Stability of Price Leadership Cartel

Yoshio Kamijo[a,*] and Noritsugu Nakanishi[b,†]

[a] *Faculty of Political Science and Economics, Waseda University*
Nishiwaseda 1-6-1, Shunjuku-ku, Tokyo 169-8050, Japan
[b] *Graduate School of Economics, Kobe University*
Rokkodai-cho 2-1, Nada-ku, Kobe 657-8501, Japan

Abstract

This paper studies the farsighted behavior of firms in an oligopolistic market with regard to the formation of a dominant cartel, which has the power to set and control the price in the market. The von Neumann-Morgenstern stable set is adopted as the solution concept. We present two different models of stability of the price leadership cartel. In the first model, the cartel chooses the optimal pricing policy. In contrast, in the second model, we do not assume *a priori* the optimal pricing behavior of the cartel; rather, we show that such behavior arises from the result of firms' consideration of the stability.
JEL classification: C71, D43, L13.
Keywords: price leadership model, cartel stability, foresight, stable set, individual move; optimal pricing; endogenous pricing

*Corresponding author. Tel: +81-3-3203-7391. Email: `kami-jo@suou.waseda.jp`
†Nakanishi acknowledges financial support from Japan Society for the Promotion of Science (Grant-in-Aid for Scientific Research (C), No. 18530175). Email: `nakanishi@econ.kobe-u.ac.jp`

1 Introduction

The study of cartel stability is a traditional topic in oligopoly theory. In particular, collusive pricing behavior, whether it is a result of overt agreement, has been viewed as "the only feasible means of assuring parallel actions among sellers" (Markham (1951, p. 901)); moreover, price leadership cartels have received considerable attention for decades. Although there is an extensive literature on price leadership models with one leader and one follower,[1] studies on the price leadership cartel with many firms are inadequate. In this paper, we study the stability of a price leadership cartel in an oligopolistic market with many (but, finite) symmetric firms.

One of the earliest contributions to the research on the stability of the price leadership cartel is d'Aspremont, Jacquemin, Gabszewicz, and Weymark (1983), which had been a starting point for subsequent studies by other authors. In their model, it is assumed that there is only one cartel in the role of the price leader that announces and sets the price (the size of the cartel in terms of the number of firms in it varies endogenously through entry-exit by firms) and that, taking the price set by the leader as given, the other fringe firms behave in a competitive fashion; in other words, they follow the price-equal-marginal-cost principle.[2] Knowing the responses of the fringe firms, the cartel can derive the residual demand function by subtracting the total supply by the fringe firms from the total demand. Taking account of the derived residual demand, the cartel members determine the price to maximize the (joint) profit.

Let the profits of cartel firms and fringe firms in the price leadership cartel model be denoted by $\pi_c^*(k) \equiv \pi_c(k, p^*(k))$ and $\pi_f^*(k) \equiv \pi_f(p^*(k))$, respectively, where $\pi_c(k,p)$ and $\pi_f(p)$ are the profits of cartel firms and fringe firms when size k cartel sets the price p, and $p^*(k)$ is the optimal price of the size k cartel. In d'Aspremont et al. (1983), a certain size k of the cartel is considered to be "stable" if (i) $\pi_c^*(k) \geqq \pi_f^*(k-1)$, i.e., no firm in the existing size k cartel finds it profitable to exit from the cartel, and (ii) $\pi_f^*(k) \geqq \pi_c^*(k+1)$, i.e., no fringe firm can be better off by entering the existing cartel. There exists a stable size of the cartel in the sense of d'Aspremont et al. (1983) because if there does not exist size $k, 1 \leqq k \leqq n$, satisfying condition (i), size 0 cartel (i.e., the situation where the cartel does not

[1]There are several studies that explore why there is a firm in the position of the price leader. Deneckere and Kovenock (1992) and Furth and Kovenock (1993) have considered a model with the firms' capacity constraints. Pastine and Pastine (2004) have used the endogenous timing model of Hamilton and Slutsky (1990) to examine endogenous role assignment of a leader and follower. van Damme and Hurkens (2004) have also used the endogenous timing model with firms' risk consideration by Harsanyi and Selten (1988).

[2]Ono (1978) regarded such behavior of fringe firms as the optimal. He argued that because, given the price set by the leader, a fringe firm can set a price infinitesimally lower than the one set by the leader and sell q_f that satisfies the price-equal-marginal-cost condition, the fringe firm can maximize its profit. However, there is some difficulty in justifying such behavior of fringe firms in a rigorous non-cooperative game model with finite players, because there must be an interaction among fringe firms. Tasnadi (2000) shows that such behavior can be justified in a non-atomic model of the fringe firms.

exist) is stable since for $k = 0$, (i) automatically holds and (ii) holds by $k = 1$ not satisfying condition (i). If there exists size k, $0 \leqq k \leqq n - 1$, satisfying condition (i) but this k does not satisfy condition (ii), the latter implies size $k+1$ also satisfies condition (i). By repeating this process, we can find size k satisfying (i) and (ii) because the number of the firms, n, is finite and for n, condition (ii) automatically holds.

Although the model in d'Aspremont et al. (1983) is simple and their results are clear, in their analysis, there remains an inadequacy concerning the foresight of the firms, pointed out by Diamantoudi (2005). She argued that the analysis by d'Aspremont et al. (1983) exhibited a certain inconsistency between an implicit assumption of the firms' brightness embedded in the model and the stability criterion that assumes the firms' myopic view. Consider a firm in the cartel consisting of k firms. When the firm contemplates the deviation (exiting from the cartel), it compares the current profit $\pi_c^*(k) = \pi_c(k, p^*(k))$ (the profit of a firm in the size k cartel) with the profit $\pi_f^*(k-1) = \pi_f(p^*(k-1))$ under a new price $p^*(k-1)$ set by a new cartel established after its deviation (the profit of a firm in the fringe with size $k-1$ cartel) but not with the profit $\pi_f(p^*(k))$ under the price $p^*(k)$ set by the current cartel. Since the cartel's pricing behavior is restricted to the optimal pricing at the very outset of the model, the deviating firm should correctly expect the response of the *readjusting* price by the new cartel against its deviation. In this sense, a firm in their model should have the ability to foresee the reaction of the other firms (in particular, those remaining in the cartel) against its deviation. To the contrary, the stability criterion adopted by d'Aspremont et al. (1983) implies that a firm contemplating deviation does not take account of possible subsequent deviations by other firms after its own deviation. That is, the stability criterion assumes that a firm's view is myopic, undermining the foresight of the firm that is assumed by the model.

In view of such inconsistency in the analysis of d'Aspremont et al. (1983), Diamantoudi (2005) has reconsidered the stability of the price leadership cartel by adopting a different stability criterion that incorporates the farsighted perspective of firms. As the stability concept, she adopts von Neumann and Morgenstern (1953) stable set with dominance relations that capture the foresight of the firms. She has shown that there exists a unique set of stable sizes of the cartel in the price leadership model. However, because her existence result of the set of stable sizes of the cartel relies on the general existence theorem of the stable set by von Neumann and Morgenstern (1953), the properties of the stable sizes of the cartel as well as the relation with stable sizes of d'Aspremont et al. (1983) are unclear from her analysis. Recently, we show, in our another paper (Nakanishi and Kamijo 2008), that the minimal stable sizes of the cartel in the sense of Diamantoudi (2005) is also stable in the sense of d'Aspremont et al. (1983) and the maximal stable size of the cartel is large enough to be Pareto-efficient for firms.

However, there is still an inadequacy in the analysis of Diamantoudi (2005), concerning cartel identification; in her model, cartels are identified by their sizes (in terms of the number of firms) and the two distinct cartels with different members are regarded as the same if their sizes are equal. This does not matter much in the case of d'Aspremont et al. (1983), because of the myopia of the firm that is embedded in their stability criterion. This, however, can become a more serious problem when we fully take account of the farsightedness of the firms as in Diamantoudi (2005). Suppose that each firm can foresee a chain reaction of further deviations by other firms after its own deviation. Consequently, it may be the case that one firm in the cartel finds it profitable to exit from the existing cartel and actually do so, expecting that another fringe firm would enter the cartel after its deviation and that the resulting cartel would be stable.

To illustrate this point, consider the following example. Suppose that some size k is considered to be stable from Diamantoudi's discussion and this k satisfies $\pi_c^*(k) > \pi_f^*(k-1)$. Let cartel C^1 be the initial cartel with size k and take i from the members in the cartel and j from the fringe. Note that in general, $\pi_f^*(k) > \pi_c^*(k)$ holds because the fringe firms are able to free ride on the price-raising effort of the cartel. Therefore, firm i wants to replace its position from the cartel members to the fringe members without changing the size of the current cartel; this can be done as follows: first, firm i exits from C^1 and changes the cartel to size $k-1$ cartel C^2, and then, firm j enters the cartel and changes C^2 to size k cartel C^3.

$$C^1 \xrightarrow{i \text{ exits}} C^2 \xrightarrow{j \text{ enters}} C^3$$

In cartel C^2, firm j actually has an incentive to join the cartel because its current profit $\pi_f^*(k-1)$ is smaller than the profit after joining the cartel, $\pi_c^*(k)$, and the resulting size k cartel C^3 is "stable." Expecting the reponse of firm j, firm i actually exits from C^1 because firm i belongs to the fringe at C^3 and in the cartel at C^1. Thus, size k cartel C^1 is considered to be "unstable" if other size k cartel C^3 is "stable."

In the above example, there are two cartels of the same size involved: the initial cartel and the resulting cartel. The former is *not* considered to be stable, while the latter is. That is, when firms are farsighted, two distinct cartels of the same size can have different stability properties; when cartels of equal size are treated the same, this possibility is ignored. Therefore, cartels should be identified by their members (not by the numbers of members). Kamijo and Muto (2008) have argued the same and constructed an appropriate model in which cartels are identified by their members. Next, they have shown that any Pareto-efficient cartel can always be stable with respect to the stability criterion incorporating the firms' farsighted view. However, Kamijo and Muto (2008) allow the sumultanous (or, coalitional) move of the firms when they consdier the stability of the cartel, in contrast to Diamantoudi (2005) that only considers the individual move of firms. Therefore what kinds of problem occures due to size identification of cartels is still unclear.

In this paper, we shall adopt the von Neumann and Morgenstern (1953) stable set as the basis of our stability concept, similar to Diamantoudi (2005) and Kamijo and Muto (2008). Stability offered by the stable set is free of contradictions inside the set of "stable" outcomes and at the same time, accouts for every "unstable" outcome it excludes. Moreover, as pointed out by Harsanyi (1974), the stable set incorporated with players' foresight improves the inconsistency of the stable set itself. The stable set is defined for the pair of a set of outcomes and a dominance relation defined over the set of outcomes, which describe the current market structure. The dominance relation is extended to capture the farsighted view of the firms. The stable set according to the set of outcomes and this extended dominance relation (called the indirect dominance) constitutes our solution concept known as the *farsighted stable set*.

First, we present a model that reveals the cartel identification problem. That is, the price leadership cartel is identified by its members, and each firm has the ability to foresee not only an immediate outcome, but also the ultimate outcome after its deviation. Thus, an outcome of our first model is the set of firms that belong to the cartel. Next, we show that there exist farsighted stable sets in the first model and the stable sets show a complicated figure. We prove this by means of a constructive approach wherein we actually construct a farishgted stable set by some algorism, that is inspired by Nakanishi (2007), who analyzes an n-person prisoners' dilemma game. A difference beween our first result and the results of Diamantoudi is that in our model, a uniqune pattern of farisghed stable sets is not guaranteed. One critical finding from our analysis is that even though certain sizes of cartels are judged to be stable by both Diamantoudi's and our first models, whether all or one of the cartels of this size is stable depends on the profit functions of firms. This point can not be discovered when we identify the cartel with respect to its size.

There still exist certain problems in our first model. As mentioned earlier, the cartel's pricing policy is restricted to the *optimal pricing* in the sense that the cartel sets the price along the residual demand to maximize the joint-profit of the members. Restricting the cartel's pricing to the optimal pricing may seem to constitute an innocuous assumption, but actually it is not. From several fields in economics, we can draw several pieces of evidence that some observed outcomes that satisfy certain criteria of rationality, efficiency and/or optimality, can often be sustained through some irrational, inefficient and/or non-optimal behavior.[3] In sum, non-optimal behavior of a player can work as "punishment" and/or "reward" to other players and, therefore, induce other players' *optimal* responses. Taking account of the possibility of non-optimal behavior has a significant influence on the final outcomes of the model.

In our second model, the cartel is allowed to choose not only the optimal price,

[3]From theoretical perspective, consider the well-known folk theorem: nearly efficient and cooperative outcomes can be maintained through the "punishment" behavior after one player's deviation, which is irrational (at least, in the one shot game) even if the continuation game satisfies the subgame perfection (see Fudenberg and Tirole (1991, Chap. 5)). From empirical standpoint, among the growing literature on experimental economics, consider Fehr and Gachter (2000); they have examined a two-stage game composed of a voluntary contribution game in the first stage and a punishment stage in the second, and shown that the high contributions of the subjects in the voluntary contribution game are realized by the actual use of the punishment option wherein punishing the other subject is irrational for the subject, because it requires a decrease of one unit of his payoff to decrease some units of the other.

but also any positive price; such a flexible pricing policy can be interpreted as punishment and/or reward to the fringe firms and, by this, the cartel can induce the fringe firms to behave *optimally*. While an outcome of our first model is the set of firms that belong to the cartel, an outcome of our second model is a pair of a cartel and a quoted price set by the cartel. The dominance relation in the second model is extended not only to capture the farsighted view of the firms but also to address the endogenous pricing by the cartel. In this setting, we show that the farsighted stable sets become a simple form. Any outcome such that it is Pareto-efficient and the cartel chooses the optimal pricing is a one-point farsighted stable set. Thus, we obtain an efficiency result similar to that of Kamijo and Muto (2008), because of the endogeneity of the price set by the cartel. Further, we also show that although we do not restrict out analysis to the optimal pricing of the cartel, the optimal pricing behavior of the stable cartel emerges as a result of the stability consideration. Therefore, by considering flexible pricing policies, the set of stable cartels undergoes a complete change.

There exist studies that analyze the stability of price leadership using a different approach from that adopted in this paper. Donsimoni, Economides, and Polemarchakis (1986) analyze the stability of a price leadership cartel in a linear demand and quadratic cost setting using the same stability criterion as d'Aspremont et al. (1983). Prokop (1999) considers two non-cooperative games wherein firms form a dominant price leadership cartel. Thoron (1998) also considers the formation of a cartel, which is sufficiently general but slightly different from the price leadership cartel, using several equilibrium concepts, including the coalition-proof Nash equilibrium introduced by Bernheim, Peleg, and Whiston (1987).

The rest of this paper is organized as follows. In the next section, we present a price leadership model and summarize the basic properties of the price leadership model. In Section 3, we present our first model of cartel stability. In Section 4, we present our second model wherein the endogeneity of the pricing is embedded in the definition of our indirect dominance relation. Section 5 constitutes the conclusion. All the proofs of lemmas are relegated to the Appendices.

2 Model

We consider an industry composed of n $(n \geqq 2)$ identical firms, which produce a homogeneous good. The demand for the good is represented by a function $d\colon \mathbb{R}_{++} \to \mathbb{R}_+$:[4]

$$Q = d(p),$$

where p is the price and Q is the total demand for the good. We assume that $d'(p) < 0$ for all $p > 0$.

Each firm has an identical cost function $c(q_i)$, wherein q_i is the output level of a firm (firm i). We assume that c is increasing, twice continuously differentiable in

[4] $\mathbb{R}_{++} = \{a \in \mathbb{R} | a > 0\}$ and $\mathbb{R}_+ = \{a \in \mathbb{R} | a \geqq 0\}$.

q_i, and it satisfies $c(0) = 0$, $c'(0) = 0$, $c'(q_i) > 0$ for $q_i > 0$, and $c''(q_i) > 0$ for $q_i > 0$.

Once k firms have decided to combine and form a cartel, the cartel can exercise its power to determine the market price of the good. The remaining $n - k$ firms constitute a competitive fringe, whose members behave competitively. That is, each firm in the fringe regards the price determined by the cartel as given, and chooses its output level to maximize its own profit. Given the price p, the supply function of a fringe firm, $q_f(p)$, is determined by means of the well-known price-equal-marginal-cost condition:

$$p \equiv c'(q_f(p)).$$

On the basis of the assumptions on the cost function, $q_f(p) > 0$ for all $p \in \mathrm{R}_{++}$ and $\lim_{p\to 0} q_f(p) = 0$.

Given the responses by the fringe firms, the residual demand for the size k cartel can be written as follows:

$$R(k, p) \equiv \max\left\{d(p) - (n-k)q_f(p),\ 0\right\}.$$

To simplify the exposition, we assume that members in the cartel divide their total quantity of production equally. Thus, the production per firm in the cartel can be written as follows:

$$r(k, p) \equiv \frac{R(k, p)}{k}.$$

As a result, the profit of a firm in the cartel can be written as a function of the cartel size k and the price p:

$$\pi_c(k, p) \equiv pr(k, p) - c\left(r(k, p)\right).$$

On the other hand, the profit of a fringe firm can be written as a function of p:

$$\pi_f(p) \equiv pq_f(p) - c\left(q_f(p)\right).$$

$\pi_f(p) > 0$ for all $p > 0$.

The optimal price for the size k cartel is determined by

$$p^*(k) = \arg\max_{p>0} \pi_c(k, p).$$

The profits of a cartel firm and a fringe firm evaluated at the optimal price $p^*(k)$ can be written as functions of the cartel size k: For $k = 1, \ldots, n$,

$$\pi_c^*(k) \equiv \pi_c(k, p^*(k)),$$

and for $k = 1, \ldots, n-1$,

$$\pi_f^*(k) \equiv \pi_f(p^*(k)).$$

If $k = 0$, that is, if there is no cartel, then it is assumed that the market structure is competitive. The competitive equilibrium price, denoted by p^{comp}, is determined by $d(p^{\mathrm{comp}}) = nq_f(p^{\mathrm{comp}})$. Consequently, we have $\pi_f^*(0) = \pi_f(p^{\mathrm{comp}})$. Note that for any k, $r(k, p^{comp}) = q_f(p^{comp})$. This implies that for any k, $\pi_c(k, p^{comp}) = \pi_f(p^{comp}) = \pi_f^*(0)$.

The following proposition is concerned with the profits of firms in the cartel and in the fringe.

Proposition 1. *π_c and π_f satisfy the following properties:*

(i) *If $p \neq p^{\mathrm{comp}}$ and $\pi_c(k,p) > 0$, then $\pi_c(k,p)$ is strictly increasing in k*—**[Size monotonicity of $\boldsymbol{\pi_c}$]**. *If $\pi_c(k,p) \leqq 0$, $\pi_c(k+1,p) \geqq \pi_c(k,p)$ holds. Further, if $p = p^{\mathrm{comp}}$, $\pi_c(k,p)$ is invariant against changes in k;*

(ii) *$\pi_f(p)$ is strictly increasing in p;*

(iii) *$\pi_f(p) \geqq \pi_c(k,p)$ for all p and for all $k = 1, \dots, n-1$ with strict inequality when $p \neq p^{\mathrm{comp}}$.*

Proof. Properties (ii) and (iii) follow immediately from the definition of q_f. Thus, it suffices to show property (i). Partially differentiating $\pi_c(k,p)$ with respect to k, we obtain

$$\frac{\partial \pi_c}{\partial k} = pr_k(k,p) - c'(r(k,p))r_k(k,p) = r_k(k,p)\left\{p - c'(r(k,p))\right\},$$

where

$$r_k(k,p) \equiv \frac{\partial r(k,p)}{\partial k} = \begin{cases} -\dfrac{d(p) - nq_f(p)}{k^2} & \text{if } d(p) - (n-k)q_f(p) \geqq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Note that $q_f(p) > r(k,p)$ if and only if $q_f(p) > d(p)/n$. We first consider the case where $d(p) - (n-k)q_f(p) \geqq 0$. If $r_k(k,p) > 0$ or, equivalently, $q_f(p) > d(p)/n$, then $q_f(p) > r(k,p)$. Because c' is strictly increasing, this result implies $p = c'(q_f(p)) > c'(r(k,p))$. Hence, we have $\partial\pi_c/\partial k > 0$. In turn, if $r_k(k,p) < 0$ or, equivalently, $q_f(p) < d(p)/n$, then $q_f(p) < r(k,p)$. Since $p = c'(q_f(p)) < c'(r(k,p))$, we have $\partial\pi_c/\partial k > 0$ again. If $r_k(k,p) = 0$, then p satisfies $d(p) - nq_f(p) = 0$. Thus, p must be p^{comp}. In this case, $\partial\pi_c/\partial k = 0$

Next, consider the case where $d(p)-(n-k)q_f(p) < 0$. In this case, $r_k(k,p) = 0$ and thus $\partial\pi_c/\partial k = 0$. This, in conjunction with the fact that $d(p)-(n-k)q_f(p)$ is increasing in k and $\partial\pi_c/\partial k \geqq 0$ when $d(p) - (n-k)q_f(p) \geqq 0$ implies that $\pi_c(k,p) \leqq \pi_c(k+1,p)$.

Because $\pi_c(k,p) > 0$ implies $d(p) - (n-k)q_f(p) > 0$, $\pi^c(k,p)$ is increasing in k if $p \neq p^{comp}$ and $\pi_c(k,p) > 0$. $\square$

The next proposition for optimal profits is based on d'Aspremont et al. (1983) and Kamijo and Muto (2008).

Proposition 2. *π_c^* and π_f^* satisfy the following properties:*

(i) *$\pi_c^*(k)$ is increasing in k*—**[Size monotonicity of $\boldsymbol{\pi_c^*}$]***;*

(ii) *$\pi_f^*(k) > \pi_c^*(k)$ for all $k = 1, \ldots, n-1$;*

(iii) *$\pi_c^*(k) > \pi_f^*(0)$ for all $k = 1, \ldots, n$;*

(iv) *$\pi_f^*(k) > \pi_c^*(0)$ for all $k = 1, \ldots, n-1$.*

The first property says that the profit of each cartel firm increases as the cartel size increases. The second property of this proposition says that the profit of a cartel member is less than the profit of associated fringe members. The third and forth properties say that both cartel and fringe firms prefer a situation involving a dominant cartel of any size to one without it.

3 Stability of collusive cartel with optimal pricing

3.1 The model of cartel with optimal pricing

Let $N = \{1, 2, \ldots, n\}$ denote the set of firms (players). Consider an n-vector $x = (x_1, x_2, \ldots, x_n)$ such that for each i, x_i is equal to 0 or 1. Here, $x_i = 1$ means that firm i belongs to the existing cartel, whereas $x_i = 0$ means that firm i does not belong to the cartel. That is, an n-vector x represents a cartel structure. Let $X \equiv \{0, 1\}^n$ be the set of all possible cartel structures. By definition, $x^f \equiv (0, \ldots, 0)$ represents a situation without an actual cartel and $x^c \equiv (1, \ldots, 1)$ represents a situation containing the largest cartel that consists of all the firms. Given $x \in X$, $C(x)$ denotes the set of firms belonging to the cartel at x, that is, $C(x) \equiv \{i \in N |\ x_i = 1\}$. We identify $C(x)$ with the cartel at x. Given $x, y \in X$, $x \wedge y$ denotes a cartel structure z such that $z_i = \min\{x_i, y_i\}$ for $i = 1, \ldots, n$. We can easily verify that $C(x \wedge y) = C(x) \cap C(y)$. For $x \in X$, let us define $|x| \equiv \sum_{i=1}^n x_i$, which signifies the cartel size at x in terms of the number of the participating firms.

Let $Z \equiv \{0, 1, 2, \ldots, n\}$ be the set of possible cartel sizes. For each $h \in Z$, $V(h)$ denotes the set of all cartels of size h: $V(h) \equiv \{x \in X \mid |x| = h\}$. For non-empty $W \subseteq Z$, let $V(W)$ be defined by $V(W) = \cup_{h \in W} V(h)$.

The payoff to a firm depends on the current cartel x and its status (i.e., whether the firm is a member of x or not). The (real-valued) payoff function $f_i(x)$ for firm $i \in N$ is written as follows:

$$f_i(x) = \begin{cases} \pi_c^*(|x|) & \text{if} \quad x_i = 1, \\ \pi_f^*(|x|) & \text{if} \quad x_i = 0. \end{cases}$$

Let $x \in X$ and $y \in X$ be two distinct cartels. We say that a cartel x *Pareto-dominates* y if $f_i(x) \geqq f_i(y)$ holds for all $i \in N$ and strict inequality holds for some j. If x is not Pareto-dominated by any other cartel, the x is called a *Pareto-efficient* cartel. The set of all the Pareto-efficient cartels is denoted by $X^{\text{PE}} \subseteq X$. Since the grand cartel $x^c = (1, \ldots, 1)$ is Pareto-efficient by the size monotonicity of π_c^* and Proposition 2-(iii), X^{PE} is not empty. On the other hand, since $\pi_c^*(n) > \pi_f^*(0)$ by Proposition 2-(iii), $x^f \notin X^{\text{PE}}$. The following lemma characterizes the set of Pareto-efficient cartels.

Lemma A1 (Kamijo and Muto (2008))**.** $X^{\text{PE}} = \{x^c\} \cup \{x \in X \,|\, \pi_f^*(|x|) > \pi_c^*(n)\}$

By Lemma A1, $x \neq x^c$ is Pareto-efficient if and only if the fringe firm of x enjoys a greater profit than that obtained at the grand cartel x^c.

If a firm enters a current cartel or exits from it, the current cartel changes to another. When a cartel $x \in X$ changes to another $y \in X$ through the entry-exit behavior of an individual firm i, we write $x \xrightarrow{i} y$. Formally,

Definition A1 (Inducement Relation with Optimal Pricing)**.** *For $x, y \in X$ and $i \in N$, we have $x \xrightarrow{i} y$ if either*

(i) *$i \in C(x)$ and $C(y) = C(x) \setminus \{i\}$ or*

(ii) *$i \notin C(x)$ and $C(y) = C(x) \cup \{i\}$.*

The first line means that firm i exits from x; the second means that firm i enters x and forms a new cartel y.

The farsightedness of firms is captured by the notion of indirect domination:

Definition A2 (O-Domination)**.** *For outcomes $x, y \in X$, we say that "y indirectly dominates x through optimal pricing," or simply "y O-dominates x," which we write $y \rhd x$ or $x \lhd y$, if and only if there exists a sequence of cartels and firms*

$$x = x^0 \xrightarrow{i_1} x^1 \xrightarrow{i_2} \cdots \xrightarrow{i_M} x^M = y$$

such that for each $m = 1, 2, \ldots, M$,

$$f_{i_m}(x^{m-1}) < f_{i_m}(x^M) = f_{i_m}(y).$$

The pair $(X, \rhd)$ of the set of all possible cartels X and the O-dominance relation $\rhd$ is the abstract system associated with the price-leadership cartel with optimal pricing. Note that, in general, given a set Y and binary relation $\gg$ defined over Y, a pair $(Y, \gg)$ is called an abstract system. A stable set for an abstract system $(Y, \gg)$ is defined as follows:

Definition 1 (Stable Set)**.** *A subset K of Y is said to be a stable set for $(Y, \gg)$ if and only if it satisfies the following two conditions:*

(i) *for any $a \in K$, there does not exist another $b \in K$ such that $b \gg a$,*

(ii) *for any $a \in Y \setminus K$, there exists another $b \in K$ such that $b \gg a$.*

These conditions are called "internal stability" and "external stability," respectively.

A stable set presumes the following stability interpretation of K. Suppose that outcomes in set K are commonly considered to be "stable" and outcomes outside K to be "unstable" by all the individuals. Then, once an outcome a in K is reached, any deviation from a never occurs because there exists no stable outcome that "dominates" a, and if in time an outcome b outside K is reached, there exists stable outcome $a \in K$ that "dominates" y.

The following is our stability concept.

Definition A3 (Farsighted Stable Set with Optimal Pricing)**.** *The farsighted stable set for the price leadership cartel with Optimal Pricing, or simply the farsighted stable set for $(X, \triangleright)$, is a stable set for the abstract system $(X, \triangleright)$.*

3.2 New concepts and properties

Definition A4 (Attractor)**.** *An integer $k \in Z$, $k \geqq 1$, is said to be an "attractor" if $\pi_c^*(k) > \pi_f^*(k-1)$.*[5] *The set of all attractors is denoted by*

$$Z^A \equiv \{k \in Z |\ k \textit{ is an attractor }\}.$$

Let $\bar{a}$ be the *smallest* integer that satisfies (i) $\bar{a}$ is an attractor and (ii) $\bar{a}+1$ is not an attractor. Because "1" is an attractor by Proposition 2-(iii), we can easily verify that an arbitrary h with $1 \leqq h \leqq \bar{a}$ is also an attractor. Let Z^L be the set of such attractors:

$$Z^L \equiv \{h \in Z^A |\ 1 \leqq h \leqq \bar{a}\},$$

which we call the set of "leading attractors." Z^L is always nonempty.

Remark A1. In d'Aspremont et al. (1983), a certain size k of the cartel is considered to be stable if (i) $\pi_c^*(k) \geqq \pi_f^*(k-1)$ and (ii) $\pi_f^*(k) \geqq \pi_c^*(k+1)$. It is easily vefified that $\bar{a}$ is the minimal stable size of the cartel in the sense of d'Aspremont et al. (1983). In fact, $\bar{a}$ is stable because by its definition, $\pi_c^*(\bar{a}) > \pi_f^*(\bar{a}-1)$ and $\pi_f^*(\bar{a}) \geqq \pi_c^*(\bar{a}+1)$, and it is minimal because for any k, $0 \leqq k < \bar{a}$, $\pi_c^*(k+1) > \pi_f^*(k)$ and this implies that k does not satisfy condition (ii).

The following lemma shows an important aspect of attractors.

Lemma A2. *Take an arbitrary attractor $h \in Z^A$. For any distinct cartels $x, y \in V(h)$, we have $x \triangleright y$.*

[5]The condition given in the definition of "attractor" provides a firm with a myopic incentive to enter the size $k-1$ cartel and form a new size k cartel; in a sense, the size k cartel is *attracting* the entering firm.

Thus, two distinct cartels in $V(h)$ O-dominate each other if h is an attractor. An immediate and important consequence of Lemma A2 is that the abstract system $(X, \triangleright)$ contains an infinite chain of dominance: $x \triangleleft x' \triangleleft x'' \triangleleft \cdots$ *ad infinitum*. It is a well-known theorem, due to von Neumann and Morgenstern (1953), that an abstract system that contains *no* infinite chain of dominance admits a stable set. Unfortunately, because our abstract system *does* contain an infinite chain of dominance, we cannot resort to the von Neumann-Morgensters' theorem to establish the existence of farsighted stable sets for our system.

The following lemmas exhibit certain properties of attractors.

Lemma A3. *Let h be an attractor and $h' \in Z$ be an integer with $h < h'$. Then, $x \in V(h)$ O-dominates $y \in V(h')$ if and only if $\pi_f^*(h) > \pi_c^*(h')$.*

Lemma A4. *Suppose $x \in V(h)$, $y \in V(h')$, and $x \triangleright y$. If $h \geqq h'$, then h is an attractor.*

Lemma A3 shows the necessity and sufficient condition for the O-domination of one cartel whose size is an attractor to other greater cartels. An interesting feature is that under the conditions of the lemma, *any* cartel of size h O-diminates *any* cartel of size h'. Lemma A4 says that if one cartel O-dominates another smaller cartel, the size of the dominating cartel must be an attractor.

The next lemma shows a certain property of the set of leading attractors:

Lemma A5. *Take arbitrary cartels $x, y \in X$. If $|x| \in Z^L$ and $|x| \geqq |y|$, then $x \triangleright y$.*

The following lemma shows a close relationship between the farsighted stable sets and the set of attractors.

Lemma A6. *Let $K \subset X$ be an farsighted stable set for $(X, \triangleright)$. Then, we have $K \cap V(Z^A) \neq \emptyset$, where Z^A is the set of all attractors.*

Remark A2. Suppose K is a farsighted stable set for $(X, \triangleright)$. If $K \cap V(a) \neq \emptyset$ for some attractor $a \in Z^A$, $K \cap V(a)$ must be a singleton. In this case, every cartel in $V(a)$ can be a candidate for an element in K. To make the exposition simple and to avoid some notational complexities, we specify a particular cartel $x_a \in V(a)$ for each $a \in Z^A$. That is, if a cartel in $V(a)$ for some $a \in Z^A$ were to be included in a farsighted stable set, we always choose x_a.

Remark A3. Let $\mathcal{V} \equiv \{V(0), V(1), \ldots, V(n)\}$. For $V(h), V(h') \in \mathcal{V}$, if there exist cartels $x \in V(h)$ and $y \in V(h')$ such that $x \triangleright y$, then we write $V(h) \triangleright V(h')$ for notational convenience. It should be noted that, unlike the case for individual cartels, we can have $V(h) \triangleright V(h)$ if h is an attractor. When $V(h) \triangleright V(h')$, we simply say that "$h$ O-dominates h'."

Remark A4. If $V(h) \triangleright V(h')$ holds for h and h' with $h \neq h'$, h O-dominates any integer between h and h'. In other words, if $h > h'$, $V(h) \triangleright V(k)$ holds for any $k = h', h'+1, \ldots, h-1$, and if $h < h'$, $V(h) \triangleright V(k)$ holds for any $k = h-1, h-2, \ldots, h'$.

Remark A5. Diamantoudi (2005) considers the dominance relation defined over Z. For two distinct sizes $k, h \in Z$, "k D-dominates h," which we write $k \blacktriangleright h$, if either of the following two conditions is satisfied:

(i) if $k < h$, $\pi_f^*(k) > \pi_c^*(m)$ for all $m = k+1, k+2, \ldots, h$, or

(ii) if $k > h$, $\pi_c^*(k) > \pi_f^*(m)$ for all $m = h, h+1, \ldots, k-1$.

The farsighted stable set for $(Z, \blacktriangleright)$ is the stable set for abstract system $(Z, \blacktriangleright)$.

An interesting relation between the two dominance relations is that $k \blacktriangleright h$ implies $V(k) \rhd V(h)$. This is checked as follows: Assume $k < h$ for example, take $x \in V(k)$ and $y \in V(h)$ such that $C(x) \subset C(y)$, and put $C(y) \setminus C(x) = \{i_1, i_2, \ldots, i_M\}$, where $M = |C(y) \setminus C(x)|$. Then, the following sequence of deviation

$$y = y^0 \xrightarrow{i_1} y^1 \xrightarrow{i_2} \ldots \xrightarrow{i_M} y^M = x,$$

where y^m is a cartel such that $C(y^m) = \{i_{m+1}, \ldots, i_M\} \cup C(x)$, realizes $x \rhd y$. However, in general, the converse is not true, that is, $V(k) \rhd V(h)$ does not imply $k \blacktriangleright h$.

Because we identify each cartel by its members, $V(k) \rhd V(h)$ does not imply, in general, that for any cartel $x \in V(k)$ and for any $y \in V(h)$, x O-dominates y. However, under certain conditions, this statement holds true. The next lemmas explores the relation between the O-dominance relation over X and the O-dominance relation over Z.

Lemma A7. *Take $k, h \in Z$ with $k > h$ and $\pi_c^*(k) > \pi_f^*(h)$. If $V(k) \rhd V(h)$, then any $x \in V(k)$ O-dominates any $z \in V(h')$, where $h \leqq h' \leqq k$.*

3.3 Procedure (#)

In this subsecition, we construct a procedure, called procedure (#), that chooses the subset of Z that is a candidate of farsighted stable set for $(X, \rhd)$. The procedure consists of two parts: in the first phase, subset of Z is selected by a recursive algorithm, and in the second phase, superfluous elements in the subset obtained from the first phase are deleted according to dominance relation $\rhd$. Let $k = 0, 1, \ldots, n$, be given.

Selection Phase. Let us define a set of integers $\alpha_1(k), \alpha_2(k), \ldots$ according to the following recursive procedure:

- $\alpha_1(k) \equiv k$,
- $\alpha_{j+1}(k) \equiv \min \left\{ h \in Z \,\middle|\, \pi_c^*(h) \geqq \pi_f^*(\alpha_j(k)) \right\}, \quad j = 1, 2, \ldots.$

Because π_c^* is increasing and $\pi_f^*(h) > \pi_c^*(h)$ for all $h = 1, \ldots, n-1$, the above procedure is defined well. We denote the number of integers determined through the above procedure by $J(k)$. Let $H(k)$ be the set of integers determined through the above recursive procedure:

$$H(k) \equiv \left\{\alpha_1(k), \alpha_2(k), \ldots, \alpha_{J(k)}(k)\right\}.$$

It is easy to verify that for each k, $\alpha_j(k)$ is increasing in j. For each k, $H(k)$ can be partitioned into two subsets: one that contains all attractors in $H(k)$, denoted by $H_1(k) \equiv H(k) \cap Z^A$, and the other that contains all non-attractors in $H(k)$, denoted by $H_2(k) \equiv H(k) \setminus H_1(k)$.

Deletion Phase. Further, based on $H(k)$, we define another set of integers, which is the subset of $H(k)$, as follows:

- Set $H^{(1)}(k) \equiv H(k)$ and $\alpha^{(1)}(k) \equiv \alpha_{J(k)}(k)$
- Delete all $h \in H^{(1)}(k)$ satisfying (i) $h < \alpha^{(1)}(k)$ and (ii) $V(h) \lhd V(\alpha^{(1)}(k))$,
- Let $H^{(2)}(k)$ be the resulting set of integers and let $\alpha^{(2)}(k)$ be the *second* largest integer in $H^{(2)}(k)$ (the largest is $\alpha^{(1)}(k)$),
- Delete all $h \in H^{(2)}(k)$ satisfying (i) $h < \alpha^{(2)}(k)$ and (ii) $V(h) \lhd V(\alpha^{(2)}(k))$,
- Let $H^{(3)}(k)$ be the resulting set of integers and let $\alpha^{(3)}(k)$ be the *third* largest integer in $H^{(3)}(k)$ (the largest is $\alpha^{(1)}(k)$ and the second largest is $\alpha^{(2)}(k)$),
- In general, given $H^{(\ell)}(k)$ and $\alpha^{(\ell)}(k)$ (i.e., the ℓth largest integer in $H^{(\ell)}(k)$), delete all $h \in H^{(\ell)}(k)$ satisfying (i) $h < \alpha^{(\ell)}(k)$ and (ii) $V(h) \lhd V(\alpha^{(\ell)}(k))$. Let $H^{(\ell+1)}(k)$ be the resulting set of integers.

Because $H(k)$ is finite, the above deletion phase stops in finite steps, say $T(k)$ steps. We denote the eventual set of integers generated by the deletion procedure as $H^*(k)$. For $i = 1, \ldots, T(k)$, we write each element in $H^*(k)$ as $\alpha_i^*(k) \equiv \alpha_i^{(T(k)-i+1)}(k)$. Then, $\alpha_i^*(k)$ can be lined in an increasing order in i: $\alpha_1^*(k) < \alpha_2^*(k) < \cdots < \alpha_{T(k)}^*(k)$. By definition, we have $\emptyset \neq H^*(k) \subseteq H(k)$ for any $k = 0, 1, \ldots, n$ and, in particular, $H^*(n) = H(n) = \{n\}$. Similar to the partition of $H(k)$, $H^*(k)$ can be partitioned into $H_1^*(k)$ and $H_2^*(k)$.

Procedure (#) can be applicable to other models, as explained in the next two remarks.

Remark A6. If we replace "$V(h) \lhd V(\alpha^{(\ell)}(k))$" in the Deletion Phase of procedure (#) by "$h \blacktriangleleft \alpha^{(\ell)}(k)$," and set $k = \bar{a}$, we have the procedure introduced by Nakanishi and Kamijo (2008), that characterizes the farsighted stable set for $(Z, \blacktriangleright)$. In Nakanishi and Kamijo (2008), it is shown that the resulting subset of Z, $H^*(\bar{a})$, is the unique farsighted stable set for $(Z, \blacktriangleright)$.

Remark A7. Our first model can be seen as the model of an n-person prisoners' dilemma game if we interpret $x_i = 1$ and $x_i = 0$ as "player i cooperates" and "player i defects", respectively, and profit functions of cartel firms and fringe firms are replaced by the payoff functions of cooperators and defectors, respectively.[6] Thus, procedure (#) works for an n-person prisoners' dilemma game. In fact, the resulting set $H^*(0)$ that is obtained by procedure (#) for $k = 0$ is the same as the set obtained from procedure considered by Nakanishi (2007). Because there is no attractor in the prisoners' dilemma game, $H^*(0) = H(0)$. Nakanishi (2007) shows that $\cup_{h \in H(0)} V(h)$ is the unique farsighted stable set for n-person prisoners' dilemma game.

3.4 Results

The next lemma shows a relationship between the set of leading attractors and $H^*(k)$:

Lemma A8. *For any leading attractor $a \in Z^L$, we have $H^*(a) = H^*(\bar{a})$, where $\bar{a}$ is the largest leading attractor.*

Let us define three subsets of Z relating to the largest leading attractor $\bar{a}$:

$$\begin{aligned}
P &\equiv \{\bar{a}\} \cup \left\{ h \in Z \;\middle|\; \begin{array}{l} \exists x \in V(h),\ \exists y \in V(\bar{a}) \text{ such that} \\ x \text{ Pareto-dominates } y \end{array} \right\}, \\
D &\equiv \{h \in Z \mid V(h) \rhd V(\bar{a})\}, \\
F &\equiv \{h \in Z \mid h \in H^*(h)\}.
\end{aligned}$$

P is nonempty by definition. Because $\bar{a}$ is an attractor, we have $V(\bar{a}) \gg V(\bar{a})$; therefore, D is nonempty, too. In general, we have $k \in H(k)$ by definition. However, $k \in H^*(k)$ may fail to be true for some k; that is, k itself may be deleted in the procedure generating $H^*(k)$ from $H(k)$. The following lemma guarantees the existence of k such that $k \in H^*(k)$ and, thereby, shows the nonemptiness of F:

Lemma A9. *Let us define $d^* \equiv \max D$. Then, $d^* \in H^*(d^*)$.*

Remark A8. It is possible to have $d^* = \bar{a}$.

Remark A9. $H^*(k)$ naturally defines a *correspondence* from Z to itself (for completeness, we have to assume $H^*(0) = \emptyset$). Then, Lemma A9 shows the existence of a fixed point of the correspondence $H^* : Z \rightarrow\!\!\rightarrow Z$.

All of P, D, and F are nonempty. Further, we can show that their intersection is also nonempty:

Lemma A10. $P \cap D \cap F \neq \emptyset$.

[6] In an n-person prisoners' dilemma game considered by Okada (1993), Suzuki and Muto (2005), and Nakanishi (2007), the payoff functions of cooperaters and defectors satisfy: (i) $\pi_c^*(k)$ is increasing in k, (ii) $\pi_f^*(k)$ is increasing in k, (iii) $\pi_f^*(k) > \pi_c^*(k+1)$ for all $k = 0, \ldots, n-1$, and (iv) $\pi_c^*(n) > \pi_f^*(0)$.

Now we show the existence of farsighted stable sets in our first model.

Theorem A1 (Existence of farsighted stable sets)**.** *For any $h \in P \cap D \cap F$, the set $K^*(h)$ represented by the following formula is a farsighted stable set for $(X, \rhd)$:*

$$K^*(h) = \{x_a \mid a \in H_1^*(h)\} \cup \bigcup_{k \in H_2^*(h)} V(k). \qquad (*)$$

where $x_a \in V(a)$ (see Remark A2)

Proof. Take an arbitrary $h \in P \cap D \cap F$ and fix it throughout the proof. Note that h is the smallest element in $H^*(h)$ by definition and satisfies $\bar{a} \leqq h$.

[External stability]: Take an arbitrary $x \in X \setminus K^*(h)$. We distinguish four cases: case 1 where $0 \leqq |x| \leqq \bar{a}$, case 2 where $\bar{a} < |x| < h$, case 3 where $\alpha_j^*(h) = |x|$ for some $\alpha_j^*(h) \in H^*(h)$, and case 4 where $\alpha_j^*(h) < |x| < \alpha_{j+1}^*(h)$ for some $\alpha_j^*(h), \alpha_{j+1}^*(h) \in H^*(h)$ or $\alpha_{T(h)}^*(h) < |x|$.

Case 1. We have $V(|x|) \lhd V(\bar{a})$ by Lemma A5 and $V(\bar{a}) \lhd V(h)$ by definition. Moreover, by Lemma A4, h must be an attractor. Take cartels $y \in V(\bar{a})$ and $y' \in V(h)$ such that $x \lhd y$ and $y \lhd y'$. y Pareto-dominates x by the definition of Z^L. If $h > \bar{a}$, then y' Pareto-dominates y; if $h = \bar{a}$, then $y, y' \in V(\bar{a})$. In any case, y' Pareto-dominates x. Then, by simply connecting the sequence realizing $y \lhd y'$ to the one realizing $x \lhd y$, we can construct an appropriate sequence that realizes $x \lhd y'$. Thus, we have $x \lhd x_h$ by Lemma A7.

Case 2. $h \in P$ implies that $\pi_f^*(\bar{a}) < \pi_c^*(h)$. Thus, by Lemma A7, $V(\bar{a}) \lhd V(h)$ implies $x \lhd x_h$.

Case 3. If $\alpha_j^*(h)$ is not an attractor, we cannot have $x \in X \setminus K^*(h)$ by the formula $(*)$. Then, $\alpha_j^*(h)$ must be an attractor. By Lemma A2, x_a, where $a = \alpha_j^*(h) \in H_1^*(h)$, O-dominates x.

Case 4. Note that $\alpha_j^*(h) = \alpha_s(h)$ and $\alpha_{j+1}^*(h) = \alpha_t(h)$ for some $\alpha_s(h), \alpha_t(h) \in H(h)$ with $s < t$. We distinguish two subcases: (a) $\alpha_s(h) < |x| < \alpha_{s+1}(h)$ and (b) $\alpha_{s+1}(h) \leqq |x| < \alpha_t(h) = \alpha_{j+1}^*(h)$.

Case 4-(a). Take a cartel $y \in V(\alpha_s(h)) = V(\alpha_j^*(h))$ such that $C(y) \subset C(x)$ and write $C(x) \setminus C(y) = \{i_1, i_2, \ldots, i_M\}$. Consider the following *decreasing* sequence from x to y:

$$x = x^0 \xrightarrow{i_1} x^1 \xrightarrow{i_2} \ldots \xrightarrow{i_M} x^M = y,$$

in which each firm in $C(x) \setminus C(y)$ exits from x one by one. On the basis of the construction of $H(h)$ and the monotonicity of π_c^*, we have $f_{i_m}(x^{m-1}) = \pi_c^*(|x^{m-1}|) < \pi_f^*(|y|) = f_{i_m}(y)$ for all $m = 1, \ldots, M$. If $\alpha_s(h)$ is not an attractor, $y \in V(\alpha_s(h)) = V(\alpha_j^*(h)) \subset K^*(h)$. If $\alpha_s(h)$ is an attractor, by Lemma A3, x_a O-dominates y where $a = \alpha_s(h) = \alpha_j^*(h)$.

Case 4-(b). By the definition of $H^*(h)$, we have $V(\alpha_{s+1}(h)) \lhd V(\alpha_{j+1}^*(h))$; this implies $V(|x|) \lhd V(\alpha_{j+1}^*(h))$. Similar to case 2, by Lemma A7, we can construct an appropriate sequence that realizes $x \lhd x_a$ where $a = \alpha_{j+1}^*(h) \in H_1^*(h)$.

[Internal stability]: Take two distinct cartels $x, y \in K^*(h)$. If $|x| = |y| = k$ for some $k \in H^*(h)$, k cannot be an attractor. Then, neither $x \rhd y$ nor $y \rhd x$ can hold by Lemma A4.

In turn, let us assume $|x| < |y|$. Note that $|x|, |y| \in H^*(h)$. By the definition of $H^*(h)$, we cannot have $y \rhd x$. Because y Pareto-dominates x by the construction of $H^*(h)$, then $x \rhd y$ cannot hold true, either. $\square$

The next theorem provides the sufficient condition for the uniqueness of the farsighted stable set for $(X, \rhd)$.

Theorem A2. *If $Z^L = Z^A$, then the farsighted stable set for $(X, \rhd)$ is determined uniquely, which is represented by the formula $(*)$.*

Proof. By the definition of Z^L, if $V(h) \rhd V(\bar{a})$ and $h \neq \bar{a}$ for some $h \in Z$, then $h > \bar{a}$ must hold. By Lemma A4, h must be an attractor; this contradicts $Z^L = Z^A$. Therefore, there is no h such that $V(h) \rhd V(\bar{a})$ other than $\bar{a}$. This implies $D = \{\bar{a}\}$. Then, by Lemma A10, we have $P \cap D \cap F = \{\bar{a}\}$. Moreover, because there is no attractor greater than $\bar{a}$, $H(\bar{a}) = H^*(\bar{a})$ holds by Lemma A4.

Let K be a farsighted stable set for $(X, \rhd)$. In order to prove the uniqueness, it suffices to show that $K = K^*(\bar{a})$.

Note that, by Lemma A6 and assumption, we have $K \cap V(Z^L) \neq \emptyset$. Then, there exists $h \in Z^L$ such that $K \cap V(h) \neq \emptyset$. We now show that $h = \bar{a}$. Suppose, in negation, that $h \neq \bar{a}$. Note that $h < \bar{a}$. Take cartels $x \in K \cap V(h)$ and $y \in V(h+1)$. By Lemma A5, y O-dominates x. Accordingly, by the internal stability of K, we have $y \notin K$. Then, by the external stability of K, there must exists a cartel $y' \in K$ that O-dominates y. Because y Pareto-dominates any y'' with $|y''| < |y|$, by the definition of O-domination, $|y'| \geqq |y|$. This implies $y \in Z^L = Z^A$ by Lemma A4. Thus, we must have $h < h + 1 \leqq |y'| \leqq \bar{a}$. Then, by Lemma A5 again, y' O-dominates x; this contradicts the internal stability of K. Hence, $K \cap V(h) \neq \emptyset$ and $h \in Z^L$ together imply $h = \bar{a}$. This implies $K \cap V(\bar{a}) = \{x_{\bar{a}}\}$ and $K \cap V(h) = \emptyset$ for all h with $0 \leqq h < \bar{a}$.

Now, consider an integer h such that $\bar{a} = \alpha_1(\bar{a}) = \alpha_1^*(\bar{a}) < h < \alpha_2^*(\bar{a}) = \alpha_2(\bar{a})$. Since $\pi_f^*(\bar{a}) > \pi_c^*(h)$, Lemma A3 implies that any $x \in V(h)$ is O-dominated by $x_{\bar{a}}$. Then, for all such h, we have $V(h) \cap K = \emptyset$.

Next, consider a case where $h = \alpha_2(\bar{a})$. Note that h is not an attractor. Suppose that there exists a cartel $x \in V(h)$ not included in K. By the external stability of K, there must exist a cartel $y \in K$ that O-dominates x. By construction, $x_{\bar{a}}$ cannot O-dominate x. Then, we have $|y| \geqq |x| = h$. In turn, this implies $|y|$ is an attractor by Lemma A4; this contradicts to $Z^L = Z^A$. Hence, $V(h) \subseteq K$.

Repeatedly applying similar arguments, we can show that for $j = 1, \dots, J(\bar{a})$, (i) $V(h) \cap K = \emptyset$ for any h with $\alpha_j(\bar{a}) = \alpha_j^*(\bar{a}) < h < \alpha_{j+1}^*(\bar{a}) = \alpha_{j+1}(\bar{a})$ or $h > \alpha_{J(\bar{a})}(\bar{a}) = \alpha_{T(\bar{a})}^*(\bar{a})$, and (ii) $V(h) \subseteq K$ for any $h = \alpha_j(\bar{a})$. Hence, $K = K^*(\bar{a})$. $\square$

Finally, we derive certain properties of the farsighted stable set represented by the formula $(*)$.

Theorem A3. *For any $k \in P \cap D \cap F$, (i) $\alpha_1^*(k) = k$ is stable size of the cartel in the sense of d'Aspremont et al. (1983), and (ii) $\alpha_{T(k)}^*(k) \in X^{PE}$. Therefore, $K^*(k) \cap X^{PE} \neq \emptyset$.*

Proof. (i). By the definition of D and Lemma A4, k must be an attractor. Thus, $\pi_c^*(k) > \pi_f^*(k-1)$, the condition (i) of d'Aspremont et al.'s stability. We will show that $k+1$ is not an attractor. We assume, in negation, that $k+1$ is also attractor. This implies that $\alpha_2(k) = k+1 \in H(k)$. If $k+1$ is deleted in the process generating $H^*(k)$, there exists some $\alpha^*(t) \in H^*(k)$ such that $\alpha^*(t) > k+1$ and $V(\alpha^*(t)) \rhd V(k+1)$. Since any $x \in V(\alpha^*(t))$ Pareto-dominates any $y \in V(k+1)$, this implies $V(\alpha^*(t)) \rhd V(k)$ — a contradiction to $k \in F$. On the other hand, if $k+1 \in H^*(k)$, $k \notin H^*(k)$ because $V(k+1) \rhd V(k)$ — a contradiction to $k \in F$. Therefore, $k+1$ is not an attractor, and thus, $\pi_f^*(k) \geqq \pi_c^*(k+1)$, condition (ii) of d'Aspremont et al.'s stability, holds.
(ii). It suffices to show that $x \in V(\alpha_{T(k)}^*(k))$ is Pareto-efficient. By construction, either $\alpha_{T(k)}^*(k) = \alpha_{J(k)}(k)$ is n, or $\alpha_{T(k)}^*(k) = \alpha_{J(k)}(k)$ satisfies $\pi_f^*(\alpha_{T(k)}^*(k)) > \pi_c^*(n)$. By Lemma A1, x is Pareto-efficient. $\square$

From Theorem A3, we know that the minimal size cartels in the stable set $K^*(k)$ is also stable in the sense of d'Aspremont et al. (1983) and the maximal size cartels in the stable set are Pareto-efficient.

3.5 Example

Table 1 shows the relationship between the profits per firm and size of cartel for the selection of certain parameters.[7]

Perfect competition $k=0$		$\swarrow$ $\pi_f^*(0) = 61.7$
$k=1$	$\pi_c^*(1) = 62.5$	$\swarrow$ $\pi_f^*(1) = 63.2$
$k=2$	$\pi_c^*(2) = 64.9$	$\swarrow$ $\pi_f^*(2) = 68.3$
$k=3$	$\pi_c^*(3) = 69.4$	$\pi_f^*(3) = 78.1$
$k=4$	$\pi_c^*(4) = 76.9$ $\nearrow$	$\pi_f^*(4) = 95.9$
$k=5$	$\pi_c^*(5) = 89.3$ $\nearrow$	$\pi_f^*(5) = 129.1$
$k=6$	$\pi_c^*(6) = 111.1$ $\nearrow$	$\pi_f^*(6) = 200.0$
$k=7$	$\pi_c^*(7) = 156.3$ $\nearrow$	$\pi_f^*(7) = 395.5$
Full cooperation $k=8$	$\pi_c^*(8) = 294.1$ $\nearrow$	

Table 1: The relationship between the profits and the cartel size ($n = 8$)

[7] Here, we consider a market with a linear demand function $d(p) = 100 - p$ and identical quadratic cost function of the firms $c_i(q_i) = \frac{1}{2}q_i^2$.

In the example depicted in Table 1, it is easily confirmed that the set of attractors is $\{1, 2, 3\}$, which coincides with the set of leading attractors. The unique stable size of the cartels in the sense of d'Aspremont et al. (1983) is size 3, and on the basis of Lemma A1, the set of Pareto-efficient cartels is $X^{\text{PE}} = V(7) \cup V(8)$.

Now we explore the stable set of this example. Because there is no attractor greater than size 3 and the size 3 cartel Pareto-dominates any other cartel that is smaller than 3, $D = \{h \in Z | V(h) \triangleright \bar{a}\}$ is the set of one element $\bar{a} = 3$. Moreover, by Lemma A10, $P \cap D \cap F = \{3\}$. Hence, Theorem A2 implies that $K^*(3)$ is the unique farsighted stable set. Since k, $k > 3$, is not an attractor, we have $\alpha_1(3) = \alpha_1^*(3) = 3$, $\alpha_2(3) = \alpha_2^*(3) = 5$, and $\alpha_3(3) = \alpha_3^*(3) = 7$. Therefore,

$$K^*(3) = \{x_3\} \cup V(5) \cup V(7).$$

Interestingly, in the model of Diamantoudi (2005), the set of stable sizes of cartels is also $\{3, 5, 7\}$. Thus, both Diamantoudi's and our first models indicate the same set of stable size of cartels in this example. In fact, the coincidence of stable seizes of the cartels between two models generally holds if $Z^L = Z^A$. As the corollary of Theorem A2, by Remark A6, we have:

Corollary A1. *If $Z^L = Z^A$, the farsighted stable set for $(X, \triangleright)$ is represented by*

$$K^*(\bar{a}) = \{x_{\bar{a}}\} \cup \bigcup_{k=2}^{J(\bar{a})} V(\alpha_k(\bar{a})).$$

Moreover, $\{\bar{a}, \alpha_2(\bar{a}), \ldots, \alpha_{J(\bar{a})}(\bar{a})\}$ is the farsighted stable set for $(Z, \blacktriangleright)$.

However, from our analysis, one stable set that represents one standards of behavior in the society, allow only the one cartel of some size k to be stable if k is an attractor and all the cartels of some size k' to be stable if k' is not an attractor. Therefore, even though a certain size k is considered to be stable from Diamantoudi (2005), two distinct cartels of size k can demonstrate a different stability property.

4 Stability of collusive cartel with endogenous pricing

4.1 The model of cartel with endogenous pricing

In the second model, a pair of a cartel structure $x \in X$ and a quoted price $p \in \mathbb{R}_{++}$ describes a market structure; it specifies the current price and the firms in the cartel (and, implicitly, the firms in the fringe). Incidentally, what will happen to the market structure if there is no actual cartel (i.e., if $x = x^f$)? In this case, we assume that (x^f, p^{comp}) will be realized. That is, if there is no actual price-leader, only the competitive equilibrium price p^{comp} will prevail in the market. In other words, any market structure such as (x^f, p) with $p \neq p^{\text{comp}}$ is meaningless. Excluding such meaningless market structures, we now define the set A of all possible market structures:

$$A \equiv \left\{ (x, p) \in \{0, 1\}^n \times \mathbb{R}_{++} \,\middle|\, x \neq x^f \text{ or } (x, p) = (x^f, p^{\text{comp}}) \right\}.$$

We shall call an element in A as an "outcome."[8]

Let g_i be the payoff function of firm i defined on A: For $(x, p) \in A$,

$$g_i(x, p) = \begin{cases} \pi_c(|x|, p) & \text{if} \quad x_i = 1, \\ \pi_f(p) & \text{if} \quad x_i = 0. \end{cases}$$

For a fringe firm, only the quoted price p matters; it does not matter who are the members of the current cartel or how many firms there are in the cartel.

Let us define a set of outcomes where a cartel charges the optimal price, denoted by $A^{\text{OP}} = \{(x, p) \in A \,|\, p = p^*(|x|)\}$. For two distinct market structures $(x, p), (y, w) \in A$, we say that "(y, w) Pareto-dominates (x, p)" if $g_i(y, w) \geqq g_i(x, p)$ for all $i \in N$ and $g_i(y, w) > g_i(x, p)$ for some $i \in N$. The set of Pareto-efficient market structures, denoted by A^{PE}, is a set of outcomes that are not Pareto-dominated. Let us define another subset of A, denoted by A^*, which will turn out to be a subset of A^{PE}:

$$A^* \equiv \left\{ (x, p^*(|x|)) \in A \,\middle|\, x = x^c \text{ or } \pi_f^*(|x|) > \pi_c^*(n) \right\}.$$

Because the largest-cartel optimal-pricing outcome $(x^c, p^*(|x^c|))$ always exists, A^* is nonempty.

Lemma B1. *A^* coincides with the intersection of A^{OP} and A^{PE}.*

Next, we define the inducement relation on A. We assume that each individual firm can enter or exit from the existing cartel freely and, thereby, change the current market structure to another. In the course of entry-exit, only individual moves are allowed, while coalitional (simultaneous) moves are not. Furthermore, we assume that the cartel members can change the current price to another through a unanimous agreement. By changing the price, the cartel can induce another market structure from the current market structure. In general, when a nonempty subset S of N changes a given market structure (x, p) to another (y, w), we write $(x, p) \xrightarrow{S} (y, w)$. The relation $\{\xrightarrow{S}\}_{S \subset N}$ is formally defined as follows:

Definition B1 (Inducement Relation with Endogenous Pricing)**.** *For $(x, p) \in A$, $(y, w) \in A$, and nonempty $S \subset N$, we have $(x, p) \xrightarrow{S} (y, w)$ if either one of the following conditions is satisfied:*

(i) $S = C(x)$ *and* $x = y$,

(ii) $S = \{i\} \neq C(x)$, $x_j = y_j$ *for all* $j \neq i$, *and* $p = w$,

(iii) $S = \{i\} = C(x)$ *and* $(y, w) = (x^f, p^{\text{comp}})$.

[8] If the price p is high enough, the demand $d(p)$ becomes zero and the supply by the fringe firms becomes strictly positive. Therefore, in an outcome (x, p) with a sufficiently high price, the market clearing condition can be violated; in this sense, such an outcome is not feasible. Although we can redefine the set of possible outcomes such that it only includes *feasible* outcomes, this will render the model unnecessarily complicated.

Part (i) means that cartel $C(x)$ can change the current price p to another w through a unanimous agreement between the members. Part (ii) means that a single player i can change the current market structure to another by means of entry-exit from the cartel without affecting the current price. Part (iii) means that if a single player i is the last one member of the current cartel, it can change the current outcome to the competitive equilibrium outcome by exiting from the cartel.

The indirect dominance relation is defined as follows.

Definition B2 (E-Domination)**.** *For $(x,p) \in A$ and $(y,w) \in A$, we say that "(y,w) indirectly dominates (x,p) through endogenous pricing," or simply, "(y,w) E-dominates (x,p)," which we shall write $(y,w) \gtrdot (x,p)$ or $(x,p) \lessdot (y,w)$, if and only if there exists a sequence of outcomes and nonempty coalitions*

$$(x,p) = (x^0,p^0) \xrightarrow{S^1} (x^1,p^1) \xrightarrow{S^2} \cdots \xrightarrow{S^M} (x^M,p^M) = (y,w)$$

such that for each $m = 1,\dots,M$,

$$g_i(x^{m-1},p^{m-1}) < g_i(x^M,p^M) = g_i(y,w)$$

for all $i \in S^m$.

A pair $(A, \gtrdot)$ is called the abstract system associated with the price leadership model with endogenous pricing.

Definition B3 (Farsighted Stable Set with Endogenous Pricing)**.** *A subset K of A is said to be a farsighted stable set for the price leadership cartel with endogenous pricing, or simply the farsighted stable set for $(A, \gtrdot)$, if K is a stable set for abstract system $(A, \gtrdot)$.*

4.2 Results

We first show the following lemma.

Lemma B2. *The largest-cartel optimal-pricing outcome $(x^c, p^*(|x^c|))$ E-dominates any other outcome.*

An immediate consequence of Lemma B2 is that $\{(x^c, p^*(|x^c|))\}$ is a farsighted stable set for $(A, \gtrdot)$. Thus, the existence of the farsighted stable set in the second model is guaranteed.

The next lemma provides the sufficient conditions for E-Domination between the two outcomes.

Lemma B3. *Take distinct outcomes $(x,p),(y,w) \in A$. Assume $\pi_c(|x|,p) > 0$. Then, (x,p) E-dominates (y,w) if either one of the following conditions is satisfied:*

(i) $C(x) \cap C(y) = \emptyset$*, and* $\pi_f(p) > \pi_c(|y|,w)$*;*

(ii) $C(x) \cap C(y) \neq \emptyset$*,* $\pi_f(p) > \pi_c(|y|,w)$ *and* $\pi_c(|x|,p) > \pi_c(|x \wedge y|,w)$*;*

(iii) $C(x) \cap C(y) \neq \emptyset$, $C(x) \not\subset C(y)$, $C(x) \not\supset C(y)$, $p \neq p^{\text{comp}}$, $w \neq p^{\text{comp}}$, *and* $\pi_c(|x|, p) \geqq \pi_c(|y|, w)$.

The following theorem shows that any market structure such that it is Pareto-efficient and the cartel chooses the optimal pricing constitute a one-point farsighted stable set for $(A, \gg)$.

Theorem B1. *For any outcome* $(x, p) \in A^*$, *the singleton set* $\{(x, p)\}$ *constitutes a farsighted stable set.*

Proof. Because the internal stability is satisfied automatically, it suffices to show the external stability. If $(x, p) = (x^c, p^*(|x^c|))$, then the external stability follows from Lemma B2 immediately. Then, let us assume $x \neq x^c$ and $p = p^*(|x|)$. Take an arbitrary $(y, w) \in A$ with $(x, p) \neq (y, w)$. We distinguish three cases: (i) $x = y$; (ii) $x \neq y$ and $|x| \geqq |y|$; (iii) $x \neq y$ and $|x| < |y|$.

Let us consider case (i). By the definition of the inducement relation, we have $(y, w) = (x, w) \xrightarrow{C(y)} (x, p^*(|x|)) = (x, p)$. Further, by the definition of p^*, we can show the following relation: for all $i \in C(y)$,

$$g_i(y, w) = \pi_c(|y|, w) < \pi_c(|y|, p^*(|y|)) = \pi_c(|x|, p^*(|x|)) = g_i(x, p).$$

Then, we obtain $(x, p) \gg (y, p)$.

Next, let us consider case (ii). By the size-monotonicity of π_c, the definition of p^*, and Proposition 2-(ii), we have the following relation:

$$\pi_c(|y|, w) \leqq \pi_c(|x|, w) \leqq \pi_c(|x|, p^*(|x|)) < \pi_f(p^*(|x|)) = \pi_f(p).$$

This relation and the fact $\pi_c(|x|, p^*(|x|)) > 0$ imply both $\pi_f(p) > \pi_c(|y|, w)$ and $\pi_f(p) > 0$. Therefore, if $C(x) \cap C(y) = \emptyset$, then the conditions in Lemma B3-(i) are satisfied. On the other hand, if $C(x) \cap C(y) \neq \emptyset$, we have $\pi_c(|x \wedge y|, w) \leqq \pi_c^*(|x \wedge y|) < \pi_c^*(|x|) = \pi_c(|x|, p^*(|x|)) = \pi_c(|x|, p)$ by the size-monotonicity of π_c^*. Then, the conditions in Lemma B3-(ii) are satisfied. Thus, we obtain the desired result.

Finally, let us consider case (iii). Since $(x, p) \in B$ and $(x, p) \neq (x^c, p^*(|x^c|))$, we have $0 < \pi_c^*(|x^c|) < \pi_f^*(|x|) = \pi_f(p^*(|x|)) = \pi_f(p)$. By the definition and the size-monotonicity of π_c^*, we have $\pi_c(|y|, w) \leqq \pi_c^*(|y|) \leqq \pi_c^*(|x^c|) = \pi_c^*(n)$. Combining these inequalities, we obtain $\pi_f(p) > \pi_c(|y|, w)$ and $\pi_f(p) > 0$. If $C(x) \cap C(y) = \emptyset$, then the conditions in Lemma B3-(i) are satisfied. If $C(x) \cap C(y) \neq \emptyset$, then we have $\pi_c(|x \wedge y|, w) \leqq \pi_c(|x|, w) < \pi_c(|x|, p^*(|x|)) = \pi_c(|x|, p)$. Consequently, the conditions in Lemma B3-(ii) are satisfied. □

As shown in Lemma B1, any outcome in A^* is Pareto-efficient. As a result, Theorem B1 shows that an efficient outcome can be attained as an *ultimate* outcome in an essentially *noncooperative* circumstance through the solution concept of the farsighted stable set . A similar result to Theorem B1 has been attained by

Kamijo and Muto (2008).[9] However, in their model, it is assumed that even firms that are not the members of the current cartel can make joint deviations and that the current cartel sets the price at the optimal, joint-profit-maximizing level automatically. Because the cooperative actions by the firms are embedded in their model at the very outset, it is natural to attain the efficiency result. On the other hand, because in our model, it is assumed that joint entry or exit by a group of firms are not allowed and only the members of the current cartel can make joint moves (of changing price) through a unanimous agreement, it is somewhat surprising to obtain the efficiency result.

The key in our model is the endogeneity of the price. Let us consider, for example, the largest-cartel optimal-pricing outcome $(x^c, p^*(|x^c|))$, which constitutes a farsighted stable set, and another nonstable outcome (x, p) with a smaller size cartel $C(x)$. Even if $(x^c, p^*(|x^c|))$ is better than (x, p) for the members of $C(x)$, the members of $C(x)$ can do nothing except for waiting for entry by other firms when the price in (x, p) is determined automatically through the optimal pricing rule as in Kamijo and Muto (2008). On the other hand, if $C(x)$ can control the price, it can force the remaining fringe firms to enter the cartel by decreasing the price to zero and, thereby, form the largest cartel $C(x^c)$. Once the largest cartel $C(x^c)$ has been formed, it can choose the optimal monopoly price and render its members (i.e., all firms) better-off.

To compare with the results of the first model in the previous section, the farsighted stable set in the second model becomes a simple form. It is characterized only by Pareto-efficiency and the optimal pricing. While one cartel in the farsighted stable set in the first model is Pareto-efficient, all of the cartels in the farsighted stable sets are Pareto-efficient in the second model. Moreover, as the next theorem shows, the unique pattern of the farsighted stable sets is guaranteed in the second model without additional conditions.

We have to prepare the additional lemma to show the uniqueness of our farsighted stable set mentioned in Theorem B1.

Lemma B4. *Let K be a farsighted stable set. Then, for any $(x, p) \in K$, we have $\pi_c(|x|, p) > 0$.*

The next theorem shows that there is no other type of a farsighted stable set.

Theorem B2. *There is no other type of farsighted stable sets than the one described in Theorem B1.*

Proof. Let K be a farsighted stable set. If $K \cap A^* \neq \emptyset$, then K must be a singleton; otherwise, it violates the internal stability. In this case, K is of the type just described in Theorem B1. Then, we can assume $K \cap A^* = \emptyset$. In the following, we prove by contradiction that this cannot be the case. Specifically, we show that, under the condition $K \cap A^* = \emptyset$, there is an *infinite* sequence (x^1, p^1), $(x^2, p^2), \ldots$

[9] Suzuki and Muto (2005) have also shown a similar result to Kamijo and Muto (2008) in an n-person prisoners' dilemma.

of outcomes in K such that $|x^1| > |x^2| > \cdots$. This contradicts the finiteness of the number of firms.

The fact $(x^c, p^*(|x^c|)) \in A^*$ implies $(x^c, p^*(|x^c|)) \notin K$. By the external stability of K, there must exist an outcome $(x^1, p^1) \in K$ that E-dominates $(x^c, p^*(|x^c|))$. For outcome $(x^1, p^1) \in K$, we show the following claim:

Claim 1. (i) $|x^1| \neq 0$ and $p^1 \neq p^{comp}$, (ii) $|x^1| < |x^c|$, and (iii) $p^1 \neq p^*(|x^1|)$.

(i). Suppose, in negation, that $x^1 = x^f$. Because we have $g_i(x^c, p^*(|x^c|)) = \pi_c^*(|x^c|) > \pi_f^*(0) = g_i(x^f, p^{\text{comp}})$ for all $i \in N$ by Proposition 2-(iii), then no player wants to deviate from $(x^c, p^*(|x^c|))$ toward $(x^1, p^1) = (x^f, p^{\text{comp}})$—a contradiction. By the same reason, $p^1 \neq p^{comp}$.
(ii). If $x^1 = x^c$, we must have $p^1 \neq p^*(|x^c|)$. Then, by the definition of p^*, we have

$$g_i(x^1, p^1) = \pi_c(|x^c|, p^1) < \pi_c(|x^c|, p^*(|x^c|)) = g_i(x^c, p^*(|x^c|))$$

for all $i \in N = C(x^c)$. This implies that (x^1, p^1) cannot E-dominate $(x^c, p^*(|x^c|))$. Hence, $x^1 \neq x^c$ must hold and thus, $|x^1| < |x^c|$.
(iii). Let S be the first coalition in a sequence that realizes $(x^c, p^*(|x^c|)) \lessdot (x^1, p^1)$ and suppose, in negation, that $p^1 = p^*(|x^1|)$. Because $(x^1, p^1) \in K$ implies $(x^1, p^1) \notin A^*$, we have

$$\pi_c^*(|x^c|) \geqq \pi_f^*(|x^1|) = \pi_f(p^*(|x^1|)) = \pi_f(p^1).$$

Further, for any firm $i \in N \setminus C(x^1) = C(x^c) \setminus C(x^1)$, we have[10]

$$g_i(x^c, p^*(|x^c|)) = \pi_c^*(|x^c|) \geqq \pi_f(p^1) = g_i(x^1, p^1).$$

Then, $S = C(x^c)$ cannot be true. Therefore, S must be a singleton $\{i_1\}$ for some $i_1 \in C(x^1)$; otherwise, the definition of the E-domination will be violated. For i_1, we have $\pi_c^*(|x^c|) = g_{i_1}(x^c, p^*(|x^c|) < g_{i_1}(x^1, p^1) = \pi_c(|x^1|, p^1) = \pi_c(|x^1|, p^*(|x^1|)) = \pi_c^*(|x^1|)$. However, since $x^c \neq x^1$ implies $|x^1| < |x^c| = n$, the inequality $\pi_c^*(|x^c|) < \pi_c^*(|x^1|)$ contradicts the size-monotonicity of π_c^*. Hence, $p^1 \neq p^*(|x^1|)$.

Let us consider an outcome $(x^1, p^*(|x^1|))$. Because $C(x^1)$ can induce $(x^1, p^*(|x^1|))$ from (x^1, p^1) by changing price and, because we have

$$g_i(x^1, p^1) = \pi_c(|x^1|, p^1) < \pi_c(|x^1|, p^*(|x^1|)) = g_i(x^1, p^*(|x^1|))$$

for all $i \in C(x^1)$, $(x^1, p^*(|x^1|))$ E-dominates (x^1, p^1). By the internal stability of K, $(x^1, p^*(|x^1|))$ cannot be in K. As a result, there must exist an outcome $(x^2, p^2) \in K$ that E-dominates $(x^1, p^*(|x^1|))$.

For this (x^2, p^2), we show the following claim:

[10] Firm i is a cartel member at x^c, but a fringe firm at x^1.

Claim 2. There is at least one firm in $C(x^1)$ that is worse-off in (x^2, p^2) than in $(x^1, p^*(|x^1|))$.

Let $S^1, S^2, \ldots, S^M$ be the sequence of coalitions that appear in a sequence that realizes $(x^1, p^*(|x^1|)) \lessdot (x^2, p^2)$:

$$(x^1, p^*(|x^1|)) = (y^0, w^0) \xrightarrow{S^1} (y^1, w^1) \xrightarrow{S^2} \cdots$$
$$\cdots \xrightarrow{S^{M-1}} (y^{M-1}, w^{M-1}) \xrightarrow{S^M} (y^M, w^M) = (x^2, p^2).$$

Suppose, in negation, that every firm in $C(x^1)$ is not worse-off in (x^2, p^2) than in $(x^1, p^*(|x^1|))$. Because $(x^1, p^1) \xrightarrow{C(x^1)} (x^1, p^*(|x^1|))$ and $(x^1, p^1) \lessdot (x^1, p^*(|x^1|))$, every firm in $C(x^1)$ is strictly better-off in $(x^1, p^*(|x^1|))$ than in (x^1, p^1) and, therefore, also strictly better-off in (x^2, p^2) than in (x^1, p^1). Moreover, we have the following inducement relation:

$$(x^1, p^1) \xrightarrow{C(x^1)} (x^1, p^*(|x^1|)) \xrightarrow{S^1} \cdots \xrightarrow{S^M} (x^2, p^2).$$

Thus, (x^2, p^2) E-dominates (x^1, p^1); however, this contradicts the internal stability of K. Hence, there must be at least one firm in $C(x^1)$ that becomes worse-off in (x^2, p^2) than in $(x^1, p^*(|x^1|))$. We denote the set of such firms in $C(x^1)$ by T; note that $\emptyset \neq T \subset C(x^1)$.

Next, we show certain properties of (x^2, p^2) similar to ones of (x^1, p^1) described in Claim 1.

Claim 3. (i) $|x^2| \neq 0$ and $p^2 \neq p^{comp}$, (ii) $|x^2| < |x^1|$, and (iii) $p^2 \neq p^*(|x^2|)$.

(i) Suppose, in negation, that $x^2 = x^f$. Because we have $g_i(x^1, p^*(|x^1|)) = [\pi_c^*(|x^1|) \text{ or } \pi_f^*(|x^1|)] > \pi_f^*(0) = g_i(x^f, p^{\text{comp}})$ for all $i \in N$ by Proposition 2-(iii) and -(iv), then no player wants to deviate from $(x^1, p^*(|x^1|))$ toward $(x^2, p^2) = (x^f, p^{\text{comp}})$—a contradiction. By the same reason, $p^2 \neq p^{comp}$.
(ii) Again, let $S^1, S^2, \ldots, S^M$ be the sequence of coalitions that appear in a sequence that realizes $(x^1, p^*(|x^1|)) \lessdot (x^2, p^2)$. We have to distinguish two cases: case (a) where $S^m \cap T = \emptyset$ for all $m = 1, \ldots, M$ and case (b) where $S^m \cap T \neq \emptyset$ for some m.

Let us consider case (a). In this case, no firm in T exits from $C(x^1)$. In other words, all firms in T remain inside the cartel all the way along the sequence. Then, if a certain coalition S in the sequence were to change the price from $p^*(|x^1|)$ to another one, then S must include T by the definition of the E-domination. However, this contradicts $S^m \cap T = \emptyset$ for all $m = 1, \ldots, M$. Therefore, the price remains unchanged along the sequence. Then, for each $i \in T$, we have

$$\pi_c(|x^1|, p^*(|x^1|)) = g_i(x^1, p^*(|x^1|)) > g_i(x^2, p^2) = \pi_c(|x^2|, p^*(|x^1|)).$$

By the size-monotonicity of π_c, the inequality $|x^1| > |x^2|$ follows.

In turn, let us consider case (b). We first show that $T = C(x^1) \cap C(x^2)$ and, then, we proceed to the proof of $|x^1| > |x^2|$. Let S^{k+1} be the first coalition in the sequence that contains at least one firm in T (that is , $S^{k+1} \cap T \neq \emptyset$ and $S^m \cap T = \emptyset$ for all $m \leqq k$) and let (y^k, w^k) be the outcome from which S^{k+1} deviates. Further, by the same reason just described in the above paragraph, $w^k = p^*(|x^1|)$ must hold. Take an arbitrary firm $i \in S^{k+1} \cap T$; note that $i \in C(x^1)$ and $i \in C(y^k)$. Then, for firm i, we have

$$
\begin{aligned}
&g_i(x^1, p^*(|x^1|)) = \pi_c(|x^1|, p^*(|x^1|)),\\
&g_i(y^k, w^k) = \pi_c(|y^k|, w^k) = \pi_c(|y^k|, p^*(|x^1|)),\\
&g_i(x^1, p^*(|x^1|)) > g_i(x^2, p^2) > g_i(y^k, w^k).
\end{aligned}
$$

Combining these [in]equalities, we obtain $\pi_c(|x^1|, p^*(|x^1|)) > \pi_c(|y^k|, p^*(|x^1|))$. By the size-monotonicity of π_c, we have $|x^1| > |y^k|$. This implies that some firms in $C(x^1) \setminus T$ have to exit from the cartel before (y^k, w^k) is reached; in other words, we must have $C(x^1) \setminus T \neq \emptyset$.

Consider arbitrary firms i and j such that $i \in C(x^1) \setminus T$ and $j \in T$. By the definition of T, the status of firm i at (x^2, p^2) must be different from that of firm j at (x^2, p^2). There are two cases: one where $i \in C(x^2)$ and $j \in N \setminus C(x^2)$ and the other where $i \in N \setminus C(x^2)$ and $j \in C(x^2)$. In the former, we have

$$
\begin{aligned}
\pi_c(|x^1|, p^*(|x^1|)) &= g_i(x^1, p^*(|x^1|)) \leqq g_i(x^2, p^2) = \pi_c(|x^2|, p^2),\\
\pi_c(|x^1|, p^*(|x^1|)) &= g_j(x^1, p^*(|x^1|)) > g_j(x^2, p^2) = \pi_f(p^2).
\end{aligned}
$$

Combining these inequalities and taking account of Proposition 1-(iii), we arrive at a contradiction:

$$\pi_f(p^2) < \pi_c(|x^1|, p^*(|x^1|)) \leqq \pi_c(|x^2|, p^2) < \pi_f(p^2).$$

Thus, the former case is not possible and the latter case must hold. The latter case produces two implications: one is that $j \in T$ implies $j \in C(x^2)$ and the other is that $i \in C(x^2)$ implies $i \in T$ *or* $i \notin C(x^1)$. From the former implication, we obtain $T \subset C(x^1) \cap C(x^2)$. Similarly, from the latter implication, we obtain $T \supset C(x^1) \cap C(x^2)$. Hence, $T = C(x^1) \cap C(x^2)$.

Now, we prove $|x^1| > |x^2|$ for case (b). We assume, to the contrary, that $|x^1| \leqq |x^2|$. The facts $C(x^1) \setminus T \neq \emptyset$, $T = C(x^1) \cap C(x^2)$, and $|x^1| \leqq |x^2|$ imply $C(x^2) \setminus T \neq \emptyset$. Then, we have both $C(x^1) \not\subset C(x^2)$ and $C(x^1) \not\supset C(x^2)$. Further, by Lemma B4 we have $\pi_c(|x^1|, p^1) > 0$ and $\pi_c(|x^2|, p^2) > 0$. Therefore, by Lemma B3-(iii), one of (x^1, p^1) and (x^2, p^2) E-dominates the other outcome. This contradicts the internal stability of K. Hence, $|x^1| > |x^2|$.

(iii) We will show that $p^2 \neq p^*(|x^2|)$. The proof varies slightly from that in the case of (x^1, p^1). Assume, in negation, that $p^2 = p^*(|x^2|)$. By the very definition of (x^2, p^2), there is a dominance sequence from $(x^1, p^*(|x^1|))$ to $[(x^2, p^2) = (x^2, p^*(|x^2|))]$. Let S be the first coalition that appears in the dominance sequence.

By the fact $|x^1| > |x^2|$, the size monotonicity of π_c^*, and Proposition 2-(ii), we have

$$\begin{aligned}\pi_c(|x^2|, p^*(|x^2|) = \pi_c^*(|x^2|) &< \pi_c^*(|x^1|) \\ &= \pi_c(|x^1|, p^*(|x^1|)) \\ &< \pi_f^*(|x^1|) = \pi_f(|x^1|, p^*(|x^1|)).\end{aligned}$$

Thus, any player i in S must be in the fringe position at the final outcome $(x^2, p^*(|x^2|))$. When the firm belongs to the cartel at $(x^1, p^*(|x^1|))$, $\pi_f(p^*(|x^2|)) > \pi_c(|x^1|, p^*(|x^1|))$ must hold by the incentive of deviation, and when the firm belongs to the fringe at $(|x^1|, p^*(|x^1|))$, $\pi_f(p^*(|x^2|)) > \pi_f(p^*(|x^1|)) = \pi_f^*(|x^1|) > \pi_c^*(|x^1|) = \pi_c(|x^1|, p^*(|x^1|))$. By Proposition 2-(ii) and the definition of p^*, in both cases, we have

$$\pi_f(p^*(|x^2|)) > \pi_c(|x^1|, p^*(|x^1|)) \geqq \pi_c(|x^1|, p^1)).$$

In the case of $C(x^1) \cap C(x^2) \neq \emptyset$,

$$\pi_c(|x^2|, p^*(|x^2|)) = \pi_c^*(|x^2|) > \pi_c^*(|x^1 \wedge x^2|) \geqq \pi_c(|x^1 \wedge x^2|, p^1),$$

where the second inequality is by the size monotonicity of π_c^* and the third is by the definition of p^*. Hence, by Lemma B3-(i) and (ii), $(x^2, p^*(|x^2|))$ E-dominates (x^1, p^1). This contradicts the internal stability of K and thus $p^2 \neq p^*(|x^2|)$.

Then, the outcome $(x^2, p^*(|x^2|))$ E-dominates (x^2, p^2) and it is not in K; therefore, there must exist $(x^3, p^3) \in K$ that E-dominates $(x^2, p^*(|x^2|))$. In addition, (x^3, p^3) must satisfy $|x^2| > |x^3|$ and $x^3 \neq x^f$. Generally, we have the following two claims:

Claim 4. Assume that $(x^k, p^k) \in K$, $0 < |x^k| < n$, and $p^k \neq p^*(|x^k|)$. Then, there exits $(x^{k+1}, y^{k+1}) \in K$ such that (x^{k+1}, y^{k+1}) E-dominates $(x^k, p^*(|x^k|))$.

Claim 5. For $(x^{k+1}, p^{k+1}) \in K$ described in Claim 4, the followings hold: (i) $x^{k+1} \neq x^f$ and $p^{k+1} \neq p^{comp}$, (ii) $|x^{k+1}| < |x^k|$, and (iii) $p^{k+1} \neq p^*(|x^{k+1}|)$.

Repeatedly applying Claims 4 and 5 alternatively, we obtain an infinite sequence of outcomes $(x^1, p^1), (x^2, p^2), \ldots$ such that $|x^1| > |x^2| > \cdots$. This contradicts the finiteness of the number of firms. Hence, finally, we obtain the desired result: $K \cap A^* \neq \emptyset$. □

5 Conclusion

In this paper, we considered the stability of price leadership cartel when each firm has the ability to foresee the future and only individual moves are allowed to the firms. In such a situation, we present two different models. In the first, the price set by the cartel is restricted to the optimal one, and in the second, the cartel can choose any positive price.

In Table 2, the farsighted stable sets described in this paper and the relevant one are shown. In the first model, we present procedure (#) that constructs a farsighted stable set for this model, and show that the minimal size of the stable cartels is also stable in the sense of d'Aspremont et al. (1983) and the maximal stable cartels are Pareto-efficient. Moreover, we provide a sufficient condition for the uniqueness of the farsighted stable set and show that under this condition, both Diamantoudi (2005) and the first model indicate the same set of stable sizes of cartels.

In the second model, we show that any market structure such that it is Pareto-efficient and the cartel chooses the optimal pricing constitute a one-point farsighted stable set. Moreover, this is unique pattern of the farsighted stable sets. It should be emphasized that even if we allow the cartel to choose any positive price, the optimal pricing is obtained on the basis of a stability consideration. The efficiency result obtained in the second model is similar to Kamijo and Muto (2008), in which the cartel's pricing is restricted to optimal one and coalitional or joint move of firms is allowed. If both coalitional moves and price endogeneity are allowed, the result is obvious. As our Theorems B1 and B2 show, only the one-point stable sets are admitted and, thus, the internal stability does not play a part. Because one situation is more likely to be dominated by another if we allow coalitional deviations, the outcomes described in Theorem B1 constitute a unique pattern of stable sets in the coalitional move cases.

	Optimal Pricing	Endogenous Pricing
Individual Move	First Model $K^*(h)$ for any $h \in P \cap D \cap F$	Second Model $\{(x, p)\}$ for any $(x, p) \in B$
Coalitional Move	Kamijo and Muto (2008) $\{x\}$ for any $x \in X^{\mathrm{PE}}$	Second Model' $\{(x, p)\}$ for any $(x, p) \in B$

Table 2: Farishgted stable sets in the various models

References

Bernheim, B. D., Peleg, B., Whiston, M. D., 1987. Coalition-proof nash equilibria I. Concepts. Journal of Economic Theory 42, 1–12.

d'Aspremont, C., Jacquemin, A., Gabszewicz, J. J., Weymark, J. A., 1983. On the stability of collusive price leadership. Canadian Journal of Economics 16, 17–25.

Deneckere, R. J., Kovenock, D., 1992. Price leadership. Review of Economic Studies 59, 143–162.

Diamantoudi, E., 2005. Stable cartels revisited. Economic Theory 26, 907–921.

Donsimoni, M.-P., Economides, N., Polemarchakis, H., 1986. Stable cartels. International Economic Review 27, 317–327.

Fehr, E., Gachter, S., 2000. Cooperation and punishment in public goods experiments. American Economic Review 90, 980–994.

Fudenberg, D., Tirole, J., 1991. Game Theory. The MIT Press.

Furth, D., Kovenock, D., 1993. Price leadership in a duopoly with capacity constraints and product differentiation. Journal of Economics 57, 1–35.

Hamilton, J. H., Slutsky, S. M., 1990. Endogenous timing in duopoly games: Stackelberg or Cournot equilibria. Games and Economic Behavior 2, 29–46.

Harsanyi, J. C., 1974. Interpretation of stable sets and a proposed alternative definition. Management Science 20, 1472–1495.

Harsanyi, J. C., Selten, R., 1988. A general theory of equilibrium selection in games. MIT Press.

Kamijo, Y., Muto, S., 2008. Farsighted coalitional stability of price leadership cartel. G-COE GLOPE II Working Paper Series 6, Waseda University.

Markham, J. W., 1951. The nature and significance of price leadership. American Economic Review 41, 891–905.

Nakanishi, N., 2007. Purely noncooperative farsighted stable set in an n-player prisoners' dilemma. Discussion Paper 0707, Kobe University.

Nakanishi, N., Kamijo, Y., 2008. A constructive proof of existence and a characterization of the farsighted stable set in a price-leadership cartel mode under the optimal pricing. Discussion Paper 0722, Kobe University.

Okada, A., 1993. The possibility of cooperation in an n-person prisoners' dilemma with institutional agreements. Public Choice 77, 629–656.

Ono, Y., 1978. The equilibrium of duopoly in a market of homogeneous goods. Economica 45, 287–295.

Pastine, I., Pastine, T., 2004. Cost of delay and endogeneous price leadership. International Journal of Industrial Organization 22, 135–145.

Prokop, J., 1999. Process of dominant-cartel formation. International Journal of Industrial Organization 17, 241–257.

Suzuki, A., Muto, S., 2005. Farsighted stability in an n-person prisoner's dilemma. International Journal of Game Theory 33, 441–445.

Tasnadi, A., 2000. A price-setting game with a nonatomic fringe. Economics Letters 69, 63–69.

Thoron, S., 1998. Formation of a coalition-proof stable cartel. Canadian Journal of Economics 31, 63–76.

van Damme, E., Hurkens, S., 2004. Endogenous price leadership. Games and Economic Behavior 47, 404–420.

von Neumann, J., Morgenstern, O., 1953. Theory of games and economic behavior. Princeton University Press, third edn.

Appendix A

Proof of Lemma A2. Note that both $C(y) \setminus C(x)$ and $C(x) \setminus C(y)$ contain the same number of firms. Let $\{i_1, i_2, \ldots, i_M\} = C(y) \setminus C(x)$ and $\{j_1, j_2, \ldots, j_M\} = C(x) \setminus C(y)$. Consider the following sequence:

- first, firm i_1 exits from $C(y)$ and changes y to y^1 such that $C(y^1) = C(y) \setminus \{i_1\}$,
- firm j_1 enters the existing cartel y^1 and forms a new cartel y^2 such that $C(y^2) = C(y^1) \cup \{j_1\}$,
- in general, firm i_m exits from y^{2m-2} and changes it to y^{2m-1} such that $C(y^{2m-1}) = C(y^{2m-2}) \setminus \{i_m\}$,
- firm j_m enters y^{2m-1} and form a new cartel y^{2m} such that $C(y^{2m}) = C(y^{2m-1}) \cup \{j_m\}$.

By repeating this replacement of players (i_m and j_m) M times, we obtain $y^{2M} = x$ from $y^0 = y$. When $i_m \in C(y) \setminus C(x)$ exits from the existing cartel, it is a member of size h cartel; in addition, it ends up with being a fringe firm at the last of the sequence. Therefore, we have $f_{i_m}(y^{2m-2}) = \pi_c^*(h) < \pi_f^*(h) = f_{i_m}(x)$. When $j_m \in C(x) \setminus C(y)$ moves, the size of the existing cartel is $h-1$; and j_m ends up with being a member of $C(x)$. Therefore, by the definition of attractor, $f_{j_m}(y^{2m-1}) = \pi_f^*(h-1) < \pi_c^*(h) = f_{j_m}(x)$. Hence, $x \rhd y$. $\square$

Proof of Lemma A3. [Sufficiency]: Suppose $\pi_f^*(h) > \pi_c^*(h')$. Let $M = |C(x) \setminus C(y)|$ and write $C(x) \setminus C(y) \equiv \{j_1, j_2, \ldots, j_M\}$. Note that $M = 0$ is possible. Let $C(y) \setminus C(x) = \{i_1, i_2, \ldots, i_M, i_{M+1}, \ldots, i_{M+h'-h}\}$. Because $|y| = h' > h = |x|$, we can always write in this way. Consider the following sequence:

- firm i_{M+1} exits from y; then, i_{M+2} exits from the resulting cartel after i_{M+1}'s exit; in this way, i_{M+k} exits from the resulting cartel after i_{M+k-1}'s exit;
- let y' be the resulting cartel of size h after $i_{M+h'-h}$'s exit (note that $|x| = h$);
- if $y' = x$ (i.e., if $M = 0$), then the sequence ends;
- if $y' \neq x$, add a sequence from y' to x analogous to the one in the proof of Lemma A2, in which firm i_k's exit from the existing cartel is followed by j_k's entry ($k = 1, \ldots, M$).

In the former part of the sequence from y to y', the size of the cartel from which firm i_{M+k} is just going to exit is $h' - k + 1$. Then, firm i_{M+k}'s payoff as a member of a cartel of size $h' - k + 1$ is $\pi_c^*(h' - k + 1)$. By the assumption of the lemma and by the monotonicity of π_c^*, we have $f_{i_{M+k}}(x) = \pi_f^*(h) > \pi_c^*(h') \geqq \pi_c^*(h' - k + 1)$ for all $k = 1, \ldots, h' - h$. In the latter part of the sequence from y' to x, similar to the proof of Lemma A2, all the related players (i.e., i_m and j_m for $m = 1, \ldots, M$) can be made better-off eventually. Hence, $x \rhd y$.

[Necessity]: Suppose $x \rhd y$ with $|x| = h < h' = |y|$. Consider a sequence of cartels and corresponding firms that realizes $x \rhd y$. There must be at least one firm who exits from a cartel of size h' in the sequence; let firm i be such a firm in the sequence and let y' be the size h' cartel from which firm i is just going to exit. By the definition of the O-domination, we must have $\pi_c^*(h') = f_i(y') < f_i(x)$. By the size monotonicity of π^*, $\pi_c^*(h') > \pi_c^*(h) = \pi_c^*(|x|)$. This implies firm i belongs to the fringe of cartel x, and thus, $\pi_c^*(h') = f_i(y) < f_i(x) = \pi_f^*(h)$. $\square$

Proof of Lemma A4. Suppose, in negation, that h is not an attractor. Then, by definition, $\pi_f^*(h-1) \geqq \pi_c^*(h)$. Consider a sequence of cartels and corresponding firms realizing $x \rhd y$ such that

$$y = y^0 \xrightarrow{i_1} y^1 \xrightarrow{i_2} \ldots \xrightarrow{i_{M-1}} y^{M-1} \xrightarrow{i_M} y^M = x.$$

We distinguish two cases: case 1 where $h = h'$ and case 2 where $h > h'$.

Case 1. Because $\pi_f^*(h-1) \geqq \pi_c^*(h)$, the last firm in the sequence (i.e., i_M) cannot be an entering firm when it moves. Firm i_M must realize x *by exiting* from some size $h+1$ cartel. Then, in the course of the sequence, the cartel size must exceed h. Accordingly, there must exist at least one firm, say firm i_k, who enter a size h cartel and form a new size $h+1$ cartel. Let y' be the size h cartel that firm i_k is just going to enter. Then, $f_{i_k}(y') = \pi_f^*(h) \geqq [\pi_f^*(h)$ or $\pi_c^*(h)] = f_{i_k}(x)$. This contradicts the definition of the O-domination.

Case 2. If there exists a cartel $y^m \neq x$ in the sequence such that $|y^m| = h$, then, similar to case 1, we immediately obtain a contradiction. Then, we can assume that $|y^m| < h$ for all $m = 0, 1, \ldots, M-1$. The last firm in the sequence, i_M, must realize $x = y^M$ *by entering* y^{M-1}. However, we have $f_{i_M}(y^{M-1}) = \pi_f^*(h-1) \geqq \pi_c^*(h) = f_{i_M}(x)$. This, again, contradicts the definition of the O-domination. $\square$

Proof of Lemma A5. If $|y| = |x|$, then Lemma A2 applies. Suppose $|y| < |x|$. By the definition of Z^L, any h with $|y| < h \leqq |x|$ is a leading attractor; $|y|$ itself is either a leading attractor or zero. Consider an *increasing* sequence of cartels and corresponding firms

$$y = y^0 \xrightarrow{i_1} y^1 \xrightarrow{i_2} \ldots \xrightarrow{i_M} y^M$$

such that for $m = 1, \ldots, M$, (i) $i_m \in C(x) \setminus C(y)$, (ii) $|y^m| = |y^{m-1}| + 1$, and (iii) $|y^M| = |x|$. Because $0 < |x| - |y| \leqq |C(x) \setminus C(y)|$, we can always choose, from $C(x) \setminus C(y)$, an appropriate set of $M \equiv |x| - |y|$ firms appeared in the sequence. Note that $|y^m|$ is an attractor for all $m = 1, \ldots, M$.

If $y^M = x$, then, by the definition of attractor and by the monotonicity of π_c^*, we have $f_{i_m}(y^{m-1}) = \pi_f^*(|y^{m-1}|) < \pi_c^*(|y^{m-1}| + 1) = \pi_c^*(|y^m|) \leqq \pi_c^*(|x|) = f_{i_m}(x)$ for all $m = 1, \ldots, M$. If $y^M \neq x$, then we only have to add an additional sequence from y^M to x analogous to the one in the proof of Lemma A2. Hence, $x \rhd y$. $\square$

Proof of Lemma A6. Suppose, in negation, that $K \cap V(Z^A) = \emptyset$. Because $Z^L \subseteq Z^A$, we have $K \cap V(Z^L) = \emptyset$. Take an arbitrary cartel x with $|x| \in Z^L$. x is not in K. By the external stability of K, there must exist a cartel $y \in K$ that O-dominates x. If $|y| \leqq |x|$, then $|y|$ is either an attractor or zero. By assumption, $|y|$ cannot be an attractor. However, if $|y| = 0$, y cannot O-dominate x. Therefore, we must have $|y| > |x|$. By Lemma A4, $|y|$ must be an attractor—a contradiction.

Proof of Lemma A7. The proof of this lemma is immediate consequence of the following three claims:

Claim (a)[Anonymity] For $x \in V(k)$ and $y \in V(h)$, if $x \rhd y$, then $z \rhd w$ for any $z \in V(k)$ and $w \in V(h)$ with $|C(x) \cap C(y)| = |C(z) \cap C(w)|$.

Claim (b) Take $k, h \in Z$ with $k > h$ and $\pi_c^*(k) > \pi_f^*(h)$. If $x \in V(k)$ O-dominates some $y \in V(h)$, then this x O-dominates any $z \in V(h)$.

Claim (c) Take $k, h \in Z$ with $k > h$ and $\pi_c^*(k) > \pi_f^*(h)$. If $x \in V(k)$ O-dominates some $y \in V(h)$, then this x O-dominates any $z \in V(h')$, where $h \leqq h' \leqq k$.

Proof of Claim (a). Before proving this claim, we prepare some notations. Given a permutation θ on N and cartel x, let $\theta(x)$ denote a permuted cartel such that $C(\theta(x)) = \theta(C(x))$.

Consider a permutation θ on N such that $\theta(x) = z$, $\theta(y) = w$, and $\theta(x \wedge z) = y \wedge w$. Such θ does exist becase of the assumotions on x, y, z and w. Since x O-dominates z, there exists a sequence of cartels and firms

$$y = y^0 \xrightarrow{i_1} y^1 \xrightarrow{i_2} \cdots \xrightarrow{i_M} y^M = x$$

that realize the domination. For this sequence, let define a sequence of permuted cartels and permuted firms from w to z as follows:

$$w = \theta(y) = \theta(y^0) \xrightarrow{\theta(i_1)} \theta(y^1) \xrightarrow{\theta(i_2)} \cdots \xrightarrow{\theta(i_M)} \theta(y^M) = \theta(x) = z.$$

By the symmetry of the firms, the above sequence realizes the O-domination from w to z.

Proof of Claim (b). If $h = k - 1$, this is obvious from the proof of Lemma A2. Therefore, we assume that $h < k - 1$.

Let $s = |C(x) \cap C(y)|$ and $t = |C(x) \cap C(z)|$. We separate three cases: (i) $s = t$, (ii) $s > t$, and (iii) $s < t$.
(i). By Claim (a), $x \rhd z$.
(ii). $C(x) \setminus C(z) = \{i_1, \ldots, i_{k-t}\}$ and $C(z) \setminus C(x) = \{j_1, \ldots, j_{h-t}\}$. Note that $k - t > h - t \geqq s - t$.

Consider the following sequence from z to w, where $|w| = h$:

- first, firm i_1 joins the cartel z and changes z to z^1 such that $C(z^1) = C(z) \cup \{i_1\}$,
- firm j_1 exits from the existing cartel z^1 and forms a new cartel z^2 such that $C(z^2) = C(z^1) \setminus \{j_1\}$,
- in general, firm i_m enters z^{2m-2} and changes it to z^{2m-1} such that $C(z^{2m-1}) = C(z^{2m-2}) \cup \{i_m\}$,
- firm j_m exits from z^{2m-1} and form a new cartel z^{2m} such that $C(z^{2m}) = C(z^{2m-1}) \setminus \{j_m\}$.

By repeating this replacement of players (i_m and j_m) $s - t$ times, we obtain $w \equiv z^{2(s-t)}$ from $z^0 = z$.

By construction, $|w| = h$ and $|C(x) \cap C(w)| = s$. Therefore, by Claim (a), $x \rhd w$. Now we show that x O-dominates z by connecting the sequence from z to w described above to the one realizing $w \lhd x$. To show this, it is enough to show that in each step in the sequence from z to w, each firm prefers x to the current situation. For i_q, $1 \leq q \leq s-t$, since $i_q \notin C(z^{2q-2})$ and $|z^{2q-2}| = h$,

$$f_{i_q}(z^{2q-2}) = \pi_f^*(h) < \pi_c^*(k) = f_{i_q}(x).$$

For j_q, For i_q, $1 \leq q \leq s - t$, since $i_q \in C(z^{2q-1})$ and $|z^{2q-1}| = h + 1$,

$$f_{j_q}(z^{2q-1}) = \pi_c^*(h + 1) < \pi_c^*(k) < \pi_f^*(k) = f_{j_q}(x).$$

Hence, $x \rhd z$.

(iii). $C(x) \cap C(z) = \{i_1, \ldots, i_t\}$ and $N \setminus (C(z) \cup C(x)) = \{j_1, \ldots, j_r\}$, where

$$r = |N \setminus (C(z) \cup C(x))| \geqq |C(y) \setminus (C(z) \cup C(x))|$$
$$= |(C(y) \setminus C(x)) \setminus (C(z) \setminus C(x))| \geqq (h-s) - (h-t) = t-s.$$

Consider the following sequence from z to w, where $|w| = h$:

- first, firm i_1 exits from the cartel z and changes z to z^1 such that $C(z^1) = C(z) \setminus \{i_1\}$,
- firm j_1 enters the existing cartel z^1 and forms a new cartel z^2 such that $C(z^2) = C(z^1) \cup \{j_1\}$,
- in general, firm i_m exits from z^{2m-2} and changes it to z^{2m-1} such that $C(z^{2m-1}) = C(z^{2m-2}) \setminus \{i_m\}$,
- firm j_m enters z^{2m-1} and form a new cartel z^{2m} such that $C(z^{2m}) = C(z^{2m-1}) \cup \{j_m\}$.

By repeating this replacement of players (i_m and j_m) $t-s$ times, we obtain $w \equiv z^{2(t-s)}$ from $z^0 = z$. Note that by construction, $|w| = h$ and $|C(x) \cap C(w)| = t$. Thus, by Claim (a), $x \rhd w$. Similar to case (ii), we can show that x O-dominates z by connecting the sequence from z to w described above to the one realizing $w \lhd x$.

Proof of Claim (c). Since k is an attractor by Lemma A4, the case where $h' = k$ holds by Lemma A2. When $h' = h$, by Claim (b), this case holds. So we consider the case where $h < h' < k$.

Take any $z \in V(h')$. Let $C(z) = \{i_1, \ldots, i_{h'}\}$. Consider the following sequence of deviation such that firms in $C(z)$ exit from the cartel until the size of cartel becomes h:

$$z = z^0 \xrightarrow{i_1} z^1 \xrightarrow{i_2} \ldots \xrightarrow{i_{h'-h}} z^{h'-h} = w.$$

Because $|w| = h$, by Claim (b), x O-dominates w. Then, we now show that $x \rhd z$ by the sequence connecting the one from z to w described above to the one realizing $w \lhd x$. Since for each i_r, $1 \leq r \leq h' - h$,

$$f_{i_r}(z^{r-1}) = \pi_c^*(h' - r + 1) < \pi_c^*(k) \leqq f_{i_r}(x),$$

the incentives of deviating firms in the above sequence holds and thus $x \rhd z$. □

Proof of Lemma A8. If $a = \bar{a}$, the proof ends. Then, take an arbitrary $a \in Z^L$ with $a < \bar{a}$. Because any h with $a \leqq h \leqq \bar{a}$ is an attractor, we have $\alpha_1(a) = a$, $\alpha_2(a) = a+1$, $\alpha_3(a) = a+2, \ldots, \alpha_M(a) = \bar{a}-1$, $\alpha_{M+1}(a) = \bar{a}$, where $M = \bar{a} - a$. Therefore, we can write $H(a)$ as follows:

$$H(a) \equiv \{\alpha_1(a), \ldots, \alpha_M(a)\} \cup H(\bar{a}).$$

Suppose that the largest leading attractor $\bar{a}$ is not deleted from $H(\bar{a})$ in the procedure generating $H^*(\bar{a})$; that is $\bar{a} \in H^*(\bar{a})$. Then, by the definition of the procedure, $\bar{a}$ will never be deleted in the procedure generating $H^*(a)$ from $H(a)$; that is, $\bar{a} \in H^*(a)$. Because, by Lemma A5, an arbitrary $x \in V(\bar{a})$ O-dominates any distinct y with $|y| \leqq |x|$, then all of $\alpha_1(a), \ldots, \alpha_M(a)$ are deleted in the procedure generating $H^*(a)$. In this case, $H^*(a) = H^*(\bar{a})$.

On the other hand, suppose $\bar{a}$ is deleted from $H(\bar{a})$ in the procedure generating $H^*(\bar{a})$. Then, there exists an integer $h \in H^*(\bar{a})$ and a cartel $x \in V(h)$ that O-dominates a cartel $y \in V(\bar{a})$. Take a cartel $y' \in V(a)$. Without loss of generality, we can assume $C(y') \subset C(y)$. Let $C(y) \setminus C(y') = \{i_1, i_2, \ldots, i_R\}$. Further, let $\{i_{R+1}, i_{R+2}, \ldots, i_{R+S}\}$ be the set of firms that appear in the sequence realizing $x \rhd y$. Consider the following sequence of cartels and corresponding firms:

$$y' = y^0 \xrightarrow{i_1} y^1 \xrightarrow{i_2} \ldots \xrightarrow{i_R} \underbrace{[y^R = y] \xrightarrow{i_{R+1}} \ldots \xrightarrow{i_{R+S}} [y^{R+S} = x]}_{\text{the sequence realizing } x \,\rhd\, y}.$$

For $r = 1, \ldots, R$, we have $f_{i_r}(y^{r-1}) = \pi_f^*(|y^{r-1}|)$ and $f_{i_r}(x) = \pi_c^*(h)$ or $f_{i_r}(x) = \pi_f^*(h)$. Note that $a \leqq |y^{r-1}| < \bar{a}$ for all $r = 1, \ldots, R$. By the definitions of $H^*(\bar{a})$ and Z^L and the properties of π^* functions, we have $\pi_c^*(a) < \pi_f^*(a) < \pi_c^*(a+1) < \pi_f^*(a+1) < \cdots < \pi_c^*(\bar{a}) < \pi_f^*(\bar{a})$ and $\pi_f^*(\bar{a}) \leqq \pi_c^*(h) < \pi_f^*(h)$. In any case, $f_{i_r}(y^{r-1}) < f_{i_r}(x)$ for all $r = 1, \ldots, R$. By the definition of O-domination, we have $f_{i_{R+s}}(y^{R+s-1}) < f_{i_{R+s}}(x)$ for all $s = 1, \ldots, S$. That is, $x \rhd y'$. Again, all of $\alpha_1(a), \ldots, \alpha_M(a)$ are deleted in the procedure generating $H^*(a)$. Hence, $H^*(a) = H^*(\bar{a})$. $\square$

Proof of Lemma A9. Suppose, in negation, that $d^* \notin H^*(d^*)$. Then, there exists $h \in H^*(d^*)$ such that $V(h) \rhd V(d^*)$ and $h > d^*$. By construction, $\pi_c^*(h) > \pi_f^*(d^*)$. Take cartels x, y, and y' such that $x \in V(h)$, $y \in V(d^*)$, and $y' \in V(\bar{a})$; we can assume $x \rhd y$ and $y \rhd y'$.

Because x Pareto-dominates y by construction, we can construct an appropriate sequence realizing $x \rhd y'$ by simply connecting the sequence realizing $x \rhd y$ to the sequence realizing $y \rhd y'$. That is, $V(h) \rhd V(\bar{a})$; this contradicts the definition of d^*. $\square$

Proof of Lemma A10. If $\bar{a} \in F$, then the proof ends. Suppose $\bar{a} \notin F$. Let m^* be the minimum element in $H^*(\bar{a})$. (Note that m^* exists.) By the definition of $H^*(\bar{a})$, $V(m^*) \rhd V(\bar{a})$ because if other integer $m' \in H^*(\bar{a})$ O-dominates $\bar{a}$, it also O-dominates m^*—a contradiction. Then, $m^* \in D$. By construction and the properties of π functions, we have $\pi_c^*(\bar{a}) < \pi_f^*(\bar{a}) \leqq \pi_c^*(m^*) < \pi_f^*(m^*)$; this implies $m^* \in P$. Further, by construction, $m^* \in H^*(m^*)$. Hence, $m^* \in P \cap D \cap F$. $\square$

Appendix B

Proof of Lemma B1. We first show $A^* \supseteq A^{\mathrm{OP}} \cap A^{\mathrm{PE}}$. Because A^* is a subset of A^{OP}, it suffices to show that any outcome $(x, p^*(|x|)) \in A^{\mathrm{OP}}$ with $\pi_f^*(|x|) \leqq \pi_c^*(n)$ is Pareto-dominated by another outcome. Fortunately, it is obvious that such an outcome $(x, p^*(|x|))$ is Pareto-dominated by $(x^c, p^*(n))$.

Next, we show $A^* \subseteq A^{\mathrm{OP}} \cap A^{\mathrm{PE}}$. Because A^* is a subset of A^{OP}, we will show that A^* is a subset of A^{PE}. Take an arbitrary $(x, p) \in A^*$. We have to show that (x, p) cannot be Pareto-dominated. We distinguish two cases: case 1 where $(x, p) = (x^c, p^*(|x^c|))$ and case 2 where $(x, p) \neq (x^c, p^*(|x^c|))$.

Let us consider case 1. Take an arbitrary $(y, w) \in A$ other than (x, p). If $C(y) = \emptyset$ or, equivalently, $y = x^f$, then we have $g_i(x, p) = \pi_c^*(n) > \pi_f^*(0) = \pi_f(p^{\mathrm{comp}}) = g_i(y, w)$ for all $i \in N$ by Proposition 2-(iii). On the other hand, if $C(y) \neq \emptyset$, we have $g_i(x, p) = \pi_c^*(n) > \pi_c^*(|y|) = \pi_c(|y|, p^*(|y|)) \geqq \pi_c(|y|, w) = g_i(y, w)$ for all $i \in C(y)$ by the size-monotonicity of π_c^* and the definition of p^*. That is, (y, w) cannot Pareto-dominate $(x^c, p^*(|x^c|))$.

Next, let us consider case 2. By the inequality $\pi_f^*(|x|) > \pi_c^*(n)$, neither $|x| = 0$ nor $|x| = n$ can be true. Therefore, we have both $C(x) \neq \emptyset$ and $N \setminus C(x) \neq \emptyset$. Suppose, in negation, that there exists an outcome $(y, w) \in A$ that Pareto-dominates (x, p).

If there is a player i such that $i \in N \setminus C(x)$ and $i \in C(y)$, then, by the definition of the Pareto-domination, we have

$$\pi_c(|y|, w) = g_i(y, w) \geqq g_i(x, p) = \pi_f^*(|x|).$$

On the other hand, by the definitions of π_c^* and A^*, we have

$$\pi_c^*(|y|) \geqq \pi_c(|y|, w) \quad \text{and} \quad \pi_f^*(|x|) > \pi_c^*(n).$$

Combining the above inequalities, we obtain $\pi_c^*(|y|) > \pi_c^*(n)$. This contradicts the size-monotonicity of π_c^*. Such player i cannot exist. Hence, $i \in N \setminus C(x)$ implies $i \in N \setminus C(y)$; equivalently, $C(y) \subset C(x)$.

In turn, if there is a player j such that $j \in C(x)$ and $j \in C(y)$, then, similar to the above paragraph, we obtain the following inequalities:

$$\pi_c^*(|y|) \geqq \pi_c(|y|, w) = g_j(y, w) \geqq g_j(x, p) = \pi_c^*(|x|).$$

By the size-monotonicity of π_c^*, the fact $\pi_c^*(|y|) \geqq \pi_c^*(|x|)$ implies $|y| \geqq |x|$. This, together with $C(y) \subset C(x)$, implies $C(y) = C(x)$ or, equivalently, $x = y$. Then, by the definition of p^*, we obtain

$$g_j(y, w) = g_j(x, w) = \pi_c(|x|, w) < \pi_c(|x|, p^*(|x|)) = \pi_c^*(|x|) = g_j(x, p).$$

This contradicts the definition of the Pareto-domination. Such player j cannot exist. Hence, $j \in C(x)$ implies $j \in N \setminus C(y)$; equivalently, $C(x) \subset N \setminus C(y)$. Therefore, we have $C(y) \subset C(x) \subset N \setminus C(y)$. This can be possible only if $C(y) = \emptyset$, but, as already shown, $C(y) = C(x) \neq \emptyset$—a contradiction. No outcome can Pareto-dominate $(x, p) \in A^*$. $\square$

Proof of Lemma B2. Take an arbitrary outcome (y, w) other than $(x^c, p^*(|x^c|))$. We distinguish three cases: case 1 where $y = x^c$, case 2 where $y = x^f$, and case 3 where $y \neq x^c$ and $y \neq x^f$.

First, let us consider case 1: $y = x^c$. Clearly, the cartel $C(y) = C(x^c)$ can change the current price w to the optimal price $p^*(|y|)$, that is, $(y, w) \xrightarrow{C(y)} (y, p^*(|y|)) = (x^c, p^*(|x^c|))$. Further, by the definition of p^*, we have

$$g_i(y, w) = \pi_c(|y|, w) = \pi_c(|x^c|, w) < \pi_c(|x^c|, p^*(|x^c|)) = g_i(x^c, p^*(|x^c|))$$

for all $i \in C(y)$. The desired result is obtained.

Next, let us consider case 2: $y = x^f$. Consider a sequence of deviations in which (a) each player enters the cartel one by one and (b) after all the players enter the cartel, the largest cartel changes the price to $p^*(|x^c|)$:

$$\begin{aligned}(y, w) = (x^0, p^{\text{comp}}) &\xrightarrow{\{i_1\}} (x^1, p^{\text{comp}}) \xrightarrow{\{i_2\}} \cdots \\ &\cdots \xrightarrow{\{i_n\}} [(x^n, p^{\text{comp}}) = (x^c, p^{\text{comp}})] \xrightarrow{C(x^c)} (x^c, p^*(|x^c|)).\end{aligned}$$

For each i_k in the above sequence, we have $g_{i_k}(x^{k-1}, p^{\text{comp}}) = \pi_f(p^{\text{comp}}) = \pi_f^*(0) < \pi_c^*(n) = \pi_c^*(|x^c|) = g_{i_k}(x^c, p^*(|x^c|))$ by Proposition 2-(iii). Further, in the last step, we have $g_i(x^n, p^{\text{comp}}) = g_i(x^c, p^{\text{comp}}) < \pi_c(|x^c|, p^*(|x^c|)) = \pi_c^*(|x^c|) = g_i(x^c, p^{\text{comp}})$ for all $i \in C(x^c)$. Again, the desired result is obtained.

Lastly, let us consider case 3: $y \neq x^c$ and $y \neq x^f$. It immediately follows that $0 < |y| < n$. Let $\hat{p} > 0$ be a price such that $\pi_f(\hat{p}) < \pi_c^*(n)$ and $\hat{p} \neq p^{comp}$. Such $\hat{p}$ exists because $\pi_f(p)$ is decreasing and $\lim_{p \to 0} \pi_f(p) = 0$.

Now, consider a sequence of deviations in which (a) cartel $C(y)$ decreases the price down to $\hat{p}$, (b) each firm in $N \setminus C(y)$ enters the cartel one by one until all the firms enter the cartel, and (c) after establishing the largest cartel, the cartel $C(x^c)$ changes the price to $p^*(|x^c|)$:

$$\begin{aligned}(y, w) &\xrightarrow{C(y)} \left[(y, \hat{p}) = (x^0, \hat{p})\right] \xrightarrow{\{j_1\}} (x^1, \hat{p}) \xrightarrow{\{j_2\}} \cdots \\ &\cdots \xrightarrow{\{j_r\}} [(x^r, \hat{p}) = (x^c, \hat{p})] \xrightarrow{C(x^c)} (x^c, p^*(|x^c|)),\end{aligned}$$

where $N \setminus C(y) \equiv \{j_1, j_2, \ldots, j_r\}$. In the first (price-cutting) step, we have

$$g_i(y, w) = \pi_c(|y|, w) \leqq \pi_c^*(|y|) < \pi_c^*(n) = \pi_c^*(|x^c|) = g_i(x^c, p^*(|x^c|))$$

for all $i \in C(y)$ by the size-monotonicity of π_c^*. In each of the intermediate (entry) steps, we have

$$g_{j_k}(x^{k-1}, \hat{p}) = \pi_f(\hat{p}) < \pi_c^*(|x^c|) = g_{j_k}(x^c, p^*(|x^c|))$$

for j_k ($k = 1, 2, \ldots, r$). In the last (price-increasing) step, we have

$$g_i(x^c, \hat{p}) = \pi_c(|x^c|, \hat{p}) < \pi_f(\hat{p}) < \pi_c(|x^c|, p^*(|x^c|)) = g_i(x^c, p^*(|x^c|))$$

for all $i \in C(x^c)$ by Proposition 1-(iii) and the definition of $\hat{p}$. Hence, the desired result is obtained. □

Proof of Lemma B3. We first prove case (ii) and, then, turn to case (i) and case (iii). Let $\hat{p} \in \mathbb{R}_{++}$, $\hat{p} \neq p^{comp}$, be a price level that satisfies $\hat{p}$

$$\pi_f(\hat{p}) < \pi_c(|x|, p).$$

Such a price level $\hat{p}$ exists since $\pi_c(|x|, p) > 0$ and $\lim_{p \to 0} \pi_f(p) = 0$.
Case (ii). Consider the following steps that form an appropriate sequence of deviations from (y, w) to (x, p):

Step 1 If $C(y) \setminus C(x) = \emptyset$, then go to the next step. Otherwise, consider a sequence in which each firm in $C(y) \setminus C(x)$ exits from the cartel in turn. Let $C(y) \setminus C(x) = \{i_1, i_2, \ldots, i_r\}$. Then,

$$(y, w) = (x^0, w) \xrightarrow{\{i_1\}} (x^1, w) \xrightarrow{\{i_2\}} (x^2, w) \xrightarrow{\{i_3\}} \ldots \xrightarrow{\{i_r\}} (x^r, w),$$

where $x^k \in X$ is defined to satisfy $C(x^k) = C(y) \setminus \{i_1, \ldots, i_k\}$ for each $k = 1, \ldots, r$. Note that $x^r = x \wedge y$.

Step 2 Cartel $C(x \wedge y)$ changes the price from w to $\hat{p}$. (Note that $C(x \wedge y) \neq \emptyset$ by assumption.) Thus,

$$(x \wedge y, w) \xrightarrow{C(x \wedge y)} (x \wedge y, \hat{p}).$$

Step 3 If $C(x) \setminus C(y) = \emptyset$, then go to the next step. Otherwise, consider a sequence in which each firm in $C(x) \setminus C(y)$ enters the cartel in turn. Let $C(x) \setminus C(y) = \{j_1, j_2, \ldots, j_{r'}\}$. Then,

$$(x^r, \hat{p}) \xrightarrow{\{j_1\}} (x^{r+1}, \hat{p}) \xrightarrow{\{j_2\}} (x^{r+2}, \hat{p}) \xrightarrow{\{j_3\}} \ldots \xrightarrow{\{j_{r'}\}} (x^{r+r'}, \hat{p}),$$

where $x^{r+k} \in X$ is defined to satisfy $C(x^{r+k}) = C(x \wedge y) \cup \{j_1, \ldots, j_k\}$ for each $k = 1, \ldots, r'$. Note that $x^{r+r'} = x$.

Step 4 Cartel $C(x)$ changes the price from $\hat{p}$ to p.

$$(x, \hat{p}) \xrightarrow{C(x)} (x, p).$$

Now we check firms' incentive of deviation. For each i_k in Step 1, we have

$$g_{i_k}(x^{k-1}, w) = \pi_c(|x^{k-1}|, w) \leqq \pi_c(|y|, w) < \pi_f(p) = g_{i_k}(x, p),$$

where the second inequality follows from the size-monotonicity of π_c ((i) in Proposition 1) and the penultimate strict inequality is due to the condition given in this lemma. Thus, all the deviating firms in Step 1 have incentives to deviate toward the ultimate outcome (x, p).

In Step 2, we have

$$g_i(x \wedge y, w) = \pi_c(|x \wedge y|, w) < \pi_c(|x|, p) = g_i(x, p)$$

for all $i \in C(x \wedge y)$. (Note that $C(x \wedge y) \subset C(x)$.) The above inequality follows from the condition given in the lemma. Therefore, cartel $C(x \wedge y)$ has an incentive to change the price as in Step 2.

Moreover, for each deviating firm j_k in Step 3, we have

$$g_{j_k}(x^{r+k-1}, \hat{p}) = \pi_f(\hat{p}) < \pi_c(|x|, p) = g_{j_k}(x, p)$$

by the definition of $\hat{p}$. Thus, j_k is better off in (x, p) than in $(x^{r+k-1}, \hat{p})$.

For $\hat{p}$, we have $\pi_c(|x|, \hat{p}) < \pi_f(\hat{p}) < \pi_c(|x|, p)$ by the definition of $\hat{p}$ and Proposition 1-(iii). Then, in Step 4, we have

$$g_i(x, \hat{p}) = \pi_c(|x|, \hat{p}) < \pi_c(|x|, p) = g_i(x, p)$$

for all $i \in C(x)$. Cartel $C(x)$ has an incentive to change their price to p. Hence, $(x, p) \gg (y, w)$ holds through this sequence of deviations.

Case (i). Take an arbitrary $i \in C(y)$. Consider the following finite sequence of deviations:

Step 1 Cartel $C(y)$ changes its price from p to $\hat{p}$.

Step 2 Firms in $C(y) \setminus \{i\}$ exit from the cartel in turn.

Step 3 Firms in $C(x)$ enter the cartel in turn.

Step 4 Firm i exits from the cartel.

Step 5 Cartel $C(x)$ changes its price from $\hat{p}$ to p.

Applying almost the same argument as the proof of "Case (ii)," we can show the incentives of the deviating firms in each step.
Case (iii). Let $z = x \wedge y$; then $C(z) = C(x) \cap C(y)$. By the conditions given in the lemma, we have both $C(x) \setminus C(z) \neq \emptyset$ and $C(y) \setminus C(z) \neq \emptyset$. Consider the following sequence of deviations:

Step 1 Firms in $C(y) \setminus C(z)$ exit from $C(y)$ one by one until the cartel $C(z)$ is realized:

$$(y, w) = (x^0, w) \xrightarrow{\{i_1\}} (x^1, w) \xrightarrow{\{i_2\}} \cdots \xrightarrow{\{i_r\}} (x^r, w) = (z, w),$$

where $C(y) \setminus C(z) = \{i_1, \ldots, i_r\}$ and $r \equiv |y| - |z|$.

Step 2 $C(z)$ decreases the price down to $\hat{p}$: $(z, w) \xrightarrow{C(z)} (z, \hat{p})$.

Step 3 Firms in $C(x) \setminus C(z)$ enter the cartel until $C(x)$ is established:

$$\begin{aligned}(z, \hat{p}) = (x^{r+1}, \hat{p}) \xrightarrow{\{j_1\}} (x^{r+2}, \hat{p}) \xrightarrow{\{j_2\}} \cdots \\ \cdots \xrightarrow{\{j_{r'}\}} (x^{r+1+r'}, \hat{p}) = (x, \hat{p}),\end{aligned}$$

where $C(x) \setminus C(z) = \{j_1, \ldots, j_{r'}\}$ and $r' = |x| - |z|$.

Step 4 $C(x)$ increases the price up to p: $(x, \hat{p}) \xrightarrow{C(x)} (x, p)$.

In Step 1, we have

$$\begin{aligned} g_{i_k}(x^{k-1}, w) = \pi_c(|y| - k + 1, w) &< \pi_c(|y|, w) \\ &\leqq \pi_c(|x|, p) \\ &< \pi_f(p) = g_{i_k}(x, p) \end{aligned}$$

for all $k = 1, \ldots, r$, where the first inequality follows from the size-monotonicity of π_c, the second from the condition given in the lemma, and the third from Proposition 1-(iii). In Step 2, we have

$$g_i(z, w) = \pi_c(|z|, w) < \pi_c(|y|, w) \leqq \pi_c(|x|, p) = g_i(x, p)$$

for all $i \in C(z)$. Thus, in each case, $g_i(x, p) > g_i(z, w)$ for all $i \in C(z)$. In Step 3, we have

$$g_{j_k}(x^{r+k}, \hat{p}) = \pi_f(\hat{p}) < \pi_c(|x|, p) = g_{j_k}(x, p)$$

for all $k = 1, \ldots, r'$. And, in Step 4, we have

$$g_i(x, \hat{p}) = \pi_c(|x|, \hat{p}) < \pi_f(\hat{p}) < \pi_c(|x|, p) = g_i(x, p)$$

for all $i \in C(x)$. Hence, (x, p) E-dominates (y, w). $\square$

Proof of Lemma B4. We distinguish two cases: case 1 where $(x^f, p^{\text{comp}}) \in K$ and case 2 where $(x^f, p^{\text{comp}}) \notin K$. Note that $g_i(x^f, p^{\text{comp}}) = \pi_f(p^{\text{comp}}) = \pi_f^*(0) > 0$ for all $i \in N$.

First we consider case 1 where $(x^f, p^{\text{comp}}) \in K$. Take any $(y, w) \in K$ such that $\pi_c(|y|, w) \leqq 0$. Consider the following finite sequence of deviations from (y, w) to (x^f, p^{comp}): Firms in $C(y)$ exit from the cartel in turn. Thus,

$$(y, w) = (y^0, w) \xrightarrow{\{i_1\}} (y^1, w) \xrightarrow{\{i_2\}} (y^2, w) \ldots \xrightarrow{\{i_r\}} (y^r, p^{comp}) = (x^f, p^{comp})$$

where $C(y) = \{i_1, \ldots, i_r\}$, $r = |y|$, and y^k is such that $C(y^k) = \{i_{k+1}, \ldots, i_r\}$.

For each deviant firm i_m, $\pi_c(|y^{m-1}|, w) \leqq \pi_c(|y|, w) < \pi_f(0)$. Therefore, (x^f, p^{comp}) E-dominates (y, w) through the above sequence of deviations and this contradicts the internal stability of K.

Next we consider case 2 where $(x^f, p^{\text{comp}}) \notin K$. In this case, there must exist $(x, p) \in K$ such that $(x, p) \gg (x^f, p^{\text{comp}})$ to assure the external stability of K. Then, $x \neq x^f$ and $p \neq p^{comp}$. We show that in the outcome (x, p), the firms in the cartel obtain a positive profit. In a dominance sequence that realizes $(x, p) \gg (x^f, p^{\text{comp}})$, there must be at least one firm, say firm i, who joins the cartel at some step k of the sequence and remains in the cartel at the final outcome because the initial outcome has no actual cartel. By the definition of the E-dominance, we have

$$0 < \pi_f(p^k) = g_i(x^k, p^k) < g_i(x, p) = \pi_c(|x|, p).$$

The first strict inequality follows from the definition of π_f. Thus, we obtain $0 < \pi_f(p^k) < \pi_c(|x|, p)$.

Recall that $\pi_f(p) > 0$. Thus, we have $g_i(x, p) > 0$ for all $i \in N$. Finally, we show that for any $(y, w) \in K$ such that $\pi_c(|y|, w) \leqq 0$, (x, p) E-dominates (y, w). This is done by Lemma B3-(i) and (ii) because $\pi_c(|x|, p) > 0 \geqq \pi_c(|y|, w) \geqq \pi_c(|x \wedge y|, w)$ and $\pi_f(p) > \pi_c(|y|, w)$. This contradicts the internal stability of K. So we have the desired result. $\square$