Rev. Econ. Design (2017) 21:177–229
DOI 10.1007/s10058-017-0204-x

ORIGINAL PAPER

Two-agent collusion-proof implementation with correlation and arbitrage

Dawen Meng[1] · **Guoqiang Tian**[2] · **Zhe Yang**[1]

Received: 19 January 2016 / Accepted: 20 July 2017 / Published online: 1 August 2017

Abstract This paper characterizes the optimal collusion-proof mechanism in a two-agent nonlinear pricing environment. Our model allows agents to have correlated types and to reallocate their total purchases among themselves. We show that, under strongly negative correlation, the coalition will, sometimes, be torn apart at no cost. Under positive or weakly negative correlations, however, the threat of collusion forces the principal to distort allocation away from the first-best level obtained without collusion. We also show that, in contrast to the result of Laffont and Martimort (Econometrica

D. Meng: Financial supports from the National Science Foundation of China (NSFC-71301094) is gratefully acknowledged.
G. Tian: Financial supports from the National Science Foundation of China (NSFC-71371117) and the Key Laboratory of Mathematical Economics (SUFE) at Ministry of Education of China is gratefully acknowledged.
Z. Yang: Financial supports from the National Science Foundation of China (NSFC-11501349) is gratefully acknowledged.

We thank the associate editor and two anonymous referees for their constructive comments. We also thank David Martimort, Jie Zheng, Huihua Nie, Zhendong Yin, Lin Gui, seminar participants at Remin University of China, Jiangxi University of Finance and Economics, South China Normal University for helpful comments. All errors are our own.

✉ Dawen Meng
devinmeng@hotmail.com

✉ Zhe Yang
zheyang211@163.com

Guoqiang Tian
gtian@tamu.edu

1 School of Economics, Shanghai University of Finance and Economics and Key Laboratory of Mathematical Economics (SHUFE), Shanghai, China

2 Department of Economics, Texas A&M University, College Station, TX, USA

68:309–342, 2000), when the correlation is almost perfectly positive, the possibility of arbitrage prevents the principal from approaching the first-best efficiency.

Keywords Nonlinear pricing · Collusion-proof implementation · Mechanism design · Arbitrage · Correlation

JEL Classification D42 · D62 · D82

1 Introduction

A central topic of mechanism design theory concerns the ability of agents to earn information rents. Both casual observation and economic intuition suggest that possession of relevant private information confers a positive rent. However, this insight is at odds with the finding of Crémer and McLean (henceforth CM) (1985, 1988). They show that, in models having common priors supported on a fixed finite number of types, the set of priors which admit full surplus extraction (FSE) is generic.

The analysis of CM has been challenged on several grounds: their conclusion is not robust to the cases where the agents are risk averse or are protected by limited liability (Roberts 1991; Demougin and Garvie 1991), or to the case with competition among principals (Peters 2003). Heifetz and Neeman (2006) show that CM's genericity result hinges on their implicit common-knowledge assumption that each agent has a fixed finite number of types. FSE is generically impossible in both a geometric and a measure-theoretical sense when convex combination of priors is allowed. Another major critique towards CM comes from its vulnerability to collusion among agents. The intuition is simple. In the FSE mechanism, payments to and from agents depend on the reports of other agents. The agents have strong incentives to collude, especially in nearly independent environments where these payments are very large.

Collusion is a widespread and noxious phenomenon in reality. Typically, it imposes severe limits on what can be achieved by the mechanism designer, and thus it is generally regarded as a factor that reduces the principal's payoff in addition to asymmetric information. The pioneering work that studies collusion in principal-multiagent setting is due to Laffont and Martimort (hereafter LM) (1997, 2000). They offer a tractable modeling framework for analyzing the role of colluders' information asymmetry in collusion-proof mechanism design. A difference is found for independent and correlated types. In procurement/public good settings with two agents, they show that the optimal outcome can be made collusion-proof at no cost to the principal if the agents' types are independent (LM 1997), but if the types are correlated, preventing collusion entails a strict cost to the principal (LM 2000). In a duopoly model, Pouyet (2002) shows that under strongly negative correlation, the principal can prevent collusion at no cost. But he does not consider the possibility of reallocation/arbitrage.

In LM's procurement and public good settings, two agents may consume certain amount of goods in a non-excludable way. As such, there is no need and it is technologically impossible to divide the goods between them. However, in a private goods setting, e.g., in monopoly pricing problem, buyers have incentives to reallocate their total purchases obtained from the principal. Thus, the mechanism designer should make an optimal contractual response preventing the agents from (i) manipulating their reports, (ii) exchanging side transfers, and (iii) conducting arbitrage.[1] Jeon and Menicucci (hereafter JM) (2005) extend LM's model by incorporating arbitrage. They show that collusion is preventable at no cost with uncorrelated types in a nonlinear pricing model that allows collusive consumers to conduct reallocations on their initial purchases. They do not, however, consider a more interesting case where agents' types are correlated.

Che and Kim (hereafter CK) (2006) advance on these fronts by developing a more general method for collusion-proofing a mechanism. They show that agents' collusion, including both reporting manipulation and arbitrage, is harmless to the principal in a broad class of circumstances. Any payoff the principal can attain in the absence of collusion, including the second-best efficiency is attainable with uncorrelated types, and the first-best efficiency is also attainable for cases with correlated types and more than three agents.[2] Their analysis is quite general in terms of the number of colluders, the distribution of types, and the production technology. They also allow collusion to take place between a subgroup of agents rather than being pervasive. However, while they give a satisfactory answer in a broad class of environments, they leave unanswered an important question about whether collusion is harmless in the two-agent correlated-type environment.[3] It is still unknown what outcome could be implemented in a two-agent environment when types are correlated and arbitrage is allowed. We are trying to fill this gap in the present paper.

Our results depart from and contribute to the existing literature in the following aspects. Firstly, our two-agent result complements CK's work and gives a more general answer to the question whether or not collusion with both reports manipulation and arbitrage is harmful.[4] Our findings are that collusion can sometimes be prevented at no cost if correlation is strongly negative; but it always incurs a strict cost to the principal if correlation is positive or weakly negative. From the perspective of rents extraction, we extend CM's work by showing that the FSE result is immune to collusion in the environment with strongly negative correlation.

Secondly, we extend the result of LM (2000) by considering both arbitrage and negative correlation. LM (2000) characterize the collusion-proof mechanism in procurement/public good environments. It is unnecessary and impossible to split the goods between consumers. In contrast, we discuss the private good problem. Consumers could conduct arbitrage on their total purchases. Moreover, LM's model considers only positive correlation, while we consider negative correlation as well. We find that a strongly negative correlation between agents may greatly facilitate the principal's fighting against collusion.

[1] A number of contributions, notably Mookherjee and Tsumagari (2004), Dequiedt (2007) and Pavlov (2008), have noted that agents can coordinate not only on the way they play the grand mechanism, but also on their participation decisions.

[2] An additional requirement in their paper is that at least one agent has more than two types if $n = 3$.

[3] In mechanism design literature, the two-agent case is usually important and different from its multi-agent counterpart. It usually needs a separate discussion. See Maskin (1999), Moore and Repullo (1990), Dutta and Sen (1991), Danilov (1992) and Sjöström (1991), among many others, for detailed discussion.

[4] Admittedly, our result does not cover all possible cases, so it is still not a full characterization.

Lastly, we also extend the work of JM (2005). They consider information manipulation and arbitrage with only independent types. We extend their model to correlated environments, which is obvious more practically applicable. Their result could be regarded as a special case of ours when the correlation approaches zero.

The rest of this paper is organized as follows. Section 2 describes the economic environment studied and reviews as a benchmark the optimal pricing mechanism without collusion. Section 3 characterizes the coalitional incentive and no-arbitrage constraints which must be satisfied by an optimal weakly collusion-proof mechanism. Section 4 describes the collusion-proof implementation of the first-best allocation under strongly negative correlation. Section 5 characterizes the optimal collusion-proof mechanism with weak (both negative and positive) correlations. Section 6 discusses the case with an almost perfectly positive correlation. Section 7 gives conclusions.

2 The model

2.1 Preferences, information, and mechanisms

A monopolistic seller can produce any amount of homogeneous goods at a constant marginal cost c and sells the goods to two buyers whose consumptions are $q_i, i \in \{1, 2\}$. Buyer i obtains utility $\theta_i V(q_i) - t_i$ from consuming q_i units of goods and paying t_i units of money to the seller. $V(\cdot)$ is an increasing concave function with $V(0) = 0, V'(x) > 0, V''(x) < 0, \forall x > 0$, and satisfies the Inada conditions: $\lim_{x\to+\infty} V'(x) = 0, \lim_{x\to 0} V'(x) = +\infty$. A consumer privately observes his own type $\theta_i \in \Theta \equiv \{\theta_L, \theta_H\}$, with $\Delta\theta \equiv \theta_H - \theta_L$. The probabilities $p(\theta_1, \theta_2)$ of each state $(\theta_1, \theta_2) \in \Theta^2$ are common knowledge prior beliefs. For simplicity, we write $p_{LL} = p(\theta_L, \theta_L), p_{LH} = p(\theta_L, \theta_H) = p(\theta_H, \theta_L), p_{HH} = p(\theta_H, \theta_H)$. We also denote by $\rho \equiv p_{LL} p_{HH} - p_{LH}^2$ the degree of correlation between the agents' types.[5]

The monopolistic seller designs a grand sale mechanism $\mathbf{M}$ to maximize her expected profit. Considering the Revelation Principle, we can restrict our attention to a direct revelation mechanism which maps any pair of reports $(\widehat{\theta_1}, \widehat{\theta_2})$ into a combination of consumptions and payments: $\mathbf{M} = \{q_1(\widehat{\theta_1}, \widehat{\theta_2}), q_2(\widehat{\theta_1}, \widehat{\theta_2}), t_1(\widehat{\theta_1}, \widehat{\theta_2}), t_2(\widehat{\theta_1}, \widehat{\theta_2})\}$, $\forall(\widehat{\theta_1}, \widehat{\theta_2}) \in \Theta^2$.[6] We assume that buyers are ex ante identical, for notational sim-

[5] Note that $\rho \in [-1/4, 1/4]$, ρ attains its maximum at $p_{LH} = 0, p_{HH} = p_{LL} = 1/2$; it attains its minimum at $p_{LH} = 1/2, p_{HH} = p_{LL} = 0$. We refer these extreme cases, respectively, as perfectly positive and negative correlations.

[6] One may argue that allowing a stochastic grand mechanism would increase the efficiency of principal. If the utility function is of general form, $U_i(\theta_i, q_i, t_i)$, and quantities q_i is chosen among a discrete set, $\mathbb{Q} = \{Q_1, \ldots, Q_n\}$, allowing randomization/convexification does make some differences. But, remember that in our model $U_i(\theta_i, q_i, t_i) = \theta_i V(q_i) - t_i$ (θ_i and $V(q_i)$ are multiplicatively separable, $\theta_i V(q_i)$ and t_i are additively separable), and $\mathbb{Q} = [0, \infty)$ is a continuum, the stochastic grand mechanism makes no substantial difference. We assume that the grand mechanism $\mathbf{M}$ is stochastic, i.e., $\mathbf{M} = \langle (q_i, t_i) : \Theta^2 \to \Delta(\mathcal{X}) \times \mathbb{R} \rangle_{i=1}^n$, where $\mathcal{X} \equiv [0, a], a$ is a sufficiently large number. $\Delta(\mathcal{X})$ denotes the set of all probability measures supported on $\mathcal{X}$. Agent θ_i's expected payoff when he reports $\tilde{\theta}_i$, his opponent reports θ_{-i} is $\sum_{\theta_{-i}} p(\theta_i, \theta_{-i}) \left[\int_{\mathcal{X}} \theta_i V(x) dq(\tilde{\theta}_i, \theta_{-i})(x) - t_i(\tilde{\theta}_i, \theta_{-i})\right]$. Under stochastic grand mechanism, all expressions in our paper are the same except that $V(q_i(\theta_i, \theta_{-i}))$ is replaced by $\int_{\mathcal{X}} V(x) dq_i(\theta_i, \theta_{-i})(x)$. Given $V(0) = 0, V(+\infty) = +\infty$, there exists a unique $x_i^*(\theta_i, \theta_{-i})$

plicity, we focus on anonymous mechanism in which the consumption and payment of a buyer depend only on the reports and not on his identity.[7] Then we denote by t_{kl} for $k, l \in \{H, L\}$ the tax paid by an agent whose report is θ_k and the other agent's report is θ_l, and q_{kl} is defined analogously, the rent obtained by agent is denoted by $\pi_{kl} = \theta_k V(q_{kl}) - t_{kl}$. Let $\mathbf{q} = (q_{LL}, q_{LH}, q_{HL}, q_{HH}) \in \mathbb{R}^4_+$, $\mathbf{t} = (t_{LL}, t_{LH}, t_{HL}, t_{HH}) \in \mathbb{R}^4$ and $\boldsymbol{\pi} = (\pi_{LL}, \pi_{LH}, \pi_{HL}, \pi_{HH}) \in \mathbb{R}^4$ denote, respectively, the vectors of quantities, transfers and rents.

2.2 Coalition formation

Applying the methodology of LM (1997, 2000), we model the buyers' coalition formation by a side-contract, denoted by **S**, offered by a benevolent uninformed third party, whose aim is to maximize the total payoff of agents. The maximizing problem is subject to the buyers' incentive compatibility and participation constraints written with respect to the utility they obtain when the grand mechanism **M** is played non-cooperatively. We study a collusive arrangement that allows the agents (i) to collectively manipulate their reports to the principal and to exchange transfers in a budget-balanced way, and (ii) to reallocate quantities assigned by the grand contract. The timing of the overall game of contract offer and coalition formation is the following:

- *Stage* 1 Buyers learn their respective "types" $\theta_i, i = 1, 2$.
- *Stage* 2 The seller proposes a grand sale mechanism **M**. If any buyer vetoes it, all buyers get their reservation utility normalized exogenously at zero and the following stages do not occur.
- *Stage* 3 The third party proposes a side mechanism **S** to the buyers. If anyone refuses this side mechanism, **M** is played non-cooperatively. If both buyers accept **S**, they report their types to the third party who enforces manipulation of report into **M**, and commits to enforce the corresponding side transfers and reallocation within the coalition.
- *Stage* 4 Reports are sent into the grand mechanism. Quantities and payments specified in **M** are enforced. Quantities reallocation and side transfers specified in **S**, if any, are implemented.

Formally, a side mechanism **S** takes the following form:

$$\mathbf{S} = \left\{\phi(\tilde{\theta}_1, \tilde{\theta}_2), x_1(\tilde{\theta}_1, \tilde{\theta}_2, \tilde{\phi}), x_2(\tilde{\theta}_1, \tilde{\theta}_2, \tilde{\phi}), y_1(\tilde{\theta}_1, \tilde{\theta}_2), y_2(\tilde{\theta}_1, \tilde{\theta}_2)\right\}, \forall(\tilde{\theta}_1, \tilde{\theta}_2) \in \Theta^2.$$

$\tilde{\theta}_i$ is buyer i's report to the third party. $\phi(\cdot)$ is the report manipulation function which maps any pair of reports $(\tilde{\theta}_1, \tilde{\theta}_2)$ submitted by the buyers to the third-party into a pair of

Footnote 6 continued
satisfying $V(x_i^*(\theta_i, \theta_{-i})) = \int_{\mathcal{X}} V(x) dq_i(\theta_i, \theta_{-i})(x)$. Therefore, choosing an optimal random allocation (a probability measure) $q_i(\theta_i, \theta_{-i})(x)$ is equivalent to choosing a deterministic function $x_i^*(\theta_i, \theta_{-i})$.

[7] We make this anonymous/symmetric assumption for tractability reasons following the conventions of LM (2000) and JM (2005). Absent this, the principal possess more flexibilities and then achieves a surplus at least as much as under the symmetric assumption. In this sense, our main result in Proposition 3 that FSE is achievable is robust since it is obtained in the worst case for the principal.

reports to the principal. To convexify the third-party's feasible set, stochastic manipulations are allowed. Let $\tilde{\phi} \in \Theta^2$ denote an outcome of $\phi(\cdot)$. Then, $\phi(\cdot)$ specifies the probability $p^{\phi}(\tilde{\theta}_1, \tilde{\theta}_2, \tilde{\phi})$ in which the third party, after receiving reports $(\tilde{\theta}_1, \tilde{\theta}_2)$, requires the buyers to report $\tilde{\phi}$ to the principal. When $p^{\phi}(\tilde{\theta}_1, \tilde{\theta}_2, \cdot)$ is a degenerated lottery that assigns probability one to some $\tilde{\phi} \in \Theta^2$, we get a deterministic manipulation. $y_i(\tilde{\theta}_1, \tilde{\theta}_2)$ denotes the monetary transfer from the third party to buyer i. y_i does not need to depend on $\tilde{\phi}$ because of quasi linearity of a buyer's payoff in money. $x_i(\tilde{\theta}_1, \tilde{\theta}_2, \tilde{\phi})$ represents the quantity of goods buyer i receives from the third party when $\tilde{\phi}$ is reported to the seller and $(\tilde{\theta}_1, \tilde{\theta}_2)$ are reported to the third party. Such a reallocation rule maximizes the buyers' joint surplus subject to the total amount of goods being allocated to them by an incentive compatible grand mechanism.[8] Since the third party is neither a source of goods nor money, we assume that a side mechanism should satisfy the ex post budget-balance constraints for the reallocation of goods and for the side transfers, respectively

$$\sum_{i=1}^{2} x_i(\tilde{\theta}_1, \tilde{\theta}_2, \tilde{\phi}) = 0 \text{ and } \sum_{i=1}^{2} y_i(\tilde{\theta}_1, \tilde{\theta}_2) = 0, \forall(\tilde{\theta}_1, \tilde{\theta}_1) \in \Theta^2 \text{ and } \forall\tilde{\phi} \in \Theta^2.$$

Let $U^M(\theta_i)$ denote the expected payoff of θ_i in truthful equilibrium of $\mathbf{M}$. The side mechanism must guarantee to an agent a utility level at least as large as what he expects from playing non-cooperatively the grand mechanism and then getting a utility $U^M(\theta_i)$.

2.3 The optimal grand-mechanism without coalition

We consider, as a benchmark, the optimal grand-mechanism without side-contracting. Absent collusion, a mechanism $\mathbf{M} = (\boldsymbol{\pi}, \mathbf{q})$ is *feasible* if it is individually rational,

$$BIR_L\colon \quad p_{LL}\pi_{LL} + p_{LH}\pi_{LH} \geqslant 0, \tag{1}$$
$$BIR_H\colon \quad p_{LH}\pi_{HL} + p_{HH}\pi_{HH} \geqslant 0; \tag{2}$$

and incentive compatible,

$$\begin{aligned} BIC_L\colon & p_{LL}\pi_{LL} + p_{LH}\pi_{LH} \geqslant p_{LL}\pi_{HL} \\ & \quad +p_{LH}\pi_{HH} - \Delta\theta[p_{LL}V(q_{HL}) + p_{LH}V(q_{HH})], \end{aligned} \tag{3}$$
$$\begin{aligned} BIC_H\colon & p_{LH}\pi_{HL} + p_{HH}\pi_{HH} \geqslant p_{LH}\pi_{LL} + p_{HH}\pi_{LH} \\ & \quad +\Delta\theta[p_{LH}V(q_{LL}) + p_{HH}V(q_{LH})]. \end{aligned} \tag{4}$$

[8] Here we implicitly assume that buyers could only reallocate their goods for at most what they receive from the seller, i.e., $x_i(\tilde{\theta}_1, \tilde{\theta}_2, \phi) + q_i(\phi) \geqslant 0$, $\forall\phi \in \Theta^2$, $\forall(\tilde{\theta}_1, \tilde{\theta}_2) \in \Theta^2$, $\forall i = 1, 2$. For cases when both types have positive virtual valuations, $x_i(\theta_1, \theta_2, \phi) + q_i(\phi) > 0$, $i = 1, 2$ are guaranteed by the Inada conditions $V(0) = +\infty$, $V(+\infty) = 0$. For case with very small p_{LH} and thus the low-type's virtual valuation is nonpositive, i.e., $\theta_L - p_{HH}\epsilon\Delta\theta/p_{LH} \leqslant 0$, a corner solution $x_1(\theta_L, \theta_H, \phi) + q_1(\phi) = 0$ may arise.

Let $\mathcal{M} \equiv \{(\boldsymbol{\pi}, \mathbf{q})|\textit{subject to } (1) \textit{ to } (4)\}$ be the set of all feasible mechanisms. We represent the principal's payoff as $\mathbf{\Pi}(\boldsymbol{\pi}, \mathbf{q}) = 2\sum_k \sum_l p_{kl} [\theta_k V(q_{kl}) - cq_{kl} - \pi_{kl}]$. Let $\mathcal{V} \equiv \{V \in \mathbb{R}^+ | V = \mathbf{\Pi}(\boldsymbol{\pi}, \mathbf{q}), (\boldsymbol{\pi}, \mathbf{q}) \in \mathcal{M}\}$ denote the set of all implementable payoffs for the principal. Of special interest is the highest implementable payoff $\mathbf{\Pi}^{SB}(\mathbf{p}) \equiv \sup \mathcal{V}$, which is represented as a function of probability distribution $\mathbf{p} \equiv (p_{LL}, p_{HH})$ and is referred to as noncollusive optimal or the second-best payoff.

Obviously, in the complete information case, the seller could implement the first-best payoff $\mathbf{\Pi}^{FB}(\mathbf{p}) \equiv 2\sum_k \sum_l p_{kl} \left[\theta_k V(q_{kl}^{FB}) - cq_{kl}^{FB}\right]$, where q_{kl}^{FB} is given by $\theta_k V'(q_{kl}^{FB}) = c, \forall k, l \in \{H, L\}$. CM's FSE result shows that under incomplete information, the first-best payoff is still achievable (i.e., $\mathbf{\Pi}^{SB}(\mathbf{p}) = \mathbf{\Pi}^{FB}(\mathbf{p})$) if $\rho \neq 0$. They show that there is a vector of rents $\boldsymbol{\pi}$, so that $(\boldsymbol{\pi}, \mathbf{q}^{FB})$ satisfies all BICs and binding BIRs. Representing π_{LH} and π_{HL} by π_{LL} and π_{HH} from BIRs written with equalities, then substituting these expressions into BICs written with $\mathbf{q} = \mathbf{q}^{FB}$ yields

$$BIC_L'\colon \quad \frac{\pi_{HH}\rho}{p_{LH}} + \Delta\theta(p_{LH} + p_{LL})V\left(q_{HH}^{FB}\right) \geqslant 0, \tag{5}$$

$$BIC_H'\colon \quad \frac{\pi_{LL}\rho}{p_{LH}} - \Delta\theta(p_{HH} + p_{LH})V\left(q_{LL}^{FB}\right) \geqslant 0. \tag{6}$$

We denote by $\mathcal{M}^*(\mathbf{p}) \equiv \left\{(\pi_{LL}, \pi_{HH}) \in \mathbb{R}^2 | \textit{subject to } BIC_L' \textit{ and } BIC_H'\right\}$ the reduced feasible region, within which the first-best (FSE) result is achieved. It is easy to find that the first-best outcome is implementable if and only if $\rho \neq 0$, because $\mathcal{M}^*(\mathbf{p}) \neq \varnothing$ for $\rho \neq 0$ and $\mathcal{M}^*(\mathbf{p}) = \varnothing$ if $\rho = 0$.

3 The third party's optimization program

In this section, we study formally the third party's optimization problem and derive the coalitional incentive and no-arbitrage constraints which must be satisfied by an optimal collusion-proof grand mechanism.

The third-party's optimal problem is given by:

$$[\mathcal{P}_T] \max_{\phi(\cdot), x_i(\cdot), y_i(\cdot)} \sum_{(\theta_1, \theta_2) \in \Theta^2} p(\theta_1, \theta_2) \left[U^1(\theta_1) + U^2(\theta_2)\right]$$

subject to:

$$U^i(\theta_i) = \sum_{\theta_j \in \Theta} p(\theta_j|\theta_i) \left\{ \sum_{\tilde{\phi} \in \Theta^2} p^{\phi}(\theta_i, \theta_j, \tilde{\phi}) \left[\theta_i V\left(x_i\left(\theta_i, \theta_j, \tilde{\phi}\right) + q_i\left(\tilde{\phi}\right)\right) \right.\right.$$
$$\left.\left. + y_i(\theta_i, \theta_j) - t_i\left(\tilde{\phi}\right)\right] \right\}$$

for any $\theta_i \in \Theta$ and $i, j = 1, 2$ with $i \neq j$;

$$\left(BIC_i^S\right) : U^i(\theta_i) \geqslant U^i(\widetilde{\theta_i} \mid \theta_i)$$

where

$$U^i(\widetilde{\theta_i} \mid \theta_i) = \sum_{\theta_j \in \Theta} p(\theta_j|\theta_i) \left\{ \sum_{\tilde{\phi} \in \Theta^2} p^{\phi}(\tilde{\theta}_i, \theta_j, \tilde{\phi}) \left[\theta_i V \left(x_i \left(\tilde{\theta}_i, \theta_j, \tilde{\phi} \right) + q_i \left(\tilde{\phi} \right) \right) \right.\right.$$
$$\left.\left. + y_i(\tilde{\theta}_i, \theta_j) - t_i \left(\tilde{\phi} \right) \right] \right\}$$

for any $\left(\theta_i, \tilde{\theta}_i\right) \in \Theta^2$ and $i, j = 1, 2$ with $i \neq j$;

$$\left(BIR_i^S\right) : U^i(\theta_i) \geqslant U^M(\theta_i)$$

for any $\theta_i \in \Theta$ and $i = 1, 2$;

$$(BB : y)\text{: } \sum_{i=1}^{2} y_i(\theta_1, \theta_2) = 0$$
$$(BB : x)\text{: } \sum_{i=1}^{2} x_i(\theta_1, \theta_2, \widetilde{\phi}) = 0$$

for any $(\theta_1, \theta_2) \in \Theta^2$ and any $\widetilde{\phi} \in \Theta^2$.

Definition 1 A side mechanism **S** is coalition-interim-efficient with respect to an incentive-compatible grand mechanism **M** providing a reservation utility $U^M(\theta)$[9] if and only if it solves the above program $[\mathcal{P}_T]$.

Let $\mathbf{S}^0 \equiv \{\phi(\cdot) = Id(\cdot), x_1(\cdot) = x_2(\cdot) = 0, y_1(\cdot) = y_2(\cdot) = 0\}$ denote the null contract that implements no manipulation of reports, no reallocation of quantities, and no side transfers. A weakly collusion-proof mechanism is such that the third party's optimal response to it is to offer a null side mechanism.

Definition 2 An incentive-compatible grand mechanism **M** is weakly collusion-proof if and only if it is a truthtelling direct mechanism and the null side mechanism $\mathbf{S}^0$ is coalition-interim-efficient with respect to **M**.

Proposition 1 (Weak collusion-proofness principle, WCP) *Any Bayesian perfect equilibrium of the two-stage game of grand and side contract offer $\boldsymbol{M} \circ \boldsymbol{S}$ can be achieved by a weakly collusion-proof mechanism.*

Proof The proof is a straightforward adaptation of the proof of Proposition 3 of LM (2000) and hence is omitted. □

[9] Following LM (2000), we assume that, if buyer i vetoes **S**, then the other buyer still has prior beliefs about θ_i. Therefore, if we denote by $U^M(\theta_i)$ the expected payoff of a θ_i agent in the truthful equilibrium of **M**, then his reservation utility upon rejection of **S** is still $U^M(\theta_i)$ [see LM (2000) for more general analysis].

According to the WCP, to characterize the optimal weakly collusion-proof mechanism, it suffices to add the coalitional incentive constraints ($CICs$) and no-arbitrage constraint (**NAC**), under which the third party's best response is to offer a null side contract, to the principal's optimization problem. Our next proposition characterizes these additional constraints.

Proposition 2 *A symmetric Bayesian incentive compatible grand mechanism* ***M*** *is weakly collusion-proof if and only if there exists* $\epsilon \in [0, 1)$ *such that:*

- *The following coalitional incentive constraints are satisfied:*[10]

$$CIC_{LL,LH}: 2\pi_{LL} \geqslant \pi_{LH} + \pi_{HL} + 2h(\epsilon)\Delta\theta V(q_{LL}) - g(q_{LH} + q_{HL}, \epsilon) - \frac{p_{HH}\epsilon\Delta\theta V(q_{LH})}{p_{LH}} \tag{7}$$

$$CIC_{LL,HH}: \pi_{LL} \geqslant \pi_{HH} - \Delta\theta V(q_{HH}) - h(\epsilon)\Delta\theta[V(q_{HH}) - V(q_{LL})] \tag{8}$$

$$CIC_{LH,LL}: \pi_{LH} + \pi_{HL} \geqslant 2\pi_{LL} + g(2q_{LL}, \epsilon) + \frac{p_{HH}\epsilon\Delta\theta}{p_{LH}} V(q_{LH}) - 2h(\epsilon)\Delta\theta V(q_{LL}) \tag{9}$$

$$CIC_{LH,HH}: \pi_{LH} + \pi_{HL} \geqslant 2\pi_{HH} - f(2q_{HH}, \epsilon) + \frac{p_{HH}\epsilon\Delta\theta}{p_{LH}} V(q_{LH}) \tag{10}$$

$$CIC_{HH,LH}: 2\pi_{HH} \geqslant \pi_{HL} + \pi_{LH} + f(q_{LH} + q_{HL}, \epsilon) - \frac{p_{HH}\epsilon\Delta\theta}{p_{LH}} V(q_{LH}) \tag{11}$$

$$CIC_{HH,LL}: \pi_{HH} \geqslant \pi_{LL} + \Delta\theta V(q_{LL}) \tag{12}$$

where

$$f(x, \epsilon) = 2\theta_H V\left(\frac{x}{2}\right) - \max_{x_1, x_2 \geqslant 0, x_1 + x_2 = x} \left[\left(\theta_L - \frac{p_{HH}\epsilon\Delta\theta}{p_{LH}}\right) V(x_1) + \theta_H V(x_2)\right],$$

$$g(x, \epsilon) = \max_{x_1, x_2 \geqslant 0, x_1 + x_2 = x} \left[\left(\theta_L - \frac{p_{HH}\epsilon\Delta\theta}{p_{LH}}\right) V(x_1) + \theta_H V(x_2)\right] - 2\left[\theta_L - \Delta\theta h(\epsilon)\right] V\left(\frac{x}{2}\right),$$

$$h(\epsilon) = \frac{p_{LH}^2 \epsilon}{p_{LL} p_{LH} + \rho\epsilon}.$$

- *The following no-arbitrage constraint is satisfied:*

$$\boldsymbol{NAC}: q_{LH} = \varphi_1(q_{LH} + q_{HL}), q_{HL} = \varphi_2(q_{LH} + q_{HL}), \tag{13}$$

[10] Since our attention is restricted to the symmetric/anonymous grand mechanism, coalitions HL and LH are identical, so we don't need to consider constraints $CIC_{LH,HL}$ or $CIC_{HL,LH}$.

where

$$(\varphi_1(x), \varphi_2(x)) = arg \max_{x_1, x_2 \geqslant 0, x_1 + x_2 = x} \left[\left(\theta_L - \frac{p_{HH}\epsilon\Delta\theta}{p_{LH}} \right) V(x_1) + \theta_H V(x_2) \right]$$

is the optimal splitting rule within a heterogenous coalition.

- *If $\epsilon > 0$, the θ_H type's Bayesian incentive compatibility constraint BIC_H (4) is binding.*

Proof See "Appendix". □

The coalitional incentive constraints prevent the third party from manipulating the agents' reports. For instance, $CIC_{LL,LH}$ requires that a (θ_L, θ_L) coalition prefers truthtelling to reporting (θ_L, θ_H). Each coalitional incentive constraint takes into account the possibility of reallocation: if both agents report the same types to the third party, each of them receives half of the total quantities available; otherwise, the total quantities are reallocated so as to maximize the coalitional total payoff evaluated at $(\theta_H, \theta_L - p_{HH}\epsilon\Delta\theta/p_{LH})$. The symmetric assumptions $q_1(\theta_k, \theta_k) = q_2(\theta_k, \theta_k)$, for all $k \in \{H, L\}$ imply that there is no reallocation within homogenous (i.e., LL or HH) coalitions. In heterogeneous (i.e., LH) coalitions, however, the third party has an incentive to reallocate the goods bought from the seller unless **NAC** is satisfied. Therefore, the optimal weakly collusion proof mechanism maximizes the seller's expected payoff subject to constraints $BICs$, $BIRs$, $CICs$ and **NAC** [(1)–(4), (7)–(13)].

The variable ϵ in coalitional incentive constraints can be interpreted as a transaction cost of side contracting due to asymmetric information. If the θ_H type's incentive compatibility constraint is binding in the third party's program, the principal has flexibility in choosing ϵ, since S^0 is optimal for the third party if and only if it satisfies conditions $CICs$ and **NAC** for at least one $\epsilon \in [0, 1)$. An agent usually cannot fully trust and share his private information with his collusive partners, then the third party has to face the same incentive problem faced by the principal and thus a frictional transaction cost arises within their coalition. This transaction cost is a major impediment to collusive efficiency. The principal, although can not necessarily implement the first-best allocation, is able to exploit the agents' divergence and mutual distrust to increase the transaction cost of side contracting and thus tear apart their coalition. In constraints (7)–(13), true valuations are replaced by virtual valuations. For high-type, the virtual and true valuations are the same, i.e., $\theta_H^v = \theta_H$; whereas for low-type, the virtual valuation is lower: $\theta_{L,1}^v \equiv \theta_L - p_{LH}^2\epsilon\Delta\theta/(p_{LL}p_{LH} + \rho\epsilon)$ in a homogeneous (LL) coalition and $\theta_{L,2}^v \equiv \theta_L - p_{HH}\epsilon\Delta\theta/p_{LH}$ in a heterogeneous (LH) coalition.

Given LM's result that collusion incurs efficiency loss for positive correlation and JM's result that collusion is preventable at no cost for independent types, a natural question to explore is what would happen if we allow for negative correlation.

4 The case with strongly negative correlation

In this section, we will show that the principal can implement the first-best allocation under strongly negative correlation. For the first-best allocations q_{LH}^{FB} and q_{HL}^{FB} to be resistant to arbitrage, the principal needs to set $\epsilon = 0$.[11] Also, both BIR_L and BIR_H need to be binding, since no information rent can accrue to either type. We thus obtain $\pi_{LH} = -p_{LL}\pi_{LL}/p_{LH}$ and $\pi_{HL} = -p_{HH}\pi_{HH}/p_{LH}$ from binding $BIRs$, then the remaining constraints evaluated at $\mathbf{q} = \mathbf{q}^{FB}$ reduce to the following conditions.

$$BIC_L(\pi_{LL}, \pi_{HH}) \equiv \frac{\pi_{HH}\rho}{p_{LH}} + \Delta\theta\,(p_{LH} + p_{LL})\,V\left(q_{HH}^{FB}\right) \geqslant 0, \tag{14}$$

$$BIC_H(\pi_{LL}, \pi_{HH}) \equiv \frac{\pi_{LL}\rho}{p_{LH}} - \Delta\theta\,(p_{HH} + p_{LH})\,V\left(q_{LL}^{FB}\right) \geqslant 0, \tag{15}$$

$$CIC_{LL,LH}(\pi_{LL}, \pi_{HH}) \equiv (1 - p_{HH})\,\pi_{LL} + p_{HH}\pi_{HH} + p_{LH}g\left(q_{HL}^{FB} + q_{LH}^{FB}, 0\right) \geqslant 0, \tag{16}$$

$$CIC_{LL,HH}(\pi_{LL}, \pi_{HH}) \equiv \pi_{HH} - \pi_{LL} - \Delta\theta V\left(q_{HH}^{FB}\right) \leqslant 0, \tag{17}$$

$$CIC_{LH,LL}(\pi_{LL}, \pi_{HH}) \equiv (1 - p_{HH})\pi_{LL} + p_{HH}\pi_{HH} + p_{LH}g(2q_{LL}^{FB}, 0) \leqslant 0, \tag{18}$$

$$CIC_{LH,HH}(\pi_{LL}, \pi_{HH}) \equiv p_{LL}\pi_{LL} + (1 - p_{LL})\pi_{HH} - p_{LH}f(2q_{HH}^{FB}, 0) \leqslant 0, \tag{19}$$

$$CIC_{HH,LL}(\pi_{LL}, \pi_{HH}) \equiv \pi_{HH} - \pi_{LL} - \Delta\theta V\left(q_{LL}^{FB}\right) \geqslant 0, \tag{20}$$

$$CIC_{HH,LH}(\pi_{LL}, \pi_{HH}) \equiv p_{LL}\pi_{LL} + (1 - p_{LL})\pi_{HH} - p_{LH}f(q_{LH}^{FB} + q_{HL}^{FB}, 0) \geqslant 0, \tag{21}$$

We denote by $\widehat{\mathcal{M}}(\mathbf{p}) = \left\{(\pi_{LL}, \pi_{HH}) \in \mathbb{R}^2 | \textit{subject to: } (14) \textit{ to } (21)\right\}$ the set of rent vectors which could support the first-best quantities $\mathbf{q}^{FB}$. It follows directly that the first-best outcome is achievable if and only if $\widehat{\mathcal{M}}(\mathbf{p}) \neq \varnothing$. In Fig. 1, the points within parallelogram $ABCD$ satisfy adjacent coalitional conditions $CIC_{HH,LH}$, $CIC_{LH,LL}$, $CIC_{LH,HH}$ and $CIC_{LL,LH}$.[12] The coordinates of points

[11] Remember that $\theta_L V'(q_{LH}^{FB}) = \theta_H V'(q_{HL}^{FB}) = c$ and **NAC**: $(\theta_L - p_{HH}\epsilon\Delta\theta/p_{LH})V'(q_{LH}) = \theta_H V'(q_{HL})$.

[12] If local coalitional constraints hold, then the global coalitional constraints $CIC_{LL,HH}$ and $CIC_{HH,LL}$ are automatically satisfied:

$$\begin{aligned} CIC_{HH,LL}(\pi_{LL}^A, \pi_{HH}^A) &= \frac{f(q_{LH}^{FB} + q_{HL}^{FB}, 0) + g(2q_{LL}^{FB}, 0)}{2} - \Delta\theta V(q_{LL}^{FB}) \\ &> \frac{f(2q_{LL}^{FB}, 0) + g(2q_{LL}^{FB}, 0)}{2} - \Delta\theta V(q_{LL}^{FB}) = 0 \\ CIC_{LL,HH}(\pi_{LL}^C, \pi_{HH}^C) &= \frac{g(q_{LH}^{FB} + q_{HL}^{FB}, 0) + f(2q_{HH}^{FB}, 0)}{2} - \Delta\theta V\left(q_{HH}^{FB}\right) \\ &< \frac{f(2q_{HH}^{FB}, 0) + g(2q_{HH}^{FB}, 0)}{2} - \Delta\theta V\left(q_{HH}^{FB}\right) = 0. \end{aligned}$$

These two inequalities follow from the monotonicity of functions $f(x, 0)$, $g(x, 0)$, conditions $q_{LL}^{FB} < [q_{LH}^{FB} + q_{HL}^{FB}]/2 < q_{HH}^{FB}$ and the identity $f(x, 0) + g(x, 0) \equiv 2\Delta\theta V(x/2)$.

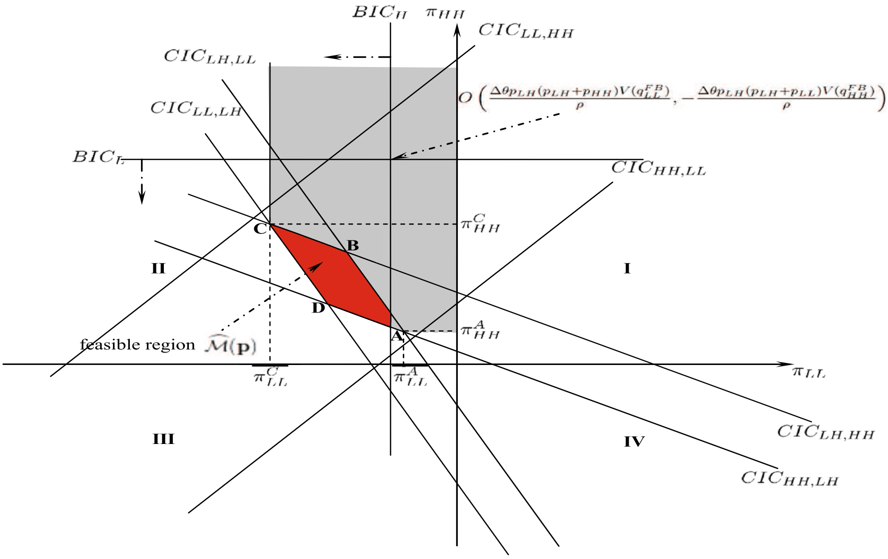


Fig. 1 Collusion-proof implementation of the first-best allocation

A to D are given as follows:[13]

$$\pi_{LL}^A = -\frac{(1-p_{LL})g\left(2q_{LL}^{FB},0\right)+p_{HH}f\left(q_{HL}^{FB}+q_{LH}^{FB},0\right)}{2} < 0,$$
$$\pi_{HH}^A = \frac{p_{LL}g\left(2q_{LL}^{FB},0\right)+(1-p_{HH})f\left(q_{HL}^{FB}+q_{LH}^{FB},0\right)}{2} > 0;$$
$$\pi_{LL}^B = -\frac{(1-p_{LL})g\left(2q_{LL}^{FB},0\right)+p_{HH}f\left(2q_{HH}^{FB},0\right)}{2} < 0,$$
$$\pi_{HH}^B = \frac{p_{LL}g\left(2q_{LL}^{FB},0\right)+(1-p_{HH})f\left(2q_{HH}^{FB},0\right)}{2} > 0;$$
$$\pi_{LL}^C = -\frac{p_{HH}f\left(2q_{HH}^{FB},0\right)+(1-p_{LL})g\left(q_{HL}^{FB}+q_{LH}^{FB},0\right)}{2} < 0$$

[13] Again, from the monotonicity of $f(x,0)$, $g(x,0)$ and inequalities $q_{LL}^{FB} < [q_{LH}^{FB}+q_{HL}^{FB}]/2 < q_{HH}^{FB}$, we have $\pi_{LL}^C < \min\{\pi_{LL}^B, \pi_{LL}^D\} < \max\{\pi_{LL}^B, \pi_{LL}^D\} < \pi_{LL}^A$ and $\pi_{HH}^A < \min\{\pi_{HH}^B, \pi_{HH}^D\} < \max\{\pi_{HH}^B, \pi_{HH}^D\} < \pi_{HH}^C$.

$$\pi_{HH}^{C} = \frac{(1-p_{HH}) f\left(2q_{HH}^{FB}, 0\right) + p_{LL} g\left(q_{HL}^{FB} + q_{LH}^{FB}, 0\right)}{2} > 0;$$
$$\pi_{LL}^{D} = -\frac{(1-p_{LL}) g\left(q_{LH}^{FB} + q_{HL}^{FB}, 0\right) + p_{HH} f\left(q_{HL}^{FB} + q_{LH}^{FB}, 0\right)}{2} < 0,$$
$$\pi_{HH}^{D} = \frac{p_{LL} g\left(q_{LH}^{FB} + q_{HL}^{FB}, 0\right) + (1-p_{HH}) f\left(q_{HL}^{FB} + q_{LH}^{FB}, 0\right)}{2} > 0.$$

To guarantee the nonemptiness of $\widehat{\mathcal{M}}(\mathbf{p})$, the intersection O of lines BIC_H and BIC_L must lie within the union of the grey and the red regions, that is:

$$\pi_{HH}^{O} \equiv -\frac{\Delta\theta(p_{LH} + p_{LL}) p_{LH} V\left(q_{HH}^{FB}\right)}{\rho} \geqslant \pi_{HH}^{A}, \tag{22}$$

$$\pi_{LL}^{O} \equiv \frac{\Delta\theta(p_{LH} + p_{HH}) p_{LH} V(q_{LL}^{FB})}{\rho} \geqslant \pi_{LL}^{C}, \tag{23}$$

$$CIC_{HH,LH}(\pi_{LL}^{O}, \pi_{HH}^{O}) \equiv p_{LL}\pi_{LL}^{O} + (1-p_{LL})\pi_{HH}^{O} - p_{LH} f(q_{LH}^{FB} + q_{HL}^{FB}, 0) \geqslant 0, \tag{24}$$

$$CIC_{LL,LH}(\pi_{LL}^{O}, \pi_{HH}^{O}) \equiv (1-p_{HH})\,\pi_{LL}^{O} + p_{HH}\pi_{HH}^{O} + p_{LH} g\left(q_{HL}^{FB} + q_{LH}^{FB}, 0\right) \geqslant 0. \tag{25}$$

To characterize the first-best implementation, we start with the following lemmas.

Lemma 1 *If* $2\Delta\theta V(q_{LL}^{FB}) \leqslant g(q_{LH}^{FB} + q_{HL}^{FB}, 0)$*, then the first-best outcome is achievable for distributions* $(p_{LL}, p_{HH}) \in \mathcal{F} \equiv \{(x, y) \in [0, 1]^2 | \rho(x, y) \leqslant \rho^*(x, y)\}$*; if* $2\Delta\theta V(q_{LL}^{FB}) > \max\{g(q_{LH}^{FB} + q_{HL}^{FB}, 0), f(2q_{HH}^{FB}, 0)\}$*, we have* $\mathcal{F} = \varnothing$*, then the first-best outcome is unachievable for any feasible distribution, where*

$$\rho(x, y) \equiv xy - \left(\frac{1-x-y}{2}\right)^2,$$
$$\rho^*(x, y) \equiv \frac{-\Delta\theta(1-x+y)(1-x-y)V(q_{LL}^{FB})}{2\left[(1-x)g(q_{LH}^{FB} + q_{HL}^{FB}, 0) + y f(2q_{HH}^{FB}, 0)\right]}.$$

Proof See "Appendix". □

The following lemma shows that the ranking between $g(q_{LH}^{FB} + q_{HL}^{FB}, 0)$ and $f(2q_{HH}^{FB}, 0)$ depends on the agents' risk attitude.

Lemma 2 $g(q_{LH}^{FB} + q_{HL}^{FB}, 0) < (resp. =, >) f(2q_{HH}^{FB}, 0)$ *if the absolute risk aversion* $r_a(x) \equiv -V''(x)/V'(x)$ *is increasing (resp. constant, decreasing) in* x.

Proof See "Appendix". □

The proof in "Appendix" shows that $g(q_{LH}^{FB} + q_{HL}^{FB}, 0)/f(2q_{HH}^{FB}, 0) = V(z^*(\xi, \theta_L))/V(z^*(\xi, \theta_H))$ for some $\xi \in (\theta_L, \theta_H)$, where $z^*(\theta_1, \theta_2) \equiv \arg\max_{z \in [0, q_{HH}^{FB} + q^*(\theta_2)]} \theta_1$

$V(z) + \theta_2 V(q_{HH}^{FB} + q^*(\theta_2) - z)$ and $q^*(\theta) \equiv \arg\max_q[\theta V(q) - cq]$. So, the comparison between $g(q_{LH}^{FB} + q_{HL}^{FB}, 0)$ and $f(2q_{HH}^{FB}, 0)$ depends on the monotonicity of function $z^*(\xi, \theta_2)$ with respect to θ_2. For fixed ξ, an increase of agent 2's valuation has two opposite effects on agent 1's allocation z^*: he will capture a smaller share of the total quantities when facing a more efficient opponent (rivalry effect); but at the same time, the total size of cake to be divided between them, i.e., $q_{HH}^{FB} + q^*(\theta_2)$, will increase (expansion effect). The net outcome hinges on the trade-off between these two counteracting effects.

With increasing absolute risk aversion, agents is inclined to take less risk as they become wealthier. This requires a more egalitarian resource distribution among them to reduce inequality and risk. The expansion effect outweighs the rivalry effect as the opponent's valuation increases. So, $\partial z^*(\xi, \theta_2)/\partial\theta_2 > 0$ and thus $g(q_{LH}^{FB} + q_{HL}^{FB}, 0) < f(2q_{HH}^{FB}, 0)$. The arguments for constant and decreasing absolute risk aversions are analogous.

The above Lemma 1 provides a sufficient condition for $\mathcal{F} \neq \varnothing$, and also a necessary condition for $\mathcal{F} \neq \varnothing$. Lemma 2 gives a condition under which these two conditions coincide. As an immediate corollary of Lemmas 1 and 2, we obtain the following result.

Proposition 3 *If the agent's preference exhibits nonincreasing absolute risk aversion, the first-best allocation is implementable for probability distributions with strongly negative correlation, i.e.,* $(p_{LL}, p_{HH}) \in \mathcal{F}$ *if and only if* $g(q_{LH}^{FB} + q_{HL}^{FB}, 0) \geqslant 2\Delta\theta V(q_{LL}^{FB})$.[14]

The result of this proposition could be depicted by Fig. 2 in (p_{LL}, p_{HH}) space. Letting X and Y denote, respectively, the intercepts of curve $\rho(p_{LL}, p_{HH}) = \rho^*(p_{LL}, p_{HH})$ with the horizontal and vertical axes, we have

$$X = \frac{g(q_{LH}^{FB} + q_{HL}^{FB}, 0) - 2\Delta\theta V(q_{LL}^{FB})}{g(q_{LH}^{FB} + q_{HL}^{FB}, 0)},$$

$$Y = \frac{2\left[g(q_{LH}^{FB} + q_{HL}^{FB}, 0) - 2\Delta\theta V(q_{LL}^{FB})\right]}{\sqrt{\begin{array}{l}\left[g(q_{LH}^{FB} + q_{HL}^{FB}, 0) + 2\Delta\theta V(q_{LL}^{FB}) - f(2q_{HH}^{FB}, 0)\right]^2 \\ +4f(2q_{HH}^{FB}, 0)\left[g(q_{LH}^{FB} + q_{HL}^{FB}, 0) - 2\Delta\theta V(q_{LL}^{FB})\right]\end{array}} + \left[\begin{array}{c} g(q_{LH}^{FB} + q_{HL}^{FB}, 0) + 2\Delta\theta V(q_{LL}^{FB}) \\ -f(2q_{HH}^{FB}, 0)\end{array}\right]}.$$

Since $g(q_{LH}^{FB} + q_{HL}^{FB}, 0) - f(2q_{HH}^{FB}, 0) \geqslant 0$ whenever $r_a(x)$ is nonincreasing, we can find easily that if $g(q_{LH}^{FB} + g_{HL}^{FB}, 0) \geqslant 2\Delta\theta V(q_{LL}^{FB})$, both X and Y are nonnegative and region $\mathcal{F}$ (the red region in Fig. 2) is nonempty, so the first-best allocation is achievable for distributions $(p_{LL}, p_{HH}) \in \mathcal{F}$. If $g(q_{LH}^{FB} + g_{HL}^{FB}, 0) < 2\Delta\theta V(q_{LL}^{FB})$, both X and

[14] Since $\rho^*(x, y)$ itself depends on probabilities (x, y), so some readers may argue that it is imprecise to interpret $\rho(x, y) \leqslant \rho^*(x, y)$ as a condition of strongly negative correlation. Please note that curve $\rho(x, y) = \rho^*(x, y)$ is not parallel to but is sandwiched by two contours $\rho(x, y) = \rho_l$ and $\rho(x, y) = \rho_h$, for some $\rho_l < \rho_h < 0$. Region $\{(x, y) \in [0, 1]^2 | \rho(x, y) \leqslant \rho^*(x, y)\}$ lies in a region of strongly negative correlation $\{(x, y) \in [0, 1]^2 | \rho(x, y) \leqslant \rho_h\}$; and it contains a region of even stronger negative correlation $\{(x, y) \in [0, 1]^2 | \rho(x, y) \leqslant \rho_l\}$. In this sense, we term the case with $\rho(x, y) \leqslant \rho^*(x, y)$ as "strongly negative correlation". The authors appreciate one referee for reminding us of this point.

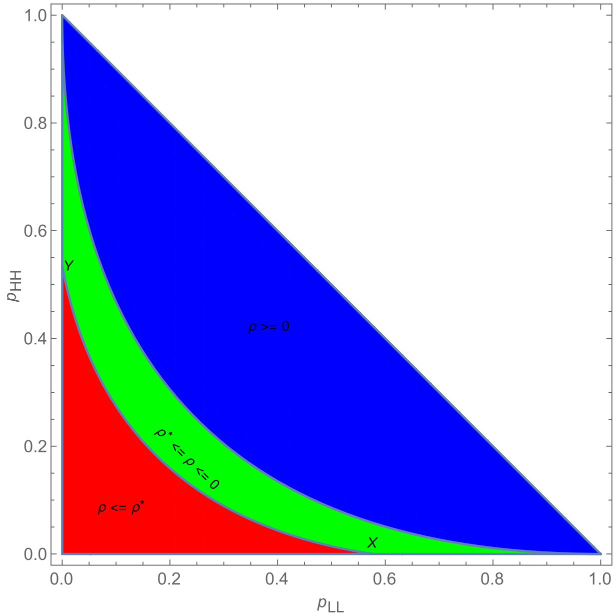


Fig. 2 Regions of probability distributions for nonincreasing absolute risk aversion

Y are negative, so region $\mathcal{F}$ vanishes, the first-best allocation is thus unachievable for any feasible distribution (p_{LL}, p_{HH}).

For utility with constant absolute risk aversion $V(x) = 1 - e^{-rx}$, expression $g(q_{LH}^{FB} + q_{HL}^{FB}, 0) \geqslant 2\Delta\theta V(q_{LL}^{FB})$ is equivalent to

$$\left(\frac{\theta_H}{\theta_L} - 1\right)\left(\frac{2c}{\theta_L r} - 1\right) + \frac{2c}{\theta_L r}\left(\sqrt{\frac{\theta_L}{\theta_H}} - 1\right) \geqslant 0.$$

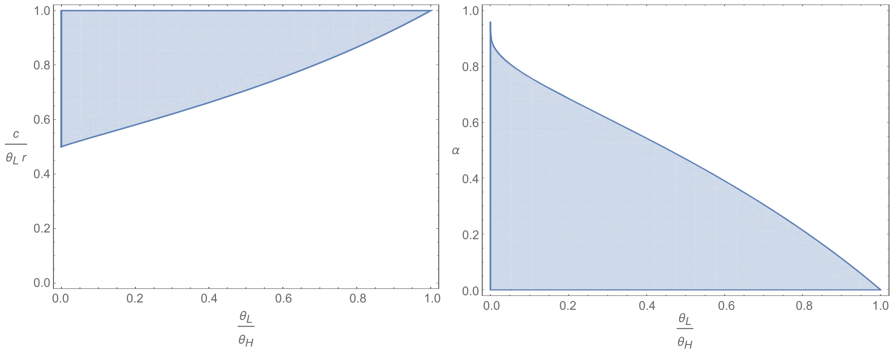
For utility with decreasing absolute risk aversion $V(x) = \frac{x^{1-\alpha}}{1-\alpha}$, $g(q_{LH}^{FB} + q_{HL}^{FB}, 0) \geqslant 2\Delta\theta V(q_{LL}^{FB})$ is equivalent to

$$\left\{\left[1 + \left(\frac{\theta_L}{\theta_H}\right)^{\frac{1}{\alpha}}\right]^{\alpha} - 2^{\alpha}\left(\frac{\theta_L}{\theta_H}\right)\right\}\left[1 + \left(\frac{\theta_L}{\theta_H}\right)^{\frac{1}{\alpha}}\right]^{1-\alpha} \geqslant 2\left(1 - \frac{\theta_L}{\theta_H}\right)\left(\frac{\theta_L}{\theta_H}\right)^{\frac{1-\alpha}{\alpha}}.$$

The region of parameters for the first-best implementation in these two cases can be depicted by the following Fig. 3.

Expression $\rho(p_{LL}, p_{HH}) = \rho^*(p_{LL}, p_{HH})$ is equivalent to

$$p_{LL}p_{HH} - \left(\frac{1 - p_{LL} - p_{HH}}{2}\right)^2 = -\frac{\frac{1}{2}\Delta\theta\left(1 - \frac{c}{r\theta_L}\right)(1 - p_{LL} - p_{HH})}{\Delta\theta - \frac{2r}{c}\left(1 - \sqrt{\frac{\theta_L}{\theta_H}}\right)}.$$



(a) **(b)**

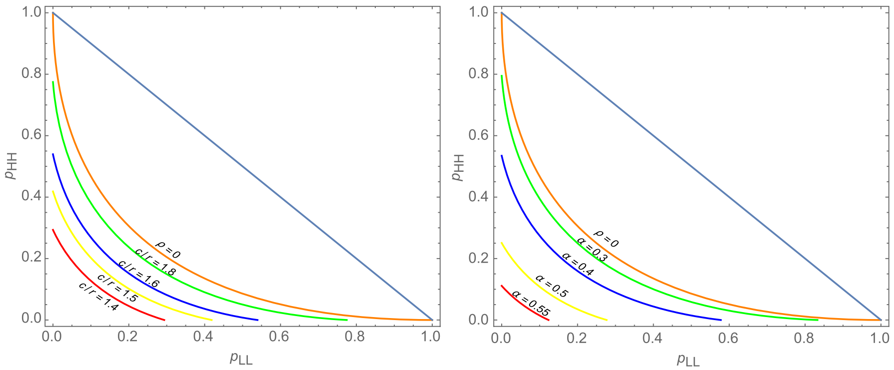
Fig. 3 The parameter regions of $g(q_{LH}^{FB} + q_{HL}^{FB}, 0) \geqslant 2\Delta\theta V(q_{LL}^{FB})$ for nonincreasing $r_a(x)$. **a** $V(x) = 1 - e^{-rx}$, **b** $V(x) = \frac{x^{1-\alpha}}{1-\alpha}$

for $V(x) = 1 - e^{-rx}$; and is equivalent to

$$p_{LL}p_{HH} - \left(\frac{1 - p_{LL} - p_{HH}}{2}\right)^2 = \frac{-(1 - \frac{\theta_L}{\theta_H})\left(\frac{\theta_L}{\theta_H}\right)^{\frac{1-\alpha}{\alpha}}(1 - p_{LL} - p_{HH})(1 - p_{LL} + p_{HH})}{2\left\{\begin{array}{c}(1 - p_{LL})\left[\left(1 + \left(\frac{\theta_L}{\theta_H}\right)^{\frac{1}{\alpha}}\right)^{\alpha} - 2^{\alpha}\left(\frac{\theta_L}{\theta_H}\right)\right]\left[1 + \left(\frac{\theta_L}{\theta_H}\right)^{\frac{1}{\alpha}}\right]^{1-\alpha} \\ +p_{HH}\left[2^{\alpha} - \left[1 + \left(\frac{\theta_L}{\theta_H}\right)^{\frac{1}{\alpha}}\right]^{\alpha}\right]2^{1-\alpha}\end{array}\right\}}$$

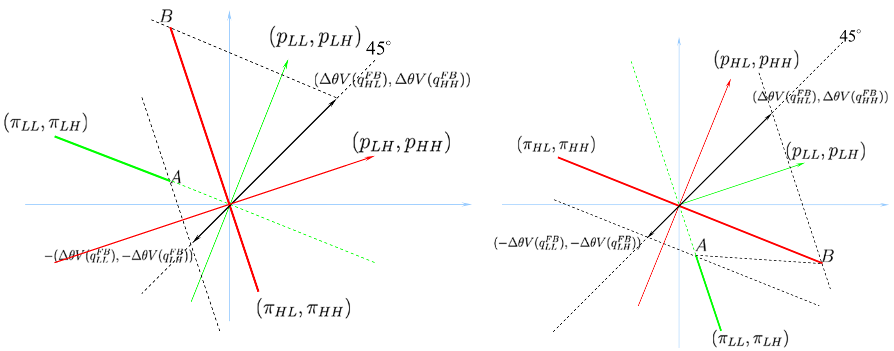
for $V(x) = x^{1-\alpha}/(1-\alpha)$. Figure 4a (resp. Fig. 4b) depicts the contours of $\rho = \rho^*$ for different values of c/r (resp. α) in (p_{LL}, p_{HH}) space with particular values $\theta_H = 6, \theta_L = 2$. It is easy to see that the larger is c/r (resp. the smaller is α), the larger is region $\mathcal{F}$, and thus the more likely will the first-best allocation be obtained.

The economic intuition behind Proposition 3 could be explained as follows. With correlated types, an agent's report contains additional information about the other agent's valuation. The mechanism designer could exploit this statistical interdependence to cross-check agents' reports, thereby inducing each agent to reveal his type truthfully without leaving any informational rent to him. Naturally, such a mechanism is not ex-post budget-balanced. The uninformed mechanism designer plays the important role of a budget-breaker. She collects transfers from the agents in some states of the world, and may also have to pay them in some other states. Figure 5a, b illustrate the penalties and rewards in setups with respective negative and positive correlations. The horizontal axis displays the first, and the vertical axis the second component of a vector. When $\rho < (resp. >)0$, (p_{LH}, p_{HH}) is flatter (resp. steeper) than (p_{LL}, p_{LH}). Vectors $(\pi_{LL}, \pi_{LH})(resp.(\pi_{HL}, \pi_{HH}))$ is perpendicular to $(p_{LL}, p_{LH})(resp.(p_{LH}, p_{HH}))$ to guarantee that a L-type (resp. H-type) agent



(a) **(b)**

Fig. 4 Contours of $\rho(p_{LL}, p_{HH}) = \rho^*(p_{LL}, p_{HH})$ with $\theta_H = 6, \theta_L = 2$. **a** $V(x) = 1 - e^{-rx}$, **b** $V(x) = \frac{x^{1-\alpha}}{1-\alpha}$



(a) **(b)**

Fig. 5 Full surplus extraction for correlated types. **a** Negative correlation, **b** positive correlation

will only get a zero expected rent. With negative (resp. positive) correlation, in order to elicit truth-telling, the principal needs to impose a penalty (resp. reward) on the agents if they both announce low types (i.e., $\pi_{LL} < (resp. >)0$). Geometrically, (π_{LL}, π_{HH}) lies to the northwest of point A in Fig. 5a, whereas it lies to the southeast of point A in Fig. 5b.

The determination of π_{LL} is the outcome of two opposing forces. First, the principal is inclined to impose a penalty on the agents when they both report θ_L to prevent their collective downward manipulations since π_{LL} is on the right-hand sides of $CIC_{LH,LL}$ and $CIC_{HH,LL}$. Second, the principal may also have incentive to impose a reward in the same state to prevent collective upward manipulations since π_{LL} is also on the left-hand sides of $CIC_{LL,LH}$ and $CIC_{LL,HH}$. As in the standard nonlinear pricing model, in order to receive information rent from the principal/seller, agents may inherently be more willing to underreport than to overreport their valuations. So, the principal is more pressured by the task of preventing downward misreport than that of preventing upward misreport. As a result, the principal will impose a punishment rather than a

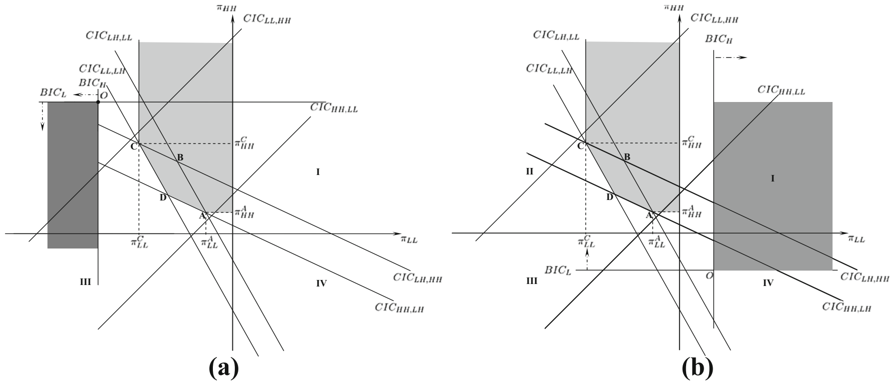


Fig. 6 Cases in which the first-best allocation is not achievable. **a** $\rho < 0$ and close enough to zero, **b** $\rho > 0$

reward on agents when they both report low-types ($\pi_{LL} \leqslant \pi_{LL}^{A} < 0$). Nevertheless, this punishment is up to the extent that the upward coalitional constraints $CIC_{LL,LH}$ and $CIC_{LH,HH}$ are binding, i.e., $\pi_{LL} \geqslant \pi_{LL}^{C}$. Analogously, a moderate reward when both agents report θ_H (i.e., $\pi_{HH} \in [\pi_{HH}^{A}, \pi_{HH}^{C}] \subset [0, +\infty)$) is required to deter them from collective manipulations.

We now come to the individual incentive constraints. With a positive correlation, the high-type agent's individual incentive constraint BIC_H requires the principal to reward both agents when they report consistent messages (θ_L, θ_L) (see Fig. 5b). As stated above, the coalitional constraints, however, requires a punishment in the same state, i.e., $\pi_{LL} < \pi_{LL}^{A} < 0$. This conflict between individual and coalitional constraints therefore prevents the principal from achieving the first-best allocation. Will the first-best allocation always be achievable for cases with negative correlation? The answer is no. As the negative correlation becomes weaker, the informativeness of one agent's report on the other's type becomes smaller. The ex post penalty and award necessary to implement the first-best allocation are both extremely large. In Fig. 5a, as the angle between (p_{LL}, p_{LH}) and (p_{HL}, p_{HH}) becomes smaller and smaller, point A goes to infinity in the north-west direction. But, conditions $CIC_{LL,LH}$ and $CIC_{LH,HH}$ together will impose a limited liability restriction on the agents which requires a finite upper bound of the level of penalty, i.e., $\pi_{LL} \geqslant \pi_{LL}^{C}$. The conflict between individual and coalitional incentive constraints thus prevents the first-best outcome from being achieved for weakly negative correlation. Geometrically, since the area $ABCD$ lies within the north-west quadrant, so $\widehat{\mathcal{M}}(\mathbf{p}) = \varnothing$ for eigher $\rho \geqslant 0$ or $\rho < 0$ but close enough to zero (see Fig. 6).

If $g(q_{LH}^{FB} + q_{HL}^{FB}, 0) < 2\Delta\theta V(q_{LL}^{FB})$, then even under perfectly negative correlation (i.e., $p_{LL} = p_{HH} = 0, \rho = -1/4$), the minimum penalty imposed on the agents for their "consistent" announcements will exceed their maximum liability imposed by the upward adjacent coalitional incentive compatibility constraints $CIC_{LH,HH}$ and $CIC_{LL,LH}$, i.e., $-\Delta\theta V(q_{LL}^{FB}) < \pi_{LL}^{C} = -g(q_{LH}^{FB} + q_{HL}^{FB}, 0)/2$. The first-best allocation is thus unattainable. Notice that, the question about the necessary and sufficient condition for the first-best implementation is still open because a grey interval $(g(q_{LH}^{FB} + q_{HL}^{FB}, 0), \max\{g(q_{LH}^{FB} + q_{HL}^{FB}, 0), f(2q_{HH}^{FB}, 0)\}]$ of $2\Delta\theta V(q_{LL}^{FB})$

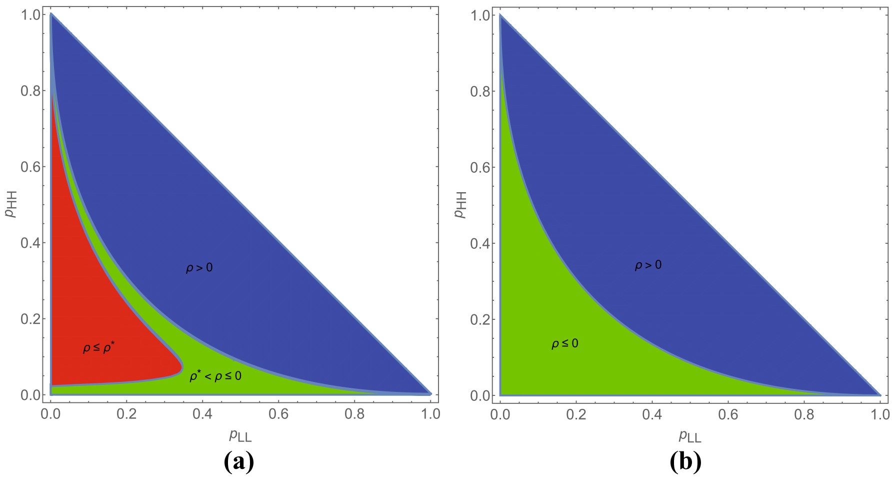


Fig. 7 $V(x) = \int_0^x e^{-5t^2} dt, \theta_H = 40, c = 1$. **a** $\mathcal{F} \neq \varnothing, \theta_L = 39, g(q_{LH}^{FB} + q_{HL}^{FB}, 0) = 0.39 < 2\Delta\theta V(q_{LL}^{FB}) = 0.79 < f(2q_{HH}^{FB}, 0) = 15.75$, **b** $\mathcal{F} = \varnothing, \theta_L = 20, g(q_{LH}^{FB} + q_{HL}^{FB}, 0) = 7.86 < 2\Delta\theta V(q_{LL}^{FB}) = 15.63 < f(2q_{HH}^{FB}, 0) = 15.71$

exists for the case with increasing absolute risk aversion. Within this interval, the nonemptiness of $\mathcal{F}$ is inconclusive. Figure 7 provides two examples, both with $g(q_{LH}^{FB} + q_{HL}^{FB}, 0) < 2\Delta\theta V(q_{LL}^{FB}) < f(2q_{HH}^{FB}, 0)$, in Fig. 7a $\mathcal{F} \neq \varnothing$, in Fig. 7b $\mathcal{F} = \varnothing$.[15]

In a related paper, Pouyet (2002) gets a similar conclusion that the first-best outcome is achievable under strongly negative correlation. However, his model does not consider the possibility of arbitrage, so it does not fit into our problem. To facilitate the comparison, we now apply Pouyet's method to a nonlinear pricing model.[16] Absent arbitrage, for the first-best allocation to be implementable, the principal will choose a parameter $\epsilon \in [0, 1)$ and a vector of rents $(\pi_{LL}, \pi_{LH}, \pi_{HL}, \pi_{HH})$ to satisfy the following conditions:[17]

$$CIC_{LL,LH}: 2\pi_{LL} \geqslant \pi_{LH} + \pi_{HL} - \Delta\theta V(q_{HL}^{FB}) - \Delta\theta h(\epsilon)\left[V(q_{HL}^{FB}) - V(q_{LL}^{FB})\right]$$

[15] Notice that in this example, function $V(x) = \int_0^x e^{-5t^2} dt$ is increasing and concave, the first order condition remains valid provided that $\theta_H > \theta_L > c$ though the Inada condition $V(0) = +\infty$ fails to hold.

[16] He considers a regulation problem of a duopoly under incomplete information.

[17] Since there is no arbitrage, $g(x_1 + x_2, \epsilon)$ in expressions (7)–(12) is replaced by

$$\theta_H V(x_1) + \left(\theta_L - \frac{p_{HH}\epsilon\Delta\theta}{p_{LH}}\right) V(x_2) - (\theta_L - h(\epsilon)\Delta\theta)\,[V(x_1) + V(x_2)],$$

and $f(x_1 + x_2, \epsilon)$ is replaced by

$$\theta_H\,[V(x_1) + V(x_2)] - \left[\theta_H V(x_1) + \left(\theta_L - \frac{p_{HH}\epsilon\Delta\theta}{p_{LH}}\right) V(x_2)\right].$$

$$CIC_{LL,HH}\colon \pi_{LL} \geqslant \pi_{HH} - \Delta\theta V\left(q_{HH}^{FB}\right) - h(\epsilon)\Delta\theta[V\left(q_{HH}^{FB}\right) - V(q_{LL}^{FB})]$$
$$CIC_{LH,LL}\colon \pi_{LH} + \pi_{HL} \geqslant 2\pi_{LL} - \Delta\theta V(q_{LL}^{FB})$$
$$CIC_{LH,HH}\ :\ \pi_{LH} + \pi_{HL} \geqslant 2\pi_{HH} - \Delta\theta V\left(q_{HH}^{FB}\right) - \frac{p_{HH}\epsilon\Delta\theta}{p_{LH}}\left[V\left(q_{HH}^{FB}\right) - V(q_{LH}^{FB})\right]$$
$$CIC_{HH,LH}\colon 2\pi_{HH} \geqslant \pi_{HL} + \pi_{LH} + \Delta\theta V(q_{LH}^{FB})$$
$$CIC_{HH,LL}\colon \pi_{HH} \geqslant \pi_{LL} + \Delta\theta V(q_{LL}^{FB})$$

Pouyet (2002) shows that if the correlation is strongly negative, i.e., $\rho + p_{LL}p_{LH} < 0$, then there exists a $\epsilon^* \in (0, 1)$ such that $\rho\epsilon^* + p_{LL}p_{LH} = 0$. The principal will choose ϵ close enough to but strictly less than ϵ^* to make $h(\epsilon)$ infinitely large. With this choice, $CIC_{LL,LH}$ and $CIC_{LL,HH}$ can be arbitrarily satisfied. Then, the principal recovers some degrees of freedom. and uses them to make both participation constraints binding and to ensure that all the other constraints are satisfied. In our setup, however, for the first-best allocation to meet the **NAC** condition, the principal needs to set $\epsilon = 0$. She is thus deprived of any flexibility of choosing ϵ. In this sense, though literally similar, our result is different from and is more striking than his.

In the following section, we will discuss the cases with weak correlations, where collusion is still detrimental to the principal.

5 The cases with weak correlations

When the first-best outcome is not implementable, the standard techniques of implementation theory suggest us to focus on the cases in which BIC_H and BIR_L are binding. The difficulty, as usual, is to determine the binding coalitional constraints. To simplify the system of constraints, it is useful to give the following implementability conditions.

Lemma 3 *For weak correlation (ρ is close enough to but is not zero), the schedule of weakly collusion-proof implementable consumptions satisfies the following monotonicity conditions:*

$$\textbf{\textit{[M]}} : q_{LL} \leqslant \frac{q_{LH} + q_{HL}}{2} \leqslant q_{HH}; \tag{26}$$

conversely, if these inequalities hold, the local coalitional incentive constraints $CIC_{LL,LH}$ (7) and $CIC_{LH,HH}$ (10) [or $CIC_{LH,LL}$ (9) and $CIC_{HH,LH}$ (11)] are binding, then all the remaining coalitional incentive constraints are indeed satisfied.

Proof See "Appendix". □

Given this result, we could focus in the sequel only on the θ_L agent's individual rationality constraint (1); the θ_H agent's Bayesian incentive constraint (4); the adjacent coalitional incentive constraints (9), (11) or (7), (10); no-arbitrage constraint (13) and the implementability condition (26). Then we can simplify the principal's problem as the following program $[\mathbf{P}_{-}^{CP}]$ or $[\mathbf{P}_{+}^{CP}]$.

$$[\mathbf{P}_{-}^{CP}]:\begin{cases}\max\limits_{\{\boldsymbol{\pi},\mathbf{q},\epsilon\in[0,1)\}} \boldsymbol{\Pi}(\boldsymbol{\pi},\mathbf{q})\\ \text{subject to:}\\ BIR_L, BIC_H, CIC_{LL,LH}, CIC_{LH,HH}, \mathbf{NAC}, \mathbf{M}\\ [(1),(4),(7),(10),(13),(26)],\end{cases}$$

$$[\mathbf{P}_{+}^{CP}]:\begin{cases}\max\limits_{\{\boldsymbol{\pi},\mathbf{q},\epsilon\in[0,1)\}} \boldsymbol{\Pi}(\boldsymbol{\pi},\mathbf{q})\\ \text{subject to:}\\ BIR_L, BIC_H, CIC_{HH,LH}, CIC_{LH,LL}, \mathbf{NAC}, \mathbf{M}\\ [(1),(4),(9),(11),(13),(26)].\end{cases}$$

Next, we present a geometric argument to show that $[\mathbf{P}_{+}^{CP}]$ and $[\mathbf{P}_{-}^{CP}]$ correspond to, respectively, the cases with weakly positive and negative correlations. Getting $\pi_{LH} = -\pi_{LL}p_{LL}/p_{LH}$ and $\pi_{HL} = [\Delta\theta p_{HH}V(q_{LH}) + \Delta\theta p_{LH}V(q_{LL}) - \pi_{HH}p_{HH}]/p_{LH} - \rho\pi_{LL}/p_{LH}^2$ from BIC_H and BIR_L written with equalities then inserting them into the remaining constraints yields:

$$BIR'_H(\pi_{LL},\pi_{HH}) \equiv \Delta\theta p_{HH}V(q_{LH}) + \Delta\theta p_{LH}V(q_{LL}) - \frac{\pi_{LL}\rho}{p_{LH}} \geqslant 0 \tag{27}$$

$$\begin{aligned}BIC'_L(\pi_{LL},\pi_{HH}) &\equiv p_{LH}\Delta\theta\left[p_{LH}^2V(q_{HH}) - p_{HH}p_{LL}V(q_{LH})\right]\\ &\quad+\Delta\theta p_{LH}^2p_{LL}[V(q_{HL}) - V(q_{LL})] + \rho p_{LH}\pi_{HH} + \rho p_{LL}\pi_{LL} \geqslant 0\end{aligned} \tag{28}$$

$$\begin{aligned}CIC'_{LL,LH}(\pi_{LL},\pi_{HH}) &\equiv p_{LH}p_{HH}\pi_{HH} + [\rho + p_{LH}(1-p_{HH})]\pi_{LL}\\ &\quad+p_{LH}^2g(q_{HL}+q_{LH},\epsilon)\\ &\quad-\Delta\theta(1-\epsilon)p_{HH}p_{LH}V(q_{LH}) - \Delta\theta p_{LH}^2[1+2h(\epsilon)]V(q_{LL}) \geqslant 0\end{aligned} \tag{29}$$

$$\begin{aligned}CIC'_{LL,HH}(\pi_{LL},\pi_{HH}) &\equiv \pi_{LL} - \pi_{HH} + \Delta\theta V(q_{HH})[h(\epsilon)+1]\\ &\quad-\Delta\theta V(q_{LL})h(\epsilon) \geqslant 0\end{aligned} \tag{30}$$

$$\begin{aligned}CIC'_{LH,LL}(\pi_{LL},\pi_{HH}) &\equiv p_{LH}p_{HH}\pi_{HH} + [\rho + p_{LH}(1-p_{HH})]\pi_{LL}\\ &\quad+g(2q_{LL},\epsilon)\,p_{LH}^2\\ &\quad-\Delta\theta(1-\epsilon)p_{HH}p_{LH}V(q_{LH}) - \Delta\theta p_{LH}^2[1+2h(\epsilon)]V(q_{LL}) \leqslant 0\end{aligned} \tag{31}$$

$$\begin{aligned}CIC'_{LH,HH}(\pi_{LL},\pi_{HH})\ equiv p_{LH}&(1-p_{LL})\pi_{HH} + (\rho + p_{LH}p_{LL})\pi_{LL}\\ &-p_{LH}^2f(2q_{HH},\epsilon) - \Delta\theta(1-\epsilon)p_{HH}p_{LH}V(q_{LH}) - \Delta\theta p_{LH}^2V(q_{LL}) \leqslant 0\end{aligned} \tag{32}$$

$$CIC'_{HH,LL}(\pi_{LL},\pi_{HH}) \equiv \pi_{HH} - \pi_{LL} - \Delta\theta V(q_{LL}) \geqslant 0 \tag{33}$$

$$\begin{aligned}CIC'_{HH,LH}(\pi_{LL},\pi_{HH}) &\equiv p_{LH}(1-p_{LL})\pi_{HH} + (\rho + p_{LH}p_{LL})\pi_{LL}\\ &\quad-p_{LH}^2f(q_{HL}+q_{LH},\epsilon)\\ &\quad-p_{HH}p_{LH}(1-\epsilon)\Delta\theta V(q_{LH}) - p_{LH}^2\Delta\theta V(q_{LL}) \geqslant 0\end{aligned} \tag{34}$$

The feasible set $\widehat{\mathcal{M}}(\mathbf{p}) \equiv \{(\pi_{LL},\pi_{HH}) \in \mathbb{R}^2 | \textit{subject to } (27) \textit{ to } (34)\}$ can be depicted by the shaded area $A'B'C'D'$ in Fig. 8. When ρ is negative and close sufficiently to zero,

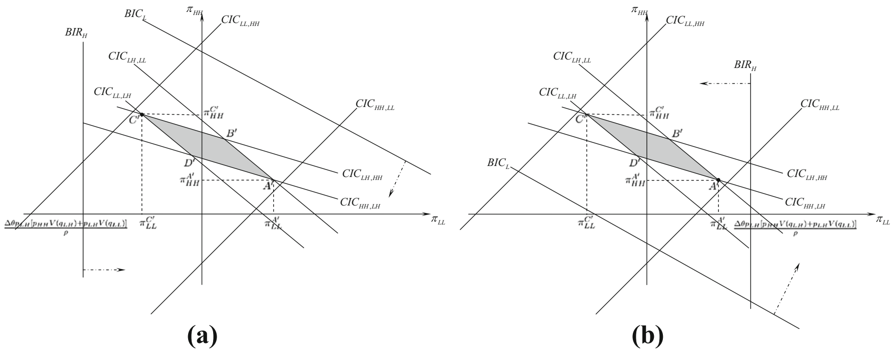


Fig. 8 The case with weak correlations. **a** $\rho < 0$, **b** $\rho > 0$

the feasible region lies to the right of line BIR_H and is under line BIC_L.[18] To minimize expected information rent conceded to the high-type agent, the principal needs to choose the leftest point C', where adjacent upward coalitional constraints $CIC_{LL,LH}$ and $CIC_{LH,HH}$ are binding. When ρ is weakly positive, however, the feasible region $A'B'C'D'$ lies to the left of line BIR_H. For the purpose of reducing information rent, the principal will obviously choose the rightest point A' at the optimum, implying binding adjacent downward coalitional constraints $CIC_{HH,LH}$ and $CIC_{LH,LL}$.

Proposition 4 *Assuming that the correlation ρ is negative and close enough to zero, $\theta_L - p_{HH}\Delta\theta/p_{LH} > 0$[19], then $\epsilon^* = 1$ is the principal's optimal choice. The optimal weakly collusion-proof mechanism entails:*

- *a monotonic schedule of consumptions represented as functions of $\boldsymbol{p}$: $q_{LL}^{CP}(\boldsymbol{p}) < [q_{LH}^{CP}(\boldsymbol{p}) + q_{HL}^{CP}(\boldsymbol{p})]/2 < q_{HH}^{CP}(\boldsymbol{p})$ given by:*

$$\left(\frac{\theta_H p_{LH}}{\rho + p_{LH}}\right) V'(q_{HH}) + \frac{\rho\theta_H V'(\varphi_2(2q_{HH}))}{\rho + p_{LH}} = c, \tag{35}$$

$$\theta_H V'(q_{HL}) - \frac{\rho(1 - p_{LL})\begin{bmatrix} \left(\theta_L - \frac{p_{HH}\Delta\theta}{p_{LH}}\right) V'(q_{LH}) \\ -\left(\theta_L - \frac{p_{LH}^2\Delta\theta}{p_{LL}p_{LH}+\rho}\right) V'\left(\frac{q_{LH}+q_{HL}}{2}\right) \end{bmatrix}}{2p_{LH}(\rho + p_{LH})} = c \tag{36}$$

[18] Notice that, though the slopes of lines representing $CICs$ change with distributions (p_{LL}, p_{HH}), for weak correlation $CIC_{LL,LH}$, $CIC_{LH,LL}$ has larger absolute slope than $CIC_{LH,HH}$, $CIC_{HH,LH}$ since

$$\frac{\rho + p_{LH}(1 - p_{HH})}{p_{LH}p_{HH}} - \frac{\rho + p_{LH}p_{LL}}{p_{LH}(1 - p_{LL})} = \frac{2(\rho + p_{LH})}{p_{HH}(1 - p_{LL})} > 0.$$

[19] This condition is imposed to avoid the tedious computation in corner solutions. If it fails, then $\varphi_1(x) = 0$, $\varphi_2(x) = x$, $\forall x$. We must have $q_{LH} = 0$, then the monotonicity condition takes the form $q_{LL} \leqslant q_{HL}/2 \leqslant q_{HH}$, this condition is very difficult to pass the ex-post check.

$$\left(\theta_L - \frac{p_{HH}\Delta\theta}{p_{LH}}\right) V'(q_{LH})$$
$$-\frac{\rho(1-p_{LL})\left[\begin{array}{c}\left(\theta_L - \frac{p_{HH}\Delta\theta}{p_{LH}}\right) V'(q_{LH}) \\ -\left(\theta_L - \frac{p_{LH}^2\Delta\theta}{p_{LL}p_{LH}+\rho}\right) V'\left(\frac{q_{LH}+q_{HL}}{2}\right)\end{array}\right]}{2p_{LH}(\rho+p_{LH})} = c \tag{37}$$
$$\left(\theta_L - \frac{\Delta\theta p_{LH}^2}{\rho + p_{LL}p_{LH}}\right) V'(q_{LL}) = c. \tag{38}$$

- *the consumptions exhibit a two-way distortion away from the first-best levels:* $q_{HH}^{CP}(\boldsymbol{p}) > q_{HH}^{FB}(\boldsymbol{p}), q_{HL}^{CP}(\boldsymbol{p}) > q_{HL}^{FB}(\boldsymbol{p}), q_{LH}^{CP}(\boldsymbol{p}) < q_{LH}^{FB}(\boldsymbol{p}), q_{LL}^{CP}(\boldsymbol{p}) < q_{LL}^{FB}(\boldsymbol{p})$.
- *a vector of rents* $\boldsymbol{\pi}^{CP} \in \mathbb{R}^4$ *such that* BIR_L, BIC_H, $CIC_{LL,LH}$, $CIC_{LH,HH}$ *are binding.*

Proof See "Appendix". □

Figure 8a provides a geometric interpretation of results in this proposition. The rents leave to the agent can be measured by the distance between point C' and line BIR_H. So for the purpose of minimizing information rents, it is optimal for the principal to move C' leftward and shift BIR_H rightward. The coordinates of points A' and C' are given as follows:

$$\pi_{LL}^{A'} = -\frac{p_{LH}\left[\begin{array}{c} p_{HH}f(q_{HL}+q_{LH},\epsilon) - (1-p_{LL})f(2q_{LL},\epsilon) \\ -2\Delta\theta p_{HH}(1-\epsilon)V(q_{LH}) + \Delta\theta(p_{LH}+p_{HH})V(q_{LL})\end{array}\right]}{2(p_{LH}+\rho)},$$
$$\pi_{HH}^{A'} = \frac{\left[\begin{array}{c}[\rho-(p_{HH}-1)p_{LH}]f(q_{HL}+q_{LH},\epsilon) - (p_{LH}p_{LL}+\rho)f(2q_{LL},\epsilon) \\ +2\Delta\theta[-p_{LL}(p_{HH}+p_{LH})V(q_{LL}) + (1-\epsilon)p_{HH}p_{LH}V(q_{LH})]\end{array}\right]}{2(p_{LH}+\rho)},$$
$$\pi_{LL}^{C'} = \frac{p_{LH}\left[\begin{array}{c}p_{HH}[2\Delta\theta(1-\epsilon)V(q_{LH}) - f(2q_{HH},\epsilon)] - (1-p_{LL})g(q_{HL}+q_{LH},\epsilon) \\ -2h(\epsilon)(p_{LL}-1)\Delta\theta V(q_{LL}) + 2\Delta\theta p_{LH}V(q_{LL})\end{array}\right]}{2(p_{LH}+\rho)},$$
$$\pi_{HH}^{C'} = \frac{\left[\begin{array}{c}[(1-p_{HH})p_{LH}+\rho]f(2q_{HH},\epsilon) + (p_{LH}p_{LL}+\rho)g(q_{HL}+q_{LH},\epsilon) \\ -2h(\epsilon)\Delta\theta V(q_{LL})(p_{LH}p_{LL}+\rho) + 2\Delta\theta p_{LH}[(1-\epsilon)p_{HH}V(q_{LH}) + p_{LH}V(q_{LL})]\end{array}\right]}{2(p_{LH}+\rho)}.$$

The sign of $\partial\pi_{LL}^{C'}/\partial q_{LL}$ and $\partial\pi_{LL}^{C'}/\partial q_{LH}$ are ambiguous, but for a weak correlation they are largely outweighed by $\partial BIR_H/\partial q_{LL} = \Delta\theta p_{LH}^2 V'(q_{LL})/\rho$ and $\partial BIR_H/\partial q_{LH} = \Delta\theta p_{LH}p_{HH}V'(q_{LH})/\rho$, which are both negative. Also, $\partial BIR_H/\partial q_{HL} = \partial BIR_H/\partial q_{HH} = 0$, $\partial\pi_{LL}^{C'}/\partial q_{HL} < 0$ and $\partial\pi_{LL}^{C'}/\partial q_{HH} < 0$. So it requires downward distortions for q_{LL} and q_{LH} and upward distortions for q_{HL} and q_{HH}.

To make things more transparent, we give a numerical example. Suppose that $V(x) = x^{1-\alpha}/(1-\alpha)$, $\theta_H = 4$, $\theta_L = 2$. The regions where the optimal mechanisms are characterized by Proposition 4 ($\mathcal{S}$) (see "Appendix" for the detailed description of

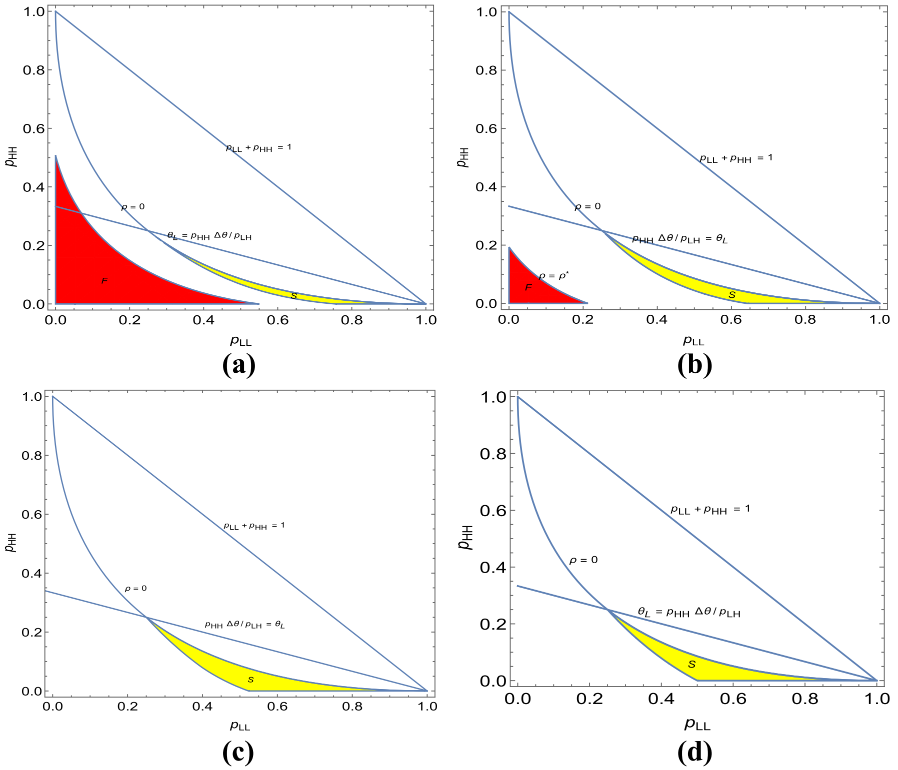


Fig. 9 $V(x) = \frac{x^{1-\alpha}}{1-\alpha}$, $\theta_H = 4$, $\theta_L = 2$ **a** $\alpha = 0.3$, **b** $\alpha = 0.4$, **c** $\alpha = 0.6$, **d** $\alpha = 0.8$

$\mathcal{S}$) and Proposition 3 ($\mathcal{F}$) are depicted in Fig. 9 for different values of α. $\mathcal{F}$ vanishes for large α (Fig. 9c, d).

Things are quite different for weakly positive correlation. We analyze this situation in the following proposition.

Proposition 5 *If the correlation ρ is positive and close sufficiently to zero, $\theta_L - p_{HH}\Delta\theta/p_{LH} > 0$, then the optimal weakly collusion-proof mechanism $\boldsymbol{M}^{CP}(\boldsymbol{p})$ entails:*

- *an $\epsilon^* \in (0, 1)$ and a monotonic schedule of consumptions: $q_{LL}^{CP}(\boldsymbol{p}) < [q_{LH}^{CP}(\boldsymbol{p}) + q_{HL}^{CP}(\boldsymbol{p})]/2 < q_{HH}^{CP}(\boldsymbol{p})$ given by*

$$\frac{\rho(1 - p_{LL})[V(\varphi_1(2q_{LL})) - V(q_{LH})]}{2(\rho + p_{LH})} + \lambda(\epsilon^*)V'(q_{LH}) = 0 \tag{39}$$

$$\left[\frac{\theta_L p_{LH} - p_{HH}\Delta\theta}{\rho + p_{LH}} + \frac{\rho\theta_H}{(\rho + p_{LH})\, p_{LL}}\right] V'(q_{LL}) - \frac{\rho(1 - p_{LL})\theta_H V'(\varphi_2(2q_{LL}))}{p_{LL}(\rho + p_{LH})} = c, \tag{40}$$

$$\left[\theta_L - \Delta\theta \frac{p_{HH}(\rho\epsilon^* + p_{LH})}{(\rho + p_{LH})p_{LH}}\right] V'(q_{LH}) - \left[\begin{array}{c} \frac{\rho p_{HH}\theta_H\left[V'\left(\frac{q_{LH}+q_{HL}}{2}\right) - V'(q_{HL})\right]}{2p_{LH}(\rho+p_{LH})} \\ +\lambda(\epsilon^*)\left(\frac{\theta_L - \frac{p_{HH}\epsilon\Delta\theta}{p_{LH}}}{p_{LH}}\right) V''(q_{LH}) \end{array}\right] = c, \tag{41}$$

$$\theta_H V'(q_{HL}) - \frac{\rho p_{HH}\theta_H \left[V'\left(\frac{q_{LH}+q_{HL}}{2}\right) - V'(q_{HL})\right]}{2p_{LH}(\rho + p_{LH})} + \lambda(\epsilon^*)\frac{\theta_H}{p_{LH}} V''(q_{HL}) = c, \tag{42}$$

$$\theta_H V'(q_{HH}) = c; \tag{43}$$

where nonnegative parameter

$$\lambda(\epsilon^*) = \frac{-\frac{\Delta\theta(1-\epsilon^*)p_{HH}}{p_{LH}+\rho}\left[c p_{LH} + \frac{\rho p_{HH}\theta_H\left(V'\left(\frac{q_{HL}+q_{LH}}{2}\right) - V'(q_{HL})\right)}{2(p_{LH}+\rho)}\right]}{\left(\theta_L - \frac{\Delta\theta\epsilon^* p_{HH}}{p_{LH}}\right)^2 V''(q_{LH}) + \theta_H\left[\theta_L - \frac{\Delta\theta p_{HH}(p_{LH}+\rho\epsilon^*)}{p_{LH}(p_{LH}+\rho)}\right] V''(q_{HL})}$$

is the Lagrangian multiplier of ***NAC*** *written with* $\epsilon = \epsilon^*$*;*

- *the consumptions except* q_{HH}^{CP} *are distorted away from their respective first-best levels:* $q_{HH}^{CP}(\boldsymbol{p}) = q_{HH}^{FB}(\boldsymbol{p}), q_{HL}^{CP}(\boldsymbol{p}) < q_{HL}^{FB}(\boldsymbol{p}), q_{LH}^{CP}(\boldsymbol{p}) < q_{LH}^{FB}(\boldsymbol{p}), q_{LL}^{CP}(\boldsymbol{p}) < q_{LL}^{FB}(\boldsymbol{p})$*;*
- *a vector of rents* $\boldsymbol{\pi}^{CP} \in \mathbb{R}^4$ *such that* BIC_H, BIR_L, $CIC_{HH,LH}$, $CIC_{LH,LL}$ *are binding.*

Proof See "Appendix". □

At this point, it is worth pausing to discuss how positive correlation differs from its negative counterpart in its influence on the principal's choice of transaction cost. Notice that two effects jointly determine the optimal ϵ^*. On the one hand, the traditional efficiency versus rent extraction tradeoff calls for a larger downward distortion of q_{LH} than that of q_{HL} relative to their respective first-best levels.[20] So, in order to meet **NAC**, the principal needs to set a larger ϵ^* to discriminate H-type from L-type. On the other hand, ϵ also enters directly into the expected information rent $\mathbb{E}\boldsymbol{\pi}$ through the binding coalitional constraints. Expressions (94) and (95) in "Appendix" show that the coalitional constraints are tightened as ϵ decreases (resp. increases) for negative (resp. positive) correlation, since

$$\frac{\partial \mathbb{E}\boldsymbol{\pi}}{\partial \epsilon} = \frac{-\rho}{2(\rho + p_{LH})}\left\{\begin{array}{c} 2(1 - p_{LL})h'(\epsilon)\Delta\theta\left[V(q_{LL}) - V\left(\frac{q_{HL}+q_{LH}}{2}\right)\right] \\ +\frac{p_{HH}^2\Delta\theta[V(q_{LH}) - V(\varphi_1(2q_{HH}))]}{p_{LH}} \end{array}\right\} < 0$$

[20] As shown in expressions of $\mathbb{E}\pi$ [(94) and (95) for respective cases with weakly negative and weakly positive correlations in the "Appendix"], if $\rho < 0$, q_{LH} affects terms β_{BIC_H}, $\beta_{CIC_{LH,HH}}$ and $\beta_{CIC_{LL,LH}}$, q_{HL} affects only $\beta_{CIC_{LL,LH}}$; if $\rho > 0$, q_{LH} affects terms β_{BIC_H}, $\beta_{CIC_{LH,LL}}$ and $\beta_{CIC_{HH,LH}}$, q_{HL} affects only $\beta_{CIC_{HH,LH}}$. Both cases require a larger distortion of q_{LH} than q_{HL}.

for $\rho < 0$ (resp. $\partial \mathbb{E}\boldsymbol{\pi}/\partial \epsilon = [\rho(1-p_{LL})p_{HH}\Delta\theta][V(q_{LH}) - V(\varphi_1(2q_{LL}))]/[2p_{LH}(\rho + p_{LH})] > 0$ for $\rho > 0$). For negative correlation, these two effects are aligned, so the principal chooses $\epsilon^* = 1$ at the optimum.[21] For positive correlation, however, the determination of ϵ^* hinges on the comparison between two opposite effects, the trade-off is optimally resolved by setting ϵ strictly below 1. The smaller is the correlation ρ, the weaker is the second effect, and hence the larger is ϵ^*. In the degenerate case of no correlation, the second effect disappears, we would have $\epsilon^* = 1$.[22]

Figure 8b give a geometric explanation to the above proposition. The principal needs to move point A' rightward and shift BIR_H leftward to minimize information rents. This requires a downward distortion for q_{LL} q_{LH} and q_{HL}, but no distortion for q_{HH} since it affects neither $\pi_{LL}^{A'}$ nor BIR_H.

For positive correlation, both individual and coalitional incentive constraints are binding for downward manipulation, the efficiency versus rent extraction trade-off calls for a downward distortions of q_{LL}, q_{LH} and q_{HL}, but no distortion for q_{HH}, this is the standard "no distortion at the top" result. For negative correlation, however, the individual incentive constraint is binding for a downward manipulation while the coalitional incentive constraints are binding for upward manipulations. Hence, an issue similar to countervailing incentives arises and this calls for two-way distortions for quantities: q_{HL} and q_{HH} are distorted upward, whereas q_{LH} and q_{LL} are distorted downward compared to their respective first-best levels.[23]

As correlation vanishes, it is easy to find from (35) to (43) that

$$\lim_{\rho \uparrow 0} q_{kl}^{CP}(\mathbf{p}) = \lim_{\rho \downarrow 0} q_{kl}^{CP}(\mathbf{p}) = q_{kl}^{CP}(\rho = 0) = q_{kl}^{SB}(\rho = 0), k, l \in \{H, L\},$$

where $q_{kl}^{SB}(\rho = 0)$ are given as follows

$$\left(\theta_L - \frac{p_{HH}\Delta\theta}{p_{LH}}\right) V'(q_{LL}) = \left(\theta_L - \frac{p_{HH}\Delta\theta}{p_{LH}}\right) V'(q_{LH})$$
$$= \theta_H V'(q_{HL}) = \theta_H V'(q_{HH}) = c.$$

Therefore, JM's result that the principal can achieve her payoff without collusion in a collusion-proof way for independent types could be regarded as a limit case of ours.[24]

[21] Although ϵ belongs to $[0, 1)$, we allow ϵ to take the value equal to one since we are interested in the supremum of the seller's profit.

[22] The working paper version of JM (2005) (available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.202.5958&rep=rep1&type=pdf), gives a brief discussion of the determination of ϵ for the case of small and positive correlation (in page 22). We provide a more elaborate and formal analysis for this problem. We appreciate one referee for reminding us this version of JM (2005).

[23] See Lewis and Sappington (1989), Maggi and Rodriguez-Clare (1995) for detailed discussions of countervailing incentives problem.

[24] JM (2005) also extend their 2×2 result to $n \times 2$ and 2×3 environments.

6 The case of almost perfectly positive correlation

With an almost perfectly positive correlation (p_{LH} is very close to zero, p_{LL} and p_{HH} are both very close to 1/2), agents have very likely the same type, and ρ is close to its supremum 1/4. Constraints $CIC_{LH,HH}$ (10) and $CIC_{HH,LH}$ (11) implies $f(2q_{HH},\epsilon) \geqslant f(q_{LH}+q_{HL},\epsilon)$, constraints $CIC_{LH,LL}$ (9) and $CIC_{LL,LH}$ (7) implies $g(q_{LH}+q_{HL},\epsilon) \geqslant g(2q_{LL},\epsilon)$. We know that $f(x,\epsilon)$ is always increasing in x for any ϵ, so we have $q_{HH} \geqslant (q_{LH}+q_{HL})/2$. If $g(x,\epsilon)$ is decreasing in x, we have $q_{LL} \geqslant (q_{LH}+q_{HL})/2$. Things are different if $g(x,\epsilon)$ is increasing in x. On the one hand, the principal needs to set $(q_{LH}+q_{HL})/2 \geqslant q_{LL}$ for the contract to be implementable; on the other hand, she needs to set q_{LH} and q_{HL} as small as possible due to tradeoff between efficiency and rent extraction. To see this, note first that the contributions of q_{HL} and q_{LH} to efficiency (i.e., term $p_{LH}[\theta_H V(q_{HL})-cq_{HL}+\theta_L V(q_{LH})-cq_{LH}]$ in the principal's payoff) is negligible for small p_{LH}, but their contributions to the expected rent $\mathbb{E}\boldsymbol{\pi}$ is nonnegligible.[25] The intuition is quite simple. If there is only a small probability that agents have different valuations, quantities in this state of nature have much smaller contribution to the principal's expected payoff than their aid to the agents to secure information rents. The principal will shutdown production in this state at almost no cost. This tightens constraint $(q_{LH}+q_{HL})/2 \geqslant q_{LL}$ and thus entails partial pooling consumptions: $q_{HH} \geqslant (q_{LH}+q_{HL})/2 = q_{LL}$. To simplify the analysis, we assume in this section that the consumer's utility function is $V(x)=x^{1-\alpha}/(1-\alpha)$, $\alpha \in [0,1)$. We will show that, with an almost perfect correlation, the principal will choose a parameter ϵ such that the virtual valuation of low-type is nonpositive, i.e., $\theta^v_{L,2} \equiv \theta_L - p_{HH}\epsilon\Delta\theta/p_{LH} \leqslant 0$. Then $g(x,\epsilon) = \max_{z\in[0,x]}[\theta_H V(z)+\theta^v_{L,2}V(x-z)] - 2[\theta_L - h(\epsilon)\Delta\theta]V(x/2) = [\theta_H - 2^\alpha(\theta_L - h(\epsilon)\Delta\theta)]\,V(x)$. If $\theta_H \geqslant 2^\alpha\theta_L$, then $g(x,\epsilon)$ is increasing in x; if $\theta_H \leqslant 2^\alpha(\theta_L-\Delta\theta)$, then $g(x,\epsilon)$ is decreasing in x. The following two propositions characterize the cases with increasing and decreasing $g(x,\epsilon)$, respectively.

Proposition 6 *With an almost perfectly positive correlation and $\theta_H \geqslant 2^\alpha\theta_L$, the principal will choose an arbitrary ϵ^* in $[\theta_L p_{LH}/(p_{HH}\Delta\theta), 1]$, the optimal weakly collusion-proof mechanism entails*

- *partial pooling quantities $q^{CP}_{HH} = q^{FB}_{HH}$, $q^{CP}_{LH} = 0$, $q^{CP}_{HL} = 2q^{CP}_{LL}$, with q^{CP}_{LL} given by*

$$q^{CP}_{LL} = \left[\max\left(0, \frac{p_{LL}}{(1-p_{HH})}\left(\frac{(2\rho+2^{1-\alpha}p_{LH})p_{LH}\theta_H}{(\rho+p_{LH})p_{LL}} + \theta_L - \frac{\Delta\theta(p_{HH}-\rho)}{\rho+p_{LH}}\right)/c\right)\right]^{1/\alpha} \tag{44}$$

- *a vector of information rents/transfers such that constraints BIC_H, BIR_L, $CIC_{LH,LL}$, $CIC_{HH,LH}$, $CIC_{LL,LH}$ and $CIC_{HH,LL}$ are all binding.*

[25] In this section, we still have binding constraints BIC_H, BIR_L, $CIC_{HH,LH}$, $CIC_{LH,LL}$, so $\mathbb{E}\boldsymbol{\pi}$ is still given in (95). This will be verified later on in the proof of Proposition 6.

Proof See "Appendix". □

With an almost perfect correlation and binding constraints BIC_H, BIR_L, $CIC_{LH,LL}$ and $CIC_{HH,LH}$, there is only a negligible small probability that agents have different types. For the purpose of reducing information rents conceded to the agent, the principal optimally chooses q_{LH} and q_{HL} as small as possible. Meanwhile, the monotonic implementability constraints $q_{HH} \geqslant (q_{LH} + q_{HL})/2 \geqslant q_{LL}$ still need to be satisfied since functions $f(x, \epsilon)$ and $g(x, \epsilon)$ are both increasing in x when $\theta_H \geqslant 2^{\alpha}\theta_L$. As a result, the principal offers a less responsive contract to the reported messages, i.e., $q_{HH} \geqslant (q_{LH} + q_{HL})/2 = q_{LL}$.

When a fixed amount $2q_{LL}$ is available to and is distributed in a LH coalition, the **NAC** implies that the larger is ϵ, the smaller share will be taken by the L agent, i.e., $dq_{LH}/d\epsilon < 0$. Notice that

$$
\begin{aligned}
\frac{\partial r_{BIC_H}}{\partial \epsilon} &= \Delta\theta p_{HH} V'(q_{LH}) \frac{dq_{LH}}{d\epsilon} < 0,\\
\frac{\partial r_{CIC_{LH,LL}}}{\partial \epsilon} &= g'_{\epsilon}(2q_{LL}, \epsilon) - 2h'(\epsilon)\Delta\theta V(q_{LL}) + \frac{p_{HH}\Delta\theta}{p_{LH}} V(q_{LH})\\
&\quad + \frac{p_{HH}\epsilon\Delta\theta}{p_{LH}} \frac{dq_{LH}}{d\epsilon} V'(q_{LH})\\
&= \frac{p_{HH}\epsilon\Delta\theta}{p_{LH}} \frac{dq_{LH}}{d\epsilon} V'(q_{LH}) < 0,\\
\frac{\partial r_{CIC_{HH,LH}}}{\partial \epsilon} &= f'_{\epsilon}(2q_{LL}, \epsilon) - \frac{p_{HH}\Delta\theta}{p_{LH}} V(q_{LH}) - \frac{p_{HH}\epsilon\Delta\theta}{p_{LH}} \frac{dq_{LH}}{d\epsilon} V'(q_{LH})\\
&= -\frac{p_{HH}\epsilon\Delta\theta}{p_{LH}} \frac{dq_{LH}}{d\epsilon} V'(q_{LH}) > 0.
\end{aligned}
$$

$\partial r_{CIC_{LH,LL}}/\partial\epsilon + \partial r_{CIC_{HH,LH}}/\partial\epsilon = 0$, and the shadow cost of $CIC_{LH,LL}$ (i.e., $\rho(1 - p_{LL})/2(\rho + p_{LH})$) is larger than that of $CIC_{HH,LH}$ (i.e., $\rho p_{HH}/2(\rho + p_{LH})$). Hence, to minimize the cost of these constraints, the principal needs to increase ϵ. Meanwhile, a larger ϵ requires allocating more resources to the θ_H-type agent, and less to the θ_L-type. So the expected revenue $2\sum_k \sum_l p_{kl}\,[\theta_k V(q_{kl}) - cq_{kl}]$ increases with ϵ. As a result, the principal will increase ϵ to the extend that the virtual valuation of low type agent is non-positive, i.e., $\epsilon^* \geqslant (\theta_L p_{LH})/(p_{HH}\Delta\theta)$. Any positive amount of goods initially allocated to the low-type agent will be reallocated to his high-type partner. To avoid reallocation, the principal thus gives zero amount to low-type agent in state (θ_L, θ_H), i.e., $q_{LH} = 0$.

The set of binding constraints could be depicted in Fig. 10, in which the feasible region degenerates to line segment PQ, the coordinates of points P and Q are given as follows:

$$\pi^P_{LL} = \frac{p^2_{LH} f\,(2q_{LL}, \epsilon) - \Delta\theta p_{LH}\,(p_{HH} + p_{LH})\,V\,(q_{LL})}{\rho + p_{LH}}, \tag{45}$$

$$\pi^P_{HH} = \frac{p^2_{LH} f\,(2q_{LL}, \epsilon) + \Delta\theta p_{LL}\,(p_{HH} + p_{LH})\,V\,(q_{LL})}{\rho + p_{LH}}, \tag{46}$$

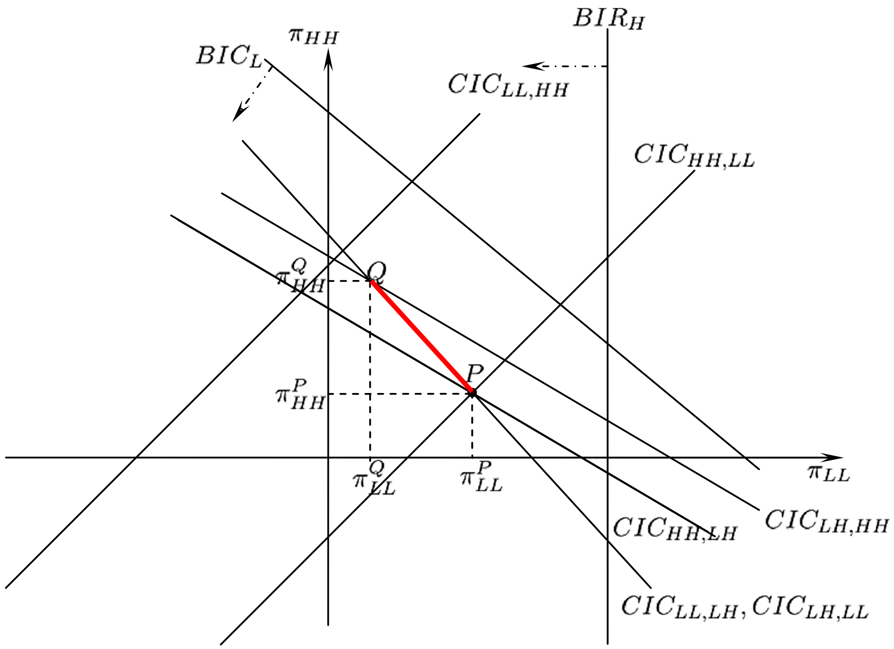


Fig. 10 The case with an almost perfect correlation and $\theta_H > 2^{\alpha}\theta_L$

$$\pi^Q_{LL} = \frac{\left[\begin{array}{c} p_{LH}\left(p_{HH}+2p_{LH}\right) f\left(2q_{LL},\epsilon\right) - p_{LH}p_{HH} f\left(2q_{HH},\epsilon\right) \\ -2\Delta\theta p_{LH}\left(p_{HH}+p_{LH}\right) V\left(q_{LL}\right) \end{array}\right]}{2\left(p_{LH}+\rho\right)}, \tag{47}$$

$$\pi^Q_{HH} = \frac{\left[\begin{array}{c} \left[\rho + p_{LH}(1-p_{HH})\right] f\left(2q_{HH},\epsilon\right) - \left(\rho + p_{LL}p_{LH}\right) f\left(2q_{LL},\epsilon\right) \\ +2\Delta\theta p_{LL}\left(p_{HH}+p_{LH}\right) V\left(q_{LL}\right) \end{array}\right]}{2\left(p_{LH}+\rho\right)}. \tag{48}$$

Note that the lines representing coalitional constraints $CIC_{HH,LL}, CIC_{HH,LH}$, $CIC_{LH,LL}$ and $CIC_{LL,LH}$ all pass through the optimal point P. Together with the presumed binding constraints BIC_H and BIR_L, we have that the set of binding constraints includes $BIC_H, BIR_L, CIC_{HH,LL}, CIC_{HH,LH}, CIC_{LH,LL}$ and $CIC_{LL,LH}$.

If the function $g(x,\epsilon)$ is decreasing in x, the optimal contract requires a non-monotonic schedule of consumptions $q_{HH} > q_{LL} > (q_{LH}+q_{HL})/2$. The binding constraints are then $BIC_H, BIR_L, CIC_{LH,LL}$ and $CIC_{HH,LL}$. To see this, we can represent the relationships between $CICs$ in Fig. 11. Area $B'C'D'E'F'$ represents the feasible region. Notice that given binding $CIC_{HH,LH}$ and $CIC_{LH,LL}$ (at point A'), we have $\ell_{CIC_{HH,LL}} - r_{CIC_{HH,LL}} = \pi_{HH} - \pi_{LL} - \Delta\theta V(q_{LL}) = [f(q_{LH}+q_{HL},\epsilon) + g(2q_{LL},\epsilon)]/2 - [h(\epsilon)+1]\Delta\theta V(q_{LL}) < [f(2q_{LL},\epsilon) + g(2q_{LL},\epsilon)]/2 - [h(\epsilon)+1]\Delta\theta V(q_{LL}) = 0$, A' is below line $CIC_{HH,LL}$; given binding $CIC_{LH,LL}$ and $CIC_{LH,HH}$ (at point B'), we have $\ell_{CIC_{HH,LL}} - r_{CIC_{HH,LL}} =$

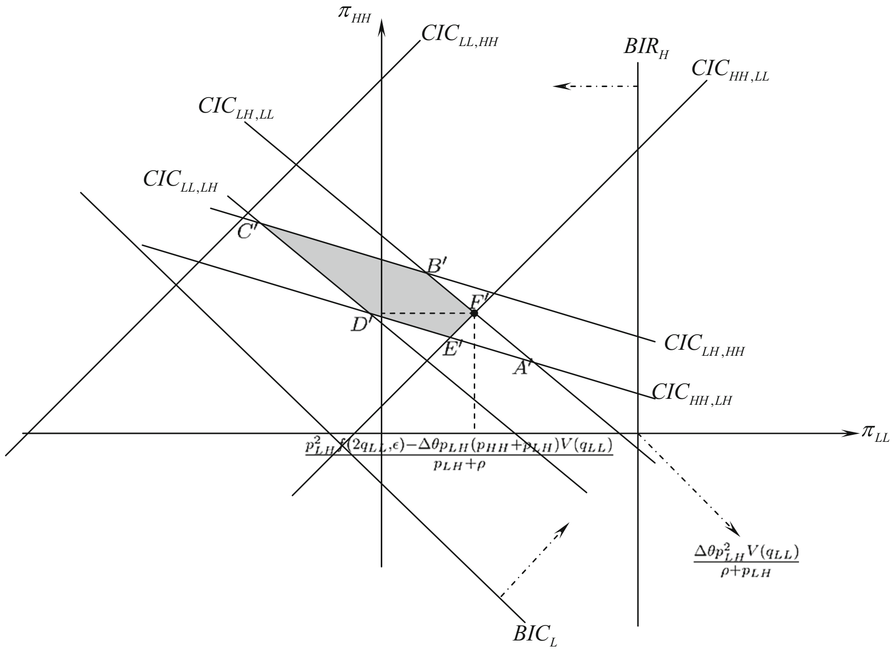


Fig. 11 The case with an almost perfect correlation and $\theta_H < 2^\alpha(\theta_L - \Delta\theta)$

$\pi_{HH} - \pi_{LL} - \Delta\theta V(q_{LL}) = [f(2q_{HH}, \epsilon) + g(2q_{LL}, \epsilon)]/2 - \Delta\theta[h(\epsilon)+1]V(q_{LL}) > [f(2q_{LL}, \epsilon) + g(2q_{LL}, \epsilon)]/2 - \Delta\theta[h(\epsilon)+1]V(q_{LL}) = 0$, B' is above line $CIC_{HH,LL}$. Hence, the optimum is attained at the rightest point F' of feasible region $B'C'D'E'F'$, where BIC_H, BIR_L, $CIC_{LH,LL}$ and $CIC_{HH,LL}$ are binding.

Proposition 7 *With an almost perfect correlation and* $\theta_H - 2^\alpha(\theta_L - \Delta\theta) < 0$*, the principal will choose an arbitrary* ϵ^* *in* $[\theta_L p_{LH}/(p_{HH}\Delta\theta), 1]$*, the optimal weakly collusion-proof mechanism* $\boldsymbol{M}^{CP}$ *entails:*

- $q_{LH}^{CP} = 0, q_{HL}^{CP} = q_{HH}^{CP} = q_{HH}^{FB} < 2q_{LL}^{CP}$, *where*

$$q_{LL}^{CP} = \left[\max\left(0, \left(\frac{\left(2 - 2^{1-\alpha}\right)\rho\theta_H p_{LH}}{p_{LL}(p_{LH}+\rho)} + \frac{\Delta\theta(\rho - p_{HH})}{p_{LH}+\rho} + \theta_L\right)/c\right)\right]^{1/\alpha} \quad (49)$$

- *a vector of rents* $\boldsymbol{\pi}$ *such that* $BIR_L, BIC_H, CIC_{HH,LL}, CIC_{LH,LL}$ *are binding.*

Proof See "Appendix". □

For a strong correlation and a decreasing $g(x, \epsilon)$, the principal can offer non-monotonic schedules of consumptions in a collusion-proof way. To a low-type agent, she offers him no good when the other agent has high type, and a nonnegative amount when the other agent is also of low-type. For a high-type agent, however, the principal always offers the first-best quantity, since neither q_{HL} nor q_{HH} affects the expected information rent $\mathbb{E}\boldsymbol{\pi}$.

Given $2q_{LL} > q_{LH} + q_{HL}$, a larger ϵ, on the one hand, make constraint $CIC_{LH,LL}$ (9) easier to be satisfied since $\partial r_{CIC_{LH,LL}}/\partial\epsilon = \Delta\theta p_{HH}[V(\varphi_1(q_{LH} + q_{HL})) - V(\varphi_1(2q_{LL}))]/p_{LH} < 0$. On the other hand, a larger ϵ demands a larger deviation away from the principal's first-best decision (notice that in the first-best allocation $\theta_H V'(q_{HL}) = \theta_L V'(q_{LH}) = c$, so $\epsilon = 0$), and thus reduces the term $2\sum_k \sum_l p_{kl}[\theta_k V(q_{kl}) - cq_{kl}]$ in the principal's profit. Since p_{LH} is very small, the first effect dominates. It is not very costly from an ex ante allocative point of view for the principal to choose a large ϵ. It will decrease the virtual valuation of a low-type agent to zero and thus has a tendency to favor the high-type agent to the extend that any positive amount of quantities will be reallocated from his low-type partner to him, so $q_{LH} = 0$.

With an almost perfectly positive correlation, a point worth discussing is the difference between cases with and without arbitrage. LM (2000) show that when the correlation becomes almost positively perfect, the first-best efficiency is approached. In our model, however, expression (49) converges to $[\max(0, (\theta_L - p_{HH}\Delta\theta/p_{LL})/c)]^{1/\alpha}$ as p_{LH} goes to zero, which implies that the allocation does not approaches its full information value q_{LL}^{FB}. The difference hinges on the role of arbitrage. When arbitrage is not allowed, an agent is only endowed with the goods initially sold to him by the principal. So, in the coalitional incentive constraints $g(2q_{LL}, \epsilon)$ and $f(2q_{HH}, \epsilon)$ need to be replaced, respectively, by $\theta_H V(q_{LL}) + (\theta_L - p_{HH}\Delta\theta\epsilon/p_{LH})V(q_{LL}) - 2[\theta_L - h(\epsilon)\Delta\theta]V(q_{LL})$, and $2\theta_H V(q_{HH}) - \theta_H V(q_{HH}) - (\theta_L - p_{HH}\Delta\theta\epsilon/p_{LH})V(q_{HH})$. The right-hand sides of constraints $CIC_{LH,HH}$ (10) and $CIC_{LH,LL}$ (9) become $2\pi_{HH} - \Delta\theta V(q_{HH}) + p_{HH}\Delta\theta\epsilon/p_{LH}[V(q_{LH}) - V(q_{HH})]$ and $2\pi_{LL} + \Delta\theta V(q_{LL}) + p_{HH}\Delta\theta\epsilon/p_{LH}[V(q_{LH}) - V(q_{LL})]$. As p_{LH} goes to zero, they both tends to $-\infty$ for quantities satisfying $0 = q_{LH} < q_{LL} < q_{HH}$ and a positive ϵ. Therefore, these two constraints can be arbitrarily satisfied, the principal then recovers some degrees of freedom. This enables him to implement the first-best allocation in the limit case of a perfectly positive correlation. But this is not the case if arbitrage would take place. Notice that, as $p_{LH} \to 0$, a corner solution emerges in the maximization problem within the coalition with $\varphi_1(x) = 0, \varphi_2(x) = x, \forall x > 0$. We thus have $r_{LH,LL} = 2\pi_{LL} + \theta_H V(\varphi_2(2q_{LL})) + \theta_L V(\varphi_1(2q_{LL})) - 2\theta_L V(q_{LL}) + p_{HH}\epsilon\Delta\theta/p_{LH}[V(\varphi_1(q_{LH} + q_{HL})) - V(\varphi_1(2q_{LL}))] \to 2\pi_{LL} + \theta_H V(2q_{LL}) - 2\theta_L V(q_{LL})$, $r_{LH,HH} = 2\pi_{HH} + \theta_H V(\varphi_2(2q_{HH})) + \theta_L V(\varphi_1(2q_{HH})) - 2\theta_H V(q_{HH}) + p_{HH}\epsilon\Delta\theta/p_{LH}[V(\varphi_1(q_{LH} + q_{HL})) - V(\varphi_1(2q_{HH}))] \to 2\pi_{HH} + \theta_H V(2q_{HH}) - 2\theta_H V(q_{HH})$. The possibility of arbitrage therefore prevents the right-hand sides of constraints $CIC_{LH,LL}$ and $CIC_{LH,HH}$ from tending to $-\infty$, they are now both bounded from below. So, these two constraints cannot be satisfied freely, that is why the first-best allocation is not achievable for an almost perfectly positive correlation in our model.

7 Conclusion

This paper explores the collusion-proof mechanism design problem in a two-agent environment with both correlation and arbitrage. CM's FSE mechanism shows that the principal may generically exploit the correlation between agents' private information to elicit their truthtelling at no cost. For the purpose of protecting their rents, agents may collude at the principal's loss by coordinating their reports and conducting arbitrage on their total purchases. As such, the principal needs to fight off collusion by designing her grand mechanism. This raises natural questions: whether and to what extend does collusion prevent the principal from implementing the first-best allocation. CK (2006) have shown that the principal can always fight off collusion at no cost in a broad class of environments with $n \geqslant 2$ agents for uncorrelated types and $n \geqslant 3$ agents for correlated types.

We extended CK's analysis to the two-agent case with arbitrage and correlation. It is shown that collusion is sometimes preventable at no cost when the correlation is strongly negative. Collusion calls for distortion away from the noncollusive efficiency if correlation is close to zero. Moreover, we find that the distortionary patterns are quite different for weakly positive and negative correlations. For weakly positive correlation, the classical no distortion at the top property is preserved; for weakly negative correlation, however, the optimal collusion-proof mechanism entails two-way distortions for consumptions. For almost perfect positive correlation, in contrast to the result of LM (2000), the possibility of arbitrage prevents the principal from implementing an efficiency close sufficiently to the first-best level.

Notice that, our model does not cover all the possible cases. The cases between weak and strong correlations, and $\theta_H \in [2^{\alpha}(\theta_L - \Delta\theta), 2^{\alpha}\theta_L)$ for almost perfectly positive correlation are not discussed. It appears to be quite challenging to obtain a complete description of the optimal collusion-proof mechanism in all the possible cases. This analysis is left for future researches.

Appendix

Proof of Proposition 2

Notice that we are finding conditions under which S^0 is optimal for the third party. Under S^0, $BIC_i^S(\theta_L), i = 1, 2$ are as same as BIC_L. Therefore, for the grand mechanisms such that θ_L-type's incentive constraint holds, the incentive constraint of θ_L-type will be satisfied in the side mechanism as well.[26] The third-party's problem can be written as:

$$\max_{\phi(\cdot),x_i(\cdot),y_i(\cdot)} \left\{ \begin{array}{l} p_{LL}\left\{\sum_{\tilde{\phi}\in\Theta^2} p^{\phi}(\theta_L,\theta_L,\tilde{\phi}) \left[\begin{array}{l} \left(\theta_L V\left(x_1(\theta_L,\theta_L,\tilde{\phi})+q_1(\tilde{\phi})\right)-t_1\left(\tilde{\phi}\right)\right) \\ +\left(\theta_L V\left(x_2(\theta_L,\theta_L,\tilde{\phi})+q_2(\tilde{\phi})\right)-t_2\left(\tilde{\phi}\right)\right) \end{array} \right]\right\} \\ +p_{LH}\left\{\sum_{\tilde{\phi}\in\Theta^2} p^{\phi}(\theta_L,\theta_H,\tilde{\phi}) \left[\begin{array}{l} \left(\theta_L V\left(x_1(\theta_L,\theta_H,\tilde{\phi})+q_1(\tilde{\phi})\right)-t_1\left(\tilde{\phi}\right)\right) \\ +\left(\theta_H V\left(x_2(\theta_L,\theta_H,\tilde{\phi})+q_2(\tilde{\phi})\right)-t_2\left(\tilde{\phi}\right)\right) \end{array} \right]\right\} \\ +p_{LH}\left\{\sum_{\tilde{\phi}\in\Theta^2} p^{\phi}(\theta_H,\theta_L,\tilde{\phi}) \left[\begin{array}{l} \left(\theta_H V\left(x_1(\theta_H,\theta_L,\tilde{\phi})+q_1(\tilde{\phi})\right)-t_1\left(\tilde{\phi}\right)\right) \\ +\left(\theta_L V\left(x_2(\theta_H,\theta_L,\tilde{\phi})+q_2(\tilde{\phi})\right)-t_2\left(\tilde{\phi}\right)\right) \end{array} \right]\right\} \\ +p_{HH}\left\{\sum_{\tilde{\phi}\in\Theta^2} p^{\phi}(\theta_H,\theta_H,\tilde{\phi}) \left[\begin{array}{l} \left(\theta_H V\left(x_1(\theta_H,\theta_H,\tilde{\phi})+q_1(\tilde{\phi})\right)-t_1\left(\tilde{\phi}\right)\right) \\ +\left(\theta_H V\left(x_2(\theta_H,\theta_H,\tilde{\phi})+q_2(\tilde{\phi})\right)-t_2\left(\tilde{\phi}\right)\right) \end{array} \right]\right\} \end{array} \right\},$$

[26] In the sequel, we will verify that BIC_L is satisfied by collusion-proof grand mechanisms in all environments considered. So readers don't need to worry about the neglected constraints $BIC_i^S(\theta_L), i = 1, 2$.

subject to the following constraints:

- Budget balance:

$$(BB : y) \sum_{i=1}^{2} y_i(\theta_1, \theta_2) = 0, \forall (\theta_1, \theta_2) \in \Theta^2 \tag{50}$$

$$(BB : x) \sum_{i=1}^{2} x_i(\theta_1, \theta_2, \tilde{\phi}) = 0, \forall (\theta_1, \theta_2) \in \Theta^2, \forall \tilde{\phi} \in \Theta^2; \tag{51}$$

- Bayesian incentive constraints for respectively the θ_H agents 1 and 2:

$$\begin{aligned} BIC_1^S(\theta_H) : & \ p_{LH} \sum_{\tilde{\phi} \in \Theta^2} p^{\phi}(\theta_H, \theta_L, \tilde{\phi}) \left[\theta_H V\left(x_1(\theta_H, \theta_L, \tilde{\phi}) + q_1(\tilde{\phi})\right) + y_1(\theta_H, \theta_L) - t_1(\tilde{\phi})\right] \\ & + p_{HH} \sum_{\tilde{\phi} \in \Theta^2} p^{\phi}(\theta_H, \theta_H, \tilde{\phi}) \left[\theta_H V\left(x_1(\theta_H, \theta_H, \tilde{\phi}) + q_1(\tilde{\phi})\right) + y_1(\theta_H, \theta_H) - t_1(\tilde{\phi})\right] \\ & \geqslant p_{LH} \sum_{\tilde{\phi} \in \Theta^2} p^{\phi}(\theta_L, \theta_L, \tilde{\phi}) \left[\theta_H V\left(x_1(\theta_L, \theta_L, \tilde{\phi}) + q_1(\tilde{\phi})\right) + y_1(\theta_L, \theta_L) - t_1(\tilde{\phi})\right] \\ & + p_{HH} \sum_{\tilde{\phi} \in \Theta^2} p^{\phi}(\theta_L, \theta_H, \tilde{\phi}) \left[\theta_H V\left(x_1(\theta_L, \theta_H, \tilde{\phi}) + q_1(\tilde{\phi})\right) + y_1(\theta_L, \theta_H) - t_1(\tilde{\phi})\right], \end{aligned} \tag{52}$$

$$\begin{aligned} BIC_2^S(\theta_H) : & \ p_{LH} \sum_{\tilde{\phi} \in \Theta^2} p^{\phi}(\theta_L, \theta_H, \tilde{\phi}) \left[\theta_H V\left(x_2(\theta_L, \theta_H, \tilde{\phi}) + q_2(\tilde{\phi})\right) + y_2(\theta_L, \theta_H) - t_2(\tilde{\phi})\right] \\ & + p_{HH} \sum_{\tilde{\phi} \in \Theta^2} p^{\phi}(\theta_H, \theta_H, \tilde{\phi}) \left[\theta_H V\left(x_2(\theta_H, \theta_H, \tilde{\phi}) + q_2(\tilde{\phi})\right) + y_2(\theta_H, \theta_H) - t_2(\tilde{\phi})\right] \\ & \geqslant p_{LH} \sum_{\tilde{\phi} \in \Theta^2} p^{\phi}(\theta_L, \theta_L, \tilde{\phi}) \left[\theta_H V\left(x_2(\theta_L, \theta_L, \tilde{\phi}) + q_2(\tilde{\phi})\right) + y_2(\theta_L, \theta_L) - t_2(\tilde{\phi})\right] \\ & + p_{HH} \sum_{\tilde{\phi} \in \Theta^2} p^{\phi}(\theta_H, \theta_L, \tilde{\phi}) \left[\theta_H V\left(x_2(\theta_H, \theta_L, \tilde{\phi}) + q_2(\tilde{\phi})\right) + y_2(\theta_H, \theta_L) - t_2(\tilde{\phi})\right]; \end{aligned} \tag{53}$$

- Bayesian participation constraints for respectively the θ_H agents 1 and 2:

$$\begin{aligned} BIR_1^S(\theta_H) : & \ p_{LH} \sum_{\tilde{\phi} \in \Theta^2} p^{\phi}(\theta_H, \theta_L, \tilde{\phi}) \left[\theta_H V\left(x_1(\theta_H, \theta_L, \tilde{\phi}) + q_1(\tilde{\phi})\right) + y_1(\theta_H, \theta_L) - t_1(\tilde{\phi})\right] \\ & + p_{HH} \sum_{\tilde{\phi} \in \Theta^2} p^{\phi}(\theta_H, \theta_H, \tilde{\phi}) \left[\theta_H V\left(x_1(\theta_H, \theta_H, \tilde{\phi}) + q_1(\tilde{\phi})\right) + y_1(\theta_H, \theta_H) - t_1(\tilde{\phi})\right] \\ & \geqslant (p_{LH} + p_{HH}) U_1^M(\theta_H), \end{aligned} \tag{54}$$

$$\begin{aligned} BIR_2^S(\theta_H) : & \ p_{LH} \sum_{\tilde{\phi} \in \Theta^2} p^{\phi}(\theta_L, \theta_H, \tilde{\phi}) \left[\theta_H V\left(x_2(\theta_L, \theta_H, \tilde{\phi}) + q_2(\tilde{\phi})\right) + y_2(\theta_L, \theta_H) - t_2(\tilde{\phi})\right] \\ & + p_{HH} \sum_{\tilde{\phi} \in \Theta^2} p^{\phi}(\theta_H, \theta_H, \tilde{\phi}) \left[\theta_H V\left(x_2(\theta_H, \theta_H, \tilde{\phi}) + q_2(\tilde{\phi})\right) + y_2(\theta_H, \theta_H) - t_2(\tilde{\phi})\right] \\ & \geqslant (p_{LH} + p_{HH}) U_2^M(\theta_H); \end{aligned} \tag{55}$$

- Bayesian participation constraints for respectively the θ_L agents 1 and 2:

$$BIR_1^S(\theta_L): p_{LL}\sum_{\tilde{\phi}\in\Theta^2} p^{\phi}(\theta_L,\theta_L,\tilde{\phi})\left[\theta_L V\left(x_1(\theta_L,\theta_L,\tilde{\phi})+q_1(\tilde{\phi})\right)+y_1(\theta_L,\theta_L)-t_1(\tilde{\phi})\right]$$
$$+p_{LH}\sum_{\tilde{\phi}\in\Theta^2} p^{\phi}(\theta_L,\theta_H,\tilde{\phi})\left[\theta_L V\left(x_1(\theta_L,\theta_H,\tilde{\phi})+q_1(\tilde{\phi})\right)+y_1(\theta_L,\theta_H)-t_1(\tilde{\phi})\right]$$
$$\geqslant (p_{LL}+p_{LH})U_1^M(\theta_L), \tag{56}$$

$$BIR_2^S(\theta_L): p_{LL}\sum_{\tilde{\phi}\in\Theta^2} p^{\phi}(\theta_L,\theta_L,\tilde{\phi})\left[\theta_L V\left(x_2(\theta_L,\theta_L,\tilde{\phi})+q_2(\tilde{\phi})\right)+y_2(\theta_L,\theta_L)-t_2(\tilde{\phi})\right]$$
$$+p_{LH}\sum_{\tilde{\phi}\in\Theta^2} p^{\phi}(\theta_H,\theta_L,\tilde{\phi})\left[\theta_L V\left(x_2(\theta_H,\theta_L,\tilde{\phi})+q_2(\tilde{\phi})\right)+y_2(\theta_H,\theta_L)-t_2(\tilde{\phi})\right]$$
$$\geqslant (p_{LL}+p_{LH})U_2^M(\theta_L). \tag{57}$$

Let us introduce the following multipliers $\rho(\theta_1,\theta_2)$, $\tau(\theta_1,\theta_2,\tilde{\phi})$, δ_1, δ_2, $\overline{\nu}_1$, $\overline{\nu}_2$, $\underline{\nu}_1$, $\underline{\nu}_2$, associated with constraints (50)–(57) respectively. We write the Lagrangian function of the above maximization problem as:

$$\mathscr{L}=\mathbb{E}(U^1+U^2)+\sum_{i=1}^{2}\delta_i BIC_i^S(\theta_H)+\sum_{i=1}^{2}\overline{\nu}_i BIR_i^S(\theta_H)+\sum_{i=1}^{2}\underline{\nu}_i BIR_i^S(\theta_L)$$
$$+\sum_{(\theta_1,\theta_2)\in\Theta^2}\rho(\theta_1,\theta_2)(BB:y)(\theta_1,\theta_2)$$
$$+\sum_{(\theta_1,\theta_2)\in\Theta^2}\sum_{\tilde{\phi}\in\Theta^2}\tau(\theta_1,\theta_2,\tilde{\phi})(BB:x)(\theta_1,\theta_2,\tilde{\phi}).$$

- *Step 1: Optimal side transfers* Optimizing with respect to $y_1(\cdot,\cdot)$, $y_2(\cdot,\cdot)$ yields

$$y_1(\theta_L,\theta_L): \rho(\theta_L,\theta_L)-p_{LH}\delta_1+p_{LL}\underline{\nu}_1=0, \tag{58}$$
$$y_2(\theta_L,\theta_L): \rho(\theta_L,\theta_L)-p_{LH}\delta_2+p_{LL}\underline{\nu}_2=0, \tag{59}$$
$$y_1(\theta_L,\theta_H): \rho(\theta_L,\theta_H)-p_{HH}\delta_1+p_{LH}\underline{\nu}_1=0, \tag{60}$$
$$y_2(\theta_L,\theta_H): \rho(\theta_L,\theta_H)+p_{LH}(\delta_2+\overline{\nu}_2)=0, \tag{61}$$
$$y_1(\theta_H,\theta_L): \rho(\theta_H,\theta_L)+p_{LH}(\delta_1+\overline{\nu}_1)=0, \tag{62}$$
$$y_2(\theta_H,\theta_L): \rho(\theta_H,\theta_L)+p_{LH}\underline{\nu}_2-p_{HH}\delta_2=0, \tag{63}$$
$$y_1(\theta_H,\theta_H): \rho(\theta_H,\theta_H)+p_{HH}(\delta_1+\overline{\nu}_1)=0, \tag{64}$$
$$y_2(\theta_H,\theta_H): \rho(\theta_H,\theta_H)+p_{HH}(\delta_2+\overline{\nu}_2)=0. \tag{65}$$

Expressions (58) and (59) imply

$$-p_{LH}\delta_1+p_{LL}\underline{\nu}_1=-p_{LH}\delta_2+p_{LL}\underline{\nu}_2. \tag{66}$$

(60) and (61) imply

$$\delta_2 + \overline{\nu}_2 = \underline{\nu}_1 - \frac{p_{HH}}{p_{LH}}\delta_1. \tag{67}$$

(62) and (63) imply

$$\delta_1 + \overline{\nu}_1 = \underline{\nu}_2 - \frac{p_{HH}}{p_{LH}}\delta_2. \tag{68}$$

(64) and (65) imply

$$\delta_1 + \overline{\nu}_1 = \delta_2 + \overline{\nu}_2. \tag{69}$$

In what follows, we restrict our attention to symmetric multipliers $\delta_1 = \delta_2 \equiv \delta$, $\underline{\nu}_1 = \underline{\nu}_2 \equiv \underline{\nu}$, $\overline{\nu}_1 = \overline{\nu}_2 \equiv \overline{\nu}$. [27]

- *Step 2: The optimal rule of reallocation* Optimizing with respect to $x_i(\theta_1, \theta_2, \tilde{\phi})$ for given $p^{\phi}(\theta_1, \theta_2, \tilde{\phi})$ yields:

$$\left\{\begin{array}{c} (x_1(\theta_L,\theta_L,\tilde{\phi}), x_2(\theta_L,\theta_L,\tilde{\phi})) \\ maximizes \\ \left[\begin{array}{c} \tau(\theta_L,\theta_L,\tilde{\phi})\left[x_1(\theta_L,\theta_L,\tilde{\phi}) + x_2(\theta_L,\theta_L,\tilde{\phi})\right] \\ +\left(\begin{array}{c} p_{LL}\theta_L \\ -p_{LH}\delta\theta_H + p_{LL}\underline{\nu}\theta_L \end{array}\right) p^{\phi}(\theta_L,\theta_L,\tilde{\phi})\left[\begin{array}{c} V(x_1(\theta_L,\theta_L,\tilde{\phi}) + q_1(\tilde{\phi})) \\ +V(x_2(\theta_L,\theta_L,\tilde{\phi}) + q_2(\tilde{\phi})) \end{array}\right] \end{array}\right] \end{array}\right\},$$

$$\left\{\begin{array}{c} (x_1(\theta_L,\theta_H,\tilde{\phi}), x_2(\theta_L,\theta_H,\tilde{\phi})) \\ maximizes \\ \left[\begin{array}{c} \tau(\theta_L,\theta_H,\tilde{\phi})\left[x_1(\theta_L,\theta_H,\tilde{\phi}) + x_2(\theta_L,\theta_H,\tilde{\phi})\right] \\ +p^{\phi}(\theta_L,\theta_H,\tilde{\phi})\left[\begin{array}{c} (p_{LH}\theta_L - p_{HH}\delta\theta_H + p_{LH}\underline{\nu}\theta_L)\, V(x_1(\theta_L,\theta_H,\tilde{\phi}) + q_1(\tilde{\phi})) \\ +(p_{LH}\theta_H + p_{LH}\delta\theta_H + p_{LH}\overline{\nu}\theta_H)\, V(x_2(\theta_L,\theta_H,\tilde{\phi}) + q_2(\tilde{\phi})) \end{array}\right] \end{array}\right] \\ = \\ \left[\begin{array}{c} \tau(\theta_L,\theta_H,\tilde{\phi})\left[x_1(\theta_L,\theta_H,\tilde{\phi}) + x_2(\theta_L,\theta_H,\tilde{\phi})\right] \\ +p^{\phi}(\theta_L,\theta_H,\tilde{\phi})p_{LH}(1+\delta+\overline{\nu})\left[\begin{array}{c} \left(\theta_L - \frac{p_{HH}\epsilon\Delta\theta}{p_{LH}}\right) V(x_1(\theta_L,\theta_H,\tilde{\phi}) + q_1(\tilde{\phi})) \\ +\theta_H V(x_2(\theta_L,\theta_H,\tilde{\phi}) + q_2(\tilde{\phi})) \end{array}\right] \end{array}\right] \end{array}\right\},$$

where $\epsilon \equiv \delta/(1 + \delta + \overline{\nu})$, the equality follows from (67). Since $x_1(\theta_L, \theta_L, \tilde{\phi}) + x_2(\theta_L, \theta_L, \tilde{\phi}) = 0$, $x_1(\theta_L, \theta_H, \tilde{\phi}) + x_2(\theta_L, \theta_H, \tilde{\phi}) = 0$ for a budget-balance constraint, we have

$$(x_1(\theta_L,\theta_L,\tilde{\phi}), x_2(\theta_L,\theta_L,\tilde{\phi})) \in \arg\max \left[\begin{array}{c} V(x_1(\theta_L,\theta_L,\tilde{\phi}) + q_1(\tilde{\phi})) \\ +V(x_2(\theta_L,\theta_L,\tilde{\phi}) + q_2(\tilde{\phi})) \end{array}\right], \tag{70}$$

$$(x_1(\theta_L,\theta_H,\tilde{\phi}), x_2(\theta_L,\theta_H,\tilde{\phi})) \in \arg\max \left[\begin{array}{c} \left(\theta_L - \frac{p_{HH}\epsilon\Delta\theta}{p_{LH}}\right) V(x_1(\theta_L,\theta_L,\tilde{\phi}) + q_1(\tilde{\phi})) \\ +\theta_H V(x_2(\theta_L,\theta_H,\tilde{\phi}) + q_2(\tilde{\phi})) \end{array}\right]. \tag{71}$$

[27] We make these assumptions for tractability following the tradition of LM (2000) and JM (2005). When no restrictions are imposed on parameters δ_i, $\underline{\nu}_i$, $\overline{\nu}_i$, the principal could possess more flexibilities and thus obtains a surplus at least as much as under the symmetric assumptions.

By analogy, we also have

$$(x_1(\theta_H, \theta_H, \tilde{\phi}), x_2(\theta_H, \theta_H, \tilde{\phi})) \in \arg\max \begin{bmatrix} V(x_1(\theta_H, \theta_H, \tilde{\phi}) + q_1(\tilde{\phi})) \\ +V(x_2(\theta_H, \theta_H, \tilde{\phi}) + q_2(\tilde{\phi})) \end{bmatrix}, \tag{72}$$

$$(x_1(\theta_H, \theta_L, \tilde{\phi}), x_2(\theta_H, \theta_L, \tilde{\phi})) \in \arg\max \begin{bmatrix} \theta_H V(x_1(\theta_H, \theta_L, \tilde{\phi}) + q_1(\tilde{\phi})) \\ +\left(\theta_L - \frac{p_{HH}\epsilon\Delta\theta}{p_{LH}}\right) V(x_2(\theta_H, \theta_L, \tilde{\phi}) + q_2(\tilde{\phi})) \end{bmatrix}. \tag{73}$$

A weakly collusion-proof mechanism requires the coalition to report truthfully and conduct no arbitrage/reallocation: in state (θ_1, θ_2), $\tilde{\phi} = (\theta_1, \theta_2)$ is required to be reported in probability one, and $x_i(\theta_1, \theta_2, \tilde{\phi}) = 0$. Then (70) and (72) are equivalent to $q_1(\theta_L, \theta_L) = q_2(\theta_L, \theta_L)$ and $q_1(\theta_H, \theta_H) = q_2(\theta_H, \theta_H)$, which are trivially satisfied. Expression (71) (or (73)) implies $q_{LH} = \varphi_1(q_{LH} + q_{HL}), q_{HL} = \varphi_2(q_{LH} + q_{HL})$, where

$$(\varphi_1(x), \varphi_2(x)) = \underset{x_1, x_2 \geqslant 0, x_1 + x_2 = x}{\arg\max} \left[\left(\theta_L - \frac{p_{HH}\epsilon\Delta\theta}{p_{LH}}\right) V(x_1) + \theta_H V(x_2)\right]. \tag{74}$$

If problem (74) has interior solutions, then q_{LH} and q_{HL} satisfies

$$\left(\theta_L - \frac{p_{HH}\epsilon}{p_{LH}}\Delta\theta\right) V'(q_{LH}) = \theta_H V'(q_{HL}). \tag{75}$$

This condition states that the agents' initial allocations must maximize their total surplus evaluated at virtual valuations, otherwise, a reallocation will be made by the third party. We call it "no-arbitrage constraint (**NAC**)".

- *Step 3: The optimal manipulation of reports* We now give the conditions under which the third party finds it optimal to require any coalition to truthfully report, i.e.,

$$p^{\phi}(\theta_1, \theta_2, \tilde{\phi}) = \begin{cases} 1 & if\, \tilde{\phi} = (\theta_1, \theta_2) \\ 0 & otherwise \end{cases}, \forall (\theta_1, \theta_2) \in \Theta^2.$$

(i) When $(\theta_1, \theta_2) = (\theta_L, \theta_L)$, $(\theta_L, \theta_L) \in \arg\max_{\tilde{\phi} \in \Theta^2} I(\tilde{\phi})$, where,

$$\begin{aligned} I(\tilde{\phi}) &= \left[\theta_L V\left(x_1(\theta_L, \theta_L, \tilde{\phi}) + q_1(\tilde{\phi})\right) - t_1(\tilde{\phi}) + \theta_L V\left(x_2(\theta_L, \theta_L, \tilde{\phi}) + q_2(\tilde{\phi})\right) - t_2(\tilde{\phi})\right] \\ &\quad - \frac{p_{LH}\Delta\theta\delta}{p_{LL} + p_{LL}\underline{\nu} - p_{LH}\delta} \begin{bmatrix} V\left(x_1(\theta_L, \theta_L, \tilde{\phi}) + q_1(\tilde{\phi})\right) \\ +V\left(x_2(\theta_L, \theta_L, \tilde{\phi}) + q_2(\tilde{\phi})\right) \end{bmatrix} \\ &= \left(\theta_L - \frac{p_{LH}^2\epsilon\Delta\theta}{p_{LL}p_{LH} + \rho\epsilon}\right) \begin{bmatrix} V\left(x_1(\theta_L, \theta_L, \tilde{\phi}) + q_1(\tilde{\phi})\right) \\ +V\left(x_2(\theta_L, \theta_L, \tilde{\phi}) + q_2(\tilde{\phi})\right) \end{bmatrix} - t_1(\tilde{\phi}) - t_2(\tilde{\phi}) \\ &= 2\left(\theta_L - \frac{p_{LH}^2\epsilon\Delta\theta}{p_{LL}p_{LH} + \rho\epsilon}\right) V\left(\frac{q_1(\tilde{\phi}) + q_2(\tilde{\phi})}{2}\right) - t_1(\tilde{\phi}) - t_2(\tilde{\phi}). \end{aligned}$$

The second equality follows by (67), the third equality is a result of (70). Therefore,

$$(\theta_L, \theta_L) \in \arg\max_{\tilde{\phi}} \left\{ 2\left(\theta_L - \frac{p_{LH}^2 \epsilon \Delta\theta}{p_{LL}p_{LH} + \rho\epsilon}\right) V\left(\frac{q_1(\tilde{\phi}) + q_2(\tilde{\phi})}{2}\right) - t_1(\tilde{\phi}) - t_2(\tilde{\phi}) \right\}. \tag{76}$$

(ii) When $(\theta_1, \theta_2) = (\theta_L, \theta_H)$,

$$\begin{aligned}(\theta_L, \theta_H) &\in \arg\max_{\tilde{\phi}} \left\{ \begin{array}{c} \left(\theta_L - \frac{p_{HH}\epsilon\Delta\theta}{p_{LH}}\right) V\left(x_1(\theta_L, \theta_H, \tilde{\phi}) + q_1(\tilde{\phi})\right) \\ +\theta_H V\left(x_2(\theta_L, \theta_H, \tilde{\phi}) + q_2(\tilde{\phi})\right) - t_1(\tilde{\phi}) - t_2(\tilde{\phi}) \end{array} \right\} \\ &= \arg\max_{\tilde{\phi}} \left\{ \begin{array}{c} \left(\theta_L - \frac{p_{HH}\epsilon\Delta\theta}{p_{LH}}\right) V\left(\varphi_1\left(q_1(\tilde{\phi}) + q_2(\tilde{\phi})\right)\right) \\ +\theta_H V\left(\varphi_2(q_1(\tilde{\phi}) + q_2(\tilde{\phi}))\right) - t_1(\tilde{\phi}) - t_2(\tilde{\phi}) \end{array} \right\}. \end{aligned} \tag{77}$$

(iii) When $(\theta_1, \theta_2) = (\theta_H, \theta_L)$,

$$\begin{aligned}(\theta_H, \theta_L) &\in \arg\max_{\tilde{\phi}} \left\{ \begin{array}{c} \theta_H V\left(x_1(\theta_H, \theta_L, \tilde{\phi}) + q_1(\tilde{\phi})\right) \\ +\left(\theta_L - \frac{p_{HH}\epsilon\Delta\theta}{p_{LH}}\right) V\left(x_2(\theta_H, \theta_L, \tilde{\phi}) + q_2(\tilde{\phi})\right) - t_1(\tilde{\phi}) - t_2(\tilde{\phi}) \end{array} \right\} \\ &= \arg\max_{\tilde{\phi}} \left\{ \begin{array}{c} \theta_H V\left(\varphi_2\left(q_1(\tilde{\phi}) + q_2(\tilde{\phi})\right)\right) \\ +\left(\theta_L - \frac{p_{HH}\epsilon\Delta\theta}{p_{LH}}\right) V\left(\varphi_1(q_1(\tilde{\phi}) + q_2(\tilde{\phi}))\right) - t_1(\tilde{\phi}) - t_2(\tilde{\phi}) \end{array} \right\}. \end{aligned} \tag{78}$$

(iv) When $(\theta_1, \theta_2) = (\theta_H, \theta_H)$,

$$\begin{aligned}(\theta_H, \theta_H) &\in \arg\max_{\tilde{\phi}} \left\{ \begin{array}{c} \theta_H V\left(x_1(\theta_H, \theta_H, \tilde{\phi}) + q_1(\tilde{\phi})\right) - t_1(\tilde{\phi}) \\ +\theta_H V\left(x_2(\theta_H, \theta_H, \tilde{\phi}) + q_2(\tilde{\phi})\right) - t_2(\tilde{\phi}) \end{array} \right\} \\ &= \arg\max_{\tilde{\phi}} \left\{ 2\theta_H V\left(\frac{q_1(\tilde{\phi}) + q_2(\tilde{\phi})}{2}\right) - t_1(\tilde{\phi}) - t_2(\tilde{\phi}) \right\}. \end{aligned} \tag{79}$$

(76)–(79) imply the following coalitional incentive constraints: for a (θ_L, θ_L) coalition,

$$\begin{aligned} CIC_{LL,LH}\colon\ & 2\left(\theta_L - \frac{p_{LH}^2 \epsilon \Delta\theta}{p_{LL}p_{LH} + \rho\epsilon}\right) V(q_{LL}) - 2t_{LL} \\ & \geqslant 2\left(\theta_L - \frac{p_{LH}^2 \epsilon \Delta\theta}{p_{LL}p_{LH} + \rho\epsilon}\right) V\left(\frac{q_{LH} + q_{HL}}{2}\right) - t_{LH} - t_{HL} \end{aligned} \tag{80}$$

$$CIC_{LL,HH}\colon\ 2\left(\theta_L - \frac{p_{LH}^2\epsilon\Delta\theta}{p_{LL}p_{LH}+\rho\epsilon}\right)V(q_{LL}) - 2t_{LL} \geqslant 2\left(\theta_L - \frac{p_{LH}^2\epsilon\Delta\theta}{p_{LL}p_{LH}+\rho\epsilon}\right)V(q_{HH}) - 2t_{HH}; \tag{81}$$

for a (θ_L, θ_H) coalition,

$$CIC_{LH,LL}\colon \left(\theta_L - \frac{p_{HH}\epsilon\Delta\theta}{p_{LH}}\right)V\big(\varphi_1(q_{LH}+q_{HL})\big) + \theta_H V\big(\varphi_2(q_{LH}+q_{HL})\big) - t_{LH} - t_{HL} \geqslant \left(\theta_L - \frac{p_{HH}\epsilon\Delta\theta}{p_{LH}}\right)V\big(\varphi_1(2q_{LL})\big) + \theta_H V\big(\varphi_2(2q_{LL})\big) - 2t_{LL} \tag{82}$$

$$CIC_{LH,HH}\colon \left(\theta_L - \frac{p_{HH}\epsilon\Delta\theta}{p_{LH}}\right)V\big(\varphi_1(q_{LH}+q_{HL})\big) + \theta_H V\big(\varphi_2(q_{LH}+q_{HL})\big) - t_{LH} - t_{HL} \geqslant \left(\theta_L - \frac{p_{HH}\epsilon\Delta\theta}{p_{LH}}\right)V\big(\varphi_1(2q_{HH})\big) + \theta_H V\big(\varphi_2(2q_{HH}\big) - 2t_{HH}; \tag{83}$$

for a (θ_H, θ_H) coalition,

$$CIC_{HH,LL}\colon 2\theta_H V(q_{HH}) - 2t_{HH} \geqslant 2\theta_H V(q_{LL}) - 2t_{LL} \tag{84}$$

$$CIC_{HH,LH}\colon 2\theta_H V(q_{HH}) - 2t_{HH} \geqslant 2\theta_H V\left(\frac{q_{LH}+q_{HL}}{2}\right) - t_{LH} - t_{HL}. \tag{85}$$

Substituting **NAC** ($q_{LH} = \varphi_1(q_{LH}+q_{HL}), q_{HL} = \varphi_2(q_{LH}+q_{HL})$) and $\pi_{kl} \equiv \theta_k V(q_{kl}) - t_{kl}$ into expressions (80)–(85) yields expressions (7)–(12) in the main text.

- Note that $\epsilon = \frac{\delta}{1+\delta+\nu} \in [0, 1)$. Moreover, $\delta = 0$ when the Bayesian incentive constraints (52) and (53) are slack in the third party's optimizing problem.
- Note that participation constraints (54)–(57) are binding for a weakly collusion-proof mechanism. Hence the slackness condition obtained from the Lagrangian optimization does not give any information on ϵ. Therefore, ϵ is a free variable in the principal's programme. □

Proof of Lemma 1

(22) is equivalent to

$$\tau_1(p_{LL}, p_{HH}) \equiv p_{LL}p_{HH} - \left(\frac{1-p_{LL}-p_{HH}}{2}\right)^2 + \frac{\Delta\theta(1+p_{LL}-p_{HH})(1-p_{LL}-p_{HH})V\left(q_{HH}^{FB}\right)}{2\left[(1-p_{HH})f(q_{LH}^{FB}+q_{HL}^{FB},0)+p_{LL}g(2q_{LL}^{FB},0)\right]} \geqslant 0; \tag{86}$$

(23) is equivalent to

$$\tau_2(q_{LL}, q_{HH}) \equiv -\frac{\Delta\theta(1-p_{LL}+p_{HH})(1-p_{LL}-p_{HH})V(q_{LL}^{FB})}{2\left[(1-p_{LL})g(q_{LH}^{FB}+q_{HL}^{FB},0)+p_{HH}f(2q_{HH}^{FB},0)\right]} - p_{LL}p_{HH} + \left(\frac{1-p_{LL}-p_{HH}}{2}\right)^2 \geqslant 0; \tag{87}$$

(24) is equivalent to

$$\tau_3(q_{LL}, q_{HH}) \equiv p_{LL}p_{HH} - \left(\frac{1-p_{LL}-p_{HH}}{2}\right)^2 - \frac{\Delta\theta\begin{bmatrix} p_{LL}(1-p_{LL}+p_{HH})V(q_{LL}^{FB}) \\ -(1-p_{LL})(1+p_{LL}-p_{HH})V(q_{HH}^{FB}) \end{bmatrix}}{2f\left(q_{LH}^{FB}+q_{HL}^{FB},0\right)} \geqslant 0; \tag{88}$$

(25) is equivalent to

$$\tau_4(p_{LL}, p_{HH}) \equiv \frac{\Delta\theta\begin{bmatrix} p_{HH}(1+p_{LL}-p_{HH})V\left(q_{HH}^{FB}\right) \\ -(1-p_{HH})(1-p_{LL}+p_{HH})V(q_{LL}^{FB}) \end{bmatrix}}{2g\left(q_{LH}^{FB}+q_{HL}^{FB},0\right)} - p_{LL}p_{HH} + \left(\frac{1-p_{LL}-p_{HH}}{2}\right)^2 \geqslant 0. \tag{89}$$

We next prove that (86) and (88) are satisfied for all feasible distributions $(p_{LL}, p_{HH}) \in \{(x, y) \in [0, 1]^2 | x + y \leqslant 1\}$, and (89) is implied by (87).

$$\begin{aligned}\tau_1(p_{LL}, p_{HH}) &\equiv p_{LL}p_{HH} - \left(\frac{1-p_{LL}-p_{HH}}{2}\right)^2 + \frac{\begin{bmatrix} \Delta\theta(1+p_{LL}-p_{HH}) \\ \times(1-p_{LL}-p_{HH})V\left(q_{HH}^{FB}\right) \end{bmatrix}}{2\left[(1-p_{HH})f(q_{LH}^{FB}+q_{HL}^{FB},0)+p_{LL}g(2q_{LL}^{FB},0)\right]} \\ &\geqslant p_{LL}p_{HH} - \left(\frac{1-p_{LL}-p_{HH}}{2}\right)^2 + \frac{(1+p_{LL}-p_{HH})(1-p_{LL}-p_{HH})}{4} \\ &= \frac{p_{LL}(1-p_{LL}+p_{HH})}{2} \geqslant 0.\end{aligned}$$

The first inequality follows from $(1-p_{HH})f(q_{HL}^{FB}+q_{LH}^{FB},0)+p_{LL}g(2q_{LL}^{FB},0)\leqslant f(q_{HL}^{FB}+q_{LH}^{FB},0)+g(2q_{LL}^{FB},0)\leqslant f(2q_{HH}^{FB},0)+g(2q_{HH}^{FB},0)=2\Delta\theta V\left(q_{HH}^{FB}\right)$.

$$\begin{aligned}\tau_3(p_{LL},p_{HH})&=p_{LL}p_{HH}-\left(\frac{1-p_{LL}-p_{HH}}{2}\right)^2\\&\quad+\frac{\Delta\theta\left[\begin{array}{c}-p_{LL}(1-p_{LL}+p_{HH})V(q_{LL}^{FB})\\(1-p_{LL})(1+p_{LL}-p_{HH})V(q_{HH}^{FB})\end{array}\right]}{2f\left(q_{LH}^{FB}+q_{HL}^{FB},0\right)}\\&\quad\geqslant p_{LL}p_{HH}-\left(\frac{1-p_{LL}-p_{HH}}{2}\right)^2+\frac{1-p_{LL}-p_{HH}}{2}\\&=p_{LL}p_{HH}+\frac{(1+p_{LL}+p_{HH})(1-p_{LL}-p_{HH})}{4}\geqslant 0.\end{aligned}$$

The first inequality follows from $(1-p_{LL})(1+p_{LL}-p_{HH})V(q_{HH}^{FB})-p_{LL}(1-p_{LL}+p_{HH})V(q_{LL}^{FB})\geqslant[(1-p_{LL})(1+p_{LL}-p_{HH})-p_{LL}(1-p_{LL}+p_{HH})]V(q_{HH}^{FB})=(1-p_{LL}-p_{HH})V(q_{HH}^{FB})\geqslant 0$ and $f(q_{LH}^{FB}+q_{HL}^{FB},0)\leqslant\Delta\theta V\left(\frac{q_{LH}^{FB}+q_{HL}^{FB}}{2}\right)<\Delta\theta V(q_{HH}^{FB})$.[28]

$$\begin{aligned}&\frac{\tau_4(p_{LL},p_{HH})-\tau_2(p_{LL},p_{HH})}{\Delta\theta}\\&=\frac{p_{HH}\left[\begin{array}{c}\left[-(1-p_{LL}+p_{HH})p_{LL}V(q_{LL}^{FB})+(1-p_{LL})(1+p_{LL}-p_{HH})V\left(q_{HH}^{FB}\right)\right]g(q_{LH}^{FB}+q_{HL}^{FB},0)\\+\left[p_{HH}(1+p_{LL}-p_{HH})V\left(q_{HH}^{FB}\right)-(1-p_{HH})(1-p_{LL}+p_{HH})V(q_{LL}^{FB})\right]f(2q_{HH}^{FB},0)\end{array}\right]}{2\left[(1-p_{LL})g(q_{LH}^{FB}+q_{HL}^{FB},0)+p_{HH}f(2q_{HH}^{FB},0)\right]g(q_{LH}^{FB}+q_{HL}^{FB},0)}\\&=\frac{p_{HH}\left[\begin{array}{c}(1-p_{LL}-p_{HH})V(q_{LL}^{FB})\left[g(q_{LH}^{FB}+q_{HL}^{FB},0)-f(2q_{HH}^{FB},0)\right]\\+\left[\begin{array}{c}(1-p_{LL})(1+p_{LL}-p_{HH})\Delta Vg(q_{LH}^{FB}+q_{HL}^{FB},0)\\+p_{HH}(1+p_{LL}-p_{HH})\Delta Vf(2q_{HH}^{FB},0)\end{array}\right]\end{array}\right]}{2\left[(1-p_{LL})g(q_{LH}^{FB}+q_{HL}^{FB},0)+p_{HH}f(2q_{HH}^{FB},0)\right]g(q_{LH}^{FB}+q_{HL}^{FB},0)}\\&\geqslant\frac{p_{HH}\left[\begin{array}{c}(1-p_{LL})(1+p_{LL}-p_{HH})\Delta V\Delta\theta V(q_{LL}^{FB})\\+p_{HH}(1+p_{LL}-p_{HH})\Delta Vf(2q_{HH}^{FB},0)\\-(1-p_{LL}-p_{HH})V(q_{LL}^{FB})\Delta\theta\Delta V\end{array}\right]}{2\left[(1-p_{LL})g(q_{LH}^{FB}+q_{HL}^{FB},0)+p_{HH}f(2q_{HH}^{FB},0)\right]g(q_{LH}^{FB}+q_{HL}^{FB},0)}\\&=\frac{p_{HH}\Delta V\left[\begin{array}{c}\Delta\theta V(q_{LL}^{FB})p_{LL}(1-p_{LL}+p_{HH})\\+p_{HH}(1+p_{LL}-p_{HH})f(2q_{HH}^{FB},0)\end{array}\right]}{2\left[(1-p_{LL})g(q_{LH}^{FB}+q_{HL}^{FB},0)+p_{HH}f(2q_{HH}^{FB},0)\right]g(q_{LH}^{FB}+q_{HL}^{FB},0)}\geqslant 0,\end{aligned}$$

where $\Delta V\equiv V(q_{HH}^{FB})-V(q_{LL}^{FB})$. The first inequality follows from $g(q_{LH}^{FB}+q_{HL}^{FB},0)-f(2q_{HH}^{FB},0)\geqslant-\Delta\theta\Delta V$, since $g(q_{LH}^{FB}+q_{HL}^{FB},0)\geqslant\Delta\theta V\left((q_{LH}^{FB}+q_{HL}^{FB})/2\right)\geqslant\Delta\theta V(q_{LL}^{FB})$ and $f(2q_{HH}^{FB})\leqslant\Delta\theta V(q_{HH}^{FB})$. Therefore, the first-best allocation is implementable if and only if the probability distribution (p_{LL},p_{HH}) falls in region

[28] Notice that $f(x,0)-\Delta\theta V(x/2)=(\theta_H+\theta_L)V(x/2)-\max_{x_1+x_2=x,x_1,x_2\geqslant 0}[\theta_H V(x_1)+\theta_L V(x_2)]\leqslant 0$, whereas $g(x,0)-\Delta\theta V(x/2)=\max_{x_1+x_2=x,x_1,x_2\geqslant 0}[\theta_H V(x_1)+\theta_L V(x_2)]-(\theta_L+\theta_H)V(x/2)\geqslant 0$.

$$
\begin{aligned}
\mathcal{F} &\equiv \left\{ (p_{LL}, p_{HH}) \in [0,1]^2 \middle| \tau_i(p_{LL}, p_{HH}) \geqslant 0, i = 1,2,3,4 \right\} \\
&= \left\{ (p_{LL}, p_{HH}) \in [0,1]^2 \middle| \tau_2(p_{LL}, p_{HH}) \geqslant 0 \right\} \\
&= \left\{ (p_{LL}, p_{HH}) \in [0,1]^2 \,\middle|\, \rho(p_{LL}, p_{HH}) \leqslant \rho^*(p_{LL}, p_{HH}) \right\},
\end{aligned}
$$

where $\rho^*(x, y) \equiv -\Delta\theta(1-x+y)(1-x-y)V(q_{LL}^{FB})/2\left[(1-x)g(q_{LH}^{FB}+q_{HL}^{FB},0) +yf(2q_{HH}^{FB},0)\right]$.

If $g(q_{LH}^{FB}+q_{HL}^{FB},0) \geqslant 2\Delta\theta V(q_{LL}^{FB})$, it is easy to see that $(0,0) \in \mathcal{F}$. So $\mathcal{F} \neq \varnothing$, the first-best allocation is thus implementable. Conversely, if $\mathcal{F} \neq \varnothing$, then

$$
2\Delta\theta V(q_{LL}^{FB}) \leqslant \frac{[(1-x-y)^2 - 4xy][(1-x)g(q_{LH}^{FB}+q_{HL}^{FB},0) + yf(2q_{HH}^{FB},0)]}{(1-x-y)(1-x+y)}
$$

holds for some feasible distribution in $\left\{ (x,y) \in [0,1]^2 \middle| x+y \leqslant 1 \right\}$. In turn, we have

$$
\begin{aligned}
&\frac{[(1-x-y)^2 - 4xy][(1-x)g(q_{LH}^{FB}+q_{HL}^{FB},0) + yf(2q_{HH}^{FB},0)]}{(1-x-y)(1-x+y)} \\
&\leqslant \frac{[(1-x-y)^2 - 4xy]\max\{g(q_{LH}^{FB}+q_{HL}^{FB},0), f(2q_{HH}^{FB},0)\}}{1-x-y} \\
&\leqslant \max\left\{ g(q_{LH}^{FB}+q_{HL}^{FB},0), f(2q_{HH}^{FB},0) \right\}.
\end{aligned}
$$

The proof is finished. □

Proof of Lemma 2

Let

$$
\psi(\theta_1,\theta_2) \equiv \max_{z \in [0, q_{HH}^{FB}+q^*(\theta_2)]} \theta_1 V(z) + \theta_2 V(q_{HH}^{FB} + q^*(\theta_2) - z), \tag{90}
$$

$$
z^*(\theta_1,\theta_2) \equiv \arg\max_{z \in [0, q_{HH}^{FB}+q^*(\theta_2)]} \theta_1 V(z) + \theta_2 V(q_{HH}^{FB} + q^*(\theta_2) - z), \tag{91}
$$

where $\theta_1, \theta_2 \in [\theta_L, \theta_H]$. Then

$$
\frac{g(q_{LH}^{FB}+q_{HL}^{FB},0)}{f(2q_{HH}^{FB},0)} = \frac{\psi(\theta_H,\theta_L) - \psi(\theta_L,\theta_L)}{\psi(\theta_H,\theta_H) - \psi(\theta_L,\theta_H)} = \frac{\psi_{\theta_1}(\xi,\theta_L)}{\psi_{\theta_1}(\xi,\theta_H)} = \frac{V(z^*(\xi,\theta_L))}{V(z^*(\xi,\theta_H))}, \tag{92}
$$

for some $\xi \in (\theta_L, \theta_H)$, where $\psi_{\theta_i}(\theta_1,\theta_2)$ denotes the partial derivative of ψ with respect to θ_i. The second equality follows from the Cauchy's mean-value theorem, the third equality is implied by the envelop theorem. The comparison between $g(q_{LH}^{FB} + q_{HL}^{FB},0)$ and $f(2q_{HH}^{FB},0)$ depends on the monotonicity of $z^*(\xi,\theta_2)$ on θ_2.

The first order condition for (90) is $\xi V'(z^*) = \theta_2 V'(q_{HH}^{FB} + q^*(\theta_2) - z^*)$. It implies that

$$\frac{\partial z^*}{\partial \xi} = \frac{-V'(z^*)}{\xi V''(z^*) + \theta_2 V''(q_{HH}^{FB} + q^*(\theta_2) - z^*)} > 0$$

$$\begin{aligned}\frac{\partial z^*}{\partial \theta_2} &= \frac{V'(q^*(\theta_2))V'(q_{HH}^{FB} + q^*(\theta_2) - z^*)}{\left[\xi V''(z^*) + \theta_2 V''(q_{HH}^{FB} + q^*(\theta_2) - z^*)\right] V''(q^*(\theta_2))} \\ &\times \left[\frac{V''(q^*(\theta_2))}{V'(q^*(\theta_2))} - \frac{V''(q_{HH}^{FB} + q^*(\theta_2) - z^*)}{V'(q_{HH}^{FB} + q^*(\theta_2) - z^*)}\right] \\ &= \frac{V'(q^*(\theta_2))V'(q_{HH}^{FB} + q^*(\theta_2) - z^*)\left[r_a(q_{HH}^{FB} + q^*(\theta_2) - z^*) - r_a(q^*(\theta_2))\right]}{\left[\xi V''(z^*) + \theta_2 V''(q_{HH}^{FB} + q^*(\theta_2) - z^*)\right] V''(q^*(\theta_2))}. \end{aligned} \tag{93}$$

- If $r_a(x)$ is increasing, we have $\partial z^*(\xi, \theta_2)/\partial \theta_2 > 0$. Assume that $\partial z^*(\xi, \theta_2)/\partial \theta_2 \leqslant 0$, (93) implies $r_a(q_{HH}^{FB} + q^*(\theta_2) - z^*) \leqslant r_a(q^*(\theta_2))$, then $q_{HH}^{FB} \leqslant z^*(\xi, \theta_2) \leqslant z^*(\xi, \theta_L) < z^*(\theta_H, \theta_L) = q_{HH}^{FB}$, a contradiction. The first inequality follows from the assumption that $r_a(x)$ is increasing, the second and third inequalities follow from $\partial z^*(\xi, \theta_2)/\partial \theta_2 \leqslant 0$ and $\partial z^*(\xi, \theta_2)/\partial \xi > 0$.
- If $r_a(x)$ is constant, it is obvious that $\partial z^*(\xi, \theta_2)/\partial \theta_2 = 0$;
- If $r_a(x)$ is decreasing, then $\partial z^*(\xi, \theta_2)/\partial \theta_2 < 0$. Assume that $\partial z^*(\xi, \theta_2)/\partial \theta_2 \geqslant 0$, (93) implies $r_a(q_{HH}^{FB} + q^*(\theta_2) - z^*) \geqslant r_a(q^*(\theta_2))$, then $q_{HH}^{FB} \leqslant z^*(\xi, \theta_2) \leqslant z^*(\xi, \theta_H) < z(\theta_H, \theta_H) = q_{HH}^{FB}$, a contradiction. The first inequality follows from the assumption that $r_a(x)$ is decreasing, the second and third inequalities follow from $\partial z^*(\xi, \theta_2)/\partial \theta_2 \geqslant 0$ and $\partial z^*(\xi, \theta_2)/\partial \xi > 0$.

It follows directly from (92) that: if $r_a(x)$ is increasing (resp. constant, decreasing) then $g(q_{LH}^{FB} + q_{HL}^{FB}, 0) < (resp. =, >) f(2q_{HH}^{FB}, 0)$. The proof is thus finished. □

Proof of Lemma 3

Letting ν denotes the multiplier of the constraint $x_1 + x_2 = x$, applying the envelop theorem, we have

$$f_x'(x, \epsilon) = \theta_H V'\left(\frac{x}{2}\right) + \nu = 0.$$

By the first-order condition

$$\theta_H V'(\varphi_2(x)) = \left(\theta_L - \frac{p_{HH}\epsilon\Delta\theta}{p_{LH}}\right) V'(\varphi_1(x)) = -\nu$$

we have

$$f_x'(x, \epsilon) = \theta_H V'\left(\frac{x}{2}\right) - \theta_H V'(\varphi_2(x)).$$

Analogously, we have

$$g'_x(x,\epsilon) = \left(\theta_L - \frac{p_{HH}\epsilon\Delta\theta}{p_{LH}}\right) V'(\varphi_1(x)) - (\theta_L - h(\epsilon)\Delta\theta)\, V'\left(\frac{x}{2}\right).$$

Note that $\varphi_1(x) < \frac{x}{2} < \varphi_2(x)$, hence, $f'_x(x,\epsilon) > 0$, $g'_x(x,\epsilon) > 0$ if ρ is close enough to zero.[29]

- ($\Rightarrow$) Summing constraints $CIC_{LH,HH}$ (10) and $CIC_{HH,LH}$ (11) yields $f(q_{LH} + q_{HL}, \epsilon) \leqslant f(2q_{HH}, \epsilon)$; summing constraints $CIC_{LL,LH}$ (7) and $CIC_{LH,LL}$ (9) yields: $g(2q_{LL}, \epsilon) \leqslant g(q_{LH} + q_{HL}, \epsilon)$. Therefore, $q_{LL} \leqslant (q_{LH} + q_{HL})/2 \leqslant q_{HH}$.
- ($\Leftarrow$) We assume that $q_{LL} \leqslant (q_{LH} + q_{HL})/2 \leqslant q_{HH}$ holds. If $CIC_{LH,LL}$(9) is binding, then $\ell_{CIC_{LL,LH}} - r_{CIC_{LL,LH}} = g(q_{LH} + q_{HL}, \epsilon) - g(2q_{LL}, \epsilon) \geqslant 0$, $CIC_{LL,LH}$ (7) holds. If $CIC_{HH,LH}$ (11) is binding, then $\ell_{CIC_{LH,HH}} - r_{CIC_{LH,HH}} = f(2q_{HH}, \epsilon) - f(q_{LH} + q_{HL}, \epsilon) \geqslant 0$, $CIC_{LH,HH}$ (10) holds. Summing the $CIC_{LH,LL}$ (9) $CIC_{HH,LH}$ (11) written with equalities yields $\pi_{HH} - \pi_{LL} = [f(q_{LH} + q_{HL}, \epsilon) + g(2q_{LL}, \epsilon)]/2 - h(\epsilon)\Delta\theta V(q_{LL})$, then $\ell_{CIC_{LL,HH}} - r_{CIC_{LL,HH}} = -[f(q_{LH} + q_{HL}, \epsilon) + g(2q_{LL}, \epsilon)]/2 + \Delta\theta[h(\epsilon) + 1]V(q_{HH}) \geqslant -[f(2q_{HH}, \epsilon) + g(2q_{HH}, \epsilon)]/2 + \Delta\theta[h(\epsilon) + 1]V(q_{HH}) = 0$, hence $CIC_{LL,HH}$ (8) holds. $\ell_{CIC_{HH,LL}} - r_{CIC_{HH,LL}} = [f(q_{LH} + q_{HL}, \epsilon) + g(2q_{LL}, \epsilon)]/2 - h(\epsilon)\Delta\theta V(q_{LL}) - \Delta\theta V(q_{LL}) \geqslant [f(2q_{LL}, \epsilon) + g(2q_{LL}, \epsilon)]/2 - h(\epsilon)\Delta\theta V(q_{LL}) - \Delta\theta V(q_{LL}) = 0$, so $CIC_{HH,LL}$ (12) holds.
  Conversely, if $CIC_{LL,LH}$ (7) and $CIC_{LH,HH}$ (10) are binding, then the remaining constraints are satisfied since $\ell_{CIC_{LH,LL}} - r_{CIC_{LH,LL}} = g(q_{LH} + q_{HL}, \epsilon) - g(2q_{LL}, \epsilon) \geqslant 0$, $\ell_{CIC_{HH,LH}} - r_{CIC_{HH,LH}} = f(2q_{HH}, \epsilon) - f(q_{LH} + q_{HL}, \epsilon) \geqslant 0$, $\ell_{CIC_{HH,LL}} - r_{CIC_{HH,LL}} = [f(2q_{HH}, \epsilon) + g(q_{LH} + q_{HL}, \epsilon)]/2 - \Delta\theta[h(\epsilon) + 1]V(q_{LL}) \geqslant [f(2q_{LL}, \epsilon) + g(2q_{LL}, \epsilon)]/2 - \Delta\theta[h(\epsilon) + 1]V(q_{LL}) = 0$, $\ell_{CIC_{LL,HH}} - r_{CIC_{LL,HH}} = -[f(2q_{HH}, \epsilon) + g(q_{LH} + q_{HL}, \epsilon)]/2 + \Delta\theta h(\epsilon)V(q_{LL}) + \Delta\theta V(q_{HH}) + h(\epsilon)\Delta\theta[V(q_{HH}) - V(q_{LL})] \geqslant -[f(2q_{HH}, \epsilon) + g(2q_{HH}, \epsilon)]/2 + [h(\epsilon) + 1]\Delta\theta V(q_{HH}) = 0$. The proof is finished. □

Proof of Proposition 4

- *Binding constraints and optimal* ϵ We first write BIR_L, BIC_H, $CIC_{LL,LH}$ and $CIC_{LH,HH}$ [(1), (4), (7) (10)] as binding constraints by inserting nonnegative parameters δ_i, $i = BIR_L, BIC_H, CIC_{LL,LH}, CIC_{LH,HH}$, then show that it is optimal for the principal to set $\delta_i = 0$ and thus verify that the corresponding constraints are binding at the optimum for all i. From the system of equations

[29] Note that when ρ is close enough to zero $p_{HH}\epsilon/p_{LH} \approx h(\epsilon) \equiv p^2_{LH}\epsilon/(\rho\epsilon + p_{LL}p_{LH})$ and $\varphi_1(x) < x/2$.

$$\begin{bmatrix} p_{LL} & p_{LH} & 0 & 0 \\ -p_{LH} & -p_{HH} & p_{LH} & p_{HH} \\ 2 & -1 & -1 & 0 \\ 0 & 1 & 1 & -2 \end{bmatrix} \begin{bmatrix} \pi_{LL} \\ \pi_{LH} \\ \pi_{HL} \\ \pi_{HH} \end{bmatrix} = \begin{bmatrix} \beta_{BIR_L} + \delta_{BIR_L} \\ \beta_{BIC_H} + \delta_{BIC_H} \\ \beta_{CIC_{LL,LH}} + \delta_{CIC_{LL,LH}} \\ \beta_{CIC_{LH,HH}} + \delta_{CIC_{LH,HH}} \end{bmatrix},$$

we get the expected rent

$$\begin{aligned} \mathbb{E}\boldsymbol{\pi} &\equiv \sum_i \sum_j p_{ij}\pi_{ij} \\ &= \frac{1}{2(p_{LH}+\rho)} \begin{bmatrix} 2p_{LH}(\beta_{BIC_H}+\delta_{BIC_H}) - \rho(1-p_{LL})(\beta_{CIC_{LL,LH}}+\delta_{CIC_{LL,LH}}) \\ +2(p_{LH}+p_{HH})(\beta_{BIR_L}+\delta_{BIR_L}) - \rho p_{HH}(\beta_{CIC_{LH,HH}}+\delta_{CIC_{LH,HH}}) \end{bmatrix}, \end{aligned} \tag{94}$$

where

$$\begin{aligned} \beta_{BIR_L} &= 0 \\ \beta_{BIC_H} &= \Delta\theta[p_{LH}V(q_{LL}) + p_{HH}V(q_{LH})] \\ \beta_{CIC_{LL,LH}} &= 2h(\epsilon)\Delta\theta V(q_{LL}) - g(q_{LH}+q_{HL},\epsilon) - \frac{p_{HH}\epsilon\Delta\theta V(q_{LH})}{p_{LH}} \\ \beta_{CIC_{LH,HH}} &= \frac{p_{HH}\epsilon\Delta\theta V(q_{LH})}{p_{LH}} - f(2q_{HH},\epsilon). \end{aligned}$$

To minimize the expected rent, the seller will set $\epsilon^* = 1, \varepsilon^*_{BIR_L} = \varepsilon^*_{BIC_H} = \varepsilon^*_{CIC_{LH,HH}} = \varepsilon^*_{CIC_{LL,LH}} = 0$, since

$$\frac{\partial \mathbb{E}\boldsymbol{\pi}}{\partial \epsilon} = \frac{-\rho}{2(\rho+p_{LH})} \begin{Bmatrix} 2(1-p_{LL})h'(\epsilon)\Delta\theta\left[V(q_{LL}) - V\left(\frac{q_{HL}+q_{LH}}{2}\right)\right] \\ +\frac{p^2_{HH}\Delta\theta[V(q_{LH})-V(\varphi_1(2q_{HH}))]}{p_{LH}} \end{Bmatrix} < 0,$$

$\partial\mathbb{E}\boldsymbol{\pi}/\partial\delta_{BIC_H} = p_{LH}/(\rho+p_{LH}) > 0, \partial\mathbb{E}\boldsymbol{\pi}/\partial\delta_{BIR_L} = (p_{LH}+p_{HH})/2(\rho+p_{LH}) > 0, \partial\mathbb{E}\boldsymbol{\pi}/\partial\delta_{CIC_{LL,LH}} = -\rho(1-p_{LL})/2(\rho+p_{LH}) > 0, \partial\mathbb{E}\boldsymbol{\pi}/\partial\delta_{CIC_{LH,HH}} = -\rho p_{HH}/2(\rho+p_{LH}) > 0$. The first inequality follows from the monotonicity condition $q_{LL} < (q_{HL}+q_{LH})/2 < q_{HH}$, which will be checked ex post.

- *Optimal quantities* Maximizing $\boldsymbol{\Pi}(\mathbf{q}) \equiv 2\left[\sum_k \sum_l p_{kl}[\theta_k V(q_{kl})) - cq_{kl}] - \mathbb{E}\boldsymbol{\pi}\right]$ with respect to q_{kl} yields (35)–(38). (36) and (37) imply that

$$\left(\theta_L - \frac{p_{HH}\Delta\theta}{p_{LH}}\right)V'(q_{LH}) = \theta_H V'(q_{HL}).$$

Hence, **NAC** (13) holds automatically for $\epsilon^* = 1$. The only work left is to verify the implementability condition. Since $\lim_{\rho\uparrow 0} q^{CP}_{kl}(\mathbf{p}) = q^{CP}_{kl}(0) = q^{SB}_{kl}(0), \forall k,l \in \{H,L\}$ and $q^{SB}_{LL}(0) = q^{SB}_{LH}(0) < q^{SB}_{HL}(0) = q^{SB}_{HH}(0)$, following the sign-preserving property of continuous function, we have $2q^{CP}_{LL}(\mathbf{p}) < q^{CP}_{LH}(\mathbf{p}) + q^{CP}_{HL}(\mathbf{p}) < 2q^{CP}_{HH}(\mathbf{p})$ when ρ is sufficiently close to zero.

- *Distortions of quantities* Notice that $\theta_H V'(q_{HH}^{CP}) < \frac{p_{LH}\theta_H V'(q_{HH}^{CP})}{\rho+p_{LH}} + \frac{\rho\theta_H V'(\varphi_2(2q_{HH}^{CP}))}{\rho+p_{LH}} = c$, therefore $q_{HH}^{CP} > q_{HH}^{FB}$. Similarly, (36) implies $\theta_H V'(q_{HL}^{CP}) < c$, hence $q_{HL}^{CP} > q_{HL}^{FB}$. $\lim_{\rho\uparrow 0} q_{LL}^{CP}(\mathbf{p}) = q_{LL}^{CP}(0) = q_{LL}^{SB}(0) < q_{LL}^{FB}(0)$, and $\lim_{\rho\uparrow 0} q_{LH}^{CP}(\mathbf{p}) = q_{LH}^{CP}(0) = q_{LH}^{SB}(0) < q_{LH}^{FB}(0)$, it follows from the sign-preserving property that $q_{LL}^{CP}(\mathbf{p}) < q_{LL}^{FB}(\mathbf{p})$, $q_{LH}^{CP}(\mathbf{p}) < q_{LH}^{FB}(\mathbf{p})$ for ρ close enough to zero. The proof is finished. □

Region $\mathcal{S}$

Let $\rho(x,y) = xy - \left(\frac{1-x-y}{2}\right)^2$, $V(x) = x^{1-\alpha}/(1-\alpha)$, then consumptions (35) to (38) can be represented as functions of (p_{LL}, p_{HH}) (or (x,y)).

$$q_{HH}(x,y) = \frac{\left[\rho(x,y)\left(\frac{\left(\theta_H^{1/\alpha} + \left(\theta_L - \frac{2y\Delta\theta}{1-x-y}\right)^{1/\alpha}\right)}{2}\right)^{\alpha} + \frac{1-x-y}{2}\theta_H\right]^{1/\alpha}}{\left[\left(\rho(x,y) + \frac{1-x-y}{2}\right)c\right]^{1/\alpha}};$$

$$q_{LH}(x,y) = \frac{q_{HL}(x,y)\left(\theta_L - \frac{2y\Delta\theta}{1-x-y}\right)^{1/\alpha}}{\theta_H^{1/\alpha}},$$

$$q_{HL}(x,y) = \left[\begin{array}{c}\frac{\left(\frac{\theta_H^{1/\alpha} + \left(\theta_L - \frac{2y\Delta\theta}{1-x-y}\right)^{1/\alpha}}{2\theta_H^{1/\alpha}}\right)^{-\alpha}(1-x)\rho(x,y)\left(\theta_L - \frac{\left(\frac{1-x-y}{2}\right)^2\Delta\theta}{\rho(x,y)+\frac{x(1-x-y)}{2}}\right)}{(1-x-y)\left(\rho(x,y)+\frac{1-x-y}{2}\right)} \\ +\theta_H\left(1 - \frac{(1-x)\rho(x,y)}{(1-x-y)\left(\rho(x,y)+\frac{1-x-y}{2}\right)}\right)\end{array}\right]^{\frac{1}{\alpha}}\left(\frac{1}{c}\right)^{1/\alpha};$$

$$q_{LL}(x,y) = \frac{\left[\theta_L - \frac{\left(\frac{1-x-y}{2}\right)^2\Delta\theta}{\rho(x,y)+\frac{x(1-x-y)}{2}}\right]^{1/\alpha}}{c^{1/\alpha}}.$$

Let

$$d(x,y) = \left[\theta_H^{1/\alpha} + \left(\theta_L - \frac{2y\Delta\theta}{1-x-y}\right)^{1/\alpha}\right]^{\alpha} - 2^{\alpha}\left[\theta_L - \frac{\left(\frac{1-x-y}{2}\right)^2\Delta\theta}{\rho(x,y) + \frac{x(1-x-y)}{2}}\right]$$

$$u(x,y) = 2^{\alpha}\theta_H - \left[\theta_H^{1/\alpha} + \left(\theta_L - \frac{2y\Delta\theta}{1-x-y}\right)^{1/\alpha}\right]^{\alpha},$$

then $g(Q,1)=G(x,y,Q)\equiv V(Q)d(x,y)$, $f(Q,1)=F(x,y,Q)\equiv V(Q)u(x,y)$, the coordinates of C' in Fig. 8 could be represented as

$$\pi_{LL}^{C'}=\pi_{LL}(x,y)=\frac{\left(\frac{1-x-y}{2}\right)\begin{bmatrix}-yF\left(x,y,2q_{HH}(x,y)\right)-(1-x)G\left(x,y,q_{HL}(x,y)+q_{LH}(x,y)\right)\\ +\frac{2(1-x)\left(\frac{1-x-y}{2}\right)^2\Delta\theta V(q_{LL}(x,y))}{\rho(x,y)+\frac{x(1-x-y)}{2}}+(1-x-y)\Delta\theta V\left(q_{LL}(x,y)\right)\end{bmatrix}}{2\left(\rho(x,y)+\frac{1-x-y}{2}\right)},$$

$$\pi_{HH}^{C'}=\pi_{HH}(x,y)=\frac{\begin{bmatrix}\left(\rho(x,y)+\frac{(1-y)(1-x-y)}{2}\right)F\left(x,y,2q_{HH}(x,y)\right)+\\ \left(\rho(x,y)+\frac{x(1-x-y)}{2}\right)G\left(x,y,q_{HL}(x,y)+q_{LH}(x,y)\right)\end{bmatrix}}{2\left(\rho(x,y)+\frac{1-x-y}{2}\right)}.$$

Being represented as functions of $p_{LL}(x)$ and $p_{HH}(y)$, BIR_H' and BIC_L' can be written as

$$\begin{aligned}BIR_H'(x,y)&=y\Delta\theta V\left(q_{LH}(x,y)\right)+\frac{1-x-y}{2}\Delta\theta V\left(q_{LL}(x,y)\right)\\&\quad-\frac{2\pi_{LL}(x,y)\rho(x,y)}{1-x-y};\\ BIC_L'(x,y)&=\frac{1-x-y}{2}\Delta\theta\left(\left(\frac{1-x-y}{2}\right)^2V\left(q_{HH}(x,y)\right)-xyV\left(q_{LH}(x,y)\right)\right)\\&\quad+x\left(\frac{1-x-y}{2}\right)^2\Delta\theta\left(V\left(q_{HL}(x,y)\right)-V\left(q_{LL}(x,y)\right)\right)\\&\quad+\rho(x,y)\left(\frac{1-x-y}{2}\pi_{HH}(x,y)+x\pi_{LL}(x,y)\right).\end{aligned}$$

So the region $\mathcal{S}$ where Proposition 4 applies could be represented as:

$$\mathcal{S}\equiv\left\{(x,y)\;\middle|\;\begin{array}{c}q_{LL}(x,y)\leqslant\frac{q_{LH}(x,y)+q_{HL}(x,y)}{2}\leqslant q_{HH}(x,y)\\ q_{kl}(x,y)\geqslant 0,k,l\in\{H,L\},\rho(x,y)\leqslant 0\\ BIR_H(x,y)\geqslant 0,BIC_L(x,y)\geqslant 0\\ \theta_L-\frac{2y\Delta\theta}{1-x-y}\geqslant 0,x+y\leqslant 1,x,y\in[0,1]\end{array}\right\},$$

it depends on parameter α.[30]

Proof of Proposition 5

- *Binding constraints* We first write BIR_L, BIC_H, $CIC_{LH,LL}$, $CIC_{HH,LH}$ [(1), (4), (9), (11)] as binding constraints by inserting nonnegative variables δ_{BIR_L},

[30] Notice that parameter c can be eliminated so it does not affect $\mathcal{S}$.

δ_{BIC_H}, $\delta_{CIC_{LH,LL}}$ and $\delta_{CIC_{HH,LH}}$ into them. From the system of equations

$$\begin{bmatrix} p_{LL} & p_{LH} & 0 & 0 \\ -p_{LH} & -p_{HH} & p_{LH} & p_{HH} \\ -2 & 1 & 1 & 0 \\ 0 & -1 & -1 & 2 \end{bmatrix} \begin{bmatrix} \pi_{LL} \\ \pi_{LH} \\ \pi_{HL} \\ \pi_{HH} \end{bmatrix} = \begin{bmatrix} \beta_{BIR_L} + \delta_{BIR_L} \\ \beta_{BIC_H} + \delta_{BIC_H} \\ \beta_{CIC_{LH,LL}} + \delta_{CIC_{LH,LL}} \\ \beta_{CIC_{HH,LH}} + \delta_{CIC_{HH,LH}} \end{bmatrix},$$

we obtain the following expected rent

$$\begin{aligned} \mathbb{E}\boldsymbol{\pi} &\equiv \sum_i \sum_j p_{ij}\pi_{ij} \\ &= \frac{\begin{bmatrix} 2(\beta_{BIC_H} + \delta_{BIC_H})p_{LH} + 2(\beta_{BIR_L} + \delta_{BIR_L})(p_{HH} + p_{LH}) \\ \rho(\beta_{CIC_{HH,LH}} + \delta_{CIC_{HH,LH}})p_{HH} + \rho(\beta_{CIC_{LH,LL}} + \delta_{CIC_{LH,LL}})(1 - p_{LL}) \end{bmatrix}}{2(p_{LH} + \rho)}, \end{aligned} \tag{95}$$

where,

$$\begin{aligned} \beta_{BIR_L} &= 0 \\ \beta_{BIC_H} &= \Delta\theta\,[p_{LH}V(q_{LL}) + p_{HH}V(q_{LH})] \\ \beta_{CIC_{LH,LL}} &= g(2q_{LL}, \epsilon) + \frac{p_{HH}\epsilon\Delta\theta}{p_{LH}}V(q_{LH}) - 2h(\epsilon)\Delta\theta V(q_{LL}) \\ \beta_{CIC_{HH,LH}} &= f(q_{LH} + q_{HL}, \epsilon) - \frac{p_{HH}\epsilon\Delta\theta}{p_{LH}}V(q_{LH}). \end{aligned}$$

δ_i is set optimally at zero since $\partial\mathbb{E}\pi/\partial\delta_i > 0$ for $i = BIR_L, BIC_H, CIC_{LH,LL}, CIC_{HH,LH}$.

- *Optimal ϵ and quantities* We write the Lagrangian function of the principal's maximization problem as:

$$\begin{aligned} \mathscr{L}(\mathbf{q}, \epsilon, \lambda) = \sum_{k,l} & p_{kl}[\theta_k V(q_{kl}) - cq_{kl}] - \mathbb{E}\boldsymbol{\pi} \\ & + \lambda\left[\theta_H V'(q_{HL}) - (\theta_L - p_{HH}\epsilon\Delta\theta/p_{LH})\,V'(q_{LH})\right]. \end{aligned}$$

Optimizing with respect to ϵ and $\mathbf{q}$ yields expressions (39)–(43). Combining (41), (42) and the **NAC** condition, we get:

$$\lambda(\epsilon) = \frac{-\frac{\Delta\theta(1-\epsilon)p_{HH}}{p_{LH}+\rho}\left(cp_{LH} + \frac{\rho p_{HH}\theta_H\left(V'\left[\frac{q_{HL}+q_{LH}}{2}\right] - V'(q_{HL})\right)}{2(p_{LH}+\rho)}\right)}{\left(\theta_L - \frac{\Delta\theta\epsilon p_{HH}}{p_{LH}}\right)^2 V''(q_{LH}) + \theta_H\left(\theta_L - \frac{\Delta\theta p_{HH}(p_{LH}+\rho\epsilon)}{p_{LH}(p_{LH}+\rho)}\right)V''(q_{HL})} \geqslant 0.$$

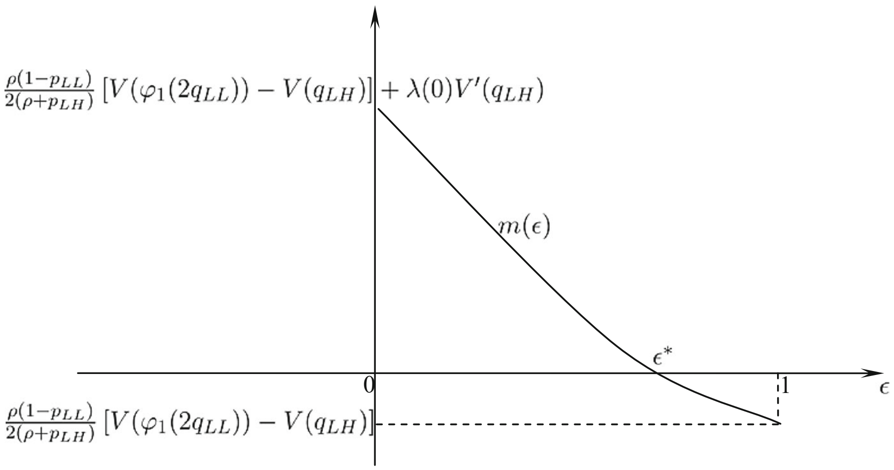


Fig. 12 Determination of ϵ^* for weakly positive correlation

Let

$$m(\epsilon) \equiv \frac{\rho(1-p_{LL})}{2(\rho+p_{LH})}[V(\varphi_1(2q_{LL}))-V(q_{LH})]+\lambda(\epsilon)V'(q_{LH}),$$

then

$$m(1) = \frac{\rho(1-p_{LL})}{2(\rho+p_{LH})}[V(\varphi_1(2q_{LL}))-V(q_{LH})] < 0$$

and

$$m(0) = \frac{\rho(1-p_{LL})}{2(\rho+p_{LH})}[V(\varphi_1(2q_{LL}))-V(q_{LH})]+\lambda(0)V'(q_{LH}) > 0.$$

Note that ρ is sufficiently small, so the sign of $m(0)$ is determined by $\lambda(0)V'(q_{LH})$, which is obviously positive. The intermediate value theorem implies there is a $\epsilon^* \in (0, 1)$ where $m(\epsilon^*) = 0$. Figure 12 depicts the determination of ϵ^*. As $\rho \to 0$, we have $\epsilon^* \to 1, \lambda \to 0$, then expressions (40) to (43) imply that $q_{kl}^{CP}(\mathbf{p}) \to q_{kl}^{CP}(0) = q_{kl}^{SB}(0), \forall k, l \in \{H, L\}$. Therefore, the implementability conditions $q_{LL}^{CP}(\mathbf{p}) < [q_{LH}^{CP}(\mathbf{p}) + q_{HL}^{CP}(\mathbf{p})]/2 < q_{LL}^{CP}(\mathbf{p})$ hold when ρ is close enough to zero.

- *Distortions of quantities* (42) implies

$$\theta_H V'(q_{HL}^{CP}) = c + \frac{\rho p_{HH}\theta_H\left[V'\left(\frac{q_{LH}+q_{HL}}{2}\right)-V'(q_{HL})\right]}{2p_{LH}(\rho+p_{LH})} - \frac{\lambda(\epsilon^*)\theta_H V''(q_{HL})}{p_{LH}} > c,$$

hence $q_{HL}^{CP} < q_{HL}^{FB}$. Furthermore, from **NAC** we get

$$\frac{V'(q_{HL}^{CP})}{V'(q_{LH}^{CP})} = \frac{\left(\theta_L - \frac{p_{HH}\epsilon^* \Delta\theta}{p_{LH}}\right)}{\theta_H} < \frac{\theta_L}{\theta_H} = \frac{V'(q_{HL}^{FB})}{V'(q_{LH}^{FB})}.$$

It implies $V'(q_{LH}^{CP}) > V'(q_{LH}^{FB})$, therefore $q_{LH}^{CP} < q_{LH}^{FB}$. As $\rho \to 0$, (40) approaches $(\theta_L - p_{HH}\Delta\theta/p_{LH})\, V'(q_{LL}) = c$, so $q_{LL}^{CP} < q_{LL}^{FB}$ for ρ close enough to zero. $q_{HH}^{CP} = q_{HH}^{FB}$ is obvious. Hence the proof is completed. □

Proof of Proposition 6

Given the utility function $V(x) = x^{1-\alpha}/(1-\alpha)$, **NAC** implies $\varphi_1(x) = s(\epsilon)x$, $\varphi_2(x) = [1 - s(\epsilon)]x$, where

$$s(\epsilon) = \frac{\left[\max\left(0, \theta_L - \frac{p_{HH}\epsilon\Delta\theta}{p_{LH}}\right)\right]^{\frac{1}{\alpha}}}{\left[\max\left(0, \theta_L - \frac{p_{HH}\epsilon\Delta\theta}{p_{LH}}\right)\right]^{\frac{1}{\alpha}} + \theta_H^{\frac{1}{\alpha}}}.$$

Given $\theta_H - 2^\alpha \theta_L \geqslant 0$,

$$\begin{aligned} g_x'(x,\epsilon) = &\left\{\left[\theta_H^{\frac{1}{\alpha}} + \left(\max\left(0, \theta_L - \frac{p_{HH}\epsilon\Delta\theta}{p_{LH}}\right)\right)^{\frac{1}{\alpha}}\right]^\alpha \right.\\ &\left. -2^\alpha \left(\theta_L - \frac{p_{LH}^2 \epsilon\Delta\theta}{\rho\epsilon + p_{LL}p_{LH}}\right)\right\} V'(x) \\ &\geqslant (\theta_H - 2^\alpha \theta_L) V'(x) \geqslant 0. \end{aligned}$$

Constraints $CIC_{LH,LL}$ (9) and $CIC_{LL,LH}$ (7) thus imply $q_{LH} + q_{HL} \geqslant 2q_{LL}$. But this condition does not hold strictly in the optimum when p_{LH} is small enough, since the principal's gains from q_{LH} and q_{HL} are very small relative to the information costs incurred by them. So pooling arises at the optimum, i.e., $q_{LH} + q_{HL} = 2q_{LL}$. Writing BIC_H, BIR_L, $CIC_{LH,LL}$ and $CIC_{HH,LH}$ as binding constraints, and incorporating $q_{LH} = 2s(\epsilon)q_{LL}$, $q_{HL} = 2[1 - s(\epsilon)]q_{LL}$, we then get the expected information rent

$$\mathbb{E}\boldsymbol{\pi} = \frac{\left[\begin{array}{c} 2p_{LH}\Delta\theta\,[p_{LH}V(q_{LL}) + p_{HH}V(q_{LH})] + \rho p_{HH}\left[f(q_{LH} + q_{HL}, \epsilon) - \frac{p_{HH}\epsilon\Delta\theta V(q_{LH})}{p_{LH}}\right] \\ \rho(1-p_{LL})\left[g(2q_{LL}, \epsilon) + \frac{p_{HH}\epsilon\Delta\theta}{p_{LH}} V(q_{LH}) - 2h(\epsilon)\Delta\theta V(q_{LL})\right] \end{array}\right]}{2(\rho + p_{LH})}$$

$$= \frac{\left[\begin{array}{c} 2p_{LH}\Delta\theta\,[p_{LH}V(q_{LL}) + p_{HH}V(q_{LH})] + \rho p_{HH}\left[f(2q_{LL}, \epsilon) - \frac{p_{HH}\epsilon\Delta\theta V(q_{LH})}{p_{LH}}\right] \\ \rho(1-p_{LL})\left[g(2q_{LL}, \epsilon) + \frac{p_{HH}\epsilon\Delta\theta}{p_{LH}} V(q_{LH}) - 2h(\epsilon)\Delta\theta V(q_{LL})\right] \end{array}\right]}{2(\rho + p_{LH})}$$

$$= \frac{\Delta\theta p_{HH} V(q_{LH})(p_{LH} + \rho\epsilon) - p_{LL}\Delta\theta V(q_{LL})(\rho - p_{HH}) - \rho p_{LH} f(2q_{LL}, \epsilon)}{p_{LH} + \rho},$$

and the principal's profit

$$\begin{aligned}\mathbf{\Pi}(q_{LL}, q_{HH}, \epsilon) \equiv\ & 2\sum_k \sum_l p_{kl}[\theta_k V(q_{kl}) - cq_{kl}] - 2\mathbb{E}\boldsymbol{\pi} \\ =\ & 2p_{LL}[\theta_L V(q_{LL}) - cq_{LL}] + 2p_{LH}[\theta_L[2s(\epsilon)]^{1-\alpha} V(q_{LL}) \\ & - 2s(\epsilon)cq_{LL}] \\ & + 2p_{LH}\left\{\theta_H[2(1 - s(\epsilon))]^{1-\alpha} V(q_{LL}) - 2(1 - s(\epsilon))cq_{LL}\right\} \\ & + 2p_{HH}[\theta_H V(q_{HH}) - cq_{HH}] - 2\mathbb{E}\boldsymbol{\pi}.\end{aligned}$$

Since

$$\frac{\partial \mathbf{\Pi}}{\partial \epsilon} = 2\left\{\frac{p_{LH} d\left[\theta_L[2s(\epsilon)]^{1-\alpha} + \theta_H[2(1 - s(\epsilon))]^{1-\alpha}\right]}{d\epsilon} - \frac{\Delta\theta p_{HH}(\rho\epsilon + p_{LH})\frac{d(2s(\epsilon))^{1-\alpha}}{d\epsilon}}{\rho + p_{LH}}\right\} V(q_{LL}) \quad (96)$$

$$\begin{cases} > 0 \text{ if} \epsilon \in [0, (\theta_L p_{LH})/(p_{HH}\Delta\theta)) \\ = 0 \text{ if} \epsilon \in ((\theta_L p_{LH})/(p_{HH}\Delta\theta), 1] \end{cases},$$

the principal will therefore choose ϵ^* arbitrarily in $[(\theta_L p_{LH})/(p_{HH}\Delta\theta), 1]$ so that $s(\epsilon) = 0$, $[1 - s(\epsilon)] = 1$. Then,

$$\begin{aligned}\mathbf{\Pi}(q_{LL}, q_{HH}) =\ & 2p_{LL}[\theta_L V(q_{LL}) - cq_{LL}] + 2p_{LH}[\theta_H 2^{1-\alpha} V(q_{LL}) - 2cq_{LL}] \\ & + 2p_{HH}[\theta_H V(q_{HH}) - cq_{HH}] \\ & - 2\frac{p_{LL}(p_{HH} - \rho)\Delta\theta V(q_{LL}) - \rho p_{LH}\left[(2 - 2^{1-\alpha})\theta_H V(q_{LL})\right]}{\rho + p_{LH}}.\end{aligned}$$

Optimizing with respect to q_{LL} and q_{HH} yields:

$$\begin{aligned}\frac{\partial \mathbf{\Pi}}{\partial q_{LL}} =\ & 2p_{LL}\left[\theta_L - \frac{\Delta\theta(p_{HH} - \rho)}{\rho + p_{LH}} + \frac{(2\rho + 2^{1-\alpha} p_{LH})p_{LH}\theta_H}{(\rho + p_{LH})p_{LL}}\right] V'(q_{LL}) \\ & - (1 - p_{HH})c, \\ \frac{\partial \mathbf{\Pi}}{\partial q_{HH}} =\ & 2p_{HH}\theta_H[V'(q_{HH}) - c].\end{aligned}$$

Therefore,

$$q_{LL} = \left[\max\left(0, \frac{p_{LL}}{(1 - p_{HH})}\left(\frac{(2\rho + 2^{1-\alpha} p_{LH})p_{LH}\theta_H}{(\rho + p_{LH})p_{LL}} + \theta_L - \frac{\Delta\theta(p_{HH} - \rho)}{\rho + p_{LH}}\right)/c\right)\right]^{1/\alpha}, \quad (97)$$

$$q_{HH} = \left(\frac{\theta_H}{c}\right)^{1/\alpha}. \quad (98)$$

It is obvious that $q_{HH} > q_{LL}$ when $(p_{LL}, p_{LH}, p_{HH}) \approx (1/2, 0, 1/2)$. Still, we need to verify the remaining constraints. As $p_{LH} \to 0$, $\ell_{BIR_H} - r_{BIR_H} = [\Delta\theta(p_{HH} - \rho)p_{LL}V(q_{LL}) - \rho p_{LH} f(2q_{LL}, \epsilon)]/(\rho + p_{LH}) \to \Delta\theta(1 - p_{LL})V(q_{LL}) > 0$,

$$\ell_{BIC_L} - r_{BIC_L} = \frac{\begin{bmatrix} \rho f(2q_{LL}, \epsilon) + \Delta\theta(p_{LH} + p_{HH})[p_{LH}V(q_{HH}) + p_{LL}V(q_{HL})] \\ -\Delta\theta p_{LL}(p_{LH} + p_{HH})V(q_{LL}) \end{bmatrix}}{p_{HH} + p_{LH}}$$
$$\to p_{LL} f(2q_{LL}, \epsilon) + \Delta\theta p_{LL}[V(q_{HL}) - V(q_{LL})] > 0,$$

$\ell_{LL,LH} - r_{LL,LH} = g(q_{HL} + q_{LH}, \epsilon) - g(2q_{LL}, \epsilon) = 0$, $\ell_{LL,HH} - r_{LL,HH} = \Delta\theta[h(\epsilon) + 1][V(q_{HH}) - V(q_{LL})] > 0$, $\ell_{LH,HH} - r_{LH,HH} = f(2q_{HH}, \epsilon) - f(q_{HL} + q_{LH}, \epsilon) > 0$, $\ell_{HH,LL} - r_{HH,LL} = \frac{1}{2} f(q_{HL} + q_{LH}, \epsilon) - \frac{1}{2} f(2q_{LL}, \epsilon) = 0$. Therefore, all the remaining constraints are satisfied, among them, $CIC_{LL,LH}$ and $CIC_{HH,LL}$ are binding. The proof is completed. □

Proof of Proposition 7

We assume momentarily then verify ex post that constraints BIC_H, BIR_L, $CIC_{LH,LL}$, $CIC_{HH,LL}$ are binding, and condition $q_{HH} > q_{LL} > (q_{LH} + q_{HL})/2$ holds. From the binding constraints, we obtain the expected rent:

$$\mathbb{E}\boldsymbol{\pi} \equiv \sum_i \sum_j p_{ij}\pi_{ij}$$
$$= \frac{\begin{bmatrix} \rho p_{LH} g(2q_{LL}, \epsilon) - \Delta\theta V(q_{LL})\left[2\rho p_{LH} h(\epsilon) - p_{LH}^2 - \rho p_{HH}\right] \\ +\Delta\theta p_{HH} V(q_{LH})(p_{LH} + \rho\epsilon) \end{bmatrix}}{p_{LH} + \rho}.$$

Then,

$$\frac{\partial \boldsymbol{\Pi}}{\partial \epsilon} = -2\frac{\partial \mathbb{E}\boldsymbol{\pi}}{\partial \epsilon} = 2\frac{\rho p_{HH}\Delta\theta[V(2s(\epsilon)q_{LL}) - V(q_{LH})]}{\rho + p_{LH}}.$$

Taking into account the **NAC** condition $q_{LH} = s(\epsilon)(q_{LH} + q_{HL})$, $q_{HL} = [1 - s(\epsilon)](q_{LH} + q_{HL})$, we can rewrite the above expression as

$$\frac{\partial \boldsymbol{\Pi}}{\partial \epsilon} = 2\frac{\rho p_{HH}\Delta\theta[s(\epsilon)]^{1-\alpha}[V(2q_{LL}) - V(q_{LH} + q_{HL})]}{\rho + p_{LH}} \begin{cases} > 0 \text{ if } \epsilon \in [0, (\theta_L p_{LH})/(p_{HH}\Delta\theta)) \\ = 0 \text{ if } \epsilon \in [(\theta_L p_{LH})/(p_{HH}\Delta\theta), 1] \end{cases}.$$

The principal will optimally choose an arbitrary $\epsilon^* \in [(\theta_L p_{LH})/(p_{HH}\Delta\theta), 1]$. Therefore, $q_{LH} = s(\epsilon^*)(q_{LH} + q_{HL}) = 0$. Maximizing with respect to q_{LL}, q_{HL}, q_{HH} yields: $q_{HL} = q_{HH} = (\theta_H/c)^{1/\alpha}$ and

$$q_{LL} = \max\left[0, \left(\theta_L - \frac{\rho p_{LH}\left(2^{1-\alpha}\theta_H - 2\theta_L\right) + \Delta\theta(p_{LH}^2 + \rho p_{HH})}{p_{LL}(\rho + p_{LH})}\right)/c\right]^{1/\alpha}.$$

Now we need only to verify the implementability condition $q_{HH} > q_{LL} > (q_{LH} + q_{HL})/2$. The first inequality is obvious. As $(p_{LL}, p_{LH}, p_{HH}) \to (1/2, 0, 1/2)$, $q_{LL} \to \max[0, ((\theta_L - \Delta\theta)/c)^{1/\alpha}]$. From $\theta_H < 2^\alpha(\theta_L - \Delta\theta)$, we get immediately that $2q_{LL} > q_{HL} = q_{HH}^{FB}$.

We verify the remaining constraints as follows. For $\epsilon \in [(\theta_L p_{LH})/(p_{HH}\Delta\theta), 1]$ and sufficiently small p_{LH}, we have

$$\begin{aligned} g'_x(x,\epsilon) &= \left\{\left[\theta_H^{\frac{1}{\alpha}} + \max\left(0, \left(\theta_L - \frac{p_{HH}\epsilon\Delta\theta}{p_{LH}}\right)^{\frac{1}{\alpha}}\right)\right]^\alpha - 2^\alpha(\theta_L - h(\epsilon)\Delta\theta)\right\} V'(x) \\ &= [\theta_H - 2^\alpha(\theta_L - h(\epsilon)\Delta\theta)]V'(x) < [\theta_H - 2^\alpha(\theta_L - \Delta\theta)]V'(x) < 0. \end{aligned}$$

As $p_{LH} \to 0$,

$$\begin{aligned} \ell_{BIR_H} - r_{BIR_H} &= \frac{\rho p_{LH} g(2q_{LL},\epsilon) - \Delta\theta V(q_{LL})[2\rho p_{LH}h(\epsilon) - p_{LH}^2 - \rho p_{HH}]}{p_{LH}+\rho} \\ &\to \Delta\theta p_{HH} V(q_{LL}) > 0, \\ \ell_{BIC_L} - r_{BIC_L} &= \frac{\left[\begin{array}{c} \Delta\theta V(q_{LL})\left(p_{LL}((2h(\epsilon)+1)p_{HH} - p_{LH}) - 2(h(\epsilon)+1)p_{LH}^2\right) \\ +\Delta\theta(p_{HH}+p_{LH})(p_{LH}V(q_{HH}) + p_{LL}V(q_{HL})) - \rho g(2q_{LL},\epsilon) \end{array}\right]}{p_{HH}+p_{LH}} \\ &\to p_{LL}\{f(2q_{LL},\epsilon) + \Delta\theta[V(q_{HL}) - V(q_{LL})]\} > 0. \end{aligned}$$

$\ell_{LL,LH} - r_{LL,LH} = g(q_{HL} + q_{LH}, \epsilon) - g(2q_{LL}, \epsilon) > 0, \ell_{LL,HH} - r_{LL,HH} = \Delta\theta[1 + h(\epsilon)][V(q_{HH}) - V(q_{LL})] > 0, \ell_{LH,HH} - r_{LH,HH} = f(2q_{HH}, \epsilon) - f(2q_{LL}, \epsilon) > 0, \ell_{HH,LH} - r_{HH,LH} = f(2q_{LL}, \epsilon) - f(q_{LH} + q_{HL}, \epsilon) > 0$. All these constraints hold strictly. The proof is finished. □

References

Che Y, Kim J (2006) Robustly collusion-proof implementation. Econometrica 74:1063–1107

Crémer J, McLean R (1985) Optimal selling strategies under uncertainty for a discriminating monopolist when demands are interdependent. Econometrica 53:345–361

Crémer J, McLean R (1988) Full extraction of the surplus in bayesian and dominant strategy auction. Econometrica 56:1247–1258

Danilov V (1992) Implementation via Nash equilibrium. Econometrica 60:43–56

Demougin D, Garvie DA (1991) Contractual design with correlated information under limited liability. Rand J Econ 22:477–489

Dequiedt V (2007) Efficient collusion in optimal auctions. J Econ Theory 136:302–323

Dutta B, Sen A (1991) A necessary and sufficient condition for two-person Nash implementation. Rev Econ Stud 58:121–128

Heifetz A, Neeman Z (2006) On the generic (im)possibility of full surplus extraction in mechanism design. Econometrica 74:213–233

Jeon D, Menicucci D (2005) Optimal second degree price discrimination under collusion: on the role of asymmetric information among buyers. Rand J Econ 36:337–360

Laffont J, Martimort D (1997) Collusion under asymmetric information. Econometrica 65:875–911

Laffont J, Martimort D (2000) Mechanism design with collusion and correlation. Econometrica 68:309–342

Lewis T, Sappington D (1989) Countervailing incentives in agency problems. J Econ Theory 49:294–313

Maskin E (1999) Nash equilibrium and welfare optimality. Rev Econ Stud 66:23–38

Maggi G, Rodriguez-Clare A (1995) On countervailing incentives. J Econ Theory 66:238–263

Mookherjee D, Tsumagari M (2004) The organization of supplier networks: effects of delegation and intermediation. Econometrica 72:1179C1219

Moore J, Repullo R (1990) Nash implementation: a full characterization. Econometrica 58:1083–1099
Pavlov G (2008) Auction design in the presence of collusion. Theor Econ 3:409–455
Peters M (2003) Surplus extraction and competition. Rev Econ Stud 68:613–631
Pouyet J (2002) Collusion under asymmetric information: the role of the correlation. J Public Econ Theory 4:543–572
Roberts J (1991) Continuity in auction design. J Econ Theory 55:169–179
Sjöström T (1991) On the necessary and sufficient conditions for Nash implementation. Soc Choice Welf 8:333–340