Working Paper

Regulatory Mechanism Design with Extortionary Collusion[1]

Dilip Mookherjee[2] and Masatoshi Tsumagari[3]

This Version: December 29, 2022

Abstract

We study regulatory mechanism design with extortionary collusion between a privately informed agent and a less well-informed supervisor, incorporating 'extortion' which permits redistribution of rents within the coalition. We show the Collusion-proofness Principle holds, and that relative bargaining power of the supervisor and agent matters. Specifically, the Principal does not benefit from hiring the supervisor if the latter has less ex ante bargaining power vis-a-vis the agent. We provide sufficient conditions for the supervisor to be valuable if she has greater bargaining power. These results suggest the importance of anti-collusion strategies that augment bargaining power of supervisors vis-a-vis agents.

KEYWORDS: mechanism design, supervision, collusion, bargaining power

[1]We thank Associate Editor and Referees of this journal, Alessandro Bonatti, Robert Gibbons, Alberto Motta, Juan Ortner, Silvia Vanutelli and participants in seminars at Boston University, MIT and New York University for useful discussions and comments.

[2]Department of Economics, Boston University, 270 Bay State Road Boston MA 02215; dilipm@bu.edu

[3]Department of Economics, Keio University, 2-15-45 Mita Minato-ku Tokyo 108-8345, Japan; tsuma@econ.keio.ac.jp

1 Introduction

The design of mechanisms to limit the harmful effects of collusion between supervisors and agents in adverse selection settings has been studied by many authors following Tirole (1986) and Laffont and Tirole (1993), with many applications to design of procurement, regulation and internal organization of firms. Subsequent literature has explored the implications of enlarging the severity of the collusion problem, such as soft information which provides wider scope for manipulation of reports (Faure-Grimaud et al. (2003), Celik (2009)), and collusion over both reporting and participation decisions (Mookherjee et al. (2020)).

All these papers however assume that collusion is 'weak' in the sense that the supervisor and agent play non-cooperatively if either vetoes the offered side-contract. This ensures collusion occurs only to realize joint gains for the colluding parties. The allocation of bargaining power within the coalition then does not matter for the Principal's capacity to control corruption. This implication of weak collusion follows from the Collusion-Proof Principle which applies quite generally in this class of models (Tirole (1992)).[4] It is important to note that the irrelevance of bargaining power (among parties engaging in weak collusion) applies much more generally than conditions usually needed for the Coase Theorem to apply. For instance, bargaining takes place with asymmetric information between the supervisor and agent, either at the ex ante or interim stage (as defined by Holmstrom and Myerson (1983)). These

[4]For details of the argument, see Faure-Grimaud et al. (2003) in a setting where the supervisor and the agent collude only over reporting, or Mookherjee et al. (2020) where they collude both on participation and reporting decisions. The Principle implies that attention can be confined to revelation mechanisms which do not allow scope for any deviating (feasible) side contract for the coalition to increase the welfare-weighted sum of their payoffs. Given transferability of utility (owing to presence of side payments and absence of binding budget constraints for either colluding party), it follows that there cannot exist a deviating feasible side contract which makes both of them strictly better off. This property is independent of the relative welfare weights, i.e., if it holds for one set of weights it also holds for any other.

models are therefore incapable of explaining the observed success of anti-corruption policies which lowered bargaining power of the agent vis-a-vis the supervisor by preventing agents from selecting their own auditor in India (Duflo et al. (2013)) and Italy (Vanutelli (2020)) .[5]

In this paper we argue that a strengthening of the nature of collusion to allow for 'extortion' can explain why relative bargaining power of the supervisor could matter. Extortion permits the colluding partner with greater bargaining power to extract rents from the other party, by threatening to send reports to the Principal which would hurt the latter if (s)he did not agree to the offered side-contract. Colluding parties can commit to such threats at the time of bargaining. This alters the sub-game following refusal of an offered side-contract by one party, from a simultaneous move noncooperative game, to one where the report of the accepting party is stipulated by the side-contract (with the other party choosing a best response). Earlier work by Dequiedt (2007) and Che and Kim (2009) has studied an analogous notion in an auction design setting, using the term 'strong collusion'. As some other authors (e.g., Quesada (2003)) have used the term 'strong collusion' to mean something different, we refer to the combination of extortion and collusion as 'extortionary collusion'.

While we do not endogenize the source of commitment power, we have in mind settings where S and A interact with one another in many other transactions with many principals, either at the same time or at later dates, and reputational concerns provide the required enforcement.[6] It should be added that this is consistent

[5]See Section 2 for more details of these policies and resulting outcomes.

[6]Consider a setting with many customers, each of whom wants to contract with an agent to commission a required production task (such as remodeling a kitchen), along with a supervisor who is better informed than customers about the cost of remodeling their respective kitchens. There are many ex ante identical potential agents and brokers/supervisors in the industry which constitutes a close-knit community, in which members know one another, enter into side-contracts whose outcomes are observed by the rest of the community. If they are far-sighted, any agent and supervisor appointed by a given customer have an incentive to develop and maintain a reputation for following through on promises and threats they make during side contracting.

with the rest of the mechanism design literature which examines the consequence of alternative assumptions regarding commitment among players, rather than providing an explicit microfoundation for these assumptions. The existing literature which has studied weak collusion is based on a specific assumption of lack of commitment among colluding players. Here we are interested in understanding the consequence of the opposite assumption where they can commit to reporting threats when they fail to agree on a side-contract.

In a law review article, Ayres (1997) refers to 'bribery' and 'extortion' as the 'twin faces of corruption'. In the language of mechanism design theory, bribery corresponds to weak collusion, while the combination of bribery and extortion corresponds to extortionary collusion. The importance of extortion has been stressed by numerous authors in descriptive accounts of corruption, from medieval England (Cam (1930)), to more contemporary accounts of corruption in Burma (Furnivall (1956)) or other developing countries (Klitgaard (1988)). Ayres (1997) and Andrionova and Melissas (2008) discuss extortion from a legal standpoint. Some papers studying tax evasion, regulations or intra-firm organization in a moral hazard setting (Mookherjee (1997), Hindricks et al. (1999), Khalil et al. (2010)) have shown that anti-corruption policy design is significantly altered by the presence of extortion. By contrast our focus is on an adverse selection context, where the modeling issues as well as results are quite different. Section 2 provides a more detailed discussion of the relation to existing literature.

Section 3 describes the model of extortionary collusion in which S obtains a noisy signal of A's cost, and A also observes the realization of this signal. A and S enter into a collusive side contract at the ex ante stage, when neither have received their respective signals. Moreover, the allocation of bargaining power between the supervisor (S) and agent (A) is given and known by the Principal (P). The primary question studied in this paper is how relative bargaining power matters in the presence of extortion. In some contexts P may have no control over the allocation of bargaining power between

S and A. Our results will show that the consequences of extortion on P's welfare and the value of hiring S vary with bargaining power allocation. In other contexts P may be able to influence bargaining power: in these settings our results imply that raising relative bargaining power of S can be an important policy instrument.

For instance, P may be able to influence the process by which S and A are appointed, which affects their relative bargaining power. Consider for instance a setting where A is allowed to appoint her own supervisor, from a set of potential supervisors. This would give A the opportunity to exercise monopsony power over S in the choice of side contract, resulting in a higher welfare weight on A's payoff. An alternative institutional rule (as in the policy reforms in India and Italy studied by Duflo et al. (2013) and Vanutelli (2020)) is one where the supervisor is selected by P instead, and assigned to a given agent. This would alter the side contract negotiation to a bilateral monopoly where bargaining power is more equally divided between S and A. In some settings (e.g., in construction, or procurement of a particular service) there could be many potential agents available to carry out the project for P, where P could appoint a supervisor S and delegate the choice of A to S. This would confer monopoly power to S over A.

Given a certain allocation of bargaining power within the coalition, Section 5 verifies that the Collusion Proofness Principle continues to apply in a extortionary collusion setting, once message spaces are augmented to include some non-type messages. This result is of some independent interest, as it contrasts with the moral hazard setting studied by Khalil et al. (2010). Other differences in our results from the moral hazard setting are described in Section 2. The result is used to characterize the class of feasible allocations in terms of a set of coalition incentive compatibility constraints.

This characterization is used in Sections 5 and 6 to provide results for alternative ranges of bargaining power allocation. First, if A has a greater bargaining power than

S,[7] appointing a supervisor is worthless for P (Proposition 2). The intuition for this result is that when A has greater bargaining power, extortion allows A to extract all of S's rents, since A knows everything that S knows. Hence A becomes the residual claimant on the coalition's total surplus, and is able to push S down to her true outside option. This eliminates all bargaining frictions within the coalition, and the presence of S becomes redundant (by having no choice other than to 'rubber-stamp' whatever A wants to report).

Our next result (Proposition 3) considers a more restricted setting where the good to be procured by P is indivisible. Here we provide sufficient conditions for hiring S to be valuable when S has greater bargaining power than A, and for P's welfare to be increasing in S's relative bargaining power. We also show that P's welfare is always strictly lower in extortionary collusion compared with weak collusion. Besides, this section provides an illustration how our preceding results can be used to analyze the corresponding optimal contracting problem with extortionary collusion.[8]

The Concluding Section discusses consequences of extending the model in different directions, while the Appendix provides details of proofs omitted from the text.

2 Related Literature

In the mechanism design literature, implications of a similar notion of extortionary collusion have been studied in the context of auction design featuring collusion among bidders, by Dequiedt (2007) and Che and Kim (2009). Quesada (2003) uses the term 'strong collusion' to refer to a different concept, which requires every Bayesian equilibrium of the game induced by P's contract to be weakly collusion proof. We are not aware of any previous study of the implications of extortionary collusion

[7] By this we mean A is assigned a higher welfare weight in the selection from the set of feasible side-contract payoff pairs.

[8] However we restrict attention to deterministic contracts in this setting, which may entail some loss of generality.

in settings of supervision with a privately informed agent.[9] In such a setting, our analysis shows how standard characterizations of feasible and optimal allocations in the existing literature on weak collusion (Faure-Grimaud et al. (2003), Celik (2009), Mookherjee et al. (2020)) need to be modified, besides providing results concerning the costs imposed by extortion and how they depend on intra-coalition bargaining power allocation. Most importantly, manipulating the bargaining power of the supervisor becomes an important tool of the Principal in curbing the costs of corruption, by increasing asymmetric information frictions within the supervisor-agent coalition.[10]

The practical relevance of these results is highlighted by recent policy experience. Duflo et al. (2013) study a controlled experiment in India during 2009-10 where a treatment group of firms were no longer allowed to appoint their own pollution auditors, but were randomly assigned auditors instead. They found significant increases in pollution reports by the assigned auditors, and corresponding decline in actual pollution levels (verified from special backchecks conducted by the research team). Vanutelli (2020) studies a related policy reform in Italy introduced in 2011, where auditors of municipal budgets were randomly assigned instead of being appointed by local mayors. This resulted in increased property tax collections, larger budget surpluses and debt repayments. These effects were significantly larger in places with higher 'risk of corruption' (measured by prior investigations of corruption-related crimes).[11]

[9] However there have been a number of studies of extortionary collusion in a moral hazard setting, described below.

[10] This is broadly similar though different in details from the strategy stiudied by Ortner and Chassang (2018) and von Negenborn and Pollrich (2020) in which P deliberately creates asymmetric information between agent and monitor by randomizing the latter's incentive contract and not letting the agent observe the monitor's contract with P. This particular tool is presumed unavailable in the settings we examine, e.g., the incentive contract for both parties are required to be in the public domain, as is commonly the case for public sector procurement or regulatory settings.

[11] Using suitable extensions of our model, an earlier version of this paper illustrated how random

The consequences of extortion for agent incentives in a supervision setting with moral hazard have been studied by a class of models (Mookherjee (1997), Hindricks et al. (1999), Khalil et al. (2010)). The most closely related paper in this group is Khalil et al. (2010), which also finds that extortion imposes a larger cost to the Principal's welfare when the agent has larger bargaining power. However, many features and results of their model differ from ours. The information structure is different (the supervisor is either perfectly informed or perfectly uninformed; the supervisor's information is hard unless the agent agrees to collude), and the Collusion-Proofness Principle does not hold in their setting. Most of their attention is subsequently devoted to the question whether bribery (i.e., resulting in mutual gain to both supervisor and agent) or extortion is the 'greater evil'. This question is not meaningful in our setting as the Collusion-Proof Principle applies in our setting and thereby implies the optimality of eliminating both bribery and extortion. Moreover, in contrast to our results, hiring the supervisor is still valuable in their model even if the agent has all the bargaining power within the coalition (when the supervisor's signal is informative enough). Extortion is costless in their setting if the supervisor has all the bargaining power; this may not be the case in our model (as verified in the indivisible good setting of Section 6).

3 Model

An appointed agent A delivers an output q to the Principal P at a personal cost of θq. Let $Q \subseteq \Re_+$ denote the set of feasible outputs. We do not impose any restriction on Q except for $0 \in Q$. In Section 6 we consider the specific case where $Q = \{0, 1\}$. P's return from q is denoted by $V(q)$ which is increasing on Q, with $V(0) = 0$.

assignment of a supervisor would raise bargaining power of the supervisor relative to a context where the agent appoints the supervisor. The idea is simple: if there are many competing potential agents and supervisors on each side of the market, giving a chosen agent or supervisor the power to appoint the other party would tilt relative bargaining power in its favor.

The realization of θ is privately observed by A. Θ, which denotes the support of θ, constitutes an interval $[\underline{\theta}, \bar{\theta}] \subset (0, \infty)$. It is common knowledge that everybody shares a common distribution function $F(\theta)$ over Θ. It has a density function $f(\theta)$ which is continuously differentiable and everywhere positive on Θ.

An appointed supervisor S costlessly acquires an informative signal $\eta \in \Pi \equiv \{\eta_1, \eta_2, ..., \eta_m\}$ about A's cost θ with $m \geq 2$.[12] The realization of S's signal is observed by A. $a(\eta \mid \theta) \in [0, 1]$, which denotes the likelihood function of η conditional on θ, is continuously differentiable and positive-valued on Θ.[13] We assume that for any $\eta \in \Pi$, $a(\eta \mid \theta)$ is not a constant function on Θ, and there are some subsets of θ with positive measure satisfying $a(\eta \mid \theta) \neq a(\eta' \mid \theta)$ for every $\eta, \eta' \in \Pi$. In this sense each possible signal realization conveys information about the agent's cost. The information conveyed is partial, since Π is finite. The distribution function over θ conditional on η is denoted $F(\theta|\eta)$. Conditional on η, the density function and distribution function are respectively denoted by $f(\theta \mid \eta) \equiv f(\theta)a(\eta \mid \theta)/p(\eta)$ and $F(\theta \mid \eta) \equiv \int_{\underline{\theta}}^{\theta} f(\theta \mid \eta) d\theta$, where $p(\eta) \equiv \int_{\underline{\theta}}^{\bar{\theta}} f(\theta)a(\eta \mid \theta) d\theta$. Let $K \equiv \Theta \times \Pi$ denote the set of possible states.

All players are risk neutral. S's payoff is $u_S = X_S + t_S$ where X_S is the transfer from P to S, and t_S is a transfer received by S within the coalition. A's payoff is $u_A = X_A + t_A - \theta q$ where X_A is the transfer from P to A and t_A is a transfer received by A within the coalition. Transfers within the coalition are subject to a budget balance condition $t_A + t_S \leq 0$. P's objective is to maximize the expected value of

[12] Even though our result of the collusion-proof principle does not depend on the discrete property of Π, this assumption simplifies the exposition. If S incurs a fixed cost c to acquire the signal, transfers received by S must be replaced by transfers net of this fixed cost while measuring S's payoff. Increases in c will of course lower the value of appointing the supervisor, but it is easy to see how the results will be modified.

[13] This assumes that the support of θ given η is Θ for all η, in which sense θ has full support. We adopt this assumption purely to simplify the exposition; our results extend to the case of non-full support. See Concluding Section 7.

profit $V(q) - X_A - X_S$.[14]

4 Mechanism, Collusion Game and Equilibrium Concept

P designs a grand contract (GC) played by A and S, describing transfers and production decisions (X_A, X_S, q) made by P in response to message sent by A and S. It is possible that (X_A, X_S, q) is randomized conditional on messages. Owing to the risk neutrality of A and S and linearity of their payoff functions, grand contracts generating the same expected value of (X_A, X_S, q) for any given message generate the same payoffs and incentives for A and S. Hence they induce the same expected allocation (i.e., expected value of (X_A, X_S, q) in each state of the world) in any equilibrium of the collusion game (to be specified below). Moreover, if some expected allocation is achieved in an equilibrium, any randomized allocation with the same expected value is also achievable with the use of a stochastic grand contract. Define $\bar{Q} \equiv [0, \sup Q]$, which is the set of feasible values of expected output. This argument implies that without loss of generality, P can restrict attention to grand contracts with deterministic transfers, which assigns $(X_A, X_S, q) \in \Re^2 \times \bar{Q}$ for any possible message combination:

$$GC = (X_A(m_A, m_S), X_S(m_A, m_S), q(m_A, m_S); M_A, M_S).^{15}$$

[14] Our analysis applies to the case that P maximizes a weighted average of profit $(V(q) - X_A - X_S)$ and welfare of A and S $(u_A + u_S)$, with a lower relative weight on the latter.

[15] Note that the allocations with the same expected value may not always perceived to be identical by P if $V(q)$ is not linear. Moreover, outputs may need to be randomized. If Q is an interval of the real line and V is concave, P would not benefit from randomizing the output, in which case attention can be restricted to deterministic contracts. In the indivisible good case $Q = \{0, 1\}$ examined in Section 6, P may conceivably benefit from randomizing the output. However, in Section 6 we restrict attention to deterministic contracts, and explain why this does not cause any problems for our main results.

M_A (resp. M_S) denotes a message set for A (resp. S). Message spaces include exit options for A and S respectively ($e_A \in M_A, e_S \in M_S$), where $X_A = q = 0$ whenever $m_A = e_A$, and $X_S = 0$ whenever $m_S = e_S$. The set of grand contracts satisfying these restrictions is denoted by $\mathcal{GC}$. As a special case, P has the option to not hire S, which we denote by *No Supervision (NS)*, where M_S is null and $X_S = 0$. We assume that GC is observable publicly, which makes it infeasible to combat with a collusion through the creation of asymmetric information about the mechanism between A and S, in contrast to Ortner and Chassang (2018) and von Negenborn and Pollrich (2020).

We allow message choices to be randomized.[16] Let $\Delta(M_A)$, $\Delta(M_S)$ and $\Delta(M)$ denote the set of probability measures on M_A, M_S and $M \equiv M_A \times M_S$ respectively. For $(\mu_A, \mu_S) \in \Delta(M_A) \times \Delta(M_S)$ and $\mu \in \Delta(M)$, the expected values of allocations assigned in grand contracts are as follows:[17]

$$\begin{aligned}\bar{GC} &\equiv (\bar{X}_A(\mu_A, \mu_S), \bar{X}_S(\mu_A, \mu_S), \bar{q}(\mu_A, \mu_S)) \\ &= \int_{M_A}\int_{M_S} (X_A(m_A, m_S), X_S(m_A, m_S), q(m_A, m_S)) d\mu_A(m_A) d\mu_S(m_S)\end{aligned}$$

and

$$\hat{GC} \equiv (\hat{X}_A(\mu), \hat{X}_S(\mu), \hat{q}(\mu)) = \int_M (X_A(m), X_S(m), q(m)) d\mu(m).$$

$\bar{GC}$ denotes the expected value of the allocation (assigned for any given message combination) when A and S do not collude, and thus select their messages independently. $\hat{GC}$ denotes the corresponding expected value of the allocation when A and S collude, by coordinating on their respective messages according to the joint distribution μ.

Collusion between A and S takes the form of a side contract (SC) which is unobserved by P. As explained in the Introduction, we treat the allocation of bargaining

[16]This helps with the analysis in two ways. First we can apply a standard minimax theorem. Second, we can justify the use of the standard Lagrangean multiplier approach in the characterization of the optimal side-contracting problem to be considered later.

[17]In order to avoid technical complications, we assume that M_A and M_S are compact subsets of finite dimensional Euclidean spaces, and $(\bar{X}_A, \bar{X}_S, \bar{q})$ are continuous for each of μ_A and μ_S and $(\hat{X}_A, \hat{X}_S, \hat{q})$ for each of μ. These assumptions enable us to apply the minimax theorem (Nikaido (1954)) and guarantee the existence of an optimal side contract.

power as a parameter, and represent it by relative welfare weights on the ex ante payoffs of S and A at the time that the side contract is chosen by the coalition in response to the contract GC offered by P. Formally, the side contract is selected at the ex ante stage by a fictional (uninformed) third party acting as a mediator, who maximizes ex ante expected value of $\alpha u_A + (1-\alpha) u_S$ ($\alpha \in [0,1]$) in response to choice of GC made by P. The third party does not play any budget breaking role, hence the coalition needs to satisfy a budget balance condition $t_A + t_S \leq 0$.[18] No side payments can be exchanged at the *ex ante* stage; they can only be exchanged at the *ex post* stage after payments from P have been received. The side contract cannot be renegotiated at the interim or *ex post* stage, while either party can decide to withdraw from the agreement at the interim stage. The side contract allows A and S to exchange messages privately among one another, which determine a side payment and joint set of messages they respectively send to P. Since message spaces include exit as well as type reports, collusion takes the *ex ante* form studied in Mookherjee et al. (2020).

The stages of the game are as follows. Figure 1 also illustrates the timeline of the game. Following the choice of GC by P, at stage 1 (the ex ante stage), the third party offers a side contract SC to S and A. A null side contract (NSC) could also be offered.

Next at stage 2 (the interim stage) S observes η and A observes (θ, η). If a NSC was offered, they play the GC noncooperatively based on their prior beliefs, just as in a game without any collusion. If a non-null side contract was offered, S and A independently decide whether to accept it. Specifically, the game proceeds as follows. $i = A, S$ selects a message $d_i \in D_i$ ($i = A, S$) where D_i is i's message set specified in the side-contract. A and S observe the other's message each other. D_i includes i's exit option $\hat{e}_i$ from the side-contract. If $d_A \neq \hat{e}_A$ and $d_S \neq \hat{e}_S$, their reports to P are selected according to $\mu(d_A, d_S) \in \Delta(M)$, and side payments to A and S are

[18] It is possible that the third party receives the positive payment, making $t_A + t_S < 0$. But we argue later that it does not arise on the equilibrium path.

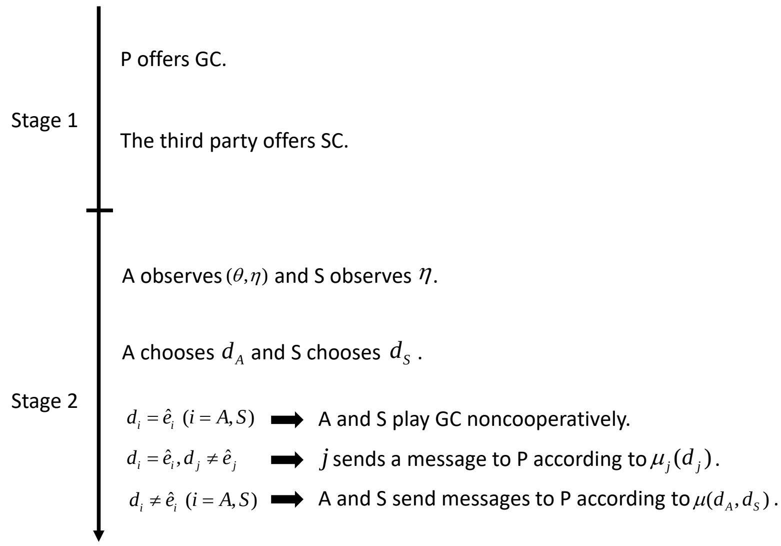

Figure 1: The Timeline of the Game

determined according to functions $t_A(d_A, d_S)$ and $t_S(d_A, d_S)$ respectively.

If $d_A = \hat{e}_A$ and $d_S = \hat{e}_S$, A and S play GC non-cooperatively. What happens when one accepts and the other does not? Then SC specifies a reporting strategy of the party that accepted it, which can be interpreted as a threat that party commits to. The party that rejected it then plays a best response to this threat. Hence if $d_i \neq \hat{e}_i$ and $d_j = \hat{e}_j$ $(i, j = A, S)$, i's message to P is selected according to $\mu_i(d_i) \in \Delta(M_i)$, and the side payment to i is $t_i(d_i)$.[19] On the other hand, j plays GC without any constraint imposed by the side contract, and without any side transfer.[20]

We focus on Perfect Bayesian Equilibrium (PBE) of this extortionary collusion game induced by the grand contract GC and bargaining weight parameter α.[21] How-

[19] Owing to the condition that the coalition cannot make a deficit, $t_i(d_i) \leq 0$.

[20] When collusion is weak instead, the side contract ceases to apply for the subsequent messages for either player when one of them exits — S and A play GC noncooperatively.

[21] For definition of PBE, see Fudenberg and Tirole (1991).

ever, there may be multiple PBE in a given game. We assume collusion permits parties to coordinate the choice of a PBE, hence the third party can specify a selected PBE to maximize the welfare-weighted sum of ex ante payoffs of S and A in the event of multiple PBE. The resulting equilibrium concept is denoted by PBE(c). In case there are two PBE(c) where the third party receives the same payoff, we assume that P can select the more desirable one.

In the economic environment, an allocation evaluated in the expected value conditional on (θ, η) (simply called an allocation in later part) is denoted by

$$(X, u_A, u_S, q) = \{(X(\theta, \eta), u_A(\theta, \eta), u_S(\theta, \eta), q(\theta, \eta)) \mid (\theta, \eta) \in K\}$$

where $X \equiv X_A + X_S$, which is the total payment from P to the coalition. The budget balance condition implies $X(\theta, \eta) \geq u_A(\theta, \eta) + u_S(\theta, \eta) + \theta q(\theta, \eta)$ for any $(\theta, \eta) \in K$.

Achievable allocations in extortionary collusion can now be defined:

Definition 1 *An allocation (X, u_A, u_S, q) is achievable in extortinary collusion with α if it is realized in PBE(c) under α for some $GC \in \mathcal{GC}$.*

5 Extortionary Collusion-Proof (ECP) Allocations

5.1 Collusion-Proof Principle

We now explain the notion of an extortionary-collusion-proof (ECP) allocation, and establish a version of the Collusion Proof Principle where, given any arbitrary mechanism chosen by P and any allocation resulting from a PBE(c) in that mechanism, there exists a (suitably augmented) revelation mechanism which does not leave scope for extortion or collusion.

We start with defining some notations that are necessary to describe S and A's outside payoffs in side-contract design problem. For GC, by applying the minimax theorem, there exists $(\underline{\mu}_A, \bar{\mu}_S)$ which satisfies

$$\begin{aligned} \underline{w}_S(GC) \equiv \bar{X}_S(\underline{\mu}_A, \bar{\mu}_S) &= \min_{\mu_A \in \Delta(M_A)} \max_{\mu_S \in \Delta(M_S)} \bar{X}_S(\mu_A, \mu_S) \\ &= \max_{\mu_S \in \Delta(M_S)} \min_{\mu_A \in \Delta(M_A)} \bar{X}_S(\mu_A, \mu_S). \end{aligned}$$

where $\underline{\mu}_A$ is A's minmax strategy, and $\bar{\mu}_S$ is S's maxmin strategy. $\underline{w}_S(GC)$ is the minmax value of the S's payoff which can be at least secured for any message choice of A in GC. Since S always has the option to exit from the grand contract, $\underline{w}_S(GC) \geq 0$ for any GC. Given grand contract GC, for a report of S: $\mu_S \in \Delta(M_S)$ and a type of A: $\theta \in \Theta$, we also define the maximum payoff that A can attain in response:

$$\hat{u}_A(\theta, \mu_S, GC) \equiv \max_{\mu_A \in \Delta(M_A)} \bar{X}_A(\mu_A, \mu_S) - \theta \bar{q}(\mu_A, \mu_S)$$

and $\mu_A(\theta, \mu_S, GC)$ as a maximizer of the above problem.

Given $GC \in \mathcal{GC}$ and $\alpha \in [0, 1]$, let (X, u_A, u_S, q) be an allocation achieved in extortionary collusion. Evidently, there exists $\mu(\theta, \eta) \in \Delta(M)$ such that $q(\theta, \eta) = \hat{q}(\mu(\theta, \eta))$ and $X(\theta, \eta) = \hat{X}_A(\mu(\theta, \eta)) + \hat{X}_S(\mu(\theta, \eta))$. Here we define the following problem $P^E(\alpha : \eta, GC)$ with

$$(\tilde{u}_A(\theta, \eta), \tilde{u}_S(\theta, \eta), \tilde{\mu}(\theta, \eta), \tilde{\mu}_S(\eta))$$

as the set of control variables, chosen to maximize the objective function

$$E[\alpha \tilde{u}_A(\theta, \eta) + (1 - \alpha)\tilde{u}_S(\theta, \eta) \mid \eta]$$

subject to the constraint that for all $\theta \in \Theta$:

(i) $\tilde{u}_A(\theta, \eta) \in \Re, \tilde{u}_S(\theta, \eta) \in \Re, \tilde{\mu}(\theta, \eta) \in \Delta(M_A \times M_S), \tilde{\mu}_S(\eta) \in \Delta(M_S)$

(ii) $\tilde{u}_A(\theta, \eta) \geq \tilde{u}_A(\theta', \eta) + (\theta' - \theta)\hat{q}(\tilde{\mu}(\theta', \eta))$ for any $\theta' \in \Theta$

(iii) $\tilde{u}_A(\theta, \eta) + \tilde{u}_S(\theta, \eta) \leq \hat{X}_A(\tilde{\mu}(\theta, \eta)) + \hat{X}_S(\tilde{\mu}(\theta, \eta)) - \theta \hat{q}(\tilde{\mu}(\theta, \eta))$

(iv) $\tilde{u}_A(\theta, \eta) \geq \hat{u}_A(\theta, \tilde{\mu}_S(\eta), GC)$ and

(v) $E[\tilde{u}_S(\theta,\eta) \mid \eta] \geq \underline{w}_S(GC)$.

It is shown below that the side-contract (SC) design faced by the third party for GC reduces to solving this problem for each state η. As shown in the proof of Lemma 1, even though SC is designed at the ex ante stage when η is not realized, the third party can design SC as if it is tailored for each η, through the A and S's interim-stage reports about η. The third party coordinates a report sent to P that maximizes the expected value of $\alpha u_A + (1-\alpha) u_S$, associated with the choices of threat $\underline{\mu}_A$ for the S's exit from SC, and threat $\tilde{\mu}_S(\eta)$ for the A's exit from SC.[22] The use of these threats determines outside payoffs of A and S in the SC design, reflected respectively to the right-hand sides of participation constraints (iv) and (v). The A's privacy about θ within the coalition requires feasible SC to satisfy A's truthful telling constraint (ii). (iii) represents the budget balance condition within the coalition.

The following lemma shows that the solution of this problem for any given $\eta \in \Pi$ characterizes achievable allocations in extortionary collusion, and that the coalitional budget is strictly balanced.

Lemma 1 *If (X, u_A, u_S, q) is achieved as a PBE(c) outcome in GC, then for any $\eta \in \Pi$:*

$$(\tilde{u}_A(\theta,\eta), \tilde{u}_S(\theta,\eta), \tilde{\mu}(\theta,\eta)) = (u_A(\theta,\eta), u_S(\theta,\eta), \mu(\theta,\eta))$$

solves $P^E(\alpha : \eta, GC)$ for some $\tilde{\mu}_S(\eta) = \mu_S(\eta) \in \Delta(M_S)$, and

$$X(\theta,\eta) - \theta q(\theta,\eta) = u_A(\theta,\eta) + u_S(\theta,\eta).$$

The balanced budget property implies there is no loss of generality in representing achievable allocation by (u_A, u_S, q). The lemma also implies that an achievable allocation satisfies:

(a) (IC_A): truthful telling condition for A

$$u_A(\theta,\eta) \geq u_A(\theta',\eta) + (\theta' - \theta) q(\theta',\eta)$$

[22] The third party can design SC which selects the threat to A contingent on the S's report at the interim stage. It creates the dependency of threat $\tilde{\mu}_S(\eta)$ on η.

for any $\theta, \theta' \in \Theta$ and any $\eta \in \Pi$;

(b) (PC_A): interim participation constraint for A

$$u_A(\theta, \eta) \geq 0$$

for any $(\theta, \eta) \in K$; and

(c) (PC_S): interim participation constraint for S

$$E[u_S(\theta, \eta) \mid \eta] \geq 0$$

for any $\eta \in \Pi$.

We say that (u_A, u_S, q) satisfies *individual incentive compatibility (IIC)* if and only if it satisfies IC_A, PC_A and PC_S.

Now we argue that the Collusion-proof principle holds in the context of extortionary collusion. This principle requires every achievable allocation to be achieved in a collusion-proof way by the revelation mechanism which corresponds to this allocation. Given an allocation (u_A, u_S, q) achieved for some GC, consider the *standard revelation mechanism* $GC^R = (X_A^R(m_A, m_S), X_S^R(m_A, m_S), q^R(m_A, m_S)) \in \mathcal{GC}$ with $M_A^R = \{e_A\} \cup \Pi \times \Theta$ and $M_S^R = \{e_S\} \cup \Pi$ defined as follows:

$$(X_A^R, X_S^R, q^R) = (u_A(\theta_A, \eta_S) + \theta_A q(\theta_A, \eta_S), u_S(\theta_A, \eta_A) - L(\eta_A, \eta_S), q(\theta_A, \eta_S))$$

for messages $((\theta_A, \eta_A), \eta_S) \in M_A^R \times M_S^R$, while $(X_A^R, X_S^R, q^R) = (0, 0, 0)$ if either $m_A = e_A$ or $m_S = e_S$, where $L(\eta_A, \eta_S) = 0$ when $\eta_A = \eta_S$ and a large positive number for $\eta_A \neq \eta_S$. Figure 2(a) outlines this revelation mechanism GC^R where elements in the matrix denote (X_A^R, X_S^R, q^R) for different combinations of S and A's messages.

Suppose that GC^R is played by A and S non-cooperatively. Truthful reporting of η is guaranteed by the property of GC^R that cross-checks reports of η between A and S, and imposing a penalty L on S when these do not match.[23] IIC ensures that A does

[23]Since A's payoff does not depend on η_A, A does not have an incentive to misreport η. And given truthful reporting of η by A, S does not have an incentive to misreport owing to the penalty imposed on S for inconsistent reports.

(a)

$S \backslash A$	e_A	$(\theta_A, \eta_A) \in K$
e_S	$(0,0,0)$	$(0,0,0)$
$\eta_S \in \Pi$	$(0,0,0)$	$(u_A(\theta_A,\eta_S)+\theta_A q(\theta_A,\eta_S), u_S(\theta_A,\eta_A)-L(\eta_A,\eta_S), q(\theta_A,\eta_S))$

(b)

$S \backslash A$	e_A	$(\theta_A, \eta_A) \in K$
e_S	$(0,0,0)$	$(0,0,0)$
η_0	$(0,\omega,0)$	$(0,\omega,0)$
$\eta_S \in \Pi$	$(0,0,0)$	$(u_A(\theta_A,\eta_S)+\theta_A q(\theta_A,\eta_S), u_S(\theta_A,\eta_A)-L(\eta_A,\eta_S), q(\theta_A,\eta_S))$

(c)

$S \backslash A$	e_A	$(\theta_A, \eta_A) \in K$
e_S	$(0,0,0)$	$(\bar{X}_A(\mu_A(\theta_A,\bar{\mu}_S,GC),\bar{\mu}_S), 0, \bar{q}(\mu_A(\theta_A,\bar{\mu}_S,GC),\bar{\mu}_S))$
η_0	$(0,\omega,0)$	$(\bar{X}_A(\mu_A(\theta_A,\bar{\mu}_S,GC),\bar{\mu}_S), \omega, \bar{q}(\mu_A(\theta_A,\bar{\mu}_S,GC),\bar{\mu}_S))$
$\eta_S \in \Pi$	$(0,0,0)$	$(u_A(\theta_A,\eta_S)+\theta_A q(\theta_A,\eta_S), u_S(\theta_A,\eta_A)-L(\eta_A,\eta_S), q(\theta_A,\eta_S))$

Figure 2: Revelation Mechanism GC^R and Its Modification

not benefit from misreporting θ, and both parties have an incentive to participate in the mechanism. Thus, allocation (u_A, u_S, q) is achieved on the continuation game following a null side-contract between S and A.

Next, collusion proofness requires the third party side-contract designer to not want to deviate to a non-null side contract. This requires the scope for collusion and extortion in GC^R should not be larger when compared to the original GC.

Note first that the coalition would never benefit from reporting $\eta_A \neq \eta_S$, since it causes a large loss $L(\eta_A, \eta_S)$ to the sum of the transfers received by A and S from P. Restricting $\eta_A = \eta_S$, for any message choice in GC^R there exists a message in GC which induces the same allocation. Thus the change from GC to GC^R does not expand the scope for collusion.

Second, the following two conditions (A) and (B) ensure that the scope for extortion does not expand:

(A) $\underline{w}_S(GC^R) \geq \underline{w}_S(GC)$

(B) For any $\mu_S \in \Delta(M_S^R)$, there exists $\mu_S^{'} \in \Delta(M_S)$ such that $\hat{u}_A(\theta, \mu_S, GC^R) \geq \hat{u}_A(\theta, \mu_S^{'}, GC)$ for all $\theta \in \Theta$.

(A) and (B) imply that the third party does not have access to a threat strategy which relaxes A and S's participation constraints (iv) and (v) in GC^R as compared to GC. Hereafter, let $\underline{w}_S(GC)$ be denoted by ω to simplify the exposition.

In order to ensure (A) and (B), we augment message spaces and modify GC^R as follows.

Start with (A). In the original GC, S's maximin strategy $\bar{\mu}_S \in \Delta(M_S)$ option helped secure a minimum payoff ω. This option may not actually be used on the equilibrium path under GC. Hence it may not be available in the revelation mechanism GC^R. This creates the scope for stronger punishments that can be imposed on S in GC^R.[24] In order to protect S from the possibility of more severe threats, we augment S's message space with an auxiliary message η_0 option, which is a counterpart of the minmax strategy $\bar{\mu}_S$ in GC. Defining $\bar{\Pi} \equiv \Pi \cup \{\eta_0\}$, S's message space is thus modified to $M_S^R = \{e_S\} \cup \bar{\Pi}$. For the moment, let us select (X_A^R, X_S^R, q^R) equal to $(0, \omega, 0)$ for any (m_A, m_S) such that $m_S = \eta_0$. With this modified GC^R, the minmax value of S's payoff is equal to ω, ensuring condition (A). This modified GC^R is illustrated in Figure 2(b).

Next we turn to (B). Observe that

$$(X_A^R(m_A, e_S), q^R(m_A, e_S)) = (X_A^R(m_A, \eta_0), q^R(m_A, \eta_0)) = (0, 0)$$

in this modified GC^R (illustrated in Figure 2(b)). It means that the choice of either e_S or η_0 by S pushes A's outside payoff down to zero. This may cause (B) to be violated, unless $\hat{u}_A(\theta, \mu_S^{'}, GC) = 0$ for some $\mu_S^{'} \in \Delta(M_S)$ in the original GC. It necessitates an additional modification of GC^R in order to protect A from the

[24]This problem may be aggravated by the large punishments imposed on S in GC^R whenever reports η are inconsistent.

use of such a threat in SC. Hence we modify $(X_A^R((\theta_A, \eta_A), e_S), q^R((\theta_A, \eta_A), e_S))$ and $(X_A^R((\theta_A, \eta_A), \eta_0), q^R((\theta_A, \eta_A), \eta_0))$ from $(0, 0)$ to

$$(\bar{X}_A(\mu_A(\theta_A, \bar{\mu}_S, GC), \bar{\mu}_S), \bar{q}(\mu_A(\theta_A, \bar{\mu}_S, GC), \bar{\mu}_S))$$

where it may be recalled $\mu_A(\theta_A, \bar{\mu}_S, GC)$ denotes the best response of type θ of A in GC. In words, A receives a transfer and output assignment that would have resulted in the original GC if S were to play his maxmin strategy $\bar{\mu}_S$ and A were to select a best response. The resulting mechanism is illustrated in Figure 2(c).

We can check that GC^R with these two modifications satisfies all the required conditions for the Collusion-proofness principle to hold. It is evident that these modifications do not affect the non-cooperative equilibrium of the mechanism, since Lemma 1 implies $E[u_S(\theta, \eta) \mid \eta] \geq \omega$ and S does not benefit from selecting η_0 when A reports truthfully. Moreover, since the minimax theorem implies

$$\bar{X}_S(\mu_A(\theta, \bar{\mu}_S, GC), \bar{\mu}_S) \geq \bar{X}_S(\underline{\mu}_A, \bar{\mu}_S) = \omega \geq 0$$

for any θ, the coalition's report of $(m_A, m_S) = ((\theta, \eta), e_S)$ or $((\theta, \eta), \eta_0)$ in this modified GC^R does not result in a higher total payment $\hat{X}_A^R + \hat{X}_S^R$ to the coalition, as compared to the report of $(\mu_A(\theta, \bar{\mu}_S, GC), \bar{\mu}_S)$ in the original GC. It means that this modification of the mechanism does not expand the scope for collusion, as compared to the original GC.

Finally we show that condition (B) holds. Since the use of e_S and η_0 as a threat in GC^R has the same impact on A's payoff, it suffices to consider what can be achieved in terms of lowering A's participation constraint in side contracting by supplementing S's message options by η_0 alone. Let any such threat be represented by a probability function $P(\cdot)$ defined over $\bar{\Pi}$, i.e., where $P(\eta)$ is the probability that S reports $\eta \in \bar{\Pi}$.

Now observe that conditional on S reporting $\eta_S = \eta \in \Pi$, type θ of A ends up with the same payoff that she would have attained in the original GC if S had reported according to $\mu_S(\eta)$ and she played a best response. And conditional on S reporting $\eta_S = \eta_0$, she would end up with a payoff that she would have attained in GC if S had reported $\bar{\mu}_S$ and she chose a best response. In particular, reporting truthfully is a dominant strategy for A in GC^R given any threat $P(.)$. It follows that the expected payoff that A attains in GC^R when S executes his threat represented by $P(.)$, is at least as large as what she would have attained in the original GC if S reported according to a mixed strategy which corresponds to the following two stage lottery: (1) at the first stage η_S is chosen according to $P(\cdot)$ over $\bar{\Pi}$, and then (2) at the second stage S reports according to $\mu_S(\eta)$ if at the first stage $\eta_S = \eta \in \Pi$, and according to $\bar{\mu}_S$ if instead $\eta_S = \eta_0$. Hence (B) holds, establishing the validity of the Collusion-proofness principle.

In the preceding argument, the mechanism augmented S's message space to include η_0 in order to protect S from additional scope of extortion relative to the original GC. No such augmentation was needed for A's message space, owing to the absence of any penalty on A for inconsistencies between A and S's reports in the underlying revelation mechanism. As explained above, A can protect herself from the possibility of any additional extortion, by reporting truthfully in response to any threat $P(.)$.

The argument also implies that for any allocation achieved for some GC, there exists a revelation mechanism where S's message space is augmented as described above, which achieves the same allocation in a collusion-proof manner. By applying Lemma 1, an allocation is achievable in extortionary collusion if and only if there exists an augmented revelation mechanism GC^R which induces the same allocation in the solution of problem $P^E(\alpha : \eta, GC^R)$. Therefore we can characterize achievable allocations using necessary and sufficient conditions for the existence of such an augmented mechanism. This allows us to define extortionary collusion-proof allocation as follows.

Definition 2 *Allocation* (u_A, u_S, q) *is extortionary collusion-proof (or ECP) for* $\alpha \in [0, 1]$*, if* (u_A, u_S, q) *is IIC, and there exists* $\omega \geq 0$*, and an augmentation of* (u_A, u_S, q) *on* $\bar{K} \equiv \Theta \times \bar{\Pi}$ *with* $u_S(\theta, \eta_0) = \omega$ *for any* $\theta \in \Theta$ *and* $(u_A(\theta, \eta_0), q(\theta, \eta_0))$ *satisfying*

(IC_A) and (PC_A), such that for any $\eta \in \Pi$,

$$(\mu(\theta,\eta), \tilde{u}_A(\theta,\eta), \tilde{u}_S(\theta,\eta), P(\cdot \mid \eta)) = (I_1(\theta,\eta), u_A(\theta,\eta), u_S(\theta,\eta), I_2(\eta))$$

solves problem $P^E(\alpha : \eta)$*:*

$$\max E[\alpha \tilde{u}_A(\theta,\eta) + (1-\alpha)\tilde{u}_S(\theta,\eta) \mid \eta]$$

subject to $(\mu(\theta,\eta), \tilde{u}_A(\theta,\eta), \tilde{u}_S(\theta,\eta), P(\cdot \mid \eta))$ *satisfies for all* $\theta \in \Theta$*:*

(i) $\mu(\theta,\eta) \in \Delta(\bar{K} \cup \{e\}), \tilde{u}_A(\theta,\eta) \in \Re, \tilde{u}_S(\theta,\eta) \in \Re, P(.|\eta) \in \Delta(\bar{\Pi})$

(ii) $\tilde{u}_A(\theta,\eta) \geq \tilde{u}_A(\theta^{'},\eta) + (\theta^{'} - \theta)q(\mu(\theta^{'},\eta))$ *for any* $\theta^{'} \in \Theta$

(iii) $\tilde{u}_A(\theta,\eta) + \tilde{u}_S(\theta,\eta) \leq X(\mu(\theta,\eta)) - \theta q(\mu(\theta,\eta))$

(iv) $\tilde{u}_A(\theta,\eta) \geq \Sigma_{\eta^{'} \in \bar{\Pi}} P(\eta^{'} \mid \eta) u_A(\theta,\eta^{'})$

(v) $E[\tilde{u}_S(.,\eta) \mid \eta] \geq \omega$,

where $X(\theta,\eta) \equiv u_A(\theta,\eta) + u_S(\theta,\eta) + \theta q(\theta,\eta)$ *for* $(\theta,\eta) \in \bar{K}$ *and* $(X(e), q(e)) \equiv (0,0)$*, and* $I_1(\theta,\eta)$ *denotes a probability measure on* $\bar{K} \cup \{e\}$ *where* (θ,η) *is selected with probability one, while* $I_2(\eta)$ *denotes a probability function on* $\bar{\Pi}$ *where* η *is selected with probability one (i.e.,* $P(\eta \mid \eta) = 1$ *and* $P(\eta^{'} \mid \eta) = 0$ *for any* $\eta \neq \eta^{'}$*).*

Note that an ECP allocation is described by properties that do not depend on the original GC. Moreover, for any ECP allocation satisfying Definition 2, we can easily construct a (modified) revelation mechanism GC^R as described above by adding the cross-checking scheme about reports of η by A and S, and then augmenting S's message space. This leads us to the main result of this section.

Proposition 1 *An allocation* (u_A, u_S, q) *is achievable in extortionary collusion with* α *if and only if it is extortionary collusion-proof (ECP) for* α*.*

It is useful to contrast this characterization of ECP allocations with allocations achievable with weak collusion (see e.g., Mookherjee et al. (2020)) where extortion is absent. In the weak collusion setting, failure of A and S to agree on a non-null side contract always triggers noncooperative play of GC. Let A and S's noncooperative payoffs be denoted respectively by $u_A^N(\cdot, GC)$ and $u_S^N(\cdot, GC)$. These serve as outside options in the bargaining over the side-contract. Hence the characterization of allocations achievable in weak collusion differs with respect to the right-hand sides of constraints (iv) and (v) in problem $P^E(\alpha : \eta, GC)$: these are replaced respectively by $u_A^N(\theta, \eta, GC)$ and $E[u_S^N(\theta, \eta, GC) \mid \eta]$. Moreover, there is no need to augment message spaces in weak collusion: owing to the absence of concerns for extortion, there is no need to 'protect' S from the possibility of additional extortion in the constructed mechanism GC^R relative to GC.

5.2 Effects of Varying Bargaining Power

As explained in the Introduction, earlier studies have established that a set of achievable allocation under weak collusion does not depend on α. The second main result of this paper is that the allocation of bargaining power does matter in extortionary collusion. In particular, there is a sharp dichotomy between the set of ECP allocations that can be achieved between cases where S has greater ($\alpha < 1/2$) and lower ($\alpha \geq 1/2$) bargaining power than A.

5.2.1 When S has Less Bargaining Power than A

We begin with the case where S has lower bargaining power. The following lemma shows that the set of ECP allocations reduces to an extremely simple and restricted class in this case.

Lemma 2 *Suppose $\alpha \geq 1/2$. For each ECP allocation (u_A, u_S, q), there exist real-valued functions $(X_A(\theta), Q(\theta))$ defined on Θ, and a nonnegative constant c and such that, for any (θ, η):*

(i) $u_S(\theta, \eta) = c$

(ii) $u_A(\theta, \eta) = X_A(\theta) - \theta Q(\theta) = \max_{\theta' \in \Theta}[X_A(\theta') - \theta Q(\theta')] \geq 0$

(iii) $q(\theta, \eta) = Q(\theta)$.

This lemma leads straightforwardly to the following proposition, since any allocation satisfying (i)-(iii) with $c = 0$ is can be achieved in a setting where S is not hired by P at all (referred to as NS).

Proposition 2 *If $\alpha \geq 1/2$, P's payoff in ECP allocation cannot be greater than in NS where S is not hired.*

When A has at least as much bargaining power as S, it is optimal for the coalition to pin S down to her constant payoff and provide all residual rents to A. Reports by the coalition are then chosen to maximize A's payoffs (i.e, S's role is restricted to rubber-stamping whatever report is in A's interest). Evidently, P cannot derive any benefit from appointing S.

The intuition underlying this result is that when A has at least as much bargaining power as S, the optimal side contract is chosen to maximize the payoff of A, subject to S's participation constraint — i.e., as if A is the 'sub-principal' within the (A-S) coalition. This side contract problem is not subject to any information friction, since A is better informed than S. The only constraint that effectively matters is S's participation constraint, i.e., that S must be assured a payoff of at least ω. The optimal side contract from A's perspective is then one in which S is provided a lumpsum payment of ω, all the residual rents accrue to A, and a joint reporting strategy is selected which maximizes the latter. The same is true when $\alpha \in (\frac{1}{2}, 1)$, since a transfer of surplus from S to A still raises the objective of the side-contract designer. Hence it is optimal for the side contract to be designed so that the reporting strategy maximizes the sum of A and S's payoffs, i.e., the coalition will achieve efficient collusion:

$$\mu^*(\theta) \in \arg\max_{\mu \in \Delta(\bar{K} \cup \{e\})}[X(\mu) - \theta q(\mu)] \quad (1)$$

This reporting strategy is evidently independent of the actual realization of η, implying that S merely acts as a 'rubberstamp' and submits the same audit report irrespective of his true signal which best serves A's interest. Hence the presence of S adds no value to the principal P.

The detailed argument (presented in the Appendix) also needs to check that efficient collusion is feasible in the side contracting problem. Here we provide an outline of the reasoning. Feasibility requires the existence of threats which ensure that S and A's participation constraints are satisfied in the side contracting problem when efficient collusion is chosen. Consider any general grand contract GC, and let $\mu^{GC}(\theta)$ denote the reporting strategy μ that maximizes $\hat{X}_A(\mu) + \hat{X}_S(\mu) - \theta\hat{q}(\mu)$ subject to $\mu \in \Delta(M_A \times M_S)$. Then observe that using threats $\bar{\mu}_S$ for A and $\underline{\mu}_A$ for S (corresponding to S threatening to sending message η_0 if A were to reject the side-contract), the sum of their respective payoffs is at least as large as the sum of their corresponding outside options:

$$\begin{aligned}
& \hat{X}_A(\mu^{GC}(\theta)) + \hat{X}_S(\mu^{GC}(\theta)) - \theta\hat{q}(\mu^{GC}(\theta)) \\
\geq \; & \hat{X}_A(\mu_A(\theta, \bar{\mu}_S, GC), \bar{\mu}_S) + \hat{X}_S(\mu_A(\theta, \bar{\mu}_S, GC), \bar{\mu}_S) - \theta\hat{q}(\mu_A(\theta, \bar{\mu}_S, GC), \bar{\mu}_S) \\
\geq \; & \hat{X}_A(\mu_A(\theta, \bar{\mu}_S, GC), \bar{\mu}_S) + \hat{X}_S(\underline{\mu}_A, \bar{\mu}_S) - \theta\hat{q}(\mu_A(\theta, \bar{\mu}_S, GC), \bar{\mu}_S) \\
= \; & \hat{u}_A(\theta, \bar{\mu}_S, GC) + \underline{w}_S(GC)
\end{aligned}$$

The first inequality follows from the definition of $\mu^{GC}(\theta)$, while the second inequality is the result of the minimax theorem. Hence with appropriate lumpsum transfers, efficient collusion is attainable without violating S and A's participation conditions, associated with the specified threats.

5.2.2 When S has Greater Bargaining Power

Now consider the opposite extreme, where S has all the bargaining power ($\alpha = 0$), i.e., S becomes the 'sub-principal' within the coalition. As S is less informed than A, the side contracting problem is subject to asymmetric information frictions. As a consequence, the coalition does not select an allocation with efficient collusion (even though it continues to be feasible). The resulting weakening of collusion would then be likely to benefit P.

One might conjecture that a similar result would obtain when α is positive but strictly lower than a half. However, it is not easy to verify this conjecture, given the current level of generality of the model. In the next section we show this is true in a specific version of the model where the good to be supplied is indivisible, S has access to a signal with two possible realizations, and the signal satisfies some monotone likelihood ratio properties. In Appendix B, we show a similar result holds when the good is perfectly divisible.

One result which nevertheless can be established quite generally provides an indication of how optimal side contracts can vary as the bargaining power allocation shifts in favor of S. Recall that when A has more bargaining power, side contracts achieve efficient collusion and S's participation constraint is binding, with all residual rents accruing to A. We now show that when S has greater bargaining power, S's participation constraint will not bind and can therefore be dropped in the formulation of the side contracting problem.

Lemma 3 *If $\alpha \in [0, 1/2)$, ECP allocation (u_A, u_S, q) solves the relaxed version of $P^E(\alpha : \eta)$ where S's participation constraint $E[\tilde{u}_S(\theta, \eta) \mid \eta] \geq \omega$ is dropped.*

The reasoning is as follows. If this lemma were false, the solution to the relaxed version of problem $P^E(\alpha : \eta)$ must violate S's participation constraint, implying that S ends up with an expected payoff below his minmax payoff ω. The coalition then has the option of switching to the side-contract (used in the proof of Lemma 2) in which S receives a constant payoff of ω and A receives the rest of the aggregate coalitional rent. This side-contract induces ex post efficient reporting strategies, thereby (weakly) expanding the aggregate rent in every state. Given $\alpha < \frac{1}{2}$, A must also benefit from deviating to this side-contract. But this leads to a contradiction, as S and A both benefit from this deviation.

This result implies that if $\alpha < 1/2$, the auxiliary message η_0 in GC^R introduced to protect S from A's minmax reporting strategy is not needed anymore. Thus for $\alpha < 1/2$, there is no loss of generality in defining ECP allocation for modified problem $P^E(\alpha : \eta)$ where ω is set to zero and constraint (v) is dropped.[25] Even with this result, the collusion formation may be still constrained by A's participation constraint (iv).

6 Procuring an Indivisible Good

In this section we restrict attention to the case where the output and the signal are both binary, and S's signal satisfies standard monotone likelihood ratio properties. The quantity to be procured is $q \in Q \equiv \{0, 1\}$. P obtains a zero gross benefit if $q = 0$, and a benefit of $V > 0$ otherwise. There are two possible signal realizations $\eta_i, i = L, H$, and $a_i(\theta) \equiv a(\eta_i \mid \theta) > 0$ on Θ for $i \in \{L, H\}$, and $a_H(\theta)$ is increasing in θ. Thus H (or L) is more likely to occur under high (or low) θ. $F_i(\theta) \equiv F(\theta \mid \eta_i)$ now denotes the distribution of θ conditional on signal realization i, which has a positive density $f_i(\theta) \equiv f(\theta \mid \eta_i)$. $\kappa_i \equiv p(\eta_i) = \int_{\underline{\theta}}^{\bar{\theta}} f(\theta) a_i(\theta) d\theta \in (0, 1)$ denotes the probability of signal i. Our assumptions imply $F_L(\theta) > F(\theta) > F_H(\theta)$ and $F_L(\theta)/f_L(\theta) > F(\theta)/f(\theta) > F_H(\theta)/f_H(\theta)$ on $(\underline{\theta}, \bar{\theta})$. In addition, we assume that (i) $h_i(\theta) = \theta + \frac{F_i(\theta)}{f_i(\theta)}$ and $l_i(\theta) = \theta + \frac{F_i(\theta)-1}{f_i(\theta)}$ are increasing in θ for each $i = L, H$, and (ii) $V \in (\underline{\theta}, \bar{\theta} + \frac{1}{f(\bar{\theta})})$. We also assume $\theta + \frac{F(\theta)}{f(\theta)}$ is increasing in θ.

Then, in the *No Supervision (NS)* case, P offers a non-contingent price p^{NS} to maximize $F(p)[V - p]$. Assumption (ii) above guarantees an interior solution $p^{NS} \in (\underline{\theta}, \bar{\theta})$ in the optimal NS contract. Let $W^{NS} \equiv F(p^{NS})[V - p^{NS}]$ denote the resulting expected payoff of P, which is positive. The *second-best allocation* results when P can costlessly access S's signal, removing any scope for collusion between A and S. Here for each $i \in \{L, H\}$, P offers A a price p_i^{SB} which maximizes $(V - p_i)F_i(p_i)$ subject

[25]Note that $(u_A(\theta, \eta_0), q(\theta, \eta_0))$ also plays the role of the A's outside option in the case that S's exit e_S is used as a threat, as depicted in Figure 2(c). Therefore η_0 in this modified problem $P^E(\alpha : \eta)$ needs to be interpreted as S's exit e_S.

to $p_i \in \Theta$. Then (ii) implies $p_H^{SB} > p^{NS} > p_L^{SB}$.

With binary output $Q = \{0, 1\}$, an allocation with an expected output $q(\theta, \eta) \in (0, 1)$ following any message combination is achievable with a stochastic grand contract. However in this section we restrict attention to deterministic mechanisms, i.e., where the output assignment in the direct revelation mechanism corresponding to the GC is deterministic following any message combination. Our main purpose is to show that hiring S can be valuable when $\alpha < \frac{1}{2}$, and it suffices to show this while using deterministic contracts. We also analyze optimal deterministic contracts. While it would have been desirable to characterize optimal stochastic contracts, we have found this quite difficult, and leave it as an open question. Moreover, the restriction to deterministic contracts could be justified if it is difficult for P to commit to randomized contracts, e.g., if the underlying lottery chosen is unverifiable.

6.1 ECP Allocation

We now provide a detailed characterization of ECP allocations in this setting. Given Proposition 2, we confine attention to the case $\alpha < 1/2$. Consider any ECP allocation (u_A, u_S, q) with deterministic output $q(\theta, \eta) \in \{0, 1\}$ for any $(\theta, \eta) \in K$. Since this allocation satisfies IIC, there exists $p_i \in \Theta$ such that $q(\theta, \eta_i) = 1$ on $[\underline{\theta}, p_i)$ and 0 on $(p_i, \bar{\theta}]$ for any $i \in \{L, H\}$, resulting in $u_A(\theta, \eta_i) = \max\{p_i - \theta, 0\} + u_i$ with $u_i \geq 0$. Following Proposition 1, we augment the mechanism by providing S an additional message option η_0, with $(u_A(\theta, \eta_0), q(\theta, \eta_0))$ satisfying (IC_A) and (PC_A) and $q(\theta, \eta_0) \in \{0, 1\}$, and $u_S(\theta, \eta_0) = 0$. This implies the existence of $p_0 \in \Theta$ and $u_0 \geq 0$ such that $u_A(\theta, \eta_0) = \max\{p_0 - \theta, 0\} + u_0$.

For an augmented allocation (u_A, u_S, q) on $\bar{K}$, consider problem $P^E(\alpha : \eta)$ for $\eta \in \{\eta_L, \eta_H\}$. By virtue of Lemma 3, S's participation constraint (v) can be dropped and ω can be set equal to zero. For a given $q \in \{0, 1\}$, the coalition maximizes the total payment $X(m)$ among $m \in \bar{K} \cup \{e\}$ such that $q(m) = q$. Let X_q denote a maximum value of $X(m)$ for $q(m) = q \in \{0, 1\}$. The total payment to the coalition depends only on the output choice. Define $b \equiv X_1 - X_0$, which is interpreted as a bonus to the coalition for the delivery of unit output. $(X(e), q(e)) \equiv (0, 0)$ implies $X_0 \geq 0$. We can then reformulate transfers as follows: X_0 is a lumpsum aggregate transfer paid by P, and b is the additional payment made when the output is delivered. Since $X(\theta, \eta_0) = p_0 + u_0$ if $q(\theta, \eta_0) = 1$ and u_0 if $q(\theta, \eta_0) = 0$, the definition of (X_0, b) implies $X_0 + b \geq p_0 + u_0$ and $X_0 \geq u_0$.

Since $q \in \{0, 1\}$ on $\bar{K} \cup \{e\}$, the coalition's choice of $\mu(\theta, \eta)$ in $P^E(\alpha : \eta)$ assigns a lottery between two outcomes: $(X, q) = (X_0 + b, 1)$ and $(X, q) = (X_0, 0)$. Let the probability of the former outcome be denoted by $\tilde{q}(\theta, \eta) \in [0, 1]$. With $q(\mu(\theta, \eta)) = \tilde{q}(\theta, \eta)$, constraint (ii) in $P^E(\alpha : \eta)$ is equivalent to (a) $\tilde{u}_A(\theta, \eta) = \int_\theta^{\bar{\theta}} \tilde{q}(y, \eta)dy + \tilde{u}_A(\bar{\theta}, \eta)$ and (b) $\tilde{q}(\theta, \eta)$ is non-increasing in θ. Since constraint (iii) is satisfied with the equality in the solution, there is no loss of generality in rewriting it as $\tilde{u}_S(\theta, \eta) + \tilde{u}_A(\theta, \eta) = \tilde{q}(\theta, \eta)(b - \theta) + X_0$.

Define $\beta = (1 - 2\alpha)/(1 - \alpha) \in (0, 1]$. Note that β is decreasing in α on $[0, 1/2)$. Then for $i \in \{L, H\}$, $P^E(\alpha : \eta_i)$ reduces to

$$\max E[\tilde{q}(\theta, \eta_i)(b - \theta) + X_0 - \beta \tilde{u}_A(\theta, \eta_i) \mid \eta_i]$$

subject to $\tilde{q}(\theta, \eta_i) \in [0, 1]$ is non-decreasing in θ,

$$\tilde{u}_A(\theta, \eta_i) \geq \Sigma_{\eta' \in \bar{\Pi}} P(\eta' \mid \eta_i) u_A(\theta, \eta') \tag{2}$$

and

$$\tilde{u}_A(\theta, \eta_i) = \int_\theta^{\bar{\theta}} \tilde{q}(\theta', \eta_i)d\theta' + \tilde{u}_A(\bar{\theta}, \eta_i).$$

ECP requires $\tilde{q}(\theta, \eta_i) = q(\theta, \eta_i)$, $\tilde{u}_A(\bar{\theta}, \eta_i) = u_i$ and $P(\eta' \mid \eta_i) = I(\eta_i)$ to solve this problem.

We employ Lagrangean methods to obtain a detailed characterization of ECP allocations.[26] ECP requires existence of some non-negative and non-decreasing $\Lambda_i(\theta)$

[26]This extends the approach followed in Mookherjee et al. (2020) to characterize weak collusion proof allocations (which in turn are based on methods used by Jullien (2000) to handle contracting problems with type-dependent participation constraints).

on Θ with $\Lambda_i(\underline{\theta}) = 0$ and $\Lambda_i(\bar{\theta}) = 1$ such that $\tilde{q}(\theta, \eta_i) = q(\theta, \eta_i)$, $\tilde{u}_A(\theta, \eta_i) = u_i$ and $P(\cdot \mid \eta_i) = I(\eta_i)$ maximize the following Lagrangean (in which we rescale the multipliers by β for expositional convenience):

$$\begin{aligned}
\mathcal{L} &\equiv \int_{\underline{\theta}}^{\bar{\theta}} [\tilde{q}(\theta, \eta_i) b - \beta \int_{\theta}^{\bar{\theta}} \tilde{q}(\tilde{\theta}, \eta_i) d\tilde{\theta}] dF_i(\theta) + X_0 - \beta \tilde{u}_A(\bar{\theta}, \eta_i) \\
&+ \int_{\underline{\theta}}^{\bar{\theta}} [\int_{\theta}^{\bar{\theta}} \tilde{q}(\tilde{\theta}, \eta_i) d\tilde{\theta} + \tilde{u}_A(\bar{\theta}, \eta_i) - \Sigma_{j \in \{L,H,0\}} P(\eta_j \mid \eta_i)[\max\{p_j - \theta, 0\} + u_j]] d(\beta \Lambda_i(\theta)) \\
&= \int_{\underline{\theta}}^{\bar{\theta}} \tilde{q}(\theta, \eta_i)[b - W_i(\theta : \beta)] dF_i(\theta) + X_0 - \beta \Sigma_{j \in \{L,H,0\}} P(\eta_j \mid \eta_i)[\int_{\underline{\theta}}^{p_j} \Lambda_i(\theta) d\theta + u_j]
\end{aligned}$$

where[27]

$$W_i(\theta : \beta) \equiv \theta + \beta \frac{F_i(\theta) - \Lambda_i(\theta)}{f_i(\theta)}.$$

$\beta \Lambda_i'(\theta) \geq 0$ (whenever it exists) can be interpreted as the Kuhn-Tucker multiplier for the constraint corresponding to type θ. Hence $\Lambda_i(\underline{\theta}) = 0$. Moreover, an exogenous uniform increase in A's outside option payoff (for all θ) by $\Delta > 0$ induces a decrease of the objective function by $\beta\Delta$ in the optimal solution; hence $\Lambda_i(\bar{\theta}) = 1$. The expression $W_i(\theta : \beta)$ represents the modified 'virtual' cost to the coalition of delivering the output in state (θ, η_i): the standard expression for cost of information rents $\beta \frac{F_i(\theta)}{f_i(\theta)}$ (when A has a zero outside option) is reduced to $\beta \frac{F_i(\theta) - \Lambda_i(\theta)}{f_i(\theta)}$ because of the need to provide type θ of A with the required outside option payoff.

Thus ECP implies the following two conditions. First, in each state i:

$$q(\theta, \eta_i) = \arg\max \int_{\underline{\theta}}^{\bar{\theta}} \tilde{q}(\theta, \eta_i)[b - W_i(\theta : \beta)] dF_i(\theta)$$

---

[27] In this derivation, we use integration by parts:

$$\int_{\underline{\theta}}^{\bar{\theta}} \int_{\theta}^{\bar{\theta}} \tilde{q}(\tilde{\theta}, \eta_i) d\tilde{\theta} dF_i(\theta) = \int_{\underline{\theta}}^{\bar{\theta}} \tilde{q}(\theta, \eta_i) F_i(\theta) d\theta$$

and

$$\int_{\underline{\theta}}^{p_j} (p_j - \theta) d\Lambda_i(\theta) = \int_{\underline{\theta}}^{p_j} \Lambda_i(\theta) d\theta.$$

subject to $\tilde{q}(\theta, \eta_i)$ non-decreasing in θ. This condition turns out to be equivalent to

$$p_i = \arg\max_{\tilde{p}_i \in \Theta} \int_{\underline{\theta}}^{\tilde{p}_i} [b - W_i(\theta : \beta)] dF_i(\theta).\text{[28]} \tag{3}$$

Thus paying the price p_i to A for delivering the good to P must maximize the interim expected value to the coalition, equal to the additional payment b received from P, less the modified virtual cost associated with delivery. Moreover if there is no scope for coalitional deviation about price $\tilde{p}_i$, the coalition does not benefit from randomizing reports to P for a given θ.

Second, in order to guarantee the optimality of $P(\cdot \mid \eta_i) = I(\eta_i)$,

$$\int_{\underline{\theta}}^{p_i} \Lambda_i(\theta) d\theta + u_i \leq \int_{\underline{\theta}}^{p_j} \Lambda_i(\theta) d\theta + u_j \tag{4}$$

for each $i \in \{L, H\}$ and for each $j \in \{L, H, 0\}$ $(j \neq i)$.

The above arguments characterizes the set of achievable allocations in this setting, as summarized in the following lemma.

Lemma 4 *Suppose $\alpha < 1/2$. An allocation (u_A, u_S, q) is achievable by an extortionary-collusion-proof deterministic revelation mechanism if and only if there exist*

$$(b, X_0, p_H, p_L, p_0, u_H, u_L, u_0)$$

and $(\Lambda_L(\theta), \Lambda_H(\theta))$ such that

---

[28] It is evident that the former condition implies the latter one. The latter also implies

$$\begin{aligned}
\int_{\underline{\theta}}^{\bar{\theta}} \tilde{q}(\theta, \eta_i)[b - W_i(\theta : \beta)] dF_i(\theta) &= \int_{\underline{\theta}}^{\bar{\theta}} \tilde{q}(\theta, \eta_i) \frac{d \int_{\underline{\theta}}^{\theta} [b - W_i(\theta' : \beta)] dF_i(\theta')}{d\theta} d\theta \\
= \quad & \tilde{q}(\bar{\theta}, \eta_i) \int_{\underline{\theta}}^{\bar{\theta}} [b - W_i(\theta : \beta)] dF_i(\theta) - \int_{\underline{\theta}}^{\bar{\theta}} \int_{\underline{\theta}}^{\theta} [b - W_i(\theta' : \beta)] dF_i(\theta') d\tilde{q}(\theta, \eta_i) \\
\leq \quad & \tilde{q}(\bar{\theta}, \eta_i) \int_{\underline{\theta}}^{p_i} [b - W_i(\theta : \beta)] dF_i(\theta) - \int_{\underline{\theta}}^{\bar{\theta}} \int_{\underline{\theta}}^{p_i} [b - W_i(\theta' : \beta)] dF_i(\theta') d\tilde{q}(\theta, \eta_i) \\
= \quad & \tilde{q}(\underline{\theta}, \eta_i) \int_{\underline{\theta}}^{p_i} [b - W_i(\theta : \beta)] dF_i(\theta) \leq \int_{\underline{\theta}}^{p_i} [b - W_i(\theta : \beta)] dF_i(\theta) \\
= \quad & \int_{\underline{\theta}}^{\bar{\theta}} q(\theta, \eta_i)[b - W_i(\theta : \beta)] dF_i(\theta).
\end{aligned}$$

The inequality uses (3) and the non-decreasing property of $\tilde{q}(\theta, \eta_i)$.

(i) $(u_A(\theta,\eta_i), u_S(\theta,\eta_i), q(\theta,\eta_i)) = (p_i - \theta + u_i, b - p_i + X_0, 1)$ *for* $\theta \le p_i$ *and* $(u_i, X_0, 0)$ *for* $\theta > p_i$.

(ii) $u_i \ge 0$ *and* $p_i \in \Theta$ *for* $i \in \{H, L, 0\}$

(iii) $X_0 \ge u_0$ *and* $X_0 + b \ge p_0 + u_0$

(iv) (3) and (4) are satisfied for $\Lambda_i(\theta)$ *which is non-decreasing in* $\theta \in \Theta$ *with* $\Lambda_i(\underline{\theta}) = 0$ *and* $\Lambda_i(\bar{\theta}) = 1$.

P's payoff equals

$$(\kappa_L F_L(p_L) + \kappa_H F_H(p_H))(V - b) - X_0. \tag{5}$$

Note that in the setting NS where S is not hired, the allocation takes the following form: p^{NS} is chosen to maximize $F(p)(V - p)$, while $b = p_L = p_H = p_0 = p^{NS}$, $X_0 = u_H = u_L = u_0 = 0$ and $\Lambda_i(\theta) = F_i(\theta)$.

6.2 Optimal ECP Allocation

Next we study the allocation (called optimal ECP allocation) which maximizes P's payoff among the set of allocations characterized in Lemma 4. Let us denote the optimal allocation by $(b^*, X_0^*, p_L^*, p_H^*, u_L^*, u_H^*)$, and the corresponding payoff of P by W^*. S is valuable if $W^* > W^{NS}$.

We begin with some useful properties of the optimal allocation.

Lemma 5 *In the optimal allocation,* $p_H^* \ge p_L^*$, $\bar{\theta} > p_L^* > \underline{\theta}$, $p_H^* \ge b^*$ *and* $F_L(p_L^*) > F_H(p_H^*)$. *If S is valuable,* $p_H^* > p_L^*$ *and* $p_H^* > b^*$.

We can show allocations not satisfying these properties are dominated by the optimal NS allocation. Given the distributional assumptions, the 'supply elasticity' is lower in state L, so it is intrinsically desirable for A to be offered p_H in state H which is not lower than p_L. And any allocation where b is larger than p_H is dominated by the NS allocation where A is offered $\hat{p} = p_H$, as it would raise the likelihood of supply in state L, leave it unchanged in state H, while lowering the price paid by P for delivery of the good. In particular, the condition $p_H \geq b$ implies the use of 'countervailing incentives', wherein S and A's interests (with regard to supply) are opposed in state H: S would be worse off while A would be better off if the good is supplied (which would happen whenever $\theta < p_H$).

The problem of selecting the optimal allocation can be broken down into two stages as follows. First take (p_L, p_H, b) as given, and optimize over $(X_0, p_0, u_0, u_L, u_H)$. Then at the second stage, optimize over (p_L, p_H, b).

Consider the first stage problem, given (p_L, p_H, b). Observe that choosing higher values of u_0 and p_0 strengthens A's participation constraint in the optimization problem $P^E(\alpha : \eta)$ faced by the coalition. So it helps to deter coalitional deviations to the S's exit from GC by choosing u_0 and p_0 as high as is allowed by condition (iii) of Lemma 4: i.e., it is optimal to set $p_0 = b, u_0 = X_0$ which maximizes $\int_{\underline{\theta}}^{p_0} \Lambda_i(\theta)d\theta + u_0$ for any $i \in \{L, H\}$. Then (4) reduces to

$$\int_{p_L}^{p_H} \Lambda_H(\theta)d\theta \leq u_L - u_H \leq \int_{p_L}^{p_H} \Lambda_L(\theta)d\theta$$

and

$$X_0 \geq \max\{\int_b^{p_L} \Lambda_L(\theta)d\theta + u_L, \int_b^{p_H} \Lambda_H(\theta)d\theta + u_H\}.$$

There exists $(u_L, u_H) \geq 0$ which satisfies the first inequality if and only if

$$\int_{p_L}^{p_H} \Lambda_H(\theta)d\theta \leq \int_{p_L}^{p_H} \Lambda_L(\theta)d\theta. \tag{6}$$

When (6) is satisfied, the minimum value of $X_0 \geq 0$ subject to (4) and $(u_L, u_H) \geq 0$ is

$$X_0 = \max\{0, \int_b^{p_L} \Lambda_L(\theta)d\theta + u_L, \int_b^{p_H} \Lambda_H(\theta)d\theta + u_H\}, \tag{7}$$

combined with choice of $u_L = \max\{0, \int_{p_L}^{p_H} \Lambda_H(\theta)d\theta\}$ and $u_H = \max\{0, \int_{p_H}^{p_L} \Lambda_L(\theta)d\theta\}$.[29]

[29] If $p_H \geq p_L$ (resp. $p_H < p_L$), X_0 is minimized at $u_L = \int_{p_L}^{p_H} \Lambda_H(\theta)d\theta$ and $u_H = 0$ (resp. $u_L = 0$ and $u_H = \int_{p_H}^{p_L} \Lambda_L(\theta)d\theta$).

Proceed now to the second stage problem. This involves selecting $(p_L, p_H, b, \Lambda_L, \Lambda_H)$ to maximize P's payoff (5) subject to the following constraints: (p_i, b) satisfies (3) for each $i \in \{L, H\}$, besides (6), while X_0 equals (7).

By Lemma 5, there is no loss of generality in restricting attention to (p_H, p_L, b) which satisfies $p_H \geq p_L$, $\bar{\theta} > p_L > \underline{\theta}$, $p_H \geq b$ and $F_L(p_L) > F_H(p_H)$. With $p_H \geq p_L$, (7) reduces to

$$X_0 = \max\{0, \int_b^{p_L} \Lambda_L(\theta)d\theta + \int_{p_L}^{p_H} \Lambda_H(\theta)d\theta, \int_b^{p_H} \Lambda_H(\theta)d\theta\}. \tag{8}$$

In what follows, Λ_i is said to be *feasible* if $\Lambda_i(\theta)$ is non-decreasing in θ with $\Lambda_i(\underline{\theta}) = 0$ and $\Lambda_i(\bar{\theta}) = 1$. The following lemma considers the problem of finding the feasible Λ_i that minimizes (8) subject to (3) and (6).

Lemma 6 *Define* $\hat{l}_i(p : \beta) = l_i(p : \beta) \equiv p + \beta\frac{F_i(p)-1}{f_i(p)}$ *for* $p > \underline{\theta}$ *and* $-\infty$ *for* $p = \underline{\theta}$ *and* $\hat{h}_i(p : \beta) = h_i(p : \beta) \equiv p + \beta\frac{F_i(p)}{f_i(p)}$ *for* $p < \bar{\theta}$ *and* $+\infty$ *for* $p = \bar{\theta}$, *and*

$$L_i(p_i, \tilde{p} : \beta) \equiv \frac{F_i(p_i)(p_i - \tilde{p}) - (1-\beta)\int_{\tilde{p}}^{p_i} F_i(\theta)d\theta}{\beta}.$$

(i) There exists feasible Λ_i *such that* (p_i, b) *satisfies (3) and (6), if and only if* (p_i, b) *satisfies*

$$\hat{l}_i(p_i : \beta) \leq b \leq \hat{h}_i(p_i : \beta), \tag{9}$$

and

$$L_L(p_H, b : \beta) - L_L(p_L, b : \beta) \geq L_H(p_H, b : \beta) - L_H(p_L, b : \beta). \tag{10}$$

(ii) Consider the problem of minimizing expression (8) for X_0, *over the set of all feasible* $\Lambda_i(.)$ *functions with the property that* (p_i, b) *satisfies (3) and (6). The minimum value of* X_0 *equals*

$$\max\{0, L_H(p_H, b, \beta), L_L(p_L, b, \beta) + L_H(p_H, b, \beta) - L_H(p_L, b, \beta)\}.$$

These results can be explained as follows. (3) implies (for interior $p_H, p_L \in (\underline{\theta}, \bar{\theta})$) the necessary condition: $b = W_i(p_i : \beta) \equiv p_i + \beta\frac{F_i(p_i)-\Lambda_i(p_i)}{f_i(p_i)}$. Since $\Lambda_i(p_i) \in [0, 1]$, this requires (9) for $i = L, H$. This condition restricts the extent to which the coalitional bonus b can deviate from the price p_i offered to A for delivering the good.

Next, we see a lower bound of X_0 equal to (8). By condition (3), the coalition should not benefit from altering the price p_i to any alternative price p':

$$\int_{p'}^{p_i} [b - W_i(\theta : \beta)] dF_i(\theta) \geq 0$$

which reduces to

$$\begin{aligned} \int_{p'}^{p_i} \Lambda_i(\theta) d\theta &\geq -\frac{1}{\beta} \int_{p'}^{p_i} [b - \theta - \frac{F_i(\theta)}{f_i(\theta)} + (1-\beta)\frac{F_i(\theta)}{f_i(\theta)}] dF_i(\theta) \\ &= L_i(p_i, b : \beta) - L_i(p', b : \beta) \end{aligned}$$

where $\beta L_i(p, b : \beta) \equiv [(p - b)F_i(p) - (1-\beta)\int_b^p F_i(\theta)d\theta]$ represents the interim (welfare weighted) loss of rents to the coalition from offering delivery bonus of p to A (rather than b the bonus received by the coalition), measured in units of S's rents.[30] This condition can thus be interpreted as saying that the reduction in coalitional rent loss resulting from a deviation from p_i to p' should be outweighed by the corresponding increase in shadow cost of meeting A's participation constraint. In particular, putting $p' = b$ we obtain the following lower bound for $\int_b^{p_i} \Lambda_i(\theta)d\theta$:

$$\int_b^{p_i} \Lambda_i(\theta)d\theta \geq L_i(p_i, b : \beta)$$

for each $i = L, H$ since $L_i(b, b : \beta) \equiv 0$.

Thus the lower bound on the fixed payment X_0 needed to implement the required allocation as ECP outcome is

$$X_0(p_L, p_H, b : \beta) \equiv \max\{0, L_H(p_H, b : \beta), L_L(p_L, b : \beta) + L_H(p_H, b : \beta) - L_H(p_L, b : \beta)\} \tag{11}$$

The proof of Lemma 6 ensures that the lower bound (11) can indeed be achieved with the appropriate construction of $\Lambda_i(\theta)$, provided the 'monotonicity' condition (10) is satisfied, which follows from the necessity of (6) in ECP.

[30] S's interim loss is $(p - b)F_i(p)$, from which we subtract the associated welfare-weighted expected rent $\int_b^p F_i(\theta)d\theta$ earned by A.

It is instructive to compare this with the corresponding expression for the minimum X_0 in weak-collusion (see Mookherjee et al. (2020, Online Appendix)[31]):

$$X_0(p_L, p_H, b) = \max\{0, F_H(p_H)(p_H - b), F_L(p_L)(p_L - b)\} \tag{12}$$

Intuitively, $F_i(p_i)(p_i - b)$ is the interim loss borne by S owing to countervailing incentives $p_i > b$ in state i. To ensure S wants to participate, P must provide at least this much lump-sum compensation to the coalition in every state under weak collusion. For state H, this minimum compensation $F_H(p_H)(p_H - b)$ is weakly smaller than the corresponding expression $L_H(p_H, b : \beta)$ appearing in the second argument on the right-hand-side of (11).[32] The lower bound on required compensation in weak collusion results from the greater ability of P to manipulate A's outside options.

The following lemma summarizes the above arguments.

Lemma 7 *The optimal allocation is characterized by* $(b^*, X_0^*, p_H^*, p_L^*)$ *which maximizes P's payoff (5) subject to (9), (10) and*

$$X_0 = X_0(p_L, p_H, b : \beta), \tag{13}$$

in addition to: $p_H \geq p_L$, $\bar{\theta} > p_L > \underline{\theta}$, $p_H \geq b$, $F_L(p_L) > F_H(p_H)$, *and* (u_L^*, u_H^*) *satisfying*

$$u_L^* = L_H(p_H^*, b^* : \beta) - L_H(p_L^*, b^* : \beta)$$

and

$$u_H^* = 0.$$

6.3 Properties of Optimal ECP Allocation

Now we provide the main results of this section.

[31] http://people.bu.edu/dilipm/publications/OnlineApp-gebrevJan2020v1.pdf.

[32] If $\beta = 1$ these expressions are exactly the same. When $\beta < 1$, it follows from recalling that we can restrict attention to allocations with countervailing incentives in state H (i.e., $p_H \geq b$), and then observing that $L_H(p_H, b : \beta) \geq F_H(p_H)(p_H - b)$ if and only if $F_H(p_H)(p_H - b) \geq \int_b^{p_H} F_H(\theta) d\theta$.

Proposition 3 *Suppose P procures an indivisible good and all the assumptions stated so far hold.*

(1) Suppose $F(p^{NS})a_H(p^{NS}) > F_H(p^{NS})$. Then hiring S is valuable for $\alpha \in [0, 1/2)$.

(2) If S is valuable at $\alpha \in [0, \frac{1}{2})$, P's payoff is locally decreasing in α, and converges to W^{NS} as α approaches $1/2$. Otherwise, P's payoff equals W^{NS} for all α.

(3) P's payoff in extortionary collusion is always strictly lower than in weak collusion.

Part (1) provides a sufficient condition for hiring of S to be valuable when $\alpha \in [0, 1/2)$. Part (2) states that P's payoff is decreasing in α over the range where S is valuable and has greater bargaining power than A. It then follows that P's payoff is non-increasing in α over the entire range $[0, 1]$ and that it is maximized at $\alpha = 0$. A specific example where these sufficient conditions hold is that of a uniform prior of θ on $[0, 1]$, with $a_H(\theta)$ satisfying $a_H(\theta) = a_L(1 - \theta)$, $a_H(\underline{\theta}) = 0$ and $a_H(\theta)$ is strictly convex on $[0, 1/2]$.[33] With a uniform prior, the condition stated in (1) reduces to

$$p^{NS} a_H(p^{NS})/2 > \int_0^{p^{NS}} a_H(\theta) d\theta.$$

Figure 3 shows an example of $a_H(\theta)$ satisfying these sufficient conditions. For some $p^{NS} \in (0, 1)$, $p^{NS} a_H(p^{NS})/2$ is equal to the area of the triangle A in Figure 3, and $\int_0^{p^{NS}} a_H(\theta) d\theta$ is the area under $a_H(\theta)$ between 0 and p^{NS}. We can confirm that the latter is less than the area A for any choice of $p^{NS} = V/2 \in (0, 1)$. Thus, in this example, S is valuable for any $V \in (0, 2)$.

In order to understand various parts of this Proposition, it is helpful to start by highlighting the contrast between settings of weak collusion and extortionary collusion, i.e., implications of extortion for how the optimal contract is designed and the value of hiring S. In weak collusion, the value of hiring S is represented by the property $b < p_L < p_H$ of the optimal contract, i.e., the use of 'countervailing incentives' in both

[33] $a_H(\theta)$ also needs to satisfy assumption (i) listed at the beginning of this section. These conditions are satisfied with $a_H(\theta) = 2\theta^2$ for $\theta \in [0, 1/2]$ and $1 - 2(1 - \theta)^2$ for $\theta \in (1/2, 1]$.

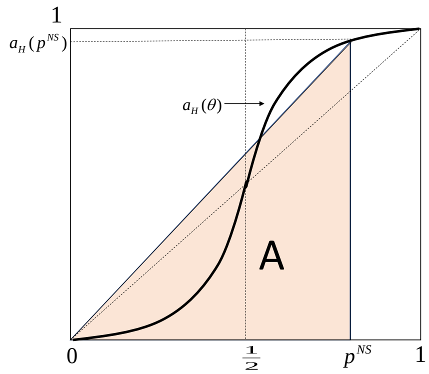

Figure 3: Example where hiring S is valuable

L and H states. As b is the aggregate bonus of the coalition for delivering output, while p_i is the additional payment received by A for delivering output, the inequality $b < p_i$ implies that in state i A has an ex post incentive to deliver output (in states where $\theta < p_i$) while S has the opposite incentive (as the net transfer to S goes down if the output is delivered). The reason for this is the following. Suppose the opposite were to hold in state L, and both S and A are provided a positive incentive to deliver output: $p_L < b$. Then the expression (12) for X_0 for local changes in p_L reduces to $F_H(p_H)(p_H - b)$, which is independent of p_L. The marginal cost of raising p_L slightly would be zero, while the marginal benefit (resulting from a higher likelihood of the good being supplied in state L) is positive, making it worthwhile to raise p_L above b in weak collusion. This argument does not apply in the case of extortionary collusion, owing to the different expression for X_0. Consider the case of $\alpha = 0$ (or $\beta = 1$). Then (13) reduces to

$$X_0 = \max\{0, F_H(p_H)(p_H - b), F_L(p_L)(p_L - b) + F_H(p_H)(p_H - b) - F_H(p_L)(p_L - b)\}.$$

To explain this expression, observe that choosing $p_L < p_H$ induces the coalition to design SC with S threatening to misreport L in state H in order to lower A's outside option payoff. In order to prevent such a manipulation, u_L (the fixed salary paid to A in L) is raised to $F_H(p_L)(b - p_L) - F_H(p_H)(b - p_H)$, equal to S's benefit from a price decline from p_H to p_L in H. This is reflected in the additional term $F_H(p_H)(p_H - b) - F_H(p_L)(p_L - b)$ in X_0, as compared to (12) in weak collusion. Since this term is decreasing in b, it is costly to select b lower than p_L. Lemma 8 in the Appendix shows that $b < p_L$ cannot be optimal, i.e., countervailing incentives are not used in state L in the presence of extortion. Given that part (2) of the proposition holds and α does not matter in weak collusion, this explains part (3) of the Proposition, i.e., why P attains a lower payoff in the presence of extortion.

Even when countervailing incentives are not utilized in state L in the presence of extortion, S may still remain valuable as countervailing incentives are utilized in state H. To see this, suppose S were not hired and we are in the NS setting. Starting with the optimal NS allocation $b = p_L = p_H = p^{NS}$, consider a small increase in p_H. Since (11) reduces to $L_H(p_H, p^{NS} : \beta)$ for $p_H > p_L = b = p^{NS}$ and P's payoff equals $(\kappa_L F_L(p^{NS}) + \kappa_H F_H(p_H))(V - p^{NS}) - L_H(p_H, p^{NS} : \beta)$, it has two opposite effects on P's payoff. It improves P's payoff by alleviating the underproduction problem in NS,

$$\frac{d[\kappa_H F_H(p_H)(V - b)]}{dp_H} |_{p_H = b = p^{NS}} = \kappa_H f_H(p^{NS})(V - p^{NS}) = a_H(p^{NS})F(p^{NS}),$$

using $V = p^{NS} + F(p^{NS})/f(p^{NS})$. On the other hand, it lowers P's payoff by increasing X_0, since

$$dX_0/dp_H|_{p_H = p^{NS}} = \frac{F_H(p_H) + f_H(p_H)(p_H - p^{NS}) - (1 - \beta)F_H(p_H)}{\beta}|_{p_H = p^{NS}} = F_H(p^{NS}).$$

It follows that this variation improves P's payoff, if the sufficient condition in part (1) of the proposition is satisfied.

To explain part (2) of the proposition, suppose S is valuable (and $p_H^* > \max\{b^*, p_L^*\}$) for some β. Moreover suppose the maximum for $X_0(p_L^*, p_H^*, b^* : \beta)$ is achieved at $L_H(p_H^*, b^* : \beta)$, which strictly exceeds $[L_L(p_L^*, b^* : \beta) + L_H(p_H^*, b^* : \beta) - L_H(p_L^*, b^* : \beta)]$. Also for convenience suppose that the monotonicity constraint (10) does not bind. Now suppose A's welfare weight falls, causing β to go up to $\beta' > \beta$. The proof is based on showing that (i) the same allocation (p_L^*, p_H^*, b^*) continues to be feasible at

β', while (ii) the cost X_0 of implementing it goes down. The allocation continues to satisfy constraint (9) at β', since this constraint is weakened further. Intuitively, the decrease in A's welfare weight increases the conflict of interest between S and A, thus widening the feasible gap between their respective interim supply incentives. And part (2) follows from observing that $L_H(p_H^*, b^* : \beta') < L_H(p_H^*, b^* : \beta)$, as $L_H(p, b : \beta)$ is strictly decreasing in β whenever there are countervailing incentives at state H, i.e., $b < p_H$ (since the slope of L_H with respect to β has the same sign as $\int_b^{p_H}[F_H(\theta) - F_H(p_H)]d\theta < 0$). As S shares rents with A when A has greater bargaining power, the compensation that P must pay the coalition for the costs imposed by countervailing incentives in state H goes up.

7 Concluding Remarks

We have not discussed the question of optimality of delegation, which has been studied in a number of preceding papers on mechanism design with weak collusion that ignored the possibility of extortion (e.g., Faure-Grimaud et al. (2003), Celik (2009), Mookherjee et al. (2020)). Delegation refers to a setting where P contracts only with S, and delegates the authority to (sub-)contract with A to S. This corresponds to a special case of our analysis where P offers a null side-contract to A and α is equal to zero. It is easy to show in the setting of our current model that there is no value of hiring S in a delegation setting, and is accordingly dominated by a non-delegation arrangement whenever S has all the bargaining power vis-a-vis A. The reasoning is as follows. In the setting of *ex ante* weak collusion, Mookherjee et al. (2020) show that there is no value of hiring S in the organization with delegation to S. Since A's outside payoff while contracting with S is identically equal to zero under delegation, any allocation that is achievable with weak collusion does not leave any scope for extortion of A by S. Hence, with delegation to S, any allocation that is achievable in weak collusion is also achievable in extortionary collusion. It follows that the presence of extortion leaves the set of achievable allocations with delegation unaffected, and S continues to have no value under delegation. Therefore whenever hiring S is valuable (e.g., conditions of Proposition 3 hold), delegation is strictly dominated by a non-delegation arrangement in which P offers a non-null contract to A.

Next we explain consequences of relaxing some of the assumptions of our model. So far we have assumed S's signal has full support, i.e., every $\eta \in \Pi$ occurs with positive probability for any $\theta \in \Theta$. If we relax this, the Collusion-proof principle can be shown to hold. Let $\Theta(\eta)$ denote the support of θ given η. An achievable allocation (u_A, u_S, q) is defined only on (θ, η) such that $\theta \in \Theta(\eta)$. This can aggravate the problem of extortion, necessitating a variation in how revelation mechanisms need to be augmented. Specifically, the revelation mechanism needs to specify an allocation assigned for messages $(m_A, m_S) = ((\theta_A, \eta_A), \eta_S)$ such that $\theta_A \notin \Theta(\eta_S)$. If A is penalized for such inconsistent messages, the coalition would take effective use of this situation as an additional instrument of extortion from A. However, it is still possible to augment the mechanism to deal with this problem. Moreover, Proposition 2 continues to hold, and Proposition 3 can also be shown to hold for specific information structures without full support. The details are available upon request.

What happens when A cannot observe S's signal? In this case, it is difficult to establish the Collusion-proof principle for the following reason. Our proof of this principle relied on the fact that $u_A(\theta, \eta) \geq \hat{u}_A(\theta, \mu_S(\eta), GC)$ for each η, which was used to ensure condition (B). It means that for the use of message η as a threat to A in GC^R, there exists a corresponding threat $\mu_S(\eta)$ in original GC which is at least as severe, to ensure that the scope of extortion is not enlarged when the mechanism is altered from GC to GC^R. When the signal is privately observed by S, this property becomes difficult to ensure with the kind of augmentation employed in this paper. It is possible that some more complicated augmentation may suffice to generate the same result, but we leave this as an open question.

Our results extend to the context of interim collusion, in which SC is offered by the uninformed third party at the interim stage, after A and S have decided to participate in GC. The Collusion-Proof principle continues to hold. If $\alpha \geq 1/2$, we can apply the same reasoning to show that side contracts will yield efficient collusion, and it is not valuable for P to hire S. With $\alpha < 1/2$, the characterization of collusion-proof allocation in interim collusion differs from that in ex-ante collusion in two respects. First, X_0 could be negative as S and A can no longer coordinate their participation decisions. Second, since collusion is designed after the entry to GC, S's exit option from GC cannot be used to threaten A, implying that we do not need to add an auxiliary message η_0 in constructing a collusion-proof revelation mechanism. If $\alpha = 0$, we can establish almost the same results as in the weak collusion setting of Mookherjee et al. (2020): the optimal allocation differs between interim collusion and ex-ante collusion if and only if delegation to S is optimal under interim collusion. Since A's outside payoff is identically equal to zero with delegation to S, the optimal allocation achieved under weak collusion is also ECP. On the other hand, when the optimal allocation does not differ between interim collusion and ex ante collusion, result (3) in Proposition 3 shows that extortionary collusion lowers P's payoff as compared to weak collusion. Thus extortionary collusion expands the range of environments where delegation to S becomes optimal.

Finally, our model assumed that the information structure is exogenous. In some settings, P may have the capacity to control information available to S and A respectively, which can be an important instrument for controlling corruption, as studied by Ortner and Chassang (2018), Asseyer (2020) and von Negenborn and Pollrich (2020). Extending their analyses to contexts with extortion seems like an interesting question for future research.

References

Andrianova, S. and N. Melissas (2008), "Corruption, Extortion and Boundaries of the Law," *Journal of Law, Economics and Organization*, 25(2), 442-471.

Asseyer, A. (2020), "Collusion and Delegation under Information Control," *Theoretical Economics*, 15(4), 1547-1586.

Ayres, I. (1997), "The Twin Faces of Judicial Corruption: Extortion and Bribery," *Denver University Law Review*, 74, 1231-1254.

Cam H. (1930), *The Hundred and the Hundred Rolls,* London Methuen and Co.

Celik, G. (2009), "Mechanism Design with Collusive Supervision," *Journal of Economic Theory*, 144(1), 69-95.

Che, Y. K. and J. Kim (2009), "Optimal Collusion-Proof Auctions," *Journal of Economic Theory*, 144(2), 565-603.

Dequiedt, V. (2007), "Efficient Collusion in Optimal Auctions," *Journal of Economic Theory*, 136(1), 302-323.

Duflo E., M. Greenstone, R. Pande and N. Ryan (2013), "Truth-Telling by Third-Party Auditors and the Response of Polluting Firms: Experimental Evidence from India," *Quarterly Journal of Economics*, 128(4), 1499-1545.

Faure-Grimaud, A., J. J. Laffont and D. Martimort (2003), "Collusion, Delegation and Supervision with Soft Information," *Review of Economic Studies*, 70(2), 253-279.

Furnivall J.S. (1956), *Colonial Policy and Practice: A Comparative Study of Burma and Netherlands India.* New York University Press.

Fudenberg, D. and J. Tirole (1991), "Perfect Bayesian Equilibrium and Sequential Equilibrium," *Journal of Economic Theory*, 53(2), 236-260.

Hindricks J, M. Keen and A. Muthoo (1999), "Corruption, Extortion and Evasion," *Journal of Public Economics*, 74(3), 395-430.

Holmstrom B. and R. Myerson (1983), "Efficient and Durable Decision Rules with Incomplete Information," *Econometrica*, 51(6), 1799-1819.

Khalil F. , J. Lawarree and S. Yun (2010), "Bribery versus Extortion: Allowing the Lesser of the Two Evils," *Rand Journal of Economics*, 41(1), 179-198.

Klitgaard R. (1988). *Controlling Corruption.* University of California Press.

Laffont, J. J. and J. Tirole (1993), *A Theory of Incentives in Procurement and Regulation*, MIT Press, Cambridge, MA.

Mookherjee D. (1997), "Incentive Reforms in Developing Country Bureaucracies: Lessons from Tax Administration," in B. Pleskovic and J. Stiglitz eds, *Proceedings of the Annual World Bank Conference on Development Economics*, Washington DC: World Bank Press.

Mookherjee, D., A. Motta and M. Tsumagari (2020), "Consulting Collusive Experts," *Games and Economic Behavior*, 122, 290-317.

Nikaido, H. (1954), "On von Neumann's Minimax Theorem," *Pacific Journal of Mathematics*, 4(1), 65-72.

Ortner, J. and S. Chassang (2018), "Making Corruption Harder: Asymmetric Information, Collusion, and Crime," *Journal of Political Economy*, 126(5), 2108-2133.

Quesada L. (2003), "Modeling Collusion as an Informed Principal Problem," Game Theory and Information 0304002, University Library of Munich.

Tirole, J. (1986), "Hierarchies and Bureaucracies: On the Role of Collusion in Organizations," *Journal of Law, Economics, and Organization*, 2(2), 181-214.

Tirole, J. (1992), "Collusion and the Theory of Organizations," In *Advances in Economic Theory, Sixth World Congress, Vol 2*, edited by J.J Laffont, 151-206. Cambridge: Cambridge University Press.

Vanutelli, S. (2020), "From Lapdogs to Watchdogs: Random Auditor Assignment and Municipal Fiscal Performance in Italy," working paper, Boston University.

von Negenborn, C., and M. Pollrich (2020), "Sweet Lemons: Mitigating Collusion in Organizations," *Journal of Economic Theory*, 189, 105074.

Appendix A: Proofs

Proof of Lemma 1

Since (X, u_A, u_S, q) is an achievable allocation, it is straightforward to check that it is feasible in $P^E(\alpha : \eta, GC)$. Here for a reporting strategy $\mu_S(\eta)$ for S in this GC, $\hat{u}_A(\theta, \mu_S(\eta), GC)$ is interpreted as the A's maximum payoff in the event that A is of type θ and exits from the side-contract, whence S chooses $\mu_S(\eta)$.

If $(u_A(\theta, \eta), u_S(\theta, \eta), \mu(\theta, \eta), \mu_S(\eta))$ does not solve problem $P^S(\eta : \alpha, GC)$ for some η, we shall now show that there exists another side-contract and a continuation equilibrium in which the third party can achieve a higher payoff, which will contradict the hypothesis that the allocation resulted from a PBE(c) of GC. Suppose that for some η, the solution of $P(\eta : \alpha, GC)$ is instead some $(\tilde{u}^*_A(\theta, \eta), \tilde{u}^*_S(\theta, \eta), \tilde{\mu}^*(\theta, \eta), \tilde{\mu}^*_S(\eta)) \neq (u_A(\theta, \eta), u_S(\theta, \eta), \mu(\theta, \eta), \mu_S(\eta))$.

Construct a side-contract SC' as follows. Conditioned on the acceptance of SC' by both S and A, the third party requests a report from A of $(\theta_A, \eta_A) \in K$, and report from S of $\eta_S \in \Pi$. The report to P is subsequently selected according to $\tilde{\mu}^*(\theta_A, \eta_S)$, while side-transfers are selected as follows.

$$t_A(\theta_A, \eta_A, \eta_S) = \tilde{u}^*_A(\theta_A, \eta_S) - [\hat{X}_A(\tilde{\mu}^*(\theta_A, \eta_S)) - \theta_A \hat{q}(\tilde{\mu}^*(\theta_A, \eta_S))] - l(\eta_A, \eta_S)$$

and

$$t_S(\theta_A, \eta_A, \eta_S) = \tilde{u}_S^*(\theta_A, \eta_A) - \hat{X}_S(\tilde{\mu}^*(\theta_A, \eta_S))$$

where $l(\eta_A, \eta_S)$ is zero for $\eta_A = \eta_S$ and a large positive number for $\eta_A \neq \eta_S$. These transfers satisfy the budget balance conditions: $t_A(\theta_A, \eta_A, \eta_S) + t_S(\theta_A, \eta_A, \eta_S) \leq 0$ for any $(\theta_A, \eta_A, \eta_S)$, because of constraint (iii) for $\eta_A = \eta_S$ and sufficiently large $l(\eta_A, \eta_S)$ for $\eta_A \neq \eta_S$.

If A were to accept and S were to reject SC', A would threaten to play $\underline{\mu}_A$. Conversely, if S accepts and reports η_S while A rejects SC', S threatens to play $\tilde{\mu}_S^*(\eta_S)$. It is easy to check that there exists a continuation equilibrium where nobody rejects SC' on the equilibrium path, and both A and S report truthfully to the third party, resulting in the allocation $(\tilde{u}_A^*(\theta, \eta), \tilde{u}_S^*(\theta, \eta))$. The third party attains a higher payoff, contradicting the hypothesis that we started with a PBE(c), completing the proof of the first part in the lemma.

Finally, if $X(\theta, \eta) - \theta q(\theta, \eta) > u_A(\theta, \eta) + u_S(\theta, \eta)$, we can find the choice of feasible control variables with $u_S^{**}(\theta, \eta) = X(\theta, \eta) - \theta q(\theta, \eta) - u_A(\theta, \eta)$ instead of $u_S(\theta, \eta)$, taking the other parts of solution as given, which improves the third party's payoff. It implies $X(\theta, \eta) - \theta q(\theta, \eta) = u_A(\theta, \eta) + u_S(\theta, \eta)$. ■

Proof of Proposition 1

Proof of Necessity

Suppose that (u_A, u_S, q) is an achievable allocation in GC in extortionary collusion with α. Evidently (u_A, u_S, q) is IIC. Consider an augmentation of (u_A, u_S, q) to the domain $\bar{K}$ with the selection of

$$(u_A(\theta, \eta_0), u_S(\theta, \eta_0), q(\theta, \eta_0)) \equiv (\hat{u}_A(\theta, \bar{\mu}_S, GC), \omega, \bar{q}(\mu_A(\theta, \bar{\mu}_S, GC), \bar{\mu}_S))$$

where $\mu_A(\theta, \bar{\mu}_S, GC)$ maximizes $\bar{X}_A(\mu_A, \bar{\mu}_S) - \theta \bar{q}(\mu_A, \bar{\mu}_S)$ subject to $\mu_A \in \Delta(M_A)$, and $\omega \equiv \underline{w}_S(GC)$. By the definition, $\omega \geq 0$ and $(u_A(\theta, \eta_0), q(\theta, \eta_0))$ satisfies (IC_A) and (PC_A).

Now consider the problem $P^E(\alpha : \eta)$ defined by the augmented allocation (u_A, u_S, q) on $\bar{K}$. Note that this problem differs from the one considered in Lemma 1 ($P^E(\alpha : \eta, GC)$), as it no longer refers to the original GC.

We show that

$$(\mu(\theta,\eta), \tilde{u}_A(\theta,\eta), \tilde{u}_S(\theta,\eta), P(\cdot \mid \eta)) = (I_1(\theta,\eta), u_A(\theta,\eta), u_S(\theta,\eta), I_2(\eta))$$

solves problem $P^E(\alpha : \eta)$.

It is straightforward to check that $(I_1(\theta,\eta), u_A(\theta,\eta), u_S(\theta,\eta), I_2(\eta))$ satisfies all constraints of $P^E(\alpha : \eta)$, and generates a payoff for the third party of

$$E[\alpha u_A(\theta,\eta) + (1-\alpha)u_S(\theta,\eta) \mid \eta].$$

Suppose otherwise: that there exists some alternative choice of controls

$$(\mu^*(\theta,\eta), u_A^*(\theta,\eta), u_S^*(\theta,\eta), P^*(\cdot \mid \eta))$$

which is feasible in $P^E(\alpha : \eta)$, such that

$$E[\alpha u_A^*(\theta,\eta) + (1-\alpha)u_S^*(\theta,\eta) \mid \eta] > E[\alpha u_A(\theta,\eta) + (1-\alpha)u_S(\theta,\eta) \mid \eta].$$

We show that in such a case there would exist

$$\tilde{\mu}(\theta,\eta) \in \Delta(M_A \times M_S), \tilde{u}_A(\theta,\eta), \tilde{u}_S(\theta,\eta), \tilde{\mu}_S(\eta) \in \Delta(M_S)$$

which would be feasible in $P^E(\alpha : \eta, GC)$ and generate a higher value in that problem compared to $(\mu(\theta,\eta), u_A(\theta,\eta), u_S(\theta,\eta), \mu_S(\eta))$, thereby contradicting the result established at Lemma 1.

$\mu^*(\theta,\eta)$, which is a probability measure on $\bar{K} \cup \{e\}$, divides its weight between reports either in $K \cup \{e\}$ or satisfying $\eta = \eta_0$. The former event corresponds to an outcome of GC that results when S and A's reports are chosen from M_S and M_A respectively. And the latter event corresponds (by specification of $(u_A(\theta,\eta_0), u_S(\theta,\eta_0), q(\theta,\eta_0)))$ to an outcome of GC resulting when S reports $\bar{\mu}_S \in \Delta(M_S)$ and A reports according to $\mu_A(\theta, \bar{\mu}_S) \in \Delta(M_A)$. In this case,

$$\bar{q}(\mu_A(\theta,\bar{\mu}_S), \bar{\mu}_S) = q(\theta,\eta_0)$$

while

$$X(\theta, \eta_0) = \omega + \bar{X}_A(\mu_A(\theta, \bar{\mu}_S), \bar{\mu}_S) \leq \bar{X}_S(\mu_A(\theta, \bar{\mu}_S), \bar{\mu}_S) + \bar{X}_A(\mu_A(\theta, \bar{\mu}_S), \bar{\mu}_S)$$

since ω is S's minmax payoff in GC. Hence the outcome of $\mu^*(\theta, \eta)$ in $P^E(\alpha : \eta)$ can be attained by the coalition as an outcome of GC resulting from some reporting strategy $\tilde{\mu}(\theta, \eta) \in \Delta(M_A \times M_S)$ that satisfies

$$\hat{X}_A(\tilde{\mu}(\theta, \eta)) + \hat{X}_S(\tilde{\mu}(\theta, \eta)) \geq X(\mu^*(\theta, \eta))$$

and

$$\hat{q}(\tilde{\mu}(\theta, \eta)) = q(\mu^*(\theta, \eta)).$$

Let $\mu_S(\eta)$ denote the optimal threat chosen by S in the event that A does not participate in the side-contract, in the solution to problem $P^E(\alpha : \eta, GC)$. Let us select $\mu_S(\eta_0) \equiv \bar{\mu}_S$. Then $u_A(\theta, \eta) \geq \hat{u}_A(\theta, \mu_S(\eta), GC)$ for any $\eta \in \bar{\Pi}$. Define $\tilde{\mu}_S(\eta) \in \Delta(M_S)$ as the composite of the measures $\mu_S(\eta')$ and $P^*(\eta' \mid \eta)$. Then by the definition of $\hat{u}_A(\theta, \mu_S, GC)$,

$$\Sigma_{\eta' \in \bar{\Pi}} P^*(\eta' \mid \eta) u_A(\theta, \eta') \geq \Sigma_{\eta' \in \bar{\Pi}} P^*(\eta' \mid \eta) \hat{u}_A(\theta, \mu_S(\eta'), GC) \geq \hat{u}_A(\theta, \tilde{\mu}_S(\eta), GC).$$

Since $u_A^*(\theta, \eta) \geq \Sigma_{\eta' \in \bar{\Pi}} P^*(\eta' \mid \eta) u_A(\theta, \eta')$, it follows that $u_A^*(\theta, \eta) \geq \hat{u}_A(\theta, \tilde{\mu}_S(\eta), GC)$. Defining $\tilde{u}_A(\theta, \eta) \equiv u_A^*(\theta, \eta)$ and

$$\tilde{u}_S(\theta, \eta) \equiv \hat{X}_A(\tilde{\mu}(\theta, \eta)) + \hat{X}_S(\tilde{\mu}(\theta, \eta)) - \theta \hat{q}(\tilde{\mu}(\theta, \eta)) - u_A^*(\theta, \eta),$$

we infer that $(\tilde{u}_A(\theta, \eta), \tilde{u}_S(\theta, \eta), \tilde{\mu}(\theta, \eta), \tilde{\mu}_S(\eta))$ is feasible in the problem $P^E(\alpha : \eta, GC)$, and $\tilde{u}_S(\theta, \eta) \geq u_S^*(\theta, \eta)$. Hence it generates a higher payoff for the third party than $E[\alpha u_A(\theta, \eta) + (1-\alpha) u_S(\theta, \eta) \mid \eta]$, and we obtain a contradiction to the result of Lemma 1. So $(I_1(\theta, \eta), u_A(\theta, \eta), u_S(\theta, \eta), I_2(\eta))$ must be a solution of $P^E(\alpha : \eta)$, establishing the necessity of the statement.

Proof of Sufficiency

Let (u_A, u_S, q) be an augmentation of allocation for which the latter satisfies the ECP property. By Definition 2, $u_S(\theta, \eta_0) = \omega$ for any $\theta \in \Theta$ and $(u_A(\theta, \eta_0), q(\theta, \eta_0))$ satisfies (IC_A) and (PC_A). P can construct a revelation grand contract GC^R as follows:

$$(X_A(m_A, m_S), X_S(m_A, m_S), q(m_A, m_S) : M_A, M_S)$$

where $M_A = K \cup \{e_A\}$, $M_S = \bar{\Pi} \cup \{e_S\}$ and

- for any $(\theta, \eta) \in K$ and $\eta' \in \Pi$, choose $(X_A((\theta, \eta), \eta'), X_S((\theta, \eta), \eta'), q((\theta, \eta), \eta')) = (u_A(\theta, \eta') + \theta q(\theta, \eta'), u_S(\theta, \eta) - L(\eta, \eta'), q(\theta, \eta'))$ where $L(\eta, \eta') = 0$ for $\eta = \eta'$ and $L > 0$ (and sufficiently large) for $\eta \neq \eta'$
- $(X_A((\theta, \eta), e_S), X_S((\theta, \eta), e_S), q((\theta, \eta), e_S)) = (u_A(\theta, \eta_0) + \theta q(\theta, \eta_0), 0, q(\theta, \eta_0))$.
- $(X_A((\theta, \eta), \eta_0), X_S((\theta, \eta), \eta_0), q((\theta, \eta), \eta_0)) = (u_A(\theta, \eta_0) + \theta q(\theta, \eta_0), \omega, q(\theta, \eta_0))$.
- $(X_A(e_A, m_S), X_S(e_A, m_S), q(e_A, m_S)) = (0, 0, 0)$ for any $m_S \neq \eta_0$
- $(X_A(e_A, \eta_0), X_S(e_A, \eta_0), q(e_A, \eta_0)) = (0, \omega, 0)$.

It is easy to check that $(\mu_A, \mu_S) = ((\theta, \eta), \eta)$ is a non-cooperative equilibrium of GC, and S's minmax payoff in GC is ω. The ECP property of (u_A, u_S, q) implies there is no room for the third party to improve its payoff by offering a deviating side-contract, so (u_A, u_S, q) is realized as the outcome of a PBE(c) under GC. ■

Proof of Lemma 2 and Proposition 2

Consider ECP allocation (u_A, u_S, q) for $\alpha \geq \frac{1}{2}$. By Proposition 1, there exists $\omega \geq 0$ and an incentive compatible augmentation of this allocation satisfying $u_S(\theta, \eta_0) = \omega$, such that for any η, $(I_1(\theta, \eta), u_A(\theta, \eta), u_S(\theta, \eta), I_2(\eta))$ solves $P^E(\alpha : \eta)$. For (X, q) in Definition 2, we define

$$\mu^*(\theta) \in \arg \max_{\mu \in \Delta(\bar{K} \cup \{e\})} [X(\mu) - \theta q(\mu)]$$

i.e., a reporting strategy that maximizes the *ex post* joint payoff of A and S in every state.

We claim that

$$(\mu(\theta,\eta),u_A(\theta,\eta),u_S(\theta,\eta))=(\mu^*(\theta),X(\mu^*(\theta))-\theta q(\mu^*(\theta))-\omega,\omega)$$

is a solution of $P^E(\alpha:\eta)$ for any η. Upon setting $\omega=c$, $X_A(\theta)=X(\mu^*(\theta))-\omega$ and $Q(\theta)=q(\mu^*(\theta))$, this claim will imply that ECP allocation satisfies (i)-(iii) in the proposition.

To establish the claim, we first derive an upper bound for the objective function in the problem $P^E(\alpha:\eta)$. From the constraint $E[\tilde{u}_S(\theta,\eta)\mid\eta]\geq\omega$ and the assumption that $\alpha\geq 1/2$, for any reporting strategy $\mu(\theta,\eta)$, the following is true:

$$\begin{aligned}
& E[\alpha\tilde{u}_A(\theta,\eta)+(1-\alpha)\tilde{u}_S(\theta,\eta)\mid\eta]\\
\leq\ & E[\alpha\{X(\mu(\theta,\eta))-\theta q(\mu(\theta,\eta))\}+(1-2\alpha)\tilde{u}_S(\theta,\eta)\mid\eta]\\
\leq\ & \alpha E[X(\mu^*(\theta))-\theta q(\mu^*(\theta))\mid\eta]+(1-2\alpha)\omega.
\end{aligned}$$

This upper bound can be attained in $P^E(\alpha:\eta)$ by choosing $\mu(\theta,\eta)=\mu^*(\theta)$,

$$\tilde{u}_A(\theta,\eta)=X(\mu^*(\theta))-\theta q(\mu^*(\theta))-\omega$$

and $\tilde{u}_S(\theta,\eta)=\omega$ for any $\theta\in\Theta$, and $P(\eta_0\mid\eta)=1$ and $P(\eta'\mid\eta)=0$ for any $\eta'\neq\eta_0$. This allocation satisfies A's participation constraint (v), since

$$\begin{aligned}
& \tilde{u}_A(\theta,\eta)=X(\mu^*(\theta))-\theta q(\mu^*(\theta))-\omega\\
\geq\ & X(\theta,\eta_0)-\theta q(\theta,\eta_0)-u_S(\theta,\eta_0)=u_A(\theta,\eta_0).
\end{aligned}$$

As the other constraints are obviously satisfied, the claim is established. It is also evident that any allocation satisfying (i)-(iii) with $c=0$ is achieved in NS. ■

Proof of Lemma 3

Suppose that for some η, $(I_1(\theta,\eta), u_A(\theta,\eta), u_S(\theta,\eta), I_2(\eta))$ does not solve the relaxed version of $P^E(\alpha:\eta)$ where the constraint $E[\tilde{u}_S(\theta,\eta)\mid\eta]\geq\omega$ is dropped. It implies $E[\tilde{u}^r_S(\theta,\eta)\mid\eta]<\omega$ in the optimal solution of the relaxed problem represented by

$$(\mu^r(\theta,\eta), \tilde{u}^r_A(\theta,\eta), \tilde{u}^r_S(\theta,\eta), P^r(\cdot\mid\eta)).$$

As shown in the proof of Proposition 2, side contract $\tilde{SC}$ defined as follows is feasible in $P^E(\alpha:\eta)$, hence also in the relaxed problem:

- $\tilde{\mu}(\theta,\eta)=\mu^*(\theta)$ which maximizes $X(\mu)-\theta q(\mu)$ subject to $\mu\in\Delta(\bar{K}\cup\{e\})$
- $P(\eta_0\mid\eta)=1$ and $P(\eta'\mid\eta)=0$ for any $\eta'\neq\eta_0$
- $\tilde{u}_A(\theta,\eta)=X(\mu^*(\theta))-\theta q(\mu^*(\theta))-\omega$ (denoted by $u^+_A(\theta,\eta)$ in later part)
- $\tilde{u}_S(\theta,\eta)=\omega$

Hence

$$\begin{aligned}
& E[\alpha\tilde{u}^r_A(\theta,\eta)+(1-\alpha)\tilde{u}^r_S(\theta,\eta)\mid\eta] \\
= \; & E[(1-\alpha)\{X(\mu^r(\theta,\eta))-\theta q(\mu^r(\theta,\eta))\}-(1-2\alpha)\tilde{u}^r_A(\theta,\eta)\mid\eta] \\
\geq \; & E[(1-\alpha)\{X(\mu^*(\theta))-\theta q(\mu^*(\theta))\}-(1-2\alpha)u^+_A(\theta,\eta)\mid\eta].
\end{aligned}$$

But since $E[X(\tilde{\mu}^r(\theta,\eta))-\theta q(\tilde{\mu}^r(\theta,\eta))\mid\eta]\leq E[X(\mu^*(\theta))-\theta q(\mu^*(\theta))\mid\eta]$ by the definition of $\mu^*(\theta)$, $\alpha<\frac{1}{2}$ implies that $E[u^+_A(\theta,\eta)\mid\eta]\geq E[\tilde{u}^r_A(\theta,\eta)\mid\eta]$. This implies that the side contract $\tilde{SC}$ creates a Pareto improvement over the solution to the relaxed problem, yielding a strictly higher value of the third party's expected payoff, a contradiction. ■

Proof of Lemma 4

Provided in the text.

Proof of Lemma 5

Consider the optimal NS allocation which generates payoff W^{NS} to P. ECP implies S's participation constraints:

$$F_i(p_i^*)(b^* - p_i^*) + X_0^* - u_i^* \geq 0 \tag{14}$$

for $i \in \{L, H\}$. This implies

$$W^* = \Sigma_i \kappa_i F_i(p_i^*)[V - b^*] - X_0^* \leq \Sigma_i \kappa_i [F_i(p_i^*)(V - p_i^*) - u_i^*] \leq \Sigma_i \kappa_i F_i(p_i^*)(V - p_i^*).$$

To show $p_H^* \geq p_L^*$, suppose otherwise: $p_L^* > p_H^*$. Consider the maximization of $\Sigma_i \kappa_i F_i(p_i)(V - p_i)$ subject to $p_L \geq p_H$. We claim that the constraint must be binding at the solution. Otherwise the solution to this problem will be the same when the constraint is dropped (because the objective function is concave and the feasible set is convex). But our distributional assumptions imply the solution to the unconstrained problem will involve a lower price when $i = L$. Therefore $p_L = p_H$ at the solution, and this solution must equal p^{NS}. Since $p_L^* > p_H^*$, $\Sigma_i \kappa_i F_i(p_i^*)(V - p_i^*) < W^{NS}$, so the inequality above implies that $W^{NS} > W^*$, a contradiction. Moreover, if S is valuable we have $W^* > W^{NS}$, so it must be the case that $p_H^* > p_L^*$ (otherwise $\Sigma_i \kappa_i F_i(p_i^*)(V - p_i^*) = F(p_L^*)(V - p_L^*)$, which cannot exceed W^{NS}).

To show $p_L^* > \underline{\theta}$, suppose otherwise $p_L^* = \underline{\theta}$. Then $W^* \leq \Sigma_i \kappa_i F_i(p_i^*)(V - p_i^*) = \kappa_H F_H(p_H^*)(V - p_H^*) < F_H(p_H^*)(V - p_H^*) \leq F(p_H^*)(V - p_H^*) \leq W^{NS}$, a contradiction. To show $p_L^* < \bar{\theta}$, suppose otherwise $p_L^* = \bar{\theta}$. Then $p_L^* = p_H^* = \bar{\theta}$, and $\Sigma_i \kappa_i F_i(p_i^*)(V - p_i^*) = V - \bar{\theta}$. This is less than W^{NS}, since $p^{NS} \in (\underline{\theta}, \bar{\theta})$, and we again get a contradiction.

Next, to show $b^* \leq p_H^*$, suppose otherwise $b^* > p_H^*$. Then

$$W^* = \Sigma_i \kappa_i F_i(p_i^*)[V - b^*] - X_0^* < \Sigma_i \kappa_i F_i(p_i^*)[V - p_H^*] \leq F(p_H^*) \max\{V - p_H^*, 0\} \leq W^{NS}.$$

The first inequality uses $X_0^* \geq 0$ and $b^* > p_H^*$, and the second inequality uses $p_L^* \leq p_H^*$ (which implies $\kappa_L F_L(p_L^*) + \kappa_H F_H(p_H^*) \leq \kappa_L F_L(p_H^*) + \kappa_H F_H(p_H^*) \equiv F(p_H^*)$). This contradicts $W^* \geq W^{NS}$. If $b^* = p_H^*$, we obtain $W^* \leq W^{NS}$ by replacing $<$ by $\leq$ in the first inequality. Thus if S is valuable, $b^* < p_H^*$.

Suppose $F_L(p_L^*) \leq F_H(p_H^*)$. Since $p_L^* \in (\underline{\theta}, \bar{\theta})$, it also implies $p_L^* < p_H^*$. Then $p_H^* \geq b^*$ implies $\max\{F_L(p_L^*)(p_L^* - b^*), 0\} \leq F_H(p_H^*)(p_H^* - b^*)$. Using (14), $X_0^* \geq 0$ and $u_i^* \geq 0$:

$$\begin{aligned} & X_0^* \geq \max\{0, F_L(p_L^*)(p_L^* - b^*) + u_L^*, F_H(p_H^*)(p_H^* - b^*) + u_H^*\} \\ \geq \quad & \max\{0, F_L(p_L^*)(p_L^* - b^*), F_H(p_H^*)(p_H^* - b^*)\} = F_H(p_H^*)(p_H^* - b^*). \end{aligned}$$

Then

$$\begin{aligned} W^* &= \Sigma_i \kappa_i F_i(p_i^*)[V - b^*] - X_0^* \leq \Sigma_i \kappa_i F_i(p_i^*)[V - b^*] - F_H(p_H^*)(p_H^* - b^*) \\ &\leq \Sigma_i \kappa_i F_i(p_i^*)[V - p_H^*] \leq F(p_H^*) \max\{V - p_H^*, 0\} \leq W^{NS} \end{aligned}$$

The second inequality uses $F_L(p_L^*) \leq F_H(p_H^*)$, and the third inequality uses $p_L^* < p_H^*$. If $V > p_H^*$, the third inequality is strict, and if $V \leq p_H^*$, the fourth inequality is strict, implying $W^* < W^{NS}$ in either case, so we obtain a contradiction. ■

Proof of Lemma 6

Proof of (i):

Suppose there exists feasible Λ_i such that (p_i, b) satisfies (3) and (6). We first show that (9) is satisfied. Suppose otherwise. If $b < \hat{l}_i(p_i : \beta)$, it implies $p_i > \underline{\theta}$ and $l_i(p_i : \beta) = p_i + \beta \frac{F_i(p_i) - 1}{f_i(p_i)} > b$. Let p' be the smallest $p \in \Theta$ such that $p + \beta \frac{F_i(p) - 1}{f_i(p)} \geq b$. Since $p + \beta \frac{F_i(p) - 1}{f_i(p)}$ is continuous and increasing in p on Θ, it follows that $p' < p_i$ and $W_i(p : \beta) \geq p + \beta \frac{F_i(p) - 1}{f_i(p)} > b$ for any $p \in (p', p_i)$ for any feasible Λ_i. Then

$$\int_{\underline{\theta}}^{p'} (b - W_i(\theta : \beta)) dF_i(\theta) > \int_{\underline{\theta}}^{p_i} (b - W_i(\theta : \beta)) dF_i(\theta),$$

which implies that p_i does not solve (3) for any feasible Λ_i. We can apply a similar argument when $b > \hat{h}_i(p_i : \beta)$. Therefore (9) must hold.

Next, we show that (10) must also hold. If (p_i, b) satisfies (3) for some feasible Λ_i,

$$\int_{\underline{\theta}}^{p_i} [b - W_i(\theta : \beta)] dF_i(\theta) \geq \int_{\underline{\theta}}^{p'} [b - W_i(\theta : \beta)] dF_i(\theta)$$

for any p'. This inequality can be rewritten as

$$0 \leq \int_{p'}^{p_i} [b - W_i(\theta : \beta)] dF_i(\theta) = (b - p_i)F_i(p_i) - (b - p')F_i(p') + (1-\beta) \int_{p'}^{p_i} F_i(\theta)d\theta + \beta \int_{p'}^{p_i} \Lambda_i(\theta)d\theta,$$

or equivalently

$$\int_{p'}^{p_i} \Lambda_i(\theta)d\theta \geq L_i(p_i, b : \beta) - L_i(p', b : \beta). \tag{15}$$

Therefore

$$\int_{p_L}^{p_H} \Lambda_L(\theta)d\theta \leq L_L(p_H, b : \beta) - L_L(p_L, b : \beta)$$

and

$$\int_{p_L}^{p_H} \Lambda_H(\theta)d\theta \geq L_H(p_H, b : \beta) - L_H(p_L, b : \beta).$$

Then using (6), it follows that (10) must hold. Hence the 'only if' statement is correct.

The proof of the reverse 'if' direction is included in the proof of (ii) below.

Proof of (ii):

By the argument above, if (p_i, b) satisfies (3) and (6), then it also satisfies (9) and (10). Using (15) and $L_i(b, b : \beta) = 0$, it follows that a lower bound of X_0 is

$$\max\{0, L_H(p_H, b : \beta), L_L(p_L, b : \beta) + L_H(p_H, b : \beta) - L_H(p_L, b : \beta)\}.$$

Next we construct feasible Λ_i ($i = L, H$) such that this lower bound of X_0 is achieved, and (p_i, b) satisfies (3) and (6). More specifically, we construct feasible Λ_i for $i = L, H$ which satisfies the following three conditions for each $i = L, H$:

$$\int_{p_L}^{p_H} [b - W_i(\theta : \beta)] dF_i(\theta) = 0,$$

$$\int_{b}^{p_i} [b - W_i(\theta : \beta)] dF_i(\theta) = 0,$$

and $\tilde{p}_i = p_i$ maximizes

$$\int_{\underline{\theta}}^{\tilde{p}_i} [b - W_i(\theta : \beta)] dF_i(\theta).$$

Since the first two conditions are equivalent to

$$\int_{p_L}^{p_H} \Lambda_i(\theta)d\theta = L_i(p_H, b : \beta) - L_i(p_L, b : \beta)$$

and

$$\int_b^{p_i} \Lambda_i(\theta)d\theta = L_i(p_i, b : \beta)$$

for $i = L, H$, Λ_L and Λ_H achieve the lower bound of X_0, and since (10) holds it follows that (6) is satisfied.

Now we explain how to construct Λ_i. By Lemma 5, there are either of two cases to consider: $b \le p_L \le p_H$ or $p_L \le b \le p_H$. We explain the construction in the former case, a similar approach can be used in the other case.

If $p_H = \bar{\theta}$, $p_H + \beta\frac{F_i(p_H)-1}{f_i(p_H)} = \bar{\theta} \le b$ implies $b = p_L = p_H = \bar{\theta}$. In this case, the above three conditions are satisfied upon choosing $\Lambda_i(\theta) = F_i(\theta)$.

Next, suppose that $\underline{\theta} < p_L \le p_H < \bar{\theta}$. Define $\hat{\Lambda}_i(b)$, $\hat{\Lambda}_i(p_L)$ and $\hat{\Lambda}_i(p_H)$ as follows:

$$\hat{\Lambda}_i(b) \equiv F_i(\max\{\underline{\theta}, b\}),$$

$$\hat{\Lambda}_i(p_L) \equiv F_i(p_L) + (p_L - b)f_i(p_L)/\beta,$$

$$\hat{\Lambda}_i(p_H) \equiv F_i(p_H) + (p_H - b)f_i(p_H)/\beta.$$

Given any $\theta^* \in [\max\{b, \underline{\theta}\}, p_L]$ and $\theta^{**} \in [p_L, p_H]$, Λ_i is constructed as follows: $\Lambda_i(\underline{\theta}) = 0$, $\Lambda_i(\theta) = \hat{\Lambda}_i(b)$ on $(\underline{\theta}, \theta^*)$, $\Lambda_i(\theta) = \hat{\Lambda}_i(p_L)$ on $[\theta^*, \theta^{**})$, $\Lambda_i(\theta) = \hat{\Lambda}_i(p_H)$ on $[\theta^{**}, \bar{\theta})$ and $\Lambda_i(\bar{\theta}) = 1$. We now show that this Λ_i will satisfy all the requirements upon selecting (θ^*, θ^{**}) appropriately in the following (a)-(c).

(a) Feasibility of Λ_i:

Since (p_k, b) satisfies (9) for $k = L, H$, $F_i(p_k) + (p_k - b)f_i(p_k)/\beta \in [0, 1]$ if $p_k \in (\underline{\theta}, \bar{\theta})$. Then $\hat{\Lambda}_i(p_k) \in [0, 1]$ for $k = L, H$. Since $h_i(\theta)$ and $l_i(\theta)$ are increasing in θ, $\theta + \beta\frac{F_i(\theta)-\gamma}{f_i(\theta)}$ is increasing in θ on Θ for any given $\gamma \in [0, 1]$. Then if $\underline{\theta} < p_L \le p_H < \bar{\theta}$, the definitions of $\hat{\Lambda}_i(p_L), \hat{\Lambda}_i(p_H)$ above imply

$$b = p_L + \beta\frac{F_i(p_L) - \hat{\Lambda}_i(p_L)}{f_i(p_L)} = p_H + \beta\frac{F_i(p_H) - \hat{\Lambda}_i(p_H)}{f_i(p_H)}$$

which in turn implies $\hat{\Lambda}_i(p_L) \le \hat{\Lambda}_i(p_H)$. Then $b \le p_L \le p_H$ implies

$$0 \le \hat{\Lambda}_i(b) \le \hat{\Lambda}_i(p_L) \le \hat{\Lambda}_i(p_H) \le 1$$

establishing that Λ_i is feasible.

(b) Achievability of the lower bound:

Define $G_i(p,p^{'}) \equiv \int_p^{p^{'}} [b - W_i(\theta : \beta)] dF_i(\theta)$. We show there exist $\theta^* \in [\max\{\underline{\theta}, b\}, p_L]$ and $\theta^{**} \in [p_L, p_H]$ such that

$$G_i(p_L, p_H) = \int_{p_L}^{p_H} [b - W_i(\theta : \beta)] dF_i(\theta) = 0$$

and

$$G_i(b, p_H) = \int_{b}^{p_H} [b - W_i(\theta : \beta)] dF_i(\theta) = 0.$$

$G_i(p_L, p_H)$ is non-increasing in θ^{**} on $[p_L, p_H]$, but does not depend on θ^*. If $\theta^{**} = p_L$,

$$b - W_i(\theta : \beta) = b - (\theta + \beta \frac{F_i(\theta) - \hat{\Lambda}_i(p_H)}{f_i(\theta)})$$

on $[p_L, p_H]$. Since $l_i(\theta)$ and $h_i(\theta)$ are increasing, this is decreasing in θ and is equal to zero at $\theta = p_H$. Hence $G_i(p_L, p_H) \geq 0$. If $\theta^{**} = p_H$,

$$b - W_i(\theta : \beta) = b - (\theta + \beta \frac{F_i(\theta) - \hat{\Lambda}_i(p_L)}{f_i(\theta)})$$

is decreasing in θ on $[p_L, p_H]$ and is equal to zero if $\theta = p_L$. Thus $G_i(p_L, p_H) \leq 0$. These arguments guarantee the existence of $\theta^{**} \in [p_L, p_H]$ such that $G_i(p_L, p_H) = 0$.

With θ^{**} selected as above, $G_i(b, p_H) = G_i(b, p_L)$. The right-hand side depends only on θ^* and is non-increasing in θ^* on $[\max\{\underline{\theta}, b\}, p_L]$. If $\theta^* = \max\{\underline{\theta}, b\}$,

$$b - W_i(\theta : \beta) = b - (\theta + \beta \frac{F_i(\theta) - \hat{\Lambda}_i(p_L)}{f_i(\theta)})$$

is decreasing in θ on $[\max\{\underline{\theta}, b\}, p_L]$, and is equal to zero at $\theta = p_L$, and $G_i(b, p_L) \geq 0$. If $\theta^* = p_L$,

$$b - W_i(\theta : \beta) = b - (\theta + \beta \frac{F_i(\theta) - \hat{\Lambda}_i(b)}{f_i(\theta)})$$

is decreasing in θ on $[\max\{\underline{\theta}, b\}, p_L]$, and is nonpositive at $\theta = \max\{\underline{\theta}, b\}$, and $G_i(b, p_L) \leq 0$. These arguments guarantee the existence of $\theta^* \in [\max\{\underline{\theta}, b\}, p_L]$ such that $G_i(b, p_L) = 0$.

(c): $\tilde{p}_i = p_i$ solves (3).

For Λ_i with the above selection of θ^* and θ^{**}, $\int_{\underline{\theta}}^{\tilde{p}_i}[b - W_i(\theta : \beta)]dF_i(\theta)$ is non-decreasing in $\tilde{p}_i$ on $[\underline{\theta}, \max\{\underline{\theta}, b\}]$, non-increasing on $[\max\{\underline{\theta}, b\}, \theta^*]$, non-decreasing on $[\theta^*, p_L]$, non-increasing on $[p_L, \theta^{**}]$, non-decreasing on $[\theta^{**}, p_H]$ and non-increasing on $[p_H, \bar{\theta}]$. Then

$$\int_{p_L}^{p_H}[b - W_i(\theta : \beta)]dF_i(\theta) = 0$$

and

$$\int_{b}^{p_i}[b - W_i(\theta : \beta)]dF_i(\theta) = 0$$

imply (p_i, b) satisfies (3). ■

Proof of Lemma 7

Provided in the text.

Proof of Proposition 3

Proof of (2)

We start with the proof of (2). The second part of (2) is trivial, so we focus on the first part, and presume that hiring S is valuable.

For $\beta \in (0,1)$, suppose that (p_H^*, p_L^*, b^*) solves the problem and $W^* > W^{NS}$. Then $p_H^* > p_L^*$ and $p_H^* > b^*$ by Lemma 5. Since $p_H^* > b^*$ implies $L_H(p_H^*, b^* : \beta) > 0$,

$$X_0^* = \tilde{X}_0(p_H^*, p_L^*, b^* : \beta) \tag{16}$$

where $\tilde{X}_0(p_H, p_L, b : \beta) \equiv \max\{L_H(p_H, b : \beta), L_L(p_L, b : \beta) + L_H(p_H, b : \beta) - L_H(p_L, b : \beta)\}$.

Step 1: (p_L^*, p_H^*, b^*) satisfies (9) and (10) for any $\tilde{\beta} > \beta$.

Since (p_L^*, p_H^*, b^*) satisfies (9) for β, it evidently satisfies it for any $\tilde{\beta} > \beta$. On the other hand, if $L_L(p_H^*, b^* : \beta) - L_L(p_L^*, b^* : \beta) - [L_H(p_H^*, b^* : \beta) - L_H(p_L^*, b^* : \beta)]$ is non-negative, the same is true for any $\tilde{\beta} > \beta$, since

$$\begin{aligned} & L_L(p_H^*, b^* : \beta) - L_L(p_L^*, b^* : \beta) - [L_H(p_H^*, b^* : \beta) - L_H(p_L^*, b^* : \beta)] \\ = \; & \frac{1}{\beta}[F_L(p_H^*) - F_H(p_H^*))(p_H^* - b^*) - (F_L(p_L^*) - F_H(p_L^*))(p_L^* - b^*) \\ - \; & (1-\beta)\int_{p_L^*}^{p_H^*} (F_L(\theta) - F_H(\theta))d\theta] \end{aligned}$$

and since $F_L(\theta) > F_H(\theta)$ for all $\theta \in (\underline{\theta}, \bar{\theta})$:

$$\int_{p_L^*}^{p_H^*} (F_L(\theta) - F_H(\theta))d\theta > 0.$$

Therefore (p_L^*, p_H^*, b^*) satisfies (10) for any $\tilde{\beta} > \beta$.

Step 2: $\tilde{X}_0(p_H^*, p_L^*, b^* : \beta) > \tilde{X}_0(p_H^*, p_L^*, b^* : \tilde{\beta})$ *for any* $\tilde{\beta} > \beta$.

The proof is divided into two cases; Case A: $b^* < p_L^*$ and Case B: $b^* \geq p_L^*$.

Case A: $b^* < p_L^*$.

The proof proceeds through the following sequence of claims. Define $A(b) \equiv \sum_i \kappa_i F_i(p_i^*)[V - b] - L_H(p_H^*, b; \beta)$ and $B(b) \equiv \sum_i \kappa_i F_i(p_i^*)[V - b] - [L_L(p_L^*, b : \beta) + L_H(p_H^*, b : \beta) - L_H(p_L^*, b : \beta)]$, so P's payoff as b is varied while prices are set at their optimal values can be expressed as $\Pi(b) \equiv \min\{A(b), B(b)\}$.

Claim 1: $\Pi(b)$ *is increasing in* b *over the range* $b < p_L^*$.

To establish this, observe first that $B(b)$ is increasing in b over the range $b < p_L^*$, because its slope over this range equals

$$\begin{aligned} & -(\kappa_L F_L(p_L^*) + \kappa_H F_H(p_H^*)) + \frac{1}{\beta}(F_H(p_H^*) + F_L(p_L^*) - (1-\beta)F_L(b) - F_H(p_L^*)) \\ > \; & -\kappa_L F_L(p_L^*) - \kappa_H F_H(p_H^*) + \frac{1}{\beta}(F_H(p_H^*) + F_L(p_L^*) - (1-\beta)F_L(p_L^*) - F_H(p_H^*)) \\ = \; & \kappa_H(F_L(p_L^*) - F_H(p_H^*)) > 0. \end{aligned}$$

Hence the claim is true at any $b < p_L^*$ where $\Pi(b) = B(b) < A(b)$.

It is also true when $\Pi(b) = A(b) \leq B(b)$ and $A'(b) > 0$. So suppose that $\Pi(\tilde{b}) = A(\tilde{b}) \leq B(\tilde{b})$ and $A'(\tilde{b}) \leq 0$ at some $\tilde{b} < p_L^*$. Observe that the slope of $A(b)$ which equals

$$\frac{1}{\beta}[(1-\beta\kappa_H)F_H(p_H^*) - \beta\kappa_L F_L(p_L^*) - (1-\beta)F_H(b)],$$

is decreasing in b. This implies $A(p_L^*) < A(\tilde{b}) \leq B(\tilde{b}) < B(p_L^*)$. This leads to a contradiction, since $A(b)$ and $B(b)$ are equal at $b = p_L^*$: $L_L(p_L^*, p_L^* : \beta) = 0 = L_H(p_L^*, p_L^* : \beta)$. This establishes Claim 1.

Claim 2: If $b^* < p_L^*$ then constraint (10) binds.

Otherwise we can raise b slightly from b^* without violating any constraint, and by Claim 1 P's profit would increase.

We are now in a position to complete the argument for Case A. By Claim 2, if $b^* < p_L^*$ then (10) binds. Hence

$$\begin{aligned}
&\tilde{X}_0(p_H^*, p_L^*, b^* : \beta) = \max\{L_H(p_H^*, b^* : \beta), L_L(p_H^*, b^* : \beta)\} \\
> \;& \max\{L_H(p_H^*, b^* : \tilde{\beta}), L_L(p_H^*, b^* : \tilde{\beta})\} \\
\geq \;& \max\{L_H(p_H^*, b^* : \tilde{\beta}), L_L(p_L^*, b^* : \tilde{\beta}) + L_H(p_H^*, b^* : \tilde{\beta}) - L_H(p_L^*, b^* : \tilde{\beta})\} \\
= \;& \tilde{X}_0(p_H^*, p_L^*, b^* : \tilde{\beta})
\end{aligned}$$

for $\tilde{\beta} > \beta$. The first inequality holds because $L_i(p_H^*, b^* : \tilde{\beta})$ is decreasing in $\tilde{\beta}$ given $p_H^* > b^*$. The second inequality follows from

$$L_L(p_H^*, b^* : \tilde{\beta}) - L_L(p_L^*, b^* : \tilde{\beta}) - [L_H(p_H^*, b^* : \tilde{\beta}) - L_H(p_L^*, b^* : \tilde{\beta})] \geq 0$$

using the argument in the second part of Step 1 combined with the property that constraint (10) is satisfied at β.

Case B: $b^* \geq p_L^*$.

Since $L_H(p_H^*, b^* : \tilde{\beta})$ is decreasing in $\tilde{\beta}$, it suffices to show that $[L_H(p_H^*, b^* : \tilde{\beta}) + L_L(p_L^*, b^* : \tilde{\beta}) - L_H(p_L^*, b^* : \tilde{\beta})]$ is also decreasing in $\tilde{\beta} \in (\beta, 1]$ when

$$X_0 = L_H(p_H^*, b^* : \tilde{\beta}) + L_L(p_L^*, b^* : \tilde{\beta}) - L_H(p_L^*, b^* : \tilde{\beta}) \geq L_H(p_H^*, b^* : \tilde{\beta}). \tag{17}$$

(17) reduces

$$L_L(p_L^*, b^* : \tilde{\beta}) \geq L_H(p_L^*, b^* : \tilde{\beta}).$$

This implies that

$$\begin{aligned}
& \frac{\partial[L_H(p_H^*, b^* : \tilde{\beta}) + L_L(p_L^*, b^* : \tilde{\beta}) - L_H(p_L^*, b^* : \tilde{\beta})]}{\partial \tilde{\beta}} \\
= \; & -\frac{1}{\tilde{\beta}^2}[F_H(p_H^*)(p_H^* - b^*) - \int_{b^*}^{p_H^*} F_H(\theta)d\theta + F_L(p_L^*)(p_L^* - b^*) - \int_{b^*}^{p_L^*} F_L(\theta)d\theta \\
- \; & F_H(p_L^*)(p_L^* - b^*) + \int_{b^*}^{p_L^*} F_H(\theta)d\theta] \\
= \; & -\frac{1}{\tilde{\beta}^2}[F_H(p_H^*)(p_H^* - b^*) - \int_{b^*}^{p_H^*} F_H(\theta)d\theta + \tilde{\beta} \int_{p_L^*}^{b^*} (F_L(\theta) - F_H(\theta))d\theta \\
+ \; & \tilde{\beta}(L_L(p_L^*, b^* : \tilde{\beta}) - L_H(p_L^*, b^* : \tilde{\beta}))] < 0.
\end{aligned}$$

This completes the proof of Step 2.

By Steps 1 and 2, when $\tilde{\beta} > \beta$, (p_L^*, p_H^*, b^*) continues to be a feasible choice and leads to a higher payoff for P. Therefore W^* is increasing in β whenever $W^* > W^{NS}$. Combining constraint (9) with the definitions of $l_i(p_i : \beta)$ and $h_i(p_i : \beta)$, it also follows that $\lim_{\beta \searrow 0} p_i^* - b^* = 0$. Therefore $\lim_{\beta \searrow 0} W^* = W^{NS}$. This completes the proof of part (2) of Proposition 3.

Proof of (1)

Now we turn to part (1). Consider a price vector with $p_H > p_L = b = p^{NS}$. Since $p^{NS} \in (\underline{\theta}, \bar{\theta})$ and $\beta > 0$,

$$p^{NS} + \beta \frac{F_H(p^{NS}) - 1}{f_H(p^{NS})} < p^{NS} < p^{NS} + \beta \frac{F_H(p^{NS})}{f_H(p^{NS})}.$$

Then (9) is satisfied for p_H sufficiently close to p^{NS}. With $b = p_L = p^{NS}$, (10) reduces to

$$L_L(p_H, p^{NS} : \beta) \geq L_H(p_H, p^{NS} : \beta).$$

This inequality is also satisfied for p_H sufficiently close to p^{NS}, since

$$L_L(p_H, p^{NS} : \beta) = L_H(p_H, p^{NS} : \beta) = 0$$

and

$$\frac{\partial L_L(p_H, p^{NS} : \beta)}{\partial p_H}|_{p_H = p^{NS}} = F_L(p^{NS}) > F_H(p^{NS}) = \frac{\partial L_H(p_H, p^{NS} : \beta)}{\partial p_H}|_{p_H = p^{NS}}.$$

(13) also reduces to $X_0 = L_H(p_H, p^{NS} : \beta)$. Thus, for p_H sufficiently close to p^{NS}, P's payoff

$$(\kappa_L F_L(p^{NS}) + \kappa_H F_H(p_H))(V - p^{NS}) - L_H(p_H, p^{NS} : \beta)$$

is achievable in extortinary collusion for $\alpha \in [0, 1/2)$. When $p_L = p_H = p^{NS}$, P's payoff equals W^{NS}. Now

$$\begin{aligned}
& \frac{\partial[(\kappa_L F_L(p^{NS}) + \kappa_H F_H(p_H))(V - p^{NS}) - L_H(p_H, p^{NS} : \beta)]}{\partial p_H}|_{p_H = p^{NS}} \\
= \; & \kappa_H f_H(p^{NS})(V - p^{NS}) - F_H(p^{NS}) \\
= \; & \kappa_H f_H(p^{NS})(H(p^{NS}) - p^{NS}) - F_H(p^{NS}) = F(p^{NS}) a_H(p^{NS}) - F_H(p^{NS}).
\end{aligned}$$

Hence if $F(p^{NS})a_H(p^{NS}) > F_H(p^{NS})$, a small increase in p_H from p^{NS}, while setting $p_L = b = p^{NS}$ as given, improves P's payoff without violating any constraint. Hence (1) is established.

Proof of (3)

Finally we turn to part (3). We begin with the following lemma which describes the optimal allocation in the case of $\beta = 1$.

Lemma 8 *Suppose $\beta = 1$. Then the optimal allocation has $b^* = p_L^*$, where (p_H^*, p_L^*) solves*

$$\max(\kappa_L F_L(p_L) + \kappa_H F_H(p_H))(V - p_L) - F_H(p_H)(p_H - p_L)$$

subject to $\max\{l_L(p_H), \underline{\theta}\} \leq p_L \leq p_H \leq \bar{\theta}$.

Proof of Lemma 8. When $\beta = 1$, (10) reduces to

$$F_L(p_H)(p_H - b) - F_L(p_L)(p_L - b) \geq F_H(p_H)(p_H - b) - F_H(p_L)(p_L - b).$$

Since we can restrict attention to $b \leq p_H$, and $F_L(p) \geq F_H(p)$, this inequality is automatically satisfied if $b \geq p_L$. Since the argument (Claim 1 in Case A) in the proof of (2) showed that P's profit is increasing in b over the range $b \leq p_L$, it follows that any b below p_L would be dominated by $b = p_L$. Hence we can confine attention to the range where $b \geq p_L$. By the argument above, constraint (10) is automatically satisfied over this range, so we can ignore it hereafter.

Note also that over the range $p_H \geq b \geq p_L$, X_0 reduces to $F_H(p_H)(p_H - b)$. Using Lemma 7, it follows that (p_L^*, p_H^*, b^*) maximizes

$$(\kappa_L F_L(p_L) + \kappa_H F_H(p_H))(V - b) - F_H(p_H)(p_H - b)$$

subject to $\max\{l_L(p_H), p_L\} \leq b \leq \min\{h_L(p_L), p_H\} \leq \bar{\theta}$, $F_L(p_L) > F_H(p_H)$ and $p_L \in (\underline{\theta}, \bar{\theta})$.

Evidently we can restrict attention to $p_L = b$, since otherwise a small increase in p_L improves P's payoff without violating any constraints. Hence the problem reduces to the maximization of

$$(\kappa_L F_L(p_L) + \kappa_H F_H(p_H))(V - p_L) - F_H(p_H)(p_H - p_L)$$

subject to $l_L(p_H) \leq p_L \leq p_H \leq \bar{\theta}$, $F_L(p_L) > F_H(p_H)$ and $p_L \in (\underline{\theta}, \bar{\theta})$. It is easy to check that we can drop the latter two constraints. ■

We invoke the following results regarding properties of the optimal weak-collusion-proof allocation (p_L^W, p_H^W, b^W) from the online Appendix[34] to Mookherjee et al. (2020):

[34] http://people.bu.edu/dilipm/publications/OnlineApp-gebrevJan2020v1.pdf

(a) S is always valuable (under the maintained assumption $V \in (\underline{\theta}, H(\bar{\theta}))$ made in this paper), (b) analogous to Lemma 5, $\max\{p_L^W, b^W\} < p_H^W$ and $p_L^W \in (\underline{\theta}, \bar{\theta})$, and (c) it maximizes $\Sigma_i \kappa_i F_i(p_i)[V - b] - X_0$ subject to (9),

$$X_0 = \max\{0, F_H(p_H)(p_H - b), F_L(p_L)(p_L - b)\} \tag{18}$$

besides $p_H \geq p_L$, $\bar{\theta} > p_L > \underline{\theta}$, $p_H \geq b$ and $F_L(p_L) > F_H(p_H)$. The main difference from the corresponding problem of finding the optimal ECP allocation is that constraint (18) is replaced by (13).

If the result is false, there must exist an allocation (p_L^W, p_H^W, b^W) which is simultaneously optimal under both weak collusion, and extortionary collusion for some value of α. Part (2) implies we can focus on the case of extortionary collusion with $\alpha = 0$. From Lemma 8, it follows that $b^W = p_L^W$.

Now we show that $b^W < p_L^W$ must hold in the optimal allocation with weak collusion, and thereby obtain a contradiction. Suppose otherwise that $b^W \geq p_L^W$. Since $p_H^W > \max\{p_L^W, b^W\}$, we have $X_0^W = F_H(p_H^W)(p_H^W - b^W) > 0 \geq F_L(p_L^W)(p_L^W - b^W)$. Then P's payoff is

$$\Sigma_i \kappa_i F_i(p_i^W)[V - b^W] - F_H(p_H^W)(p_H^W - b^W).$$

Since P achieves a payoff which is bounded below by $W^{NS} > 0$, it follows that $V > b^W$. Moreover, $\underline{\theta} < p_L^W < \bar{\theta}$ and (9) imply $l_L(p_L^W) < p_L^W \leq b^W$. Hence we can increase p_L slightly from p_L^W to $p_L^{'}$ such that $p_L^{'}$ satisfies $F_H(p_H^W)(p_H^W - b^W) > F_L(p_L^{'})(p_L^{'} - b^W)$, $l_L(p_L^{'}) \leq b^W$ and $p_L^{'} < \min\{p_H^W, \bar{\theta}\}$. Then $(p_L^{'}, p_H^W, b^W, X_0^W)$ satisfies all constraints of the P's problem in weak collusion, and generates a higher payoff for P (since $V > b^W$), a contradiction. ■

Appendix B: Divisible Good Model

Here we consider the case of divisible good with $Q \equiv \Re_+$. We also assume that $V(q)$ is assumed to be a twice continuously differentiable, increasing and strictly concave function satisfying the Inada condition ($\lim_{q\to 0} V'(q) = +\infty$ and $\lim_{q\to+\infty} V'(q) = 0$) and $V(0) = 0$. $H(\theta) \equiv \theta + \frac{F(\theta)}{f(\theta)}$ is also assumed to be strictly increasing in θ. As in Section 6, S's cost signal has two possible signal realizations η_L and η_H satisfying a Monotone Likelihood Ratio Property (MLRP) such that $a_L(\theta) \equiv a(\eta_L \mid \theta)$ is decreasing (while $a_H(\theta) \equiv a(\eta_H \mid \theta)$ is increasing) in θ. Let us use $F_i(\theta) \equiv F(\theta \mid \eta_i)$, $f_i(\theta) \equiv f(\theta \mid \eta_i)$ and $\kappa_i \equiv \int_{\underline{\theta}}^{\bar{\theta}} a_i(\theta)f(\theta)\theta$ for $i = L, H$, as in Section 6.

Our attention is restricted to the case of $\alpha \in [0, 1/2)$, since Proposition 2 is also applied to this setting. Our main result is that P can derive positive value from appointing S for $\alpha \in [0, 1/2)$, for a generic set of information structures. In this setting, our formal statement is as follows.

Proposition 4 *If there do not exist $(\rho, \nu, \gamma) \in \Re^3$ such that $a_L(\theta) = \rho + \nu F(\theta)^\gamma$ for all $\theta \in \Theta$, P can attain a strictly higher expected payoff by appointing S, compared to not appointing S.*

Let us provide rough interpretations before the formal proof of this statement. By Lemma 3, we can drop S's participation condition $E[\tilde{u}_S(\theta, \eta) \mid \eta] \geq \omega$ from problem $P^S(\alpha : \eta)$. P augments the mechanism in the manner described in Definition 1, where the auxiliary message η_0 is identified with the high-cost signal report η_H (i.e., results in the same outcomes). Hence we can confine attention to two possible signal reports η_L and η_H for A and S. If both report η_H, P selects the optimal allocation $(u_A^{NS}(\theta), u_S^{NS}(\theta), q^{NS}(\theta))$ in NS:

$$u_A^{NS}(\theta) = \int_\theta^{\bar{\theta}} \bar{q}(y)dy, u_S^{NS}(\theta) = 0$$

$$q^{NS}(\theta) = \bar{q}(\theta) \equiv \arg\max_q [V(q) - H(\theta)q],$$

The P's optimal payoff in NS is represented by $W^{NS} \equiv E[V(\bar{q}(\theta)) - H(\theta)\bar{q}(\theta)]$.

When both S and A report the low-cost signal η_L, let P select the following variation on the optimal allocation in NS. Let $\beta \equiv \frac{1-2\alpha}{1-\alpha}$, which lies in the interval

(0, 1). Let $\Lambda(\cdot) : \Theta \rightarrow \Re$ be such that (i) $\Lambda(\theta)$ is non-decreasing in θ with $\Lambda(\underline{\theta}) = 0$ and $\Lambda(\bar{\theta}) = 1$, and (ii) the function $z_\beta(\theta)$ defined by

$$z_\beta(\theta) = \theta + \beta \frac{F_L(\theta) - \Lambda(\theta)}{f_L(\theta)} \tag{19}$$

is nondecreasing. The choice of $\Lambda(.)$ is akin to choosing a set of outside option payoffs for A in the design of the side-contract. As this function takes non-negative values, A's outside option is raised. This lowers the virtual cost to the coalition of asking A to produce one more unit of the good, from $\theta + \beta \frac{F_L(\theta)}{f_L(\theta)}$ to $z_\beta(\theta)$. Hence *ceteris paribus* the coalition is induced to deliver a larger quantity of output. This adjustment enables P to control the 'double marginalization of rent' problem (wherein the coalition under-produces the good owing to S's desire to limit A's information rents).

Below we explain in further detail the exact manner in which P can select $\Lambda(.)$. Given this function and thereby the coalitional virtual cost function $z_\beta(.)$, the corresponding output schedule is set at $q(\theta, \eta_L) = \bar{q}(z_\beta(\theta))$. $z_\beta(\theta)$ exceeds or falls below θ according as $\Lambda(\theta)$ is smaller or larger than $F_L(\theta)$, implying in turn that $q(\theta, \eta_L)$ is smaller or larger than $q^{NS}(\theta)$. So (conditional on $\eta = \eta_L$) the contracted output with A will be expanded over some ranges of reported cost θ, and shrunk for other values of θ. Specifically, the output and payoffs are altered as follows:

$$q(\theta, \eta_L) = \bar{q}(z_\beta(\theta)) \tag{20}$$

$$u_A(\theta, \eta_L) = \int_\theta^{\bar{\theta}} \bar{q}(z_\beta(y))dy \tag{21}$$

$$u_S(\theta, \eta_L) = \bar{X}(z_\beta(\theta)) - \theta \bar{q}(z_\beta(\theta)) - \int_\theta^{\bar{\theta}} \bar{q}(z_\beta(y))dy \tag{22}$$

where

$$\bar{X}(z) \equiv z\bar{q}(z) + \int_z^{\bar{\theta}} \bar{q}(y)dy. \tag{23}$$

This can be interpreted as follows. P behaves 'as if' she is contracting with a single composite agent (representing the coalition) with a unit cost z of delivering the good. The coalition submits a report of (θ, η_L) which then determines a report of z given by (19) and an output order of $q(\theta, \eta_L) = \bar{q}(z_\beta(\theta))$. The corresponding total payment from P to the coalition is given by (23), so as to induce the coalition to report θ and hence z truthfully. The contract also specifies the division of this aggregate payment between A and S as per (21) and (22) to insure that A's individual incentive constraint to report θ is satisfied, with the rest going to S.

Finally, when S and A submit different reports $\eta^S \neq \eta^A$, A is offered the same allocation as in the case where the submitted η reports are η^S for both S and A, while S receives a payment equal to what he would have received if their η reports had been η^A for both S and A, minus a large penalty. This will ensure that the side contract will always involve submission of a common report by S and A.

The aim is to construct $\Lambda(\cdot)$ with the properties stated above, such that the resulting allocation is ECP and improves P's payoff in state η_L relative to the allocation resulting when S is not hired:

$$\begin{aligned} & E[V(\bar{q}(z_\beta(\theta))) - \bar{X}(z_\beta(\theta)) \mid \eta_L] \\ > \;& E[V(q^{NS}(\theta)) - \bar{X}(\theta) \mid \eta_L]. \end{aligned} \tag{24}$$

Since the allocation is unchanged in state η_H, P will achieve a higher payoff than in NS.

As shown below in the proof, such a variation is indeed ECP provided the following two conditions are satisfied:

(a)$E[u_A(\theta, \eta_L) - u_A(\theta, \eta_H) \mid \eta_H] \geq 0$

(b) $\int_{\underline{\theta}}^{\bar{\theta}} [u_A(\theta, \eta_H) - u_A(\theta, \eta_L)] d\Lambda(\theta) \geq 0.$

These two conditions are shown to jointly imply that the coalition does not benefit from manipulating A's outside option by S threatening to report $\eta_j, j \neq i$ different from the true signal η_i, if A were to refuse the offered side contract.

Conditions (a) and (b) can be rewritten as follows:

$$E[\{\bar{q}(z_\beta(\theta)) - \bar{q}(\theta)\} \frac{F_H(\theta)}{f_H(\theta)} \mid \eta_H] \geq 0 \tag{25}$$

and

$$E[(z_\beta(\theta) - h_\beta(\theta \mid \eta_L))(\bar{q}(z_\beta(\theta)) - \bar{q}(\theta)) \mid \eta_L] \geq 0 \tag{26}$$

where $h_\beta(\theta \mid \eta_i) = \theta + \beta \frac{F_i(\theta)}{f_i(\theta)}$ for $i = L, H$.[35]

Consider a small variation of the $z_\beta(\theta)$ function around the identity map $\theta^{'}(\theta) = \theta$. The corresponding point-wise variations in the left-hand-sides of (24)-(26) are as follows[36]

$$[V^{'}(\bar{q}(z))\bar{q}^{'}(z) - \bar{X}^{'}(z)]_{z=\theta} f_L(\theta) = \bar{q}^{'}(\theta)\frac{F(\theta)}{f(\theta)} f_L(\theta), \tag{27}$$

$$[\bar{q}^{'}(z)\frac{F_H(\theta)}{f_H(\theta)}]_{z=\theta} f_H(\theta) = \bar{q}^{'}(\theta) F_H(\theta), \tag{28}$$

and

$$[(\bar{q}(z) - \bar{q}(\theta)) + (z - h_\beta(\theta \mid \eta_L))\bar{q}^{'}(z)]_{z=\theta} f_L(\theta) = -\beta \bar{q}^{'}(\theta) F_L(\theta). \tag{29}$$

A sufficient condition for a variation which locally preserves the value of the left-hand-sides of (25) and (26), while increasing the value of the left-hand-side of (24), is that $\frac{F(\theta)}{f(\theta)} f_L(\theta)$ does not lie in the space spanned linearly by $F_H(\theta)$ and $-F_L(\theta)$. This is ensured by the generic property stated in Proposition 4. So the variation ends up expanding the output procured over some ranges and contracting it over others (compared to the NS allocation) when the low-cost signal η_L is reported. As is well known, the optimal allocation in NS involves 'under-procurement' owing to standard adverse selection distortions: hence expanding (resp. contracting) output procured increases (resp, decreases) P's ex post payoff. The ranges over which expansion and contraction respectively take place can be chosen to ensure that P's ex ante payoff increases (the difference being proportional to the ex ante expected value of (24) which will be positive), while ensuring that ECP conditions are preserved, i.e., (25) and (26) are preserved.

Now we provide the proof of Proposition 4.

Proof of Proposition 4

[35] We use $E[\int_{\underline{\theta}}^{\bar{\theta}} \bar{q}(y)dy \mid \eta_i] = E[\frac{F_i(\theta)}{f_i(\theta)}\bar{q}(\theta) \mid \eta_i]$ to derive these equations.

[36] We use $V^{'}(\bar{q}(z)) = H(\theta)$ to obtain (27).

We consider the specific mechanism described above; we establish this allocation is ECP provided conditions (a) and (b) are satisfied. Owing to Lemma 3, we can drop S's participation constraint (v) from problem $P^S(\alpha : \eta)$. So consider the relaxed problem denoted by $\bar{P}^S(\alpha : \eta)$, for this allocation defined on $\Theta \times \{\eta_L, \eta_H\}$, which selects $(\mu(\theta, \eta), \tilde{u}_A(\theta, \eta), p(\eta))$ to maximize

$$E[X^e(\mu(\theta, \eta)) - \theta q^e(\mu(\theta, \eta)) - \beta \tilde{u}_A(\theta, \eta) \mid \eta]$$

subject to $\mu(\theta, \eta) \in \Delta(\Theta \times \{\eta_L, \eta_H\} \cup \{e\})$ and $p(\eta) \in [0, 1]$,

$$\tilde{u}_A(\theta, \eta) \geq p(\eta) u_A(\theta, \eta) + (1 - p(\eta)) u_A(\theta, \eta^{'})$$

and

$$\tilde{u}_A(\theta, \eta) \geq \tilde{u}_A(\theta^{'}, \eta) + (\theta^{'} - \theta) q^e(\mu(\theta^{'}, \eta))$$

for any $\theta, \theta^{'} \in \Theta$.

Specifically, we aim to show that $(\mu(\theta, \eta), \tilde{u}_A(\theta, \eta), p(\eta)) = ((\theta, \eta), u_A(\theta, \eta), 1)$ solves $\bar{P}^S(\alpha : \eta)$, if

(a)$E[u_A(\theta, \eta_L) - u_A(\theta, \eta_H) \mid \eta_H] \geq 0$

(b) $\int_{\underline{\theta}}^{\bar{\theta}} [u_A(\theta, \eta_H) - u_A(\theta, \eta_L)] d\Lambda(\theta) \geq 0$.

Upon choosing $\Lambda(., \eta_L) \equiv \Lambda(.)$ and $\Lambda(., \eta_H) \equiv F(.|\eta_H)$, we can combine (a) and (b) into the following single condition

$$\int_{\underline{\theta}}^{\bar{\theta}} [u_A(\theta, \eta^{'}) - u_A(\theta, \eta)] d\Lambda(\theta, \eta) \geq 0$$

when $\eta, \eta^{'} \in \{\eta_L, \eta_H\}$ and $\eta \neq \eta^{'}$.

Since $\Lambda(\theta, \eta)$ is non-decreasing in θ, this condition implies that

$$\begin{aligned} 0 &\leq \int_{\underline{\theta}}^{\bar{\theta}} [\tilde{u}_A(\theta, \eta) - p(\eta) u_A(\theta, \eta) - (1 - p(\eta)) u_A(\theta, \eta^{'})] d\Lambda(\theta, \eta) \\ &\leq \int_{\underline{\theta}}^{\bar{\theta}} [\tilde{u}_A(\theta, \eta) - u_A(\theta, \eta)] d\Lambda(\theta, \eta) \end{aligned}$$

for any $(\tilde{u}_A(\theta,\eta), p(\eta))$ satisfying constraints of $\bar{P}^S(\alpha:\eta)$. This result can be used to obtain an upper bound of the objective function in $\bar{P}^S(\alpha:\eta)$. First note that

$$\begin{aligned}
& E[X^e(\mu(\theta,\eta)) - \theta q^e(\mu(\theta,\eta)) - \beta\tilde{u}_A(\theta,\eta) \mid \eta] \\
\leq\ & E[X^e(\mu(\theta,\eta)) - \theta q^e(\mu(\theta,\eta)) - \beta\tilde{u}_A(\theta,\eta) \mid \eta] + \beta\int_{\underline{\theta}}^{\bar{\theta}} [\tilde{u}_A(\theta,\eta) - u_A(\theta,\eta)] d\Lambda(\theta,\eta) \\
=\ & E[X^e(\mu(\theta,\eta)) - z_\beta(\theta,\eta) q^e(\mu(\theta,\eta)) \mid \eta] - \beta\int_{\underline{\theta}}^{\bar{\theta}} u_A(\theta,\eta) d\Lambda(\theta,\eta).
\end{aligned}$$

The second equality uses the fact that

$$\tilde{u}_A(\theta,\eta) = \tilde{u}_A(\bar{\theta},\eta) + \int_{\theta}^{\bar{\theta}} q^e(\tilde{\mu}(y,\eta)) dy.$$

Next, note that $\tilde{\mu} = (\theta,\eta)$ maximizes $X^e(\tilde{\mu}) - z_\beta(\theta,\eta) q^e(\tilde{\mu})$. This implies that an upper bound to the value of the objective function is given by:[37]

$$E[\bar{X}(z_\beta(\theta,\eta)) - \theta\bar{q}(z_\beta(\theta,\eta)) - \beta u_A(\theta,\eta) \mid \eta].$$

But this is attainable with $(\mu(\theta,\eta), \tilde{u}_A(\theta,\eta), p(\eta)) = ((\theta,\eta), u_A(\theta,\eta), 1)$ (which satisfies all constraints) in $\bar{P}^S(\alpha:\eta)$, implying that it is the optimal solution of this problem. This implies the allocation is ECP.

Let $Z(\eta_L)$ denote the set of non-decreasing functions $z:\Theta\rightarrow\Re$ such that $z(\theta) = \theta + \beta\frac{F_L(\theta)-\Lambda(\theta)}{f_L(\theta)}$ for some $\Lambda(\theta)$ which is non-decreasing in θ with $\Lambda(\underline{\theta}) = 0$ and $\Lambda(\bar{\theta}) = 1$. In order to prove Proposition 4, it suffices to construct $z_\beta(\cdot) \in Z(\eta_L)$ where (24), (25) and (26) are satisfied at the same time. The rest of the proof is devoted to this construction.

Step 1: Under the hypothesis of Proposition 4, there exist (λ_1, λ_2) and closed intervals on Θ ($\Theta_1 = [\underline{\theta}_1, \bar{\theta}_1]$, $\Theta_2 = [\underline{\theta}_2, \bar{\theta}_2]$ and $\Theta_3 = [\underline{\theta}_3, \bar{\theta}_3]$) such that $\underline{\theta} < \underline{\theta}_i < \bar{\theta}_i <$

[37] By definition of $(X^e(\mu), q^e(\mu))$,

$$X^e(\theta,\eta) - z_\beta(\theta,\eta) q^e(\theta,\eta) = \bar{X}(z_\beta(\theta,\eta)) - z_\beta(\theta,\eta)\bar{q}(z_\beta(\theta,\eta)).$$

$\underline{\theta}_{i+1} < \bar{\theta}_{i+1} < \bar{\theta}$ $(i = 1, 2)$, and the sign of

$$\frac{F(\theta)}{f(\theta)} f_L(\theta) + \lambda_1 F_H(\theta) - \lambda_2 F_L(\theta)$$

alternates among the interiors of Θ_1, Θ_2 and Θ_3.

Proof of Step 1

We begin with the proof of the following statement: There exists $(\lambda_1, \lambda_2) \neq 0$ such that

$$\frac{F(\theta)}{f(\theta)} f_L(\theta) + \lambda_1 F_H(\theta) - \lambda_2 F_L(\theta) = 0$$

for all $\theta \in \Theta$, if and only if if there exist $(\rho, \nu, \gamma) \in \Re^3$ such that $a_L(\theta) = \rho + \nu F(\theta)^\gamma$ for all $\theta \in \Theta$.

Proof of (If)

$a_L(\theta) = \rho + \nu F(\theta)^\gamma$ implies

$$\frac{F(\theta)}{f(\theta)} f_L(\theta) = \frac{\rho F(\theta) + \nu F(\theta)^{\gamma+1}}{\rho + \frac{\nu}{\gamma+1}},$$

$$F_L(\theta) = \frac{\rho F(\theta) + \frac{\nu}{\gamma+1} F(\theta)^{\gamma+1}}{\rho + \frac{\nu}{\gamma+1}},$$

and

$$F_H(\theta) = \frac{1}{1 - \rho - \frac{\nu}{\gamma+1}} [(1-\rho)F(\theta) - \frac{\nu}{\gamma+1} F(\theta)^{\gamma+1}].$$

Then by choosing

$$\lambda_1 = \rho\gamma \frac{1 - \rho - \frac{\nu}{\gamma+1}}{\rho + \frac{\nu}{\gamma+1}}$$

and

$$\lambda_2 = 1 + (1-\rho)\gamma,$$

we obtain

$$\frac{F(\theta)}{f(\theta)} f_L(\theta) + \lambda_1 F_H(\theta) - \lambda_2 F_L(\theta) = 0$$

for any $\theta \in \Theta$.

Proof of (Only if)

Suppose that there exists $(\lambda_1, \lambda_2) \neq 0$ such that

$$\frac{F(\theta)}{f(\theta)} f_L(\theta) + \lambda_1 F_H(\theta) - \lambda_2 F_L(\theta) = 0$$

for any $\theta \in \Theta$. Using $\frac{F(\theta)}{f(\theta)} f_L(\theta) = F(\theta) a_L(\theta)/\kappa_L$, and taking the derivative of both sides of the above equation with respect to θ, we obtain

$$\frac{F(\theta)}{\kappa_L} \frac{d a_L(\theta)}{d\theta} + \lambda_1 f_H(\theta) + (1 - \lambda_2) f_L(\theta) = 0.$$

for any θ. This can be rewritten as

$$\frac{\frac{d a_L(\theta)}{d\theta}}{(\frac{\lambda_1 \kappa_L}{\kappa_H} - (1 - \lambda_2)) a_L(\theta) - \frac{\lambda_1 \kappa_L}{\kappa_H}} = \frac{f(\theta)}{F(\theta)}.$$

Solving this differential equation, we obtain

$$a_L(\theta) = \frac{1}{(\frac{\lambda_1 \kappa_L}{\kappa_H} - (1 - \lambda_2))} [F(\theta)^{\frac{\lambda_1 \kappa_L}{\kappa_H} - (1 - \lambda_2)} C + \frac{\lambda_1 \kappa_L}{\kappa_H}].$$

for some constant C. It implies that there exists $(\rho, \nu, \gamma) \in \Re^3$ such that $a_L(\theta) = \rho + \nu F(\theta)^\gamma$. ■

Using this result, now we prove the statement in Step 1. Under the conditions of Proposition 4, there exists $(\theta_1, \theta_2, \theta_3)$ with $\underline{\theta} < \theta_1 < \theta_2 < \theta_3 < \bar{\theta}$ such that

$$A(\theta_1, \theta_2, \theta_3) \equiv \begin{pmatrix} \frac{F(\theta_1)}{f(\theta_1)} f_L(\theta_1) & F_H(\theta_1) & -F_L(\theta_1) \\ \frac{F(\theta_2)}{f(\theta_2)} f_L(\theta_2) & F_H(\theta_2) & -F_L(\theta_2) \\ \frac{F(\theta_3)}{f(\theta_3)} f_L(\theta_3) & F_H(\theta_3) & -F_L(\theta_3) \end{pmatrix}$$

is non-singular. To see this, for arbitrary θ' and θ'' ($\theta' \neq \theta''$ and $\theta', \theta'' \in (\underline{\theta}, \bar{\theta})$), consider

$$\frac{|A(\theta, \theta', \theta'')|}{|B(\theta', \theta'')|} = \frac{F(\theta)}{f(\theta)} f_L(\theta) + \lambda_1 F_H(\theta) - \lambda_2 F_L(\theta)$$

with

$$\lambda_1 \equiv -\frac{1}{|B(\theta', \theta'')|} \begin{vmatrix} \frac{F(\theta')}{f(\theta')} f_L(\theta') & -F_L(\theta') \\ \frac{F(\theta'')}{f(\theta'')} f_L(\theta'') & -F_L(\theta'') \end{vmatrix}$$

and

$$\lambda_2 \equiv -\frac{1}{|B(\theta', \theta'')|} \begin{vmatrix} \frac{F(\theta')}{f(\theta')} f_L(\theta') & F_H(\theta') \\ \frac{F(\theta'')}{f(\theta'')} f_L(\theta'') & F_H(\theta'') \end{vmatrix}$$

where

$$B(\theta', \theta'') \equiv \begin{pmatrix} F_H(\theta') & -F_L(\theta') \\ F_H(\theta'') & -F_L(\theta'') \end{pmatrix}.$$

Since $|B(\theta', \theta'')| \neq 0$ because of the monotone likelihood ratio property, the expressions above are well-defined. Our presumption and the above statement imply that we can find $\theta \neq \theta', \theta''$ ($\theta \in (\underline{\theta}, \bar{\theta})$) such that the above equation is not zero, i.e., $A(\theta, \theta', \theta'')$ is non-singular.

Next for $\bar{\theta} < \theta_1 < \theta_2 < \theta_3 < \bar{\theta}$ such that $|A(\theta_1, \theta_2, \theta_3)| \neq 0$ and for arbitrary $(b_1, b_2, b_3) \neq 0$ such that Sign $b_1 =$ Sign $b_3 \neq$ Sign b_2, consider the set of equations

$$A(\theta_1, \theta_2, \theta_3) \begin{pmatrix} \tilde{\lambda}_0 \\ \tilde{\lambda}_1 \\ \tilde{\lambda}_2 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}.$$

Since $|A(\theta_1, \theta_2, \theta_3)| \neq 0$, these equations have a unique solution for $(\tilde{\lambda}_0, \tilde{\lambda}_1, \tilde{\lambda}_2)$. Moreover we can show $\tilde{\lambda}_0 \neq 0$. Otherwise, suppose that $\tilde{\lambda}_0 = 0$. Then there must exist $(\tilde{\lambda}_1, \tilde{\lambda}_2)$ such that the sign of $\tilde{\lambda}_1 F_H(\theta) - \tilde{\lambda}_2 F_L(\theta)$ alternates between $\theta_1, \theta_2, \theta_3$. However this contradicts the monotone likelihood ratio property which states that $\frac{F_L(\theta)}{F_H(\theta)}$ is monotone in θ. So we can define $\lambda_1 \equiv \tilde{\lambda}_1 / \tilde{\lambda}_0$ and $\lambda_2 \equiv \tilde{\lambda}_2 / \tilde{\lambda}_0$, and the sign of

$$\frac{F(\theta_i)}{f(\theta_i)} f_L(\theta_i) + \lambda_1 F_H(\theta_i) - \lambda_2 F_L(\theta_i) = b_i / \tilde{\lambda}_0$$

alternates among $i = 1, 2, 3$.

By the continuity of $\frac{F(\theta)}{f(\theta)} f_L(\theta) + \lambda_1 F_H(\theta) - \lambda_2 F_L(\theta)$ for θ, we can choose closed intervals Θ_1, Θ_2 and Θ_3 ($\Theta_i \cap \Theta_{i+1} = \phi$ and $\underline{\theta} < \underline{\theta}_1 < \bar{\theta}_3 < \bar{\theta}$) such that

$$\frac{F(\theta)}{f(\theta)} f_L(\theta) + \lambda_1 F_H(\theta) - \lambda_2 F_L(\theta)$$

has the same sign as at θ_i on the interior of Θ_i ($i = 1, 2, 3$). ■

In later analysis, our focus is restricted to the case that there exists (λ_1, λ_2) such that

$$\frac{F(\theta)}{f(\theta)} f_L(\theta) + \lambda_1 F_H(\theta) - \lambda_2 F_L(\theta)$$

is negative on the interior of Θ_1 and Θ_3, and positive on the interior of Θ_2. We can adopt the same analysis for the opposite case.

Step 2: For any closed interval $[\theta', \theta''] \subset \Theta$ such that $\underline{\theta} < \theta' < \theta'' < \bar{\theta}$, there exists $\delta > 0$ so that $z(\cdot) \in Z(\eta_L)$ for any function $z(\cdot)$ satisfying the following properties:

(i) $z(\cdot)$ is increasing and differentiable with $|z(\theta) - \theta| < \delta\beta$ and $|z'(\theta) - 1| < \delta\beta$ for any $\theta \in \Theta$

(ii) $z(\theta) = \theta$ for any $\theta \notin [\theta', \theta'']$.

Proof of Step 2:

(i) and (ii) means that a function $z(\cdot)$ is sufficiently close to identity function $\hat{\theta}(\cdot)$ (with $\hat{\theta}(\theta) = \theta$) in both distance and the slope. For arbitrary closed interval $[\theta', \theta''] \subset \Theta$ such that $\underline{\theta} < \theta' < \theta'' < \bar{\theta}$, we choose ϵ_1 and ϵ_2 such that

$$\epsilon_1 \equiv \min_{\theta \in [\theta', \theta'']} f_L(\theta)$$

and

$$\epsilon_2 \equiv \max_{\theta \in [\theta', \theta'']} |f_L'(\theta)|.$$

From our assumptions that $f_L(\theta)$ is continuously differentiable and positive on Θ, $\epsilon_1 > 0$, and ϵ_2 is non-negative and bounded above. We choose $\delta > 0$ such that

$$\delta \in (0, \frac{\epsilon_1}{\epsilon_1 + \epsilon_2}).$$

For this δ, consider a function $z(\cdot)$ which satisfies the condition (i) and (ii) of the statement. Define

$$\Lambda(\theta) \equiv \frac{(\theta - z(\theta))}{\beta} f_L(\theta) + F_L(\theta).$$

Since $z(\theta)$ is differentiable on Θ, $\Lambda(\theta)$ is also so. It is equal to $\Lambda(\theta) = F_L(\theta)$ on $\theta \notin [\theta^{'}, \theta^{''}]$. For $\theta \in [\theta^{'}, \theta^{''}]$,

$$\begin{aligned}\frac{\partial \Lambda(\theta)}{\partial \theta} &= (\frac{1 - z^{'}(\theta)}{\beta} + 1) f_L(\theta) + \frac{(\theta - z(\theta))}{\beta} f_L^{'}(\theta) \\ &> (1-\delta) f_L(\theta) - \delta |f_L^{'}(\theta)| \geq (1-\delta)\epsilon_1 - \delta\epsilon_2.\end{aligned}$$

This is positive by the definition of $(\epsilon_1, \epsilon_2, \delta)$. Then $\Lambda(\theta)$ is increasing in θ on Θ with $\Lambda(\underline{\theta}) = 0$ and $\Lambda(\bar{\theta}) = 1$. Since $z(\theta)$ is increasing in θ by the definition, $z(\cdot) \in Z(\eta_L)$ by the definition of $Z(\eta_L)$. ■

Step 3: There exists $z_\beta(\cdot) \in Z(\eta_L)$ satisfying (24) – (26), and for which

$$\begin{aligned}& E[V(\bar{q}(z_\beta(\theta))) - \bar{X}(z_\beta(\theta)) \mid \eta_L] \\ > \ & E[V(\bar{q}(\theta)) - \bar{X}(\theta) \mid \eta_L].\end{aligned}$$

.

Proof of Step 3:

To simplify the notation, we use $z(\cdot)$ instead of $z_\beta(\cdot)$ in later argument. The construction of $z(\theta)$ involves the following four steps.

(a) Construction of $\bar{z}(\cdot)$

First let us define $\Phi(z, \theta)$ by

$$\begin{aligned}& \Phi(z,\theta) \equiv [H(z) - z + \frac{\lambda_2}{\beta}(z - h_\beta(\theta \mid \eta_L)) + \lambda_1 \frac{F_H(\theta)}{f_L(\theta)}]\bar{q}^{'}(z) \\ + \ & \frac{\lambda_2}{\beta}[\bar{q}(z) - \bar{q}(\theta)]\end{aligned}$$

where $h_\beta(\theta \mid \eta_L) \equiv \theta + \beta \frac{F_L(\theta)}{f_L(\theta)}$. With $z = \theta$,

$$\Phi(\theta, \theta) \equiv \frac{1}{f_L(\theta)}[\frac{F(\theta)}{f(\theta)} f_L(\theta) + \lambda_1 F_H(\theta) - \lambda_2 F_L(\theta)]\bar{q}^{'}(\theta).$$

Since $\Phi(z, \theta)$ is differentiable in z and θ, the statement in Step 1 guarantees the existence of $\bar{z}(\theta)$ such that (i) $\bar{z}(\theta)$ is differentiable on Θ, (ii) $\bar{z}(\theta) > \theta$ on $(\underline{\theta}_1, \bar{\theta}_1)$ and

$\Phi(z,\theta) > 0$ for any $z \in [\theta, \bar{z}(\theta)]$ and any $\theta \in (\underline{\theta}_1, \bar{\theta}_1)$, (iii) $\bar{z}(\theta) < \theta$ on $(\underline{\theta}_2, \bar{\theta}_2)$ and $\Phi(z,\theta) < 0$ for any $z \in [\bar{z}(\theta), \theta]$ and any $\theta \in (\underline{\theta}_2, \bar{\theta}_2)$, (iv) $\bar{z}(\theta) > \theta$ on $(\underline{\theta}_3, \bar{\theta}_3)$ and $\Phi(z,\theta) > 0$ for any $z \in [\theta, \bar{z}(\theta)]$ and any $\theta \in (\underline{\theta}_3, \bar{\theta}_3)$, and (v) $\bar{z}(\theta) = \theta$ elsewhere.

(b) Construction of $z_1(\cdot)$

For $\hat{\theta}_1 \in (\bar{\theta}_1, \underline{\theta}_2)$ and $\hat{\theta}_2 \in (\bar{\theta}_2, \underline{\theta}_3)$ (chosen arbitrary), ρ_1 and ρ_2 are defined by

$$\rho_1 \equiv \frac{F_H(\hat{\theta}_1)}{F_L(\hat{\theta}_1)}$$

and

$$\rho_2 \equiv \frac{F_H(\hat{\theta}_2)}{F_L(\hat{\theta}_2)}.$$

Then define

$$\Psi_i(z,\theta) \equiv [\frac{F_H(\theta)}{f_L(\theta)} + \frac{\rho_i}{\beta}(z - h_\beta(\theta \mid \eta_L))]\bar{q}^{'}(z) + \frac{\rho_i}{\beta}(\bar{q}(z) - \bar{q}(\theta)).$$

$z_1(\theta)$ is defined such that $\Psi_1(z_1(\theta),\theta) = 0$ is satisfied. There always exists such a $z_1(\theta)$, since for each θ, $\Psi_i(z,\theta)$ is continuous for z and is negative for $z > \max\{\theta, h_\beta(\theta \mid \eta_L) - \frac{\beta F_H(\theta)}{\rho_1 f_L(\theta)}\}$ and is positive for $z < \min\{\theta, h_\beta(\theta \mid \eta_L) - \frac{\beta F_H(\theta)}{\rho_1 f_L(\theta)}\}$. It also implies that $z_1(\theta) < h_\beta(\theta \mid \eta)$ for any θ. If there are multiple z which satisfies $\Psi_1(z,\theta) = 0$, we choose one which is the closest to θ. Then rewriting $\Psi_1(z_1(\theta),\theta) = 0$, we obtain

$$z_1(\theta) - \theta + \frac{\bar{q}(z_1(\theta)) - \bar{q}(\theta)}{\bar{q}^{'}(z_1(\theta))} = \frac{\beta F_L(\theta)}{\rho_1 f_L(\theta)}[\rho_1 - \frac{F_H(\theta)}{F_L(\theta)}].$$

Since $\frac{F_H(\theta)}{F_L(\theta)}$ is increasing in θ by the monotone likelihood ratio assumption, $z_1(\theta) > \theta$ for $\theta < \hat{\theta}_1$ and $z_1(\theta) < \theta$ for $\theta > \hat{\theta}_1$. Since $\Psi_1(\theta,\theta) > 0$ (or < 0) for $\theta < \hat{\theta}_1$ (or $\theta > \hat{\theta}_1$), $\Psi_1(z,\theta) > 0$ for any $z \in (\theta, z_1(\theta))$ and for any $\theta < \hat{\theta}_1$ and $\Psi_1(z,\theta) < 0$ for any $z \in (z_1(\theta), \theta)$ and for any $\theta > \hat{\theta}_1$. On the other hand, $\Psi_2(z,\theta)$ is positive for (θ, z) such that $z < \theta < \hat{\theta}_2$ and negative for (θ, z) such that $\hat{\theta}_2 < \theta < z$ from the definition of $\Psi_2(z,\theta)$ and θ_2. Then the argument is summarized as

- For $z \in (\theta, z_1(\theta))$, $\Psi_1(z,\theta) > 0$ for any $\theta \in \Theta_1$

- For $z \in (z_1(\theta), \theta)$, $\Psi_1(z,\theta) < 0$ and $\Psi_2(z,\theta) > 0$ for any $\theta \in \Theta_2$
- For $z > \theta$, $\Psi_2(z,\theta) < 0$ for any $\theta \in \Theta_3$.

(c) Construction of $z_2(\cdot)$

Next let us define

$$\Gamma(z,\theta) \equiv \frac{d[(z - h_\beta(\theta \mid \eta_L))(\bar{q}(z) - \bar{q}(\theta))]}{dz} = \bar{q}(z) - \bar{q}(\theta) + (z - h_\beta(\theta \mid \eta_L))\bar{q}^{'}(z).$$

$\Gamma(z,\theta) > 0$ for $z \leq \theta$ and $\Gamma(z,\theta) < 0$ at $z = h_\beta(\theta \mid \eta_L)$. Then we can choose $z_2(\theta)(> \theta)$ which is the minimum z such that $\Gamma(z,\theta) = 0$. Therefore $(z - h_\beta(\theta \mid \eta_L))(\bar{q}(z) - \bar{q}(\theta))$ is increasing in z on $z < z_2(\theta)$.

(d) Construction of $z(\cdot)$

Finally let us construct $z(\cdot)$, based on $\bar{z}(\cdot)$, $z_1(\cdot)$ and $z_2(\cdot)$. According to the procedure in Step 2, for $[\theta^{'}, \theta^{''}] = [\underline{\theta}_1, \bar{\theta}_3]$, choose $\delta > 0$. We construct $z(\theta)$ as follows:

(i) $z(\theta)$ is differentiable and increasing in θ on Θ with $|z(\theta) - \theta| < \delta\beta$ and $|z^{'}(\theta) - 1| < \delta\beta$

(ii) $z(\theta) \in (\theta, \min\{\bar{z}(\theta), z_1(\theta), z_2(\theta)\})$ on $(\underline{\theta}_1, \bar{\theta}_1)$

(iii) $z(\theta) \in (\max\{\bar{z}(\theta), z_1(\theta)\}, \theta)$ on $(\underline{\theta}_2, \bar{\theta}_2)$

(iv) $z(\theta) \in (\theta, \min\{\bar{z}(\theta), z_2(\theta)\})$ on $(\underline{\theta}_3, \bar{\theta}_3)$

(v) $z(\theta) = \theta$ elsewhere

(vi) $E[(z(\theta) - h_\beta(\theta \mid \eta_L))(\bar{q}(z(\theta)) - \bar{q}(\theta)) \mid \eta_L] = 0$

(vii) $E[(\bar{q}(\theta) - \bar{q}(z(\theta)))\frac{F_H(\theta)}{f_H(\theta)} \mid \eta_H] = 0$.

(i) implies $z(\theta) \in Z(\eta_L)$. We argue that there exists $z(\theta)$ which satisfies (i)-(vii). It is evident that there exists $z(\cdot)$ which satisfies (i)-(v). In addition, since $(z - h_\beta(\theta \mid \eta_L))(\bar{q}(z) - \bar{q}(\theta))$ is increasing in z for $z < z_2(\theta)$, $z(\theta) > \theta$ on Θ_1 and Θ_3 (or $z(\theta) < \theta$ on Θ_2) has the effect on raising (or reducing) $E[(z(\theta) - h_\beta(\theta \mid \eta_L))(\bar{q}(z(\theta)) - \bar{q}(\theta)) \mid \eta_L]$ away from zero. By balancing the two effects, $z(\cdot)$ can also satisfy (vi).

Suppose $z(\cdot)$ which satisfies (i)-(vi), but does not satisfy (vii). It is shown that we can construct a new function which satisfies all of (i)-(vii) with small adjustment of $z(\cdot)$. First we define $\tilde{z}(\cdot, \epsilon)$ ($\epsilon = (\epsilon_1, \epsilon_2, \epsilon_3)$) as $\tilde{z}(\theta, \epsilon) \equiv \theta + \epsilon_i(z(\theta) - \theta)$ on Θ_i ($i = 1, 2, 3$) and $\tilde{z}(\theta, \epsilon) = \theta$ elsewhere. It is evident that for any $\epsilon_i \in (0, 1]$ ($i = 1, 2, 3$), $\tilde{z}(\cdot, \epsilon)$ satisfies (i)-(v), since $\tilde{z}(\cdot, \epsilon)$ is closer to $\hat{\theta}(\cdot)$ than $z(\cdot)$ in both the distance and the slope. For the convenience of the exposition, define $\Pi(\epsilon_1, \epsilon_2, \epsilon_3)$ as

$$\Pi(\epsilon_1, \epsilon_2, \epsilon_3) \equiv E[(\tilde{z}(\theta, \epsilon) - h_\beta(\theta \mid \eta_L))(\bar{q}(\tilde{z}(\theta, \epsilon)) - \bar{q}(\theta)) \mid \eta_L].$$

It is evident that $\Pi(1, 1, 1) = 0$, since $z(\cdot)$ satisfies (vi), and $\Pi(0, 0, 0) = 0$. $\Pi(\epsilon_1, \epsilon_2, \epsilon_3)$ is continuous for each ϵ_i ($i = 1, 2, 3$), increasing in ϵ_1 and ϵ_3 and decreasing in ϵ_2. Then since $\Pi(1, 0, 0) > 0$ and $\Pi(1, 1, 0) < 0$, there exists $\epsilon_2^{'} \in (0, 1)$ such that $\Pi(1, \epsilon_2^{'}, 0) = 0$. Similarly since $\Pi(0, 0, 1) > 0$ and $\Pi(0, 1, 1) < 0$, there exists $\epsilon_2^{''} \in (0, 1)$ such that $\Pi(0, \epsilon_2^{''}, 1) = 0$. Define $\epsilon^{'} \equiv (1, \epsilon_2^{'}, 0)$ and $\epsilon^{''} \equiv (0, \epsilon_2^{''}, 1)$. It is shown that there exists a function $\epsilon(t)$ on $t \in [0, 1]$ such that $\epsilon(t)$ is continuous and monotonic function with $\epsilon(0) = \epsilon^{'}$ and $\epsilon(1) = \epsilon^{''}$, and $\Pi(\epsilon(t)) = 0$ for any $t \in [0, 1]$. Evidently $\epsilon(t) \neq 0$ for any $t \in [0, 1]$. Suppose the case that $\epsilon_2^{'} < \epsilon_2^{''}$. (The same argument is applied for the case of $\epsilon_2^{'} \geq \epsilon_2^{''}$, and so we omit the argument for the latter case.) We choose arbitrary continuous and monotonic functions $(\epsilon_1(t), \epsilon_2(t))$ with $(\epsilon_1(0), \epsilon_2(0)) = (1, \epsilon_2^{'})$ and $(\epsilon_1(0), \epsilon_2(0)) = (0, \epsilon_2^{''})$. $\epsilon_2(t)$ is increasing in t. Then for $t \in (0, 1)$,

$$\Pi(\epsilon_1(t), \epsilon_2(t), 0) < \Pi(1, \epsilon_2^{'}, 0) = 0 = \Pi(0, \epsilon_2^{''}, 1) < \Pi(\epsilon_1(t), \epsilon_2(t), 1).$$

It implies that there exists $\epsilon_3(t) \in (0, 1)$ such that $\Pi(\epsilon_1(t), \epsilon_2(t), \epsilon_3(t)) = 0$. The continuity of $\Pi(\epsilon), \epsilon_1(t), \epsilon_2(t)$ implies that $\epsilon_3(t)$ is continuous. For $t, t^{'} \in [0, 1]$ such that $t < t^{'}$, and for any $\epsilon_3 \in (0,1)$, $\Pi(\epsilon_1(t), \epsilon_2(t), \epsilon_3) > \Pi(\epsilon_1(t^{'}), \epsilon_2(t^{'}), \epsilon_3)$, implying that $\epsilon_3(t)$ is increasing in t. For $\epsilon^{'} \equiv (1, \epsilon_2^{'}, 0)$,

$$\begin{aligned}
& E[\frac{F_H(\theta)}{f_H(\theta)}(\bar{q}(\tilde{z}(\theta,\epsilon^{'})) - \bar{q}(\theta)) \mid \eta_H] \\
= \; & E[\frac{F_H(\theta)}{f_H(\theta)}(\bar{q}(\tilde{z}(\theta,\epsilon^{'})) - \bar{q}(\theta)) \mid \eta_H] + \frac{\rho_1}{\beta}E[(\tilde{z}(\theta,\epsilon^{'}) - h_\beta(\theta \mid \eta_L))(\bar{q}(\tilde{z}(\theta,\epsilon^{'})) - \bar{q}(\theta)) \mid \eta_L] \\
= \; & E[\int_\theta^{\tilde{z}(\theta,\epsilon^{'})} [\{\frac{F_H(\theta)}{f_L(\theta)} + \frac{\rho_1}{\beta}(z - h_\beta(\theta \mid \eta_L))\}\bar{q}^{'}(z) + \frac{\rho_1}{\beta}(\bar{q}(z) - \bar{q}(\theta))]dz \mid \eta_L] \\
= \; & E[\int_\theta^{\tilde{z}(\theta,\epsilon^{'})} \Psi_1(z,\theta)dz \mid \eta_L] > 0,
\end{aligned}$$

since $\Psi_1(z,\theta) > 0$ for any $z \in (\theta, z(\theta))$ and any $\theta \in \Theta_1$ and $\Psi_1(z,\theta) < 0$ for any $z \in (\theta + \epsilon_2^{'}(z(\theta) - \theta), \theta)$ and any $\theta \in \Theta_2$. Similarly for $\epsilon^{''} \equiv (0, \epsilon_2^{''}, 1)$.

$$\begin{aligned}
& E[\frac{F_H(\theta)}{f_H(\theta)}(\bar{q}(\tilde{z}(\theta,\epsilon^{''})) - \bar{q}(\theta)) \mid \eta_H] \\
= \; & E[\frac{F_H(\theta)}{f_H(\theta)}(\bar{q}(\tilde{z}(\theta,\epsilon^{''})) - \bar{q}(\theta)) \mid \eta_H] + \frac{\rho_2}{\beta}E[(\tilde{z}(\theta,\epsilon^{''}) - h_\beta(\theta \mid \eta_L))(\bar{q}(\tilde{z}(\theta,\epsilon^{''})) - \bar{q}(\theta)) \mid \eta_L] \\
= \; & E[\int_\theta^{\tilde{z}(\theta,\epsilon^{''})} \Psi_2(z,\theta)dz \mid \eta_L] < 0,
\end{aligned}$$

since $\Psi_2(z,\theta) > 0$ for any $z \in (\theta + \epsilon_2^{''}(z(\theta) - \theta), \theta)$ and any $\theta \in \Theta_2$ and $\Psi_2(z,\theta) < 0$ for any $z \in (\theta, z(\theta))$ and any $\theta \in \Theta_3$. Moreover $E[\frac{F_H(\theta)}{f_H(\theta)}(\bar{q}(\tilde{z}(\theta,\epsilon)) - \bar{q}(\theta)) \mid \eta_H]$ is continuous for ϵ. Therefore there exists $t \in (0,1)$ such that

$$E[\frac{F_H(\theta)}{f_H(\theta)}(\bar{q}(\tilde{z}(\theta,\epsilon(t))) - \bar{q}(\theta)) \mid \eta_H] = 0.$$

This argument implies that there exists $\epsilon \neq 0$ such that both (vi) and (vii) are satisfied under $\tilde{z}(\cdot,\epsilon)$. For this $\tilde{z}(\cdot,\epsilon)$, all conditions (i)-(vii) are satisfied.

Finally we check that under $z(\theta)$ which is constructed above, and for (λ_1, λ_2)

specified in Step 1,

$$
\begin{aligned}
& E[V(\bar{q}(z(\theta))) - \bar{X}(z(\theta)) \mid \eta_L] \\
- \quad & E[V(\bar{q}(\theta)) - \bar{X}(\theta) \mid \eta_L] \\
= \quad & E[V(\bar{q}(z(\theta))) - \bar{X}(z(\theta)) \mid \eta_L] \\
+ \quad & \frac{\lambda_2}{\beta_1} E[(z(\theta) - h_\beta(\theta \mid \eta_L))(\bar{q}(z(\theta)) - \bar{q}(\theta)) \mid \eta_L] \\
+ \quad & \lambda_1 [E[\frac{F_H(\theta)}{f_H(\theta)}(\bar{q}(z(\theta)) - \bar{q}(\theta)) \mid \eta_H] \\
- \quad & E[V(\bar{q}(\theta)) - \bar{X}(\theta) \mid \eta_L] = E[\int_\theta^{z(\theta)} \Phi(z, \theta) dz \mid \eta_L] > 0.
\end{aligned}
$$

The first equality comes from (vi) and (vii). ■

Step 3 implies that P's payoff is improved over the optimal NS, and the proof is completed. ■