The Lieberman Chemical FIVE Data

by

Marvin Lieberman

Professor of Strategy
Anderson School of Management
UCLA

Contents

1. Conditions of Use

The data sets described in this document are part of the FIVE Project on Firm and Industry Evolution and Entrepreneurship, organized by Constance Helfat (Tuck School of Business at Dartmouth) and Steven Klepper (Carnegie-Mellon University). If you download or otherwise obtain any of the data sets described in this document, or obtain modified versions of these data sets, you agree to abide by the conditions of use set forth here.

The organizers of the FIVE Project, Constance Helfat and Steven Klepper, are not responsible for the content of any of the FIVE Project data sets. The FIVE Project organizers and the FIVE Project data set authors make no representations or warranties of any kind concerning any of the FIVE Project data, including with regard to the absence or presence of errors.

All FIVE Project data sets are protected by copyright. The copyright holders for each data set are indicated on the first page of the documentation for each data set. Permission is granted to reproduce FIVE Project data for non-profit educational and research use only. Requests to reproduce FIVE Project data for other uses should be addressed to: Professor Constance Helfat, Tuck School of Business at Dartmouth, Hanover, NH 03755, constance.helfat@dartmouth.edu.

This document describes the Lieberman Chemical FIVE Data, which should be referred to by this name in any derivative works. In any written and published work, users of these data should cite this document, one of the data set author's original articles (see below, Section 4), and the FIVE Project: Data Overview (Helfat & Klepper, 2007).

You may create a new data set by modifying a FIVE Project data set, including but not limited to adding or removing data or merging data from different FIVE Project data sets. If you utilize any FIVE Project data to create new data sets, within six years from the date that you obtain the FIVE Project data, you must provide the FIVE Project with a copy of your modified data set by contacting: Constance Helfat, constance.helfat@dartmouth.edu.

Any written documents or statistical estimates that use FIVE Project data, including but not limited to working papers and publications, must be reported to: Constance Helfat, constance.helfat@dartmouth.edu.

Additional conditions of use (specified by the author of this document):
Users of the Lieberman Chemical FIVE data must not redistribute the data without written permission from IHS Inc.

2. Data Set Author

Marvin Lieberman
Professor of Strategy
Anderson School of Management
UCLA

3. Data Set Summary

The data describe 41 homogeneous, commodity-type chemicals or related products, between 1948 and 1983. In total, the data set consists of product level information for 1,177 product-years, with an average observation period of 28 years (Data File 1), firm-level data indicating which products each firm produced (Data File 2), and plant-level capacity data by product and firm for the aforementioned chemical industry products (Data Files 3 and 4). Data Files 3 and 4 contain the same information, but are organized differently. Data File 4 consists of pivot tables in an Excel spreadsheet. See Data Description for further explanation. Different articles by Lieberman (see Author's Research Using This Data Set, below) use various subsets of the data, and some products in these articles are not included in the FIVE data.

4. Author's Research Using This Data Set

Lieberman M. 1982. The Learning Curve, Pricing, and Market Structure in the Chemical Processing Industries. Doctoral dissertation, Harvard University.

Lieberman M. 1984. The Learning Curve and Pricing in the Chemical Processing Industries. *Rand Journal of Economics* 15: 213-228.

Lieberman M. 1987. "Patents, Learning by Doing, and Market Structure in the Chemical Processing Industries," *International Journal of Industrial Organization* 5: 257-276.

Lieberman M, Gilbert R. 1987. Investment and Coordination in Oligopolistic Industries. *Rand Journal of Economics* 18: 17-33.

Lieberman M. 1987. Excess Capacity as a Barrier to Entry: An Empirical Appraisal. *Journal of Industrial Economics* 35: 607-627.

Lieberman M. 1987. Post-Entry Investment and Market Structure in the Chemical Processing Industries.. *Rand Journal of Economics* 18: 533-549.

Lieberman M. 1987. Market Growth, Economies of Scale, and Plant Size in the Chemical Processing Industries. *Journal of Industrial Economics* 36: 175-191.

Lieberman M. 1989. "Capacity Utilization: Theoretical models and Empirical Tests." *European Journal of Operational Research* 40: 155-168.

Lieberman M. 1989. "The Learning Curve, Technology Barriers to Entry, and Competitive Survival in the Chemical Processing Industries," *Strategic Management Journal* 10: 431-447.

Lieberman M. 1990. Exit from Declining Industries: 'Shakeout' or 'Stakeout'? *Rand Journal of Economics* 21: 538-554.

Lieberman M. 1991. Determinants of Vertical Integration: An Empirical Test. *Journal of Industrial Economics* 39: 451-466.

5. Additional References

Helfat, CE & Klepper, S. 2007. FIVE Project: Data Overview. http://papers.ssrn.com/paper=1028022.

6. Data Set Sources

Capacity and output data come from the *Directory of Chemical Producers** (capacity data, annual issues, originally published by SRI International, now owned by IHS Inc.), *Synthetic Organic Chemicals* (output data, annual issues, published by the U.S. International Trade Commission), *Current Industrial Reports* (output data, published by the Census Bureau of the U.S. Department of Commerce), *Minerals Yearbook* (capacity data, annual issues, published by the U.S. Bureau of Mines), and *Textile Organon* (output data). Additional sources include Compustat (firm size, liquidity, R&D expenditures), *SRI Process Economics Program Handbook*, and various chemical trade journals.

7. Data Description

The products in the sample are all homogeneous, commodity-type chemicals or related products. For each product, the raw data collected include: (1) firm entry and exit from industry, (2) output volume and unit sales value at the industry level, (3) market concentration, (4) individual plant capacities by plant, firm, and year, (5) engineering data on capital investment costs and plant-level scale economies, (6) issued chemical patents. The data pertain to the US market and comprise a full sample of US producers.

7.1. Product/Industry Definition

The full sample includes more than 1,000 plants operated by about 200 individual firms. All products in the sample had positive net output growth from

the earliest year of coverage through at least 1975. Thus, the sample represents products with growing demand, although in a number of cases output declined after 1975. Rapid growth led to considerable entry.

The data include production capacity by plant and firm, observed on January 1 of each year, and total industry output of each product, observed over the course of each year. Thus, the capacity data are quite detailed at the individual plant level, whereas the output data are industry-level aggregates for each product.

Production capacities are well defined; except for chlorine, chemicals with production processes involving significant joint products have been excluded, as have those where capacity can be switched from one product to another in response to shifts in market demand. Plants are typically operated on a round-the-clock basis, assuming adequate demand. High storage costs generally prohibit the use of inventory to smooth out demand fluctuations extending over long periods (i.e., more than several weeks or months).

Vertical integration is common in the chemical industry, and many entrants in the sample were pursuing policies of forward or backward integration. Despite the presence of integration, all products in the sample had well-defined output markets. Although a sizable fraction of output was manufactured for captive use in firms' downstream operations, in all cases at least 25 percent of industry output was sold to independent parties through arms-length channels.

The products in the sample have cost functions that resemble the stylized examples of economic theory. Marginal cost for any given plant tends to be relatively constant up to the output level defined by the plant's rated production capacity. Plants for a given product may differ in cost, however, owing to differences in technology and age of plant. The learning curve is a salient factor contributing to long-term cost and price reductions in many industries, including the chemical products in these data. The sample spans a wide range of market concentration; the number of producers for any given product ranges from one to 64.

7.2. Excluded Fields

The data sets do not include R&D/sales, some cumulative industry process patent variables, and some firm level data used in the articles on vertical integration (Journal of Industrial Economics, 1991) and exit from declining industries (Rand Journal of Economics, 1990).

8. Variable List and Definitions: File 1: LiebermanChemProdFIVEdata

8.1. Year of Observation:

Year

[Variable name: *Year*; 2 digit numeric variable]: Year of observation, ranging from 1948 to 1983. The starting year varies by product. The range of years per product includes four years prior to the start of the annual capacity data.

8.2. Industry Submarket or Product Identifiers:

Product Name

[Variable name: *Product*; character string variable]: Name of the chemical product.

Product ID (product identifier assigned in the original data set)

[Variable name: *ProductID*; 2 digit numeric variable]: Unique identifier that was created for each product.

8.3. Entry/Exit Variables:

Number of Entrants

[Variable name: *NENTR*; numeric variable ranging from 0 to 10, -99.9 indicates missing data]: Number of new entrants during the year.

Entrant Capacity

[Variable name: *ENTRCAP*; numeric variable ranging from 0 to 1800, -99.9 indicates missing data]: Total annual capacity of entrants. All capacity data in the data set are measured in millions of pounds, with the following exceptions: Cyclohexane, ethyl alcohol, and methanol are measured in millions of gallons; chlorine, sodium hydrosulfite, sodium chlorate, titanium dioxide, ammonia, urea, aluminum, and magnesium are measured in thousands of tons (short tons, equal to 2000 pounds).

Number Exited

[Variable name: *NEXIT*; 1 digit numeric variable ranging from 0 to 9, -99.9 indicates missing data]: Number of firms that exited during the year.

Exit Capacity

[Variable name: *EXITCAP*; numeric variable ranging from 0 to 1000, -99.9 indicates missing data]: Total capacity of firms that exited during the year.

8.4. Industry Output Variables:

Industry Production

[Variable name: *Q*; numeric variable ranging from 0 to 29904, -99.9 indicates missing data]: Total annual industry production per product. All production data in the data set are measured in millions of pounds, with the following exceptions: Cyclohexane, ethyl alcohol, and methanol are measured in millions of gallons; chlorine, sodium hydrosulfite, sodium chlorate, titanium dioxide, ammonia, urea, aluminum, and magnesium are measured in thousands of tons (short tons, equal to 2000 pounds).

Cumulative Industry Production

[Variable name: *CUMQ*; numeric variable ranging from 1 to 429122, -99.9 indicates missing data]: Cumulative industry production per product through start of year. Includes estimated production prior to the start of the sample (see Lieberman 1982, Appendix A).

Average Capacity Utilization

[Variable name: *CU*; numeric variable ranging from 0 to 1.183, -99.9 indicates missing data]: Average annual capacity utilization per product (total industry production / average of beginning and end of year capacity).

8.5. Market Concentration Variables:

Number of Firms

[Variable name: *NFIRMS*; numeric variable ranging from 1 to 64, -99.9 indicates missing data]: Number of firms at start of year per product.

Number of Plants

[Variable name: *NPLTS*; numeric variable ranging from 0 to 98, -99.9 indicates missing data]: Number of plants at start of year per product.

Herfindahl Index

[Variable name: *HERF*; numeric variable ranging from 1 to 1, -99.9 indicates missing data]: Herfindahl index of producer concentration per product (based on producer capacities at start of year).

Entropy Index

[Variable name: *ENTROPY*; numeric variable ranging from 0 to 3.801, -99.9 indicates missing data]: Entropy index of producer concentration per product (based on producer capacities at start of year).

Capacity Share Largest Firm

[Variable name: *C1*; numeric variable ranging from 0 to 1, -99.9 indicates missing data]: Capacity share of largest firm per product (at start of year).

Capacity Share Two Largest Firms

[Variable name: *C2*; numeric variable ranging from 0 to 1, -99.9 indicates missing data]: Capacity share of the two largest firms per product (at start of year).

Capacity Share Three Largest Firms

[Variable name: *C3*; numeric variable ranging from 0 to 1, -99.9 indicates missing data]: Capacity share of the three largest firms per product (at start of year).

Capacity Share Four Largest Firms

[Variable name: *C4*; numeric variable ranging from 0 to 1, -99.9 indicates missing data]: Capacity share of the four largest firms per product (at start of year).

Multi-Plant Firms Capacity Share

[Variable name: *PCTMLT*; numeric variable ranging from 0 to 1, -99.9 indicates missing data]: Fraction of total industry capacity (per product) in multi-plant firms.

Multi-Plant Firms Proportion

[Variable name: *PCTMLT2*; numeric variable ranging from 0 to 1, -99.9 indicates missing data]: Fraction of total industry firms (per product) that are multi-plant.

8.6. Capacity Expansion Variables:

Industry Capacity

[Variable name: *TCAP*; numeric variable ranging from 0 to 39705, -99.9 indicates missing data]: Total industry capacity (per product) at start of year.

Number of New Plants

[Variable name: *NNEW*; numeric variable ranging from 0 to 12, -99.9 indicates missing data]: Number of new plants opened during year per product.

Incremental Plant Expansion

[Variable name: *NINC*; numeric variable ranging from 0 to 24, -99.9 indicates missing data]: Number of incremental plant expansions completed during year per product.

New Plant Capacity

[Variable name: *NEWPLT*; numeric variable ranging from 0 to 1950, -99.9 indicates missing data]: Total capacity of new plants completed during year per product.

Incremental Capacity

[Variable name: *INCCAP*; numeric variable ranging from 0 to 4055, -99.9 indicates missing data]: Total incremental capacity added to existing plants during year per product.

Capacity Closed

[Variable name: *CLOSED*; numeric variable ranging from 0 to 2322, -99.9 indicates missing data]: Total capacity closed by producers during the year per product.

Incremental Capacity Proportion

[Variable name: *INCNEW*; numeric variable ranging from 0.353 to 1, -99.9 indicates missing data]: Capacity added through incremental expansion as a fraction of total capacity added over entire sample period per product.

Incremental Capacity Proportion Prior 4 Years

[Variable name: *INCNEW4*; numeric variable ranging from 0 to 1, -99.9 indicates missing data]: Capacity added through incremental expansion as a fraction of total capacity added over prior four years per product.

Cumulative Industry Capacity Investment

[Variable name: *CUMCAP*; numeric variable ranging from 0 to 45825, -99.9 indicates missing data]: Cumulative industry capacity investment through start of year per product. Includes estimated cumulative capacity investment for the first year of capacity data per product (see Lieberman 1982, Appendix A, for further explanation).

"Benchmark" Plant Capacity

[Variable name: *PEPCAP*; numeric variable ranging from 0 to 1000]: Capacity of "benchmark" (typical) plant per product in 1976, based on estimates by SRI International. (This variable is called K76 in Lieberman, Journal of Industrial Economics, 1987.)

8.7. Production Cost Variables:

Fixed Investment Cost

[Variable name: *ABSCOST*; numeric variable ranging from 1 to 60, -99.9 indicates missing data]: Total fixed investment cost of a typical plant built in the mid-1970s

per product, in millions of dollars, based on SRI estimates. (This variable is called FIXED in Lieberman, Strategic Management Journal, 1987.)

Scale Economy Exponent

[Variable name: *SCOEF1*; numeric variable ranging from 0 to 0.88]: Scale economy exponent per product, based on SRI estimates. A larger value implies more extensive scale economies. (See description in Lieberman, Journal of Industrial Economics, 1987, page 182).

Cost Disadvantage

[Variable name: *CDR*; numeric variable ranging from 0 to 0.25]: Cost disadvantage ratio of small scale plants per product, based on SRI estimates. It is the average total unit cost of a small scale plant (one half benchmark scale) divided by the average total unit cost of a large scale plant (two times the benchmark scale). (See Lieberman, Journal of Industrial Economics, 1987, page 182. Note that the CDR measures are computed differently in Lieberman, Strategic Management Journal, 1989.)

8.8. Product Price Variables:

Unit Sales Value

[Variable name: *USV*; numeric variable ranging from 1.451 to 130.061, -99.9 indicates missing data]: Unit sales value (price), measured in cents per pound with the exception of cyclohexane, which is measured in cents per gallon.

Percent Change Unit Sales Value

[Variable name: *DELUSV*; numeric variable ranging from 0.69 to 2.18, -99.9 indicates missing data]: Unit sales value (USV) in the year of observation divided by unit sales value in the previous year. Percent change in Unit Sales Value from previous year.

Percent Change List Price

[Variable name: *DELIST*; numeric variable ranging from 0.40 to 1.87, -99.9 indicates missing data]: Percent change in list price from previous year.

Deflator

[Variable name: *DEFLATOR*; numeric variable ranging from 86.40 to 164, -99.9 indicates missing data]: Weighted average price deflator for chemical products. 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. (See Lieberman, Rand Journal of Economics, 1984, footnote 9.)

8.9. Patent Variables:

Number of Process Patents

[Variable name: *NPATENTS*; numeric variable ranging from 0 to 99]: Number of process patents (relating to production of the product) awarded to producers in the observation year.

Cumulative Process Patents from 1937

[Variable name: *CUMPAT37*; numeric variable ranging from 0 to 1325]: Cumulative number of process patents (relating to production of the product) awarded to producers from 1937 to start of year.

8.10. Other Variables:

Capital Investment Cost

[Variable name: *KINT*; numeric variable ranging from 0 to 71.5]: Estimate of capital investment cost as a percentage of total production cost per product, based on engineering data for new plants built in the late 1960s. (See Lieberman, Rand Journal of Economics, 1984 and Lieberman 1982, Appendix A.)

Imports

[Variable name: *IMPORTS*; numeric variable ranging from 0 to 794, -99.9 indicates missing data]: Total U.S. imports during the year (if available – partial coverage), measured in millions of pounds, with the following exceptions: Cyclohexane, ethyl alcohol, and methanol are measured in millions of gallons; chlorine, sodium hydrosulfite, sodium chlorate, titanium dioxide, ammonia, urea, aluminum, and magnesium are measured in thousands of tons (short tons, equal to 2000 pounds).

Exports

[Variable name: *EXPORTS*; numeric variable ranging from 0 to 1128, -99.9 indicates missing data]: Total U.S. exports during the year (if available – partial coverage), measured in millions of pounds, with the following exceptions: Cyclohexane, ethyl alcohol, and methanol are measured in millions of gallons; chlorine, sodium hydrosulfite, sodium chlorate, titanium dioxide, ammonia, urea, aluminum, and magnesium are measured in thousands of tons (short tons, equal to 2000 pounds).

9. Variable List and Definitions: File 2: LiebermanChemFirmFIVEdata

9.1. Firm Identifiers:

Firm Name

[Variable name: *firmname*; 40 character string variable]: Name of firm in the original data set.

FIVE Firm ID (unique firm identifier for the FIVE project)

[Variable name; *fivefirmID*; 7 digit numeric variable]: Unique identifier for each *Firm Name* that appears in the FIVE data sets. Firm names and FIVE firm identifiers may appear in more than one FIVE data set. If a firm's name changed over time, the FIVE Firm ID for the firm sometimes changes as well. In addition, different FIVE data set authors may have coded different Firm Names for the same company, which could result in different FIVE Firm IDs for the same firm in different data sets.

ORIGINAL CUSIP ID (firm identifier in original data set)

[Variable name: *originalCUSIPID*; 6 digit numeric variable]: Unique identifier per firm assigned in the original data set. ***There is no guarantee that these IDs are accurate.*** Not all firms had a CUSIP number in the original data set. Many of the firms without an original CUSIP ID were privately held or were foreign firms or joint ventures at the time.

CUSIP ID (firm identifier assigned by CRSP)

[Variable name: *CUSIPID*; 6 digit numeric variable]: Unique identifier for a firm as shown in the CRSP database as of 2013. Not all firms with a CUSIP ID are currently in operation. For a given Firm Name, this ID is constant across all years in the data set. These IDs come from matching each Firm Name in the FIVE data sets to an identical or similar firm name in CRSP. ***There is no guarantee that these matches and IDs are accurate.*** If no match was found, the CUSIP ID is blank. Joint ventures do not have CUSIP IDs in this file. Some companies have CUSIP IDs but not ORIGINAL CUSIP IDs.

PERMNO (permanent identifier assigned by CRSP)

[Variable name: *permno;* 5 digit numeric variable]: Unique permanent identification number assigned by CRSP to each security. Permno does not change during an issue's trading history, nor is it reassigned after an issue stops trading.

CUSIP Firm Name (company name assigned by CRSP)
[Variable name: *CUSIPname*; character variable]: Name of company given in CRSP for the CUSIP ID.

CUSIP Name Date (company name assigned by CRSP)
[Variable name: *CUSIPnamedate*; 4 digit numeric variable]: Year the current CUSIP Firm Name was assigned by CRSP.

Ownership Notes (information on company ownership)
[Variable name: *ownershipnotes;* character variable]: Notes from the original data set indicating changes in ownership, joint ventures, and foreign ownership.

Product-Firm Numbers (firm numbers by product)
[Variable name: *productfirmno*; 4 digit numeric variable]: These variables indicate which products each firm produced. Because many firms produced more than one product, there are multiple Product-Firm Number variables per firm. For each product-firm number, the numbers to the left of the decimal point correspond to a Product ID from File 1; the numbers to the right of the decimal point indicate the number assigned to that firm for the product to the left of the decimal point in Files 3 and 4 (which contain plant capacity data by firm for each product). Note that these firm numbers per product do not correspond to the FIVE Firm IDs. The firm numbers for each product begin at 1, and an individual firm that produced more than one product has more than one firm number in Files 3 and File 4.

10. Variable List and Definitions: File 3: LiebermanChemPlantFIVEdata

10.1. Industry Submarket or Product Identifiers:

Product ID (product identifier assigned in the original data set)
[Variable name: *ProductID*; 2 digit numeric variable]: Unique identifier that was created for each product.

10.2. Year of Observation:

Year
[Variable name: *Year*; 2 digit numeric variable]: Year of observation, ranging from 1948 to 1983.

10.3. Firm Identifiers:

Firm Number by Product (firm number within each product category)

[Variable name: *firm*; 2 digit numeric variable]: Firm number within each Product ID. The firm numbers are the same as those in the corresponding pivot table for each product in File 4. Many firms produced more than one product, and often have different firm numbers for different products. The firm numbers in this file and in File 4 differ from the FIVE Firm IDs. In addition, the firm numbers for each product are not necessarily continuous, because this file does not include firms that produced a particular product but for which capacity data were not available.

Plant Number by Firm and Product

[Variable name: *plant*; 2 digit numeric variable ranging from 1 to 12]: Numerical indicator for an individual plant, by firm, within a given product category. For each product category, the plant numbers for an individual firm start at 1.

Plant Capacity

[Variable name: *capacity*; 3 digit numeric variable ranging from 0 to 2900]: Annual capacity per plant, measured in millions of pounds, with the following exceptions: Cyclohexane, ethyl alcohol, and methanol are measured in millions of gallons; chlorine, sodium hydrosulfite, sodium chlorate, titanium dioxide, ammonia, urea, aluminum, and magnesium are measured in thousands of tons (short tons, equal to 2000 pounds). Note that this file does not include firms that produced a particular product but for which capacity data were not available.

11. Pivot Tables: File 4: LiebermanChemPivotFIVEdata

This file contains pivot tables from the original firm level data in an Excel spreadsheet. The first worksheet lists each product and product ID. Each subsequent worksheet contains a pivot table with capacity data for a single product, organized by year, firm, and plant. The names of the firms that produced each product are listed below each pivot table, along with the SIC Code for the firm (not the product), the stock market ticker symbol for the company (if publicly-held) in the original data set, and the original CUSIP IDs. For joint ventures, the SIC codes and CUSIPs IDs in the original data for both joint venture partners are given in the pivot tables where available. The original CUSIP IDs in File 2 do not contain any CUSIP IDs for joint ventures. Values of -9, 8888, 9999, or blanks indicate missing or unavailable data. Not all firms listed in the pivot tables as producing a particular product have capacity data for that product. By clicking on the pivot table entries, the plant level data can be

aggregated to the firm level. Note that the firm numbers given in each pivot table are not the same as the FIVE Firm IDs in File 2. The firm numbers in each pivot table begin at 1. Therefore, an individual firm that produced more than one product could have different firm numbers in different pivot tables. Note that for product 69, there is no firm with a firm number of 8; for product 82, there is no firm with a firm number of 24. The final worksheet contains a combined list of the firms in each of the pivot tables with their identifying information (product ID, firm number by pivot table, SIC code, original CUSIP ID, stock market ticker symbol, and firm name where available).

