ABSTRACT

Title of dissertation: ENERGY PRICES AND SUBSTITUTION IN U.S.
MANUFACTURING PLANTS

Cheryl Grim, Doctor of Philosophy, 2006

Dissertation directed by: Professor John C. Haltiwanger
Department of Economics

Persistent regional disparities in electricity prices, growth in wholesale power

markets, and recent deregulation attempts have intensified interest in the performance of

the U.S. electric power industry, while skyrocketing fuel prices have brought renewed

interest in the effect of changes in prices of all energy types on the U.S. economy. This

dissertation examines energy prices and substitution between energy types in U.S.

manufacturing.

I use a newly constructed database that includes information on purchased

electricity and electricity expenditures for more than 48,000 plants per year and

additional data on the utilities that supply electricity to study the distribution of electricity

prices paid by U.S. manufacturing plants from 1963 to 2000. I find a large compression

in the dispersion of electricity prices from 1963 to 1978 due primarily to a decrease in

quantity discounts for large electricity purchasers. I also find that spatial dispersion in

retail electricity prices among states, counties and utility service territories is large, rises

over time for smaller purchasers, and does not diminish as wholesale power markets

expand in the 1990s.

In addition, I examine energy type consumption patterns, prices, and substitution

in U.S. manufacturing plants. I develop a plant-level dataset for 1998 with data on

consumption and expenditures on energy and non-energy production inputs, output, and

other plant characteristics. I find energy type consumption patterns vary widely across manufacturing plants. Further, I find a large amount of dispersion across plants in the prices paid for electricity, oil, natural gas, and coal. These high levels of dispersion are accounted for by the plant's location, industry, and purchase quantity. Finally, I present estimates of own- and cross-price elasticities of demand for both the energy and non-energy production inputs.

ENERGY PRICES AND SUBSTITUTION IN U.S. MANUFACTURING PLANTS

by

Cheryl Grim

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2006

Advisory Committee:

        Professor John Haltiwanger, Chair
        Professor Clopper Almon
        Professor John Horowitz
        Dr. Ronald Jarmin
        Professor John Shea

PREFACE

This document contains research and analysis conducted at the Center for Economic Studies at the U.S. Census Bureau. The analysis and results presented in this document are attributable to the author and do not necessarily reflect concurrence by the U.S. Bureau of the Census. This document has undergone a more limited review by the Census Bureau than its official publications. It has been screened to ensure that no confidential data are revealed.

This dissertation is dedicated to my husband.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

Persistent regional disparities in electricity prices, growth in wholesale power markets, and recent deregulation attempts have intensified interest in the performance of the U.S. electric power industry, while skyrocketing fuel prices have brought renewed interest in the effect of changes in prices of all energy types on the U.S. economy. This dissertation utilizes plant-level data to examine energy prices and substitution between energy types in U.S. manufacturing.

Specifically, Chapter 2 examines the distribution of electricity prices paid by U.S. manufacturing plants from 1963 to 2000.[1] Chapter 3 looks at electricity, oil, natural gas, and coal consumption patterns, prices paid for these energy types, and substitution between both energy and non-energy inputs to production by manufacturing plants. Chapter 4 gives a detailed description of the construction of the electricity database utilized in Chapter 2.[2] Finally, Chapter 5 provides concluding remarks for this dissertation.

Despite the intensified interest in the U.S. electric power industry, we lack broad empirical studies of electricity prices paid by end users, and there are major gaps in our knowledge of retail pricing patterns and their evolution over time. These gaps hamper efforts to place recent developments in historical perspective, to evaluate the impact of regulatory changes on electricity users, and to assess theories of public utility pricing.

---

[1] Chapter 2 draws heavily upon Davis, Grim, Haltiwanger, and Streitwieser (2006a).

[2] Chapter 4 draws heavily upon Davis, Grim, Haltiwanger, and Streitwieser (2006b).

In Chapter 2, I use a newly constructed database that includes information on purchased electricity and electricity expenditures for more than 48,000 plants per year and additional data on the utilities that supply electricity to study the distribution of electricity prices paid by U.S. manufacturing plants from 1963 to 2000. The data show tremendous cross-sectional dispersion in the electricity prices paid by manufacturing plants, reflecting spatial price differences and quantity discounts. Price dispersion declined sharply between 1967 and 1977 because of erosion in quantity discounts. Differences among utilities in the purchases distribution of their customers are exploited to estimate the role of cost factors and markups in quantity discounts. The estimation results reveal that supply costs per watt-hour decline by more than half over the range of customer-level purchases in the data, regardless of time period. Prior to the mid 1970s, marginal price and marginal cost schedules with respect to annual purchase quantity are remarkably similar, in line with efficient pricing. In later years, marginal supply costs exceed marginal prices for smaller manufacturing customers by 10% or more. Spatial dispersion in retail electricity prices among states, counties and utility service territories is large, rises over time for smaller purchasers, and does not diminish as wholesale power markets expand in the 1990s.

Considerable effort has gone into the study of energy substitution in manufacturing. Previous empirical research focused primarily, though not solely, on the energy aggregate with the bulk of earlier studies using industry level data. In Chapter 3, I examine energy type consumption patterns, prices, and substitution in U.S. manufacturing plants for multiple energy types using plant-level data.

I develop a plant-level dataset for 1998 with data on consumption and expenditures on energy types from the 1998 Manufacturing Energy Consumption Survey (MECS) and data on non-energy production inputs, output, and other plant characteristics from the 1998 Annual Survey of Manufactures (ASM), the Longitudinal Business Database (LBD), and the Chiang-Haltiwanger dataset. I find energy type consumption patterns vary widely across manufacturing plants with over half of plants, 55.6%, consuming only electricity and natural gas and only 0.22% of plants consuming all four energy types. Further, I find a large amount of dispersion across plants in the prices paid for electricity (43%), oil (46%), natural gas (27%), and coal (43%). These high levels of dispersion are accounted for by plant location, industry, and purchase quantity. Finally, I exploit the cross-sectional plant-level variation in input prices to estimate own- and cross-price elasticities of demand for both the energy and non-energy production inputs.

Development of the Prices and Quantities of Electricity in Manufacturing (PQEM) database and corresponding public release statistics is a primary contribution of this dissertation. In Chapter 4, I describe the construction of the PQEM database in detail. The PQEM, utilized in Chapter 2, contains plant-level observations on electricity purchases, prices, and suppliers for the U.S. manufacturing sector from 1963 to 2000. In constructing the PQEM, considerable effort is devoted to treating anomalous data on electricity prices and quantities in the ASM. A number of coding errors in the ASM data are identified; the raw ASM data contain high error rates in 1983 and from 1989 to 1991. Procedures to correct or impute values for the erroneous data are developed, paying special attention to the years with high error rates. Several other measurement issues pertaining to ASM sample weights in 1963 and 1967, erroneous geographic indicators,

and the creation of consistent industry and geography codes over time are also addressed.

Finally, Chapter 4 includes a description of planned public release statistics based on the

PQEM database.

Chapter 2

Electricity Pricing to U.S. Manufacturing Plants, 1963-2000

2.1   Introduction

Longstanding concerns and recent developments have combined to intensify

interest in the performance of the U.S. electric power industry. These include persistent

regional disparities in retail prices, growth in wholesale power markets, a wave of

restructuring and deregulation initiatives in the 1990s, difficulties in the transition to a

more competitive electricity sector, and, perhaps most spectacularly, the California

electricity crisis of 2000-2001.[3] Despite these concerns and developments, we lack broad

empirical studies of electricity prices paid by end users, and there are major gaps in our

knowledge of retail pricing patterns and their evolution over time. These gaps hamper

efforts to place recent developments in historical perspective, to evaluate the impact of

regulatory changes on electricity users, and to assess theories of public utility pricing.

To help address these issues, we construct a rich micro database – Prices and

Quantities of Electricity in Manufacturing (PQEM) – and use it to study electricity

pricing to U.S. manufacturing plants from 1963 to 2000.[4] The PQEM includes data on

electricity expenditures, purchases (watt-hours) and other variables for more than 48,000

manufacturing plants per year, linked to additional data on the utilities that supply

electricity. Our customer-level data are limited to manufacturers, but they are informative

---

[3] Hirsh (1999), EIA (2000b), Besanko et al. (2001), Borenstein (2002), and Joskow (2005), among others, describe and analyze these matters. Joskow and Schmalensee (1983) anticipate many of the pitfalls and challenges that have confronted reform efforts in the electricity sector.

[4] This work in this chapter draws heavily upon Davis, Grim, Haltiwanger, and Streitwieser (2006a).

about pricing practices for a broader class that includes other industrial customers and large and mid-size commercial customers.[5]

Figure 2.1 displays several measures of dispersion in the distribution of log electricity prices from 1963 to 2000.[6] The price measure is the ratio of the plant's annual expenditures on purchased electricity to its annual purchases (watt-hours). The figure shows purchase-weighted and shipments-weighted price distributions, where the former weights each plant-level observation by watt-hours of electricity purchases, and the latter weights by output as measured by shipments.[7] As seen in Figure 2.1, there is tremendous dispersion in the electricity prices paid by manufacturing plants. The purchase-weighted standard deviation exceeds 38% in all years and reaches 55% in some years. By way of comparison, the hours-weighted standard deviation of log hourly production worker wages among manufacturing plants in the PQEM ranges from 39% to 43% between 1975 and 1993.[8] In other words, the dispersion in electricity prices among manufacturing plants is at least as great as the dispersion in their average hourly wages.

---

[5] We inspected electricity tariffs for several utilities and found that they offered the same menu of electricity pricing terms to manufacturers, other industrial customers, and large and mid-size commercial customers. In addition, average electricity prices for the manufacturing sector behave similarly to average prices for the industrial sector as a whole, as we show below. Industrial purchasers account for 45% of retail electricity sales (watt-hours) in 1963 and 31% in 2000 (EIA, 2003a, Table 8.5). In turn, manufacturing plants account for the lion's share of electricity purchases by the industrial sector.

[6] The natural log transformation is convenient for characterizing the magnitude of price differences and price dispersion. In addition, electricity transmission over power lines and the process of transforming voltage levels involve costs in the form of electrical energy dissipated as heat energy. The dissipation of electrical energy rises with transmission distance, other things equal, so that spatial price differentials are aptly described in log terms. For these reasons, we often consider log price differentials in this chapter, but we also consider prices measured in natural units.

[7] These weighting methods mirror the use of input-weighted and output-weighted distributions in studies that quantify between-plant and within-plant components of productivity growth. Examples include Foster et al. (2001) and van Biesebroeck (2004).

[8] The PQEM lacks clean measures of hourly wages before 1975 or after 1993. See Davis and Haltiwanger (1991) for a detailed study of between-plant wage dispersion in the U.S. manufacturing sector.
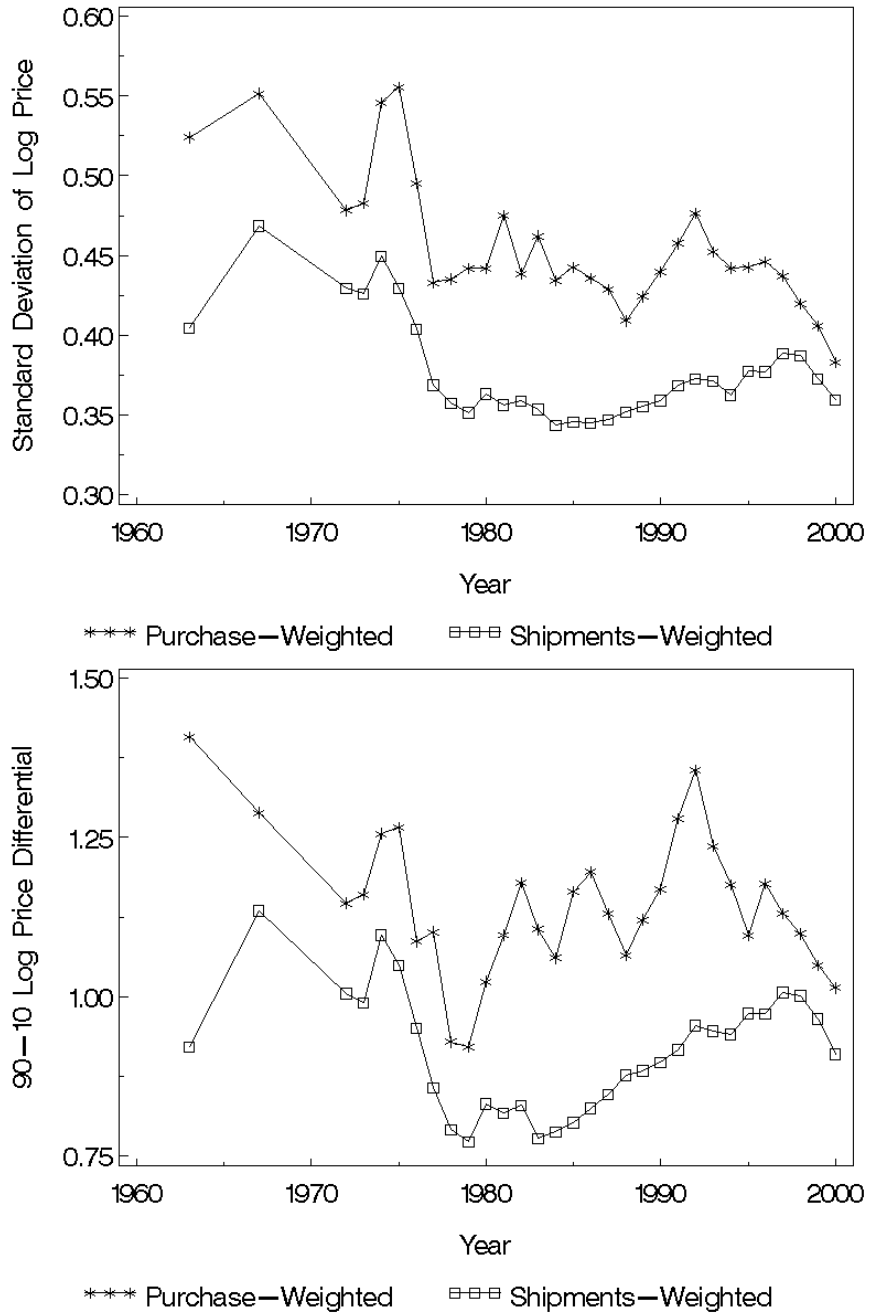
Source: Authors' calculations on PQEM data.

Figure 2.1: Electricity Price Dispersion Among U.S. Manufacturing Plants, 1963-2000

Figure 2.1 also reveals that the log price distribution underwent a great compression from 1967 to the late 1970s. The between-plant standard deviation fell from 55% in 1967 to 44% in 1979 on a purchase-weighted basis and from 47% to 35% on a shipments-weighted basis. Over the same time frame, the 90-10 price differential shrank by about 37 log points under both weighting methods. The 90-10 differential later widened but never returned to the peaks of the 1960s. To the best of our knowledge, this study is the first to quantify the remarkable extent of electricity price dispersion for U.S. manufacturing end-users and the first to document the great compression that played out during the 1970s.

We show below that the great compression episode reflects a sharp erosion of quantity discounts. On a purchase-weighted basis, the average elasticity of price with respect to a plant's annual purchase quantity declined sharply in magnitude from about -22% in 1967 to about -9% in the late 1970s, partially recovering after the mid 1980s. Because the range of electricity purchases among manufacturers is enormous, these elasticities translate into very large price differentials. For example, prices for the biggest purchasers are two-thirds below the median price in the 1960s. Plant-level differences in purchase amounts account for 75% of overall price dispersion among manufacturers in 1963 but only 30% by 1978.

Quantity discounts in the form of declining-block tariffs are a well-known feature of retail electricity pricing for industrial and commercial customers and a sometimes contentious topic in ratemaking proceedings and legislative hearings.[9] They are also the

---

[9] Cudahy and Malko (1976) discuss quantity discounts and other aspects of rate design from the perspective of public utility regulators in a prominent case involving the Madison Gas & Electric Company. Hirsh (1999) recounts the political struggles over federal legislative efforts to reform rate-making practices,

object of careful analysis in theoretical treatments of nonlinear pricing (e.g., Wilson, 1993) and public utility pricing in particular (e.g., Brown and Sibley, 1986). Insofar as the cost of supplying electrical power declines with a customer's purchase quantity, an efficient two-part tariff or other marginal-cost pricing scheme requires quantity discounts. If demand is also more elastic at higher purchase levels, Ramsey pricing by a revenue-constrained public utility entails lower markups for bigger customers and, hence, is another potential explanation for quantity discounts.

These cost and demand determinants of quantity discounts are well understood as a matter of theory, but their importance in practice is unclear. Brown and Sibley (1986) and Borenstein and Holland (2003), for example, argue that the approach to rate setting by electric utilities and their regulators, and the resulting tariff schedules, do not seem well designed to achieve efficient pricing. Moreover, previous research offers no quantitative, theoretically grounded explanation for the sharp erosion in quantity discounts. To address these matters, we propose and implement a novel method for estimating the contribution of cost factors and price markups to quantity discounts. In particular, we exploit the considerable variation across electric utilities in the size distribution of customer purchases to estimate how supply costs per watt-hour vary with customers' annual purchase quantities. The results reveal that supply costs fall by more than half in moving from smaller to bigger purchasers. This pattern holds throughout the past four decades, providing a clear cost-based rationale for quantity discounts.

We use the estimated price and supply cost schedules to construct marginal prices and marginal costs with respect to customer purchase quantity. Comparing the marginal

efforts that culminated in the Public Utilities Regulatory Policies Act (PURPA) of 1978, a major component of President Carter's National Energy Plan.

schedules, we find no support for the Ramsey-pricing view that quantity discounts reflect smaller markups for more elastic demanders. However, the evidence is highly consistent with efficient pricing in the early years of our sample. Indeed, marginal cost and marginal price schedules are nearly identical prior to the mid 1970s. In the upper half of the customer purchase distribution, they are nearly identical from 1967 to 2000. Among smaller manufacturing customers, however, the pricing structure begins to deviate from efficiency after 1973. From 1981 onwards, marginal supply costs for smaller manufacturing customers exceed marginal prices by 10% or more.

We also consider the dispersion in average electricity prices among states, counties and utility service territories. We show that spatial price differentials are large, and that they display three interesting and somewhat surprising time-series patterns. First, in the lower deciles of the purchases distribution, spatial price dispersion widened over time. Second, and in glaring contrast, spatial price dispersion in the top deciles of the purchases distribution fell sharply from the 1960s to the late 1980s. Third, in the 1990s – when wholesale power markets grew rapidly – spatial price dispersion at the retail level did not diminish and even rose modestly over much of the purchases distribution.

The chapter proceeds as follows. Section 2.2 reviews selected economic and regulatory developments in the electric power industry, and Section 2.3 describes the PQEM database. Section 2.4 quantifies the dispersion of electricity prices between and within industries, states, counties, utilities, and purchase size classes. Section 2.5 discusses cost and demand influences on electricity pricing, describes key features of the tariff schedules, and develops evidence on electricity price-quantity schedules. Section 2.6 considers behavioral responses by customers that contribute to a negative relationship

between electricity price and purchase quantity. Section 2.7 estimates supply costs as a function of customer purchase levels, then applies the supply schedules to evaluate whether cost factors can explain quantity discounts and their evolution over time. Section 2.8 constructs marginal price and marginal cost schedules and asks whether they comport with efficient pricing and Ramsey pricing. Section 2.9 summarizes our main findings and identifies several issues for future research.

2.2   Background

From its inception in the 1880s until the mid 1960s, the electric power industry enjoyed a "golden era" in which generating technology improved rapidly, capacity was plentiful, and electricity prices fell.[10] Utilities offered promotional block pricing whereby the price per kilowatt-hour (kWh) declined with purchase amounts. Stimulated by falling real prices, quantity discounts, and new electrical appliances and machinery, electricity consumption grew rapidly after World War II (Hirsh, 1989, Chapter 4). This golden era drew to a close by the late 1960s as unforeseen technological and metallurgical barriers hampered progress in the creation of better electric generators.[11]

Economic factors in the 1970s exacerbated the technological problems facing the industry. Uncertain demand, the high cost of electricity storage and, historically, the absence of peak-load pricing at the retail level made it difficult to project electricity consumption and generating requirements. Accurate projections became more difficult in

---

[10] This view, widely shared by knowledgeable observers, is articulated at length in Hirsh (1999). Joskow (1989) puts it this way: "During the 1950s and most of the 1960s the electric power industry attracted little attention from public policy makers. It experienced high productivity growth, falling nominal and real prices, excellent financial performance, and little regulatory or political controversy."

[11] Chapters 7 and 8 in Hirsh (1989) provide a detailed discussion of the technological difficulties that confronted the electric power industry in the late 1960s and the 1970s.

the 1970s because of large fluctuations in economic activity and in energy input costs. Prices rose sharply for coal and oil, major fuel sources for electricity generation, and there were big disruptions in petroleum supplies. The OPEC Oil Embargo of 1973 precipitated a dramatic rise in oil prices, as did the Iranian Revolution of 1979.

Several regulatory developments added to cost pressures and tightened capacity constraints. Concerns about pollution from conventional power plants and about safety at nuclear power plants led to several pieces of legislation in the late 1960s and 1970s that raised costs and hampered the operation and development of the industry.[12] The National Environmental Policy Act of 1969 required utilities to prepare and defend environmental impact statements for new generator sites. The Clean Air Act of 1970 restricted air pollutants at electricity-generating plants and encouraged utilities to switch from coal to cleaner burning oil or natural gas. The Federal Water Pollution Control Act of 1972 limited waste discharge, and the Resource Conservation and Recovery Act of 1976 set forth standards for utility waste products. Following the 1973 energy crisis, the Energy Supply and Environmental Coordination Act of 1974 authorized the federal government to prohibit purchases of natural gas and petroleum by utilities.[13] The Clean Air Act Amendments of 1977 imposed more stringent restrictions on emissions from electricity-generating plants.

In 1978, several major pieces of legislation passed as part of President Carter's National Energy Plan. The plan included the gradual removal of price controls on oil and

---

[12] See Appendix A of EIA (2000b) for a detailed description of legislation summarized in this paragraph.

[13] Post-energy crisis legislation was intended to reduce U.S. dependency on other countries for fuel. Some of the post-energy crisis legislation was in direct opposition to earlier legislation intended to increase environmental cleanliness.

natural gas, restrictions on the use of oil and natural gas by generating plants, and rate reform provisions for electric utilities. The Public Utilities Regulatory Policies Act (PURPA) of 1978 had the biggest impact on the electricity sector.[14] Its rate-reform provisions were hotly contested in Congress (Joskow, 1979 and Hirsh, 1999) but, in their final form, required that state regulatory authorities merely "consider" various reforms that included an end to promotional pricing. PURPA Section 210 required utilities to buy from and sell power to "qualifying facilities." The goal was to draw non-utilities, such as cogeneration plants and renewable resource plants, into the electric power market. In this respect, PURPA and later legislation had a major impact. By 1999, non-utilities owned 19.8 percent of the electric generating capacity in the U.S. (EIA, 2000a, p.1).

The effect of these technological, economic and regulatory developments on retail electricity prices can be seen in Figure 2.2, which plots the average real price per kilowatt-hour (kWh) for major end-user sectors.[15] Real electricity prices ceased falling in 1970, and they began to rise after 1973, partly because of sharply higher costs for the fossil fuels that powered many of the generating plants. Real electricity prices continued to rise for about ten years, before resuming the historical pattern of steady declines.

---

[14] See Joskow (1989, pages 127-128), White (1996, pages 206-207), and Chapters 4 and 5 in Hirsh (1999). Hirsh (1999) and Gordon (1982) provide extensive discussion of PURPA.

[15] The electricity price series in Figure 2.2 for the residential, commercial and industrial sectors are from the Energy Information Administration (EIA), and the two series for the manufacturing sector are constructed from the PQEM. The EIA data rely on reports from electric utilities, and the PQEM data rely on reports from electricity customers (manufacturing plants). EIA prices are calculated as revenue from retail electricity sales divided by kilowatt hours delivered to retail customers. Real prices are calculated using the BEA implicit price deflator for GDP (1996 = 100). In the EIA data, the industrial sector encompasses manufacturing, mining, construction and agriculture.

Figure 2.2: Real Electricity Prices by End-Use Sector, 1960-2000
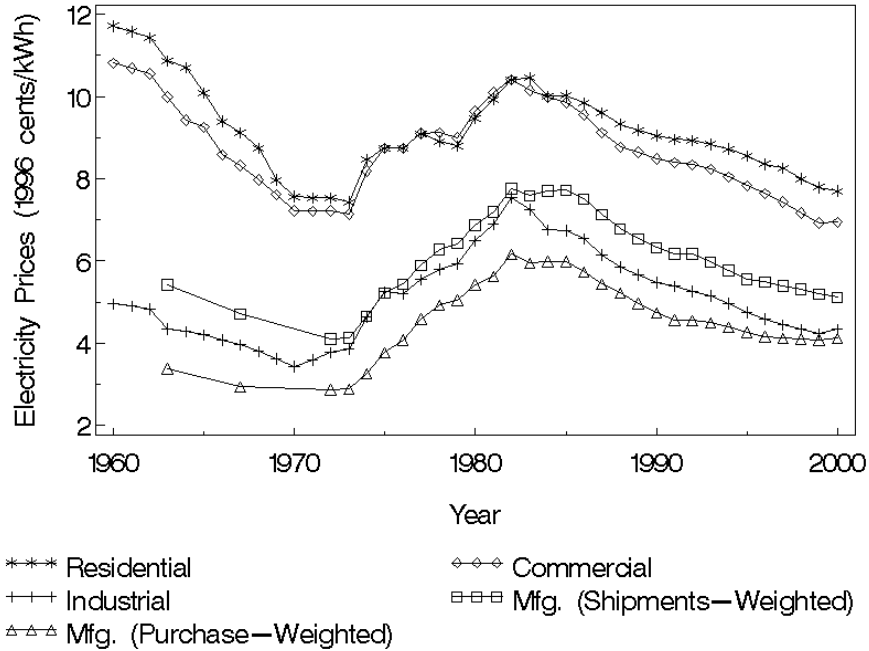
Source: Energy Information Administration for Residential, Commercial and Industrial series; authors' calculations on PQEM data for Manufacturing.

Wholesale trade in electric power expanded rapidly in the 1990s, stimulated by legislative and regulatory policy changes. The Energy Policy Act of 1992 (EPACT) sought to promote greater competition and participation in wholesale markets and to unbundle the sale of electric power from transmission and distribution services (White, 1996 and Besanko et al., 2001). PURPA Section 210, FERC Orders 888 and 889 (issued in 1996) and various state-level actions during the 1990s also stimulated growth in wholesale power markets. These legislative and regulatory actions helped to create a new class of power producers (non-utility qualifying facilities) with secure access to transmission facilities and exemption from many restrictions on public utilities. Sales of

electricity for resale rose from 41% of generated power in 1991 to 61% in 2000 (EIA, 2003b, Tables ES and 6.2).

In recent years, several states have undertaken efforts, not always successful, to introduce greater retail competition in the electricity sector. According to Joskow (2005), the "first retail competition programs began operating in Massachusetts, Rhode Island and California in early 1998 and spread to about a dozen states by the end of 2000." These developments on the retail side occur at the end of the period covered by our data.

2.3   The PQEM Database

The PQEM database derives principally from the U.S. Census Bureau's Annual Survey of Manufactures (ASM) and various data files provided by the Energy Information Administration (EIA). We draw our data on electricity prices and quantities and other variables for individual manufacturing plants from ASM micro files for 1963, 1967, and 1972-2000. The ASM is a series of nationally representative, five-year panels that are refreshed by births as a panel ages. Large manufacturing plants with at least 250 employees are sampled with certainty, and smaller plants with at least 5 employees are sampled randomly with probabilities that increase with the number of employees.[16] ASM plants account for about one-sixth of all manufacturing plants and about three-quarters of manufacturing employment. Our statistics make use of ASM sample weights, so that our results are nationally representative.

---

[16] The number of employees required to be a certainty case is lower in 1963 and 1967. In 1963, all plants in a multi-plant firm with 100 or more employees were sampled with certainty. The same was true in 1967 except for plants in apparel (SIC 23) and printing and publishing (SIC 27), which had certainty thresholds of 250 employees.

ASM plants report expenditures for purchased electricity during the calendar year and annual purchases (kWh). As mentioned above, we calculate the plant-level price as expenditures on purchased electricity divided by quantity purchased. The ASM also contains county and state codes that we use to assign manufacturing plants to electricity suppliers. As described in Chapter 4, we identified and resolved several issues with ASM electricity price and quantity measures in the course of preparing this study. We also cross-checked the ASM data against the Manufacturing Energy Consumption Survey, another plant-level data source at the U.S. Bureau of the Census that relies on a different survey.[17]

We merged ASM plants to their electricity suppliers using the Annual Electric Power Industry Reports, also known as the EIA 861 files. These files include each utility's revenue from sales to industrial customers (by state) and a list of the counties in which the utility has industrial customers. For most counties, the EIA 861 data do not determine a unique assignment of manufacturing plants to electricity suppliers.[18] To address this issue we created a "best-match" utility indicator for each county. Given a list of utilities with industrial customers in the county, the indicator selects the utility with the most statewide revenues from sales to industrial customers. Based on each manufacturing plant's county of operation, we then assign it to the utility selected by the best-match

---

[17] The Manufacturing Energy Consumption Survey is conducted by the EIA. The U.S. Bureau of the Census collects and compiles the data for the EIA.

[18] In the PQEM, 459 counties are served by a single utility, 776 are served by 2 utilities, 791 are served by 3 utilities, 536 are served by 4 utilities, 440 are served by 5-7 utilities, and the remaining 29 counties are served by 8-12 utilities. To the best of our knowledge, data on the list of counties served by each electric utility are not available prior to 1999. Hence, we apply each utility's county list for 2000 to all years. See Chapter 4 for more information on the construction of best-match utility.

indicator. We introduce a separate utility code for each state in which a utility operates, because state laws and state-level public utility commissions govern rate setting.

We also exploit publicly available information on the identity of those plants that purchase electricity directly from six large public power authorities.[19] Direct purchasers from public power authorities typically consume large quantities of electricity, and they often accept high-voltage power, operate their own transformers, and obtain electric power at heavily discounted rates. While few in number, these direct purchasers account for a large fraction of electricity purchases in some counties, and they constitute a distinct segment of the retail electricity market. We identified between 58 and 94 direct purchasers from public power authorities per year.

It should be noted that our utility-matching procedures are imperfect, because incorrect assignments can occur in counties served by more than one utility. Matching errors between plants and utilities have no impact on much of our analysis, but they may affect our characterization of price and cost differences among utilities. In work underway, we are refining our matching procedures by drawing on utility service territory maps and zip code data for utility service areas and manufacturing plants. At this time, we have incorporated zip code level data on utility service areas for three states: California, Kentucky, and Ohio.[20] If possible, we plan to incorporate zip code level data for additional states.

---

[19] They are the Tennessee Valley Authority, Bonneville Power Administration, Santee Cooper, New York Power Authority, Grand River Dam Authority, and Colorado River Commission of Nevada. Fourteen public power authorities supplied electricity directly to industrial customers in 2000, but the six largest accounted for nearly 98% of the revenues from direct sales to industrial customers (EIA 861 file).

[20] In the 2000 PQEM, California represents 5.3% of the quantity of purchased electricity, while Kentucky and Ohio represent 3.9% and 7.7%, respectively, of the quantity of purchased electricity. Note, as we did for counties served by multiple utilities, we assign the plants in zip codes served by multiple utilities to the

Finally, we incorporated information from the State Energy Data System 2000 files into the PQEM.[21] These files include annual data on fuel sources used for electricity generation by state from 1960 to 2000. We use this data source to create annual state-level fuel shares of electricity generation for the following five categories: coal, petroleum and natural gas, hydropower, nuclear power, and other (includes geothermal, wind, wood and waste, photovoltaic, and solar).

Table 2.1 reports selected characteristics of the PQEM. The database contains more than 1.8 million plant-level observations over the period from 1963 to 2000. There are 3,031 counties with manufacturing plants and 362 utilities, counting multi-state utilities once for each state in which they sell to industrial customers. The table shows that electricity purchases and cost shares vary enormously across manufacturing plants. For example, the 90th quantile of the purchases distribution is 381 times the 10th quantile on a shipments weighted basis and 736 times on a purchase-weighted basis. The median ratio of electricity costs to labor costs is 4.7% on a shipments-weighted basis and 17.2% on a purchase-weighted basis. While electricity costs are a modest percentage of labor costs for most plants, those for which electricity costs exceed 62% (201%) of labor costs account for one-fourth (one-tenth) of all electricity purchases. In other words, a large fraction of electricity is purchased by plants for which electric power is a primary or major cost of production.

---

utility with the most statewide revenues from sales to industrial customers. We can use our one-to-one zip code to utility matches to check the accuracy of our original county matches for California, Ohio, and Kentucky. In 2000, we have one-to-one zip-utility matches for 69% of California plants, 47% of Ohio plants, and 24% of Kentucky plants. Our original county-based utility matches are correct for 90% of one-to-one zip-utility matched plants in Calfornia, 56% in Ohio, and 91% in Kentucky. We also use the zip code level information to calculate an overall estimated match accuracy rate of 78% in the year 2000. See Chapter 4 for more information on the calculation of the estimation match accuracy rate.

[21] This data is from the State Energy Data System (SEDS) on the Energy Information Administration Internet site, http://www.eia.doe.gov.

Table 2.1: Selected Characteristics of the PQEM Database

Years covered	1963, 1967, 1972-2000
Number of plant-level observations per year	48,164 to 72,128
Total number of annual plant-level observations[a]	1,816,720
Number of counties with manufacturing plants	3,031
Number of 4-digit SIC industries (1972 / 1987)[b]	447 / 458
Number of best-match utilities[c]	362
Mean annual electricity purchases, Gigawatt hours (GWh)[d]	99.7 (860.4)
Standard deviation of annual electricity purchases (GWh)[d]	334.0 (2,400.0)

Weighting Method	Quantiles of Annual Electricity Purchases, Gigawatt-hours[e]								
	1	5	10	25	50	75	90	95	99
Shipments	.07	.30	.70	3.22	16.4	89.2	267	444	1,500
Purchases	.20	1.08	2.84	13.58	85.9	452	2,100	4,185	14,241
Weighting Method	Quantiles of Electricity Costs as a Percent of Total Labor Costs[e]								
Shipments	0.4	1.1	1.5	2.5	4.7	10.2	25.7	46.3	197.2
Purchases	1.1	2.1	3.0	6.1	17.2	61.7	201.0	305.3	3,461

Notes:
[a] The initial sample contains 1,945,813 records. We drop 107 records because of invalid geography codes and 128,058 (6.6%) because of missing values for electricity price, total employment, value added or shipments. We also trim the bottom 0.05% of the electricity price distribution in each year (928 observations over all years).
[b] We use 1972 SIC codes in 1963, 1967, and 1972-1986 and 1987 SIC codes in 1987-2000. See Davis et al. (2004) for additional information.
[c] There are 349 best-match utilities not counting public power authorities: Tennessee Valley Authority, Bonneville Power Administration, New York Power Authority, Santee Cooper, Grand River Dam Authority, and the Colorado River Commission of Nevada. By construction, a best-match utility does not cross state lines.
[d] Weighted by shipments (electricity purchases).
[e] For disclosure reasons, the quantiles shown above are averages of plant-level observations in three quantiles, the quantile shown and the two surrounding quantiles (e.g., quantile 50 as shown is the average of observations in quantiles 49, 50, and 51).

2.4 Price Dispersion Between and Within Groups of Plants

We decompose the variance of electricity prices into within-group and between-group components using indicators for industry, geography, electricity supplier, and purchase quantity. Indexing plants by e and groups by g, write the overall variance as

$$V = \sum_e s_e (p_e - \bar{p})^2 = \sum_g \sum_{e \in g} s_e (p_e - \bar{p})^2$$

$$V = \sum_g s_g \left(\sum_{e \in g} s_e (p_e - \bar{p}_g)^2 \right) + \sum_g s_g (\bar{p}_g - \bar{p})^2 \qquad (2.1)$$

$$V = \sum_g s_g V_g^W + V^B = V^W + V^B$$

where p_e is the log price of electricity for plant e, s_e is the weight for plant e, $\bar{p}$ is the overall weighted mean log price, $\bar{p}_g$ is the weighted mean log price for group g,

$s_g = \sum_{e \in g} s_e$ is the sum of weights for plants in group g, V_g^W is the weighted variance within

group g, and V^B is the between-group variance. Table 2.2 reports the shipments-weighted version of (2.1) and its components for selected years, with s_e set to the product of the plant's ASM sample weight and its shipments value. Table 2.3 reports analogous purchase-weighted statistics.[22]

---

[22] The statistics in Tables 2.2 and 2.3 are presented as descriptive statistics. These tables do not include test statistics for whether the variation is significantly different from zero since the variances shown are based on a very large sample of around 50,000 plants per year.

Table 2.2: The Shipments-Weighted Distribution of Log Electricity Prices Paid by U.S.
Manufacturing Plants, Dispersion and Variance Decompositions

	1963	1967	1972	1977	1982	1987	1992	1997	2000
Overall Standard Deviation	.409	.468	.429	.369	.359	.347	.373	.388	.360
Price Dispersion Between Industries									
4-Digit SIC Industries (447/458)[+]									
Between Variance as % of Total	36.6	36.3	28.0	20.6	19.4	23.1	26.4	25.1	23.8
Between Standard Deviation	.248	.282	.227	.167	.158	.167	.192	.194	.175
Price Dispersion Between Geographic Areas									
NERC Regions (12)									
Between Variance as % of Total	9.0	9.7	12.6	13.2	17.8	14.9	22.1	20.8	21.1
Between Standard Deviation	.123	.146	.152	.134	.152	.134	.175	.177	.165
States (51)									
Between Variance as % of Total	11.9	13.6	17.3	34.8	46.5	36.7	42.7	39.4	38.0
Between Standard Deviation	.141	.173	.179	.218	.245	.210	.243	.244	.222
Utilities (362)									
Between Variance as % of Total	19.9	21.5	22.6	42.5	56.3	44.4	50.7	47.3	45.4
Between Standard Deviation	.183	.217	.204	.240	.269	.231	.265	.267	.242
Counties (3,031)									
Between Variance as % of Total	31.4	32.0	32.2	53.0	67.2	54.3	61.6	57.6	56.3
Between Standard Deviation	.230	.265	.244	.269	.294	.256	.292	.295	.270
Price Dispersion Between Groups Defined by Annual Electricity Purchases									
Purchase Deciles (10)									
Between Variance as % of Total	57.2	54.2	33.2	16.4	19.3	26.2	29.0	30.6	25.6
Between Standard Deviation	.310	.345	.247	.150	.158	.177	.201	.215	.182
Purchase Centiles (100)									
Between Variance as % of Total	61.1	57.2	35.8	18.6	21.6	28.7	31.9	32.7	29.0
Between Standard Deviation	.320	.354	.257	.159	.167	.186	.210	.222	.194
Price Dispersion Between Groups Defined by Utility and Purchase Level									
Utility x Purchase Decile (2,670)									
Between Variance as % of Total	74.0	69.7	55.4	59.2	72.1	65.6	72.7	70.7	68.7
Between Standard Deviation	.352	.391	.320	.284	.305	.281	.318	.326	.298
Utility x Purchase Centile (22,230)									
Between Variance as % of Total	83.4	78.8	66.6	69.9	81.3	76.9	83.0	82.0	80.6
Between Standard Deviation	.374	.416	.351	.308	.324	.304	.340	.352	.323

[+] Years prior to 1987 are classified using the 1977 SIC system (447 4-digit industries).
  Years 1987 and later are classified using the 1987 SIC system (458 4-digit industries).

  Source: Authors' calculations on PQEM data.

Table 2.3: The Purchases-Weighted Distribution of Log Electricity Prices Paid by U.S. Manufacturing Plants, Dispersion and Variance Decompositions

	1963	1967	1972	1977	1982	1987	1992	1997	2000
Overall Standard Deviation	.524	.552	.478	.433	.439	.429	.477	.437	.383
Price Dispersion Between Industries									
4-Digit SIC Industries (447/458)[+]									
Between Variance as % of Total	71.3	61.4	48.8	40.9	37.9	46.8	59.0	44.5	37.5
Between Standard Deviation	.443	.432	.334	.277	.270	.293	.366	.292	.234
Price Dispersion Between Geographic Areas									
NERC Regions (12)									
Between Variance as % of Total	22.1	18.9	19.5	9.2	10.2	8.4	10.3	9.8	13.5
Between Standard Deviation	.246	.240	.211	.131	.140	.124	.153	.137	.141
States (51)									
Between Variance as % of Total	43.8	40.5	37.5	40.0	45.7	38.3	39.3	37.5	39.5
Between Standard Deviation	.347	.351	.293	.274	.297	.265	.299	.268	.240
Utilities (362)									
Between Variance as % of Total	64.9	56.2	50.0	56.1	62.0	54.9	57.2	53.0	51.2
Between Standard Deviation	.422	.414	.338	.324	.346	.318	.360	.318	.274
Counties (3,031)									
Between Variance as % of Total	77.9	69.6	64.9	73.5	78.6	74.9	77.5	69.9	65.4
Between Standard Deviation	.462	.460	.385	.371	.389	.371	.419	.365	.310
Price Dispersion Between Groups Defined by Annual Electricity Purchases									
Purchase Deciles (10)									
Between Variance as % of Total	62.8	56.3	36.2	27.4	24.7	38.0	49.5	41.3	38.1
Between Standard Deviation	.415	.414	.288	.227	.218	.264	.335	.281	.236
Purchase Centiles (100)									
Between Variance as % of Total	74.7	65.5	41.5	33.8	31.8	45.0	60.8	45.9	43.4
Between Standard Deviation	.453	.446	.308	.252	.247	.288	.372	.296	.252
Price Dispersion Between Groups Defined by Utility and Purchase Level									
Utility x Purchase Decile (2,670)									
Between Variance as % of Total	88.8	82.3	69.8	73.0	78.2	75.5	80.8	75.5	73.0
Between Standard Deviation	.494	.500	.400	.370	.388	.372	.428	.380	.327
Utility x Purchase Centile (22,230)									
Between Variance as % of Total	94.2	90.5	80.7	83.3	87.9	87.3	90.6	86.9	84.9
Between Standard Deviation	.509	.525	.430	.395	.411	.400	.454	.408	.353

[+] Years prior to 1987 are classified using the 1977 SIC system (447 4-digit industries).
 Years 1987 and later are classified using the 1987 SIC system (458 4-digit industries).

Source: Authors' calculations on PQEM data.

According to Table 2.2, the shipments-weighted standard deviation of log electricity prices across manufacturing plants stood at 47% in 1967, fell sharply to 37% by 1977, and then changed little over the next 23 years. Price dispersion also fell sharply on a purchase-weighted basis (Table 2.3), from 55% in 1967 to 43% in 1977 and then further in the 1990s to stand at 38% in 2000. Following a similar path, the between-industry dispersion of electricity prices fell rapidly through 1982 and to even lower levels in the 1990s on a purchase-weighted basis. All told, the purchase-weighted dispersion of industry prices fell by almost half over the past four decades.

Tables 2.2 and 2.3 also document several other facts. First, spatial price differentials are large. County effects, for example, never account for less than 65% of the overall price variance on a purchase-weighted basis. About 80-90% of the county-level price dispersion is accounted for by average price differences among the 362 utilities. Second, customer groups defined by electricity purchase quantities also account for a high percentage of overall price dispersion, especially in the 1960s.[23] Price dispersion among purchase-level groups fell by nearly half during our sample period, mostly between 1967 and 1977. Third, purchase level and utility jointly account for a high percentage of price dispersion throughout the past four decades. Groups defined by utility crossed with purchase deciles account for 55-74% of price dispersion on a shipments-weighted basis and 70-89% on a purchase-weighted basis. This grouping involves about 450 fewer cells than grouping by counties, but always accounts for at least as much of the price variation.

---

[23] We group plants by where they fit into the distribution of electricity purchases in the indicated year, allowing the decile and centile boundaries to vary over time.

There is also an apparent paradox in Tables 2.2 and 2.3. Spatial price dispersion declined sharply over time on a purchase-weighted basis but rose on a shipments-weighted basis. Focusing on counties, the purchase-weighted standard deviation fell by nearly one-third from 1963 to 2000, while the analogous shipments-weighted measure rose by one-fifth. Closer examination of the data resolves the paradox: spatial price dispersion diminished dramatically in the top decile of the purchases distribution (more heavily weighted in Table 2.3), but it rose in the bottom five deciles (more heavily weighted in Table 2.2). We highlight this somewhat surprising pattern in Figure 2.3, which shows the evolution of spatial price dispersion for three selected deciles of the purchases distribution. To control for purchase quantity differences within deciles, we construct Figure 2.3 using residuals from annual customer-level regressions of log price on a polynomial in log purchases. As seen in Figure 2.3, there is an enormous decline in spatial price dispersion in Decile 10, comprising the biggest electricity purchasers, from the 1960s to the late 1980s. A similar, but more muted, pattern holds for Decile 9. The remaining deciles exhibit little trend change in spatial price dispersion, as illustrated by Decile 6, or trend increases, as illustrated by Decile 1. Another noteworthy pattern highlighted by Figure 2.3 is the lack of a downward trend in spatial price dispersion during the 1990s, when wholesale power markets grew rapidly.

We summarize the empirical findings to this point in three statements. One, there is tremendous dispersion among manufacturing plants in price per kWh of electricity. Two, the plant-level distribution of electricity prices underwent a great compression through the late 1970s. Three, readily observed plant characteristics such as utility and purchase quantity capture most of the cross-sectional variation in electricity prices. The

rest of the chapter more fully explores the role of utility characteristics and purchase

quantity in electricity pricing and supply costs.
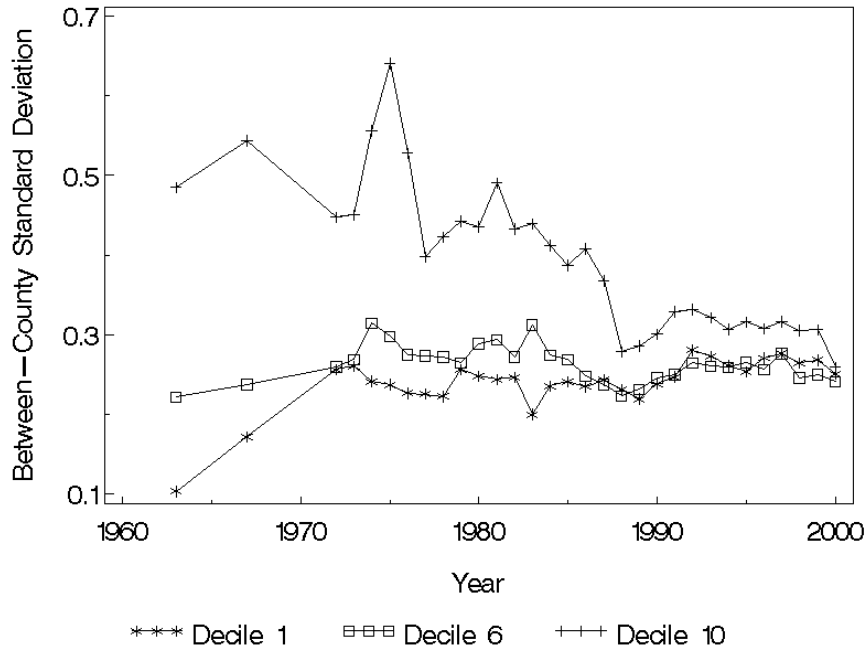
Source: Authors' calculations on PQEM data with part-year observations excluded.

Note: The between-county standard deviations are calculated in a purchase-weighted manner using residuals from annual customer-level regressions of log price on a fifth-order polynomial in log purchases.

Figure 2.3: Spatial Price Dispersion by Selected Deciles of the Purchases Distribution, 1963-2000

2.5 Electricity Price-Quantity Schedules

2.5.1 Cost and Demand Influences on Electricity Pricing

The delivery of electricity to end users requires generating facilities, transmission

lines and transformers. Marginal costs of generation depend on several factors including

power source and generator efficiency. Marginal costs of delivery depend on the physical

characteristics of the transmission grid and distribution system.

Electrical energy dissipates as heat energy during transmission and in the process of transforming voltage levels. One way to lower energy losses and transmission costs is to reduce the resistance to electric current.[24] The resistance of a wire depends on its physical characteristics such as material type, length, and thickness. Resistance increases with the length of the wire and decreases with thickness. Another way to lower transmission costs is to rely on high-voltage power lines that involve less dissipation as heat energy. However, high voltage levels are dangerous, so transformer stations near the final delivery point are typically used to step down voltage levels for end users. The process of transforming voltage levels also involves some dissipation of electrical energy.[25]

Supply costs per kWh of electricity tend to be lower for larger industrial and commercial customers for several reasons. Large purchasers are more likely to locate near generating facilities to minimize transmission losses. High-voltage transmission lines can lead all the way to the customer's doorstep, further reducing transmission costs. A large power user is also more likely to operate equipment at high voltage levels, circumventing or reducing the need for step-down transformers and complex distribution networks. Large power users may operate and maintain their own step-down transformers as well, relieving the utility of this task and associated costs. Larger electricity customers also have stronger incentives to respond to pricing structures that discourage volatile consumption patterns and peak-period consumption. In turn, these incentive responses

---

[24] See Halliday et al. (1992) for a discussion of the basic physics of electricity.

[25] Transformers are used to convert high voltage electricity to low voltage electricity and vice versa. Those that convert from high to low voltage are often called "step-down" transformers. Some power is lost when the electricity is transformed due to eddy currents, which are currents induced by the magnetic field in the iron core of the transformer. The eddy currents heat up the core of the transformer and the energy in that heat is lost.

economize on generating and transmission facilities and mute the effect of system-wide

demand fluctuations on marginal generating costs. Similarly, larger customers have

stronger incentives to consider provisions for interruptible and curtailable power as a

means of lowering electricity costs. These customer supply cost characteristics provide

cost-based rationales for quantity discounts in electricity pricing.

Customer demand characteristics also lead to quantity discounts under plausible

conditions. Consider a utility that prices electricity to maximize consumer surplus subject

to the constraint that its revenues equal its costs. As shown by Goldman et al. (1984),

Brown and Sibley (1986) and Wilson (1993), among others, the optimal nonlinear pricing

schedule for successive increments of electrical power satisfies the Ramsey pricing rule:

$$\frac{M(q) - C(q;Q)}{M(q)} = \frac{-\alpha}{\eta[M(q),q]} \tag{2.2}$$

where $M(q)$ is the marginal price for the customer's qth unit of electricity, $C(q;Q)$ is the

marginal cost of the qth unit when the utility's total quantity supplied is Q, $\eta[M(q),q]$ is

the elasticity of demand for the qth unit with respect to the marginal price, and the

Ramsey number $\alpha \in [0,1]$ is chosen to satisfy the revenue constraint. Note that

$\alpha = 0$ corresponds to marginal cost pricing, and $\alpha = 1$ corresponds to the standard inverse

elasticity rule for a profit-maximizing multi-product monopolist.[26]

---

[26] It is worth pointing out that the revenue constraint does not preclude marginal cost pricing, even for a utility with declining costs over the relevant range. For example, consider a two-part tariff with a fixed access fee for each customer and marginal price set to marginal cost. Set the access fees so that total revenues cover total costs. Then, provided that the access fees are not so high as to deter participation by any consumer who values (some) electricity at more than its marginal cost, this type of two-part tariff is fully efficient (Brown and Sibley, 1986). In this case, $\alpha = 0$ and the Ramsey-pricing condition (2.2) reduces to a form of marginal cost pricing. When efficient pricing is infeasible, the Ramsey pricing rule (2.2) minimizes the allocative distortions induced by pricing above marginal cost.

According to the Ramsey pricing formula (2.2), the markup of price over marginal cost declines with the purchase level provided that demand becomes more price elastic for successive units. Under this condition, Ramsey pricing leads to quantity discounts even when marginal costs are invariant with respect to purchase amount. If marginal costs also decline with purchases, then Ramsey pricing implies that the marginal price schedule declines more steeply than the marginal cost schedule.

2.5.2 Electricity Tariffs for Industrial Customers

Electricity tariffs for industrial customers usually include separate energy and "demand" charges.[27] The energy charge depends on total kilowatt-hours of consumption during the billing period, and the demand charge depends on the highest consumption over 15- or 30-minute intervals within the billing period or longer time period. Roughly speaking, the demand charge reflects the customer's maximal requirements for power. By discouraging uneven and erratic patterns of power consumption, the separate demand charge economizes on the need for generating, transmission and transformer facilities. Eligibility for the most favorable tariff schedules is usually limited to large customers who make long term commitments to minimum contract demand levels that place a high floor on monthly charges.

Traditionally, electric utilities have offered declining-block rate schedules, whereby the marginal price per kWh of energy and the marginal price per kW of demand decline as step functions (Caywood, 1972). For bigger purchasers, in particular, electricity tariffs also depend on other factors such as voltage level and willingness to

---

[27] See Cowern (2001) for a concise introduction to electricity tariffs for industrial customers. Caywood (1972) provides a detailed description of electricity tariffs and rate-setting practices.

accept power interruptions or curtailments. Differential rates by time of day and other

applications of peak-load pricing principles came into wider use after the mid 1970s

(ELR, 1975, and Cudahy and Malko, 1976). Moves toward more finely differentiated

tariff schedules for industrial customers continued through at least the late 1980s

(Wilson, 1993, pages 36-38). The California Electricity Crisis of 2000-2001 intensified

interest in retail pricing structures (Borenstein and Holland, 2003).

     As an illustration of current and past practice, Table 2.4 summarizes the menu of

electricity tariff schedules offered to industrial customers by Santee Cooper Power.[28] The

tariffs contain three main charges: a monthly customer charge, monthly demand charges,

and monthly energy charges. Larger customers face smaller energy charges per kWh and

smaller demand charges per kW but higher monthly minimum charges. For example, the

Medium General Service schedule offers an energy charge of 2.6¢ per kWh, a demand

charge of $11.85 per kW, and a minimum monthly payment of $29. The Large Power and

Light schedule offers a lower energy charge of 2.19¢ per kWh and a lower demand

charge of $10.76 per kW, but a much higher minimum monthly payment of $11,960.[29]

---

[28] Santee Cooper is also known as the South Carolina Public Service Authority. Among utilities with positive industrial revenue, Santee Cooper is close to average size with industrial sales of $238 million in 2000. The Santee Cooper schedules reflected in Table 2.4 are in effect as of July 2004 and date back to 1996. They are available for download at http://www.santeecooper.com/.

[29] This minimum holds for a customer who contracts for at least 1,000 kW of firm power. Lower minimum charges are available to customers who accept interruptible or curtailable power.

Table 2.4: Menu of Electricity Tariff Schedules Offered to Industrial Customers by Santee Cooper Power as of July 2004

Service Type and Schedule	Energy Charge Per kWh	Monthly Demand Charge Per kW	Minimum Monthly Demand Charge	Own Trans-Former Discount?	Monthly Customer Charge	Customer Profile
General Service, GN-96	6.56¢	None	None	No	$6.85	Less than 90 MWh per year
Medium General Service, GS-96	2.60¢	$11.85	$11.85	No	$16.15	Greater than 90 MWh and less than 1,080 MWh per year
Large General Service, GL-96 (Optional provision for interruptible power)	2.32¢	$13.20 ($8.57 for interruptible portion)	$3,960	Yes, $0.50 per kW	$24.00	Greater than 1,080 MWh per year, and delivery points near transmission line
General Service Time of Use, GT-96	2.32¢	$13.20 peak, $3.87 off-peak		No	$24.00	Greater than 90 MWh per year
Large Power and Light, L-96 (Requires 5-year contract with high floor on demand charges)	2.19¢	$10.76 (extra $6.00 per kW in excess of contract level)	$10,760 (for 1,000 kW of Firm Power)	Yes, $0.50 per kW	$1,200	Demand greater than 1,000 kW and delivery points near transmission lines; minimum 5-year commitment.
Optional Riders to Large Power and Light Schedule						
Curtailable Supplemental Power, L-97	Different energy charges and a discount of 72% on demand charges for supplemental power that is subject to temporary or permanent curtailment or interruption with six months notice.					
Interruptible Power, L-02-I	Discount of 36% on demand charges for power subject to curtailment or interruption on short notice (2.5 hours); limitations on frequency and duration of curtailments and interruptions; one-year advance notice required by customer to reduce interruptible portion of demand.					
Off-Peak Service, L-96-OP	Discount of 80% on demand charges for off-peak power in excess of contracted levels for Firm, Supplemental and Interruptible Demands; subject to curtailment or interruption on short notice.					
Economy Power, L-02-EP	Discounted energy charges offered, at Santee Cooper's sole discretion, to customers with Contract Demand greater than 2,000 kW. Available on short notice during specified clock hours.					
Standby Power, L-96-SB	Available at Santee Cooper's discretion to customers with alternative non-emergency power sources.					

(Table 2.4 Notes on Next Page)

Notes:

1. The charges listed above exclude South Carolina Sales Tax and other taxes and fees levied by governmental authorities.
2. Electricity is metered and billed separately for each delivery point and voltage level, so that the Monthly Customer Charge and Minimum Monthly Demand Charge apply per delivery point and voltage level.
3. All service types are subject to a Fuel Adjustment Clause (FAC-96) whereby the energy charge per kWh is adjusted by an additive factor that depends on Santee Cooper's fuel costs in the preceding three months, an allowance for its capital improvements and distribution losses, and other considerations. The energy charge adjustment per kWh is similar for all service types, but the adjustment is less sensitive to capital improvements and distribution losses under the Large Power and Light schedule. Under all schedules, standard "firm-requirements" service is also subject to a Demand Sales Adjustment Clause (DSC-96) that credits Santee Cooper customers with specified shares of its demand-related and capacity-related revenues. The Demand Sales Adjustment can be positive or negative. It is applied as a proportional adjustment to the monthly demand charge under the Large Power and Light schedule and as a proportional adjustment to the monthly energy charge under the General Service schedules.
4. The kW level used to calculate the Monthly Demand Charge can be greater than "Measured Demand" during the billing period, defined as "the maximum 30-minute integrated kW demand recorded by suitable measuring device during each billing period." For example, the Medium General Service schedule states that the "monthly Billing Demand shall be the greater of (i) the Measured Demand for the current billing period or (ii) fifty percent (50%) of the greatest Firm Billing Demand computed for the preceding eleven months." The Large General Service schedule specifies a 70% figure.
5. The discounted Demand Charge under the General Service Time-of-Use Schedule applies to the difference between the customer's Off-Peak Measured Demand and the customer's On-Peak Measured Demand.
6. The transformer discount requires that the customer take delivery at available transmission voltage (69kV or greater).
7. Customers that opt for curtailable or interruptible power forfeit all discounts previously received during the calendar year for such power in the event they fail to meet a request for power curtailment or interruption. In addition, future discounts for curtailable and interruptible power can be withdrawn.
8. Under the Large Power and Light schedule, the customer must commit to a Firm Contract Demand level for a five-year period. The Firm Contract Level places a floor on the demand level used to compute the Monthly Demand Charge. Lower minimum monthly demand charges are available under certain conditions. The Large Light and Power Schedule also includes an Excess Demand Charge of $6.00 per kW for Measured Demand in excess of the Firm Contract Demand, a charge of $0.44 per kVAr of Excess Reactive Demand, and a Monthly Facilities Charge equal to 1.4% of the original installed cost of any facilities that Santee Cooper provides in addition to the facilities it normally provides to its customers.

Source: Santee Cooper tariff schedules for commercial and industrial customers at http://www.santeecooper.com/ (20 July 2004).

Large Santee Cooper customers who locate near transmission lines and provide their own transformers receive discounts of roughly 4% on demand charges. Optional riders to the Large Power and Light schedule offer big discounts on demand charges for off-peak power and power subject to curtailment or interruption. The Large Power and Light schedule and its optional riders require a five-year customer commitment to a contract demand level of at least 1,000 kW and the implied demand charges. These basic features of the Santee Cooper tariff schedules are similar to the tariff menu offered to industrial customers by Pacific Gas & Electric in 1988, as described in Wilson (1993), and to the illustrative tariff schedule for industrial customers reported by Caywood in the 1956 and 1972 editions of *Electric Utility Rate Economics.*

Recall that the PQEM contains the average price per kWh paid by a plant during the calendar year, so it does not capture the full complexity of the underlying electricity tariff schedules. In this respect, the PQEM is analogous to household and establishment-level data sets that report workers' average hourly or annual wages but not the details of the underlying compensation arrangements. To be sure, the lack of data on the underlying tariff schedules (or compensation terms) is a limitation, but it does not preclude an informative analysis. Despite the complexity of real-world compensation arrangements, there is a vast body of informative research on wage structure and labor demand that fruitfully exploits simple data on wage rates for individual workers and employers. Our empirical analysis of the retail pricing structure for electricity is in the same spirit.

2.5.3 Empirical Price-Quantity Schedules

We now present evidence on empirical price-quantity schedules for electricity, and changes in these schedules over time. When a plant operates for only part of the

calendar year, the PQEM measure of annual kWh does not accurately indicate where the plant fits into the purchases distribution. For this reason, we henceforth exclude part-year observations.[30] We also exclude observations that display extreme seasonality or variation in production activity within the year, because customers with highly variable loads typically face special tariff schedules with higher charges.[31]

Figure 2.4 shows the mean log real price of electricity by purchase decile from 1963 to 2000. The purchase deciles are almost perfectly rank ordered by price during the past four decades. Price differentials peak in 1967, when the gap in mean price between the top and bottom deciles exceeds 100 log points. Purchase-level price differentials shrink dramatically from 1967 through the first half of the 1970s, and they continue to shrink through the end of the decade. The gap between mean prices in the top and bottom deciles of the purchase distribution remains large throughout the past four decades, amounting to about 50 log points in 2000.

---

[30] We define part-year observations as those for which the number of production workers in any single quarter is less than 5 percent of the annual average number of production workers. These part-year observations represent less than 2 percent of shipments and electricity purchases in each year.

[31] For example, Santee Cooper tariff schedule TP for temporary service (e.g., ballpark lighting) specifies a flat rate of 7.23¢ per kWh. Schedule GV for Seasonal General Service specifies energy charges of 2.34¢ per kWh and demand charges of $14.35 per kW.
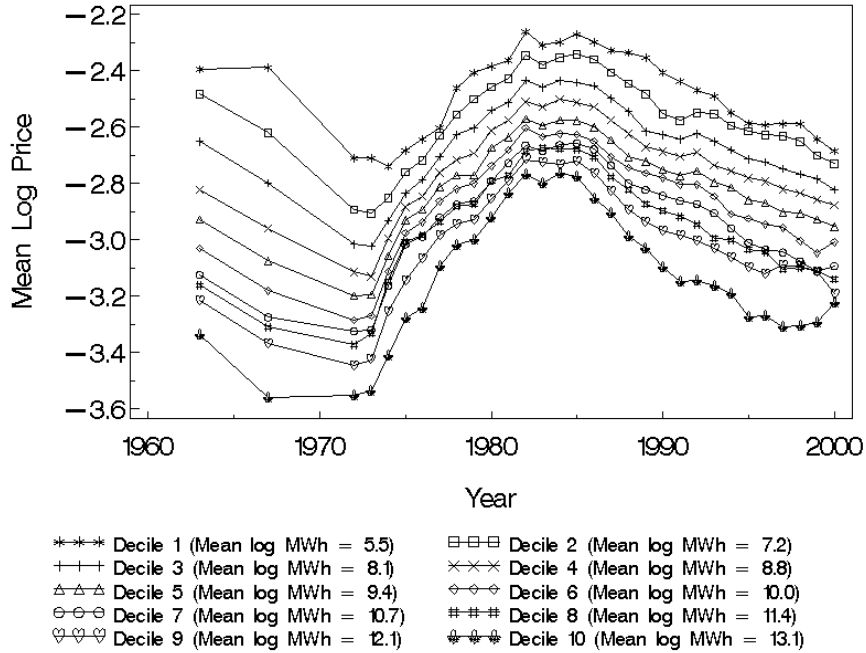
Figure 2.4: Mean of Log Real Electricity Prices by Purchase Deciles, 1963-2000

Figure 2.5 presents a more detailed empirical price-quantity schedule for selected years. It shows the fit from plant-level regressions of log price on a fifth-order polynomial in the log of annual purchases (MWh).[32] We run the regressions separately by year, weighting each observation by its shipments value and ASM sample weight. The regression fits show a dramatic flattening of the price-quantity schedule between 1967 and 1978. According to Figure 2.5, the price differential between the 25th and 75th quantiles of the purchase distribution shrinks from 46 log points in 1967 to 26 log points in 1978, and the gap between the 5th and 95th purchase quantiles shrinks from 103 to 51

---

[32] We also considered nonparametric regression fits for the price-quantity schedule using the SAS GAM procedure (spline option, 100 degrees of freedom). Except at the extreme upper end of the purchase distribution, accounting for less than one percent of shipments, the nonparametric fits are highly similar to the fifth-order polynomial fits. Given this similarity and the much longer run times for the nonparametric fits, especially when we add covariates, we focus on polynomial fits throughout the chapter.

log points.[33] In short, there was a remarkably sharp erosion of quantity discounts between 1967 and the late 1970s. We turn next to potential explanations for these strikingly large quantity discounts and their evolution over time.
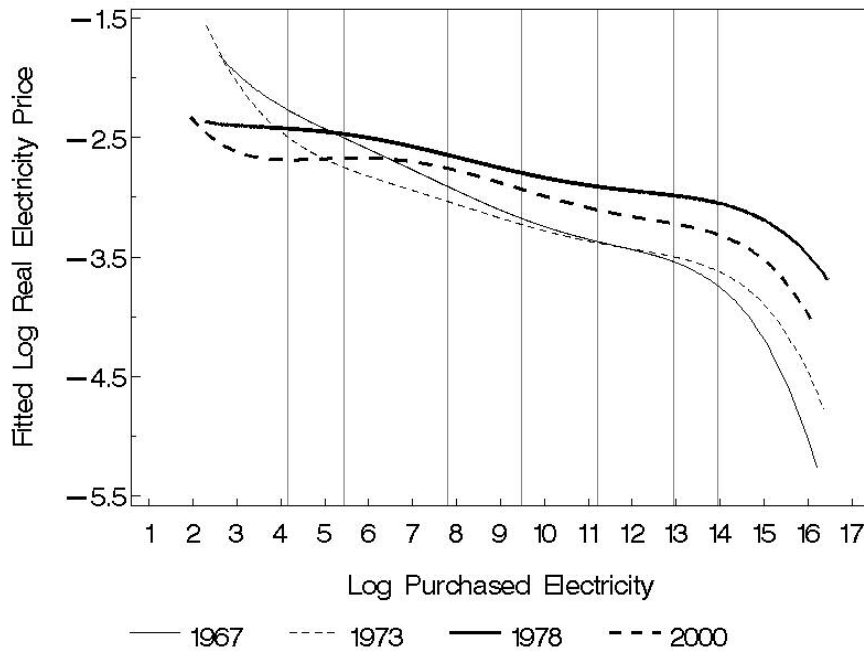
Source: Authors' calculations on PQEM data with part-year observations excluded.

Note: Vertical lines depict the simple average of the 1st, 5th, 25th, 50th, 75th, 95th and 99th percentiles of the shipments-weighted distribution of annual purchases for 1967, 1973, 1978, and 2000.

Figure 2.5: Log Electricity Price Fit to Fifth-Order Polynomials in Log Purchases, Selected Years

---

[33] We also created analogs to Figure 2.5 for the five utilities with the largest number of customer-level observations (several hundred per year). All five utilities show the same basic pattern as in Figure 2.5.

2.6 Behavioral Responses by Customers as a Source of Quantity Discounts

2.6.1 Spatial Sorting of Production Activity

If bigger purchasers locate in areas with cheaper electricity, the pooled data will show a negative relationship between price and purchase level even if all utilities offer flat price-quantity schedules. More generally, any tendency by bigger purchasers to buy from utilities with cheaper electricity contributes to a negative price-quantity relationship. This type of spatial sorting potentially explains much of the pricing structure seen in Figures 2.4 and 2.5. To evaluate this explanation, we fit two plant-level regressions of log price on a fifth-order polynomial in log purchases for each year. One regression specification includes utility fixed effects to control for the identity of the plant's electricity supplier, and the other specification omits utility effects. We then use the fitted regressions to calculate the average elasticity of electricity price with respect to customers' annual purchase levels. To isolate the role of spatial sorting, we compare the elasticity values calculated from regressions with and without utility fixed effects.

Figure 2.6 shows the results. It confirms a dramatic flattening of price-quantity schedules through the late 1970s, and it conveniently summarizes the magnitude of quantity discounts. In the 1960s, the average price-quantity elasticity is -22% on a purchase-weighted basis, and it ranges from -12% to -14% on a shipments-weighted basis. Bigger values for purchase-weighted elasticities reflect the steeper slopes of the price-quantity schedules at the upper end of the purchase distribution (Figure 2.5).
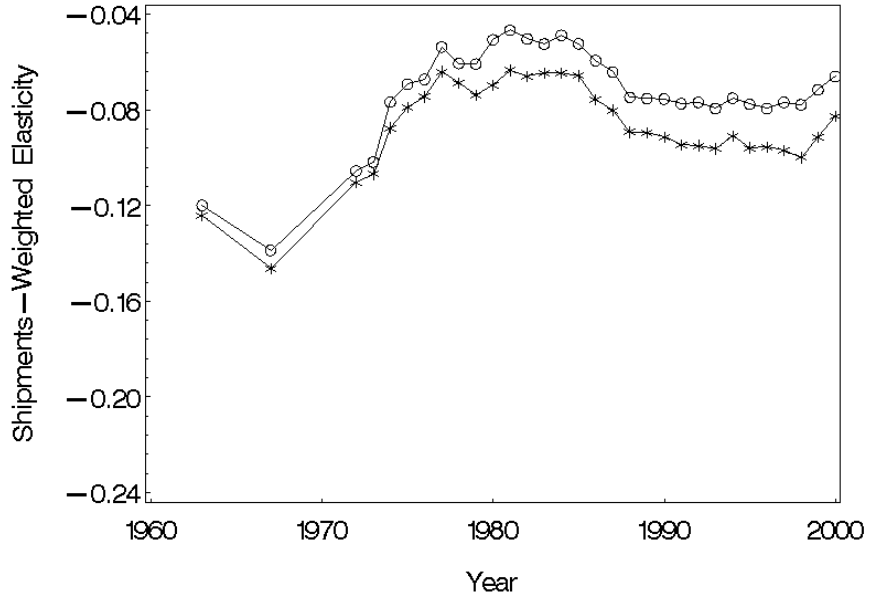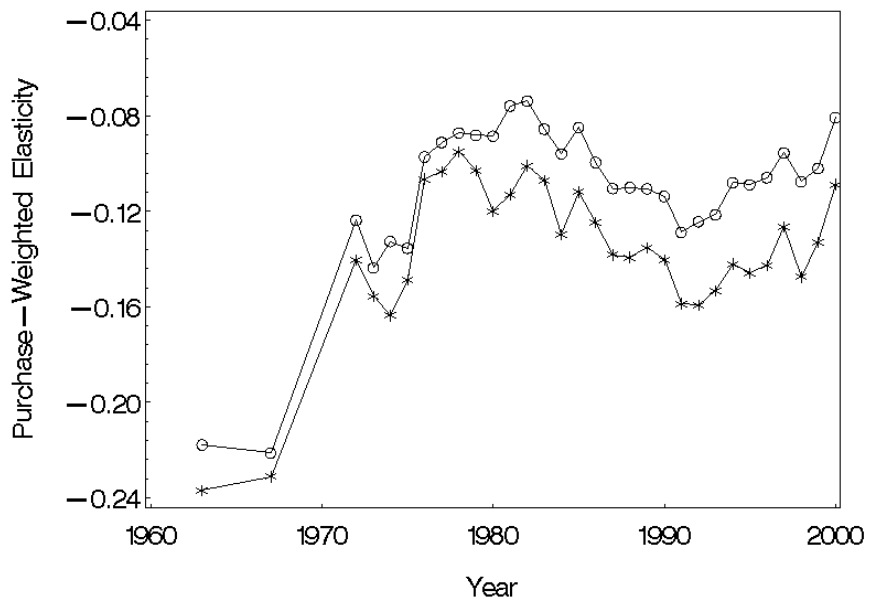
Source: Authors' calculations on PQEM data with part-year observations excluded.

Note: Elasticity values are calculated from shipments-weighted regressions of the log price on a fifth-order polynomial in log purchases.

Figure 2.6: Average Elasticity of Price with Respect to Purchase Quantity, 1963-2000

The inclusion of utility fixed effects has only a modest impact on the elasticity values prior to 1974. That is, in the early part of our sample period the huge purchase-level price differentials in Figures 2.4 and 2.5 reflect within-utility price variation, not spatial sorting of manufacturing customers. Spatial sorting plays a bigger role after 1973, especially on a purchase-weighted basis. Evidently, the onset of rising real electricity prices in 1973 (Figure 2.2) encouraged the migration of electricity-intensive manufacturing activity to areas served by utilities with cheaper electricity. The bigger role for spatial sorting on a purchase-weighted basis suggests that bigger purchasers are more sensitive to spatial price differences in their choice of location.

The evolution of the price-quantity elasticity also provides evidence about the impact of PURPA on a key dimension of electricity pricing. Figure 2.6 shows that the dramatic flattening of price-quantity schedules had already unfolded by 1978, the year of PURPA's enactment. The within-utility elasticity (i.e., controlling for utility fixed effects) fell only slightly in the first few years after 1978 on a shipments-weighted basis, and it actually rose on a purchase-weighted basis. This evidence demonstrates that the contentious rate-reform provisions in PURPA did little to restrain quantity discounts in electricity tariff schedules – at least for manufacturing customers. Instead, PURPA merely ratified rate structure changes that had already occurred.

2.6.2 Other Behavioral Responses to Electricity Tariffs

In addition to location choice, several other behavioral responses by customers influence the empirical price-quantity schedule. Bigger purchasers have greater opportunity and incentive to reduce price per kWh by managing load factors (ratio of average to peak demand), taking high-voltage power, responding to peak-load pricing

structures, and accepting curtailable or interruptible power. To help assess the importance

of these behaviors for the observed quantity discounts, we compare the empirical price-

quantity schedule in the PQEM data to the schedule for "firm" power implied by the

Santee Cooper tariff summarized in Table 2.4. In calculating the implied price-quantity

schedule for firm power, we fix the load factor at 50% and assume no discounts for off-

peak or high-voltage power.[34] These assumptions serve to foreclose quantity discounts

that arise from behavioral responses to pricing incentives and to thereby isolate a pure

customer size effect.[35] In contrast, the empirical price-quantity schedule reflects the pure

size effect *and* the behavioral responses by electricity customers.

Figure 2.7 plots the implied Santee Cooper price-quantity schedule and the

within-utility price-quantity schedule in the 2000 PQEM data. (We do not have enough

customer observations to estimate an empirical price-quantity schedule for Santee Cooper

alone.) As in Figure 2.5, the fitted empirical schedule is based on a fifth-order

polynomial specification. We now include utility fixed effects in the plant-level

regression to isolate within-utility price variation. Figure 2.7 delivers three results. First,

the Santee Cooper and empirical schedules are both rather flat in the lower quartile of the

---

[34] Mechanically, we compute the lower envelope of the price-quantity schedules implied by the General Service, Medium General Service, Large General Service, and Large Power and Light schedules. Recall that the tariff schedules described in Table 2.4 do not include taxes or adjustments specified by the Fuel Adjustment Clause and the Demand Sales Adjustment Clause.

[35] We should note here that the utility likely anticipates responses of certain size customers and builds a discount into the price schedule for these customers. For example, built in quantity discounts for larger purchasers reflect the drop in energy charge for larger purchasers. The customer must commit to maintain a certain level of demand over a long period of time to receive this discount. Since demand is measured as the maximum demand during a given time period, the customer will have strong incentives to smooth demand. Certainty about the level of demand reduces generation costs to the utility since it limits the amount of "extra" power they have to have available (either through deals with other utilities or through maintaining additional, usually less efficient, power generators that can be brought on-line as needed). On the other hand, the built in quantity discounts for relatively small purchasers simply reflect that the fixed administrative and metering costs for each customer are spread out over more kWh of consumption.

purchase distribution, except at the extreme bottom end. Second, over the middle part of

the distribution that roughly spans the interquartile range of purchases by manufacturing

customers, the price per kWh declines with annual purchase quantity by 30 to 40 log

points. Over this range, quantity discounts are essentially "built into" the tariff schedule

according to the evidence in Figure 2.7.[36] Third, the large quantity discounts in the upper

quartile of the distribution reflect behavioral responses to pricing incentives. "Built in"

quantity discounts do not underlie the negative price-quantity relationship in this segment

of the purchase distribution. Instead, the story is one of customer responses to pricing

incentives.
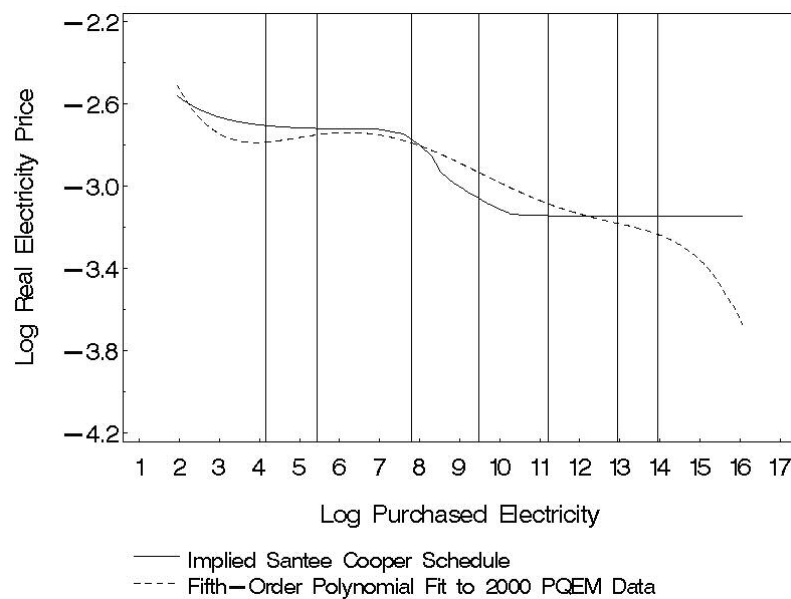
Source: Authors' calculations on PQEM data and Santee Cooper tariff schedules.

Notes: The regression fit on the PQEM data controls for utility fixed effects. Vertical lines depict the simple average of the 1st, 5th, 25th, 50th, 75th, 95th and 99th percentiles of the shipments-weighted distribution of annual purchases for 1967, 1973, 1978, and 2000.

Figure 2.7: Comparison of Empirical and Implied Price-Quantity Schedules, 2000

---

[36] The implied schedule declines more rapidly than the empirical schedule over this range, which indicates that the Santee Cooper tariff menu involves bigger "built in" quantity discounts than the average utility.

2.6.3 Summary

This section establishes that the negative price-quantity relation evident in Figures 2.4 and 2.5 reflects a combination of customer responses to pricing incentives and mechanical discounts built into electricity tariff schedules. Both aspects are important, but they are relevant for different segments of the purchase distribution. Mechanical discounts are important in the middle of the distribution, and behavioral responses to pricing incentives are important in the upper quartile. Both the responses to pricing incentives reflected in Figure 2.7 and the spatial sorting response documented in Figure 2.6 are concentrated among larger purchasers. This evidence reinforces the view – often expressed in the public utility and Ramsey-pricing literatures – that demand is more price elastic at higher purchase levels.

2.7   Customer Purchase Quantity and Electricity Supply Costs

2.7.1   A Method for Estimating Supply Costs as a Function of Purchase Amount

As discussed above, customer cost and demand characteristics can both lead to quantity discounts. Figure 2.7 implies that cost characteristics play an important role, because discounts in the upper quartile of the purchase distribution reflect behaviors that reduce supply costs. Insofar as bigger customers have higher load factors, higher voltage levels, closer proximity to transmission lines, their own transformers, greater response to peak-load pricing incentives, and greater willingness to accept power curtailments and interruptions, they are cheaper to supply. The negative "mechanical" relationship between price and quantity embodied in the implied Santee Cooper schedule may also reflect lower supply costs for bigger purchasers. Thus, in line with prior views, the

evidence in Section 2.6 implies that electricity supply costs decline with purchase quantity.

To more precisely evaluate the role of supply costs in quantity discounts, we now develop a method for estimating the supply schedules that exploits the cross-sectional richness of the PQEM. To the best of our knowledge, our method offers a novel approach to estimating supply cost schedules as a function of customer size. The method involves three main steps. Step one uses customer-level data on purchase quantities to calculate utility-level statistics for the location and shape of the purchase distribution. Step two exploits the utility's revenue constraint, which states that average cost per kWh equals average price per kWh.[37] Step three exploits cross-utility variation in the purchase distribution to estimate how costs per kWh of delivered electricity vary with customers' annual purchases. We carry out step three using regression methods to control for other factors that affect supply costs. We now develop the method in detail.

A portion of a utility's costs are common to all customers, and the remaining portion can be allocated to particular customers. Let θ_g be the common cost per kWh at utility g.[38] Write the allocable portion of costs per kWh for customer e that purchases q_e as $C_g(q_e) + k_e$, where the first term captures cost differences that vary systematically by purchase level and the second term captures idiosyncratic supply cost differences unrelated to purchase level. By construction, $\sum s_e k_e = 0,$ where s_e is the share of

---

[37] Recall, the utility's revenue constraint is consistent with Ramsey pricing. In fact, meeting the revenue constraint is a requirement for Ramsey pricing (see Section 2.5.1).

[38] Note that utility common costs include the return on equity they pay their investors.

purchases from utility g by plant e. Thus, letting TC denote total cost, we can write the average cost per kWh at utility g as shown in (2.3).

$$AC_g \equiv \frac{TC_g}{\sum_{e \in g} q_e} = \theta_g + \sum_{e \in g} s_e C_g(q_e) \tag{2.3}$$

The revenue constraint implies that a utility's average cost per kWh equals its average price per kWh. Imposing this requirement in (2.3) yields

$$P_g = \theta_g + \sum_{e \in g} s_e C_g(q_e) + v_g^P \tag{2.4}$$

where P_g is the purchase-weighted mean price per kWh at utility g, and v_g^P is an error term introduced by sampling variation in P_g. We do not directly observe the utility's average price per kWh in the PQEM, but we can estimate it using price and quantity observations on the utility's manufacturing customers.

To obtain an estimable specification from (2.4), we adopt three assumptions. First, we postulate that the $C_g(q)$ functions are the same for all g up to an additive term; i.e., $C_g(q) = C(q) + \alpha_g$. Second, we approximate $C(q)$ as a polynomial in log (q). Third, we model the sum of the utility's additive and common cost components as a linear function of observable utility characteristics X; namely, $\alpha_g + \theta_g = X_g b + u_g$ Applying these assumptions to (2.4) yields an estimating equation with four error components:

$$P_g = X_g b + \sum_{n=1}^{N} \gamma_n \sum_{e \in g} s_e \left[\log(q_e) \right]^n + u_g + v_g^P + v_g^q + \xi_g \tag{2.5}$$

where N is the order of approximation to the C function, $\sum_{e \in g} s_e \left[\log(q_e) \right]^n$ is the nth uncentered sample moment of the log purchase distribution at g, and the γ's are the key parameters of interest for the supply cost schedule. The error component v_g^q arises from sampling variation in the moments of the purchase distribution, and ξ_g arises from the polynomial approximation to C. Though not our main focus, the b parameters are also interesting, because they provide estimates of how average costs vary with utility characteristics when we control for the size distribution of customer purchases.

We estimate (2.5) by weighted ordinary least squares (WLS) and instrumental variables (IV) regression. We then use the γ estimates to trace out the supply cost schedule as a function of customer purchase quantity. Before turning to the results, three econometric issues require some discussion.

First, consider the error term u_g in (2.5) that arises from unobserved determinants of the additive and common costs. If these unobserved cost determinants vary

44

systematically with the size distribution of customer purchases, they give rise to an omitted variables problem that biases the estimates of γ. As a case in point, municipal and cooperatively owned utilities tend to serve smaller manufacturing customers.[39] If these same utilities also have lower supply costs conditional on customer size, then the regression (2.5) understates the extent to which costs per kWh decline with purchase amount, unless we control for utility type. Hence, we include the utility's organizational form in the X vector, distinguishing among cooperative and municipal utilities, state and federal power authorities, and private investor-owned utilities. For similar reasons, we include controls for the size of the utility and for the shares of electrical power generated from hydro, nuclear, coal, and petroleum and natural gas. A potential omitted variables problem also arises in connection with non-sampling components of the error term v_g^P in (2.4) and (2.5). In particular, the revenue constraint might fail for manufacturing customers as a group because of cross-subsidization between classes of customers within the utility. To control for this possibility, we include in the X vector the fraction of the utility's revenues derived from sales to industrial customers.

Second, the error term v_g^q that arises from sampling variation in the moments of the purchase distribution creates a standard errors-in-variables problem. To address this potential source of bias, we exploit the fact that consecutive ASM panels are independently drawn from the universe of manufacturing plants. It follows that the sampling error in the purchase distribution statistics for utility g at time t is uncorrelated

---

[39] We have examined the distribution of mean log purchases by manufacturing customers for private investor owned utilities and the analogous distribution for municipal and cooperatively owned utilities. A comparison of these distributions confirms that average customer size tends to be considerably smaller at municipal and cooperatively owned utilities.

with the sampling error at $t+k$, provided that a new ASM panel has commenced

between t and $t+k$. Thus, we instrument the moments of the utility's log (q) distribution

with the corresponding statistics for the same utility calculated from a nearby year that

draws on a different ASM sample.[40]

Third, the number of annual customer-level observations per utility in the PQEM

ranges widely from a small handful to hundreds. Hence, the sampling error components

in (2.5) have a heteroscedastic structure. To improve the efficiency of our estimates, we

weight each observation in the regression (2.5) by the square root of the number of

manufacturing plants used to calculate the utility-level quantities. As a side benefit, this

weighting method mitigates the errors-in-variable problem under least squares.

---

[40] For k=1, we can construct instruments across ASM panels for 12 years. For k=5, we can construct
instruments across ASM panels for all years. We tried both approaches.

2.7.2 Supply Cost Estimation Results

Table 2.5 reports the WLS regressions of the form (2.5) on the utility-level data. We approximate the supply cost schedule $C(q)$ as a third-order polynomial in $\log(q)$. We normalize the purchase-weighted mean price per kWh to 100 in each year, so that slope coefficients on the indicator variables reflect percentage differences from the omitted category. We report results for selected years to economize on space, but our discussion below draws on results for all years.

Municipal and cooperative utilities have lower estimated supply costs in the 1960s and early 1970s, after controlling for other factors, and the cost advantage over private investor-owned utilities re-emerges in the 1990s. Relative to coal-powered electricity generation, greater reliance on nuclear power yields higher supply costs; hydro power yields lower supply costs until the 1990s; and petroleum and natural gas yield higher supply costs after the 1970s. The estimated effects of power source are sizable. For example, the 1967 estimates imply that shifting 10% of power generation from coal to hydro involves a 3.3% reduction in cost per kWh. The estimates also imply that bigger utilities have lower supply costs in the 1960s, but the effects are small.

Table 2.5: Regression Results for Electricity Supply Costs, Selected Years

Dependent Variable: Purchase-weighted mean price per kWh for the utility's manufacturing customers

	1967	1973	1978	2000
Public Ownership	37**	24**	15	18
	(9)	(9)	(14)	(12)
Private Ownership	40**	22**	13**	9*
	(4)	(3)	(4)	(4)
Fraction of Utility Total Revenue from Industrial Customers	-3	-30**	-37**	-7
	(9)	(10)	(12)	(11)
Share of Power From Hydro	-33**	-45**	-55**	15
	(5)	(5)	(6)	(8)
Share of Power From Nuclear	372**	48**	12	46**
	(83)	(13)	(7)	(8)
Share of Power From Oil and Natural Gas	-4	-5	8	39**
	(3)	(4)	(6)	(8)
Adjusted R-Square	0.80	0.69	0.63	0.66
Test: Utility Size Measures = 0	0.00	0.88	0.34	0.62
Test: Customer Size Measures = 0	0.00	0.00	0.00	0.00
Test: Ownership Measures = 0	0.00	0.00	0.00	0.05
N	227	235	242	235

* $p<0.05$, ** $p<0.01$

Notes:
1. Regressions are on utility-level data by weighted least squares. Weights are proportional to the square root of the number of customer observations used to calculate the utility-level statistics. The sample is limited to utilities for which there are at least 8 customer-level observations. The dependent variable is normalized so that the purchase-weighted mean price over utilities equals 100.
2. In addition to the variables shown in the table, the regression also includes the first three uncentered moments of the utility's log customer size distribution and a quadratic polynomial in the log of the utility's electricity sales to industrial customers.
3. The ownership variables and the fraction of total revenue from industrial customers are from the 2000 EIA 861 file. Public and private ownership variables are dummy variables, and the omitted category is cooperative and municipal ownership. Fuel share variables are state-level data from the State Energy Data 2000 files. Both coal and "other" (includes geothermal, wind, wood and waste, photovoltaic, and solar) are omitted since "other" is always very small. Moments of the customer size distribution are constructed from the PQEM.

Source: Authors' calculations on data from the PQEM, EIA 861 files, and State Energy Data 2000.

Turning to our main focus, the moments of the customer purchase distribution are jointly significant at the 0.1% level in all years, strongly confirming the statistical significance of purchase quantity as a determinant of supply costs. Table 2.6 and Figure 2.8 report the estimated supply cost schedules. Figure 2.8 also shows, for each utility g, the coordinates of the mean log purchases of its customers, $\overline{lq_{1g}}$, and the sum of the implied cost value $\tilde{P}_g = \left(\overline{X_g \hat{b}}\right) + \hat{\gamma}_1 \overline{lq_{1g}} + \hat{\gamma}_2 \left(\overline{lq_{1g}}\right)^2 + \hat{\gamma}_3 \left(\overline{lq_{1g}}\right)^3$ and the utility's regression residual. As seen in Figure 2.8 and Table 2.6, supply costs per kWh fall by a factor of 2 or 3 over the range of purchases spanned by the utilities in our sample. This pattern holds in all years from 1963 to 2000. In unreported results, we re-estimated the supply cost regressions by IV using the approach described above, and obtained essentially the same findings. These results provide strong evidence of powerful, cost-based reasons for large quantity discounts in electricity pricing to industrial customers.

Table 2.6: Estimated Electricity Supply-Cost Schedules as a Function of Customer
Purchase Quantity, Selected Years

Annual Purchase Amount (GWh)	Percentile of Purchases Distribution	Supply Cost per kWh in 1996 Cents			
		1967	1973	1978	2000
0.53	10	8.25	6.47	9.96	10.89
2.43	25	5.75	4.90	7.58	7.65
13.1	50	4.20	3.81	6.07	5.45
73.9	75	3.42	3.19	5.33	4.29
229	90	3.09	2.94	5.05	3.91
422	95	2.90	2.83	4.92	3.78
1,130	99	2.50	2.65	4.66	3.59

Notes:
1. The supply-cost schedules are derived from the regressions reported in Table 2.5 and described in Section 2.7.1. The schedules are evaluated at sample mean values of the other regression covariates.
2. The percentiles of the purchases distribution are the simple average of the percentiles of the shipments-weighted purchase distribution in 1967, 1973, 1978, and 2000.
3. We do not report supply costs for the bottom tail of the purchases distribution, because small purchase values are outside the range we used to fit the utility-level regressions in Table 2.5.

Source: Authors' calculations on PQEM data.
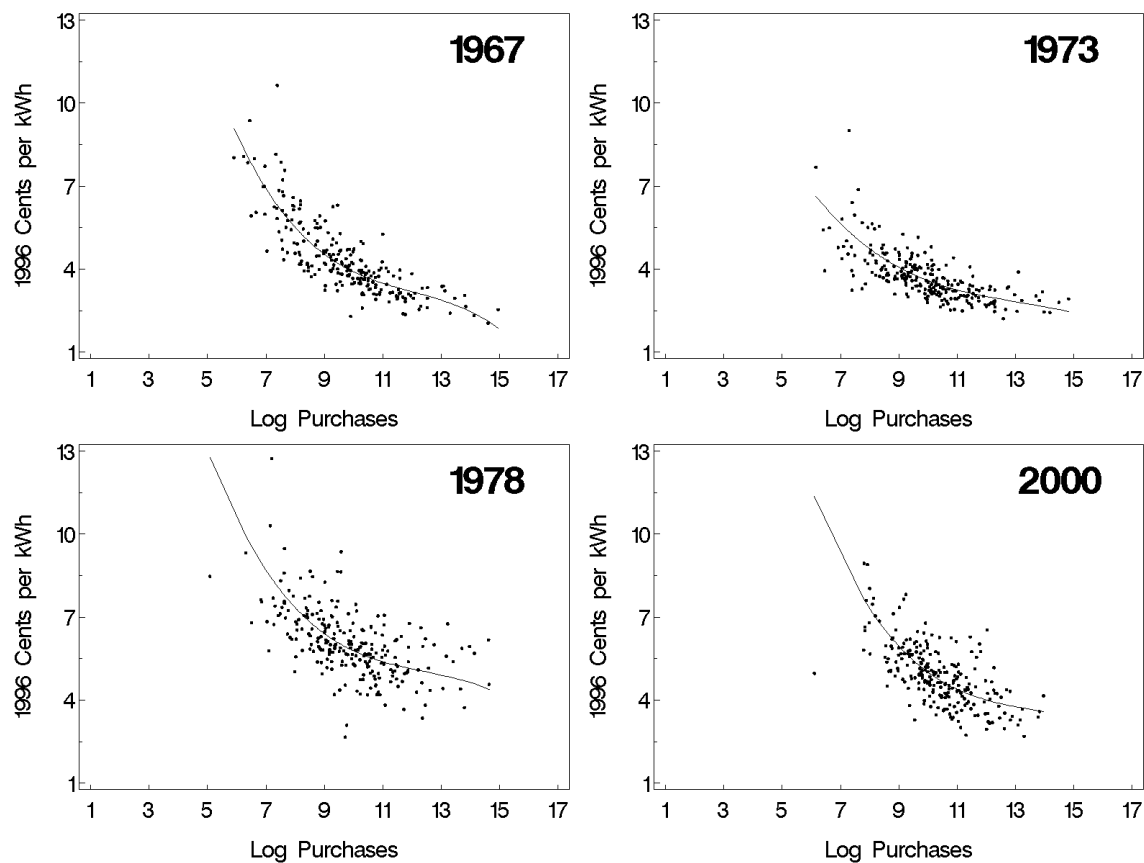
Source: Authors' calculations on PQEM data with part-year observations excluded.

Notes: Each curve shows the fitted relationship between supply costs per kWh and annual customer purchases, evaluated at sample means of other covariates in the regression. The vertical coordinate for each plotted point is the sum of the fitted supply cost and the regression residual for a particular utility in the sample, as described in the text.

Figure 2.8: Electricity Supply Costs per kWh as a Function of Annual Customer Purchase Level, Selected Years

We also computed the average supply-cost elasticity with respect to customer purchase quantity for each year and compared it to the average price elasticity with respect to purchase quantity (Figure 2.6). The comparison yields two interesting results. First, the average cost elasticity is consistently somewhat larger in magnitude than the average price elasticity, indicating that supply costs fall more rapidly with purchase quantity than price per kWh. Second, longer term swings in the average cost elasticity closely mirror the swings in the average price elasticity in Figure 2.6. This time-series pattern reinforces the inference derived from the cross-sectional evidence that cost factors drive large quantity discounts in electricity pricing.

2.8   Evaluating the Pricing Structure

2.8.1   Is Pricing Efficient on the Purchase Quantity Margin?

Pricing efficiency requires that marginal prices for successive increments of electrical power equal marginal supply costs at all points on the customer purchase distribution. This is a demanding requirement in our context, because the range of purchases is enormous. We now test whether this condition holds in the data. Earlier empirical studies also consider retail pricing efficiency in the electric power industry. Examples include Meyer and Leland (1980), Hayashi, Sevier and Trapani (1985) and Nelson, Roberts and Trump (1987). However, these studies evaluate pricing differences across classes of customers – residential, industrial and commercial – from the vantage point of efficiency, Ramsey pricing, and rate of return regulation. They do not consider pricing efficiency on the purchase quantity margin. Indeed, no previous empirical study evaluates efficiency or Ramsey pricing on this margin, but the issue receives much

attention in theoretical works.[41] See Brown and Sibley (1986) and Wilson (1993) and references therein. Our empirical assessment of pricing on the purchase quantity margin complements the well-developed theoretical literature on the topic.

For purposes of comparing the marginal curves, we first re-estimate the price-quantity schedules by regressing price per unit (not logged) on a third-order polynomial in log customer purchases. We include utility fixed effects to isolate within-utility price variation. In re-estimating the price schedules, we omit plants with annual purchases outside the range of mean log purchases in the utility-level data. These modifications to the specification and samples used in Sections 2.5 and 2.6 provide for an apples-to-apples comparison of the marginal curves. Given a fitted price-quantity schedule, it is easy to calculate the corresponding marginal price schedule. Let $T(q) = qP(q)$ be the total electricity tariff paid by a plant that purchases q kWh, where $P(q)$ is the average price per kWh. We compute the marginal price schedule as

$$\hat{M}(q) = \hat{P}(q) + (q/\varepsilon)\left[\hat{P}\left(q + (\varepsilon/2)\right) - \hat{P}\left(q - (\varepsilon/2)\right)\right] \qquad (2.6)$$

where $\hat{P}(q)$ is the fitted value of the price-quantity schedule at q, and ε is a small positive number. We follow the same approach in calculating marginal cost schedules from estimated supply cost schedules of the type displayed in Figure 2.8.

Recall that each ASM panel is an independently drawn random sample. To exploit this sample design feature, we pool customer-level observations over year-pairs that straddle ASM panel changeovers prior to constructing the utility-level data. This pooling method yields more customer-level observations per utility and a larger number

---

[41] Peltzman (1971) considers electricity pricing on the purchase quantity margin, but he lacks the cost data needed for an assessment of pricing efficiency on this margin.

of usable utility-level observations, thereby improving estimation efficiency in the supply

cost regressions. We estimate these regressions using the same specification and

weighted least squares method as before except for the addition of a year control.

Figure 2.9 displays the marginal schedules for selected years, along with

bootstrapped standard error bands for the marginal cost schedules. (Standard errors for

marginal prices are extremely small, and we abstract from them in the discussion that

follows.) The marginal schedules are remarkably similar in 1967 and 1973/74, strongly

confirming the central implication of pricing efficiency on the purchase quantity margin.

After 1973/74, however, a gap opens between marginal cost and marginal price in the

lower deciles of the purchases distribution. The gap is sizable, with marginal cost

exceeding marginal price by 10% or more for smaller purchasers.

Figure 2.10 shows the evolution of pricing efficiency over time. Outcomes at the

20th, 50th and 80th percentiles accurately reflect outcomes more generally in the lower,

middle and upper portions of the purchase quantity distribution. Except for 1963, Figure

2.10 provides essentially no indication of departures from pricing efficiency in the middle

or upper portions of the purchase distribution in the four decades covered by our sample.

That is, the data strongly favor the hypothesis of pricing efficiency on the quantity

margin for mid-sized and large industrial customers. The story for small purchasers is

different. As Figure 2.10 shows, marginal prices fall short of marginal costs at the 20th

percentile after 1973/74, and the gap continues to widen though 1983/84.
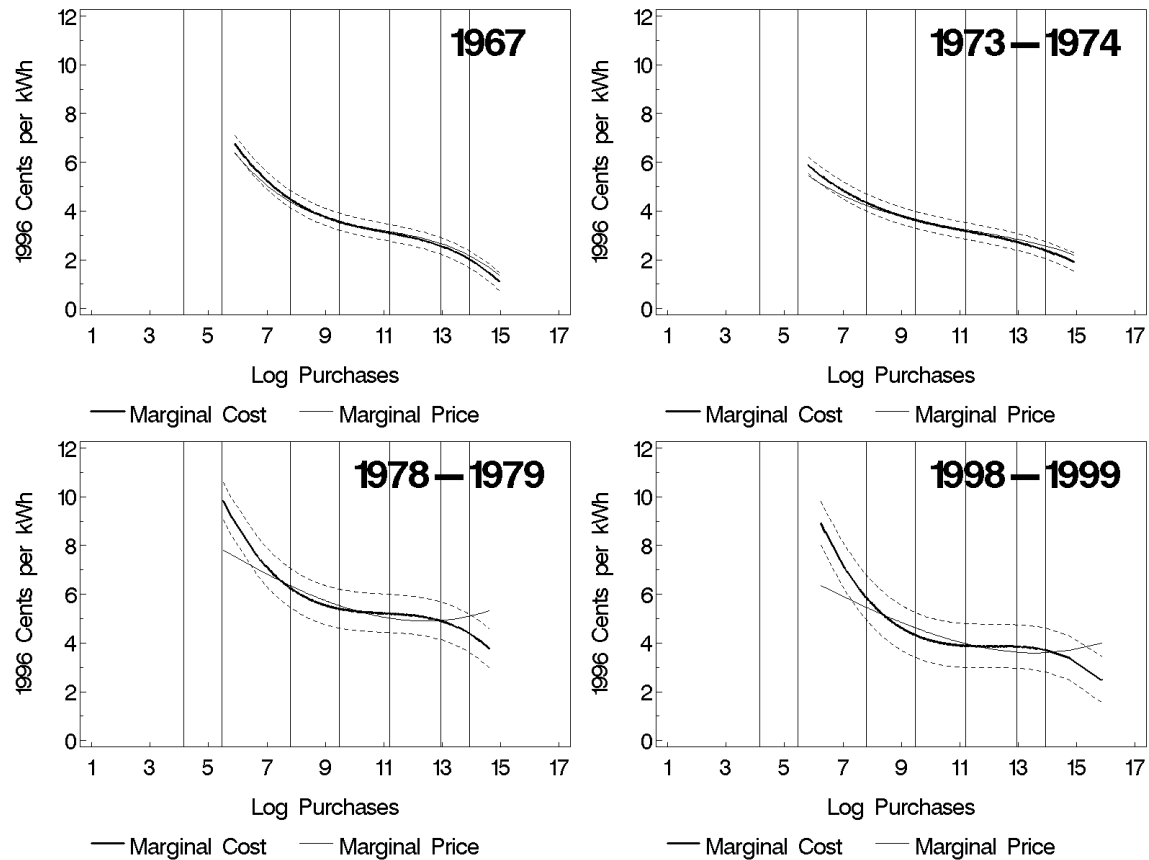
Source: Authors' calculations on PQEM data with part-year observations excluded.

Note: Vertical lines depict the simple average of the 1st, 5th, 25th, 50th, 75th, 95th and 99th percentiles of the shipments-weighted distribution of annual purchases for 1967, 1973, 1978, and 2000. Dashed curves show bootstrapped standard error bands.

Figure 2.9: Marginal Cost and Marginal Price Schedules Compared, Selected Years
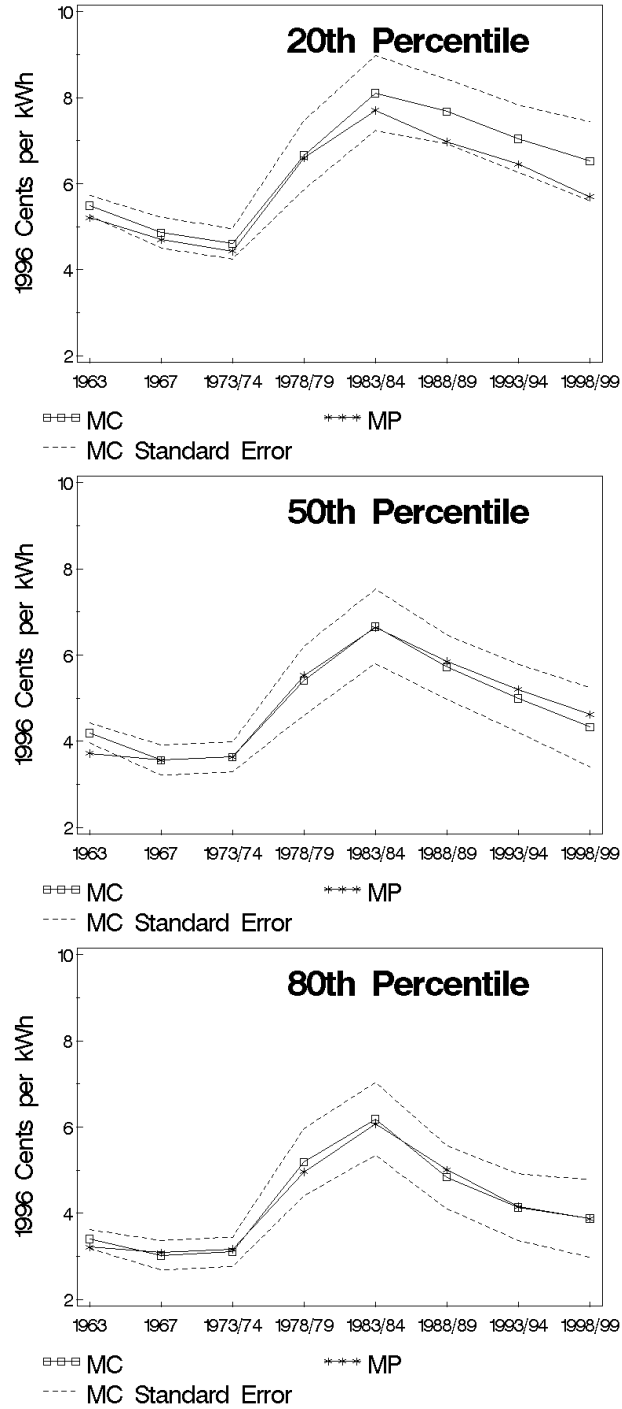
Source: Authors' calculations on PQEM data with part-year observations excluded.

Note: Annual purchases are 7.4 log MWh at the 20th percentile, 9.5 log MWh at the 50th, and 11.5 log MWh at the 80th.

Figure 2.10: Marginal Cost and Marginal Price, Selected Percentiles of the Purchases Distribution

To construct a more powerful formal test for the null hypothesis of pricing efficiency, we now pool the data over several years. We evaluate pricing efficiency in the "early years" 1963, 1967, 1973 and 1974 and the "recent years" 1988, 1993, 1998 and 1999. The early years predate the departures from pricing efficiency suggested by 2.9 and 2.10, and the late years postdate them. We selected these particular years because they involve eight independently drawn random samples of manufacturing plants. In pooling the data over years, we introduce year controls that allow for marginal costs to shift over time in a manner that is uniform with respect to purchase quantity.

Table 2.7 reports the pooled-sample estimates and bootstrapped standard errors for early and late years. The upper panel extends our previous pooling method for calculating utility-level statistics from pooled customer-level observations. This method results in many customer-level observations per utility but only one observation per utility in the supply cost regression. This first method exploits only between-utility variation to estimate the cost schedules. The lower panel calculates utility-level statistics from customer-level data first and then pools over years. This method results in fewer customer-level observations per utility but up to four observations per utility in the supply cost regression. Under this method, we assume that utility-level error terms in the supply cost regression are uncorrelated over time. This second method exploits between- and within-utility variation to estimate the cost schedules.

Table 2.7: Tests of Pricing Efficiency with Alternative Pooling Methods

A. Customer-Level Data Pooled over Years before Calculating Utility-Level Statistics

	Marginal Price (1996 ¢ / kWh)	Marginal Cost (1996 ¢ / kWh)	Standard Error of Marginal Cost (1996 ¢ / kWh)	Difference: MP - MC (1996 ¢ / kWh)
1963, 1967, 1973, 1974	***N* = 292 in the Supply Cost Regression**			
20th Percentile	4.57	4.81	0.24	-0.24
50th	3.67	3.70	0.23	-0.03
80th	3.25	3.10	0.23	0.15
1988, 1993, 1998, 1999	***N* = 305 in the Supply Cost Regression**			
20th Percentile	6.26	7.24	0.71	-0.98
50th	5.07	5.05	0.71	0.02
80th	4.20	3.90	0.69	0.30

B. Utility-Level Statistics Calculated from Customer-Level Data before Pooling

	Marginal Price (1996 ¢ / kWh)	Marginal Cost (1996 ¢ / kWh)	Standard Error of Marginal Cost (1996 ¢ / kWh)	Difference: MP - MC (1996 ¢ / kWh)
1963, 1967, 1973, 1974	***N* = 902 in the Supply Cost Regression**			
20th Percentile	4.67	4.92	0.11	-0.25
50th	3.68	3.69	0.11	-0.01
80th	3.18	3.11	0.10	0.07
1988, 1993, 1998, 1999	***N* = 961 in the Supply Cost Regression**			
20th Percentile	6.30	7.04	0.28	-0.74
50th	5.12	4.90	0.28	0.22
80th	4.29	4.24	0.28	0.05

Note: See text for a description of the underlying specifications and estimation methods.

Source: Authors' calculations on PQEM data.

The two pooling methods yield a similar pattern of point estimates that shows sizable departures from pricing efficiency in the later years for smaller customers. Marginal prices are more than 10% below marginal costs at the 20th percentile of the purchase quantity distribution in the later years, but the gap is statistically significant only under the second pooling method. This method also yields evidence against marginal cost pricing for smaller customers in the early years, but the deviation from pricing efficiency is much smaller, amounting to less than 5% of marginal cost. In line with 2.9 and 2.10, Table 2.7 yields no evidence of departures from pricing efficiency in the middle and upper portions of the purchase quantity distribution. These results hint at cost shifting from smaller to larger industrial customers. In future work, we plan to run an empirical test for cost shifting among industrial customers.

What caused the departure from pricing efficiency for smaller customers after the mid 1970s? An answer to that question is beyond the scope of this chapter, but we suggest two avenues for future investigation. First, sizable deviations from marginal cost pricing began to emerge at the same time as real electricity prices began to rise (Figure 2.2). As discussed in Section 2.2, the rise in real electricity prices from 1973 to 1983 reversed a decades-long trend. Perhaps utility companies or their regulators deliberately sought to insulate smaller industrial customers from the full impact of rising energy costs. Small industrial customers likely purchase electricity from the commercial schedule. As seen in Figure 2.2, residential and commercial electricity prices are similar in magnitude. Hence, populous pressure on utilities to ease residential prices may have created a cap for prices paid by small industrial customers. A difficulty with this story is its failure to

explain the persistence of deviations from marginal cost pricing after real electricity prices resumed a downward trend.

Second, during the 1970s public utility commissions began to focus greater effort on the review and design of electricity tariff schedules, as discussed by Cudahy and Malko (1976) in their treatment of the landmark Madison Gas and Electric case. The Madison case, initiated in 1972, stimulated similar reviews in other states. "By 1977, 12 state commissions had held generic hearings on retail electric rate structure reform." (Joskow, 1979, page 794). Ironically, these moves toward more aggressive intervention in rate design were often presented as efforts to implement marginal-cost pricing principles. Our evidence shows that greater involvement in the review and design of rate structures by public utility commissions coincided with significant steps away from efficient pricing on the margin we measure. A careful study of whether intervention by public utility commissions caused the deviations from efficient pricing merits investigation.

2.8.2   Is There Any Role for Ramsey Pricing?

Our results provide no support for the standard Ramsey-pricing explanation of quantity discounts. According to this explanation, the markup of marginal price over marginal cost is positive, and it declines with the elasticity of demand. By all accounts, and consistent with our evidence in Section 2.6, demand is more price sensitive in the upper segments of the purchases distribution. Hence, the standard Ramsey-pricing perspective predicts that marginal price exceeds marginal cost, and that the markup shrinks with purchase level. The pattern we have seen is more nearly the opposite.

That the data do not conform perfectly to Ramsey pricing is no surprise. However, we are struck by the utter failure of the standard Ramsey-pricing view to account for *any* portion of the large quantity discounts in electricity pricing. Evidently, cost differences and not markup differences are the predominant reason for quantity discounts. When the pricing structure deviates from efficiency, it does so in the opposite direction from the prediction of the standard Ramsey-pricing view.

It is worth remarking, however, that the data might be reconciled with Ramsey pricing under the unusual premise that marginal cost pricing raises too *much* revenue; i.e., that efficient pricing raises more revenue than required to cover costs and a normal return on equity. In this circumstance, Ramsey-pricing logic implies that the second-best pricing structure involves bigger *markdowns* of marginal prices relative to marginal costs in the less elastic portion of the purchases distribution. That is essentially the marginal pricing structure that emerges after 1973. The premise that yields this rationalization is greatly at odds with the traditional view that electric utilities operate with declining costs. However, it resonates with evidence that changes in the regulatory environment over the course of the 1970s led to tighter capacity constraints and higher costs of expanding capacity. In any event, the role of tighter constraints on capacity in the move away from efficient pricing is another topic worthy of future investigation.

2.9   Concluding Remarks

Let us summarize our main empirical findings:

1. There is tremendous dispersion among manufacturers in the prices they pay per kWh of electricity. The purchase-weighted standard deviation of log electricity prices exceeds 40% in the cross section.

2. The log price distribution underwent a great compression from 1967 to the late 1970s because of a dramatic flattening of price-quantity schedules. During this period, the average elasticity of price per kWh with respect to annual purchase quantity shrank from -22% to -9%.

3. Supply costs per kWh decline by more than half over the range of customer purchases in the cross section. This finding and other results provide evidence of a powerful cost-based rationale for large quantity discounts in electricity pricing.

4. Among smaller and mid-sized manufacturing customers, quantity discounts are built into electricity tariff schedules in a mechanical way. Among customers in the upper quartile of the purchases distribution, even deeper quantity discounts arise from behavioral responses to pricing incentives.

5. Prior to the mid 1970s, marginal costs and marginal prices with respect to customer purchase quantities are nearly identical, remarkably in line with efficient pricing. After the mid 1970s, a gap opens between marginal prices and marginal costs for smaller purchasers. In the 1980s and 1990s, the marginal cost of incremental purchases for these customers exceeds the marginal price by more than 10%.

6. The data provide no support for the Ramsey-pricing view that quantity discounts reflect smaller markups over marginal cost for more price-sensitive customers.

7. Spatial price dispersion declined sharply from the late 1960s to the late 1980s for the largest purchasers, but it rose over time in the lower half of the purchases distribution. The expansion of wholesale power markets in the 1990s had no apparent impact on spatial price dispersion at the retail level for manufacturing customers.

These findings considerably expand our knowledge of electricity pricing to industrial customers, especially in connection with the magnitude and sources of quantity discounts. They strengthen the empirical basis for theorizing about public utility pricing, for evaluating the impact of regulatory expansion in the 1970s, and for assessing the effects of growth in wholesale markets on spatial price dispersion at the retail level.

Our study also points to several questions for future research: Why has the rapid expansion of wholesale markets in the 1990s had such a limited effect on spatial price dispersion at the retail level (Figure 2.3)? Why did price-quantity schedules flatten so sharply between 1967 and 1977 (Figures 2.5 and 2.6)? Why did electricity pricing to manufacturing customers deviate from an efficient structure after mid 1970s, and why are the deviations limited to smaller purchasers (Figures 2.9 and 2.10 and Table 2.7)? How big are the costs of these departures from pricing efficiency? These questions can be fruitfully attacked with the help of the PQEM database developed in the course of our study.

Price dispersion for electricity and other inputs also has potentially important implications for the study and interpretation of productivity differences. Productivity dispersion in the cross section is "extremely large" (Bartlesman and Doms, 2000). Lacking data on input prices, productivity studies typically rely on input expenditures in place of input quantities. In this regard, we stress that Tables 2.2 and 2.3 report large input price differences within narrowly defined industries. We have traced these price differences to cost-based quantity discounts. Casual empiricism suggests that quantity discounts are prevalent for many other intermediate inputs including office supplies, computer software, legal services, information goods, and airline travel. Systematic

evidence supports this impression. In a field study of 39 manufacturing and service firms, Munson and Rosenblatt (1998) find that 83% receive quantity discounts for *most* of the items they purchase. If larger businesses are better positioned to exploit quantity discounts, then most previous studies overstate the relative physical productivity of bigger producers. These observations suggest that input price variation among producers merits greater attention in future research on productivity differences.

Chapter 3

Energy Type Substitution in U.S. Manufacturing: A Cross-sectional Study

Using the 1998 Manufacturing Energy Consumption Survey

3.1   Introduction

Recent skyrocketing fuel prices have brought renewed interest in the effect of changes in energy prices on the U.S. economy. Davis and Haltiwanger (2001) find that oil price shocks account for between 20 and 25 percent of the cyclical variation in employment growth in U.S. manufacturing. In this chapter, I examine energy type consumption patterns, prices, and substitution in U.S. manufacturing plants. Despite a recent decline, manufacturing is still an important part of the U.S. economy, and energy is a vital input into manufacturing. In fact, every manufacturing plant uses at least one type of energy in their production process.

I first examine energy type consumption patterns of manufacturing plants and dispersion in prices paid by plants for four energy types: electricity, oil, natural gas and coal.[42] I then estimate price elasticities for these four energy types and three non-energy inputs (capital, labor, and materials). For my estimation, I use cross-sectional plant-level energy type consumption and expenditures data from the 1998 Manufacturing Energy Consumption Survey (MECS) along with data on non-energy inputs and plant

---

[42] Oil is defined as the aggregate of three MECS energy types: residual fuel oil, diesel and distillate fuel oil, and liquefied petroleum gases (LPG) / natural gas liquids (NGL). For detailed definitions of these energy types, see the glossary on pages 377-396 of EIA, 2005b.

characteristics from the 1998 Annual Survey of Manufactures (ASM), the Longitudinal

Business Database (LBD) and the Chiang-Haltiwanger dataset.[43]

Energy types differ in cost, environmental cleanliness, and ease of use for a given

function. As shown in Table 3.1, electricity is considerably more expensive per Btu of

raw energy output than other energy types. In terms of raw energy, electricity is more

than four times more costly than natural gas and more than seven times more costly than

coal. However, for many purposes electricity is easier and environmentally cleaner to use

than other energy types.[44] In fact, despite the relatively high cost, nearly all

manufacturing plants purchase at least some electricity. In 1998, 99.8% of U.S.

manufacturing plants purchased electricity. While electricity use is pervasive throughout

manufacturing, many plants also use other energy types. In 1998, 16.3% of U.S.

manufacturing plants purchased oil, 64.3% purchased natural gas, and 0.5% purchased

coal.[45]

---

[43] The MECS, ASM, LBD, and Chiang-Haltiwanger dataset are described in detail in Section 3.3.

[44] I consider electricity "easier" to use because nearly all plant locations are wired for electricity. Additionally, many common types of equipment are powered by electricity (e.g., computers). The phrase "environmentally cleaner" here refers to the emissions of the manufacturing plant alone. The emissions of the electricity generating plant clearly affect the overall environmental cleanliness of using electricity.

[45] These values are the author's calculations on the 1998 MECS micro data weighted by MECS sample weight to be roughly nationally representative.

Table 3.1: Mean Nominal Energy Type Prices Paid by U.S. Manufacturing Plants, 1998

Energy Type	Units	Mean	SD
Physical Units			
Electricity	$ / MWh	44.51	21.41
Oil	$ / bbl	16.40	8.40
Natural Gas	$ / 1000 cu ft	2.82	1.05
Coal	$ / short ton	37.14	13.88
Btu Measures			
Electricity	$ / Million Btu	13.05	6.27
Oil	$ / Million Btu	4.25	2.22
Natural Gas	$ / Million Btu	2.74	1.01
Coal	$ / Million Btu	1.68	0.62

Notes: Mean prices are purchase-weighted (weighted by MECS sample weight times the purchased quantity of the energy type). See Appendix B for the physical unit to Btu conversion factors used to create this table.

Source: Author's calculations on the 1998 MECS.

Energy prices vary both across time and in the cross-section. Oil has been famous for its price volatility since the energy crisis in 1973. Figure 3.1 displays the volatile nature of energy prices. Electricity prices depend heavily on the prices of the fuels used to generate electricity and hence are nearly as volatile as fuel prices. In addition to the volatility of electricity prices across time seen in Figure 3.1, Chapter 2 shows there is huge cross-sectional dispersion in electricity prices paid by U.S. manufacturing plants. The purchase-weighted standard deviation of electricity prices ranges from 38% to 55% in the 1963 to 2000 period. In this chapter, I find prices paid by U.S. manufacturing plants for other energy types also show considerable cross-sectional dispersion. For example, in 1998, the purchase-weighted standard deviation of log prices ranges from 27% for natural gas to 46% for oil. I find that the cross-sectional variation in energy type prices is due to a combination of the plant's industry, location, and purchase quantity.
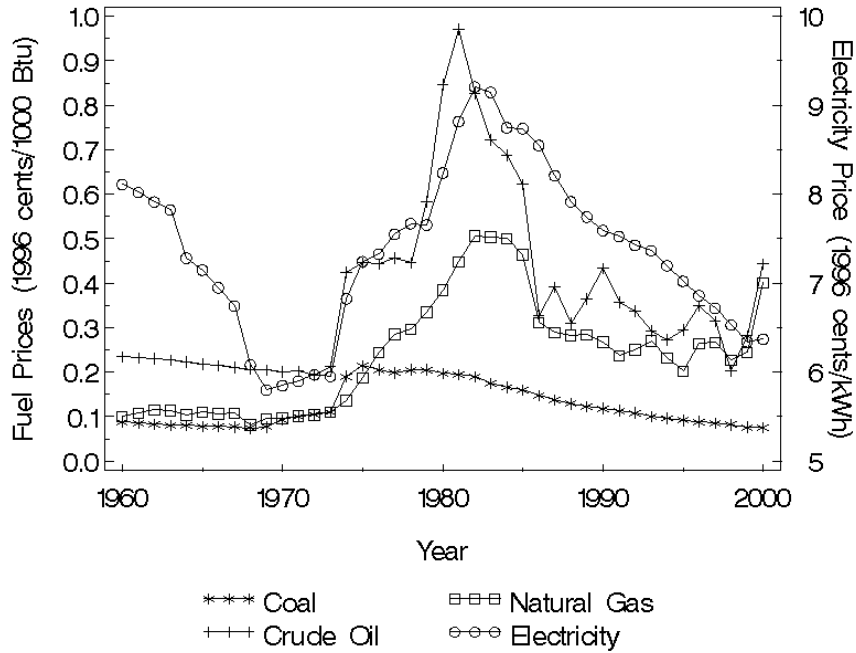
Source: EIA, 2004, Tables 5.19, 6.7, 7.8, and 8.10.

Figure 3.1: Real Prices for Coal, Natural Gas, Crude Oil, and Electricity, 1960-2000

Not surprisingly, given the considerable cross-sectional price dispersion and the inherent physical differences between energy types, there is heterogeneity in consumption of energy types across manufacturing plants. In 1998, only 0.22% of plants consume all four energy types. The majority of plants consume two of the four energy types. The variation in energy type consumption occurs both within and between industries. For example, while 25.3% of 4-digit Standard Industrial Classification (SIC) industries in 1998 contain at least one plant that consumes coal, only 0.5% of plants consume coal.[46]

The varying energy type consumption patterns of plants reflect differences in the technologies used by the plants. Plants face a two-step decision process. First, the plant

---

[46] An industry is counted as consuming an energy type if one or more plants in the industry consume the energy type. These percentages are the author's calculations on the 1998 MECS.

must choose their technology based on their physical requirements, the up-front cost of the technology, and the expected prices of the inputs required to run the technology for the time period they plan to use it. Once the plant's technology choice is in place, they decide how much of each input to consume.

There is potential for plants to substitute both between energy inputs and other inputs and between energy inputs themselves. There are a number of studies of factor substitution in manufacturing, including Griffin and Gregory (1976), Halvorsen (1977), and Woodland (1993). In this study, I exploit cross-sectional variation in input prices and consumption to estimate own- and cross-price demand elasticities for seven inputs.

I take advantage of two features of the 1998 MECS data not present in the types of data used in the bulk of earlier studies of factor substitution in manufacturing. Many previous studies, including Berndt and Wood (1975) and Nguyen and Streitwieser (1997), treat energy as an aggregate rather than examining individual energy types.[47] While this is a reasonable approach for many purposes, it fails to account for the possibility of substitution between energy types. I allow for some level of substitution between energy types by examining four distinct energy types. Additionally, the majority of earlier studies use aggregate industry-level data to estimate price elasticities.[48] These studies miss the within industry between plant variation that can only be adequately

---

[47] While many studies focus on energy as an aggregate, there are studies that look at individual energy types including Baxter and Rees (1968), Halvorsen (1977), Woodland (1993) and Bjørner and Jensen (2001).

[48] Woodland (1993), Nguyen and Streitwieser (1997), and Bjørner and Jensen (2001) are exceptions to this statement.

accounted for with plant-level micro data.[49] Two-digit industry, the level of aggregation

utilized in many previous studies, accounts for only between 5% and 28% of the variation

in energy type prices in 1998.[50] Even four-digit industry accounts for just 13% to 54% of

the variation in energy type prices in 1998. In this chapter, I exploit the cross-sectional

variation in input consumption and expenditures in the plant-level 1998 MECS data to

estimate elasticities of demand, accounting for the between plant variation missed by

most other studies.

This paper proceeds as follows. Section 3.2 provides background on previous

studies. Section 3.3 describes the data. Section 3.4 examines the energy type consumption

patterns of plants. Section 3.5 describes the magnitude and nature of variation in energy

type prices paid by U.S. manufacturing plants. I describe my elasticity estimation

methodology and results in Sections 3.6 and 3.7. Finally, Section 3.8 contains concluding

remarks.

3.2   Background

Considerable effort has gone in to the study of energy substitution in

manufacturing. Baxter and Rees (1968) examine the industrial demand for electricity

using industry-level quarterly time series data.[51] Unlike many studies before and since,

they do not combine energy types into an aggregate energy measure. They point out that

---

[49] The use of plant-level micro data in economic studies has grown in the last couple of decades. For example, there are many studies that use plant-level micro data to look at productivity dynamics. Haltiwanger (1997) provides a review of some of these studies and discusses the importance of plant-level data for this purpose.

[50] Section 3.5 contains more information on the sources of variation in energy type prices.

[51] They examine 16 industry groups including: Textiles, Leather, Non-ferrous Metals, Mining and Quarrying, and Food, Drink, Tobacco.

earlier studies calculate the demand for individual fuel types by estimating demand for an energy aggregate and then calculating the demand for the fuel type from its previous share of energy demand. Baxter and Rees find fault with this method since it does not reflect substitution between fuel types. They conclude relative price changes are not the only cause of changes in electricity consumption; both technology and output changes also affect electricity consumption.

A body of literature on the relationship between capital and energy emerged after the energy crisis in 1973 sharpened interest in the role of energy as an input into production. This literature focuses on whether capital and energy are substitutes or complements. The capital-energy debate has important policy implications. If energy and capital are substitutes, energy policies that result in raising the price of energy sources will have the effect of raising the demand for energy-saving capital. On the other hand, if energy and capital are complements, rising energy prices will reduce the demand for such capital. Unlike the pre-energy crisis study by Baxter and Rees (1968), the bulk of the capital-energy literature treats energy as an aggregate of electricity and fuel types. Papers by Hudson and Jorgenson (1974) and Berndt and Wood (1975) found that energy and capital are complements, forming the basis for the capital-energy literature. Both of these studies used time series data. However, studies by Humphrey and Moroney (1975), Griffin and Gregory (1976), and Halvorsen (1977) used cross-sectional data and found that energy and capital are substitutes. Later studies, including Field and Grabenstein (1980), Nguyen and Andrews (1989), and Morrison (1993) found mixed results.

While the capital-energy literature is both interesting and important, all of the studies mentioned above treat energy as an aggregate and ignore issues of inter-fuel

substitution such as those discussed by Baxter and Rees (1968). However, there are studies of energy in manufacturing that look at substitution between energy types. For example, Halvorsen (1977) and Taylor (1981) both study substitution between energy types in U.S. manufacturing at the 2-digit level. Additionally, Woodland (1993) and Bjørner and Jensen (2001) examine energy type substitution in Australian and Danish manufacturing, respectively.

Halvorsen (1977) estimates elasticities for most of the 20 2-digit SIC manufacturing sectors. He uses cross-sectional state data from the 1972 Census of Manufactures with the translog cost function to create his estimates.[52] Halvorsen's own-price elasticity results vary significantly across industries. They range from -0.124 to -1.096 for electricity, from -1.151 to -4.300 for fuel oil, from -0.425 to -2.134 for natural gas, and from -0.656 to -2.531 for coal.

Taylor (1981) uses the generalized Leontief function to estimate price effects for energy and transportation.[53] In the energy section of his study, he calculates elasticities for 17 of the 20 2-digit manufacturing sectors. He uses cross-sectional data for 48 states (excluding Texas and Louisiana) from the 1975 Annual Survey of Manufacturers. Taylor includes labor and capital in his estimations along with coal, oil, natural gas, and electricity. As in Halvorsen (1977), Taylor's own-price elasticity results vary significantly over industries. They range from -0.1 to -3.9 for electricity, from -0.4 to -5.8 for fuel oil, from -0.4 to -2.6 for natural gas, and from -0.1 to -1.9 for coal.

---

[52] See Christensen, Jorgenson, and Lau (1973) for details on the development of the translog function.

[53] The generalized Leontief function is described in Diewert (1971).

Woodland (1993) uses longitudinal establishment level Australian manufacturing data from 1977-1985 to estimate price elasticities for coal, oil, gas, electricity, labor, and capital. He estimates a separate set of elasticities for each set of establishments that consume the same pattern of energy types with a translog demand system. Additionally, he gives very careful consideration to the declining block price structures of electricity and gas.[54] Woodland's overall own-price elasticity estimates vary by industry sub-division, ranging from -0.782 to -1.062 for coal, -1.215 to -2.333 for oil, -1.250 to -2.540 for gas, and -1.218 to -1.520 for electricity.[55]

Bjørner and Jensen (2001) use company level Danish industrial data from 1983 to 1997 to estimate price elasticities for electricity, "district heating", and other fuels.[56] Like Woodland (1993), they estimate their demand models individually for each set of companies that use the same energy type patterns, although they do not include non-energy inputs in their estimation. They estimate elasticities based on both translog and generalized Leontief demand systems. Bjørner and Jensen's demand elasticity estimates for energy types are much lower than estimates of many other studies. Their translog-based own-price elasticity estimate for electricity is –0.27, and their own-price elasticity estimate for other fuels is –0.14.

---

[54] While a detailed study of the declining block price structures of electricity and natural gas are outside of the scope of this chapter, this type of pricing structure does exist in the U.S. as shown for electricity in Chapter 2.

[55] Woodland looks at 12 manufacturing industry sub-divisions. They are roughly similar to U.S. 2-digit SIC industries.

[56] "District heating" is publicly supplied heat in Denmark; it is similar in nature to the public water supply system in the U.S. The "other fuels" category is a composite of coal, oil, and natural gas. The authors note that Danish regulations tie the price of natural gas to the price of oil products.

With some exceptions, several of which are discussed above, previous studies use either industry-level data or aggregate data for all of manufacturing to study factor substitution. In a review of the debate over whether capital and energy are substitutes or complements, Solow (1987) questioned the suitability of aggregate data for this purpose. He states the argument can never be resolved using aggregate data because changes in the composition of products in aggregate output and changes in the composition of inputs will occur as energy prices change. Solow concludes:

> "Factor substitution is a microeconomic phenomenon, and is best examined by looking at microeconomic data." (Solow, 1987, p. 612)

3.3 Data

This study uses plant-level data from the 1998 Manufacturing Energy Consumption Survey (MECS), the Longitudinal Business Database (LBD), the 1998 Annual Survey of Manufactures (ASM), and the Chiang-Haltiwanger dataset. I use plant-level data on energy type purchases and expenditures from the 1998 MECS and information on plant age from the LBD for the discussion on energy type consumption and price dispersion in Sections 3.4 and 3.5. My substitution analysis requires information on output, plant characteristics, and non-energy production inputs not available in the 1998 MECS. I obtain this data by matching the 1998 MECS data to the 1998 ASM, LBD, and the Chiang-Haltiwanger dataset.

3.3.1   1998 Manufacturing Energy Consumption Survey

The 1998 MECS sample is a subset of the 1997 Census of Manufactures (CM).[57]

MECS estimates represent approximately 98% of total manufacturing payroll (EIA,

2005). The MECS collects plant-level data on purchases and expenditures for many

energy types.[58] I focus on four major energy types: electricity, oil, natural gas and coal.

The 1998 MECS micro data file contains 18,115 plant-level observations. Of

these, 14,197 observations (78%) have a positive MECS sample weight (*MECSWT*).[59]

The 1998 MECS is based on the North American Industrial Classification System

(NAICS).[60] However, I use SIC industries in this paper to facilitate my use of data from

the Chiang-Haltiwanger dataset. The NAICS set of manufacturing industries differs

slightly from the SIC set of manufacturing industries.[61] There are 134 (0.9%)

observations in the positive *MECSWT* sample that have 4-digit SIC codes outside of the

---

[57] The CM is collected every five years for years ending in "2" and "7". The CM includes the full universe of manufacturing plants.

[58] See Appendix A for the exact wording of the purchases and expenditures questions from the 1998 MECS survey form.

[59] The MECS sample weight for each plant is the inverse probability of selection times an adjustment factor to account for non-response. The probability of selection is proportional to a plant-level measure of size, total energy expenditures. This information on the construction of MECS sample weights was obtained from Appendix B of the 1994 MECS publication (EIA, 1997). There is no similar information available specifically for the 1998 MECS. The MECS sample weights are constructed in the same way for the 1994 and 1998 MECS. The electronically published 1998 MECS tables link to the 1994 published information on sample design. Observations on the micro data file with non-positive MECS sample weights have sample weights of zero. These are plants that were added as 1998 births but are not included in the final sample, plants that were mailed surveys in error, or plants that did not respond and are not included in the final sample.

[60] Since 1998 was the first year the NAICS basis was used for the MECS, SIC industries are also provided on the MECS micro data file. See Table B.1 in Appendix B for a list of covered SIC industries in the 1998 MECS. Note all plants in ten energy intensive SIC industries were selected for the 1998 MECS sample.

[61] Publishing and logging plants are in SIC manufacturing and not in NAICS manufacturing, while bakeries (that bake on the premises) and custom manufacturing plants are in NAICS manufacturing and not in SIC manufacturing.

SIC definition of manufacturing. Additionally, there are 14 observations in the positive

MECSWT sample that do not consume any of the major energy types. Observations with

industry codes outside of the SIC manufacturing industries and observations that do not

consume any of the major energy types are dropped, leaving 14,049 plant-level

observations.

3.3.2 Longitudinal Business Database

I obtain plant age data for 1998 MECS plants from the LBD. Ron Jarmin and

Javier Miranda created the LBD by linking data from the Census Bureau's annual

business register files. Currently, the LBD contains longitudinal linkages for the period

1975-2000. A detailed description of the LBD is available in Jarmin and Miranda (2002).

There are 13 observations from the remaining 14,049 MECS observations that cannot be

matched to the LBD. I drop these observations, leaving the 1998 MECS dataset used in

the analyses in 3.4 and 3.5 with 14,036 plant-level observations.

3.3.3 1998 Annual Survey of Manufactures

I obtain additional information on MECS manufacturing plants from the 1998

ASM. The ASM is a series of five-year panels that are refreshed by births as a panel ages.

Large manufacturing plants with at least 250 employees are sampled with certainty, and

smaller plants with at least 5 employees are sampled randomly with probabilities that

increase with the number of employees. ASM plants account for about one-sixth of all

manufacturing plants and about three-quarters of manufacturing employment. I use plant-

level data on shipments, inventories, labor, and materials from the 1998 ASM. I create a

plant-level output measure (Y_e) from 1998 ASM variables as shown in (3.1).

$$Y_e = TVS_e + \left(FIE_e - FIB_e\right) + \left(WIE_e - WIB_e\right) \qquad (3.1)$$

where e indexes the plant, TVS_e = total value of shipments, FIB_e = finished goods

inventory beginning of year, FIE_e = finished goods inventory end of year, WIB_e = work-

in-progress inventory beginning of year, and WIE_e = work-in-progress inventory end of

year.

I make use of three labor-related measures available in the 1998 ASM: total

salaries and wages for all workers (SW), number of non-production employees (OE), and

production worker hours (PH). Clearly, I would like to have hours for all workers rather

than just production worker hours. Unfortunately, this measure is not available in the

ASM. Following Davis and Haltiwanger (1991), I calculate the mean annual hours per

non-production worker by 2-digit SIC industry from the Current Population Survey

(CPS).[62] I call this mean NPH_WORKER_d, where d indexes 2-digit industries and

calculate total hours as shown in (3.2).

$$QL_e = PH_e + \left(OE_e \times NPH\_WORKER_d\right) \qquad (3.2)$$

I then calculate a plant-level wage as shown in (3.3).

$$PL_e = \frac{SW_e}{QL_e} \qquad (3.3)$$

The plant-level wage in (3.3) does not reflect worker heterogeneity. I exploit the

fact that the available skill mix varies across industries and location to create a proxy for

---

[62] I use the 1999 CPS March supplement data for this calculation. Note the 1999 March supplement asks
questions about 1998. I restrict my sample to workers who are full time (>= 35 hours per week), between
18-64 years old, not self-employed, and make at least half of minimum wage. More detailed information,
such as the breakdown of major occupation codes into production and non-production workers, is available
upon request from the author.

the wage rate the plant actually faces. I run a regression of the plant-level wage rate, calculated as in (3.3), on county and industry fixed effects using the entire 1998 ASM. I use the plant's predicted value from this regression as the wage in my elasticity estimation models.

In addition to the data on output and labor, I need data on quantity and price for materials and capital. I calculate plant-level expenditures on materials as the sum of cost of materials and parts (*CP*), cost of resales (*CR*), and cost of contract work (*CW*) from the 1998 ASM. The 1998 ASM does not have data on materials quantities or prices nor does it contain capital stock or capital rental price data. I obtain that information from the Chiang-Haltiwanger dataset.

3.3.4 Chiang-Haltiwanger Dataset

The Chiang-Haltiwanger dataset contains data on output, labor, capital stock, and materials for 1972 to 2000. Hyowook Chiang and John Haltiwanger created this dataset using the 1972-2000 ASM and CM microdata files, the NBER-CES productivity database, and various publicly available measures from the Bureau of Economic Analysis and the Bureau of Labor Statistics.[63] The measures vary in level from plant-level to 2-digit SIC level. I use a plant-level measure of capital stock, a 2-digit SIC industry-level measure of the rental price of capital, and a 4-digit SIC industry-level measure of materials prices.[64]

---

[63] This dataset is used in Becker et al (2005) and Chiang (2005), Chapter 2 "Learning By Doing, Worker Turnover and Productivity Dynamics".

[64] The 1998 ASM does not contain information on capital stocks. However, some earlier years contain capital stock information. Chiang and Haltiwanger use this capital stock information and a perpetual inventory method to create plant-level capital stocks for years without capital stock information. In the Chiang-Haltiwanger dataset, capital stock and rental price are split into structures and equipment

3.3.5   ASM-MECS Matched Dataset

Of the 14,036 observations in the 1998 MECS, 8,990 observations (64%) can be matched to the 1998 ASM. I further require that observations can be matched to the Chiang-Haltiwanger dataset, have no extreme outliers in energy type prices, and have positive values for total value of shipments (*TVS*), output (*Y*), payroll (*SW*), production worker wages (*WW*), production worker hours (*PH*), total employment (*TE*), electricity expenditures (*E10*), and electricity consumption (*Q10*).[65] These requirements reduce the sample size by an additional 1.3% to 8,872 observations.[66] Finally, I create adjusted sample weights based on ASM sample weights for the ASM-MECS matched dataset.[67]

3.4   Energy Type Consumption Patterns

Table 3.2 highlights the fact that not all plants use all energy types, showing the percent of plants consuming each energy type. Nearly all plants, 99.8% on a sample-weighted basis, consume electricity. More than half of plants consume natural gas. Oil is consumed by 16.3% of plants, while coal is only consumed by 0.5% of plants. Comparison of the unweighted and sample-weighted percentages in Table 3.2 shows the

---

components. Since energy types are my primary focus, I combine these into single measures of capital stock and rental price.

[65] I look for price outliers among six energy types: electricity, residual fuel oil, total diesel fuel and distillate fuel oil, total liquefied petroleum gases (LPG) and natural gas liquids (NGL), natural gas, and coal. I flag extreme energy type price outliers as observations with energy type prices greater than 10 times the full MECS mean for LPG/NGL and coal, greater than 5 times the full MECS mean for other energy types. A total of 29 observations are dropped for this reason. I do this to avoid problems with outliers in the estimation procedure. Further, there are four observations with positive shipments and payroll that have negative values of output. There is no particular reason to consider these observations to be invalid. However, negative output values will cause problems in the estimation procedure. Since there are only four observations of this type, I drop them.

[66] Note there are no observations in the final sample with positive consumption of an energy type and missing or zero expenditures of an energy type. The alternative is also true; there are no observations in the final sample with positive expenditures of an energy type and missing or zero consumption of that energy type.

[67] See Appendix D for a detailed description of the construction of adjusted sample weights.

MECS survey sample is skewed toward large, energy intensive plants that are more likely to use energy types other than electricity. For example, 5.0% of plants consume coal on an unweighted basis, while only 0.5% of plants consume coal on a sample-weighted basis.

Table 3.2: Percent of Plants Consuming Individual Production Inputs, 1998

	1998 MECS	1998 ASM-MECS Matched
	Unweighted	
Number of Observations	14,036	8,872
	Percent of Observations With Positive Input Consumption	
Electricity	99.7	100.0
Oil	36.0	42.7
Natural Gas	80.8	86.0
Coal	3.7	5.0
Capital	-	100.0
Labor	-	100.0
Materials	-	100.0
	Sample weighted	
Number of Observations	213,661	160,318
	Percent of Observations With Positive Input Consumption	
Electricity	99.8	100.0
Oil	16.3	14.7
Natural Gas	64.3	71.5
Coal	0.5	0.8
Capital	-	100.0
Labor	-	100.0
Materials	-	100.0

Source: Author's calculations on the 1998 MECS and the 1998 ASM-MECS matched dataset.

It is rare for a single plant to consume all four energy types. Table 3.3 shows the sample-weighted percent of plants and the percent of industries that consume one to four energy types. On a sample-weighted basis, only 0.22% of plants consume all four energy types. The majority of plants, 63.3%, consume just two energy types.

Table 3.3: Percent of Plants and Industries Consuming 1-4 Energy Types, 1998

	1998 MECS	
	Percent	
Number of Fuels	Plants	4-Digit SIC Industries
1	27.97	0.22
2	63.34	12.56
3	8.47	62.56
4	0.22	24.67
	1998 ASM-MECS Matched	
	Percent	
Number of Fuels	Plants	4-Digit SIC Industries
1	23.90	0.69
2	65.45	14.68
3	10.36	61.24
4	0.29	23.39

Notes: The percent of plants is calculated on a MECS sample-weighted basis. The percent of industries is calculated on a completely unweighted basis (i.e., an industry is listed as consuming two energy types regardless of whether all plants consume the same two energy types or some plants consume one energy type and the rest of the plants consume another energy type).

Source: Author's calculations on the 1998 MECS.

There are 16 possible patterns of energy type consumption for plants; only 11 patterns appear in the data. The top panel of Figure 3.2 shows the top five energy type patterns consumed by U.S. manufacturing plants on a MECS sample-weighted basis. The majority of plants, 55.6%, consume natural gas and electricity and no other energy types. The segments within each of the bars in Figure 3.2 indicate the fraction of the total quantity purchased (Btu) of each energy type for all of the plants purchasing that energy type pattern. For example, looking at the sample-weighted distribution of plants that purchase just electricity and natural gas, 29% of the total Btu purchases of these plants are electricity and 71% are natural gas.

The second largest group of plants, comprising 27.8% of plants, consumes electricity alone. The energy type consumption pattern of electricity, natural gas, and oil comes in third with 8.3% of plants, followed closely by the pattern of electricity and oil with 7.7% of plants. There is a sharp drop off from fourth to fifth place. Plants that consume oil alone are the fifth largest group of plants and account for only 0.22% of plants.
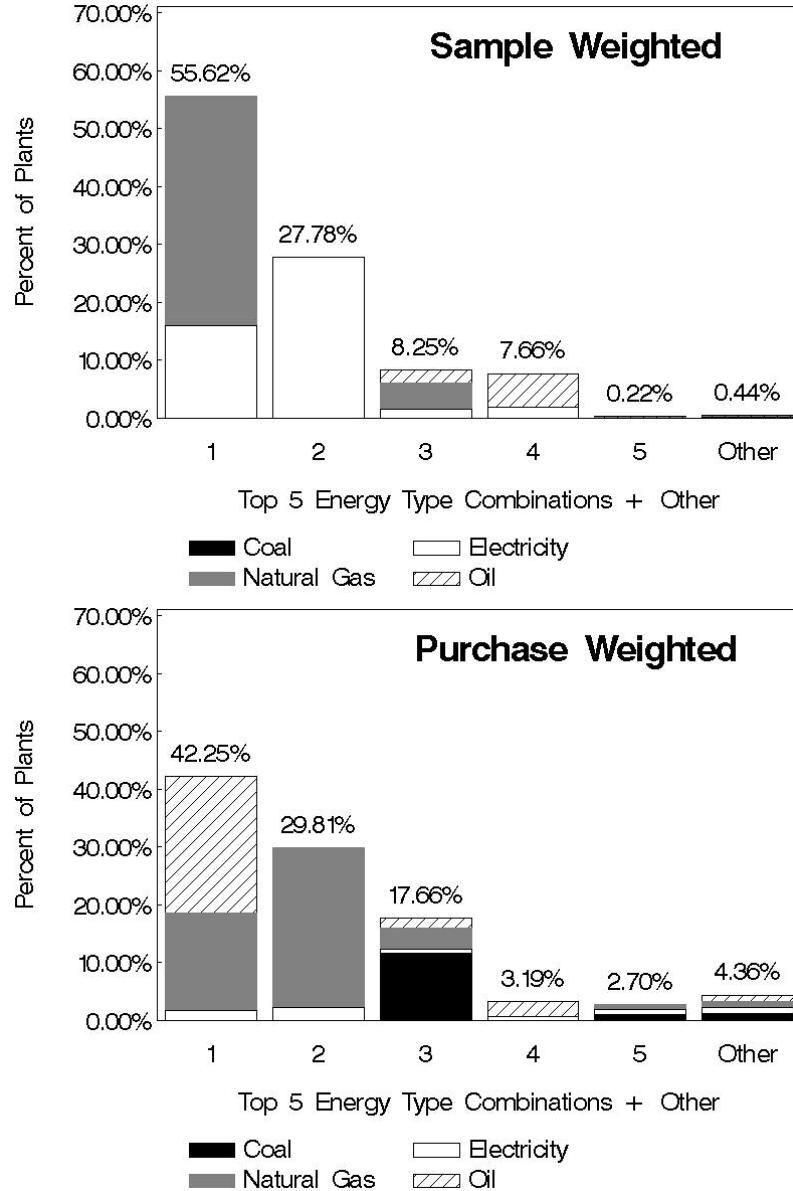
Source: Author's calculations on the 1998 MECS microdata with non-manufacturing plants (on a 1987 SIC) basis excluded.

Notes: The fifth place energy type combination in the top panel is oil alone. The sizes of very small bar segments within larger bars are exaggerated for clarity. Purchase weighting indicates weighting by sample weight times the quantity of energy type purchases in Btu.

Figure 3.2: Top Five Energy Type Consumption Patterns for Manufacturing Plants on a MECS Sample-weighted Basis (Top Panel) and on a Purchase-weighted Basis (Bottom Panel), 1998 MECS

The lower panel of Figure 3.2 shows the top five energy type patterns consumed by U.S. manufacturing plants weighted by sample weight times purchase quantity (Btu). In the purchase-weighted panel, first place is taken over by the pattern of electricity, natural gas, and oil, while the pattern of electricity and natural gas moves to second place. Therefore, plants that purchase higher amounts of energy overall are more likely to consume the combination of electricity, natural gas, and oil than just electricity and natural gas. It is also interesting to note that electricity alone drops out of the top five, falling to number 6 with 2% of plants. This implies plants that purchase large amounts of energy are less likely to purchase electricity alone than plants that purchase smaller amounts of energy. This makes intuitive sense since electricity is the easiest of the energy types to access and use. For example, a plant that uses very little energy overall will likely consume electricity alone; it is not worthwhile for the plant to invest in a boiler that burns oil, natural gas, or coal.

Figure 3.3 is analogous to Figure 3.2, but it is for the ASM-MECS matched dataset. Comparing the sample-weighted panels of Figures 3.2 and 3.3, the top five energy type consumptions patterns are the same. The purchase-weighted panels of Figures 3.2 and 3.3 also show the same top five patterns, in the same order with roughly the same percent of plants. Overall, comparison of Figures 3.2 and 3.3 show the fuel use characteristics of the ASM-MECS matched dataset are a reasonably close match to the fuel use characteristics of the full MECS dataset.
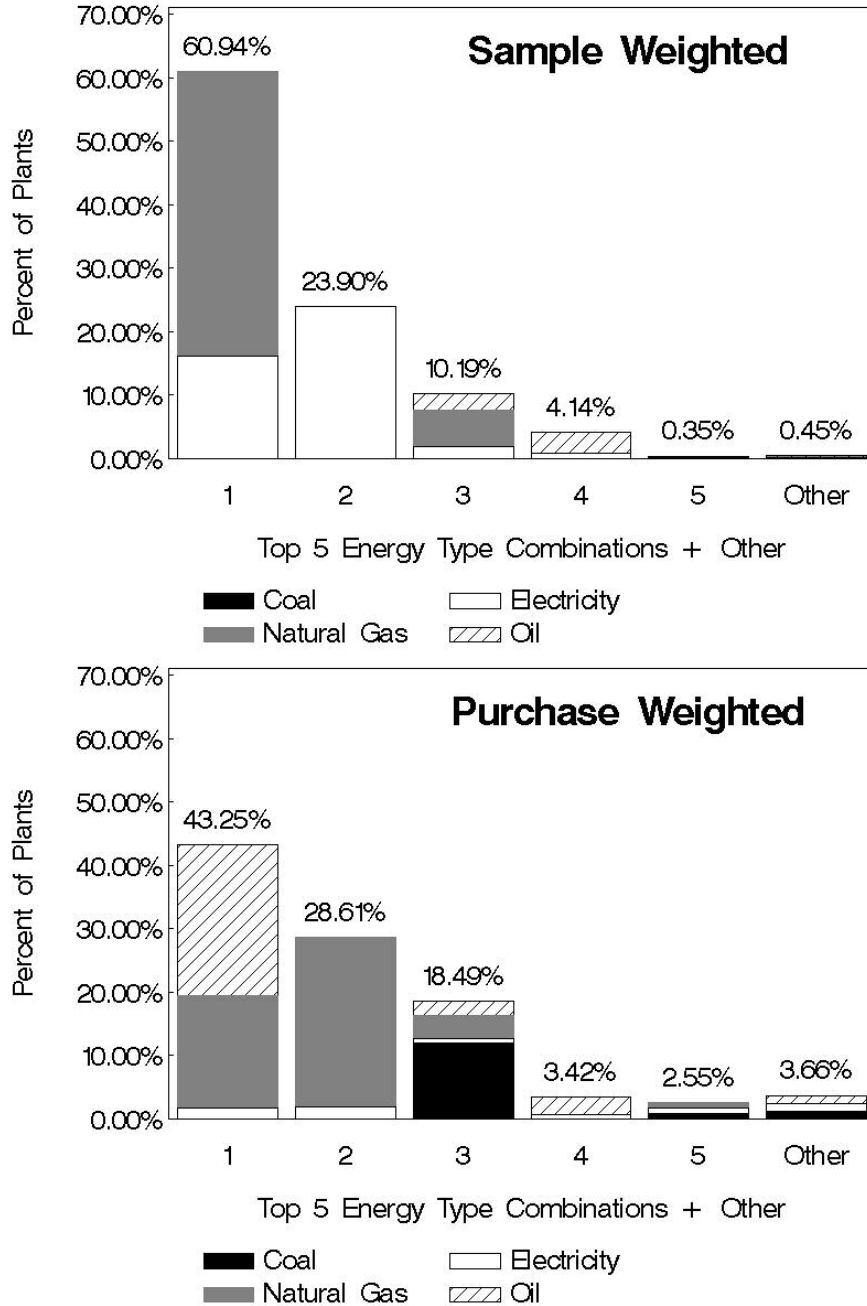
Source: Author's calculations on the 1998 ASM-MECS matched dataset.

Notes: The fifth place energy type combination in the top panel is oil alone. The sizes of very small bar segments within larger bars are exaggerated for clarity. Purchase weighting indicates weighting by sample weight times the quantity of energy type purchases in Btu.

Figure 3.3: Top Five Energy Type Consumption Patterns for Manufacturing Plants on an Adjusted ASM Sample-weighted Basis (Top Panel) and on a Purchase-weighted Basis (Bottom Panel), 1998 ASM-MECS Matched Dataset

I explore manufacturing energy type consumption patterns further by running probit regressions of energy type consumption on plant characteristics using the full 1998 MECS dataset. I run probit regressions for the energy types oil, natural gas, and coal. I do not run a probit regression for electricity since nearly all plants in the 1998 MECS consume electricity.

The dependent variable in the probit regressions is a 0-1 dummy variable for energy type consumption; it is equal to one if the plant consumes the energy type and zero otherwise. Plant characteristics included as explanatory variables in the probit regressions are: size, location, age and industry. Plant size is represented by a categorical variable based on the plant's total value of shipments (*TVS*).[68] The Census regions (Northeast, Midwest, South, and West) are used to represent the location of the plant. Plant age is included using a categorical variable based on plant age as calculated from the LBD.[69] Finally, I include 2-digit SIC industry dummies as explanatory variables in the probit regression.

---

[68] The size categories are *TVS* (in million $) from 0-19, 20-49, 50-99, 100-249, 250-499, and $\geq$ 500. I use the categorical shipments variable included on the 1998 MECS micro data file (and used in MECS published tables) since I do not have detailed shipments information for all 1998 MECS plants. The 1998 MECS also contains a categorical employment variable that could be used as a measure of plant size. I ran a version of the probit regression using the employment-based size variable and found very similar results.

[69] The plant age categories are plant age (in years) from 0-4, 5-9, 10-19, and $\geq$ 20. The plant age measure is based on the LBD. In particular, age is counted from the first year the plant appears in the LBD. Since the LBD starts in 1975, the age variable is censored at 23 years old. Therefore, I use a categorical variable for plant age.

Table 3.4 shows marginal effects from the three energy type consumption probit regressions described above.[70] All of the plant size dummy marginal effects are significant. As plants get larger, they are more likely to consume coal, oil, and natural gas. Plants with 500 million dollars or more in total shipments are over 38% more likely to consume oil, over 27% more likely to consume natural gas, and over 7% more likely to consume coal than plants with less than 20 million dollars in shipments. Additionally, older plants are more likely than younger plants to consume oil, natural gas, and coal. For example, plants 20 years old or older are 8% more likely to consume oil than plants between 0 and 4 years of age. Plant location also affects the plant's choice of whether or not to consume oil, natural gas, and coal. Plants in the located in the Northeast are more likely to consume oil than plants in other areas. There is also weaker evidence that Northeastern plants are the most likely to consume coal.[71] Finally, Table 3.4 shows Midwestern plants are the most likely to consume natural gas.

---

[70] The 2-digit industry Leather and Leather Products (31) contains no plants that consume coal. Therefore, 2-digit industry exactly identifies coal consumption, requiring that plants in the 2-digit industry 21 be left out of the coal consumption probit regression. Note, in Table 3.4, only 13,940 observations are used in the coal probit regression.

[71] Note the negative marginal effect on the South region dummy is not significant at the 5% level.

Table 3.4: Energy Type Use Probit Marginal Effects, 1998 MECS

Variable	Marginal Effects		
	Oil	Natural Gas	Coal
Midwest	-0.1117**	0.2631**	-0.0004*
	(0.0112)	(0.0221)	(0.0002)
South	-0.0716**	-0.0087	-0.0002
	(0.0114)	(0.0258)	(0.0002)
West	-0.0987**	-0.0107	-0.0004**
	(0.0100)	(0.0282)	(0.0002)
Plant Age: 5-9 Years	0.0429*	0.0504	0.0016**
	(0.0215)	(0.0321)	(0.0011)
Plant Age: 10-19 Years	0.0508**	0.0621*	0.0002
	(0.0198)	(0.0295)	(0.0002)
Plant Age: $\geq$ 20 Years	0.0825**	0.1370**	0.0011**
	(0.0166)	(0.0291)	(0.0004)
TVS: 20-49 Million $	0.0798**	0.1260**	0.0016**
	(0.0120)	(0.0231)	(0.0006)
TVS: 50-99 Million $	0.1612**	0.1689**	0.0042**
	(0.0177)	(0.0334)	(0.0013)
TVS: 100-249 Million $	0.2087**	0.2119**	0.0146**
	(0.0192)	(0.0200)	(0.0039)
TVS: 250-499 Million $	0.3316**	0.2346**	0.0406**
	(0.0290)	(0.0196)	(0.0099)
TVS: $\geq$ 500 Million $	0.3875**	0.2768**	0.0735**
	(0.0326)	(0.0142)	(0.0174)
Log Pseudo Likelihood	-4529.808	-7943.827	-309.534
Pseudo R^2	0.275	0.131	0.277
N	14036	14036	13940

* $p < 0.05$, ** $p < 0.01$

Notes: The marginal effects for the dummy variables are for a discrete change of the dummy variable from 0 to 1. The probit regressions are weighted by MECS sample weight. Two-digit SIC industry coefficients are suppressed. The omitted region is Northeast. The omitted age category is 0-4 years, and the omitted shipments category is 1-19 Million $.

Source: Author's calculations on the 1998 MECS.

3.5   Energy Type Price Variation

In this section, I describe the magnitude and nature of variation in energy type prices paid by U.S. manufacturing plants. I decompose the variance of energy type prices into within- and between-group components for the 1998 MECS. The decomposition methodology exactly follows the methodology in Chapter 2. Indexing plants by e and groups by g, write the overall variance as

$$V = \sum_e s_e (p_e - \bar{p})^2 = \sum_g \sum_{e \in g} s_e (p_e - \bar{p})^2$$

$$V = \sum_g s_g \left(\sum_{e \in g} s_e (p_e - \bar{p}_g)^2 \right) + \sum_g s_g (\bar{p}_g - \bar{p})^2 \tag{3.4}$$

$$V = \sum_g s_g V_g^W + V^B = V^W + V^B$$

where p_e is the log price of the energy type for plant e, s_e is the weight for plant e, $\bar{p}$ is the overall weighted mean log price, $\bar{p}_g$ is the weighted mean log price for group g, $s_g = \sum_{e \in g} s_e$ is the sum of weights for plants in group g, V_g^W is the weighted variance within group g, and V^B is the between-group variance. I examine seven industry and/or geographic area groups: 4-digit industry, 2-digit industry, state, Census division, Census region, annual purchase quantity decile, and annual purchase quantity decile-state.[72] Table 3.5 contains purchase-weighted dispersion and variance decomposition measures for log energy type prices.

---

[72] Annual purchase quantity deciles are based on the MECS sample-weighted distribution of log quantity of the purchased energy type. See Appendix E for definitions of Census regions and divisions.

Table 3.5: Purchase-Weighted Distribution of Log Energy Type Prices Paid by U.S.
Manufacturing Plants, Dispersion and Variance Decompositions, 1998

	Electricity	Oil	Natural Gas	Coal
Overall SD	0.43	0.46	0.27	0.43
Number of Observations	13,979	5,046	11,347	524
Price Dispersion Between Industries				
4-Digit SIC (454)				
Between SD	0.32	0.22	0.17	0.15
Between Variance as % of Total	54.0	23.0	37.3	12.7
2-Digit SIC (20)				
Between SD	0.23	0.12	0.13	0.10
Between Variance as % of Total	28.0	7.0	23.3	5.3
Price Dispersion Between Geographic Areas				
State (51)				
Between SD	0.24	0.16	0.16	0.32
Between Variance as % of Total	30.3	11.6	35.1	56.3
Census Division (9)				
Between SD	0.14	0.12	0.15	0.25
Between Variance as % of Total	11.0	6.5	30.7	35.3
Census Region (4)				
Between SD	0.12	0.09	0.25	0.23
Between Variance as % of Total	8.2	3.5	35.3	28.3
Price Dispersion Between Annual Purchase Quantity Deciles				
Purchase Deciles (10)				
Between SD	0.29	0.12	0.15	0.11
Between Variance as % of Total	45.2	7.1	30.8	6.8
Price Dispersion Between State-Annual Purchase Quantity Decile Groups				
State x Purchase Decile				
Between SD	0.35	0.18	0.20	0.38
Between Variance as % of Total	64.4	15.7	54.0	76.7

Source: Author's calculations on the 1998 MECS.

Table 3.5 shows the overall standard deviation of electricity prices is 43% in 1998. The other energy types also have notable levels of price dispersion. Coal and oil prices have levels of dispersion comparable to electricity. The overall standard deviations of coal and oil prices stand at 43% and 46%, respectively. The dispersion of natural gas prices is the lowest at 27%. The differences in prices paid for energy types reflect the combined effects of differences between plants including industry, location, and purchase amount.

Industry accounts for a large fraction of the variation in the prices of electricity and natural gas. Four-digit industry accounts for roughly 54% of the variation in electricity prices and 37% of the variation in natural gas prices. More modestly, four digit-industry accounts for 13% and 23% of the price variation in coal and oil.

Location also helps explain the dispersion in energy type prices. With the exception of oil, state dummies never explain less than 30% of the price variation for any energy type. There are a couple of reasons to expect energy type prices to vary with location. First, environmental regulations vary widely across states. These regulations can affect energy type prices both directly and indirectly. California, for example, is well known for having more stringent environmental regulations and higher electricity prices than other states. The plant's distance from the source of the energy type could also play a role in the variation of energy type prices across locations through differences in transport costs.

Location has the largest impact on coal prices. State effects account for 56% of the coal price variation. A couple of factors help account for the relatively large impact of location on coal prices. First, since coal is generally transported by train, it has relatively

high transportation costs compared to many other energy types. This implies, all else

equal, that plants located close to coal mines will pay lower prices for coal. Additionally,

there is significant heterogeneity in the types of coal that make up the total coal energy

type in the MECS.[73]

Different regions of the U.S. mine different types of coal, which fetch different

prices. One major difference across coal types is the dramatic variation in heat content.

For example, anthracite has roughly 21% higher heat content than combined bituminous

and sub-bituminous coal.[74] There are also notable differences in the sulfur content of coal

types mined in different areas. As noted in both Burtraw (1996) and Bonkowski (1999),

sub-bituminous coal has significantly lower sulfur content than does bituminous coal.

According to Bonkowski (1999), Western mines tend to produce more sub-bituminous

coal, while Appalachian mines tend to produce more bituminous coal. Plants may be

willing to pay more for lower sulfur content coal if they are attempting to lower sulfur

dioxide (SO_2) emissions in the face of environmental regulations.[75]

In Chapter 2, I found evidence of large quantity discounts in the electricity prices

paid by U.S. manufacturers. Table 3.5 supports this conclusion; purchase decile accounts

for 45% of the variation in electricity prices. There is also strong evidence of quantity

discounts in natural gas where purchase decile accounts for 31% of the price variation.

---

[73] In the 1998 MECS, total coal is the sum of three types of coal: anthracite, bituminous and sub-bituminous coal, and lignite.

[74] This percentage was calculated based on information from Table C.1 in Appendix C.

[75] See Burtraw (1996) and Bonkowski (1999) for detailed descriptions of the coal industry in the 1990s and the effect of SO_2-related environmental regulations on the industry. This discussion indicates it would be interesting to examine the dispersion of coal sub-type prices individually. The comparison of primary interest would be between bituminous and sub-bituminous coal. However, these two types of coal are combined in the 1998 MECS. Recall, total coal is the sum of three types of coal: anthracite, bituminous and sub-bituminous coal, and lignite. Unfortunately, there are not enough plants that purchase anthracite and lignite to do an interesting and disclosable analysis of their price dispersion.

Coal and oil show evidence of much smaller quantity discounts, with just 6-7% of their price variation accounted for by purchase decile.

Table 3.6 compares price dispersion measures of log electricity prices from the 1998 MECS and the 1998 cross-section of the Prices and Quantities of Electricity in Manufacturing (PQEM) database utilized in Chapter 2 and described in detail in Chapter 4.[76] It is reassuring to note the overall standard deviations of log electricity prices are quite similar across the two data sources. Electricity prices in the 1998 MECS have a slightly higher standard deviation of 43% compared to 42% in the PQEM. In general, the between group variation numbers are quite similar. The largest difference occurs for 4-digit SIC. In the 1998 MECS, 4-digit SIC accounts for 54% of the electricity price variation while it only accounts for 43% of the electricity price variation in the 1998 PQEM.

---

[76] The 1998 ASM is the source of the 1998 PQEM electricity price information. The electricity purchases and expenditures questions in the MECS and ASM are equivalent.

Table 3.6: Purchase-Weighted Distribution of Log Electricity Prices Paid by U.S. Manufacturing Plants, Dispersion and Variance Decompositions, MECS and PQEM, 1998

	Electricity	
	MECS	**PQEM**
Overall SD	0.43	0.42
Number of Observations	13,979	58,852
Price Dispersion Between Industries		
4-Digit SIC (454)		
Between SD	0.32	0.27
Between Variance as % of Total	54.0	42.5
2-Digit SIC (20)		
Between SD	0.23	0.20
Between Variance as % of Total	28.0	23.0
Price Dispersion Between Geographic Areas		
State (51)		
Between SD	0.24	0.25
Between Variance as % of Total	30.3	36.1
Census Division (9)		
Between SD	0.14	0.15
Between Variance as % of Total	11.0	12.0
Census Region (4)		
Between SD	0.12	0.12
Between Variance as % of Total	8.2	8.9
Price Dispersion Between Annual Purchase Quantity Deciles		
Purchase Deciles (10)		
Between SD	0.29	0.27
Between Variance as % of Total	45.2	42.6
Price Dispersion Between State-Annual Purchase Quantity Decile Groups		
State x Purchase Decile		
Between SD	0.35	0.34
Between Variance as % of Total	64.4	66.9

Source: Author's calculations on the 1998 MECS and 1998 PQEM.

3.6 Elasticity Estimation Methodology

3.6.1 Translog Demand System

In this chapter, I use a translog demand system to estimate elasticities of demand for factors of production based on the 1998 ASM-MECS matched dataset.[77] Originally developed by Christensen, Jorgenson, and Lau (1971, 1973) the translog function is a flexible functional form that allows for substitution between inputs. The non-homothetic translog cost function for plant e is specified in (3.5):[78]

$$\ln C_e = \ln \alpha_0 + \sum_{i=1}^{n} \alpha_i \ln p_{ie} + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij} \ln p_{ie} \ln p_{je} + \alpha_Y \ln Y_e + \frac{1}{2} b_{YY} \left(\ln Y_e\right)^2$$
$$+ \sum_{i=1}^{n} b_{iY} \ln p_{ie} \ln Y_e + \sum_{k=1}^{20} d_k IND_{ke} + \sum_{z=1}^{4} r_z REG_{ze}$$

(3.5)

where p_i is the price of input i, p_j is the price of input j, Y is output, IND_k are 2-digit industry dummies, REG_z are Census region dummies, and n is the total number of inputs. Industry dummies are included to account for industry-specific effects such as differences in technology across industries. Region dummies are included to account for region-specific effects such as differences in the availability and prices of energy types across regions.

The restrictions $b_{ij} = b_{ji}$, $\sum_{i=1}^{n} \alpha_i = 1$, and $\sum_{i=1}^{n} b_{ij} = \sum_{j=1}^{n} b_{ij} = \sum_{j=1}^{n} b_{iY} = 0$ are implied from assumptions of symmetry and linear homogeneity in input prices. There are n

---

[77] My choice of the translog demand system has important implications. If the estimated coefficients are close to zero, the translog demand system implies large amounts of substitution between inputs. Other models could yield very different results based on the same data. For example, the Generalized Leontief (Diewert, 1971) demand system implies there is little substitution between inputs if the estimated coefficients are close to zero.

[78] See pages 469-476 of Berndt (1990) for a detailed discussion of the non-homothetic translog function.

inputs. Logarithmically differentiating equation (3.5) with respect to p_i yields equation

(3.6).

$$\frac{\partial \ln C_e}{\partial \ln p_{ie}} = \frac{\partial C_e}{\partial p_{ie}} \frac{p_{ie}}{C_e} = \alpha_i + \sum_{j=1}^{n} b_{ij} \ln p_{je} + b_{iY} \ln Y_e, \qquad i = 1, ..., n \qquad (3.6)$$

Substituting in Shephard's Lemma, $\frac{\partial C_e}{\partial p_{ie}} = X_{ie}$ where X_i is the factor demand for i, along

with $C_e = \sum_{j=1}^{n} X_{je} p_{je}$ yields the cost share form of the translog equation shown in (3.7).

$$S_{ie} = \frac{X_{ie} p_{ie}}{\sum_{j=1}^{n} X_{je} p_{je}} = \alpha_i + \sum_{j=1}^{n} b_{ij} \ln p_{je} + b_{iY} \ln Y_e, \qquad i = 1, ..., n \qquad (3.7)$$

In equation (3.7), S_{ie} is the cost share of input i in the total cost of producing

output for plant e. The system of cost share equations and restrictions is linearly

dependent. In order to estimate the system of equations, one cost share equation must be

dropped.[79] Dropping the share equation for input n, the plant-level share equations to be

estimated are shown in (3.8) where e indexes plants and ε_{ie} is an error term.

$$S_{ie} = \alpha_i + \sum_{j=1}^{n} b_{ij} \ln\left(\frac{p_{je}}{p_{ne}}\right) + b_{iY} \ln Y_e + \varepsilon_{ie}, \qquad i = 1, ..., n-1 \qquad (3.8)$$

In order to obtain estimates for α_o, α_Y, and b_{YY}, the cost function shown in (3.5) is

estimated in conjunction with the share equations in (3.8).

---

[79] As long as the system of cost share equations is estimated using maximum likelihood, it does not matter
which cost share equation is dropped.

Own- and cross-price elasticities of demand can be calculated for each plant using the estimated coefficients and the predicted factor shares. The plant-level own-and cross-price elasticities are calculated as shown in (3.9) and (3.10). [80]

$$\hat{\varepsilon}_{ii,e} = \frac{\hat{b}_{ii} + \hat{S}_i^2 - \hat{S}_i}{\hat{S}_i}, \quad i = 1, \dots, n \tag{3.9}$$

$$\hat{\varepsilon}_{ij,e} = \frac{\hat{b}_{ij} + \hat{S}_i \hat{S}_j}{\hat{S}_i}, \quad i, j = 1, \dots, n \text{ where } i \neq j \tag{3.10}$$

3.6.2 Estimation Methodology

There is an important economic selection problem to be addressed when estimating elasticities of demand for individual energy types using plant-level data. This issue is not confronted in the large part of the earlier literature, which primarily looks at data aggregated to the industry level and/or at aggregate energy. As I discussed at length in Section 3.4, all plants do not consume all energy types. This leads to a major complication for the estimation of the translog demand system using the ASM-MECS matched dataset. There are a non-trivial number of zeros in the oil, natural gas, and coal cost share dependent variables. Table 3.2 shows the number of plants with non-zero consumption of each input in the ASM-MECS matched dataset. All of the plants in the matched dataset consume electricity, capital, labor, and materials. On a completely unweighted basis, only 5% of plants consume coal, and only 43% of plants consume oil. A higher percentage, 86%, of plants consume natural gas.

---

[80] See Berndt (1990), p. 475.

The varying energy type consumption patterns of plants reflect differences in the technologies used by the plants. Plants face a two-step decision process.[81] First, the plant makes a choice of technology. This decision is based on their physical requirements, the up-front cost of the technology, and the expected prices of the inputs required to run the technology for the time period they plan to use it. In some cases, the plant's physical requirements alone will determine the technology they purchase and the input(s) they consume. Take an extreme example where there is only one type of machine the plant can purchase to perform a necessary function. If that machine uses oil as an energy source, the plant will end up consuming oil, regardless of their expectations for future oil prices. However, if the plant has a menu of available technology options based on their physical requirements, they will be able to include the relative up-front costs of the technologies and expected prices of input(s) used by the technologies when making their technology choice. Once the plant's technology choice is in place, they decide how much of each input to consume.

The issue of censoring in the dependent variables that results from the differing technology choices of plants is not a simple issue to address.[82] I follow the basic methodology of Woodland (1993) and Bjørner and Jensen (2001) by estimating a separate translog demand system for each energy type consumption pattern. I augment their methodology by including a correction for the plant's choice of technology and resulting energy type consumption pattern.

---

[81] A detailed model of the plant's technology choice is outside the scope of this chapter, but is an interesting avenue for possible future work.

[82] It is technically possible to estimate the translog demand system ignoring the censoring that occurs in the oil, natural gas, and coal equations. I do this in Section 3.7.2. However, the resulting estimates are both biased and inconsistent.

This approach fixes the plant's menu of inputs after the plant makes an initial technology choice. The model is not exactly like a traditional putty-clay model since substitution between all inputs on the plant's menu of inputs is allowed after the initial technology choice. However, the model is similar in nature to putty-clay in that the plant cannot add an input to their menu of input choices after they make their initial technology choice. As a simple example, take a plant with all electric-powered technology. If electricity prices rise relative to coal prices, the plant might like to switch to coal-powered technology. By estimating separate translog models for each energy type consumption pattern, I implicitly assume the plant would not be able to make the switch to coal-powered technology.

A drawback of this approach is that the estimated elasticities will not represent potential switches in plant technology and resulting energy type consumption patterns.[83] However, both Woodland and Bjørner and Jensen observe persistence in energy type consumption patterns for their plants/companies over time. I find 67% of plants in both the 1994 and 1998 MECS have the same energy type consumption pattern in both years.[84] Persistence in energy type consumption patterns is highest for plants consuming all four energy types, with 80% of these plants having the same energy type consumption pattern in both 1994 and 1998.

---

[83] Recently, alternative methods have been developed to address the issue of censoring in the dependent variables. However, these methods involve computationally intensive estimation of complex likelihood functions and are outside the scope of this chapter. See Yen, Lin, and Smallwood (2003) for discussions of some of these methods. Also, Golan, Perloff, and Shen (2001) present a generalized maximum entropy method for solving a demand system with binding non-negativity constraints. Their method is also quite computationally intensive; they are forced to limit their sample to 1,000 household observations due to a computer space limitation.

[84] There are 5,368 plants in both the 1994 and 1998 MECS micro data files. Since the MECS samples plants with a probability proportional to their energy consumption, the 1994 to 1998 continuers are likely to be plants that consume relatively higher amounts of energy.

I account for the plant's choice of technology in my estimation by including a selection effect control variable in the translog model. To obtain the selection effect control variable, I first estimate a univariate probit regression for each energy type consumption pattern.[85] I then calculate the inverse Mills ratio for each energy type consumption pattern based on the probit estimates. Finally, I include the inverse Mills ratio as a selection correction control in each of the translog model equations, (3.5) and (3.8).[86]

The dependent variable in each univariate probit regression is a 0-1 dummy for whether or not the plant consumes that energy type pattern.[87] As in my individual energy type consumption probit models in Section 3.4, I include categorical plant characteristic variables as explanatory variables. Specifically, I include the following plant characteristics: region, plant age, shipments-based plant size, and 2-digit industry.[88] I include plant size and age as proxies for technology. Region is included as a proxy for the plant's expected prices of inputs. For some inputs, where quantity discounts are available, plant size also acts as a proxy for the plant's expected input price. Electricity, studied in detail in Chapter 2, is a good example of such an input. The quantity of purchased

---

[85] My selection control is not perfect. I account for selection (or non-selection) of each energy type consumption pattern individually, rather than in a system wide manner. Ideally, I would like to estimate a multivariate probit for all energy type consumption patterns rather than individual probits for each energy type consumption pattern. However, I do not currently have the computing capability required to estimate a multivariate probit for all of the energy type consumption patterns together. It should be possible to use a multinomial logit to calculate a selection effect control variable that reflects system wide selections of energy type consumption patterns. I plan to pursue this possibility in future work.

[86] The inverse Mills ratio is equal to the standard normal probability density function divided by the standard normal cumulative density function.

[87] I treat these as exclusive choices (e.g., a plant that consumes electricity, natural gas, and coal has a value of one in the dependent variable for the electricity, natural gas, and coal probit regression and a value of zero in all other probits – including the electricity and natural gas probit).

[88] The plant characteristic variables are described in detail in Section 3.4.

electricity is positively correlated with plant shipments, and Chapter 2 showed evidence of large quantity discounts for electricity. Industry is included as a proxy for both technology and expected input prices.

There are eight possible energy type consumption patterns in the 1998 ASM-MECS matched dataset. All of these patterns appear in the data. I estimate a translog demand system for six of the eight energy type consumption patterns. I do not have enough observations to estimate the translog demand system for the two energy type consumption patterns of electricity and coal and electricity, oil, and coal. Combined, these two energy type consumption patterns are seen in only 57 plants (0.6%), which account for 1.3% of energy expenditures and 1.8% of energy consumption on a sample-weighted basis.[89]

Table 3.7 shows descriptive statistics by energy type consumption pattern and for all plants combined. Not surprisingly, plants that use electricity alone are the smallest plants in terms of both employment and shipments. They also have the lowest overall energy expenditures and energy intensity. On the other hand, plants that consume all four energy types are the largest plants with the highest energy expenditures.

---

[89] Also note, these 57 plants account for 0.5% of shipments, 0.9% of total employment, and 0.9% of payroll in the 1998 ASM-MECS matched dataset.

Table 3.7: Descriptive Statistics by Energy Type Consumption Pattern, 1998 ASM-MECS Matched Dataset

Statistic	Electricity	Electricity and Natural Gas	Electricity and Oil	Electricity, Natural Gas, and Coal	Electricity, Oil, and Natural Gas	Electricity, Oil, Natural Gas, and Coal	All Plants
Number of Observations	734	4,248	451	91	2,991	300	8,872
Number of Employees							
Mean	47	82	82	851	259	932	95
Standard Deviation	98	227	167	1,309	760	1,841	334
Total Value of Shipments (1000 $)							
Mean	6,502	17,000	15,162	465,966	81,464	364,438	22,391
Standard Deviation	38,362	121,942	65,841	1,107,011	362,689	724,481	162,706
Energy Expenditures (1000 $)							
Mean	95	265	452	10,908	1,856	16,130	453
Standard Deviation	1,587	1,835	5,214	13,634	13,474	31,268	5,141
Energy Intensity							
Mean	0.019	0.068	0.043	0.067	0.032	0.109	0.052
Standard Deviation	0.023	0.095	0.067	0.071	0.068	0.174	0.083

Notes: Energy intensity is calculated as energy expenditures divided by total value of shipments. There are two additional energy type consumption patterns that appear in the data: (1) electricity and coal and (2) electricity, oil, and coal. Statistics for these two energy type consumption patterns are withheld for disclosure reasons.

Source: Author's calculations on the 1998 ASM-MECS matched dataset.

3.7 Elasticity Results

In this section, I present the results of my elasticity estimations. First, I aggregate the four energy types together and run a simple four-input KLEM (capital, labor, energy, and materials) model, both to get results comparable to other studies and as a reality check of the data in the ASM-MECS matched dataset. Second, I run a version of the translog model ignoring the zero consumption problems. Finally, I estimate the translog model described in Section 3.6, with and without selection controls, for each of the six empirically relevant energy type consumption patterns.

3.7.1 Aggregate Energy: A Simple KLEM Model

I aggregate energy types together and run a simple four-input KLEM model. Since all plants use at least one energy type, there are no zero consumption problems to consider in this model. I solve the model using full information maximum likelihood.[90] Table 3.8 shows elasticity of demand estimates for the KLEM model.

---

[90] I solve all models described in this section using full information maximum likelihood. All models in this section are sample-weighted; Nguyen and Streitwieser (1997) find that an unweighted model underestimates the elasticities of demand.

Table 3.8: Elasticity Estimates $\left(\hat{\varepsilon}_{ij} \right)$ for the KLEM model, 1998

i \ j	Energy	Capital	Labor	Materials
Energy	-1.841**	0.236**	0.336**	1.268
	(0.674)	(0.068)	(0.052)	(0.658)
Capital	0.053**	-0.309**	0.129**	0.128**
	(0.000)	(0.001)	(0.000)	(0.001)
Labor	0.030**	0.041**	-0.472**	0.401**
	(0.000)	(0.005)	(0.006)	(0.001)
Materials	0.098**	0.040**	0.393**	-0.531**
	(0.000)	(0.000)	(0.000)	(0.000)
Log Likelihood	-15,422			
N	8,872			

* $p < 0.05$, ** $p < 0.01$

Notes: The model is sample-weighted. The elasticity estimates shown are sample-weighted means of the plant-level elasticity estimates. Standard errors for the elasticity estimates are calculated by bootstrapping with 1,000 iterations. Standard errors shown as '(0.000)' are less than 0.001.

Source: Author's calculations on the 1998 ASM-MECS matched dataset.

As expected, all of the own-price elasticities in Table 3.8 are negative. Energy has the most elastic demand with an own-price elasticity of -1.841. This estimate of energy's own-price elasticity falls between the estimate of -3.584 from Nguyen and Streitwieser (1997), which is based on 1991 MECS cross-sectional microdata, and the estimate (for 1971) of -0.49 from Berndt and Wood (1975), which is based on 2-digit industry-level time series data. Like Nguyen and Streitwieser (1997), I find positive cross-price elasticities between capital and energy indicating that capital and energy are substitutes.

3.7.2 Model Ignoring the Zero Consumption Problem

I also run a version of the translog model ignoring the zero consumption problems. Despite the obvious problems with this model, it serves as a benchmark for the corrected model. I must address the problem of missing plant-level prices for some inputs to run this version of the model. I address this problem by replacing the missing prices with industry-location mean prices.[91] The replacement prices are my best guess of the price the plant would face if they purchased the energy type. Since I do not address the problem of zero consumption of oil, natural gas, and coal, the estimated coefficients and the resulting elasticity estimates in Table 3.9 are suspect.

As in the KLEM model, all of the own-price elasticity estimates from this model are negative. The energy types are considerably more elastic than capital, labor, and materials. The own-price elasticity estimate of electricity is a very elastic -8.553, and the oil own-price elasticity comes in at a huge -18.279. However, the standard error for the oil estimate is also very large, 28.744, indicating that this estimate is not to be trusted, even putting aside the issues raised by ignoring the zero consumption problems. The own-price elasticity estimates for natural gas, -1.747, and coal, -1.476, in Table 3.9 are within the range of the estimates of Halvorsen (1977), Taylor (1981) and Woodland (1993). Their natural gas own-price elasticity estimates range from -0.4 to -2.6, and their coal own-price elasticity estimates range from -0.1 to -2.5.

---

[91] Yen, Lin, and Smallwood (2003) and Golan, Perloff, and Shen (2001) take this approach. I calculate the industry-location means based on the full 1998 MECS sample. I replace missing energy type prices with the first available mean energy type price from the following ordered list: 4-digit SIC industry-state mean, 3-digit SIC industry-state mean, 2-digit SIC industry-state mean, 4-digit SIC industry-division mean, 3-digit SIC industry-division mean, 2-digit SIC industry-division mean, 4-digit SIC industry mean, 3-digit SIC industry mean, 2-digit SIC industry mean, state mean, or division mean.

Table 3.9: Elasticity Estimates $(\hat{\varepsilon}_{ij})$, Model Ignoring Zero Consumption Problem, 1998

i \ j	Electricity	Oil	Natural Gas	Coal	Capital	Labor	Materials
Electricity	-8.553*	1.270*	-2.104*	0.221*	2.913*	-2.460	8.713*
	(3.901)	(0.653)	(1.085)	(0.114)	(1.424)	(1.474)	(4.269)
Oil	20.132	-18.279	-4.713	0.084	-3.303	5.717	0.361**
	(33.951)	(28.744)	(8.695)	(0.159)	(5.637)	(8.947)	(0.112)
Natural Gas	-0.790**	-0.114*	-1.747**	-0.020**	0.417**	1.168**	1.085**
	(0.030)	(0.005)	(0.026)	(0.001)	(0.010)	(0.027)	(0.024)
Coal	1.626	0.040	-0.380	-1.476**	-1.006	2.090*	-0.894
	(0.870)	(0.022)	(0.209)	(0.259)	(0.630)	(0.917)	(0.710)
Capital	0.070**	-0.002**	0.016**	-0.002**	-0.341**	0.153**	0.107**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.003)	(0.005)	(0.003)
Labor	0.012**	0.003**	0.017**	0.002**	0.048**	-0.479**	0.397**
	(0.001)	(0.000)	(0.001)	(0.000)	(0.009)	(0.010)	(0.001)
Materials	0.076**	0.001**	0.015**	-0.001**	0.031**	0.390**	-0.512**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
Log Likelihood	41,342						
N	8,872						

* $p < 0.05$, ** $p < 0.01$

Notes: The model is sample-weighted. The elasticity estimates shown are sample-weighted means of the plant-level elasticity estimates. Standard errors for the elasticity estimates are calculated by bootstrapping with 1,000 iterations. Standard errors shown as '(0.000)' are less than 0.001.

Source: Author's calculations on the 1998 ASM-MECS matched dataset.

3.7.3 Individual Energy Type Consumption Pattern Models

Finally, I estimate the translog model described in Section 3.6 for each of the six empirically relevant energy type consumption patterns. The models include a selection correction control variable based on energy type consumption pattern probits as described in Section 3.6.2.[92] The marginal effects from the first stage probits for each energy type consumption pattern are shown in Table 3.10. The results are consistent with expectations based on the oil, natural gas, and coal consumption probits in Section 3.4. Larger, older plants are less likely to consume electricity alone. The largest plants are the most likely to consume all four energy types. Tables 3.11 - 3.16 show the estimated elasticities for each of the six energy type consumption patterns.

---

[92] I also run the individual energy type consumption pattern models without correcting for selection. Tables of estimated elasticities from these models can be found in Appendix F.

Table 3.10: Energy Type Consumption Pattern Probit Marginal Effects,
            1998 ASM-MECS Matched Dataset

	Marginal Effects		
Variable	Electricity	Electricity and Natural Gas	Electricity and Oil
Midwest	-0.009	0.201	-0.022**
	(0.095)	(0.106)	(0.006)
South	0.095	0.015	-0.006**
	(0.097)	(0.108)	(0.002)
West	-0.063	0.190	-0.005**
	(0.084)	(0.095)	(0.002)
Plant Age: 5-9 Years	-0.168	0.324*	-0.005
	(0.084)	(0.123)	(0.003)
Plant Age: 10-19 Years	-0.182*	0.279*	0.000
	(0.066)	(0.102)	(0.004)
Plant Age: ≥ 20 Years	-0.236**	0.273*	-0.001
	(0.081)	(0.112)	(0.003)
TVS: 20-49 Million $	-0.124**	-0.012	-0.002
	(0.037)	(0.052)	(0.002)
TVS: 50-99 Million $	-0.148**	-0.053	0.001
	(0.038)	(0.056)	(0.003)
TVS: 100-249 Million $	-0.194**	-0.102	-0.001
	(0.039)	(0.091)	(0.001)
TVS: 250-499 Million $	-0.186**	-0.179**	-0.002
	(0.037)	(0.058)	(0.001)
TVS: ≥ 500 Million $	-0.186**	-0.211**	-0.003*
	(0.037)	(0.058)	(0.001)
Log Pseudo Likelihood	-4151.069	-4919.098	-905.079
Pseudo R^2	0.147	0.171	0.409
N	8,851	8,872	8,872

* $p < 0.05$, ** $p < 0.01$

Notes: The marginal effects for the dummy variables are for a discrete change of the dummy variable from 0 to 1. The probit regressions are weighted by MECS sample weight. Two-digit SIC industry coefficients are suppressed. The omitted region is Northeast. The omitted age category is 0-4 years, and the omitted shipments category is 1-19 Million $.

Source: Author's calculations on the 1998 MECS.

Table 3.10: (Continued) Energy Type Consumption Pattern Marginal Effects, 1998 ASM-MECS Matched Dataset

Variable	Electricity, Natural Gas, and Coal	Electricity, Oil, and Natural Gas	Electricity, Oil, Natural Gas, and Coal
Midwest	0.0003**	-0.008	0.000002
	(0.0002)	(0.018)	(0.000010)
South	0.0002**	0.005	-0.000011
	(0.0002)	(0.018)	(0.000009)
West	0.0002*	-0.017	0.000007
	(0.0002)	(0.014)	(0.000018)
Plant Age: 5-9 Years	-0.0000	-0.006	0.000016
	(0.0001)	(0.023)	(0.000027)
Plant Age: 10-19 Years	0.0000	0.003	0.000015
	(0.0001)	(0.020)	(0.000024)
Plant Age: ≥ 20 Years	0.0001	0.032	0.000015
	(0.0001)	(0.023)	(0.000017)
TVS: 20-49 Million $	0.0004**	0.135**	0.003582**
	(0.0002)	(0.037)	(0.001090)
TVS: 50-99 Million $	0.0010**	0.175**	0.015028**
	(0.0005)	(0.035)	(0.005048)
TVS: 100-249 Million $	0.0037**	0.288**	0.032364**
	(0.0017)	(0.080)	(0.008152)
TVS: 250-499 Million $	0.0062**	0.293**	0.099470**
	(0.0033)	(0.044)	(0.019890)
TVS: ≥ 500 Million $	0.0151**	0.304**	0.152033**
	(0.0064)	(0.048)	(0.027770)
Log Pseudo Likelihood	-48.255	-2105.925	-129.950
Pseudo R^2	0.310	0.279	0.403
N	7265	8872	8168

* $p < 0.05$, ** $p < 0.01$

Notes: The marginal effects for the dummy variables are for a discrete change of the dummy variable from 0 to 1. The probit regressions are weighted by MECS sample weight. Two-digit SIC industry coefficients are suppressed. The omitted region is Northeast. The omitted age category is 0-4 years, and the omitted shipments category is 1-19 Million $.

Source: Author's calculations on the 1998 MECS.

Table 3.11: Elasticity Estimates $\left(\hat{\varepsilon}_{ij}\right)$ for the Electricity Energy Type Consumption
Pattern, 1998

i \ j	Electricity	Capital	Labor	Materials
Electricity	-2.126**	0.155**	0.176**	1.795**
	(0.002)	(0.001)	(0.002)	(0.004)
Capital	0.017**	-0.681**	-0.029**	0.693**
	(0.000)	(0.001)	(0.001)	(0.002)
Labor	0.009**	-0.018	-0.701**	0.710**
	(0.002)	(0.033)	(0.022)	(0.057)
Materials	0.081**	0.320**	0.790**	-1.191**
	(0.000)	(0.001)	(0.001)	(0.001)
Log Likelihood	-1,157			
N	734			

* $p < 0.05$, ** $p < 0.01$

Notes: Selection corrections are included in this model. The model is sample-weighted. The elasticity estimates shown are sample-weighted means of the plant-level elasticity estimates. Standard errors for the elasticity estimates are calculated by bootstrapping with 1,000 iterations. Standard errors shown as '(0.000)' are less than 0.001.

Source: Author's calculations on the 1998 ASM-MECS matched dataset.

Table 3.12: Elasticity Estimates $\left(\hat{\varepsilon}_{ij}\right)$ for the Electricity and Natural Gas Energy Type Consumption Pattern, 1998

i \ j	Electricity	Natural Gas	Capital	Labor	Materials
Electricity	-1.781**	-0.090	1.097	-2.190	2.965
	(0.533)	(0.063)	(0.633)	(1.716)	(1.680)
Natural Gas	-0.165**	-1.159**	0.142**	0.675**	0.507**
	(0.002)	(0.001)	(0.002)	(0.001)	(0.003)
Capital	0.138**	0.007**	-0.357**	0.455**	-0.244**
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
Labor	-0.039	0.010	0.135**	-0.556**	0.450**
	(0.124)	(0.006)	(0.009)	(0.014)	(0.095)
Materials	0.135**	0.009**	-0.103**	0.514**	-0.554**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Log Likelihood	1,573				
N	4,248				

* $p < 0.05$, ** $p < 0.01$

Notes: Selection corrections are included in this model. The model is sample-weighted. The elasticity estimates shown are sample-weighted means of the plant-level elasticity estimates. Standard errors for the elasticity estimates are calculated by bootstrapping with 1,000 iterations. Standard errors shown as '(0.000)' are less than 0.001.

Source: Author's calculations on the 1998 ASM-MECS matched dataset.

Table 3.13: Elasticity Estimates $\left(\hat{\varepsilon}_{ij}\right)$ for the Electricity and Oil Energy Type
Consumption Pattern, 1998

i \ j	Electricity	Oil	Capital	Labor	Materials
Electricity	-0.861**	0.077*	0.063	0.374**	0.347**
	(0.069)	(0.034)	(0.076)	(0.062)	(0.089)
Oil	0.096	-2.623	-1.184	0.989	2.723
	(0.098)	(2.171)	(1.816)	(0.949)	(2.939)
Capital	0.003	-0.132**	-0.310**	0.521**	-0.082
	(0.003)	(0.027)	(0.090)	(0.045)	(0.105)
Labor	0.030**	0.066**	0.337**	-0.764**	0.330**
	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)
Materials	0.015**	0.096**	0.012**	0.180**	-0.303**
	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)
Log Likelihood	622				
N	451				

* $p < 0.05$, ** $p < 0.01$

Notes: Selection corrections are included in this model. The model is sample-weighted. The elasticity estimates shown are sample-weighted means of the plant-level elasticity estimates. Standard errors for the elasticity estimates are calculated by bootstrapping with 1,000 iterations. Standard errors shown as '(0.000)' are less than 0.001.

Source: Author's calculations on the 1998 ASM-MECS matched dataset.

Table 3.14: Elasticity Estimates $(\hat{\varepsilon}_{ij})$ for the Electricity, Natural Gas, and Coal Energy Type Consumption Pattern, 1998

j i	Electricity	Natural Gas	Coal	Capital	Labor	Materials
Electricity	1.393	0.210**	-0.873*	1.476*	-1.919*	-0.287
	(0.921)	(0.071)	(0.357)	(0.557)	(0.867)	(0.324)
Natural Gas	-0.215**	-2.023**	-0.252**	-0.455*	1.373**	1.572**
	(0.034)	(0.280)	(0.056)	(0.196)	(0.304)	(0.251)
Coal	3.048**	-0.685**	-3.122**	0.128**	3.925**	-3.294**
	(0.370)	(0.082)	(0.266)	(0.010)	(0.461)	(0.476)
Capital	-0.211**	-0.076**	0.024**	0.223**	0.055**	-0.016**
	(0.001)	(0.001)	(0.001)	(0.003)	(0.002)	(0.003)
Labor	0.439**	0.186**	0.238**	0.051**	-1.493**	0.579**
	(0.001)	(0.000)	(0.001)	(0.002)	(0.004)	(0.003)
Materials	0.087**	0.056**	-0.035**	0.015**	0.186**	-0.310**
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
Log Likelihood	703					
N	91					

* $p < 0.05$, ** $p < 0.01$

Notes: Selection corrections are included in this model. The model is sample-weighted. The elasticity estimates shown are sample-weighted means of the plant-level elasticity estimates. Standard errors for the elasticity estimates are calculated by bootstrapping with 1,000 iterations. Standard errors shown as '(0.000)' are less than 0.001.

Source: Author's calculations on the 1998 ASM-MECS matched dataset.

Table 3.15: Elasticity Estimates $\left(\hat{\varepsilon}_{ij}\right)$ for the Electricity, Oil, and Natural Gas Energy Type Consumption Pattern, 1998

j i	Electricity	Oil	Natural Gas	Capital	Labor	Materials
Electricity	-1.674**	-0.006**	-0.638**	-0.457**	1.327**	1.447**
	(0.026)	(0.000)	(0.024)	(0.023)	(0.040)	(0.032)
Oil	-12.280	-302.330	-304.603	132.183	604.543	-117.514
	(249.061)	(6101.161)	(6167.401)	(2672.779)	(12235.552)	(2390.709)
Natural Gas	-0.616	-0.218	-2.116	0.165**	1.296	1.489
	(4.096)	(1.424)	(7.286)	(0.009)	(6.841)	(5.964)
Capital	-0.065*	0.017**	0.012**	0.078	0.273**	-0.316
	(0.030)	(0.005)	(0.000)	(0.328)	(0.018)	(0.321)
Labor	0.105**	0.039**	0.098**	0.186**	-0.762**	0.334**
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
Materials	0.047**	0.001**	0.042**	-0.045**	0.141**	-0.186**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log Likelihood	14,924					
N	2,991					

* $p < 0.05$, ** $p < 0.01$

Notes: Selection corrections are included in this model. The model is sample-weighted. The elasticity estimates shown are sample-weighted means of the plant-level elasticity estimates. Standard errors for the elasticity estimates are calculated by bootstrapping with 1,000 iterations. Standard errors shown as '(0.000)' are less than 0.001.

Source: Author's calculations on the 1998 ASM-MECS matched dataset.

Table 3.16: Elasticity Estimates $\left(\hat{\varepsilon}_{ij}\right)$ for the Electricity, Oil, Natural Gas, and Coal Energy Type Consumption Pattern, 1998

i ╲ *j*	Electricity	Oil	Natural Gas	Coal	Capital	Labor	Materials
Electricity	-1.878**	-0.146**	0.390**	0.440**	-0.432**	0.709**	0.916**
	(0.046)	(0.007)	(0.018)	(0.020)	(0.027)	(0.032)	(0.020)
Oil	4.673	3.237	-2.148	1.201	-2.138	-0.126	-4.700
	(66.590)	(61.108)	(30.651)	(16.193)	(5.180)	(33.192)	(74.868)
Natural Gas	-0.505	-0.101	0.016	0.050	0.226	0.001	0.313
	(4.691)	(0.886)	(8.471)	(0.060)	(1.544)	(0.357)	(1.767)
Coal	0.602**	-0.049**	0.014**	-1.991**	0.412**	1.157**	-0.145**
	(0.026)	(0.003)	(0.001)	(0.051)	(0.046)	(0.011)	(0.030)
Capital	-0.161**	0.028**	0.012**	0.097**	0.444**	-0.100**	-0.318**
	(0.002)	(0.000)	(0.000)	(0.001)	(0.003)	(0.012)	(0.008)
Labor	0.171**	0.005**	0.050**	0.208**	-0.016**	-1.301**	0.884**
	(0.001)	(0.000)	(0.000)	(0.001)	(0.002)	(0.001)	(0.002)
Materials	0.078**	0.017**	0.033**	0.002**	-0.023**	0.309**	-0.416**
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)
Log Likelihood	2,774						
N	300						

* $p < 0.05$, ** $p < 0.01$

Notes: Selection corrections are included in this model. The model is sample-weighted. The elasticity estimates shown are sample-weighted means of the plant-level elasticity estimates. Standard errors for the elasticity estimates are calculated by bootstrapping with 1,000 iterations. Standard errors shown as '(0.000)' are less than 0.001.

Source: Author's calculations on the 1998 ASM-MECS matched dataset.

For the most part, my estimated elasticities look sensible. I discuss exceptions to this statement at the end of this section. Considering only estimates that are significant at least at the 5% level, my estimates of the own-price elasticity of electricity range from -0.861 to -2.126. None of my oil own-price elasticities are significant at the 5% level. The oil own-price elasticity for plants that consume only electricity and oil is -2.623 and is significant at the 25% level. There are only two own-price elasticity estimates of natural gas that are significant, -1.159 for plants that consume only electricity and natural gas and -2.023 for plants that consume only electricity, natural gas, and coal. My estimates of the own-price elasticity of coal are -3.122 for plants that consume only electricity, natural gas, and coal and -1.991 for plants that consume all four energy types. These estimated own-price elasticities have reasonable magnitudes. Since I am estimating the elasticities based on cross-sectional data, in some sense I am actually estimating long-run elasticities. My elasticity estimates are fairly close in magnitude to those estimated for U.S. manufacturing by Halvorsen (1977) and Taylor (1981) using state-industry level cross-sectional data. In general, my own-price elasticity estimates for energy types are a bit more elastic than the estimates of Woodland (1993) for Australian manufacturing plants.

Since all plants in the 1998 ASM-MECS matched dataset consume electricity, it is possible to compare estimates of the own-price elasticity of electricity across all energy type consumption patterns. You might expect to see more elastic demand for electricity as the number of energy types consumed increases since the plant seemingly has more possibilities for substitution away from electricity. However, there does not appear to be any clear relationship between the magnitude of the electricity own-price elasticity and

the number of energy types consumed. In fact, the estimated electricity demand is most elastic, -2.126, for plants that consume electricity only. Since electricity prices are strongly correlated with other energy type prices, a rise in electricity prices may mean the prices of other energy types have also risen, making them less attractive substitutes. It is also notable that the own-price elasticity of electricity varies with the plant's energy type consumption pattern even when the number of inputs is the same. For example, electricity demand is more elastic for plants that consume only electricity and natural gas, -1.781, than for plants that consume only electricity and oil, -0.861.

The differences in elasticities across energy type consumption patterns may be the result of differences in the underlying technologies of the plants. Consumption of a larger number of inputs does not necessarily mean it is easier to substitute between those inputs. Take a simple example of two plants. Plant A performs three different processes with three machines each powered by a different energy type. Plant B performs only one process and has a single machine powered by a variable combination of two energy types. While plant A consumes more energy types than plant B, it is clearly easier for plant B to substitute between the two energy types they consume.

In addition, I find including the selection correction in the demand model does affect the elasticity estimates. At least half of the coefficients on the inverse Mills ratios are significant at the 5% level for four of the six models. Only two of the five Mills ratio coefficients are significant in the model of the electricity and oil energy type consumption pattern, and none of the inverse Mills ratio coefficients are significant in the model of the electricity, natural gas, and coal energy type consumption pattern. Appendix F shows tables of estimated elasticities for the six energy type consumption

patterns without correcting for selection. In general, the selection correction has more of an effect on the estimated elasticities for the energy inputs than it does on the estimated elasticities for the non-energy inputs. For example, for plants that consume only electricity and natural gas, the selection adjusted own-price elasticity of materials is -0.554, while the non-selection adjusted own-price elasticity of materials is -0.521. On the other hand, the own-price elasticity of electricity is -1.781 in the selection adjusted case and –1.038 in the non-selection adjusted case.

Finally, while the magnitudes of the elasticity estimates in Tables 3.11-3.16 are sensible in general, I should note that the models do produce some problematic results. The estimated elasticity results, particularly those in Tables 3.14 - 3.16, should be considered with caution given these problematic results. First, all of the models with more than five inputs produce some positive values for own-price elasticities. For example, Table 3.14 contains positive own-price elasticity values for both electricity and capital. The positive electricity own-price elasticity estimate is not significant at the 5% level, while the positive capital own-price elasticity estimate is significant at the 1% level. The only other statistically significant (at least at the 5% level) own-price elasticity estimate comes for capital in Table 3.16. The models underlying the elasticity estimates in Tables 3.14 and 3.16 have relatively small numbers of observations, only 91 and 300, respectively. It is possible the small sample sizes in combination with the large number of parameters to be estimated are causing problems with the capital own-price elasticity estimates.

Also noteworthy are the large elasticities and even larger standard errors for oil in Table 3.15. These are driven in part by the large number of plants with oil shares close to

zero. Among plants that consume only electricity, oil, and natural gas, there are 1,170

observations (39.1%) with oil cost shares of less than 0.0001. These tiny oil cost shares

do not occur because the plants spend unrealistically small amounts of money on oil, but

because they are large plants that spend a lot of money on other inputs.[93] There is no

reason to consider these tiny oil cost shares invalid, but they do explain how it is possible

to see huge elasticity estimates and standard errors for oil, such as those in Table 3.15.

The denominator of the oil elasticity is the estimated oil cost share; see equations (3.9)

and (3.10). If this is close to zero, the elasticity will be very large. Further, seemingly

small fluctuations in the estimated oil cost share will cause large fluctuations in the

estimated elasticity and hence the bootstrap standard error. For example, an increase of

0.005 in the level of the predicted oil cost share for all plants changes the overall

estimated oil own-price elasticity in Table 3.15 from -302.33 to -1.45, a drop of over 99%

in absolute magnitude.[94] The inability to generate reasonable estimates of oil elasticities

---

[93] There are several possible reasons for the existence of tiny oil cost shares. For example, the MECS data on oil consumption includes consumption of diesel fuel that can be used to run both emergency generators and small vehicles such as forklifts. Hence it is possible for plants to consume some oil even if they do not power their main machinery with oil. Consumption of oil as a lubricant is also included in the MECS data. There is MECS data on how much of the oil is consumed as a non-fuel. Only 5% of plants that consume oil use oil as a non-fuel. That percentage is the same for plants with tiny (<0.0001) oil cost shares so I do not think use of oil as a lubricant is driving the existence of tiny oil cost shares. There is no obvious cutoff point for to define "tiny" oil cost shares. Most of the plants have small oil cost shares. Among plants with positive oil consumption, 34% of plants spend less than 0.01% of their total costs on oil. An additional 39% of these plants spend between 0.01% and 0.1% of their total costs on oil. As a sensitivity test, I take all plants with oil cost shares of less than 0.01% and set their oil consumption and expenditures to zero. I then recalculate total costs and cost shares and run the model for each adjusted energy type consumption pattern. The results, not shown, are mixed. For example, the problem oil elasticities and standard deviations in Table 3.15 have somewhat more realistic magnitudes but are still not statistically significant. I plan to explore this issue further in future work.

[94] The effect of very small oil cost shares can be seen to a lesser degree in Table 3.16. Also, the selection correction has an unpredictable effect on whether or not there are very large estimated oil elasticities in Tables 3.15 and 3.16. Comparing Table F.5, which does not include the selection correction, and Table 3.15, the oil elasticities are much more reasonable (though still not significant) when the selection correction is not included. The reverse is seen comparing Tables F.6 and 3.16. The very small oil cost shares cause the model to be quite sensitive regardless of whether or not selection corrections are included.

is a limitation of using this model with this data because the plant-level data contains a large number of observations with tiny oil cost shares.

3.8   Conclusions

In this chapter, I examine energy type consumption patterns, prices, and substitution in U.S. manufacturing. My major findings are summarized below.

- Energy type consumption patterns vary widely across manufacturing plants, with over half of plants, 55.6% on a sample-weighted basis, consuming only electricity and natural gas and only 0.22% of plants consuming all four energy types.

- There is a large amount of dispersion in the prices paid by U.S. manufacturing plants for electricity (43%), oil (46%), natural gas (27%), and coal (43%).

- There are significant quantity discounts given to manufacturing plants for both electricity (discussed in detail in Chapter 2) and natural gas.

- Location accounts for a large part of the variation in electricity, natural gas, and coal prices.

- Capital and aggregate energy are substitutes.

- The varying energy type consumption patterns of electricity introduce complications in estimating demand models with plant-level microdata.

- Correcting for the plant's initial choice of technology when estimating elasticities of demand for production inputs does affect the estimated elasticity results.

- Statistically significant (at the 1% level) estimates of the own-price elasticity of demand range from -0.861 to -2.126 for electricity, from -1.159 to -2.023 for natural gas, and from -1.991 to -3.122 for coal. I obtain an estimate of the own-price elasticity of demand for oil of -2.623 that is significant at the 25% level.

- Elasticities of demand for inputs vary with the plant's energy type consumption pattern.

This chapter provides evidence to support Solow's opinion that one needs to use caution when interpreting industry-level elasticity estimates. As discussed in Solow (1987), industry-level elasticity estimates reflect a combination of micro-level elasticities and within industry differences in the micro-level compositions of inputs. I observe large within industry differences in prices paid for energy types (see Table 3.2) and differences in the energy type consumption patterns of plants (see Section 3.4). While some of these differences in energy type consumption patterns across plants can be explained by industry, most cannot. In fact, all six major energy type consumption patterns for which I estimated elasticities are found in at least 14 of 20 2-digit industries.[95]

While the capital-energy debate is not the focus of this chapter, the simple KLEM model in Section 3.7.1 allows me to examine the nature of the relationship between capital and energy using microdata. I find statistically significant positive cross-price elasticities for capital and aggregate energy (Table 3.8) indicating that capital and aggregate energy are substitutes rather than complements.

The variation in energy type consumption patterns across plants is due, at least in part, to differences in the technologies in place at the plants. The work in this chapter indicates it would be interesting to model technology switching by looking at longitudinal data. There are several interesting questions to address. How frequently do plants change

---

[95] Specifically, the energy type consumption patterns of electricity and natural gas, electricity and oil, and electricity, oil, and natural gas are found in all 20 2-digit industries. The energy type consumption pattern of electricity only is found all but one 2-digit industry. Plants that consume all four energy types appear in 17 of 20 2-digit industries, and plants that consume electricity, natural gas, and coal appear in 14 of 20 2-digit industries.

technologies in a way that results in energy type consumption pattern changes? Do plants have the same energy type consumption pattern from birth? Or do they switch technologies over time? Is something like the type of machine replacement examined in Cooper and Haltiwanger (1993) for the auto industry going on for technology switches that affect the plant's energy type consumption pattern?[96] Addressing questions like these with MECS data presents a challenge since the number of continuers across MECS panels is relatively small.[97] Additionally, the continuers will be skewed towards larger energy users with higher probabilities of selection. However, such a study could still provide interesting results and is one possibility for future research.

The elasticity estimation work in this chapter is a first attempt to estimate demand elasticities for individual energy types and non-energy inputs with U.S. plant-level data. To the best of my knowledge, this is the first work estimating individual energy type elasticities with plant-level data that attempts to incorporate the effect of the plant's initial technology choice. I find that incorporating the effect of the plant's technology choice does affect the estimated elasticities. This chapter takes a somewhat simplistic approach to address the complications introduced into the estimation by the varying energy type consumption patterns of plants. Future work will make use of new methods, such as those described in Yen, Lin, and Smallwood (2003) and Golan, Perloff, and Shen (2001), to address these complications.

---

[96] Cooper and Haltiwanger (1993) examine the aggregate effects of machine replacement focusing on the auto industry. Annual machine replacement occurs in the auto industry when they revise and/or completely change the model of auto produced in the plant.

[97] For example, there are only 5,368 (38%) continuers from the 1994 to 1998 MECS. MECS micro data is available for 1985, 1988, 1991, 1994, and 1998. MECS micro data for 2002 will soon be available. The MECS was a subset of the ASM through 1991. In order to increase coverage of some large energy using industries, the MECS became a subset of the most recent Census of Manufactures (updated with plant births) in 1994.

Chapter 4

Prices and Quantities of Electricity in the U.S. Manufacturing Sector:

A Plant-Level Database and Public-Release Statistics, 1963-2000

4.1   Introduction

This chapter describes the Prices and Quantities of Electricity in Manufacturing

(PQEM) database, which contains plant-level observations on electricity purchases,

electricity prices and electricity suppliers for the U.S. manufacturing sector.[98] To

construct the database, we link plant-level data on electricity prices and quantities in the

Annual Survey of Manufactures (ASM) to information on electricity suppliers from the

Energy Information Administration and other sources. The resulting database contains

about 1.8 million annual observations over the period from 1963 to 2000. Table 4.1

summarizes the scope and coverage of the PQEM.

In constructing the PQEM, we devote considerable effort to treating anomalous

data on electricity prices and quantities in the ASM. We identify a number of coding

errors in the ASM data, and we find that the raw ASM data contain high error rates in

1983 and from 1989 to 1991. As explained below, we develop procedures to correct or

impute values for the erroneous data, paying special attention to the years with high error

rates. We also address several other measurement issues pertaining to ASM sample

weights in 1963 and 1967, erroneous geographic indicators, and the creation of consistent

industry and geography codes over time.

---

[98] This chapter draws heavily upon Davis, Grim, Haltiwanger, and Streitwieser (2006b).

Table 4.1: Scope and Coverage of the PQEM

A. Manufacturing Plants	
Plant-level electricity variables	Annual expenditures on purchased electricity, Annual electricity purchases (watt-hours), Price per unit of purchased electricity, Identity of electricity supplier
Years covered	1963, 1967, 1972-2000
Number of annual plant-level observations on the quantity and price of purchased electricity[a]	1,816,720
Number of plant-level observations per year, range	48,164 to 72,128
Number of counties with manufacturing plants	3,031
Other plant-level variables include state, industry, shipments, value added, fuel costs, employment, labor costs, and capital stock measures.	
B. Electricity Suppliers	
Electric Utilities: revenues from electricity sales by state and purchaser category (residential, commercial, industrial, municipal); list of counties served; indicator for whether the utility is investor-owned, publicly owned or a cooperative.	
Public Power Authorities: list of direct purchasers in the manufacturing sector.[b]	
Best-match Utilities: mean and dispersion of electricity prices paid by covered manufacturing plants, summary statistics on electricity purchases by covered plants, elasticity of price with respect to annual electricity purchases by covered plants, and other variables[c]	
Number of distinct best-match utilities that supply electricity to manufacturing plants[d]	362
C. State-Level Data on Electrical Power Sources	
Electricity generation from coal, petroleum and natural gas, hydropower, nuclear power, and other fuels, annually from 1960 to 2000	

Notes:

a. Our initial sample contains 1,945,813 records. We drop 107 records because of invalid geography codes and 128,058 (6.6%) because of missing values for electricity price, total employment, value added or shipments. We also trim the bottom 0.05% (five one-hundredths of one percent) of the electricity price distribution in each year (928 observations over all years).

b. We draw upon news articles and other public sources to compile a list of manufacturing plants that purchase electricity directly from the following six public power authorities: the Tennessee Valley Authority (TVA), the Bonneville Power Authority (BPA), the New York Power Authority (NYPA), Santee Cooper (SC), the Grand River Dam Authority (GRDA), and the Colorado River Commission of Nevada (CRC). Direct purchasers typically use large quantities of electricity, often operate their own transformers and sometimes obtain electricity for much lower prices than other plants.

c. Best-match utility is a constructed variable that reflects our efforts to assign a unique electricity supplier to each manufacturing plant in the database. We make these assignments according to the following algorithm. First, we compile a list of the electric utilities that supply a given county. Second, among utilities on that list, we determine which one accounts for the largest share of state-wide electricity sales to industrial customers. We designate this utility as the best-match utility serving industrial customers in the county. Third, we assign that utility to all manufacturing plants located in the county barring specific data to the contrary. For example, if the manufacturing plant is on our list of direct purchasers from a public power authority, we assign that power authority as the plant's best-match utility.

d. By construction, no best-match utility crosses state boundaries. Thus a single electric utility or public power authority can map to multiple best-match utilities. Among the 362 distinct best-match utilities in the database, 13 are accounted for by the 6 public power authorities.

To match manufacturing plants in the ASM to their electricity suppliers, we proceed as follows. First, we draw upon news articles and other public sources to compile a list of manufacturing plants that purchase electricity directly from one of six large public power authorities in the United States. Second, we rely on the Annual Electric Power Industry Report to identify the set of electric utilities that serve each county in the United States.[99] Third, among the utilities that supply electricity to a given county, we use the Annual Report to determine which one accounts for the largest share of electricity sales to industrial customers in the same state. We designate this utility as the best-match utility serving industrial customers in that county. Fourth, we assign that best-match utility to all manufacturing plants located in the county unless specific information to the contrary appears in our list of direct purchasers from public power authorities. By construction, no best-match utility operates in more than one state. We adopt this

---

[99] To the best of our knowledge, such data are unavailable prior to 1999, so we apply the data for 2000 to all years covered by the PQEM.

approach because of the important role played by state authorities in electric utility regulation and electricity pricing. Among the 362 distinct best-match utilities in the PQEM, 13 are accounted for by the 6 public power authorities.

Our procedure for assigning electricity suppliers to manufacturing plants is imperfect – only 460 of the 3,031 counties covered by the PQEM are supplied by a single electric utility.[100] Despite the potential for assignment errors, our procedure has proved useful for the study of spatial variation in electricity prices and for the study of differences among utilities in pricing schedules. See Chapter 1. We are currently attempting to use additional information on electric utility service areas to improve the assignment of manufacturing plants to their electricity suppliers in the PQEM. At this time, we have incorporated zip code level data on utility service areas for three states: California, Kentucky, and Ohio.[101] If possible, we plan to incorporate zip code level data for additional states.

---

[100] Note we exclude electric utilities with zero statewide industrial revenue. In the PQEM, 459 counties are served by a single utility, 776 are served by 2 utilities, 791 are served by 3 utilities, 536 are served by 4 utilities, 440 are served by 5-7 utilities, and the remaining 29 counties are served by 8-12 utilities.

[101] In the 2000 PQEM, California represents 5.3% of the quantity of purchased electricity, while Kentucky and Ohio represent 3.9% and 7.7%, respectively, of the quantity of purchased electricity. Note, as we did for counties served by multiple utilities, we assign the plants in zip codes served by multiple utilities to the utility with the most statewide revenues from sales to industrial customers. We consider a plant to have a well-defined best-match utility if it is a one-to-one county match, a one-to-one zip code match, or a hand-coded direct purchaser from a public power authority. Incorporating the zip code level data increased our percent of plants with well-defined best-match utilities to 70.0% from 10.4% in California, to 28.4% from 5.3% in Kentucky, and to 48.0% from 4.5% in Ohio. The percent of electricity purchases represented by plants with well-defined best-match utilities increased to 67.6% from 11.8% in California, to 44.5% from 35.4% in Kentucky, and to 38.0% from 5.8% in Ohio.

The plant-level data in the PQEM can be usefully grouped by state, county, industry, plant size, electricity price, electricity purchases, electricity supplier and other classification variables. We exploit these classification variables to produce a large number and variety of electricity price and quantity statistics for public release. Table 4.2 summarizes the contents of the PQEM public-release statistics, which will be made available for download once the statistics have passed U.S. Census Bureau disclosure requirements.[102]

---

[102] We are in the process of completing a disclosure analysis of the PQEM public release statistics. Some cells will be suppressed to maintain confidentiality.

Table 4.2: Contents of the PQEM Public-Release Tabulations

Classification Var.(s)	Public-Release Statistics
4-digit Industry	Mean and standard deviation of plant-level electricity prices and electricity purchases in logs and natural units, GWh and log GWh
2-digit Industry	Mean, standard deviation and quintile mean of plant-level electricity prices and electricity purchases in logs and natural units, GWh and log GWh
State	Same as 2-digit industry
County	Mean and standard deviation of electricity prices and electricity purchases in logs and natural units, GWh and log GWh
Employment size class	Same as county
Electricity purchases	Mean and standard deviation of plant-level electricity prices and electricity purchases in logs and natural units by quintiles of both the shipments-weighted electricity purchases distribution and the purchase-weighted electricity purchases distribution, GWh and log GWh
Value of shipments	Mean and standard deviation of plant-level electricity prices and electricity purchases in logs and natural units by centiles of the shipments distribution, GWh and log GWh
Best-match utility	Mean and standard deviation of electricity prices and quantities in logs and natural units, elasticity of sales price with respect to customer's annual purchases, number of counties in which the utility operates, number of counties for which the utility is designated as the best-match utility and average number of utilities that supply electricity in those counties plus indicators for the state in which the best-match utility operates and its ownership type (private, public or cooperative)
State and power source	Electricity generation from coal, petroleum and natural gas, hydropower, nuclear power, and other fuels (1960 to 2000)
Annual	Mean and standard deviation of plant-level electricity prices and electricity purchases in logs and natural units by centiles of the shipments distribution, GWh and log GWh; Coefficients from fifth-order polynomial fit of log price of electricity on log purchases with and without utility fixed effects.

Notes: The public-release statistics are computed from the PQEM database for 1963, 1967 and annually from 1972 to 2000. Part-year observations are excluded from the PQEM. Most statistics are computed with weighting by the value of shipments and with weighting by the quantity of purchased electricity. Statistics are suppressed in certain cells to maintain confidentiality.

Tables 4.3 - 4.6 report selected summary statistics on electricity prices and quantities for manufacturing plants and best-match utilities. Not surprisingly, Table 4.3 shows tremendous heterogeneity among manufacturing plants in the quantity of purchased electricity. Of perhaps greater interest, Table 4.3 also reveals large differences in electricity prices among manufacturing plants. For example, the price paid by manufacturing plants in 1963 ranges from 3.45 cents per KWh at the 10th centile of the shipments-weighted distribution to 8.82 cents at the 90th centile. Similarly, Tables 4.4 - 4.6 show considerable variation among utilities in average electricity prices. Figure 4.1 highlights large changes over time in the dispersion of electricity prices among manufacturing plants, a phenomenon that we study in some detail in Chapter 2.

Electricity's percentage of variable costs, intermediate input costs, and total energy costs all rose from 1963 to 2000 as shown in Figure 4.2.[103] Electricity represents the majority of total energy costs for the U.S. manufacturing industry from 1963 to 2000. The percentage of total energy costs accounted for by electricity rose dramatically from 57% in 1979 to 69% in 1987. Electricity costs are a much smaller percentage of variable costs and intermediate input costs. At the maximum, in 1983, electricity costs only account for 2.0% of variable costs. The maximum percent of intermediate input costs represented by electricity is 3.4% in 1986.

---

[103] Variable costs are defined as *CP* (cost of materials and parts) + *CR* (cost of resales) + *CW* (cost of contract work) + *EE* (electricity expenditures) + *CF* (cost of fuels) + *SW* (salaries and wages). Intermediate input costs are defined as *CP* + *CW* + *EE* + *CF*. Energy costs are defined as *EE* + *CF*.

Table 4.3: Electricity Prices and Quantities in the PQEM,
Selected Summary Statistics for Manufacturing Plants

	Year(s)	Weighted by Plant Output (Value of Shipments)	Weighted by Plant Electricity Purchases
Mean annual electricity purchases, GWh	All	99.76	860.39
Standard deviation of annual purchases	All	333.97	2,399.99
Mean real price of purchased electricity, 1996 cents per kilowatt-hours (KWh)	1963	5.93	3.88
	2000	5.47	4.44
Standard deviation of real electricity prices, 1996 cents per KWh	1963	2.87	2.16
	2000	2.11	1.80
Standard deviation of log real electricity prices	1963	0.409	0.524
	2000	0.360	0.383

Quantiles of Annual Electricity Purchases (GWh), Shipments Weighted, All Years								
1	5	10	25	50	75	90	95	99
.07	.30	.70	3.22	16.37	89.23	267.35	443.94	1,499.81

Quantiles of Real Electricity Prices in 1996 cents per KWh, Shipments Weighted, 1963								
1	5	10	25	50	75	90	95	99
2.01	2.97	3.45	4.14	5.30	7.20	8.82	10.74	18.88

Quantiles of Real Electricity Prices in 1996 cents per KWh, Shipments Weighted, 2000								
1	5	10	25	50	75	90	95	99
2.26	2.86	3.30	3.99	5.08	6.59	8.22	9.21	13.31

Notes: Statistics calculated from the PQEM database. For disclosure reasons, the quantiles shown above are averages of plant-level observations in three quantiles, the quantile shown and the two surrounding quantiles (e.g., quantile 50 as shown is the average of observations in quantiles 49, 50, and 51).

Table 4.4: Electricity Prices and Quantities in the PQEM,
Selected Sample-Weighted Summary Statistics for Best-match Utilities

	Year(s)	Unweighted	Weighted by Electricity Sales (GWh)
Mean annual electricity sales, GWh	All	2,048	10,873
Standard deviation of best-match utility mean price per KWh in 1996 cents	1963	2.22	2,310.95
	2000	1.61	2,242.21
Standard deviation of best-match utility mean price per KWh in 1996 cents, excluding public power authorities	1963	2.02	769.83
	2000	1.58	2,024.78

	Mean	25th centile	Median	75th centile
Number of covered manufacturing plants per best-match utility, excluding public power authorities				
	184	7	24	127

Quantiles of Annual Electricity Sales (GWh) by Best-match Utilities, Unweighted, All Years								
1	5	10	25	50	75	90	95	99
0.19	1.74	5.95	34.61	237.55	1.79E3	6.11E3	1.25E4	2.05E4

Quantiles of Annual Electricity Sales (GWh) by Best-match Utilities, Sales Weighted, All Years								
1	5	10	25	50	75	90	95	99
1.68E2	8.62E2	1.62E3	4.07E3	1.04E4	1.69E4	2.07E4	2.23E4	2.79E4

Quantiles of Mean Electricity Prices in 1996 cents per KWh, Sales Weighted, 2000								
1	5	10	25	50	75	90	95	99
2.26	3.30	4.16	4.84	5.73	6.57	7.68	8.27	8.79

Notes: Statistics calculated from the PQEM database. Best-match utility level statistics are ASM sample-weighted. Number of covered plants per best-match utility is not weighted.

Table 4.5: Electricity Prices and Quantities in the PQEM,
Selected Shipments-Weighted Summary Statistics for Best-match Utilities

	Year(s)	Unweighted	Weighted by Electricity Sales (GWh)
Mean annual electricity sales, GWh	All	6.65E8	6.70E9
Standard deviation of best-match utility mean price per KWh in 1996 cents	1963	2.05	4.19E5
	2000	1.78	1.38E6
Standard deviation of best-match utility mean price per KWh in 1996 cents, excluding public power authorities	1963	1.96	2.18E5
	2000	1.76	1.34E6

	Mean	25th centile	Median	75th centile
Number of covered manufacturing plants per best-match utility, excluding public power authorities	184	7	24	127

Quantiles of Annual Electricity Sales (GWh) by Best-match Utilities,
Unweighted, All Years

1	5	10	25	50	75	90	95	99
135	6.22E3	4.52E4	7.22E5	1.50E7	2.84E8	1.78E9	3.63E9	1.02E10

Quantiles of Annual Electricity Sales (GWh) by Best-match Utilities,
Sales Weighted, All Years

1	5	10	25	50	75	90	95	99
8.10E7	3.82E8	8.01E8	2.10E9	5.18E9	9.34E9	1.55E10	1.82E10	2.14E10

Quantiles of Mean Electricity Prices in 1996 cents per KWh, Sales Weighted, 2000

1	5	10	25	50	75	90	95	99
2.57	3.21	3.69	4.36	4.78	5.38	6.96	7.10	8.06

Note: Statistics calculated from the PQEM database. Best-match utility level statistics are shipments-weighted. Number of covered plants per best-match utility is not weighted.

Table 4.6: Electricity Prices and Quantities in the PQEM,
Selected Purchase-Weighted Summary Statistics for Best-match Utilities

	Year(s)	Unweighted	Weighted by Electricity Sales (GWh)		
Mean annual electricity sales, GWh	All	1.76E9	1.03E11		
Standard deviation of best-match utility mean price per KWh in 1996 cents	1963	2.14	8.09E5		
	2000	1.69	1.11E6		
Standard deviation of best-match utility mean price per KWh in 1996 cents, excluding public power authorities	1963	2.08	6.60E5		
	2000	1.68	9.52E5		
Number of covered manufacturing plants per best-match utility, excluding public power authorities	Mean	25th centile	Median		75th centile
	184	7	24		127

Quantiles of Annual Electricity Sales (GWh) by Best-match Utilities, Unweighted, All Years								
1	5	10	25	50	75	90	95	99
17.5	1.01E3	9.75E3	2.58E5	7.54E6	2.24E8	1.61E9	4.74E9	2.81E10

Quantiles of Annual Electricity Sales (GWh) by Best-match Utilities, Sales Weighted, All Years								
1	5	10	25	50	75	90	95	99
2.18E8	1.19E9	2.73E9	1.14E10	7.08E10	1.77E11	2.62E11	2.90E11	2.95E11

Quantiles of Mean Electricity Prices in 1996 cents per KWh, Sales Weighted, 2000								
1	5	10	25	50	75	90	95	99
2.14	2.17	2.41	2.48	2.72	4.01	4.57	4.92	6.31

Note: Statistics calculated from the PQEM database. Best-match utility level statistics are purchase-weighted. Number of covered plants per best-match utility is not weighted.
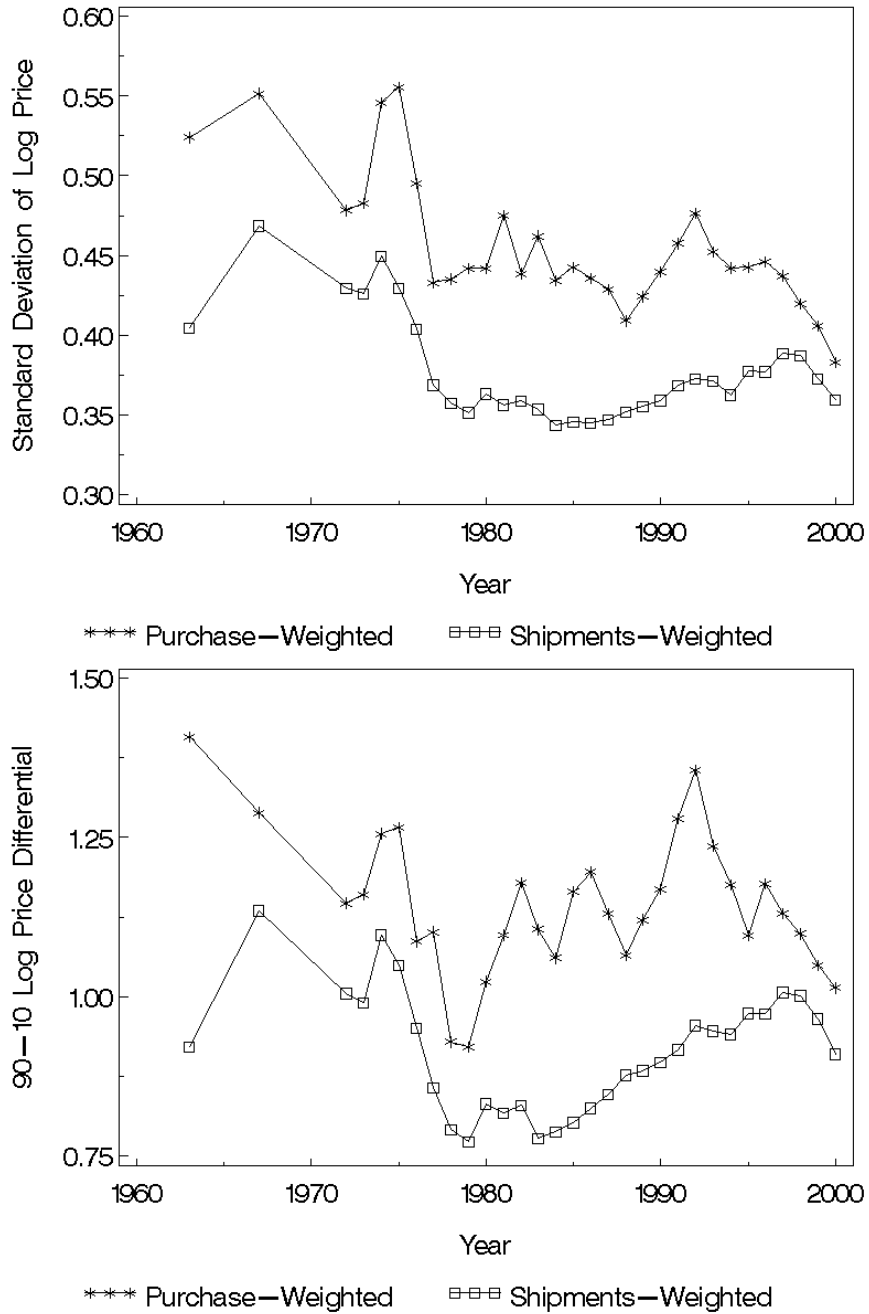
Source: Authors' calculations on PQEM data.

Figure 4.1: Electricity Price Dispersion Among U.S. Manufacturing Plants, 1963-2000
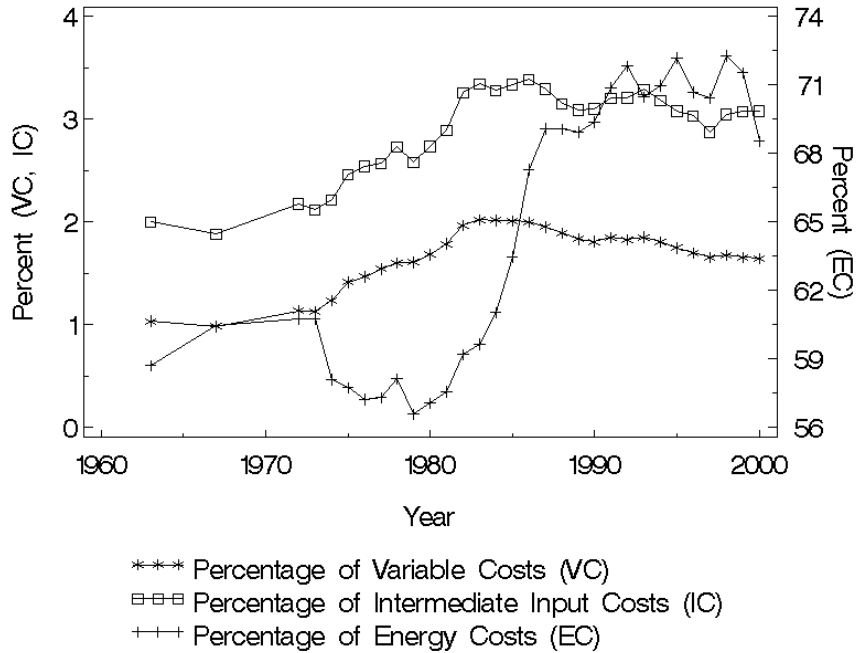
Figure 4.2: Shipments-Weighted Annual Electricity Percentage of Variable Costs, Intermediate Input Costs, and Energy Costs, 1963-2000

The paper proceeds as follows. Section 4.2 describes the data sources for the PQEM, and Section 4.3 describes the creation of industry codes in the PQEM. Section 4.4 describes the identification of ASM plants in 1967 and the creation of ASM sample weights for 1963 and 1967. Section 4.5 describes the geography codes in the PQEM. Section 4.6 discusses the construction of electricity purchase level variables. Sections 4.7 and 4.8 explain the imputations we used for observations with unreasonable electricity prices. Section 4.9 discusses our method of handling of total value of shipments outliers in the PQEM. Finally, Section 4.10 provides brief conclusions.

4.2 Sources for the PQEM Database

4.2.1 The Annual Survey of Manufactures

The ASM is a large sample of manufacturing plants with five or more employees. Larger plants are sampled with certainty, and sampling probabilities for other plants vary inversely with plant size. The ASM gathers data on roughly 60,000 plants per year, which jointly account for about three-quarters of employment in the manufacturing sector. ASM data are available in electronic form from 1973, and similar data are available from the Census of Manufactures (CM) starting in 1963 and 1967.[104] See the technical appendix of Davis, Haltiwanger and Schuh (1996) for a detailed discussion of ASM coverage and sampling procedures.

4.2.2 The Annual Electric Power Industry Report

The Energy Information Administration collects data from participants in the electric power industry through its EIA 861 form, also known as the Annual Electrical Power Industry Report. Electric utilities, wholesale power marketers, energy service providers and electric power producers are required by law to complete the EIA 861 form. The Annual Report includes data on revenues from electricity sales to industrial customers by state for each utility and a list of the counties served by the utility.[105] It also

---

[104] The ASM and Census files we use to construct the PQEM are the versions maintained in the Center for Economic Studies Data Warehouse as of March 2005. The Center for Economic Studies is part of the U.S. Census Bureau. Plant-level data in the ASM, Census and PQEM can only be accessed at the Center and the several U.S. Census Bureau Research Data Centers located in the United States. Information on the process for accessing the plant-level data is available at http://www.ces.census.gov/.

[105] The Annual Report also reports the North American Reliability Council (NERC) regions in which a utility operates. NERC regions do not necessarily follow county or state boundaries. See pages 6-7 of EIA (1998) for more information on NERC regions.

specifies whether a utility is an investor-owned, public or cooperative enterprise.[106] Electronic versions of the Annual Report are available from the Energy Information Administration at www.eia.doe.gov/cneaf/electricity/page/eia861.html.

4.2.3   Direct Purchasers from Public Power Authorities

According to the 2000 Annual Electric Power Industry Report (EIA 861 file), there are 35 public power authorities in the U.S., 14 of which have positive industrial revenue. Manufacturers that purchase directly from public power authorities typically consume large quantities of electricity, often operate their own transformers, and often obtain electrical power at lower prices than other plants. These direct purchasers are few in number, but they account for a large fraction of electricity purchases in some counties, and they constitute a distinct demand-side segment of the retail electricity market. For these reasons, we sought to separately identify direct purchasers. Unfortunately, the EIA 861 data file does not contain county coverage information for all public power authorities, nor does it contain other information that identifies direct purchasers.

To address this issue, we draw on a variety of public sources that specifically identify direct purchasers from the four public power authorities with the largest industrial revenues in 2000: the Tennessee Valley Authority (TVA), the Bonneville Power Administration (BPA), Santee Cooper (SC), and the New York Power Authority (NYPA). Additionally, we identify direct purchasers from the Grand River Dam

---

[106] The eight ownership categories in the EIA 861 data file are: private, power marketer, cooperative, federal, municipal, sub-division, municipal marketer, and state. The EIA 861 survey respondent provided this ownership information. Note that utilities with power marketer and municipal marketer ownership categories do not appear in the PQEM because these types of utilities do not sell power directly to end-use customers.

Authority (GRDA) and the Colorado River Commission of Nevada (CRC).[107] In 2000, these six public power authorities account for nearly 98% of all public power authority industrial revenue and 3% of all utility revenue from industrial customers (source: 2000 EIA 861 files). We identify an average of 84 plants per year that are direct serve customers of one of our six major public power authorities.[108] Pooling across years in the PQEM, the direct-serve customers of our six major public power authorities account for 3.6% of all electricity purchases (on a shipments-weighted basis). The shipments-weighted mean of electricity purchases for direct-serve customers of our six public power authorities is 1,030.58 GWh.

4.2.4 Electric Utility Customers

We link plants that are not direct serve customers of public power authorities to their electric utilities using the 2000 EIA 861 file. We rely on the EIA 861 file to identify the set of electric utilities that serve each county in the United States. First, we use the EIA 861 file to determine which electric utility accounts for the largest share of electricity sales to industrial customers in the same state. We designate this utility as the best-match utility serving industrial customers in that county. By construction, no best-match utility operates in more than one state.

---

[107] Specifically, we use public information to identify large aluminum smelting plants in the Pacific Northwest as direct purchasers from the BPA. For TVA, because it has a huge service area, we concentrate on certain industries (Aluminum, Steel, Other Primary Metals, Chemicals, Paper and Forest, Motor Vehicles and Motor Equipment) when searching for direct serve customers. We do not focus on specific industries when identifying direct-serve industrial customers for the other public power authorities. For CRC and NYPA, we found publicly available lists of direct-serve power customers. A detailed technical note with a full description of our sources and methods for identifying direct purchasers is available to researchers who have permission to access the PQEM micro data.

[108] The number of plants identified as public power authority customer plants ranges between 58 and 94 plants per year.

As discussed earlier, our procedure for assigning electricity suppliers to manufacturing plants is imperfect because many counties are served by multiple electric utilities. We are currently attempting to use additional information on electric utility service areas to improve the assignment of manufacturing plants to their electricity suppliers. At this time, we have incorporated zip code level data on utility service areas for three states: California, Kentucky, and Ohio. If possible, we plan to incorporate zip code level data for additional states.

Introduction of the zip code level data for California, Kentucky, and Ohio allows us to calculate an overall estimated match accuracy rate for the PQEM. Let Z denote the set of plants with information on utility service territories by zip code, and let $^{\wedge}Z$ be the complement; i.e., all other plants. Let $n(\cdot)$ be a function that returns the number of utilities serving industrial customers in the county according to the EIA-861 files. Lump all counties served by five or more utilities into a single category, called "5+". Define $N(i)$ as the number of ASM-sampled plants for which $n(e) = i$ for $i = 1, 2, 3, 4, 5+$. Define $N(i; Z)$ and $N(i; ^{\wedge}Z)$ as analogous quantities for Z and $^{\wedge}Z$. For present purposes, treat customers that purchase directly from public power authorities as belonging to a county served by a single utility.

First, for plants in Z, we construct the share of plants that are accurately matched by county-level matching in counties served by i utilities as shown in (4.1).

$$p(i) = \left[N(i; Z)\right]^{-1} \sum_{e \in \{Z:\, n(e)=i\}} I(e) \quad \text{for } i = 1, 2, 3, 4, 5+. \tag{4.1}$$

where $I(\cdot)$ is an indicator function that returns a value of one if our county-based matching algorithm produces a correct match, and zero otherwise. We then construct the

share of plants in counties served by *i* utilities that cannot be matched using zip code

matching, scaled down by the inverse of *p(i)* as shown in (4.2).

$$q(i) = [N(i;Z)]^{-1} \sum_{e \in \{Z:\, n(e)=i\}} \tilde{I}(e)p(z(e)) \quad \text{for i = 1, 2, 3, 4, 5+} \tag{4.2}$$

where $z(e)$ is the number of utilities that serve the zip code in which *e* is located and $\tilde{I}(\cdot)$

returns a value of 1 if zip code information does not unambiguously determine the correct

match, and if the utility selected by the county-level algorithm is among the set serving

the zip code, and 0 otherwise. That is, $\tilde{I}(\cdot)$ returns a value of zero even when zip code

level information is ambiguous, provided that the subset of possible utility matches

excludes the utility selected by the county-based algorithm.

Finally, we construct the overall estimated rate of an accurate match, $\hat{A}$, as shown

in (4.3).

$$\hat{A} = \left(\sum_{i=1}^{5+} \sum_{E \in \wedge Z} s(e)[p(i)+q(i)] \right) + \left(\sum_{e \in Z} s(e)p(z(e)) \right) \tag{4.3}$$

where the *s(e)* are plant-level weights that sum to one. The first term on the right side of

(4.3) captures plants with county-level matching, and the second term captures plants

with zip code matching. We calculate sample-weighted, shipments-weighted, and

purchase-weighted versions of $\hat{A}$. Table 4.7 shows utility match statistics for the year

2000. The sample-weighted overall estimated match accuracy rate for 2000 is 78%.

Table 4.7: PQEM Utility Match Statistics, 2000

	All	Number of Possible Utility Matches				
		1	**2**	**3**	**4**	**5+**
Percent of Total Plants	-	25.4	23.0	21.8	13.6	16.2
Percent of Plants in Z	18.6	59.6	28.0	9.0	2.9	0.5
Percent of Plants in ^Z	81.4	17.6	22.8	24.7	16.1	19.8

	County-Level Utility Match Category (i)				
	1	**2**	**3**	**4**	**5+**
$p(i)$	1.00	0.70	0.46	0.38	0.30
$q(i)$	0.00	0.16	0.24	0.22	0.27

	Sample Weighted	Purchase Weighted	Shipments Weighted
$\hat{A}$	0.78	0.75	0.76

Source: Authors' calculations on the PQEM.

Notes:

1. The percent of plants values are calculated on a completely unweighted basis.

2. $p(i)$ is the percentage (divided by 100) of PQEM plants accurately matched by county-level matching in counties served by i utilities. The $p(i)$ values are calculated on a completely unweighted basis. There are a small number of our one-to-one county matches that would be assigned a different utility using the one-to-one zip match. Therefore, we must assign all plants with one-to-one county matches a value of 1 for $I(\cdot)$ to ensure the condition $p(1) = 1$ is met.

3. $q(i)$ is the percentage (divided by 100) of PQEM plants in counties served by i utilities that cannot be uniquely matched using zip code matching, scaled down by the inverse of the predicted percent correct for the number of utilities that serve the plant's zip code. The $q(i)$ are calculated on a completely unweighted basis. All plants with one-to-one county matches do not also have one-to-one zip code matches. We assign all plants with one-to-one county matches a value of 0 for $\tilde{I}(\cdot)$ to ensure the condition $q(1) = 0$ is met.

4.2.5 State-Level Data on Power Sources for Generated Electricity

The Energy Information Administration makes annual state-level data on fuels and other sources of power for the generation of electricity available to the public. These data are available in the State Energy Data System (SEDS) for all 50 states and the District of Columbia from 1960 to 2000.

We use the SEDS data to create annual state-level shares of fuels used for electricity generation for the following five categories: coal, petroleum and natural gas, hydropower, nuclear power, and other (includes geothermal, wind, wood and waste, photovoltaic, and solar).[109] Figure 4.3 shows the annual national fuel shares of electricity generation calculated from the SEDS data. Coal is used to generate more electricity than any other fuel in every year from 1960 to 2000. Also note, "other" fuels account for very little electricity generation throughout the entire period.

Figures 4.4 - 4.8 show comparisons of national fuel shares for electricity generation calculated from SEDS data and fuel shares calculated from Table 7.2b of EIA (2003c).[110] The fuel shares we calculate from SEDS data follow the pattern of and are close in magnitude to the EIA Table 7.2b data for all fuels except "other". Also of note, the gap between the SEDS-based data and the EIA Table 7.2b data gets larger in 1989 because the SEDS data does not contain non-utility generation for any fuels other than

---

[109] We use the SEDS updated through 2000. Current SEDS data is available at http://www.eia.doe.gov/emeu/states/_seds.html. A detailed technical note on the creation of fuel shares from the EIA SEDS data is available from the authors upon request.

[110] Table 7.2b can be found in EIA (2003c). Note the fuel shares shown in Figures 4.4 - 4.8 are calculated simply by adding up kWh generated by the fuels in each of the five categories and dividing by "total" kWh generated.

hydropower. There is a particularly large difference for "other" fuels because power

generated by wind, geothermal, and other sources is, in large part, generated by non-

utilities. However, since the "other" category is such a small percentage of electricity

generation, we do not address this difference.
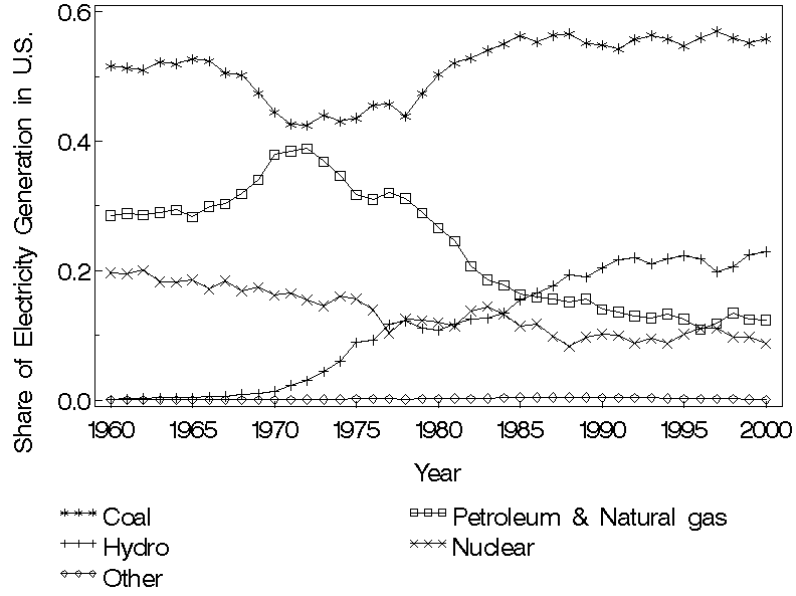
Source: Authors' calculations from EIA (2003c) state-level data.

Figure 4.3: Fuel Shares of Electricity Generation in the U.S.
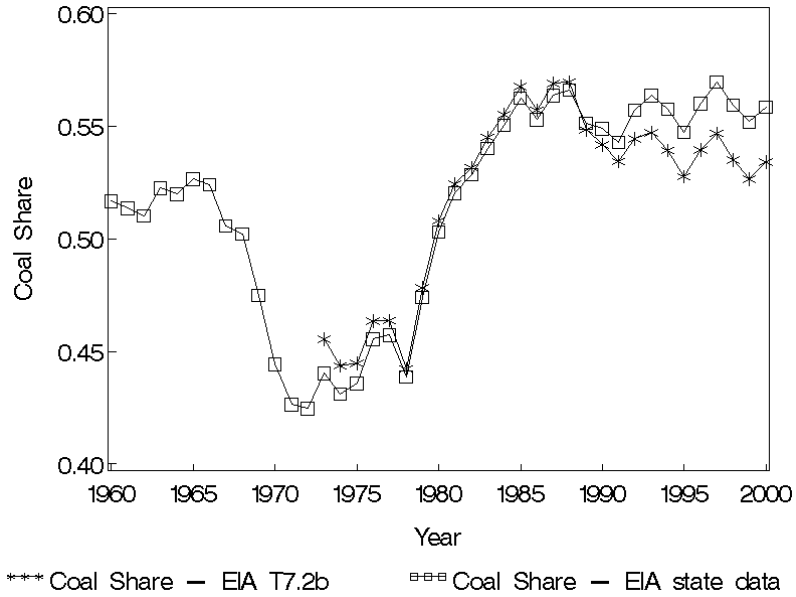
Sources: Authors' calculations from EIA SEDS data and authors' calculations from EIA (2003c) Table 7.2b.

Figure 4.4: Coal Fuel Shares of Electricity Generation in the U.S.
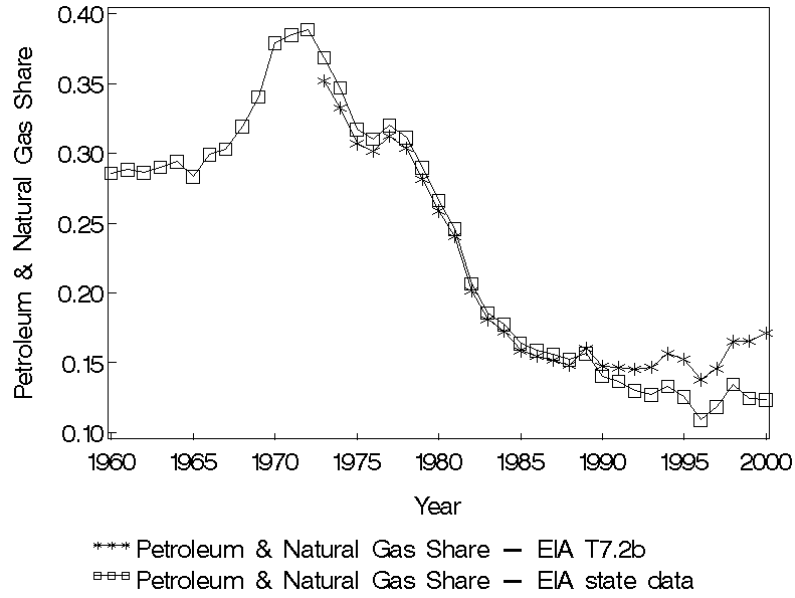
Sources: Authors' calculations from EIA SEDS data and authors' calculations from EIA (2003c) Table 7.2b.

Figure 4.5: Petroleum and Natural Gas Fuel Shares of Electricity Generation in the U.S.
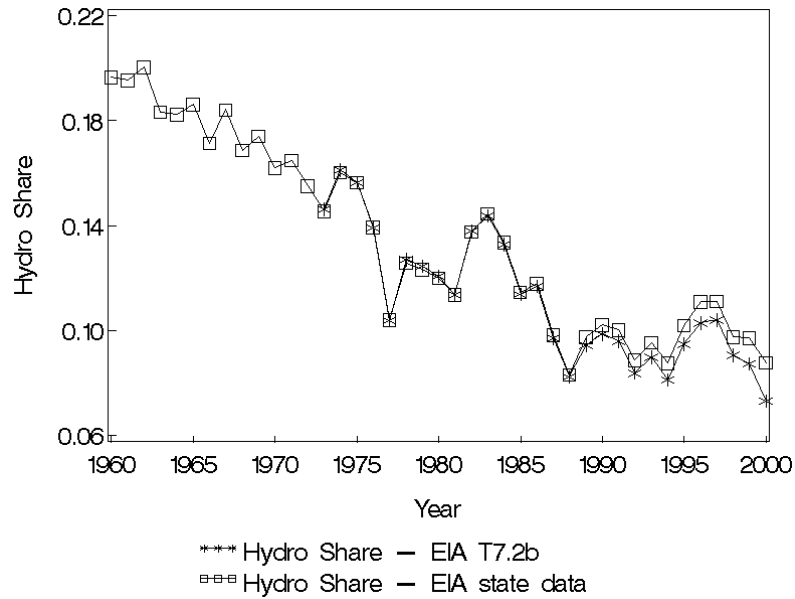
Sources: Authors' calculations from EIA SEDS data and authors' calculations from EIA (2003c) Table 7.2b.

Figure 4.6: Hydro Fuel Shares of Electricity Generation in the U.S.
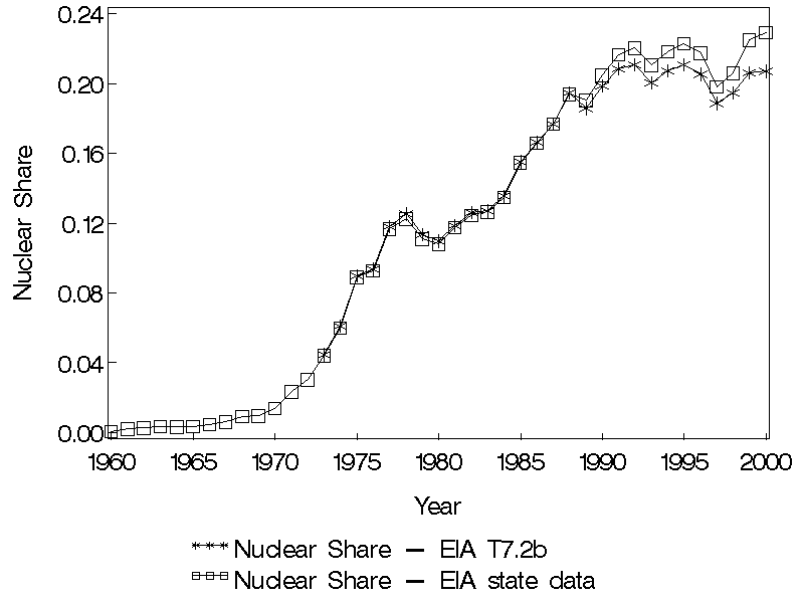
Sources: Authors' calculations from EIA SEDS data and authors' calculations from EIA (2003c) Table 7.2b.

Figure 4.7: Nuclear Fuel Shares of Electricity Generation in the U.S.
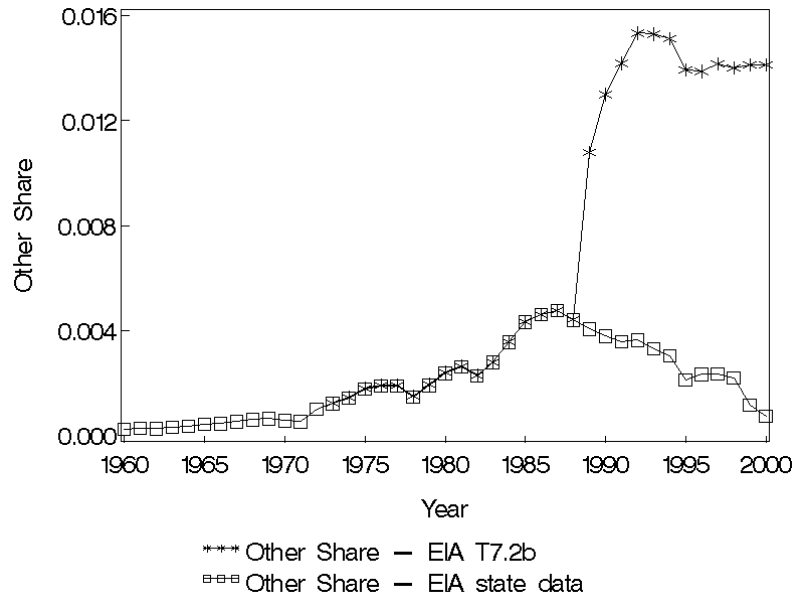
Sources: Authors' calculations from EIA SEDS data and authors' calculations from EIA (2003c) Table 7.2b.

Figure 4.8: Other Fuel Shares of Electricity Generation in the U.S.

4.3   Industry Codes

The PQEM contains a revised industry variable, *newind*. The industry codes in the variable *newind* are 4-digit SIC industries. We use a combination of 1972 and 1987 SIC codes across years. Years prior to 1987 are coded with 1972 SIC codes, and years 1987 and later are coded with 1987 SIC codes.[111] Overall, the PQEM contains 447 1972 SIC codes and 458 1987 SIC codes.

We want researchers to have the option of using the NBER-CES producer price indices in conjunction with the PQEM.[112] Therefore, any industries that cannot be matched to the NBER-CES producer price index database are considered invalid.[113] If possible, corrections are made to invalid industry codes using data from surrounding years. Observations for which corrections cannot be made are dropped. As can usually be expected with microdata, there are some complexities involved in creating the *newind* variable from the existing ASM industry variables. Appendix G contains a detailed description of how we create the *newind* variable.

---

[111] This does not cause problems for analysis as long as all analyses involving the *newind* variable are either done by year or for groups of years that do not straddle 1987.

[112] The NBER-CES producer price indices are available at http://www.nber.org/nberces/nbprod96.htm. There are price indices for each year and 4-digit SIC code.

[113] We also drop industry 2794, which is in the 1972 NBER-CES database, because it is listed as a discontinuous industry in Davis, Haltiwanger, and Schuh (1996). We follow Davis, Haltiwanger, and Schuh (1996) and recode industry 2794 into industry 2793. We should also note that, for confidentiality reasons, Census combined industry 2067 (Chewing Gum) into industry 2064 (Candy and Other Confectionary Products). Therefore the industry 2067 does not exist in the PQEM.

4.4 Identification of Annual Survey of Manufactures Plants in Census of Manufactures Years and Sample Weights

The *ET* (establishment type) variable is used to identify ASM plants in all years except for 1967. *ET* = 0 indicates a plant is an ASM plant. Unfortunately, there is no simple way to identify ASM plants in the 1967 CM. The variable *ET* is equal to one for all observations on the 1967 CM file. We combine information on certainty cases from the 1967 CM publication and information on typical imputation patterns for non-ASM plants to identify ASM plants in the 1967 CM. Appendix H describes our methodology for identifying ASM plants in the 1967 CM in detail.

As discussed in Section 4.2.1, the ASM is a sub-sample of the universe of manufacturing plants. Hence, sample weights are required to create nationally representative statistics. Generally, ASM sample weights are available on the ASM and CM microdata files in the variable *WT*. However, ASM sample weights are not available on the 1963 and 1967 CM plant-level data files. Appendix I describes the method we use to estimate ASM sample weights for the 1963 and 1967 CM surveys.

4.5 Geography Codes

The PQEM contains a time consistent Federal Information Processing Standards (FIPS) combined county-state identifier, *cyfstn*.[114] Unfortunately, there are problems with the FIPS county and state code identifiers on some of the ASM and CM microdata files. Further, FIPS county codes have changed over time. Appendix J describes changes we

---

[114] The National Institute of Standards and Technology (NIST) issues FIPS codes. NIST publications related to FIPS codes can be found at http://www.itl.nist.gov/fipspubs/.

make to the state and county FIPS codes to make them correct and consistent across time in the PQEM.

4.6 Creation of Purchase Level Variables

In addition to log purchased electricity itself, we create four variables to measure the electricity purchase levels of plants by where they fit into the distribution of purchased electricity. First, we compute two unweighted decile and centile purchase level variables to use in our imputation procedures. Then, we compute shipments-weighted decile and centile purchase level variables for use in our analysis.

We compute the unweighted distribution of log purchased electricity across all plants by year. Then we compute deciles of purchased electricity and assign each plant-year observation a value in the variable, *decile_uu,* based on where it fits into the distribution. A value of *decile_uu* equal to one implies that the plant's quantity of purchased electricity for that year falls in the lowest decile of users for that year. Additionally, we create a variable, *centile_uu*, representing the unweighted electricity purchase level centile of the plant in that year. We create this variable by ordering plants from lowest to highest quantity of purchased electricity and dividing them into 100 equally sized groups.[115] A value of *centile_uu* equal to one implies that the plant's quantity of purchased electricity for that year falls into the group containing the smallest electricity purchasers in that year.

Further, we create two measures of where a plant fits in the shipments-weighted distribution of electricity purchases. For one measure, we pool observations over all

---

[115] We also created a centile variable based on percentiles of the distribution by year. The results were very similar. The use of equally sized groups allows us to avoid potential disclosure problems.

plants within a year and compute shipment-weighted deciles of the resulting distribution of electricity purchases. We then assign each plant-year observation a decile rank from 1 to 10 based on where it fits in the pooled distribution for that year and call this variable *decile*. For a second shipments-weighted measure, we assign centile ranks from 1 to 100 to the variable *centile* in the same manner.

4.7   Imputations for Observations with Unreasonable Electricity Prices

This section describes our procedures for identifying and imputing values for electricity variables for observations with unreasonable electricity prices, purchases, and/or expenditures. We impute values for the following two electricity variables: EE = electricity expenditures and PE = quantity of purchased electricity. Annual electricity prices are calculated as EE divided by PE.

We employ two data filters to identify observations with unreasonable values of the electricity variables. The first data filter identifies observations with specific problems we identified in the electricity variables. We apply an imputation algorithm to the observations identified as problem observations by the first data filter. Following this, we apply a second, more general, data filter to the electricity variables to identify remaining problem observations. We use the same imputation algorithm to correct as many of these problem observations as possible.

4.7.1 Data Filter 1: Specific Problems

Data Filter 1 marks observations with one or more of the following characteristics as problem observations.

- The observation has a real price of electricity $\geq \$0.30$ per kWh.[116]

- The observation has a nominal price of electricity equal to exactly $0.001 per kWh.

- The observation has missing or zero *PE* and/or *EE*.

There is a major data anomaly in the 1983 ASM. The variables *EE* and *PE* are equal for many plants. Although this phenomenon occurs across all purchase deciles, it occurs primarily with plants that are smaller, where size is defined in terms of their amount of purchased electricity. Ninety percent of all *EE = PE* observations occur in the three smallest plant unweighted purchase deciles, representing eight percent of all small plants.[117] Nearly 1,500 plants (3 percent of all ASM plants in 1983) have a nominal price of electricity equal to $1 per kWh (i.e., *EE = PE* for that plant) in 1983. Other years have only a trivial number of occurrences: less than 0.2 percent.[118] Further, 1983 contains a relatively large number of plants with real price of electricity greater than 30 1996 cents

---

[116] Nominal prices were deflated to real 1996 prices using the annual implicit GDP deflator from the Bureau of Economic Analysis.

[117] We create an unweighted purchase decile variable, *decile_uu*, as described in Section 4.6 for use in our imputation process.

[118] There are also observations for which the price of electricity is missing. These are observations for which *PE* and/or *EE* are zero or missing so the price of electricity could not be calculated.

per kWh. According to EIA data, these prices are very high; average industrial electricity prices were $0.0485 per kWh in 1983.[119]

For the purpose of examining and correcting the log price of electricity in the PQEM, we consider all plants with a real price of electricity $\geq$ $0.30 per kWh to be "problem" plants. We feel this is a reasonable cutoff for problem plants for a couple of reasons. First, this cutoff is six times the highest average U.S. industrial electricity price as reported by the EIA. Second, a quick look at the Manufacturing Energy Consumption Survey (MECS) plant-level data for 1988, 1991, 1994 and 1998 shows that less than one fifth of one percent of the plants have real electricity prices $\geq$ $0.30 per kWh.

There are over 2,900 plants in 1983 with real electricity price $\geq$ $0.30 per kWh. This is close to 6 percent of all plants in the 1983 ASM sample. While this may not seem like a large percentage, you can see in Figure 4.9 that a small percentage of plants can have a big effect on important characteristics of the data, such as the standard deviation of log electricity prices. Specifically, Figure 4.9 shows the completely unweighted standard deviation of the log price of electricity from the 1963, 1967 and 1972-2000 ASM. It is easy to see that the standard deviation of the log price of electricity in 1983 is disproportionately high relative to surrounding years. Further, problem observations with a real price of electricity $\geq$ $0.30 per kWh are not uniformly distributed among levels of energy users. Approximately 15 percent of plants in the three smallest plant purchase deciles have a real price of electricity $\geq$ $0.30 per kWh in 1983. Thus, failure to correct for problems with the *EE* and *PE* variables could lead to especially poor results for plants that purchase relatively smaller amounts of electricity.

---

[119] Annual average electricity prices by class of customer can be found on the EIA Internet site: http://www.eia.doe.gov.

Additionally, there are close to 40 observations among all years with nominal price of electricity equal to exactly $0.001 per kWh. Looking at the data shows that *PE* is simply *EE* multiplied by 1000. Since this is quite unlikely to be true, these observations will also be considered "problem" observations. Finally, all observations with missing or zero *PE* and/or *EE* will be considered "problem" observations.
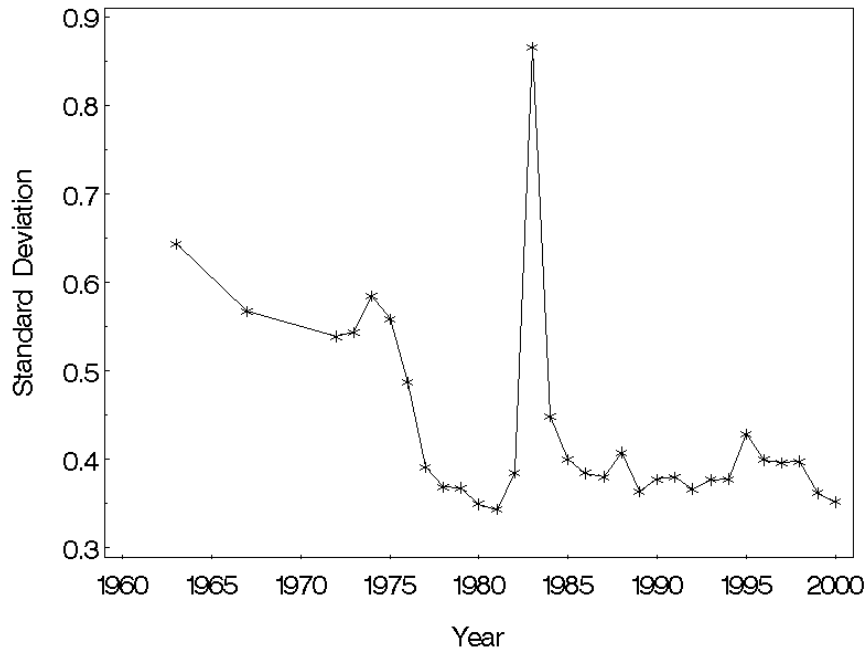
Source: Author's calculations on 1963-2000 Annual Survey of Manufactures. Statistics are not weighted.

Figure 4.9: Standard Deviation of the Log Price of Electricity Prior to Applying the Data Filters and Correction Algorithms Described in Section 4.7, 1963-2000

4.7.2 Data Filter 2: General Outliers

After we create "corrected" log price of electricity values for the maximum possible number of problem observations identified by Data Filter 1, we apply a second data filter to catch general outliers. An observation is a general outlier according to Data Filter 2 if either of the following are true.

- The log price of electricity is both a time series and cross-sectional outlier within its' best-match utility area and state.[120]

- *PE* in year *t* is greater than or equal to the 75th percentile of *PE* for year *t* and *PE* in year *t* is at least 10 times greater than *PE* in year *t*-1 and *PE* in year *t* is at least 10 times greater than *PE* in year *t*+1.[121]

An observation in year *t* is a time series outlier if the difference between the log price of electricity in year *t*-1 and the log price of electricity in year *t* is more than 0.5 and the difference between the log price of electricity in year *t*+1 and the log price of electricity in year *t* is more than 0.5. An observation in year *t* is a cross-sectional outlier within its best-match utility area and state if the difference between the observation's log price of electricity and the mean log price of electricity for its best-match utility area and state is greater than 1.

4.7.3 Imputation Algorithm

We split our imputation algorithm into two parts. First, we correct the log price of electricity. Once we have corrected the log price of electricity, we correct the values of *EE* and *PE*. The algorithm to correct the log price of electricity is described below.

(1) Estimate *decile*-county log nominal price of electricity (*lpe*) changes using non-problem observations. Apply this price change to the non-problem previous (or subsequent) year log price of electricity to calculate a corrected log price of electricity.

---

[120] As described in Section 4.1, the best-match utility area of the plant is based on EIA data that is merged to the ASM data by county.

[121] This condition on *PE* is required in addition to the condition on the log price of electricity because if both *PE* and *EE* are outliers for the plant, the electricity price may still be reasonable.

(2) If step 1 does not provide a corrected log price of electricity, estimate county log price of electricity changes using non-problem observations. Apply this price change to the non-problem previous (or subsequent) year log price of electricity to calculate a corrected log price of electricity.

(3) If steps 1 and 2 do not provide a corrected log price of electricity, run the following regression by year on non-problem observations

$$lpe_{et} = \alpha_0 + \alpha_1 TE_{et} + \alpha_2 COUNTY_{et} \tag{4.4}$$

where lpe = log price of electricity, TE = total employment, $COUNTY$ = county where the plant is located, e = plant and t = $YEAR$. Use the predicted lpe from regression (4.4) as the corrected log price of electricity for all remaining problem observations in that county for that year.

Recall, we first identify problem observations with Data Filter 1 and create corrected log price of electricity values (*lpe_corrected*) using the algorithm above. Using the corrected observations, we then identify additional problem observations with Data Filter 2 and create corrected log price of electricity values using the algorithm above. Finally, we calculate corrected *PE* and *EE* values as follows.

- If *EE* is non-zero and non-missing, calculate corrected *PE* as *PE_corrected = EE / lpe_corrected*.

- If *EE* is zero or missing, but *PE* is non-zero and non-missing, calculate corrected *EE* as *EE_corrected = PE * lpe_corrected*.

- If both *EE* and *PE* are missing or zero, but *lpe* was corrected using either the previous or subsequent year data, assign corrected *PE* as the value of the current year mean of the unweighted purchase centile (*centile_uu*) of the previous or

subsequent year that was used to correct *lpe*. [122] Then calculate corrected *EE* as

*EE_corrected = PE * lpe_corrected*. Note that observations with both *PE* and *EE*

equal to zero or missing and no non-problem previous or subsequent year

observations cannot have *PE* and *EE* corrected though they might have a

corrected *lpe*.

4.7.4   Results of Imputation

Figure 4.10 shows the unweighted standard deviation of the log price of

electricity after the correction algorithm has been applied to all years. There is no huge

peak in the standard deviation in 1983 in Figure 4.10 as there is in Figure 4.9. The total

number of "problem" observations in 1983 is close to 6,000. Of these, nearly half are

fixed using Step 1 of our log price of electricity correction algorithm with the previous

year. In total, we are able to fix approximately two-thirds of the problem observations in

1983. We drop observations we cannot fix from the PQEM.

---

[122] The *centile_uu* variable is calculated as described in Section 4.6.
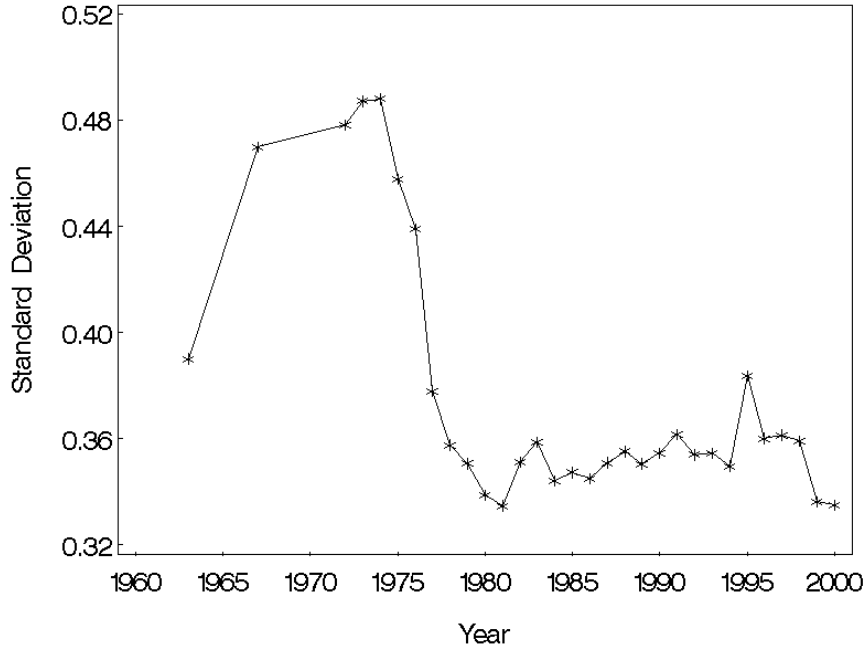
Source: Author's calculations on 1963-2000 Annual Survey of Manufactures. Statistics are not weighted.

Figure 4.10: Standard Deviation of the Log Price of Electricity After Applying the Data Filters and Correction Algorithms Described in Section 4.7, 1963-2000

4.8 Imputation of Electricity Prices in 1989-1991

The data contain a noticeable dip in electricity prices from 1989 to 1991 in the lower end of the electricity purchase distribution. Figure 4.11 shows this dip, which appears to reflect a measurement problem in the ASM. We apply the imputation model described below only to plants in the lower six deciles of the unweighted electricity purchases distribution, because the upper four deciles show no sign of measurement problems.

4.8.1 Step 1: Fit the Imputation Model to the 1988 and 1992 ASM Data

For each of 1988 and 1992, we run the following plant level regression

$$lpe\_asm_e = \gamma_0 + \gamma_1\big(centile\_uu_e\big) + \gamma_2\big(FIPSST_e\big) + \gamma_3\big(bigutil_e\big) + \varepsilon_e \qquad (4.5)$$

where *lpe_asm* is the ASM log price of electricity, *centile_uu* is the unweighted purchase centile of the plant, *FIPSST* is the state indicator, *bigutil* is the best-match utility code, and ε is a residual. In equation (4.5), *centile_uu*, *FIPSST*, and *bigutil* are vectors of fixed effects.

··*· Decile 1 (Mean log MWh = 3.9)   ⊟⊟⊟ Decile 2 (Mean log MWh = 5.0)
+++ Decile 3 (Mean log MWh = 5.7)   ××× Decile 4 (Mean log MWh = 6.3)
△△△ Decile 5 (Mean log MWh = 6.8)   ◇◇◇ Decile 6 (Mean log MWh = 7.4)
⊖⊖⊖ Decile 7 (Mean log MWh = 7.9)   ### Decile 8 (Mean log MWh = 8.5)
♡♡♡ Decile 9 (Mean log MWh = 9.2)   ✳✳✳ Decile 10 (Mean log MWh = 11.4)
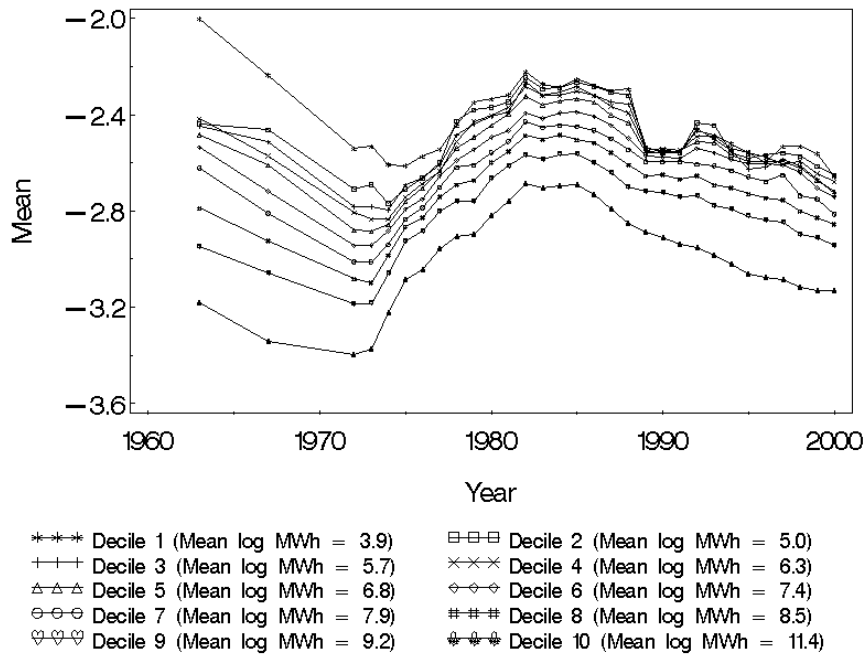
Source: Authors' calculations on plant-level data in the ASM. Statistics computed on a shipments-weighted basis.

Figure 4.11: Mean Log Real Price of Electricity by Unweighted Purchase Decile in the ASM Prior to the 1989-1991 Imputation Described in Section 4.8, 1963-2000

4.8.2 Step 2: Interpolate Coefficients for 1989-1991

Regression (4.2) yields estimated coefficients $\hat{\gamma}_0$, $\hat{\gamma}_1$, $\hat{\gamma}_2$, and $\hat{\gamma}_3$ for 1988 and 1992. We use these coefficients to interpolate coefficients for 1989, 1990, and 1991.[123] First, we calculate the difference between the 1988 and 1992 coefficients. Then we divide this difference evenly over time between 1989, 1990, and 1991. For example, the 1990 intercept coefficient is defined as shown in (4.6).

$$\hat{\gamma}_{0,1990} = \hat{\gamma}_{0,1988} + 2\left(\frac{\hat{\gamma}_{0,1992} - \hat{\gamma}_{0,1988}}{4}\right) \tag{4.6}$$

4.8.3 Step 3: Apply the Imputation Model to Plants in 1989-1991

Using the interpolated coefficients and the residual distribution, we impute a value for log electricity price for the ASM plants in 1989-1991 in the lower six deciles of the unweighted electricity purchases distribution. The formula for the imputed value is given in (4.7).

---

[123] Note there is a slight complication when interpolating the best-match utility fixed effect coefficients. All 362 best-match utilities do not appear in all the years in our sample. Recall best-match utility is a county-based notion, and we do not require county to be consistent for a single plant across our time period. Further, 1989 is the start of a new ASM panel so some counties that were represented in 1988 may not be represented in 1989. We calculate estimated 1988 and 1992 coefficients for best-match utilities in 1989, 1990, and 1991 that are not in one or both of 1988 and 1992 as the unweighted average (plant-level) of the coefficients for all of the best-match utilities in that state in that year. The estimated coefficients are calculated for 4 best-match utilities (16 plants) in 1989, 4 best-match utilities (24 plants) in 1990, and 5 best-match utilities (25 plants) in 1991.

$$lpe\_imp_e = \hat{\gamma}_0 + \hat{\gamma}_1(centile\_uu_e) + \hat{\gamma}_2(FIPSST_e) + \hat{\gamma}_3(bigutil_e) + Z(decile\_uu_e) \quad (4.7)$$

where *Z(decile_uu)* is a random draw from the pooled 1988 and 1992 distribution of

residuals in (4.2). We allow the distribution of residuals to vary freely across deciles of

electricity purchasers. We then impute *EE* and *PE* as described in Section 4.7.3.

Overall, close to 110,000 PQEM observations (60%) in 1989-1991 have

electricity variables imputed using the methodology described above. Figure 4.12 shows

the mean log real price of electricity by unweighted purchase decile and year after the

imputation. There is a big improvement from Figure 4.11 to Figure 4.12. Further, Figure

4.13 shows the mean log real price of electricity by shipments-weighted purchase decile,

where the small remaining 1989-1991 price dip is nearly gone.

4.9   Total Value of Shipments Outliers

Finally, the ASM files contain some outliers in the total value of shipments (*TVS*)

variable. For example, there is a single plant in 1963 that accounts for approximately

3.6% of total *TVS* in 1963. The same plant appears in later years with much more

reasonable *TVS* values. The discovery of this major outlier and clear data error led us to

create a *TVS* data filter to identify outliers and an imputation method to "correct" these

outliers. Appendix K describes the *TVS* data filter and imputation method we use in

creating the PQEM database.
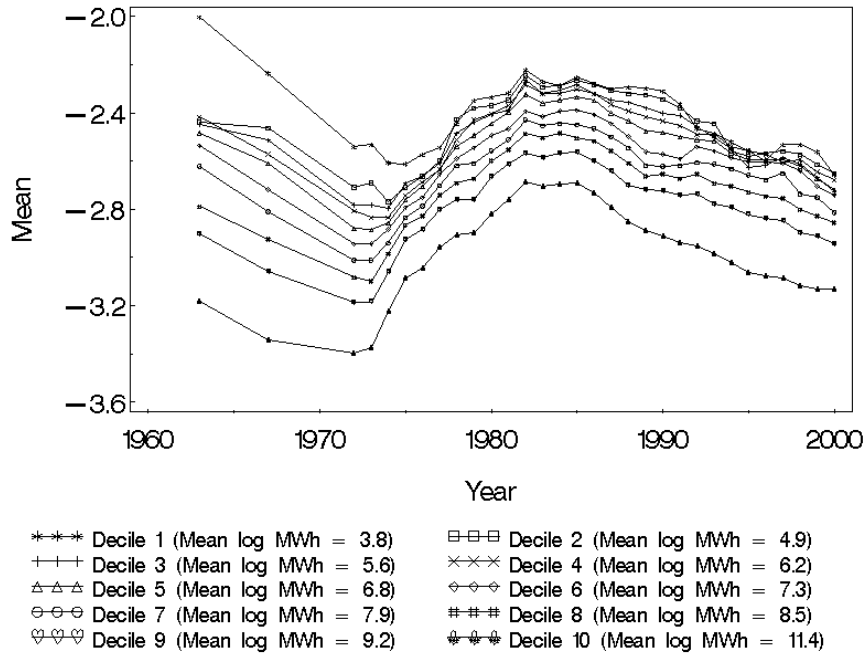
Source: PQEM database. Statistics computed on a shipments-weighted basis.

Figure 4.12: Mean Log Real Price of Electricity by Unweighted Purchase Decile, 1963-2000
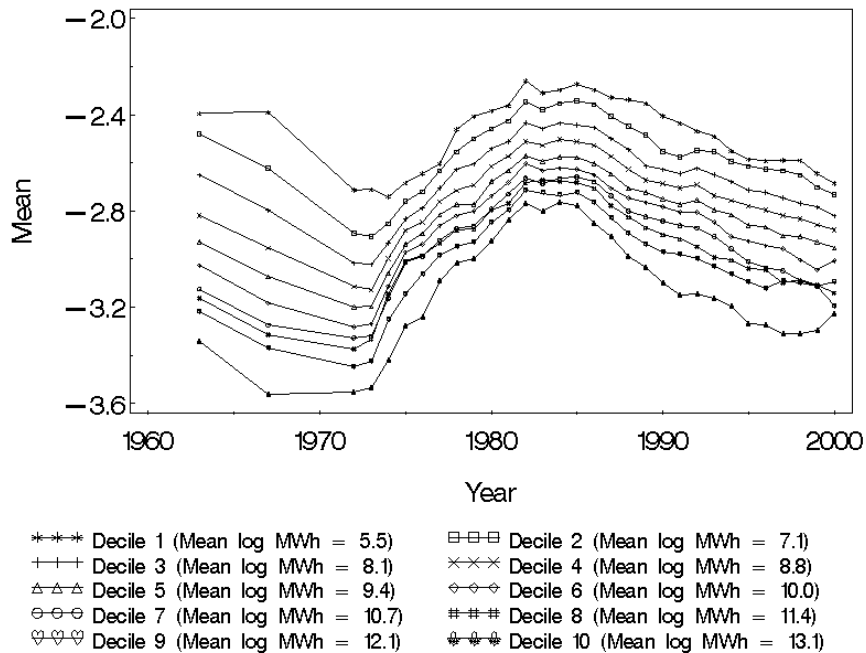
Source: PQEM database. Statistics computed on a shipments-weighted basis.

Figure 4.13: Mean Log Real Price of Electricity by Purchase Decile, 1963-2000

4.10 Conclusions

Development of the PQEM database and corresponding public release statistics is a primary contribution of this dissertation. In this chapter, we described the construction of the PQEM database in detail and provided some basic statistics of the PQEM database. The PQEM database can be used to examine many aspects of electricity prices. For example, electricity price dispersion is examined in detail in Chapter 2. We also create summary statistics for public release from the PQEM database. These statistics should be useful to researchers for a variety of purposes, including examining or controlling for electricity prices or purchases in studies of U.S. manufacturing.

Chapter 5

Conclusions

This dissertation studies the energy consumption of U.S. manufacturing plants using plant-level microdata. Chapter 2 presents a detailed examination of electricity prices paid by manufacturing plants. Chapter 3 looks at electricity, oil, natural gas, and coal consumption patterns, prices paid for these energy types, and substitution between both energy and non-energy inputs to production by manufacturing plants. Chapter 4 provides a detailed description of the construction of the PQEM database utilized in Chapter 2.

In Chapter 2, I use a newly constructed database that includes information on purchased electricity and electricity expenditures for more than 48,000 plants per year and additional data on the utilities that supply electricity to study the distribution of electricity prices paid by U.S. manufacturing plants from 1963 to 2000. The data show tremendous cross-sectional dispersion in the electricity prices paid by manufacturing plants, reflecting spatial price differences and quantity discounts. Price dispersion declined sharply between 1967 and 1977 because of erosion in quantity discounts. Differences among utilities in the purchases distribution of their customers are exploited to estimate the role of cost factors and markups in quantity discounts. The estimation results reveal that supply costs per watt-hour decline by more than half over the range of customer-level purchases in the data, regardless of time period. Prior to the mid 1970s, marginal price and marginal cost schedules with respect to annual purchase quantity are remarkably similar, in line with efficient pricing. In later years, marginal supply costs exceed marginal prices for smaller manufacturing customers by 10% or more. Spatial

dispersion in retail electricity prices among states, counties and utility service territories is large, rises over time for smaller purchasers, and does not diminish as wholesale power markets expand in the 1990s.

In Chapter 3, I examine energy type consumption patterns, prices, and substitution in U.S. manufacturing plants in 1998. I find energy type consumption patterns vary widely across manufacturing plants with over half of plants, 55.6%, consuming only electricity and natural gas and only 0.22% of plants consuming all four energy types. Further, I find a large amount of dispersion across plants in the prices paid for electricity (43%), oil (46%), natural gas (27%), and coal (43%). These high levels of dispersion are accounted for by plant location, industry, and purchase quantity. I find statistically significant (at the 1% level) estimates of the own-price elasticity of demand that range from -0.861 to -2.126 for electricity, from -1.159 to -2.023 for natural gas, and from -1.991 to -3.122 for coal. I obtain an estimate of the own-price elasticity of demand for oil of -2.623 that is significant at the 25% level. Finally, I find that energy type elasticities of demand vary with energy type consumption patterns.

The work in Chapter 3 is a first attempt at creating estimates of elasticities of demand for production inputs, including individual energy types, with U.S. plant-level data. To the best of my knowledge, this is the first work estimating individual energy type elasticities with plant-level data that attempts to incorporate the effect of the plant's initial technology choice. I find that incorporating the effect of the plant's technology choice does affect the estimated elasticities.

Development of the PQEM database and corresponding public release statistics is a primary contribution of this dissertation. In Chapter 4, I describe the construction of the

PQEM. The PQEM database is utilized in Chapter 2 and contains plant-level observations on electricity purchases, prices, and suppliers for the U.S. manufacturing sector from 1963 to 2000. In constructing the PQEM, considerable effort is devoted to treating anomalous data on electricity prices and quantities in the ASM. A number of coding errors in the ASM data are identified; the raw ASM data contain high error rates in 1983 and 1989 to 1991. Procedures to correct or impute values for the erroneous data are developed, paying special attention to the years with high error rates. Several other measurement issues pertaining to ASM sample weights in 1963 and 1967, erroneous geographic indicators, and the creation of consistent industry and geography codes over time are also addressed. Finally, Chapter 4 includes a description of planned public release statistics based on the PQEM database.

Overall, the work in this dissertation increases our knowledge of dispersion in prices paid by U.S. manufacturing plants for energy inputs, the energy type consumption patterns of plants, and electricity pricing by utilities to U.S. manufacturing plants. This dissertation also provides two data products, the plant-level PQEM database for use by approved researchers at Census Research Data Centers and the public-use data based on the PQEM database.

Electricity price dispersion is examined particularly closely, using two different sources of plant-level data on electricity consumption and expenditures, both of which show high levels of dispersion in prices paid for electricity across plants. I also find high levels of dispersion in the prices of other energy types.

There is also evidence of large quantity discounts for both electricity and natural gas. Lacking data on input prices, productivity studies typically rely on input

expenditures in place of input quantities. If larger businesses are better positioned to exploit quantity discounts, then most previous studies overstate the relative physical productivity of bigger producers. Future research on productivity differences should pay attention to input price variation among producers.

Appendix A: 1998 MECS Form Purchases and Expenditures Questions

There are three versions of the 1998 MECS form: EIA-846A, EIA-846B, and EIA-846C. All three versions of the form contain purchase and expenditure questions about the energy types discussed in this paper. The following is an example of a general purchases question (source: 1998 MECS Form EIA-846C, p. 17, Question 58).

"Enter quantity purchased by, and delivered to, this establishment in 1998, regardless of when payment was made.

Include quantities that were purchased for any onsite use, e.g., process heating or cooling, building heating or cooling, machine drive, or as raw material input to any manufacturing operation (feedstock).

Exclude quantities: 1) purchased centrally within your company, 2) delivered from other establishments of your company, or 3) for which payment was made in-kind."

The following is an example of a general expenditures question (source: 1998 MECS Form EIA-846C, p. 17, Question 59).

"Enter total expenditures, including taxes and delivery charges, for the quantity reported in question 58.

Include all expenditures regardless of when payment was made."

Appendix B: 1998 MECS SIC Coverage

Table B.1: SIC Industries Covered by the 1998 MECS

SIC Code	Industry
20	Food and Kindred Products
21	Tobacco Products
22	Textile Mill Products
23	Apparel and Other Textile Products
24	Lumber and Wood Products
25	Furniture and Fixtures
26	Paper and Allied Products
2621	Paper Mills
2631	Paperboard Mills
27	Printing and Publishing
28	Chemicals and Allied Products
2819	Industrial Inorganic Chemicals, nec.
2821	Plastics Materials and Resins
2869	Industrial Organic Chemicals, nec.
29	Petroleum and Coal Products
2911	Petroleum Refining
30	Rubber and Misc. Plastics Products
308	Miscellaneous Plastics Products, nec.
31	Leather and Leather Products
32	Stone, Clay, and Glass Products
33	Primary Metal Industries
3312	Blast Furnaces and Steel Mills
3334	Primary Aluminum
34	Fabricated Metal Products
35	Industrial Machinery and Equipment
36	Electronic and Other Electric Equipment
37	Transportation Equipment
3714	Motor Vehicle Parts and Accessories
38	Instruments and Related Products
39	Misc. Manufacturing Industries

Note: All eligible establishments in the 3- and 4-digit SIC industries shown
above were selected to receive forms. These industries were targeted for
complete coverage because they are very energy intensive.

Source: 1998 MECS Tables published on the EIA Internet site
(See http://www.eia.doe.gov/emeu/mecs/)

Appendix C: Conversion Factors for Energy Type Consumption

Table C.1: Conversion Factors for Energy Type Consumption, Physical Units to
  Million Btu

Energy Type	MECS Code	Physical Units	Conversion Factor to Million Btu
Electricity	10	1000 kWh	3.412
Residual Fuel Oil[1]	21	42 gallon barrel	6.287
Diesel Fuel[1,2]	28	42 gallon barrel	5.825[5]
Distillate Fuel Oil[1,2]	29	42 gallon barrel	5.825
Mixtures of Butane, Ethane, and Propane[1,3]	34	gallons	3.606 / 42[6]
Other LPG and NGL[1,3]	35	gallons	3.614 / 42[7]
Butane[1,3]	36	gallons	4.326 / 42
Ethane[1,3]	37	gallons	3.082 / 42
Propane[1,3]	38	gallons	3.836 / 42
Natural Gas	30	1000 cubic feet	1.031
Anthracite[4]	40	short tons	26.280
Bituminous and Sub-bituminous Coal[4]	41	short tons	22.036
Lignite[4]	42	short tons	22.036

[1] Total oil is the sum of residual fuel oil (21), diesel and distillate fuel oil (22), and LPG/NGL (24).

[2] Total diesel and distillate fuel oil (22) is the sum of diesel fuel (28) and distillate fuel oil (29).

[3] Total LPG and NGL (24) is the sum of mixtures of butane, ethane, and propane (34), other LPG and NGL (35), butane (36), ethane (37), and propane (38).

[4] Total coal is the sum of anthracite (40), bituminous and sub-bituminous coal (41), and lignite (42).

[5] EIA (2005) did not contain a conversion factor for diesel fuel. I use the conversion factor for distillate fuel oil.

[6] EIA (2005) did not contain a conversion factor for mixtures of butane, ethane, and propane. I use the simple average of the conversion factors for butane-propane mixture and ethane-propane mixture.

[7] EIA (2005) did not contain a conversion factor for other LPG and NGL. I use the conversion factor for LPG.

Source: The conversion factors for the components of coal are from the 1994 Manufacturing Energy Consumption Survey publication (EIA, 1997). The conversion factors for other energy types are from Appendix A of the 2004 Annual Energy Review (EIA, 2005).

Appendix D: Constructing Adjusted Sample Weights

Since the MECS-ASM matched dataset is considerably smaller than either the full MECS sample or the full ASM sample, I create adjusted sample weights for the matched dataset. I consider the ASM as the full sample and use 4-digit SIC industry-total employment class cells with shipments (*TVS*) to adjust the ASM sample weight.

A post-stratification using data from the 2002 CM was done for the 2002 MECS by Richard Hough and Stacey Cole of the U.S. Census Bureau (Hough and Cole, 2004). I follow their methodology, adjusting for the fact that the ASM is not a true population.

I want the weighted mean of shipments for each industry-employment class stratum for the whole sample to be equal to the weighted mean of shipments for each industry-employment class stratum for the sub-sample. In other words, (D.1) should be an equality, but it is not.

$$\sum_{e=1}^{N_{is}} w_e \times TVS_e \neq \sum_{e=1}^{n_{is}} w_e \times TVS_e \qquad\qquad (D.1)$$

where i = 4-digit SIC industry

s = total employment class

e = establishment

N_{is} = number of establishments in stratum i-s in full ASM

n_{is} = number of establishments in stratum i-s in ASM-MECS

      matched sub-sample

WT_e = ASM sample weight of establishment e

TVS_e = Total value of shipments of establishment e

w_e = ASM sample weight for establishment e

The goal is to create an adjusted weight, $\hat{w}_e$, such that

$$\sum_{e=1}^{N_{is}} w_e \times TVS_e = \sum_{e=1}^{n_{is}} \hat{w}_e \times TVS_e \tag{D.2}$$

Step 1: Decompose the RHS of (D.1).

$$\sum_{e=1}^{n_{is}} w_e \times TVS_e = \sum_{e=1}^{n_{is}} w_e \times TVS_e + \sum_{e=1}^{n_{is}} TVS_e - \sum_{e=1}^{n_{is}} TVS_e$$

$$\sum_{e=1}^{n_{is}} w_e \times TVS_e = \sum_{e=1}^{n_{is}} TVS_e + \sum_{e=1}^{n_{is}} (w_e - 1)TVS_e \tag{D.3}$$

Step 2: Resolve the inequality in (D.1).

$$\sum_{e=1}^{N_{is}} w_e \times TVS_e = \sum_{e=1}^{n_{is}} TVS_e + K_{is} \sum_{e=1}^{n_{is}} (w_e - 1)TVS_e$$

$$\sum_{e=1}^{N_{is}} w_e \times TVS_e = \sum_{e=1}^{n_{is}} TVS_e + \sum_{e=1}^{n_{is}} K_{is} (w_e - 1)TVS_e$$

$$\sum_{e=1}^{N_{is}} w_e \times TVS_e = \sum_{e=1}^{n_{is}} TVS_e + K_{is} (w_e - 1)TVS_e \tag{D.4}$$

$$\sum_{e=1}^{N_{is}} w_e \times TVS_e = \sum_{e=1}^{n_{is}} [1 + K_{is} (w_e - 1)]TVS_e$$

Step 3: The last line in (D.4) implies

$$\hat{w}_e = 1 + K_{is} (w_e - 1) \tag{D.5}$$

Step 4: Solve for K_{is} from the first line of (D.4).

$$\sum_{e=1}^{N_{is}} w_e \times TVS_e = \sum_{e=1}^{n_{is}} TVS_e + K_{is} \sum_{e=1}^{n_{is}} (w_e - 1)TVS_e$$

$$\sum_{e=1}^{N_{is}} w_e \times TVS_e - \sum_{e=1}^{n_{is}} TVS_e = K_{is} \sum_{e=1}^{n_{is}} (w_e - 1)TVS_e \tag{D.6}$$

$$K_{is} = \frac{\displaystyle\sum_{e=1}^{N_{is}} w_e \times TVS_e - \sum_{e=1}^{n_{is}} TVS_e}{\displaystyle\sum_{e=1}^{n_{is}} (w_e - 1)TVS_e}$$

K_{is} can be undefined if the industry-employment size class stratum contains only certainty

cases. Following Hough and Cole and adjusting for the fact that we are starting with a

sample rather than a population, I define the adjusted weight as shown in (D.7) if K_{is} is undefined.

$$\hat{w}_e = \frac{\displaystyle\sum_{e=1}^{N_{is}} w_e \times TVS_e}{\displaystyle\sum_{e=1}^{n_{is}} TVS_e} \tag{D.7}$$

Appendix E: Census Regions and Divisions

Northeast Region

New England Division

  Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont

Middle Atlantic Division

  New Jersey, New York, Pennsylvania

South Region

South Atlantic Division

  Delaware, District of Columbia, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, West Virginia

East South Central Division

  Alabama, Kentucky, Mississippi, Tennessee

West South Central Division

  Arkansas, Louisiana, Oklahoma, Texas

Midwest Region

East North Central Division

  Illinois, Indiana, Michigan, Ohio, Wisconsin

West North Central Division

  Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, South Dakota

West Region

Mountain Division

  Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, Wyoming

Pacific Division

  Alaska, California, Hawaii, Oregon, Washington

Appendix F: Elasticity Estimates for the Individual Energy Type

Consumption Pattern Models Without Selection Correction

Table F.1: Elasticity Estimates $\left(\hat{\varepsilon}_{ij}\right)$ for the Electricity Energy Type Consumption Pattern, No Selection, 1998

j i	Electricity	Capital	Labor	Materials
Electricity	-1.619**	-0.283**	0.164**	1.739**
	(0.007)	(0.009)	(0.006)	(0.013)
Capital	-0.026**	-0.345**	-0.202**	0.573**
	(0.003)	(0.033)	(0.043)	(0.014)
Labor	0.008**	-0.089**	-0.586**	0.666**
	(0.000)	(0.003)	(0.005)	(0.003)
Materials	0.086**	0.272**	0.749**	-1.107**
	(0.001)	(0.004)	(0.004)	(0.006)
Log Likelihood	-1,061			
N	734			

* $p < 0.05$, ** $p < 0.01$

Notes: Selection corrections are not included in this model. The model is sample-weighted. The elasticity estimates shown are sample-weighted means of the plant-level elasticity estimates. Standard errors for the elasticity estimates are calculated by bootstrapping with 1,000 iterations. Standard errors shown as '(0.000)' are less than 0.001.

Source: Author's calculations on the 1998 ASM-MECS matched dataset.

Table F.2: Elasticity Estimates $\left(\hat{\varepsilon}_{ij}\right)$ for the Electricity and Natural Gas Energy Type
Consumption Pattern, No Selection, 1998

i \ j	Electricity	Natural Gas	Capital	Labor	Materials
Electricity	-1.038	-0.007	0.217	0.208	0.619
	(0.882)	(0.164)	(1.060)	(2.756)	(2.742)
Natural Gas	-0.286**	-1.264**	0.240**	0.634**	0.677**
	(0.004)	(0.008)	(0.005)	(0.007)	(0.015)
Capital	0.144**	0.012**	-0.325**	0.492**	-0.323**
	(0.000)	(0.000)	(0.004)	(0.002)	(0.005)
Labor	-0.057**	0.010**	0.148**	-0.555**	0.453**
	(0.005)	(0.000)	(0.001)	(0.002)	(0.004)
Materials	0.138**	0.012**	-0.133**	0.503**	-0.521**
	(0.000)	(0.000)	(0.002)	(0.001)	(0.002)
Log Likelihood	1,473				
N	4,248				

* $p < 0.05$, ** $p < 0.01$

Notes: Selection corrections are not included in this model. The model is sample-weighted. The elasticity estimates shown are sample-weighted means of the plant-level elasticity estimates. Standard errors for the elasticity estimates are calculated by bootstrapping with 1,000 iterations. Standard errors shown as '(0.000)' are less than 0.001.

Source: Author's calculations on the 1998 ASM-MECS matched dataset.

Table F.3: Elasticity Estimates $\left(\hat{\varepsilon}_{ij}\right)$ for the Electricity and Oil Energy Type
Consumption Pattern, No Selection, 1998

j i	Electricity	Oil	Capital	Labor	Materials
Electricity	-0.843 (2.769)	0.071 (1.743)	0.036 (4.350)	0.380 (3.036)	0.355 (3.197)
Oil	0.060 (0.160)	-1.811 (3.252)	-0.875 (5.627)	0.812 (2.648)	1.815 (6.359)
Capital	0.006 (0.004)	-0.132** (0.027)	-0.382** (0.067)	0.470** (0.038)	0.038 (0.075)
Labor	0.030** (0.001)	0.072** (0.002)	0.329** (0.003)	-0.760** (0.006)	0.330** (0.005)
Materials	0.016** (0.001)	0.089** (0.002)	0.023** (0.004)	0.174** (0.003)	-0.303** (0.006)
Log Likelihood	587				
N	451				

* p < 0.05, ** p < 0.01

Notes: Selection corrections are not included in this model. The model is sample-weighted. The elasticity estimates shown are sample-weighted means of the plant-level elasticity estimates. Standard errors for the elasticity estimates are calculated by bootstrapping with 1,000 iterations. Standard errors shown as '(0.000)' are less than 0.001.

Source: Author's calculations on the 1998 ASM-MECS matched dataset.

Table F.4: Elasticity Estimates $\left(\hat{\varepsilon}_{ij}\right)$ for the Electricity, Natural Gas, and Coal Energy Type Consumption Pattern, No Selection, 1998

i \ j	Electricity	Natural Gas	Coal	Capital	Labor	Materials
Electricity	-1.001	0.014	0.048	0.149	0.219	0.572
	(1.403)	(0.044)	(0.535)	(0.832)	(1.285)	(0.396)
Natural Gas	0.305	1.096	0.344	0.661	-1.084	-1.322
	(0.833)	(11.222)	(1.624)	(7.488)	(3.899)	(15.613)
Coal	59.827*	-8.520**	-42.025	-2.241	68.759	-75.800*
	(22.361)	(2.773)	(23.827)	(4.788)	(40.001)	(31.032)
Capital	-0.164**	-0.025**	0.019*	0.208**	-0.002	-0.036
	(0.014)	(0.006)	(0.009)	(0.029)	(0.026)	(0.033)
Labor	0.357**	0.117**	0.230**	-0.002	-1.374**	0.672**
	(0.012)	(0.009)	(0.011)	(0.025)	(0.026)	(0.038)
Materials	0.067**	0.059**	-0.039**	0.008	0.212**	-0.308**
	(0.005)	(0.002)	(0.004)	(0.009)	(0.011)	(0.019)
Log Likelihood	697					
N	91					

* $p < 0.05$, ** $p < 0.01$

Notes: Selection corrections are not included in this model. The model is sample-weighted. The elasticity estimates shown are sample-weighted means of the plant-level elasticity estimates. Standard errors for the elasticity estimates are calculated by bootstrapping with 1,000 iterations. Standard errors shown as '(0.000)' are less than 0.001.

Source: Author's calculations on the 1998 ASM-MECS matched dataset.

Table F.5: Elasticity Estimates $\left(\hat{\varepsilon}_{ij}\right)$ for the Electricity, Oil, and Natural Gas Energy Type Consumption Pattern, No Selection, 1998

i \ j	Electricity	Oil	Natural Gas	Capital	Labor	Materials
Electricity	-1.268**	-0.032	-0.258	-0.092	0.724	0.926*
	(0.308)	(0.035)	(0.292)	(0.274)	(0.524)	(0.386)
Oil	-0.439	-2.202	-1.357	1.103	2.206	0.689**
	(0.624)	(1.697)	(1.928)	(1.433)	(2.775)	(0.064)
Natural Gas	7.738	3.063	13.337	0.346	-13.807	-10.676
	(12.273)	(4.662)	(22.746)	(0.383)	(22.222)	(17.841)
Capital	-0.059	0.026	0.010	0.020	0.279**	-0.276
	(0.184)	(0.054)	(0.006)	(1.996)	(0.105)	(1.965)
Labor	0.111**	0.032**	0.104**	0.191**	-0.757**	0.319**
	(0.000)	(0.000)	(0.000)	(0.001)	(0.002)	(0.001)
Materials	0.047**	0.004**	0.042**	-0.047**	0.133**	-0.179**
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
Log Likelihood	14,861					
N	2,991					

* $p < 0.05$, ** $p < 0.01$

Notes: Selection corrections are not included in this model. The model is sample-weighted. The elasticity estimates shown are sample-weighted means of the plant-level elasticity estimates. Standard errors for the elasticity estimates are calculated by bootstrapping with 1,000 iterations. Standard errors shown as '(0.000)' are less than 0.001.

Source: Author's calculations on the 1998 ASM-MECS matched dataset.

Table F.6: Elasticity Estimates $\left(\hat{\varepsilon}_{ij}\right)$ for the Electricity, Oil, Natural Gas, and Coal Energy Type Cons. Pattern, No Selection, 1998

i \ j	Electricity	Oil	Natural Gas	Coal	Capital	Labor	Materials
Electricity	-0.806	0.028	-0.033	-0.020	0.313	0.083	0.436
	(2.926)	(0.483)	(1.026)	(1.211)	(2.548)	(1.982)	(1.739)
Oil	21.986	17.723	-9.272	6.723	-7.561	-3.439	-26.160
	(746.880)	(635.032)	(308.623)	(223.568)	(261.550)	(124.823)	(910.503)
Natural Gas	0.917*	0.173**	-2.601**	0.091**	-0.012**	0.586**	0.847**
	(0.358)	(0.065)	(0.661)	(0.019)	(0.094)	(0.166)	(0.162)
Coal	-0.126	0.024	0.013	-0.695	0.192	-0.182	0.774
	(23.157)	(2.453)	(0.272)	(30.257)	(1.669)	(41.746)	(30.796)
Capital	-0.158**	0.017**	0.002	0.038**	-0.209**	0.117**	0.194**
	(0.002)	(0.001)	(0.002)	(0.004)	(0.008)	(0.006)	(0.015)
Labor	0.199**	0.009**	0.060**	0.246**	0.120**	-1.398**	0.764**
	(0.003)	(0.000)	(0.002)	(0.004)	(0.006)	(0.012)	(0.015)
Materials	0.091**	0.019**	0.032**	-0.007*	0.073**	0.266**	-0.475**
	(0.002)	(0.001)	(0.001)	(0.003)	(0.005)	(0.005)	(0.009)
Log Likelihood	2,728						
N	300						

* $p < 0.05$, ** $p < 0.01$

Notes: Selection corrections are not included in this model. The model is sample-weighted. The elasticity estimates shown are sample-weighted means of the plant-level elasticity estimates. Standard errors for the elasticity estimates are calculated by bootstrapping with 1,000 iterations. Standard errors shown as '(0.000)' are less than 0.001.

Source: Author's calculations on the 1998 ASM-MECS matched dataset.

Appendix G: Industry Codes

G.1 1963 and 1967

The variable *newind* for 1963 and 1967 is initially created as: *newind = OIND* (original industry code). However, the codes in *OIND* contain a combination of 1963/67 SIC codes and 1972 SIC codes. For this reason, we apply codes based on permanent plant number (*PPN*) matches to 1972-1975 for as many observations as possible. We can match in industry codes for close to 47% of the observations in the 1963 CM and just over 62% of the observations in the 1967 CM.

Codes that cannot be matched in by *PPN* using 1972-1975 data are assigned as *newind = OIND* with corrections. Two sets of corrections are made to the *OIND* codes. First, corrections are made to *OIND* according to information from Davis, Haltiwanger, and Schuh, 1996 (DHS). Table G.1 shows the DHS industry corrections. Note we do not use two of the DHS corrections: 3391 into 3399 and 3392 into 3399. Additional corrections are added to account for changes from the 1963/67 to the 1972 SIC classification system that are not included in the DHS table, which is designed for use with 1972-1986 ASM and CM data. Table G.2 shows these corrections for industries that can be assigned in a simple way.[124] Only about 270 (0.2%) plant-year observations required corrections from these tables in 1963 and 1967.

---

[124] The corrections in Table G.2 are based on the industry concordance table for 1967 to 1972 SIC codes found in Appendix B of the *Census of Manufactures, 1972* publication.

Table G.1: DHS Industry Corrections

Old Industry	Corrected Industry
Discontinuous Industries	
2794	2793
3672	3671
3673	3671
Miscoded Industries	
2015	2016
2031	2091
2042	2048
2071	2065
2072	2066
2093	2076
2094	2077
2317	3317
2433	2439
2443	2449
2689	2649
3323	3325
3391[*]	3399
3392[*]	3399
3461	3466
3472	3743
3481	3496
3578	3579
3614	3674
3642	3646
3716	3713
3722	3724
3729	3728
3741	3743
3791	2451

Source: Page 222 of Davis, Haltiwanger, and Schuh (1996).

[*] Note there are two DHS corrections not used in the PQEM. According to the 1967 to 1972 SIC code mapping, 1967 industry 3391 compares directly to 1972 industry 3462 and 1967 industry 3392 compares directly to 1972 industry 3463.

Table G.2: Industry Corrections (1963/67 to 1972 SIC Codes)[125]

Old Industry	Corrected Industry
1911	3489
1925	3761
1929	3483
1931	3795
1951	3484
1961	3482
1999	3489
2036	2092
2073	2067
2096	2079
2442	2449
2445	2449
2644	2649
2815	2865
2872	2875
3391	3462
3392	3463
3491	3412
3492	3499
3571	3574
3572	3579
3717	3711
3726	3728
3831	3832
3871	3873
3872	3873
3912	3915
3913	3915
3941	3944
3943	3944
3981	3991
3987	3999
3988	3995
3992	2371
3994	3995

---

[125] There are three industries that cannot be corrected if they are 1963/67 SIC codes because they also exist as 1972 SIC codes and may be correct already in *OIND*: 2091, 2092, and 3496

For splitter industries, we randomly assign plants from the 1963/67 industry that splits to the 1972 industry based on the fraction of plants in the industry in 1972. A simple example of this, where 1963/67 industry A splits into 1972 industries B and C, is laid out below.

- Let N_X = the number of plants in industry X in 1972.

- Let $Z = \dfrac{N_B}{\left(N_B + N_C\right)}$.[126]

- Let Q = random number between in the range [0,1] assigned to a plant in industry A in 1963/67.

- If $Q \leq Z$, then assign the plant to industry B. Otherwise, assign the plant to industry C.

Approximately 3.6% (roughly 4,100) of total plant-year observations in 1963 and 1967 have industry codes assigned with the methodology described above.

G.2 1972-1986

In 1972-1986, the variable *newind* is also initially created as: *newind = OIND*. As in 1963 and 1967, additional corrections are made to *OIND* according to the DHS corrections shown in Table G.1. 1972 also contains a significant number of industry codes in *OIND* that are 1987 SIC codes. Corrections shown in Table G.3 are applied to

---

[126] Note that simple (shown in Tables A.1 and A.2) corrections are made to 1972 SIC codes and only valid industry codes are kept when creating the Z variable.

these codes.[127] We make industry corrections from these tables to approximately 1,400

(0.2%) plant-year observations from 1972 to 1986.

Table G.3: Industry Corrections (1987 SIC Codes in 1972)

Old Industry	Corrected Industry
2053	2038
2064	2065
2326	2328
2656	2654
2671	2641
2672	2641
2673	2643
2674	2643
2675	2645
2676	2647
2677	2642
2678	2648
2835	2831
2836	2831
3052	3041
3053	3293
3061	3069
3081	3079
3082	3079
3084	3079
3089	3079
3363	3361
3365	3361
3366	3362
3556	3551
3596	3576
3663	3662
3669	3662
3827	3832
3845	3693

---

[127] The 1987 SIC manual is the source of the corrections in Table G.3. It is not clear how 1987 SIC codes ended up in the *OIND* variable in 1972.

G.3 1987-2000

In 1987-1997, the variable *newind* is created as: *newind = IND* (tabulated industry code). However, North American Industrial Classification System (NAICS) codes are introduced to the ASM in 1998. The *IND* variable contains NAICS codes in 1998-2000. Therefore, the 1998-2000 values of *newind* are created using different methods. The method for 1998 and 1999 is simple; the variable *OIND* contains the 1987 SIC code so *newind = OIND*. The method we use to create *newind* in 2000 is a bit more complicated. The 2000 values of *newind* are based on *PIND* (processing industry code) instead of *IND* or *OIND*. The variable *newind* is equal to *PIND* if *PIND* is less than 10,000. If *PIND* is greater than or equal to 10,000, then *newind* is equal to the substring containing the first four digits of *PIND*. Further, if this method yields an invalid industry code, *newind* is taken from the most recent year containing a valid 1987 SIC code for that plant. We only do this for roughly 400 (0.2%) of plant-year observations in 1998 to 2000.

The Asbestos industry (3292) completely disappears from the ASM in 1994-1996 and then reappears in 1997. We correct for this in the PQEM by assigning all plants that are in industry 3292 in any of 1990-2000 to be in 3292 in all of 1990-2000. This is not a perfect solution. There are very few plants in the Asbestos industry from 1992 to 1997. For some purposes, a researcher may want to consider combining the Asbestos industry with another industry similar in nature for post-1986 years.

There are 458 industries in years 1987 through 1998. In 1999, four industries drop out. They are 2411 (Logging), 2711 (Newspapers: Publishing, or Publishing and Printing), 2721 (Periodicals: Publishing, or Publishing and Printing), and 2741 (Miscellaneous Publishing). In 2000, one additional industry drops out, 2731 (Books:

Publishing, or Publishing and Printing). These industries drop out of manufacturing as a

result of the switch from SIC to NAICS. Therefore, they are no longer included in the

population for the ASM and cannot be included in the PQEM in these years.

Appendix H: Identification of ASM Plants in the 1967 CM

H.1 Introduction

ASM plants are not readily identifiable in the 1967 CM. The ET variable generally used to identify ASM plants in CM years contains no useful information. We examined the 1967 CM publications for information on ASM sample selection and the number of plants in the actual ASM sample. We found the following useful information.

- The 1967 ASM reporting panel is comprised of approximately 61,000 plants.

- The ASM sample constitutes "about one-fifth of the plants but with complete coverage of all large ones."[128] (Source: *Census of Manufactures, 1967* publication)

- The certainty cases in the ASM include, in general, "all plants of companies with at least one manufacturing plant with 100 employees or more (250 employees or more in the apparel manufacturing and printing and publishing industries)." (Source: *Census of Manufactures, 1967* publication)

We define ASM plants in the 1967 CM in the following way. A plant is an ASM plant if one or more of the following three conditions are met:

1. The plant is a certainty case.

2. The plant has unique values within a four-digit SIC industry for the following ratios of CM variables: *CP/TVS*, *WW/SW*, *EE/CP*, *PH/PW* and *PW/WW*.[129]

---

[128] There are 305,643 manufacturing plants in the 1967 CM, one-fifth of this is 61,128, which matches the estimate given above.

[129] The definitions of the variables in these ratios are given in Section H.3.

3. If, after applying criteria 1 and 2, the plant's four-digit SIC industry - total

   employment class category still contains too few plants and the plant has unique

   values within a four-digit SIC industry for *any* four of the following six ratios of

   ASM variables: *CP/TVS*, *WW/SW*, *EE/CP*, *PE/SW*, *PH/PW* and *PW/WW*. [130]

H.2 Certainty Cases

We use the criteria listed above with the 1967 CM value of *TE* to determine if a

plant is a certainty case. Clearly, the creators of the 1967 ASM sample did not have this

value of *TE* when they chose the ASM cases. However, the 1967 CM value of *TE* is

probably a reasonably good proxy for the total employment variable used to choose the

ASM certainty cases. We identify 45,741 certainty cases. This is a good core of plants

that are likely ASM plants. However, this number is roughly 15,000 plants short of the

true number of ASM plants in 1967.

H.3 Dunne Ratio Method

Dunne, in a 1998 internal CES memo on 1987 and 1992 imputes, offers a possible

suggestion of how to identify imputed (non-ASM) cases in the CM. He discusses

"typical" imputation patterns in the data. Further, he says that in 1987 and 1992, Census

imputed a "common ratio to non-respondent non-ASM cases for each four-digit

industry". Dunne examines several ratios in 1987 and 1992, focusing on the ratio

CM/TVS (cost of materials / total value of shipments). He looks for repeated occurrences

---

[130] Total employment classes are defined from 1 to 5 as follows: 1-19 employees, 20-49 employees, 50-99 employees, 100-249 employees and 250 or more employees. The number of plants in a four-digit SIC industry-total employment class category must be high enough that the created ASM sample weight is less than or equal to 200. See Appendix I for information on the creation of ASM sample weights in 1967.

of the same value of the ratios within four-digit industries. He states that plants with these repeated ratios are likely imputed cases.

We examine six ratios as part of our procedure to identify 1967 ASM plants.

1. *CP/TVS* – cost of materials and parts / total value of shipments[131]

2. *WW/SW* – production workers wages / total salaries and wages

3. *PH/PW* – total plant hours / average number of production workers

4. *PE/SW* – quantity of purchased electricity / total salaries and wages

5. *EE/CP* – expenditures on electricity / cost of materials and parts

6. *PW/WW* – average number of production workers / production workers wages

We examine ratios 1-3 based on the results in Tim Dunne's 1998 memo, and we examine ratio 4 because we know *PE* is imputed for non-ASM cases. We examine ratios 5 and 6 because they do not include any potential administrative record variables.[132] Further, we think that *CP*, *EE*, *PW*, and *WW* are likely to have low item impute rates because they should be fairly easy for the manager of a plant to learn and report. While we make use of the Dunne ratio method to identify possible impute cases, we recognize that imputation procedures at the U.S. Census Bureau may have been very different in 1967 than in 1987 and 1992.

---

[131] Note the variable *CM* is not available in the 1967 CM so we use *CP* (cost of materials and parts) instead

[132] *TVS* and *SW* could potentially be in the administrative record. If administrative record information for "whole" impute cases was added to the 1967 CM data, it is possible that unique ratios will occur even for "whole" impute cases if the ratios involve administrative record data such as *TVS* or *SW*. [Recall that the standard imputation method is $X_{et} = PAYROLL_{et} (X_{it}/PAYROLL_{it})$.] Thus, ratios 1, 2, and 4 may not yield clear results for identifying ASM plants.

H.4 Coverage Considerations

ASM sample weights have a maximum value of 200 in each of the several years following 1967 (1972-1993). Therefore, we force our sample to be such that there are no sample weights over 200. We also develop a method of creating ASM sample weights for 1967.[133]

H.5 Choosing the Final Set of 1967 ASM Plants

First, we create several variables to examine repeats of our six ratios within four-digit SIC industries. Then, we create an ASM sample based on the method described in criteria 1 and 2 in Part H.1 of this appendix. However, this sample contains several plants with sample weights of over 200. These high weights are the results of selecting too few plants in the smallest total employment class in some four-digit SIC industries. To solve this problem, we add plants from the smallest total employment class in problematic four-digit SIC industries. Plants are added to the ASM sample if they have unique ratios for *any* four of the following six ratios of ASM variables: *CP/TVS*, *WW/SW*, *EE/CP*, *PE/SW*, *PH/PW* and *PW/WW*. In total, this method yields 62,900 ASM plants.

---

[133] See Appendix I for a more detailed description of this method.

Appendix I: Creation of ASM Sample Weights for 1963 and 1967

I.1 Background

ASM sample weights are not available on the 1963 and 1967 CM plant-level data files. This section describes the method we use to estimate ASM sample weights for the 1963 and 1967 CM surveys. Plant-level data files for all other years already contain ASM sample weights in the variable *WT*.

I.2 1963 Sample Weights

We use a simple algorithm to estimate ASM sample weights in 1963. The steps of the algorithm are listed below.

1. Identify ASM cases: ASM cases in the 1963 CM are identified using the establishment type (*ET*) flag. A value of *ET* equal to zero indicates an ASM case.

2. Assign weights to ASM certainty cases: All ASM certainty cases are given a weight equal to one. The 1963 CM publication states that all plants of companies with 100 or more employees were ASM certainty cases. (Source: *Census of Manufactures, 1963* publication)

3. Assign weights to ASM non-certainty cases: Non-certainty cases are assigned a weight of one over the probability of selection. We define the probability of selection based on the following three total employment (*TE*) classes, by 4-digit industry: 1-19, 20-49, and 50-99 employees. For example, if there are 20 ASM plants and 100 CM plants with *TE* of 1-19 in industry 2011, the probability of selection of each of the 20 ASM plants is 1/5.

I.3 1967 Sample Weights

The algorithm we use to estimate ASM sample weights in1967 is very similar to the algorithm used for 1963. The steps of the 1967 algorithm are as follows:

1. Identify ASM cases: Unlike the 1963 ASM cases, ASM cases in the 1967 CM are not immediately identifiable; all plants have *ET* equal to one. Our criteria for choosing ASM cases in the 1967 CM are described in Appendix G.

2. Assign weights to ASM certainty cases: All ASM certainty cases are given a weight equal to one. The 1967 CM publication states that the certainty cases in the ASM include, in general, "all plants of companies with at least one manufacturing plant with 100 employees or more (250 employees or more in the apparel manufacturing and printing and publishing industries)." (Source: *Census of Manufactures, 1967* publication)

3. Assign weights to ASM non-certainty cases: Non-certainty cases are assigned a weight of one over the probability of selection. We define the probability of selection based on the following four total employment (*TE*) classes, by 4-digit industry: 1-19, 20-49, 50-99, and 100-249 employees.

Appendix J: Geography Codes

J.1 Introduction

This appendix describes changes we made to the state and county FIPS codes to make them correct and consistent in the PQEM.

J.2 Creation of a Unique County Identifier

The PQEM contains FIPS county and state codes for each plant-year observation. The FIPS county code is repeated across states. Therefore, we combine the FIPS state and county codes to get a single county identifier. We create a variable called *cyfstn* that is unique to each county-state combination by concatenating the 2-digit FIPS state code to the end of the 3-digit FIPS county code. The *cyfstn* variable is a character variable of length five.[134]

J.3 Hawaii FIPS State Code Correction: 1963-1988

There are several state identifiers on the ASM and CM plant-level data files: FIPS state code (*FIPSST*), Census state code (*CENST*), state name (*STATE*), and state postal abbreviation (*POSTALST*). For the years 1963-1988, the FIPS state code for Hawaii (*FIPSST* = 15) never appears on the ASM or CM files.[135] While Hawaii is not a big manufacturing state, it does have some manufacturing plants, such as food canning plants. This problem was identified on previous versions of the ASM and CM plant-level

---

[134] The *cyfstn* variable is re-created after the corrections described below to the FIPS county and state code variables.

[135] The *STATE* for Hawaii ("Hawaii"), the *POSTALST* for Hawaii ("HI"), and the *CENST* for Hawaii ("15") also do not appear in 1963-1988.

data files by Randy Becker. Becker (1999) describes the Hawaii problem and a method for correcting it on the 1963-1987 CM files.

Following Becker (1999), we use the algorithm below to correct state codes from 1963-1988.

- For 1963-1981 and 1983-1985, *CENST* is correct on previous versions of the ASM and CM files. For Hawaii, *CENST* = 95. The fix is straightforward: drop *CENST* from the PQEM (for the problem years), merge on *CENST* from the previous versions of the files by year and *PPN*, and then if *CENST* = 95 set *FIPSST* = 15.

- For 1982 and 1986-1988, the fix is more complicated. In these years, *FIPSST*, *CENST*, and *REG* (census region) all have problems. There is a fix file, *STATEFIX.sas7bdat*, available for the CM years. This file contains the variables shown in Table J.1 for the years 1972, 1977, 1982, and 1987.

- In 1982 and 1987, the values of *FIPSST* are corrected directly by merging the *STATEFIX* file to the PQEM by year and *PPN*. This match is checked by comparing the *TVS*, *TE*, and *IND* variables. [136]

- We use data from surrounding years to correct the bad *FIPSST* variable in 1986 and 1988.

---

[136] The real *SMSA* code is "lost" in 1986 (missing in the PQEM). However, it could be mapped in using the correct state and county codes.

Table J.1: Variables Contained in the *STATEFIX* Data File

Variable	Definition
CFN	Census file number
COUNTY	County
FST	FIPS state code
IND	4-digit SIC code
PLACE	Place code
PPN	Permanent plant number
ST	Census state code
ST_POSTAL	Postal state code (2-letter)
TE	Total employment
TVS	Total value of shipments
YR	Year (2-digit)

J.4 FIPS County Code Problem: 1986

The FIPS county code (*COU*) variable in the 1986 ASM plant-level data file does not contain the FIPS county code (as of March 2005). It actually contains the last three digits of the place code variable (*PLAC*). The correct county code is contained in the first three digits of the *SMSA* field.

J.5 County Concordance Over Time

According to Becker (1999), there are two main areas of concern related to concordance over time for FIPS county codes. First, FIPS county codes have changed over time. For instance, between 1967 and 1972, there was a major change in the codes. To make room for future additions, "space" was added between FIPS county codes. The new codes were usually, though not always, defined according to (J.1).

$$\text{New FIPS County Code} = (\text{Original FIPS County Code})*2 - 1 \qquad (J.1)$$

There are also cases of counties being combined or split up. The second major concern with county data is the changing of boundaries over time. Unfortunately, if this boundary change is not reflected in a simple FIPS county code change, we do not correct for it.

Randy Becker provided us with a county concordance program for the CM years. We modified this program in two ways. First, we modified the program to include ASM years. Second, we added corrections for Alaska, described below, to the program. We then applied the program to the PQEM database to achieve the best possible county concordance over time.

The original program does not include corrections for Alaska beyond the general correction for the 1967 to 1972 FIPS county code change. As of 2000, there are 27 FIPS county-equivalents in Alaska. Table J.2 shows a list of the Alaska county-equivalents. Some of these county-equivalents were split out from other county-equivalents. Therefore corrections need to be made for concordance over time. Following the method in Randy Becker's program, the counties are put back together into their former county. Additionally, there are several county-equivalents in the early years (1963, 1967) of the CM that are later added to other county-equivalents. Table J.3 shows the corrections we include in our database. Finally, we should note in the 1960s there are plants listed in county "001" for Alaska. Most of the occurrences are in 1963, though there are a few in 1967. County "001" is a large division that is later separated into several county-equivalents. We feel it is too large to put all of the county-equivalents back together. Therefore, we use later years to determine the location of the plant. If the plant does not appear in later years (there are only around 30 such plants), we drop the plant-year observations from the PQEM.

Table J.2: Alaska FIPS County-Equivalents

FIPS County Code	County
013	Aleutians East
016	Aleutians West
020	Anchorage
050	Bethel
060	Bristol Bay
068	Denali
070	Dillingham
090	Fairbanks North Star
100	Haines
110	Juneau
122	Kenai Peninsula
130	Ketchikan Gateway
150	Kodiak Island
164	Lake and Peninsula
170	Matanuska-Sustina
180	Nome
185	North Slope
188	Northwest Arctic
201	Prince of Wales-Outer Ketchikan
220	Sitka
232	Skagway-Hoonah-Angoon
240	Southeast Fairbanks
261	Valdez-Cordova
270	Wade-Hampton
280	Wrangell-Petersburg
282	Yakutat
290	Yukon-Koyukuk

Table J.3: Corrections to Alaska FIPS County-Equivalents

Original FIPS County Code	Original County	Corrected FIPS County Code	Corrected County
013	Aleutians East	010	Aleutians
016	Aleutians West	010	Aleutians
030	Angoon	231	Skagway-Yakutat-Angoon
040	Barrow-North Slope	185	North Slope
080	Cordova-McCarthy	261	Valdez-Cordova
164	Lake and Peninsula	070	Dillingham
230	Skagway-Yakutat	231	Skagway-Yakutat-Angoon
232	Skagway-Hoonah-Angoon	231	Skagway-Yakutat-Angoon
282	Yakutat	231	Skagway-Yakutat-Angoon

J.6 Concordance with EIA Utility Data Counties

EIA utility data, described in Section 4.2.2, is merged to the PQEM by county. We want utility data for every plant-year observation. Therefore, we need concordance between the counties in the PQEM database and the counties in the EIA data file. There are 13 counties in the PQEM that do not have matches to the EIA data file. These counties are actually independent cities with their own FIPS county codes. We examined each of these individually to determine the utility that most likely provides their electricity.[137] We then added them to the EIA data file. Table J.4 shows a list of these counties and the utilities they were assigned. There are roughly 11,500 (0.63%) plant-year observations located in these counties over all years of the PQEM.

---

[137] We looked the cities up on the Internet to determine the utility currently serving them. As a check, we looked at the utilities serving the surrounding counties as well. In a couple of cases, there was more than one possible utility. For these cases, we picked the utility that served the largest number of surrounding counties.

Table J.4: Utilities Individually Assigned to Counties

State	FIPSST	County	COU	Assigned Utility (Utility Code)
MO	29	St. Louis city	510	Union Electric Company (19436)
VA	51	Bristol city	520	Bristol Utilities Board (2248)
VA	51	Danville city	590	Danville City (4794)
VA	51	Franklin city	620	Franklin City (6715)
VA	51	Galax city	640	Appalachian Power Company (733)
VA	51	Harrisonburg city	660	Harrisonburg City (8198)
VA	51	Lynchburg city	680	Appalachian Power Company (733)
VA	51	Martinsville city	690	Martinsville City (11770)
VA	51	Norton city	720	Old Dominion Electric Power (10171)
VA	51	Radford city	750	Radford City (15619)
VA	51	Roanoke city	770	Appalachian Power Company (733)
VA	51	South Boston city	780	Virginia Electric & Power (19876)
VA	51	Williamsburg city	830	Virginia Electric & Power (19876)

J.7 Consistent State Code Requirement for Plants

Finally, we require that a plant does not change states over time in the PQEM. We use permanent plant number (*PPN*) as the plant identifier. A *PPN* should be assigned to a single plant location regardless of ownership or industry. If a manufacturing plant relocates, it should be assigned a new *PPN*. Therefore, if county and state boundaries are stable over time, a given *PPN* should have the same FIPS county and state codes throughout our sample period. However, we know that county boundaries do change over time. As noted previously, we cannot capture these boundary changes if they are not reflected in a simple FIPS county code change. Therefore, we allow the county of a given *PPN* to vary over our sample period. However, since state boundary changes are very rare, we do not allow the state of a given *PPN* to change over our sample period. There were a small number of occurrences of state changes over time for a given *PPN* in our

database. We used the following algorithm to create consistent FIPS state codes over time in our database.

1. Flag all plants for which the FIPS state code changes during our sample period.

2. For the flagged plants, assign the *cyfstn* that appears in the most years to that plant in all years.[138] If there is a tie, assign the *cyfstn* that appeared most recently to the plant in all years.

3. For flagged plants, correct the FIPS state code to match the corrected *cyfstn*.

This correction changes the *cyfstn* and FIPS state code variable for 0.57% (roughly 10,000) of our plant-year observations. The number of corrected observations is spread fairly evenly across years, ranging from around 200 to 500 corrected observations per year.

---

[138] We change the *cyfstn* variable rather than the FIPS state code alone because changing the FIPS state code alone would lead to non-existent county-state combinations in *cyfstn*.

Appendix K: *TVS* Outliers

K.1 Data Filter

A plant-year observation is flagged for *TVS* imputation if the following criteria are met:

(1) Payroll (SW_t) < 2% of TVS_t

and

(2) TVS_t > median of 4-digit SIC industry *TVS* for year t

and

(3)

- If $t = 1963$:

    Absolute value of $(\log(TVS_{t+5}) - \log(TVS_t)) > 5$

- If $t = 1967$:

    Absolute value of $(\log(TVS_{t+5}) - \log(TVS_t)) > 5$

    and[139]

    Absolute value of $(\log(TVS_{t-5}) - \log(TVS_t)) > 5$

- If $t = 1972$:

    Absolute value of $(\log(TVS_{t+1}) - \log(TVS_t)) > 1$

    and

    Absolute value of $(\log(TVS_{t-5}) - \log(TVS_t)) > 5$

---

[139] We also tried using an "or" condition in place of all of the "and" conditions shown in step (3). Unfortunately, we flag some cases we should not flag when we do this. For example, we flag a lot of cases for imputation where there is a big jump in TVS from the previous year where TVS is at approximately the same level in later years (e.g., the year after entry).

- If $1973 \leq t \leq 1999$:

    Absolute value of $(\log(TVS_{t+1}) - \log(TVS_t)) > 1$

    and

    Absolute value of $(\log(TVS_{t-1}) - \log(TVS_t)) > 1$

- If $t = 2000$:

    Absolute value of $(\log(TVS_{t-1}) - \log(TVS_t)) > 1$

We flag a total of 105 plant-year observations for *TVS* imputation, including the 1963 case described in Section 4.9. Not surprisingly, we flag more observations for *TVS* imputation in the two end years, 1963 and 2000.

K.2 Imputation Method

For observations we flag for *TVS* imputation, we perform a simple imputation based on the plant's *TVS* to *SW* ratio in the two (if available) surrounding years. We use the following procedure to impute *TVS*.

(1) Calculate the ratio of *TVS* to *SW* in each year:

$$rat\_TVS\_SW_t = \frac{TVS_t}{SW_t} \qquad \text{(K.1)}$$

(2) Calculate *TVS_SW_avg*.

- If $t = 1963$:

$$TVS\_SW\_avg_t = TVS\_SW\_avg_{t+5} \qquad \text{(K.2)}$$

- If $t = 1967$:

$$TVS\_SW\_avg_t = \frac{1}{2}\left(TVS\_SW\_avg_{t-5} + TVS\_SW\_avg_{t+5}\right) \qquad \text{(K.3)}$$

- If $t = 1972$:

$$TVS\_SW\_avg_t = \frac{1}{5}\left(TVS\_SW\_avg_{t-5}\right) + \frac{4}{5}\left(TVS\_SW\_avg_{t+1}\right) \quad \text{(K.4)}$$

- If $1973 \leq t \leq 1999$:

$$TVS\_SW\_avg_t = \frac{1}{2}\left(TVS\_SW\_avg_{t-1} + TVS\_SW\_avg_{t+1}\right) \qquad \text{(K.5)}$$

- If $t = 2000$:

$$TVS\_SW\_avg_t = TVS\_SW\_avg_{t-1} \qquad \text{(K.6)}$$

(3) Calculate corrected TVS, *TVS_CORR* as shown in (K.7).

$$TVS\_CORR_t = SW_t * TVS\_SW\_avg_t \qquad \text{(K.7)}$$

References

Bartelsman, Eric J. and Mark Doms. "Understanding Productivity: Lessons from Longtitudinal Microdata." *Journal of Economic Literature*, 2000, *38*(3), pp. 569-594.

Bartelsman, Eric J. and Wayne Gray. "The NBER Manufacturing Productivity Database." National Bureau of Economic Research, Inc., NBER Technical Working Paper Series: No. 205, 1996.

Baxter, R.E. and R. Rees. "Analysis of the Industrial Demand for Electricity." *The Economic Journal*, 1968, *78*(310), pp. 277-298.

Becker, Randy, John Haltiwanger, Ron Jarmin, Shawn Klimek, and Dan Wilson. "Micro and Macro Data Integration: The Case of Capital." Center for Economic Studies, U.S. Census Bureau, CES Working Paper: CES-WP-05-02, 2005.

Becker, Randy A. "Geographical Coding in the Censuses of Manufactures: 1963-1997." Center for Economic Studies mimeo, 1999.

Berndt, Ernst R. *The Practice of Econometrics – Classic and Contemporary*. Reading, MA: Addison-Wesley Publishing Company, 1990.

Berndt, Ernst R. and David O. Wood "Technology, Prices, and the Derived Demand for Energy." *The Review of Economics and Statistics*, 1975, *57*(3), pp. 259-268.

Besanko, David, Julia D'Souza and S. Ramu Thiagarajan. "The Effect of Wholesale Market Deregulation on Shareholder Wealth in the Electric Power Industry." *Journal of Law and Economics*, 2001, *44*(1), pp. 65-88.

Bjørner Thomas Bue and Henrik Holm Jensen. "Interfuel Substitution within Industrial Companies – An Analysis Based on Panel Data at Company Level." Danish Local Government Research Institute (AKF Forlaget) Discussion Paper, 2001.

Bonkowski, Richard. "The U.S. Coal Industry in the 1990's: Low Prices and Record Production." Energy Information Administration, U.S. Department of Energy, 1999.

Borenstein, Severin. "The Trouble with Electricity Markets: Understanding California's Restructuring Disaster". *The Journal of Economic Perspectives*, 2002, *16*(1), pp. 191-211.

Borenstein, Severin and Stephen P. Holland. "On the Efficiency of Competitive Electricity Markets with Time-Invariant Retail Prices." *The RAND Journal of Economics*, 2005, *36*(3), pp. 469-493.

Brown, Stephen J. and David S. Sibley. *The Theory of Public Utility Pricing.* Cambridge: Cambridge University Press, 1986.

Burtraw, Dallas. "Cost Savings Sans Allowance Trades? Evaluating the SO2 Emissions Trading Program to Date." Resources for the Future Discussion Paper: 95-30-REV, 1996.

Caywood, Russell E. *Electric Utility Rate Economics.* New York: McGraw-Hill Book Company, 1972.

Chiang, Hyowook. "Essays on Employer-Employee Relationships and Firm Performance." Ph.D. diss. University of Maryland, 2005.

Christensen, Laurits R., Dale W. Jorgenson and Lawrence J. Lau. "Transcendental Logarithmic Production Frontiers." *The Review of Economics and Statistics*, 1973, *55*(1), pp. 28-45.

Christensen, Laurits R., Dale W. Jorgenson and Lawrence J. Lau. "Duality in the Theory of Production." *Econometrica*, 1971, *39*(4), pp. 255-256.

Cooper, Russell and John Haltiwanger. "The Aggregate Implications of Machine Replacement: Theory and Evidence." *The American Economic Review*, 1993, *83*(3), pp. 360-382.

Cowern, Edward H. "Power and Energy: Factors that Determine Industrial Electric Bills." The Cowern Papers presented by Motors and Drives, LLC (http://www.motorsanddrives.com/), 2001.

Cudahy, Richard D. and J. Robert Malko. "Electric Peak-Load Pricing: Madison Gas and Beyond." *Wisconsin Law Review*, 1976, *47*, pp. 47-78.

Davis, Steven J., Cheryl Grim, John Haltiwanger, and Mary Streitwieser. "Electricity Pricing to U.S. Manufacturing Plants, 1963-2000." Unpublished Paper, 2006a.

Davis, Steven J., Cheryl Grim, John Haltiwanger, and Mary Streitwieser. "Prices and Quantities of Electricity in the U.S. Manufacturing Sector: A Plant-Level Database and Public-Release Statistics, 1963-2000." Unpublished Paper, 2006b.

Davis, Steven J. and Haltiwanger, John. "Sectoral Job Creation and Destruction Responses to Oil Price Changes." *Journal of Monetary Economics*, 2001, *48*(3), pp. 465-512.

Davis, Steven J., John C. Haltiwanger and Scott Schuh. *Job Creation and Destruction.* Cambridge, MA: The MIT Press, 1996.

Davis, Steven J. and John Haltiwanger. "Wage Dispersion Between and Within U.S. Manufacturing Plants, 1963-1986," in Martin Neil Baily and Clifford Winston, eds., *Brookings Papers on Economic Activity: Microeconomics,* Washington, D.C.: The Brookings Institution, 1991, pp. 115-200.

Diewert, W.E. (1971). "An Application of the Shephard Duality Theorem: A Generalized Leontief Function." *Journal of Political Economy,* 1971, *79*(3), pp. 481-507.

Dunne, Timothy. "1987 & 1992 Imputes." Center for Economic Studies, CES DATA ISSUES Memorandum 98-1 (Access restricted to CES staff and approved data users), 1998.

EIA [Energy Information Administration]. "2002 Manufacturing Energy Consumption Survey Methodology and Data Quality: Survey Design, Implementation, and Estimates." Energy Information Administration, U.S. Department of Energy, Available from: http://www.eia.doe.gov/, 2005a.

EIA [Energy Information Administration]. "Annual Energy Review 2004." Energy Information Administration, U.S. Department of Energy: No. DOE/EIA-0384(2004), 2005b.

EIA [Energy Information Administration]. "Annual Energy Review 2003." Energy Information Administration, U.S. Department of Energy: No. DOE/EIA-0384(2003), 2004.

EIA [Energy Information Administration]. "Manufacturing Energy Consumption of Energy 1994." Energy Information Administration, U.S. Department of Energy: No. DOE/EIA-0512(94), 1997.

EIA [Energy Information Administration]. "Inventory of Nonutility Electric Power Plants in the United States 1999." Energy Information Administration, U.S. Department of Energy: No. DOE/EIA-0095(99)/2, 2000a.

EIA [Energy Information Administration]. "The Changing Structure of the Electric Power Industry 2000: An Update." Energy Information Administration, U.S. Department of Energy: No. DOE/EIA-0562(00), 2000b.

EIA [Energy Information Administration]. "Annual Energy Review 2002." Energy Information Administration, U.S. Department of Energy: No. DOE/EIA-0384(2002), 2003a.

EIA [Energy Information Administration]. "Electric Power Annual 2002." Energy Information Administration, U.S. Department of Energy: No. DOE/EIA-0348(2002), 2003b.

EIA [Energy Information Administration]. "Monthly Energy Review, October 2003." Energy Information Administration, U.S. Department of Energy: Available on-line at http://www.eia.doe.gov/mer/, 2003c.

EIA [Energy Information Administration] "Electric Trade in the United States 1996." Energy Information Administration, U.S. Department of Energy: No. DOE/EIA-0531(96), 1998.

ELR [Environmental Law Reporter]. "Electric Utility Rate Design: The Move Toward Peak-Load Pricing." *Environmental Law Reporter*, 1975, *5*(ELR 10084).

Executive Office of the President, Office of Management and Budget. *Standard Industrial Classification Manual, 1987*, Indianapolis, IN: JIST Works, 1989.

Field, Barry C. and Charles Grebenstein. "Capital-Energy Substitution in U.S. Manufacturing." *The Review of Economics and Statistics*, 1980, *62*(2), pp. 207-212.

Foster, Lucia, John Haltiwanger and C. J. Krizan. "Aggregate Productivity Growth: Lessons from Microeconomic Evidence," in Edward Dean, Michael Harper and Charles Hulten, eds., *New Directions in Productivity Analysis*. Chicago: The University of Chicago Press, 2001, pp. 373-414.

Golan, Amos, Jeffrey M. Perloff, and Zhihua Shen. "Estimating a Demand System with Nonnegativity Constraints: Mexican Meat Demand." *The Review of Economics and Statistics*, 2001, *83*(3), pp. 541-550.

Goldman, M. Barry, Hayne E. Leland, and David S. Sibley. "Optimal Nonuniform Prices." *Review of Economic Studies*, 1984, *51*(2), pp. 305-319.

Gordon, Richard L. *Reforming the Regulation of Electric Utilities*. Lexington Books, Lexington, MA, 1982.

Griffin, James M. and Paul R. Gregory. "An Intercountry Translog Model of Energy Substituion Responses." *The American Economic Review*, 1976, *66*(5), pp. 845-857.

Halliday, David, Robert Resnick, and Kenneth S. Krane. *Physics*. Volume 2 (extended), 4th ed., John Wiley and Sons, New York, 1992.

Haltiwanger, John. "Measuring and Analyzing Aggregate Fluctuations: The Importance of Building from Microeconomic Evidence." *Federal Reserve Bank of St. Louis Economic Review*, 1997, *79*(3), pp. 55-77.

Halvorsen, Robert (1977). "Energy Substitution in U.S. Manufacturing." *The Review of Economics and Statistics*, 1977, *59*(4), pp. 381-388.

Hayashi, Paul M., Melanie Sevier and John M. Trapani. "Pricing Efficiency under Rate-of-Return Regulation: Some Empirical Evidence for the Electric Utility Industry." *Southern Economic Journal*, 1985, *51*(3), pp. 776-792.

Hirsh, Richard F. *Technology and Transformation in the American Electric Utility Industry*. Cambridge University Press, Cambridge, 1989.

Hirsh, Richard F. *Power Loss: The Origins of Deregulation and Restructuring in the American Electric Utility System*. Cambridge, Massachusetts: The MIT Press, 1999.

Hough, Richard and Stacey Cole. "Post-Stratification Methodology for the 2002 Manufactures Energy Consumption Survey." Unpublished Paper, 2004..

Hudson, Edward A. and Dale W. Jorgenson. "U.S. Energy Policy and Economic Growth, 1975-2000." *Bell Journal of Economics*, 1974, *5*(2), pp. 461-514.

Humphrey, David B. and J.R. Moroney. "Substitution among Capital, Labor, and Natural Resource Products in American Manufacturing." *The Journal of Political Economy*, 1975, *83*(1), pp.57-82.

Jarmin, Ron S. and Javier Miranda. "The Longitudinal Business Database." Center for Economic Studies, U.S. Census Bureau, CES Discussion Paper: CES-WP-02-17, 2002.

Joskow, Paul L. "The Difficult Transition to Competitive Electricity Markets in the United States," in J. Griffin and S. Puller, eds., *Electricity Deregulation: Choices and Challenges*, Chicago: University of Chicago Press, 2005.

Joskow, Paul L. "Regulatory Failure, Regulatory Reform, and Structural Change in the Electric Power Industry," in Martin Neil Baily and Clifford Winston, eds., *Brookings Papers on Economic Activity: Microeconomics*, Washington, D.C.: The Brookings Institution, 1989, pp. 125-208.

Joskow, Paul L. "Public Utility Regulatory Policies Act of 1978: Electric Utility Rate Reform." *Natural Resources Journal*, 1979, *19*(October), pp. 787-809.

Joskow, Paul L. and Richard Schmalensee. *Markets for Power: An Analysis of Electric Utility Deregulation*. Cambridge, Massachusetts: The MIT Press, 1983.

Meyer, Robert A. and Hayne E. Leland. "The Effectiveness of Price Regulation." *Review of Economics and Statistics*, 1980, *62*(4), pp. 555-566.

Morrison, Catherine. "Energy and Capital: Further Exploration of E-K Interactions and Economic Performance." *The Energy Journal*, 1993, *14*(1), pp. 217-243.

Munson, Charles L. and Meir J. Rosenblatt. "Theories and Realities of Quantity Discounts: An Exploratory Study," *Production and Operations Management*, 1998, *7*(4), pp. 352-369.

Nelson, Jon P., Mark J. Roberts and Emsley P. Tromp. "An Analysis of Ramsey Pricing in Electric Utilities," in Michael Crew, ed., *Regulating Utilities in an Era of Deregulation*. New York: Macmillan, 1987, pp.85-110.

Nguyen, Sang V. and Stephen H. Andrews. "The Effect of Energy Aggregation on Energy Elasticities: Some Evidence from U.S. Manufacturing Data." *The Energy Journal*, 1989, *10*(1), pp. 149-156.

Nguyen, Sang V. and Mary L. Streitwieser. "Capital-Energy Substitution Revisited: New Evidence from Micro Data." Center for Economic Studies, U.S. Census Bureau, CES Discussion Paper: CES-WP-97-4, 1997.

Peltzman, Sam. "Pricing in Public and Private Enterprises: Electric Utilities in the United States." *Journal of Law and Economics*, 1971, *14* (April), pp. 109-147.

Solow, John L. "The Capital-Energy Complementarity Debate Revisited." *The American Economic Review*, 1987, *77*(4), pp. 605-614.

Taylor, Peter (1981). "Estimating Price Effects on Input-Output Coefficients." Ph.D. diss. University of Maryland, 1981.

U.S. Department of Commerce, U.S. Bureau of the Census. *Census of Manufactures, 1972*, Washington D.C.: U.S. Government Printing Office, 1976.

U.S. Department of Commerce, U.S. Bureau of the Census. *Census of Manufactures, 1967*, Washington D.C.: U.S. Government Printing Office, 1971.

U.S. Department of Commerce, U.S. Bureau of the Census. *Census of Manufactures, 1963*, Washington D.C.: U.S. Government Printing Office, 1966.

Van Biesebroeck, Johannes. "Robustness of Productivity Estimates." National Bureau of Economic Research, Inc., NBER Working Papers: No. 10303, 2004.

White, Matthew W. "Power Struggles: Explaining Deregulatory Reform in Electricity Markets," in Martin Neil Baily, Peter C. Reiss and Clifford Winston, eds., *Brookings Papers on Economic Activity: Microeconomics*, Washington, D.C.: The Brookings Institution, 1996, pp. 201-225.

Wilson, Robert B. *Nonlinear Pricing*. New York: Oxford University Press, 1993.

Woodland, Alan D. "A Micro-Economic Analysis of the Industrial Demand for Energy in NSW." *Energy Journal*, *14*(2), 1993.

Yen, Steven T., Biing-Hwan Lin, and David M. Smallwood. "Quasi- and Simulated-Likelihood Approaches to Censored Demand Systems: Food Consumption by Food Stamp Recipients in the United States." *American Journal of Agricultural Economics*, *85*(2), 2003.