Rand Journal of Economics
Vol. 15, No. 2, Summer 1984

The learning curve and pricing in the chemical processing industries

Marvin B. Lieberman*

Data on 37 chemical products are used to test a number of hypotheses about the learning curve and industrial price behavior. The results document a strong and consistent learning effect. Learning is found to be a function of cumulated industry output and cumulated investment rather than calendar time. Standard economies of scale appear significant but small in magnitude relative to the learning effect. Variations in the slope of the learning curve are linked to differences in R&D expenditures and capital intensity. Market concentration is found to be a strong influence on price flexibility and the timing of learning-related price changes.

1. Introduction

■ The learning curve model is used extensively in industry as a tool for production planning and cost forecasting. Strategic planners and strategy consultants frequently base their analysis on the learning curve and related concepts. Despite this widespread practical acceptance, theoretical research on the learning curve has been limited, and there have been relatively few published empirical studies. This article reports some empirical results on the learning curve on the basis of comprehensive data for 37 chemical products.

The article is organized as follows. Section 2 describes the learning curve and summarizes prior research. Section 3 introduces the regression model and Section 4 describes the data sample and explanatory variables. In Section 5 the data are used to test a series of alternative learning index measures, including cumulated industry output, cumulated investment, and calendar time. Standard economies of scale are also examined. Section 6 documents the existence of small but significant variations in the slope of the learning curve across products in the sample. Section 7 examines the timing of learning-induced price changes. Market structure is found to have a strong influence on short-run price behavior. Section 8 tests a series of possible interaction effects that might account for the observed variations in the slope of the learning curve. The results reveal that R&D expenditures and capital intensity influence the slope of the learning curve. Finally, Section 9 summarizes the findings of the study and suggests directions for future research.

---

* Stanford University.

The author would like to thank Richard Caves, A. Michael Spence, Michael Porter, Robert Stobaugh, Myron Scholes, James Patell, the Editorial Board, and numerous other colleagues for their helpful comments. Financial support was provided by the Strategic Management Program at Stanford Business School and by the Harvard Business School Division of Research.

2. Prior research on the learning curve

■ In the 1930s T. P. Wright (1936) observed that the direct labor cost of producing an airframe declined with the accumulated number of airframes produced, according to the function:

$$Y = aX^{-b}, \tag{1}$$

where Y is the direct labor cost, X is the cumulated production of airframes, a is the direct labor cost of the first airframe, and b is the learning "elasticity," which defines the slope of the learning curve. Wright's findings were corroborated by additional studies of aircraft produced during World War II. During the 1950s and 1960s, most research in the field addressed the question of whether this learning principle applied to the manufacture of other products. Although some variations were found in the learning curves for different products and firms, Wright's learning curve model received general confirmation.[1]

Work by the Boston Consulting Group (BCG) in the late 1960s led to further recognition of the wide applicability of the learning curve concept. BCG demonstrated that the learning curve (called the "experience curve" by BCG) encompassed not only labor costs, but also capital, marketing, and administrative costs. The learning curve was the aggregate result of labor learning, process improvement, product standardization, and economies of scale. BCG did not, however, attempt to decompose the learning curve into these elements. BCG simply plotted prices and costs as a function of cumulated industry output and found remarkably strong regularities in a wide range of industries.

Despite these numerous industry studies, evidence on the precise nature of the learning process is still quite limited. Several investigators have attempted to distinguish between static scale economies and dynamic learning effects (Preston and Keachie, 1964; Stobaugh and Townsend, 1975; Sultan, 1975; Hollander, 1965). In general, these studies have found static scale economies to be statistically significant but small in magnitude relative to learning-based economies. A few researchers have compared the performance of alternative proxies for the rate of learning. Rapping (1965) and Sheshinski (1967) found that cumulated output gave much stronger results than calendar time. Sheshinski also compared cumulated output and cumulated investment and obtained slightly better results for the latter. The appropriate functional form of the learning curve has never been rigorously tested, but many studies have demonstrated close conformance to the "log linear" model of equation (1), and this has emerged as the standard empirical approach.

Recent theoretical work has focused on the nature of competitive equilibria when learning effects are present. Spence (1981) has demonstrated that when learning occurs, part of the firm's short-run marginal cost can be regarded as an investment, which reduces the cost of production in future periods. Spence shows that in the absence of entry, a profit-maximizing firm lowers prices more slowly than costs, thereby causing short-run margins to widen over time. In the extreme case of no discounting, the model implies that firms should hold prices constant as costs fall. This counterintuitive result contrasts with the empirical finding that in most industries prices decline as learning proceeds, and profit margins remain approximately constant (Boston Consulting Group, 1972). The discrepancy can be resolved if the information acquired through learning diffuses across firms. Information spillovers reduce the proprietary "investment value" of additional output and shift the firm's optimal pricing policy towards one of maintaining constant profit margins (Lieberman, 1982). Recent empirical evidence suggests that such spillovers are extensive in most industries (Boston Consulting Group, 1978; Lieberman, 1982).[2]

---

[1] Many of the World War II airframe studies are summarized in Asher (1956); the most comprehensive airframe study is Alchian (1963). Early studies which extended the airframe results to other industries include Hirsch (1952, 1956), Rapping (1965), and Hirschmann (1964).

[2] These spillovers also diminish the competitive cost advantages potentially generated by the learning curve. This fact was initially overlooked by strategic planners and consultants who promoted aggressive learning curve-based strategies to their business clients during the late 1960s and early 1970s.

3. Regression model

■ The model used in the study is:

$$P_{i,t} = a_i e^{b_0 t}(X_{1,i,t})^{b_1}(X_{2,i,t})^{b_2} \dots (X_{n,i,t})^{b_n} e^{u_{i,t}} \tag{2}$$

$$\log P_{i,t} = \log a_i + b_0 t + b_1 \log X_{1,i,t} + b_2 \log X_{2,i,t} + \dots + b_n \log X_{n,i,t} + u_{i,t},$$

where $P_{i,t}$ is the average market price for product i at time t, a_i is a product-specific constant term, $u_{i,t}$ is a random error term, and the X's correspond to the learning indexes or other variables used to control for changes in the level of price/cost margins. With a single learning index, equation (2) reduces to the conventional learning curve model shown in equation (1). The model is linear in logarithms and can be estimated by conventional regression techniques.

Strictly speaking, the learning curve applies to cost rather than price. Unfortunately, cost data are generally proprietary. The present study uses publicly available price data as a surrogate for cost. Reliable estimates of the learning curve can be obtained from historical price data, given any of the following conditions: price/cost margins remained constant over time; price/cost margins changed, but in a manner controlled for in the analysis; or changes in margins were small relative to changes in production cost. There are reasons to believe that one or more of these conditions held at least approximately for products in the sample over the coverage period.[3] Nevertheless, it is important to recognize that the price changes analyzed here stemmed partly from learning-based cost reductions and scale economies, and partly from shifts in profit margins.

4. Data

■ The data sample covers 37 chemical products over the years indicated in Table 1. Of the 37 products in the sample, 26 are organic chemicals, 7 are inorganic chemicals, 2 are synthetic fibers, and 2 are metals. For most products coverage begins in the late 1950s or early 1960s; coverage ends uniformly in 1972 to avoid problems associated with increases in the price of petroleum inputs. The average period of coverage for any given product is 13.3 years.

For each product, annual data were collected on nameplate production capacity by plant and firm, total industry output, and average market price. In addition, data were obtained on the average capital investment cost of new plants and the average R&D expenditures of the major firms included in the sample.[4]

The sample is limited to homogeneous, commodity-type chemical products.[5] The data sample includes all such products for which data could be obtained from public sources and which met the following criteria: at least 25% of industry output was sold through arms-length "merchant market" channels; joint products were not a major factor in production; and the annual rate of output had positive net growth over the course of the sample coverage period.

---

[3] If the number of producers and the elasticity of demand both remain constant, and if information acquired through learning tends to diffuse across firms, then profit maximization implies constant profit margins over time (Lieberman, 1982; Boston Consulting Group, 1974). Entry and changes in capacity utilization may influence margins; variables to control for these effects are incorporated in the analysis in Section 7. Some changes in margins probably persist in the data even after the use of these controls. But one would expect these margin changes to be small relative to the magnitude of price changes for most products in the sample. Deflated product prices declined by an average of 52% over the sample period. (Undeflated prices declined by 40%.)

[4] Data sources are described in detail in an appendix available directly from the author.

[5] Several products in the sample such as polyethylene and polyester fibers are slightly differentiated across producers.

TABLE 1 List of Products Included in Data Sample

Organic Chemicals	Inorganic Chemicals
Acrylonitrile (1959–1972)	Ammonia (1960–1972)
Aniline (1961–1972)	Carbon Black (1964–1972)
Bisphenol A (1959–1972)	Hydrofluoric Acid (1962–1972)
Caprolactam (1962–1972)	Sodium (1957–1972)
Carbon Disulfide (1963–1972)	Sodium Chlorate (1957–1972)
Cyclohexane (1956–1972)	Sodium Hydrosulfite (1964–1972)
Ethanolamines (1955–1972)	Titanium Dioxide (1964–1972)
Ethyl Alcohol (1958–1972)	
Ethylene (1964–1972)	*Synthetic Fibers*
Ethylene Glycol (1960–1972)	
Formaldehyde (1962–1972)	Acrylic Fibers (1960–1972)
Isopropyl Alcohol (1964–1972)	Polyester Fibers (1960–1972)
Maleic Anhydride (1959–1972)	
Methanol (1957–1972)	*Metals*
Neoprene Rubber (1960–1972)	
Pentaerythritol (1952–1972)	Aluminum (1956–1972)
Phenol (1959–1972)	Magnesium (1956–1972)
Phthalic Anhydride (1955–1972)	
Polyethylene-LD (1958–1972)	
Polyethylene-HD (1958–1972)	
Sorbitol (1965–1972)	
Styrene (1958–1972)	
1,1,1-Trichloroethane (1966–1972)	
Urea (1962–1972)	
Vinyl Acetate (1960–1972)	
Vinyl Chloride (1962–1972)	

The dependent variable is average market price measured in dollars per pound. This price is a "unit sales value" computed annually for each product by dividing the total dollar volume of noncaptive sales by the same sales volume measured in pounds.[6]

For each of the 37 products, the raw data on capacity and output were used to compute the following alternative learning indexes:

(1) *TIME*: The observation year serves as a time trend in the analysis.

(2) *CUMULATED INDUSTRY OUTPUT*: Historical data on total industry output of each product were cumulated through the end of each observation year.

(3) *CUMULATED INDUSTRY CAPACITY*: All additions to industry capacity (including expansions of existing facilities as well as new "greenfield" plants) were cumulated through the end of each observation year. This cumulated capacity measure provides a proxy for cumulated industry investment.

(4) *ANNUAL RATE OF INDUSTRY OUTPUT*: This is the absolute level of industry output during each observation year.

(5) *AVERAGE SCALE OF PLANT*: For each observation year, average plant scale was calculated by dividing total industry capacity by the total number of plants.

(6) *RATE OF NEW PLANT INVESTMENT ("NEWPLANT")*: This variable was computed in the same manner as industry cumulated investment except that capacity added through incremental expansion was excluded.[7]

A correlation matrix for these variables is given in Table 2. Although the variables are highly correlated, the large data sample contains sufficient variance to distinguish their independent effects.

[6] For synthetic fibers, the dependent variable is the published market price of a representative denier of fiber.

[7] For the regressions reported in Tables 4 and 5, *NEWPLANT* was computed by dividing the plant capacity added during each three-year observation period by the capacity in place at the beginning of the period.

TABLE 2 Correlation Matrix of Alternative Learning Indexes[1]

	TIME	*CUMULATED INDUSTRY OUTPUT*	*CUMULATED INDUSTRY CAPACITY*	*ANNUAL RATE OF OUTPUT*	*AVERAGE PLANT SCALE*	*NEWPLANT*
TIME	1.00					
CUMULATED INDUSTRY OUTPUT	.90	1.00				
CUMULATED INDUSTRY CAPACITY	.87	.95	1.00			
ANNUAL RATE OF OUTPUT	.91	.98	.93	1.00		
AVERAGE PLANT SCALE	.76	.82	.84	.80	1.00	
NEWPLANT	.79	.83	.93	.82	.70	1.00

[1] First-order correlation coefficients. All variables except time were converted to logarithms and then taken as deviations from means within each product group. The correlations are substantially lower when the data are analyzed in difference form.

In addition to these learning indexes, the following variables were used to control for shifts in price/cost margins:

(7) *RATE OF NEW MARKET ENTRY*: For each observation year, *ENTRY* is the capacity of plants built by firms that entered during the year, computed as a fraction of total industry capacity in place at the beginning of the year.
(8) *CAPACITY UTILIZATION LEVEL ("CU")*: For each observation year, industry average capacity utilization was obtained for each product by dividing total industry output by average total capacity.

5. Comparison of alternative learning indexes

■ Table 3 reports the regression results comparing the various learning indexes. The regressions were estimated by OLS based on the model in equation (2), assuming identical slope coefficients across the 37 products but allowing each product to have a separate constant term.[8] The Cochrane-Orcutt procedure was used to correct for serial correlation in the data. The entry and capacity utilization control variables are not included in these regressions, but their omission does not substantially affect the results.

The dependent variable in Table 3 is the average market price, deflated by a weighted index of input factor prices.[9] Deflation is appropriate, assuming that changes in factor input prices were passed through to output prices (net of the effect of learning curve economies). If undeflated output prices are used in the regressions, the results are virtually identical, except that the time trend is shifted upward by roughly 1.7% per year.

Figure 1 gives the correspondence between the learning curve "slope" and the learning "elasticity" represented by the regression coefficients. As conventionally defined, the learning curve "slope" is the level to which costs fall each time cumulated output doubles. For example, an "80% learning curve" implies that costs fall to 80% of their previous level for each doubling of cumulated output.

The results in Table 3 reveal that price (and by inference, cost) reductions in the chemical industry were linked to growth in cumulated industry output, investment in improved capital equipment, and scale economies at the plant level. Cumulated industry output is the best single proxy for learning. Learning does not appear to be a function of time; the time trend becomes insignificant once cumulated industry output is included in the regression. Similarly, the current output rate becomes insignificant once cumulated output is included.

The cumulated output and cumulated capacity indexes appear individually significant when included in the same regression, namely, regressions (5) and (6). This indicates

[8] The regressions were also estimated with individual constant terms for each observation year as well as for each product. Results were virtually identical to those reported in Table 3. The assumption that the residuals are distributed independently across products was verified by computing correlation coefficients between the residuals for each pair of products in the sample. The distribution of these 666 correlation coefficients conformed closely to the t-distribution expected under the null hypothesis of independence. Given the small number of observations per product, however, it was impossible to perform the more conclusive likelihood ratio test of independence proposed by Box (1949). Several additional tests were made to determine whether the log-linear specification represented the most accurate functional form for the learning curve. Quadratic terms were evaluated in the regressions in an effort to detect nonlinearities, and separate slope coefficients were estimated depending on the growth rate of industry cumulated output. No significant departures from log-linearity could be detected, although a conclusive test requires use of cost data.

[9] The deflator is a weighted average of four price series: the producer price index of "fuels and related products" (30%); the GNP deflator for "industrial inorganic and organic chemicals" (30%); the Chemical Engineering Plant Construction Cost Index (30%); and the index of average hourly earnings of production workers in the chemical and allied products industries (10%). These weights are believed to correspond roughly to the average factor proportions of products included in the sample.

TABLE 3 Comparison of Alternative Learning Indexes* (Dependent Variable: *AVERAGE MARKET PRICE*)[1]

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
TIME	−.055[a]	−.003	−.014[a]	−.029[a]	.002	.001	−.002	.002	.002
	(−37.7)	(−.8)	(−5.3)	(−12.0)	(.5)	(.4)	(−.6)	(.6)	(.6)
CUMULATED INDUSTRY OUTPUT		−.471[a]			−.303[a]	−.283[a]	−.387[a]	−.335[a]	−.355[a]
		(−17.0)			(−8.8)	(−6.8)	(−12.9)	(−11.3)	(−9.9)
CUMULATED INDUSTRY CAPACITY			−.413[a]		−.234[a]	−.230[a]			.064
			(−16.3)		(−7.5)	(−7.3)			(1.0)
ANNUAL RATE OF INDUSTRY OUTPUT				−.286[a]		−.025			
				(−12.1)		(−.9)			
AVERAGE PLANT SCALE							−.173[a]	−.141[a]	−.162[a]
							(−6.0)	(−5.1)	(−4.6)
NEWPLANT								−.132[a]	−.161[a]
								(−6.9)	(−4.6)
R^2	.743	.836	.829	.800	.853	.853	.847	.861	.861
D.F.	455	454	454	454	453	452	453	452	451

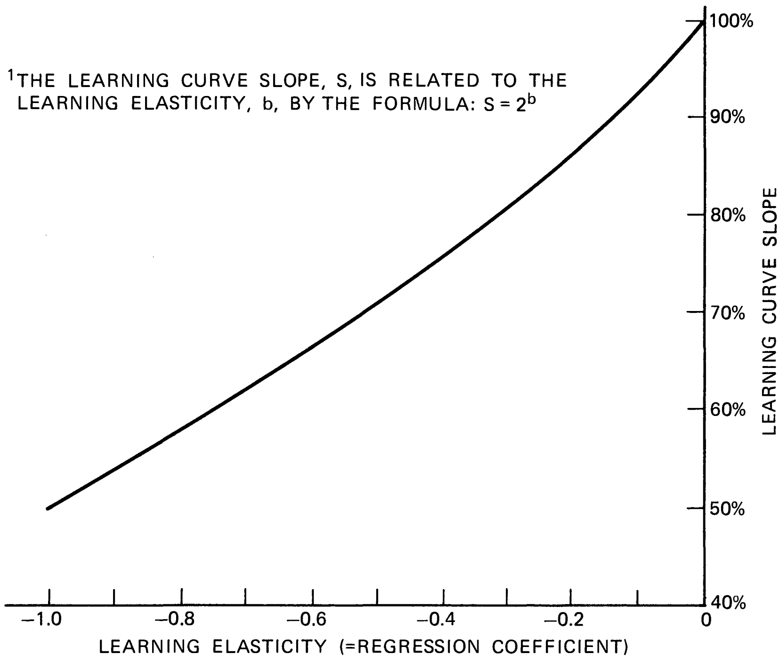
* All variables (except time) are specified in log form. Numbers in parentheses are *t*-statistics. Table omits individual constant terms estimated for each product.

[1] Adjusted for input price changes as described in text.

[a] Significant at .01 level, one-tailed test.

FIGURE 1

CORRESPONDENCE BETWEEN "LEARNING ELASTICITY" AND "LEARNING CURVE SLOPE"[1]



that they are linked to independent components of the overall learning effect.[10] The cumulated capacity index captures efficiency improvements stemming from both increases in average plant scale and technical improvements embodied in new state-of-the-art plants. These effects can be evaluated separately by using the average plant scale and *NEWPLANT* variables as in regressions (7) and (8). Regression (9) reveals that the cumulated capacity index has no explanatory power after controlling for the scale and new plant effects.

The average plant scale coefficient in regression (7) indicates that on average each doubling of plant scale was accompanied by an 11% reduction in unit costs. This finding is consistent with the "six-tenths rule" commonly used as a rule of thumb to assess scale economies in the chemical industry.[11]

It is instructive to compare the magnitude of plant scale economies with that of cost reductions attributable to the learning curve. This can be accomplished by multiplying the regression coefficients (which represent point elasticities) by the average rates of change of the underlying variables. For the overall sample, prices declined at an average rate of 5.5%

[10] Cumulated capacity remains highly significant in the regressions even if all joint effects are assigned to cumulated output. (This was accomplished by regressing cumulated capacity on cumulated output and by using the corresponding residuals in lieu of the capacity index in regression (5).) The average plant scale and *NEWPLANT* variables also remain highly significant when tested in this manner.

[11] The "six-tenths rule" dictates that capital investment costs rise in proportion to plant capacity raised to the .6 power. To put the 11% figure in perspective, note that the "six-tenths rule" implies a 24% reduction in unit capital costs for each plant scale doubling. In the chemical industry, labor costs are sometimes assumed to follow a "two-tenths rule," which suggests a 42% reduction in unit labor costs for each doubling of plant scale. Fuel and material cost savings may also be forthcoming, but at a much lower rate. Given that fuel and materials typically account for the largest proportion of total chemical product costs, the measured scale coefficient is consistent with the prevailing rules of thumb. For example, assume that for a given product, capital costs account for 30% of total production costs, labor costs account for 10%, and fuel and materials account for the remainder. Using the conventional rules of thumb, each doubling of plant scale would induce capital cost savings equal to 7% of total costs and labor cost savings equal to 4%.

per year. Assuming constant margins, production costs declined at an equivalent rate. Based on the coefficients in regression (7), about 15% of this cost reduction can be allocated to increases in average plant scale and about 85% to increases in cumulated industry output. Thus, the effect of plant scale economies was comparatively small.

6. Estimation of individual product learning curves

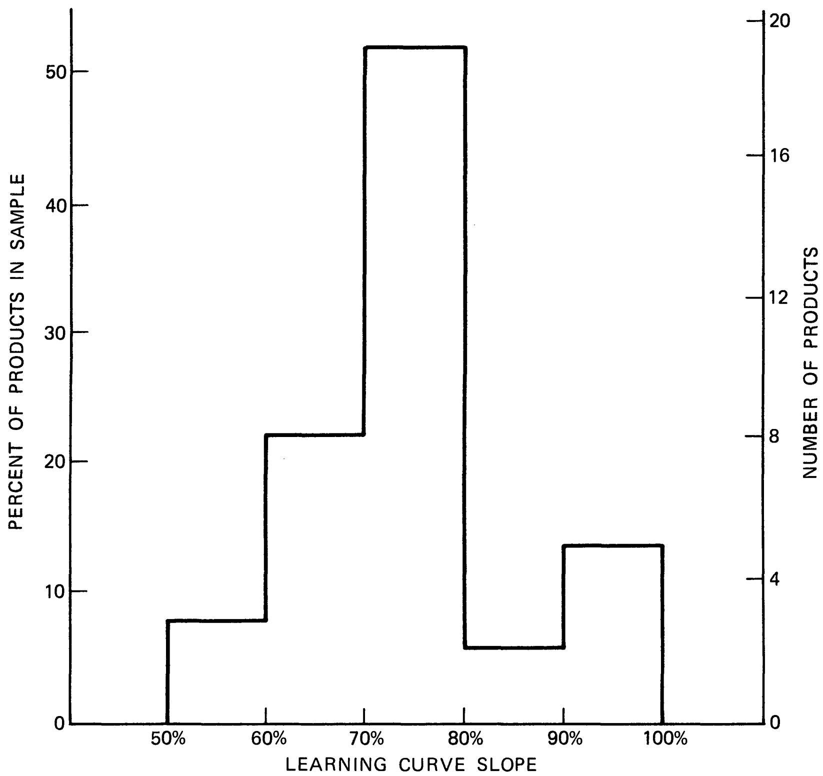
■ The regressions in Table 3 assume that all products in the sample share a common learning curve slope. To test this uniform slope assumption, regression (7) was estimated in a manner that allowed the industry cumulated output variable to have separate slope coefficients for each of the 37 products in the sample.

Figure 2 is a histogram that summarizes the resulting distribution of learning curve slope estimates for the 37 products in the sample. Sampling errors and shifts in price/cost margins cause this histogram to have greater dispersion than the underlying "true" slope distribution. Nonetheless, most of the estimated slopes fall in the expected range. More than half of the products in the sample have estimated slope coefficients between 70% and 80%.

In general, the uniform slope model performs remarkably well; it accounts for more than 90% of the variance explained by the differential slope model.[12] Given the large sample size, however, the hypothesis that the slopes are all identical can easily be rejected at the

FIGURE 2

DISTRIBUTION OF ESTIMATED LEARNING CURVE SLOPES



[12] R-squared equals .85 for the uniform slope model (regression (7)), versus .90 when each product is allowed a separate slope.

.01 level.[13] Factors which might explain these small but significant slope variations are investigated in Section 8.

7. Effect of producer concentration on the timing of price changes

■ The results in Table 3 reveal that over the long term, prices in the chemical industry declined in close conformance with the learning curve model. Over the short term, however, one might expect the timing of price changes to have been influenced by market concentration and other structural factors. Indeed, scanning the data revealed a sharp contrast in short-run price behavior between high and low concentration markets. Using the Herfindahl index (H) as a measure of producer concentration, the change in price behavior appeared at a concentration level of roughly $H = .2$.[14] This $H = .2$ threshold also divided the sample into two approximately equal parts.

Table 4 gives the results when separate regression coefficients are fitted for the high and low concentration subsamples. The procedure in Table 4 is equivalent to estimating individual regressions for the high and low concentration subsamples, while constraining the scale and time trends to have identical coefficients for both subsamples (since no significant differences in these coefficients were expected or found).

For the regressions in Table 4 the data were differenced over three-year nonoverlapping time periods. Differencing the data in this manner puts greater emphasis on short-term price changes, while eliminating serial correlation of the residuals and the need for product-specific intercept terms. The dependent variable is the percentage change in average market price over each three-year interval. The price data were not deflated; consequently, the average rate of input price inflation appears as part of the time trend.

Regression (10) reveals the absence of any significant difference in learning curve slope between the high and low concentration subsamples, based on cumulated output as the learning index. When the rate of new plant construction is added to the regression (regression (11)), however, the results for the low concentration ($H < .2$) subsample reveal that substantial price reductions occurred when new plants were opened. If the rate of entry is added to the regression to distinguish between plants built by entrants and those built by incumbents (regression (12)), the entry measure proves insignificant for the low concentration subsample. This reveals that in low concentration markets, price reductions were triggered by the completion of new plant facilities regardless of whether these were built by incumbents or new entrants.

For the high concentration subsample, only the entry effect proves statistically significant. Price reductions occurred when plants were completed by new entrants; plants added by incumbents had no effect on the average price level. Nevertheless, the average effect of entry was less than half of that observed in the low concentration subsample.[15] The coefficients in regression (12) indicate that on average, prices in the high concentration subsample declined by about .3% for each 1% of total industry capacity added by new entrants. In the low concentration subsample, prices declined by about .8% for each 1% of industry capacity added by either entrants or incumbents.

---

[13] The likelihood ratio statistic is 187, which exceeds the critical chi-squared value of 56 (36 *D.F.*) required to reject the homogeneous slope hypothesis at the .01 level. Properties of this test statistic are discussed in Savin (1976).

[14] The Herfindahl index is defined by the formula

$$H = \Sigma s_i^2,$$

where s_i is the market share of firm i. Note that the $H = .2$ concentration level implies five firms of equal size (or more firms if sizes differ). Varying the concentration threshold within the range $.15 < H < .30$ had little effect on the results.

[15] The effect of entry is obtained by adding the *ENTRY* and *NEWPLANT* coefficients.

TABLE 4 Effect of Seller Concentration on Price Flexibility and the Timing of Learning-Related Price Changes* (Dependent Variable: *AVERAGE MARKET PRICE;* Data in 3-Year Difference Form)

		(10)	(11)	(12)	(13)
TIME[a]		.016[b]	.024[c]	.025[c]	.028[c]
		(.009)	(.008)	(.008)	(.008)
AVERAGE PLANT SCALE		−.163[c]	−.092[b]	−.101[b]	−.122[b]
		(.064)	(.056)	(.055)	(.055)
INDUSTRY CUMULATED OUTPUT	$H < .2$	−.460[c]	−.218[c]	−.217[c]	−.376[c]
		(.085)	(.082)	(.081)	(.087)
	$H > .2$	−.390[c]	−.458[c]	−.439[c]	−.481[c]
		(.083)	(.075)	(.075)	(.073)
NEWPLANT	$H < .2$		−.799[c]	−.763[c]	−.521[c]
			(.113)	(.131)	(.138)
	$H > .2$		−.096	.044	.134
			(.119)	(.134)	(.138)
ENTRY	$H < .2$			−.076	.102
				(.150)	(.153)
	$H > .2$			−.322[b]	−.338[c]
				(.147)	(.144)
CU $H < .2$	$0 < K < 15$				.018
					(.317)
	$15 < K < 30$				.716[c]
					(.256)
	$30 < K < 60$				1.50[c]
					(.39)
CU $H > .2$	$0 < K < 15$				.102
					(.208)
	$15 < K < 30$				.303
					(.193)
	$30 < K < 60$				.265
					(.184)
R^2		.278	.476	.495	.577
D-W		2.18	2.05	2.03	2.05
D.F.		135	133	131	125

* Numbers in parentheses are standard errors. All variables except *NEWPLANT, ENTRY,* and *CU* are specified in log form.

[a] Time trend equals constant term divided by 3.

[b] Significant at .05 level, one-tailed test.

[c] Significant at .01 level, one-tailed test.

Thus, it appears that prices in highly concentrated markets declined relatively smoothly, perturbed only slightly by the appearance of new entrants. In less concentrated markets, prices declined sharply each time new (and generally more efficient) plants came on stream, regardless of whether these were built by entrants or by incumbents. Despite these short-term differences in pricing behavior, the long-term rate of price reduction was not significantly affected by market concentration.

Conceivably, the *NEWPLANT* and *ENTRY* effects observed in regressions 11 and 12 might simply be the result of depressed capacity utilization. To check this possibility, controls for capacity utilization were added to the analysis, as shown in regression (13). The capacity utilization terms were obtained by computing three-way interaction variables using industry capacity utilization (*CU*), the Herfindahl index of market concentration

(H), and the capital intensity of the production process (K).[16] The capacity utilization terms reveal a significantly greater degree of price flexibility in the low concentration industries. Moreover, the extent of price fluctuation increases with capital intensity, as basic microeconomic theory would predict.

Inclusion of the capacity utilization terms causes the *NEWPLANT* coefficient for low concentration industries to decline to about two-thirds of the magnitude indicated in regressions (11) and (12). This suggests that on average about one-third of the "new plant effect" is linked to depressed capacity utilization and is therefore transitory.

8. Factors influencing the learning curve slope: tests for interaction effects

■ The previous section demonstrated that market concentration affects the short-run timing of learning-related price reductions, but has no discernable effect on the slope of the learning curve. This section examines a number of potential interaction effects, which might explain the variations in learning curve slope documented in Section 6.

A series of learning curve interaction variables were defined on the basis of differences in the level of industry R&D expenditures and differences in product and process type. Interaction terms were computed either multiplicatively or by defining a series of cut points in the range of each interaction variable and allowing the industry cumulated output variable to have a separate slope coefficient over each part of the range. As several of the interaction variables varied across products and also over time, it was necessary to analyze the data in difference form. Regression (13), with the Herfindahl split for the industry cumulated output variable omitted, was used as a control equation to which interaction terms were added. Table 5 gives the results for some of the more significant interaction tests.

The following interaction variables were tested:

Level of research and development expenditures. Regressions (14) and (15) test the hypothesis that the rate of cost reduction is related to the level of expenditures on process-oriented R&D. R&D expenditures might reduce costs either directly or through an interaction effect with the learning curve. To test these two related hypotheses, data on yearly R&D expenditures and total company sales were obtained from the COMPUSTAT data files for the period 1960 to 1969. An average R&D to sales ratio was computed for each company for which data were available. A weighted average industry R&D to sales ratio was then calculated for each annual product observation by using firms' capacity shares as weights. This procedure is obviously fairly crude, and the resulting R&D to sales ratios can be considered only rough estimates. Nevertheless the procedure does capture major differences in R&D intensity across product lines as well as shifts in R&D intensity over time due to the entry and exit of research-intensive firms.

In regression (14) the R&D to sales ratio is included as an explanatory variable to test whether the rate of price reduction depends directly on the level of research and development expenditures. The resulting R&D/sales coefficient is not statistically significant, although it appears with a negative sign which is consistent with the hypothesis.

In regression (15), the R&D to sales ratio and the cumulated output variable are combined in a multiplicative interaction term. This allows a test of the hypothesis that research and development expenditures accelerate the rate of cost reduction associated with

[16] The variable K is capital cost expressed as a percentage of total production cost, based on engineering data for new plants built in the late 1960s. The capacity utilization measure used in the regression applies to the intermediate year of each three-year observation period. This specification implies a one-year lag in price adjustment, which is consistent with the fact that one-year supply contracts were in effect for many products in the sample over the coverage period.

TABLE 5 Learning Curve Interaction Tests* (Dependent Variable: *AVERAGE MARKET PRICE;* Data in 3-Year Difference Form)

			(14)	(15)	(16)	(17)
TIME[a]			.042[b]	.025[b]	.025[b]	.014
			(.013)	(.007)	(.008)	(.009)
INDUSTRY CUMULATED OUTPUT			−.429[b]	−.141		
			(.067)	(.144)		
R&D/S			−1.49			
			(1.05)			
R&D/S × *INDUSTRY CUMULATED OUTPUT*				−8.16[b]		
				(3.46)		
INDUSTRY CUMULATED OUTPUT	$0 < K < 15$				−.338[b]	
					(.096)	
	$15 < K < 30$				−.439[b]	
					(.067)	
	$30 < K < 60$				−.476[b]	
					(.094)	
INDUSTRY CUMULATED OUTPUT	Primary Organic					−.180
						(.161)
	Intermediate Organic					−.327[b]
						(.096)
	Final Organic					−.357[b]
						(.091)
	Synthetic Fibers					−.441[b]
						(.073)
	Inorganic					−.185
						(.163)
	Metallic					.046
						(.216)
AVERAGE PLANT SCALE			−.108[c]	−.103[c]	−.110[c]	−.110[c]
			(.055)	(.054)	(.055)	(.055)
NEWPLANT	$H < .2$		−.464[b]	−.515[b]	−.441[b]	−.465[b]
			(.119)	(.120)	(.116)	(.119)
	$H > .2$		.078	.059	.094	.099
			(.134)	(.132)	(.134)	(.135)
ENTRY	$H < .2$		.123	.072	.127	.150
			(.150)	(.150)	(.150)	(.149)
	$H > .2$		−.371[b]	−.394[b]	−.386[b]	−.384[b]
			(.143)	(.142)	(.144)	(.143)
CU	$H < .2$	$0 < K < 15$	.103	.152	.107	.023
			(.317)	(.314)	(.317)	(.314)
		$15 < K < 30$	.819[b]	.729[b]	.830[b]	.839[b]
			(.231)	(.232)	(.230)	(.235)
		$30 < K < 60$	1.533[b]	1.431[b]	1.552[b]	1.541[b]
			(.379)	(.377)	(.377)	(.375)
	$H > .2$	$0 < K < 15$	.112	.097	.126	.079
			(.208)	(.205)	(.209)	(.207)
		$15 < K < 30$	.279	.273	.273	.262
			(.191)	(.189)	(.192)	(.191)
		$30 < K < 60$	.249	.289	.251	.190
			(.182)	(.181)	(.188)	(.183)
R^2			.577	.588	.579	.597
D-W			2.09	2.14	2.06	2.17
D.F.			125	125	124	121

* Numbers in parentheses are standard errors.
[a] Constant term divided by 3.
[b] Significant at the .01 level, one-tailed test.
[c] Significant at the .05 level, one-tailed test.

the learning curve. The interaction coefficient proves significant at the .01 level, indicating strong confirmation of the interaction hypothesis.

Capital-intensity of the production process. Opportunities for long-run cost reduction may be related to the capital intensity of the production process. If the scope for efficiency improvement is positively related to process complexity, then the learning curve slope should increase with capital intensity. On the other hand, if the learning effect stems primarily from learning by direct labor, then the learning curve slope should diminish as capital intensity rises and labor intensity falls.

This hypothesis is tested in regression (16). The results indicate a significant increase in the slope of the learning curve as capital intensity rises.[17]

An alternative measure of capital intensity is the total value of equipment contained within a typical plant, rather than the amount used per unit of output. A variable corresponding to the average capital investment cost of new plants was tested in the same manner as the capital intensity variable in regression (16). Similar results were obtained.

Type of manufacturing process: continuous vs. batch. The sample was divided into two groups depending on whether most plants operated in continuous or batch mode. No significant difference in learning curve slope could be detected between the continuous and batch production processes.

Extent of multiplant operation. If process improvements acquired through learning remain proprietary, then one might expect a link between the rate of cost reduction and the extent of multiplant operation. Assume that information acquired through learning diffuses rapidly across multiple plants belonging to a single firm, but only slowly (if at all) across plants belonging to different firms. For a given number of industry plants, cost reduction should be more rapid if production is controlled by a smaller number of firms.

A set of interaction variables was defined to test this multiplant hypothesis. No evidence of an interaction effect between multiplant operation and learning-based economies could be detected. This suggests a high rate of information diffusion across firms.

Category of product. Products were divided into six groups: (1) primary organic chemicals, (2) intermediate organic chemicals, (3) end product organic chemicals, (4) synthetic fibers, (5) inorganic chemicals, and (6) metals. Separate learning curve slopes were estimated for each product group, as shown in regression (17).

The results for the organic chemicals and fibers suggest that the slope of the learning curve may increase as products move further along the chemical processing chain. But this pattern is not statistically significant and may simply result from inadequate controls for differences in factor inputs and the inflation rate of raw materials costs. The metals product group (which includes only two products) is the only category for which the slope coefficient is significantly different from that of the aggregate sample. No rationale for this difference in slope could be identified.

These interaction tests suggest that the slope of the learning curve is affected by R&D intensity, capital intensity, and product type. The R&D effect is of greatest interest. The results suggest that the level of R&D expenditure does not translate directly into an average rate of cost reduction. Rather, R&D tends to accelerate the learning process and to increase the steepness of the learning curve. One interpretation is that the learning curve relation

[17] The learning curve slope for low capital intensity (defined as capital costs accounting for less than 15% of product price) is significantly less steep than the slopes for intermediate and high capital intensity. If the interaction term corresponding to low capital intensity is added to the model specified in regression (13), the t-value of the resulting coefficient is 1.91. This exceeds the value of 1.65 required to reject the null hypothesis of no effect at the .05 level.

defines the overall potential for process improvement and hence the manner in which returns to R&D diminish over time.

The capital intensity result contradicts the commonly held belief (based on early studies from the 1940s and 1950s) that the learning curve results primarily from increases in direct labor efficiency within a given plant. The long-run learning curves observed here stem from many accumulated improvements embodied in plant and equipment. Such improvements have clearly been important in the chemical industry, which is quite capital-intensive overall.

9. Summary and conclusions

■ This study has disaggregated the learning curve into some of its component elements and examined several determinants of industrial price behavior. The results document the pervasiveness of the learning curve in the chemical industry. The learning curve appears to be a function of cumulated output and cumulated investment rather than calendar time, and it is distinct from standard economies of scale. Scale economies are important in the chemical industry, but they account for only a small fraction of the cost reductions indicated.

Although the individual learning curves identified for the 37 products in the sample are remarkably uniform, there are some small but significant differences in learning curve slope across products. R&D expenditures (or the underlying technological opportunities which give rise to them) appear to steepen the learning curve. Higher capital intensity has a similar effect. No relation could be detected between the slope of the learning curve and process type or the extent of multiplant operation.

The R&D result is of particular interest because it provides a new perspective on the dynamics of productivity improvement and technical change. Economic research on R&D and innovation has tended to ignore the learning curve relation despite widespread evidence documenting its importance as a powerful empirical regularity. "Learning-by-doing" and "learning-by-spending" on R&D are closely linked in practice, as some industrial case studies have revealed.[18] Additional research is needed to explore the nature of this interaction in greater detail.

The regression results on short-run price behavior provide dramatic illustration of the workings of competitive markets versus oligopoly. Price behavior in the chemical industry is strongly linked to market structure. For products with numerous producers, prices tend to be flexible and respond rapidly to the appearance of new, more efficient plants. For products with fewer than five or six producers, price changes tend to be more gradual, except when price reductions are triggered by new entry. These results represent a rare opportunity to observe market price behavior across a large sample of well-defined products.

References

ALCHIAN, A. "Reliability of Progress Curves in Airframe Production." *Econometrica,* Vol. 31 (October 1963), pp. 679–693.

ASHER, H. *Cost-Quantity Relationships in the Airframe Industry.* The Rand Corporation, 1956.

BOSTON CONSULTING GROUP. "Perspectives on Experience." Technical Report. Boston: 1972.

———. "The Experience Curve Reviewed: Price Stability." Technical Report. Boston: 1974.

———. "Cross Sectional Experience Curves." Technical Report. Boston: 1978.

BOX, G.E.P. "A General Distribution Theory for a Class of Likelihood Criteria." *Biometrica,* Vol. 36 (1949), pp. 317–346.

HIRSCH, W.Z. "Manufacturing Progress Function." *Review of Economics and Statistics,* Vol. 34, No. 2 (1952), pp. 143–155.

———. "Firm Progress Ratios." *Econometrica,* Vol. 24, No. 2 (April 1956), pp. 136–144.

HIRSCHMANN, W.B. "Profit from the Learning Curve." *Harvard Business Review,* Vol. 42, No. 1 (January 1964), pp. 125–139.

[18] For an insightful study on the chemical industry, see Hollander (1965).

HOLLANDER, S. *The Sources of Increased Efficiency: A Study of DuPont Rayon Plants.* Cambridge: MIT Press, 1965.

LIEBERMAN, M.B. "The Learning Curve, Pricing, and Market Structure in the Chemical Processing Industries." Ph.D. Dissertation, Harvard University, 1982.

PRESTON, L. AND KEACHIE, E.C. "Cost Functions and Progress Functions: An Integration." *American Economic Review,* Vol. 54, No. 2 (March 1964), pp. 100–107.

RAPPING, L. "Learning and World War II Production Functions." *Review of Economics and Statistics,* Vol. 48, No. 1 (1965), pp. 81–86.

SAVIN, N.E. "Conflict among Testing Procedures in a Linear Regression Model with Autoregressive Disturbance." *Econometrica,* Vol. 44, No. 6 (November 1976), pp. 1303–1315.

SHESHINSKI, E. "Tests of the Learning-by-Doing Hypothesis." *Review of Economics and Statistics,* Vol. 49, No. 4 (1967), pp. 568–578.

SPENCE, A.M. "The Learning Curve and Competition." *Bell Journal of Economics,* Vol. 12, No. 1 (Spring 1981), pp. 49–70.

STOBAUGH, R.B. AND TOWNSEND, P.L. "Price Forecasting and Strategic Planning: The Case of Petrochemicals." *Journal of Marketing Research,* Vol. 12 (February 1975), pp. 19–29.

SULTAN, R.G.M. *Pricing in the Electrical Oligopoly,* Vol. 2. Division of Research, Graduate School of Business Administration, Harvard University, 1975.

WRIGHT, T.P. "Factors Affecting the Cost of Airplanes." *Journal of Aeronautical Sciences,* Vol. 3, No. 4 (1936), pp. 122–128.