DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

NONCOOPERATIVE COLLUSION UNDER IMPERFECT
PRICE INFORMATION*

Edward J. Green
California Institute of Technology

Robert H. Porter
University of Minnesota

SOCIAL SCIENCE WORKING PAPER 367

January 1981
Revised December 1982

NONCOOPERATIVE COLLUSION UNDER IMPERFECT PRICE INFORMATION*

Edward J. Green
California Institute of Technology

and

Robert H. Porter
University of Minnesota and Bell Laboratories

November 1982

ABSTRACT

Recent work in game theory has shown that, in principle, it may be possible for firms in an industry to form a self-policing cartel to maximize their joints profits. This paper studies the applicability of that work to empirical industrial organization. A model of a noncooperatively supported cartel is presented, and the aspects of industry structure which would make such a cartel viable are discussed.

Mailing Address: Robert Porter
Bell Laboratories, Room 2C-123
600 Mountain Avenue
Murray Hill, NJ 07974

NONCOOPERATIVE COLLUSION UNDER IMPERFECT PRICE INFORMATION*

Edward J. Green
California Institute of Technology

and

Robert H. Porter
University of Minnesota and Bell Laboratories

November 1982

1. INTRODUCTION

Long-standing questions about how widespread is the occurrence of collusion in industries having several firms, and about the extent to which the performance of industries experiencing such collusion departs from the competitive norm, continue to provoke spirited debate. In this paper we offer a theory of collusive industry equilibrium which will provide a means of clarifying these questions.

In his classic paper "A Theory of Oligopoly" [14], George Stigler appealed to dynamic considerations to explain how apparently cooperative industry performance might result from noncooperative motives. According to this theory, the firms of an industry form a cartel, which is designed to enforce monopolistic conduct in a self-policing way. "Self-policing" means precisely that the agreed-upon conduct is noncooperatively viable and that it remains so over time.

Stigler's theory differs markedly from traditional oligopoly theories based on static equilibrium concepts (e.g., Cournot and Stackelberg). This difference is particularly striking in the case of an industry structure which is essentially immune from entry. The traditional theories would suggest that the performance of such an industry should be largely determined by its degree of concentration — the number of firms in the industry and their relative sizes — and by the extent to which substitute goods are available. In contrast, Stigler suggested that the greatest obstacle to collusion in the absence of entry would be what he characterized as "secret price cutting." By informally relating concentration and various other features of industry structure to the immunity of a cartel from entry and to its ability to deter inimical firm behavior, and by assuming that industry profitability reflects successful operation of a cartel, he justified the use of cross-industry regressions to test his theory.

The obvious interpretation of Stigler is that he made explicit a theory of oligopoly which implicitly conceived of a cartel as a "policeman" which with some frequency is required to punish destabilizing "offenses" of individual cartel members. The somewhat different interpretation of this paper is that Stigler had a view of cartel organization as an instance of an optimization problem: to design an institution which achieves an efficient equilibrium outcome subject to the constraint that agents in the institution behave noncooperatively. On this interpretation, the optimal cartel structure may be one which provides member firms with strong positive incentives which make collusive behavior attractive, rather than one which provides insufficient incentives and which severely punishes defecting firms after the fact.

In fact, two formulations of the cartel problem exist already which treat noncooperative collusion in a rigorous way. Osborne [8] proposes a reaction function equilibrium in which firms respond to changes in output by other firms in order to maintain their proportionate share of industry output. (See also the extensions of Spence [12,13].) Knowing that other firms will respond in this manner, each firm will realize that it does not pay to deviate from the collusive output level.

Friedman [3], on the other hand, outlines a strategy in which firms respond to suspected cheating, which they infer from a drop in the market price below the price that obtains when all firms produce at agreed-upon levels, by producing at Cournot levels thereafter. If future profit streams are discounted sufficiently slowly, than a firm would reduce the discounted value of its returns by failing to collude. Therefore, for all firms to adopt the collusive strategy would be a noncooperative equilibrium.

The trouble with these formulations, from an applied industrial organization viewpoint, is that

incentives in these equilibria are so perfect that the deterrent mechanisms are never observed. Then it may be difficult to infer from econometric time-series evidence whether the observed market data is the outcome of a quasicompetitive or collusive equilibrium (cf. T. Bresnahan [2]). The substance of the present contribution is that this perfection is an artifact of the certainty world in which these models are formulated. When the considerations of imperfect information, which played a decisive role in Stigler's theory, are reintroduced, optimal incentive structures may involve episodic recourse to the kind of short-run unprofitable conduct which would have been characterized as "price wars" or "punishment" previously.

Our argument has three parts. First, we frame a precise definition of collusion in terms of industry conduct. Second, we show that collusive conduct may, in a particular industry structure, result in a pattern of industry performance marked by recurrent episodes in which price and profit levels sharply decrease. Thus we reject the received view that performance of this type necessarily indicates an industry where firms are engaging in a sequence of abortive attempts to form a cartel. Since this opinion is often used as a basis to deny the need for intervention to promote competition in such industries (because the market purportedly is already withstanding the collusive assaults), our argument suggests the need to reexamine a widely-held assumption about policy.

Third, we point out that the distinctive character of the phenomenon just discussed and the necessary appearance of this phenomenon if collusion is to take place (given the particular industry structure in question) make it possible to draw clear-cut conclusions about the presence or absence of collusion in some specific industries on the basis of market data. This is a singular opportunity to learn about whether collusion does indeed exist in situations where it might plausibly occur, without having to face the many problems of interpretation surrounding the usual cross-industry tests of its extent.[1]

2. COLLUSION UNDER UNCERTAINTY

Collusive equilibria exhibiting stable performance may possibly characterize some industries. For instance, a market might be segmented geographically because firms have divided it. As long as this agreement was adhered to, each firm would be a monopolist within its area. Moreover, poaching by one firm in another's territory would be quickly and surely detected, and would invite retaliation. In that situation, no one would poach. All that would ever be "observed" is monopolistic conduct.[2]

Similarly, in an industry in which contracts are awarded by competitive bidding, a scheme to rotate winning bids might be perfectly enforceable. Each firm would act as a monopolist when its turn came, and would clearly see that bidding low out of turn would jeopardize a profitable arrangement. Again, only monopolistic conduct would ever be "observed."[3]

We will study a model in which demand fluctuations not directly observed by firms lead to unstable industry performance. Intuitively firms will act monopolistically while prices remain high, but they will revert for a while to Cournot behavior when prices fall. Specifically, it will be assumed that firms agree on a "trigger price" to which they compare the market price when they set their production.[4] Whenever the market price dips below the trigger price while they have been acting monopolistically, they will revert to Cournot behavior for some fixed amount of time before resuming monopolistic conduct.

Suppose that, at a given time, firms are supposed to be colluding (i.e., they expect one another to collude). If a firm produces more than its share of the monopoly output, its net return at that time will increase. However, by increasing the probability that the market price will fall below the trigger price, the firm incurs a greater risk that the industry will enter a reversionary episode during which profits will be low for everyone. For producing its monopolistic share to be the firms' noncooperatively optimal action, the marginal expected loss in future profits from possibly triggering a Cournot reversion must exactly balance (in terms of present discounted value) the marginal gain from over-producing. Of course, a reversionary episode will

sometimes occur without any firm defecting, simply because of low demand. Thus, over a long period, both Cournot behavior and collusive behavior will be observed at various times. In this respect, collusion under uncertainty differs markedly from the collusive equilibria under certainty discussed earlier. The fact that both monopolistic and Cournot performance are observed will make it possible to identify statistically the collusive equilibrium under uncertainty.

We now address the question of exactly what sort of industry our model might appropriately describe. Such an industry would have a structure possessing four features.

First, the industry is presumed to be stable over time. Temporal stability is required if the assumption that firms have rational expectations — an assumption which underlies the use of Nash equilibrium — is to be credible. On a more technical level, it justifies the use of stationary dynamic programming to characterize equilibrium.

Second, output quantity is assumed to be the only decision variable which firms can manipulate. In particular, firms should not be able to engage in product differentiation or have ability to divide their market regionally. With firm decisions so restricted, noncooperative cartel incentive schemes other than Cournot reversion are ruled out. In particular, even if one firm were suspected of violating a cartel agreement, other firms would have no way of isolating it and punishing it differentially.

Third, except for each firm's private knowledge about its present and past production, information about the industry and its environment is public. The Nash equilibrium assumption presupposes that firms have an accurate idea of their competitor's cost functions, for example. Also, for firms to coordinate effectively in keeping track of whether the industry is in a collusive or a reversionary state, they must all observe the realization of a common variable.

Fourth, the information which firms use to monitor whether the cartel is in a collusive or reversionary state must be imperfectly correlated with firms' conduct. Otherwise, if compliance

were optimal for firms in collusive periods, reversion would never occur. Price is not the only information variable which could be used for monitoring — price data with correction for a systematic demand component, or market share information, would also be subject to error. However, this assumption of imperfect information is incompatible with transactions in the industry being few and publicly announced, (e.g., with individual contracts being awarded on the basis of sealed-bid auctions) or with completely accurate and current market-share information being available to firms.

In our model firms monitor market price, which imperfectly reflects the output levels of other firms. We assume that the products of the firms are of homogeneous quality, and so they face a common market price. This structure is adopted for expositional ease. An environment in which firms monitored their own market share, which imperfectly reflected the price choices of other firms, would be more in the spirit of Stigler's paradigm.

We now give a formal description of collusion under uncertainty as a Nash equilibrium in contingent strategies. Consider an oligopoly of n firms which produce an undifferentiated product in a stationary and time separable environment. This environment is like that described in Friedman [3], except that demand is subject to multiplicative uncertainty. Specifically, i,j range over *firms* $1,...,n$. $\pi_i:R_+^2 \to R$ is the *return function* of i. $\pi_i(x_i,p)$ is i's net return from producing x_i units and selling at price p. β is the discount rate. Firms are risk neutral and maximize $E[\sum_{t=0}^{\infty} \beta^t \pi_i(x_{it},p_t)]$. Observed price $p_t = \theta_t p(\sum_{i=1}^{n} x_{it})$, where $p:R_+ \to R_+$. The random variables θ_t are i.i.d. with c.d.f. F having continuous density f. $E(\theta) = 1$. Each θ_t is a demand shock which firms cannot observe directly.[5]

A *contingent strategy* for firm i is an infinite sequence $s_i = (s_{i0}, s_{i1},...)$, where s_{i0} is a determinate initial output level x_{i0}, and $s_{it+1}:R_+^{t+1} \to R_+$ determines i's output level at time $t+1$ as a function of past prices by $s_{it+1}(p_0,...,p_t) = x_{it+1}$. The choice of domain reflects the assumption that firms do not observe rivals' production levels directly.

A strategy profile $(s_1,...s_n)$ determines recursively a stochastic process of prices, which in turn induces a probability distribution on the space of infinite sequences of prices. Expectation with respect to this distribution will be denoted by $E_{s_1...s_n}$.

A Nash equilibrium is a strategy profile $(s_1^*,...s_n^*)$ which satisfies

$$E_{s_1^*...s_i...s_n^*}[\sum_{t=0}^{\infty} \beta^t \pi_i(s_{it}(p_0,...,p_{t-1}),p_t)] \tag{1}$$

$$\leq E_{s_1^*...s_i^*...s_n^*}[\sum_{t=0}^{\infty} \beta^t \pi_i(s_{it}^*(p_0,...,p_{t-1}),p_t)]$$

for all firms i and feasible strategies s_i.

Now consider how the industry might produce at a monopolistic level most of the time (i.e., except during reversionary episodes) in a Nash equilibrium in trigger price strategies. Firms will initially produce their respective shares of this restricted industry output, and will continue to do so until the market price falls below a trigger price $\bar{p}$. Then they will produce Cournot outputs for the duration (we will specify this to be $T-1$ periods) of a reversionary episode, regardless of what happens to prices during this time. At the conclusion of the episode, T periods after the price drop, they will resume monopolistic production. This will continue until the next time that $p_t < \bar{p}$, and so forth.[6]

Formally, let $y = (y_1,...y_n)$ be a profile of restricted outputs, and let $z = (z_1,...,z_n)$ be a Cournot output profile. Choose a price level $\bar{p}$ and a length of time T. Define time t to be *normal* if (a) $t = 0$, or (b) $t-1$ was normal and $\bar{p} \leq p_{t-1}$, or (c) $t-T$ was normal and $p_{t-T} < \bar{p}$. Define t to be *reversionary* otherwise. Define strategies for firms by

$$x_{it} = \begin{cases} y_i & \text{if } t \text{ is normal} \\ z_i & \text{if } t \text{ is reversionary} \end{cases}$$

These are well-defined policy strategies.

Each firm faces a stationary two-state (normal and reversionary) $T-$ stage Markov dynamic programming problem. Its optimal policy is to produce z_i in reversionary periods, and to

produce some fixed quantity r in normal periods. Let $V_i(r)$ be the expected discounted present value of firm i if it sets $x_{it} = r$ in normal periods. Define

$$w_i = \sum_{j \neq i} y_j, \quad \gamma_i(r) = E_\theta \pi_i(r,\theta p(r + w_i)), \quad \delta_i = E_\theta \pi_i\left[z_i,\theta p\left[\sum_{j \leq n} z_j\right]\right]$$

In normal periods, i anticipates that the aggregate output of the other firms will be w_i, and so $\gamma_i(r)$ is the expected profit of then producing r. The expected profit in reversionary periods is δ_i. Let $Pr(\cdot)$ denote probability with respect to the distribution of θ. We assume that $\gamma_i(y_i) > \delta_i$ for each firm i. Then V_i satisfies the functional equation

$$V_i(r) = \gamma_i(r) + \beta Pr(\bar{p} \leq \theta p(r + w_i))V_i(r) \tag{2}$$

$$+ Pr(\theta p(r + w_i) < \bar{p})\left[\sum_{t=1}^{T-1} \beta^t \delta_i + \beta^T V_i(r)\right]$$

$Pr(\theta p(r + w_i) < \bar{p}) = F(\bar{p}/p(r + w_i))$, so (2) is equivalent to

$$V_i(r) = \frac{\gamma_i(r) + F(\bar{p}/p(r+w_i))\,((\beta-\beta^T)/(1-\beta))\delta_i}{1 - \beta + (\beta - \beta^T)F(\bar{p}/p(r + w_i))} \tag{3}$$

$$= \frac{\gamma_i(r) - \delta_i}{1 - \beta + (\beta - \beta^T)F(\bar{p}/p(r + w_i))} + \frac{\delta_i}{1 - \beta}$$

Thus the expected discounted present value of firm i equals what it would be in a Cournot environment, plus the single-period gain in returns to colluding, appropriately discounted. Inequality (1), the defining condition for Nash equilibrium, can now be rewritten

$$V_i(r) \leq V_i(y_i) \quad \text{for all } r \text{ and } i \tag{4}$$

The first-order condition for (4) is

$$V_i'(y_i) = 0 \quad \text{for all } i. \tag{5}$$

Using the fact that $(f/g)' = 0$ if and only if $f'g - fg' = 0$, (5) is equivalent to

$$0 = [1-\beta+(\beta-\beta^T)F(\bar{p}/p(\sum_{j\leq n}y_j))]\gamma_i'(y_i) \qquad (6)$$

$$+ (\beta-\beta^T)f(\bar{p}/p(\sum_{j\leq n}y_j))[\bar{p}\,p'(\sum_{j\leq n}y_j)/(p(\sum_{j\leq n}y_j))^2](\gamma_i(y_i) - \delta_i)$$

for all i.

Equation (6) states that the marginal return to a firm from increasing its production in normal periods ($\gamma_i'(y_i)$) must be offset exactly by the marginal increase in risk of suffering a loss in returns ($\gamma_i(y_i) - \delta_i$) by triggering a reversionary episode. When this condition holds for all firms, n differential constraints are placed on the n-dimensional vector y of restricted outputs in equilibrium. Thus, the assertion that an equilibrium which satisfies an additional constraint exists will require careful justification. In particular, the output profile which maximizes total returns to the industry may not be supportable in equilibrium.[7]

There are two related final observations about the formal model of collusion under uncertainty. First, no firm ever defects from the cartel. More precisely, no firm i has any private information that would lead it to assess its return function π_i more accurately than its competitors do. Thus, every competitor is able to figure out what i will do to maximize profits. The market price reveals information about demand only, and never leads i's competitors to revise their beliefs about how much i has produced. In equilibrium, the frequency of reversion from normal states will be given by $F(\bar{p}/p(\Sigma y_j))$.

Second, despite the fact that firms know that low prices reflect demand conditions rather than overproduction by competitors, it is rational for them to participate in reversionary episodes.[8] Basically, a reversionary episode is just a temporary switch to a Nash equilibrium in noncontingent strategies. It doesn't pay any firm to deviate unilaterally from its Nash strategy in this temporary situation, any more than it would if the industry were permanently a Cournot industry. It might be asked why Cournot equilibrium is appropriate at all. If firms know at a particular time that a low price has been observed in the past, and that the cartel has had a perfect record of monopolistic conduct, why do firms not disregard the price and continue to act

monopolistically? The answer is that everyone understands the incentive properties of equilibrium. If firms did not revert to Cournot behavior in response to low prices, equation (5) would not hold the rest of the time, so monopolistic behavior would cease to be individually optimal for firms.

We realize that the assumptions about industry structure are quite restrictive. We emphasize that the particular Nash equilibrium we are studying is not the only sort of Nash equilibrium which would be collusive according to the definition offered in this section, and that evidence that this particular Nash equilibrium operates in a specific industry is not the only evidence relevant to forming an opinion about the extent of collusion in various sectors of the economy. However, even though the direct applicability of our model is severely limited, it would be valuable to examine an industry for which it would be appropriate. We believe that the American rail freight industry in the 1880s was one example of an industry which satisfies our structural conditions quite well. Studies of that industry by Paul MacAvoy [7] and Thomas Ulen [15,16] have produced qualitative conclusions which are consistent with our model. Recent econometric work by Porter [10] (based on the extensive time series data collected by Ulen) strengthens these conclusions.

3. PRICE PROCESSES GENERATED BY COLLUSION

The equilibrium discussed in the preceding section is noteworthy because it reverses the traditional interpretations of a certain kind of industry price pattern. According to these traditional interpretations, an episode in which price drops sharply, remains low for some time, and then sharply rises again without there being an apparent cost or demand shock would indicate one of two possible events. The episode might be a symptom of the predatory reaction of incumbent firms to a threatened entry. Alternatively, it might signal (as in Stigler's theory) a breakdown of a cartel agreement followed by the reestablishment of the agreement. In either case, such evidence would indicate the fragility of collusion among the incumbents. Thus, in the formulation of policy, it has sometimes been argued that intervention to promote

competition would likely be redundant in markets where these episodes are already occurring.

In marked contrast, such episodes play an essential role in the maintenance of an ongoing scheme of collusive incentives in the model presented here. While the traditional views would predict the transience of collusion in a market marked by these episodes of price depression, and with the demise of collusion also the cessation of the price instability which it engendered, our model suggests that industries having certain structural characteristics (i.e., the four characteristics enumerated in the previous section) will exhibit price instability as a feature of a stable, time-stationary pattern of prices if its member firms are colluding. This observation raises the question of whether it is possible to estimate consistently, from the stochastic process of prices generated by a collusive equilibrium of the form described in Section 2, the trigger price $\bar{p}$ and the reversionary length T which determine that equilibrium. The answer to this question is affirmative. Moreover, there also exists an estimator which is computationally attractive and which has only a small asymptotic bias if the interval between price observations is short relative to both the length of reversionary episodes and the expected length of normal episodes — the situation which one would expect to encounter in an industry where collusion actually did confer significant market power on firms.

While a discussion of estimation *per se* lies beyond the scope of this paper, we characterize in an Appendix[9] the stochastic process of prices which arises in the equilibrium of the model presented in Section 2. It can be shown that any data series of prices may be treated as a sample path of a stationary ergodic process. This result provides a foundation for the study of asymptotic properties of estimation of the model, because it justifies the use of the ergodic theorem [1,Thm.6.28] to generalize the role which the law of large numbers plays in the estimation theory of independent processes.[10] (In particular, the existence of consistent estimators of $\bar{p}$ and T is a consequence of the ergodic theorem.) In the Appendix, the price process will be compared to an alternative process which is a Markov version of the well-known Bernoulli switching process (cf. [6]). In his study of the rail freight industry [10], Porter uses an estimator which is a maximum likelihood estimator for a Bernoulli process, rather than for

the process predicted by our theoretical model. It can be shown that a data series of prices may be regarded as a "contaminated sample path" of the alternative process, and the degree of contamination will be computed as a function of the true parameters of the equilibrium. This analysis provides a justification for the interpretation of Porter's empirical findings in terms of the present theory.

FOOTNOTES

1. These problems, involving both the nature of the cross-industry data and also the logical difficulties of using it as a basis for inference are described in the essays by J. McGee, H. Demsetz and L. Weiss in [4].

2. A referee has suggested that the U.S. steel industry employed such an enforcement device in the first half of this century.

3. For example, a "phases of the moon" system has been used to allocate low-bidding privileges in the high voltage switchgear industry. (See Scherer [11,Ch.6].)

4. It is logically possible for this agreement to be a tacit one which arises spontaneously. Nevertheless, in view of the relative complexity of the conduct to be specified by this particular equilibrium and of the need for close coordination among its participants, it seems natural to assume here that the equilibrium arises from an explicit agreement.

5. James Friedman has suggested to us that the variables θ_t might alternatively be specified to be a martingale, so that the prices p_t would also be a martingale. This property ought to be satisfied if the good is a durable, or if consumption is perfectly substitutable across times. We retain the i.i.d. specification which makes the analysis simpler, but acknowledge that it is restrictive.

6. For simplicity, we are considering only the simplest variant of a trigger price strategy. For example, firms might condition T on the amount by which $\bar{p}$ exceeds the observed market price.

7. In [9] it is shown that, for symmetric firms under imperfect price information, the output profile for normal periods which will maximize discounted industry profits in a noncooperative equilibrium in trigger price strategies is different from the profile which would be chosen if the industry were a monopoly. I.e., firms forego some profits in normal periods in order to reduce the frequency and duration of reversion needed to provide appropriate incentives, if $\bar{p}$ and T are chosen to maximize expected discounted profits subject to the incentive compatibility constraint (5).

8. To be precise, we argue here that the equilibrium is perfect or sequentially rational. A formal statement and proof of this assertion are given in [5].

9. This appendix is available from the authors upon request.

10. A stochastic process is ergodic if every event definable in terms of the tails of sample paths (e.g., the set of sample points having convergent paths) has probability zero or one. The ergodic theorem extends the strong law of large numbers to such processes.

REFERENCES

[1] Breiman, L., *Probability*, Addison-Wesley, Reading, 1968.

[2] Bresnahan, T., "The Oligopoly Solution Concept is Identified," *Economics Letters* 10 (1982): 87-92.

[3] Friedman, J. W., "A Non-cooperative Equilibrium for Supergames," *Review of Economic Studies* 28 (1971): 1-12.

[4] Goldschmid, H. J., H. M. Mann, and J. F. Weston (eds.), *Industrial Concentration: The New Learning*, Little, Brown and Co., Boston 1974.

[5] Green, E. J., "Non-cooperative Price Taking in Large Dynamic Markets," *Journal of Economic Theory* 22 (1980): 155-182.

[6] Kiefer, N. M., "A Note on Switching Regressions and Logistic Discrimination," *Econometrica* 48 (1980): 1065-1069.

[7] MacAvoy, P. W., *The Economic Effects of Regulation*, MIT Press, Cambridge, 1965.

[8] Osborne, D. K., "Cartel Problems," *American Economic Review* 66 (1976): 835-844.

[9] Porter, R. H., "Optimal Cartel Trigger-Price Strategies," *Journal of Economic Theory*, to appear.

[10] Porter, R. H., "A Study of Cartel Stability: The Joint Executive Committee 1880-1886," University of Minnesota, 1981.

[11] Scherer, F. M., *Industrial Market Structure and Economic Performance*, Second Ed. Rand McNally, Chicago, 1980.

[12] Spence, M., "Tacit Coordination and Imperfect Information," *Canadian Journal of Economics* 11 (1978): 490-505.

[13] Spence, M., "Efficient Collusion and Reaction Functions," *Canadian Journal of Economics* 11 (1978): 527-533.

[14] Stigler, G. J., "A Theory of Oligopoly," *Journal of Political Economy* 72 (1964): 44-61.

[15] Ulen, T. S., *Cartels and Regulation*, unpublished Ph.D. dissertation, Stanford University, 1978.

[16] Ulen, T. S., "The Market for Regulation: The ICC from 1887 to 1920," *American Economic Review, Papers and Proceedings* 70 (1980): 306-310.

APPENDIX

To begin, consider a very general definition of the class of stochastic processes which will be under consideration. The observed price process $\{X_t\}_{t \in \mathbb{N}}$ will be determined by two processes $\{Y_t\}_{t \in \mathbb{N}}$, the price process which would ensue if all periods were normal (i.e., if the industry were to produce the restricted output vector y at all times), and $\{Z_t\}_{t \in \mathbb{N}}$, the price process which would ensue if all periods were reversionary (i.e., if the industry were to operate in Cournot equilibrium at all times, producing the output vector z). Whether the observed price is drawn from the normal or the reversionary distribution is determined by a process $\{W_t\}_{t \in \mathbb{N}}$, which specifies whether the industry is in a normal or a reversionary state. Note that $\{X_t\}_{t \in \mathbb{N}}$ is the only component of the joint process $\{(W_t, X_t, Y_t, Z_t)\}_{t \in \mathbb{N}}$ which is observed.

Formally, define a *switching process* to be determined by a probability space (Ω, β, m), a state space S, a subset $N \subseteq S$, and four sequences of random variables $\{W\} = \{W_t : \Omega \to S\}_{t \in \mathbb{N}}$, $\{X\} = \{X_t : \Omega \to \mathbb{R}\}_{t \in \mathbb{N}}$, $\{Y\} = \{Y_t : \Omega \to \mathbb{R}\}_{t \in \mathbb{N}}$ and $\{Z\} = \{Z_t : \Omega \to \mathbb{R}\}_{t \in \mathbb{N}}$ which satisfy

$$\{Y\} \cup \{Z\} \text{ is a set of independent r.v.'s,} \tag{A1}$$

$$\{Y\} \text{ is identically distributed with c.d.f. G,} \tag{A2}$$

$$\{Z\} \text{ is identically distributed with c.d.f. H,} \tag{A3}$$

$$\{W\} \text{ is a Markov process with stationary transition probabilities,} \tag{A4}$$

$$\forall t \ S_t \in N \Rightarrow X_t = Y_t \ \text{w.p.1,} \tag{A5}$$

$$\forall t \ S_t \notin N \Rightarrow X_t = Z_t \ \text{w.p.1.} \tag{A6}$$

Note that the special case of a switching process usually studied occurs when $S = \{0,1\}$, $N = \{0\}$, and $\{W\}$ is a Bernoulli process which is independent of $\{Y\} \cup \{Z\}$.

In the case of a collusive price process, G and H are the c.d.f.'s of the normal and reversionary

price distributions, respectively. $S = \{0,...,T-1\}$ and $N = \{0\}$. (I.e., $W_t = 0$ signifies that the industry is in a normal period at time t.) The Markov process $\{W\}$ is defined recursively by starting with an arbitrary $W_0: \Omega \rightarrow S$, and then imposing

if $W_t(\omega) = 0$ and $Y_t(\omega) \geq \bar{p}$, then $W_{t+1}(\omega) = 0$, (A7)

if $W_t(\omega) = 0$ and $Y_t(\omega) < \bar{p}$, then $W_{t+1}(\omega) = 1$, (A8)

if $W_t(\omega) = k$, $1 \leq k < T - 1$, then $W_{t+1}(\omega) = k + 1$, (A9)

if $W_t(\omega) = T - 1$, then $W_{t+1}(\omega) = 0$. (A10)

The process $\{W\}$ defined by (A7)-(A10) is Markov with stationary transition probabilities because, by (A1) and (A2), $\{Y\}$ is i.i.d. The transition graph of $\{W\}$ is shown in Figure 1, in which each arrow is labeled with its transition probability.

The aim is to show that W_0 can be chosen in such a way that $\{X\}$ will be a stationary ergodic process. Conditions (A5) and (A6) show that $X_t(\omega)$ is a function of $(W_t(\omega), Y_t(\omega), Z_t(\omega))$, so by [1,Prop.6.32] it is sufficient to show that the joint process $\{W,Y,Z\}$ is ergodic. By [1,Thm.7.16], this process is ergodic if it is a stationary Markov process having a unique invariant distribution (i.e., a unique distribution such that, if W_1 is defined by (A7)-(A10), then $\{W_0,Y_0,Z_0\}$ and $\{W_1,Y_1,Z_1\}$ have identical joint distributions). This follows from [1,Thm.7.18], completing the proof that $\{X\}$ is ergodic.

In [10], a maximum likelihood estimator for a switching process somewhat different from (A1)-(A10) is used to study the pre-ICC rail freight cartel in the U.S. That process is obtained by replacing (A1), (A2) and (A7)-(A10) with

$\{Y\}$ is identically distributed with c.d.f. (A11)

$J(p) = \max[(1-G(\bar{p}))^{-1}(G(p)-G(\bar{p})),0]$.

(I.e., J is the distribution of p according to G, conditional on $p \geq \bar{p}$.)

$\{W\} \cup \{Y\} \cup \{Z\}$ is a set of independent random variables, (A12)

and

$\{W\}$ is a stationary Markov process having the transition probabilities specified by Figure 2.[*] (A13)

That is, this process is defined by relaxing the usual assumption that the switching process is Bernoulli, while retaining the assumption that it is independent of the underlying variables which determine the observed prices. Call the process defined by (A1)-(A10) the *price process*, and that defined by (A3)-(A5), (A11)-(A13) the *approximating process*.

The advantage of the approximating process over the price process is that it permits adaptation of much of the work which has been done on maximum-likelihood estimation of the Bernoulli switching process. In particular, it is possible both to compute the ML estimator economically and to appeal to theoretical results asserting its consistency and asymptotic normality. The crucial question raised by use of the approximating process is of how seriously misspecified it is as a model for data actually generated by the price process. We now address this question.

The basis for comparing the two processes is that, given a stationary price process $\{W,X,Y,Z\}$ with parameters $(\bar{p},T,G,H)$, a stationary approximating process $\{W',X',Y',Z'\}$ with the same parameters can be obtained by a kind of censoring. Looking at the matter from the opposite perspective, the sample paths of Z can be viewed as a contaminated (by reinsertion of the censored observations) version of the approximating process. The extent of the contamination is easily computable from the parameters of the process. If it is slight, and if the ML estimator is regarded as robust, then the ML estimator of the approximating process should also be considered to have small asymptotic bias as an estimator of the price process.

The approximating process $\{W',X',Y',Z'\}$ is defined from $\{W,X,Y,Z\}$ simply by censoring the

---

[*] Using extensive information including industry prices, macroeconomic variables, and firm-specific quantity data, Porter estimates the structural equations of a detailed industry model. His method may be viewed as an imposition of prior constraints on the reduced-form estimation described here.

triggering events (i.e., the events in which $W_t = 0$ and $Y_t < \bar{p}$). Formally, this is done by means of a sequence of stopping times $\{\tau_t: \Omega \to \mathbb{N}\}_{t \in \mathbb{N}}$. Define

$$\tau_0(\omega) = \begin{cases} 1 & \text{if } W_0(\omega) = 0 \text{ and } Y_0(\omega) < \bar{p} \\ 0 & \text{otherwise} \end{cases} \tag{A14}$$

and

$$\tau_{t+1}(\omega) = \begin{cases} \tau_t(\omega) + 2 & \text{if } W_{\tau_t(\omega)+1}(\omega) = 0 \text{ and } Y_{\tau_t(\omega)+1}(\omega) < \bar{p} \\ \tau_t(\omega) + 1 & \text{otherwise.} \end{cases} \tag{A15}$$

Then define

$$W_t'(\omega) = W_{\tau_t(\omega)}(\omega) \,, \tag{A16}$$

$$X_t'(\omega) = X_{\tau_t(\omega)}(\omega) \text{ and } Z_t'(\omega) = Z_{\tau_t(\omega)}(\omega) \,.$$

Finally, take a set $\{Y_t''\}_{t \in \mathbb{N}}$ which are identically distributed with c.d.f. J and such that $\{Y\} \cup \{Z\} \cup \{Y''\}$ is independent, and define

$$Y_t'(\omega) = \begin{cases} Y_{\tau_t(\omega)}(\omega) & \text{if } W_t'(\omega) = 0 \\ Y_t''(\omega) & \text{if } W_t'(\omega) > 0 \,. \end{cases} \tag{A17}$$

(N.B. The definition of the observed component $\{X'\}$ of the approximating process is the same whether $\{Y'\}$ is defined by (A17) or by $Y_t'(\omega) = Y_{\tau_t(\omega)}(\omega)$ for all ω. The reason for using (A17) is both to satisfy (A11) and to keep $\{W'\}$ and $\{Y'\}$ independent so that (A12) is satisfied. Under the simpler definition, (A15) would have introduced dependency between them.)

The effect of (A15) and (A16) is to continue to let a low realization of Y_t be the event which causes the state to change from zero to one, but to censor this event if it occurs. Thus the dependence of W_{t+1} on $\{W_t, Y_t\}$ in the price process is removed, and (A12) holds. By the strong Markov property [1,Prop.7.8], the censored process is a stationary Markov process, so (A13) holds. I.e., $\{W', X', Y', Z'\}$ is an approximating process with parameters $(\bar{p}, T, J, H)$.

It remains to calculate how much censoring of the price series $X(\omega)$ is required to construct the approximating series $X'(\omega)$. (Alternatively, how much contamination of $X'(\omega)$ is required to reconstruct $X(\omega)$?) Formally, what is $\lim_{t \to \infty} \dfrac{\tau_t(\omega) - t}{\tau_t(\omega)}$? If this quotient is close to zero for almost every ω, then the asymptotic bias of the approximating-process ML estimator applied to data generated by the price process should be small.

To calculate the quotient, first define $\sigma(W, Y) = 1$ if $W = 0$ and $Y < \bar{p}$, and $\sigma(W, Y) = 0$ otherwise. By (A14) and (A15), $\tau_t(\omega) = t + \sum_{u=0}^{\tau_t(\omega)} \sigma(W_u(\omega), Y_u(\omega))$, or

$$\frac{\tau_t(\omega) - t}{\tau_t(\omega)} = \frac{1}{\tau_t(\omega)} \sum_{u=0}^{\tau_t(\omega)} \sigma(W_u(\omega), Y_u(\omega)) \,. \tag{A18}$$

By the ergodic theorem,

$$\lim_{t \to \infty} \frac{1}{\tau_t(\omega)} \sum_{u=0}^{\tau_t(\omega)} \sigma(W_u(\omega), Y_u(\omega)) = m(\{W_0 = 0, Y_0 < \bar{p}\}) \, G(\bar{p}) \text{ a.s.} \tag{A19}$$

(Recall that m is the stationary measure on Ω.) Combining (A18) and (A19), and appealing to the fact that the stationarity of the price process forces W_0 and Y_0 to be independent, yields

$$\lim_{t \to \infty} \frac{\tau_t(\omega) - t}{\tau_t(\omega)} = m(\{W_0 = 0\}) G(\bar{p}) \text{ a.s.} \tag{A20}$$

The calculation of $m(\{W_0 = 0\})$ is an easy matter. For $1 \le k < T - 1$, by (A9) and stationarity, we have

$$m(\{W_0 = k\}) = m(\{W_1 = k + 1\}) = m(\{W_0 = k + 1\}) \,. \tag{A21}$$

Also, by (A7) and stationarity, we have

$$m(\{W_0 = 1\}) = m(\{W_1 = 1\}) = m(\{W_0 = 0\}) G(\bar{p}) \,. \tag{A22}$$

Since the probabilities of the states sum to unity, (A21) and (A22) yield

$$m(\{W_0 = 0\}) = [1 + (T - 1)G(\bar{p})]^{-1} \,. \tag{A23}$$

Thus, by (A22) and (A23),

$$\lim_{t \to \infty} \frac{\tau_t(\omega) - t}{\tau_t(\omega)} = G(\bar{p})[1 + (T-1)G(\bar{p})]^{-1} . \qquad (A24)$$

For example, consider a hypothetical industry in which a trade association disseminates weekly price data to its members. I.e., the appropriate interpretation of a period in the discrete-time model is one week. Suppose that the parameters of this industry were estimated using the ML estimator for the approximating process, with the results that $\hat{G}(\hat{p}) = .025$ and $\hat{T} = 11$. Since the expected duration of an episode of normal conduct is $(G(\bar{p}))^{-1}$, these estimates indicate that a reversionary episode occurs once a year on average, and lasts ten weeks. Thus there is (on average) one price observation a year (that being the observation of the price which triggers the reversionary episode), which would not be included if the approximating process were really generating the data. This is a contamination ratio of one in fifty, or $(.025)[1 + .25]^{-1}$ which is the expression which is obtained from (A24).

The ML estimator is computed by dividing the data into two sub-samples, one of which is presumed to have been drawn from distribution G and the other from H, and then estimating these distributions from the respective subsamples. If the "contaminating" observations were to comprise equal proportions of the two subsamples, then each subsample is being estimated with 2 percent contamination, and one might reasonably suppose the discrepancy between the price process and the approximating process to be rather small. If all of the "contaminating" observations were assigned to the subsample presumed to be generated by normal conduct, then this subsample would have 2-1/2 percent contamination, which still might reasonably be ignored. However, if the "contaminating" observations were all included in the subsample presumed to reflect reversionary conduct, then that subsample would have a 10 percent contamination level. In this worst case, it is easy to imagine that the observations actually drawn from the lower tail of G would seriously bias the estimation of H.

The parameter estimates for the example just given are approximately the same as those reported by Porter [10] for the rail freight industry. Thus, while the foregoing analysis is insufficiently precise to rule out the worst-case assumption concerning bias of his estimator

relative to the price process, it has shown that under more optimistic assumptions the bias would plausibly be slight. While we acknowledge that there is an inevitable element of subjective judgment in a situation such as this, we suggest that Porter's study provides presumptive evidence that the rail freight industry may have exemplified the kind of equilibrium which has been studied here.
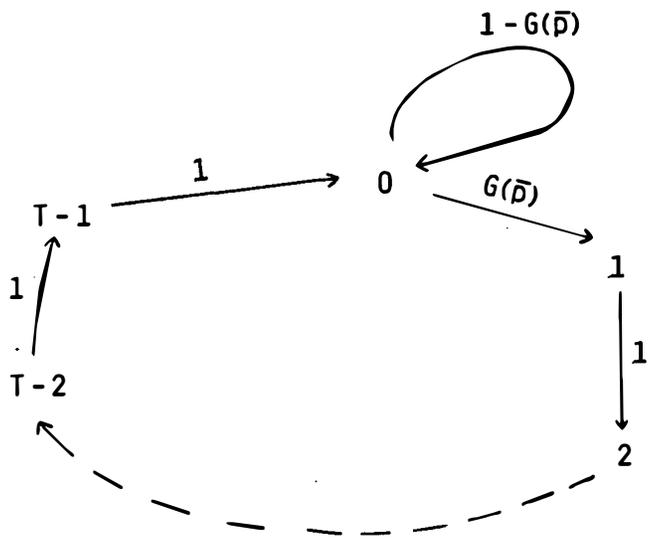
FIGURE 1

FIGURE 2