AIM

Association for Information and Image Management

1100 Wayne Avenue, Suite 1100
Silver Spring, Maryland 20910
301/587-8202

Centimeter

Inches

MANUFACTURED TO AIM STANDARDS
BY APPLIED IMAGE, INC.

1 of 1

ANL/EAIS/TM-96

Modeling Plant-Level Industrial Energy Demand with the Manufacturing Energy Consumption Survey (MECS) Database and the Longitudinal Research Database (LRD)

by G.A. Boyd, M.J. Neifer, and M.H. Ross

Environmental Assessment and Information Sciences Division,
Argonne National Laboratory, 9700 South Cass Avenue, Argonne, Illinois 60439

January 1992 (revised August 1992)

MASTER

Work sponsored by United States Department of Energy,
Energy Information Administration, Office of Integrated Analysis and Forecasting, and
Deputy Assistant Secretary of Conservation and Renewables, Office of Industrial Technology

rb

CONTENTS

ACKNOWLEDGMENTS	vi
ABSTRACT	1
SUMMARY	1
1 OVERVIEW OF THE MECS/LRD	5
1.1 Introduction	5
1.2 Census of Manufactures	7
1.3 Annual Survey of Manufactures	7
1.4 Manufacturing Energy Consumption Survey	8
1.5 Energy Data Collected by the CM, ASM, and MECS	9
1.6 Longitudinal Aspects of the MECS/LRD: Data Editing	9
2 PRELIMINARY ANALYSIS OF THE MECS/LRD	12
2.1 Summary Statistics	12
2.1.1 All Energy Inputs	12
2.1.2 Electricity	15
2.1.3 Output	16
2.2 Energy Intensities	19
2.3 Technical Efficiency	27
2.3.1 Estimates of Best Practice Energy Use	32
2.3.2 Relationship between Plant-Level Energy Intensity and Input Efficiency	33
2.4 Plant-Level Prices	34
3 DISAGGREGATION OF INDUSTRIAL CHEMICAL SECTOR	46
4 CONCLUSIONS	52
5 REFERENCES	54
APPENDIX A: Measures of Inputs and Output Used in Calculation of Technical Efficiency	55
APPENDIX B: Description of Karlson's Database on Steel Industry Plants	57

TABLES

1 Total Number of Plants in the LRD	6
2 Number of Plants in the Six Industries in the LRD	10

TABLES (Cont.)

3	Number of Plants in the Six Industries with Continuous Records for 1972 through 1988 in the LRD	11
4	Number and Percentage of Plants Using Various Types of Energy	14
5	Amount and Percentage of Various Types of Energy Used	15
6	Hirfendahl Index of Energy Diversity	16
7	Amount and Percentage Spent for Various Types of Energy	17
8	Simple Correlation between Amounts Reported in the MECS and ASM	18
9	Energy Intensities	19
10	Descriptive Statistics for Three Measures of Total Energy Intensity in the Primary Aluminum Industry	27
11	Summary Statistics on Technical Efficiency Measures	30
12	Distribution of Technical Efficiency Measures	31
13	Estimates of Energy Consumption under Best Practice Technology Assumptions	33
14	Correlation between Technical Efficiency and Total Energy Intensity	34
15	Descriptive Statistics for Plant-Level Prices for Different Fuels	36
16	Estimates of Plant-Level Variations in Energy Prices Caused by the Location of the Plant and Amount of Energy Purchased by the Plant	43
17	Product Analysis for 1987 and Earlier Census of Manufacture Years	47
18	High-Energy-Consumption Products Considered	49
19	Basic Chemical Products for Which a Substantial Amount of Production Occurred Outside the Primary Industry in 1987	50

FIGURES

1	Distribution of Total Energy Intensities by Industry	21
2	Distribution of Net Electricity Intensities by Industry	22

FIGURES (Cont.)

3	Distribution of Fossil Fuel Intensities by Industry	23
4	Distribution of Other Energy Intensities by Industry	24
5	Comparison of Total Energy Intensities Computed on the Basis of Three Measures of Output	28
6	Example of a Technical Efficiency Calculation	29
7	Distribution of Total Energy Intensities by Industry	35
8	Distribution of Electricity Prices	38
9	Distribution of Distillate Fuel Oil Prices	39
10	Distribution of Natural Gas Prices	40
11	Distribution of LPG Prices	41
12	Distribution of Coal Prices	42
13	Distribution of Production of Chlorine	51

ACKNOWLEDGMENTS

We conducted the work discussed in this report at the U.S. Bureau of the Census, Center for Economic Studies. We gratefully acknowledge the support of the staff of the Center, particularly the assistance of Mark Doms and of Bob McGuckin, the director of the Center. Results from this report were presented at a meeting of the Midwest Economic Association in Chicago, Illinois, on March 26, 1992. We benefitted from the comments made by the participants at this meeting and by our session's discussant, Richard Kosobud. The plant-level data presented herein were subject to the standard review and disclosure rules of the U.S. Bureau of the Census.

**MODELING PLANT-LEVEL INDUSTRIAL ENERGY DEMAND
WITH THE MANUFACTURING ENERGY CONSUMPTION
SURVEY (MECS) DATABASE AND THE LONGITUDINAL
RESEARCH DATABASE (LRD)**

by

G.A. Boyd, M.J. Neifer, and M.H. Ross

ABSTRACT

This report discusses Phase I of a project to help the U.S. Department of Energy determine the applicability of the Manufacturing Energy Consumption Survey (MECS) database and the Longitudinal Research Database (LRD) for industrial modeling and analysis. Research was conducted at the U.S. Bureau of the Census; disclosure of the MECS/LRD data used as a basis for this report was subject to the Bureau's confidentiality restriction. The project is designed to examine the plant-level energy behavior of energy-intensive industries. In Phase I, six industries at the four-digit standard industrial classification (SIC) level were studied. The utility of analyzing four-digit SIC samples at the plant level is mixed, but the plant-level structure of the MECS/LRD makes analyzing samples disaggregated below the four-digit level feasible, particularly when the MECS/LRD data are combined with trade association or other external data. When external data are used, the validity of using value of shipments as a measure of output for analyzing energy use can also be examined. Phase I results indicate that technical efficiency and the distribution of energy intensities vary significantly at the plant level. They also show that the six industries exhibit monopsony-like behavior; that is, energy prices vary significantly at the plant level, with lower prices being correlated with a higher level of energy consumption. Finally, they show to what degree selected energy-intensive products are manufactured outside their primary industry.

SUMMARY

The U.S. Department of Energy (DOE) needs to analyze the role that industrial production processes will play in the future, in terms of the types and amounts of energy they will consume, the types and amounts of pollution they might cause, and all related costs. Yet macroeconomic models rarely have enough information on energy-intensive industries and products to allow for detailed forecasts at the process (or product) level. Process-level forecasts require, at a minimum, data at the four-digit standard industrial classification (SIC)

level. Moreover, national data series rarely have significant degrees of freedom upon which to build models that account for material and energy flows at the process level. The last time detailed data on industrial energy use at the end-use level were collected was for the 1980 Manufacturing Industries Survey of Large Combustors, and this project was canceled by the Office of Management and Budget before DOE could complete it (DOE 1982).

To address DOE's need, Argonne National Laboratory (ANL) has been performing two tasks. The first task has been to examine the Manufacturing Energy Consumption Survey (MECS) database and the Longitudinal Research Database (LRD) and their combined role in providing input data for a revised modeling system. The second task has been to directly explore how information from the combined MECS and LRD databases (MECS/LRD) can be used in industrial energy and economic forecasting. The initial focus has been on six four-digit industries that are highly homogeneous and represent the energy-intensive sectors of major materials industries: paper mills, except building paper (SIC 2621); alkalies and chlorine (SIC 2812); nitrogenous fertilizers (SIC 2873); cement, hydraulic (SIC 3241); blast furnaces and steel mills (SIC 3312); and primary aluminum (SIC 3334).

The MECS/LRD plant-level data make many types of comparisons and analyses possible. For example, energy intensities in "best practice" plants (i.e., those with minimal energy intensities) can be compared with the average energy intensity of their primary industry. Plant-level comparisons can be used to analyze how energy intensities depend on actual energy prices. (For energy-intensive industries, energy prices vary considerably, depending on plant location.)

The MECS/LRD data also allow production relationships to be examined at the plant level, even for the least homogeneous of the four-digit SICs studied in this report (i.e., the paper and steel industries). For these industries, additional insights like those just mentioned can be obtained by categorizing (i.e., disaggregating) plants by process or product to obtain a more homogeneous sector. The focus on homogeneous sectors allows for a more direct interpretation of the relationship between energy (and other) inputs and product outputs. The magnitude of the effect of changing the product mix both over time and among establishments is reduced when the selected sector's output is narrowly defined. The issue of product mix must ultimately be dealt with for all energy-intensive industries that require forecasts.

This report presents information on the results of the two tasks to date. It consists of two parts. The first part of the report is an overview of the MECS/LRD. It discusses the relationships among the Annual Survey of Manufactures (ASM), Census of Manufactures (CM), MECS, and LRD. In working on this task, ANL researchers found the Center for Economic Studies in the U.S. Bureau of the Census to be very helpful in enabling them to analyze data and disseminate results. They also found that the plant-level energy data were typically of good quality. The Bureau's limitations on the use of its data, while important, did not prevent most results of interest from being discussed in this report, because such results involved adequate aggregations of the data. The second part of the report summarizes the work accomplished to date in selected exercises designed to assess the

usefulness of the MECS/LRD data in energy analysis and forecasting. ANL researchers have found that extensive and important analyses can result from using the MECS/LRD data series (and from combining this data series with other plant-level series, from either the Bureau or an outside organization). The exploratory exercises successfully measured best practice energy efficiencies, determined that energy prices depend on the scale of plant-level purchases, and assessed the degree to which selected energy-intensive products are produced in plants that are outside the primary industry.

1 OVERVIEW OF THE MECS/LRD

1.1 INTRODUCTION

Three primary sources were used for the plant-level¹ data considered in this project: the Census of Manufactures (CM), Annual Survey of Manufactures (ASM), and Manufacturing Energy Consumption Survey (MECS). Data for all three sources are collected by the U.S. Bureau of the Census (hereinafter referred to simply as the Bureau) to compile industrywide and economywide aggregate figures on activity in the manufacturing sector. The data are released to the public in a variety of reports.

Access to the data is provided by the Bureau's Center for Economic Studies (CES) through its program of Census Fellows and Special Sworn Employees (SSEs). SSEs have complete access to confidential plant-level data collected by the Bureau but are required to execute a statement that they will not disclose these data. Yet, by its very existence, the CES encourages research on economic and modeling issues that may involve plant-level data. Thus, the results of any analysis (including those presented in this report) based on CES-compiled plant-level data are reviewed by the CES for confidentiality disclosure. This type of review generally requires that any aggregate statistics compiled by researchers at the CES for publication or outside release cover at least four to six plants. Other types of behavioral (statistical) relationships are reviewed by CES as well to ensure they do not violate confidentiality.

Data from the CM and ASM have been linked together in the Longitudinal Research Database (LRD), which is maintained at the Bureau. The LRD contains data taken from the CM for the years 1963, 1967, 1972, 1977, 1982, and 1987 and data taken from the ASM for all years not covered by the CM from 1973 through 1989. At present, the LRD has substantially more than 2 million manufacturing plant-year records,² including information on more than 800,000 different plants for the 1963-1989 period. The existence of plant identification numbers makes it possible to link the LRD with the MECS; the database that consists of the MECS and LRD linked together is referred to as MECS/LRD. Table 1 shows the number of plants in the LRD in various years; this table is discussed in more detail later in this section.

All three sources — CM, ASM, and MECS — contain establishment-level data for manufacturing industries as defined and classified in the Office of Management and Budget's *Standard Industrial Classification Manual*. The 1987 manual defines an establishment as "an economic unit, generally at a single physical location, where business is conducted or where services or industrial operations are performed" (Office of Management and Budget 1987, p. 12). An establishment is generally a plant, factory, or mill owned by a firm. Thus,

¹ Throughout this report, the terms *establishment* and *plant* are synonymous.

² A plant-year record is one observation of a single year's worth of data for a plant.

TABLE 1 Total Number of Plants in the LRD

Year	Number	Year	Number	Year	Number	Year	Number
1963	305,747	1975	71,145	1980	55,953	1985	55,128
1967	305,611	1976	70,345	1981	55,045	1986	54,858
1972	312,398	1977	350,648	1982	348,384	1987	367,998
1973	73,460	1978	73,853	1983	51,619	1988	53,106
1974	68,262	1979	57,559	1984	56,551	1989	57,276

Source: McGuckin (1990, Table 1).

a firm may own one or more establishments. Identification numbers are assigned to each establishment so that it is possible to identify the firm to which an establishment belongs and construct firm-level data from the establishment-level data.

Manufacturing industries include "establishments engaged in the mechanical or chemical transformation of materials or substances into new products. These establishments characteristically use power-driven machines and materials handling equipment. Establishments engaged in assembling component parts of manufactured products are also considered manufacturing if the new product is neither a structure nor other fixed improvement. Also included is the blending of materials, such as lubricating oils, plastics, resins, or liquors" (Office of Management and Budget 1987, p. 67).

Data sets like the MECS/LRD that contain multiple observations of economic units (in this case, plants) over time are referred to as panel data sets. If a panel data set contains data on the same economic units over time, it is referred to as a balanced panel. If some economic entities are eliminated from the panel over time and if new entities are added, the panel data set is referred to as an unbalanced panel.

Sets of plant-level data provide a vehicle for calculating measures of gross changes that are often missed in aggregate statistics. For example, aggregate figures constructed from the MECS indicate that consumption of energy per dollar value of shipments declined in the primary aluminum industry between 1985 and 1988. Plant-level data from the MECS allow for an investigation of how that decline came about. (For example, did it occur as the result of the reallocation of production from plants that are less energy efficient to those that are more energy efficient, or did it occur as a result of all plants becoming more energy efficient?) Linking the MECS and LRD makes it possible to ask an even greater array of questions about the factors contributing to aggregate changes. (For example, has the decline in primary consumption of energy per unit of output coincided with similar declines in the consumption of other inputs, or has there been an increase in the consumption of other inputs to offset the decline in consumption of energy inputs?) Understanding the relationships among the CM, ASM, and MECS is essential for understanding the relationship between aggregate data and the plant-level data from which the aggregate data are constructed and for understanding the potential of the MECS/LRD in energy analysis and forecasting.

1.2 CENSUS OF MANUFACTURES

The CM obtains data on various measures of economic activity for every manufacturing plant in the nation with one or more employees. It is conducted at five-year intervals, covering years ending in "2" and "7." For example, Table 1 shows that there were 367,998 manufacturing plants with one or more employees operating in the United States in 1987.

The CM collects data on each plant's labor, material, and capital inputs; its product and service outputs; its location; and the legal form of the organization that owns it. Associated with each plant record is a permanent identification number and location. Both of these items stay with the plant record from the time the plant starts until it shuts down. In addition, each plant is linked to a parent firm, and detailed status codes allow one to trace ownership changes over time as well as to construct firm-level data.

Although the CM is a complete enumeration of all manufacturing plants, not all plants actually report data to the Bureau. Beginning in 1967, some small, single-plant companies were exempted from reporting their data to the Bureau. Nevertheless, even though the excluded companies do not file census forms, data for these single plants are constructed from Internal Revenue Service and Social Security Administration records and imputed from industry average ratios and included in the LRD.

Those establishments exempted from reporting data for the CM generally account for less than 2% of total value added: in 1972, 120,000 establishments accounting for 1.2% of value added were exempted; in 1977, 130,000 establishments accounting for 1.7% of value added were exempted; in 1982, 130,000 accounting for 1.3% of value added were exempted (CES 1990). Although the impact of such exclusions on industry aggregates is slight, the usefulness of CM data in studying the behavior of small, single-plant companies is diminished because of the imputations, since CM data reflect the behavior of industry averages rather than the behavior of an individual plant. Because the energy-intensive industries analyzed here are much less likely to fall in this small, single-plant category, the adverse impacts of excluding this category of industries are diminished in this report.

1.3 ANNUAL SURVEY OF MANUFACTURES

The ASM is a national survey of a sample of manufacturing plants. Since 1949, the ASM has collected data for each year that is not covered by the CM. The plants chosen for the ASM sample are drawn from those listed in the CM. The sample is selected in the year that follows the year for which the CM data are collected. (The year for which the CM data are collected is referred to in this report as the CM or census year.) Data on the ASM sample are collected for the year the sample is selected and for each of the next three years. Then a new CM is conducted, and the process is repeated; that is, a new sample is drawn from the most recent CM. As can be seen in Table 1, the number of plants listed in the LRD (which combines CM and ASM data) is five to six times larger in each CM year than it is in non-CM years because the ASM sample is a subset of the universe of plants in the CM. For example,

the 1988 ASM included about 14% (53,106) of the 367,998 plants that had reported in the previous CM year.

The ASM collects data on the same basic measures of economic activity as the CM, but the information it collects on material consumption and product output for each plant is less detailed. For example, in years for which ASM data are collected (ASM or survey years), the value of products shipped is recorded in terms of only about 1,500 product classes instead of the roughly 11,000 individual products used in CM years. Thus, any data on plants found in the ASM can be found in the CM for years the ASM is not conducted. Because the number of variables for which data are collected for the ASM and CM is large, a variety of selection keys (e.g., geographic location, industry, size, and firm) can be used to generate data extracts from the LRD.

Although there is substantial overlap with respect to the plants that are included in each ASM sample from year to year, the correspondence is not perfect because the sampling design changed several times over the years. Despite this fact, all ASM panels have included all plants with 250 or more employees. Smaller plants have been sampled at probabilities proportional to their size.

As is the case for data from the CM, data from the ASM may fail to reflect the behavior of very small plants. Thus, microeconomic studies based on LRD data can lead to inferences that are misleading to the extent that small plants behave differently and are systematically excluded from the CM and ASM. Davis et al. (1991, p. 1) notes that "new economic aggregates constructed from the LRD can yield incorrect estimates of levels and growth rates" because of such sample selection problems.

1.4 MANUFACTURING ENERGY CONSUMPTION SURVEY

The supplemental ASM Fuels and Electric Energy Survey, which began in 1972, included details on energy consumption not collected by the ASM or CM. It was discontinued in 1982. The MECS was established to resume and expand this data collection effort. The MECS is a national survey of the energy consumed by a sample of manufacturing plants. It is conducted by the Bureau for the U.S. Department of Energy (DOE) every three years. The first MECS was conducted in 1986 to collect data for 1985, and the second MECS (DOE 1991) was conducted in 1989 to collect data for 1988. All large establishments included in any ASM sample after 1984 in the 10 four-digit industries that consumed the most energy (SIC 2621, 2631, 2819, 2821 2869, 2873, 2911, 3241, 3312, and 3334) were also included in the 1985 and 1988 MECS samples. Establishments that were not in these 10 industries were sampled at probabilities proportional to an energy measure of their size (see DOE [1991, Appendix A]).

For every plant surveyed, the MECS collects detailed data on the consumption of more than 20 types of energy. In addition to common sources of energy such as electricity, natural gas, fuel oil, and coal, less common sources like biomass, wood chips, black liquor, and others are covered. Detailed data about the on-site generation of electricity, steam, and

industrial hot water are collected, as is information about the purchases and transfers of these energy sources to plants.

The major component of the MECS sample is a subset of the ASM sample. Some plants respond to both surveys, but the surveys are administered separately. The 1988 MECS sample consisted of two components. The first component included data on plants in the 1985 MECS sample (which was based on the 1985 ASM mail sample) that were still in operation in 1989. The second component included data on that part of the 1987 CM mail file not covered by the first component, and it also included data on all plants that reported generating more than 10,000 kWh of electricity on site and that were not otherwise selected for the 1988 MECS. This second component was included to "produce the best possible estimates of onsite electricity generation and associated measures" (DOE 1991, p. 138).

1.5 ENERGY DATA COLLECTED BY THE CM, ASM, AND MECS

The CM, ASM, and MECS all collect data on the consumption of energy by manufacturing plants, but the MECS provides much more detail. For example, the ASM collects data on the quantity and cost of electricity purchased, generated, and consumed as well as the quantity and cost of five types of purchased fuels. These data were part of the supplemental ASM Fuels and Electric Energy Survey data that were collected from 1972 through 1981. The MECS, in addition to collecting data on electricity that is purchased, generated, and consumed, collects data on the quantity and cost of more than 20 types of combustible energy sources that are purchased, produced, and consumed. Because there is some overlap between the MECS and LRD energy data, the responses of plants to similar questions can be compared, and the quality of the data can, to some degree, be ascertained.

1.6 LONGITUDINAL ASPECTS OF THE MECS/LRD: DATA EDITING

Table 2 presents a count of plants for the six selected industries (by standard industrial classification [SIC] code) for which data for 1972 through 1988 are available in the LRD. This table gives the number of raw, unedited data points available in the LRD. As expected, the number of plants for which data are available is higher in CM years than ASM years. For example, in 1982 there were 200 cement plants in the United States; however, in 1981 and 1983, only 81 of them were sampled. To check how complete the Table 2 data are for ASM years, compare the number of plants listed for an ASM year with the number of plants listed for the nearest CM year (i.e., 1972, 1977, 1982, or 1987). For most ASM years, the number of plants is at least half of the total number that responded to the CM. The ratio is often closer to two-thirds or three-fourths (e.g., see the most recent years for paper mills and cement).

The MECS/LRD is an unbalanced panel because it excludes and includes plants for a variety of reasons discussed below. Despite the structure of the MECS/LRD, both balanced and unbalanced panels can be constructed for various time periods. This flexibility allows

TABLE 2 Number of Plants in the Six Industries in the LRD^a

Number of Plants by SIC Code and Industry						
Year	2621 Paper Mills	2812 Alkalies and Chlorine	2873 Nitrogenous Fertilizers	3241 Cement, Hydraulic	3312 Blast Furnaces and Steel Mills	3334 Primary Aluminum
1972	372	46	77	187	334	31
1973	282	39	64	175	256	34
1974	275	38	76	179	263	33
1975	279	40	83	179	284	35
1976	281	42	85	178	284	35
1977	387	38	120	189	281	35
1978	276	35	96	165	254	33
1979	221	27	49	82	218	30
1980	219	28	49	80	215	32
1981	223	28	48	81	215	32
1982	332	44	110	200	286	34
1983	218	25	45	81	185	28
1984	239	27	77	132	177	29
1985	245	27	79	125	179	30
1986	242	29	74	124	173	27
1987	282	31	122	175	342	49
1988	243	30	73	123	167	24

^a Initially tabulated from unedited data. No administrative records for any of the industries for any year were used.

for the analysis of all plants sampled within a specific industry over time (unbalanced panel) and the analysis of the same plants in an industry from year to year (balanced panel).

Table 3 shows the number of plants that constitute a balanced panel for the period 1972-1988. These plants were in continuous operation and were included in the LRD over that time period. Because of the structure of the MECS sample, most of these plants were also included in the MECS. If one wished to examine a balanced panel from the LRD and examine information from the MECS as well, very few observations (i.e., responses) would be lost.

As noted, the Bureau gathers data for the LRD to compile industrywide and economywide aggregate data. In processing ASM and CM responses, the Bureau's primary objective is to identify significant errors that would affect the quality of aggregate statistics. The usefulness of the LRD for the study of plant behavior depends not only on the quality of the aggregate data but also on the quality of the plant-level data. As a result of this fact, the Bureau must further examine the data in the LRD to identify factors or alleged errors that could affect the quality of plant-level data.

TABLE 3 Number of Plants in the Six Industries with Continuous Records for 1972 through 1988 in the LRD

SIC Code and Industry	Number in LRD	Number in LRD and 1988 MECS	Share of 1988 ASM Sample (%)
2621 Paper mills	139	134	55
2812 Alkalies and chlorine	17	13	57
2873 Nitrogenous fertilizers	3	22	30
3241 Cement, hydraulic	47	45	37
3312 Blast furnaces and steel mills	97	93	57
3334 Primary aluminum	19	18	75

Routine editing of LRD data consists of simple checks; for example, one check is to make sure that one period's ending quantities equal the following period's beginning quantities. Ratios of inputs to outputs are also commonly examined to identify seemingly unreliable data. Editing of LRD data is complicated by the fact that when aggregate data are constructed for the LRD, missing or unreasonable plant data are reestimated on the basis of their relationships to reliable data. These reestimated data are referred to as "imputed" data. While such imputed information may be useful in the construction of aggregate statistics, it may not accurately reflect the behavior of an individual plant. Unfortunately, imputed data are not identified in the LRD. Thus, to uncover imputed data, researchers must examine the ratios of key variables. Once the imputed or seemingly unreliable data are uncovered, they must decide what to do with these data — a problem for researchers who want to use only the data that represent an original respondent's plant-level information.

2 PRELIMINARY ANALYSIS OF THE MECS/LRD

To assess the usefulness of the MECS/LRD for detailed energy analysis and forecasting, four types of data were analyzed: (1) simple summary statistics on energy inputs and outputs, (2) energy inputs per unit of output (i.e., energy intensities), (3) measures of technical efficiency, and (4) plant-level energy prices. Six energy-intensive, four-digit industries were chosen for preliminary analysis, in part because of the degree of homogeneity in their output and production processes. They are:

- 2621 — Paper mills, except building paper;
- 2812 — Alkalies and chlorine;
- 2873 — Nitrogenous fertilizers;
- 3241 — Cement, hydraulic;
- 3312 — Blast furnaces and steel mills; and
- 3334 — Primary aluminum.

These six industries together accounted for nearly 25% of total energy inputs in the manufacturing sector in 1988. Data for 1988, the second year for the MECS, were examined. As the project continues, data from 1972 through 1987 will also be considered.

2.1 SUMMARY STATISTICS

2.1.1 All Energy Inputs

To calculate aggregate figures for release to the public, the DOE Energy Information Administration has done a significant amount of editing to eliminate unreliable responses from the MECS/LRD (DOE 1991; CES 1990). Editing activities for the MECS included checking the consistency of responses on individual questionnaires, comparing responses to the MECS with responses to the ASM, and checking for outliers in the distributions of individual variables (MECS methodological report in DOE 1992). Editing activities for the ASM included checking for outliers in the distribution of key operating ratios such as cost of materials per dollar of shipments, payroll per employee, and production-worker wages per hour (CES 1990). For both the ASM and the MECS, weights were applied to each sampled plant. The weights inflated the data reported by the plant so that the data reflected that portion of the population they represented. Unreliable responses were assigned weights of zero and dropped from consideration to calculate aggregate figures. (An example of an unreliable response is one that indicates a plant for which a high level of production but no person-hours are indicated.)

Table 4 shows how many plants use various types of energy for heat, power, and electricity generation, by industry.³ The final column in Table 4 indicates the total number of plants for which MECS energy data are available. Although all plants almost certainly used electricity in some form for manufacturing, not all firms were net users of electricity according to the definition of net electricity input in Table 4. In other words, if the quantity of electricity sold and transferred off the site exceeds the sum of purchases, transfers in, and generation from nonrenewable resources, net electricity input is negative, and the plant is not categorized as a net user.⁴

As can be seen in Table 5, the patterns of energy use in the six industries vary a great deal. In terms of heat content, primary aluminum uses electricity almost exclusively. Blast furnaces and steel mills use relatively large amounts of coke and breeze, natural gas, and other fuels (primarily blast furnace gas and coke oven gas) but little electricity. The alkalies and chlorine and nitrogenous fertilizer industries use relatively large amounts of natural gas. The cement industry uses large amounts of coal. Paper mills use large amounts of other fuels (primarily pulping or black liquor).

A measure of the degree of diversity in energy usage is the Hirfendahl index. The index is calculated for each plant by squaring and summing the shares of each type of energy in the plant's total energy usage. Thus, the Hirfendahl index for a plant using only one type of energy would be one; the index for a plant using more than one type of energy would be less than one. The smaller the Hirfendahl index, the greater the diversity in the type of energy used. The mean value of the Hirfendahl index for each industry is reported in Table 6. The index confirms the patterns of energy usage evident in Table 5: Primary aluminum and nitrogenous fertilizers, which rely primarily on one type of energy in terms of Btu, have the highest Hirfendahl indexes; paper mills and blast furnaces and steel mills, which rely on several types of energy, have the lowest Hirfendahl indexes. Table 7 presents total expenditures for all purposes by energy type for each of the six industries considered. Since these expenditures are for all purposes, including energy used as a raw material input, Table 7 is not directly comparable with previous tables. Nonetheless, a comparison of Table 7 with Tables 5 and 6 is interesting. The pattern of expenditures in Table 7 contrasts sharply with the pattern of Btu input in Table 5. While Table 5 indicates that net electricity is relatively unimportant in terms of Btu input for all industries except primary aluminum, Table 7 indicates that it is the most important energy type in terms of expenditures for four of the six industries.

³ This definition of inputs for heat, power, and electricity generation is consistent with the definition used in Table 3 of the 1988 MECS (DOE 1991). It includes energy produced off the site as well as by-product fuels that are produced on the site and result from the use of other energy sources originally consumed as a raw material input. It excludes all energy sources consumed as a raw material input to the manufacturing process.

⁴ An example of negative net electricity input would be a case in which a plant cogenerates more than enough electricity to meet its own needs and sells the surplus to a utility. Since the cogenerated electricity does not add to net electricity input but rather diminishes it when it is sold, the plant would not be a net user of electricity.

TABLE 4 Number and Percentage of Plants Using Various Types of Energy

SIC Code and Industry	Number and Percent (%) of Plants That Use Various Types of Energy								Total ^b
	Net Elec. ^a	Resid. Fuel Oil	Dist. Fuel Oil	Nat. Gas	LPG	Coal	Coke and Breeze	Other	
2621 Paper mills	204 (94)	128 (59)	121 (56)	153 (71)	134 (62)	71 (33)	0 (0)	161 (74)	216 (100)
2812 Alkalies and chlorine	20 (95)	3 (14)	10 (48)	18 (86)	3 (14)	2 (1)	0 (0)	20 (95)	21 (100)
2873 Nitrogenous fertilizers	45 (96)	0 (0)	28 (60)	39 (83)	12 (26)	0 (0)	0 (0)	31 (66)	47 (100)
3241 Cement, hydraulic	104 (100)	12 (12)	89 (86)	74 (71)	35 (34)	91 (88)	9 (9)	86 (83)	104 (100)
3312 Blast furnaces and steel mills	128 (97)	21 (16)	97 (73)	126 (95)	48 (36)	19 (14)	28 (21)	101 (77)	132 (100)
3334 Primary aluminum	22 (100)	NA ^c NA	21 (95)	22 (100)	17 (77)	0 (0)	0 (0)	20 (91)	22 (100)

^a Net electricity input is defined as the sum of purchased electricity, transfers in, and generation from noncombustible renewable resources, minus quantities sold and transferred off site. It does not include quantities of electricity generated or cogenerated on site from combustible fuels.

^b Total = total number of plants for which there are data in the 1988 MECS database.

^c NA = not available.

Source: Unpublished 1988 MECS database.

TABLE 5 Amount and Percentage of Various Types of Energy Used

SIC Code and Industry	Amount (10 ⁹ Btu) and Percent (%) of Energy Used by Various Industries								Total
	Net Elec. ^a	Resid. Fuel Oil	Dist. Fuel Oil	Nat. Gas	LPG	Coal	Coke and Breeze	Other	
2621 Paper mills	98.0 (9)	97.9 (9)	5.4 (0.5)	177.7 (16.4)	2.0 (0.1)	191.4 (17.7)	0 (0)	507.7 (47)	1,080.2 (100)
2812 Alkalies and chlorine	37.5 (25)	NA ^b NA	0.1 (0.1)	85.6 (57.9)	NA NA	NA NA	0 (0)	9.8 (6.6)	147.9 (100)
2873 Nitrogenous fertilizers	10.0 (5)	0 (0)	0.1 (0.1)	172.7 (89.5)	0.0 (0.0)	0 (0)	0 (0)	10.1 (5.2)	192.9 (100)
3241 Cement, hydraulic	33.4 (10)	2.8 (1)	5.2 (1.5)	29.8 (8.8)	0.0 (0.0)	221.0 (65.4)	3.2 (0.9)	42.3 (12.5)	337.7 (100)
3312 Blast furnaces and steel mills	138.4 (7)	36.2 (2)	6.1 (0.3)	437.4 23.6	1.0 (0.1)	35.5 (1.9)	743.7 (40)	456.4 (24.6)	1,854.6 (100)
3334 Primary aluminum	224.6 (91)	NA NA	NA NA	19.7 (8.0)	0.1 (0.1)	0 (0)	0 (0)	1.0 (0.4)	245.9 (100)

^a Net electricity input is defined as the sum of purchased electricity, transfers in, and generation from noncombustible renewable resources, minus quantities sold and transferred off site. It does not include quantities of electricity generated or cogenerated on site from combustible fuels.

^b NA = not available.

Source: Unpublished 1988 MECS database.

2.1.2 Electricity

The MECS asks each plant to supply data on how much electricity it purchased from utility and nonutility sources and how much it spent for it. The sum of these two responses indicates the total quantity and cost of purchased electricity for each plant. The ASM asks plants to supply data on the quantity and cost of purchased electricity. Although the instructions for responses to the ASM are much less detailed than those for the MECS, there is a rough correspondence between responses.

Table 8 reports that for five of the six industries, there is a fairly high correlation between responses in the MECS and those in the ASM. (The exception is primary aluminum.) Although highly correlated, the mean absolute percentage by which MECS

responses deviated from ASM responses varied by industry, from 3% to 14% for cost of purchases and from 3% to 20% for quantity of purchases.

A closer examination of responses for primary aluminum suggests that most of the lack of correlation was caused by keypunch or reporting errors, which resulted in extra digits being added to or dropped from responses. An examination of the data on the cost and quantity of purchased electricity per dollar value of shipments suggests that responses from the MECS were more reasonable than those from the ASM.⁵ When plants with apparent errors were dropped from the sample, the correlations for primary aluminum rose to 0.99 for cost of purchased electricity and 0.99 for quantity of purchased electricity.

2.1.3 Output

One of the criteria for inclusion of a specific four-digit industry in the scoping study was the homogeneity of its output. If a sector is relatively homogeneous in its technology choice, product mix, etc., this homogeneity facilitates interplant comparisons. In the LRD, output measures include value of shipments and value added (i.e., value of shipments minus value of materials). Value of shipments can be viewed as the (market) price multiplied by some measure of physical production. Price and physical quantity are not reported in survey years in the LRD. Alternative sources of data on physical quantity are discussed below.

For sectors with a very narrow definition of output (i.e., ones with relatively few products), physical output can be treated as a single (scalar) value, Q , with a single price, P . Therefore, differences between value of shipments and a physical measure of output may be attributable to the inferred output price. If energy use is more directly related to the physical dimension of production, a physical output measure is preferred. If there is little variation in the implicitly reported output prices, value of shipments is a good proxy for physical production and therefore a good point of departure for energy use analysis.

TABLE 6 Hirfendahl Index of Energy Diversity

SIC Code and Industry	Hirfendahl Index
2621 Paper mills	0.55
2812 Alkalies and chlorine	0.66
2873 Nitrogenous fertilizers	0.76
3241 Cement, hydraulic	0.62
3312 Blast furnaces and steel mills	0.54
3334 Primary aluminum	0.84

Source: Unpublished 1988 MECS database.

⁵ The data on the cost and quantity of electricity purchased per dollar value of shipments based on the ASM response for plants with apparent errors differed by a factor of 10 to 100 from the rest of the industry. The data on the cost and quantity of electricity per dollar value of shipments based on the MECS response were within the range for the rest of the industry.

TABLE 7 Amount and Percentage Spent for Various Types of Energy

SIC Code and Industry	Total Expenditures (1988 \$10 ⁶) and Percentage of Total Industry Expenditures (%) for Various Types of Energy								Total
	Elec. ^a	Resid. Fuel Oil	Dist. Fuel Oil	Nat. Gas	LPG	Coal	Coke and Breeze	Other	
2621 Paper mills	1,075 (48.7)	220 (9.9)	20 (0.9)	400 (18.1)	10 (0.4)	311 (14.1)	0 (0)	173 (7.8)	2,209 (100)
2812 Alkalies and chlorine	314 (65.2)	NA ^b NA	0 (0.1)	149 (30.9)	NA NA	NA NA	0 (0)	2 (0.3)	481 (100)
2873 Nitrogenous fertilizers	95 (11.4)	0 (0)	1 (0.1)	716 (85.8)	4 (0.4)	0 (0)	0 (0)	18 (2.2)	834 (100)
3241 Cement, hydraulic	369 (48.6)	3 (0.4)	21 (2.7)	54 (7.1)	0.0 (0.0)	280 (36.8)	3.7 (0.5)	30 (3.9)	761 (100)
3312 Blast furnaces and steel mills	1,420 (28.6)	80 (1.6)	24 (0.5)	1,077 (21.7)	4.0 (0.1)	1,514 (30.5)	727 (14.7)	110 (2.2)	4,957 (100)
3334 Primary aluminum	1,462 (87.7)	NA NA	NA NA	53 (3.2)	1 (0.0)	13 (0.8)	8 (0.5)	128 (7.7)	1,667 (100)

^a Net electricity input is defined as the sum of purchased electricity, transfers in, and generation from noncombustible renewable resources, minus quantities sold and transferred off site. It does not include quantities of electricity generated or cogenerated on site from combustible fuels.

^b NA = not available.

Source: Unpublished 1988 MECS database.

The primary aluminum industry produces a very narrowly defined product. For this industry, plant-level output, measured in tons (2,000 lb), was obtained from an industry publication, *Metal Statistics 1988* (American Metal Market 1988), and compared with the value-of-shipments data in the MECS/LRD. However, American Metal Market's listing of aluminum ingot producers did not contain information on one of the plants in the MECS/LRD classified as a member of SIC 3334. Furthermore, *Primary Aluminum Plants Worldwide*, published by the U.S. Bureau of Mines, indicated that the smelter at this plant had been permanently shut down. This plant was dropped from consideration in the comparison.

The simple correlation between the tonnage data from American Metal Market and data on adjusted value of shipments from the ASM was only 0.41. If all producers reported

the value of their shipments on the basis of a common market price, there should have been a high degree of correlation between value of shipments and tonnage. There are several possible reasons for the lack of close correlation between the measures of output.

One possible explanation is that plants used different prices in valuing shipments. These different prices may have resulted from the fact that the spot price of aluminum varies considerably during a year. For example, in 1988, the spot price for a pound of aluminum varied from a low of \$0.87 to a high of \$1.42. For purposes of the ASM, a plant might arbitrarily select a price somewhere in this range and apply it to its output.

Another possible cause of the lack of correlation is that interplant transfers within the same firm, which are considered shipments, may have no explicit market price attached to them. In completing the ASM, a plant may or may not assign a value that approximates market value. According to the CM, of the 3,198,600 tons of aluminum ingot produced in 1987, 1,527,900 tons were transferred to plants owned by the same company (i.e., 48% of total ingot production was an interplant transfer for which there was not an explicit market price).

Finally, it is possible that at the plant level (the level at which data on value of shipments are collected for the ASM), more attention is paid to physical output data than value-of-shipments data. (A specialist at the U.S. Bureau of the Census, Industry Division, who is familiar with the primary metals industries suggested this as a possible cause.) In a multiplant firm, the actual price may be applied to shipments at the firm's headquarters or financial office; the valuation made at the plant level for purposes of the ASM may be somewhat arbitrary.

Is tonnage as reported by the American Metal Market or value of shipments as reported for the ASM a more accurate measure of output? Insight into this question was gained upon examination of energy intensities.

TABLE 8 Simple Correlation between Amounts Reported in the MECS and ASM

SIC Code and Industry	Correlation for Purchased Electricity	
	Quantity	Cost
2621 Paper mills	0.95	0.98
2812 Alkalies and chlorine	0.97	0.99
2873 Nitrogenous fertilizers	0.98	0.94
3241 Cement, hydraulic	0.84	0.91
3312 Blast furnaces and steel mills	0.95	0.91
3334 Primary aluminum	0.80 (0.99) ^a	0.08 (0.99) ^a

^a Correlations were made after additional editing of ASM data was done.

Sources: Unpublished 1988 MECS and 1988 ASM databases.

2.2 ENERGY INTENSITIES

The second means of assessing the usefulness of the MECS/LRD was an examination of energy input per unit of output (i.e., energy intensity) at the plant level. The measure of output was taken from the ASM. Value of shipments, adjusted for changes in inventories, and value added were used as denominators in the calculation of intensities. Since variations in measures of both energy input and output contribute to variations in intensities, sorting out the causes of variation in intensities can be complicated. Summary statistics for total energy intensities are presented in Table 9. Cement, the most energy intensive industry, is followed by a second group that includes both fertilizers and alkalies and chlorine. This

TABLE 9 Energy Intensities^a

SIC Code and Industry	Number of Plants Observed	Mean	Std. Dev.	Coeff. of Var. ^b	Percentage that Is More than 2 Std. Dev. from Mean	Skewness	Kurtosis
2621 Paper mills	212	26.1	17.0	65.1	4.7	1.11	0.91
2812 Alkalies and chlorine	21	48.9	22.3	45.6	4.7	0.34	1.02
2873 Nitrogenous fertilizers	45	62.2	58.4	93.8	2.2	0.58	-0.98
3241 Cement, hydraulic	104	83.6	45.1	54.0	3.8	0.63	2.29
3312 Blast furnaces and steel mills	126	25.3	26.2	103.3	2.4	2.43	9.29
3334 Primary aluminum	22	32.2	11.8	36.8	9.1	0.62	0.41

^a Intensity = energy input (10^3 Btu) divided by the dollar value of shipments reported to the U.S. Bureau of the Census.

^b The coefficient of variation is a unitless measure of relative variability. It is defined as the ratio of the standard deviation to the mean, expressed as a percentage.

Sources: Unpublished 1988 ASM database for adjusted value of shipments; unpublished 1988 MECS database for total energy input.

second group's high intensity is caused primarily by feedstocks, since the MECS total energy is used for the computation. The last group is composed of paper mills, steel plants, and aluminum smelters, all of which have similar energy intensities.

One of the more interesting statistics from Table 9 is the coefficient of variation. The higher the value of the coefficient of variation, the greater the relative variability. Thus, Table 9 indicates that the variability in energy intensities is greater for steel mills than for other industries. This fact is not surprising, given the range of mill types in the sample. Additional data are required to identify and isolate the plants with different technologies. The nitrogenous fertilizer industry also has a high coefficient of variation. The reasons for this have not been investigated. With the exception of aluminum, which has a relatively small coefficient of variation (as expected), the other industries in the sample exhibit a similar range of variation. One could reduce this variability by disaggregating the plant sample to the process level.

Cumulative distributions of the intensities are presented in Figures 1-4. These cumulative distributions, sometimes referred to as Salter diagrams, compactly represent a wealth of information regarding industry use of energy. The distributions for each industry were created by (1) calculating the energy intensity for each establishment, (2) calculating the share of industry output produced by each establishment, (3) ranking establishments from least to most energy intensive, (4) totalling shares of output for each establishment, and (5) plotting energy intensities from lowest to highest against total shares of output. Intensities were averaged over groups of establishments to avoid violating U.S. Bureau of the Census disclosure rules.

As an example of how the cumulative distributions were constructed, consider Figure 1, the cumulative distribution of total energy intensities, for paper mills. Step 1 in the construction of this figure was to calculate energy intensity for each of the 212 plants in the paper mills category by dividing each establishment's energy input by its shipments. Step 2 was to calculate each establishment's share of total industry output. Each establishment's reported output was scaled by its MECS sample weight and totalled over all establishments to obtain total industry output. Each establishment's share of industry output was calculated as its scaled output divided by total industry output. Step 3 was to rank all 212 plants by their total energy intensity from lowest (which was somewhere below 10,000 Btu per dollar value of shipments, adjusted for changes in inventories [hereinafter referred to as dollar value of adjusted shipments]) to highest (which was somewhere above 60,000 Btu per dollar value of adjusted shipments). Finally, the shares of output for establishments as ranked by energy intensity were "accumulated"; that is, the sum of each establishment's share of shipments was added to the shares of those firms with lower energy intensities, and the total was plotted against energy intensity.

It would be helpful to show energy intensity for each establishment in the paper mills category, but U.S. Bureau of the Census disclosure rules will not allow these individual figures to be reported. Instead, average intensities for each 10% of industry shipments are

Notes: Value of shipments is adjusted for changes in inventories and weighted by MECS weight.

Total energy input is total inputs of energy for heat, power, and electricity generation as defined in 1988 MECS, Consumption of Energy, Table 3.

Integer value at each step is the number of establishments that make up the step.

FIGURE 1 Distribution of Total Energy Intensities by Industry (Sources: Unpublished 1988 ASM database for adjusted value of shipments; unpublished 1988 MECS database for total energy input)

Notes: Value of shipments is adjusted for changes in inventories and weighted by MECS weight.

Net electricity input is the sum of purchased electricity, transfers in, and generation from noncombustible renewable resources, minus quantities sold and transferred off site. Thus, it excludes electricity generated or cogenerated from combustible fuels.

Integer value at each step is the number of establishments that make up the step.

FIGURE 2 Distribution of Net Electricity Intensities by Industry (Sources: Unpublished 1988 ASM database for adjusted value of shipments; unpublished 1988 MECS database for net electricity input)

Notes: Value of shipments is adjusted for changes in inventories and weighted by MECS weight.

Fossil fuel input is defined as the sum of residual fuel oil, distillate fuel oil, natural gas, LPG, coal, and coke and breeze.

Integer value at each step is the number of establishments that make up the step.

FIGURE 3 Distribution of Fossil Fuel Intensities by Industry (Sources: Unpublished 1988 ASM database for adjusted value of shipments; unpublished 1988 MECS database for fossil fuel input)

Notes: Value of shipments is adjusted for changes in inventories and weighted by MECS weight.

Other energy input is total energy input minus fossil fuel input and net electricity input.

Integer value at each step is the number of establishments that make up the step.

FIGURE 4 Distribution of Other Energy Intensities by Industry (Sources: Unpublished 1988 ASM database for adjusted value of shipments; unpublished 1988 MECS database for other energy input)

reported when possible. When the number of establishments is not sufficient to meet disclosure rules, intensities are averaged over the establishments that represent 20% of industry shipments or more, as required to meet disclosure rules. Thus, for the alkalies and chlorine industry, Figure 1 shows that the average energy intensity for the six least energy intensive establishments is 20% rather than 10% (as for paper mills) of total industry shipments.

Figure 1 shows that the 39 least energy intensive paper mills had shipments that added up to 10% of the industry total. The next 42 least energy intensive mills had shipments that also added up to 10% of the industry total. Thus the 81 (39 + 42) least energy intensive establishments had shipments making up 20% (10% + 10%) of total industry shipments.

This way of organizing the data provides the following information:

- A range of variation in the energy intensities.
- The form of the distribution of energy intensities. An average measure of the distribution is obtained by looking at the median output value along the horizontal axis.
- A relationship between the energy intensity distribution and the size distribution of the units.

Consider Figure 1 for paper mills again. The range in variation of intensities is roughly from 9,000 to 65,000 Btu per dollar value of adjusted shipments (i.e., the lowest and highest intensities differ by a factor of about seven). The median total intensity is about 32,000 Btu per dollar value of adjusted shipments, so half of the total output is produced at an intensity of less than 32,000 Btu, and half is produced at an intensity of more than 32,000 Btu.

At each step of the cumulative distribution, the number of establishments composing that step's share of value of shipments can be seen. Thus, in the first step for paper mills (Figure 1), a total of 42 establishments with relatively low energy intensities make up 10% of the total value of shipments. In the last step, only 16 very energy intensive establishments make up 10% of the total value of shipments. It appears that large, very energy intensive establishments produce a large proportion of total output in the paper mills category. Indeed, 172 establishments with relatively low energy intensities produce the first 50% of output, whereas only 60 establishments with relatively high intensities produce the second 50% of output.

The distributions exhibited in the Salter diagrams provide information about how the industry could be segmented for further analysis. In the example above, one might conclude that many of the 172 low-intensity plants are probably converting mills. This conclusion could be confirmed by trade associations and other sources. One might also expect

the plants on the low-intensity "steps" in the steel industry to be some type of nonintegrated producers, like rolling mills.

The electricity intensity diagrams help identify plants that use (or do not use) electricity-intensive processes. For example, there is a large gap in the alkalies and chlorine plot, suggesting that a large segment of the industry (60% of shipments in seven plants) does not use electricity-intensive processes. In the paper industry, large cogenerators can be identified by the negative net electricity values.

For the most part, fossil fuel intensity for these six industries tracks the total intensity. The major exception is found in the primary aluminum category. Significant outliers, represented by very high energy intensities and a large number of plants, are exhibited in the paper, steel, and cement industries. It is important to identify these plants and the factors that have given rise to their energy use patterns.

For other energy intensities, the Salter diagrams identify the plants that primarily make use of waste fuels. This practice is reflected as efficiency in other energy types. The linkage between by-product fuel use and energy efficiency is an important one to model explicitly rather than implicitly (as was often the case in the past).

Overall, the degree of variability in intensities revealed in Figures 1-4 seems excessive.⁶ Excessive variability in either the energy input measures or value of shipments may contribute to excessive variability in intensities. The technology issues discussed above would lead to differences in actual input (i.e., energy) differences. One way of assessing the degree of variability due to value of shipments is to calculate intensity on the basis of some other measure of output or a proxy for output.

As noted above, data on the output tonnage for primary aluminum were obtained from the American Metal Market to be compared with data on output from the LRD. The simple correlation between total energy intensity based on output tonnage and total energy intensity based on dollar value of adjusted shipments from the ASM was -0.32 and insignificant.

Table 10 reports descriptive statistics for energy intensities on the basis of three measures of output: dollar value of adjusted shipments, dollars of value added as defined in the LRD, and tonnage. The coefficient of variation, a measure of relative variability, indicates that intensity based on output tonnage is much less variable than intensity based on the other measures. This hypothesis is confirmed in Figure 5, which shows that average intensity based on tonnage data for the most energy intensive third of plants is about 1.3 times that of the least energy intensive third. On the other hand, average intensity based on dollar value of adjusted shipments for the most energy intensive third is about 1.9 times

⁶ A simple linear regression of intensities on the share of total output of each product suggested that none of the variation in intensities could be explained by product mix. Intensities based on value added (not reported here) in general revealed an even greater degree of variability.

TABLE 10 Descriptive Statistics for Three Measures of Total Energy Intensity in the Primary Aluminum Industry^a

Measure of Intensity	Mean	Std. Dev.	Coeff. of Var.
Total energy input per dollars of value added (10 ⁶ Btu)	96.7	90.7	93.8
Total energy input per dollar value of adjusted shipments (10 ³ Btu)	33.3	10.8	32.6
Total energy input per ton of output (10 ³ Btu)	59.5	9.1	15.3

^a Based on 21 observations.

that of the energy intensive third. Plants in primary aluminum employ similar production processes to produce similar output, so energy intensities across plants should be similar. Because intensities based on tonnage are more similar than those based on dollar value of adjusted shipments, more confidence can be placed in tonnage data as a measure of input.

2.3 TECHNICAL EFFICIENCY

Since energy intensities vary so widely, a screening process was applied to identify those plants in the MEC/LRD that exhibit "best practice" production. Best practice production occurs when the most output is produced with the least amount of all inputs, including energy. The energy data come from the MECS; data on other production inputs, labor, materials, plant, equipment, and output come from the LRD. This approach departs somewhat from an examination of the plants with the lowest energy output ratios. The individual plants in the selected industries are screened through the use of frontier efficiency linear programming (FELP). This linear programming technique, based on Farrell's measure (Farrell 1957), has also been referred to in the operations research literature as data envelopment analysis. The method uses a linear program to determine the establishments that make up the "best practice production frontier" (i.e., the program finds the most efficient establishments).

Technical efficiency measures were calculated for all plants in the six industries. The linear programming technique used to calculate these measures identifies those plants that are most efficient in their use of inputs and then measures the degree of relative inefficiency of all other plants.⁷ The measures indicate the extent to which plants could reduce their use of all inputs by the same proportion and still achieve the same level of

⁷ The linear programming problem used to calculate the Farrell measure of technical efficiency is formulated in Färe et al. (1985).

FIGURE 5 Comparison of Total Energy Intensities Computed on the Basis of Three Measures of Output (Sources: Unpublished 1988 ASM database for adjusted value of shipments and value added; unpublished 1988 MECS database for total energy input; American Metal Market [1988, June 8] for tonnage output)

output. The most efficient plants in a sample receive a technical efficiency measure of one; less efficient plants receive a measure of less than one.

The technique used to measure technical efficiency is illustrated for a simple case of two inputs and one output in Figure 6, a diagram similar to the standard isoquant diagram. Here observations on inputs of capital and labor per unit of output for nine plants are plotted. The solution of a series of linear programming problems allows one to construct the lower envelope of the set of observations. This envelope, QQ in Figure 6, is called the best practice unit isoquant. In the diagram, the technical efficiency measure for any plant is inversely related to the distance of that plant from the best practice isoquant. In Figure 6, technical efficiency for the plant using the input combination X is measured by the ratio Oa/OX .

FIGURE 6 Example of a Technical Efficiency Calculation

In the calculation of efficiency measures, it was assumed that plants in each industry use capital, labor, energy, and material inputs to produce a single output. Two assumptions about the nature of technology are implicit in the measure calculated: (1) strong disposability of inputs and (2) constant returns to scale in production. These assumptions are commonly made in the calculation of technical efficiency, although it is possible to calculate efficiency under alternative assumptions (Färe et al. 1985).

Measures of inputs and outputs used in the actual calculations are detailed in Appendix A. Energy inputs were constructed from the MECS; measures of material inputs, capital, labor, and outputs were taken from the ASM. Several plants in addition to those with no MECS and ASM weights were eliminated from consideration because of the absence of capital stock data in the LRD.

Table 11 reports summary statistics for the efficiency measures for each industry. These efficiency distributions divide the six industry categories into two groups. The group that has a small sample of plants and a narrow range of efficiency scores consists of alkalies and chlorine, nitrogenous fertilizers, and primary aluminum. The group that has a large sample of plants and a wider range of variation in efficiency is composed of paper mills, cement, and steel mills. The mean efficiency for the first group is between 90% and 100%. The high level of efficiency for all plants in the first group suggests that these three industries are relatively homogeneous and have few, if any, data outliers. For the latter group of industries, the mean efficiency is in the range of 80%. The coefficient of variation is also much larger for the latter group. These industries have a more diverse technology

TABLE 11 Summary Statistics on Technical Efficiency Measures

SIC Code and Industry	Mean	Std. Dev.	Coeff. of Var.	Skewness	Kurtosis
2621 Paper mills	0.84	0.15	17.5	-0.58	-0.58
2812 Alkalies and chlorine	0.99	0.03	2.8	-4.4	19.9
2873 Nitrogenous fertilizers	0.93	0.13	14.7	-2.0	0.02
3241 Cement, hydraulic	0.77	0.20	25.9	-0.27	-1.3
3312 Blast furnaces and steel mills	0.83	0.17	20.6	-0.54	-0.87
3334 Primary aluminum	0.94	0.11	11.7	-2.1	3.5

mix. To better analyze this second group of industries, more detailed subsamples arranged according to process, product, vintage, or other explanatory variables would be desirable. If this additional information were obtained but did not fully explain the wide degree of variation in these sectors, one might then question the quality of some of the outlier plant data.

Table 12 gives a different breakdown of the efficiency analysis, showing the number and percentage of plants in the sample that have different levels of technical efficiency. In the paper, cement, and steel industries, there is an almost bimodal distribution of efficiency measures. One possible reason for this result is the two relatively distinct groups of plants or subsectors in the data set. If this hypothesis could be confirmed by external trade association data, these subsectors could be analyzed separately. For example, in the steel industry, minimills could be separated from the other steel plants. In the paper sector, bleached Kraft mills could be separated from unbleached Kraft mills. If different subsamples could be constructed on the basis of the different process technologies employed by the plants and their different products, the resulting distributions of efficiency measures would probably no longer be bimodal. The interplant comparisons in the efficiency analysis would only account for real differences in efficiency; they would no longer account for the differences between disparate technologies.

TABLE 12 Distribution of Technical Efficiency Measures

SIC Code and Industry	Technical Efficiency	No. of Plants	% of Total Plants	% of Total Plants, Weighted by Value of Shipments
2621 Paper mills	<0.5	3	1.4	1
	0.5-0.59	8	3.8	3
	0.6-0.69	29	13.9	16
	0.7-0.79	43	20.7	25
	0.8-0.89	35	16.8	18
	0.9-0.99	27	13.0	16
	1.0	63	30.1	21
	Total	208	100.0	100
2812 Alkalies and chlorine	<1	2	10.0	5
	1	18	90.0	95
	Total	20	100.0	100
2873 Nitrogenous fertilizers	<0.8	6	14.6	10
	0.8-0.89	3	7.3	14
	0.9-0.99	2	4.9	4
	1.0	30	73.2	72
	Total	41	100.0	100
3241 Cement, hydraulic	<0.4	2	2.2	1
	0.4-0.49	7	7.5	6
	0.5-0.59	14	15.1	13
	0.6-0.69	15	16.2	13
	0.7-0.79	10	10.8	9
	0.8-0.89	6	6.5	6
	0.9-0.99	13	14.0	19
	1.0	26	28.0	33
	Total	93	100.0	100
3312 Blast furnaces and steel mills	<0.4	5	4.3	2
	0.5-0.59	7	6.0	14
	0.6-0.69	15	12.8	21
	0.7-0.79	27	23.1	18
	0.8-0.89	10	8.5	9
	0.9-0.99	14	12.0	11
	1.0	39	33.3	25
	Total	117	100.0	100
3334 Primary aluminum	<0.8	2	9.5	9
	0.8-0.89	2	9.5	7
	0.9-0.99	4	19.0	21
	1.0	13	61.9	63
	Total	21	100.0	100

Another observation with respect to these three industries is that several values are extremely low. These outliers may have resulted from erroneous data or other peculiarities in reporting (e.g., the output valuation issues raised in the aluminum industry analysis). These plants should be screened by hand before they are included in any additional analysis.

2.3.1 Estimates of Best Practice Energy Use

Given the caveats regarding a strict interpretation of the efficiency measure as a real technical phenomenon, it is still interesting to explore this measure fully by studying how aggregate energy use would be affected if all plants adopted best practice technology.

Multiplying a plant's actual input usage by its technical efficiency measure gives its level of input usage, had the plant adopted best practice production techniques. For example, a firm that actually used 25 million Btu of energy input and had a technical efficiency measure of 0.95 would have used only 23.75 million Btu (0.95×25 million Btu) had it adopted best practice production techniques. Thus, by (1) calculating the best practice level of energy input for each plant, (2) using the plant's MECS weight to inflate this level of energy usage to the proportion of the population that it represents, and (3) totalling MECS-weighted energy usage over all plants in the sample, one can arrive at the industrywide level of energy usage if all firms were technically efficient.

Data on actual energy input and best practice energy input by industry are reported in Table 13. They indicate that if the paper, cement, and steel industries had adopted best practice production techniques, their use of energy would have been significantly reduced.⁸

Because an examination of intensities indicated that the tonnage output data obtained from *American Metal Market* seemed more reliable, technical efficiency measures were calculated for the primary aluminum category on the basis of these tonnage data. Two plants were eliminated from consideration (one shut down its smelter; there was a lack of capital data for the other). The mean technical efficiency measure based on tonnage output was 0.943 for the aluminum industry. The Pearson correlation between technical efficiency based on tonnage output and technical efficiency based on value of shipments was 0.04 and insignificant. The Spearman rank correlation was 0.29 and insignificant. These lacks of correlation suggest that it is desirable to check the reliability of value-of-shipments data from the ASM against other measures of output. (It also suggests that the reduction in energy input from adoption of best practice production techniques reported in Table 10 may be misleading.)

⁸ As noted, several plants had been dropped because there was a lack of capital data for the calculation of efficiency scores. To calculate aggregate best practice energy input, it was assumed that all the dropped plants were technically efficient; that is, they were actually on the best practice isoquant and so had technical efficiency measures equal to one. If any of these firms were, in fact, technically inefficient, the percentage reduction in energy input would be larger than that reported in Table 10.

TABLE 13 Estimates of Energy Consumption under Best Practice Technology Assumptions

SIC Code and Industry	Actual Total Energy Input (10 ⁹ Btu)	Best Practice Total Energy Input (10 ⁹ Btu)	Percentage Decrease
2621 Paper mills	1,080.2	880.4	18.5
2812 Alkalies and chlorine	147.9	147.6	0.2
2873 Nitrogenous fertilizers	192.9	176.6	8.4
3241 Cement, hydraulic	337.7	269.4	20.2
3312 Blast furnaces and steel mills	1,854.6	1,416.5	23.6
3334 Primary aluminum	245.9	231.1	6.0

The resulting efficiency measures were applied to actual energy inputs used to calculate the amount of energy each firm would use had they adopted best practice production techniques. The 20 sample plants actually used 234 trillion Btu of energy (as defined in Table 3 of MECS). When computed efficiency measures were applied to actual energy usage by plant and totalled over all plants, energy usage fell to 202 trillion Btu, a 14% reduction. Future work might focus on explaining such inefficiency.

2.2.2 Relationship between Plant-Level Energy Intensity and Input Efficiency

It is natural to ask whether the plants with the highest energy intensity are also those that are ranked as inefficient. Table 14 computes a simple correlation between energy intensity and input efficiency. Although the results are statistically significant and have the correct sign (i.e., the intensity/efficiency relationship is an inverse relationship), the magnitude of the correlation is small. In particular, for the paper and steel industries, the sample should probably be divided into subsamples on the basis of process technology or

product line. The comparison of technical efficiency and energy intensity would therefore be improved, since the validity of the technical efficiency measure depends on the homogeneity of the sample.

Another way to represent the relationship between efficiency measures and energy intensities is shown in Figure 7. The number of plants that are efficient (i.e., those that have efficiency levels of 95% or more) are shown in parentheses on each step of the total energy Salter diagram. Although the first (lowest) step of the intensity plot has the largest number of efficient plants, there are a few efficient plants at almost all levels of energy intensity. One reason for this lack is that technical efficiency is different from allocative efficiency. Allocative efficiency is determined on the basis of cost- or price-based allocation of resources. Thus, plants that are technically efficient but have higher-than-average energy intensities may simply be otherwise well-run plants that have poor energy management practices. The other possibility is that the energy data are in error. These plants will have to be screened carefully before they are included in any further analysis.

TABLE 14 Correlation between Technical Efficiency and Total Energy Intensity

SIC Code and Industry	Correlation Coefficient
2621 Paper mills	-0.25 ^a
2812 Alkalies and chlorine	-0.10
2873 Nitrogenous fertilizers	-0.46 ^a
3241 Cement, hydraulic	-0.47 ^a
3312 Blast furnaces and steel mills	-0.27 ^a
3334 Primary aluminum	-0.50 ^a

^a Significant at 5% level.

2.4 PLANT-LEVEL PRICES

Because the MECS collects data on quantity and cost of purchases by fuel type, an implied price can be calculated for each fuel type. These plant-level prices may be important when the variation in energy intensity is modeled. (Only a small number of plants used residual fuel oil, so no statistics on this fuel are discussed here because of confidentiality restrictions.)

Data on the unweighted (arithmetic sample) mean and quantity-weighted mean of implied prices for selected fuel types by industry are shown in Table 15. Because these data represent purchases for all purposes, they do not represent purchases solely for generating heat, power, and electricity and so are not directly comparable with data in Tables 4, 5, and 6. Nonetheless, a comparison of Table 15 with Table 5 reveals that those industries that use a particular fuel most intensively to generate heat, power, and electricity generally pay the lowest price for that fuel. Anecdotal evidence of this type of monopsony power (the ability

Notes: Value of shipments is adjusted for changes in inventories and weighted by MECS weight.

Total energy input is total inputs of energy for heat, power, and electricity generation as defined in 1988 MECS, Consumption of Energy, Table 3.

Integer value at each step is the number of establishments that make up the step.

Number in parentheses is the number of establishments at that step with a technical efficiency measure that is greater than 0.95.

FIGURE 7 Distribution of Total Energy Intensities by Industry (Sources: Unpublished 1988 ASM database for adjusted value of shipments; 1988 MECS database for total energy input)

TABLE 15 Descriptive Statistics for Plant-Level Prices for Different Fuels

Fuel ^a	SIC Code and Industry ^b	Unweighted Mean	Mean, Weighted by Quantity Purchased
Electricity (\$/kWh)	2621 Paper	0.048	0.037
	2812 Alkalies	0.030	0.027
	2873 Fertilizers	0.053	0.035
	3241 Cement	0.044	0.042
	3312 Steel	0.045	0.039
	3334 Aluminum	NA ^c	NA
Distillate fuel oil (\$/gal)	2621 Paper	0.70	0.53
	2812 Alkalies	0.77	0.61
	2873 Fertilizers	0.76	0.63
	3241 Cement	0.65	0.61
	3312 Steel	0.66	0.57
	3334 Aluminum	0.67	0.65
Natural gas (\$/10 ³ ft ³)	2621 Paper	3.10	2.40
	2812 Alkalies	3.30	2.00
	2873 Fertilizers	2.60	1.90
	3241 Cement	3.10	2.10
	3312 Steel	3.20	2.70
	3334 Aluminum	NA	NA
LPG (\$/gal)	2621 Paper	0.66	0.44
	2812 Alkalies	NA	NA
	2873 Fertilizers	NA	NA
	3241 Cement	0.93	0.68
	3312 Steel	9.40	0.36
	3334 Aluminum	0.50	0.38
Coal (\$/ton)	2621 Paper	42.90	39.5
	2812 Alkalies	NA	NA
	2873 Fertilizers	NA	NA
	3241 Cement	34.40	34.1
	3312 Steel	53.20	46.4
	3334 Aluminum	276.80	176.6

^a Only a small number of plants used residual fuel oil, so no statistics are presented here because of disclosure rules.

^b Paper = paper mills; alkalies = alkalies and chlorine; fertilizers = nitrogenous fertilizers; cement = cement, hydraulic; steel = blast furnaces and steel mills; aluminum = primary aluminum.

^c NA = not available; withheld because of disclosure rules.

to influence the purchase prices of inputs — particularly energy) has come from other sources in industry. The truth of this anecdotal evidence is borne out by the fact that the weighted mean price is much lower than the unweighted mean price. Industries that use a lot of a particular form of energy pay a lower price for it. Whether they use more because it is cheaper or they pay less because of some market power is not discernable from this simple comparison. One possibility is that industries tend to locate where energy is cheaper. Larger-scale purchases may also result in lower delivery costs. The potential for pure monopsony (market) power is one that needs to be investigated.

The distributions of average energy prices by industry and fuel type are shown in Figures 8-12. An examination of the distribution of electricity prices for nitrogenous fertilizers reveals several apparent outliers. A comparison of electricity prices implied by responses to the 1988 MECS with those implied by responses to the 1988 and 1987 ASM suggests that the MECS-implied price for electricity for the outliers was too high. The ASM-implied prices were within one standard deviation of the industry sample mean.

The mean price of LPG for blast furnaces and steel mills in Table 15 suggests the presence of outliers. A closer examination of the data reveals one extreme outlier. When the outlier is eliminated, the sample mean for the industry falls to a level within the range of the other industries. Although the single outlier exerted a large influence on the sample mean, its influence on the aggregate figures reported in the 1988 MECS (DOE 1991, Table 30) was minimal. Assuming the outlier actually took the value of the sample mean and using this value to calculate the aggregate figure reported in the 1988 MECS resulted in a change of less than 2%. This example suggests that although data from the MECS may be adequate to calculate aggregate figures, the data must be carefully scrutinized when used to explore plant-level behavior.

Data for all six industries were pooled to estimate simple linear regressions for each of six energy types (i.e., electricity, residual fuel oil, distillate fuel oil, LPG, natural gas, and coal). The regressions were designed to explain variations in price that might be caused by regional factors and quantity factors. Least-squares estimates were made for each energy type, over all industries:

$$PRICE = a \times QUANTITY + \sum_{i=1}^9 b_i \times REGION_i + e$$

Here PRICE, QUANTITY, REGION, a , and b_i are of appropriate dimension and a and b_i are parameters whose values are estimated.

Results (Table 16, second column) show that there is an inverse relationship between the quantity purchased and the price paid for all energy types, which suggests that industries using a large amount of energy can exert some monopsony power over suppliers. In addition, Table 16 results generally confirm expectations about regional variations in energy prices.

FIGURE 8 Distribution of Electricity Prices (Source: Unpublished 1988 MECS database)

FIGURE 9 Distribution of Distillate Fuel Oil Prices (Source: Unpublished 1988 MECS database)

SIC 3334 / Primary Aluminum
Withheld

FIGURE 10 Distribution of Natural Gas Prices (Source: Unpublished 1988 MECS database)

FIGURE 11 Distribution of LPG Prices (Source: Unpublished 1988 MECS database)

FIGURE 12 Distribution of Coal Prices (Source: Unpublished 1988 MECS database)

TABLE 16 Estimates of Plant-Level Variations in Energy Prices Caused by the Location of the Plant and Amount of Energy Purchased by the Plant^a

		Estimated Coefficients								
		Regional Dummies ^b								
Dependent Variable, Price of Fuel	Quantity Purchased ^a	New England b_1	Mid Atlantic b_2	East North Central b_3	West North Central b_4	South Atlantic b_5	East South Central b_6	West South Central b_7	Mountain b_8	Pacific b_9
Electricity $R^2 = 0.16$	-8.6×10^{-12}	0.059 (43)	0.051 (92)	0.046 (112)	0.046 (33)	0.044 (62)	0.045 (39)	0.043 (60)	0.049 (23)	0.044 (61)
Residual fuel oil $R^2 = 0.14$	-2.2×10^{-9} (footnote c)	0.40 (40)	0.42 (36)	0.40 (29)	0.55 (3)	0.42 (19)	0.57 (10)	0.35 (3)	0.33 (4)	0.34 (16)
Distillate fuel oil $R^2 = 0.11$	-7.6×10^{-8}	0.79 (21)	0.69 (61)	0.74 (82)	0.72 (26)	0.64 (50)	0.69 (28)	0.72 (41)	0.65 (14)	0.74 (40)
LPG ^d $R^2 = 0.09$	-6.1×10^{-8}	0.79 (23)	0.83 (46)	0.69 (56)	0.54 (15)	0.75 (30)	0.61 (18)	0.71 (15)	0.73 (5)	0.60 (36)
Natural gas $R^2 = 0.13$	-3.9×10^{-8}	3.3 (11)	3.4 (70)	3.5 (99)	3.0 (29)	3.5 (49)	3.1 (37)	2.4 (55)	2.8 (19)	3.1 (54)
Coal ^e $R^2 = 0.11$	-1.2×10^{-6} (footnote c)	60.0 (4)	42.8 (30)	40.6 (58)	33.9 (17)	44.8 (33)	38.1 (19)	38.2 (14)	32.0 (13)	49.4 (12)

See next page for footnotes.

TABLE 16 (Cont.)

-
- ^a Numbers in parentheses below the values for the estimated coefficients of variation are the number of observations in the region.
- ^b New England: Maine, New Hampshire, Vermont, Rhode Island, Connecticut; Middle Atlantic: New York, New Jersey, Pennsylvania; East North Central: Ohio, Indiana, Illinois, Michigan, Wisconsin; West North Central: Minnesota, Iowa, Missouri, North Dakota, South Dakota, Nebraska, Kansas; South Atlantic: Delaware, Maryland, District of Columbia, Virginia, West Virginia, North Carolina, South Carolina, Georgia, Florida; East South Central: Kentucky, Tennessee, Alabama, Mississippi; West South Central: Arkansas, Louisiana, Oklahoma, Texas; Mountain: Montana, Idaho, Wyoming, Colorado, New Mexico, Arizona, Utah, Nevada; Pacific: Washington, Oregon, California, Alaska, Hawaii.
- ^c Estimated coefficient is insignificant at 5% level.
- ^d A single observation with an implied price more than 200 times larger than the nearest price was dropped in the estimation of the LPG equation.
- ^e Plants in the primary aluminum category (SIC 3334) were dropped in the estimation of the coal equation because of the specialized use of coal in the production of aluminum.

For example, the regression for electricity reveals that the Middle Atlantic and New England regions pay the highest prices for this energy source. The regression for natural gas reveals that the South West Central region (including Texas, Louisiana, Mississippi, and Oklahoma) pays the least for gas. Because energy prices are calculated as the ratio of total expenditures to total amount purchased, the price and quantity variables used in these regressions are simultaneously determined, which results in biased estimates.

3 DISAGGREGATION OF INDUSTRIAL CHEMICAL SECTOR

A standard industrial classification (SIC) code or number is assigned to a plant or establishment on the basis of its primary product. However, plants included in a given SIC often manufacture a variety of products, many of which can correspond to some other SIC. Chemical plants are a good example. Chemical plants in SIC 2812 (alkalies and chlorine classification) manufacture chlorine and coproducts as their primary products. Yet many plants in SIC 2869 (industrial organic compounds not elsewhere classified [n.e.c.] classification) or other SICs are also involved in producing chlorine. Chlorine is a secondary product of these plants.

An SIC code is assigned to a product category on the basis of the product's primary industry. If a portion of the products are actually manufactured in plants with other SICs (i.e., these that belong to other industries), this fact is disregarded. When the first three or four digits in the SIC of a particular product category are the same as those digits in the SIC of the plant that produces it, the product is the primary product of that plant and that plant is part of the primary industry for that product. In other words, the product and plant correspond.

The correspondence issue is a major one for high-energy-consumption (HEC) products. How much of an HEC product is produced outside its primary industry? This question can also be expressed inversely: For a given industry or SIC, what fraction of its total production is represented by its primary product? Consider the most important HEC products in terms of the distribution of production at the three-digit SIC level. Table 17 shows the main results (see the column in Table 17 that shows the primary product specialization rating). For several industries, as much as 20% of their production value is represented by products not in that SIC.

In analyzing energy use, knowing the industrial classifications (SICs) assigned to plants is less revealing than knowing the distribution of production by HEC product. The 16 five-digit SIC level HEC products selected for analysis are shown in Table 18. They fall into three three-digit SIC level industries (281, 286, and 287). Much of the energy use in these three industries is associated with these products. The products for which about 20% or more of production occurred outside their primary three-digit SIC industry are shown in Table 19. Several of the most important HEC products are in this subset. They account for perhaps 50 billion kWh of the electricity consumed each year and one-half to one quad⁹ of fuel use (excluding feedstock use). Only the ethyl alcohol category is relatively unimportant.

Only 75% of the primary products in the industrial inorganic chemicals industry are produced in their primary industry (SIC 281). Only 82% of synthetic ammonia and related compounds are produced in their primary industry (SIC 287). One-third of the selected

⁹ One quad = 10^{15} Btu.

TABLE 17 Product Analysis for 1987 and Earlier Census of Manufacture Years

SIC	Industry	Year	Value of Primary Products (\$10 ⁶)	Value of Secondary Products (\$10 ⁶)	Primary Product Specialization Rating (%) ^a	Total Value of Product Made in All Industries (\$10 ⁶)	Coverage Ratio (%) ^b
2873	Nitrogenous fertilizers	1987	2,222.90	157.30	93	2,788.30	81
		1982	2,896.80	388.00	88	3,398.80	79
		1977	2,068.10	383.30	84	2,811.50	74
2874	Phosphatic fertilizers	1987	3,472.30	388.00	92	3,810.00	96
		1982	3,515.80	224.80	94	3,742.00	84
		1977	2,328.80	172.40	93	2,002.00	91
2875	Fertilizers, mixing only	1987	1,290.20	18.40	99	1,370.00	94
		1982	1,440.30	50.10	97	1,528.90	94
		1977	1,451.20	71.60	96	1,480.00	98
2979	Agricultural chemicals not elsewhere classified	1987	4,624.40	1,390.30	77	5,574.80	83
		1982	4,148.30	900.70	81	4,704.00	87
		1977	2,071.10	420.10	81	2,704.00	77
2812	Alkalies and chlorine	1987	1,314.40	217.90	88	2,033.50	85
		1982	1,220.00	282.70	81	2,340.10	67
		1977	1,034.80	610.70	63	1,700.70	58
2813	Industrial gases	1987	2,488.70	54.00	98	2,831.10	84
		1982	1,800.00	34.40	88	2,002.20	81
		1977	1,034.80	33.00	97	1,198.10	83
2818	Inorganic pigments	1987	2,189.20	132.80	84	2,425.50	80
		1982	1,630.00	198.40	88	1,500.70	88
		1977	1,269.90	145.50	80	1,330.20	84

TABLE 17 (Cont.)

SIC	Industry	Year	Value of Primary Products (\$10 ⁶)	Value of Secondary Products (\$10 ⁶)	Primary Product Specialization Rating (%) ^a	Total Value of Product Made in All Industries (\$10 ⁶)	Coverage Ratio (%) ^b
2819	Industrial inorganic chemicals	1987	8,171.50	828.50	91	10,257.80	80
		1982	7,438.00	680.00	91	9,898.20	77
		1977	8,312.70	770.90	87	8,920.30	77

^a Primary products divided by the total of primary products plus secondary products, expressed as a percentage.

^b Primary products divided by total of those products made in all industries, expressed as a percentage.

Source: U.S. Bureau of the Census (1989).

TABLE 18 High-Energy-Consumption Products Considered^a

Industry and Product Category	SIC
Industrial inorganic chemicals	281
Chlorine	28121
Oxygen	28136
Other industrial gases	28137
Sulfuric acid	28193
Other inorganic acids	28194
Aluminum oxide	28195
Other aluminum compounds	28196
(Other) potassium and sodium compounds	28197
Industrial inorganic chemicals not elsewhere classified	28199
Industrial organic chemicals	286
Cyclic intermediates	28651
Aromatic feedstocks	2865658, ^b 2911D55 ^c
Light refinery gases (LRG) for feedstock use (noncyclics)	2869132, ^d 2911A21 ^e
Ethyl alcohol and related organics	28695
Other industrial organics	28696, 28697
Agricultural chemicals	287
Synthetic ammonia and related compounds	28731
Phosphoric acid	28741

^a Unofficial names are used for brevity in some cases.

^b Includes benzene, toluene, and xylenes not made at refineries.

^c Includes benzene, toluene, and xylenes made at refineries.

^d Includes ethylene, propylene, and butadienes not made at refineries.

^e Includes ethylene, propylene, and butadienes made at refineries.

upstream organics are manufactured in the petroleum refining industry (SIC 291) instead of the primary industry, industrial organic chemicals (SIC 286). (These results are, unfortunately, based on dollar values and not physical values. As discussed elsewhere in this report, the dollar values quoted for intermediate products of vertically integrated firms can be misleading.)

Disaggregating the two-digit SIC level HEC chemical industry (SIC 28) into three-digit SIC level product categories (SIC 281, 286, and 287) offers great advantages for modeling. The cross-production results described here do not represent a critical objection to such a disaggregation. They do suggest, however, that there are major practical limits to the accuracy achieved by combining process-specific information with three-digit SIC level disaggregation.

TABLE 19 Basic Chemical Products for Which a Substantial Amount of Production Occurred Outside the Primary Industry in 1987

Product Category	SIC	Share of Production by Industry SIC (%) ^a					
		281	282	286	287	289	291
Chlorine	28121	74.4	0.0	21.6	0.0	0.0	0.0
Major inorganic chemicals ^b	28193,4,7	74.4	2.8	10.8	6.1	0.0	0.0
Industrial inorganic chemicals not elsewhere classified	28199	81.7	0.0	5.0	0.0	0.8	4.3
Light refinery gases (LRG) for feedstock use (noncyclics) ^c	2869132 (2911A21)	0.0	0.0	68.0	0.0	0.0	32.0
Ethyl alcohol and related organics	28695	0.0	0.0	58.0	(Most of remainder in 2046)		
Synthetic ammonia and related compounds	28731	4.1	0.0	3.3	82.1	0.0	0.0

^a Because many small contributions are omitted, percentages do not add to 100. The number 281 = industrial inorganic chemicals, 282 = plastics materials and synthetics, 286 = industrial organic chemicals, 287 = agricultural chemicals, 289 = miscellaneous chemical products, and 291 = petroleum refining.

^b Sulfuric acid, other inorganic acids, and (other) potassium and sodium compounds.

^c Seven-digit product data from Table 5a on industrial organic chemicals, in the CM (U.S. Bureau of the Census 1984).

Source: Unpublished LRD database.

To allow for further insight into this matter, Figure 13 shows the shares of chlorine production in the primary industry (SIC 281), as well as in SIC 286, as indicated by the last four CMs (1972, 1977, 1982, and 1987). Fluctuations may be a result of establishments "crossing the line" between SICs as their production share changes or as a result of other causes. The variation is not large, but it is large enough to show that building an accurate historical energy series for disaggregated chemical sectors requires a considerable amount of care.

FIGURE 13 Distribution of Production of Chlorine

4 CONCLUSIONS

These are the conclusions that have emerged from the work done to date:

- *Energy data from MECS seem to be relatively well edited and reliable.* No operational difficulties were encountered in using the MECS and the LRD at the U.S. Bureau of the Census.
- *The utility of analyzing four-digit SIC data at the plant level is mixed, but the structure of the plant-level MECS/LRD database makes analysis of groupings disaggregated below the four-digit level feasible.* The cement and aluminum sectors appear to be well defined, but the paper and steel sectors exhibit internal variations that indicate these sectors should be further disaggregated before analysis. Moreover, energy-intensive products in the chemical sectors are not always produced in the four-digit SIC sector in which they are classified. Screening analysis revealed a bimodal distribution of efficiency for some sectors, and Salter diagrams revealed groups of plants with very high and very low energy intensities. These results suggest that further analysis of plants disaggregated by technology or by a more detailed product mix may be required, particularly for the steel and paper industries. The ability to match external data to the plant identifiers in the MECS/LRD makes this additional level of analysis feasible.
- *Value of shipments may not be a completely reliable measure of output for analyzing energy use in the industries selected.* It would be desirable to have some physical measure of plant output against which to compare value-of-shipments data, particularly for the energy-intensive, material-producing sectors whose output is likely to be subject to intrafirm transfers. For industries whose output is more likely to be sold directly at a market price, this issue may be less important.
- *Plant-level prices are an important dimension of the database and the MECS/LRD.* While it is not clear whether low energy prices that result from local market conditions induce more energy use or whether those who use more energy enjoy some market power over prices (making these prices lower), neither of these effects can be adequately examined through the use of aggregate data.
- *It is possible to implement fairly traditional structural models of plant-level production on the basis of plant-level data.* Plant-level models are much more likely to result in insights about the actual behavior of industries than are analyses of aggregate data.

- *The degree of diversity across some selected four-digit industries suggests that additional disaggregation according to type of technology would be useful.* For industries with active and cooperative trade associations, data on technologies at the process level could be identified, used to supplement the MECS/LRD data, and then used to further disaggregate the industry samples (e.g., steel plants could be disaggregated into those with and without electric arc furnaces). Alternative industry samples that would not be tied to the SIC system could also be constructed. For example, one could construct a sample of plants that produce ammonia as one of their products instead of a sample of just those plants that produce ammonia as their primary product (these latter plants are the ones that are included in SIC 2873).

The CES staff members have been very cooperative in supporting analyses of these and other modeling issues. They suggested several sources that could be used to extend this analysis. *Current Industrial Reports* (CIRs), which are published by the U.S. Bureau of the Census and include details on physical output and material use that are not a routine part of the LRD, could be used. Although matching CIR and LRD data could prove to be problematic, the only loss of data would be associated with a few plants in the MECS/LRD sample. Another option would be to use external data from trade associations and other publicly available sources. Possible sources include the *Lockwood-Post's Directory of the Pulp, Paper and Allied Trades* for paper mills, a database for the steel industry compiled by Steve Karlson at Northern Illinois University in DeKalb, and a similar database for the cement industry compiled by David Rosenbaum at the University of Nebraska. The National Acid Precipitation Assessment Program (NAPAP) boiler inventory for 1985 is another source that could supplement the MECS/LRD. The CES will make the software it has written available to help solve the data matching problems and minimize the amount of "hand" work required to make use of these data.

5 REFERENCES

- American Metal Market, 1988, *Metal Statistics 1988*, Fairchild Publications, New York, N.Y.
- Davis, S.J., et al., 1991, "Published Versus Sample Statistics from the ASM: Implications for the LRD," discussion paper CES 91-1, U.S. Bureau of the Census, Center for Economic Studies, Washington, D.C.
- CES, 1990, *Longitudinal Research Database Technical Documentation*, U.S. Bureau of the Census, Center for Economic Studies, Washington, D.C.
- DOE, 1982, *Methodological Report on the 1980 Manufacturing Industries Survey of Large Combustors (EIA-463)*, DOE/EIA-0306, U.S. Department of Energy, Energy Information Administration, Washington, D.C.
- DOE, 1991, *Manufacturing Energy Consumption Survey: Consumption of Energy, 1988*, DOE/EIA-0512(88), U.S. Department of Energy, Energy Information Administration, Washington, D.C., May.
- DOE, 1992, *Development of the 1991 Manufacturing Energy Consumption Survey*, DOE/EIA-0555(92)/2, U.S. Department of Energy, Energy Information Administration, Washington, D.C.
- Färe, R., et al., 1985, *The Measurement of Efficiency in Production*, Kluwer-Nijhoff Publishing, Boston, Mass.
- Farrell, M., 1957, "The Measurement of Productive Efficiency," *Journal of the Royal Statistical Society, Series A, General* 120:253-281.
- McGuckin, R.H., 1990, "Longitudinal Economic Data at the Census Bureau: A New Database Yields Fresh Insights on Some Old Issues," discussion paper CES 90-1, U.S. Bureau of the Census, Center for Economic Studies, Washington, D.C.
- Office of Management and Budget, 1987, *Standard Industrial Classification Manual, 1987*, Washington, D.C.
- U.S. Bureau of the Census, 1984, *1982 Census of Manufactures*, Washington, D.C.
- U.S. Bureau of the Census, 1989, *1987 Census of Manufactures*, Washington, D.C.

APPENDIX A:**MEASURES OF INPUTS AND OUTPUT USED IN
CALCULATION OF TECHNICAL EFFICIENCY***Input*

- PW88:** Average number of production workers (1988 ASM).
- OE88:** Nonproduction personnel (1988 ASM).
Number of nonproduction personnel employed during the pay period including March 12.
- CP88:** Cost of materials, parts, etc. (1988 ASM).
Exclusive of cost of fuels, purchased electricity, contract work, and resales.
- MAE87:** Machinery assets, end of year (1987 ASM).
Original cost of depreciable machinery assets.
- BAE87:** Building assets, end of year (1987 ASM).
Original cost of depreciable building assets.
- DEM10:** Electricity demand (1988 MECS).
Amount of electricity consumed on site, regardless of where or how produced.
Includes cogenerated and self-generated electricity.
- FOSS:** Fossil fuel inputs for heat, power, and electricity generation (1988 MECS).
- OTHER:** All other inputs for heat, power, and electricity generation (1988 MECS).

Output

- VS88:** Value of shipments (1988 ASM).

APPENDIX B:

DESCRIPTION OF KARLSON'S DATABASE ON STEEL INDUSTRY PLANTS

Plant-specific estimates of capacity, output, and employment for the 165 steel plants that have operated in the United States between 1970 and 1990 are available (not all plants operated in all 21 years). The American Iron and Steel Institute's *Directory of Iron and Steel Works of the United States and Canada* (Institute Directory), printed in 1970, 1974, 1977, 1980, and 1984, and the Association of Iron and Steel Engineers' *Directory, Iron and Steel Plants* (Association Directory), published annually, are the principal sources for identifying the plants in operation in a given year. The Institute Directory lists the iron- and steel-making furnaces present at each plant. The Association Directory lists the capacity of some plants. The complete inventory of steel-making capacity at all U.S. plants begins with information published in Deily (1980); this 1979 information has been extended with data from capacity reports published occasionally by *Iron Age* or *Iron and Steel Engineer* and from the complete capacity roundups published annually in *Iron and Steelmaker* since 1987 for electric furnaces and since 1988 for blast furnaces and basic oxygen furnaces. (The sources sometimes report different capacities for the same equipment in the same year, in which case, the smaller figure is used.)

Estimates of plant-specific production begin with raw steel production statistics published in the American Iron and Steel Institute's *Annual Statistical Report*. The Institute reports nationwide raw steel production and disaggregates it by state (for the major steel-producing states) or by groups of states (for those states that have few steel producers). The state totals are one set of controls used to check the validity of the estimates.

The second set of control totals consists of company-specific production totals that were reported in the *Iron Age* "Steel Industry Financial Analysis" until the end of 1984. These totals are extended by data from Moody's *Industrial Manuals*, Standard and Poor's *Corporation Records* and *Stock Price Reports*, and the Institute for Iron and Steel Studies' *Iron and Steel Engineer*. These figures allow the output of six or seven plants to be exactly identified. The company-specific totals make up two-thirds to four-fifths of total U.S. steel production in each year. The difference between total national production and the sum of company-specific production is an otherwise unassigned total.

First an initial estimate of production by each company's plant and by each unassigned plant is made. The initial estimate allocates the company's output or the unassigned output to each plant in the same proportion that each plant represents with respect to its company's total capacity or to unassigned capacity. These initial estimates are then adjusted to satisfy the state production totals. The revised estimates are updated iteratively until the outputs for each plant satisfy both the company (or unassigned) production total and the state production total. The method is the same method that Karlson (1980, 1983) used to estimate fully integrated steel plant outputs for 1972; modern spreadsheets reduce the computation involved in calculating and updating estimates for 21

years of data. The capacity and production data underlie empirical research reported in Karlson (1983, 1986).

An employment database is being built from public sources. *Marketing Economics Key Plants* and *Ward's Business Directory* list the employment at each steel plant. Some plants report their employment in the Association Directory; volumes for each year from 1970 forward can be used to construct a complete time series of plant employment.

BIBLIOGRAPHY FOR APPENDIX B

American Iron and Steel Institute, 1970-1990, *Annual Statistical Report*, Washington, D.C.

American Iron and Steel Institute, various years, *Directory of Iron and Steel Works of the United States and Canada*, Washington, D.C.

Association of Iron and Steel Engineers, various years, *Directory, Iron and Steel Plants*, Pittsburgh, Penn.

Deily, R.L., 1980, *Steel Plants USA 1960-1980 Annual Steelmaking Capacities*, IISS Commentary Special Report I, Institute for Iron and Steel Studies, Green Brook, N.J.

Gale Research, Inc., various years, *Ward's Business Directory of U.S. Private and Public Companies*, Detroit, Mich.

Iron Age, various years, "Iron Age Steel Industry Financial Analysis."

Iron and Steel Society, various years, "Electric Arc Furnace Roundup — United States," *Iron and Steelmaker*.

Karlson, S.H., 1980, "Industrial Location and Economic Performance: Spatial Competition Involving a Raw Material and an Econometric Model of Production in Space," Unpublished Ph.D. dissertation, University of Wisconsin, Madison.

Karlson, S.H., 1983, "Modeling Location and Production: An Application to U.S. Fully-Integrated Steel Plants," *The Review of Economics and Statistics* 65:41-50, Feb.

Karlson, S.H., 1986, "Adoption of Competing Inventions by United States Steel Producers," *The Review of Economics and Statistics* 68:415-422, Aug.

Marketing Economics Institute, Ltd., various years, *Marketing Economics Key Plants: Guide to Industrial Purchasing Power*, Marketing Economics Institute, Ltd., New York, N.Y.

Moody's Investors Service, Inc., various years, *Moody's Industrial Manual*, New York, N.Y.

Moody's Investors Service, Inc., various years, *Moody's OTC Industrial Manual*, New York, N.Y.

Standard and Poor's Corporation, various years, *Corporation Records*, New York, N.Y.

Standard and Poor's Corporation, various years, *Stock Reports: American Stock Exchange*, New York, N.Y.

Standard and Poor's Corporation, various years, *Stock Reports: New York Stock Exchange*, New York, N.Y.

Standard and Poor's Corporation, various years, *Stock Reports: Over the Counter*, New York, N.Y.

Stone, J.K., and E.M. Michaelis, various years, "North American BOF Roundup," *Iron and Steelmaker*.

**DATE
FILMED**

8 / 24 / 93

END