facilitates P's task will be further confirmed in Proposition 3 below.

small enough[9]. Observe however that, in our framework, the possibility that (PC) be slack for all types but one is ruled out due to the fact that the fixed cost decreases with type. This facilitates surplus extraction, indeed, by weakening the incentives to cheat of types that are sufficiently close to $\widehat{c}$.

Suppose the conditions described above do hold so that the cost values c_1 and c_4 exist. We shall now see how the SB output is characterized in this situation. Consider that the incentive to overstate (resp. understate) type that an agent with $c < \widehat{c}$ (resp. $c > \widehat{c}$) would display if he were to receive the sole fixed payment to produce the FB quantity gets increasingly more intense as c approaches $\underline{c}$ (resp. $\overline{c}$). To remove the incentive to mimic by means of the lottery, while keeping output at the FB level, P would need to progressively increase the wedge between rewards and losses as c moves away from $\widehat{c}$. Nevertheless, (LL) imposes a bound on how large losses can be set, for FB does not attain when $c \notin [c_2, c_3]$. Without quantity distortions, P could solicit information revelation only by raising the reward sufficiently, which would yield an information rent to the agent. This would be too costly though. The optimal strategy is thus to reduce the rent by fixing output away from the efficient level. For types with weak incentives to cheat, namely those in $[c_1, c_2)$ and $(c_3, c_4]$, P distorts output till all surplus is extracted. This further clarifies why, over these cost ranges, participation constraints are saturated, as we said above. For types with more intense incentives to misreport, namely those in $[\underline{c}, c_1]$ and $[c_4, \overline{c}]$, P distorts output to contain the rent, but it would be too costly to remove the rent entirely.

The whole SB output profile and the thresholds of the relevant cost ranges will be characterized in a moment. Before proceeding, it is however useful to make the following standard assumption.

Assumption 4 *The conditional likelihood and cumulative distribution function satisfy the following properties:*

$$\frac{d}{dc}\left(\frac{F(c|n)}{f(c|n)}\right) \geq 0, \ \forall c \in [\underline{c}, \overline{c}] \tag{15}$$

$$\frac{d}{dc}\left(\frac{1-F(c|1)}{f(c|1)}\right) \leq 0, \ \forall c \in [\underline{c}, \overline{c}]. \tag{16}$$

This assumption states the monotonicity with respect to c of the conditional hazard rates $\frac{F(c|n)}{f(c|n)}$ and $\frac{1-F(c|1)}{f(c|1)}$. According to (15), once types between $\underline{c}$ and c have been drawn, it becomes more likely that a type higher than c be drawn, conditional on signal n being observed. According to (16), once types between c and $\overline{c}$ have been drawn, it is less likely that a type higher than c be drawn, conditional on signal 1 being observed.

In the following lemma, roman numbers are appended to denote SB quantities and payoffs over the five relevant cost ranges.

Lemma 6 *Suppose condition (9) does not hold. Then, under Assumption 1 - 4, at the*

[9] See page 5 of the technical appendix to Gary-Bobo and Spiegel [4].

solution to (Γ), *quantities are characterized as follows:*

$$S'\left(q^{I}(c)\right) = c+(1-\alpha)\frac{F(c\,|n)}{f(c\,|n)},\ \forall c\in[\underline{c},c_1] \tag{17}$$

$$q^{II}(c) = \frac{dp(c,n)/dc}{p(c,n)}L-K'(c),\ \forall c\in[c_1,c_2] \tag{18}$$

$$q^{III}(c) = q^{fb}(c),\ \forall c\in[c_2,c_3] \tag{19}$$

$$q^{IV}(c) = \frac{dp(c,1)/dc}{p(c,1)}L-K'(c),\ \forall c\in[c_3,c_4] \tag{20}$$

$$S'\left(q^{V}(c)\right) = c-(1-\alpha)\frac{1-F(c\,|1)}{f(c\,|1)},\ \forall c\in[c_4,\overline{c}]. \tag{21}$$

Moreover, interim profits (rents) are given by

$$E_s\left[\pi^{I}(c,s)\right] = p(c,n)\int_{c}^{c_1}\frac{q^{I}(x)+K'(x)}{p(x,n)}dx-\left[1-\frac{p(c,n)}{p(c_1,n)}\right]L, \quad \forall c\in[\underline{c},c_1] \tag{22}$$

$$E_s\left[\pi^{k}(c,s)\right]=0,\ \forall c\in[c_1,c_2],[c_2,c_3],[c_3,c_4],\quad \forall k\in\{II,III,IV\} \tag{23}$$

$$E_s\left[\pi^{V}(c,s)\right] = -p(c,1)\int_{c_4}^{c}\frac{q^{V}(x)+K'(x)}{p(x,1)}dx-\left[1-\frac{p(c,1)}{p(c_4,1)}\right]L, \quad \forall c\in[c_4,\overline{c}]. \tag{24}$$

To begin with, (19) confirms that output is still efficiently set as long as $c\in[c_2,c_3]$. According to (17) and (21), the same occurs at both the lowest and the highest marginal cost realization. (17) further highlights that output is downward distorted for all types in $(\underline{c},c_1]$, which allows to contain the rent in (22). Moreover, under the first part of Assumption 4, q^I decreases with c all over this set. Observe that the SB quantity solution in Gary-Bobo and Spiegel [4] is characterized precisely as in (17) for all possible agent's types. This occurs because, in their context, the agent displays a systematic incentive to overstate type. (21) further evidences that output is upward distorted for all types in $[c_4,\overline{c})$, which helps limit the rent in (24). Under the second part of Assumption 4, also q^V decreases with type $\forall c\in[c_4,\overline{c})$. Lastly, (18) and (20) define how output is downward and upward distorted in the second and fourth region respectively, just enough to fully extract surplus in an incentive-compatible way.

We now define the thresholds of the relevant cost ranges, which we have only mentioned in the lemmas above but not yet characterized.

Lemma 7 *Suppose condition (9) does not hold. Then, under Assumption 1 - 4, at the*

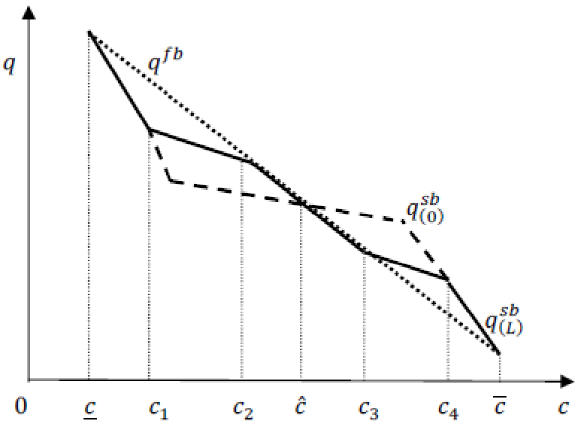

Figure 1: The FB output profile (q^{fb}; dotted line) and the output profile in the SB contract with $L > 0$ (q^{sb}_L; thick line) and $L = 0$ (q^{sb}_0; dashed line).

solution to (Γ), *the cost values* c_1, c_2, c_3 *and* c_4, *are defined as follows:*

$$q^I(c_1) + K'(c_1) = \frac{dp(c_1, n)/dc_1}{p(c_1, n)} L \tag{25}$$

$$q^{II}(c_2) = q^{fb}(c_2) \tag{26}$$

$$q^{IV}(c_3) = q^{fb}(c_3) \tag{27}$$

$$q^V(c_4) + K'(c_4) = \frac{dp(c_4, 1)/dc_4}{p(c_4, 1)} L. \tag{28}$$

Interpreting Lemma 7 together with the results previously presented, it should be clear that c_1 is the cost value at which P retains all surplus from the agent by sufficiently deflating output q^I below the FB level, c_2 is the value at which P retains all surplus by keeping output q^{II} at the FB level and similarly for c_3 and c_4.

A graphical illustration of the full profile of quantities is provided in Figure 1 with regards to both FB implementation and the SB contract defined by (17) to (21). The graph evidences that the set of cost values around $\widehat{c}$ for which FB is still enforced under tight limited liability enlarges as L raises and would collapse onto the singleton $\{\widehat{c}\}$ in the extreme case in which $L = 0$. The graph further shows that the SB quantity decreases with c all over the support, *i.e.* $\frac{dq^k(c)}{dc} \leq 0 \; \forall k \in \{I, II, III, IV, V\}, \forall c \in [\underline{c}, \overline{c}]$, with a rate of decrease that is specific to each cost interval[10]. In particular, it is $\frac{dq^I(c)}{dc} < \frac{dq^{fb}(c)}{dc} < \frac{dq^{II}(c)}{dc}$ and $\frac{dq^V(c)}{dc} < \frac{dq^{fb}(c)}{dc} < \frac{dq^{IV}(c)}{dc}$.

The following proposition lists the conditions under which the SB solution previously characterized is globally incentive compatible.

Proposition 3 *Suppose condition (9) does not hold. Then, under Assumption 1 - 4, the*

[10] That (18) and (20) decrease with c is ensured by condition (31) in Proposition 3 below.

quantity profile (17) - (21) is implemented as the solution to (Γ) *if*

$$\frac{dq^{I}(c)}{dc} \leq -\left[q^{I}(c)+K'(c)\right]\frac{dp(c,n)/dc}{p(c,n)}, \ \forall c\in[\underline{c},c_1] \quad (29)$$

$$\frac{dq^{V}(c)}{dc} \leq -\left[q^{V}(c)+K'(c)\right]\frac{dp(c,1)/dc}{p(c,1)}, \ \forall c\in[c_4,\overline{c}] \quad (30)$$

$$K"(c) \geq \frac{d^2p(c,\widetilde{s}(c))/dc^2}{p(c,\widetilde{s}(c))}L, \ \forall c\in[\underline{c},\overline{c}], \quad (31)$$

with $\widetilde{s}(c)=n$ *for* $c<\widehat{c}$ *and* $\widetilde{s}(c)=1$ *for* $c<\widehat{c}$*.*

We have previously explained that, under Assumption 4, quantities q^I and q^V decrease with type. Proposition 3 further evidences that, for the contract presented in Lemma 6 to be globally incentive compatible, it suffices that those quantities decrease sufficiently fast over the respective cost ranges (see Figure 1 again). According to condition (29) and (30), how fast q^I and q^V should decrease depends on the rate of change of the conditional likelihood that is relevant in the concerned region. To illustrate why this is the case, let us focus on (29), keeping in mind that analogous reasoning applies to (30), *mutatis mutandis.* Take $c\in[\underline{c},c_1)$. As the report r is raised above the true type c, under Assumption 2, the probability of reward increases. Because the loss that the agent might bear equals $-L$ whatever the report, over-reporting yields a higher *interim* profit, as compared to truthtelling, unless the quantity is diminished sufficiently. The incentive to over-report is removed if $q^I(c)$ decreases as fast as (29) dictates. Perfectly analogous to (29) would be the sufficient condition for global incentive compatibility in Gary-Bobo and Spiegel [4] if, in their model, the marginal cost were assumed to be constant in type, as it is in ours, rather than strictly increasing and convex[11].

Condition (31) tells that the contract described in Lemma 4 to 6 is optimal if, for all possible types, the curvature of the fixed cost function does not fall below some given bound that depends on both the conditional likelihood and L. In fact, (31) is the counterpart of (14) in the FB framework previously explored and can be interpreted in a similar fashion, *mutatis mutandis.* Yet, (31) is more stringent as compared to (14). This further reflects the circumstance that, all else equal, it is harder to induce information release when the limits on the agent's liability are tight.

Corollary 2 *Take* $L=0$ *and* $K"(c)=0$, $\forall c\in[\underline{c},\overline{c}]$*. Suppose condition (9) does not hold, whereas (29) and (30) are satisfied. Then, at the solution to* (Γ), $q^{sb}(c)=q^{fb}(\widehat{c})$, $\forall c\in[c_1,c_4]$.

The corollary refers to the specific situation in which the agent can bear no deficit *ex post* and the fixed cost is linear in type. In that case, the range of types for which FB is enforced collapses onto the singleton $\{\widehat{c}\}$. To see this, recall that c_2 and c_3 are defined by $q^{II}(c_2)=q^{fb}(c_2)$ and $q^{II}(c_3)=q^{fb}(c_3)$ respectively. Moreover, with $L=0$,

[11] Compare the inequality at the end of page 5 in the technical appendix of Gary-Bobo and Spiegel [4] with (65) in the proof of (29) in our Appendix.

$q^{II}(c) = -K'(c)$ and $q^{IV}(c) = -K'(c)$. Remembering also the definition of $\widehat{c}$, it is immediate to conclude that $c_2 \equiv \widehat{c} \equiv c_3$ when $L = 0$. Further observe that quantities $q^{II}(c)$ and $q^{IV}(c)$ are constant over types when so is $K'(c)$. Hence, all types within the set $[c_1, c_4]$, from which surplus is entirely extracted, are required to produce the same amount of output, *i.e.* the optimal contract entails pooling at $q^{fb}(\widehat{c})$ in a neighborhood of $\widehat{c}$.

The outcome in Corollary 2 is reminiscent of that Maggi and Rodriguez-Clare [8] find in a setting without correlated information. They characterize the optimal contract in the presence of countervailing incentives for different possible shapes of the agent's reservation utility. They show that, when the reservation utility is linear in type, the contract entails pooling of quantities over some interval of types that earn zero rents[12]. The case of $K" = 0$ in our model is the counterpart for the linear reservation utility in Maggi and Rodriguez-Clare [8]. Corollary 2 evidences that, when the case of $K" = 0$ arises, the optimal contract exhibits analogous features (namely, pooling and no rent in a neighborhood of $\widehat{c}$) in a correlated information framework as soon as the agent cannot be punished *ex post*. This is explained by considering that having $L = 0$ in the presence of correlated information is tantamount to assuming that the agent has to break even *ex post* (rather than at *interim*), whereas *ex post* and *interim* participation are equivalent without correlated information. Observe however that, despite the analogy in terms of structure, the optimal contract with correlated information is not simply twice a replica of the optimal contract without correlated information. Indeed, correlation allows for improvements, as usual. First, the range of types for which bunching arises is less wide. Second, the quantity distortions induced for low and high types are smaller. Third, the expected rents that accrue to those same types are lower than the rents P assigns in the absence of correlated information.

4.1 "Very concave" fixed cost

As previously said, condition (14) in Proposition 2 is taken to be satisfied all along the analysis. Even under this assumption, it is not necessarily the case that condition (31) in Proposition 3 is satisfied in turn. In what follows, we consider the situation in which (31) is violated. The following proposition describes the optimal contract under this circumstance.

Proposition 4 *Suppose condition (9) and (31) do not hold. Then, under Assumption 1 - 4, the quantity solution to* (Γ) *is given by*

$$\begin{aligned} q^{sb}(c) &= q^{I}(c), \ \forall c \in \left[\underline{c}, c^{-}\right) \\ q^{sb}(c) &= q^{fb}(\widehat{c}), \ \ \forall c \in \left[c^{-}, c^{+}\right] \\ q^{sb}(c) &= q^{V}(c), \ \forall c \in \left(c^{+}, \overline{c}\right], \end{aligned}$$

[12] An environment with reservation utility linear in type is analysed also in other works, such as that of Brainard and Martimort [1], with analogous result.

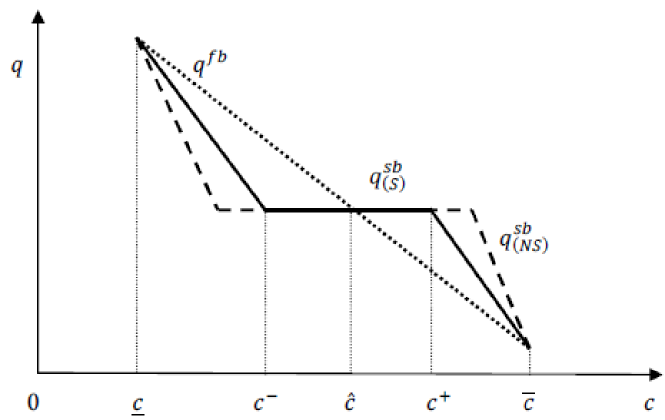

Figure 2: The FB output profile (q^{fb}; dotted line), the output profile in the SB contract with informative signals ($q^{sb}_{(S)}$; thick line) and the output profile in the SB contract with no informative signal ($q^{sb}_{(NS)}$; dashed line) with K "very" concave.

where c^- and c^+ are such that

$$\begin{aligned} q^I\left(c^-\right) &= q^{fb}\left(\widehat{c}\right) \\ q^V\left(c^+\right) &= q^{fb}\left(\widehat{c}\right). \end{aligned}$$

Moreover, (PC) is binding only for type $\widehat{c}$.

The contract described in the proposition entails pooling of quantities in a neighborhood of $\widehat{c}$. This is the sole type from which P is able to retain all surplus when (31) is violated. The contract is reminiscent of that characterized by Lewis and Sappington [6]. They study countervailing incentives in a setting without correlated signals, focusing on the case in which the agent's fixed cost function is concave in type[13]. Yet, in our environment, pooling concerns a smaller range of types, a benefit that follows from the presence of information correlation. This is shown by the graph in Figure 2, which compares the optimal output profile in the two situations.

Having (31) violated means that, as long as $L > 0$, the fixed cost must be sufficiently concave in type for the contract illustrated in Proposition 4 to be SB optimal. By contrast, in the absence of correlated signals, the optimal contract exhibits the structure aforementioned even with a slightly concave fixed cost. This shows that, whenever some loss can be inflicted to the agent *ex post*, the presence of information correlation yields an additional benefit. That is, it also enlarges the class of environments in which full separation arises.

Further observe that the structure of the contract in Proposition 4 is similar to that in Corollary 2, except that, in the former, the range of types with no rent degenerates onto a singleton. This follows from the circumstance that, as already illustrated, incentives to over/under-report are especially strong when K is very concave. In that case, information release is not induced unless a rent is given up even to types around $\widehat{c}$. To interpret this point in a unified way with the rest of our SB results, it is useful to recall that, in

[13] Maggi and Rodriguez-Clare [8] obtain the same outcome in the equivalent situation in which the agent's reservation utility is concave in type.

the absence of correlated signals, the linear-fixed-cost case (or, equivalently, the linear-reservation-utility case) with pooling and no rent for some type range can be seen as a "knife-edge" situation: pooling is removed as soon as K becomes convex; all types but one obtain a rent as soon as K becomes concave (compare Maggi and Rodriguez-Clare [8]). From our analysis, it emerges that the linear case remains a "knife-edge" situation in correlated information frameworks insofar as *ex post* deficits are unfeasible (recall the explanation after Corollary 2). In such frameworks, the relevant "knife-edge" situation becomes condition (31) as soon as the agent can be exposed to (bounded) deficits under *interim* participation.

5 Concluding remarks

We have studied the optimal contract between a principal and an agent who may have countervailing incentives to misreport the type, which is correlated with an *ex post* publicly observable signal. We have focused on the realistic case in which the agent is protected by limited liability. As an example of the situations we have represented, one may consider monopoly regulation in industries (typically, public utilities) in which overhead costs decline with marginal production costs.

Our analysis predicts that, as long as the agent's pocket is sufficiently deep, the first-best outcome is implemented by the incentive scheme that yields the smallest feasible *ex post* loss to the agent (the MFL scheme), if the latter's fixed cost is either linear or convex in type. However, the first-best outcome is unfeasible if the agent's technology does not display this property, unless the principal offers a contract that imposes higher deficit to the agent. We show that, in the presence of countervailing incentives, the contract that yields the highest sustainable loss (the MSL scheme) expands at maximum the range of cost functions that support first best.

Our analysis further predicts that, if the agent's fixed cost is not very concave in type (so that incentives to over and under-report are not too intense), the optimal incentive scheme is a separating contract under which, thanks to the presence of countervailing incentives, the first-best outcome can still be effected for some range of types even when the agent has no especially deep pocket. Otherwise, the optimal contract entails pooling of quantities. However, the concavity threshold between separating and pooling contracts does depend on the agent's liability. As the latter raises (though not to the point that first best can be implemented for any type), increasingly more concave cost functions, *i.e.* a wider class of possible technologies, sustain the separating contract.

Our study offers a clue about the achievements that would be at the principal's hand if she were to face multiple agents with correlated information. In that environment, the signal would be replaced by the type of a second agent. Our results seem to provide further scope for resorting to centralized incentive schemes in correlated information settings in which agents display countervailing incentives and can be exposed to some deficit *ex post*. Our findings further suggest that, when the characteristics of the technology the agent uses make his incentives to lie especially strong, improving contractual efficiency may require to impose on the agent as much uncertainty as feasible, hence to raise his loss. This contrasts with the usual attitude of regulators not to aggravate the regulated firms' financial burden.

References

[1] Brainard, S.L. and D. Martimort (1996), "Strategic Trade Policy Design with Asymmetric Information and Public Contracts", *The Review of Economic Studies*, 63(1), 81-105

[2] Crémer, J. and R. McLean (1988), "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions", *Econometrica*, 56(6), 1247-1258

[3] Demougin, D.M. and D.A. Garvie (1991), "Contractual design with correlated information under limited liability", *The RAND Journal of Economics*, 22(4), 477-489

[4] Gary-Bobo, R. and Y. Spiegel (2006), "Optimal State-Contingent Regulation under Limited Liability", *RAND Journal of Economics*, 37, 431-448

[5] Jullien, B. (2000), "Participation Constraints in Adverse Selection Models", *Journal of Economic Theory*, 93, 1-47

[6] Lewis, T.R. and D.E.M. Sappington (1989), "Countervailing Incentives in Agency Problems", *Journal of Economic Theory*, 49, 294-313

[7] Lewis, T.R. and D.E.M. Sappington (1989), "Inflexible Rules in Incentive Problems", *American Economic Review*, 79(1), 69-84

[8] Maggi, G. and A. Rodriguez-Clare (1995), "On Countervailing Incentives", *Journal of Economic Theory*, 66, 238-263

[9] Riordan, M. and D.E.M. Sappington (1988), "Optimal Contracts with Public Ex Post Information", *Journal of Economic Theory*, 45, 189-199

A First-best implementation

A.1 Local incentive constraint (LIC)

Let $\widetilde{\pi}(r,s) = t(r,s) - cq(r) - K(c)$ the *ex post* profit of the agent when he has type c and reports r while the state of nature is s. His *interim* profit is written

$$E_s[\widetilde{\pi}(r,s)] \equiv \sum_{s=1}^{n} \{t(r,s) - cq(r) - K(c)\} p(c,s). \tag{32}$$

From (32), the first order-condition of the programme of the agent, evaluated at $r = c$, is given by

$$\sum_{s=1}^{n} \left[\frac{dt(c,s)}{dc} - cq'(c)\right] p(c,s) = 0. \tag{33}$$

From $(2a)$ we can compute

$$\frac{dt(c,s)}{dc} = \frac{d\pi(c,s)}{dc} + cq'(c) + q(c) + K'(c). \tag{34}$$

Replacing (34) into (33), we have

$$\sum_{s=1}^{n}\left\{\frac{d\pi\left(c,s\right)}{dc}+\left[q\left(c\right)+K'\left(c\right)\right]\right\}p\left(c,s\right)=0. \tag{35}$$

At FB $\pi\left(c,s\right)=\pi^{fb}\left(c,s\right)$ and (PC) is binding for all c, so that

$$\sum_{s=1}^{n}\pi^{fb}\left(c,s\right)p\left(c,s\right)=0 \tag{36}$$

and

$$\sum_{s=1}^{n}\frac{d\pi^{fb}\left(c,s\right)}{dc}p\left(c,s\right)=-\sum_{s=1}^{n}\pi^{fb}\left(c,s\right)\frac{dp\left(c,s\right)}{dc}.$$

Using this in (35) together with $q\left(c\right)=q^{fb}\left(c\right)$, (LIC) is obtained.

A.2 Proof of Lemma 2

Suppose $\widetilde{s}\left(c\right)=\overline{s}$, $\forall c\in\left[\underline{c},\overline{c}\right]$, with $\overline{s}$ some constant in N. Take also $\frac{dp(c,\overline{s})}{dc}>0$. For $c<\widehat{c}$, the punishment is as from (4), *i.e.* $\underline{\pi}^{fb}\left(c,\overline{s}\right)<0$. Similarly, for $c>\widehat{c}$, the punishment is as from (3), *i.e.* $\overline{\pi}^{fb}\left(c,\overline{s}\right)<0$. Furthermore, because these profits belong to the MFL scheme, $\overline{s}$ must maximize both $\underline{\pi}^{fb}\left(c,\overline{s}\right)$ for $c<\widehat{c}$ and $\overline{\pi}^{fb}\left(c,\overline{s}\right)$ for $c>\widehat{c}$ at once. The former requires that $\frac{d}{dc}\left(\frac{p(c,s)}{p(c,\overline{s})}\right)<0$, $\forall s\neq\overline{s}$, the latter that $\frac{d}{dc}\left(\frac{1-p(c,s)}{p(c,\overline{s})}\right)>0$, $\forall s\neq\overline{s}$. Suppose $\frac{d}{dc}\left(\frac{p(c,s)}{p(c,\overline{s})}\right)<0$. Together with $\frac{dp(c,\overline{s})}{dc}>0$, this involves that $\frac{d}{dc}\left(\frac{1-p(c,s)}{p(c,\overline{s})}\right)\leq 0$, contradicting the hypothesis that the MFL scheme is obtained with $\widetilde{s}\left(c\right)=\overline{s}$. The proof proceeds similarly for $\frac{dp(c,\overline{s})}{dc}<0$.

A.3 Proof of Proposition 1

The profits (3) and (4) that solve (Γ') also solve (Γ) if and only if they satisfy (GIC) and (LL). In what follows, we find a sufficient condition for them to satisfy (GIC) and then we prove Proposition 1.

A.3.1 Global incentive compatibility

The *interim* profit $E_s\left[\widetilde{\pi}\left(r,s\right)\right]$ given by (32) is rewritten

$$\begin{aligned}E_s\left[\widetilde{\pi}\left(r,s\right)\right] &\equiv q^{fb}\left(r\right)\left(r-c\right)+K\left(r\right)-K\left(c\right)+\sum_{s=1}^{n}\pi\left(r,s\right)p\left(c,s\right) \\ &= q^{fb}\left(r\right)\left(r-c\right)+K\left(r\right)-K\left(c\right)+\sum_{s=1}^{n}\underline{\pi}^{fb}\left(c,\widetilde{s}\left(r\right)\right)p\left(c,s\right) \\ &\quad+\left[\overline{\pi}^{fb}\left(r,\widetilde{s}\left(c\right)\right)-\underline{\pi}^{fb}\left(r,\widetilde{s}\left(c\right)\right)\right]p\left(c,\widetilde{s}\left(r\right)\right)\end{aligned} \tag{37}$$

Substituting (3) and (4), we rewrite (37) as

$$\begin{aligned}E_s\left[\widetilde{\pi}\left(r,s\right)\right] &= \int_c^r\left\{\left[q^{fb}\left(r\right)+K'\left(r\right)\right]\left[1-\frac{dp\left(x,\widetilde{s}\left(r\right)\right)/dx}{dp\left(r,\widetilde{s}\left(r\right)\right)/dr}\right]\right. \\ &\quad\left.+K'(x)-K'(r)\right\}dx.\end{aligned} \tag{38}$$

From (38), from the condition for global incentive compatibility $E_s[\widetilde{\pi}(r,s)] \leq 0$ and taking into account that $\widetilde{s}(r) = n$ if $r < \widehat{c}$ and $\widetilde{s}(r) = 1$ if $r > \widehat{c}$, we deduce the following conditions:

$$K'(r) - K'(c) \geq \left[q^{fb}(r) + K'(r)\right]\left[1 - \frac{dp(c,\widetilde{s}(r))/dc}{dp(r,\widetilde{s}(r))/dr}\right] \text{ if } r \geq c \tag{39}$$

$$K'(r) - K'(c) \leq \left[q^{fb}(r) + K'(r)\right]\left[1 - \frac{dp(c,\widetilde{s}(r))/dc}{dp(r,\widetilde{s}(r))/dr}\right] \text{ if } r \leq c. \tag{40}$$

These conditions are satisfied if $K"(c) \geq 0$.

A.3.2 Proposition 1

From (3) and (4), from $\widetilde{s}(r) = n$ if $r \leq \widehat{c}$ and $\widetilde{s}(r) = 1$ if $r > \widehat{c}$ and from Assumption 1 to 3, we deduce that $\underline{\pi}^{fb}(c,\widetilde{s}(r)) < 0 < \overline{\pi}^{fb}(c,\widetilde{s}(c))\ \forall r \in [\underline{c},\overline{c}]$. Moreover, replacing (3) and (4) into (LL), we find that (LL) is satisfied by condition (9) for any feasible report. As (3) and (4) satisfy both (LIC) and (GIC) whenever $K"(c) \geq 0$ (see the proof above), FB is implemented as the solution to (Γ).

A.4 Proof of Corollary 1

It follows immediately from (39) and (40), with $q^{fb}(r) + K'(r) \neq 0\ \forall r \in [\underline{c},\overline{c}]$.

A.5 Proof of Lemma 3

We hereafter show that the expected value of the lottery, *i.e.* $\sum_{s=1}^{n}\pi(r,s)\,p(c,s)$, is minimized (with (LIC), (PC) and (LL) all satisfied) when P assigns one reward and losses that are all equal to $-L$. We proceed as follows. We first calculate the expected value of the lottery with one reward and losses all equal to $-L$. We then calculate the expected value of the lottery with three distinct profits, the smallest of which equal to $-L$. We finally compare the expected value of the lottery in the two cases and show that it is higher in the latter case.

As a first step, assume that, when the agent has type c and reports r, he receives $\overline{\pi}(r,\widetilde{s}(r)) > 0$ if the state is some $s = \widetilde{s}(r)$ and $\underline{\pi}(r,\widetilde{s}(r)) = -L$ in any state $s \neq \widetilde{s}(r)$. P seeks to minimize

$$\sum_{s=1}^{n}\pi(r,s)\,p(c,s) = \underline{\pi}(r,\widetilde{s}(r)) + \left[\overline{\pi}(r,\widetilde{s}(r)) - \underline{\pi}(r,\widetilde{s}(r))\right]p(c,\widetilde{s}(r)). \tag{41}$$

With (PC) binding for type r, we have

$$\overline{\pi}(r,\widetilde{s}(r)) = -\frac{1-p(r,\widetilde{s}(r))}{p(r,\widetilde{s}(r))}\underline{\pi}(r,\widetilde{s}(r)).$$

Replacing this expression into (41) together with $\underline{\pi}(r,\widetilde{s}(r)) = -L$, we get

$$\sum_{s=1}^{n}\pi(r,s)\,p(c,s) = -L\left[1 - \frac{p(c,\widetilde{s}(r))}{p(r,\widetilde{s}(r))}\right]. \tag{42}$$

Assume next that P implements FB with three distinct profit levels, namely $\underline{\pi}(c,\widetilde{s}(c))$, $\widehat{\pi}(c,\widehat{s}(c))$ and $\overline{\pi}(c,\widetilde{s}(c))$, such that $\underline{\pi}(c,\widetilde{s}(c)) = -L$, $\underline{\pi}(c,\widetilde{s}(c)) < \widehat{\pi}(c,\widehat{s}(c)) < \overline{\pi}(c,\widetilde{s}(c))$

and $\widehat{s}(c) \in N \setminus \{\widetilde{s}(c)\}$. The expected value of the lottery becomes

$$\begin{aligned}\sum_{s=1}^{n} \pi(r,s)\, p(c,s) &= [\overline{\pi}(r,\widetilde{s}(r)) - \underline{\pi}(r,\widetilde{s}(r))]\, p(c,\widetilde{s}(r)) \\ &\quad + \underline{\pi}(r,\widetilde{s}(r)) + [\widehat{\pi}(c,\widehat{s}(c)) - \underline{\pi}(r,\widetilde{s}(r))]\, p(c,\widehat{s}(r)),\end{aligned} \tag{43}$$

whereas the binding (PC) is now written

$$\overline{\pi}(r,\widetilde{s}(r)) - \underline{\pi}(r,\widetilde{s}(r)) = -\frac{\underline{\pi}(r,\widetilde{s}(r)) + [\widehat{\pi}(r,\widehat{s}(r)) - \underline{\pi}(r,\widetilde{s}(r))]\, p(r,\widehat{s}(r))}{p(r,\widetilde{s}(r))}.$$

Replacing this expression into (43), together with $\underline{\pi}(r,\widetilde{s}(r)) = -L$, we obtain

$$\begin{aligned}\sum_{s=1}^{n} \pi(r,s)\, p(c,s) &= -L\left[1 - \frac{p(c,\widetilde{s}(r))}{p(r,\widetilde{s}(r))}\right] \\ &\quad + [\widehat{\pi}(r,\widehat{s}(r)) + L]\left[1 - \frac{p(r,\widehat{s}(r))\, p(c,\widetilde{s}(r))}{p(c,\widehat{s}(r))\, p(r,\widetilde{s}(r))}\right] p(c,\widehat{s}(r)).\end{aligned} \tag{44}$$

Calculating the difference between (42) and (44), we further obtain

$$[\widehat{\pi}(r,\widehat{s}(r)) + L]\left[-\frac{p(r,\widehat{s}(r))\, p(c,\widetilde{s}(r))}{p(c,\widehat{s}(r))\, p(r,\widetilde{s}(r))} + 1\right] p(c,\widehat{s}(r)).$$

From Proposition 2, $\widetilde{s}(r) = n$ if $r < \widehat{c}$ and $\widetilde{s}(r) = 1$ if $r > \widehat{c}$. Under Assumption 3 and because $\widehat{\pi}(r,\widehat{s}(r)) + L > 0$, the above difference is positive. Hence, the expected value of the lottery is higher with any triplet $\{\underline{\pi}(c,\widetilde{s}(c)), \widehat{\pi}(c,\widehat{s}(c)), \overline{\pi}(c,\widetilde{s}(c))\}$, such that $\underline{\pi}(c,\widetilde{s}(c)) = -L$ and $\underline{\pi}(c,\widetilde{s}(c)) < \widehat{\pi}(c,\widehat{s}(c)) < \overline{\pi}(c,\widetilde{s}(c))$, than it is with the pair of profits $\{\underline{\pi}(c,\widetilde{s}(c)), \overline{\pi}(c,\widetilde{s}(c))\}$.

A.6 Proof of Proposition 2

As $\underline{\pi}^*(c,\widetilde{s}(c),L) = -L$ and (PC) is binding, the reward is given by (12). Then, the payoff $E_s[\widetilde{\pi}(r,s)]$ described by (38) is rewritten

$$\begin{aligned}E_s[\widetilde{\pi}(r,s)] &= q^{fb}(r)(r-c) + K(r) - K(c) + \sum_{s=1}^{n} \underline{\pi}^*(r,\widetilde{s}(r),L)\, p(c,s) \\ &\quad + [\overline{\pi}^*(r,\widetilde{s}(r),L) - \underline{\pi}^*(r,\widetilde{s}(r),L)]\, p(c,\widetilde{s}(r)) \\ &= \int_c^r \left[q^{fb}(r) + K'(r) - L\frac{dp(x,\widetilde{s}(r))/dx}{p(r,\widetilde{s}(r))} + K'(x) - K'(r)\right] dx.\end{aligned}$$

Under (14), $E_s[\widetilde{\pi}(r,s)] \leq 0$ for all $x < r$. To see which value the state $\widetilde{s}(r)$ takes, suppose $c < r$. Under Assumption 3, the ratio $\frac{p(c,\widetilde{s}(r))}{p(r,\widetilde{s}(r))}$ in the expression here above is minimized. Hence, (14) is least stringent if $s(r) = n$ when $r < \widehat{c}$ and $s(r) = 1$ when $r > \widehat{c}$. The same reasoning applies for $r > c$.

B The optimal contract with tight limited liability

B.1 Proof of Lemma 4

From the definition of $\widehat{c}$ in Lemma 2, (9) holds for $c = \widehat{c}$, $\forall L \geq 0$. Take now $c < \widehat{c}$ and suppose that (9) is violated for c, in which case

$$q^{fb}(c) + K'(c) > L \frac{dp(c,n)/dc}{p(c,n)}. \tag{45}$$

(i) Suppose that

$$\frac{dq^{fb}(c)}{dc} + K"(c) < \frac{L}{p(c,n)} \left[\frac{d^2 p(c,n)}{dc^2} - \frac{(dp(c,n)/dc)^2}{p(c,n)} \right]. \tag{46}$$

and recall that $\frac{dq^{fb}(c)}{dc} + K"(c) < 0$ (Assumption 1). It follows that, as c raises, the left hand side (LHS hereafter) of (45) decreases faster than the right hand side (RHS hereafter). Because (45) does not hold for $c = \widehat{c}$, there is at most one value $c_2 \in [\underline{c}, \widehat{c}]$ such that (45) does not hold for any $c \in [\underline{c}, c_2)$ and holds for all $c \in [c_2, \widehat{c}]$. This value exists if (45) holds for $c = \underline{c}$.

(ii) Next suppose that (46) is not satisfied, so that, as c raises, the LHS of (45) decreases less fast than the RHS. Hence, if (45) does not hold for $c = \underline{c}$, then it does not hold for any $c \in [\underline{c}, \widehat{c}]$, in which case there is no c in this interval for which (9) is violated. If (45) holds for $c = \underline{c}$, then it must hold for any $c \in [\underline{c}, \widehat{c}]$, involving that (9) is violated for all types within this interval. This contradicts the definition of $\widehat{c}$, under which (9) is satisfied for $c = \widehat{c}$. Therefore, (45) does not hold for $c = \underline{c}$, so that (9) is satisfied for all $c \in [\underline{c}, \widehat{c}]$.

Considering (i) and (ii) altogether, we deduce that there exists at most one subset $[\underline{c}, c_2) \subseteq [\underline{c}, \widehat{c}]$ over which (9) is violated, with $c_2 \in [\underline{c}, \widehat{c}]$. This value exists if and only if (9) is violated for $c = \underline{c}$.

A similar reasoning applies when $c > \widehat{c}$, meaning that there exists at most one subset $(c_3, \overline{c}] \subseteq [\widehat{c}, \overline{c}]$, with $c_3 \in [\widehat{c}, \overline{c}]$, for which (9) is violated.

B.2 Proof of Lemma 5

Take $c \in [\underline{c}, c_2)$. By Lemma 4, (LL) is binding. By Lemma 3, the optimal contract is such that there is only one reward and equal losses fixed at $-L$. Furthermore, the proof of Lemma 6 shows that, for any $c \in [\underline{c}, \widehat{c}]$ to which a rent accrues with this profile of profits, the SB quantity is given by $q^I(c)$ as defined by (17) in the main text, while in case no rent accrues the SB quantity is given by $q^{II}(c)$ as defined by (18). A rent is left to type $c \in [\underline{c}, \widehat{c}]$ if and only if

$$q^I(c) + K'(c) > L \frac{dp(c,n)/dc}{p(c,n)}. \tag{47}$$

Indeed, for the types for which (47) is violated, P is better off by choosing the quantity $q^{II}(c) \geq q^I(c)$ such that all surplus is extracted.

From Assumption 1 and because $\frac{dq^I(c)}{dc} < \frac{dq^{fb}(c)}{dc}$, it is $K"(c) < -\frac{dq^I(c)}{dc}$. Using this condition, we proceed identically as in the proof of Lemma 4 but replacing $q^{fb}(c)$ with $q^I(c)$, c_2 with c_1, $\widehat{c}$ with c_2 and (45) with (47). We find that the curves $q^I(c) + K'(c)$ and $L\frac{dp(c,n)/dc}{p(c,n)}$ cross once at most, at some $c_1 \in [\underline{c}, c_2)$. This value c_1 exists if and only if (47) is satisfied for $c = \underline{c}$ (and a rent is given up at least to type $\underline{c}$).

The procedure is similar for $c \in [\widehat{c}, \overline{c}]$.

B.3 Proof of Lemma 6

B.3.1 Expected utility

Define

$$\widetilde{W}(a,b) \equiv \int_a^b \sum_{s=1}^n \left[V(q(c)) + \alpha\pi(c,s)\right] p(c,s) f(c)\, dc, \tag{48}$$

so that the objective function in (Γ) is rewritten

$$\widetilde{W} = \left[\widetilde{W}(\underline{c},c_1) + \widetilde{W}(c_1,c_2) + \widetilde{W}(c_2,c_3) + \widetilde{W}(c_3,c_4) + \widetilde{W}(c_4,\overline{c})\right].$$

As the maximization of the expected utility in each cost interval is independent of that in any other interval, we treat the various intervals separately. We have already established that, in the situation under scrutiny, FB attains $\forall c \in [c_2, c_3]$ (Lemma 4) and we shall not come back to this case.

B.3.2 The solution for $c \in [\underline{c}, c_1)$

This proof is close to that of Gary-Bobo and Spiegel [4]. We first calculate the *ex post* transfer, then the expected transfer for $c \in [\underline{c}, c_1)$, namely $E(t_1)$. We finally replace it into the expression of $\widetilde{W}(\underline{c}, c_1)$ and optimize with respect to quantity.

The *ex post* transfer when $c \in [\underline{c}, c_1)$ It is useful to define $t(c,s) \equiv g(c)$ the transfer the agent receives when $s = \widetilde{s}(c)$ and $t(c,s) \equiv h(c,s)$ the transfer he receives when $s \neq \widetilde{s}(c)$. For sake of simplicity, $h(c,s)$ is defined for any $s \in N$, although in reality $h(c,\widetilde{s}(c))$ does not exist (as the agent is not punished in state $\widetilde{s}(c)$). Replacing into (33) and rearranging, we get

$$g'(c) = \sum_{s=1}^n cq'(c)\frac{p(c,s)}{p(c,\widetilde{s}(c))} - \sum_{s=1}^n \frac{dh(c,s)}{dc}\frac{p(c,s)}{p(c,\widetilde{s}(c))} + \frac{dh(c,\widetilde{s}(c))}{dc}$$

Define $c_k \in \{c_1, c_4\}$ any type c for which $E_s[\pi(c_k,s)] = 0$. Integrating all terms above from c to c_k we obtain

$$\begin{aligned} g(c) &= g(c_k) - \int_c^{c_k} \sum_{s=1}^n xq'(x)\frac{p(x,s)}{p(x,\widetilde{s}(c))}dx \\ &\quad + \int_c^{c_k} \sum_{s=1}^n \frac{dh(x,s)}{dx}\frac{p(x,s)}{p(x,\widetilde{s}(c))}dx - h(c_k,\widetilde{s}(c)) + h(c,\widetilde{s}(c)). \end{aligned} \tag{49}$$

Integrating by parts the second and the third term in the RHS of (49), we rewrite it as

$$\begin{aligned} g(c) &= g(c_k) - h(c_k,\widetilde{s}(c)) + h(c,\widetilde{s}(c)) \\ &\quad - \sum_{s=1}^n c_k q(c_k)\frac{p(c_k,s)}{p(c_k,\widetilde{s}(c))} + \sum_{s=1}^n h(c_k,s)\frac{p(c_k,s)}{p(c_k,\widetilde{s}(c))} \\ &\quad + \sum_{s=1}^n cq(c)\frac{p(c,s)}{p(c,\widetilde{s}(c))} - \sum_{s=1}^n h(c,s)\frac{p(c,s)}{p(c,\widetilde{s}(c))} \\ &\quad + \int_c^{c_k} \sum_{s=1}^n q(x)\frac{d}{dx}\left[x\frac{p(c,s)}{p(x,\widetilde{s}(c))}\right]dx - \int_c^{c_k} \sum_{s=1}^n h(x,s)\frac{d}{dx}\left[\frac{p(x,s)}{p(x,\widetilde{s}(c))}\right]dx. \end{aligned} \tag{50}$$

Denote

$$\psi_s(c) \equiv \frac{p(c,s)}{p(c,\widetilde{s}(c))}. \tag{51}$$

Using it in (50) we obtain

$$\begin{aligned} g(c) &= \sum_{s=1}^{n} [h(c_k,s) - c_k q(c_k)]\,\psi_s(c_k) + g(c_k) \\ &\quad -h(c_k,\widetilde{s}(c)) + \sum_{s=1}^{n} q(c)\,c\psi_s(c) + \int_c^{c_k} \sum_{s=1}^{n} q(x)\,\frac{d}{dx}[x\psi_s(x)]\,dx \\ &\quad -\sum_{s=1}^{n} h(c,s)\,\psi_s(c) - \int_c^{c_k} \sum_{s=1}^{n} h(x,s)\,\psi_s'(x)\,dx + h(c,\widetilde{s}(c)). \end{aligned} \tag{52}$$

Using $(2a)$ we can group the expression

$$\begin{aligned} &\sum_{s=1}^{n} [h(c_k,s) - c_k q(c_k)]\,\psi_s(c_k) + r(c_k) - h(c_k,\widetilde{s}(c_k)) \\ =& \sum_{s=1}^{n} [\pi(c_k,s) + K(c_k)]\,\psi_s(c_k). \end{aligned}$$

Replacing into (52) returns

$$\begin{aligned} g(c) &= \sum_{s=1}^{n} [\pi(c_k,s) + K(c_k)]\,\psi_s(c_k) - \sum_{s=1}^{n} h(c,s)\,\psi_s(c) + \sum_{s=1}^{n} cq(c)\,\psi_s(c) \\ &\quad +h(c,\widetilde{s}(c)) + \int_c^{c_k} \sum_{s=1}^{n} q(x)\,\frac{d}{dx}[x\psi_s(x)]\,dx - \int_c^{c_k} \sum_{s=1}^{n} h(x,s)\,\psi_s'(x)\,dx. \end{aligned} \tag{53}$$

Using $(2a)$ as well as $t(c,s) = h(c,s)$ for $s \neq \widetilde{s}(c)$ and letting $\underline{\pi}^{sb}(c,s)$ the loss, we have $\underline{\pi}^{sb}(c,s) = h(c,s) - [cq(c) + K(c)]$. We use this to rewrite the expression

$$\begin{aligned} &\int_c^{c_k} \sum_{s=1}^{n} q(x)\,\frac{d}{dx}[x\psi_s(x)]\,dx - \int_c^{c_k} \sum_{s=1}^{n} h(x,s)\,\psi_s'(x)\,dx \\ =& \int_c^{c_k} \sum_{s=1}^{n} \left[q(x)\,\psi_s(x) + [xq(x) - h(x,s)]\,\psi_s'(x)\right] dx \\ =& \int_c^{c_k} \sum_{s=1}^{n} \left[q(x)\,\psi_s(x) - \left[\underline{\pi}^{sb}(x,s) + K(x)\right] \psi_s'(x)\right] dx. \end{aligned}$$

Replacing this into (53) yields the *ex post* transfer

$$
\begin{aligned}
g(c) &= \sum_{s=1}^{n} \left[\pi(c_k, s) + K(c_k)\right] \psi_s(c_k) + \sum_{s=1}^{n} cq(c)\, \psi_s(c) \\
&\quad - \left[\sum_{s=1}^{n} h(c,s)\, \psi_s(c) - h(c, \widetilde{s}(c))\right] \\
&\quad + \int_{c}^{c_k} \sum_{s=1}^{n} q(x)\, \psi_s(x) - \left[\underline{\pi}^{sb}(x,s) + K(x)\right] \psi_s'(x)\, dx.
\end{aligned} \tag{54}
$$

The expected transfer for $c \in [\underline{c}, c_1)$ Using the notation $h(c,s)$ and $g(c)$ as defined above, the expected transfer $E(t_1)$ when $c < c_1$ is given by

$$
E(t_1) = \int_{\underline{c}}^{c_1} \left[\sum_{s=1}^{n-1} h(c,s)\, p(c,s) + g(c)\, p(c,n)\right] f(c)\, dc
$$

Substitute $\widetilde{s}(c) = n$ and $c_k = c_1$ into (54) and then substitute $g(c)$ from (54) into the above expression. This yields

$$
\begin{aligned}
E(t_1) = \int_{\underline{c}}^{c_1} \Bigg\{ & \sum_{s=1}^{n} \left(\left[\pi(c_1, s) + K(c_1)\right] \psi_s(c_1) + cq(c)\, \psi_s(c)\right) \\
+ & \sum_{s=1}^{n} \left(q(c)\, \psi_s(c) - \left[\underline{\pi}^{sb}(c,s) + K(c)\right] \psi_s'(c)\right) \Bigg\} p(c,n)\, f(c)\, dc
\end{aligned}
$$

Define

$$
\phi(c) \equiv \int_{\underline{c}}^{c} p(x,n)\, f(x)\, dx, \forall c \in [\underline{c}, c_1) \tag{55}
$$

for any $c \in [\underline{c}, c_1)$. We calculate

$$
\begin{aligned}
& \int_{\underline{c}}^{c_1} \left\{ \int_{c}^{c_1} \left[\sum_{s=1}^{n} \left(q(x)\, \psi_s(x) - \left[\underline{\pi}^{sb}(x,s) + K(x)\right] \psi_s'(x)\right)\right] dx \right\} p(c,n)\, f(c)\, dc \\
= & \int_{\underline{c}}^{c_1} \left\{ \sum_{s=1}^{n} \left(q(c)\, \psi_s(c) - \left[\underline{\pi}^{sb}(c,s) + K(c)\right] \psi_s'(c)\right) \right\} \phi(c)\, dc
\end{aligned}
$$

We thus find

$$
\begin{aligned}
E(t_1) &= \int_{\underline{c}}^{c_1} \left\{ \sum_{s=1}^{n} \left(\left[\pi(c_1, s) + K(c_1)\right] \psi_s(c_1) + cq(c)\, \psi_s(c)\right) \right\} p(c,n)\, f(c)\, dc \\
&\quad + \int_{\underline{c}}^{c_1} \left\{ \sum_{s=1}^{n} \left(q(c)\, \psi_s(c) - \left[\underline{\pi}^{sb}(c,s) + K(c)\right] \psi_s'(c)\right) \right\} \phi(c)\, dc.
\end{aligned} \tag{56}
$$

The optimal output for $c \in [\underline{c}, c_1)$ Substituting (56) into (2a) and then (2a) into the expression of $\widetilde{W}(\underline{c}, c_1)$ from (48), we rewrite it as follows:

$$\begin{aligned}
\widetilde{W}(\underline{c}, c_1) &= \int_{\underline{c}}^{c_1} \sum_{s=1}^{n} \left[S(q(c)) - \alpha c q(c) - \alpha K(c)\right] p(c,s) f(c)\, dc \qquad (57) \\
&\quad - (1-\alpha) \int_{\underline{c}}^{c_1} \left\{ \sum_{s=1}^{n} \left(\left[\pi(c_1, s) + K(c_1)\right] \psi_s(c_1) \right.\right. \\
&\quad \left.\left. + c q(c) \psi_s(c) \right) \right\} p(c,n) f(c)\, dc \qquad (58) \\
&\quad - (1-\alpha) \int_{\underline{c}}^{c_1} \left\{ \sum_{s=1}^{n} \left(q(c) \psi_s(c) - \left[\underline{\pi}^{sb}(c,s) + K(c)\right] \psi_s'(c) \right) \right\} \phi(c)\, dc
\end{aligned}$$

From the definition of c_1 (see Lemma 5), $E_s[\pi(c_1, s)] = 0$. Also, because $\widetilde{W}(\underline{c}, c_1)$ decreases with $\underline{\pi}^{sb}(c,s)$, it is optimal to set the latter at the lowest feasible value, *i.e.* $\underline{\pi}^{sb}(c,s) = -L$. Replacing into $\widetilde{W}(\underline{c}, c_1)$, the first-order condition with respect to q, $\forall c \in [\underline{c}, c_1)$, is given by

$$\begin{aligned}
& \left[S'(q(c)) - \alpha c\right] p(c,s) f(c) \\
=\; & (1-\alpha) \left[c \psi_s(c) p(c,n) f(c) + \psi_s(c) \phi(c)\right].
\end{aligned}$$

Denoting $q^I(c)$ the quantity that satisfies the condition above together with (51) and (55), we can rewrite

$$\begin{aligned}
S'\left(q^I(c)\right) &= \alpha c + (1-\alpha) \frac{\psi_s(c)}{p(c,s) f(c)} \left[c p(c,n) f(c) + \phi(c)\right] \\
&= c + (1-\alpha) \frac{\int_{\underline{c}}^{c} p(x,s) f(x)\, dx}{p(c,n) f(c)} \\
&= c + (1-\alpha) \frac{F(c|n)}{f(c|n)},
\end{aligned}$$

with $F(c|n) = \frac{\int_{\underline{c}}^{c} p(x,n) f(x) dx}{\int_{\underline{c}}^{\overline{c}} p(x,n) f(x) dx}$ and $f(c|n) = \frac{p(c,n) f(c)}{\int_{\underline{c}}^{\overline{c}} p(x,n) f(x) dx}$.

B.3.3 The solution for $c \in [c_1, c_2)$

From Lemma 5 one has $E_s[\pi(c,s)] = 0$ whenever $c \in [c_1, c_2)$. It means that the functional form of the *ex post* profit $\pi(c,s)$ is similar to that in (3) and (4), except that $q^{fb}(c)$ is replaced by $q^{II}(c)$, the value of which we need to determine. In particular, the punishment is given by

$$\underline{\pi}^{sb}(c,s) = -\left[q^{II}(c) + K'(c)\right] \frac{p(c,n)}{dp(c,n)/dc}.$$

Moreover, by Lemma 4 (see also proof of Lemma 6), $\underline{\pi}^{sb}(c,s) = -L, \forall c \in [\underline{c}, c_2)$. Using this in the expression above, $q^{II}(c)$ is found to be as defined in (18).

The proof is identical for $c \in (c_3, c_4]$.

B.3.4 The solution for $c \in (c_4, \overline{c}]$

Define

$$\varphi(c) \equiv \int_c^{\overline{c}} p(x,1) f(x)\, dx, \forall c \in (c_4, \overline{c}].$$

Proceeding as for $c \in [\underline{c}, c_1)$, one finds the expected transfer $E(t_2)$ that corresponds to $c \in (c_4, \overline{c}]$ as follows

$$\begin{aligned} E(t_2) &= \int_{c_4}^{\overline{c}} \left\{ \sum_{s=1}^{n} \left([\pi(c_4, s) + K(c_4)] \psi_s(c_4) + cq(c) \psi_s(c) \right) \right\} p(c,1) f(c)\, dc \\ &+ \int_{c_4}^{\overline{c}} \left\{ \sum_{s=1}^{n} \left(\left[\underline{\pi}^{sb}(c,s) + K(c) \right] \psi_s'(c) - q(c) \psi_s(c) \right) \right\} \varphi(c)\, dc \end{aligned} \tag{59}$$

Substituting (59) into $\widetilde{W}(c_4, \overline{c})$, we can characterize the optimal output $q^V(c)$ as

$$S'\left(q^V(c)\right) = c - (1-\alpha) \frac{1 - F(c\,|1)}{f(c\,|1)},$$

with $[1 - F(c\,|1)] = \frac{\int_c^{\overline{c}} p(x,1) f(x) dx}{\int_{\underline{c}}^{\overline{c}} p(x,1) f(x) dx}$ and $f(c\,|1) = \frac{p(c,1) f(c)}{\int_{\underline{c}}^{\overline{c}} p(x,1) f(x) dx}$.

B.4 Proof of Proposition 3

The *interim* profit of the agent when he reports r is written similarly to (37) as

$$E_s[\widetilde{\pi}(r,s)] = q^{sb}(r)(r-c) + K(r) - K(c) + \sum_{s=1}^{n} \pi(r,s) p(c,s), \tag{60}$$

with $q^{sb} \in \left\{q^I, q^{II}, q^{III}, q^{IV}, q^V\right\}$ (see Lemma 6). The *ex post* profit $\pi(r,s)$ that appears in (60) is calculated in a different way according to the value the report r takes. We thus develop the analysis case by case.

B.4.1 Case $r \in [\underline{c}, c_1]$

We proceed as follows. We first calculate the *ex post* profit $\pi(r,s)$, $\forall r \in [\underline{c}, c_1]$. We replace into (60) so as to calculate $E_s[\widetilde{\pi}(r,s)]$. We finally state the global incentive condition $E_s[\widetilde{\pi}(r,s)] \leq E_s[\pi(c,s)]$ for any report $r \in [\underline{c}, c_1]$. Two sub-cases are considered, namely $c \in [\underline{c}, c_1]$ and $c \in [c_1, \overline{c}]$.

The *ex post* profit $\pi(r,s)$ Recall that $c_k \in \{c_1, c_4\}$ is by definition a type c for which $E_s(\pi(c_k, s)) = 0$. Using the definition of c_k and replacing $\sum_{s=1}^{n} \psi_s(c_k) = \frac{1}{p(c_k, \widetilde{s}(c_k))}$ (from (51)) into (54), we obtain

$$\begin{aligned} g(c) &= \frac{K(c_k)}{p(c_k, \widetilde{s}(c_k))} + cq^{sb}(c) \sum_{s=1}^{n} \psi_s(c) - \sum_{s=1}^{n} h(c,s) \psi_s(c) + h(c, \widetilde{s}(c)) \\ &+ \int_c^{c_k} \left\{ \sum_{s=1}^{n} \left[q^{sb}(x) \psi_s(x) - \left[\underline{\pi}^{sb}(x,s) + K(x) \right] \psi_s'(x) \right] \right\} dx. \end{aligned}$$

We further calculate

$$\begin{aligned}
& \sum_{s=1}^{n} cq^{sb}(c)\,\psi_s(c) - \sum_{s=1}^{n} h(c,s)\,\psi_s(c) + h(c,\widetilde{s}(c)) \\
= & \sum_{s=1}^{n}\left[cq^{sb}(c) - h(c,s)\right]\psi_s(c) + h(c,\widetilde{s}(c)) \\
= & \sum_{s=1}^{n}\left[-K(c) - \underline{\pi}^{sb}(c,s)\right]\psi_s(c) + \underline{\pi}^{sb}(c,\widetilde{s}(c)) + cq^{sb}(c) + K(c) \\
= & \sum_{s=1}^{n}(-K(c)+L)\,\psi_s(c) - L + cq^{sb}(c) + K(c) \\
= & \left[L - K(c)\right]\frac{1-p(c,\widetilde{s}(c))}{p(c,\widetilde{s}(c))} + cq^{sb}(c)
\end{aligned}$$

and then substitute into the expression of $g(c)$ above. Rearranging yields

$$\begin{aligned}
g(c) = {} & cq^{sb}(c) + \int_c^{c_k} q^{sb}(x)\left(\sum_{s=1}^{n}\psi_s(x)\right)dx + \frac{1-p(c,\widetilde{s}(c))}{p(c,\widetilde{s}(c))}\left[L-K(c)\right] \\
& +L\int_c^{c_k}\left(\sum_{s=1}^{n}\psi_s'(x)\right)dx - \int_c^{c_k}\sum_{s=1}^{n}\big(K(x)\,\psi_s'(x)\big)\,dx + \frac{K(c_k)}{p(c_k,\widetilde{s}(c))}.
\end{aligned}$$

Integrating by parts $\int_c^{c_k}\left(\sum_{s=1}^{n}\psi_s'(x)\right)dx$, where $\psi_s(x)$ is defined by (51), together with $\int_c^{c_k}\sum_{s=1}^{n}\big(K(x)\,\psi_s'(x)\big)\,dx$ and then replacing into the above expression of $g(c)$, we find

$$\begin{aligned}
g(c) = {} & cq^{sb}(c) + K(c) + \frac{1-p(c_k,\widetilde{s}(c))}{p(c_k,\widetilde{s}(c))}L \\
& + \int_c^{c_k}\sum_{s=1}^{n}\left[q^{sb}(x) + K'(x)\right]\psi_s(x)\,dx.
\end{aligned} \tag{61}$$

Using (61) in $(2a)$ for $t(r,\widetilde{s}(r)) = g(r)$ (knowing that $g(r)$ is the transfer that corresponds to type $\widetilde{s}(r)$), the reward of the agent when he reports $r\in[\underline{c},c_1]$ and $\widetilde{s}(r)=n$ is written

$$\pi(r,n) = \int_r^{c_1}\frac{q^I(x)+K'(x)}{p(x,n)}dx + \frac{1-p(c_1,n)}{p(c_1,n)}L. \tag{62}$$

From the proof of Lemma 6, $\pi(r,s) = -L$ whenever $r\in[\underline{c},c_1]$ and $s\neq n$.

The *interim* profit Using (62) and $\pi(r,s) = -L$ for $s\neq n$ in (60), $E_s[\widetilde{\pi}(r,s)]$ is rewritten

$$\begin{aligned}
E_s[\widetilde{\pi}(r,s)] = {} & -\int_r^{c}\left[q^I(r)+K'(x)\right]dx + p(c,n)\int_r^{c_1}\frac{q^I(x)+K'(x)}{p(x,n)}dx \\
& -\left[1-\frac{p(c,n)}{p(c_1,n)}\right]L,
\end{aligned} \tag{63}$$

whereas the *interim* profit from a truthful report $r = c$ is given by

$$E_s\left[\pi\left(c,s\right)\right] = p\left(c,n\right)\int_c^{c_1}\frac{q^I\left(x\right)+K'\left(x\right)}{p\left(x,n\right)}dx - \left[1-\frac{p\left(c,n\right)}{p\left(c_1,n\right)}\right]L. \tag{64}$$

Sub-case $c \in [\underline{c}, c_1]$ Using (63) and (64), we have $E_s\left[\pi\left(c,s\right)\right] \geq E_s\left[\widetilde{\pi}\left(r,s\right)\right]$ if and only if

$$\int_c^r\left[q^I\left(x\right)+K'\left(x\right)\right]\left[1-\frac{p\left(c,n\right)}{p\left(x,n\right)}\right]dx + \int_c^r\left[q^I\left(r\right)-q^I\left(x\right)\right]dx \leq 0. \tag{65}$$

This condition is satisfied whenever so is (29) in Proposition 3.

Sub-case $c \in [c_1, \overline{c}]$ Assume that $r = c_1$ and calculate from (63) the following

$$\begin{aligned}
\frac{dE_s\left[\widetilde{\pi}\left(c,s\right)\right]}{dc} &= -\left[q^I\left(c_1\right)+K'\left(c\right)\right] + \frac{L}{p\left(c_1,n\right)}\frac{dp\left(c,n\right)}{dc} \\
&= -\left[q^I\left(c_1\right)+K'\left(c_1\right)\right] + \frac{L}{p\left(c_1,n\right)}\frac{dp\left(c_1,n\right)}{dc_1} \\
&\quad +K'\left(c_1\right)-K'\left(c\right)+\frac{L}{p\left(c_1,n\right)}\left[\frac{dp\left(c,n\right)}{dc}-\frac{dp\left(c_1,n\right)}{dc_1}\right] \\
&= K'\left(c_1\right)-K'\left(c\right)+\frac{L}{p\left(c_1,n\right)}\left[\frac{dp\left(c,n\right)}{dc}-\frac{dp\left(c_1,n\right)}{dc_1}\right].
\end{aligned}$$

With (31) from Proposition 3 satisfied for $\widetilde{s}\left(c\right) = n$, $\frac{dE_s\left[\widetilde{\pi}\left(c_1,s\right)\right]}{dc} \leq 0$. Moreover, if $c = c_1$, then $E_s\left[\widetilde{\pi}\left(c_1,s\right)\right] = E_s\left[\pi\left(c_1,s\right)\right]$, which is zero by Lemma 5. Therefore, under (31), $E_s\left[\widetilde{\pi}\left(c_1,s\right)\right] \leq 0$ whenever $c \in [c_1, \overline{c}]$ and $r = c_1$.

Take now $r \leq c_1$ and calculate

$$\frac{dE_s\left[\widetilde{\pi}\left(r,s\right)\right]}{dr} = -\int_r^c\left\{\left[q^I\left(r\right)+K'\left(r\right)\right]\frac{dp\left(x,n\right)/dx}{p\left(r,n\right)}+\frac{dq^I\left(r\right)}{dr}\right\}dx.$$

We look for the condition under which $\frac{dE_s\left[\widetilde{\pi}\left(r,s\right)\right]}{dr} \geq 0$. Because $c \geq c_1$ and $r \leq c_1$, this inequality holds if and only if

$$\frac{dq^I\left(r\right)}{dr} \leq -\left[q^I\left(r\right)+K'\left(r\right)\right]\frac{dp\left(x,n\right)/dx}{p\left(r,n\right)}, \forall r \in [\underline{c}, c_1] \text{ and } x \geq c_1, \tag{66}$$

which is implied by (29) together with Assumption 2 and $x \geq r_i$. As $\frac{dE_s\left[\widetilde{\pi}\left(r,s\right)\right]}{dr} \geq 0$ $\forall r \in [\underline{c}, c_1]$ and $c \in [c_1, \overline{c}]$, whereas $E_s\left[\widetilde{\pi}\left(c_1,s\right)\right] \leq 0$ (as previously found), one has $E_s\left[\widetilde{\pi}\left(r,s\right)\right] \leq 0$, $\forall r \in [\underline{c}, c_1]$ and $c \in [c_1, \overline{c}]$.

Overall, (29) and (31) ensure that the agent has no incentive to report $r \in [\underline{c}, c_1]$ such that $r \neq c$, whatever his real type.

B.4.2 Case $r \in [c_1, c_2]$

As $\pi\left(r,s\right) = -L$ for $s \neq n$ and $E_s\left[\pi\left(r,s\right)\right] = 0$ $\forall r \in [c_1, c_2]$ (by Lemma 4), we have $\pi\left(r,n\right) = \frac{1-p(r,n)}{p(r,n)}L$. Substituting these values of $\pi\left(r,s\right)$ into $E_s\left[\widetilde{\pi}\left(r,s\right)\right]$, together with $q^{II}\left(r\right) = \frac{dp(r,n)/dr}{p(r,n)}L - K'\left(r\right)$ (from Lemma 6), we find that, with regards to this interval,

(60) specifies as

$$E_s\left[\widetilde{\pi}^{II}\right] = \int_c^r \left\{\frac{L}{p(r,n)}\left[\frac{dp(r,n)}{dr} - \frac{dp(x,n)}{dx}\right] + K'(x) - K'(r)\right\} dx.$$

From the expression above, condition (31) and $\widetilde{s}(c) = n$, we can establish that $E_s\left[\widetilde{\pi}^{II}\right] \leq 0$.

B.4.3 Case $r \in [c_2, c_3]$

Proposition 2 shows that, in this case, the condition for global incentive compatibility is given by (14), which is implied by (31).

B.4.4 Case $r \in [c_3, c_4]$

Proceeding as we did for $r \in [c_1, c_2]$, we find that the payoff of the agent when he reports r is written

$$\begin{aligned} E\left[\widetilde{\pi}^{IV}\right] &= \int_c^r \left[L\frac{dp(r,1)/dr}{p(r,1)} - K'(r) + K'(x) - \frac{L}{p(r,1)}\frac{dp(x,1)}{dx}\right] dx \\ &= \int_c^r \left\{\frac{L}{p(r,1)}\left[\frac{dp(r,1)}{dr} - \frac{dp(x,1)}{dx}\right] + K'(x) - K'(r)\right\} dx. \end{aligned}$$

Together with $\widetilde{s}(c) = 1$, (31) in Proposition 3 involves that $E\left[\widetilde{\pi}^{IV}\right] \leq 0$.

B.4.5 Case $r \in [c_4, \overline{c}]$

We follow the same steps as with the very first case.

The *ex post* profit $\pi(r,s)$ Using (61) in $(2a)$ for $t(r, \widetilde{s}(r)) = g(r)$, the reward of the agent when he reports $r \in [c_4, \overline{c}]$ and $\widetilde{s}(r) = 1$ is written

$$\pi(r,1) = -\int_{c_4}^r \frac{q^V(x) + K'(x)}{p(x,1)} dx + \frac{1 - p(c_4,1)}{p(c_4,1)} L, \tag{67}$$

From the proof of Lemma 6, $\pi(r,s) = -L$ whenever $r \in [\underline{c}, c_4]$ and $s \neq 1$.

The *interim* profit The *interim* profit of the agent when he reports r is given by

$$\begin{aligned} E\left[\widetilde{\pi}^{V}\right] &= \int_c^r \left[q^V(r) + K'(x)\right] dx - \int_{c_4}^r \left[q^V(x) + K'(x)\right] \frac{p(c,1)}{p(x,1)} dx \\ &\quad -L\left[1 - \frac{p(c,1)}{p(c_4,1)}\right], \end{aligned} \tag{68}$$

whereas the *interim* profit in case of truthtelling is written

$$E_s[\pi(c,s)] = -\int_{c_4}^c \left[q^V(x) + K'(x)\right] \frac{p(c,1)}{p(x,1)} dx - L\left[1 - \frac{p(c,1)}{p(c_4,1)}\right]. \tag{69}$$

Sub-case $c \in [c_4, \overline{c}]$ Using (68) and (69), we have $E_s[\pi(c,s)] \geq E_s[\widetilde{\pi}(r,s)]$ if and only if

$$\int_c^r \left[q^V(x) + K'(x)\right]\left[1 - \frac{p(c,1)}{p(x,1)}\right] dx + \int_c^r \left[q^V(r_i) - q^V(x)\right] dx \leq 0.$$

This condition is satisfied whenever so is (30) in Proposition 3.

Sub-case $c \notin [c_4, \overline{c}]$ Take first $r = c_4$ and calculate

$$\begin{aligned}
\frac{dE\left[\widetilde{\pi}^V\right]}{dc} &= -\left[q^V(c_4) + K'(c)\right] + \frac{dp(c,1)/dc}{p(c_4,1)} L \\
&= -\left[q^V(c_4) + K'(c_4)\right] + \frac{dp(c_4,1)/dc_4}{p(c_4,1)} L \\
&\quad + K'(c_4) - K'(c) + \frac{dp(c,1)/dc - dp(c_4,1)/dc_4}{p(c_4,1)} L \\
&= K'(c_4) - K'(c) + \frac{dp(c,1)/dc - dp(c_4,1)/dc_4}{p(c_4,1)} L.
\end{aligned}$$

One has $\frac{dE\left[\widetilde{\pi}^V\right]}{dc} \geq 0$ if (31) in Proposition 3 holds. Moreover, $E\left[\widetilde{\pi}^V\right] = 0$ if $c = r = c_4$. This shows that any type $c \notin [c_4, \overline{c}]$ that reports $r = c_4$ obtains $E\left[\widetilde{\pi}^V\right] \leq 0$. Furthermore,

$$\begin{aligned}
\frac{dE\left[\widetilde{\pi}^V\right]}{dr} &= \left[q^V(r) + K'(r)\right]\left[1 - \frac{p(c,1)}{p(r,1)}\right] + \int_c^r \frac{dq^V(r)}{dr} dx \\
&= \int_c^r \left\{\frac{dq^V(r)}{dr} + \left[q^V(r) + K'(r)\right]\frac{dp(x,1)/dx}{p(r,1)}\right\} dx.
\end{aligned}$$

$E\left[\widetilde{\pi}^V\right] \leq 0$ for any report $r \in [c_4, \overline{c}]$ if $\frac{dE\left[\widetilde{\pi}^V\right]}{dr} \leq 0$, which is implied by

$$\frac{dq^V(r)}{dr} \leq -\left[q^V(r) + K'(r)\right]\frac{dp(x,1)/dx}{p(r,1)}.$$

In turn, this is implied by (30) in Proposition 3 together with Assumption 2 and $x \leq r$.

Overall, the agent has no incentive to report $r \in [c_4, \overline{c}]$ such that $r \neq c$ whenever (30) and (31) are satisfied.

B.5 Proof of Proposition 4

We proceed as follows. We begin by showing that, whenever (31) is violated for any feasible c, at the SB solution, there exists no $c \neq \widehat{c}$ for which (PC) is binding. We then rewrite (Γ) for the situation in which (PC) is not binding $\forall c \neq \widehat{c}$ and show that there exists a unique cost range over which pooling arises.

Suppose (PC) is binding over some non empty interval $\left[c^L, c^H\right]$, with either $c^L \neq \widehat{c}$ or $c^H \neq \widehat{c}$ or both. Assume also that FB is not implementable over this interval at the solution to (Γ). From the proof of Lemma 6, the SB quantity would be $q^{II}(c), \forall c \in \left[c^L, c^H\right]$. Furthermore, the proof of Proposition 3 shows that the quantity $q^{II}(c)$ and the transfers that leave no rent to the agent are not implementable when (31) is not satisfied for any feasible c. This contradicts the assumption that (PC) is binding and, at the same time, FB is not at hand for types in $\left[c^L, c^H\right]$.

Now suppose that, at the solution to (Γ), (PC) is binding and FB is implemented $\forall c \in \left[c^L, c^H\right]$. From Lemma 5, it follows that there exist other cost values around $\left[c^L, c^H\right]$ for which (PC) is binding and the SB quantity is $q^{II}(c)$. Once again, as (31) is not satisfied for any feasible c, this contradicts also the assumption that FB is enforced $\forall c \in \left[c^L, c^H\right]$.

Overall, there exists no subset $\left[c^L, c^H\right]$, with either $c^L \neq \widehat{c}$ or $c^H \neq \widehat{c}$ or both, in which (PC) is binding. It follows that (PC) is slack for $\forall c \neq \widehat{c}$. Hence, the interval $[c_2, c_4]$ defined by Lemma 4 reduces to the singleton $\{\widehat{c}\}$. From the proof of Proposition 3, the scheme is globally incentive compatible whenever

$$\frac{dq^{sb}(c)}{dc} \leq -\left[q^{sb}(c) + K'(c)\right] \frac{dp(c, \widetilde{s}(c))/dc}{p(c, \widetilde{s}(c))}, \ \forall c \in [\underline{c}, \overline{c}], c \neq \widehat{c}, \tag{70}$$

$q^{sb}(c)$ being the SB quantity for type c. We can thus rewrite (Γ) as

$$\max_{q(c)} \widetilde{W} \equiv \left[\widetilde{W}(\underline{c}, \widehat{c}) + \widetilde{W}(\widehat{c}, \overline{c})\right] \quad s.t. \ (70),$$

where $\widetilde{W}(\underline{c}, \widehat{c})$ and $\widetilde{W}(\widehat{c}, \overline{c})$ as defined in the proof of Lemma 6 are such that, beside (70), all other relevant constraints are satisfied. In particular, $\widetilde{W}(\underline{c}, \widehat{c})$ is defined by (48) with c_1 replaced by $\widehat{c}$ and $\underline{\pi}^{sb}(c, s) = -L$. For low and high types, the optimal quantities are still $q^I(c)$ and $q^V(c)$ respectively, as characterized by (17) and (21) in Lemma 6. However, such quantities do not satisfy (70) in a neighborhood of $\widehat{c}$. To see this, rewrite (70) as

$$\begin{aligned} q^{sb}(c) &\geq q^{sb}(\widehat{c}) + \int_c^{\widehat{c}} \left[q^{sb}(x) + K'(x)\right] \frac{dp(x,n)/dx}{p(x,n)}, \ \forall c \in [\underline{c}, \widehat{c}) \\ q^{sb}(c) &\leq q^{sb}(\widehat{c}) - \int_{\widehat{c}}^{c} \left[q^{sb}(x) + K'(x)\right] \frac{dp(x,1)/dx}{p(x,1)}, \ \forall c \in (\widehat{c}, \overline{c}]. \end{aligned}$$

In either inequality, the RHS approaches $q^{sb}(\widehat{c})$ for c close to $\widehat{c}$. Moreover, at the solution to (Γ), $q^{sb}(\widehat{c}) = q^{sb}(\widehat{c})$ and $q^I(c) < q^{fb}(c) < q^V(c)$ $\forall c \in [\underline{c}, \overline{c}]$. Hence, pooling of output arises around $\widehat{c}$. The pooling interval is unique for the same reasons as in Lewis and Sappington [6] (compare pp.309-310 in their article) and the proof is here omitted. As the unique pooling interval includes $\widehat{c}$ and as $q^{sb}(\widehat{c}) = q^{fb}(\widehat{c})$, we have $q^{sb}(c) = q^{fb}(\widehat{c})$ $\forall c \in [c^-, c^+]$, where c^- is defined by $q^I(c^-) = q^{fb}(\widehat{c})$ and c^+ by $q^V(c^+) = q^{fb}(\widehat{c})$.