Review of Industrial Organization (2023) 63:239–273
https://doi.org/10.1007/s11151-023-09910-4

Cartel Leniency and Settlements: A Joint Perspective

Peter T. Dijkstra[1] · Jacob Seifert[2]

Published online: 18 August 2023

Abstract
This paper studies the interactions between a cartel leniency program and a settlement procedure. The EU settlement procedure gives colluding firms that do not apply for leniency an additional opportunity to cooperate with the competition authority in exchange for a reduced fine after a preliminary case has been established against them. We derive the conditions under which colluding firms apply for leniency, settle, or refuse to cooperate with the authority in equilibrium. Our policy results show that settlements can act as a complement or as a substitute to the leniency program. We also study the welfare-optimal policy and highlight a novel interdependence between the fine reductions that should be offered to leniency applicants and to settling firms.

Keywords Cartels · Leniency · Settlements · Welfare

JEL Classification K21 · L31 · L41

1 Introduction

Collusive agreements in which firms act in a coordinated manner so as to set prices that are above the competitive level represent arguably the clearest violation of competition laws existing across the globe. Collusion is per se illegal under Section 1 of the US Sherman Act and constitutes both a 'hardcore restriction' and a 'restriction

✉ Peter T. Dijkstra
peter.dijkstra@acm.nl

✉ Jacob Seifert
jacob.seifert@leicester.ac.uk

[1] Chief Economist Office, Netherlands Authority for Consumers & Markets, PO Box 16326, 2500 BH The Hague, The Netherlands

[2] Department of Economics, Finance and Accounting, University of Leicester, School of Business, Leicester LE1 7RH, UK

by object' under Article 101 TFEU.[1] The question of how to prevent collusion is therefore fundamental to the operations of competition authorities and to economic research that seeks to guide policymakers in this area. A number of tools are used as part of enforcement efforts against cartels. These include not only fines but also possible fine reductions for colluding firms that cooperate with the competition authority by applying for leniency or settling their case. This paper contributes to our understanding of optimal competition law enforcement against cartels by studying the effect that the joint use of a leniency program and a settlement procedure has on welfare.

Cartel leniency programs have been a feature of US antitrust enforcement since 1978 (US Department of Justice, 1993). The first leniency program in Europe was introduced in 1996 (European Commission, 1996). In their current form, these leniency programs grant firms that report their involvement in a cartel to the competition authority partial or full immunity from antitrust fines. While the first reporting firm is typically granted full immunity, in the EU later leniency applicants may also be eligible for substantial fine reductions (European Commission, 2002, 2006).

The settlement procedure was introduced by the EU in 2008 with the aim of persuading suspected cartels that do not apply for leniency to admit their guilt once a preliminary case has been formed against them—thereby avoiding the need for a lengthy prosecution (European Commission, 2008).[2] Settling firms are awarded a fine reduction, which currently stands at 10%. The main distinction between the leniency program and the settlement procedure is therefore one of timing. While leniency applications require evidence that will be supplied of an as-yet-undetected cartel, settlements are concluded after the competition authority has formed a preliminary case against a cartel. At that point, the conditions for a leniency application can no longer be met.[3]

The paper most closely related to ours is Motta and Polo (2003), who use a repeated-game framework to study the effects of leniency in the absence of a settlement procedure. These authors show that leniency programs exert a pro-competitive effect by encouraging cartels to desist from collusion during the course of their leniency application. Leniency may also have an anti-competitive effect, however, when lower expected fines encourage more cartels to form ex ante. A second related paper is Ghebrihiwet and Motchenkova (2010), who introduce the possibility of pretrial settlements into the Motta and Polo (2003) model.[4] An important difference

[1] The term hardcore restriction implies that price-fixing agreements are excluded from the EU's block exemption regulations. A restriction by object is presumed illegal without consideration of economic effects in the EU. This should be distinguished from strict per se illegality as any agreement may, in principle, benefit from an individual exemption under Article 101(3) TFEU (Whish and Bailey, 2021).

[2] Barennes (2017) and Laina and Bogdanov (2019) provide evidence of the successful uptake of the EU's settlement procedure in practice.

[3] Note that the settlement procedure is slightly different in the US, where fine reductions are determined on a case-by-case basis and firms enter a plea agreement rather than acknowledging their liability. See O'Brien (2008) and Motchenkova and Spagnolo (2019). Our focus in this paper is on settlements that follow the European approach.

[4] Another important aspect that is introduced by Ghebrihiwet and Motchenkova (2010)—which we do not consider—is the possibility that the authority may commit a type I decision error.

between their work and ours is that—unlike the present paper—Ghebrihiwet and Motchenkova (2010) consider settlements only as part of firms' deviations from an agreed collusive strategy profile. They do not consider agreements as part of a scenario in which colluding firms plan to settle their case on the equilibrium path.

This paper extends Motta and Polo (2003) to allow for settlements. In order to reflect the interplay between leniency and settlements that we described above, we adopt the following timing in the stage game: First, the firms that make up a given industry decide whether or not to collude. If a cartel is formed and the industry becomes the subject of a competition investigation, firms decide individually whether or not to apply for leniency, which involves revealing the existence of the cartel to the competition authority. If firms do not apply for leniency, the authority proceeds to form a preliminary case against the cartel, which may be strong or weak. The cartel learns the strength of this preliminary case and, thereafter, decides whether or not to settle. Only if no settlement is reached does the authority proceed to a full prosecution of the cartel, which is more likely to result in a conviction if the authority's preliminary case is strong rather than weak.

In this context, we study the stability—incentive-compatibility—of three collusive strategy profiles: Under the first, firms agree to collude and not to cooperate with the competition authority through either the leniency program or the settlement procedure. Under the second, firms agree to collude, but cooperate with the authority by applying for leniency whenever they are investigated. Finally, under the third, firms agree to collude and do not to apply for leniency, but they do settle if the authority forms a strong preliminary case against them. As part of all three collusive strategies, any deviation at the collusion, leniency, or settlement stage is punished by non-cooperative Nash equilibrium play in every subsequent period. Following Motta and Polo (2003), we determine the equilibrium outcome of the game by supposing that, whenever multiple equilibria exist, firms select the Pareto-dominant equilibrium.

We use this framework to address two important policy questions: The first concerns the effect that the addition of a settlement procedure has on equilibrium outcomes. We show that settlements can act as a complement or a substitute to the leniency program. In the former case, cartels that would not have cooperated with the authority in the absence of the settlement procedure take up the option to settle whenever the authority forms a strong preliminary case against them. In the latter, firms that would otherwise have applied for leniency no longer do so but settle instead—conditional on the authority's forming a strong preliminary case. We also identify important deterrence effects that are associated with the settlement procedure. These can be pro-competitive—when deviation opportunities at the settlement stage undermine otherwise stable cartels—as well as anti-competitive: when the added flexibility of the settlement procedure leads to the formation of new stable cartels.

The second policy question concerns the joint design of the leniency program and the settlement procedure when the authority's enforcement approach is chosen to maximize welfare. We highlight a novel and important interaction in this context: Whenever cartel deterrence is not feasible due to the limited budgetary resources of the competition authority, the optimal policy maximizes the frequency with which colluding firms cooperate with the authority in equilibrium. It does so by encouraging participation in the leniency program: This involves offering maximal fine reductions to leniency applicants and minimal reductions to settling firms.

Whenever this approach does not induce colluding firms to apply for leniency, however, the optimal policy offers maximal fine reductions to settling firms as well. This improves welfare by increasing the likelihood that colluding firms that cannot be reached through the leniency program instead cooperate with the authority by settling whenever it forms a strong preliminary case against them. Thus the optimal level of the fine reduction that should be offered to settling firms depends crucially on whether colluding firms can, under any circumstances, be persuaded to participate in the leniency program.

The remainder of the paper is organized as follows. Section 2 reviews the related literature. Section 3 presents our model, while Sect. 4 derives the equilibrium outcomes. Section 5 presents our policy analysis before Sect. 6 discusses robustness and extensions of the model. Section 7 concludes. Proofs and supplementary calculations are contained in the accompanying Online Appendix.[5]

2 Related Literature

There is a large literature on the effect of leniency programs on cartel stability (for a recent overview, see Marvão and Spagnolo, 2018). Motta and Polo (2003) identify a potential side-effect of the leniency program. While leniency can reduce the frequency of collusion by allowing firms to reveal information once an investigation has been opened, it can also lead to the formation of more stable cartels ex ante because expected fines are lower. Spagnolo (2004) shows that, when the first leniency applicant is rewarded with the sum of the fines that are paid by the remaining cartel members, the first-best outcome of complete cartel deterrence can be achieved.

Even 'moderate' leniency programs that are limited to fine reductions rather than rewards can increase cartel deterrence relative to the case without a leniency program, however. Spagnolo (2004) and Chen and Rey (2013) argue that fine reductions should be offered only to the first leniency applicant, since this incentivizes deviations from collusive agreements and therefore maximizes cartel deterrence. By contrast, Motta and Polo (2003) show that offering fine reductions to later leniency applicants can improve welfare when cartel deterrence is not feasible. These welfare gains arise as a result of non-deterred cartels' cooperating with the competition authority in the course of its investigation.

In terms of the level of the fine reduction, Harrington (2008) shows that firms are more likely to apply for leniency when the associated fine reduction is higher. When only the first applicant benefits from leniency, this increases the expected fine of cartel members that do not apply for leniency, which further incentivizes leniency applications. Nevertheless, a third, offsetting effect of leniency is that it lowers the expected fine, which increases cartel stability.

[5] The Online Appendix is accessible at https://doi.org/10.25392/leicester.data.23674800.v2.

Leniency may also increase cartel stability if it is used as a disciplining device. Any firm that deviates from the collusive price level but does not apply for leniency may be punished by the remaining cartel members that submit a leniency application (Ellis and Wilson, 2003). Moreover, according to Marx and Mezzetti (2013), firms may be incentivized to conceal their anti-competitive agreement more effectively in the presence of a leniency program, which makes it more difficult for a competition authority to prove the existence of the cartel.

Chen and Harrington (2007) consider the effect of leniency programs on the optimal cartel price. Their model allows for changes in the price that is set by the cartel to affect the probability of detection. They show that varying the collusive price over time can improve cartel stability. However, cartel prices may be lower with a leniency program than without one. This result also holds in a setting where cartel fines are price dependent; see Houba et al. (2015).

Several scholars have investigated the effect of leniency programs experimentally. The conclusions of these studies are varied: Some report positive results in terms of the effect of leniency on cartel formation and collusive prices (Hinloopen and Soetevent, 2008); others report more mixed or weak results (Bigoni et al., 2012, 2015; Andres et al., 2021; Dijkstra et al., 2021).

Brenner (2009) investigates the effectiveness of the 1996 EU leniency program from an empirical perspective and concludes that this program was successful in reducing prosecution costs by promoting the flow of information from investigated cartels to the competition authority. On the other hand, there is no evidence to suggest that the program succeeded in deterring cartels. This contrasts with Miller (2009), who does identify a deterrence effect of leniency in the US context. More recently, Jochem et al. (2020) show that the 2002 reform of the EU leniency program made it more effective in deterring cartels.

In terms of settlements, Fotis and Tselekounis (2020) develop a theoretical model to investigate the relationship between the likelihood of investigation by the competition authority and the fine reduction that is available to settling firms. In a Cournot duopoly setting, they find that firms will settle for a 10% fine reduction if the probability of cartel detection is high enough (over 40%). While Fotis and Tselekounis (2020) consider a static model, we adopt a dynamic approach to investigating collusion. We also complement their analysis by studying settlements in the context of an existing leniency program rather than in isolation.

Most existing research into the settlement procedure has adopted an empirical approach. Using a panel data set of all EC cartel cases between 2000 and 2014, Hüschelrath and Laitenberger (2017) find that settlements reduce the average case duration by over 8 months. Hellwig et al. (2018) estimate that the number of appeals of Commission decisions decreased by over 50% due to the implementation of the settlement procedure. Relatedly, Laina and Bogdanov (2019) show that conviction decisions reached on the basis of settlements are robust in the EU context: These cartel decisions tend to be confirmed by the EU courts on appeal.

3 The Model

Our model is based on Motta and Polo (2003), hereafter MP, whose framework we extend to include a settlement procedure. In particular, we study a group of perfectly symmetric firms (an industry) that may decide to collude. As in MP, we will consider symmetric industries in the equilibrium analysis. This implies that all firms in all industries will adopt identical strategies with respect to collusion in equilibrium. Each individual firm is active in only one industry.[6]

Firms' decisions to collude or not are made in the presence of a competition authority (CA). The objective of the CA is to maximize a utilitarian welfare function with equal weights on producer and consumer surplus. To that end, the CA is able to commit to an enforcement approach as part of which it conducts investigations that can lead to the conviction of cartels. Convicted cartel members are obliged to pay a fine, though fine reductions may be offered to firms that cooperate with the CA by applying for leniency or settling their case. Importantly, in line with MP, an application for leniency or a settlement by any individual firm is assumed to result in a certain conviction of all firms that make up the cartel.

We first set out the CA's enforcement approach in detail before describing the firms' collusive strategies and the timing of the game.

3.1 Policy Parameters

The CA's enforcement approach is captured in the following policy parameters, which are chosen by the CA to maximize its utilitarian welfare function.

Cartel Fine (F). The fine that is payable by a colluding firm that is convicted by the CA and that did not cooperate through the leniency program or the settlement procedure is equal to $F \in (0, \overline{F}]$, where the maximum possible fine $\overline{F}$ is exogenously specified by competition law.

Fine Reductions ($\gamma_1, \gamma_L, \gamma_S$). Firms that are subject to a CA investigation may cooperate with the CA by applying for leniency or, if no leniency application is made, by settling before a decision is reached in their case. We suppose that the expected fine that is payable by a colluding firm when all cartel members apply for leniency is equal to $\gamma_L F$, where $\gamma_L \in [0, 1]$. Consistent with practice in both the EU and US, where earlier leniency applicants receive more generous fine reductions, we allow the fine that is paid by the first leniency applicant to differ from later applicants. In particular, the first firm to apply for leniency is assumed to pay an amount $\gamma_1 F$, where $\gamma_1 \in [0, 1]$.[7]

If firms do not apply for leniency, the CA proceeds to form a preliminary case against the cartel. This preliminary case may be either 'strong' or 'weak', with a

[6] As in MP, while the existence of more than one symmetric industry is important in deriving the profits of collusion, the precise number of industries does not affect the equilibrium outcome of the game for a given probability of investigation (see Sect. 4).

[7] This adds an element of realism relative to MP, who assume that all leniency applicants receive the same fine reduction ($\gamma_1 = \gamma_L$). While we do not restrict γ_1 to lie below γ_L, we show in Sect. 5 that the optimal policy sets $\gamma_1 = 0$.

strong preliminary case implying a higher likelihood of subsequent conviction (see further discussion below).[8] We assume that it is equally likely that the CA will form a strong or weak preliminary case and, in order to ensure that the game is stationary, that industries are symmetric in terms of the strength of the preliminary case that is formed against them in any given period. The strength of the CA's preliminary case is common knowledge.[9] After learning the strength of the CA's preliminary case, colluding firms can settle, in which case they pay a fine equal to $\gamma_S F$, where $\gamma_S \in [0, 1]$.[10]

In what follows, we restrict the fine reduction that is offered under the settlement procedure to be less generous than that awarded to leniency applicants: $\gamma_S > \max\left\{\gamma_1, \gamma_L\right\}$. This ensures that the leniency program is economically meaningful, in the sense that colluding firms always prefer to apply for leniency rather than settling regardless of the strength of the CA's preliminary case (see Appendix A). This implies that, in line with practice, settlements are not used as a means of replicating the leniency program, but rather of encouraging cartels that choose not to apply for leniency to cooperate with the CA (European Commission, 2008).

Probability of Investigation (α). The probability that the firms in a given industry become the subject of a CA investigation is denoted by $\alpha \in [0, 1]$.

Baseline Conviction Probability (p). If colluding firms neither apply for leniency nor settle, the CA proceeds to the prosecution phase. This results in a conviction of the cartel with probability p_0 if its preliminary case was weak and $p_1 > p_0$ if its preliminary case was strong. The expected probability of conviction, before the CA's preliminary case is formed, is therefore given by

$$\overline{p} = \frac{1}{2}(p_0 + p_1). \tag{1}$$

An important objective of this paper is to explore the effect that the addition of a settlement procedure has on equilibrium outcomes. To do so, we compare the results of this model with MP who—absent settlements but allowing for leniency—plot equilibrium outcomes as a function of the probability of investigation α and the average conviction probability $\overline{p}$ (in our notation). Since $\overline{p}$ is itself determined as a function of p_0 and p_1 in our extended model, in order to facilitate this comparison we will assume that

$$p_0 = p - d, \tag{2}$$

$$p_1 = p + d, \tag{3}$$

[8] The strength of this preliminary case may, for instance, reflect the nature of the evidence that first comes to the CA's attention following a dawn raid; see Blatter et al. (2018).

[9] This is consistent with the EU settlement procedure, under which settlement opportunities are extended equally to all investigated firms (Dekeyser and Roques, 2010).

[10] More broadly, we may also think of the fine reduction that accompanies a settlement as reflecting a reduction in litigation costs that are borne by the firm.

where $0 < d < \frac{1}{2}$ is exogenous and $p \in [d, 1-d]$ measures the baseline probability of conviction. From Eqs. (1), (2), and (3) it follows that $\overline{p} = p$, so that the ex ante expected conviction probability is now determined entirely on the basis of the baseline conviction probability p. We assume that the CA selects the level of this baseline conviction probability p, which determines p_0 and p_1 through Eqs. (2) and (3).

The full list of CA policy parameters is therefore given by $(F, \gamma_1, \gamma_L, \gamma_S, \alpha, p)$. Following MP, we suppose that the choice of α and p is constrained by the CA's exogenously given budget. This budget can be allocated alternatively to investigating more cartels – increasing α – or to increasing the baseline probability of conviction – increasing p – so that we have fixed rather than variable enforcement costs. The form of the budget constraint will be made precise in the context of our discussion of the optimal policy in Sect. 5.

Finally, again following MP, we suppose that the CA can impose compliance on any convicted cartel, so that firms desist from collusion during any period in which they are convicted. This occurs regardless of whether the cartel was convicted following a successful CA prosecution, a leniency application, or a settlement. The notion that this desistence effect is achieved only after a delay whenever the CA is obliged to complete its prosecution of the cartel will be made precise in our discussion of the model's timing in Sect. 3.3 below.

3.2 Firms' Collusive Strategies

The CA's enforcement approach that was described above—particularly the opportunities that it affords colluding firms to apply for leniency or to settle once an investigation is opened—implies that any collusive agreement between firms must specify not only whether to collude or not, but also the course of action that the firms will take with respect to leniency and settlements, should they be investigated by the CA. We consider the following three collusive strategies:

– In the first, **NNN** ('no leniency application, no settlement of a strong preliminary case, no settlement of a weak preliminary case'), firms collude from the first period onwards—provided that no deviation occurs. In any period in which no investigation is opened, this earns each firm collusive profits π_C. In any period in which an investigation is opened, firms neither apply for leniency nor settle, regardless of whether the CA's preliminary case is strong or weak, again earning profits π_C in that period. In the following period, the CA moves to the prosecution phase and convicts the cartel with probability p_0 (p_1) if the strength of its preliminary case was weak (strong). If convicted, firms pay the fine F and play non-cooperatively for the current period, earning profits $\pi_N < \pi_C$, before reverting to collusion. A deviation at the collusion stage gives a firm deviation profits $\pi_D > \pi_C$ in that period. Deviations at the leniency or settlement stage will result in a certain conviction of the cartel, and hence profits π_N in that period, but will grant the deviating firm the fine reduction corresponding to the leniency program or settlement procedure that was described above.

– The second collusive strategy is **NSN** ('no leniency application, settle a strong preliminary case, no settlement of a weak preliminary case'). This parallels NNN, except that firms agree to settle whenever they are investigated and the CA forms a strong preliminary case against them. In any period in which an investigation is opened, firms therefore do not apply for leniency but await the strength of the CA's preliminary case. If it is strong, firms settle, pay their reduced fine $\gamma_S F$ and play non-cooperatively in the current period before resuming collusion. If the CA's preliminary case is weak, firms do not settle and earn the collusive profits π_C in that period. In the following period, they are convicted with probability p_0 as a result of the CA's prosecution. In that case, they pay the full fine F and earn profits π_N in that period before reverting to collusion in the following period.

– The final collusive strategy is denoted by **LXX** ('apply for leniency'). Now firms collude from the first period onwards, provided no deviation occurs. In any period in which no investigation is opened, firms earn the collusive profits π_C. If an investigation is opened, firms apply for leniency, pay the reduced fine $\gamma_L F$ and play non-cooperatively for the current period before reverting to collusion. In this case, we leave firms' plans with respect to settlements unspecified and denote these with an X. On the equilibrium path, firms will never reach the settlement stage since they always apply for leniency first. Since the cartel is convicted with certainty as soon as one firm applies for leniency, the question of a strong or weak preliminary case being formed does not arise for any firm that deviates from LXX by *not* applying for leniency. For clarity, we assume that such deviations cannot be followed by a settlement. It is important to emphasize, however, that this implies no loss in generality. We show in Sect. 4.1 that, regardless of whether deviations from LXX at the leniency stage can be followed by a settlement or not, the binding incentive-compatibility constraint on LXX arises at the collusion stage – not the leniency stage.

We assume that firms play the familiar grim trigger strategies that are common to the repeated games literature (e.g., Friedman, 1971). Accordingly, any deviation from the agreed collusion, leniency, or settlement plans is punished by firms' playing the non-cooperative Nash equilibrium in every subsequent period. We rule out the possibility that firms can collude tacitly in the absence of any collusive agreement. Tacit agreements do not generate any evidence on the basis of which the CA can prosecute firms and therefore do not depend on the design of the enforcement tools, including leniency and settlements, that we investigate in this paper.

It is important to note that the conviction of a cartel—whether as the result of a leniency application, a settlement, or a prosecution by the CA—does not necessarily imply a deviation from the above collusive strategies. Whenever a cartel is convicted but no deviation has occurred—e.g., because firms applied for leniency under collusive strategy LXX—firms play non-cooperatively for one period before resuming their collusive strategy in the following period.[11]

Finally, note that collusive strategies NSS ('no leniency application, settle a strong preliminary case, settle a weak preliminary case') and NNS ('no leniency application, no settlement of a strong preliminary case, settle a weak preliminary case') can be omitted from the analysis, since they are dominated by LXX and NSN, respectively. This is shown formally in Appendix A.

[11] This follows MP and is also consistent with Blatter et al. (2018), for example.

3.3 Timing

The timing of the game is as follows: At $t = 0$, the CA sets its policy parameters. Thereafter, firms select a collusive strategy and—provided that no deviation occurs—each period t has the following structure: Firms are investigated with probability α. If investigated, firms decide whether or not to apply for leniency. If they do not apply for leniency, the CA forms its preliminary case, and firms decide whether or not to settle. If firms apply for leniency or settle, they pay their reduced fines, play non-cooperatively for the current period, before resuming collusion in period $t + 1$. If firms neither apply for leniency nor settle, the game proceeds to period $t + 1$, in which the CA concludes its prosecution. If firms are convicted, they pay the full fine and play non-cooperatively in that period. If firms are not convicted, they earn the collusive profits in that period. In both cases, collusion resumes in $t + 2$.

We summarize the timing of the stage game in the form of a game tree in Fig. 1. This game tree depicts the two-firm case, beginning after the CA policy parameters have been set at $t = 0$. Firm 1 receives the top payoff in each payoff vector. Following the realization of the firms' payoffs in a given period t, the game restarts beginning in period $t + 1$.

4 Equilibrium Analysis

We solve for the subgame-perfect Nash equilibrium of the model by backward induction. To begin, we hold fixed the levels of the CA's policy parameters and identify the regions in (α, p) space in which each of the three collusive strategies—when played symmetrically by all firms—constitute an equilibrium. We also identify the Pareto-dominant equilibrium whenever multiple equilibria exist by comparing the value that is derived from each collusive strategy. This leads into the analysis of the optimal policy that follows in Sect. 5.

In line with the existing literature, we assume that collusion is an economically meaningful phenomenon in the sense that—absent any intervention by the CA—colluding yields higher profits to firms than not colluding. The value of the game under collusion and absent CA intervention—which is denoted by V_C—may be written as

$$V_C = \pi_C + \delta V_C \Rightarrow V_C = \frac{\pi_C}{1 - \delta}, \tag{4}$$

where $\delta \in (0, 1)$ is the discount factor. The value that is derived by any firm that deviates by not colluding—which is denoted by V_D—is equal to

$$V_D = \pi_D + \frac{\delta}{1 - \delta}\pi_N. \tag{5}$$

The cartel is stable in the absence of CA interventions if and only if $V_C > V_D$, which, given Eqs. (4) and (5), is equivalent to

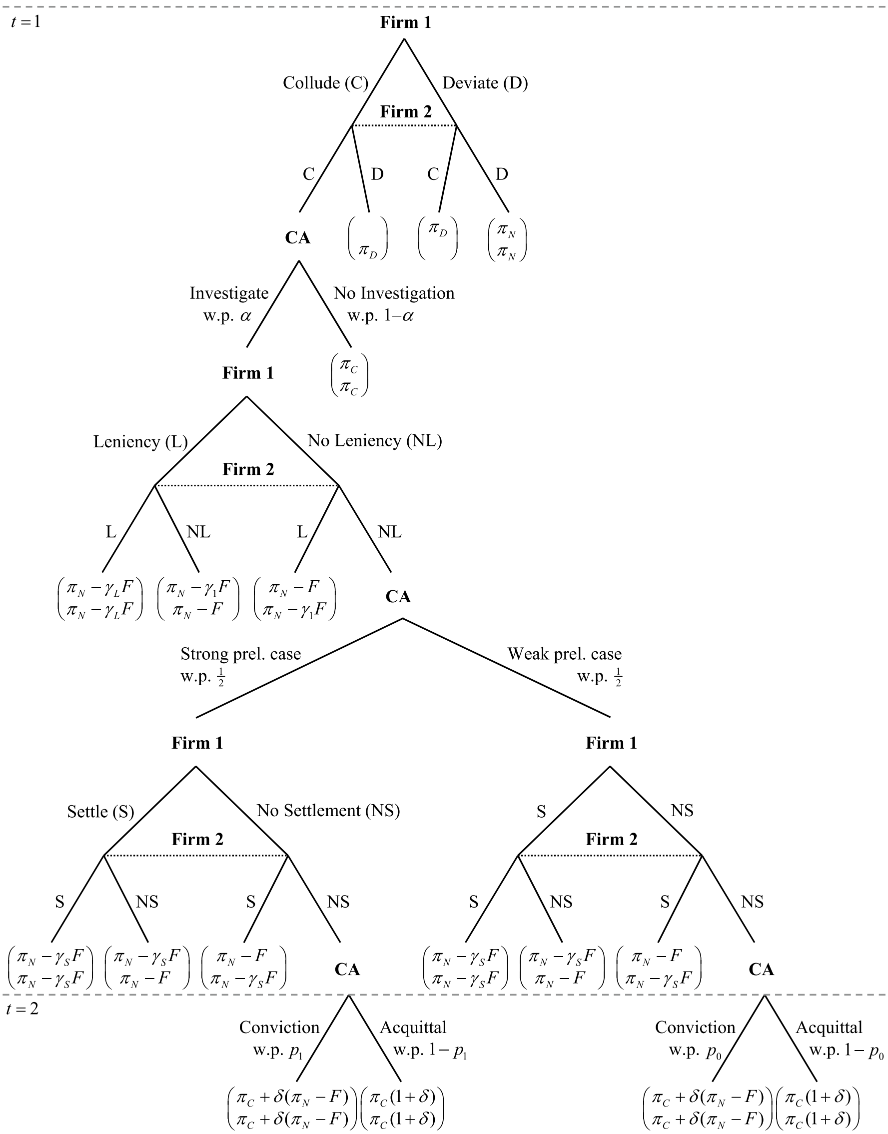


Fig. 1 Stage Game in the Two-firm Case

$$\delta > \delta_C := \frac{\pi_D - \pi_C}{\pi_D - \pi_N}. \tag{6}$$

This condition will be assumed to hold throughout the remainder of the paper.

Assumption 1 $\delta > \delta_C$.

In terms of notation, we let α_k denote the threshold value of the investigation probability α, such that a given collusive strategy $k = LXX, NNN, NSN$ is robust to deviations at the collusion stage if and only if $\alpha < \alpha_k$. We identify the incentive-compatibility thresholds that relate to deviations at the leniency and settlement stages by using a breve ($\breve{}$) to indicate the relevant deviation. Thus $\alpha_{\breve{L}NN}$ denotes the critical value of α, such that collusive strategy NNN is robust to deviations at the leniency stage if and only if $\alpha < \alpha_{\breve{L}NN}$, for example.

4.1 Collusive Strategy LXX

The value of the game when all firms play collusive strategy LXX is equal to

$$V_{LXX} = \alpha\left[\pi_N - \gamma_L F + \delta V_{LXX}\right] + (1-\alpha)\left[\pi_C + \delta V_{LXX}\right]. \tag{7}$$

The first term captures the situation in which the CA opens an investigation into the firms' industry, which happens with probability α. Firms immediately apply for leniency, earn the non-cooperative Nash profits π_N, and pay the reduced fine $\gamma_L F$ in that period before reverting to the same collusive strategy in the following period. The second term captures the situation in which the CA does not open an investigation. In this case firms earn the collusive profits without paying a fine in the current period and revert to the same collusive strategy in the next period.

The following proposition formalizes the condition under which all firms' playing LXX represents a subgame-perfect Nash equilibrium of the game:

Proposition 1 *For given policy parameters* $(F, \gamma_1, \gamma_L, \gamma_S, \alpha, p)$, *a subgame-perfect equilibrium in which all firms play collusive strategy LXX exists whenever*

$$\alpha < \alpha_{LXX} := \frac{\pi_C - (1-\delta)\pi_D - \delta\pi_N}{\pi_C - \pi_N + \gamma_L F}. \tag{8}$$

Proof All proofs can be found in the Online Appendix. □

It is interesting to see that the binding constraint on the LXX equilibrium always arises at the collusion stage as opposed to the leniency stage. Conditional on other firms' applying for leniency, it is never optimal to be the sole firm not to do so since the cartel will be convicted in any case.

4.2 Collusive Strategy NNN

Consider now the collusive strategy profile as part of which all firms play NNN, implying they neither apply for leniency nor settle their case under any circumstances. The value of the game under this strategy is equal to

$$V_{NNN} = \alpha\Big[\pi_C + \delta\big[p\,(\pi_N - F) + (1-p)\pi_C + \delta V_{NNN}\big]\Big] \\ + (1-\alpha)\Big[\pi_C + \delta\big(\pi_C + \delta V_{NNN}\big)\Big]. \tag{9}$$

The first line of Eq. (9) captures the situation in which an investigation is launched in the firms' industry. Since firms neither apply for leniency nor settle their case, they earn the collusive profits for one period before the CA proceeds to its full prosecution. This results in a conviction with probability p, in which case firms desist from collusion for one period and pay the full fine F. With probability $1-p$, the cartel is acquitted and earns the collusive profits for an additional period. Regardless of whether the cartel is convicted or not, it resumes the collusive strategy in the period that follows the conclusion of the CA's prosecution.

The second line captures the case in which no investigation is launched in the firms' industry, which implies that an investigation is instead opened in some other industries. By symmetry, the firms in these other industries are also playing NNN. This allows firms in the non-investigated industry to earn the collusive profits for two periods without the threat of being fined because the CA takes two periods to conclude its investigations in the industries that it does investigate.

The following proposition derives the condition under which all firms' playing NNN constitutes a subgame-perfect Nash equilibrium of the game:

Proposition 2 *For given policy parameters $(F, \gamma_1, \gamma_L, \gamma_S, \alpha, p)$, a subgame-perfect equilibrium in which all firms play collusive strategy NNN exists whenever*

$$\alpha < \min\left\{\alpha_{NNN}, \alpha_{\check{L}NN}, \alpha_{N\check{S}N}\right\}, \tag{10}$$

where

$$\alpha_{NNN} := \left(\frac{1+\delta}{\delta p}\right)\frac{\pi_C - (1-\delta)\pi_D - \delta\pi_N}{\pi_C - \pi_N + F}, \tag{11}$$

$$\alpha_{\check{L}NN} := \left(\frac{1+\delta}{\delta p}\right)\frac{\big(\pi_C - \pi_N\big)\big[1-(1-\delta)\delta p\big] + (1-\delta)F\big(\gamma_1 - \delta p\big)}{\delta^2\big(\pi_C - \pi_N + F\big)}, \tag{12}$$

$$\alpha_{N\check{S}N} := \left(\frac{1+\delta}{\delta p}\right)\frac{\big(\pi_C - \pi_N\big)\big[1-(1-\delta)\delta p_1\big] + (1-\delta)F\big(\gamma_S - \delta p_1\big)}{\delta^2\big(\pi_C - \pi_N + F\big)}. \tag{13}$$

We see that, in the case of the NNN equilibrium, the binding incentive-compatibility constraint may arise at the collusion, leniency, or settlement stage.

4.3 Collusive Strategy NSN

It remains to consider collusive strategy NSN, as part of which investigated cartels do not apply for leniency but agree to settle if the CA establishes a strong preliminary case. The value of collusive strategy NSN is equal to

$$V_{NSN} = \alpha \left\{ \begin{array}{c} \frac{1}{2}\left[\pi_N - \gamma_S F + \delta V_{NSN}\right] \\ +\frac{1}{2}\left[\pi_C + \delta\left[p_0\left(\pi_N - F\right) + (1-p_0)\pi_C + \delta V_{NSN}\right]\right] \end{array} \right\} + (1-\alpha)\left\{ \frac{1}{2}\left[\pi_C + \delta V_{NSN}\right] + \frac{1}{2}\left[\pi_C + \delta\left(\pi_C + \delta V_{NSN}\right)\right] \right\}. \tag{14}$$

The top line in Eq. (14) captures the case in which the CA opens an investigation into the firms' industry. Firms do not apply for leniency but instead wait for the CA's preliminary case to be established. With probability $\frac{1}{2}$ the case is strong, and firms settle: They pay their reduced fine $\gamma_S F$ and earn the non-cooperative profits π_N for one period before resuming their collusive strategy in the following period. With equal probability, the CA's case is weak, and the firms do not settle. This allows the firms to earn the collusive profits for one period before the CA proceeds to a full prosecution in the following period. Firms are convicted with probability p_0, in which case they desist from collusion and pay the fine F before resuming their collusive strategy in the following period. With probability $1 - p_0$, the cartel is acquitted, and firms earn the collusive profits for an additional period before resuming the collusive strategy.

The second line captures the case in which the CA does not open an investigation into the firms' industry, which implies that an investigation is instead opened in some other industries. By symmetry, the firms in these other industries also play NSN. With probability $\frac{1}{2}$, the CA's investigation in these other industries will be concluded in one period since a strong preliminary case is established in each of them. With equal probability, it will take two periods before the CA's investigation is concluded because the preliminary case is weak, which allows the firms in the non-investigated industry to earn two periods of collusive profits before resuming the collusive strategy.

The following proposition derives the condition under which all firms' playing NSN is a subgame-perfect equilibrium of the game:

Proposition 3 *For given policy parameters $(F, \gamma_1, \gamma_L, \gamma_S, \alpha, p)$, a subgame-perfect equilibrium in which all firms play collusive strategy NSN exists whenever*

$$\alpha < \min\left\{\alpha_{NSN}, \alpha_{\check{L}SN}, \alpha_{NS\check{S}}\right\}, \tag{15}$$

where

$$\alpha_{NSN} := (2+\delta)\,\frac{\pi_C - (1-\delta)\pi_D - \delta\pi_N}{\left(\pi_C - \pi_N\right)\left(1 + \delta p_0\right) + F\left(\gamma_S + \delta p_0\right)}, \tag{16}$$

$$\alpha_{\check{L}SN} := \frac{1}{\delta}\left(\frac{2+\delta}{1+\delta}\right)\frac{\left\{\begin{array}{c}(\pi_C-\pi_N)\left[2-(1-\delta)(1+\delta p_0)\right]\\ +(1-\delta)F(2\gamma_1-\gamma_S-\delta p_0)\end{array}\right\}}{\left(\pi_C-\pi_N\right)\left(1+\delta p_0\right)+F\left(\gamma_S+\delta p_0\right)}, \tag{17}$$

$$\alpha_{NS\check{S}} := \left(\frac{2+\delta}{\delta^2}\right)\frac{\left(\pi_C-\pi_N\right)\left[1-(1-\delta)\delta p_0\right]+(1-\delta)F\left(\gamma_S-\delta p_0\right)}{\left(\pi_C-\pi_N\right)\left(1+\delta p_0\right)+F\left(\gamma_S+\delta p_0\right)}. \tag{18}$$

We now consider the issue of equilibrium selection.

4.4 Equilibrium Selection

Propositions 1-3 establish very general conditions for collusive strategies LXX, NNN, and NSN—when played symmetrically by all firms—to constitute subgame-perfect Nash equilibria. It is apparent by inspection of these results that, for many parameter values, multiple equilibria may exist (see Fig. 2 for an illustration). Following the literature (e.g., MP and Blatter et al., 2018), we assume that, whenever multiple equilibria exist, firms are able to select the Pareto-dominant equilibrium by playing the collusive strategy that yields the highest value.

Colluding firms prefer not to cooperate with the CA if investigated—as compared with applying for leniency—whenever $V_{NNN} > V_{LXX}$. With the use of Eqs. (2), (3), (7), and (9), this is equivalent to $p < p_{LXX}^{NNN}$, where

$$p_{LXX}^{NNN} := \left(\frac{1+\delta}{\delta}\right)\frac{\pi_C-\pi_N+\gamma_L F}{\pi_C-\pi_N+F}. \tag{19}$$

In a similar way, colluding firms prefer not to cooperate with the CA – as compared with settling a strong case—whenever $V_{NNN} > V_{NSN}$. With the use of Eqs. (2), (3), (9), and (14), this is equivalent to $p < p_{NSN}^{NNN}$, where

$$p_{NSN}^{NNN} := \left(\frac{1+\delta}{\delta}\right)\frac{\left(\pi_C-\pi_N\right)(1-\delta d)+F\left(\gamma_S-\delta d\right)}{\pi_C-\pi_N+F}. \tag{20}$$

Finally, colluding firms prefer to settle a strong case—as compared with applying for leniency as soon as an investigation is launched—whenever $V_{NSN} > V_{LXX}$. With the use of (2), (3), (7), and (14), this is equivalent to $p < p_{LXX}^{NSN}$, where

$$p_{LXX}^{NSN} := \frac{\left(\pi_C-\pi_N\right)\left[1+\delta(1+d)\right]+F\left[(2+\delta)\gamma_L-\gamma_S+\delta d\right]}{\delta\left(\pi_C-\pi_N+F\right)}. \tag{21}$$

If we define the following threshold value for δ,

$$\delta_S := \frac{F}{d} \frac{(\gamma_S - \gamma_L)}{(\pi_C - \pi_N + F)}, \tag{22}$$

an algebraic comparison of (19), (20), and (21) leads to the following ranking of these profitability thresholds:

Lemma 1 *If*:

1. $\delta < \delta_S$: $p^{NSN}_{LXX} < p^{NNN}_{LXX} < p^{NNN}_{NSN}$,
2. $\delta > \delta_S$: $p^{NNN}_{NSN} < p^{NNN}_{LXX} < p^{NSN}_{LXX}$,
3. $\delta = \delta_S$: $p^{NNN}_{NSN} = p^{NNN}_{LXX} = p^{NSN}_{LXX}$.

Lemma 1 in turn allows us to rank the values of the collusive strategies as follows:

Proposition 4 *For given policy parameters* $(F, \gamma_1, \gamma_L, \gamma_S, \alpha, p)$, *the values of the three collusive strategies LXX, NNN and NSN are ranked as follows*:

1. *Given* $\delta < \delta_S$:

(a) $V_{NNN} > V_{NSN} > V_{LXX}$ *if* $p < p^{NSN}_{LXX}$,
(b) $V_{NNN} > V_{LXX} > V_{NSN}$ *if* $p \in \left(p^{NSN}_{LXX}, p^{NNN}_{LXX}\right)$,
(c) $V_{LXX} > V_{NNN} > V_{NSN}$ *if* $p \in \left(p^{NNN}_{LXX}, p^{NNN}_{NSN}\right)$,
(d) $V_{LXX} > V_{NSN} > V_{NNN}$ *if* $p > p^{NNN}_{NSN}$.

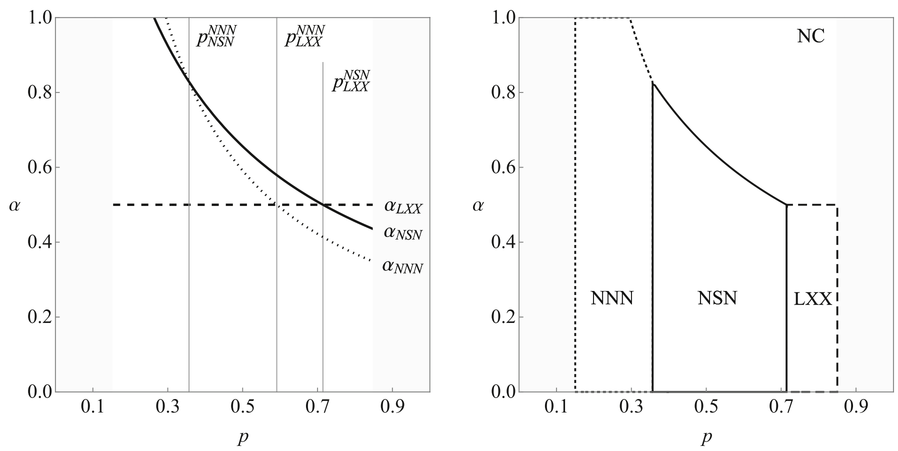
2. *Given* $\delta > \delta_S$:

(a) $V_{NNN} > V_{NSN} > V_{LXX}$ *if* $p < p^{NNN}_{NSN}$,
(b) $V_{NSN} > V_{NNN} > V_{LXX}$ *if* $p \in \left(p^{NNN}_{NSN}, p^{NNN}_{LXX}\right)$,
(c) $V_{NSN} > V_{LXX} > V_{NNN}$ *if* $p \in \left(p^{NNN}_{LXX}, p^{NSN}_{LXX}\right)$,
(d) $V_{LXX} > V_{NSN} > V_{NNN}$ *if* $p > p^{NSN}_{LXX}$.

3. *Given* $\delta = \delta_S$:

(a) $V_{NNN} > V_{NSN} > V_{LXX}$ *if* $p < p^{NNN}_{NSN} = p^{NNN}_{LXX} = p^{NSN}_{LXX}$,
(b) $V_{LXX} > V_{NSN} > V_{NNN}$ *if* $p > p^{NNN}_{NSN} = p^{NNN}_{LXX} = p^{NSN}_{LXX}$.

We are now in a position to solve for the equilibrium outcome of the model. Given the large number of cases that can arise with respect to the incentive-compatibility constraints that bind for collusive strategies NNN and NSN—as well as the ranking of the profitability thresholds that determine the Pareto-dominant equilibrium—we do not provide a complete analytical characterization of the equilibrium for all possible parameter values. Instead, we show how the relevant



(a) Stability and Profitability (b) Equilibrium Strategies

Fig. 2 Solving the Model with Settlements

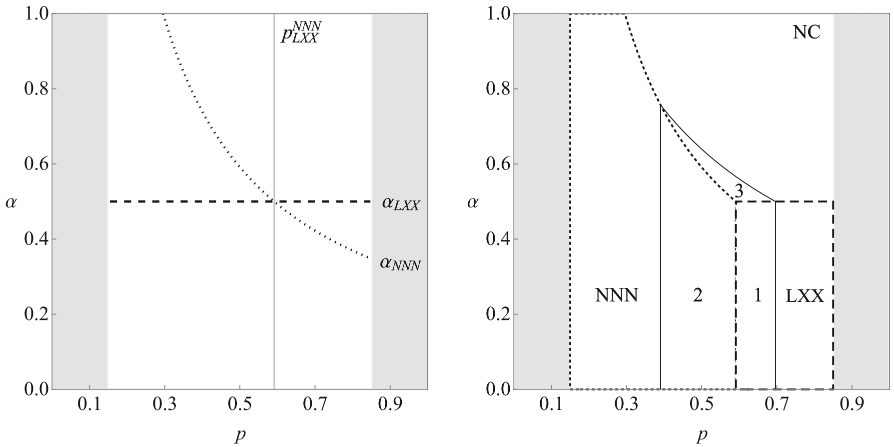
constraints are combined to identify the equilibrium for specific parameter values. This solution method and the example underlying it will nonetheless allow us to identify the effects that the addition of a settlement procedure has on equilibrium outcomes and to characterize features of the optimal policy in Sect. 5.

Consider, therefore, a scenario in which $F = 4$, $d = 0.15$, $\gamma_1 = \gamma_L = 0.1$, $\gamma_S = 0.13$, $\pi_C = 1$, $\pi_D = 3$, $\pi_N = 0$ and $\delta = 0.9$. Panel (a) of Fig. 2 plots the binding incentive-compatibility constraints for each of the three collusive strategies and the profitability thresholds that correspond to this example in (α, p) space. Panel (b) identifies the associated subgame-perfect equilibrium or, where multiple equilibria exist, the Pareto-dominant equilibrium.

In this case, the binding incentive-compatibility constraints on NNN and NSN arise at the collusion stage rather than at the leniency or settlement stages for all $p \in (0.15, 0.85)$. From Proposition 1, the same is necessarily true for collusive strategy LXX. These incentive-compatibility constraints are plotted in panel (a) of Fig. 2. Panel (a) also shows the profitability thresholds in (19)-(21). Since $\delta = 0.9 > 0.16 = \delta_S$ in this example, see (22), the ranking of these thresholds falls within case (ii) of Proposition 4.

From Proposition 4, collusive strategy NNN is value-maximizing for $p < p_{NSN}^{NNN}$. Moreover, in panel (a) we see that whenever NSN or LXX is incentive-compatible in this range of p, so is NNN. This follows because α_{NNN} lies above both α_{LXX} and α_{NSN} for all $p < p_{NSN}^{NNN}$. Therefore, for $p < p_{NSN}^{NNN}$, NNN is the equilibrium of the game – either uniquely or by Pareto dominance – whenever $\alpha < \alpha_{NNN}$. This region is highlighted by the dotted outline labeled NNN in panel (b) of Fig. 2.

Similarly, for p between p_{NSN}^{NNN} and p_{LXX}^{NSN}, NSN is value-maximizing. In panel (a) of Fig. 2 we see that whenever NNN or LXX is incentive-compatible in this range of p, so is NSN. Therefore NSN is the unique equilibrium or the Pareto-dominant equilibrium whenever α is low enough to ensure that NSN is incentive-compatible:



(a) Stability and Profitability (b) Equilibrium Strategies

Fig. 3 Solving the Model without Settlements

whenever $\alpha < \alpha_{NSN}$. This region corresponds to the solid outline labeled NSN in panel (b) of Fig. 2.

For $p > p_{LXX}^{NSN}$, LXX is value-maximizing. We see in panel (a) that, if either NNN or NSN is incentive-compatible for $p > p_{LXX}^{NSN}$, so is LXX. Therefore LXX is the equilibrium of the game for $p > p_{LXX}^{NSN}$ whenever $\alpha < \alpha_{LXX}$. This region is identified with the dashed outline that is labeled LXX in panel (b) of Fig. 2.

Finally, whenever $\alpha > \max\{\alpha_{LXX}, \alpha_{NNN}, \alpha_{NSN}\}$, so that the investigation probability α lies above all three binding incentive-compatibility constraints, none of the three collusive strategies is incentive-compatible. In this case, the cartel is deterred. This cartel-deterrence region is identified with the label NC in panel (b) of Fig. 2.

5 Policy Implications

This section considers two important policy questions that arise in the context of leniency and settlements: First, we explore the impact that the introduction of a settlement procedure has on equilibrium outcomes. We then discuss the interactions between leniency and settlements in the context of the optimal policy.

5.1 The Impact of a Settlement Procedure

In order to study the impact that the introduction of a settlement procedure has on equilibrium outcomes, we contrast the example in Fig. 2 with an amended version of the same example in which the option to settle does not exist. While settling firms are assumed to pay a reduced fine rather than an increased fine under the settlement procedure that was described in Sect. 3, in mathematical terms, we can eliminate settlements from the model by letting $\gamma_S \to \infty$. As γ_S grows, collusive strategy NSN is certain not to be incentive-compatible; see Proposition 3. Similarly, the option to deviate into a settlement can no longer constrain the incentive-compatibility of collusive strategy NNN in that case; see (13).

Holding the remaining parameters fixed at the values used in Fig. 2, the incentive-compatibility constraints, profitability thresholds, and equilibrium outcomes are shown in Fig. 3.

From Eq. (22), high values of γ_S guarantee that we are in case (i) of Proposition 4. The thresholds p_{NSN}^{NNN} and p_{LXX}^{NSN} grow arbitrarily large and small, respectively, as $\gamma_S \to \infty$, so they do not appear in Fig. 3. Proposition 4 implies that NNN is value-maximizing for all $p < p_{LXX}^{NNN}$. We see in panel (a) that, whenever LXX is incentive-compatible for p in this range, so is NNN. Therefore NNN is either the unique equilibrium or the Pareto-dominant equilibrium for $p < p_{LXX}^{NNN}$ whenever $\alpha < \alpha_{NNN}$.[12] Similarly, LXX is value-maximizing for all $p > p_{LXX}^{NNN}$ and, by identical arguments, constitutes the unique equilibrium or the Pareto-dominant equilibrium for these values of p whenever $\alpha < \alpha_{LXX}$.

The regions of the (α, p) space in which NNN and LXX represent the equilibrium of the game are again highlighted by the dotted and dashed outlines, respectively, in panel (b) of Fig. 3. In order to compare these outcomes with those arising in Fig. 2, where settlements were available, we superimpose the NSN region from panel (b) of Fig. 2 onto panel (b) of Fig. 3 with the use of a solid black line. Three distinct effects associated with the settlement procedure then emerge, which we identify with the numerals 1-3 in panel (b) of Fig. 3:

In region 1, firms would have applied for leniency if investigated in the absence of a settlement procedure. Given the option to settle, they prefer instead to play collusive strategy NSN, which implies that they cooperate with the CA only when a strong preliminary case is formed against them. In this region, settlements therefore act as a *substitute* to the leniency program.

In region 2, firms that would have played NNN in the absence of settlements prefer instead to settle whenever the CA forms a strong preliminary case. In this region, the settlement procedure acts as a *complement* to the leniency program. It encourages cooperation by cartels that cannot otherwise be persuaded to cooperate with the CA through the leniency program.

Region 3 captures a pro-collusive effect of the settlement procedure: In this region, neither LXX nor NNN is incentive-compatible, but when settlements are available, NSN is incentive-compatible. In this region, firms exploit the flexibility that the settlement procedure offers them: the option to await learning about the strength of the preliminary case before deciding whether or not to cooperate with the CA. This allows the formation of stable cartels that could not have formed in the absence of the settlement procedure.[13]

A fourth, related effect of the settlement procedure arises when the binding incentive-compatibility constraint on collusive strategy NNN arises at the settlement stage

[12] Recall that, in this example, α_{NNN} lies below $\alpha_{\check{L}NN}$ for all $p \in (0.15, 0.85)$.

[13] Of course, effects 1-3 arise only when firms choose to play NSN in equilibrium for some levels of α and p. This motivates the choice of parameter values in Fig. 2.

as opposed to the collusion stage (as it does in our example) or at the leniency stage. In that case, the binding incentive-compatibility constraint on NNN in the presence of settlements—$\alpha_{N\check{S}N}$—lies below the binding incentive-compatibility constraint on NNN in the absence of settlements: α_{NNN}. Although this effect is not present in our example because $\alpha_{NNN} < \alpha_{N\check{S}N}$, we can replicate it graphically by shifting up the α_{NNN} line in Fig. 3 while holding the NSN region in Fig. 2 fixed. In terms of deterrence, this offsets the pro-collusive effect of settlements that was discussed above by promoting the creation of stable NNN cartels when settlements are unavailable. For the purpose of clarity in Figs. 2 and 3, we show an example that omits this final effect of the settlement procedure.

5.2 Optimal Policy

We now consider how the CA's policy parameters should be set in order to maximize welfare in the presence of a settlement procedure. First, we make the CA's budget constraint precise. Then we derive the welfare impact that is associated with the various collusive strategies, before drawing conclusions for the optimal design of policy.

5.2.1 Budget Constraint

The budget constraint defines the feasible levels of α and p that can be implemented under the CA's enforcement technology. Following MP, we suppose that the CA is endowed with an exogenous, sunk per-period budget that can be allocated either towards investigating more cartels—to increasing α—or to increasing the baseline probability of conviction p (see Sect. 3).

To be precise, let T denote the total number of industries and n denote the number of industries that the CA can investigate in any given period.[14] Since firms choose collusive strategies symmetrically in equilibrium, investigations will last one period when firms choose LXX, two periods if firms choose NNN, and one or two periods with equal probability when firms choose NSN. By symmetry, the CA's investigations conclude at the same time in all investigated industries, regardless of which collusive strategy is being played. This implies that, whenever the CA launches new investigations, the probability of an industry's becoming the subject of an investigation is equal to

$$\alpha = \frac{n}{T}.$$

It follows that, for a fixed number of total industries T, the choice of a given probability of investigation α is equivalent to the decision to open investigations in $n = \alpha T$ of those industries.

We will assume that the total cost C of achieving a given probability of investigation α and baseline conviction probability p is equal to

[14] For ease of exposition, we follow MP in treating n as a continuous variable.

$$C = \alpha + mp,$$

where $m > 0$ measures the cost to the CA of increasing p relative to α.

For a given level of the budget B, this implies that the CA's budget constraint is linear and downward sloping in (α, p) space,

$$\alpha = B - mp. \tag{23}$$

This budget constraint reflects the fundamental economic trade-off that a CA faces between maximizing the probability of investigation α and maximizing the baseline conviction probability p. We may, for instance, think of the CA as allocating a fixed number of staff to either investigation or prosecution activities, although we do not model these aspects of the CA's internal operations here.[15]

5.2.2 Welfare Gains

Competition law enforcement generates a welfare gain in this setting whenever the CA deters a cartel or causes a convicted cartel to desist from collusion. As in MP, we evaluate these welfare gains with the use of a utilitarian welfare function with equal weights on consumer and producer surplus. Fines are treated as pure transfers that do not affect total welfare.

In this setting, the welfare gains that are associated with a successful CA intervention that brings about a competitive market equilibrium are reflected in the deadweight loss (DWL) that collusion generates relative to this competitive outcome. To calculate these welfare gains, we compare each possible equilibrium outcome—NC, LXX, NNN and NSN—to the outcome that arises in the absence of enforcement, which is certain to feature collusion given Assumption 1.

If cartels are deterred in every industry, the CA achieves a welfare gain of DWL in each of the T industries in every period. The total welfare gain that is achieved by the CA relative to the no enforcement case is therefore equal to

$$W_{NC} = T\left(\frac{DWL}{1-\delta}\right). \tag{24}$$

Clearly, this welfare gain is independent of both α and p.

If all of the firms play collusive strategy LXX, the CA opens n investigations in each period. Each investigation induces the cartel in that industry to apply for leniency and therefore to play non-cooperatively for one period. Thereafter, n new industries are selected randomly as the subject of a new set of CA investigations. Therefore the welfare gain in an LXX equilibrium relative to the no enforcement case is

$$W_{LXX}(\alpha) = n\left(\frac{DWL}{1-\delta}\right) = \alpha W_{NC}. \tag{25}$$

[15] MP specify a budget constraint that is more detailed in these respects, although, as those authors note, their general conclusions hold for any downward-sloping budget constraint.

In an NNN equilibrium, cartels that become the subject of an investigation do not cooperate with the CA. Instead, they desist from collusion for one period if and only if the CA succeeds in convicting them after the prosecution phase: after an investigation that takes two periods to conclude. This conviction occurs with probability p. Therefore the welfare gain relative to the no enforcement case is

$$W_{NNN}(\alpha,p) = n\,p\,\delta\left(\frac{DWL}{1-\delta^2}\right) = \alpha\,p\left(\frac{\delta}{1+\delta}\right)W_{NC}. \tag{26}$$

If all of the firms play NSN, investigated cartels cooperate with the CA if and only if it forms a strong preliminary case against them. This happens with probability $\frac{1}{2}$. In this case, the firms settle, and the welfare gain DWL is achieved in the same period. With equal probability, the CA's preliminary case in each industry is weak. Now the procedure takes an additional period to conclude, and the welfare gain DWL is achieved only when a cartel is convicted, which happens with probability p_0. Therefore the welfare gain relative to the no enforcement case can be written implicitly as

$$W_{NSN} = \frac{1}{2}\left[nDWL + \delta W_{NSN}\right] + \frac{1}{2}\left[\delta\, n\, p_0\, DWL + \delta^2 W_{NSN}\right].$$

Solving this expression for W_{NSN}, we have

$$W_{NSN}(\alpha,p) = n\left[\frac{DWL(1+\delta p_0)}{(1-\delta)(2+\delta)}\right] = \alpha\left[\frac{1+\delta(p-d)}{2+\delta}\right]W_{NC}. \tag{27}$$

A comparison of the welfare gains in (24)-(27) reveals that cartel deterrence achieves the first-best welfare outcome: It generates the highest welfare gain for all feasible parameter values.

Proposition 5 $W_{NC} > \max\{W_{LXX}, W_{NNN}, W_{NSN}\}$ *for all parameter values, so that cartel deterrence achieves the first-best welfare gain.*

We now consider how the optimal policy is determined.

5.2.3 Optimal Policy

Our analysis of the optimal policy proceeds in two steps. First, we explore how the probability of investigation α and baseline conviction probability p should be set for given values of the remaining policy parameters. Taking these results as given, we then investigate how F, γ_1, γ_L, and γ_S should be set optimally. This latter discussion will, in turn, consider the policy that is most likely to achieve the first-best—cartel deterrence—and the policy that should be chosen when deterring cartels is not feasible due to the CA's limited resources: in the setting of the second-best.

When F, γ_1, γ_L, and γ_S are held fixed, the incentive-compatibility constraints and profitability thresholds that underlie Propositions 1–4 are likewise fixed and characterize—in line with the discussion in Sect. 4.4—the equilibrium outcomes of the game in (α, p) space. The optimal policy will select from among the feasible levels of α and p—as reflected in the budget constraint in (23)—those that lead to the highest welfare gain. Consider, to that end, the iso-welfare gains curves in (α, p) space that are associated with each of the three collusive strategies:

From (25), in the region where LXX is the equilibrium of the game, the iso-welfare gains curve that is associated with a given welfare gain W is

$$\overline{\alpha}_{LXX} = \frac{W}{W_{NC}}. \tag{28}$$

Since W_{NC} is independent of p, see (24), this iso-welfare gains curve is also independent of p. Hence the iso-welfare gains curves in the LXX region are horizontal lines that are shifted upwards for higher values of W.

From (26), in the region where NNN is the equilibrium of the game, the iso-welfare gains curve for a given welfare gain W is equal to

$$\overline{\alpha}_{NNN}(p) = \left(\frac{W}{W_{NC}}\right)\left(\frac{1+\delta}{\delta p}\right). \tag{29}$$

This describes an equilateral hyperbola in (α, p) space with a horizontal (vertical) asymptote at $\alpha = 0$ ($p = 0$), which shifts upwards as W increases.

Finally, in the NSN region, the iso-welfare gains curve that is associated with a given welfare gain W is, from (27), equal to

$$\overline{\alpha}_{NSN}(p) = \left(\frac{W}{W_{NC}}\right)\left[\frac{2+\delta}{1+\delta(p-d)}\right]. \tag{30}$$

This describes another equilateral hyperbola – now with a horizontal (vertical) asymptote at $\alpha = 0$ ($p = d - \frac{1}{\delta}$) – which is again shifted upwards for higher values of W.

To fix ideas, Fig. 4 illustrates the maximum attainable iso-welfare gains curves in a case where all three collusive strategies may be played in equilibrium and in which the budget constraint, labeled BC, lies below the cartel deterrence region NC for all p, so that cartel deterrence is not feasible.[16]

In the LXX region, the iso-welfare gains curves are horizontal lines. Given the downward-sloping budget constraint, the highest possible welfare gain in this region is achieved at a corner solution at the left-hand boundary of the LXX region.[17] This point is identified by the circle in Fig. 4. Let α^*_{LXX} and p^*_{LXX} denote these welfare-maximizing parameter values and let W^*_{LXX} denote the corresponding welfare gain as calculated according to (25).

Given the shape of the iso-welfare gains curves in the NNN and NSN regions, the highest welfare gain in these regions may—depending on the slope of the budget constraint—be achieved at an interior or corner solution. The steeper is the budget

[16] Given Proposition 5, we know that whenever the NC outcome is feasible, it will necessarily be implemented under the optimal policy.

[17] This is given by $p = p^{NNN}_{LXX}$ in case (i) and by $p = p^{NSN}_{LXX}$ in case (ii) of Proposition 4.

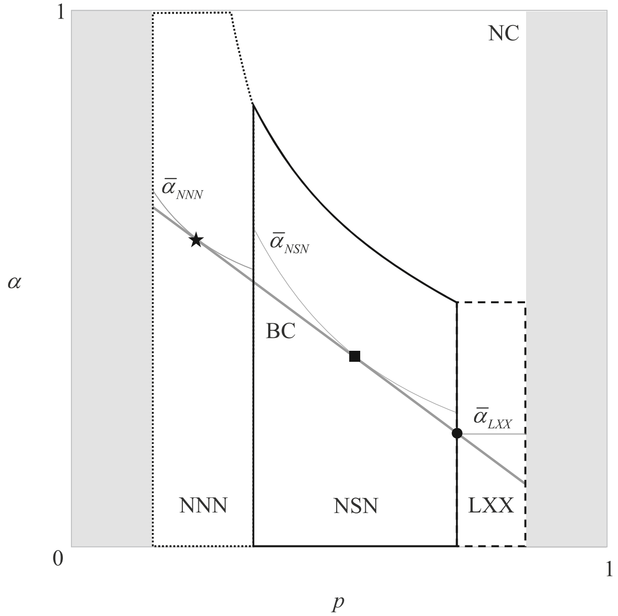


Fig. 4 Illustration of Optimal α, p in LXX, NNN and NSN Regions

constraint, the higher is the relative cost of increasing p relative to α, and consequently the more likely it is that the optimum in these regions is achieved at a corner solution that maximizes α relative to p. For illustrative purposes, Fig. 4 depicts a case in which the maximal welfare gain is achieved at an interior solution, which is denoted by a star in the NNN region and by a square in the NSN region.[18] Let α^*_{NNN} and p^*_{NNN} denote the welfare-maximizing parameter values in the NNN region and let W^*_{NNN} denote the welfare gain that is calculated at this point with the use of (26). Similarly, let α^*_{NSN} and p^*_{NSN} denote the welfare-maximizing parameter values in the NSN region and let W^*_{NSN} denote the welfare gain that is calculated at this optimum according to (27).

Given these welfare-maximizing points within the LXX, NNN and NSN regions, the final determination of the optimal policy rests on a comparison of the welfare levels achieved at each of them. Proposition 5 shows that cartel deterrence achieves the first-best welfare gain. Therefore, if the NC region is attainable, any α and p within this region will equivalently achieve the first-best since W_{NC} is independent of α and p; see (24). If cartel deterrence is not feasible, the optimal policy implements the collusive strategy that yields the highest welfare gain.

This idea is formalized in the following result:

[18] Note that this distinction between interior and corner solutions is not important for the discussion of the optimal policy that follows.

Proposition 6 Hold fixed F, γ_1, γ_L, and γ_S. If cartel deterrence is feasible, the optimal policy sets (α^*, p^*) in the NC region and attains the first-best. If cartel deterrence is not feasible, the optimal policy implements:

- LXX by setting $\alpha^* = \alpha^*_{LXX}$ and $p^* = p^*_{LXX}$ if $W^*_{LXX} > \max\left\{W^*_{NNN}, W^*_{NSN}\right\}$,
- NNN by setting $\alpha^* = \alpha^*_{NNN}$ and $p^* = p^*_{NNN}$ if $W^*_{NNN} > \max\left\{W^*_{LXX}, W^*_{NSN}\right\}$,
- NSN by setting $\alpha^* = \alpha^*_{NSN}$ and $p^* = p^*_{NSN}$ if $W^*_{NSN} > \max\left\{W^*_{LXX}, W^*_{NNN}\right\}$.

If any of these collusive strategies cannot arise in equilibrium, the corresponding optimum point is not defined, and it is dropped from the above comparison.

Consider now how the remaining policy parameters—F, γ_1, γ_L, and γ_S—should be set. Proposition 6 states that the CA selects α and p in the NC region whenever feasible since this achieves the first-best. The aim of the CA is therefore to maximize the size of the NC region so that, for a fixed budget constraint, the chances that part of the NC region lies within the feasible set of α and p – below the budget constraint – are maximized. The following result considers this optimal deterrence policy:

Proposition 7 *The policy most likely to achieve deterrence sets* $F^* = \overline{F}$ *and* $\gamma_1^* = 0$.

Intuitively, a higher fine increases the costs of collusion and therefore tends to deter cartels by shifting down the incentive-compatibility constraints. Offering the first leniency applicant the maximum fine reduction by charging that party a zero fine maximizes deviation incentives at the leniency stage, and thus reduces the stability of NNN and NSN-type cartels; see (12) and (17). Note that this result also echoes similar findings in the literature, including those of MP, Chen and Harrington (2007), and Houba et al. (2015).

The levels of γ_L and γ_S that maximize deterrence cannot be determined at a general level and must, instead, be calculated numerically for a given budget constraint. On the one hand, offering settling firms more generous fine reductions by reducing γ_S improves deterrence by increasing the incentives to deviate from collusive strategy NNN at the settlement stage; see (13). On the other hand, reducing γ_S increases the value of collusive strategy NSN as part of which firms plan to settle on the equilibrium path, which increases the stability of cartels of type NSN; see (16) and (17). This, in turn, complicates the determination of γ_L. Although higher values of γ_L improve deterrence by reducing the value of collusive strategy LXX—see (7)—γ_L is bounded above by γ_S (see Sect. 3.1).

These interactions between the leniency program and the settlement procedure become much clearer in the second-best setting: The CA's limited budget implies that it may be impossible to deter cartels—even when the policy parameters are set in order to maximize the size of the NC region. When the first-best is not attainable, the second-best policy will—following Proposition 6—implement the collusive strategy that generates the highest welfare gain. Therefore, in this second-best world, the objective of the CA shifts from attempting to deter cartels to influencing the collusive strategy that non-deterred cartels adopt in equilibrium.

Our analysis of this second-best is facilitated by the following intermediate result. Comparing (25), (26), and (27) at given α and p, we have:

Lemma 2 *For a given level of* α *and* p, *we have* $W_{LXX} > W_{NSN} > W_{NNN}$.

Intuitively, the LXX equilibrium generates the highest welfare gain for given α and p because it maximizes the frequency with which colluding firms cooperate with the CA in equilibrium. Note also that this result differs from the welfare gains comparison that underlies Proposition 6, which was carried out with the use of different values of α and p that, in each case, corresponded to the optimal values for a given collusive strategy.

The following result considers the optimal policy in the second-best setting.

Proposition 8 *If cartel deterrence cannot be achieved for any parameter values, the optimal policy sets* $F^* = \overline{F}$ *and* $\gamma_L^* = 0$. *If the resulting* (α^*, p^*) *falls within the LXX region according to Proposition 6 when* $\gamma_S = 1$, *the optimal policy also sets* $\gamma_S^* = 1$. *Otherwise, it sets* γ_S^* *arbitrarily close to zero.*

The intuition for this result is most easily understood by considering Fig. 4: When cartel deterrence is not feasible, the objective of the CA is to maximize the size of the LXX region relative to both the NNN and NSN regions. If the optimal (α^*, p^*) is (according to Proposition 6) implemented in the NNN or NSN regions, then by expanding the LXX region the CA may succeed in moving this optimum point into the LXX region, which increases welfare (according to Lemma 2). If the optimum falls within the LXX region to begin with then—given the downward-sloping budget constraint and horizontal iso-welfare gains curves in the LXX region—expanding the size of this region allows a higher iso-welfare gains curve to be reached. It is easy to see by inspection of the profitability thresholds in (19)–(21) that the size of the LXX region is maximized by setting $\gamma_L = 0$ and $\gamma_S = 1$.

If the optimal (α^*, p^*) does not fall within the LXX region when $\gamma_L = 0$ and $\gamma_S = 1$, the CA should maximize the size of the NSN region relative to NNN by offering the most generous possible fine reduction to settling firms: $\gamma_S \to 0$. This implies an increase in welfare whenever an optimum that previously fell within the NNN region now falls within the NSN region (Lemma 2). Although reducing γ_S reduces the size of the LXX region, this cannot imply a drop in welfare by moving an optimum that previously fell within the LXX region into NSN, since the optimum was not implemented in LXX – even when this region was at its maximum size.

This highlights an important feature of the optimal joint design of the leniency program and the settlement procedure: Whenever cartel deterrence is not feasible due to the CA's limited resources, the CA should maximize the frequency with which colluding firms cooperate with it in equilibrium. It does so by making the leniency program as generous as possible relative to settling: by setting $\gamma_L = 0$ and $\gamma_S = 1$. If firms do not apply for leniency in these circumstances, the CA should maximize the appeal of the settlement procedure by extending generous fine reductions to settling firms as well. This improves welfare by encouraging otherwise uncooperative NNN-type cartels to cooperate with the CA by settling.

Finally, since a higher fine F can be shown to shift all three profitability thresholds leftwards, setting the highest fine possible maximizes the size of the LXX region relative to both NNN and NSN, and maximizes the size of the NSN region relative to NNN. The fine reduction that is offered to the first leniency applicant—γ_1—does not affect the profits of colluding firms on the equilibrium path, and hence its optimal value is not pinned down by these second-best considerations.

6 Robustness & Extensions

In this section, we show that our results are robust to the inclusion of private actions for damages (PADs) and an ex ante leniency program. We also discuss extensions of the model to alternative fine structures.

6.1 Private Actions for Damages

A growing volume of research has highlighted the importance of PADs as a component of the total penalty paid by convicted cartels (see, e.g., Buccirossi et al., 2020; Buiten et al., 2018; Emons and Lenhard, 2020; Katsoulacos et al., 2020a; Spagnolo, 2004). We may integrate damages into our model in a simple way by assuming that leniency applicants, settling firms, and cartels that are convicted following a full prosecution will be found liable for damages by a court with probability σ_L, σ_S, and σ_N, respectively.[19] If we suppose that the damages payable are, in each case, equal to a proportion ξ of the cartel fine F,[20] the total expected penalties payable when one firm applies for leniency, when all firms apply for leniency, when firms settle, and when firms are convicted following a prosecution may be written, respectively, as: $\phi_1 F = (\gamma_1 + \sigma_L\xi)F$; $\phi_L F = (\gamma_L + \sigma_L\xi)F$; $\phi_S F = (\gamma_S + \sigma_S\xi)F$; and $\phi_N F = (1 + \sigma_N\xi)F$.

The integration of damages into the model is therefore achieved by replacing γ_1, γ_L, and γ_S with ϕ_1, ϕ_L, and ϕ_S, as defined above, and by increasing the fine paid by firms that do not cooperate with the CA to $\phi_N F$. This leaves our equilibrium analysis in all other respects unchanged. Moreover, since the ϕ parameters defined above are, in each case, proportional to the respective γ's, the marginal effects that underlie our analysis of the optimal fine reductions under leniency and settlements carry over to this setting, which leaves our policy results unaffected.

[19] In practice, it is reasonable to suppose that $\sigma_N > \max\{\sigma_L, \sigma_S\}$. The EU Damages Directive (Directive 2014/104/EU) prevents the information that is contained in leniency statements and settlement submissions from being used in follow-on damages claims (Motchenkova and Spagnolo, 2019).

[20] This can be motivated by the idea that cartel fines capture an important element of the economic harm of collusion. Since the optimal fine will continue to be set at $F = \overline{F}$ in this version of the model, we can think of the upper bound on fines $\overline{F}$ that is defined by competition law as reflecting the harm of cartels in symmetric industries of this type.

6.2 Ex Ante Leniency

We can also show that offering firms the opportunity to apply for leniency after a cartel forms but before an investigation is opened leaves our results unchanged. To illustrate, consider collusive strategy NNN: The value of this strategy – V_{NNN} – is given in (9). Deviating from this collusive strategy by applying for leniency before an investigation is opened yields deviation profits $\widetilde{V}_{\check{L}NN} = \frac{\pi_N}{1-\delta} - \gamma_1 F$. NNN is robust to these deviations whenever $V_{NNN} > \widetilde{V}_{\check{L}NN}$. Since $\frac{\pi_N}{1-\delta} - \gamma_1 F < V_D$, see (5), NNN must be robust to these deviations into an ex ante leniency application whenever it is robust to deviations at the collusion stage.

In a similar way, it is straightforward to show that deviating from NSN and LXX by applying for ex ante leniency does not lead to additional constraints on the incentive-compatibility of those strategies. In each case, the deviation profits lie below V_D, while the profits of remaining with NSN and LXX—V_{NSN} and V_{LXX}—must lie above V_D in order for those strategies to be robust to deviations at the collusion stage.

Moreover, the profits that are derived from the collusive strategy as part of which all firms agree to apply for ex ante leniency on the equilibrium path can be written as $\widetilde{V}_{LXX} = \pi_N - \gamma_L F + \delta \widetilde{V}_{LXX} \Rightarrow \widetilde{V}_{LXX} = \frac{\pi_N - \gamma_L F}{1-\delta}$. It follows immediately from (5) that $\widetilde{V}_{LXX} < V_D$, so that cartels that intend to apply for ex ante leniency cannot be incentive-compatible. Adding the possibility of ex ante leniency therefore does not affect the results of this paper.

6.3 Fine Structure and Recidivism

Our model carries over the assumption of a fixed fine from MP. By contrast, related research has considered the issue of optimal fines (see Katsoulacos et al., 2015; 2019a; b; 2020b). One important issue that is highlighted by this literature is the fact that revenue-based fines incentivize cartels to increase their price above the monopoly level in order to reduce their revenue and, hence, their expected fines. Since we consider the decision of firms to form a cartel and abstract from the pricing decisions that cartels adopt, this issue does not arise in our set-up.

Instead, our model focuses on the interactions between leniency and settlements, which, in turn, have not been considered in the optimal fine literature to date. While combining these strands of research remains an important objective for future work, the fundamental trade-offs between the fine reductions offered to leniency applicants and settling firms that we identify here should be robust to the particular method of calculating cartel fines.

Other related research makes the probability of investigation endogenous to the pricing behavior adopted by a cartel (see, e.g., Chen and Harrington, 2007, which was discussed in Sect. 2). These effects are also absent in our set-up that abstracts from cartel pricing decisions. By contrast, in line with MP, we make the probability of investigation endogenous to the CA's optimal budget allocation decision.

A third consideration in relation to fines concerns cartel duration. Several jurisdictions make the fine dependent on the duration of a cartel (US Department of

Justice, 1993; European Commission, 2002). If the fine increases over the lifetime of a cartel, colluding firms may abandon collusive agreements after some time in order to reduce expected fines. The idea that the CA should encourage cooperation among cartels of a given duration through the leniency program and, failing that, through the settlement procedure should still hold in this non-stationary setting, however.

Finally, in line with MP and the EU leniency program (see European Commission, 2006; Marvão, 2016), we do not restrict access to the leniency program for repeat offenders (recidivists). In this regard, it is worth highlighting the legal distinction between repeat offenders (who re-cartelize a market in which they were previously convicted) and multiple offenders (who establish cartels in markets other than the one in which they were previously convicted) (Marvão, 2022). According to Werden et al. (2011), while multiple offenders are common, legal nuances around market and firm definition imply that repeat offenders are rarely, if ever, identified in practice. If repeat offenders are excluded from the leniency program, we should expect it to become less relevant: Firms that might otherwise have applied for leniency as part of their agreed collusive strategy would instead settle or refuse to cooperate with the CA.

Moreover, denying leniency to repeat offenders reduces the number of deviations that leniency applicants have available to them in later periods, which can stabilize collusion. Indeed, Chen and Rey (2013) show that when leniency is restricted to first offenders and firms play a collusive strategy as part of which they apply for leniency once and then never again, the leniency program is rendered entirely ineffective.

7 Conclusion

This paper studies the interactions between a cartel leniency program and a settlement procedure. We derive the conditions under which colluding firms apply for leniency, settle, or refuse all cooperation with the CA in equilibrium. By comparing the equilibrium outcomes in the presence of settlements with those that arise in the absence of settlements, we show that settlements can act as a complement or substitute to the leniency program, and that settlements exert important deterrence effects.

We also study the optimal design of policy. Whenever cartel deterrence is not feasible due to the limited budgetary resources of the CA, the CA should make the leniency program as attractive as possible by offering maximal fine reductions to leniency applicants and minimal fine reductions to settling firms. This maximizes the frequency with which colluding firms cooperate with the CA in equilibrium. Only if colluding firms cannot be induced to apply for leniency under these circumstances should maximal fine reductions also be offered to settling firms. This maximizes the chances that settlements improve welfare by encouraging firms that cannot be persuaded to participate in the leniency program under any circumstances to cooperate with the CA by settling instead.

This result highlights an important interdependence between the fine reductions that should be offered to leniency applicants and to settling firms. Current antitrust enforcement practice does not explicitly link the leniency program and the settlement procedure to one another in the way that is suggested by this result, however. Past experience may facilitate this linkage, in the sense that more generous fine reductions could be offered to settling firms where take-up of the leniency program has historically been limited.

Reflecting on current practice in light of this result, the EU approach of granting settling firms a relatively modest fine reduction of 10% appears consistent with the concern that additional reductions may result in settlements' undermining the effectiveness of the leniency program.

There are several directions for future research: In line with the discussion in Sect. 6.3, it would be interesting to combine our analysis of leniency and settlements with existing work with respect to the optimal structure of cartel fines. Moreover, it could be interesting to allow the CA to reveal information about the strength of its preliminary case strategically. This might allow the CA greater control over the types of cartels that settle their cases and hence control over the circumstances under which settlements act as a complement or substitute to the leniency program.

Appendix A Dominated Strategies

We demonstrate that collusive strategies NNS ('no leniency application, no settlement of a strong preliminary case, settle a weak preliminary case') and NSS ('no leniency application, settle a strong preliminary case, settle a weak preliminary case') can be omitted from consideration.

A.1 Collusive Strategy NNS

Consider first collusive strategy NNS. Using similar arguments to those in Sect. 4.3, the value of this strategy is equal to

$$V_{NNS} = \alpha \left\{ \begin{array}{c} \frac{1}{2}\left[\pi_C + \delta\left(p_1(\pi_N - F) + (1-p_1)\pi_C + \delta V_{NNS}\right)\right] \\ +\frac{1}{2}\left[\pi_N - \gamma_S F + \delta V_{NNS}\right] \end{array} \right\} + (1-\alpha)\left\{ \frac{1}{2}\left[\pi_C + \delta\left(\pi_C + \delta V_{NNS}\right)\right] + \frac{1}{2}\left[\pi_N + \delta V_{NNS}\right] \right\}. \tag{A1}$$

Rewriting (A1) yields

$$V_{NNS} = \frac{\pi_C}{1-\delta} - \frac{\alpha}{2+\delta} \frac{(\pi_C - \pi_N)(1+\delta p_1) + F(\gamma_S + \delta p_1)}{1-\delta}. \tag{A2}$$

Comparing this to V_{NSN}, we have

$$V_{NSN} - V_{NNS} = \left[\frac{\alpha\,\delta(p_1 - p_0)}{2+\delta}\right]\left(\frac{\pi_C - \pi_N + F}{1-\delta}\right) > 0,$$

Thus, if both strategies are incentive-compatible, firms strictly prefer NSN over NNS. We will show below that, whenever NNS is incentive-compatible, so is NSN, so that NNS may indeed be omitted from consideration.

A deviation from NNS at the leniency stage yields value

$$V_{\check{L}NS} = \frac{\pi_N}{1-\delta} - \gamma_1 F.$$

These are identical to the profits of deviating into a leniency application from collusive strategy NSN (see Online Appendix). Continuing with NNS by not applying for leniency yields

$$\begin{aligned} V_{\dot{N}NS} =& \frac{1}{2}\left\{\pi_C + \delta\left[p_1(\pi_N - F) + (1-p_1)\pi_C + \delta V_{NNS}\right]\right\} \\ &+ \frac{1}{2}\left[\pi_N - \gamma_S F + \delta V_{NNS}\right]. \end{aligned}$$

Comparing the continuation profits under NSN and NNS, we have:

$$V_{\dot{N}SN} - V_{\dot{N}NS} = \frac{1}{2}\Big[(V_{NSN} - V_{NNS})(1+\delta) + (p_1 - p_0)\big(\pi_C - \pi_N + F\big)\Big] > 0.$$

Thus the continuation profits are higher under NSN than under NNS; and, hence, if NNS is robust to deviations at the leniency stage, NSN must similarly be robust.

Now consider the deviation from NNS whereby a firm unilaterally settles a strong case. Settling when the CA has a strong case yields value

$$V_{N\check{S}S} = \frac{\pi_N}{1-\delta} - \gamma_S F,$$

which is the same as the value achieved by a firm that deviates from NSN by unilaterally settling a weak case: $V_{NS\check{S}}$ (see Online Appendix).

The value of continuing with NNS by not settling the strong case, once established, is

$$V_{N\dot{N}S} = \pi_C + \delta\left[p_1(\pi_N - F) + (1-p_1)\pi_C + \delta V_{NNS}\right].$$

Since $V_{NNS} < V_{NSN}$ and $p_1 > p_0$, we have $V_{N\dot{N}S} < V_{NS\dot{N}}$. So if NNS is robust to deviations as part of which a firm unilaterally settles a strong case, then NSN must also be robust to deviations as part of which a firm unilaterally settles a weak case.

Finally, the firm could deviate from NNS by deciding unilaterally not to settle a weak case. The value of this deviation is

$$V_{NN\check{N}} = \frac{\pi_N}{1-\delta} - F. \tag{A3}$$

If the firm sticks to settling the weak case, the profits are

$$V_{NN\dot{S}} = \pi_N - \gamma_S F + \delta V_{NNS}. \tag{A4}$$

Sticking to NNS is more profitable whenever $V_{NN\dot{S}} > V_{NN\check{N}}$, which is equivalent to

$$(1-\gamma_S)F + \delta\left(V_{NNS} - \frac{\pi_N}{1-\delta}\right) > 0.$$

The first term is non-negative because $\gamma_S \leq 1$. Hence, this condition is always satisfied if the second term is positive. This must be true whenever NNS is robust to deviations at the cartel formation stage, since then $V_{NNS} > V_D > \frac{\pi_N}{1-\delta}$; see (5). Hence, this latter deviation does not impose any additional restrictions on the incentive-compatibility of NNS.

We therefore conclude that if collusive strategy NNS is incentive-compatible, so is collusive strategy NSN. Whenever both strategies are incentive-compatible, NSN yields a higher value to firms, so that NNS is Pareto-dominated.

A.2 Collusive Strategy NSS

Consider the collusive strategy NSS as part of which firms do not apply for leniency, but agree to settle any case if investigated. The value of this strategy is

$$V_{NSS} = \alpha(\pi_N - \gamma_S F) + (1 - \alpha)\pi_C + \delta V_{NSS},$$

which can be rewritten as

$$V_{NSS} = \frac{\pi_C}{1 - \delta} - \frac{\alpha\left(\pi_C - \pi_N + \gamma_S F\right)}{1 - \delta}.$$

Since $\gamma_L < \gamma_S$, this value cannot exceed the profits from a leniency application; see (7) and the Online Appendix. Hence the constraint that is implied by deviations at the collusion stage is tighter for NSS than LXX, and this is the only relevant incentive-compatibility constraint for LXX.

We therefore conclude that if collusive strategy NSS is incentive-compatible, so is collusive strategy LXX. Whenever both strategies are incentive-compatible, LXX always yields a higher value to firms, so that NSS is Pareto-dominated.

Acknowledgements We thank the Editor (Lawrence J. White), Special-Issue Editor (Adriaan Soetevent), Associate Editor and two anonymous referees for helpful comments that have substantially improved the paper. We are also indebted to Pablo Amador Sanchez and Jan Tichem for constructive feedback. The views expressed in this paper do not necessarily reflect the official opinion of the Netherlands Authority for Consumers & Markets: The authors are fully responsible for the contents of the paper.

Author Contributions This work was produced jointly and equally by both named authors.

Data Availability We do not analyze or generate any datasets because our work proceeds within a theoretical and mathematical approach.

Declarations

Conflict of interest The authors declare no competing interests

References

Andres, M., Bruttel, L., & Friedrichsen, J. (2021). The leniency rule revisited: Experiments on cartel formation with open communication. *International Journal of Industrial Organization, 76*, 102728. https://doi.org/10.1016/j.ijindorg.2021.102728

Barennes, M. (2017). The role of the settlement procedure and leniency programme in the European Commission's fight against cartels. V. Tomljenović, N. Bodiroga-Vukobrat, V. Butorac Malnar & I. Kunda (Eds.), EU competition and state aid rules: Public and private enforcement (pp. 113–131). Berlin, Heidelberg:Springer. https://doi.org/10.1007/978-3-662-47962-9_6.

Bigoni, M., Fridolfsson, S. O., Le Coq, C., & Spagnolo, G. (2012). Fines, leniency and rewards in antitrust. *RAND Journal of Economics, 43*(2), 368–390. https://doi.org/10.1111/j.1756-2171.2012.00170.x

Bigoni, M., Fridolfsson, S. O., Le Coq, C., & Spagnolo, G. (2015). Trust, leniency, and deterrence. *The Journal of Law, Economics, and Organization, 31*(4), 663–689. https://doi.org/10.1093/jleo/ewv006

Blatter, M., Emons, W., & Sticher, S. (2018). Optimal leniency programs when firms have cumulative and asymmetric evidence. *Review of Industrial Organization, 52*(3), 403–427. https://doi.org/10.1007/s11151-017-9586-8

Brenner, S. (2009). An empirical study of the European corporate leniency program. *International Journal of Industrial Organization, 27*(6), 639–645. https://doi.org/10.1016/j.ijindorg.2009.02.007

Buccirossi, P., Marvão, C., & Spagnolo, G. (2020). Leniency and damages: Where is the conflict? *The Journal of Legal Studies, 49*(2), 335–379. https://doi.org/10.1086/711392

Buiten, M. C., van Wijck, P., & Winters, J. K. (2018). Does the European damages directive make consumers better off? *Journal of Competition Law & Economics, 14*(1), 91–114. https://doi.org/10.1093/joclec/nhx029

Chen, J. & Harrington, J.E., Jr. (2007). The impact of the corporate leniency program on cartel formation and the cartel price path. V. Ghosal (Eds.) J. Stennek (Eds.), The political economy of antitrust **282**, (pp. 59–80). Amsterdam, the Netherlands:Elsevier. https://doi.org/10.1016/S0573-8555(06)82003-1.

Chen, Z., & Rey, P. (2013). On the design of leniency programs. *The Journal of Law & Economics, 56*(4), 917–957. https://doi.org/10.1086/674011

Dekeyser, K., & Roques, C. (2010). The European Commission's settlement procedure in cartel cases. *The Antitrust Bulletin, 55*(4), 819–842. https://doi.org/10.1177/0003603X1005500406

Dijkstra, P. T., Haan, M. A., & Schoonbeek, L. (2021). Leniency programs and the design of antitrust: Experimental evidence with free-form communication. *Review of Industrial Organization, 59*(1), 13–36. https://doi.org/10.1007/s11151-020-09789-5

Ellis, C.J. & Wilson, W.W. (2003). Cartels, price-fixing, and corporate leniency policy: What doesn't kill us makes us stronger (Tech. Rep.). Eugene:University of Oregon.

Emons, W. & Lenhard, S. (2020). Rebating antitrust fines to encourage private damages actions (Discussion Papers No. 20-02). Bern:University of Bern, Department of Economics. https://doi.org/10.2139/ssrn.3527422.

European Commission. (1996). Commission notice on the non-imposition or reduction of fines in cartel cases. *Official Journal of the European Communities,* C *207*, 4–6.

European Commission. (2002). Commission notice on immunity from fines and reduction of fines in cartel cases. *Official Journal of the European Communities,* C *45*, 3–5.

European Commission. (2006). Commission notice on immunity from fines and reduction of fines in cartel cases. *Official Journal of the European Communities,* C *298*, 17–22.

European Commission (2008). Antitrust: Commission introduces settlement procedure for cartels. Retrieved June 30, 2021, from https://ec.europa.eu/commission/presscorner/detail/en/IP_08_1056.

Fotis, P. & Tselekounis, M. (2020). Optimal reduction of cartel fines induced by the settlement procedure (MPRA Paper No. No. 99154). Munich Personal RePEc Archive. https://mpra.ub.uni-muenchen.de/99154/.

Friedman, J. W. (1971). A non-cooperative equilibrium for supergames. *Review of Economic Studies, 38*(1), 1–12. https://doi.org/10.2307/2296617
Ghebrihiwet, N. & Motchenkova, E. (2010). Leniency programs in the presence of judicial errors (TILEC Discussion Paper No. No. 2010-030). TILEC. https://doi.org/10.2139/ssrn.1666111.
Harrington, J. E., Jr. (2008). Optimal corporate leniency programs. *Journal of Industrial Economics, 56*(2), 215–246. https://doi.org/10.1111/j.1467-6451.2008.00339.x
Hellwig, M., Hüschelrath, K., & Laitenberger, U. (2018). Settlements and appeals in the European Commission's cartel cases: An empirical assessment. *Review of Industrial Organization, 52*(1), 55–84. https://doi.org/10.1007/s11151-017-9572-1
Hinloopen, J., & Soetevent, A. R. (2008). Laboratory evidence on the effectiveness of corporate leniency programs. *RAND Journal of Economics, 39*(2), 607–616. https://doi.org/10.1111/j.0741-6261.2008.00030.x
Houba, H., Motchenkova, E., & Wen, Q. (2015). The effects of leniency on cartel pricing. *The B.E. Journal of Theoretical Economics, 15*(2), 351–389. https://doi.org/10.1515/bejte-2013-0139
Hüschelrath, K., & Laitenberger, U. (2017). The settlement procedure in the European Commission's cartel cases: an early evaluation. *Journal of Antitrust Enforcement, 5*(3), 458–487. https://doi.org/10.1093/jaenfo/jnw015
Jochem, A., Parrotta, P., & Valletta, G. (2020). The impact of the 2002 reform of the EU leniency program on cartel outcomes. *International Journal of Industrial Organization, 71*, 102640. https://doi.org/10.1016/j.ijindorg.2020.102640
Katsoulacos, Y., Motchenkova, E., & Ulph, D. (2015). Penalizing cartels: The case for basing penalties on price overcharge. *International Journal of Industrial Organization, 42*, 70–80. https://doi.org/10.1016/j.ijindorg.2015.07.007
Katsoulacos, Y., Motchenkova, E., & Ulph, D. (2019). Penalizing cartels-a spectrum of regimes. *Journal of Antitrust Enforcement, 7*(3), 339–351. https://doi.org/10.1093/jaenfo/jnz013
Katsoulacos, Y., Motchenkova, E., & Ulph, D. (2019). Public and private antitrust enforcement for cartels: Should there be a common approach to sanctioning based on the overcharge rate? *Revista de Economia Contemporânea, 23*(02), 1–19. https://doi.org/10.1590/198055272321
Katsoulacos, Y., Motchenkova, E., & Ulph, D. (2020). Combining cartel penalties and private damage actions: The impact on cartel prices. *International Journal of Industrial Organization, 73*, 102604. https://doi.org/10.1016/j.ijindorg.2020.102604
Katsoulacos, Y., Motchenkova, E., & Ulph, D. (2020). Penalising on the basis of the severity of the offence: A sophisticated revenue-based cartel penalty. *Review of Industrial Organization, 57*(3), 627–646. https://doi.org/10.1007/s11151-019-09738-x
Laina, F., & Bogdanov, A. (2019). The EU Cartel settlement procedure: Latest developments (2017–2018). *Journal of European Competition Law & Practice, 10*(5), 322–333. https://doi.org/10.1093/jeclap/lpz015
Marvão, C. (2016). The EU leniency programme and recidivism. *Review of Industrial Organization, 48*(1), 1–27. https://doi.org/10.1007/s11151-015-9474-z.
Marvão, C. (2022). Cartel activity and recidivism. In: P. Whelan (Ed), Research handbook on cartels (chap. 19). Edward Elgar Publishing. https://doi.org/10.4337/9781839102875.00029 and available at SSRN: https://ssrn.com/abstract=3875885.
Marvão, C. & Spagnolo, G. (2018). Cartels and leniency: Taking stock of what we learnt. In: Handbook of game theory and industrial organization, volume ii (pp. 57–90). Cheltenham, UK: Edward Elgar Publishing. https://doi.org/10.4337/9781788112789.00010 and available at SSRN: https://ssrn.com/abstract=2850498.
Marx, L. M., & Mezzetti, C. (2013). Effects of antitrust leniency on concealment effort by colluding firms. *Journal of Antitrust Enforcement, 2*(2), 305–332. https://doi.org/10.1093/jaenfo/jnu008.
Miller, N. H. (2009). Strategic leniency and cartel enforcement. *American Economic Review, 99*(3), 750–768. https://doi.org/10.1257/aer.99.3.750.
Motchenkova, E. & Spagnolo, G. (2019). Leniency programs in antitrust: Practice vs theory. CPI Antitrust Chronicle**1**(2), 62–70, https://www.competitionpolicyinternational.com/wp-content/uploads/2019/01/CPI-Motchenkova-Spagnolo.pdf.
Motta, M., & Polo, M. (2003). Leniency programs and cartel prosecution. *International Journal of Industrial Organization, 21*(3), 347–379. https://doi.org/10.1016/S0167-7187(02)00057-7.
O'Brien, A. (2008). Cartel settlements in the U.S. and EU: Similarities, differences & remaining questions. Presented at the 13th Annual EU Competition Law and Policy Workshop, 6 June, Florence,

Italy. https://www.justice.gov/atr/speech/cartel-settlements-us-and-eu-similarities-differences-remaining-questions.
Spagnolo, G. (2004). Divide et impera: Optimal leniency programmes (CEPR Discussion Paper no. No. 4840). London:Centre for Economic Policy Research. http://www.cepr.org/pubs/dps/DP4840.asp.
US Department of Justice (1993). Corporate leniency policy. http://www.justice.gov/atr/public/guidelines/0091.pdf.
Werden, G.J. , Hammond, S.D. & Barnett, B.A. (2011). Recidivism eliminated: Cartel enforcement in the United States since 1999 (Tech. Rep.). prepared for the Georgetown Global Antitrust Enforcement Symposium, Washington DC, 22 September 2011. Available at SSRN: https://ssrn.com/abstract=1927864 or https://doi.org/10.2139/ssrn.1927864.
Whish, R., & Bailey, D. (2021). *Competition law* (10th ed.). Oxford: Oxford University Press.