APPENDIX 1

DETAILED REVIEW OF THE NATURAL GAS DEMAND MODELING LITERATURE

General Issues in Modeling Demand and Supply

Modeling natural gas demand and supply in local, regional, and national markets is important for a number of reasons. These models give researchers and other market observers information about the structure and composition of demand and supply. Futhermore, the results of these models inform users about the magnitude of future demand and its sensitivity to key determinants such as energy prices and income. This information is used to understand:

- Past trends and the determinants of realized demand and supply;
- The responsiveness of demand and supply to changes in its important determinants; and
- Future demand and supply under different assumptions about future scenarios.

From its most basic perspective, the relationships of demand and supply can be summarized as:

- Demand is a function of prices, income, and tastes and preferences; and
- Supply is a function of input factor prices, technology, and other factors.

Transforming these theoretical relationships into measurable statistical equations is difficult. The way empirical data is measured may not conform with the structure implied by theory. For instance, theory suggests that the quantity demanded is a function of prices and other important variables. Yet the "appropriate" prices may not be readily available or easily generated. Furthermore, in many energy pricing situations, prices are set in a multitude of different manners (i.e, average rates, two-part tariffs, increasing block rates, decreasing block rates, time of day and seasonal pricing, etc.) Data measurement problems in terms of definition, sampling, and aggregation complicate model specification and statistical estimation.

Most quantitative analyses of supply and demand are broken into two types of models: cross sectional and time series. Cross sectional models typically examine causal relationships across a collection of variables over a fixed period of time. As suggested by the nomenclature, time series models focus on time dependency.

Cross sectional models are used to examine existing determinants of either supply and demand. These models are structural in nature since they attempt to flush out causality and typicially employ many different determinants of demand or supply as independent (explanatory) variables. Thus, a model of the industrial demand for energy could consider a number of different explanatory factors that include economic characteristics (i.e., relative energy prices, output levels, etc.) and technical characteristics of the facilities (i.e., number of boilers, fuel switching abilities, heat to power ratios, etc.).

Cross sectional models provide useful information on the relative statistical importance of these variables at a given period of time but are less useful in estimating how relationships change over time. Thus, their ability to serve as a springboard for forecasting is limited. In addition, these types of approaches usually require detailed disaggregate information (usually at the firm or production unit level), that can be difficult to acquire, particularly for independent research.

Time series models, on the other hand, are more useful in examining the dynamic determinants of demand or supply. The advantage of time series models is that they can convey information about how supply or demand relationships have varied historically and where particular “structural breaks” in certain trends have occurred. These models are equally useful as a starting point for forecasting since most forecasts are developed from historical trend relationships. Their disadvantage is that data availability usually limits the range of the determinants measuring the supply or demand relationship.

Another consideration in time series models is that they can be developed in two different fashions. The first is traditionally referred to as an “econometric” approach while the second is commonly referred to more generally as a “time series” approach.[1] The econometric approach is concerned with the estimation of relationships suggested by economic theory across time. For instance, in a time series-based analysis might look at the relationship of energy demand relative to prices, income, weather, and other relevant variables. Such models serve two purposes. First, they allow economic hypotheses to be tested empirically.[2] Second, they provide a framework for making rational and consistent predictions (i.e., forecasting).

Pure time series approaches, on the other hand, are more generalized trend analyses based on statistical extrapolation techniques rather than theoretic

[1]A seminal text on the econometric analysis of time series is Andrew Harvey. (1991) *The Econometric Analysis of Time Series*. Second Edition. Cambridge, Massachusetts: The MIT Press.

[2]*Ibid.*, 1.

relationships. Traditional time series analyses forecast the time path of a variable with models that explicitly contain stochastic components to measure their dynamic relationships.[3] Difference equations, such as moving averages of either the error term, the dependent variable, or both, are at the core of these types of approaches. Uncovering the dynamic path of a series improves forecasts since the measurable components of the series can be extrapolated into the future.

There is a third option in facilitating what is known as cross-sectional/time series models. These approaches, as the name suggests, merge these two approaches to maximize the relative benefits, and minimize their relative shortcomings. The problem with these approaches is that, in many instances, they require relatively advanced statistical techniques, as well as being very data intensive.

Another important question in measuring either supply or demand relationships is the determination of which of the two general approaches should be facilitated. In many instances, this is usually done by purpose of the study as well as the practical limitations of the data. If a researcher is interested in examining the price elasticity of the residential demand for natural gas, then a cross sectional analysis of account-specific information would be a useful approach. However, many researchers outside of natural gas local distribution companies usually have limited to no access to this type of information. The US Department of Energy, however, does report aggregate information by customer class across time, thus some type of time series approach may be more readily facilitated.

Lastly, determining the appropriateness of a particular model is an important specification issue. Often, applied modeling can emphasize goodness of fit of a particular model to the expense of all other considerations. However, more balanced consideration should include such factors as:

- *Consistency with theory*. Ensuring the quantitative estimates of model parameters exhibit mathematical signs and magnitudes consistent with economic theory (i.e., negative price elasticities and positive income elasticities).

- *Consistency with goals*. Obviously specifying and measuring time series models can be more important for forecasting goals, while cross sectional models can be more important for research questions related to the relative importance of structural determinants.

- *Parsimony*. Ensuring that models that are not overly specified and are straightforward.

[3] Walter Enders. (1995). *Applied Econometric Time Series*. New York: John Wiley and Sons, Inc.

- *Robustness*. Ensuring that models are not overly dependent upon unique specifications or time periods under consideration.

The modeling of the demand for natural gas builds on a broad arena of industry-based energy modeling. The study of natural gas demand can also be linked to technical-engineering models, sociological models, economic models, and hybrid models that employ varying combinations of these factors. Econometric analysis, as opposed to time series approaches, has dominated much of the demand modeling literature. The preference for these econometric approaches is probably to be expected. First, econometric approaches are useful in explaining the changes in natural gas disposition that result from general changes in the industry—particularly, the response to shifts in price and the general degree of price volatility in the industry since the early 1970s.

Second, while data measurement and implementation is still a challenge in the analysis of energy demand, accessibility of the information has improved considerably. Reporting requirements and data collection developed at the U.S. Department of Energy gives researchers a consistent source of information to examine and corroborate existing studies in the energy industry. With the advent of the internet, the electronic availability of the information enhances the ability to concentrated important efforts in understanding empirical relationships rather than collecting basic information on industry disposition and trends.

Third, over the past twenty years, econometric approaches have become more accessible to industry practitioners as software packages have reduced the programming work needed to do the earlier models by an exceptional order of magnitude. Today, many readily available statistical packages can estimate either supply or demand models in matter of seconds. The reduction in computational difficulty has helped facilitate the development of a large body of analysis related to important energy relationships.

Empirical Studies of Natural Gas Demand

One of the pioneering authors in demand modeling, for many sectors that go beyond just energy demand modeling, is Hendrick S. Houthakker. His studies in energy demand modeling were extensive, and provided some of the first insights into the importance many structural determinants of energy demand. His work is still commonly cited in principals textbooks of microeconomic theory.[4]

[4]Hendrick S. Houthakker and Lester D. Taylor. (1966). *Consumer Demand in the United States, 1929-1970*. Cambridge: Harvard University Press.

Houthakker's work in energy demand modeling, developed in the early 1950s, was a basis for his broader work in overall demand modeling.[5]

On the more practical side, there is a considerable amount of work in natural gas demand modeling that rests outside the traditional academic literature. This work is associated with the modeling conducted within the process of regulated natural gas distribution companies, commonly referred to as local distribution companies or LDCs. These LDCs use forecasting models for internal planning process in meeting supply (commodity) and capacity (transportation and storage) needs.[6]

Many of the theoretic developments of natural gas demand modeling have come from the academic literature. A good portion of this analysis has focused on residential, and to a lesser degree commercial, demand for natural gas. These models are primarily econometric in nature since the purpose of many are to get accurate estimates of price, income, and weather related sensitivities of natural gas demand.

Another practical consideration in reviewing the literature on natural gas modeling is its relationship with its sister energy industry, electricity. A number of the earliest works in energy demand concentrated in the area of electricity (i.e., Houthakker) and not natural gas. It seems likely that one of the initial reasons for more comprehensive development of demand modeling in the electricity industry is associated with its greater degree of data availability. Thus, any survey of natural gas demand modeling will have to include some references to the development in the power industry as well.

There are a number of surveys in the literature dedicated to natural gas and energy demand modeling in general. One of the earliest and most comprehensive surveys of energy demand modeling was prepared by Douglas R. Bohi for the Electric Power Research Institute (EPRI).[7] While the overall purpose of the study was to examine price elasticities, the study is an excellent overview of demand modeling since price elasticities are usually outputs derived from an overall analysis of demand determinants. An update to this study was prepared in 1984 by Bohi and Zimmerman.[8]

---

[5]For instance see: Hendrick S. Houthakker. (1951), "Some Calculations of Electricity Consumption in Great Britain." *Journal of the Royal Statistical Socie*ty. Series A, 114, Part III, 351-71.

[6]A general primer on the role of natural gas demand forecasting and how it relates to overall LDC planning can be found in: Charles Goldman, et al. (1993). *Primer on Gas Integrated Resource Planning*. Berkeley, California: Lawrence Berkeley Laboratories.

[7]Douglas R. Bohi. *Price Elasticities of Demand for Energy: Evaluating the Estimates*. Palo Alto: Electric Power Research Institute.

[8]Douglas R. Bohi and Martin B. Zimmerman. (1984). "An Update on Econometric Studies of Energy Demand Behavior." *Annual Review of Energy*. 9: 105-54.

A more recent study, which emphasizes the development of the literature in residential energy demand modeling, was presented by Reinhard Madlener.[9] In the survey, Madlener attempts to update the earlier Bohi work, as well as breaking the existing econometric literature into a number of useful different categories. These include studies associated with log-linear functional forms, transcendental logarithmic (translog) functional forms, qualitative choice models (also know as discrete choice models), household production theory (end-use modeling), and pooled time series-cross sectional models.

This survey will follow the same lines as Madlener, since it provides such a useful frame of reference to consider the development of energy demand modeling. The following survey will differ, however, by placing a larger explanation on the methods and their advantages, and highlighting in more detail, the seminal pieces of literature within each of these modeling categories. This survey will also concentrate on the more generalized areas of: log linear and double log models, transcendental logarithmic (translog) functional forms, qualitative choice and end-use models (also know as discrete choice models).

Log-Linear and Double Log Models

The typical log-linear and double log models are relatively straightforward and tend to be the model of choice, particularly for industry practitioners. This model generally takes the form:

$$\log D = \beta_0 + \beta_1 P + \beta_2 Y + \beta_3 W + \beta_4 X \qquad \text{(eq. A.2.1)}$$

$$\log D = \beta_0 + \beta_1 \log P + \beta_2 \log Y + \beta_3 \log W + \beta_4 \log X \qquad \text{(eq. A.2.2)}$$

Where:

D = Natural gas demand
P = Price of natural gas
Y = Income
W = Weather
X = Other structural variables influencing demand
β = Estimated parameters.

The benefit of the log-linear and double log form is that coefficients can easily be translated into elasticities. In the double log form presented in equation A.2.2, the parameter for price is interpreted as the price elasticity of demand, while the parameter estimate for income can be interpreted as the income elasticity of demand.

[9] Reinhard Madlener. Econometric Analysis of Residential Energy Demand: A Survey. *Journal of Energy Literature*. 2:3-32.

The log-linear literature starts with Houthakker and continues with Balestra and Nerlove (1966), who suggested a dynamic approach to the modeling of the demand for natural gas. This model contained a pooled cross sectional approach to modeling natural gas demand since it examined residential households, across several different regions, across time. The model is important since it uses an error-components specification and demonstrates the importance of relative fuel prices in determining both natural gas demand and fuel substitution.

For instance, in their study, Balestra and Nerlove assumed that the new demand for gas was a function of the relative price of gas and the total new requirements for all types of fuel. The problem with this approach was that the concept of new energy demand was difficult to translate into observable variables. The total new demand appeared as the sum of the incremental change in consumption and "replacement" demand, which represented the portion of the total demand for fuel "freed" by the retirement and replacement of old appliances. Specific equations were developed for each type of demand model, and ultimately fed into a larger equation examining total fuel use.

This total fuel use equation facilitated data from 1950 through 1962. The fuel use variables and price information was standardized into a Btu equivalent. Usage was normalized for weather in each state, and prices and income were measured in constant dollars. There were 13 observations per state, though only 36 states had gas service over the entire period. All states were grouped together and estimations were performed on the combined sample of cross sectional and time series data. Additional equations were estimated using dummy variables for each state.

While the estimation results presented negative and significant results for the impact of own price changes on energy demand, the greatest statistical significance rested with the state-specific dummy variables. The results would tend to suggest that there were a number of state-specific implications for energy usage that could not be directly modeled (i.e., regulation, etc.) The overall predictive capabilities of the model were very good, with 99 percent of the demand for natural gas being explained by the model's independent variables.

Because the demand function was for new gas demand, the average price elasticity was attainable from the model results. According to Balestra and Nerlove, the estimated average price elasticity of new gas demand ranged from –0.58 to –0.69 given the various functional forms estimated.

Beierlein, Dunn, and McConnon (1981) took the general framework discussed by Balestra and Nerlove and applied a Cobb-Douglas framework which has a double-log component. Their specification for energy demand included specific equations for fuel oil, natural gas, and electricity. This model is also a pooled cross-section approach since it examined energy usage across fuel type, state, customer class, including residential, commercial, and industrial, and year.

The independent variables were the average deflated price of gas per 1000 therms, the average deflated price per kWh of electricity, the average deflated price per gallon of fuel oil, lagged per capita fuel consumption, and per capita deflated income represented by disposable personal income, value of retail sales, and value added by manufacturing.

The model facilitated an error component and error component/seemingly unrelated regression (SUR) approach. The Cobb-Douglas framework allowed for constant elasticity of substitution, thus the estimated parameters for price, were the elasticities for each variable. The estimated own-price elasticity of gas for the residential sector was between –0.23 and –0.35 depending on the technique and between –0.61 and –0.63 for the natural gas industrial sector. The fit on the estimations showed that between 94 and 99 percent of the variation in the fuel consumption by various sectors was attributable to their respective independent variables.

The MacAvoy-Pindyck (M-P) model (1973) used similar techniques in what was a basically a demand component in a supply model.[10] In the demand module of this model, MacAvoy and Pindyck focused on wholesale natural gas markets. Supply of production out of reserves had to be measured against demand for the production after it had been transmitted to wholesale markets by pipelines, and the quantity demanded by direct industrial consumers as well as retail consumers.

MacAvoy and Pindyck modeled demand as a function of the prices for wholesale gas contracts, the prices for alternative fuels consumed by the final buyers, and economy-wide variables that determined the overall size of energy markets. For the model, the demands for production were approximated by curves fitted on a disaggregated basis into wholesale equations for (1) gas sales for resale,[11] (2) gas sales directly off the pipelines for final consumption (mainline sales), and (3) intrastate sales by producers and pipelines to final consumers. The wholesale prices of gas were computed by adding a markup to the field price based on (1) mileage between the production district and the consuming region, and (2) volumetric capacity of the pipeline.

Before the wholesale demand equations were estimated, the M-P model looked at wholesale price markups. Markups over field prices were a function of mileage and volumetric capacity of the lines transmitting to each region. These field prices were the rolled in wellhead price for the wholesale region under

[10] The discussion of the supply model can be found in the later section of this chapter on supply modeling.

[11] Split in to commercial-residential gas and industrial gas on the basis of percentages distributed to those two groups for ultimate consumption.

investigation. The coefficient of volumetric capacity as determined by the M-P model was negative, as a larger capacity implies lower average costs. The fit of the estimated equation[12] showed that 56 percent of the variation in wholesale price of gas sales for resale could be explained the variation in the independent variables.

Gas sales for resale were broken down in to gas that ultimately is resold for residential and commercial consumption and gas for industrial consumption and the M-P model had a separate equation for each category for each of the five regions of the country. For each of these equations, new or additional demand was used as the dependant variable. The M-P model assumes that all fuel-burning equipment had an average lifespan of 14 years and chose a depreciation rate *r* equal to 0.07. Independent variables in the models included average wholesale price of gas, the wholesale price of oil, income, population, value added in manufacturing, capital investment by industry, and a price index of alternative fuels. In the South Central, Southeast, and West regions the residential and commercial sales were aggregated with industrial sales to make up for lack of stable elasticity estimates in the disaggregated form. All equations were estimated over the years 1964 through 1970.

Similar equations were developed for Northeastern region on a specific user basis. Results showed that an increase in the price of oil increases the demand for gas, additional units of value added in manufacturing increased the demand for natural gas, and additional units of capital investment increased the demand for natural gas.

MacAvoy and Pindyck, instead of using gas price for the current year, used the average wholesale price of gas for the previous two years and also did the same for the wholesale price of oil. The fit of this equation showed that 90 percent of the variation in total demand for the region was attributable to variation in the independent variables.

As noted earlier, additional units of capital investment in industry increased the total demand for natural gas. The fit of the equation showed that 80 percent of the variation in total demand for the region was attributable to the variation in the independent variables. The equation for Southeast-residential and commercial revealed that the coefficient for income is positive, which meant that additional units of income would increase the region's residential and commercial demand for natural gas. The fit of the equation showed that 26.7 percent of the variation in residential and commercial demand for the region was attributable to the variation in the independent variables. The final regional gas sales for resale equation, that for Southeast-industrial demand, revealed that the coefficients for the price index for alternative fuels and value added in manufacturing exhibited a positive relationship. The fit of the equation showed that 37.3 percent of the

[12] As taken from each equation's R^2 values.

variation in the industrial demand for the region was attributable to the variation in the independent variables.

The quantity of mainline sales to industrial buyers was estimated. The wholesale price for mainline sales was represented by the average of the wholesale price in the current year *t* and the previous year *t-1*. The same operation was also performed on the price index of alternative fuels. The coefficient of the price index of alternative fuels showed that an increase in the price index led to an increase in the quantity of mainline sales. The fit of this equation showed that only 15 percent of the variation in the quantity of mainline sales was attributable to the variation in the independent variables.

Finally, the quantity of intrastate demand was estimated. Like the mainline sales equation, the wholesale price of gas was represented by the average of the wholesale price for current year *t* and the previous year *t-1*. The fit of this equation showed that 21 percent of the variation in the quantity of intrastate demand was attributable to the variation in the independent variables.

Six of the ten demand equations had significant coefficients for the negative price effects on demand, with the strongest effects in regions closer to producing centers with more alternative sources of energy. MacAvoy and Pindyck concluded that size-of-market variables such as consumer incomes or industrial investment did not appear to be causal factors in all sectors of the natural gas market.

MacAvoy and Pindyck also calculated interregional flows of gas in order to be able to calculate excess demand of consuming regions. Estimates of interchange at an aggregate level were made using the five demand regions, West, Northeast, North Central, Southeast, and South Central, and eight production regions. Total flow, the fraction of a consuming region's demand which comes from a particular production region, and the fraction of gas from a production region going to a particular consuming region were calculated. Demand was forecasted for the period 1966 through 1970, and the mean demand error[13] was –2.5 Tcf with an RMS[14] demand error of 2.5. Estimated demand quantities for each year were about 13 percent lower than the actual values.

Lyness (1984) developed a gas demand forecasting model which focused on the temperature-gas demand relationship. He identified three regular cyclical patterns in gas demand: (1) the diurnal swing during each day, which had peaks at breakfast time and the evening and a trough during the night, (2) a weekly cycle, and (3) an annual cycle related to seasonal changes in temperature. All

[13] Mean error is the average of the errors of the predicted values. The error of a predicted value is calculated by subtracting the actual value from the predicted value.

[14] RMS error, or Root Mean Square error, is simply a quantitative measure of the deviation of model predictions from actual observations. Smaller RMS error is better.

three cycles were superimposed on each other and were treated as being related.

Lyness forecasted long-term demand almost exclusive on temperature and the underlying concept of seasonal normal temperature (SNT). For each day of the year a long-run average temperature could be derived and those could be smoothed to form a sinusoidal curve for the entire year. Thus daily, weekly, or monthly SNT's were known in advance and the forecast of demand for the remainder of the year was obtained through the insertion of the appropriate SNT values into the current forecast demand and temperature relationship.

While he provided no specific model for the forecasting of temperature, Lyness did provide two ways to look at this variable. The approach considered, within a linear framework, a number of different seasonal, daily, and temperature influences on natural gas demand. Lyness left the addition of market data to the individual modeler, as different regions had different market conditions and thus market variables. The model was broken down in to separate equations corresponding to the market sectors. For each forecast year, parameters in each market sector equation were scaled in the ratio of the forecast annual market sector demand to the current market sector demand and then re-aggregated to arrive at an equation for the forecast year that was consistent with the total forecast demand for that year.

Herbert and Kriel (1989) built on the studies by Beierlein (1981), Grady (1986), Green (1987), Blattenberger (1983), and Lin (1987) by creating a natural gas demand model which incorporated both heating degree day data as well as wealth data, and estimated the model based on monthly information. The main equation in the model estimated monthly aggregate residential sales as the function of six variables: (1) the index of changes in total personal income in constant dollars received by gas customers and changes in the number of gas customers, (2) heating degree days weighted by gas residential space-heating customers, (3) cooling degree days weighted by population, (4) household wealth in constant dollars measured by financial and non-financial asset holdings, (5) the price index of natural gas in constant dollars, and (6) the seasonal shift in residential gas demand for the one-month period from mid-December to mid-January.

Weighted heating degree days were indexed to changes in the percentage of space-heating to total gas customers. The authors also estimated regressions for real wealth, which was a function of time, and real personal income, which was a function of the number of residential customers in a given quarter and Census Division and personal income in a given quarter and Census Division.

The fit of the estimated equation showed that 99 percent of the variation in monthly aggregate residential sales was attributable to the variation in the independent variables. The model was used to forecast values for the year

1984, and the mean error was 217 Mcf and individual differences ranged from one percent to five percent.

Hsing (1992) built on the work of Taylor (1977), Blattenberger (1983), and Griffin (1979) in an exercise for estimating the own-price and income elasticities of natural gas for each of the 50 states except Hawaii for the year 1989. The model had the demand for natural gas for each state in a given period as its dependent variable. The independent variables included the price of natural gas, disposable income per capita, the price of residential electricity, and the number of heating degree days. The model also included dummy variables for the South (SO) and West (WE) as well as the years 1985 and 1986 but no reason is given for these inclusions.

Hsing estimated the elasticities from the results of the linear regression of the model. His results included Alaska-specific estimates of –0.29 for the price elasticity of demand and 0.37 for the income elasticity of demand.

Transcendental Logarithmic (Translog) Models

Translog models became popular in the 1960s with the advent of the Christensen, et al. (1973) approach of estimating industrial production, and later with utility functions.[15] This approach was applied to the electric power industry in 1976, and the approach has become commonplace for a considerable amount of energy economics research.[16]

The translog specification is a quadratic function with its elements expressed in terms of their natural logarithm. This specification is a second order approximation around a given point for the Cobb-Douglas production function. The Cobb-Douglas production function is a flexible functional form for a production function that allows declining marginal products for all inputs, and also assumes that opportunities exist to substitute inputs in production without gaining or losing output.

The advantage of the translog approach is that it provides some structure on the assumed production/utility function under investigation. The parameters associated with the own and cross-price terms provide estimates of own and cross-price elasticities of demand. In additional, the translog approach allows for a more flexible functional form that enables empirical validation of utility-function properties. For example, while the Cobb Douglas function imposes unitary elasticity of substitution among inputs, the translog enables the data to determine

[15]Laurits Christensen, Dale Jorgenson, and Lawrence Lau. (1973) “Transcendental Logarithmic Production Frontiers.” *The Review of Economics and Statistics*. 55:28-45. Laurits Christensen, Dale Jorgenson, and Lawrence Lau. (1975) “Transcendental Logarithmic Utility Functions.” *The American Economic Review* 65: 367-83.

[16]Laurits Christensen and William Greene. (1976). “Economies of Scale in U.S. Electric Power Generation.” *Journal of Political Economy*. 84 (4): 655-76.

the degree of input substitutability. In general, this flexible functional form enables the data to determine if the assumed functional form is correct, and imposes fewer a-priori restrictions on model specification.

The approach, however, is not without its potential problems. First, translog models require a significant amount of information which can be difficult to attain. Second, these models can be relatively difficult to apply and interpret. This has led many practitioners to steer clear of these approaches. Third, the parameter estimates in many instances do not tend to be robust or stable, and can lead to some erroneous results. Last, the model tends to lend itself better to cross-sectional analyses, and, as a result, is not a very useful tool for forecasting.

The translog specification[17], usually takes the form:

$$\log D = \beta_0 + \beta_1 \log P + \beta_{11}(\log P)^2 + \beta_{12}(\log P)(\log Y) + \beta_{13}(\log P)(\log W) + \beta_{14}(\log P)(\log X) + \beta_2 \log Y + \beta_{22}(\log Y)^2 + \beta_{23}(\log Y)(\log W) + \beta_{24}(\log Y)(\log X) + \beta_3 \log W + \beta_{33}(\log W)^2 + \beta_{34}(\log W)(\log X) + \beta_4 \log X + \beta_{44}(\log X)^2$$

(eq. A.2.3)

Where:

D = Natural gas demand
P = Price of natural gas
Y = Income
W = Weather
X = Other structural variables influencing demand
β = Estimated parameters.

Christensen and Jorgensen introduced the translog approach in 1969 and then again with Lau in 1973, and Pindyck (1979) used the approach extensively to analyze demand in his work on world energy demand. Estrada and Fugleberg (1989) took Pindyck's work and applied it to the natural gas markets in West Germany and France in order to determine own-price and cross-price elasticities of demand. Using a translog equation based on Pindyck's, Estrada and Fugleberg estimated a number of equations for the household and commercial sectors:

The resulting equations included estimates with lagged price variables in order to test the underlying hypothesis that long-term changes in the composition of energy demand were the result of changes in relative fuel prices, infrastructural changes in the economy, and the technology incorporated in equipment used to consume different fuels. The authors hypothesized that the response to an increase in the relative prices of fuels would take one to two years as consumers replaced their old equipment with types that were more energy efficient.

[17] From Brynjolfsson and Hitt (1995).

The actual estimation of the elasticities was done using a two-step process, the first of which was the calculation of partial own-price and cross-price elasticities: The second step was to incorporate the partial elasticities in to equations for total elasticities. The authors found that the own price elasticity for gas was much higher in Germany and believe that this was because the German government did not regulate prices as much as the French, and changes in fuel costs were more rapidly reflected in consumer prices.

Qualitative Choice and End Use Models

Most demand models prior to the early to mid 1970s, and even to this day, facilitate continuous variables for consumption. There are equally interesting empirical applications that examine not how much of a particular resource is utilized, but whether or not that resource is utilized at all. Such approaches are discrete in nature and have led to the development of qualitative choice, or discrete choice models of energy usage.

Discrete choice models are those in which the dependent variable is a discrete variable. The simplest application is one where the dependent variable is a binary choice variable that represents a simple positive or negative response. The dependent variable takes the value 1 if the choice is made, and 0 if the choice is not made. Independent variables are then used to estimate parameters influencing that choice.

Consider a generalized binary choice model that takes the form:

$$y = x\beta + e \qquad \text{(eq. A.2.4)}$$

Where:

> y = A discrete variable (eg. gas heating) that takes the value 1 if the choice is made, 0 otherwise
>
> x = A matrix of explanatory variables, such as characteristics of the alternatives or socioeconomic variables
>
> β = A vector of parameter estimates
>
> e = A sequence of error terms which can take either logistic or normal distribution

Discrete choice models can be powerful tools to examine individual customer choice behavior and the factors influencing those decisions. Sensitivities, developed through the calculation of odds ratio statistics, can then be derived. These odds ratio statistics given some indication on how the probability of making a particular discrete energy consumption decision change as the independent variables change.

These qualitative based models, however, usually require specific and relatively comprehensive end use information. Typically, data used in these types of analyses are from individual consumer surveys. Thus, such empirical approaches are limited, if customer, or decision making unit information is not available. In addition, these types of models can tend to be more static in nature making it difficult to use for long forecasting and trend analysis.

Some of the representative works in this area include the work of the State Utility Forecasting Group (1999) in Indiana, which used a logit form of discrete choice model to determine fuel choice among residential energy consumers. The dependent variable of the model was the ratio of electricity's share of the space heating market to that of all other fuels. Market share was used because it captured current activity, was independent of the rate of customer growth, and exhibited greater year-to-year variation than measures of market saturation. The group used a double-log functional form of the logit model, which allowed for easy calculation of elasticities. The national energy outlook model released by the Energy Information Administration (2001) also used discrete choice modeling for fuel choice components of the overall model.

Table A.2.1. Summary of the Strengths and Weaknesses of Modeling Approaches

Approach	*Strengths*	*Weaknesses*
Log-linear/double-log	1) Relatively easy to specify and estimate 2) Estimated coefficients are directly interpretable as short-run elasticities, and long-run elasticities are easy to calculate 3) Estimated standard errors provide measure of the variability of the estimated elasticities	1) Constant elasticity assumption often unrealistic and not justifiable 2) Sometimes problems of consistency with the underlying economic theory 3) Appropriate only when one has reason to believe that the variables enter multiplicatively in to the equation
Translog	1) Imposes a minimum of restrictions on demand behavior and is very flexible 2) Firmly based in economic theory 3) Particular demand characteristics are testable (eg. separability, homotheticity, etc.) 4) Allows the analysis of substitutional relations	1) Sometimes lack degrees of freedom due to the large number of regressors 2) Only well-behaved for a limited range of relative prices 3) Estimated elasticities are not directly interpretable 4) More complicated estimation techniques are required 5) Static formulations dominate
Qualitative choice	1) Appropriate when dependent variable comprises a finite set of discrete alternatives 2) Relatively easy to estimate 3) Flexible specification 4) Tobit models allow for observations to equal zero	1) Inefficient estimates in the case of zeros (logit, probit) 2) Theoretically not based on assumptions of utility maximization (logit) 3) Relies on rich and reliable data sets
Pooled time series/cross-section	1) Pooling enables greater efficiency of the estimates	2) Only makes sense if the cross-sectional parameters are constant over time 3) Difficult specification

Source: Madlener (1996)