Universität Trier

Optimal Destabilization of Cartels

Ludwig von Auer
Tu Anh Pham

Research Papers in Economics
No. 7/19

Optimal Destabilization of Cartels

Ludwig von Auer (Universität Trier)

Tu Anh Pham (Universität Trier)[1]

February 2, 2020

Abstract: A model-based derivation of an effective antitrust policy requires an economic framework that includes three actors: a cartel, a group of competing fringe firms, and a welfare maximizing antitrust authority. In existing models of cartel behavior, at least one of these actors is always missing. By contrast, the present paper's oligopoly model includes all three actors. The cartel is the Stackelberg quantity leader and the fringe firms are in Cournot competition with respect to the residual demand. Taking into account that the antitrust policy instruments (effort, fine, and leniency program) are not costless for society, an optimal policy is derived.

JEL-Classification: L13, L41

Keywords: antitrust, stability, Cournot fringe, oligopoly, leniency.

[1]This is a fundamentally revised version of a paper that we presented at IIOC 2017 in Boston and EARIE 2017 in Maastricht. The new manuscript benefited from suggestions by Kebin Ma, Maarten Pieter Schinkel, Armin Schmutzler, Andranik Stepanyan, and Vasily Korovkin. Any remaining errors and omissions are of course ours.

1 Introduction

The "prestressing steel cartel" operated on the European market between 1984 and 2002. It agreed to set quotas on the quantities to be supplied to shared clients. The cartel comprised eighteen members competing against six fringe firms. The collusion was complicated by new competitors and by a drop in demand in 1996. In 2002, the cartel was detected. In 2011, it was finally punished by a penalty of almost € 270 million.[2] The "methionine cartel" operated between 1986 and 1999. It agreed to limit its sales outside the USA and Japan. The cartel had four members competing against two fringe firms. When in 1991 the fringe competitor Novus introduced a successful rival product, the cooperation within the cartel became more difficult. The cartel was detected in 2001 and a year later penelized with a fine of € 127 million.[3]

These two examples suggest that the collusion of cartels is confronted by at least three external threats: the investigations of antitrust authorities, competition from fringe firms, and changes in the market environment. Therefore, the model-based derivation of an effective antitrust policy requires an economic framework that includes three actors: a cartel, a group of competing fringe firms, and an antitrust authority that incorporates into its policy the specific characteristics of the relevant market. In existing models of cartel behavior, at least one of these actors is missing. To address this oversight, the present paper develops a comprehensive oligopoly model that includes all three actors.

Allowing for fringe firms complicates the theoretical analysis, as it raises the issue of cartel stability (d'Aspremont *et al.*, 1983). A cartel is stable if no cartel member has an incentive to become a fringe firm and, at the same time, no fringe firm wants to become a cartel member. Questions of stability usually have been studied in the context of so-called leadership models.[4] However, existing leadership models are not concerned with antitrust policy and, accordingly, do not include an antitrust authority.

Therefore, we introduce a leadership model with a welfare maximizing antitrust authority that tries to deter firms from becoming cartel members. The authority can decide on its own investigative effort, on the appropriate size of the fine that detected cartels must pay, and on the discount offered to testifying firms (leniency program). Of these three instruments, the authority's own investigative effort is its most direct option to increase the probability of detection. Investigative effort is also necessary to turn fines and leniency programs into effective antitrust instruments. Nevertheless, the authority's effort has rarely been addressed in the literature.[5]

In our leadership model, a more aggressive antitrust policy reduces the size of the cartel and increases the number of fringe firms. However, increasing the policy's aggressiveness is not costless for society. This cost must be considered in the design of an optimal antitrust policy. We employ a three-stage game to derive such a policy. In the first stage,

[2] EC (2010, paras. 6, 93, 122, 142, 424, 533), EC (2011, p. 1).

[3] EC (2003, paras. 1, 36-40, 81-89, 279, 356).

[4] An alternative to the leadership approach is the so-called supergame approach. Only a few studies of that strand of literature are concerned with stability. They are discussed in Section 2.

[5] Some exceptions are discussed in Section 2. They all relate to the supergame approach.

the antitrust authority decides on its optimal policy, taking into account the reactions of all firms. In the second stage, each of the firms decides on its status (cartel member or fringe firm). In the third stage, the cartel and each fringe firm determine their optimal output quantities, given the implemented antitrust policy.

The markets in which cartels operate are not uniform and they change over time. For example, stronger demand may increase the market volume and/or new producers may enter the market. Should the antitrust authority react to these changes by an expansion or reduction of its three policy instruments (effort, fine, discount)? Should all three instruments change in the same direction or does it make sense to alter two instruments in one direction and the third in the opposite direction? Existing models of cartel behavior are not designed to address such important practical questions. With the comparative statics of our model we can tackle such issues. For example, we show that minor expansions of the market volume allow for a reduction of all three policy instruments, while the optimal response to large expansions of the market volume is a more aggressive policy that induces one or more cartel members to become fringe firms.

This paper proceeds as follows. Section 2 surveys the related literature. In Section 3 we introduce our model and discuss its assumptions. Sections 4 through 6 are devoted to the derivation of the optimal antitrust policy. Section 7 discusses the underlying economics and the resulting policy implications. Concluding remarks are offered in Section 8.

2 Related Literature

The current literature on cartel behavior is rarely concerned with a cartel's stability. Instead, the focus is on a cartel's sustainability. A cartel is sustainable, if all members adhere to the collusive agreement. Following Friedman's (1971) seminal paper, this literature usually relies on repeated oligopoly games (supergames) with grim-trigger strategies that form a symmetric subgame perfect Nash equilibrium. For example, Motta and Polo (2003, p. 353) assume that the authorities have a given budget that they must efficiently allocate for the monitoring and prosecution of cartels. Spagnolo's (2005, p. 13) supergame features an antitrust authority that can raise the conviction probability by increasing its own effort, the size of the fine, and the extent of leniency. As in our own model, none of these measures is costless.

Escrihuela-Villar (2008, p. 326; 2009, p. 138) as well as Bos and Harrington (2010, pp. 92-93; 2015, p. 133) criticize this part of the supergame literature, because it focuses on all-inclusive cartels, while in reality cartels usually compete against some fringe firms.[6] Models that neglect the role of fringe firms cannot analyze the link between antitrust policy and cartel stability.[7] Therefore, Bos and Harrington (2010, pp. 92-93) propose

[6] This point has also been made by Bos (2009, pp. 11-12). Empirical studies such as Harrington (2006) and Levenstein and Suslow (2006) confirm this position. Hellwig and Hüschelrath (2017) provide a dataset on 114 illegal cartels convicted by the European Commission between 1999 and 2016. The data reveal frequent entries into the cartel and exits from the cartel. Both instances confirm the existence of a fringe.

[7] This is also true for studies by Souam (2001) and Mouraviev and Rey (2011). The latter consider

a supergame with heterogeneous capacity-constrained firms, some of which may stay outside the cartel. They show (p. 101) that a sustainable and stable cartel is made up of the largest firms. The smallest firms prefer the status of a fringe competitor. The fringe firms produce at capacity, whereas the cartel members restrict their output below capacity. In supergames with quantity setting firms, Escrihuela-Villar (2008, p. 327-331; 2009, pp. 139-140) demonstrates that even homogeneous firms can establish a sustainable and stable cartel.

Levenstein and Suslow (2006, p. 78) argue that in the real world the breakup of cartels is pre-dominantly the result of changing economic conditions and not so much the cartel's response to the misbehavior of a cartel member. Misbehavior often results in limited retaliation rather than in the dissolution of the cartel (see also Genesove and Mullin, 2001, pp. 390-394). The grim-trigger strategies underlying most supergames are not fully consistent with this empirical observation.[8] Eaton and Eswaran (1998) and Escrihuela-Villar and Guillén (2011) propose supergames that do not rely on grim-trigger strategies. Instead, the non-cheating members of a cartel continue to operate the cartel without the cheating member.

All previously listed supergames either neglect the issue of an antitrust policy or preclude fringe firms. In a later study, however, Bos and Harrington (2015, p. 135) amend their former supergame framework by an *exogenously given* antitrust policy and investigate the impact of that policy on the properties of the cartel and the fringe. Their analysis confirms that antitrust policies affect the stability of cartels. At the same time, the authors concede that even with the exogenously given antitrust policy "the relationship between antitrust enforcement and cartel size is too complex for us to provide specific guidance for enforcement policies (p. 148)".

To derive appropriate antitrust enforcement policies, the present paper explores a completely different route. It revives the leadership approach and augments it with an *endogenously* derived antitrust policy. The leadership approach was once the backbone of stability analysis. In the price leadership model developed by d'Aspremont *et al.* (1983) the cartel is the Stackelberg price leader. The fringe firms take the leader's price as given and set their quantities such that price equals marginal cost. The price leadership model with its perfectly competitive fringe might fit industries with a large number of competing firms. In the two cartel cases described in the introduction, the number of fringe firms was six and two, respectively. Furthermore, the cartel agreement focused on admissible quantities. Such a situation is better described by a special variant of a framework that Daughety (1990) introduced to analyze the welfare effects of mergers. This special variant is the quantity leadership model advocated by Shaffer (1995). The cartel is the Stackelberg quantity leader and the fringe firms are in Cournot competition with

a cartel the members of which play sequentially instead of simultaneously. The authors show how this can facilitate collusion. Souam (2001) proposes a framework in which the antitrust authority takes the market price as a signal for the probability that a cartel exists. Since antitrust enforcement is costly, the antitrust authority should increase its effort with the observed market price. If the market price is sufficiently low, however, collusion should be tolerated. Sustainability is not an issue in his framework.

[8] See also Green and Porter (1984).

respect to the residual demand. The present paper builds on this oligopoly framework.

Leadership models have inspired additional work on the conditions required for the successful formation and stability of cartels. For example, Donsimoni (1985), Donsimoni *et al.* (1986), and Prokop (1999) utilize the price leadership model, whereas studies by Konishi and Lin (1999) and Zu *et al.* (2012) are based on the quantity leadership model.

Antitrust policy is not an issue in either type of leadership model. Instead, these studies focus on the formal conditions for the existence and uniqueness of a stable cartel. The present study combines Shaffer's (1995) quantity leadership model with an active antitrust authority that wants to maximize social welfare.

The supergame approach and the leadership approach, including our own model, have a common weakness. Studies such as Levenstein and Suslow (2006) and Harrington (2006) show that colluding firms usually develop organizational structures that ensure some degree of enforceability of the cartel agreement. By contrast, the supergame approach is built on the assumption that the cartel has no means of enforcement. In the leadership approach this extreme assumption is replaced by the opposite extreme: perfect enforceability of the cartel agreement.

3 Model

3.1 Three Stage Antitrust Game

Our model is a three stage game with a finite integer number of $n \geq 2$ identical firms and an antitrust authority. First, the antitrust authority chooses its policy, taking into account the reactions of the n firms. Then, given the implemented antitrust policy, each of the n firms decides whether it wants to become a fringe firm or a member of the cartel. The resulting number of fringe firms is denoted by n_F. The remaining $(n - n_F)$ firms form the cartel. In their choice between fringe and cartel the firms take into account the resulting equilibrium output quantities and the associated profits. Both are determined in the third stage of our antitrust game.

In that final stage, all n firms produce the same homogeneous good and have the same constant marginal cost equal to c. The inverse demand function is $P = a - bQ$, where P is the market price, Q is the aggregate quantity produced, and a and b are positive constants. The $(n - n_F)$ members of the cartel act as one company and collectively determine their profit maximizing joint output Q_C. Afterwards, each fringe firm determines its profit maximizing output q_F. In other words, the cartel acts as a Stackelberg leader, while the group of fringe firms is the Stackelberg follower.[9] If the cartel shrinks to one firm

[9] In equilibrium, the cartel always produces more than half of the complete output. Therefore, assigning the role of the Stackelberg leader to the cartel is a reasonable feature of the quantity leadership model. As an additional justification, Shaffer (1995, pp. 348-349) points out that the cartel benefits from the Stackelberg sequence and, therefore, may want to impose its will on the fringe firms. Huck *et al.* (2007) provide some experimental evidence that firms that cooperate in a binding manner show leadership behavior, whereas the remaining firms exhibit follower behavior. For additional references that support the leadership role of perfectly colluding firms see Brito and Catalão-Lopes (2011, pp. 3-4).

$(n - n_F = 1)$, this firm will no longer represent an illegal cartel, but will become a legal Stackelberg leader. Since that firm would never want to give up that position, we know that $n_F \leq n - 1$. Each fringe firm considers both the cartel's output, Q_C, and the aggregate output of the other fringe firms, Q_{-F}, as given. Therefore, the output of each fringe firm, q_F, is determined by the Cournot-Nash equilibrium concept.

If a cartel exists $(n - n_F \geq 2)$, it is detected with some probability $p \in [0, 1]$. The probability depends on the antitrust authority's policy. The policy is implemented before the n firms decide on their cartel membership and their output quantities.

Three policy instruments are available to the authority. The first instrument is the fine $f \geq 0$ that the members of a detected cartel must pay. The second instrument is the expected discount offered to some or all cartel firms that inform the authority about the cartel. The expected discount is defined by $d = r\mu \in [0, 1)$, with $r \geq 0$ denoting the percentage by which the fine of an eligible and cooperating cartel member is reduced, and $\mu \in [0, 1]$ denoting the share of cartel members eligible for that reduction.[10] Accordingly, $p\, f\, (1 - d)$ is the expected fine of each member of the cartel and $f\, (1 - d)$ is the average fine of the members of a *detected* cartel. The third policy instrument is the authority's own investigative effort, $e \geq 0$.

We define the probability of detection by the multiplicative function

$$p = h(e, f, d)\, g(n - n_F) \; . \tag{1}$$

The factor $h(e, f, d) \in [0, 1]$ captures the impact of the authority's antitrust policy. We assume that $h(e, f, d)$ is a continuous concave function that approaches 1 from below and has positive first order and negative second order derivatives. Furthermore, $h(0, f, d) = h(e, 0, d) = 0$. When $e > 0$ and $f > 0$, then $0 < h(e, f, 0) < 1$. The second factor, $g(n - n_F) \in [0, 1]$, takes care of the fact that larger cartels are more likely to be detected than smaller ones. We assume that $g(1) = 0$ and that $g(n - n_F)$ is concave and approaching 1 from below. A more elaborated justification of these assumptions is provided in Section 3.2.

Our model recognizes that the implementation of an antitrust policy is not costless. Following the law enforcement literature initiated by Becker (1968), we capture this cost by a continuous social cost function, $s(e, f, d)$, with positive first order partial derivatives.

The objective of the antitrust authority is to implement a policy (e, f, d) that maximizes welfare. This policy is denoted as the *optimal antitrust policy*. Welfare does not depend on the budgetary effects of the fines and discounts, because these are of a purely redistributional nature. Therefore, welfare is defined here as the sum of consumer and producer rent minus the social cost, $s(e, f, d)$, caused by the antitrust policy.[11]

[10] Suppose that the cartel is detected. If $\mu = 0.1$ and the number of cartel members is $n - n_F = 10$, then exactly one member is randomly drawn. This member is regarded as a cooperating firm and receives the reduction r. If $\mu = 0.1$ and $n - n_F = 5$, again one member is randomly drawn and that member has a 50% chance of being regarded as a cooperating firm.

[11] Wilson (2019) and Albæk (2013) present some explanations why antitrust authorities such as the U.S. Federal Trade Commission and the European Commission focus on consumer rent and tend to neglect producer rent. A compact discussion of these issues can be found in Motta (2004, pp. 19-22).

3.2 Discussion of Some Assumptions

In practice, fines are often linked to turnover or to profits. However, for algebraic simplicity, we assume that the fine, f, is lump-sum.

There is ample evidence that leniency programs increase the probability of detecting cartels (e.g., Aubert *et al.*, 2006, p. 1242; Brenner, 2009, pp. 642-644). In anticipation of being detected, cartel firms may apply for leniency by providing evidence of a cartel agreement. Furthermore, even if cartel members consider it unlikely that the antitrust authority will discover anything, they may worry that some fellow member will apply for leniency and, because of that worry, apply themselves. Harrington (2013, pp. 2-3) denotes these two effects as "prosecution effect" and "preemption" effect, respectively. The policy variables f and $d = r\mu$ in $h(e, f, d)$ capture these effects. An expansion of eligibility, μ, or an increase in the percentage r by which the fine of an eligible cartel member is reduced, strengthens the preemptive effect of discounts. However, it lowers the average fine of the members of a detected cartel, weakening the prosecution effect. We assume that the former effect dominates the latter effect, that is, $\partial h/\partial d > 0$. Our specification allows for a percentage $r > 1$. For plausibility reasons, however, we restrict the domain of $d = r\mu$ to the interval $[0, 1)$. Otherwise, the members of a *detected* cartel, on average, receive a reward instead of a fine: $f\,(1 - d) \leq 0$. This cannot be a sensible antitrust policy.

A positive effort, e, is necessary to turn the fine and the leniency program into effective instruments. Without any effort on the side of the antitrust authority the prosecution effect and the preemption effect do not exist, regardless of the size of the cartel and the size of the fine. Therefore, the case $e = 0$ must give $p = 0$, which requires that $h(0, f, d) = 0$. When a detected cartel never pays a fine ($f = 0$), the investigative staff is likely to be demoralized and its effort may become completely ineffective, that is, $h(e, 0, d) = 0$. If the fine for detected cartels is positive, the antitrust authority must be able to detect an existing cartel through its own investigative effort e. Therefore, we assume that $h(e, f, 0) > 0$, when $e > 0$ and $f > 0$. The preceding discussion demonstrates that the probability of detection defined by Equation (1) captures several important interdependencies between the policy instruments of the antitrust authority.[12]

The assumption $g(1) = 0$ ensures that a "cartel" with only one member cannot be detected, because this member does not form an illegal cartel, but merely represents a legal Stackelberg leader.

The function $s(e, f, d)$ represents the social cost arising from the three antitrust policy instruments. Obviously, if society desires a larger effort, e, it must provide the resources necessary to hire more and better staff and to purchase a more effective system. Less obvious is the social cost arising from the fine f. The antitrust authority has a strong incentive to choose very large fines, because this reduces the expected profits from cartel membership. Excessive fines, however, induce a social cost, because they violate the

[12]Nevertheless, some interdependencies may exist that are not fully represented by this specification. For example, the separability between the factors $g(n - n_F)$ and $h(e, f, d)$ implies that the effect of the cartel size on the probability of detection does not depend on the specific policy mix ensuring a given value of $h(e, f, d)$.

principle of proportional justice and may increase the risk of convicting innocent firms (e.g., Allain *et al.*, 2015). The discount, d, causes similar social costs. The public may dislike the idea that testifying firms that have broken the law can get away with a discount or, even worse, are rewarded. Lenient treatment of guilty firms may undermine a general respect for the law and may encourage unlawful behavior.

We will solve our three stage game by backward induction, starting with the derivation of the profit maximizing quantity reactions, q_F and Q_C, to each given antitrust policy, (e, f, d), and number of fringe firms, n_F (Section 4). Then, to each given antitrust policy, (e, f, d), we derive the equilibrium number of n_F, that is, the profit maximizing status decisions (fringe or cartel) of the n firms (Section 5). To this end, we exploit the previously derived quantity reactions, q_F and Q_C, and the concept of stability. Finally, given the equilibrium reactions of the n firms (status and output quantity), we derive the optimal antitrust policy (Section 6).

4 Third Stage: Determining the Output Quantities

In the presence of a given antitrust policy, (e, f, d), and a given number of fringe firms, $n_F \in (0, \ldots, n-1)$, there is a given expected fine, $p\, f\,(1-d)$. This fine can be interpreted as a fixed cost of the cartel members and, therefore, does not affect their profit maximizing behavior. The resulting equilibrium output of the cartel (Stackelberg leader) is

$$Q_C = \frac{a-c}{2b}\ , \tag{2}$$

while each fringe firm produces

$$q_F = \frac{a-c}{2b(n_F+1)}\ . \tag{3}$$

Therefore, total output is

$$Q = Q_C + n_F q_F = \frac{a-c}{b}\,\frac{2n_F+1}{2n_F+2} \tag{4}$$

and the market price is

$$P = c + \frac{a-c}{2\,(n_F+1)}\ . \tag{5}$$

The profit of each fringe firm is[13]

$$\pi_F\,(n_F) = \frac{(a-c)^2}{4b(n_F+1)^2}\ , \tag{6}$$

while each cartel member receives the expected profit

$$E\left[\pi_C\,(n_F)\right] = \frac{(a-c)^2}{4b(n_F+1)(n-n_F)} - p\, f\ (1-d)\ . \tag{7}$$

[13] The results (2) to (6) can also be found in Shaffer (1995, p. 745).

5 Second Stage: Choosing the Status

At this stage, the n firms decide whether they want to become a fringe firm or a member of the cartel. In their decision, they take the policy (e, f, d) as given and they anticipate the quantity reactions (2) and (3) and the associated profits (6) and (7).

Suppose that a member of the cartel is presented with an offer to become a fringe firm. If and only if

$$E\left[\pi_C\left(n_F\right)\right] > \pi_F(n_F + 1)\ , \tag{8}$$

the firm rejects the offer. If all cartel members reject the offer, the cartel is denoted as *internally* stable.[14] The cartel is *externally* stable, if each fringe firm rejects the offer to become a member of the cartel. This rejection arises, if and only if

$$E\left[\pi_C\left(n_F - 1\right)\right] \leq \pi_F\left(n_F\right)\ . \tag{9}$$

A stable cartel is a cartel that is internally and externally stable.[15]

Our model determines the final status of a firm (cartel or fringe) using the following random process. First, each of the n firms is randomly assigned its status. Then one firm is randomly drawn and given the opportunity to change its status. If the firm is a member of the cartel and if (8) is violated, the firm decides to become a fringe firm. Since all cartel members are identical, the decision is independent of which cartel member is drawn. If, instead, a fringe firm is drawn and condition (9) is violated, this firm decides to enter the cartel. Again, the decision is independent of which fringe firm is drawn. Next, another (or the same) firm is randomly drawn and allowed to change its status. This random process is continued until two consecutive draws occur in which the two firms drawn have different statuses and both decide to keep their status. After these two decisions the random process terminates, because all fringe firms are identical and all cartel members are identical and, therefore, in all additional random draws no firm would want to change its status.

For the characterization of the equilibrium solution it is useful to define the "force", A, of the given policy (e, f, d) by the following expression:

$$A := h(e, f, d) f\left(1 - d\right) \geq 0\ . \tag{10}$$

The value of A depends on the policy instruments e, f, and d, but not on n and n_F. Any antitrust policy with $e = 0$ or $f = 0$ leads to $h(e, f, d) = 0$ and, therefore, to $A = 0$ and $p = 0$. Therefore, such a policy completely eliminates the possibility of detecting an operating cartel. We denote such policies as *passive* antitrust policies. Increases in the effort, e, and the fine, f, raise the value of A. The impact of the discount d on the value of A is ambiguous, since it increases the value of the factor $h(e, f, d)$ and, therefore,

[14] In the original definition given by d'Aspremont *et al.* (1983, p. 21) and many subsequent papers a *weakly* larger profit is sufficient to reject the offer.

[15] Thoron (1998) demonstrates that the internal and external stability concepts merely reproduce a Nash equilibrium of a participation game.

the probability of detection, p, but lowers the average fine of the members of a detected cartel, $f\,(1-d)$. Since d was restricted to values smaller than 1, A cannot be negative.

The profits (6) and (7) imply that an all-inclusive cartel ($n_F = 0$) is internally stable, if and only if

$$A < \frac{(a-c)^2}{4b}\left(\frac{1}{n} - \frac{1}{4}\right) . \tag{11}$$

For $n \geq 4$, the right hand side would be non positive, while A is non negative. Therefore, an all-inclusive cartel with more than three members cannot be stable, even when $A = 0$. At least one member of the cartel would decide to change its status. For $n = 3$ or $n = 2$, however, an all-inclusive cartel is conceivable.

Furthermore, we introduce the following "threshold variable":

$$T(n_F) := \frac{(a-c)^2}{4b} \frac{n_F\,(2n_F + 1 - n) + 1}{n_F(n - n_F + 1)(n_F + 1)^2 g(n - n_F + 1)} . \tag{12}$$

It is independent of the policy (e, f, d). In Lemma 1 (see Appendix) it is shown that $\partial T(n_F)/\partial n_F > 0$.

Theorem 1 *Given some antitrust policy (e, f, d) and the quantity reactions (2) and (3), the random process that determines a firm's status leads to a unique equilibrium n_F-value. Policies that satisfy condition (11) lead to $n_F = 0$. For all other policies, the equilibrium value of n_F is given by*

$$T(n_F) \leq A < T(n_F + 1) . \tag{13}$$

Proof: See Appendix.

Theorem 1 determines the size of the stable cartel for each policy, (e, f, d). For example, a policy with the force $A = T(n_F)$ leads to a stable cartel with $(n - n_F)$ members. Since $\partial T(n_F)/\partial n_F > 0$, an increase in the equilibrium n_F-value (that is, a reduction of the cartel's size) requires an increase in the policy's force, A. Whether such an increase is desirable, is to be examined in the first stage of the antitrust game.

6 First Stage: Determining the Antitrust Policy

The sum of consumer and producer rent is equal to $(a-c)\,Q - 0.5(a-P)Q$, where the values of Q and P are defined by (4) and (5). Subtracting the social cost, $s(e, f, d)$, yields the following welfare function:

$$W(e, f, d) = \frac{(a-c)^2\,(2n_F + 1)\,(2n_F + 3)}{8b\,(n_F + 1)^2} - s(e, f, d) . \tag{14}$$

The antitrust authority chooses its policy, (e, f, d), such that welfare, $W(e, f, d)$, is maximized. This policy is denoted as the authority's optimal antitrust policy, (e^*, f^*, d^*). In the derivation of this policy, the authority anticipates the equilibrium n_F-value (determined by Theorem 1) and the corresponding quantity reactions (2) and (3).

Relationship (13) implies that for passive antitrust policies the condition for stability becomes

$$T(n_F) \leq 0 < T(n_F + 1) \; . \tag{15}$$

Only one n_F-value exists that satisfies this condition. We denote this value by $n_F^{\min}$, because an *active* antitrust policy ($e > 0$ and $f > 0$) would lead to n_F-values that are at least as large as $n_F^{\min}$ and, therefore, to cartels that are never larger than $\left(n - n_F^{\min}\right)$. Thus, we can confine our search for the optimal antitrust policy to those policies (e, f, d) that lead to $n_F \in \left(n_F^{\min}, \ldots, n-1\right)$.

To find the optimal antitrust policy (e^*, f^*, d^*), we pursue a three step procedure. First, we find $n_F^{\min}$. Then, we derive *for each given* $n_F \in \left(n_F^{\min}, \ldots, n-1\right)$ the antitrust policy $(e^*_{n_F}, f^*_{n_F}, d^*_{n_F})$ that minimizes the social cost, $s(e, f, d)$. Finally, we compute the resulting welfare for each of these cost minimizing antitrust policies. The policy that generates the largest welfare is the optimal antitrust policy (e^*, f^*, d^*). In the following, we describe these three steps in more detail.

6.1 Finding $n_F^{\min}$

For $n = 2$ or $n = 3$, we get $n_F^{\min} = 0$. When $n > 3$ and a passive antitrust policy is chosen, the number $n_F^{\min}$ is the smallest n_F-value that satisfies the left hand side inequality of (15). Using (12), this inequality simplifies to $n - n_F \geq n_F + 1 + 1/n_F$. Therefore, $n_F^{\min}$ is the largest integer for which the condition $n - n_F^{\min} \geq n_F^{\min} + 2$ is satisfied. Rearranging this condition gives $n_F^{\min} \leq (n-2)/2$.[16] Therefore,

$$n_F^{\min} = \begin{cases} (n-2)/2 & \text{for} \quad \text{even } n \\ (n-3)/2 & \text{for} \quad \text{uneven } n \; . \end{cases} \tag{16}$$

The antitrust authority can restrict its search for the optimal antitrust policy (e^*, f^*, d^*) to policies that lead to $n_F \geq n_F^{\min}$, where $n_F^{\min}$ is defined by (16).

6.2 Computing the Cost Minimizing Policies

Among all antitrust policies leading to a stable cartel with $(n - n_F^{\min})$ members, the passive policy $(e, f, d) = (0, 0, 0)$ is the cost minimizing policy $(e^*_{n_F^{\min}}, f^*_{n_F^{\min}}, d^*_{n_F^{\min}})$. Suppose that the antitrust authority wants to shrink the cartel from $\left(n - n_F^{\min}\right)$ members to $(n - n_F)$ members, where $n_F \in \left(n_F^{\min} + 1, \ldots, n-1\right)$. This requires an active antitrust policy, that is, a policy with $e > 0$, $f > 0$, and $d \geq 0$.

From condition (13) we know that an active antitrust policy pursuing a cartel with $(n - n_F)$ members must be such that the resulting A-value defined by (10) falls into the interval $[T(n_F), T(n_F + 1))$. An infinite number of active policies exist that satisfy this

[16] This is just a reformulation of Shaffer's (1995, p. 746) Proposition 4.

condition. All of these policies lead to the same given n_F-value and, therefore, to the same quantity Q and price P. Thus, they all yield the same consumer rent and producer rent. However, the social cost varies. Therefore, the authority should choose the policy that causes the lowest social cost, $s(e, f, d)$. Since $\partial s/\partial d > 0$, a welfare maximizing antitrust authority will always decide for a d-value that satisfies the condition $\partial A/\partial d > 0$. This implies that lower A-values allow for lower values of e, f, and d. In other words, lower A-values reduce the social cost, $s(e, f, d)$.

Therefore, for each given n_F-value, the antitrust authority should opt for a policy the force of which, A, reaches the lower bound of its admissible interval defined by (13):

$$A = T(n_F) \ . \tag{17}$$

Choosing a force A slightly below $T(n_F)$ would make a cartel with $(n - n_F)$ members externally instable and its size would increase to $(n - n_F + 1)$. Therefore, Equation (17) defines the smallest possible force, A, that caps the cartel size at $(n - n_F)$. We denote condition (17) as the *efficacy condition*.

An infinite number of policies (e, f, d) satisfy the efficacy condition (17). Among these policies, the authority should choose the one that causes the lowest social cost, $s(e, f, d)$. For given n_F, this cost minimization problem can be written in the following form:

$$\min_{e,f,d} s(e, f, d) \qquad \text{subject to } A = T(n_F) \ . \tag{18}$$

The solution to this cost minimization problem is denoted by $(e^*_{n_F}, f^*_{n_F}, d^*_{n_F})$. We know that this solution is characterized by $e > 0$ and $f > 0$. An interior solution would also require that $d > 0$.

To keep the model analytically tractable, we assume that the two factors of the probability of detection, p, defined by Equation (1) are given by

$$g(n - n_F) = \frac{n - n_F - 1}{n - n_F} \tag{19}$$

and

$$h(e, f, d) = w(e) \cdot k(d) \cdot m(f) \ , \tag{20}$$

where

$$w(e) = \frac{e}{e + 1} \tag{21}$$

$$k(d) = \frac{d + \rho}{d + \rho + 1} \qquad (\rho > 0) \tag{22}$$

$$m(f) = \frac{f}{f + 1} \ . \tag{23}$$

This specification is fully consistent with the postulated properties of $g(n - n_F)$ and $h(e, f, d)$ discussed in Sections 3.1 and 3.2.

Furthermore, we assume that the continuous social cost function is

$$s(e,f,d) = s(z) \qquad \text{with} \qquad z = \alpha e + \beta f m(f) + \gamma d \qquad \text{and} \qquad \partial s/\partial z > 0 \ . \tag{24}$$

The parameters α, β, and γ can be interpreted as the marginal effects of the respective policy instrument on the social cost variable z.[17]

The specifications (19) to (24) imply that a unique solution arises (though not necessarily an interior solution).[18] To characterize the cost minimizing policy $(e^*_{n_F}, f^*_{n_F}, d^*_{n_F})$, we make use of the following definitions:

$$\mathcal{E} := \frac{T(n_F)}{k(d)\,(1-d)} \tag{25}$$

$$\mathcal{F} := \mathcal{E} + \left(\frac{\alpha}{\beta}\mathcal{E}\right)^{1/2} \tag{26}$$

$$\mathcal{D} := -\frac{\partial \mathcal{E}}{\partial d}\left[\left(\frac{\alpha\beta}{\mathcal{E}}\right)^{1/2} + \beta\right]. \tag{27}$$

The three terms $\mathcal{E}$, $\mathcal{F}$, and $\mathcal{D}$ depend on d, but not on e and f.

Theorem 2 *For each* $n_F \in \left(n_F^{min}+1, \ldots, n-1\right)$*, the unique cost minimizing policy that leads to a stable cartel with* $(n-n_F)$ *members, is*

$$e^*_{n_F} = \left(\frac{\beta}{\alpha}\mathcal{E}\right)^{1/2} \tag{28}$$

$$f^*_{n_F} = \frac{1}{2}\left[\mathcal{F} + \left(\mathcal{F}^2 + 4\mathcal{F}\right)^{1/2}\right] \ . \tag{29}$$

If d *is endogenous, an interior solution,* $d^*_{n_F} > 0$*, must satisfy the condition*

$$\mathcal{D} = \gamma \ . \tag{30}$$

Proof: See Appendix.

For a given cartel size, $(n-n_F)$, Equations (28) to (30) of Theorem 2 specify the cost minimizing policy $(e^*_{n_F}, f^*_{n_F}, d^*_{n_F})$, such that the efficacy condition (17), $A = T(n_F)$, is satisfied.

An increase in e or f raises the antitrust policy's force, $A = h(e,f,d)\, f\ (1-d)$. The effort, e, exerts its positive influence only via the "probability factor" $h(e,f,d)$, while the fine, f, exerts its positive influence via both, the probability factor $h(e,f,d)$ and the average fine $f\ (1-d)$.

[17] Again, for reasons of analytical simplicity, we use the function $\beta f m(f)$ instead of the simple linear function βf. Since $\lim_{f\to\infty} m(f) = 1$, the function $\beta f m(f)$ closely approximates the function βf.

[18] Uniqueness merely requires that in e-f-d-space the plane corresponding to the efficacy condition (17) and to a given n_F-value is "more convex" than the isocost-planes of the applied social cost function.

The third policy variable is the expected discount, d. As was true for e and f, an increase in d increases the probability factor $h(e, f, d)$. However, an increase in d also reduces the average fine $f\ (1-d)$ and, therefore, counteracts the increase in $h(e, f, d)$. As a consequence, an increase in d can make sense only if it has a strong positive effect on $h(e, f, d)$. This requires that the original d-value was sufficiently small. Since the right hand side of condition (30) and the expression in square brackets in Equation (27) are positive, the inequality $\partial\mathcal{E}/\partial d^*_{n_F} < 0$ is a necessary condition for a cost efficient positive discount $d^*_{n_F}$.

Lemma 2 in the Appendix shows that $\partial\mathcal{E}/\partial d < 0$, if and only if $(d+\rho)^2 + 2d + \rho < 1$. For $\rho \geq 0.61803$, this inequality is never satisfied. Then the cost minimizing value of d is $d^*_{n_F} = 0$ and Equation (30) of Theorem 2 is redundant. The cost minimizing values $e^*_{n_F}$ and $f^*_{n_F}$ are obtained from Equations (28) and (29). In many countries, the antitrust authorities are not completely free to determine their policy (e, f, d), but are restricted by legal regulations on the fine, f, and/or the expected discount, d. For example, if the expected discount is legally fixed at $d = \bar{d}$, Equation (30) is redundant. Instead, the value $\bar{d}$ is inserted in Equation (25). Inserting the resulting values of $\mathcal{E}$ and $\mathcal{F}$ in Equations (28) and (29) yields the cost minimizing values $e^*_{n_F}$ and $f^*_{n_F}$. This process of finding the cost minimizing policy $(e^*_{n_F}, f^*_{n_F}, \bar{d})$ is executed for each given n_F.

If $\rho < 0.61803$ and, at the same time, γ is not too large, the value $d^*_{n_F}$ satisfying condition (30) is positive. Inserting this cost minimizing value $d^*_{n_F}$ in (25) to (29), yields the cost minimizing values $e^*_{n_F}$ and $f^*_{n_F}$. For each given n_F, the cost minimizing policy $(e^*_{n_F}, f^*_{n_F}, d^*_{n_F})$ is derived in this way.

6.3 Selecting the Optimal Antitrust Policy

If the expected discount is exogenously given, $d = \bar{d}$, we insert $n_F = n_F^{\min}$ and the passive policy $\left(0, 0, \bar{d}\right)$ in the welfare funtion (14) and compute the corresponding welfare level. Then we compile the welfare levels arising from active policies. To this end, we insert $\bar{d}$, (24), (28), and (29) in the welfare function (14) and maximize this expression with respect to n_F. We obtain the optimal fringe size, n_F^*, the corresponding policy $(e^*, f^*, \bar{d})$, and the resulting welfare. The policy $(e^*, f^*, \bar{d})$ is implemented, if the associated welfare is larger than the welfare arising from the passive policy $\left(0, 0, \bar{d}\right)$.

When d is endogenous, we start by inserting $n_F = n_F^{\min}$ and the passive policy $(e, f, d) = (0, 0, 0)$ in welfare function (14) and compute the resulting welfare level. Then we consider the cost minimizing active antitrust policies $(e^*_{n_F}, f^*_{n_F}, d^*_{n_F})$. The welfare levels corresponding to each integer $n_F \in \left(n_F^{\min} + 1, \ldots, n-1\right)$ are calculated. For this purpose we insert each of these n_F-values together with its corresponding cost minimizing policy $(e^*_{n_F}, f^*_{n_F}, d^*_{n_F})$ in the welfare function (14). We get a set of welfare levels. From this set we select the maximum value. If this welfare is larger than the one generated by the passive policy, the corresponding number of fringe firms is the optimal fringe size n_F^*. The cost minimizing antitrust policy leading to the stable cartel with $(n - n_F^*)$ members is the optimal antitrust policy (e^*, f^*, d^*).

7 Further Analysis and Policy Recommendations

To derive important economic implications from Theorem 2, we analyze how changes in the parameter values affect the optimal antitrust policy (e^*, f^*, d^*). We begin the analysis with small parameter changes that do not affect the optimal number of fringe firms, n_F^*. Afterwards, larger parameter changes are considered that alter n_F^*. Only interior solutions $(d^* > 0)$ are discussed.

The force of the original optimal antitrust policy, (e^*, f^*, d^*), is denoted by A^* and the corresponding threshold by $T(n_F^*)$. We analyze changes in the social cost parameters (α, β, γ), the discount parameter (ρ), the market volume parameters (a, b, c), and the number of firms (n). Changes in the parameters α, β, γ, and ρ do not alter the threshold $T(n_F^*)$. They primarily affect the *relative* cost-effectiveness of the three antitrust policy instruments. By contrast, changes in the parameters a, b, c, and n change the threshold $T(n_F^*)$ and, therefore, primarily affect the *overall* cost-effectiveness of antitrust policy.

Even though we analyze *changes* in the parameters, our findings can be interpreted in two different ways. Obviously, they show how the antitrust authority should adjust its policy to changes that occur in some given market. However, they also describe how differences between two markets should be reflected in the corresponding optimal antitrust policies. Before the formal results will be presented (see Theorem 3), we provide a brief intuitive elucidation of these results.

Social cost parameters α, β, *and* γ*:* The larger the parameter α, the more resources the antitrust authority needs to achieve a given level of effort e. The parameters β and γ measure the damage to the rule of law when disproportionate fines are imposed, innocent firms are prosecuted, or discounts are granted to guilty firms. A small change in α, β, or γ does not affect the threshold $T(n_F^*)$. Therefore, the new policy must preserve the original policy's force, A^*.

Suppose that parameter α increases (e.g., the antitrust authority must pay higher wages to attract or retain qualified personnel). We know that an optimal antitrust policy, (e^*, f^*, d^*), ensures that marginal changes to any pair of policy instruments (e.g., e and f) consistent with the efficacy condition, lead to changes in the social cost that exactly offset each other. An increase in α raises the relative cost of effort e and reduces the relative cost of the fine f and the expected discount d. More specifically, to preserve the original policy's force, A^*, the role of e within the probability factor $h(e, f, d)$ must be reduced in favor of f and d. Furthermore, the role of the probability factor $h(e, f, d)$ must be downsized in favor of the factor $f\ (1 - d)$. The latter requires an increase in f and a reduction of d. Therefore, we expect a decrease in e^* and an increase in f^*, while the identification of the overall effect on d^* requires a more formal examination. This will be provided in Theorem 3.

A small increase in β (e.g., stronger public dislike for disproportionate penalties) strengthens the role of e and d and weakens the role of f within the probability factor $h(e, f, d)$ and the factor $f\ (1 - d)$. The latter would require a reduction of f and/or d. Overall, we expect an increase in e^* and a decrease in f^*, while the effect on d^* appears to be ambiguous. In Theorem 3 we will show that this is not the case.

A small increase in γ (e.g., stronger erosion of the respect for the law when guilty firms get away with reduced fines) raises the cost of the expected discount relative to the cost of the effort and the fine. Within the factors $h(e, f, d)$ and $f\ (1-d)$ the role of d must be reduced in favor of e and f. Therefore, one expects that the new optimal antitrust policy must be characterized by a reduced value of d^* and by larger values of e^* and f^*.

Discount parameter ρ: The parameter ρ indicates the independence of the antitrust policy's efficacy from the existence and size of the leniency program (expected discount d^*). As was pointed out earlier, for $\rho \geq 0.61803$ no leniency program should be installed ($d^* = 0$). Here we consider smaller ρ-values such that $d^* > 0$. A small increase in ρ (e.g., improved ethical standards within the management of the firms) raises the probability factor $h(e, f, d)$ and, therefore, the original policy's force such that $A^* > T(n_F^*)$. Theoretically, the increase in A^* allows for reductions of all three policy instruments, e^*, f^*, and d^*. However, the instruments e and f have become more attractive relative to d. Therefore, a strong reduction of d^* accompanied by increases of e^* and f^* is another plausible result. The correct result will be derived in Theorem 3.

Market volume parameters a, b, *and* c: How should the antitrust authority react to changes in the market volume, $(a-c)/b$? A small increase in a or a small reduction in c or b increase the market volume, the sum of consumer and producer rent and, therefore, the value of $T(n_F^*)$ such that $A^* < T(n_F^*)$. To restore the efficacy condition, the policy's force, A^*, must increase. Therefore, we expect an increase in e^*, f^*, and d^*.

Number of firms n: In dynamic markets, new firms can enter. If they join the cartel, the number of fringe firms, n_F, remains constant, while n increases. The profits of the fringe firms are not affected by the additional cartel member. This is also true for the aggregate profit of the cartel. However, the profit per cartel member falls and, therefore, the attractiveness of the cartel status also falls. This allows the antitrust authority to lower the values of the three policy variables e^*, f^*, and d^*, without changing the number of fringe firms.

In this intuitive discussion of the antitrust authority's adaptation to parameter changes, some questions remained unanswered (e.g., the change of d^* when α or β change or the changes of e^* and f^* when ρ changes). The following Theorem 3 provides the missing answers and verifies all of the intuitive conclusions by a rigorous formal analysis.

Theorem 3 *Marginal changes in the social cost parameters* (α, β, γ), *the discount parameter* (ρ), *the market volume parameters* (a, b, c), *or the number of firms* (n), *affect the optimal antitrust policy* (e^*, f^*, d^*), *but not the optimal number of fringe firms,* n_F^*. *The individual effects of the parameter changes are listed in Table 1.*

Proof: See Appendix.

Consider again some optimal policy, (e^*, f^*, d^*), and the corresponding force, A^*. A sufficient increase in e, f, and d and, therefore, of A would induce one of the cartel members to become a fringe firm. This changeover increases the sum of consumer and producer rent by

$$\frac{(a-c)^2}{8b} \frac{2n_F + 3}{(n_F^2 + 3n_F + 2)^2} > 0\ . \tag{31}$$

Table 1: Comparative statics of optimal antitrust policy for given n_F^* ("+" indicates a positive and "−" a negative derivative).

parameter	effort e^*	fine f^*	exp. discount d^*
α	−	+	+
β	+	−	+
γ	+	+	−
ρ	−	−	−
a	+	+	+
b	−	−	−
c	−	−	−
n	−	−	−

We denote this beneficial welfare effect as the positive "competition effect" of the additional fringe firm. For all positive values of n_F, the competition effect is positive and falling in n_F. However, the additional fringe firm also causes a negative "cost effect", because the larger values of e, f, and d raise the social cost $s(e, f, d)$. Since the original force, A^*, was optimal, the cost effect would overcompensate the competition effect and welfare would fall.

However, after a sufficiently large change in the parameters, the original force A^* might be no longer optimal and the competition effect may outweigh the cost effect. For example, consider a significant decrease in the social cost parameters α, β, and γ. For each given n_F, the change in the parameters reduces the negative cost effect, while the competition effect is unaffected. The same is true when ρ increases. If the reduction of the cost effect is sufficiently strong, the increase in A above the original level A^* and the ensuing increase in n_F from n_F^* to n_F^*+1 would be welfare increasing. Theoretically, even a change to n_F^*+2 could be welfare increasing. Note, however, that the competition effect defined by Equation (31) is falling in n_F, while, due to the specification of $h(e, f, d)$, the size of the cost effect tends to increase.

Equation (31) also shows that a significant increase in the market volume, $(a - c)/b$, leads to a strong increase in the competition effect. However, the value of $T(n_F)$ and, therefore, the required values of the three policy instruments also increase (see Theorem 3). Unless the cost function possesses a highly exponential form, the former effect would dominate the latter, such that raising n_F to $n_F^* + 1$ would be welfare increasing.

Finally, suppose that new firms enter the market and that all of them join the cartel. Then, n increases, but n_F remains constant. From Theorem 3 we know that this parameter change allows for a reduction of all three policy instruments. Therefore, for each given n_F, the cost effect falls. The competition effect, however, remains unchanged, because it depends on n_F, but not on n. Therefore, with a sufficiently strong increase in n, raising n_F to $n_F^* + 1$ would be welfare increasing.

8 Concluding Remarks

A model-based derivation of an effective antitrust policy requires an economic framework that includes three actors: a cartel, a group of competing fringe firms, and a welfare maximizing antitrust authority. In existing models of cartel behavior, at least one of these actors is always missing. We take a first step in the present paper to address this situation. Our paper introduces a quantity leadership model with an antitrust authority that has three policy instruments at its disposal: its own effort, a fine for detected cartels, and a leniency program for cartel members that cooperate with the authority. Taking the cost of these instruments into consideration, we derive an optimal antitrust policy. We show that the antitrust authority should reduce the size of the cartel until the resulting gains in the sum of consumer and producer rent (the positive competition effect) no longer overcompensate the resulting increase in social cost (the negative cost effect).

Our analysis reveals that both, the optimal force and the optimal mix of the antitrust authority's policy depend on the characteristics of the specific market. The market characteristics include aspects such as the efficiency of the antitrust authority's operations, the public respect for the rule of law, the ethical standards of the firms' managers, the market volume, and the number of firms operating on the market. With heterogeneous markets, a one-size-fits-all antitrust policy is inappropriate. For example, suppose that there is a public attitude that collusion in the banking sector deserves particularly harsh punishment. In other words, the additional social cost from increasing the fine is low, while the social cost savings from lowering the discount are large. The antitrust authority should respond to this situation by a policy that features a larger fine and a lower discount than in other markets with, otherwise, similar characteristics.

Furthermore, our findings demonstrate that the antitrust authority should recalibrate its policy when changes in the market environment occur. For example, a small increase in the market volume should lead to small reductions of all three policy instruments. These minor adjustments would leave the size of the cartel unchanged. However, if a sufficiently strong expansion of the market volume occurs, the policy instruments should be adjusted in the opposite direction, that is, the antitrust authority should pursue a more forceful policy that induces one or more of the cartel members to become fringe firms.

As pointed out earlier, supergames of collusive behavior assume that the cartel is completely unable to enforce the cartel agreement, while our quantity leadership model assumes perfect enforceability. However, the empirical evidence shows that cartels are impressively creative in designing cartel agreements that allow for limited forms of monitoring and dispute settlement. Therefore, a promising area of future research are oligopoly models that analyze the effects of antitrust policy directed at cartels that have limited means of inducing cooperative member behavior.

Appendix

Lemma 1 *The function $T(n_F)$ defined by (12) increases in n_F and decreases in n.*

Proof: Consider the threshold

$$T(n_F) = \frac{(a-c)^2}{4b} \frac{I(n_F)}{g(n-n_F)}$$

with

$$I(n_F) = \frac{2n_F^2 + (1-n)n_F + 1}{n_F(n-n_F+1)(n_F+1)^2} \; . \tag{32}$$

Differentiation of (32) with respect to n_F (by quotient rule) gives a positive denominator and the numerator

$$(4n_F + 1 - n)n_F(n-n_F+1)(n_F+1)^2 \tag{33}$$
$$-(n-n_F+1)(n_F+1)^2[2n_F^2 + (1-n)n_F + 1] \tag{34}$$
$$-2n_F(n_F+1)(n-n_F+1)[2n_F^2 + (1-n)n_F + 1] \tag{35}$$
$$+n_F(n_F+1)^2[2n_F^2 + (1-n)n_F + 1] \; . \tag{36}$$

The expression in line (33) is equal to

$$(2n_F + 2 - n)n_F(n-n_F+1)(n_F+1)^2 + (2n_F - 1)n_F(n-n_F+1)(n_F+1)^2 \; . \tag{37}$$

We add to the first summand of (37) the expression in line (34) and obtain

$$(n_F - 1)(n-n_F+1)(n_F+1)^2 \geq 0 \; .$$

Next we add to the second summand of (37) the expressions in lines (35) and (36), factor out $n_F(n_F+1)$, and simplify the remaining term to get

$$\begin{aligned} & n_F[4n_F^2 + (2-5n)n_F + 2n^2 + 4] - 3n - 2 \\ &= n_F\left[\left(2n_F + \frac{1}{4}(2-5n)\right)^2 + \frac{7}{16}(n-2)^2 + 3n + 2\right] - 3n - 2 \; . \end{aligned}$$

For $n_F \geq 1$, this expression and, therefore, the expression in lines (33) to (36) are positive and so is the derivative of $I(n_F)$ with respect to n_F: $I(n_F) > I(n_F - 1)$. In addition, $g(n-n_F)$ is increasing in $(n-n_F)$, and therefore, decreasing in n_F: $g(n-n_F) < g(n-n_F+1)$. Therefore, we get

$$\frac{I(n_F)}{g\,(n-n_F)} > \frac{I(n_F-1)}{g\,(n-n_F+1)}$$

which is identical to $T(n_F) > T(n_F - 1)$.

The sign of the derivative of $T(n_F)$ with respect to n is equal to the sign of the derivative of

$$\frac{(n_F+1)^2 - n_F)(n-n_F+1)}{n_F(n-n_F+1)(n_F+1)^2 g(n-n_F+1)} \tag{38}$$

with respect to n. The latter derivative is

$$\frac{-n_F D - n_F(n_F+1)^2\left[\left[\partial g(n-n_F+1)/\partial n\right](n-n_F+1)+g(n-n_F+1)\right]N}{D^2}$$
$$= \frac{-n_F\left[1+(n_F+1)^2\left[\left[\partial g(n-n_F+1)/\partial n\right](n-n_F+1)+g(n-n_F+1)\right]T(n_F)\right]}{D} ,$$

where N denotes the numerator and D the denominator of the quotient (38). We know that $\partial g(n-n_F)/\partial n > 0$ and $D > 0$. Thus, for $T(n_F) \geq 0$, the derivative is negative. ■

Proof of Theorem 1: In Lemma 1 it was shown that $T(n_F)$ is monotonically increasing in n_F. Given (e, f, d), A is some non negative number. Therefore, for each policy (e, f, d) with $A \leq T(n)$, exactly one n_F-value satisfying (13) exists. The right hand inequality in (13) can be transformed into the internal stability condition (8). If it is violated, n_F and, therefore, $T(n_F)$ and $T(n_F+1)$ increase until internal stability is established. The left hand inequality of (13) can be transformed into the external stability condition (9). If it is violated, n_F and, therefore, $T(n_F)$ and $T(n_F+1)$ decrease until external stability is established. ■

The following lemmas will be used in the proofs of Theorems 2 and 3.

Lemma 2 *Differentiation of Equation (25) yields*

$$\frac{\partial^2\mathcal{E}}{\partial d^2} > 0 , \quad \frac{\partial\mathcal{E}}{\partial\rho} < 0 , \quad \frac{\partial^2\mathcal{E}}{\partial d^2} - \frac{1}{\mathcal{E}}\left(\frac{\partial\mathcal{E}}{\partial d}\right)^2 > 0 , \quad \textit{and} \quad \frac{\partial\mathcal{E}}{\partial d}\frac{\partial^2\mathcal{E}}{\partial d\partial\rho} - \frac{\partial\mathcal{E}}{\partial\rho}\frac{\partial^2\mathcal{E}}{\partial d^2} > 0 . \quad (39)$$

For

$$(d+\rho)^2 + 2d + \rho < 1 , \quad (40)$$

differentiation of Equation (25) yields

$$\frac{\partial\mathcal{E}}{\partial d} < 0 , \qquad \frac{\partial^2\mathcal{E}}{\partial d\partial\rho} > 0 , \qquad \textit{and} \qquad \frac{\partial^2\mathcal{E}}{\partial d\partial\rho} - \frac{1}{2\mathcal{E}}\frac{\partial\mathcal{E}}{\partial d}\frac{\partial\mathcal{E}}{\partial\rho} > 0 . \quad (41)$$

Proof: Differentiation of (25) with respect to d gives

$$\frac{\partial\mathcal{E}}{\partial d} = T(n_F)\left[\frac{d+\rho+1}{(d+\rho)(1-d)^2} - \frac{1}{(d+\rho)^2(1-d)}\right] = T(n_F)\frac{(d+\rho)^2+2d+\rho-1}{(d+\rho)^2(1-d)^2} . \quad (42)$$

This derivative is negative, if and only if condition (40) is satisfied.
To prove the first inequality in (39), we differentiate (42) with respect to d:

$$\frac{\partial^2\mathcal{E}}{\partial d^2} = 2T(n_F)\frac{(d+\rho+1)(d+\rho)(1-d)-(1-2d-\rho)\left[(d+\rho)^2+2d+\rho-1\right]}{(d+\rho)^3(1-d)^3}$$
$$= 2T(n_F)\frac{(d+\rho)^2(1-d)+(d+\rho)(1-d)+(2d+\rho-1)(d+\rho)^2+(2d+\rho-1)^2}{(d+\rho)^3(1-d)^3}$$
$$= 2T(n_F)\frac{(d+\rho)^3+(d+\rho)(1-d)+(2d+\rho-1)^2}{(d+\rho)^3(1-d)^3} > 0 . \quad (43)$$

To prove the second inequality in (41), we differentiate (42) with respect to ρ:

$$\frac{\partial^2 \mathcal{E}}{\partial d \partial \rho} = \frac{T(n_F)}{(1-d)^2} \left[\frac{1}{(d+\rho)^2} - 2\frac{2d+\rho-1}{(d+\rho)^3} \right] = \frac{T(n_F)\,(2-3d-\rho)}{(1-d)^2 (d+\rho)^3} > 0 \,, \qquad (44)$$

where the inequality follows from condition (40), because that condition implies that

$$2 > (d+\rho)^2 + 2d + \rho + 1/4 + 3/4 = (d+\rho-1/2)^2 + 3d + 2\rho + 3/4 > 3d + \rho \,.$$

To prove the second inequality in (39), we differentiate (25) with respect to ρ:

$$\frac{\partial \mathcal{E}}{\partial \rho} = \frac{T(n_F)}{1-d} \left[\frac{d+\rho-(d+\rho+1)}{(d+\rho)^2} \right] = \frac{-T(n_F)}{(1-d)\,(d+\rho)^2} < 0 \,. \qquad (45)$$

To prove the third inequality in (41), we insert expressions (25), (42), (44), and (45):

$$\frac{T(n_F)}{(1-d)^2 (d+\rho)^3} \left[2 - 3d - \rho + \frac{(d+\rho)^2 + 2d + \rho - 1}{2(d+\rho+1)} \right] \,.$$

Rearranging the term in square brackets yields

$$\begin{aligned} &\frac{-5d^2 - \rho^2 - 6d\rho + 3\rho + 3}{2(d+\rho+1)} \\ =&\frac{1 - [(d+\rho)^2 + 2d + \rho] + 2 + 2d + 4\rho - 4d\rho - 4d^2}{2(d+\rho+1)} \\ =&\frac{1 - [(d+\rho)^2 + 2d + \rho] + 2(1-d^2) + 2d(1-d) + 4\rho(1-d)}{2(d+\rho+1)} > 0 \,, \end{aligned}$$

where the inequality follows from condition (40).
To prove the third inequality in (39), we insert expressions (25), (42), and (43):

$$T(n_F) \frac{2\left[(d+\rho)^3 + (d+\rho)(1-d) + (2d+\rho-1)^2\right]}{(d+\rho)^3 (1-d)^3} - T(n_F) \frac{\left[(d+\rho)^2 + 2d + \rho - 1\right]^2}{(d+\rho)^3 (1-d)^3 (d+\rho+1)} > 0 \,,$$

where the inequality can be seen after expanding the left quotient by $(d+\rho+1)$. The resulting numerator of that quotient is larger than the numerator of the right quotient:

$$\begin{aligned} &2(d+\rho+1)(d+\rho)^3 + 2(d+\rho+1)(d+\rho)(1-d) + 2(d+\rho+1)(2d+\rho-1)^2 \\ &> 2(d+\rho)^4 + 2(2d+\rho-1)^2 \\ &> \left[(d+\rho)^2 + (2d+\rho-1)\right]^2 \,. \end{aligned}$$

To prove the fourth inequality in (39), we insert expressions (42) to (45):

$$T(n_F)^2 \left[\frac{\left[(d+\rho)^2 + 2d + \rho - 1\right](2-3d-\rho)}{(d+\rho)^5 (1-d)^4} + \frac{2\left[(d+\rho)^3 + (d+\rho)(1-d) + (2d+\rho-1)^2\right]}{(d+\rho)^5 (1-d)^4} \right] \,.$$

The sum of the two numerators is positive, because the first numerator yields

$$\begin{aligned}&\left[(d+\rho)^2+(d+\rho)-(1-d)\right][2(1-d)-(d+\rho)]\\&>-(d+\rho)^3+[(d+\rho)-(1-d)][2(1-d)-(d+\rho)]\\&=-(d+\rho)^3-2(1-d)^2-(d+\rho)^2+3(1-d)(d+\rho)\end{aligned}$$

and the second numerator yields

$$\begin{aligned}&2(d+\rho)^3+2(d+\rho)(1-d)+2\left[(d+\rho)-(1-d)\right]^2\\&=2(d+\rho)^3+2(1-d)^2+2(d+\rho)^2-2(1-d)(d+\rho)\ .\end{aligned}$$

■

Lemma 3 *Differentiation of Equation (27) yields*

$$\frac{\partial\mathcal{D}}{\partial d}<0\ .$$

If condition (40) is satisfied, differentiation of Equation (27) yields

$$\frac{\partial\mathcal{D}}{\partial\rho}<0\qquad\textit{and}\qquad\frac{\partial\mathcal{D}}{\partial T(n_F)}>0\ .$$

Proof: Differentiation of (27) with respect to d yields

$$\begin{aligned}\frac{\partial\mathcal{D}}{\partial d}&=-\frac{\partial^2\mathcal{E}}{\partial d^2}\left[\left(\frac{\alpha\beta}{\mathcal{E}}\right)^{1/2}+\beta\right]+\frac{(\alpha\beta)^{1/2}}{2}\left(\frac{1}{\mathcal{E}}\right)^{3/2}\left(\frac{\partial\mathcal{E}}{\partial d}\right)^2 &(46)\\&=-\beta\frac{\partial^2\mathcal{E}}{\partial d^2}-\left(\frac{\alpha\beta}{\mathcal{E}}\right)^{1/2}\frac{\partial^2\mathcal{E}}{\partial d^2}+\left(\frac{\alpha\beta}{\mathcal{E}}\right)^{1/2}\frac{1}{2\mathcal{E}}\left(\frac{\partial\mathcal{E}}{\partial d}\right)^2\\&=-\beta\frac{\partial^2\mathcal{E}}{\partial d^2}-\left(\frac{\alpha\beta}{\mathcal{E}}\right)^{1/2}\left[\frac{\partial^2\mathcal{E}}{\partial d^2}-\frac{1}{2\mathcal{E}}\left(\frac{\partial\mathcal{E}}{\partial d}\right)^2\right]<0\ , &(47)\end{aligned}$$

where the inequality follows from (39) of Lemma 2.
Differentiation of (27) with respect to ρ yields

$$\begin{aligned}\frac{\partial\mathcal{D}}{\partial\rho}&=-\frac{\partial^2\mathcal{E}}{\partial d\partial\rho}\left[\left(\frac{\alpha\beta}{\mathcal{E}}\right)^{1/2}+\beta\right]+\frac{(\alpha\beta)^{1/2}}{2}\left(\frac{1}{\mathcal{E}}\right)^{3/2}\frac{\partial\mathcal{E}}{\partial d}\frac{\partial\mathcal{E}}{\partial\rho} &(48)\\&=-\beta\frac{\partial^2\mathcal{E}}{\partial d\partial\rho}-\left(\frac{\alpha\beta}{\mathcal{E}}\right)^{1/2}\left[\frac{\partial^2\mathcal{E}}{\partial d\partial\rho}-\frac{1}{2\mathcal{E}}\frac{\partial\mathcal{E}}{\partial d}\frac{\partial\mathcal{E}}{\partial\rho}\right]<0\ , &(49)\end{aligned}$$

where the inequality follows from (41) of Lemma 2.
Inserting (42) in (27) yields

$$\begin{aligned}\mathcal{D}&=-T(n_F)\frac{(d+\rho)^2+2d+\rho-1}{(d+\rho)^2(1-d)^2}\left[\left(\frac{k\,(d)\,(1-d)\,\alpha\beta}{T(n_F)}\right)^{1/2}+\beta\right]\\&=-\frac{(d+\rho)^2+2d+\rho-1}{(d+\rho)^2(1-d)^2}\left[\left(\frac{(d+\rho)\,(1-d)\,\alpha\beta T(n_F)}{(d+\rho+1)}\right)^{1/2}+\beta T(n_F)\right].\end{aligned}$$

Therefore,

$$\frac{\partial \mathcal{D}}{\partial T(n_F)} = -\frac{(d+\rho)^2+2d+\rho-1}{(d+\rho)^2(1-d)^2}\left[\left(\frac{(d+\rho)\,(1-d)\,\alpha\beta}{(d+\rho+1)}\right)^{1/2}\frac{1}{2}T(n_F)^{-1/2}+\beta\right] > 0\ . \tag{50}$$

■

Proof of Theorem 2: Minimizing the monotonically increasing social cost function (24) is equivalent to minimizing the sum

$$\alpha e + \beta f m(f) + \gamma d\ . \tag{51}$$

From (10), (17), (20), and (25) we obtain, for $e > 0$, another formulation of the efficacy condition:

$$f m(f) = \frac{\mathcal{E}}{w(e)}\ . \tag{52}$$

Inserting the right hand side of (52) in (51), we can transform the constrained minimization problem (18) into the unconstrained minimization problem

$$\min_{e,d}\left(\alpha e + \beta\frac{\mathcal{E}}{w(e)} + \gamma d\right)\ . \tag{53}$$

Minimizing this expression with respect to e yields

$$\frac{[w(e)]^2}{w'(e)} = \frac{\beta}{\alpha}\mathcal{E}\ . \tag{54}$$

Exploiting the relationship $[w(e)]^2/w'(e) = e^{-2}$ for the left hand side of (54), taking the square root, and replacing d by $d_{n_F^*}$ gives the cost minimizing effort (28), for given n_F and $d_{n_F^*}$.

Inserting Equation (28) in Equation (21) gives

$$\frac{\mathcal{E}}{w\left(e^*_{n_F}\right)} = \mathcal{E}\frac{[(\beta/\alpha)\,\mathcal{E}]^{1/2}+1}{[(\beta/\alpha)\,\mathcal{E}]^{1/2}} = \mathcal{E} + \left[\left(\frac{\alpha}{\beta}\right)\mathcal{E}\right]^{1/2} = \mathcal{F}\ . \tag{55}$$

Substituting in (52) the function $m(f)$ by its definition (23), substituting the right hand side of (52) by the right hand side of (55), replacing f by $f^*_{n_F}$, and solving for $f^*_{n_F}$ yields the cost minimizing fine (29), for given n_F and $d^*_{n_F}$.

To find the cost minimizing expected discount, $d^*_{n_F}$, we insert the right hand sides of Equations (28) and (55) in (53) to obtain the following minimization problem:

$$\min_{d}\left[(2\alpha\beta\mathcal{E})^{1/2} + \beta\mathcal{E} + \gamma d\right]\ .$$

The first order condition is

$$\left[\left(\frac{\alpha\beta}{\mathcal{E}}\right)^{1/2} + \beta\right]\frac{\partial\mathcal{E}}{\partial d} + \gamma = 0$$

which is equivalent to condition (30). From Lemmas 2 and 3 we know that $\partial\mathcal{E}/\partial d < 0$ and $\partial\mathcal{D}/\partial d < 0$, respectively. This monotony proves the uniqueness of $d^*_{n_F}$ and, therefore, of $e^*_{n_F}$ and $f^*_{n_F}$. ■

Lemma 4 *The parameter change* $\mathrm{d}\alpha > 0$ *leads to* $\mathrm{d}\mathcal{F} > 0$.

Proof: $\mathcal{F}$ is defined in (26). Its differential is

$$\mathrm{d}\mathcal{F} = \frac{\partial\mathcal{F}}{\partial d}\mathrm{d}d + \frac{\partial\mathcal{F}}{\partial\alpha}\mathrm{d}\alpha ,$$

with

$$\frac{\partial\mathcal{F}}{\partial d} = \frac{\partial\mathcal{E}}{\partial d} + \frac{1}{2}\left(\frac{\alpha}{\beta\mathcal{E}}\right)^{1/2}\frac{\partial\mathcal{E}}{\partial d} = \frac{\partial\mathcal{E}}{\partial d}\left[1 + \frac{1}{2}\left(\frac{\alpha}{\beta\mathcal{E}}\right)^{1/2}\right] < 0$$

$$\frac{\partial\mathcal{F}}{\partial\alpha} = \frac{1}{2}\left(\frac{\mathcal{E}}{\alpha\beta}\right)^{1/2} = \frac{1}{2\beta}\left(\frac{\beta\mathcal{E}}{\alpha}\right)^{1/2} > 0 .$$

We get $\mathrm{d}\mathcal{F} > 0$, if and only if

$$\frac{\mathrm{d}\alpha}{\mathrm{d}d} > -\frac{\partial\mathcal{F}/\partial d}{\partial\mathcal{F}/\partial\alpha} = -\frac{\partial\mathcal{E}}{\partial d}\left[2\left(\frac{\alpha\beta}{\mathcal{E}}\right)^{1/2} + \frac{\alpha}{\mathcal{E}}\right] . \tag{56}$$

To determine the value of $\mathrm{d}\alpha/\mathrm{d}d$, we exploit (30) and write the differential of $\mathcal{D}$:

$$\mathrm{d}\mathcal{D} = \frac{\partial\mathcal{D}}{\partial d}\mathrm{d}d + \frac{\partial\mathcal{D}}{\partial\alpha}\mathrm{d}\alpha + \frac{\partial\mathcal{D}}{\partial\beta}\mathrm{d}\beta + \frac{\partial\mathcal{D}}{\partial\rho}\mathrm{d}\rho = 0 . \tag{57}$$

Since only d and α change, we get from (57):

$$\frac{\mathrm{d}\alpha}{\mathrm{d}d} = -\frac{\partial\mathcal{D}/\partial d}{\partial\mathcal{D}/\partial\alpha} .$$

Therefore, (56) can be written in the form

$$-\frac{\partial\mathcal{D}/\partial d}{\partial\mathcal{D}/\partial\alpha} > -\frac{\partial\mathcal{E}}{\partial d}\left[2\left(\frac{\alpha\beta}{\mathcal{E}}\right)^{1/2} + \frac{\alpha}{\mathcal{E}}\right] . \tag{58}$$

Since

$$\frac{\partial\mathcal{D}}{\partial\alpha} = -\frac{1}{2}\frac{\partial\mathcal{E}}{\partial d}\left(\frac{\beta}{\alpha\mathcal{E}}\right)^{1/2} > 0 ,$$

(58) becomes

$$-\frac{\partial\mathcal{D}}{\partial d} > \left(\frac{\partial\mathcal{E}}{\partial d}\right)^2\left[\frac{\beta}{\mathcal{E}} + \frac{1}{2\mathcal{E}}\left(\frac{\alpha\beta}{\mathcal{E}}\right)^{1/2}\right] . \tag{59}$$

Replacing in (59) the derivative $\partial\mathcal{D}/\partial d$ by (47) yields

$$\beta\frac{\partial^2\mathcal{E}}{\partial d^2}+\left(\frac{\alpha\beta}{\mathcal{E}}\right)^{1/2}\left[\frac{\partial^2\mathcal{E}}{\partial d^2}-\frac{1}{2\mathcal{E}}\left(\frac{\partial\mathcal{E}}{\partial d}\right)^2\right]-\frac{1}{2\mathcal{E}}\left(\frac{\partial\mathcal{E}}{\partial d}\right)^2\left(\frac{\alpha\beta}{\mathcal{E}}\right)^{1/2}>\left(\frac{\partial\mathcal{E}}{\partial d}\right)^2\frac{\beta}{\mathcal{E}}$$

$$\beta\frac{\partial^2\mathcal{E}}{\partial d^2}-\beta\frac{1}{\mathcal{E}}\left(\frac{\partial\mathcal{E}}{\partial d}\right)^2+\left(\frac{\alpha\beta}{\mathcal{E}}\right)^{1/2}\left[\frac{\partial^2\mathcal{E}}{\partial d^2}-\frac{1}{\mathcal{E}}\left(\frac{\partial\mathcal{E}}{\partial d}\right)^2\right]>0$$

$$\left[\beta+\left(\frac{\alpha\beta}{\mathcal{E}}\right)^{1/2}\right]\left[\frac{\partial^2\mathcal{E}}{\partial d^2}-\frac{1}{\mathcal{E}}\left(\frac{\partial\mathcal{E}}{\partial d}\right)^2\right]>0\ , \tag{60}$$

where we exploit (39) of Lemma 2. Therefore, condition (56) is satisfied. ■

Lemma 5 *The parameter change* $\mathrm{d}\beta>0$ *leads to* $\mathrm{d}e^*>0$.

Proof: Total differentiation of e^* in (28) yields

$$\mathrm{d}e^*=\frac{\partial e^*}{\partial d}\mathrm{d}d+\frac{\partial e^*}{\partial\beta}\mathrm{d}\beta\ ,$$

with

$$\frac{\partial e^*}{\partial\beta}=\frac{1}{2}\left(\frac{\mathcal{E}}{\alpha\beta}\right)^{1/2}>0\qquad\text{and}\qquad\frac{\partial e^*}{\partial d}=\frac{\partial e^*}{\partial\mathcal{E}}\frac{\partial\mathcal{E}}{\partial d}=\frac{1}{2}\left(\frac{\beta}{\alpha\mathcal{E}}\right)^{1/2}\frac{\partial\mathcal{E}}{\partial d}<0\ .$$

We get $\mathrm{d}e^*>0$, if and only if

$$\frac{\mathrm{d}\beta}{\mathrm{d}d}>-\frac{\partial e^*/\partial d}{\partial e^*/\partial\beta}=-\left(\frac{\beta}{\alpha\mathcal{E}}\right)^{1/2}\frac{\partial\mathcal{E}}{\partial d}\left(\frac{\mathcal{E}}{\alpha\beta}\right)^{-1/2}=-\frac{\beta}{\mathcal{E}}\frac{\partial\mathcal{E}}{\partial d}\ . \tag{61}$$

When only d and β change, we get from (57):

$$\frac{\mathrm{d}\beta}{\mathrm{d}d}=-\frac{\partial\mathcal{D}/\partial d}{\partial\mathcal{D}/\partial\beta}\ .$$

Then, condition (61) can be expressed in the form

$$-\frac{\partial\mathcal{D}/\partial d}{\partial\mathcal{D}/\partial\beta}>-\frac{\beta}{\mathcal{E}}\frac{\partial\mathcal{E}}{\partial d}\ . \tag{62}$$

Since

$$\frac{\partial\mathcal{D}}{\partial\beta}=-\frac{\partial\mathcal{E}}{\partial d}\left[\frac{1}{2}\left(\frac{\alpha}{\beta\mathcal{E}}\right)^{1/2}+1\right]>0\ ,$$

condition (62) becomes

$$-\frac{\partial\mathcal{D}}{\partial d}>\frac{\beta}{\mathcal{E}}\left(\frac{\partial\mathcal{E}}{\partial d}\right)^2\left[\frac{1}{2}\left(\frac{\alpha}{\beta\mathcal{E}}\right)^{1/2}+1\right]$$

which is condition (59). ■

Lemma 6 *The parameter change* $\mathrm{d}\rho > 0$ *leads to* $\mathrm{d}\mathcal{E} < 0$.

Proof: Total differentiation of $\mathcal{E}$ in (25) and dividing through by $\mathrm{d}\rho$ gives

$$\frac{\mathrm{d}\mathcal{E}}{\mathrm{d}\rho} = \frac{\partial \mathcal{E}}{\partial d}\frac{\mathrm{d}d}{\mathrm{d}\rho} + \frac{\partial \mathcal{E}}{\partial \rho} \ .$$

Exploiting (57) yields

$$\frac{\mathrm{d}\mathcal{E}}{\mathrm{d}\rho} = -\frac{\partial \mathcal{E}}{\partial d}\frac{\partial \mathcal{D}/\partial \rho}{\partial \mathcal{D}/\partial d} + \frac{\partial \mathcal{E}}{\partial \rho} = \frac{-\left(\partial \mathcal{E}/\partial d\right)\left(\partial \mathcal{D}/\partial \rho\right) + \left(\partial \mathcal{E}/\partial \rho\right)\left(\partial \mathcal{D}/\partial d\right)}{\partial \mathcal{D}/\partial d} \ . \tag{63}$$

Using (46) and (48), the two summands in the numerator on the right hand side of (63) can be expressed as

$$-\frac{\partial \mathcal{E}}{\partial d}\frac{\partial \mathcal{D}}{\partial \rho} = \frac{\partial \mathcal{E}}{\partial d}\frac{\partial^2 \mathcal{E}}{\partial d \partial \rho}\left[\left(\frac{\alpha\beta}{\mathcal{E}}\right)^{1/2} + \beta\right] - \frac{(\alpha\beta)^{1/2}}{2}\left(\frac{1}{\mathcal{E}}\right)^{3/2}\left(\frac{\partial \mathcal{E}}{\partial d}\right)^2 \frac{\partial \mathcal{E}}{\partial \rho}$$

$$\frac{\partial \mathcal{E}}{\partial \rho}\frac{\partial \mathcal{D}}{\partial d} = -\frac{\partial \mathcal{E}}{\partial \rho}\frac{\partial^2 \mathcal{E}}{\partial d^2}\left[\left(\frac{\alpha\beta}{\mathcal{E}}\right)^{1/2} + \beta\right] + \frac{(\alpha\beta)^{1/2}}{2}\left(\frac{1}{\mathcal{E}}\right)^{3/2}\left(\frac{\partial \mathcal{E}}{\partial d}\right)^2 \frac{\partial \mathcal{E}}{\partial \rho} \ .$$

Therefore, the numerator on the right hand side of (63) is equivalent with

$$\left[\left(\frac{\alpha\beta}{\mathcal{E}}\right)^{1/2} + \beta\right]\left[\frac{\partial \mathcal{E}}{\partial d}\frac{\partial^2 \mathcal{E}}{\partial d \partial \rho} - \frac{\partial \mathcal{E}}{\partial \rho}\frac{\partial^2 \mathcal{E}}{\partial d^2}\right] > 0 \ ,$$

where the inequality follows from (39) of Lemma 2. Thus, $\mathrm{d}\mathcal{E}/\mathrm{d}\rho < 0$. ■

Lemma 7 *The parameter change* $\mathrm{d}T(n_F) > 0$ *leads to* $\mathrm{d}\mathcal{E} > 0$.

Proof: Total differentiation of $\mathcal{E}$ in (25) and dividing through by $\mathrm{d}T(n_F)$ yields

$$\frac{\mathrm{d}\mathcal{E}}{\mathrm{d}T(n_F)} = \frac{\partial \mathcal{E}}{\partial d}\frac{\mathrm{d}d}{\mathrm{d}T(n_F)} + \frac{\partial \mathcal{E}}{\partial T(n_F)} = \frac{-\left(\partial \mathcal{E}/\partial d\right)\left(\partial \mathcal{D}/\partial T(n_F)\right) + \left(\partial \mathcal{E}/\partial T(n_F)\right)\left(\partial \mathcal{D}/\partial d\right)}{\partial \mathcal{D}/\partial d} \ . \tag{64}$$

From Lemma 3 we know that the denominator is negative. Using (42) and (50), the first summand in the numerator on the right hand side of (64) yields

$$-\frac{\partial \mathcal{E}}{\partial d}\frac{\partial \mathcal{D}}{\partial T(n_F)} = \frac{1}{T(n_F)}\left(\frac{\partial \mathcal{E}}{\partial d}\right)^2\left[\frac{1}{2}\left(\frac{\alpha\beta}{\mathcal{E}}\right)^{1/2} + \beta\right] \ . \tag{65}$$

Using

$$\frac{\partial \mathcal{E}}{\partial T(n_F)} = \frac{1}{k(d)(1-d)} = \frac{\mathcal{E}}{T(n_F)}$$

and (47), the second summand in the numerator of (64) yields

$$\frac{\partial \mathcal{E}}{\partial T(n_F)}\frac{\partial \mathcal{D}}{\partial d}=-\frac{\mathcal{E}}{T(n_F)}\frac{\partial^2 \mathcal{E}}{\partial d^2}\left[\left(\frac{\alpha\beta}{\mathcal{E}}\right)^{1/2}+\beta\right]+\frac{1}{2}\frac{1}{T(n_F)}\left(\frac{\alpha\beta}{\mathcal{E}}\right)^{1/2}\left(\frac{\partial \mathcal{E}}{\partial d}\right)^2 . \tag{66}$$

Adding (65) and (66) gives

$$-\frac{\partial \mathcal{E}}{\partial d}\frac{\partial \mathcal{D}}{\partial T(n_F)}+\frac{\partial \mathcal{D}}{\partial d}\frac{\partial \mathcal{E}}{\partial T(n_F)}=\frac{\mathcal{E}}{T(n_F)}\left[\left(\frac{\alpha\beta}{\mathcal{E}}\right)^{1/2}+\beta\right]\left[\frac{1}{\mathcal{E}}\left(\frac{\partial \mathcal{E}}{\partial d}\right)^2-\frac{\partial^2 \mathcal{E}}{\partial d^2}\right] .$$

From (39) of Lemma 2 we know that the term in the last square brackets is negative. Thus, $\mathrm{d}\mathcal{E}/\mathrm{d}T(n_F)>0$. ■

Proof of Theorem 3:

α: $\mathcal{E}$, $\mathcal{D}$, and $\mathcal{F}$ depend on d, but not on e and f. For given d^*, an increase in α increases the value of $\mathcal{D}$ in (27). Since $\partial\mathcal{D}/\partial d<0$ (Lemma 3), in (30) the restoration of the original $\mathcal{D}$-value γ requires an increase in the expected discount d^*. Since $\partial\mathcal{E}/\partial d<0$ (Lemma 2), this leads to a reduction of $\mathcal{E}$. In (28) both, the reduction of $\mathcal{E}$ and the increase in α reduce e^*. Furthermore, the larger α-value directly increases $\mathcal{F}$ in (26), while the d^*-induced reduction of $\mathcal{E}$ reduces $\mathcal{F}$. We know from Lemma 4 that the direct $\mathcal{F}$-increasing effect dominates. In (29) the larger $\mathcal{F}$-value raises the fine f^*.

β: The same line of reasoning as for α leads to an increase in d^* and a lower $\mathcal{E}$-value. The latter effect reinforces the β-induced decrease in $\mathcal{F}$ in (26). This leads to a reduction of f^* in (29). (28) shows that the increase in β and the reduction of $\mathcal{E}$ have opposing effects on the effort e^*. Lemma 5 shows that the former e^*-increasing effect dominates.

γ: (30) shows that the increase in γ necessitates an equal increase in $\mathcal{D}$. Since $\partial\mathcal{D}/\partial d<0$, a reduction of d^* is required. Since $\partial\mathcal{E}/\partial d<0$, the reduction of d^* increases $\mathcal{E}$. (26), (28), and (29) reveal that the larger $\mathcal{E}$-value raises e^* and f^*.

ρ: For given d^*, the increase in ρ and, therefore, $k(d)$ reduces the value of $\mathcal{E}$ in (25). This has a direct $\mathcal{D}$-increasing effect, but also an indirect $\mathcal{D}$-reducing effect, because $\partial\mathcal{E}/\partial d<0$ and $\partial^2\mathcal{E}/(\partial d\partial\rho)>0$ (Lemma 2). From Lemma 3 we know that the latter effect dominates, that is, $\mathcal{D}$ falls. From (30) we can see that d^* must be reduced to restore $\mathcal{D}$ to its former level γ ($\partial\mathcal{D}/\partial d<0$). Since $\partial\mathcal{E}/\partial d<0$, the reduction of d^* raises the value of $\mathcal{E}$, counteracting the previous decrease. In Lemma 5 it is shown that the new value of $\mathcal{E}$ remains smaller than its original value. Thus, e^* and $\mathcal{F}$ and, therefore, f^* decrease.

$(a-c)/b$: For given d^*, the increase in $T(n_F)$ leads to a larger $\mathcal{E}$-value in (25). In (27) the larger $\mathcal{E}$-value directly reduces $\mathcal{D}$. In addition, (42) reveals that the increase in $T(n_F)$ increases the value of $(-\partial\mathcal{E}/\partial d)$ in (27), increasing $\mathcal{D}$. We know from Lemma 3 that the overall effect on $\mathcal{D}$ is positive. To restore $\mathcal{D}$ to its original value γ, d^* must increase. Since $\partial\mathcal{E}/\partial d<0$, the increase in d^* reduces $\mathcal{E}$, counteracting the previous increase. From Lemma 7 we know that the new value of $\mathcal{E}$ is larger than its original value. Thus, also e^* and $\mathcal{F}$ and, therefore, f^* increase.

n: We know from Lemma 1 that $T(n_F)$ decreases in n. Therefore, we get exactly the opposite results as for an increase in the market volume $(a-c)/b$. ■

Literature

Albæk, S. (2013), Consumer Welfare in EU Competition Policy, in: Heide-Jørgensen, C. *et al.* (eds.) Aims and Values in Competition Law, Copenhagen: DJØF Publishing, 67-88.

Allain, M.L., M. Boyer, R. Kotchoni, J.P. Ponssard (2015), Are Cartel Fines Optimal? Theory and Evidence From the European Union, *International Review of Law and Economics*, 42, 38-47.

Aubert, C., P. Rey, W.E. Kovacic (2006), The Impact of Leniency and Whistle-Blowing Programs on Cartels, *International Journal of Industrial Organization*, 24(6), 1241-1266.

Becker, G.S. (1968), Crime and Punishment: An Economic Approach, *Journal of Political Economy*, 76, 169-217.

Bos, I. (2009), Incomplete Cartels and Antitrust Policy: Incidence and Detection, Amsterdam: Tinbergen Institute.

Bos, I., J.E. Harrington (2010), Endogenous Cartel Formation With Heterogeneous Firms, *RAND Journal of Economics*, 41(1), 92-117.

Bos, I., J.E. Harrington (2015), Competition Policy and Cartel Size, *International Economic Review*, 56(1), 133-153.

Brenner, S. (2009), An Empirical Study of the European Corporate Leniency Program, *International Journal of Industrial Organization*, 27(6), 639-645.

Brito, D., M. Catalão-Lopes (2011), Small Fish Become Big Fish: Mergers in Stackelberg Markets Revisited, *The B.E. Journal of Economic Analysis & Policy*, 11(1), Article 24.

d'Aspremont, C., A. Jacquemin, J.J. Gabszewicz, J.A. Weymark (1983), On the Stability of Collusive Price Leadership, *The Canadian Journal of Economics*, 16(1), 17-25.

Daughety, A.F. (1990), Beneficial Concentration, *American Economic Review*, 80(5), 1231-1237.

Donsimoni, M.P. (1985), Stable Heterogeneous Cartels, *International Journal of Industrial Organization*, 3(4), 451-467.

Donsimoni, M.P., N. Economides, H. Polenachakis (1986), Stable Cartels, *International Economic Review*, 27(2), 317-327.

Eaton, C., M. Eswaran (1998), Endogenous Cartel Formation, *Australian Economic Papers*, 37(1), 1-13.

EC (2003), Official Journal of the European Union, L 255/1, 8.10.2003, Case C.37.519 -Methionine, Decision of July 2, 2002.

EC (2010), Commission Decision of 30.06.2010, C(2010) 4387 final, COMP/38.344 - Prestressing Steel.

EC (2011), Press Release IP/11/403, 04.04.2011, https://ec.europa.eu/commission/presscorner/detail/en/IP_11_403.

Escrihuela-Villar, M. (2008), On Endogenous Cartel Size Under Tacit Collusion, *Investigaciones Económicas*, 32(3), 325-338.

Escrihuela-Villar, M. (2009), A Note On Cartel Stability and Endogenous Sequencing With Tacit Collusion, *Journal of Economics*, 96(2), 137-147.

Escrihuela-Villar, M., J. Guillén (2011), On Collusion Sustainability With Stacked Reversion, *Economic Reserach - Ekonomska Istraživanja*, 24(2), 89-98.

Friedman, J.W. (1971), A Non-Cooperative Equilibrium for Supergames, *Review of Economic Studies*, 38(1), 1-12.

Genesove, D., W.P. Mullin (2001), Rules, Communication, and Collusion: Narrative Evidence From the Sugar Institute Case, *American Economic Review*, 91(3), 379-398.

Green, E.J., R.H. Porter (1984), Noncooperative Collusion Under Imperfect Price Information, *Econometrica*, 52(1), 87-100.

Harrington, J.E. (2006), How Do Cartels Operate?, *Foundations and Trends in Microeconomics*, 2(1), 1-105.

Harrington, J.E. (2013), Corporate Leniency Programs When Firms Have Private Information: The Push of Prosecution and the Pull of Pre-Emption, *The Journal of Industrial Economics*, 61(1), 1-27.

Hellwig, M., K. Hüschelrath (2017), When Do Firms Leave Cartels? Determinants and the Impact On Cartel Survival, ZEW Discussion Paper No. 17-002.

Huck, S., K.A. Konrad, W. Müller, H.-T. Normann (2007), The Merger Paradox an why Aspiration Levels Let it Fail in the Laboratory, *The Economic Journal*, 117, 1073-1095.

Konishi, H., P. Lin (1999), Stable Cartels With a Cournot Fringe in a Symmetric Oligopoly, *KEIO Economic Studies*, 36(2), 1-10.

Levenstein, M.C., V.Y. Suslow (2006), What Determines Cartel Success?, *Journal of Economic Literature*, 44(1), 43-95.

Motta, M. (2004), Competition Policy: Theory and Practice, Cambridge University Press, New York.

Motta, M., M. Polo (2003), Leniency Programs and Cartel Prosecution, *International Journal of Industrial Organization*, 21(3), 347-379.

Mouraviev, I., P. Rey (2011), Collusion and Leadership, *International Journal of Industrial Organization*, 29(6), 705-717.

Prokop, J. (1999), Process of Dominant Cartel Formation, *International Journal of Industrial Organization*, 17(2), 241-257.

Shaffer, S. (1995), Stable Cartels With a Cournot Fringe, *Southern Economic Journal*, 61(3), 744-754.

Spagnolo, G. (2005), Divide et Impera: Optimal Leniency Programs, Stockholm School of Economics, *unpublished manuscript*.

Thoron, S. (1998), Formation of a Coalition-Proof Stable Cartel, *Canadian Journal of Economics*, 31(1), 63-76.

Wilson, C.S. (2019), Welfare Standards Underlying Antitrust Enforcement: What You Measure is What You Get, in: Luncheon Keynote Address delivered at the George Mason Law Review 22nd Annual Antitrust Symposium, Arlington, VA.

Zu, L., J. Zhang, S. Wang (2012), The Size of Stable Cartels: An Analytical Approach, *International Journal of Industrial Organization*, 30(2), 217-222.