Endogenous Cartel Formation with Free Market Entry and Firm Heterogeneity*

Xin Zhao

School of Economic Sciences
Washington State University

November 14, 2016

Abstract

Why do large firms choose to compete with a cartel rather than to cooperate? Bos and Harrington (2010) explained why small firms do not join in a cartel by introducing heterogeneous firms. But it is still unclear why large firms do not join in a cartel. This paper shows a possibility that only mid-level productive firms benefit from joining a cartel by considering endogenous choices of firm-production capacity. The reason is that low-productive firms cannot compete efficiently for production quota in cartel and staying out can yield them larger profits. High-productive firms prefer to stay out because building excess capacity in cartel lowers their profits. Additionally, this paper also contributes to the cartel literature by considering an endogenous market entry through incorporating the heterogeneous-firm model into the infinitely repeated game approach. So the results also predict that an increase in demand for output and a decrease in capacity and entry costs induce low-productive firms to enter a cartel. Technology improvements of low-productive independent firms and an increase in price elasticity of output demand entice high-productive firms to join a cartel.

JEL classification: L1, L2

Keywords: Endogenous Cartel Formation, Free Entry, Firm Heterogeneity

101 Hulbert Hall, Pullman, WA 99164.
xin.zhao@wsu.edu

*The author thanks Stephen Devadoss, Felix Munoz-Garcia, Gregmar Galinato, Ron Mittelhammer, and Thomas Dickinson for their helpful guidances and suggestions during the model derivation; the author also thanks the seminar attendees in the School of Economic Sciences: Mark Gibson, Ana Espinola-Arredondo, Alejandro Prera, Michael Brady, Jiehong Qiu, Daegoon Lee, You Zhou, and Yingzi Li for their valuable comments.

1 Introduction

Homogeneous-firm assumption implies that all firms benefit from joining a cartel, i.e., all-inclusive cartel, as studied by many research [Harrington Jr, 1991, Vasconcelos, 2004, 2005, Verboven, 1997]. However, most cartels do not include all the firms in one industry [Griffin, 1989, Harrington, 2006, Hay and Kelley, 1974]. A theoretical explanation to partial inclusive cartel was given by Bos and Harrington [2010].[1] They incorporate firm heterogeneity into an infinitely repeated game, showing that large firms benefit from joining a cartel, and small firms are better off by staying out.

In the real world, however, it is not only small but also large firms do not join a cartel. For example, Coors was a big independent vitamin B2 producer in 1990s. Powerpipe was a major supplier in the heating pipes industry but not a cartel member either. Both firms were so influential that their respective industry cartels had to take actions against challenges posed by these outsiders. Large Chinese vitamin C firms are another example of independent firms which exerted considerable influence that caused the worldwide vitamin C cartel to collapse in 1996 [Harrington, 2006]. These examples underscore that large firms, not only small firms, do not want to participate a cartel. However, the literature has not explained the reason behind the existence of large firms outside of the cartels.

Bos and Harrington [2010] assume firms are endowed with heterogeneous level of production-capacity. This assumption of exogenous capacity guarantees their result of only small firms staying out of cartel. In the short run, the exogenous capacity assumption may be justified because firms need time to adjust the production capacity levels. But, in the long run, firms can optimally choose their capacity levels to earn the maximum profit. Empirical literature has found average cartel duration varies from 3.7 to 10 years, though some cartels only last less than one year, and some cartels exist for decades [Levenstein and Suslow, 2006]. Therefore, it is conceivable that cartel-member firms have enough time to adjust their production-capacity levels. Since capacity level is a main criteria for many cartels to allocate production quotas, cartel members tend to build production capacity to secure the largest possible production quota [Harrington, 2006]. However, building production capacity does require firms to incur additional cost. Therefore, production-

[1] See Bos and Harrington [2010] for a detail review of the body of knowledge related to endogenous cartel formation.

capacity level should be an endogenous choice for cartel members in the profit maximization. In this study, we relax the assumption of exogenous production capacity, which allows us to explain the existence of large cartel outsiders.

Besides the challenges from existing outside firms, cartels are also impacted from the new entrants into the market. For example, the market entry of an independent lysine firm—ADM (Archer Daniels Midland)—caused the average price of lysine in the United States to fall from \$1.32 to \$0.68 per pound in 1990s, and 71% of U.S. lysine market was dominated by the new entrant. Eventually, ADM was included in the lysine cartel [Harrington, 2006]. Another example is with the rise of shale and sand oil firms in North America. World crude oil price dropped nearly 70% during 2014 to 2016[2]. In spite of the substantial price decline, Organization of the Petroleum Exporting Countries (OPEC)—the largest crude oil cartel—abandoned their traditional strategy of production cut and maintained their supply level to keep the oil price low with the goal of preventing market share loss to the new North American competitors.[3]

Market entry can be attributed to demand increase, technology improvement of less efficient firms, and more importantly to a high market price which could be caused by the collusion behavior of a cartel. Therefore, market entry and cartel behavior can be interdependent: cartel collusion causes a high market price, attracting market entry and market entry impacts the cartel, intensifying market concentration. Thus, market entry is an endogenous process. To my knowledge, however, the existing cartel literature has not considered an endogenous market entry. We model endogenous market entry by applying the heterogeneous-firm model of Hopenhayn [1992] to the infinitely repeated game. In doing so, the number of total market firms and cartel members can be solved endogenously, which has not been explored by the previous studies as they assumed an exogenous number of firms.

Therefore, this paper contributes to the endogenous cartel formation literature by endogenizing two firm choices, i.e., production capacity and market entry. In particular, this paper answers the research questions a) how does a partial inclusive cartel (large and small firms are not in the cartel) exist, and b) how does market entry affect the formation of a cartel?

[2] http://www.macrotrends.net/1369/crude-oil-price-history-chart (accessed Jul 26, 2016)

[3] http://www.opec.org/opec_web/en/index.htm (accessed Jun 1, 2016)

Most studies in the cartel literature dealt with price competition [Bos and Harrington, 2010, Harrington Jr, 1991, Vasconcelos, 2004], with exception of van den Berg and Bos [2011] and Paha [2010] who investigated quantity competition. In this study, we develop a model of quantity competition with firm heterogeneity because in many cartels such as lysine, citric acid, sorbates etc., firms with different costs compete in quantities [Harrington, 2006]. Then a four-stage complete information game is solved for the subgame perfect Nash equilibrium.

The findings show a possibility that in the equilibrium only the mid-level productive firms benefit from joining a cartel, but the least and most productive firms stay outside. The rationale is that low-productive firms cannot compete for sufficient amount of production quota in cartel and staying out yields them larger profits. In contrast, high-productive firms find building excess capacity in cartel lowers their profits. In addition to the cartel membership prediction, the results also illustrate that an increase in demand for output and a decrease in capacity and entry costs induce low-productive firms to enter the cartel. Technology improvements of high cost firms and a decrease in price elasticity of output demand entice high-productive firms to join in the cartel.

Section 2 discusses the model assumptions and time structure of the four-stage game. Section 3 solves the equilibrium of the game. Section 4 simulates the model and conducts the comparative statics. The final section 5 concludes the paper.

2 Model Assumption and Time Structure

In this section, we first present the various assumptions embedded in the model, then discuss a time structure involved in the analysis.

Assumption 1. *One cartel (r) and n independent firms engage in Stackelberg competition with homogeneous products.*[4] *The cartel is market leader, deciding output (q_r) at first. All the independent firms are followers and select outputs ($q_i, i = 1, 2, ..., n$) simultaneously by taking the cartel output as given.*

[4] The number of non-cartel firms n is to be solved endogenously.

Assumption 2. *Let P denote the market price. A linear inverse-demand function is defined as*

$$P = a - bQ. \tag{1}$$

The total market supply (Q) is composed of productions from all the independent firms and the cartel, i.e.,

$$Q = \sum_{i=1}^{n} q_i + q_r. \tag{2}$$

Assumption 3. *After the total output of cartel (q_r) is determined, the cartel allocates this output to m member firms as production quota (q_j), i.e., $\sum_{j=1}^{m} q_j = q_r$ where $j = 1, 2, ..., m$. Cartel production quota is assigned according to the share of production-capacity level of each member (k_j) to the whole capacity level of all members*

$$q_j = \frac{k_j}{\sum_{j=1}^{m} k_j} q_r. \tag{3}$$

Each member's output should not exceed the assigned production quota.

Assumption 4. *Because firms have various natural characteristics, unit-production cost (c_d, where $d = i, j$) of each firm is heterogeneous and follows a Pareto distribution $G(.)$ with the shape parameter κ and bound parameter c_b. The cumulative distribution function (CDF) of c_d is defined as*

$$G(c_d) = \left(\frac{c_d}{c_b}\right)^{\kappa}, \text{ where } c_d \in [0, c_b]. \tag{4}$$

Assumption 5. *Firm's total cost consists of production cost, capacity-building cost, and excess-capacity cost, i.e.,*

$$C_d(q_d, k_d) = c_d q_d + \rho c_d k_d + \delta(k_d - q_d) \quad \text{where} \quad k_d \geq q_d. \tag{5}$$

$c_d q_d$ represents the production cost. $\rho c_d k_d$ denotes capacity-building cost where k_d is dth firm's production capacity and ρ is the capacity-building cost multiplier. Production capacity refers to the maximum output that a firm is capable of producing. c_d appears in $\rho c_d k_d$ because capacity-building cost is also affected by natural characteristics of firms. ρ transforms the unit-production cost to unit capacity-building cost. The difference between production capacity and actual production $(k_d - q_d)$ in the third term of equation (5) captures the excess capacity. Excess capacity arises from production capability that are not utilized but can be ready for immediate production, which requires maintenance cost $\delta(k_d - q_d)$. δ is a multiplier that transforms the level of excess capacity to cost. Since excess-capacity cost cannot be negative, $k_d \geq q_d$.[5]

Reorganizing equation (5), total cost function can be simplified into two cost terms—one related to production and the other related to capacity:

$$C_d(q_d, k_d) = \underbrace{(c_d - \delta)q_d}_{\text{production term}} + \underbrace{(\rho c_d + \delta)k_d}_{\text{capacity term}}. \tag{6}$$

This cost function is generic and applies to all firms, both within and outside the cartel; however, we prove below that the independent firms do not have excess capacity and equation (5) becomes $C_d(q_d) = (c_d + \rho c_d)\, q_d$.

In Section 3, a four-stage complete information game is solved for the subgame perfect Nash equilibrium using backward induction.

1. In the first stage, firms decide whether to enter the market. A new entrant pays an entry cost c_e to draw a unit-production cost c_d, which has to be smaller than the cutoff cost $\hat{c}$ for the entrant to operate and earn a positive net present value of life time profit. Otherwise, the entrant fails to operate and exit the market.

2. In the second stage, the successful entrants decide whether to join the cartel or stay outside as an independent firm.

3. In the third stage, cartel-member firms compete for the largest possible production quota in cartel by selecting the optimal production-capacity levels.

4. In the final stage, cartel as a whole unit and all the independent firms perform Stackelberg competition by optimally producing their outputs.

[5] The component of excess-capacity cost in the cost function is in the same spirit of Lu and Poddar [2005] and Nishimori and Ogawa [2004]

3 Four-Stage Complete Information Game

3.1 Stage Four: Firms' Output Decisions

Using equation (1) and (6), profit function of firm d can be written as

$$\begin{aligned}\Pi_d(q_d, k_d) &= Pq_d - C_d(q_d, k_d) \\ &= (a - b(q_r + \sum_{i=1}^{n} q_i))q_d - (c_d - \delta)q_d - (\rho c_d + \delta)k_d \end{aligned} \tag{7}$$

Lemma 1. *No excess capacity exists for independent firms.*

Proof. Independent firm i can optimally choose both output and production-capacity level. However, equation (7) shows that building capacity can only reduce profit. Thus the smallest production capacity is preferred by independent firms. As demonstrated in Section 2, the smallest production capacity is the actual output i.e., $k_i = q_i$. Hence, no excess capacity should exist for independent firms. □

World crude oil industry provides a real world example of Lemma 1. The excess capacity of crude oil production in OPEC nearly reflects the total world excess capacity, and other major producers of crude oil operate nearly at full capacity.[6]

Plugging $k_i = q_i$ into equation (7), the profit function of independent firm i becomes

$$\pi_i(q_i, k_i = q_i) = (a - b(q_r + \sum_{i=1}^{n} q_i))q_i - (1+\rho)c_i q_i. \tag{8}$$

Section 3.2 below shows that cartel members carry excess capacity to compete for production quota. However, the level of production capacity is a choice of the individual member firm only and not that of the cartel because the cartel as a whole unit chooses the optimal total output without taking into account each member's excess capacity. This is similar to the independent firms selecting optimal output without accounting for excess production capacity, i.e., $k_r = q_r$. Let c_r denotes the mean unit-production cost of cartel members. Thus the cartel profit function is

[6]http://instituteforenergyresearch.org/analysis/the-significance-of-spare-oil-capacity/ (accessed Jul 26, 2016)

$$\pi_r(q_r, k_r = q_r) = (a - b(q_r + \sum_{i=1}^{n} q_i))q_r - (1+\rho)c_r q_r. \tag{9}$$

Lemma 2. *In Stackelberg competition with the cartel as market leader and all the independent firms as followers, the equilibrium outputs of cartel and independent firms can be solved as a function of the number of independent firms (n) and unit-production cost of all firms as given in equations* (11) *and* (12)*.*

Proof. Using backward induction, independent firm i chooses q_i to maximize equation (8) given the exogenous total cartel output q_r. From the first-order condition, the best response function of q_i to q_r can be written as

$$q_i = \frac{a - bq_r + (1+\rho)(\sum_{i=1}^{n} c_i - (n+1)c_i)}{(n+1)b}. \tag{10}$$

The cartel decides its total output (q_r) using the response function of all the independent firms. Plugging equation (10) into (9) and differentiating with respect to q_r, the optimal total cartel output q_r can be solved as

$$q_r = \frac{a + (1+\rho)(\sum_{i=1}^{n} c_i - (n+1)c_r)}{2b}. \tag{11}$$

Substitution of equation (11) into (10) yields the equilibrium output of independent firm i

$$q_i = \frac{a + (1+\rho)(\sum_{i=1}^{n} c_i + (n+1)(c_r - 2c_i)}{2b(n+1)}. \tag{12}$$

□

From the equation (12), as one would expect, the output of an independent firm decreases with its own unit-production cost but increases with other firms' unit costs.

By plugging (11) and (12)into (1), we can obtain the market price

$$P = \frac{a + (1+\rho)(\sum_{i=1}^{n} c_i + c_r(n+1))}{2(n+1)}. \tag{13}$$

By substituting(11) and (12) into (8), we can solve for independent firm i's profit

$$\pi_i = \frac{[a + (1+\rho)(\sum_{i=1}^{n} c_i + (n+1)(c_r - 2c_i)]^2}{4b(n+1)^2}. \tag{14}$$

The numerator in equation (14) is the square of the numerator in equation (12)—output function of independent firms. Since firm's output cannot be negative, the range of unit-production cost c_i is such that all the independent firms can operate with a positive output and profit

$$c_i \in \left(0, \frac{a + (1+\rho)\sum_{i=1}^{n} c_i + (1+\rho)(1+n)c_r}{2(1+\rho)(1+n)}\right].$$

The upper bound, $\hat{c} = \frac{a+(1+\rho)\sum_{i=1}^{n} c_i+(1+\rho)(1+n)c_r}{2(1+\rho)(1+n)}$, is the largest unit-production cost, i.e., the cutoff cost at which an independent firm earns zero profit, i.e.,$\pi_i(\hat{c}) = 0$. If $c_i > \hat{c}$, the firm will earn negative profit and does not operate. The profit function of all the independent firms (equation 14) is plotted against c_i in Figure 1. The relevant part of the curve is the first half of the quadratic curve which shows profit of the independent firms declines with unit production cost c_i.

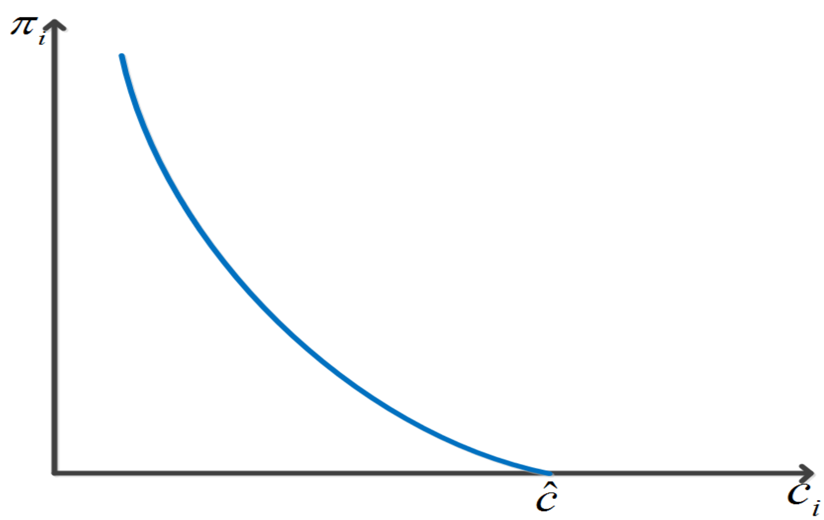

Figure 1. Profits across Unit-Production Costs for Indepednent Firms

3.2 Stage Three: Cartel Member's Capacity Choice

As elaborated in section 3.1, total market supply is jointly decided by the cartel and all the independent firms. Cartel members compete for the largest possible production quota inside the cartel by choosing the optimal capacity level based on production-capacity share of each member in the cartel.

Lemma 3. *Under the rule of production-quota allocation, i.e.,* $q_j = \frac{k_j}{\sum_{j=1}^m k_j} q_r$*, the optimal capacity level of each cartel-member firm is*

$$k_j = \frac{q_r(m-1)}{\sum_{j=1}^m \frac{\rho c_j+\delta}{P-c_j+\delta}} \left[1-(m-1)\frac{\frac{\rho c_j+\delta}{P-c_j+\delta}}{\sum_{j=1}^m \frac{\rho c_j+\delta}{P-c_j+\delta}}\right]. \tag{15}$$

Then each cartel-member firm's profit can be represented as a function of the unit-production costs of all cartel members

$$\pi(c_j) = q_r(P-c_j+\delta)\left[1-(m-1)\frac{\frac{\rho c_j+\delta}{P-c_j+\delta}}{\sum_{j=1}^m \frac{\rho c_j+\delta}{P-c_j+\delta}}\right]^2. \tag{16}$$

Proof. Substitute the production quota rule (equation 3) into the profit function (equation 7) to obtain

$$\pi(k_j) = P\frac{k_j}{\sum_{j=1}^m k_j}q_r - (c_j-\delta)\frac{k_j}{\sum_{j=1}^m k_j}q_r - (\rho c_j+\delta)k_j. \tag{17}$$

The optimal capacity level k_j for cartel member j can be solved by maximizing equation (17), which yields equation (15). Substitution of equation (15) into (17) results in equation (16). □

Using equation (15), the rule of production quota allocation, i.e., ratio of each member's capacity level to the aggregate capacity level in the cartel, can be calculated as

$$\frac{k_j}{\sum_{j=1}^m k_j} = \left[1-(m-1)\frac{\frac{\rho c_j+\delta}{P-c_j+\delta}}{\sum_{j=1}^m \frac{\rho c_j+\delta}{P-c_j+\delta}}\right]. \tag{18}$$

This result entails that a smaller unit-production cost c_j leads to a larger production-capacity ratio $\frac{k_j}{\sum_{j=1}^m k_j}$ which results in higher production quota q_j for cartel-member firm j. Production quota q_j can be computed by plugging equation (18) into (3).

The difference of equation (15)—production-capacity level—and the production quota q_j is the level of excess production capacity maintained by cartel-member firm j, i.e.,

$$k_j - q_j = q_r \left[\frac{m-1}{\sum_{j=1}^{m} \frac{\rho c_j + \delta}{P - c_j + \delta}} - 1 \right] \left[1 - (m-1) \frac{\frac{\rho c_j + \delta}{P - c_j + \delta}}{\sum_{j=1}^{m} \frac{\rho c_j + \delta}{P - c_j + \delta}} \right]. \tag{19}$$

Equation (19) implies that cartel-member firm with smaller unit cost carries larger excess production capacity, meaning that the cost from maintaining excess capacity is greater for smaller unit-cost cartel member.

The cartel member's profit function (16) can be reorganized as

$$\pi(c_j) = \frac{1}{(\sum_{j=1}^{m} \frac{\rho c_j + \delta}{P - c_j + \delta} + (m-1)\rho)^2} \cdot \frac{q_r (m-1)^2 \rho^2}{(\sum_{j=1}^{m} \frac{\rho c_j + \delta}{P - c_j + \delta})^2 (P - c_j + \delta)} \left[\frac{(P+\delta) \sum_{j=1}^{m} \frac{\rho c_j + \delta}{P - c_j + \delta} - (m-1)\delta}{\sum_{j=1}^{m} \frac{\rho c_j + \delta}{P - c_j + \delta} + (m-1)\rho} - c_j \right]^2. \tag{20}$$

The third term with the square in equation (20) causes cartel member's profit curve to be quadratic. However, the extra c_j in the denominator of second term causes the quadratic profit curve asymmetric in that the left side is flatter than the right side, implying that the speed of profit increase becomes slower as unit cost decreases for cartel-member firms. The reason is from the fact that smaller unit-cost cartel member incurs greater cost to maintain larger level of excess production capacity as implied by equation (19).

The nonnegativity condition for production-capacity level implies only the left flatter side of the quadratic curve is meaningful as in the profit curve of an independent firm, i.e., $k(c_j) \geq 0$ implies

$$c_j \leq \frac{(P+\delta) \sum_{j=1}^{m} \frac{\rho c_j + \delta}{P - c_j + \delta} - (m-1)\delta}{\sum_{j=1}^{m} \frac{\rho c_j + \delta}{P - c_j + \delta} + (m-1)\rho}.$$

The cartel member's profit curve is plotted in Figure 2, which decreases in c_j.

3.3 Stage Two: Endogenous Formation of Cartel

The cartel is open to any firm that wants to join. Firms decide whether to join the cartel or not according to the profit opportunities. Therefore, if at least two firms earn higher collusion profit

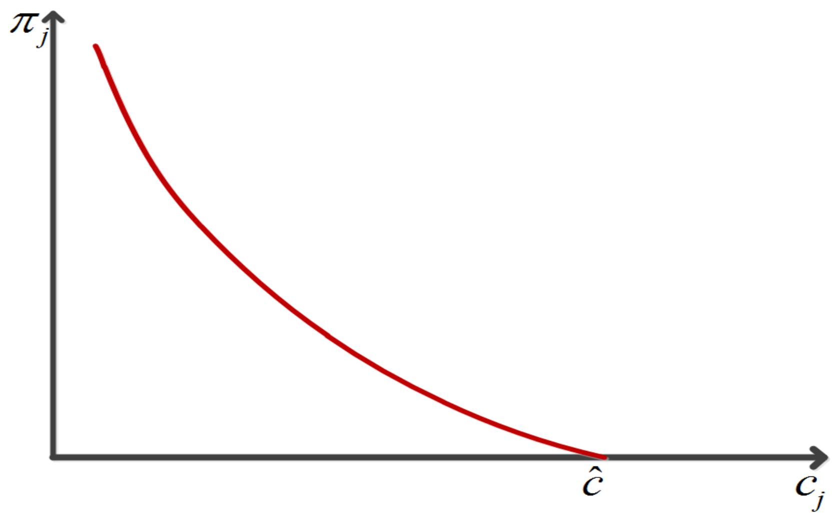

Figure 2. Profits across Unit-Production Costs for Cartel Members

than as independent firms, a cartel will be formed by these firms. If no firm benefits from joining cartel, cartel will not exist. Figure 3 combines figures 1 and 2 to show there are a total of five possibilities for a cartel to exist. First case is an all-inclusive cartel, i.e., all firms earn higher profit as a member of cartel than as an independent firm. The second case is that no cartel exists because all firms are worse off by joining the cartel. The third case is that only the lower cost firms join the cartel because of higher profits, and there exists a unit-production cost threshold c_h which is the upper bound for a firm to join the cartel. The fourth case, in contrast, is that the cartel consists of the higher cost firms with a lower bound unit-production cost threshold c_l.

The final case is that the cartel is comprised of firms with mid-level unit production costs with two cartel boundaries, i.e., c_l and c_h. For firms in the cost range between c_l and c_h, joining the cartel generates higher profit than staying out. For firms with unit-production cost less than c_l and greater than c_h, staying out of the cartel is more beneficial. The marginal firms with unit-production cost at c_l or c_h are indifferent between joining or staying out. Therefore, two profit-indifferent conditions can be implied:

$$\pi(c_j = c_l) = \pi(c_i = c_l) \tag{21}$$

$$\pi(c_j = c_h) = \pi(c_i = c_h). \tag{22}$$

In Sections 3.5 and 4, the general case 5 is used to discuss the endogenous formation of cartel.

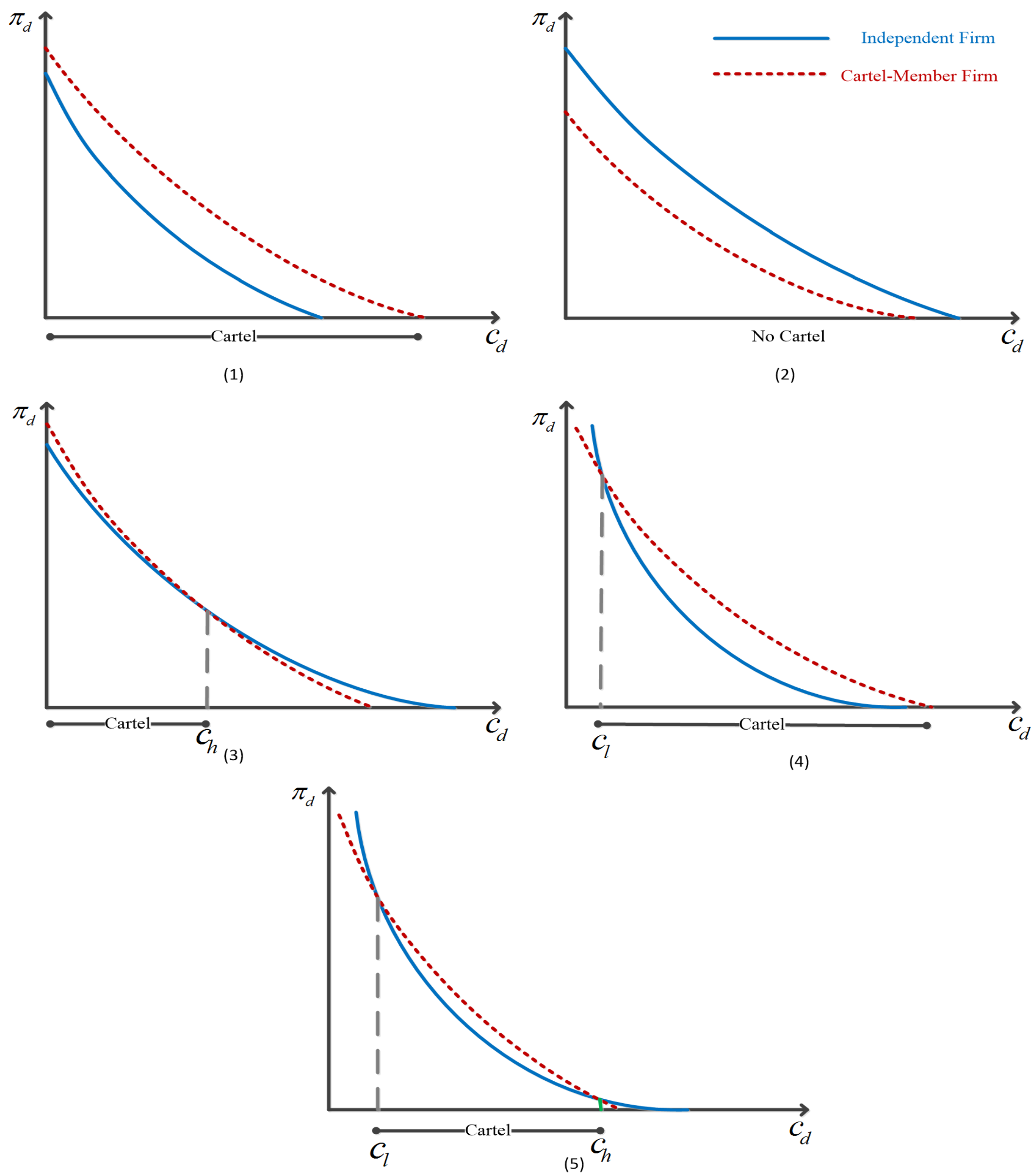

Figure 3. Profits versus Unit-Production Costs for Cartel Members and Independent Firms

3.4 Aggregate Unit Cost Variables

In equations (16) and (14), the optimal profits of cartel-member firms and independent firms have been shown to depend on the cartel members' mean production cost c_r, independent firms' aggregate production cost $\sum_{i=1}^{n} c_i$, and sum of ratio of the capacity cost to per-unit variable profit

$\sum_{j=1}^{m} \frac{\rho c_j+\delta}{P-c_j+\delta}$. This section aims at computing these three costs using the Pareto distribution (equation 4) for the unit-production cost.

The average unit-production cost of cartel c_r is

$$c_r = \int_{c_l}^{c_h} c_d d\frac{G(c_d)}{G(c_h)-G(c_l)} = \frac{\kappa(c_h^{\kappa+1} - c_l^{\kappa+1})}{(\kappa+1)(c_h^{\kappa+1} - c_l^{\kappa+1})}, \tag{23}$$

where $\frac{G(c_d)}{G(c_h)-G(c_l)}$ is the truncated cumulative distribution function for c_d in the cartel cost range between c_l and c_h.

From the fifth scenario in figure 3, independent firms exist in the two ranges of c_d: from 0 to c_l and from c_h to $\hat{c}$. Thus the summation of unit-production costs of all the independent firms is

$$\sum_{i=1}^{n} c_i = n_1 \int_0^{c_l} c_d d\frac{G(c_d)}{G(c_l)} + n_2 \int_{c_h}^{\hat{c}} c_d d\frac{G(c_d)}{G(\hat{c})-G(c_h)} = \frac{\kappa n(\hat{c}^{\kappa+1} - c_h^{\kappa+1} + c_l^{\kappa+1})}{(\kappa+1)(\hat{c}^{\kappa} - c_h^{\kappa} + c_l^{\kappa})}, \tag{24}$$

where n_1 and n_2 refer to the numbers of firms in their corresponding ranges of c_i, and thus, the total number of independent firms is $n_1 + n_2 = n$.

Using equation (4), the cumulative distribution function of $P - c_d + \delta$ can be computed as

$$G(P - c_d + \delta) = \Big(\frac{P - c_d + \delta}{P - c_b + \delta}\Big)^{\kappa},$$

where $c_d \in [0, c_b]$. Therefore,

$$H(s_j) = \Big(\frac{s_j}{P - c_b + \delta}\Big)^{\kappa},$$

where $s_j = (P - c_d + \delta) \in [P - c_b + \delta, P + \delta]$. Since $\frac{\rho c_j+\delta}{P-c_j+\delta} = \frac{\rho P+\rho\delta+\delta}{P-c_j+\delta} - \rho$, $\sum_{j=1}^{m} \frac{\rho c_j+\delta}{P-c_j+\delta}$ can be computed as

$$\begin{aligned}\sum_{j=1}^{m} \frac{\rho c_j + \delta}{P - c_j + \delta} &= m \int_{s_l}^{s_h} \Big(\frac{\rho P + \rho\delta + \delta}{P - c_j + \delta} - \rho\Big) d\frac{H(s_i)}{H(s_h) - H(s_l)} \\ &= \frac{m\kappa(\rho P + \rho\delta + \delta)}{\kappa - 1} \cdot \frac{(P - c_l + \delta)^{\kappa-1} - (P - c_h + \delta)^{\kappa-1}}{(P - c_l + \delta)^{\kappa} - (P - c_h + \delta)^{\kappa}} - \rho m,\end{aligned} \tag{25}$$

where m denotes the number of cartel members.

Substitution of equations (23), (24), and (25) into equations (14) and (16), profits of independent firms and cartel members can be solved as functions of seven parameters $(a, b, \rho, \delta, c_e, \kappa, c_b)$ and five unknowns $(n, m, c_l, c_h, \hat{c})$.

Since number of firms is proportional to the cumulative distribution function of c_d, the relation between n and m can be represented as

$$\frac{m}{n+m} = \frac{G(c_h) - G(c_l)}{G(\hat{c})}, \tag{26}$$

i.e., the share of number of cartel-member firms (m) to all the operating firms $(n+m)$ equals the ratio of probabilities of their corresponding ranges of unit-production cost.

3.5 Stage One: Firm's Market Entry Decision

In the first stage, firms decide whether or not to enter the market. To enter the market, a firm must pay an entry cost c_e to receive a unit-production cost draw c_d. The incentive for firms to pay the entry cost exists if their life time expected profit is greater than the entry cost, i.e., $E_{life}(\pi_d) > c_e$. As long as this inequality holds, more firms will enter the market until

$$E_{life}(\pi_d) = c_e, \tag{27}$$

which is the free-entry condition.

A successful entrant will operate and earn a positive profit if $c_d < \hat{c}$, the cutoff cost. In contrast, an entrant will earn negative profit and exit the industry if $c_d > \hat{c}$. At $c_d = \hat{c}$,

$$\pi(\hat{c}) = 0, \tag{28}$$

which is the zero-profit condition, and the firm will be indifferent between operating or exiting the market. This zero-profit condition can be obtained by combining equations (13), (25), and (24) and plugging into equation (14).

Following Melitz [2003], firms may die with a probability μ in each period. Firms weigh profits equally from all periods, and the life time expected profit can be expressed by summing the expected profits in each period, i.e.,

$$E_{life}(\pi_d) = \sum_{t=1}^{\infty}(1-\mu)^t E_t(\pi_d) = \frac{1-\mu}{\mu} E_t(\pi_d). \tag{29}$$

The unit-production cost draw may fall in one of the four possible segments as depicted in Figure 4. Thus, one-period expected profit for each possible segment can be computed by integrating over the respective range of the cost.

$$\begin{aligned}
&E_{c_d \in (0,c_l)}[\pi(c_i)] = \int_0^{c_l} \pi(c_i) d\frac{G(c_i)}{G(c_l)} \\
&E_{c_d \in [c_l,c_h]}[\pi(c_j)] = \int_{c_l}^{c_h} \pi(c_j) d\frac{G(c_j)}{G(c_h) - G(c_l)} \\
&E_{c_d \in (c_h,\hat{c}]}[\pi(c_i)] = \int_{c_h}^{\hat{c}} \pi(c_i) d\frac{G(c_i)}{G(\hat{c}) - G(c_h)} \\
&E_{c_d \in (\hat{c},c_b]}[\pi(c_d)] = 0
\end{aligned}$$

Firms with unit cost draw falling in $[c_l, c_h]$ will join the cartel and earn the cartel-member profit. Firms with cost in the ranges of $(0, c_l)$ and $(c_h, \hat{c}]$ will stay out and earn the independent firm profit. Firms with $c_d > \hat{c}$ will not operate and exit the market.

The probability of unit-production cost draw c_d falling in each segment can be computed using the cumulative distribution function as given below.

$$\begin{aligned}
&Pro[c_d \in (0, c_l)] = G(c_l) \\
&Pro[c_d \in [c_l, c_h]] = G(c_h) - G(c_l) \\
&Pro[c_d \in (c_h, \hat{c}]] = G(\hat{c}) - G(c_h) \\
&Pro[c_d \in (\hat{c}, c_b]] = 100\% - G(\hat{c})
\end{aligned}$$

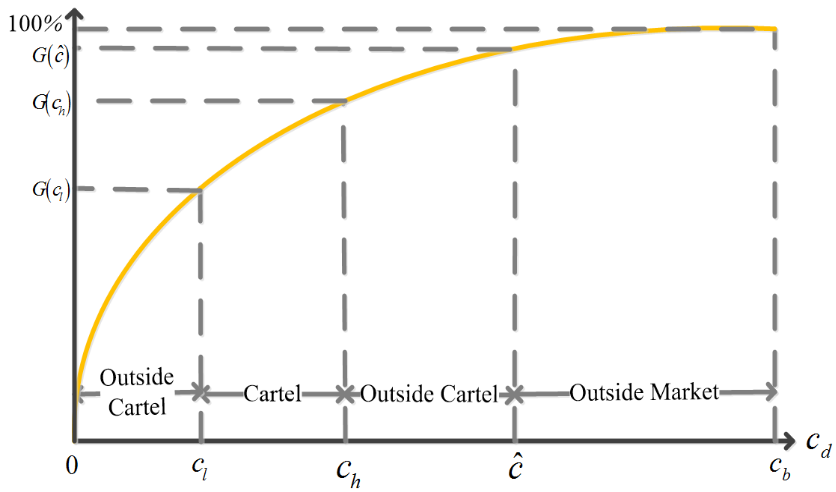

Figure 4. Cumulative Probability Distribution of Unit-Production Cost

Therefore, one-period total expected profit can be computed by summing over all the weighted expected profits of each possible segment of unit-production cost c_d:

$$E_t(\pi_d) = Pro[c_d \in (0, c_l)] \cdot E_{c_d \in (0,c_l)}[\pi(c_i)] + Pro[c_d \in [c_l, c_h]] \cdot E_{c_d \in [c_l,c_h]}[\pi(c_j)]$$
$$+Pro[c_d \in (c_h, \hat{c}]] \cdot E_{c_d \in (c_h,\hat{c}]}[\pi(c_i)] + Pro[c_d \in (\hat{c}, c_b]] \cdot E_{c_d \in (\hat{c},c_b]}[\pi(c_d)]. \quad (30)$$

Plugging equation (30) into equation (29), the life-time expected profit can be calculated.

Proposition 1. *The subgame perfect Nash equilibrium occurs when a) firms enter the market until the life-time expected profit equals the entry cost (equation 27). And the market marginal firms earn zero profit (equation 28). b) Upon entering the market, firms decide either to join the cartel or to stay as independent based on profit opportunities; marginal cartel members at the upper and lower bound are indifferent between staying in or out of the cartel (equation 21). c) Upon joining the cartel, member firms decide their production capacity (equation 15), which forms the basis for obtaining the production-quota share (equation 18). d) Cartel as a leader and independent firms as followers play Stackelberg game to select the optimal outputs, and cartel distributes the total outputs among its members (equation 3 and 12).*

Using the five equations—the zero-profit condition (equation 28), free-entry condition (equation 27), profit-indifference conditions (equation 21 and 22), and the relationship between n and m

(equation 26)—the subgame equilibrium can be solved for five unknowns: $n, m, c_l, c_h, \hat{c}$. Using the solution values of these variables, we compute the values of market price P, total industry output Q, and cartel output q_r.

Since some functional forms are highly non-linear, these equations do not lend themselves to analytical solutions, equilibrium values of these variables are solved through simulations with assumed parameter values.

4 Simulation and Comparative Statics

4.1 Base Simulation

We conduct simulation analysis with parameter values $a = 1000,\ b = 1,\ \rho = 2,\ \delta = 1,\ c_e = 10000,\ \kappa = 4,\ c_b = 1000,\ \mu = \frac{1}{2}$ to generate baseline results for the key endogenous variables (see table 1). This baseline leads to a lower bound unit cost of 61.82 and a upper bound unit cost of 136.93. These cost bounds entail that a cartel exists with firms in this mid-level production cost range, and firms with cost below the lower bound and above the upper bound do not participate in the cartel. The firms in the lower cost range 0 to 61.82 is better off not joining the cartel because their profits as independent firms are higher than the profits as cartel members (see figure 5). Similarly, firms in the higher cost range 136.93 to 160.11 will benefit from staying outside of the cartel because of the higher profit opportunities. These results highlight that only firms in the mid-level cost range belonged to the cartel. The baseline results show that the market-cutoff cost is 160.11, and any firm with cost higher than this cutoff cost will find unprofitable to operate and exit the industry.

Under the baseline parameter values, the number of independent firms is 1.98 and the number of cartel-member firms is 2.08.[7] Using the parameter and the solution values, we compute price from equation (13), which is substituted into equations (1) to obtain the market output and cartel output.

[7]Since the model is based on continuous variables, the numbers of firms are not integers.

Variables	Baseline	Lower Entry Cost c_e	Lower Capacity Cost ρ	Lower Excess Capacity Cost δ	Bigger Demand a	Smaller Demand Elasticity b	Lower Cost for High Cost Firm κ
Cartel Lower-Cost Bound c_l	61.82	73.61	166.83	61.87	163.27	41.71	42.34
Cartel Upper-Cost Bound c_h	136.93	136.57	298.85	137.13	205.32	134.42	123.15
Market-Cutoff Cost $\hat{c}$	160.11	157.17	344.74	160.34	213.98	160.28	150.09
Number of Ind. Firms n	1.98	2.15	1.66	1.97	4.94	1.86	1.29
Number of Cartel Members m	2.08	2.35	1.73	2.08	5.12	1.79	1.88
Market Price P	480.34	471.51	517.11	481.01	641.93	480.84	450.28
Total Market Supply Q	519.66	528.50	482.89	518.99	1358.07	259.58	549.72
Cartel Total Output q_r	429.01	409.06	376.67	428.60	486.45	223.67	417.07

Table 1. Simulation Results

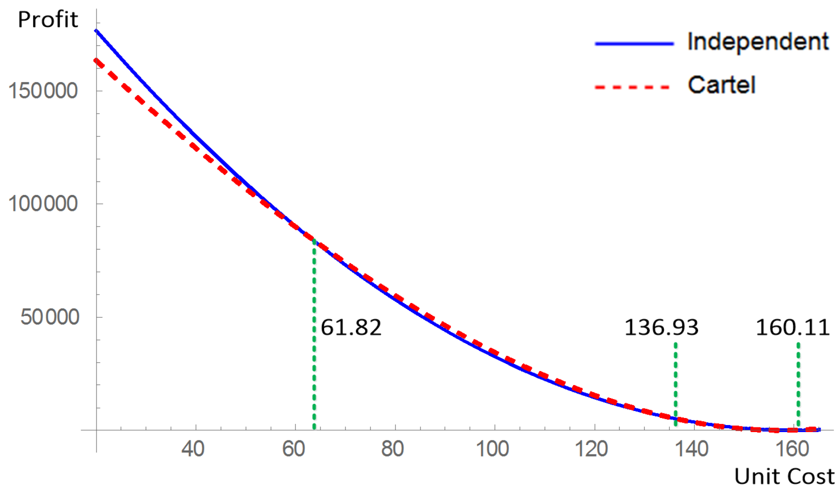

Figure 5. Profit Simulation

4.2 Comparative Statics

We conduct comparative statics analysis by changing the values of parameters to examine the impacts on the key endogenous variables and implications for the market equilibrium and industry. The results of these comparative statics simulations are reported in table 1.

The first scenario investigates the effects of smaller entry cost ($c_e : 10000 \to 5000$), i.e., easier to enter the industry. Because of the decline in entry cost, the life time expected profit will be greater than the entry cost, i.e., $E_{life}(\pi_d) > c_e$. As a result, more firms will pay the entry cost and enter the market, which will increase the number of successful entrants and both the number of independent (n) and cartel-member (m) firms will rise. As entry of more firms intensifies the competition, only more efficient firms with lower unit cost can survive, which lowers the market-cutoff unit cost ($\hat{c}$). This causes the average cost of all operating firms to decline, i.e., firms are more productive which augments total industry supply (Q), leading to a lower market price. Due to this reduction in price, lower-cost cartel members find it more beneficial to exit the cartel than to stay in the cartel and maintain the excess capacities. Further more, due to the lower price, higher-cost members can no longer compete efficiently for production quota and leave the cartel. Moreover, the increase of cartel lower bound (c_l) is greater than the decrease of cartel upper bound (c_h) because lower-cost members have to carry large excess capacity than higher-cost members if they stay in the cartel (see equation 18). Because of this composition of firm exits, average cartel cost is higher and cartel firms become less efficient, leading to a decline in total cartel output.

The second scenario considers a decease in the capacity building cost ($\rho : 2 \rightarrow 0.5$), which causes all the firms become more productive though their unit-production costs remain the same. Even the marginal market firm will earn positive profit and firms with higher production cost will operate, which raises the market-cutoff unit cost ($\hat{c}$). Since demand has not changed and firms are more productive, fewer firms operate and total market supply decreases, resulting in a higher price. With the higher price, the opportunity cost of staying in cartel becomes larger for lower cost members. Consequently, these firms leave the cartel, i.e., c_l becomes larger. Because of the increase in market price, high cost market firms find it more profitable to build excess capacity and join the cartel with expectation of securing a larger production quota and earn higher collusion profit, which increases c_h the cartel upper bound unit-production cost. As a result, the cartel cost range expands rightward.

Since the desire to maintain excess capacity varies depending on the industry, we examine in this scenario the effect of zero excess capacity cost, i.e., $\delta = 0$. The simulation results show that the value of δ does have directional impacts, even though the magnitudes of these impacts are not large due to small change in δ from 2 to 0. Since only cartel member firms carry excess capacity, only these firms benefit from the lower excess capacity cost. However, higher-cost cartel members have more advantage in competing for the production quota due to this decline in excess capacity cost, because unit-excess capacity cost is same for all cartel members. As a result, higher-cost non-cartel firms have incentive to join the cartel and lower-cost cartel members exit the cartel. Consequently, the whole cartel-cost range moves to the right, i.e., c_l and c_h increase. Furthermore, this lower excess-capacity cost decreases the average cost of all operating firms which entices higher-cost firms to operate even though these firms were not profitable before the decline in the value of δ. This causes the market-cutoff cost ($\hat{c}$) to rise. As demand remains the same, the number of firms decreases (note $m = 2.078$ before rounding two decimals), resulting in lower industry output and higher market price.

In the fourth scenario, we analyze the effect of an increase in demand ($a : 1000 \rightarrow 2000$). As demand increase boosts market price, the life-time expected profit exceeds the entry cost, i.e.,

$E_{life}(\pi_d) > c_e$. The higher expected profit induces more firms to pay the entry cost and enter the market, resulting in an increase in both independent (n) and cartel-member (m) firms. So total market supply expands. The demand increase also causes the higher cost firms that were outside the market to be profitable to operate, leading to a higher market-cutoff cost $\hat{c}$. With the higher price, the opportunity cost of staying in cartel becomes larger for lower cost members. Consequently, these firms leave the cartel, i.e., c_l becomes larger. Because of the higher market price, high cost market firms find it more profitable to build excess capacity and join the cartel with expectation of securing a larger production quota and earn a higher collusion profit, which increases c_h the cartel upper bound unit-production cost. As a result, the cartel cost range expands rightward.

The fifth scenario considers a decrease in the price elasticity of demand ($b : 1 \rightarrow 2$), which enhances firms' market power. This augments the firm's profitability and incentivizes the high-cost firms that were out of market to operate, resulting in a higher market-cutoff unit cost ($\hat{c}$). With the greater market power, low cost firms prefer to join the cartel as collusion can generate higher profits, which decreases the lower bound c_l. However, with this addition of lower-cost firms, the incumbent higher-cost members can no longer compete efficiently for quota and exit the cartel, which decreases the upper bound cost c_h. As a result, the cartel cost range ($c_l - c_h$) moves to left, implying more productive firms operate in the cartel as a result of greater market power. With demand not changing, this increase in market power decreases the number of operating firms, both independent and cartel firms. Consequently, market supply decreases and price increases.

In the final scenario, we analyze the case when the probability distribution function (PDF) of unit-production cost becomes flatter ($\kappa : 4 \rightarrow 2$), implying more low-cost firms and fewer high-cost firms, which could be a consequence of technology improvements of the high cost firms. As more efficient firms operate, it lowers the average unit-production cost of all the operating firms in market, leading to a lower cutoff cost $\hat{c}$, higher market supply, and lower price. As more low-cost firms operate, market competition is intensified. As a result, these low-cost firms earn collusion profit by joining the cartel. Consequently, cartel's lower-bound cost (c_l) decreases. With more lower-cost firms joining the cartel, high-cost members no longer can compete efficiently for production quota, and thus exit the cartel, i.e., c_h decreases. Since demand remains the same, the decrease in average unit-production cost of operating firms leads to a fewer operating firms in market.

5 Conclusion

We analyze in this paper the endogenous cartel formation by incorporating the heterogeneous-firm model into the infinitely repeated game problem. This allows us to study the market entry effect and the number of independent firms and cartel members can be solved endogenously. We also endogenize firms' production capacity choices, which predicts the possibility that only mid-level productive firms benefit from joining a cartel and the most and least productive firms stay outside the cartel. By the comparative statics analysis, we show that an increase in demand and a decrease in capacity and entry costs induce low-productive firms to enter the cartel. Technology improvements of high cost non-cartel firms and a decrease in price elasticity of demand entice high-productive firms to join cartel.

References

I. Bos and J. E. Harrington. Endogenous cartel formation with heterogeneous firms. *The RAND Journal of Economics*, 41(1):92–117, 2010.

J. M. Griffin. Previous cartel experience: Any lessons for opec? In *Economics in theory and practice: An eclectic approach*, pages 179–206. Springer, 1989.

J. E. Harrington. *How do cartels operate?* Now Publishers Inc, 2006.

J. E. Harrington Jr. The determination of price and output quotas in a heterogeneous cartel. *International Economic Review*, pages 767–792, 1991.

G. A. Hay and D. Kelley. Empirical survey of price fixing conspiracies, an. *JL & Econ.*, 17:13, 1974.

H. A. Hopenhayn. Entry, exit, and firm dynamics in long run equilibrium. *Econometrica: Journal of the Econometric Society*, pages 1127–1150, 1992.

M. C. Levenstein and V. Y. Suslow. What determines cartel success? *Journal of Economic Literature*, 44(1):43–95, 2006.

Y. Lu and S. Poddar. Mixed oligopoly and the choice of capacity. *Research in Economics*, 59(4): 365–374, 2005.

M. J. Melitz. The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71(6):1695–1725, 2003.

A. Nishimori and H. Ogawa. Do firms always choose excess capacity? *Economics Bulletin*, 12(2): 1–7, 2004.

J. Paha. Simulation and prosecution of a cartel with endogenous cartel formation. Technical report, Joint discussion paper series in economics, 2010.

A. van den Berg and I. Bos. Collusion in a price-quantity oligopoly. *Department of Quantitative Economics, Maastricht University, The Netherlands (Manuscript)*, 2011.

H. Vasconcelos. Entry effects on cartel stability and the joint executive committee. *Review of Industrial Organization*, 24(3):219–241, 2004.

H. Vasconcelos. Tacit collusion, cost asymmetries, and mergers. *RAND Journal of Economics*, pages 39–62, 2005.

F. Verboven. Collusive behavior with heterogeneous firms. *Journal of economic behavior & organization*, 33(1):121–136, 1997.