Short-Run Supply with Capacity Constraints

Author(s): Timothy F. Bresnahan and Valerie Y. Suslow

Source: *The Journal of Law & Economics*, Oct., 1989, Vol. 32, No. 2, Part 2, Empirical Approaches to Market Power: A Conference Sponsored by the University of Illinois and the Federal Trade Commission (Oct., 1989), pp. S11-S42

Published by: The University of Chicago Press for The Booth School of Business, University of Chicago and The University of Chicago Law School

Stable URL: http://www.jstor.com/stable/725587

SHORT-RUN SUPPLY WITH CAPACITY CONSTRAINTS*

TIMOTHY F. BRESNAHAN *and* *VALERIE Y. SUSLOW*
Stanford University *University of Michigan*

I. INTRODUCTION

THE behavior of industrial production and inventories over the business cycle has recently been a topic of considerable interest.[1] Though macroeconomic issues motivate the discussion, the hypotheses themselves are microeconomic. Short-run supply might reflect costs and expectations or it might be heavily influenced by strategic considerations. The strategic hypothesis is particularly attractive for those industries in which concentrated oligopolies use capital-intensive production facilities to meet cyclically sensitive demand.[2]

Thus, there is a particular set of industries whose macroeconomic importance suggests a close investigation of the microeconomics of their supply behavior. In this article, we take up the analysis of an important standard case: mass production achieved by flow processes in capital-intensive facilities, where capacity adjusts slowly and is long lived. In other words, we focus on that part of the economy whose elementary

---

* We would like to thank John Hause and participants at the Federal Trade Commission (FTC) Conference on Empirical Approaches to Market Power for helpful comments on an earlier version. We would also like to acknowledge the financial support of the FTC, of the Hoover Institution and the Center for Economic Policy Research at Stanford, and of the National Science Foundation under grant IRI-8814179.

[1] See Alan S. Blinder, Can the Production Smoothing Model of Inventory Behavior Be Saved? 101 Q. J. Econ. 431 (1986); Kenneth D. West, A Variance Bounds Test of Linear Quadratic Inventory Model, 94 J. Pol. Econ. 374 (1986); Julio Rotemberg & Garth Saloner, A Supergame-Theoretic Model of Price Wars during Booms, 76 Amer. Econ. Rev. 390 (1986).

[2] This form of industry structure is particularly prevalent in that part of the economy with cyclical employment, as pointed out in Richard L. Hall & Charles J. Hitch, Price Theory and Business Behavior, Oxford Economic Papers (1939).

[*Journal of Law & Economics*, vol. XXXII (October 1989)]
0022-2186/89/3202-0010$01.50

operations research (OR) in the short run is a linear program.[3] Much is known about individual industry production processes that could enhance the ordinary study of supply and cost. Our goal is to show that very elementary analysis of the details of the production process can have a substantial payoff in terms of the econometrics of short-run supply. Since many oligopolistic industries are highly cyclically sensitive, the benefits to understanding their quantity determination are many.

The short-run supply and short-run marginal cost (SRMC) we study are those of the North American primary aluminum industry over the period 1958–83. The aluminum production process is well documented in the public domain and has remarkably simple economics. To an excellent first approximation SRMC is linear out to capacity and vertical at capacity. Input requirements for labor, energy, and other materials are largely determined (in the short run) by well-documented features of techniques embodied in capital. Much of the modeling effort of this article goes into embedding this simple story into an econometric model of industry supply. We will focus on the equations determining industry production and shipments in the short run, here taken to be a calendar month. Only SRMC receives a structural econometric treatment. Issues such as dynamic demand, strategic interaction, and market power receive a reduced-form treatment.

II. Theory of Supply with Linear-Program Cost

In this section, we focus on the supply behavior of an industry whose short-run (SR) behavior with respect to production is as shown in Figure 1. The figure shows a short-run marginal cost, or SRMC, function appropriate for some capital-intensive, flow-process industries, including aluminum. The definition of "SR" in SRMC is the run in which one can take as fixed both the height of average variable cost (AVC)—determined by the level of technology—and the level of economically available capacity (EK)—determined by the amount and type of plant and equipment. Our approach will take these two cost determinants to be econometrically exogenous in monthly data. Further, we take capital to be completely fixed and all other factors to be completely variable.[4] In the short run,

---

[3] Though the econometrics of supply has not yet made this leap, the economic analysis of supply and cost made it long ago. See, for example, William J. Baumol, Economic Theory and Operations Analysis (1961).

[4] This assumption works rather better for aluminum—which appears to engage in no labor hoarding—than for many industries. When aluminum plants are not producing, they are manned by skeleton crews that maintain the capital equipment. Restarting production involves bringing back the crews (conversation with Mr. Jim Whitchurch, Mgr., Operations Analysis, Alcoa Smelting Div.).

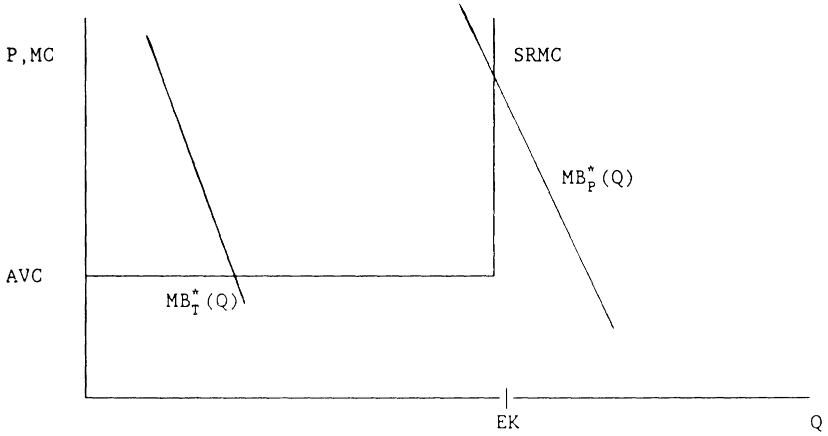

FIGURE 1.—SRMC and private value: quantity determination

AVC is thus given by the input requirements for variable factors. The price of capital never enters SRMC directly. At the peak, the height of marginal cost (MC) is determined by the shadow value of the constraint implied by not having more capacity; at the trough, capital's contribution to MC is zero.

In thinking about the downward-sloping curve in Figure 1, we will need to be robust to several different considerations. The first is production of a good that is potentially storable. The ability to hold output inventory drives a wedge between production and shipments in any given period. Thus, short-run supply describes the selection of both quantity produced (Q) and quantity shipped (S). We also need to be aware of the possibility that demand is linked over time, either because of the durable nature of the good or because of strategic interaction among firms. In either case, expectations will matter, and perceived marginal revenue (the industry acts as if it sets perceived marginal revenue equal to marginal cost) can become a function of present, past, and future variables. Having said all of the above, we hasten to add that our approach to the downward-sloping curve in Figure 1 will emphasize the reduced-form equations rather than any structure.

There are three fundamental endogenous variables in this setting: price, quantity produced, and quantity sold. Merely reading the figure, we see that the equilibrium condition for the industry during its capacity-unconstrained periods takes the general form

$$\mathrm{SRMC}(Q) = \mathrm{MB}(Q,S,P), \tag{1}$$

where MB is an unspecified marginal benefit function. What we will do is solve this in general for the reduced-form equation determining Q. We will not attempt to put an interpretation on MB(·) in equation (1) that would permit much in the way of inferences about market power, dynamics, and so on. Nor will we explicitly model the aggregation of firms to the industry that lies behind (1). Instead, we will emphasize the implications of a general, unstructured form of equation (1) when confronted with right-angle SRMC.

The first step is to solve out S and P in equation (1) so that the reduced form for Q is revealed. That is straightforward in some familiar examples. In the static perfect competition case, $MB(\cdot) = P$, and the structural system is

$$\begin{aligned} \mathrm{SRMC}(Q,W) &= P, \\ P &= D^{-1}(Q,Y), \end{aligned}$$

where W is a vector of cost factors in AVC and Y is a vector of demand-side exogenous variables. Thus MB*(·), the equilibrium MB, is $D^{-1}(Q,Y)$, and the reduced-form equation for desired production is

$$\begin{aligned} \mathrm{SRMC}(Q_1^*,W) &= D^{-1}(Q_1^*,Y), \\ => \qquad Q_1^* &= Q_1^*(W,Y). \end{aligned}$$

This defines the reduced-form equation for the capacity-unconstrained regime. When the capacity constraint binds, production is determined by the level of economically available capacity, EK (measured with some error which we will describe in detail in the next section).

If we let $X = (W, Y)$ and β_1 be unknown parameters, this leads to a two-regime model:
regime 1:

$$Q_1^* = X\beta_1, \tag{2}$$

regime 2:

$$Q^c = \mathrm{EK}, \tag{3}$$

and

$$Q = \min[Q_1^*,Q^c],$$

where Q_1^* is the unconstrained amount of production (if AVC were extended out beyond EK), Q^c is the quantity produced when the capacity constraint is binding, and MR is marginal revenue.

In the static monopoly case, MB(·) = MR, and we get

$$\begin{aligned} \mathrm{SRMC}(Q,W) &= \mathrm{MR}(P,Q,Y), \\ P &= D^{-1}(Q,Y). \end{aligned}$$

The reduced form in the unconstrained regime is

$$\mathrm{SRMC}(Q_1^*, W) = D^{-1}(Q_1^*, Y) + Q_1^* D_Q^{-1}(Q_1^*, Y).$$

Note that this takes the same form, $Q_1^* = X\beta_1$, as equation (2). The coefficients will, of course, be different. The constrained-regime specification follows, as above.

In between competition and monopoly are all of the possible oligopoly behaviors. We assume that they can be written in the form SRMC($\cdot$) = PMR($\cdot$), that is, marginal cost equals perceived marginal revenue.[5]

In the dynamic case (allowing for inventory holding), the gap between production and sales gives the industry two short-run supply functions. These are linked by the fact that selling one more unit and producing one less unit have the same effect on future inventories. We will summarize that effect through a "value function" for inventories, $V(I_{t+1}, Z_t)$, where Z_t are variables that are correlated with expected future supply and demand, and $I_{t+1} = I_t + Q_t - S_t$.

The function V arises from a dynamic optimization problem stemming from production smoothing[6] or some strategic motive.[7] Once again, if the motive is strategic, we assume that it applies roughly symmetrically across firms in equilibrium, so that aggregation to the industry level will be appropriate. Then, the short-run supply behavior of the industry will take the form

$$\begin{aligned} \mathrm{SRMC}(Q_t, W_t) &= V'(I_t + Q_t - S_t, Z_t); \\ \mathrm{MR}(S_t, Y_t) &= V'(I_t + Q_t - S_t, Z_t). \end{aligned}$$

The simultaneous solution of the above equations produces a reduced-form expression for sales and production:

$$S_1^* = S_1^*(W, Y, Z),$$

$$S_1^* = Q_1^*(W, Y, Z).$$

In the capacity-constrained regime we again have Q^c = EK. With SRMC vertical, we would expect SR supply behavior to be different from that in regime 1.[8]

---

[5] This introduces an aggregation problem if industry behavior is not symmetric, for example, if the capacity constraint binds tightly for some firms while it does not bind at all for others. This phenomenon is probably quantitatively unimportant for a technologically mature, stable oligopoly like aluminum. It would be much more troubling in a context where different firms used different technologies.

[6] Blinder, *supra* note 1. West, *supra* note 1.

[7] Rotemberg & Saloner, *supra* note 1.

[8] In these circumstances, the zero-inventory "stockout" constraint might also be relevant. Existing treatments of this constraint have emphasized prices, which we do not take

Thus we let the sales equation vary by regime as well, and write

$$S_2^* = S_2^*(W,Y,Z).$$

We see two general implications of the shape of the SRMC in Figure 1. These hold independent of the degree of competitiveness and independent of the dynamics. First, the reduced-form equation for Q takes on two different forms. In one regime, Q is affected by changes in variables that shift current costs or demand or that measure expectations about the future. In the other regime, Q is unaffected by those forces. Second, the two regimes should affect supply behavior as well. Even with inventory storage, one would expect the vertical SRMC to imply comparative statics for S very different from the flat(-ish) SRMC. We now turn to an investigation of the actual supply behavior of the aluminum industry.

III. Aluminum Industry Particulars

This section lays out the empirical specification for our study of the primary aluminum industry, Standard Industrial Classification (SIC) 3334. The details of the industry's cost structure will lead to a more complete and detailed specification of EK and of those AVC variables entering $X = (W,Y,Z)$. The nature of demand for the industry's product will further contribute to X. We also discuss competition and market definition issues in this section and comment on the definitions of some key variables. Table 1 gives brief definitions and means of variables referred to in the text.

Technology

By all accounts, aluminum smelting in any particular plant at any particular time takes place according to a fixed-coefficients production function in which alumina is reduced to crude aluminum by means of an electric current.[9] Nonmaterials inputs include capital, an important component of which is electrolytic "pots" organized into "potlines" in "potrooms" in the thirty-two primary aluminum plants operating at present in North America. Production workers' hours are dependent on the operation of the potline and on the making of anodes for the pots. When it is in operation, a line runs twenty-four hours a day, seven days a week, using four crews working twenty-one shifts. Potlines are the basic unit of pro-

---

up in this article. See Timothy F. Bresnahan & Pablo Spiller, Normal Backwardation under Risk Neutrality, 24 Econ. Inq. 429 (1986); and Timothy F. Bresnahan & Valerie Suslow, Short-Run Pricing with Capacity Constraints, forthcoming in Annales d'Economie et de Statistique.

[9] J. A. Stuckey, Vertical Integration and Joint Ventures in the Aluminum Industry (1983).

TABLE 1

Brief Variable Definitions and Means

Variable	Definition	Units	Mean	SD
MONTH	. . .	1, . . . , 12	6.21	3.58
YEAR	. . .	58, . . . , 83	72.61	6.78
Q	primary aluminum production	100,000 tons	3.17	.78
S	primary aluminum sales	100,000 tons	3.09	.85
K	primary aluminum capacity	100,000 tons	3.62	.91
RWAGE	average standard hourly rate plus COLA, aluminum production workers	index	.21	.004
RMATS	cost of materials	index	.28	.04
RPSCR	cast aluminum scrap price	1967 cents/lb	.13	.02
PPLAS	index of plastics prices	1970 = 1	.55	.08
RPSTL	steel billet price	1967 $/ton	1.06	.18
RPCOP	refined copper price	1967 cents/lb	.37	.07
LEAD	index of 12 leading indicators	1967 = 1	1.18	.23
IP	index of industrial production	1977 = 1	.90	.25
IPCA	index of automotive products	1977 = 1	.73	.18
IPCHC	index of appliances AC, TV	1977 = 1	.83	.22
IPDTTB	index of trucks and buses	1977 = 1	.59	.25
IPDT2	index of motor vehicles and parts	1977 = 1	.72	.17
IP372	index of aircraft and parts	1977 = 1	1.23	.20
HSBP	index of new private housing	1967 = 1	1.16	.32
MD072	new orders, durable goods	1972 bil$/100	.33	.06
PW	PPI, all commodities	1972 = 1	1.39	.75
TIME	. . .	[(year − 57)*12 + month]/120	1.61	.21

duction supply; they are switched on and off with some frequency, although at a cost.[10] Production is thus highly divisible, and total variable costs (in a particular plant at a particular time) are proportional to the number of potlines in operation—and thus to output.[11]

Based on these facts, we expect SRMC below capacity to be almost perfectly constant *at the level of an individual potline*. We *do not* expect constant marginal cost at the plant or firm level. There are two analytical points. First, heterogeneity of potlines at the plant level induces slope in the supply curve. The evidence shows, however, that cutbacks in production are made one potline at a time across plants.[12] This suggests that older plants have been at least partially retrofitted with new equipment so that on an industrywide basis SRMC below capacity is still flat.[13] Because of this rationalization of production, our method of aggregating across potlines and plants is defensible.

Our specification will not be able to detect sloped, but linear, SRMC. We will, however, be able to test for curved SRMC. The interesting economic case is the one in which AVC rises as production nears capacity, removing some of the sharpness of the angle in Figure 1.

Several variables that shift AVC enter X. A chronology of standard hourly wages for Alcoa has been kept by the Bureau of Labor Statistics (BLS) since the 1930s. Data are provided by plant, union, and job grade.[14] Our real-wage series is based on the average job grade and is a weighted average across membership in two unions. Cost-of-living adjustments are added to the wage series based on formulas prescribed in the contracts.

The industry has seen some labor-saving technical change over the period that can be characterized into two broad movements. First was the burst of new capacity and increase in average plant size by roughly one-

---

[10] Minimum maintenance is needed when pots are shut down. The power costs of remelting the "bath," the possibility of cracked pot lining, and the labor costs of avoiding cracked linings are the primary restart costs. Mr. Whitchurch of Alcoa (note 4 *supra*) estimated these at 1–1.5 cents per pound of capacity.

[11] The typical plant layout involves one or more potrooms and a carbon anode assembly facility, with one or two potlines per potroom and 100–250 pots per potline.

[12] We have information on individual potline closings by plant from Mineral Industry Surveys. Although the data are not sufficient to put together firm-level supply curves, they do make clear that recessions and power shortages initially result in potline closings across plants, not plant closings.

[13] Valerie Suslow, Estimating Monopoly Behavior with Competitive Recycling, an Application to Alcoa, 17 Rand J. Econ. 389 (1986), found a fairly flat SRMC for Alcoa in the interwar period.

[14] Two unions, the United Steelworkers of America and the Aluminum Workers International Union, organized Alcoa's plants and have historically negotiated contracts at the same time. The wage strucure set at Alcoa has been followed by the other major firms in the industry. An average of fifteen job grades existed over the sample period.

third in the 1950s. This first movement represents a period of embodied technical change: new equipment was installed, and plants built before World War II were retrofitted. The second major advance was the automation of all potroom operations in the late 1960s and early 1970s. The extensive use of mechanization boosted labor productivity significantly.[15] Accordingly, we include not only the real wage rate[16] but also the real-wage rate interacted with three technology variables. These are $T_1 = K_t$, for observations through 1961, and $T_1 = K_{1961}$ at later times, where K is installed industry capacity. The variable $T_2 = \text{YEAR} + (\text{MONTH} - 1)/12$ is a time trend designed to capture autonomous technical change, where YEAR = 58, . . . , 83 and MONTH = 1, . . . , 12. Finally, T_3 accounts for the diffusion of automation through the industry; it is a time trend that begins in 1967.

Materials prices are based on census data. As defined by the Census of Manufactures, "cost of materials" refers to direct charges actually paid for items consumed. It includes the cost of fuels consumed for heat, power, or generating electricity. Alumina accounts for the bulk of materials costs and has been used in constant proportions to aluminum output since the aluminum refining process was invented in 1896. Electricity demand has been affected by changes in technique since 1974. Despite investment in electricity-saving production methods, the industry average electricity requirement has decreased very slowly.[17] Thus we do not estimate any materials-demand equations in the SR; instead, we assume that materials' contribution to SRMC is unit materials cost (defined as COM/Q, where COM is cost of materials from the Annual Survey of Manufacturers [ASM]).[18]

The capacity figure in our data is monthly nameplate North American primary ingot capacity. Actual available or economic capacity may differ from nameplate capacity because of variations in the availability of electric power. In our data, large cutbacks in power supply are noted by the trade press. In particular, major cuts in power allocations to aluminum producers were made by the Bonneville Power Authority several times between 1977 and 1980. Power supply was cut anywhere from an estimated 12 to 25 percent, with these cuts lasting from two to ten months.

---

[15] See the Bureau of Labor Statistics publication, Technological Change and Manpower Trends in Six Industries, BLS Bulletin 1817 (1974).

[16] Throughout, "real" means "deflated by the Producer Price Index (PPI)."

[17] Minerals Yearbook states that the sum of all technology improvements from 1960 to 1980 decreased electricity usage from 7.7 kilowatt-hours (kwh) per pound to 5.9 kwh per pound in modern plants. 1 Minerals Yearbook 90 (1981).

[18] Since aluminum plants are large, the ASM has complete coverage even in intercensal years. We interpolate the annual unit materials cost figures to get monthly costs.

These shortages initially resulted in potline closings across plants, then plant closings in severe cases. As a result, we adjust K to reflect available capacity during these power brownouts. Our adjustment factor, BPA, is equal to Bonneville's operating rate multiplied by the fraction of aluminum capacity served by Bonneville (a constant over our sample of roughly 30 percent). We will include BPA as a regressor in regime 2:

$$\mathrm{EK}_t = \mu K_t(1 + \mu_B \mathrm{BPA}_t) + e_2. \tag{4}$$

Smaller, local variations in the power flow enter the error term e_2. These can be substantial and of either sign. For instance, published capacity figures represent normal operating rates, based on median power flows, which can be exceeded during peak periods: the maximum of Q/K in our data is 1.03. Second, although the Bureau of Mines publication, *Mineral Industry Surveys,* occasionally gives dates for major midyear capacity expansions, there may be some intrayear capacity changes not accounted for in our series. Third, nameplate capacity can differ from available capacity because the time needed for routine maintenance is a random variable and because of the random occurrence of unscheduled maintenance or breakdowns. The frequency and size of these events is uncertain, as is that of power availability. Thus the shape of the distribution of e_2 is uncertain.

Demand for aluminum is derived largely from a demand for durables. The major markets are transportation (cars, trucks), building and construction (windows, doors, residential siding), consumer durables (air conditioners, appliances), and containers and packaging (metal cans). Transportation and consumer durables have historically been the more volatile markets. The sectoral composition of demand for aluminum exhibits both a trend and a cycle within our sample period.[19]

One can imagine, then, a great many different forces bearing on the determination of Q_1^*: effects of the sectoral composition of the economy, dynamic effects arising from expectations, dynamic effects working through the state variables such as inventories, and so on. We will make no effort to disentangle the distinct effects of these forces in our specification of Q_1^*. We have, however, experimented somewhat with the list of demand-side covariates we include in Q_1^*. Our final specifications are the ones we report. They have six demand-side covariates:

the index of Leading Indicators;
housing starts;

---

[19] See Charles River Assoc., An Economic Analysis of the Aluminum Industry, prepared for the General Services Administration, Washington, D.C., March 1971, for estimation of disaggregate demand curves.

the Industrial Production Index for Aircraft and Parts;
the Industrial Production Index for Trucks and Buses;
time; and
time squared.

Our basic interpretation is that these six variables effectively span the space of the size of the economy, its expected rate of growth, and its recent history. Lengthening the list does not appear to add much to our ability to predict Q (see Table 1 for demand-side variables excluded from our final specification). Shortening the list *does* substantially degrade our ability to predict Q. Over a fairly broad range, experimenting with lags or variables has only a minor effect on fit or predicted values. This last fact is the basis for our diffuse interpretation; any half-dozen or so variables bearing on the current or recent state of the metals-durables side of the economy appears to contain roughly the information used in determining Q_1^*. As a result of the "spanning" interpretation, there are no predictions as to sign or order of magnitude on these variables' coefficients.

Aluminum is like other "smokestack" industries in that there has been a tremendous increase in competition over the sample period. It is atypical, however, in that the source of this competition has come largely from domestic entry rather than imports, as a result of tariff protection and high shipping costs.[20] In 1957, the annual average market share held by imports (to the United States) was roughly 10 percent. By 1982, it had grown only to 13.9 percent. The bulk of this imported aluminum ingot comes from Canada, in particular from Alcan, which has accounted for roughly 85 percent of Canadian capacity for the entire sample period.[21]

Concentration in the North American market, as in the world, has been steadily declining over time. Let H be the Herfindahl index, measured in capacities, over three possible market definitions: the United States, North America (the big three U.S. producers plus Alcan), and the world. As Table 2 shows, concentration in the narrower markets starts very high

---

[20] Ingot carries a tariff of one cent per pound and accounts for most of the imports, versus aluminum shapes, with tariffs of 3.5–5 cents per pound (9.5 percent of value). A 1976 Council on Wage and Price Stability publication on aluminum (hereafter referred to as COWPS) reports that shipping costs from Japan and Europe to the American market average about one-fifth the price.

[21] Other major exporters to the United States are Ghana, Norway, and Japan. Imports from the first two countries pose no threat since the capacity in those countries is owned by U.S. producers. Imports from Norway have declined dramatically since 1970, when the aluminum smelters were nationalized. Prior to 1970, the smelter capacity was owned by U.S. producers (see COWPS, at 44). For Japan, the barrier is high production costs. In 1974, Japan had estimated average costs of producing primary aluminum that were fifty percent higher than North American producers (COWPS, at 42).

TABLE 2

MARKET SHARES (Primary Capacity) FOR TOP PRODUCERS (%)

	UNITED STATES			NORTH AMERICA			WORLD		
	1957	1970	1982	1957	1970	1982	1955	1971	1979
Alcoa	43.1	31.4	28.5	29.8	24.5	22.9	20.4	17.1	15.1
Reynolds	26.6	22.4	17.6	18.4	17.5	14.1	15.0	11.8	8.8
Kaiser	27.1	16.8	13.1	18.7	13.1	10.5	14.6	8.4	7.8
Alcan	...	...	...	29.1	18.9	17.2	26.2	19.9	16.3
Pechiney	...	...	...	...	...	...	5.8	10.0	8.3
Alusuisse	...	...	...	...	...	...	3.9	5.8	5.9
C^*	96.8	70.6	59.2	96.0	74.0	64.7	85.9	73.0	62.2
H^*	.33	.18	.13	.24	.14	.11	.16	.10	.07

SOURCES.—Domestic figures from Aluminum Statistical Review, Minerals Yearbook. World figures from J. A. Stuckey, Vertical Integration and Joint Ventures in the Aluminum Industry (1983).

NOTE.—C^* = n-firm concentration ratio; H^* = n-firm Herfindahl index, where n = 3, 4, or 6.

and declines to a middling level, while world concentration declines in parallel. Within North America, the cause of this decline is the slow expansion of a competitive fringe following the government-subsidized entry of Reynolds and Kaiser after World War II.

This is not to say that we are convinced that North America is isolated from world markets for aluminum. Rather, for our purposes in this article, there is little distinction between a North American market approach and a world market approach.

Data

The aluminum industry data were compiled primarily from five sources: *American Bureau of Metal Statistics Yearbook, Aluminum Statistical Review, Metals Statistics, Metal Bulletin Handbook,* and *Mineral Industry Surveys*. Additional data came from the *Census of Manufactures* and the BLS *Employment and Earnings*. Specific sources for individual variables are given in Appendix A. In this section, we elaborate on the less straightforward variables.

The sample period runs from January 1958 through December 1983.[22] When possible, the data are monthly—for example, aluminum prices, production, shipments, and inventories. Important variables available only on an annual basis are production capacity and materials costs. Monthly information related to these annual variables was used when possible. For instance, the Bureau of Mines publication, *Mineral Industry Surveys,* occasionally gives dates for major midyear capacity expansions. These data were incorporated to make the capacity series as error free as possible.

Shipments data include all private domestic shipments of primary ingot plus the aluminum content of domestically shipped mill products. Imports and shipments of secondary aluminum are excluded. Despite the fact that we possess data on production and shipments of primary aluminum, those data cannot be easily used to create an inventory series.[23]

The data for production, nameplate capacity, and unit sales, as well as those for price, are graphed against time in Figure 2. The three quantity

---

[22] The early 1950s were eliminated because of the Korean War. Also in the early 1950s, the U.S. government instituted a subsidized expansion program for the aluminum industry. Contracts were written in 1950–52 and expired in 1958 for the big three producers. By 1958, all of the new capacity built with government aid had come on stream.

[23] Conversation with Ed Coan, Industry Div., Bureau of the Census, May 1987. Data are missing on the quantities of new scrap that flow through the system (for example, from the fabricators of aluminum shapes back to the primary producers). In fact, the inventory data we use from Current Industrial Reports are not calculated from production and shipments data within that report but are reported as a separate line item by aluminum producers.

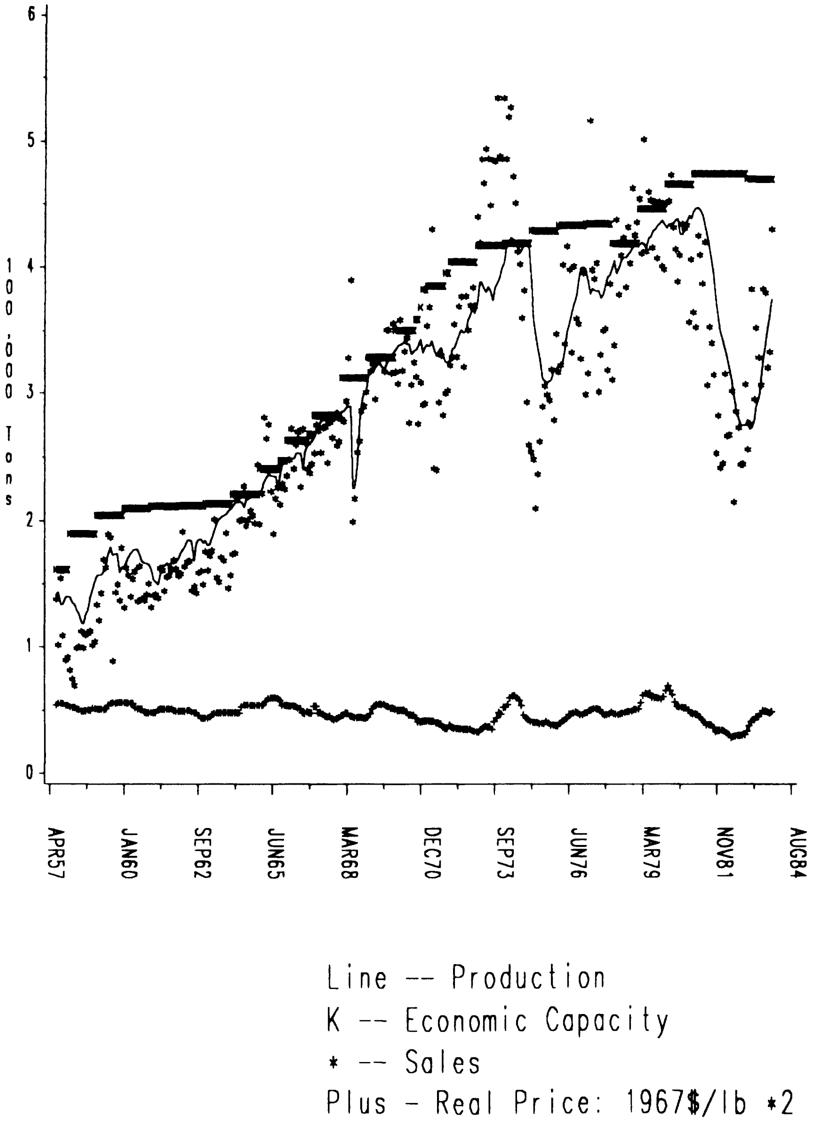

FIGURE 2.—Production, capacity, sales, and price

variables all have units of 100,000 tons. In this and later graphs, we make a small seasonal adjustment for clarity. In the data, these series are the actual monthly amounts. The fact that February is 10 percent shorter than January, that October is 3 percent larger than September, and so on, would make graphs somewhat difficult to read. We rescale the date so that all months are 31 days long; in nonleap years the February figures are 31/28 as large as the underlying date, and so on. The prices graphed are deflated by the producer price index for all quantities. We have rescaled the price series to show it on the same graph as quantities.

IV. Estimation and Testing

First, rewrite (2) and (4) with additive errors:

$$Q_1^* = X\beta_1 + e_1; \tag{2a}$$

$$\mathrm{EK} = \mu K(1 + \mu_B \mathrm{BPA}) + e_2. \tag{4a}$$

We assume that e_1 and e_2 are independent normal with variances σ_Q^2 and σ_K^2, and write the density function of the random variables (Q_1^*, EK) as $g(Q_1^*, \mathrm{EK})$.

Then the likelihood with normal errors is[24]

$$L = \int_{\mathrm{EK}=Q}^{\infty} g(Q,\mathrm{EK})d\mathrm{EK} + \int_{Q_1^*=Q}^{\infty} g(Q_1^*,Q)dQ_1^*.$$

SR Supply: Capacity

The maximum-likelihood estimates of the quantity equations are presented in Table 3, column 1. As is well known, the likelihood function for the disequilibrium model is unbounded, and it can be difficult to obtain estimates. We did not experience any difficulty, however.[25] The coefficients of the capacity equation are readily interpretable:

$$\mathrm{EK} = .97{*}K(1 - .05{*}\mathrm{BPA}) + e_2.$$

---

[24] This is the familiar "disequilibrium" model. See Stephen Goldfeld & Richard E. Quandt, Estimation of a Disequilibrium Model and the Value of Information, 3 J. Econometrics 325 (1975). Note that our treatment of capacity utilization is basically an "engineering" one; see Ernst R. Berndt & Catherine J. Morrison, Capacity Utilization Measures: Underlying Economic Theory and an Alternative Approach, 71 Amer. Econ. Rev. (1987), for a very different approach.

[25] The problem arises when a few data points are predicted perfectly and an estimated variance goes to zero. It is solved by finding parameter values such that the derivatives of the likelihood function with respect to all the parameters are zero. At the estimates we report, the vector of derivatives is zero to six digits.

TABLE 3

QUANTITY REDUCED FORM (t-Ratios in Parentheses)

Variable	Base Case	Regime 1 Tests	Regime 2 Tests
$\log(\sigma_Q)$	-1.52	-1.52	-1.49
	(-27.28)	(-33.97)	(-39.42)
$\log(\sigma_K)$	-2.57	-2.57	-2.55
	(-32.22)	(-46.08)	(-59.84)
μ	.97	.97	.90
	(408.60)	(421.44)	(74.67)
INTERCEPT	.21	.22	.06
	(1.24)	(3.19)	(1.26)
RWAGE	.61	.56	.45
	(3.43)	(7.81)	(9.19)
RMATS	-7.56	-7 .50	-7.61
	(-43.42)	(-105.54)	(-155.85)
LEAD	4.31	4.29	4.34
	(26.72)	(62.26)	(92.53)
WAGE*T1	.90	.91	.95
	(5.11)	(12.81)	(19.41)
WAGE*T2	.73	.72	.77
	(9.96)	(19.02)	(30.43)
WAGE*T3	0.21	.25	.07
	(1.38)	(3.67)	(1.45)
HSBP	-.58	-.58	-.58
	(-7.24)	(-10.93)	(-14.30)
IPDTTB	1.64	1.61	1.74
	(14.13)	(26.09)	(38.66)
IP372	3.01	3.00	3.22
	(23.06)	(47.30)	(71.13)
TIME	.22	.25	.25
	(1.73)	(3.72)	(5.83)
TIMESQ	-.83	-.83	-.88
	(-13.89)	(-25.63)	(-40.05)
μ_B	.05	.07	-.04
	(.51)	(1.10)	(-.85)
H^*	1.04	1.03	1.14
	(5.93)	(14.44)	(23.31)
K - REG. 1	...	-.01	...
		(-.29)	
BPA*K - REG. 1	...	.09	...
		(1.30)	
LEAD - REG. 2	...	...	.24
			(6.49)
RWAGE - REG. 2	...	...	1.41
			(29.34)
RMATS - REG. 2	...	...	-1.35
			(-28.63)
Log L	138	139	141

Thus, EK is, on average, about 97 percent of nameplate capacity. Brownouts do reduce economic capacity, but by somewhat less than we thought. Finally, the random departures from expected available capacity represented in e_2 are small. Its standard deviation, σ_K, is .077, about 2 percent of average capacity. (The unconstrained regime error, with σ_Q = .22, is substantially larger.) From this one can infer that, even if *individual* potlines offer tricky process-control problems, the law of large numbers leads to small industry-level variation from expected performance.

Some of the *t*-statistics in Table 3 are extremely large. The *t* of 408.6 on capacity simply reflects the fact that the hypothesis $\mu = 0$ is nonsense. The corresponding statistic for $\mu = 1$ is -12.6; we reject zero downtime for maintenance, but not at an unbelievable rate.

To understand the implications of the estimates of the coefficients in regime 1, turn to Figure 3. The solid line graphs actual production in units of 100,000 tons. Note that, in the late 1950s, the line is broken; in that period, our sample includes only the first three months of each year.[26] Note also the gap from December 1972 to August 1974. During this period, price controls led us to delete all observations.[27] We have also deleted a brief strike period in 1968.

The graph of actual production shows steady growth from the beginning of our sample, with some flattening in the late 1960s, until late 1974. There is then a marked decline until mid-1975, followed first by rapid, but then slowing, growth until 1980, and a deep drop through late 1982. The early phases of the current macroeconomic recovery can be seen in the last year graphed. Though our sample ends in 1983, a continuation of the graph would show growth throughout the 1983 to 1987 recovery.

The econometric model's predicted values are also shown in Figure 3. Regime 1 unconstrained production, Q_1^*, is graphed as a star, while economic capacity is shown as a *k*. Here Q_1^* is much more volatile; note its peaks in October 1969 and April 1979 and troughs in June 1975 and January 1983. Capacity is a much smoother, slower moving series, despite its early period of rapid growth. The economic interpretation follows from Figure 1. When $Q_1^* > \text{EK}$, pressure is placed on capacity. Of course, Q_1^* is only estimated on the data from capacity-unconstrained periods. If one is

---

[26] This is due to a data constraint on total production-worker hours for the aluminum industry in the 1950s.

[27] The Cost of Living Council imposed price controls on the aluminum industry from August 1971 until August 1974. In the early years of the freeze, however, there was excess capacity in the industry, as one can see from Figure 4, and the controls were not binding. COWPS, at 123, states that the transactions price was significantly below list for much of the 1971–72 period: "A reversal occurred in early 1973 when markets firmed and the government-controlled list price did not clear the market."

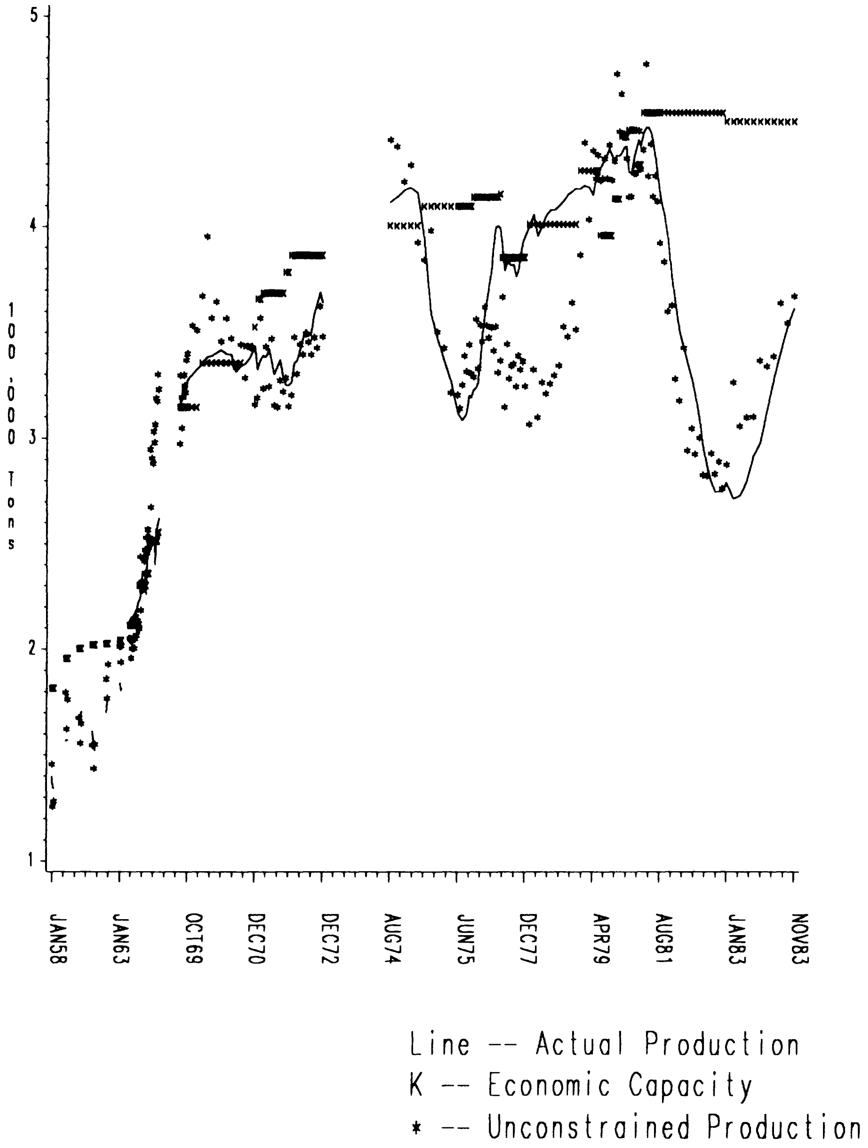

FIGURE 3.—Production, Q^*, and K

prepared to extrapolate the particular form of Q_1^* to the rest of the sample, a further inference follows. The shadow value of capacity is *proportional* to $Q_1^* - \text{EK}$.

Now consider the relationship between the two predicted values (Q_1^* and EK) and actual production. The industry is first characterized by a period of excess capacity, with output below capacity but actual and unconstrained production, Q_1^*, coinciding. Then, in early 1963, unconstrained output outstrips capacity, and actual production is determined by $Q = \text{EK}$. This pattern persists throughout: either economic capacity or desired output is an excellent predictor of actual output. It is rarely the case, however, that unconstrained output and capacity coincide. Looking at Figure 3 does not suggest any serial correlation in the errors, and the obvious calculation confirms this. Seventy-nine percent of the observations for which the probability of regime 1 exceeds one half have P_r (regime 1) $> .95$ and also P_r (regime 1) $> .95$ for the previous period. (Confining our attention to those periods, the Durbin-Watson statistic for the regime 1 errors is 1.78.)

The disequilibrium econometric model, on the evidence of Figure 3, is fitting the data very well. There is strong evidence in the figure against the theoretical possibility of having bad estimates due to the unbounded property of disequilibrium likelihood. Quantities do seem to be determined by an extremely volatile unconstrained output function, upper truncated by an unbending capacity constraint. The volatility is as marked as one would expect in the durable/metals end of the economy: Q_1^*/K is around 1.37 in the 1975 peak, around .60 in the 1982 trough.

Columns 2 and 3 of Table 3 show our tests of the exclusion restrictions across regimes. The theory says that either $dQ/dK = 0$ or $dQ/dX = 0$, depending on regime. To investigate this econometrically, we test the exclusion of the AVC and demand variables from the quantity reduced form in regime 2, and exclusion of capacity predictors from regime 1. As one can see from column 2, there is nothing in the data to suggest that K or BPA$*K$ belongs in the regime 1 quantity-determination equation. In column 3, three selected variables from the AVC and X vectors are inserted in the capacity-constrained regime. Though they appear individually significant, this is illusory; the three estimated parameters covary strongly, and the test for joint significance does not reject zero.[28] Adding

---

[28] When we include any of the variables alone, it is insignificant both according to the t test and the likelihood-ratio test. The high t's in the table reflect the very substantial collinearity between RWAGE and RMATS conditional on the other variable. If we select constrained periods, regress RWAGE and RMATS on LEAD, K, and BPA$*$K, the residuals are correlated at over .99. This is why we do not report any regime 2 exclusion tests with more variables.

more or different variables to the capacity-constrained regime does not appear to alter the story much.

What is to be made of the exclusion test results? It seems clear that quantity determination is very well described by a two-regime system. In one regime, capacity constraints completely determine Q, while in the other regime, the amount of plant capacity is irrelevant.

Moreover, in the second regime, the data do not encourage the view that the amount of capacity is endogenous to the state of demand. What does this mean? What seems clear here is that short-run strategies such as deferred maintenance are of little practical importance in determining economic capacity. If managers were responding to demand pressures by cutting corners on maintenance, that would tend to show up in our estimates as an EK which depended on demand circumstances. We detect no such effect.[29]

Further, the data seem to support the view that a relatively small number of demand-side indicators adequately predict quantity. In particular, the exclusion of K from Q_1^* implies that K does not embody any part of the forecast of future demand not captured by the variables in X. This is quite consistent with what is probably the right view of aluminum technology. Capacity in this industry is a very slowly adjusting asset. Capacity K available at any particular time reflects expectations formed years earlier. The half-dozen nearly contemporary variables in X predict future sales—and, therefore, current desired production—excellently without K.

Also, there is the test of our distributional assumption for e_2: are the disturbances around EK normal or asymmetric? One's first instinct is that the shape of the distribution should be as in Figure 4*a*. In that figure, nameplate capacity, K, corresponds to the theoretically available maximum capacity. The only random events that occur are negative, so that the distribution of e_2 is one-sided. Figure 4*a* reflects the assumption that small departures from available capacity are more likely than large ones. Most of the probability density is on available capacity near nameplate capacity.

There is no particular a priori reason, however, to rule out a distribution of e_2, as shown in Figure 4*b*. In that figure, there is a mean available economic capacity, $\overline{\text{EK}}$, somewhat below nameplate capacity, K. The distribution of random capacity disturbances around $\overline{\text{EK}}$ is drawn much more symmetrically. Once more, there is a physical theoretical maximum production, $K_{\max}$, but having available capacity near $K_{\max}$ is a rare event.

---

[29] Perhaps in other industries with less rigid technologies this would be less marked. In the long run, of course, capacity accommodates itself to demand. But that is a distinct question.

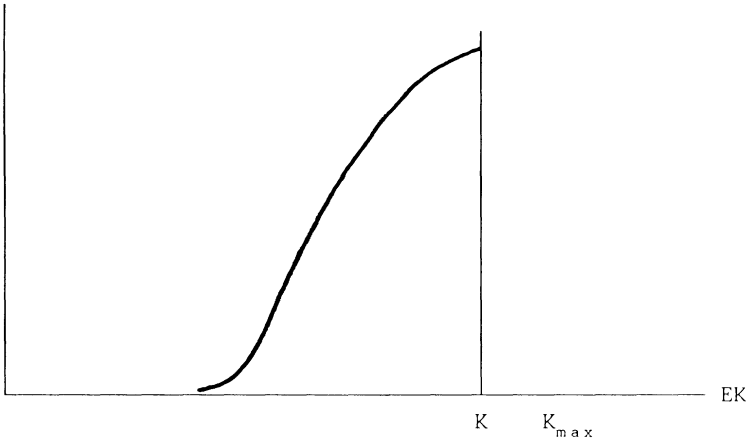

FIGURE 4*a*

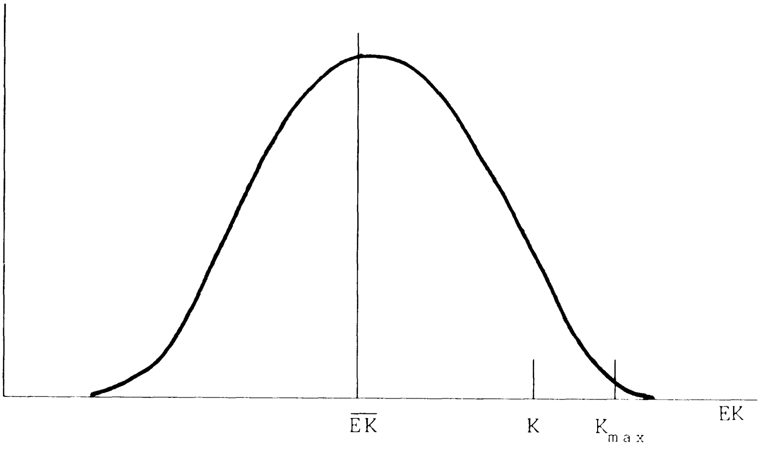

FIGURE 4*b*

Since neither shape can be ruled out on prior grounds, we make distributional assumptions for e_2 that can take either form. In particular, while we assume that e_1 is normal with mean 0 and variance σ_q^2, we assume that e_2 is the sum of a half-normal variate (like Figure 4*a*) with variance σ_1^2 and a normal variate (like Figure 4*b*) with variance σ_e^2. This econometric treatment is familiar from the "frontier production function" approach of Aigner *et al.*[30]

---

[30] Dennis J. Aigner, C. A. Knox Lovell, & Peter Schmidt, Formulation and Estimation of Stochastic Production Function Models, 6 Econometrics 21 (1977).

The likelihood function for this case is described in Appendix B. We do not present results since they rarely differ from our base case. This is because the mean of the asymmetric part of the error in capacity is about 88 tons per month. Capacity itself is, on average, 3.61 hundred thousand tons per month. Thus, it is clear that the asymmetric part of the error is essentially trivial.

We should emphasize that the absence of an asymmetric error has more implications than the obvious ones about the shape of the distribution of scheduled and unscheduled maintenance. It further provides an important test against nonlinearities in SRMC near capacity. Suppose, for example, that plants were heterogeneous, with a few plants embodying poor technology. These plants would be active only when the industry is at capacity. Such a scenario creates nonlinearities in SRMC. It would also imply an apparent asymmetric distribution for e_2. This would occur because of the frequent circumstance that all of the efficient plants operate while the inefficient ones remain shut. The absence of this phenomenon in our data argues against interplant heterogeneity of this type.

The evidence from quantity determination, then, speaks to a marked right-angle SRMC influence on supply. The implications of right-angle SRMC for SR supply are quite different from those of smoothly rising SRMC. Given this evidence, we can now proceed to examine the behavior of shipments and inventory holding across these two distinct regimes.

SR Supply: Shipments

While it is production that enters aluminum's cost function, it is shipments that generate revenue. Since the costs of holding finished ingot in inventory are quite low, one might expect occasional substantial inventory holdings. Even in the simple competitive case, inventories add significant complexities to the story of SR supply behavior. In particular, expectations about the future demand, as well as about the level of current demand, matter. The marginal benefits of building up or drawing down inventories obviously depend on these expectations, as well as on the size of current inventory. Sellers with market power would further need to take into account the strategic behavior implications of building inventory. If customers can also store the good, strategic sellers might need to take that into account as well.

Our approach to all these dynamic complications follows our modeling of Q_1^*. While this reduced-form approach keeps us from drawing structural inferences, it has the advantage of not imposing any particular view of the dynamic process driving the data. We continue in that structure-free mode by writing S_1^* and S_2^* as completely separate and independent functions of all the exogenous variables.

As outlined in Section II, we can expand the econometric model to include reduced-form equations for price and sales. These reduced-form equations are allowed to vary arbitrarily by regime. In particular, our expanded model with sales is

$$\begin{aligned} Q_1^* &= X\beta_1 + e_1 & \begin{pmatrix} e_1 \\ e_3 \end{pmatrix} &\sim n\,(0, \Sigma_1^2) \\ S_1^* &= X\Gamma_1 + e_3 & & \\ Q_2^* &= \mu K(1 + \mu_B \mathrm{BPA}) + e_2 & \begin{pmatrix} e_2 \\ e_4 \end{pmatrix} &\sim n\,(0, \Sigma_2^2) \\ S_2^* &= X\Gamma_2 + e_4 & & \end{aligned} \tag{5}$$

Several things about this specification are notable. First, it continues to assume independence of the errors across regimes. Second, it lets the equations for S and Q have arbitrary correlation within regimes. Third, Γ_1 and Γ_2 are completely unrestricted. The half-dozen demand-side variables in X that we use to predict S are the same as those used to predict Q_1^* (see Table 3). Finally, P does not appear in the system. Price equations are estimated separately using a structure analogous to equation (5).

Systems like (5) can be estimated by maximum likelihood. We found, however, that the maximum-likelihood estimate (MLE) departed only trivially from a much simpler procedure.

The simpler procedure is as follows. First, take the quantity reduced-form estimation as reported in Table 3, column 1. Calculate a dummy variable for regime 2:[31]

$$R = 1 \quad <=> \quad \mathrm{pr}(Q_1^* > \mathrm{EK}) > .5. \tag{6}$$

Then run an ordinary least squares (OLS) regression for S on the variables X and $R*X$.

Why does this simpler procedure work so well on our data? First, the probabilities in equation (6) are almost all near zero or one. The reason for this can be seen in Figure 4. Almost all observations are predicted well by one regime and badly by the other. Thus the regime classification is near certain at most observations. Second, at the MLE, the quantity-equation errors and sales-equation errors in (5) are nearly independent. Under these circumstances, there is no important stochastic difference between OLS and MLE.[32]

Estimates for the sales and price specifications of system (5) are reported in Table 4. It is quite difficult to interpret the coefficients. Various predicted values from the specifications are reported in Figures 5–7.

---

[31] This is $\mathrm{pr}(X\beta_1 - \mu K(1 + \mu_B \mathrm{BPA}) > e_2 - e_1)$, a simple cumulative normal integral.

[32] For example, the correlation coefficient for Σ_1^2 was .04, that for Σ_2^2, $-.08$. There is an important cost difference between the two estimates, as computation of the MLE for systems like (5) takes many, many hours of CPU time.

TABLE 4
Shipments Reduced Form (*t*-Ratios in Parentheses)

Variable	Shipments
INTERCEPT	2.74 (1.24)
RWAGE	5.24 (.55)
RMATS	.21 (.14)
LEAD	6.62 (5.67)
WAGE*T1	1.34 (.22)
WAGE*T2	2.56 (3.08)
WAGE*T3	−.52 (−.93)
HSBP	.18 (.65)
IPDTTB	−.42 (−.91)
IP372	2.12 (4.35)
TIME	.48 (.64)
TIMESQ	−2.06 (−5.65)
INTERCEPT*R	1.01 (.28)
RWAGE*R	6.49 (.34)
RMATS*R	−6.03 (−1.01)
LEAD*R	−1.93 (−9.50)
WAGE*T1*R	9.10 (.58)
WAGE*T2*R	−1.18 (−.88)
WAGE*T3*R	.15 (.19)
HSBP*R	−.64 (−1.49)
IPDTTB*R	.22 (.33)
IP372*R	−2.19 (−3.52)
TIME*R	−1.30 (−5.30)
TIMESQ*R	1.46 (1.89)
R^2	.87

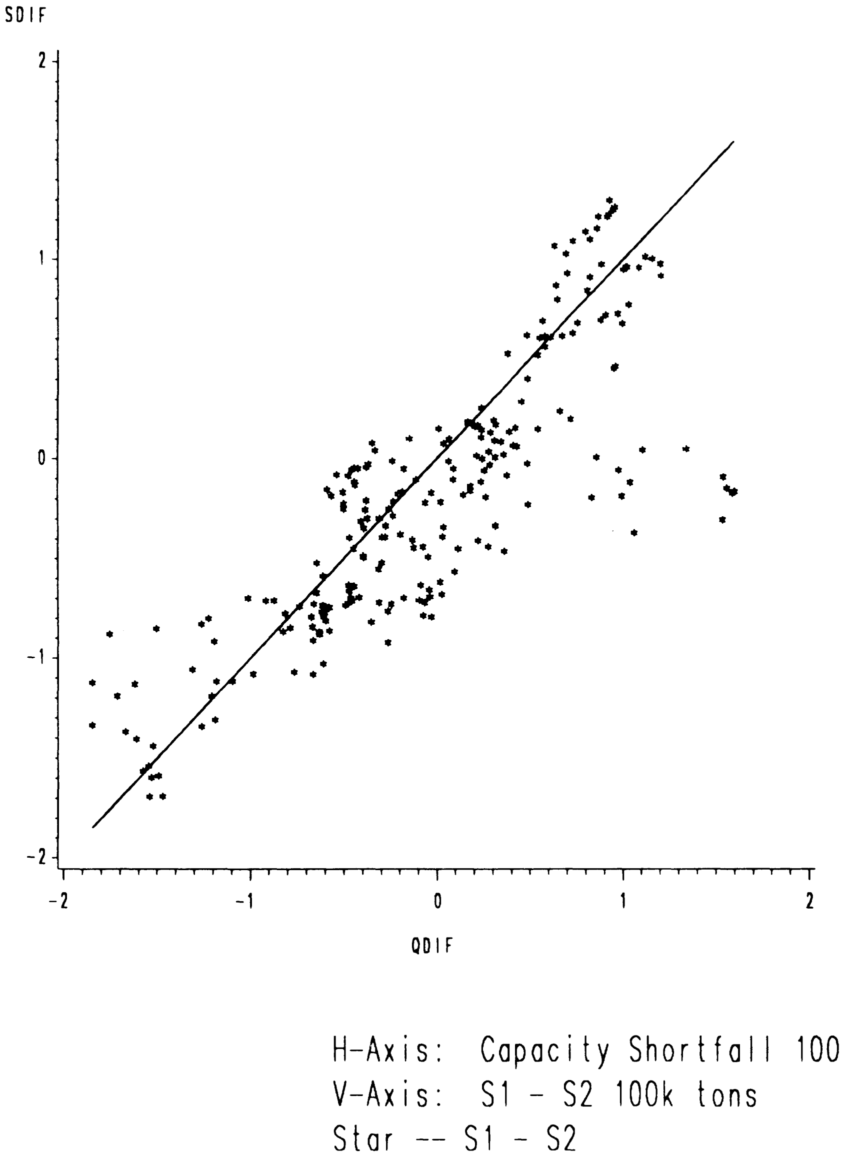

FIGURE 5.—Regime and predicted shipments; horizontal axis: $Q_1^* - EK$; vertical axis: $S_1^* - S_2^*$.

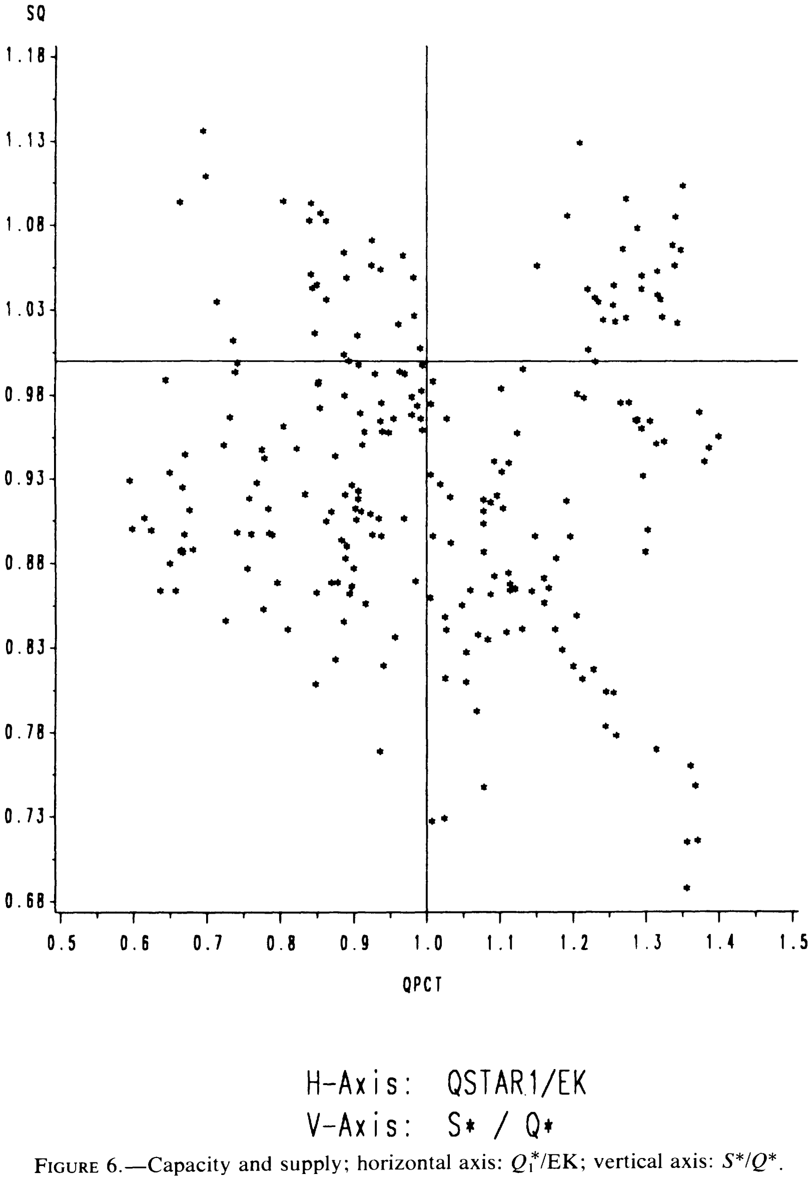

FIGURE 6.—Capacity and supply; horizontal axis: Q_1^*/EK; vertical axis: S^*/Q^*.

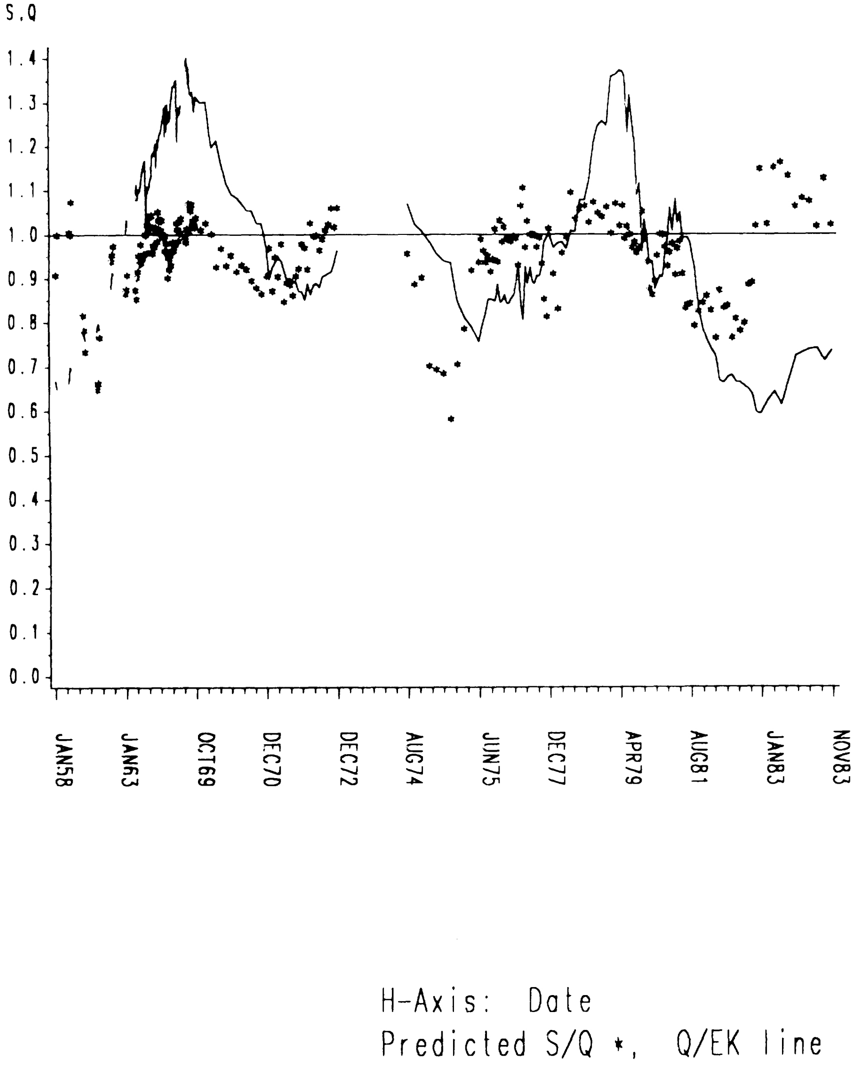

FIGURE 7.—Capacity and supply over time

The first graph to look at is Figure 5. On the horizontal axis, it shows the difference between regime 1 unconstrained production and economic capacity, $Q_1^* - \text{EK}$. Thus, starting at the left, the graph moves from very capacity-unconstrained periods to very capacity-constrained periods. On the vertical axis, we show the difference between regime 1 predicted shipments and regime 2 predicted shipments, $S_1^* - S_2^*$. Note that, with the exception of a small batch of outliers below the main body of the plot, there is a striking pattern:

$$\begin{aligned} Q_1^* &> \mathrm{EK} & &<=> & S_1^* &> S_2^*, \\ Q_1^* &< \mathrm{EK} & &<=> & S_1^* &< S_2^*, \\ Q_1^* &\approx \mathrm{EK} & &<=> & S_1^* &\approx S_2^*. \end{aligned}$$

The degree to which this coincidence holds is apparent in the graph. Our econometric procedures do not impose any restriction that would imply this; S_1^* and S_2^* are completely unrestricted linear functions of all the regressors. What the data show is that there exist two distinct regimes in the equations determining sales. Switches back and forth between these two regimes occur at the same sample points as switches between Q_1^* and EK in the quantity equation. The data therefore strongly support a switching-regimes view of the industry's supply, with both Q and S showing two-regime behavior, and with switches between regimes coincident in the two series. Thus, there is strong economic support for right-angle SRMC.

The economic interpretation of this coincidence can be seen more clearly in Figure 6. In that figure, the horizontal axis is the ratio of regime 1 production to economic capacity, Q_1^*/EK. The point where the capacity constraint just binds is marked with a vertical line. The vertical axis has a slightly more complex definition. When the capacity constraint binds, predicted shipments are S_2^* and predicted production is EK. In that case, we plot S_2^*/EK, the ratio of predicted shipments to predicted production. When the capacity constraint does not bind, the same ratio of predicted values is given by S_1^*/Q_1^*. The point where predicted shipments just equal predicted production is marked by a horizontal line. The average predicted ratio of S/Q is below one. It might appear from this that substantial inventory investment is forecast. This is slightly misleading. An equal number of shipments at 1.2 times a high level of production and at .8 times a low level of production do not average out to zero net inventory growth.

The main economic inference from Figure 6 is based on the observation that it is, on average, upward sloping. To the right of the $Q_1^* = \mathrm{EK}$ vertical line the average ratio of predicted S to predicted Q is almost exactly one. To the left of the line, it is below one on average. The upward slope here is quite consistent with the notion of *production smoothing*. The higher the amount of capacity available, the more likely the industry is to have $Q > S$, that is, to build inventory. As capacity gets scarce, it becomes much more likely that $S < Q$, that is, shipments are made out of inventory, not production. This positive slope is in marked contrast to the common finding in the literature that production smoothing does not occur in industry.[33]

The other highly visible feature of the graph is the much greater spread in S^*/Q^* to the left. In part, this is a damping phenomenon; when the capacity constraint holds, S tends to respond much less to the size of the economy. But this graph does not let us see the state of the economy and its recent history and therefore obscures the forces underlying the greater spread.

Those forces can be seen somewhat more clearly in Figure 7, which graphs the same two variables seen in Figure 6 against time. Consider, first, what happens in the long upturns of the early 1960s, of 1971–72, and of late 1975 through mid-1979. In each of these periods, S^* is initially below Q^* while both are rising. This appears on the graph as asterisks below the $S^* = Q^*$ line. This is a period of inventory building, suggesting that at least the possibility of a long upturn was foreseeable in those times. Sales, S, pass above Q at some point well into the expansion. In the early 1960s, a period of rapid capacity expansion, S passes above Q after the capacity constraint has been binding for some time. In the other two instances, S passes above Q before the industry goes to the capacity-constrained regime. In all three instances, however, $S > Q$ is a clear feature of the cyclical peak, when inventories are drawn down to meet demand.

By dropping out the price-controls period from the sample, the graph does not show the entirety of the long downturn beginning with the first fuel-price crisis. We do see all of the late-1960s downturn, as well as all of the second fuel-price-crisis downturn. In both instances, note that S falls below Q before Q falls below K. Why is this? At such times, the industry has very low inventories (just coming off a peak in demand). It may not be certain that the downturn will be permanent. The tendency is thus to have a period of S/Q below one while inventories are rebuilt. This provides a buffer against the possibility of a repeated upturn in demand.

The peculiar economics of the 1979–82 downturn and the beginnings of the long upturn that has just been completed are visible as well. Note that $S < Q$ well into the downturn; substantial inventories were built. The initial effect of the upturn, however, was not nearly so rapid a recovery of

---

[33] We do not present any formal statistical analysis of production smoothing. The existing body of methods is based on the assumptions that SRMC, adjustment costs for Q, and costs of departing from target inventory are all quadratic. (See West, *supra* note 1; and Blinder, *supra* note 1.) Since those assumptions are false for our industry, the methods will not provide a reliable formal test.

production as before. Instead, S goes above Q very quickly, implying that much of the early upturn in demand was met out of inventories. Though this upturn has turned out to last until the present day,[34] that was not foreseeable in 1983, and suppliers reacted with appropriate caution.

The evidence for a sophisticated production-smoothing explanation of production and shipments seems clear. It will take a substantial investment in new econometric/economic modeling to integrate the right-angle SRMC into a structural model of supply of S and Q with fully specified dynamics. This effort for the future seems warranted by the strong evidence for a two-regime system.

V. Conclusions

Any sensible theory of supply and industry equilibrium would suggest a discrete difference in behavior between capacity-constrained and unconstrained regimes. We modeled that discrete difference using an econometric disequilibrium model for quantity produced. This model offers a natural way to introduce the capacity constraint into the economic model. There is considerable evidence, in both the quantity-produced and quantity-shipped equations, that there are two separate supply regimes in aluminum.

The primary potential drawback with our model lies in its stark yes/no treatment of the capacity constraint. Industry SRMC might begin to rise near, but not at, capacity because of heterogeneity in firms' or plants' costs. This possible problem can be cast as a testable hypothesis in the disequilibrium framework. As one would have expected from the statements of aluminum production managers, this problem does not arise in our study.

Finally, the production-smoothing theory of inventory investment is quite promising for aluminum. Production typically exceeds shipments when capacity is available, and typically falls short of shipments when capacity is tight. Thus, we are unprepared to join the rapidly growing consensus that strategic considerations, not cost considerations, drive inventories.

---

[34] At this writing, all legal and illegal sources of supply of aluminum are stretched to their limit. Not only are primary suppliers' sources low but the recycling sector has exhausted much of the stock of easily accessed scrapped aluminum objects as well. Thieves are now the marginal supply source. They are unbolting aluminum rails from highways, dismantling farmer's irrigation systems, and so on.

APPENDIX A

DATA SOURCES[35]

Aluminum Statistical Review (a): shipments to and from the government stockpile; primary aluminum capacity (industry and individual firms)
American Bureau of Metal Statistics Yearbook (a): primary aluminum capacity (industry and individual firms)
Annual Survey of Manufactures (a): primary aluminum cost of materials, cost of electricity
BLS Wage Study (n): chronology of Alcoa production workers' nominal wage
Citibase (m): index of industrial production and its components
Current Industrial Reports (m): month-end inventories of primary aluminum ingot, mill products, and scrap
Employment and Earnings (m): average weekly production worker hours for primary aluminum
Handbook of Cyclical Indicators (m): composite index of leading indicators
Metal Bulletin Handbook (n): U.K. transaction price of primary aluminum (prices posted with no regular frequency)
Metallgesellschaft Aktiengesellschaft (m): U.K. transaction price of primary aluminum
Metal Statistics (m): primary aluminum list price, domestic production; prices of copper and steel
Metals Week Handbook (m): U.S. transaction price of primary aluminum
Mineral Industry Survey (m): shipments from the government stockpile after 1965; midyear capacity expansions
Minerals Yearbook (a): Canadian primary aluminum capacity (industry and individual firms)
Survey of Current Business (m): primary aluminum domestic shipments, imports, exports; secondary aluminum recovery

APPENDIX B

Under the assumptions in the text, the density function for e_2 is

$$g_2\,(e_2) = \frac{2}{\sigma_2}\,\phi\left(\frac{E_2}{\sigma_2}\right)\left[1 - \Phi\left(\frac{E_2\lambda}{\sigma_2}\right)\right],$$

where $\lambda = \sigma_u/\sigma_e$, and $\sigma_2^2 = \sigma_e^2 + \sigma_u^2$ (our notation follows Maddala).[36] Noting that $e_2 = Q_2 - \overline{\mathrm{EK}}\,(K, \mathrm{BPA}, \mu)$, we can write $g_2(Q_2)$ compactly. Then our likelihood function is

$$h(Q) = \phi(e_1)/\sigma_1 \int_Q^\infty g_2(Q_2)\,dQ_2 + g_2(Q)\left[1 - \Phi\left(\frac{Q - Q_1^*}{\sigma_1}\right)\right].$$

---

[35] Monthly and annual data are denoted by (m) and (a), respectively, while data with no set frequency are denoted by (n).

[36] G. S. Maddala, Limited-Dependent and Qualitative Variables in Econometrics 195 (1983).

We maximize this with respect to the parameters μ in $\overline{\mathrm{EK}}$, β in Q_1^*, and the three variance parameters σ_1, σ_u, σ_e.

We do not present estimates for this specification because the MLE of σ_u is so close to zero that none of the other parameters changes significantly. For example, in the base-case specification where Q_1^* and $\overline{\mathrm{EK}}$ are modeled as in Table 3, column 1, the MLE of σ_u is .0011. Thus, the expected value of the asymmetric part of the capacity error is $E[u] = -\sqrt{2/\pi} \cdot \sigma_u = -.000088$.