The Optimal Market Structure in an Asymmetric Oligopoly

Donald J. Smythe and Jingang Zhao*

California Western School of Law
Iowa State University

October 2001

Abstract: We characterize the optimal market structure in Cournot oligopolies using the underlying cost structure. In symmetric oligopolies, the optimal structure will be a monopoly (perfectly competitive; an oligopoly) if costs exhibit strong economies of scale (constant returns to or diseconomies of scale; U-shaped average cost). In asymmetric oligopolies with two types of firms (i.e., with identically efficient and inefficient costs), the optimal structure will be perfectly competitive (a monopoly) with no inefficient firms if the efficient firms’ costs exhibit constant returns to or diseconomies of scale (strong economies of scale); and it will be an asymmetric oligopoly with both efficient and inefficient firms if both types’ costs exhibit U-shaped average cost curves. These characterizations imply empirically estimable criteria for assessing whether mergers, entry, or exit will be welfare-maximizing as opposed to merely welfare-increasing.

JEL Classification Number: L10, L13, L40

Key Words: Oligopoly, optimal market structure, mergers, entry, exit, antitrust

* Address all correspondence to Jingang Zhao, Department of Economics, 260 Heady Hall, Iowa State University, Ames, Iowa 50011-1070, jingang@iastate.edu, Fax: (515) 294-0221, Tel: (515) 294-5245. We would like to thank Steve Heubeck, Hajime Miyazaki and seminar participants at Univ. of Missouri, NanYang Technical Univ., Ohio State Univ., National Univ. of Singapore, Univ. of Windsor, 99 FEMES and 99 EARIE conference for useful comments. All errors, of course, are our own.
I. Introduction

The welfare effects of horizontal mergers, entries, and exits have been widely studied in the economics literature, although usually in separate papers. In this paper, we adopt a unified approach to merger, entry, and exit problems by framing them as problems in welfare maximization, and characterize the optimal market structure in terms of the underlying cost structures. In this respect, we extend Mankiw and Whinston’s (1986) and Suzumura and Kiyono’s (1987) work on symmetric oligopolies in three directions.¹

First, we show that the optimal market structure in symmetric oligopolies will be a monopoly (perfectly competitive; an oligopoly) when costs exhibit strong economies of scale (constant returns to or diseconomies of scale; U-shaped average cost). Second, we extend our analysis to asymmetric oligopolies with two types of firms (i.e., those with (identical) low costs and those with (identical) high costs). The optimal structure includes an infinite number of efficient (low cost) firms and zero inefficient (high cost) firms (is a monopoly) when the efficient firms’ costs exhibit constant returns to scale or diseconomies of scale (strong economies of scale), and it includes a finite number of both efficient and inefficient firms when both types’ costs exhibit U-shaped average cost curves.

Third, we show that the second order conditions for welfare maximization hold for most of the empirically relevant specifications of market demand and costs, including linear demand and costs and most quadratic and loglinear specifications as well. Our characterizations thus provide empirically estimable criteria for determining whether horizontal mergers, entries, and exits will be welfare-maximizing as opposed to merely welfare-increasing.

The paper is organized as follows: Section II defines the welfare maximization problem. Sections III and IV characterize the solutions in symmetric and asymmetric oligopolies, respectively. Section V discusses empirical applications and optimal merger, entry, and exit policies. Finally, section VI provides conclusions, and the appendix provides proofs.

II. The Welfare Maximization Problem

¹ See also von Weizsacker (1980), Farrell and Shapiro (1990), and Nachbar et al. (1998).
Consider an oligopoly with a homogeneous good, an inverse demand \( P(X) = P(\Sigma x_j) \), and \( k \) cost functions, \( C_i(x_i), i = 1, \ldots, k \). Assume that output is the choice variable and thus focus on the Cournot equilibrium, which is a vector of outputs such that each firm \( i \)'s output maximizes its profits, as given by:

\[
\pi_i(x) = \pi_i(x_1, \ldots, x_k) = P(\Sigma x_j)x_i - C_i(x_i).
\]

In other words, each firm's output, \( x_i \), is i's best response to \( x_{-i} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k) \). The Cournot equilibrium thus solves for \( x \) in the following first order conditions:

\[
P(X) + x_iP'(X) = C'_i(x_i), i = 1, \ldots, k.
\]

This paper will assume throughout that a unique Cournot equilibrium exists. Existence is guaranteed by the assumption that \( P'(X) + x_i P''(X) \leq 0 \) for all \( x_i \leq X \) (Novshek, 1985). In general, the requirements for uniqueness are more complex; some recent results are surveyed in Zhang and Zhang (1996).

We focus on asymmetric oligopolies with two types of firms: \( m \) identical low cost (efficient) firms and \( n \) identical high cost (inefficient) firms (where \( m + n = k \)). Thus, a market structure is given by an \( (m,n) \) pair. Social welfare is defined as the sum of consumer and producer surplus:

\[
W = W(m,n) = CS + \Sigma p_i = \int_X P(t)dt - \Sigma C_j(x_j).
\]

The optimal market structure is therefore the pair \( (m^*,n^*) \) which maximizes \( W(m,n) \) over all \( m \) and \( n \).

Let \( m^*(n) \) denote the number of efficient firms that maximizes \( W \) for a given \( n \), and \( n^*(m) \) denote the number of inefficient firms that maximizes \( W \) for a given \( m \). As the discussion below will elaborate, a thorough understanding of socially optimal merger and entry policies requires an analysis of \( m^*(n) \), \( n^*(m) \), and \( (m^*,n^*) \).

### III. The Optimal Market Structure in a Symmetric Oligopoly

Mankiw and Whinston (1986) and Suzumura and Kiyono (1987) have shown that in a symmetric oligopoly the socially optimal number of firms, \( k^* \), is generally less than the

---

2 As in Mankiw-Whinston (1986) and Suzumura-Kiyono (1987), \( (m^*,n^*) \) is optimal only in the second-best sense, because it maximizes welfare under the constraint that firms are Cournot oligopolists.
free entry equilibrium number of firms, \( k^e \). In this section, we extend their results by characterizing how the optimal number of firms is determined by the properties of firms’ costs. We begin with the following assumption, A1:

**A1:** Let \( C_i(q) = C(q) \) for all \( i \), and let the following conditions hold: (i) \( P', C'' \) are continuous; (ii) \( P' < 0, C' > 0 \); (iii) \( P'(X) + x_i P''(X) < 0 \) for all \( x_i \leq X \); (iv) \( P'(X) - C''(x_i) < 0 \) for all \( x_i \leq X \); and (v) the Cournot equilibrium is unique.

By symmetry, the first order conditions become: \( P(X) + x_i P'(X) = C'(x_i) \), for all \( i \), and the firms’ equilibrium outputs can be written as \( x_i = q \) for all \( i \), where \( q \) solves

\[
P'(kq) + q P''(kq) = C'(q).
\]

As in the previous literature, we treat the number of firms as a continuous variable, assume all endogenous variables are differentiable functions of \( k \), and assume that \( P \geq AC \) in all equilibria. Under A1, it is relatively simple to show the following results:

\[
\frac{dq}{dk} = -\frac{\alpha q}{k\alpha + \beta} < 0, \quad \text{and} \quad \frac{dX}{dk} = \frac{\beta q}{k\alpha + \beta} > 0,
\]

where \( X = kq \), and \( \alpha \) and \( \beta \) are given by

\[
\alpha = P'(kq) + q P''(kq) < 0, \quad \text{and} \quad \beta = P'(kq) - C''(q) < 0.
\]

Since (4) holds for all \( k \), entry always increases total output and decreases the market price, even though it always decreases firms’ individual outputs. Let \( \gamma_0 \) be given by

\[
\gamma_0 = \gamma_0(k) = \frac{\beta}{\alpha} = \frac{P'(kq(k)) - C''(q(k))}{P'(kq(k)) + q(k)P''(kq(k))} > 0.
\]

Let \( AC = C(q)/q \) be the average cost, let \( AC' = dAC/dq = (C'(q) - AC)/q \), and let

---

3 The Suzumura-Kiyono (1987) results are identical to those in Mankiw-Whinston (1986) when the conjectural variations parameter, \( \mu \), is zero. Assuming \( W''(k) < 0 \) and \( W'(k) = \pi(k) - XqP' = 0 \) (see (D0) in the appendix), one has \( \pi(k^*) \geq 0 \). Then, \( d\pi(k)/dk \leq 0 \) leads to \( k^* \leq k^e \), where \( k^e \) solves \( \pi(k) = 0 \).

4 Part (iii) can be weakened to “\( \leq 0 \),” without harming the general results. We assume “\( < 0 \)” for simplicity. Part (iv) is the same as Assumption 4 (with \( \mu = 1 \)) in Suzumura and Kiyono (1987). Parts (iii) and (iv) together lead to \( \partial^2 \pi/\partial x_i^2 = (2P' + x_iP'' - C'(x_i)) < 0 \). These assumptions are made implicitly in Mankiw and Whinston (1986) in order to carry out the comparative statics.
\[
\theta = \theta(k) = \frac{AC(q)-C'(q)}{P(kq)-AC(q)}
\]
denote a measure of relative economies of scale (EOS). Since \( P > AC \), it follows that
\[
AC' \leq 0 \text{ for all } q(k) \iff \theta \geq 0 \text{ for all } k.
\]
Therefore, there are global dis-EOS (constant-EOS) \( \iff \theta < 0 \text{ (} \theta = 0 \text{) for all } k \); and there are
global EOS \( \iff \theta > 0 \text{ for all } k \).

We use the following concepts of weak, moderate, and strong global EOS in characterizing the optimal number of firms:

**Definition 1:** Let \( \gamma_0 \) and \( \theta \) be given by (6-7) and assume \( \theta > 0 \) for all \( k \) and \( q \). (i) \( C(q) \) exhibits weak global EOS if \( \theta < \gamma_0/k \) for all \( k>0 \); (ii) \( C(q) \) exhibits strong global EOS if \( \theta > \gamma_0/k \) for all \( k>0 \); and (iii) \( C(q) \) exhibits moderate global EOS if neither (i) nor (ii) holds.

**Proposition 1:** Let \( k^* \) denote the optimal number of firms in a symmetric oligopoly. Under A1, the following claims hold:

(i) the slope of the welfare function \( W(k) \) satisfies
\[
dW/dk > 0 \iff (\gamma_0 - k\theta) > 0;
\]
(ii) \( k^* = \infty \) if \( C(q) \) exhibits global dis-EOS, constant EOS, or weak global EOS;
(iii) \( k^* = 1 \) if \( C(q) \) exhibits strong global EOS; and
(iv) \( k^* \) is given by
\[
k^* = \frac{\gamma_0(k_0,q)}{\theta(k_0,q)} = \frac{(P'-C'')(P-AC)}{(AC-C')(P''+qP''')}
\]
if \( C(q) \) exhibits moderate global EOS or U-shaped average costs, and if \( W(k) \) has a unique peak point.

To summarize, perfect competition is socially optimal if there are global dis-EOS, constant EOS, or weak global EOS (see parts a and b of Figure 1);\(^5\) a pure monopoly is

---

\(^5\) Note that \( \theta < \gamma_0/k \iff \text{lik} [-AC''(q)] < \text{lik} [P(kq)-AC(q)] \gamma_0 \). Since total output, \( kq = kq(k) \), tends to a constant as \( k \to \infty \), the previous expression implies that with weak (strong) global EOS, the slope of average cost, \([-AC''(q)] \), will be bounded from above (below) by a small number. This is why the weak (strong) global EOS case is associated with a relatively flat (steep) AC curve.

---
socially optimal if there are strong global EOS (see part c of Figure 1); and, assuming there is a unique peak point, a market structure in between perfect competition and a pure monopoly is socially optimal if there are moderate global EOS or the AC curve is U-shaped (see parts d and e of Figure 1).

(Figure 1 about here)

Note that the constant EOS case in part (ii) includes markets with any demand and symmetric constant variable costs. Note also that Proposition 1 completely characterizes the first order conditions in Mankiw-Whinston (1986) and Suzumura-Kiyono (1987). Since \( W(k)' > 0 \) \((< 0)\) in part (ii) (part (iii)), \( W(k)' = 0 \) holds only with moderate EOS or a U-shaped AC curve. It is helpful to recompute Mankiw-Whinston’s Example 1:

**Example 1:** Let \( p = a-bQ, c(q) = c_0+cq \). Then, (10) becomes \( k = (bq^2-c_0)/c_0 \). By \( q = (a-c)/[(k+1)b] \), one derives the same formula, \( (k^*+1)^3 = (a-c)^2/(c_0b) \).

In addition, note that we provide sufficient conditions (see Lemma 1 and Claim 1 in the appendix) for the solution of \( W(k)' = 0 \) to be the unique peak point of \( W(k) \). As shown in Figure 1 and Claim 1, the welfare function in part (ii) (part (iii)) is globally increasing (decreasing), which implies \( k^* = \infty \) \((= 1)\). The welfare function in part (iv) is initially increasing and eventually decreasing; the unique peak assumption guarantees that \( k^* \) is the unique global optimal. Without the unique peak assumption, the W curve will still have a finite global maximum (i.e., optimal market structures will still exist and still lie somewhere in between perfect competition and pure monopoly) but they may simply not be unique (see part f(ii) in Figure 1) and will not be uniquely identified by equation (10).

It is important to note that the welfare function in part (iv) will have a unique peak for most, if not all, empirically relevant specifications of market demand and costs. Indeed, as shown in part (iv) of Claim 1, the conditions for a unique peak hold in all linear markets and in large classes of markets with quadratic or loglinear demand and costs. It is also important to note that our conditions for a unique peak are weaker than the global concavity assumption, since they allow piece-wise convex curves as shown in part f(i) of Figure 1. Lemma 1 is a useful contribution in itself, as it can be used to simplify the second order conditions in other studies.

The optimal market structure entails a tradeoff between economic surplus and
economies of scale. Recall that $\theta$ is a measure of relative scale economies. The parameter $\gamma_0$, on the other hand, is a weighted measure of the curvature of the inverse demand function relative to the curvature of firms' marginal costs. Hence, it provides a measure of how much total economic surplus increases with an increase in industry output. The larger (smaller) $\gamma_0$, the steeper (flatter) the demand curve relative to firms' marginal costs, and the less (more) economic surplus increases with an increase in industry output. Part (i) of Proposition 1 therefore states that entry of one or more firms will increase social welfare if and only if any scale economies that might be lost (due to a reduction of each firm’s output) are more than offset by the economic surplus gained.

In this light, the results are quite transparent. Obviously, with constant returns or diseconomies of scale, an increase in the number of firms will never cause a loss in scale economies and so the optimal market structure will entail an infinite number of firms. And if there are global EOS but they are sufficiently weak, the same calculus will apply. With strong global EOS, on the other hand, a pure monopoly will be optimal, since the economic surplus gained by adding even a single entrant would not justify the economies foregone. The truly interesting case arises, however, where there are moderate global EOS; in that case, the optimal market structure entails a balance between economies of scale and incremental economic surplus. Equation (10) shows how the appropriate balance can be calibrated using estimable parameters of demand and costs.

Proposition 1 provides further insights. Consider the following example:

Example 2: Let $P = 100 - Q$ and $C(q) \equiv 370.37$. One has $\alpha = \beta = -1$, $\gamma_0 = 1$, and $\theta = 0.5$, and $k^* = 2$. By $W(2) = 3703.70 > W(3) = 3576.39 > W(1) = 3379.63$, a duopoly is superior to either an unregulated monopoly or a three firm Cournot oligopoly. But suppose that two of the three oligopoly firms wanted to merge (assume no cost synergies). The Herfindahl-Hirschman Index (HHI) would increase from 3333.33 to 5000. Indeed, if the antitrust authorities ignored the strategic effects of the merger on the post-merger market shares of the remaining two firms they would (incorrectly) forecast that the HHI would increase to 5555.55. Regardless of how the calculations were done, the merger would fall well outside the safe harbors established under the DOJ’s and FTC’s 1992 Horizontal
Merger Guidelines,\textsuperscript{6} even though it would actually increase welfare by 3.44%.

IV. The Optimal Market Structure in Asymmetric Oligopolies

Consider now the optimal market structures in asymmetric Cournot oligopolies. Recall there are \( m \) identical efficient firms with costs, \( C_1(q) \), and \( n \) identical inefficient firms with costs, \( C_2(q) \). We begin by characterizing the optimal number of efficient firms for a given number of inefficient firms, \( m^*(n) \).

A. The optimal number of efficient firms, \( m^*(n) \).

We modify A1 as follows:

\textbf{A2:} (i) \( P' \), \( C'_1 \) and \( C'_2 \) are continuous; (ii) \( P' < 0 \), \( C'_1(q) < C'_2(q) \) for all \( q \geq 0 \), \( C'_i(x) < C'_j(y) \) holds in any equilibrium \((x,y)\) with \( x > y \); (iii) \( P'(X) + XP''(X) < 0 \); (iv) \( P'(X)- C''_i(x_i) < 0 \) for all \( i \) and all \( x_i \leq X \); and (v) the Cournot equilibrium is unique.

Part (ii) of A2 assumes that an efficient firm's marginal cost curve lies below that of an inefficient firm, and that an efficient firm's marginal cost is always less than that of an inefficient firm in any equilibrium. The Cournot equilibrium is now defined by \( m \) first-order conditions for the efficient firms and \( n \) first-order conditions for the inefficient firms. The symmetry within each class of firms implies that \( x_i = q_1 \) for all efficient firms, and that \( x_i = q_2 \) for all inefficient firms. Hence, \( q_1 \) and \( q_2 \) satisfy the following equations:

\[
\begin{align*}
&\text{(11)} \quad P(mq_1 + nq_2) + q_1 P'(mq_1 + nq_2) = C'_1(q_1), \text{ and} \\
&\quad P(mq_1 + nq_2) + q_2 P'(mq_1 + nq_2) = C'_2(q_2),
\end{align*}
\]

Note that these imply \( q_1 > q_2 \). Let \( \alpha_i, \beta_i, \) and \( A \) be given by

\[
\begin{align*}
&\text{(12)} \quad \alpha_i = P'(mq_1 + nq_2) + q_1 P''(mq_1 + nq_2) < 0, \\
&\quad \beta_i = P'(mq_1 + nq_2) - C''_i(q_i) < 0, \text{ and} \\
&\quad A = m\alpha_1\beta_2 + n\alpha_2\beta_1 + \beta_1\beta_2 > 0.
\end{align*}
\]

By differentiating (11) with respect to \( m \) and to \( n \), we obtain the following comparative static results for \( q_1, q_2, \) and \( X = mq_1 + nq_2 \):

\textsuperscript{6} Department of Justice and Federal Trade Commission, 1992 Horizontal Merger Guidelines.
\[
\begin{align*}
\frac{dq_1}{dm} &= \frac{-\alpha_1 \beta_2 q_1}{A} < 0, \quad \frac{dq_2}{dm} = \frac{-\alpha_2 \beta_1 q_1}{A} < 0, \quad \frac{dX}{dm} = \frac{\beta_1 \beta_2 q_1}{A} > 0; \quad \text{and} \\
\frac{dq_1}{dn} &= \frac{-\alpha_1 \beta_2 q_2}{A} < 0, \quad \frac{dq_2}{dn} = \frac{-\alpha_2 \beta_1 q_2}{A} < 0, \quad \frac{dX}{dn} = \frac{\beta_1 \beta_2 q_2}{A} > 0.
\end{align*}
\]

Therefore, entry always decreases firms' individual outputs and increases total output, and the entry of an efficient firm increases total output more than the entry of an inefficient firm. Let
\[
\theta_i(m, n) = \frac{AC_i - C'_i}{P - AC_i}
\]
denote a measure of relative EOS for firm i, let
\[
\mu_i(m, n) = \frac{C'_2 - C'_1}{P - AC_i} > 0
\]
denote the difference between the efficient and inefficient firms' marginal costs relative to firm i's profit margin, and let \( \gamma_i \) and \( \gamma \) be given by
\[
\gamma_1(m, n) = \frac{n \alpha_2 \beta_1}{m \alpha_1 \beta_2 + n \alpha_2 \beta_1} > 0, \quad \gamma_2(m, n) = \frac{m \alpha_1 \beta_2}{m \alpha_1 \beta_2 + n \alpha_2 \beta_1} > 0, \quad \text{and} \\
\gamma(m, n) = \frac{\beta_1 \beta_2}{m \alpha_1 \beta_2 + n \alpha_2 \beta_1} > 0.
\]

In order to characterize \( m^*(n) \), we modify the definitions of weak, moderate and strong global EOS as follows:

**Definition 2.** Let \( n \geq 1 \) be fixed, let \( \gamma, \gamma_i, \theta_i, \) and \( \mu_i \) be given by (15-17), and assume \( \theta_i > 0 \) for all \( m \). (i) \( C_i(q_1) \) exhibits weak global EOS if \( 0 < \theta_i < (\gamma + \mu_i \gamma_i) \) for all \( m > 0 \); (ii) \( C_i(q_1) \) exhibits strong global EOS if \( \theta_i > (\gamma + \mu_i \gamma_i) \) for all \( m > 0 \); (iii) \( C_i(q_1) \) exhibits moderate global EOS if neither (i) nor (ii) holds.

**Proposition 2:** Let \( m^*(n) \) denote the optimal number of efficient firms in asymmetric oligopolies where the \( m \) efficient firms have costs, \( C_1(q) \), and the \( n \) inefficient firms have costs, \( C_2(q) \). Under A2, the following claims hold:

(i) the slope of the welfare function \( W(m, n) \) with respect to \( m \) satisfies:
\[
\partial W/\partial m > 0 \iff (\gamma + \mu_i \gamma_i - \theta_i) > 0;
\]
(ii) \( m^*(n) = \infty \) if \( C_1(q) \) exhibits global dis-EOS, constant EOS, or weak EOS;
(iii) \( m^*(n) = 0 \) or 1 if \( C_1(q) \) exhibits strong global EOS;
(iv) $m^*(n)$ uniquely solves

$$f = f(m, n) = \gamma + \mu_1 \gamma_1 - \theta_1 = 0.$$  

if $C_1(q)$ exhibits moderate global EOS or has a U-shaped AC curve, and if $W(m, n)$ has a unique peak in $m$.

Note that by rearranging the formula for $\partial W/\partial m$ in the proof, one can show that the number of efficient firms will be excessive under free entry --- $m^*(n)$ in part (iv) is less than the free-entry number of efficient firms given by $P(X)q = C_1(q_1)$.

Proposition 2 can also be interpreted in terms of the underlying geometry and the trade-off between economies of scale and economic surplus. As shown in part (iii) of Claim 2 in the appendix, $\partial f/\partial m < 0$ is sufficient for $m^*(n)$ to be the unique peak of $W(m, n)$ with respect to $m$. Claim 2 also provides a set of sufficient conditions for $\partial f/\partial m < 0$; these conditions hold under linearity assumptions and in most cases where demand and costs are assumed to be quadratic or loglinear.

Part (iii) of Proposition 2 is particularly striking. As shown in Example 3 below, $m^*(1) = 0$ in a duopoly implies that it will be more efficient for an efficient firm to exit an industry, leaving an inefficient firm with a monopoly, than for the two firms to compete.

Example 3: Let $p = a - Q$, $C_1(q) = d_1 + c_1q$, $C_2(q) = d_2 + c_2q$, $c_2 > c_1$, and assume $m = n = 1$, $a = 12$, $d_1 = c_1 = d_2 = 5$, and $c_2 = 5.1$. By $f(m, n) = f(1, 1) = -7.62 < 0$, one has $m^*(1) = 0$. Indeed, welfare increases from $W(1,1) = 11.47$ to $W(0,1) = 12.85$, even though the exit of the efficient firm leaves the inefficient firm with a pure monopoly.

---

7 Rearranging terms, $\partial W/\partial m$ in the proof becomes $W = \pi_1 + (P-\pi_1) m q_1 + (P-q_2) m q_2$. By (13), $\pi_1 > 0$. The conclusion follows from the arguments of footnote 3.

8 Of course, welfare will increase more if the inefficient firm exits. Note that strong global EOS for firm 1 implies that $\mu_2$ or $(C_2 - C_1)/(P-AC_1)$ cannot be too large. Equivalently, firm 2 must also have strong global EOS.

9 In the general case, let $\varepsilon = (c_2-c_1)/(a-c_1)$. Then, $\alpha_i = \beta_i = -1$, $\gamma_1 = n/(m+n)$, $\gamma_2 = m/(m+n)$, and $\gamma = 1/(m+n)$, $q_1 = (a-c_1)(1+n\varepsilon)/(m+n+1)$, $q_2 = (a-c_1)(1-(m+1)\varepsilon)/(m+n+1)$, $\mu_i = (c_2-c_1)q_i/(q_2^2-d_i)$, $\theta_i = d_i/(q_2^2-d_i)$, and (19) becomes: $f(m,n) = [1/(m+n)] + [n(c_2-c_1)q_1/(q_2^2-d_i)(m+n)] - [d_1/(q_2^2-d_i)] = 0$, or $(m+n+1)^3d_1 = (1+n\varepsilon)^2(a-c_1)^2 + n(m+n+1)(a-c_1)^2(1+n\varepsilon)$. This becomes the formula in Example 1 when $\varepsilon = 0$ and $k = (m+n)$. An explicit formula for the solution, $m^*(n)$, can be obtained using Maple or Mathematica.
Alternatively, if the two firms merged, welfare would increase even if the merged firm inherited the inefficient firm's cost structure. In these cases with strong global EOS, the gains in consumer surplus that result from an efficient firms' entry are insufficient to offset the lost economies of scale, and a merger of two firms will be welfare-increasing even if the merged firm is inefficient.

A comparison of (18) with (9) reveals that the existence of a cost differential, $\mu_1 > 0$, between the efficient and inefficient firms makes the entry of an efficient firm more likely to raise welfare than in the symmetric case. Note also that with global dis-EOS, constant EOS, or weak EOS, welfare always increases with the entry of additional efficient firms. Indeed, the entry of efficient firms would eventually force the inefficient firms to exit, since at some point, the number of efficient firms would become so large that the inefficient firms could no longer break even.

The impact of an increase in the number of inefficient firms on the optimal number of efficient firms, $m^*(n)$, has significant policy implications. The issue is truly interesting, however, only when there are moderate global EOS or the AC curve is U-shaped (i.e., when $1 \leq m^*(n) < \infty$). In the other cases, the optimal number of efficient firms is always either infinity or zero.

Proposition 3 below provides the main comparative static results. Let $f(m,n)$ be given by (19), and $k_1$ and $k_2$ be given by

$$
k_1 = k_1(m, n) = \frac{-\alpha_2 \beta_1^2 \beta_2 q_1}{(m \alpha_1 \beta_2 + n \alpha_2 \beta_1)^2} < 0,
$$

$$
k_2 = k_2(m, n) = \frac{\alpha_1 \beta_1 \beta_2 (\beta_2 q_2 + \alpha_2 \mu_1 X)}{(m \alpha_1 \beta_2 + n \alpha_2 \beta_1)^2} > 0.
$$

**Proposition 3:** Assume $C_1(q_1)$ exhibits moderate global EOS or a U-shaped AC curve. Under A2, and assuming $\partial f / \partial m = f_m < 0$ holds at $m^*(n)$, one has

$$
dm^*/dn > 0 \iff \frac{\partial f}{\partial n} = f_n = (k_1 + k_2 + q_2 f_m)/q_1 > 0.
$$

Note that $f_m < 0$, $k_1 < 0$, and $k_2 > 0$. In general, therefore, whether an increase in the number of inefficient firms increases or decreases the optimal number of efficient firms depends on the relative magnitudes of the three terms. Consider Example 3, with $n = 2$, $a =$
12, \( d_1 = d_2 = c_1 = 1, c_2 = 1.5 \). Using the formula provided in footnote 9, one can show \( m^*(2) = 3, f_m(3,2) = -0.1778, \) and \( f_n(3,2) = -0.083 \).\(^{10}\) Hence, an increase in \( n \) will reduce \( m^*(n) \). For instance, if \( n \) increases from 3 to 4, \( m^*(n) \) decreases from \( m^*(3) = 2.53 \) to \( m^*(4) = 2.07 \).

B. The optimal number of inefficient firms, \( n^*(m) \).

We now turn to the optimal number of inefficient firms for a given number of efficient firms. The welfare effects of entry by inefficient firms differ substantially from the effects of entry by efficient firms because the inefficient firms' EOS (or dis-EOS) are only partly determinative. What really matters is the size of the inefficient firms' EOS (or dis-EOS) relative to their cost disadvantage compared to the efficient firms. In order to characterize \( n^*(m) \), we modify the earlier definitions as below.

**Definition 3.** Given \( m \geq 1, \gamma, \gamma_i, \theta_i \), let \( \mu_i \) be given by (15-17). (I) Suppose \((\gamma - \mu_2 \gamma_2) > 0\). Then, the EOS concepts for \( C_2(q_2) \) are the same as in Definition 2: (I.a) \( C_2(q_2) \) exhibits weak global EOS if \( 0 < \theta_2 < (\gamma - \mu_2 \gamma_2) \) for all \( n \); (I.b) \( C_2(q_2) \) exhibits strong global EOS if \( \theta_2 > (\gamma - \mu_2 \gamma_2) \) for all \( n \); (I.c) \( C_2(q_2) \) exhibits moderate global EOS if neither (I.a) nor (I.b) holds.

(II) Suppose \((\gamma - \mu_2 \gamma_2) \leq 0\). (II.a) \( C_2(q_2) \) exhibits modified strong dis-EOS if \( \theta_2 < (\gamma - \mu_2 \gamma_2) \) for all \( n \); (II.b) \( C_2(q_2) \) exhibits modified weak dis-EOS if \( \theta_2 > (\gamma - \mu_2 \gamma_2) \) for all \( n \). (II.b) \( C_2(q_2) \) exhibits modified moderate dis-EOS if neither (II.a) nor (II.b) holds.

Parts (II.a)-(II.c) respectively cover the cases in which \( \theta_2 \) is a large negative number, \( \theta_2 \) is a small negative or non-negative number, and \( \theta_2 \) is a medium sized negative number. These new EOS concepts are used to determine the sign of \( \partial W/\partial n \). Intuitively, as the inefficient firms' cost disadvantage increases, it becomes less likely that an increase in their number will increase welfare, even if their costs exhibit dis-EOS, because as the number of inefficient firms increases the total output of the efficient firms decreases. Even though the inefficient firms might be able to produce at lower average costs, a smaller share of industry output will be produced by the more efficient firms.

\(^{10}\) One can confirm \( f_n(3,2) = -0.083 \) by using \( (21) \). By \( q_1(3,2) = 2, q_2(3,2) = 1.5, k_1(3,2) = -0.08, k_2(3,2) = 0.18, \) and \( f_m(3,2) = -0.1778 \), one can derive \( f_n = (k_1 + k_2 + q_2 f_m)/q_1 = -0.083 \).
Proposition 4: Let \( n^*(m) \) denote the optimal number of inefficient firms in the same market of Proposition 2, and let (I.a-I.c), (II.a-II.c) denote the assumptions in Definition 3. Under A2, the following claims hold:

(i) the slope of \( W(m, n) \) with respect to \( n \) satisfies:
\[
\frac{\partial W}{\partial n} > 0 \iff (\gamma - \mu_2 \gamma_2 - \theta_2) > 0;
\]
(ii) \( n^*(m) = \infty \) in each of the following two cases: (ii.1) \((\gamma - \mu_2 \gamma_2) > 0 \) and \( C_2(q_2) \) exhibits global dis-EOS, or constant EOS, or weak EOS; (ii.2) assumption (II.a) holds;
(iii) \( n^*(m) = 0 \) or 1 if either assumption (I.b) or assumption (II.b) holds;
(iv), \( n^*(m) \) uniquely solves
\[
g = g(m, n) = \gamma - \mu_2 \gamma_2 - \theta_2 = 0
\]
if \( W(m,n) \) has a unique peak in \( n \), and if either assumption (I.c) or assumption (II.c) holds or \( C_2(q) \) has a U-shaped AC curve.

As in the previous case, one can show that the number of inefficient firms will be excessive under free-entry. Owing to the negative sign of “\(-\mu_2 \gamma_2\)” in (22), it is less likely (than in the previous case of (18)) that the entry of an inefficient firm will increase welfare, and it is more difficult to find tractable conditions for \( \frac{\partial g}{\partial n} = g_n < 0 \), which would be sufficient for \( W \) to have a single peak in \( n \). Nonetheless, Claim 3 in the appendix provides a set of (somewhat less tractable) conditions sufficient for \( g_n < 0 \).

The “\(-\mu_2 \gamma_2\)” term in (22) also makes it extremely difficult to characterize the sign of \( \frac{dn^*/dm}{dm} \), which is determined by the signs of \( g_n = \frac{\partial g}{\partial n} \) and \( g_m = \frac{\partial g}{\partial m} \). A preliminary result on the sign of \( \frac{dn^*/dm}{dm} \) is provided below. Assume either (I.c) or (II.c) holds, and assume that \( A2 \) and \( g_n = \frac{\partial g}{\partial n} < 0 \) both hold. Then,
\[
\frac{dn^*/dm}{dm} > 0 \iff g_m > 0.
\]

Although (24) is less attractive than the condition on \( f \) as given in (21), it is simpler than \( \frac{\partial^2 W}{\partial n \partial m} > 0 \), and it may provide a shortcut for deriving comparative static results for particular empirical specifications. It can be proved by totally differentiating (23), and by using Lemma 1 in the appendix.

It is not difficult to solve equation (23) under the assumptions of Example 3.\footnote{Using footnote 9, (23) becomes \( g(m,n) = \frac{1}{m+n} - \left[ m(c_2-c_1)q_2/d_2 \right] \left[ \frac{(q_2^2-d_2)(m+n)}{d_2} \right] = 0 \).} The
“$-\mu_2\gamma_2$” term allows $d_2 = 0$ and thus constant economies of scale for the inefficient firms.

**Example 4:** Consider $p = a-Q$, $C_1(q) = d_1+c_1q$, $C_2(q) = c_2q$, $c_2 > c_1$. Then,

$$g(m,n) = \frac{1}{m+n} \cdot \frac{m(c_2-c_1)}{(m+n)q_2} = \frac{1}{m+n} \cdot \frac{m\epsilon}{(m+n)(1-(m+1)\epsilon)}.$$  

Since each of the two terms after the “$-$” sign increases when $m$ increases, $g_n < 0$. By (24) and $g_n = -[1-(2m+1)\epsilon]/[(1-(m+1)\epsilon)(m+n)^2] < 0$, $dn^*/dm < 0$. Note that $g(m,n) = 0$ implies

$$(25) \quad n^*(m) = n^*(m, \epsilon) = [1-\epsilon(m+1)^2]/(m\epsilon).$$

(recall that $n^*(m)$ is defined only if $\epsilon \leq 1/(m+1)^2$, because $q_2 \geq 0$ requires $\epsilon \leq 1/(m+1)$). Differentiating (25), one has $dn^*/dm = [(1-m^2)\epsilon-1]/(m\epsilon)^2 < 0$. Either way, an increase in $m$ decreases the optimal number of inefficient firms.

**C. The optimal market structure, $\mathbf{(m^*, n^*)}$.

Finally, we characterize the optimal market structure, $(m^*, n^*)$. Since this really just combines the results of Proposition 2 with those of Proposition 4, we only report the characterizations for an interior solution and for the main corner solutions\(^{12}\) (other corner solutions are reported in footnote 12). Recall that $f$ and $g$ are defined by (19) and (23).

**Proposition 5:** Under the conditions of Proposition 4, $(m^*, n^*)$ is given as follows:

(i) $(m^*, n^*) = (\infty, 0)$ if $C_1(q_1)$ exhibits global dis-EOS, constant EOS or weak EOS.

(ii) Assume $C_1(q_1)$ exhibits strong global EOS. Then, $(m^*, n^*) = (1, 0)$ or $(1, 1)$ if either assumption (I.b) or assumption (II.b) holds.

(iii) Assume $C_1(q_1)$ exhibits moderate global EOS or a U-shaped AC curve. Then, $(m^*, n^*)$ uniquely solves $f(m,n) = 0$ and $g(m,n) = 0$ if $W$ is strictly concave and if either

\(^{12}\) Let $n^*(m)$ and $m^*(n)$ be defined as before. There are two other corner solutions in part (ii): $(m^*, n^*) = (0, n^*(0))$ or $(1, n^*(1))$ if $g_m < 0$ and if either (I.c) or (II.c) holds (ii.1); and $(m^*, n^*) = (0, \infty)$ or $(1, \infty)$ in each of the following cases: A) (II.a) holds; B) ($\gamma_2\mu_2\gamma_1 > 0$ and (I.a) holds; and C) ($\gamma_2\mu_2\gamma_1 > 0$ and $C_2(q_2)$ exhibits global dis-, or constant or weak EOS.

On the other hand, the other solutions of part (iii) are: $(m^*, n^*) = (m^*(1), 1)$ or $(m^*(0), 0)$ if $f_m < 0$ and if either (I.b) or (II.b) holds; and $(m^*, n^*) = (m^*(\infty), \infty)$ if $f_m < 0$ and if any one of the following cases holds: A) (II.a) holds; B) ($\gamma_2\mu_2\gamma_1 > 0$ and (I.a) hold; and C) ($\gamma_2\mu_2\gamma_1 > 0$ and $C_2(q_2)$ exhibits global dis-, or constant or weak EOS.
(I.c) or (II.c) holds or $C_2(q)$ has a U-shaped average cost.

By part (i), the optimal market structure will entail an infinite number of efficient firms and no inefficient firms when the efficient firms' costs exhibit global dis-EOS, constant EOS, or weak global EOS -- essentially, when their cost structures allow them to achieve the greatest cost efficiencies at small output levels. In this case, competition between the efficient firms is preferable, and such competition would drive the inefficient firms out of the industry. By part (ii), the optimal structure will entail a single efficient firm and zero or a single inefficient firm, when $C_1(q_1)$ exhibits strong EOS and $C_2(q_2)$ exhibits strong EOS or modified weak dis-EOS (note that $\partial W/\partial n < 0$ in each case).

Part (iii) is in some respects the most interesting case. In that case, the optimal number of efficient and inefficient firms will both be finite, as illustrated in Example 5 below, where $C_1(q_1)$ and $C_2(q_2)$ both have U-shaped AC curves. An increase in the number of either the efficient or inefficient firms causes an increase in their average costs, but it also provides incremental economic surplus. Hence, there is a real trade-off between scale economies and incremental economic surplus. In fact, there was a similar trade-off in parts (iv) of Propositions 1, 2, and 4. Note that part (iii) assumes the global concavity of $W$. As shown in the proof, $W$ is locally strictly concave if $(f_{m}g_{n} - f_{n}g_{m}) > 0$, $f_{m} < 0$, and $g_{n} < 0$ all hold at $(m^*, n^*)$. It remains to be seen whether tractable, global second order conditions can be found.

**Example 5:** Consider Example 3 with $p = 12-Q$, $C_1(q) = 2+q$, $C_2(q) = 1.4+1.1q$. Using $f(m,n)$ and $g(m,n)$ in footnotes 9 and 11, and solving $f(m,n) = g(m,n) = 0$, one gets $m^* = 13.8$ and $n^* = 2.5$, which is a local maximal solution for $W(m,n)$. Indeed, one can verify numerically that $(m^*, n^*) = (13.8, 2.5)$ is a local maximum of $W$.

**V. Policy Implications**

One advantage of our approach is that it provides a basis for determining whether merger, entry, and exit policies are welfare-maximizing as opposed to merely welfare-increasing. Indeed, in a model without capacity constraints, a merger generating "no synergies" is equivalent to the exit of all but the most efficient member of the group of
merging firms.\textsuperscript{13} From this perspective, a lax merger policy may be viewed as a means of encouraging the exit of firms from an industry, and a strict merger policy may be viewed as a means of impeding the exit of firms from an industry. Moreover, the entry of an efficient or inefficient firm will have a welfare effect exactly opposite to the exit of an efficient or inefficient firm.

Our analysis has provided empirically estimable criteria for determining whether mergers, entries, and exits will be welfare-increasing or decreasing. In light of the foregoing discussion, we will focus on mergers. Of course, since our criteria assume that infinitesimal changes in the number of firms are possible, they may not provide accurate predictions of the welfare effects of particularly large mergers. This caveat aside, (9) shows that in symmetric oligopolies mergers will be welfare-increasing if and only if \(dW/dk<0\), or equivalently,

\[
(\gamma_0 - k\theta) < 0, 
\]

where \(\gamma_0\) and \(\theta\) are determined by the estimable parameters of demand and costs given in (6) and (7).

In asymmetric oligopolies, the welfare effects of mergers will depend on the types of firms involved. From Propositions 2 or from (18), a merger between two or more efficient firms will be welfare-increasing if and only if \(dW/dm<0\), or equivalently,

\[
(\gamma + \mu_1\gamma_1 - \theta_1) < 0, 
\]

where \(\gamma\), \(\mu_1\), \(\gamma_1\), and \(\theta_1\) are determined by the estimable parameters of demand and costs given in (15), (16), and (17). From Proposition 4 or from (22), a merger between an efficient firm and two or more inefficient firms, or between two or more efficient firms, will be welfare-increasing if and only if \(dW/dn<0\), or equivalently,

\[
(\gamma - \mu_2\gamma_2 - \theta_2) < 0, 
\]

where \(\gamma\), \(\mu_2\), \(\gamma_2\), and \(\theta_2\) are also determined by (15), (16), and (17).

Although the inefficiency of free entry is well known, our results precisely characterize a catalogue of cases in which entry will be welfare-decreasing. Indeed, our

\textsuperscript{13} A merger generates “no synergies” if, when a group of firms merge, the merged firm simply acquires the cost function of the most efficient member of the group. See Farrell and Shapiro (1990) for further discussion.
analysis suggests that if a merger of two efficient firms is welfare-increasing, the subsequent entry of a new efficient (or inefficient) firm will be welfare-decreasing. This is interesting because it is often assumed that mergers will have more positive welfare effects when industry conditions are highly favorable to new entry. Of course, this view may be grounded in a presumption that mergers will usually be welfare-decreasing, but in our view, the rationale for taking entry conditions into account should be placed on firmer ground than this.

VI. Conclusions

We have completely characterized the optimal market structure in a broad class of oligopolies in terms of firms' underlying cost structures. As the economies of scale change in a symmetric oligopoly from strong economies, to moderate economies, and to diseconomies, the optimal market structure will change from monopoly to oligopoly and eventually to perfectly competition. Similar results hold for asymmetric oligopolies with two types of firms. Although we have focused on an asymmetric Cournot model, we hope that our approach will be extended to other important industrial organization models.

Appendix

We first present a lemma on simplifying the second order sufficient conditions for a class of unconstrained maximization problems.

Lemma 1: Consider \( F(x), -\infty < x < \infty \), suppose \( F(x)'' \) is continuous and \( F(x)' = h(x)g(x) \) with \( g(x) > 0 \). (i) If \( h(x^*) = 0 \) and \( h(x)' < 0 \) for all \( x \), then \( x^* \) is the unique global peak point of \( F(x) \); (ii) if \( x^* \) is a maximum point, then \( F(x^*)'' < 0 \Leftrightarrow h(x^*)' < 0 \).

Proof of Lemma 1: Part (i). By \( F(x)' = h(x)g(x) \) and \( h(x^*) = 0 \), one has \( F(x^*)' = 0 \). By \( h(x)' < 0 \), \( F(x)' = h(x)g(x) \), \( g(x) > 0 \), one has \( F(x)' > 0 \) for \( x < x^* \), and \( < 0 \) for \( x > x^* \), which implies that \( x^* \) is the unique global peak point of \( F(x) \). Part (ii) follows from \( h(x^*) = 0 \), \( g(x^*) > 0 \), and \( F(x)'' = h(x)g(x)+h(x)g(x)' \). Q.E.D

Proof of Proposition 1: Part (i) By \( \frac{dw}{dk} = W' = (kq)' P-(kC(q))' \), one has

---

14 See the discussion of entry conditions in DOJ and FTC, 1992 Horizontal Merger Guidelines.
\[(D0) \quad W' = Pq - C \cdot (P - C') \cdot kq' = \pi(k) - Xq'P'.\]

Using (4), \(W'\) becomes

\[W' = q \cdot (P-AC) - (P-C') \cdot k \cdot \frac{\alpha q}{k\alpha + \beta} = \frac{\alpha q (P-AC)}{k\alpha + \beta} \cdot \frac{\beta}{\alpha} \cdot (AC-C').\]

Using (6) and (7), one gets (9).

Parts (ii-iii). By (9), the sign of \(dW/dk\) equals that of \((\gamma_0 - k\theta)\). Hence, \(dW/dk > 0\) holds in part (ii), \(dW/dk < 0\) holds in part (iii), this implies \(k^* = \infty\) and \(k^* = 1\), respectively.

Part (iv). By the assumptions, \(dW/dk = 0\) has a solution \(k^*\) given by (10), which is the unique peak point of \(W(k)\) or the unique optimal number of firms. \(Q.E.D\)

Define assumptions A3 and A4 as below:

**A3:** \[\beta^2 (P'' + q P''') \leq 2\alpha \beta P'' + \alpha^2 C'''.\]

**A4:** \[\gamma_0 (C' - AC + kq\theta P') \leq k [qC'' - (\theta + 2)(C'-AC)].\]

**Claim 1:** Under A1, the following four conclusions holds: (i) \(W(k)\) is globally increasing (decreasing) with global dis-EOS or constant EOS or weak EOS (strong EOS);

(ii) \(W(k)\) is initially increasing and eventually decreasing in \(k\) with either moderate global EOS or a U-shaped AC curve.

(iii) \(k^*\) given by (10) is the unique global maximum of \(W(k)\) if any of the following holds: (a) \(C(q)\) exhibits moderate global EOS, A3 and \(C'' \geq 0\) hold; (b) \(C(q)\) has a U-shaped AC curve, A3 and A4 hold;

(iv) the conditions of part (iii) are satisfied in the following classes of markets: (a) all linear markets; (b) quadratic markets with moderate EOS, \(2\alpha \leq \beta\), and \(P', C'' \geq 0\); (c) quadratic markets with a U-shaped AC curve, \(2\alpha \leq \beta\), \(P' \geq 0\), and A4; (d) loglinear markets with moderate EOS and \(2\alpha \leq \beta\); (e) loglinear markets with a U-shaped AC curve, \(2\alpha \leq \beta\), and A4; (f) all markets with moderate EOS, \(2\alpha \leq \beta\), \(P'' \geq 0\), \(C'' \geq 0\), \(C''' \leq 0\), and (g) all markets with a U-shaped AC curve, \(2\alpha \leq \beta\), \(P'' \geq 0\), \(P''' \leq 0\), \(C''' \geq 0\), and A4.

Note that part (i) of Claim 1 implies part (ii-iii) of Proposition 1. It is useful to note that A3 and A4 are sufficient conditions for \(k^*\) to be a unique peak, which is weaker than global concavity. As can be seen in the following proof, A3 guarantees \(d\gamma_0/dk \leq 0\), A4 (or moderate EOS) plus \(C'' \geq 0\) guarantees \(d[k\theta]/dk > 0\).
**Proof of Claim 1:** Part (i). It follows from (9).

Part (ii). Consider \(dW/dk = W'(k)\). By (6) and (9), one has \(W'(0) > 0\), which implies that \(W(k)\) is initially increasing. (9) can be reformulated as

\[
W' = \frac{q}{k + \gamma_0}[(P-AC)\gamma_0 - k(AC-C')] = \frac{q}{k + \gamma_0}[(P-AC)\gamma_0 + kq AC'].
\]

Since \((P-AC) \to 0\) as \(k \to \infty\), \((P - AC)\gamma_0 \to 0\) as \(k \to \infty\). On the other hand, individual supply \(q \to 0\), total supply \(kq \to \) a constant, and the slope of AC curve \(AC' \to \) a large (with U-shaped AC curve) or medium sized (with moderate EOS) negative number as \(k \to \infty\). Therefore, \((P-AC)\gamma_0 + kq AC')\) \to \) a negative number as \(k \to \infty\). Hence, \(W(k)\) will eventually be decreasing.

Part (iii). Using (4) and (5), the marginal impact of \(k\) on \(\alpha\) and \(\beta\) are:

\[
\alpha' = (P'' + qP''')X' + P''q' = \frac{q}{k\alpha + \beta}[(\beta - \alpha)P'' + q\beta P'''], \text{ and }
\beta' = P''X' - C'''q' = \frac{q}{k\alpha + \beta}((\beta P'' + \alpha C''').
\]

Using (6), one has

\[
\frac{d\gamma_0}{dk} = \frac{\beta'\alpha - \beta\alpha'}{\alpha^2} = \frac{q}{\alpha^2(k\alpha + \beta)}\{[2\alpha\beta P'' + \alpha^2 C'''] - [\beta^2(P'' + qP''')]\}.
\]

By (D3) and \((k\alpha + \beta) < 0\), one sees that assumption A3 implies

\[
\gamma'_0 = \frac{d\gamma_0}{dk} < 0.
\]

Now consider \(d(k\theta)/dk = (k\theta)'\). By differentiating with respect to \(k\), we have

\[
(AC-C')' = \frac{C'-AC}{q}C'q' = \frac{\alpha}{k\alpha + \beta}[qC''' - (C'-AC)], \text{ and }
(P-AC)' = P'X' - \frac{C'-AC}{q}q' = \frac{\beta qP' + \alpha(C'-AC)}{k\alpha + \beta}.
\]

Using (7) and the above expressions, we have

\[
(k\theta)' = \theta + k\theta' = \frac{AC-C'}{P-AC} + k\frac{(AC-C')(P-AC)-(AC-C')(P-AC)'}{(P-AC)^2} = \frac{\alpha}{(k\alpha + \beta)(P-AC)}\{k[qC'' - (\theta + 2)(C'-AC)] - \gamma_0 (C'-AC + kqP')\}.
\]

By the above expression, one sees that A4 implies

\[
(k\theta)' \geq 0.
\]

If there exist global moderate EOS, \(\theta \geq 0\) or \((C'-AC) \leq 0\). By (D5), \(C'' \geq 0\) and moderate
EOS guarantee \((k\theta')^{'} \geq 0\). Now, define
\[
(D7) \quad h(k) = (\gamma_0 - k\theta).
\]
By (D4), (D6), and (D7), and by the above discussions, \(h_k = dh/dk < 0\) holds for all \(k\). By Lemma 1, in both parts (iii.a) and (iii.b), \(k^*\) is the unique global maximal point of \(W(k)\). This completes the proof of part (iii).

Part (iv). In a linear market, there are global EOS, and \(P'' = C'' = P' = C' = 0\), therefore both \(A3\) and \(A4\) are satisfied, hence, part (a) holds; parts (b-c) can be similarly proved. In loglinear markets, one must have \(P' = -b_0 P < 0\), \(C' = b_1 C > 0\). Using these and the assumptions, one can prove parts (d-e). The proof for parts (f-g) are similar. \(Q.E.D\)

**Proof of Proposition 2:** Part (i) By \(W = \partial W/\partial m = PdX/dm - (MC_1 + nC_2)/dm\), one has \(W' = Pq_1c_1 + (P-C_1')mq_1 + (P-C_2')nq_2\). By (12-13) and (15-17), \(W'\) becomes
\[
W' = \frac{\gamma_1}{A} ((P-AC_1)\beta_1\beta_2 + (C_1'-AC_1)(mA\beta_2 + n\alpha_2\beta_1) + n\alpha_2\beta_1(C_2'-C_1'))
\]
\[
= (\gamma + \mu_i\gamma'_1 - \theta_1) \frac{q_1(P-AC_1)(m\alpha_1\beta_2 + n\alpha_2\beta_1)}{A}.
\]
Parts (ii-iii). By (18), the sign of \(\partial W/\partial m\) is the same as that of \(f(m, n) = (\gamma + \mu_i\gamma'_1 - \theta_1)\) given by (19). Hence, \(\partial W/\partial m > 0\) holds in part (ii), \(\partial W/\partial m < 0\) holds in part (iii), this implies \(m^*(n) = \infty\) and \(m^*(n) = 0\) or 1, respectively.

Part (iv). By the assumptions, \(\partial W/\partial m = 0\) has a solution \(m^*(n)\) given by (19), which is the unique peak point of \(W(m, n)\) in \(m\). \(Q.E.D\)

Define assumptions \(A5-A7\) as below:

**A5:** \(P'' \geq 0\), \(P''' \leq 0\), and \(2\alpha_i \leq \beta_i\), \(C_i'' \geq 0\), all \(i\).

**A6:** \(C_i'' \geq 0\), and \(P'\theta_1(\alpha_i + \beta_1) \geq \alpha_i C_i''\).

**A7:**

\[(i) \quad P'' \geq 0, \quad P''' \leq 0, \quad (q_1 - q_2)\beta_1\beta_2 P'' \leq 2\alpha_i \alpha_2 (C_2'' - C_1'');\]

\[(ii) \quad \alpha_i \beta_2^2 C_1'' \leq \alpha_2 \beta_2 C_2'';\]

\[(iii) \quad \beta_2 q_1(\alpha_i C_i'' - \beta_1 \mu_i P') \leq \alpha_2 \beta_1 q_1 C_2'' + \alpha_1 \beta_2 \mu_1 (C_1' - AC_1).\]

**Claim 2:** Assume \(A2\) holds. (i) \(W(m, n)\) is globally increasing (decreasing) in \(m\) with global dis- or constant-EOS or weak EOS (strong EOS); (ii) \(m^*(n)\) given by (19) is the
unique global maximum of \( W(m, n) \) in \( m \) if the above A5-A7 hold.

Similar to Claim 1, A5-A7 hold in large classes of markets with quadratic or loglinear demand and costs. In particular, they hold in all liner markets.

**Proof of Claim 2:** The proof of part (i) is similar to that in Claim 1.

Part (ii). Using (12) and (13), the marginal impact of \( m \) on \( \alpha_1 \) and \( \beta_1 \) are:

\[
\begin{align*}
\alpha'_1 &= \partial \alpha_1 / \partial m = (P'' + q_1 P''')X' + P'' q'_1 = \frac{B_2 q_1}{A} [(\beta_1 - \alpha_1)P'' + q_1 \beta_1 P'''], \\
\alpha'_2 &= \partial \alpha_2 / \partial m = (P'' + q_2 P''')X' + P'' q'_2 = \frac{B_1 q_1}{A} [(\beta_2 - \alpha_2)P'' + q_2 \beta_2 P'''], \\
\beta'_1 &= \partial \beta_1 / \partial m = P''X' - C'' q'_1 = \frac{B_2 q_1}{A} (\beta_1 P'' + \alpha_1 C'''), \text{ and} \\
\beta'_2 &= \partial \beta_2 / \partial m = P''X' - C'' q'_2 = \frac{B_1 q_1}{A} (\beta_2 P'' + \alpha_2 C''').
\end{align*}
\]

Using (D8), one has

\[
\begin{align*}
(\beta_1 \beta_2)' &= \frac{q_1}{A} [\beta_1 \beta_2 (\beta_1 + \beta_2) P'' + (\alpha_1 \beta_2^2 C''' + \alpha_2 \beta_1^2 C''')], \\
(\alpha_1 \beta_2)' &= \frac{q_1}{A} [\beta_2 (\beta_1 - \alpha_1) + \alpha_1 \beta_1] P'' + \beta_1 \beta_2 q_1 P''' + \alpha_1 \alpha_2 \beta_1 C''], \\
(\alpha_2 \beta_1)' &= \frac{q_1}{A} [\beta_1 (\beta_2 - \alpha_2) + \alpha_2 \beta_2] P'' + \beta_2 \beta_2 q_2 P''' + \alpha_1 \alpha_2 \beta_2 C''].
\end{align*}
\]

Using (17), (D9), and collecting the terms for \( P''', P'''', C''' \) and \( C'''', \) one has

\[
\gamma' = \frac{(\beta_1 \beta_2')}{m \alpha_1 \beta_2 + n \alpha_2 \beta_1} = \frac{(\beta_1 \beta_2') (m \alpha_1 \beta_2 + n \alpha_2 \beta_1) - \beta_1 \beta_2 (m \alpha_1 \beta_2 + n \alpha_2 \beta_1)^2}{(m \alpha_1 \beta_2 + n \alpha_2 \beta_1)^2} \\
= - \alpha_1 \beta_2 \beta_2^2 A + q_1 (\delta_1 + \delta_2 + \delta_3 + \delta_4),
\]

where

\[
\begin{align*}
\delta_1 &= \beta_1 \beta_2 [m \beta_2^2 (2 \alpha_1 - \beta_1) + n \beta_1^2 (2 \alpha_2 - \beta_2) P''], \\
\delta_2 &= - (\beta_1 \beta_2)^2 [m \beta_2 q_1 + \beta_1 q_2] P''', \\
\delta_3 &= m \alpha_2 \beta_2^3 C''', \text{ and } \delta_4 = n \alpha_2 \beta_2^3 C''''.
\end{align*}
\]

By (12) and (11), A5 implies \( \delta_i \leq 0 \) for all \( i \). By (10) and (D10), A5 guarantees that \( \gamma' < 0 \).

Now, consider the sign of \( \theta_1' = \partial \theta_1 / \partial m \). By (13), we have

\[
(AC_1 - C'_1)' = \left[ \frac{C'_1 - AC_1}{q_1} - C''_1 \right] q_1 = \frac{\alpha_1 \beta_2}{A} [q_1 C''_1 - (C'_1 - AC_1)], \text{ and}
\]
\[(P-AC_1)' = P'X' - \frac{C_1'-AC_1}{q_i} q_i' = \frac{\beta_1\beta_2 q_i P' + \alpha_1\beta_2(C_1'-AC_1)}{A}.\]

Using (15), the above expressions and \((1+\theta_1) = (P-C_1)/(P-AC_1) = - q_1 P'/(P-AC_1)\), one has

\[
\theta_i' = \frac{\alpha_1\beta_2 q_i C_i'' - \alpha_1\beta_2(1+\theta_1)(C_1'-AC_1) - \beta_1\beta_2 q_i \theta_1 P'}{(P-AC_1)A} = \frac{\beta_2 q_i}{(P-AC_1)A} [\alpha_1 C_i'' - (\alpha_1 + \beta_2)\theta_1 P'].
\]

By (12) and the above expression, A6 implies

\[(D13) \quad \theta_i' \geq 0.\]

Finally, consider the sign of \((\mu, \gamma_1)' = \partial(\mu, \gamma_1)/\partial m\). By (13), (16-17), we have

\[
(C_2'-C_1')' = C_2''q_i' - C_1''q_i' = \frac{q_1}{A} [\alpha_1\beta_2 C_i'' - \alpha_2\beta_1 C_i''], \text{ and}
\]

\[
\mu_1' = \left(\frac{(C_2'-C_1')'}{(P-AC_1)}\right)' = \frac{(C_2'-C_1')' - \mu_1(P-AC_1)'}{(P-AC_1)},
\]

\[
\gamma_1' = \left(\frac{n\alpha_2\beta_1}{m\alpha_1\beta_2 + n\alpha_2\beta_1}\right)' = \frac{\gamma_2(n\alpha_2\beta_1)' - \gamma_1(m\alpha_1\beta_2)'}{m\alpha_1\beta_2 + n\alpha_2\beta_1}.
\]

Using the above expressions and (D9), and collecting terms, one has

\[(D14) \quad \mu_1' = \frac{[\gamma_2(n\alpha_2\beta_1)' - \gamma_1(m\alpha_1\beta_2)'] (C_2'-C_1') + n\alpha_2\beta_1[(C_2'-C_1')' - \mu_1(P-AC_1)']}{(m\alpha_1\beta_2 + n\alpha_2\beta_1)(P-AC_1)} = \frac{-\gamma_1\alpha_1\beta_2(C_2'-C_1')A + (\Delta_1 + \Delta_2 + \Delta_3 + \Delta_4)}{(m\alpha_1\beta_2 + n\alpha_2\beta_1)(P-AC_1)A}, \text{ where}
\]

\[(D15) \quad \Delta_1 = q_1 P''\{\eta_2\beta_1[\beta_1(\beta_2-\alpha_2) + \alpha_2\beta_2] - \eta_1\beta_1[\beta_2(\beta_1-\alpha_1) + \alpha_1\beta_1]\}(C_2'-C_1')
\]

\[
= \frac{mn\beta_2\beta_1 q_1 P''(C_2'-C_1')}{m\alpha_1\beta_2 + n\alpha_2\beta_1} [\beta_1\beta_2(q_1-q_2)P'' + 2\alpha_1\alpha_2(C_i''-C_i)];
\]

\[
\Delta_2 = q_1 P''\{\eta_2\beta_2^2\beta_2 q_2 - \eta_1\beta_1\beta_2^2 q_1\}(C_2'-C_1') = \frac{mn\beta_1^2\beta_2^2(C_2'-C_1')}{m\alpha_1\beta_2 + n\alpha_2\beta_1} q_1(q_2-q_1)P''';
\]

\[
\Delta_3 = q_1 \{\eta_2\alpha_1\alpha_2\beta_2 C_i''' - \eta_1\alpha_1\alpha_2\beta_1 C_i''\}(C_2'-C_1')
\]

\[
= \frac{mn\alpha_1\alpha_2(C_2'-C_1')}{m\alpha_1\beta_2 + n\alpha_2\beta_1} (\alpha_1\beta_2^2 C_i''' - \alpha_2\beta_1^2 C_i''');
\]

\[
\Delta_4 = - n\alpha_2\beta_1^2\beta_2\mu_1 q_1 P' + n\alpha_2\beta_1 q_1[\alpha_1\beta_2 C_i'' - \alpha_2\beta_1 C_i''] - n\alpha_1\alpha_2\beta_1\beta_2\mu_1(C_i'-AC_1)
\]

22
\[= n\beta_1(\beta_1q_1(\alpha_1(\gamma_1-\beta_1\mu_1)P') - [\alpha_2\beta_1q_1C_1'' + \alpha_1\beta_2\mu_1(C_1''-\Delta_1)])].\]

By (12) and (D15), A7 implies \(\Delta_i \leq 0\) for all \(i\). By (D14), A7 implies

\[(D16) \quad (\mu, \gamma_1) < 0.\]

It follows from (19), (D12-13), and (D16) that \(\partial\theta_1/\partial m = f_m = \gamma' + (\mu, \gamma_1)' - \theta'_1 < 0\) holds for all \(m\). By Lemma 1, \(m^*(n)\) is the unique global maximal point of \(W(m, n)\). This completes the proof of part (iii).

**Q.E.D**

**Proof of Proposition 3:** Totally differentiating (19), we have

\[(D17) \quad \frac{dm^*}{dn} = -\frac{f_n}{f_m}.\]

Now, consider \(\partial\theta_1/\partial n = f_n\). Using (12) and (14), one has

\[(D18) \quad \partial\alpha_1/\partial n = (P''+q_1P''')X' + P''q_1' = \frac{\beta_2q_2}{A}[(\beta_1-\alpha_1)P'' + q_1\beta_1P'''],\]

\[(D19) \quad \partial(\beta_1\beta_2)/\partial n = \frac{q_2}{A}[\beta_1\beta_2(\beta_1+\beta_2)P'' + (\alpha_1\beta_2^2C_1''' + \alpha_2\beta_1^2C_1'')],\]

\[\partial(\alpha_1\beta_2)/\partial n = \frac{q_2}{A} \{\beta_2[\beta_2(\beta_1-\alpha_1) + \alpha_2\beta_2]P'' + \beta_1^2q_1P'' + \alpha_1\alpha_2\beta_2C_1''}]]], and

\[\partial(\alpha_2\beta_1)/\partial n = \frac{q_2}{A} \{\beta_2[\beta_2(\beta_2-\alpha_2) + \alpha_2\beta_2]P'' + \beta_2^2q_2P'' + \alpha_1\alpha_2\beta_2C_1''}]].\]

Note that (D18-19) are the same as (D8-9) except that outside \(q_1\) becomes \(q_2\). Using (17), (D19), and collecting the terms for \(P'', P''', C_1'''\) and \(C_2''\), one has

\[\gamma_n = \partial(\beta_1\beta_2)/\partial n = \frac{-\beta_1^2\beta_2A - q_2(\gamma_1 + \delta_1 + \delta_2 + \delta_3 + \delta_4)}{(m\alpha_1\beta_2 + n\alpha_2\beta_1)^2A},\]

where \(\delta_i\) are the same as in (D11). Using the formula for \(\gamma'\) in (D10), \(\gamma_n\) becomes

\[(D20) \quad \gamma_n = \frac{\alpha_1\beta_1\beta_2q_2 - \alpha_2\beta_1^2\beta_2q_1}{(m\alpha_1\beta_2 + n\alpha_2\beta_1)^2q_1} + \frac{q_2}{q_1}\gamma'.\]

Now, consider \(\partial\theta_1/\partial n\). By (14), we have
Proof of Proposition 4: Part (i)  By $W = \frac{\partial W}{\partial n} = PdX/dn - (mC_1 + nC_2)/dn$, one has $W' = Pq_2 - C_2 + (P - C_2') q_1 + (P - C_2') q_2$. By (12) and (14-17), $W'$ becomes

$$W' = \frac{q_2}{A} ((P - AC_2) \beta_1 \beta_2 + (C_2' - AC_2)(m\alpha_1 \beta_2 + n\alpha_2 \beta_1) - m\alpha_1 \beta_2 (C_2' - C_2'))$$

By (15) and the above expressions, by collecting terms and by using expressions for $\theta_1'$ preceding (D13), we have

(D21) $\frac{\partial \theta_1}{\partial n} = \frac{q_2}{q_1} \theta_1'$.

Finally, consider $\frac{\partial (\mu_1, \gamma_1)}{\partial n}$. By (13), (16-17), using the related expressions in the proof of Claim 2, we have

$$\frac{\partial (C_2' - C_2)}{\partial n} = \frac{q_2}{A} [\alpha_1 \beta_2 C_2'' - \alpha_2 \beta_1 C_2''] = \frac{q_2}{q_1} \frac{\partial (C_2' - C_2)}{\partial m}, \quad \frac{\partial \mu_1}{\partial n} = \frac{q_2}{q_1} \mu_1', \quad \text{and} \quad \frac{\partial \gamma_1}{\partial n} = \frac{1}{m\alpha_1 \beta_2 + n\alpha_2 \beta_1} [\alpha_2 \beta_1 \gamma_1 + \frac{q_2}{q_1} (m\gamma_1 - \theta_1')] = \frac{q_2}{q_1} (\mu_1, \gamma_1)'.

Using the above expressions and (D9), and collecting terms, one has

$$\frac{\partial (\mu_1, \gamma_1)}{\partial n} = \gamma_1 \alpha_2 \beta_1 (C_2' - C_2') A + (\Delta_1 + \Delta_2 + \Delta_3 + \Delta_4) q_2 / q_1,$$

where $\Delta_i$ are the same as in (D15). Using $(\mu_1, \gamma_1)'$ in (D14), the above expression becomes

(D22) $\frac{\partial (\mu_1, \gamma_1)}{\partial n} = \frac{(C_2' - C_2') (\alpha_2 \beta_1 \gamma_1 q_1 + \alpha_1 \beta_2 \gamma_1 q_2) / q_1}{(m\alpha_1 \beta_2 + n\alpha_2 \beta_1) (P - AC_2) A} + \frac{q_2}{q_1} (\mu_1, \gamma_1)'.

By (19), (D20-21), and (D16), one has

(D22) $f_n = \gamma_1 + \frac{\partial (\mu_1, \gamma_1)}{\partial n} - \frac{\partial \theta_1}{\partial n}$

$$= \frac{(C_2' - C_2') (\alpha_2 \beta_1 \gamma_1 q_1 + \alpha_1 \beta_2 \gamma_1 q_2)}{(m\alpha_1 \beta_2 + n\alpha_2 \beta_1) (P - AC_2) q_1} + \frac{\alpha_1 \beta_1 \beta_2 q_2 - \alpha_2 \beta_1 \beta_2 q_1}{(m\alpha_1 \beta_2 + n\alpha_2 \beta_1)^2 q_1} + \frac{q_2}{q_1} f_n = \frac{1}{q_1} (k_1 + k_2 + q_2 f_n).

By (D17), (D22) and $f_n < 0$, $dm*/dn > 0 \iff (k_1 + k_2 + q_2 f_n) > 0$.  \textbf{Q.E.D.}
\[ \frac{\partial W}{\partial n} = \left( \gamma - \mu_2 \gamma_2 - \theta_2 \right) \frac{q_2 (P - AC_2) (m \alpha_1 \beta_2 + n \alpha_2 \beta_1)}{A}. \]

Parts (ii-iii). By (22), the sign of \( \frac{\partial W}{\partial n} \) is the same as that of \( g(m, n) = \gamma - \mu_2 \gamma_2 - \theta_2 \) given by (23). Hence, \( \frac{\partial W}{\partial n} > 0 \) holds in part (ii), \( \frac{\partial W}{\partial n} < 0 \) holds in part (iii), this implies \( n^*(m) = \infty \) and \( n^*(m) = 0 \) or 1, respectively.

Part (iv). By the assumptions, \( \frac{\partial W}{\partial n} = 0 \) has a solution \( n^*(m) \) given by (23), which is the unique peak point of \( W(m, n) \) in \( n \).

Define assumptions A8-A9 as below:

**A8:** \[ \beta_1 \theta_2 \leq q_2 - \left( \frac{\alpha_1 \beta_1 \gamma_1 C''_1 + \alpha_2 \beta_1 \gamma_1 C''_2 - \beta_1 \gamma_2 \mu_2 P'}{q_2 (\alpha_2 + \beta_2) P' - \alpha_2 \gamma_2 (C'_2 - C'_1)} \right). \]

**A9:**

(i) \( m \alpha_1 \mu_2 \geq \beta_1, n \beta_1 \leq \mu_2 \alpha_1; \)

(ii) \( P'' \geq 0, P''' \leq 0, C''_i \geq 0, \) all \( i; \)

(ii) \( \mu_2 \beta_1 \beta_2 (q_1 - q_2) P'' \leq 2 \alpha_1 \alpha_2 \mu_2 (C''_2 - C''_1) + m \beta_2^2 (\beta_1 - 2 \alpha_1) + n \beta_1^2 (\beta_2 - 2 \alpha_2). \)

**Claim 3:** Assume A2 holds. \( n^*(m) \) given by (23) is the unique global maximum of \( W(m, n) \) in \( n \) if the above A8-A9 hold.

**Proof of Claim 3:** Using (23), we have

\[ g_n = (\gamma - \mu_2 \gamma_2 - \theta_2)' = (\gamma' - \mu_2 \gamma_2' - \theta_2'). \]

Using the expressions or techniques in the preceding proofs, one can show the following: A8 guarantees \( (\gamma' - \mu_2 \gamma_2') \leq 0, \) and A9 guarantees \( (\gamma' - \mu_2 \gamma_2') > 0. \) Therefore, the conclusion holds. As already mentioned, A8 and A9 are less tractable than the preceding assumptions, because the negative sign of the term “-\( \mu_2 \gamma_2 \)” in (22) makes it extremely difficult to derive tractable conditions, as seen from the complexity in the proof for Claim 2.

**Q.E.D**

**Proof of Proposition 5:** Parts (i)-(ii), (iii.2) and (iii.3) follow from Propositions 2 and 4, and part (iii.1) from the concavity of \( W. \) By (18) and (22), by Lemma 1, and by the first order conditions (19) and (23), \( (f_m g_n - f_n g_m) > 0, f_m < 0, \) and \( g_n < 0 \) will guarantee that \( W \) is strictly concave near \( (m^*, n^*). \)

**Q.E.D**
References


Figure 1. The relation between shapes of the AC curve and the optimal number of firms.