On Oligopolistic Markets for

Nonrenewable Natural Resources

Tracy R. Lewis
University of Arizona

Richard Schmalensee
Massachusetts Institute of Technology

WP #1052-79 February 1979

ABSTRACT

Noncooperative oligopoly behavior in nonrenewable resource markets is analyzed under stationary conditions assuming perfect information. The existence of Cournot-Nash equilbria in output paths is established under standard cost and demand assumptions, and a number of comparative dynamic results are obtained. If all suppliers have the same costs, for instance, and total reserves are fixed, either increasing the number of suppliers or equalizing their reserve holdings causes more rapid resource use. If suppliers' costs differ, it is shown that equilbrium involves inefficient production; high-cost reserves may even be exhausted before low-cost reserves.

On Oligopolistic Markets for Nonrenewable Natural Resources

Tracy R. Lewis and Richard Schmalensee

Considerable effort has been devoted to the theoretical analysis of the effects of supply-side market structure on the pattern of exploitation of nonrenewable resources. Weinstein and Zeckhauser (1975), Stiglitz (1976), and Sweeney (1977) have compared competitive and monopolistic exploitation policies, while Salant (1976), Lewis and Schmalensee (1978, 1979), and Gilbert (1978) have considered markets in which a cartel faces price-taking buyers and competitive fringe suppliers. But, aside from some remarks by Salant (1976, Appendix B) and some preliminary analysis by Lewis and Schmalensee (1979, Section 4), no theoretical analysis of oligopolistic market structures seems to have been done. This neglect can hardly be defended by arguing that such structures are of no empirical importance.[1]

The present essay describes some of the central properties of noncooperative equilibria (described more fully below) in oligopolistic markets for nonrenewable natural resources under simplifying assumptions. The stationary model analysed here seems the most natural intertemporal extension of the standard static Cournot model of oligopoly equilibrium: output level choice in the static framework is replaced here by choice of an output <u>path</u> over time. Neither the Cournot model nor the noncooperative model of this paper can be defended as universally descriptive. But both are tractable representations of oligopoly situations, both have at least a degree of coherence and plausibility, and both may yield useful insights and intuitions. Given the state of static oligopoly theory, we see no defensible, tractable alternative to the Cournot/Nash cooperative approach taken here, but we would not argue that this approach should be taken as the last word. In light of these considerations, it may or may not be reassuring that the general message of our analysis seems to be that natural analogs of many of the properties of the static Cournot model hold in the dynamic situation considered here.

We begin by presenting our assumptions and notation, describing some general properties of equilibria, and establishing the existence of equilibria. Section II provides some comparative dynamic properties of equilibria in which all sellers have the same costs. We show there, among other things, that increases in the number of sellers or equalization of their reserve holdings (total reserves constant) tends to speed up resource use. Section III deals with the case of unequal costs. Equilibria are characterized in somewhat more detail, and it is shown that the industry equilibrium output path is produced inefficiently. It is demonstrated by example that noncooperative equilibrium may entail the exhaustion of high-cost deposits of a resource before low-cost deposits.

I. Existence and Properties of Equilibrium

In order to see with some clarity how different oligopoly structures interact with the dynamics inherent in natural resource markets, we follow the papers cited above and make the sort of simplifying assumptions common in the literature begun by Hotelling (1931).[2] First, exploration does not occur in the world of this model, and all reserve stocks are known. There exist N "firms", which may be ordinary firms, nations, cartels, or other resource owners, with firm i's reserves at time zero equal to I_i. Let I be the sum of the I_i. Second, each firm i can extract its reserves (and transport them to market) at a constant unit cost $C_i \geq 0$; there are neither depletion effects nor capacity constraints. Let $q_i(t)$ be firm i's rate of production at time t, and let Q(t) be the sum of the $q_i(t)$.

Third, the model is partial equilibrium and stationary; consumers are passive price-takers and their behavior can be summarized by the smooth inverse demand function $P(t) = P[Q(t)]$, with $P' < 0$. There is a single, constant discount rate, r. Following Salant (1976) and others, the inverse demand curve is assumed to satisfy the following:

(D.1) $P(0) = F$, a finite constant greater than the largest of the C_i;

(D.2) Demand elasticity, $E = -(1/P')(P/Q)$, is strictly decreasing in Q (or increasing in P);

(D.3) There exists a price $\underline{P}$ such that $\underline{P}\{1-[1/NE(\underline{P})]\}$ equals the smallest of the C_i.

The "choke price", F, may be thought of as determined by the cost and other properties of some sort of backstop, with demand declining smoothly to zero as the price of the resource becomes high enough. If F does not exceed the largest of the C_i, the set of firms must re-defined, since price can never rise above F. Assumption (D.2) is sufficient but not necessary for industry marginal revenue to be declining in Q whenever it is positive; it is thus a strong convexity assumption. Assumption (D.3) plays much the same role here as Salant's (1976) assumption of a point of unit elasticity; it serves to rule out pathological and technically troublesome behavior of P and E as Q becomes large. It is a bounding assumption and is implied by the plausible assumption that demand is finite at zero price.

Finally, the nature of equilibrium must be clarified. We shall be concerned with solutions to a noncooperative game with full information, in which each firm chooses its time-path of output, $q_i(t)$, to maximize its net wealth subject to the reserve constraint, taking as fixed the output plans of all other producers. An equilibrium exists when no $q_i(t)$ can profitably be changed. The presumption of noncooperative behavior is strong, as is the assumption that firms know each other's output paths in advance. The latter comes close to the assumption of complete futures markets.[3] This sort of equilibrium has been explored widely enough in the natural resources context and elsewhere, however, to make our application of it here of some interest, if not to establish its decriptive value.

Under these assumptions, we can partially characterize an equilibrium reasonably simply, relying on the fact that each firm basically faces a relatively simple constrained maximization problem. Let $a(P) = -dP/dQ$, the absolute value of the slope of the market demand curve. Then at equilibrium there must exist a set of constants, λ_i, and functions, $q_i(t)$, such that for $i = 1,\ldots,N$ and $t \geq 0$,

(1a) $$e^{-rt}\{P(t) - q_i(t)a[P(t)] - C_i\} = \lambda_i, \text{ if } q_i(t) > 0;$$

(1b) $$e^{-rt}\{P(t) - C_i\} \leq \lambda_i, \text{ if } q_i(t) = 0$$

(1c) $$\lambda_i \geq 0; \quad I_i \geq \int_0^\infty q_i(t)dt; \quad \lambda_i[I_i - \int_0^\infty q_i(t)dt] = 0;$$

where $P(t)$ satisfies the demand constraint.

The second inequality in (1c) is the reserve constraint. The left-hand side of (1a) is the discounted net marginal revenue of firm i when $q_i(t)$ is positive and the other firms' outputs are taken as fixed. Demand assumption (D.2) implies that this is a decreasing function of $q_i(t)$, all else constant. It then follows from first principles that unless firm i's discounted net marginal revenue is equal at each instant when it is producing (and non-negative at all such instants) wealth can be increased by redistributing total production over time (or by reducing total production). The constant λ_i is then the marginal value, in units of wealth at time zero, of additional reserves. Condition (1b) simply says that if any firm rationally elects not to produce at any instant, discounted net marginal revenue at that instant with $q = 0$ cannot exceed discounted net marginal revenue at times when production optimally occurs. The third part of condition (1c) is a standard complementary slackness statement: reserves will be exhausted unless the marginal value of additional reserves is zero.

Suppose that at some instant m firms are producing, with $1 \leq m \leq N$. Differentiation of (1a) totally with respect to time and addition across producing firms yields an equation employed frequently in what follows:[4]

(2) $$\dot{P} = \frac{r[P(mE-1) - \bar{C}(mE)]}{E[(m+1)-Qa']}$$

where $\bar{C}$ is the average of the C_i for producing firms. The assumption that E rises with P can be seen to imply

(3) $$Qa' < (E+1)/E,$$

from which it follows that the denominator on the right of (2) is positive as long as $Em > 1$. For given P, $\dot{P}$ is easily seen to be decreasing in C and increasing in m.

Using assumptions that resemble (ii) and (iii) above, Frank and Quandt (1963) establish the existence of equilibrium in a static Cournot oligopoly model. The work of Roberts and Sonnenshein (1976) indicates that assumptions of this sort are almost necessary for existence, in the sense that equilibria apparently exist only by chance or in very special cases if they are violated. Since our model is related to the Cournot model studied by these authors, worries about non-existence of equilbria would seem to have more than usual force here. If our assumption did not suffice to establish existence, one might suspect them of either inconsistency or incompleteness. If equilibria cannot be shown to exist, it is hard to take the study of their properties very seriously.

As it happens, the assumptions made above suffice to ensure existence. Let T be the time at which all production of the resource ceases. Then the Appendix establishes[5]

Proposition 1 A noncooperative equilibrium satisfying conditions (1) exists and involves (a) $Q(t) > 0$ for $0 \leq t < T$, (b) $\lambda_i > 0$ for $i = 1,\ldots,N$, (c) T finite, (d) $I_i = \int_0^\infty q_i(t)\,dt$ for $i = 1,\ldots,N$, (e) $P(t)$ continuous for $0 \leq t \leq T$, and (f) $P(t) \geq \underline{P}$ for $0 \leq t \leq T$.

From part (a) of the Proposition, some firms are operating at all times. Adding condition (1a) across the operating firms, the numerator on the right of (2) is seen to equal $rm\bar{\lambda}Ee^{rt}$, where $\bar{\lambda}$ is the average of the λ_i of the m operating firms. By (b) of Proposition 1, this means that the numerator on the right of (2) is always positive. This in turn directly implies that $mE > 1$ at all times, which along with (3) implies that the denominator on the right of (2) is positive also. Thus $\dot{P}$ is always strictly increasing. Price starts out at some level not below $\underline{P}$ and rises continuously and monotonically until it reaches F in finite time, whereupon demand and the resource market vanish forever as the last remaining reserves are exhausted. Total output is positive and decreasing at all time, and it follows from (e) of Proposition 1 that Q(t) declines smoothly to zero as t approaches T.

Let us define market marginal revenue in the usual fashion:

(4) $$MR = P - aQ.$$

The arguments of the preceding paragraphs, along with (1a), (2), and (3), above, then serve to establish

Proposition 2 In equilibrium,

(a) $0 < \dot{P} \leq r(P-\bar{C})$, with strict inequality for finite m, and

(b) $\dot{MR} \geq r(MR-\bar{C})$, with strict inequality for $m > 1$.

Hotelling (1931) established that (price - marginal cost) grows at the rate of interest under competition in models of this sort, while (marginal revenue - marginal cost) grows at the same rate under monopoly. Proposition 2 indicates, as one might expect, that noncooperative oligopoly equilibria are intermediate between these polar cases on this dimension.

II. Equal Costs

In this Section, we make the simplifying assumption that $C_i = C$ for all i. Proposition 2 implies that (P-C) is always growing at a rate less than r in this case. This, along with conditions (1a) and (1b) imply that all firms with positive reserves will produce at time zero; as long as net marginal revenue when $q = 0$, $(P-C)$, grows at a rate less than the rate of interest, so that discounted net marginal revenue declines, it does not pay to postpone the start of production. On the other hand some firms may exhaust their reserves and cease producing before P reaches F. From these remarks and Proposition 1, it follows that equilibrium in the equal costs case can be characterized by a set of positive, finite exhaustion times $\{T_i\}$ and shadow values of reserves $\{\lambda_i\}$ such that for $i = 1,\ldots,N$,

$$\text{(5a)} \quad e^{-rt}\{P(t) - q_i(t)a[P(t)] - C\} = \lambda_i, \quad 0 \leq t \leq T_i,$$

$$\text{(5b)} \quad q_i(t) = 0, \ t > T_i,$$

$$\text{(5c)} \quad \int_0^{T_i} q_i(t)dt = I_i.$$

From Proposition 1, $P(t)$ is continuous. It is then immediate that for (5a) and (5b) to hold, the $q_i(t)$ must also be continuous at all times. If not, some firm's net marginal revenue would change discontinuously at some instant, and this cannot happen in equilibrium. (Such changes imply the desirability of shifting production either before or after the discontinuity until the jump is eliminated.) Each of the $q_i(t)$ must thus decline smoothly to zero as t approaches T_i.

It follows from (5a) that if firms i and j are producing at any instant and if $\lambda_i < \lambda_j$, then $q_i > q_j$. This holds even in the limit as $q_j \to 0$, so it must be that $T_i > T_j$. But it is then clear that firm i's total production exceeds firm j's, so that $I_i > I_j$. If firms have unequal reserves, it must be that the firm with the larger reserves has the smaller value of λ, produces more at each instant, and exhausts later.[6] The first of these suggests that firms with smaller reserves have greater exploration incentives, all else equal.

The preceding paragraph also indicates that if $I_i = I/N$ for all i, it must be that $\lambda_i = \lambda$ for all i. If firms have equal reserves in addition to equal costs, equilibrium is thus unique and is described by (5) with all i subscripts removed.[7]

For this symmetric case, a large number of comparative dynamic results can be established:

Proposition 3 If $C_i = C$ and $I_i = I/N$ for $i = 1,\ldots,N$, the noncooperative equilibrium is unique. Holding I constant, increases in N (a) increase $\dot{P}(P)$, (b) lower T, (c) reduce price at every level of reserves encountered, thus lowering $P(0)$ and leading to more rapid resource use, (d) raise , the marginal value of reserves to each firm, (e) lower the present value of total industry profits, and (f) raise discounted producers' plus consumers' surplus.

These comparative dynamic results are established in Lewis and Schmalensee (1979, Sect. 4) for the case C = 0, and only trivial modificaitons are required to generalize that proof.[8] In the symmetric case, Proposition 3 indicates that increases in N tend to make the market more nearly competitive, as in the standard static Cournot model.[9] Increases in N serve to speed up resource use, in the sense that the total time to exhaustion is reduced and total production is larger at any level of reserves. The latter assertion follows directly form a proof that with I constant, increases in N lower the initial price. Because reserves are drawn down faster with more firms, though, such increases will always make the price higher than it otherwise would have been at some future times. Part (d) of Proposition 3 indicates that more intense rivalry, even though it reduces prices and profits, may enhance exploration incentives. Parts (e) and (f) of that Proposition support the standard intuition that when rivalry intensifies, consumers' gains outweight producers' losses.

We now want to consider the implications of unequal reserves, still assuming cost equality. Let $P^s(t)$ and T^s be the price path and exhaustion time, respectively, in the symmetric case. Consider another market with the same demand curve and the same values of r, C, I, and N, but in which $I_i \neq I_j$ for some i and j. Let the price path and exhaustion date in this asymmetric case be $P^a(t)$ and T^a, respectively. We now show that $P^s(0) < P^a(0)$ and $T^s < T^a$.

From the discussion above, $T_i \neq T_j$ in the asymmetric case, so that the number of firms operating decreases over time; an instant before T^a, fewer than N firms are operating. In contrast, all N firms operate until T^s in the symmetric case. Let T_1^a be the time at which the first firm (the one with the smallest of the I_i) runs out of reserves in the asymmetric case.

Suppose $P^s(0) \geq P^a(0)$. Since the motion of price until time T_1^a in both cases is given by equation (2) with m = N and $\bar{C} = C$, it must be that $P^s(T_1^a) \geq P^a(T_1^a)$. But after T_1^a, fewer firms are operating in the asymmetric case, so that equation (2) implies $\dot{P}^s(P) > \dot{P}^a(P)$ for all relevant values of P. (Recall that the right-hand side of (2) is increasing in m, all else fixed.) It follows that $P^s(t) > P^a(t)$ for all $t > T_1^a$. Since $P^a(T^a) = P^s(T^s) = F$ and prices are monotone increasing, $T^s < T^a$. But this implies that total production in the asymmetric case exceeds that in the symmetric case and this is impossible. The supposition at the start of the paragraph must therefore be false. Thus $P^s(0) < P^a(0)$. If the price paths never cross, equality of total produciton will be violated. It follows from the arguments just above that they can only cross after T_1^a and that they can only cross once. After the crossing, $P^s(t) > P^a(t)$, so that $T^s < T^a$, as was to be shown. We have thus completed the proof of

<u>Proposition 4</u> If $C_i = C$ for $i = 1,\ldots,N$, and $I_i > I_j$ for some i and j, then (a) firm i produces more at each instant than firm j and ceases production later, and (b) P(0) and T are greater than if the I_i were all equal, so that resource use is less rapid.

There are many measures of concentration in the industrial organization literature, the most appealing of which have the property that they increase as the number of firms falls and as some measure of the inequality of firm sizes rises.[10] Propositions 3 and 4 taken together indicate that if firm size is defined as total reserves, changes of both these sorts tend to produce higher initial prices and thus higher prices at any given level of total reserves. These results are thus consistent with the usual notion that increases in concentration worsen market performance, with the problem here being excessive conservation of the resource.

Both these Propositions depend on the equal cost assumption, under which the level of reserves provides a natural, exogenous measure of firm size. The association between concentration and market performance in the static Cournot theory is of a rather different sort. Rader (1972, pp. 270-3), for instance, shows that the H index of concentration (equal to the sum of squared market shares) relates market demand elasticity to the percentage markup over (an average across the firms of) marginal cost. But in that model, market shares are endogenously determined as a function of marginal costs; H is thus logically endogenous. If all firms have equal costs in the Cournot structure, all will have equal market shares. In natural resource markets, even with equal costs, market shares vary over time if reserve holdings are unequal, so that reserve holdings provide the only reasonable basis for judgements about concentration. If costs differ, however, it seems apparent that any measure of concentration that relates to pricing behavior must reflect both costs and reserve holdings, but it is not yet clear to us how any such measure might be constructed.

III. Unequal costs

We now allow both costs and reserve levels to differ among firms. Let us order fims so that $C_1 \leq C_2 \leq \cdots \leq C_N$, with at least one of the inequalities strict. We first investigate whether or not firms with different C_i produce simultaneously in equilibrium.

If production were efficient, in the sense that the discounted cost of producing the actual total output stream, Q(t), were minimized, this could not occur. Suppose, for simplicity, that the C_i are all different. Then it is reasonably well-known that as long as the rate of interest is positive, efficiency requires exploiting only firm 1's reserves until they are exhausted at time T_1, then exploiting only firms 2's reserves until some later time T_2, and so on. It is never efficient to produce from any but the lowest cost currently available reserves; see Solow (1974) for a discussion. To see what drives this, suppose that extraction of a small amount, dq, of I_1 were postponed from t_1 to t_2, where $t_1 < T_1 < t_2$, with the same amount of I_2 extracted at t_1 rather than at t_2. The change in the present value of cost is then

$$dq\{e^{-rt_1}(C_2-C_1) + e^{-rt_2}(C_1-C_2)\} = dq\{e^{-rt_1}(C_2-C_1)[1-e^{-r(t_2-t_1)}]\} > 0.$$

Under monopoly, with all reserves in the hands of a single firm, discounted costs are minimized, and reserves with different costs are not simultaneously exploited. Under competition, with many small firms operating at each cost level, production is also efficient.

But under oligopoly, efficient exploitation is impossible because equilibrium must involve periods during which reserves with different costs are simultaneously exploited. To show this, let us first note that the $q_i(t)$ are continuous in any oligopoly equilibrium along with (from Proposition 1) P(t) and Q(t). Given continuity of P(t), continuity of the $q_i(t)$ follows directly from conditions (1a) and (1b). Now assume the contrary: suppose oligopoly equilibrium output path Q(t) is produced efficiently. Suppose there are N_1 firms with unit cost equal to C_1 and N_2 with unit cost equal to C_2. By the assumption of efficiency, there must be a time T_1 such that all firms in the first group cease production at T_1 and all firms in the second group start at that time. By the continuity of the $q_i(t)$, however, this must mean that Q(t) approaches zero as t approaches T_1 from either above or below. This means that price falls after T_1, and that is impossible. The only way that Q(t) can be kept smooth at T_1 with a shift of all production from firms in the low-cost group to those in the next higher cost group would be for some of the $q_i(t)$ to be discontinuous. But, given the continuity of P(t), conditions (1) rule out the discontinuous change in discounted marginal revenue that would result. We have thus established

<u>Proposition 5</u> If $C_i \neq C_j$ for some i and j, the discounted cost of producing the equilibrium total output stream, Q(t), is not minimized, because firms with different costs are simultaneously producing at some times.

This inefficiency result is the intertemporal analog of the standard finding of inefficiency in asymmetric static Cournot oligopoly; see Schmalensee (1976b) for a simple example. In that context, efficiency requires equating marginal costs of all plants operated at any time, but only in symmetric cases does noncooperative behavior produce this result. These inefficiency results indicate that to think of oligopolistic markets as intermediate between monopoly and competition in all respects is to overlook potentially important qualitative differences.

A number of important properties of equilibria when costs are unequal are given by Propositions 1 and 2. It turns out to be difficult to say much more. In particular, many of the properties of equilibria with equal costs do not generalize. In order to illustrate the variety of possible outcomes when costs differ, the rest of this section is devoted to establishing and discussing

<u>Proposition 6</u> If $C_i > C_j$ for some i and j, then in noncooperative equilibrium it is <u>possible</u> that (a) some firms do not begin production at time zero, (b) firm i's output is greater than firm j's at some times and less at others, and (c) production inefficiency is sufficiently severe that firm i's reserves are exhausted before firm j's.

We argue for the first possibility in general terms; the second and third are illustrated by a simple example.

Suppose there is initially only one firm on the market: firm 1 with unit cost C_1 and reserves I_1. Acting as a monopolist, that firm would choose price path $P^m(t)$. Now add a second firm to the market, with cost C_2 between $P^m(0)$ and F and reserves I_2 less than I_1. Let $P^d(t)$ be the noncooperative duopoly equilibrium price trajectory. It is clear that by making C_2 sufficiently close to F and I_2 sufficiently small relative to I_1, $P^d(0)$ can be made to be below C_2. But this means that firm 2 must wait until price has risen (a finite distance) above C_2 before it begins production. (Recall that $q_2(t)$ must then rise continuously from zero; it cannot jump to a positive value.) There is nothing pathological about such behavior, of course. Inefficiency arises only because both firms produce for some time, not because firm 2 waits.

At any instant, conditions (1a) and (1b) describe a Cournot equilibrium in the q_i, based on unit costs equal to $(C_i+\lambda_i e^{rt}) \equiv C_i'$. It follows from condition (1a) that $q_i > q_j$ if and only if $C_i' < C_j'$. When the C_i are all equal, the rankings of the C_i' cannot switch over time, and, as Proposition 4 establishes, the rankings of the q_i cannot change either. But if the C_i are not equal, it would appear possible for $(C_i' - C_j')$, and thus (q_i-q_j), to switch sign once before T for some i and j. To show that appearances are not deceiving in this case and to establish part (c) of Proposition 6, we turn to an example.

Suppose N = 2, $0 < C_1 < C_2 < F$, and demand is linear with $P(Q) = F - aQ$. With this demand structure, $\underline{P} = (F+2C_1)/3$. Let $P_0 = (F-C_1)/2$, and assume $C_2 < C_1 + P_0$. Then it is straightforward to show from conditions (1) that if $P \geq P_0$ and firm 1 [firm 2] is acting as a monopolist, $(P-C_2)[(P-C_1)]$ is growing at a rate less than r. This means that if $P_0 \leq P \leq F$ and either firm is producing, the other will rationally elect to produce also. If $F < 7C_1$, it follows that $P_0 < \underline{P}$, so that both firms will begin production at time zero because P will always exceed P_0 in equilibrium (from Proposition 1).[11]

At any time when both firms are producing, conditions (1a) can be written as

$$d_1 \equiv F - C_1 = \lambda_1 e^{rt} + 2aq_1 + aq_2,$$

$$d_2 \equiv F - C_2 = \lambda_2 e^{rt} + aq_1 + 2aq_2.$$

Let $s_2 = q_2/Q$, the high-cost firm's share of total output. By solving the two equations above for s_2, differentiating, and manipulating the result, one obtains

(8a) $\dot{s}_2 = [r(d_1 + d_2)/3aQ][s_2 - s_2^*]$, where

(8b) $s_2^* = (2d_2 - d_1)/(d_1 + d_2)$.

It is clear that $s_2 = s_2^*$ is an unstable node of (8a). If s_2 begins above s_2^*, s_2 increases and reaches unity in finite time; if $s_2(0) < s_2^*$, s_2 declines with time and reaches zero in finite time. Holding C_1 fixed, s_2^* is a decreasing function of C_2; it is easy to show that $C_2 > C_1$ implies $s_2^* < 1/2$.

In order to establish part (b) of Proposition 6, it must be shown that equilibria exist in which $s_2(t)$ takes on values greater than and less than 1/2. Suppose we have $s_2^* < I_2/I < 1/2$. Then if $s_2(0) \leq s_2^*$, firm 2's share of production at each instant will be less than its share of total reserves. Similarly, if $s_2(0) \geq 1/2$, q_2/Q exceeds I_2/I at all times. Thus, in order to meet the reserve constraint, $s_2(0)$ must be between s_2^* and 1/2, so that $s_2(t)$ is below 1/2 for a finite time and above 1/2 thereafter, and the proof is complete.

This example illustrates the difficulty, discussed at the end of Section II, of measuring concentration in a sensible way in this sort of market. In the preceding paragraph, firm 2 has higher costs and lower reserves then firm 1, but after a finite period of time its output comes to exceed that of firm 1. Indeed, since s_2 goes to unity, firm 2 is a monopolist for the latter part of the resource market's existence.

Part (c) of Proposition 6 is established by showing that equilibrium may involve $s_2^* > s_2(0)$, since then s_2 goes to zero in finite time, and this means that firm 2 exhausts its reserves while firm 1 is still producing at a positive rate. (We noted above that if this happens and if C_1 and C_2 are close enough together, condition (1b) is satisfied thereafter for firm 2.) Suppose $0 < I_2/I < s_2^*$. If $s_2(0)$ exceeds s_2^*, it follows that q_2/Q always exceeds I_2/I. But this is impossible, so that $s_2(0)$ must be less than s_2^* (and greater than I_2/I). Firm 2, the high-cost firm, then exhausts first, and the proof of Proposition 6 is complete.

Notice that in this example, as in general, the patterns described by (b) and (c) of that Proposition cannot occur when $C_1 = C_2$. Then $s_2^* = 1/2$, (b) is impossible because I_2/I cannot be between s_2^* and 1/2, and (c) is impossible because there is no high-cost firm. The example explored here suggests that these patterns are, at least, more likely when the higher-cost firm has the smaller reserves. Moreover, (b) may be more likely when reserve differences are moderate relative to cost differences, while (c) may require reserve differences that are substantial relative to cost differences. Whether these features of this example hold in general is not known, however. Finally, it is worth noting that while exhaustion of high-cost reserves before low-cost reserves is a particularly vivid manifestation of production inefficiency, we know of nothing that indicates that it signals especially severe waste in any quantitative sense.

Appendix: Proof of Proposition 1

It will first be shown that conditions (1) imply (a) - (d) directly. A fixed-point argument using the λ_i is then used to establish the existence of equilibrium, and (e) and (f) are proven in the course of that demonstration.

To prove (a), suppose the contrary. First, suppose that Q(t) jumps to zero at some t'. From (1a) and (1b) in the text, this would involve a discontinuous increase in the net marginal revenue of all firms. In the face of such a jump, all firms operating just before t' would have an incentive to delay some production. Thus Q(t) cannot jump to zero in equilibrium. Second, suppose that Q(t) declines smoothly to zero at some time t' before T. This means that Q(t") is positive for some t" > t'. Thus for at least one firm, conditions (1) imply

$$e^{-rt'}[F - C_i] \leq \lambda_i = e^{-rt''}[P - aq_i - C_i].$$

But since t" > t' and $(P - aq_i) < F$, this is clearly not possible. Thus Q(t) cannot jump to zero and cannot decline to zero before T. It follows that $Q(t) > 0$ for $0 \leq t < T$, as was to be shown, and that Q(t) declines smoothly to zero as t approaches T.

It follows also that equation (2) holds until time T, since $m \geq 1$. Adding conditions (1a) for operating firms, it is easily seen that the numerator on the right of (2) equals $rmE\bar{\lambda}e^{rt}$, where $\bar{\lambda}$ is the average of the λ_i for the operating firms. If $\bar{\lambda} = 0$ at any instant, price will be constant forever afterwards, as (1b) makes it clear that no entry will ever occur. But total production must then eventually come to exceed total reserves for all operating firms, so this is impossible. It is clear from (1b) that if $(C_i+\lambda_i e^{rt})$ exceeds F, firm i will not produce. If $\lambda_i > 0$, this must occur in finite time. If any of the λ_i are zero, it then follows that the corresponding firms will be alone in the market after finite time. But we just showed that this implied that those firms would produce constant q_i forever after, violating (1c), so that λ_i cannot be zero for any i in equilibrium, establishing (b) of the Proposition. The argument just above then establishes (c), and (d) follows from (1c) in the text.

Let λ be a vector of the N λ_i, and let Λ be the set of all λ such that $0 \leq \lambda_i \leq (F-C_i)$. This set is clearly compact and convex. We construct a continuous function γ mapping Λ into itself. By the Brouwer fixed point theorem, this function must have a fixed point, a vector λ^* such that $\lambda^* = \gamma(\lambda^*)$. The function γ is defined so that any such fixed point corresponds to an equilibrium satisfying conditions (1).

The first step is to obtain the path P(t) satisfying (1a) and (1b) for any $\lambda \epsilon \Lambda$. Let us define

$$\text{(A1)} \qquad Z_i(P,\lambda_i e^{rt}) = \max[0,\ P-C_i-\lambda_i e^{rt}], \qquad i = 1,\ldots,N.$$

These functions are continuous, but not differentiable when $P = C_i + \lambda_i e^{rt} \equiv P_i^c(t)$. Next define the continuous function $\phi(P,\lambda e^{rt})$:

$$\text{(A2)} \qquad \phi(P,\lambda e^{rt}) = \sum_{i=1}^{N} Z_i(P,\lambda_i e^{rt}) - a(P)Q(P).$$

It is apparent that ϕ is everywhere continuous, though it is not differentiable when $P = P_i^c$ for any i. Comparing equations (1) in the text, it is clear that Z_i is positive in equilibrium if and only if firm i is producing, and $\phi[P(t),\lambda e^{rt}] = 0$ for all t.

For given λ, P, and t, let m be the number of positive Z_i. If P = F, Q(F) = 0 implies that no matter what m is, $\phi(F,\lambda e^{rt}) \geq 0$. Now suppose $P = \underline{P}$, defined in assumption (D.3) in the text. If m = 0 at this price, $\phi = -aQ < 0$. If m > 0, (A1) and (A2) imply

$$\text{(A3)} \qquad \phi(P,\lambda e^{rt}) = m\{P[1-(1/mE)] - \bar{C} - \bar{\lambda} e^{rt}\},$$

where $\bar{C}$ and $\bar{\lambda}$ are the averages of the C_i and λ_i, respectively, for firms with positive Z_i. When $P = \underline{P}$,

$$P[1-(1/mE)] \leq \underline{P}\{1-[1/NE(\underline{P})]\} = \min_i C_i \leq \bar{C},$$

so that $\phi(\underline{P},\lambda e^{rt}) \leq 0$ by (A3). Because ϕ is continuous, there thus exists a scalar $P^e \varepsilon [\underline{P},F]$ such that $\phi[P^e,\lambda e^{rt}]=0$.

Now let us show that P^e thus defined is unique. As long as $P \neq P_i^c$ for all i, the derivative $\partial\phi/\partial P$ exists and satisfies

$$\text{(A4)} \qquad \partial\phi/\partial P = m + 1 - a'Q > m[1-(1/mE)] = [\phi+\bar{C}+\bar{\lambda}e^{rt}]/P,$$

from (3) in the text and (A3) above. It follows that if $\phi[P_i^c,\lambda e^{rt}] \neq 0$, $i = 1,\ldots,N$, then $\partial\phi/\partial P$ is positive whenever $\phi = 0$, and P^e is unique. Suppose, then that $\phi[P_i^c,\lambda e^{rt}] = 0$ for some i. From the first equality in (A4), this means that $\partial\phi/\partial P$ increases discontinuously at a point which $\phi = 0$ as m increases by one. If this derivative is positive in the limits as P approaches P_i^c from above and below, the continuity of ϕ establishes that ϕ is increasing for all P in a neighborhood of P^e, and uniqueness still holds. But $(\partial\phi/\partial P)$ cannot be negative in the limit as P approaches P_i^c from below, since (A4) indicates that ϕ cannot decline to zero as P increases. Since $(\partial\phi/\partial P)$ increases at P_i^c, it cannot be negative in the limit as P approaches P_i^c from above either, so that it must be positive in both limits. There is thus one and only one P that solves $\phi(P,\lambda e^{rt}) = 0$ for any $\lambda\varepsilon\Lambda$ and $t \geq 0$. This equation thus defines an implicit function $P^e(\lambda e^{rt})$, with range contained in the interval $[\underline{P},F]$.

We must next demonstrate that this function is continuous. Suppose the contrary: suppose there exist sequences $\{t^i\}$ and $\{\lambda^i\}$, such that $t^i \geq 0$ and $\lambda^i\varepsilon\Lambda$ for all i and such that the sequence $\{\lambda^i e^{rt^i}\}$ approaches $\lambda^0 e^{rt^0}$, with $\lim P^e[\lambda^i e^{rt^i}] = P^e[\lambda^0 e^{rt^0}] + K$, where $K \neq 0$. By construction, $\phi[P^e(\lambda^i e^{rt^i}), \lambda^i e^{rt^i}] = 0$ for all i, so that by the continuity of ϕ,

$$\lim \phi[P^e(\lambda^i e^{rt^i}),\lambda^i e^{rt^i}] = \phi[P^e(\lambda^0 e^{rt^0})+K,\lambda^0 e^{rt^0}] = 0.$$

But this implies that there are two P's that solve $\phi[P,\lambda^0 e^{rt^0}]=0$, and we showed above that this could not occur. The continuity of P^e is thus established by contradiction. Since P^e depends on λe^{rt}, it follows that in equilibrium, P(t) is continuous, establishing part (e) of the Proposition.

We can now define the function γ on which the existence proof turns. Pick a vector $\lambda \varepsilon \Lambda$. For all $t \geq 0$, compute $\hat{P}(t) = P^e(\lambda e^{rt})$. Then use (1a) in the text and (A1) to obtain for all i and t,

$$\hat{q}_i(t;\lambda) = Z_i[\hat{P}(t),\lambda_i e^{rt}]/a[\hat{P}(t)].$$

Integration then yields total resource production for each firm implied by λ and conditions (1a) and (1b):

$$\hat{I}_i(\lambda) = \int_0^\infty \hat{q}_i(t;\lambda)dt, \qquad i = 1,\ldots,N.$$

It follows from the continuity of P^e that the $\hat{I}_i$ are continuous. Next, define

$$\text{(A5)} \qquad D_i(\lambda) = \max[\hat{I}_i(\lambda)-I_i,0], \text{ and } E_i(\lambda) = \max[I_i-\hat{I}_i(\lambda),0], \qquad i = 1,\ldots,N.$$

The components of the vector-valued function γ are then given by

$$\text{(A6)} \qquad \gamma_i(\lambda) = \lambda_i + (F-C_i-\lambda_i)\{1-\exp[-D_i(\lambda)]\} - (\lambda_i)\{1-\exp[-E_i(\lambda)]\}, \qquad i = 1,\ldots,N.$$

This continuous function takes points in the compact, convex set Λ into points in that same set, so by the Brouwer theorem it has at least one fixed point, one vector $\lambda^* \varepsilon \Lambda$ such that $\gamma_i(\lambda^*) = \lambda_i^*$ for $i = 1,\ldots,N$.

By construction, conditions (1a) and (1b) are satisfied by the $\hat{q}_i(t;\lambda^*)$. To prove that fixed points of γ are equilibria, it remains only to show that conditions (1c) are satisfied. The first inequalities in (1c) are satisfied by the definition of Λ. Suppose that $D_i > 0$ at a fixed point for some i, so that $\hat{I}_i$ exceeds I_i and the inequality constraint on total production in (1c) is violated. Since E_i must then be zero for that i by (A5), it follows that $\lambda_i^* = F - C_i$. But since P can never exceed F, it follows from (A1) that Z_i is never positive, thus $\hat{q}_i$ is never positive, $\hat{I}_i$ is zero, and $\hat{I}_i$ cannot exceed I_i. This contradicts the assumption $D_i > 0$, so D_i must equal zero at a fixed point and the reserve inequality constraints in (1c) are satisfied at fixed points of γ. Now suppose that $E_i > 0$ at a fixed point. As above, this implies that total production falls short of total reserves for firm i with $\lambda_i^* = 0$. This would not violate (1c). But the argument for (b) of the Proposition shows that if $\lambda_i = 0$, firm i's total production as given by (1a) and (1b) in the text is infinite. Thus E_i cannot be positive at a fixed point. At such points, as at equilibria, $E_i = D_i = 0$; initial reserves are exactly used up by all firms. The proof of Proposition 1 is complete.

REFERENCES

Arrow, K.J., "Optimal Pricing, Use and Exploration of an Uncertain Natural Resource Stock," in P.T. Liu, ed. Dynamic Optimization and Applications to Economics (New York: Plenum Press, 1979).

Gilbert, R.J., "Dominant Firm Pricing Policy in a Market for an Exhaustible Resource," Bell Journal of Economics, 9 (Autumn 1978), 385-96.

Hart, P.E., "Moment Distributions in Economics: An Exposition," Journal of the Royal Statistical Society, Series A, 138 (1975), 423-34.

Hotelling, H., "The Economics of Exhaustible Resources," Journal of Political Economy, 39 (April 1931), 137-75.

Lewis, T.R., and R. Schmalensee, "Cartel and Oligopoly Pricing of Nonrenewable Natural Resources," in P.T. Liu, ed., Dynamic Optimization and Applications to Economics (New York: Plenum Press, 1979).

__________, and __________, "Cartel Deception in Nonrenewable Resource Markets," Working Paper 1010-78, Sloan School of Management, MIT, July 1978.

Loury, G.C., "The Optimal Exploitation of an Unknown Reserve," Review of Economic Studies, 45 (October 1978), 621-36.

Okuguchi, K., "Quasi-Competitiveness and Cournot Oligopoly," Review of Economic Studies, 40 (January 1973), 145-48.

Pindyck, R.S., "Cartel Pricing and the Structure of the World Bauxite Market," Bell Journal of Economics, 8 (Autumn 1977), 343-60.

__________, "Gains to Producers from the Cartelization of Exhaustible Resources," Review of Economics and Statistics, 60 (May 1978a), 238-51.

__________, "The Optimal Exploration and Production of Nonrenewable Resources," Journal of Political Economy, 86 (October 1978b), 841-62.

Rader, T., Theory of Microeconomics (New York: Academic Press, 1972).

Roberts, J., and H. Sonnenschein, "On the Existence of Cournot Equilibrium Without Concave Profit Functions," Journal of Economic Theory, 13 (1976), 112-17.

Salant, S.W., "Exhaustible Resources and Industrial Structure: A Nash-Cournot Approach to the World Oil Market," Journal of Political Economy, 84 (October 1976), 1079-93.

Schmalensee, R., "Is More Competition Necessarily Good?" Industrial Organization Review, 4 (1976b), 120-21.

__________, "Resource Exploitation Theory and the Behavior of the Oil Cartel," European Economic Review, 7 (April 1976a), 257-79.

Solow, R.M., "The Economics of Resources and the Resources of Economics," American Economic Review, 64 (May 1974), 1-14.

Stiglitz, J.E., "Monopoly and the Rate of Extraction of Exhaustible Resources," American Economic Review, 66 (September 1976), 655-61.

Sweeney, J., "Economics of Depletable Resources: Market Forces and Intertemporal Bias," Review of Economic Studies, 44 (February 1977), 125-42.

Weinstein, M., and R.J. Zeckhauser, "The Optimal Consumption of Depletable Natural Resources," Quarterly Journal of Economics, 89 (August 1975), 371-92.

FOOTNOTES

1. The markets considered by Pindyck (1977, 1978a), for instance, are either actually or potentially (in the event of cartel breakdown) oligopolistic.

2. Most of the assumptions that follow have been relaxed somewhere in the literature, though usually only one at a time and never in the context of an oligopoly model. In addition to the papers cited thus far, see Schmalensee (1976a), Loury (1978), Pindyck (1978b), Arrow (1979), and the references they cite.

3. The validity of this equilibrium concept under conditions of incomplete futures markets with a dominant firm has been explored by Lewis and Schmalensee (1978, 1979). Under that structure, the ability accurately to estimate the current output of the dominant firm may serve to support an equilibrium of this basic type.

4. Dots indicate time derivatives.

5. In unpublished work, Stephen Salant has independently developed a fixed-point argument like that used in the Appendix and employed it both to study the existence of equilibrium and to compute equilibria under more general assumptions than those made here.

6. Salant (1976, Appendix B) notes the latter two points in a special case of this model with only two firm types.

7. In fact, we have been able to demonstrate uniqueness without the assumption of equal reserves (but using the assumption of equal costs) by using the same sort of proof by contradiction employed to show uniqueness in a dominant extractor model in Lewis and Schmalensee (1979). As the proof is somewhat tedious and the result neither very surprising nor useful for any of our comparative statements, we do not present it.

8. By setting $q_i = Q/N$, conditions (5) can be defined for non-integer N. The proof of (e) and (f) in Lewis and Schmalensee (1979) relies on demand being sufficiently smooth that $Q(t)$ is a differentiable function of N for all t.

9. On the relations between competition and monopoly in natural resource markets under these demand assumptions, see Weinstein and Zeckhauser (1975), Stiglitz (1976), and Sweeney (1977). On the effects of changing N in static Cournot models, see Okuguchi (1973).

10. See, for instance, Hart (1975) and the references he cites.

11. This illustrates, in an inverse fashion, the discussion just below Proposition 6: if cost differences are minor, there is a presumption that all firms will begin production at the same time.