Graduate Theses, Dissertations, and Problem Reports

1999

Forecasting electricity demand using regression and Monte Carlo simulation under conditions of insufficient data

Kathleen Ann Cullen
West Virginia University

Follow this and additional works at: https://researchrepository.wvu.edu/etd

Forecasting Electricity Demand using Regression and Monte Carlo Simulation Under Conditions of Insufficient Data

Kathleen Ann Cullen

Thesis submitted to the
College of Agriculture, Forestry, and Consumer Sciences
At West Virginia University
In partial fulfillment of the requirements for the degree of

Master of Science
In
Agriculture and Natural Resource Economics

Thomas F. Torries, Ph.D., Chair
Peter V. Schaeffer, Ph.D.
Gerard D'Souza, Ph.D.

Division of Resource Management

Morgantown, West Virginia
1999

Abstract

Forecasting Electricity Demand using Regression and Monte Carlo Simulation Under Conditions of Insufficient Data

Kathleen Ann Cullen

The problem studied is that of a summer peak residential energy demand model for Appalachian Power Company's service area in West Virginia. By restricting the forecast to a region smaller than the state, serious data problems result due to insufficient data to obtain reliable forecasts.

Regression analysis and Monte Carlo Simulation are the two methods used to forecast energy demand. Both methods incorporate risk into the analysis in different ways. Regression analysis yields a measure of the reliability of the coefficients of the variables and of the reliability of the forecast. The resulting forecast and confidence limits of the forecast values give an indication of the risk using regression analysis and Monte Carlo Simulation. Monte Carlo Simulation uses a probabilistic range of input values rather than a single discrete value, which accounts for future uncertainty to determine the probabilistic future summer peak.

Acknowledgments

I would like to thank my committee, Dr. Torries, Dr. Schaeffer, and Dr. D'Souza for their support, encouragement, and understanding during the long thesis process. A special thank you to Dr. Torries for his gentle nudges when I need it to make sure I finished my thesis and for all of his help when I needed it the most.

I would like to thank my supervisor at the West Virginia Public Service Commission, Michael Fletcher, for the concept of my thesis which is an expansion on the work I did there. The data collection and research I did there during the summer of 1996 helped a great deal with my thesis.

I would like to thank my supervisors at the National Research Center for Coal and Energy, Carl Irwin and Frank Saus, for their support and for giving me the time to work on my thesis when I needed it. I would also like to thank all of my friends and co-workers at the National Research Center for Coal and Energy for their support and encouragement.

A big thank you to my friends Mary, Sarah, Beth, Dan, and Elise for their help and support. You kept me laughing and listened to my problems when I needed it during the trials of writing my thesis.

A very special thank you to my family for their constant support, love, understanding, and patience while I was writing my thesis. Your support helped me when I needed it the most.

Table of Contents

List of Tables

List of Figures

1.0   Introduction

Energy demand forecasting is an essential activity of electricity providers. Without an accurate picture of the future, which may be based upon the past, over capacity or shortages in power may result producing unexpected high costs. An accurate forecast requires adequate data, for without it, the results of the forecast are unreliable. However, the data available for regional electricity demand studies are often insufficient to meet the needs of the forecast. This lack of data presents a number of theoretical and practical problems.

Baum (1993) says that forecasters believe electricity will remain one of the fastest growing energy sectors. A large portion of the increase in demand for electricity is due to an increase in residential demand. With the overall increase in electricity demand in the residential sector combined with the increased use of electricity for residential cooling during the summer months, a summer peak residential model is needed to avoid future shortages or to determine how much electricity is available for sale to other companies.

A model to determine residential summer peak electricity demand is normally a function of variables that measure temperature, electricity price, the price of substitutes, such as natural gas price, the number of customers, and economic indicators, such as income. Such a model for Appalachian Power Company's service area in West Virginia is developed in this research.

Data were collected to identify and to test the importance of each independent variable that contributes to electricity demand for the West Virginia service area of Appalachian Power Company. For a number of reasons, data are inadequate to easily construct a statistically rigorous forecasting model. Therefore, forecasting energy demand with insufficient data is analyzed using two evaluation methods.

A forecast model may be constructed by using either regression analysis or Monte Carlo simulation. Both methods arrive at similar answers, but present the data and the results differently. Regression analysis produces only an expected value of future demand with a corresponding standard error of the estimate that indicates the precision of the forecast. Monte Carlo Simulation produces a range of values for future demand based on a range of values for the input variables. It produces a probabilistic picture of the distribution of future demand and the input variables in a manner that are sometimes obscured in regression analysis. Monte Carlo Simulation also provides a framework for decision making that easily incorporates the risk tolerance of investors, which then yields a certainty equivalence for electricity demand.

1.1   Problem Statement

According to Bartels and Fiebig (1996), the residential sector is usually one of the main contributors to electricity generation system peaks; the summer is no exception. The increase in electricity demand during the summer months is related to the recent increased use of cooling devices in West Virginia, such as air conditioning, to ward off the heat of the summer months. Le (1983) asserts that the net system summer peak load is strongly related to the air conditioning load that is caused by the weather effect of the summer months. Therefore, a summer peak residential energy demand model is needed to measure this new trend in energy demand.

A summer-peak model is needed because regional summer and winter demand in West Virginia are quite different. A summer peak energy model predicts the energy used on the days of greatest demand during the summer months—usually the warmest days. Forecasting peak residential energy demand will aid power companies in determining what changes will be needed

whether it is building new power plants or instituting conservation measures. The forecast will also aid in determining how much energy will be available for sale to other electric companies.

Appalachian Power Company supplies power to the southern part of West Virginia. This thesis develops a summer-peak residential demand model for energy usage by Appalachian Power Company's service area. Considering only the southern part of West Virginia presented serious data collection problems. Regional data were limited, and the more abundant state-level data were often inappropriate for use in the forecast model.

West Virginia is also developing its plan for the deregulation of the generation sector of the electricity industry, which adds an additional element of uncertainty to forecasts. Deregulation makes the need for accurate long-term forecasts even greater.

The forecast period considered in this paper is ten years, which is considered a long-term forecast. According to Kher, Sioshansi, and Sorooshian (1987), "energy is a long lead-time, capital intensive industry, hence the avid interest in long-term forecasting and planning" (p. 133).

The analysis of the long-term summer peak energy demand model for Appalachian Power Company's service area in West Virginia is accomplished by using both regression analysis and Monte Carlo Simulation. Both methods of forecasting energy demand reach the same goal, but the methods use different means to achieve the same goal. Regression analysis consists of the development of a model that studies the impact of the independent variables upon the dependent variable. The initial model for summer peak energy demand includes all variables that theoretically should have an impact upon energy demand—temperature, electricity price, price of substitutes, and economic indicators. Those variables that have no effect upon energy demand are identified and are omitted from future models as appropriate.

Regression analysis yields a measure of the reliability of the coefficients of the variables and the reliability of the forecast. The resulting forecast and confidence limits about the forecast values give an indication of the risk involved in using these forecasts. However, regression analysis presents only one view of the risk, while probabilistic analysis presents another view.

Monte Carlo Simulation is accomplished by using commercial risk analysis software. With probabilistic analysis all possible combinations of variations of input variables and the subsequent outcomes are considered simultaneously. Therefore, a probabilistic range of input values is used to determine the probabilistic future level of the summer peak. The reason for this use of a range of values is that the future is uncertain, and the use of a range of values rather than a discrete single value accounts for this uncertainty.

1.2   Objectives

The goal of this thesis is to develop a summer peak residential energy demand model for Appalachian Power Company's service area in West Virginia using available but largely insufficient data. Regression analysis and Monte Carlo Simulation will be used to test the model. The tests are conducted to determine how insufficient data affects the forecast by comparing the results of the regression analysis and the Monte Carlo Simulation. Therefore, the objectives of the thesis are:

1.) Develop a residential summer peak model applicable to Appalachian Power Company's service area in West Virginia.

2.) Use regression analysis to analyze the summer peak model

3.) Use Monte Carlo Simulation to analyze the summer peak model.

4

4.)   Compare and contrast the assumptions from the results of the regression analysis and

        Monte Carlo Simulation.

1.3   Organization of Thesis

Chapter 2 discusses the history and development of energy demand forecasting, summarizes the different types of energy forecasts, and provides a literature review of energy forecasting models. The theoretical concepts used in the forecast, the functional forms for energy modeling, regression analysis, probabilistic analysis, risk analysis, and the use of the consumer price index are discussed in Chapter 3. Chapter 4 presents the methodology for the development of the summer peak forecast, and a description of the data variables used for the forecast. Chapter 5 presents the results and their implications, and Chapter 6 presents the conclusions, policy implications, and the limitations of the research.

2.0   Background

Energy demand forecasting has developed over time from a very basic and simplistic exercise into a complex procedure. Numerous methods have been developed over the history of energy forecasting.

2.1   History

The need for accurate energy forecasts began during the industrial revolution when human and animal power were replaced with machinery and the efficient use of energy (Burns, 1984).   Prior to the 1970's, the demand for electricity was very predictable; thus energy forecasting was a simplistic process. Reliable forecasts were achieved using simple trend extrapolation (see section 2.3.1). The "7% rule" was also used to forecast energy demand. The "7% rule" assumed that electricity consumption doubled every ten years (Crow, Robinson, & Squitieri, 1981). Forecasting was not perceived to be complex, because forecasters thought that it was immune to the laws of supply and demand (Burns, 1984).

Suddenly this dream of simplistic forecasting was shattered due to rising energy prices resulting from the oil crisis of the 1970's, the severe recession of 1974-75, the growth of energy conservation, the slowdown of economic growth, and the change of end uses. These events made energy demand forecasting much more uncertain and complex. The volatility of energy prices following the 1973 oil embargo and the unexpected elasticity of demand to higher prices caught most energy forecasters off guard (Leung & Miklius, 1994).

Energy demand forecasting came into prominence during the energy crisis of the 1970's. During this time period, changes in the market and political turmoil brought about the need for electricity forecasts. "The convergence of improved economic theory, and electronic data

processing capability in the 1970's resulted in the heyday of models that represented major steps in understanding the complex interrelationships between energy and economic theory" (Burns, 1984, p. 1438). The energy crisis also produced a situation where "not only was it difficult to predict rates of long-term growth, it was difficult to tell whether there would ever be any growth" (Crow *et al.*, 1981, p. 2). In other words, the situation during the 1970 energy crisis was so uncertain that forecasters could not even forecast long-term energy demand, let alone be sure that there could ever be any growth.

Once the oil crisis of the 1970's had begun, energy forecasting became a complex and time-consuming process. Energy demand is now regarded as an economic good that follows the laws of supply and demand. Modern energy demand forecasting requires more analysis, thought, and informed judgement. Modern energy forecasting differs in many ways from past methods, because more complete data and statistics exist which give practitioners a better understanding of the recent past (Burns, 1984).

2.2   Importance of Energy Demand Forecasting

Electricity demand forecasting is important because the future is uncertain. Also, electricity forecasting attempts to predict what the future electricity demand will be. Forecasting electricity demand helps to determine if there will be a shortage of electricity, and the need for new power plants or the implementation of conservation measures, or an overabundance of electricity and the need for shutting down of some of the power plants. Thus, "over forecasting can lead to a decade of over capacity. Under forecasting can mean a long and explosive period while capacity catches up with loads" (Charles River Associates, 1986, p. 1).

Another important reason for energy demand modeling is costs. "If forecasts are too low, energy shortages may develop whose costs are usually a large multiple of the volume of energy not supplied; but if forecasts are too high opportunity costs might be uselessly tied up for long periods of time" (Labys, 1999, p. 40). In other words, if the forecast results in a shortage of electricity, prices would increase and the consumer would pay more for energy. Yet, if the forecast results in an overabundant supply, the costs associated with shutting down power plants and other ways of decreasing supply would be passed on to the consumer.

According to Walter Labys (1999), three important reasons exist for modeling energy demand. The first reason is that the timely and reasonable reliable availability of energy supplies is vital for the functioning of a modern economy. Secondly, the expansion of energy supply systems usually requires many years, and the third reason is that investments in such systems generally are highly capital intensive, on average, accounting for some 30% of gross investments in most countries.

Another important aspect of energy modeling is in the area of deregulation. In many states, including West Virginia, deregulation of the transmission sector of the electricity industry is taking place. With deregulation the future is even more uncertain as the market will decide price and price will determine demand. Therefore the importance of electricity demand forecasting grows. Electricity generators will have competition from out-of-state generators, while existing generators in West Virginia will have expanding territories. Thus, the lines are no longer clearly drawn as to who services what customer.

2.3   Methods of Energy Demand Modeling

A variety of methods exist for forecasting energy demand. These methods represent the path from the relatively simple methods that existed prior to the 1970's to the complicated models that exist today. The various methods also account for the differences between the long run and the short-run forecasts.

2.3.1 Time-Trend Method

The time-trend method determines the overall trend in historical kWh sales or kW peaks, and develops the forecast based upon that trend known as extrapolation of historical trend. In its most simplistic form the method consists of plotting the data on graph paper and laying a ruler through the points. Prior to the 1970's time-trend was the leading method of forecasting used by electric utilities, but today it is used mainly for short-term modeling. Clearly, this methodology does not recognize structural changes and is vulnerable to errors. The advantages and disadvantages of time trend follow (Charles River Associates, 1986, p. 59):

Advantages:

- Little skill required

- Inexpensive and quick

- Can be upgraded by adjusting data

- Useful for minor load categories

- Minimal data requirements

Disadvantages:

- Vulnerable to changes in underlying influences

- No explicit attention to underlying influences

- No explicit audit trail for interpreting reasons for load forecasting error

2.3.2 Time-Series Method

The time-series method forecasts energy demand by the patterns and trends found in the data. When using a time-series method, the researcher uses statistical extrapolation of loads based upon historical data for the loads being forecast—hourly loads, peaks, or energy sales. Time-series is especially useful when projecting load shapes, seasonal patterns, daily or seasonal loads, etc. The advantages and disadvantages of this method are (Charles River Associates, 1986, p. 59):

Advantages:

- Tracks historical patterns closely: often accurate for short term

- Low cost

- Minimal data requirements

- Provides statistical evaluation of forecast uncertainty, especially for short term

Disadvantages:

- Does not treat underlying factors explicitly

- Sources of errors difficult to interpret

- Difficult to allow for conservation or change

- Historical time-series patterns unlikely to apply to long-term load growth

2.3.3 Expert/Informed Opinion Method

The expert/informed opinion method uses information obtained from experts in the field rather than relying on numerical data. This method is interesting, because it uses the knowledge of experts or users to develop the forecast, e.g. plant managers, members of local planning and zoning boards, developers, or other knowledgeable sources within the service area. The expert/informed opinion method can be used to forecast where other forecasting methods are ineffective or to provide a check on a forecast done using another method. It is most effective for forecasting new technologies. The advantages and disadvantages of this method are listed below (Charles River Associates, 1986, p. 60):

Advantages:

- Utilizes available knowledge

- Usually low cost

- Helpful for forecasting new end uses where no historical data are available

Disadvantages:

- No consensus opinion on most loads

- No audit trail for forecast

- Informed source not always available to utility

- Opinions sometimes biased

2.3.4 Identity Method

The identity method is a very basic form of forecasting energy demand using separate forecasts of load factor and kWh sales and definitional relationships between them. It is simple because no special skills are required and the forecast can be computed in a very short period of

time. The data required for this method are kWh sales for forecast year generated by another

model, and the load forecast for the forecast year, usually obtained from observing recent

movements in load factor and adjusted for anticipated future trends, although it may come from

another model. This model uses the equation for the annual load factor which is sales in kWh

divided by peak in kWh times 8760 hours per year. The formula can be rearranged to determine

peak demand. The problem with this method is that it does not account for underlying factors.

The advantages and disadvantages of this method follow (Crow *et al.*, 1981, p. 8):

Advantages:

- Simple

- Uses other independent forecasts

Disadvantages:

- Forecasting kWh sales and load factor may be difficult and subject to great
  uncertainty.

2.3.5 End-Use Method

The end-use method determines energy demand through total kWh use from all of the

electrical appliances used in the home. The forecast is "built from the sum of end-using

activities" (Charles River Associates, 1986, p. 27). End-use models must include kWh

consumption data by type of equipment or process. This method is most readily applied to the

residential sector, because the data required for this forecast include: forecast year, number of

residential customers, residential housing stock or commercial buildings, industrial process data,

major appliances, and kWh use per appliance.   The basic form of this model are simple

accounting procedures which "enumerate the end uses and add the electricity use for each end

use for its components" (Charles River Associates, 1986, p. 20). The problem is that in the real world nothing is that simple and everything is a difficult process. The advantages and disadvantages of these models are (Charles River Associates, 1986, p. 60):

Advantages:

- Focuses on components of electricity use: can trace true location of forecasting error

- Good structure for incorporating load management and conservation impacts and for analyzing marketing impacts

- Intermediate technical and computer skills required

- Easy to explain results to others

Disadvantages:

- Requires large amount of detailed data

- Data assembly costly and difficult

- Static framework: technology must be explicitly specified

- Components (saturation rates and use per appliance) are difficult to forecast, particularly for long-term

- Requires knowledge of end-use technologies and practices

2.3.6 Econometric Method

The econometric method determines energy demand by considering the influence of independent variables, such as, population, employment, income, weather, appliance ownership, and rates. Econometric models are estimated equations that relate electricity demand to external factors (Charles River Associates, 1986). Where end-use models use kWh consumption data by

type of equipment or process, econometric models use time-series data. Econometric models are among the most complex forms of energy forecasting. They are used for all areas of service—residential, industrial, and commercial. The data required for this method of forecasting include observations of customer income, commercial and industrial activity, electricity rates, gas and heating oil prices, weather, and housing stock. The advantages and disadvantages of this method follow (Charles River Associates, 1986, p. 60):

Advantages:

- Explicitly measures effect of underlying causes of trends and patterns
- Provides statistical evaluation of forecast uncertainty
- Combines well with economic and demographic information on service territory
- Can incorporate an end-use structure or time-series method
- Models can be readily re-estimated

Disadvantages:

- Requires skill and experience in econometrics and computer programs
- Extensive data required for detailed disaggregated model
- Costs can be relatively high

2.3.7 Hybrid Econometric and End-Use Method

The Hybrid Econometric and End-Use method is another form of energy modeling that combines end-use structure with econometric estimation to forecast energy demand. The data required for this method are forecast years, the number of customers, independent variables to estimate housing and appliance stocks, kWh use per appliance, and industrial process

information. Hybrid models have been developed mainly for the commercial and residential sectors. These models use "econometric methods to produce parameter values and input data" (Charles River Associates, 1986, p. 41). When doing a forecast using a hybrid end-use and econometric method, the procedure is first to develop the model specification, both for core end-use model and for the econometric equations used to estimate input values. The second step is the estimation of the econometric equations. The third step is the assembly of all data for the end-use model, except econometric estimates. The fourth step is the development of econometric forecast values, which include forecasting values for the exogenous variables and simulating the econometric equations for forecast years. The fifth step is actually making the forecast. The advantages and disadvantages of this method are (Charles River Associates, 1986, p. 60):

Advantages:

- Combines strengths of econometric and end-use methodologies
- Most comprehensive forecasting approach available
- Relatively easy to update and maintain

Disadvantages:

- Highest-cost method
- High skill and experience requirements in econometrics, computer programming, and end-use technologies and practices
- Very large amount of data required

2.4   Types of Energy Forecasts

Different types of energy forecasts exist due to the shifts in consumption patterns that occur over time. Electricity consumption varies according to the season and the time of day. Energy consumption varies from season to season, because more electricity is used during the winter and summer months when it is hotter or colder, than in the spring and fall months when the temperatures are usually moderate. Residential energy consumption varies from day to day because more electricity is used during the weekdays than on weekends.   Electricity consumption also varies from hour to hour as more electricity is used during the morning hours when people are waking up and getting ready to start their day. Electricity consumption then declines during the late morning and afternoon hours when most people are at work and school. Electricity consumption increases again during the evening hours when the work day ends and people come home and cook dinner, watch television, and need light to see by. Figure 1 shows this phenomenon.
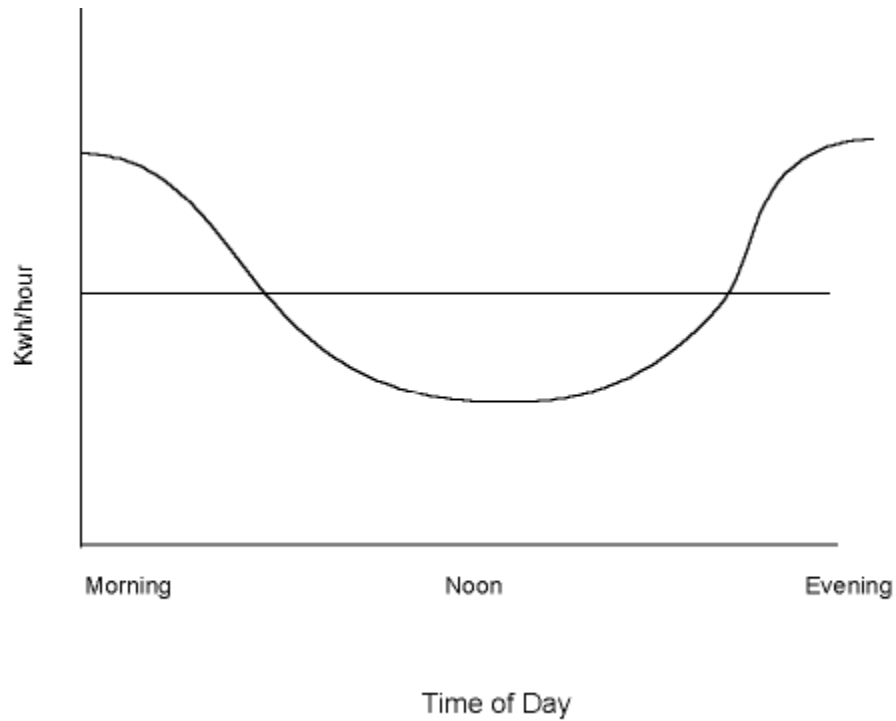
Figure 1:   Electricity Demand by Time of Day

Therefore, different forecasts are needed to fulfill the different requirements of energy demand: seasonal demand, short-run demand, long-run demand, and peak demand. In other words, electricity generation must always meet demand since it is not possible to store any large amounts. Thus, the complexities of energy demand forecasting are apparent in the various types of forecasts that are used to determine energy demand at various stages.

The three different sectors for which energy demand can be forecasted include the industrial, the residential and the commercial sectors. This thesis will focus on forecasting residential energy demand because that is the sector under consideration in Appalachian Power Company's service area. Also the residential sector is one of the main contributors to system peak.

2.4.1 Winter & Summer Peak Forecasts

A winter peaking forecast is one that predicts the maximum amount of power required on one day during the winter months. The northern United States is the area that experiences a winter peak, because it experiences bitterly cold winters. The following equation is an example of a winter peak model (Le, 1983, p. 224):

$$Ln(WPK) = c + b_1 Ln(PI) + b_2 Ln(GASP) + b_3 Ln(ELECP) + b_4 Ln(WTX) + b_5(DEC)$$

Where:

PK represents winter peak load

PI represents personal income

GASP represents price index of natural gas

ELECP represents price of electricity (typical residential and commercial electric bill)

WTX represents weather factor = (RSH = CISH) × HDD

RSH represents number of residential customers with electric space heating

CISH represents number of commercial and industrial customers with electric space heating

HDD represents heating degree days of system peak day

DEC represents dummy variable for month of December

In an area that experiences a winter peak, freezing temperatures can dramatically increase the demand for energy. Therefore, a dependable forecast is necessary to ensure that there is enough electricity to heat homes.

Summer is the warmest time of the year, and in the southern part of the United States it is stifling. The southern United States is the area that typically experiences a summer peak. A summer peak forecast predicts energy usage for the summer months (see Section 3.2 for an

example of Le's summer peak model). A summer peak model is used to ensure that there is enough energy to meet demand during the summer with the increasing use of cooling devices. Summer forecasts are also used to determine how much electricity will be available for sale to other electric companies.

Although the southern United States is the area that experiences a summer peak, electricity consumption is increasing during the summer all across the United States with the increased use of air conditioning. This trend has prompted the need for a summer peak forecast for areas of the country that never had need of one before. Figure 2 depicts a winter and summer peak. The figure shows that they are opposite as to when their peak occurs.
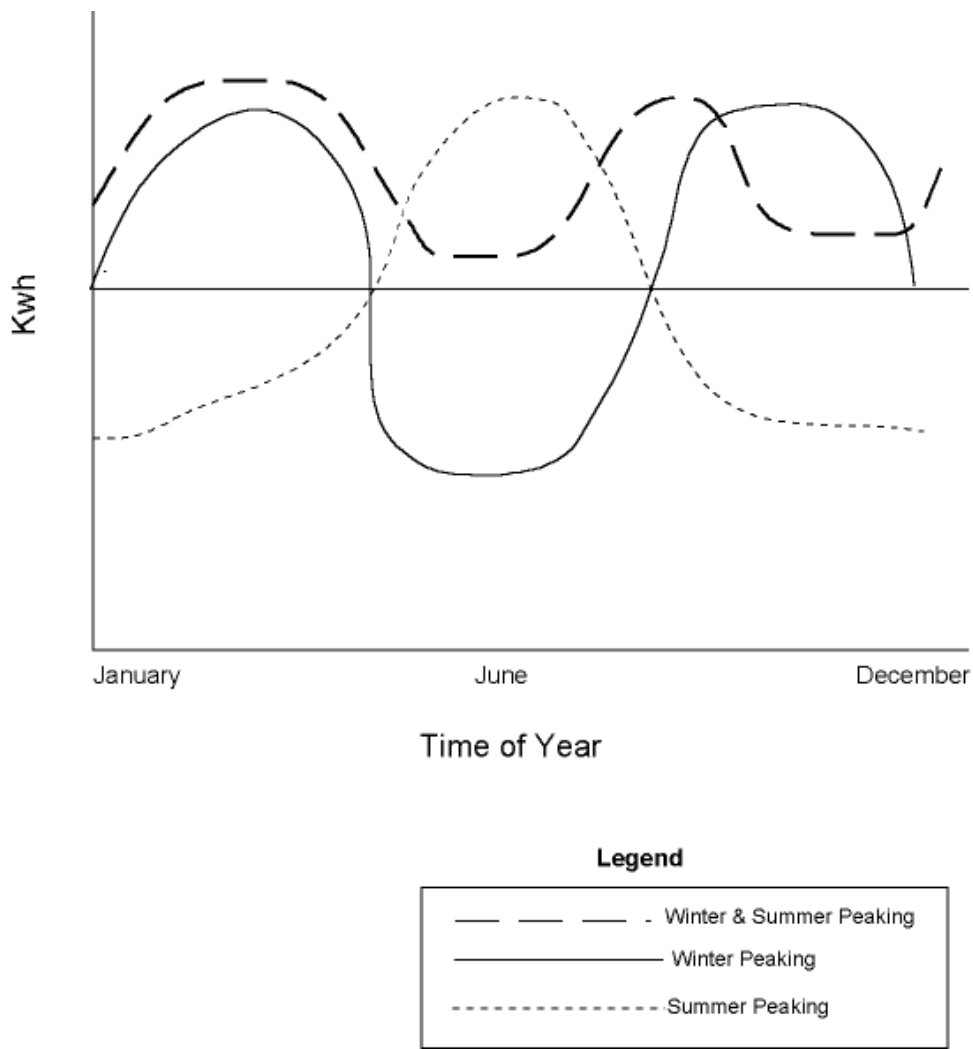
Figure 2:   Electricity Demand for Peaking Forecasts

2.4.2 Short-Run Forecasts

Short-run forecasts are intended to be valid for only a few minutes to a few hours into the future and are needed by electric utility operators for the purpose of scheduling and dispatching generating units. According to Donnelly (1987), a short-term forecast is important for "unit commitment, economic dispatch, hydro-thermal co-ordination, load management, etc." (p. 139). Ackerman (1985) says that a short-run forecast plays an important role in the day-to-day

operations of a utility, and it is typically used for optimizing system operation and scheduling of hydro units and other peaking plants, such as gas turbines. The objective of the operators is to minimize variable costs without jeopardizing the electric system to power failures.

Ackerman's (1985) short-term forecast can be given by the nonlinear function:

$$Y_t = f\left(y_t, x_t, t, \varepsilon_t\right)$$

Where:

Y_t represents actual system load in megawatts (expressed as integrated hourly load in the case of

the hourly predictor)

y_t represents $\{Y_t; -\infty < T \geq t\text{-}1\}$; i.e. all past observed actual system loads

x_t represents $\{X_T; -\infty < T \geq t\text{-}1\}$; i.e. all past observed exogenous variables

t represents time index

ε_t represents additive random disturbances representing all unobserved effects on the system load

The hourly short-term forecast can be represented by P.C. Gupta's (1985) model:

$$Z(i, j) = T(i, j) + WC(i, j) + X(i, j)$$

Where:

Z(i, j) represents system load measure in hourly mWh at hour j and day i;

T(i, j) represents basic component of the load hour j and day i;

WC(i, j) represents weekly cycle component (day-of-the-week effect) at hour j and day i;

X(i, j) represents residual component containing the effect of weather variations at hour j and day

i

Gupta also formulated a weather model because, an hourly forecast is dependent upon the variability of the weather. The weather model can be represented by:

$$Y(i) = B(i) + S(i) + W(i) + \varepsilon(i)$$

Where:

Y(i) represents peak load on the ith day;

B(i) represents basic load component of the peak load on the ith day;

S(i) represents weakly pattern component of the peak load on the ith day;

W(i) represents weather-sensitive component of the peak load on the ith day;

ε(i) represents random component of the peak load on the ith day.

"Strictly speaking, B(i) and S(i) should be defined as elements of the T(i, j) and WC (i, j).

However, better estimates can be expected to result if B(i) and S(i) are estimated independently

of T(i, j) and WC (i, j) respectively" (p. 45).

2.4.3 Medium-Term Forecasts

Medium-term forecasts predict future energy usage a few months to a few years into the

future, and are necessary for planning fuel procurement, scheduling unit maintenance, energy

training, and revenue assessment. Other uses for medium-term forecasting are decisions on

capital repairs, inventory control of coal and liquid fuels, the purchase of fuel quantities and the

assessment of revenue impacts due to changes in electricity tariffs. A medium term forecast is

similar in structure to the long-term forecast presented in the next section.

2.4.4 Long-Term Forecasts

A long-term peak load forecast is designed to be valid from five to twenty-five years into

the future. The earliest method of long-term forecasting, prior to the 1970's, involved trend line

extrapolation (see section 2.3.1) (Lakshminarayana, 1991). Modern methods of long-term

energy forecasting use sophisticated statistical based analysis as the forecasting tool. It is important in plans for system generation and transmission and allows for planning engineers to determine the type and size of new power installations to minimize fixed and variable costs. When determining a long-term forecast outside factors such as governmental policy, and development policy should be considered.

Donnelly (1987) describes a typical functional form of a long term forecast as:

$$Q_E = f\left(P_E, P_S, P_C, Y, X\right)$$

Where:

Q_E represents Quantity of electricity

P_E represents Electricity Price

P_E represents Price of substitute good

P_E represents Price of complementary good

Y represents Consumer Income

X represents Other factors

3.0   Literature Review

The literature review section presents the various methods of energy modeling, the types of energy modeling that exist, and the variety of methods that have been developed through the years to determine energy demand. The literature review shows the gap in the amount of literature available for statistical methods of forecasting energy demand versus the probabilistic method of forecasting energy demand.

3.1   Early Energy Demand Models

As seen in section 2.1, energy demand forecasting has grown from a relatively simplistic procedure into a more complex one with many different factors impacting energy demand. Early energy demand models from the 1960's assumed that energy was tied to output (Labys, 1999, p. 32):

$$Q = \min\left(\frac{VA(K,L)}{\gamma_1}, \frac{E}{\gamma_2}, \frac{M}{\gamma_3}\right)$$

The equation was representative of a typical Leontief production output function—output (Q) depends on value added (VA(K,L)), energy (E), and materials (M). Inputs γ_1, γ_2, γ_3 are technical coefficients which "measure the Leontief recipe for the dollar value of any given input necessary to produce a dollars worth of output" (Labys, 1999, p. 40). This model is dependent entirely upon GNP. GNP is used as a proxy for output and energy is tied to GNP through a technically determined coefficient (γ_2). Therefore,

$$E = \partial_2 GNP$$

3.2　Energy Demand Models that use the Statistical Method

Al-Garni, Zubair, and Nizami (1994) demonstrate a modern method of forecasting energy demand for Saudi Arabia. Al-Garni *et al.* use a statistical method for forecasting peak energy demand. A statistical summer peak model must account for temperature and other factors affected by the warm weather. Although the climate of West Virginia is vastly different from that of Saudi Arabia, the study of a warm climate provides useful information as to what variables are important for a summer peak demand model for West Virginia. Al-Garni *et al.* include variables representing weather data, global solar radiation, and population. The model is represented by the equation:

$$E_P = E(T, H, S, P)$$

Where:

E represented the measured energy consumption,

E_P represents the predicted energy consumption,

T represents the monthly mean air temperature (°C),

H represents the monthly mean relative humidity (%),

S represents the monthly mean global solar radiation (w * H/m$_2$ * day), and

P represents population.

Variable selection was accomplished by using a step-wise regression technique, and evaluation of model accuracy was done by analyzing the residuals.

Leung and Miklius (1994) compare different statistical methods of energy forecasting using Hawaii as a test case. Leung and Miklius show the different forms that exist for statistical forecasting. Their study is also relevant, because it highlights what factors influence energy demand in Hawaii's warm climate which gives an indication of what may impact summer

demand in West Virginia with its hot humid summers. Since the forecast used to compare the statistical and probabilistic forecasting methods is a summer peak forecast, this forecast will help determine which factors are important for a West Virginia summer peak forecast.

The authors compared extrapolation based on linear (average) growth, linear exponential smoothing, quadratic exponential smoothing, causal relationship between electricity consumption and de facto population, causal relationship between electricity consumption and per capita income, and the Hawaii Energy Demand Forecasting Model (HEDFM).

The final method discussed in the comparison is the Hawaii Energy Demand Forecasting Model (HEDFM). HEDFM is an econometric-based simulation system that was developed by the Research and Economic Analysis Division of the Hawaii Department of Planning and Economic Department in conjunction with the Energy Analysis Program of the University of California Lawrence Berkeley Laboratory. The purpose of the HEDFM is to provide detailed forecasts for various types of fuels up to the year 2005. The HEDFM "follows a very widely used dynamic flow adjustment model structure to capture the lagged adjustment process of energy demand due to (1) the existing stock of energy consuming devices that cannot be replaced immediately, and/or (2) the unwillingness of consumers to view changes as permanent until they have continued for some time. The model relates consumption in a given year to price, income, other relevant variables, and previous year's consumption" (p. 293).

The results of the comparisons of the six different types of forecasting methods show that the quadratic exponential smoothing method produced the best results. HEDFM performed fairly well, but it tended to underestimate actual electricity consumption for the entire period. A possible explanation for these results could be attributed to the inaccurate forecasts of the independent variables. Linear exponential smoothing performed reasonably well also. The

results of the study were analyzed by comparing different "goodness of fit" criterion, the R^2, the root-mean-squared error (RMSE), mean absolute error (MAE), Theil's fifth inequality coefficient (μ) and the turning point errors.

S.V. Le (1983) formulates a statistical model to forecast summer residential energy demand. Le uses the statistical method to study which factors influence summer peak residential energy demand and then uses those results to forecast demand. Le's model indicates what influences residential energy demand in a climate very similar to West Virginia's. Unlike some of the other models discussed which were developed for either a tropical or desert climate, Le's model accounts for the uniqueness of West Virginia's economy and climate, because income and price are included along with a temperature variable. Le's model is also unusual in that it measures the use of air conditioning saturation, which suggests that residential energy demand in the summer is tied to air conditioning use, which is based upon the temperature and technology adaptation.

The equation developed by Le for determining residential summer peak demand is:

$$Ln(SPK) = c + b_1 Ln(PI) + b_2 Ln(GASP) + b_3 Ln(ELECP) + b_4 Ln(WT)$$

Where:

SPK represents net system summer peak load,

PI represents personal income in the service area,

GASP represents the price index of natural gas,

ELECP represents the price electricity (typical residential and commercial electric bill)

WT represents the weather factor that is composed of air conditioning saturation multiplied by

the weighted average of cooling degree hours of peak day and three previous days

multiplied by the number of residential customers.

Crow, Robinson, and Squitieri (1981) discuss a variation on the residential energy demand model. They assume that the use of electricity is a derived demand because it is not consumed "as an end itself, but for other purposes, such as to run equipment" (p. 90). This type of model is in a simple linear statistical form that is represented by the equation:

$$KWH = a_1 - a_2 PELEC + a_3 PGAS + a_4 INC + a_5 HDD + a_6 CDD - a_7 CONS + a_8 R + a_9 S$$

Where:

PELEC represents the price of electricity,

PGAS represents the price of gas,

INC represents income level,

HDD represents heating degree days,

CDD represents cooling degree days,

R represents the number of rooms per house, and

S represents average family size

Crow *et al.* provide an excellent example of a linear statistical model that gives insight into what form the model for Appalachian Power Company's service area in West Virginia may take. The variable representing average family size is interesting because the larger the family, the more likely they are to use more electricity. The problem with Crow *et al.*'s model is that it is a model to forecast energy demand, therefore, it is not divided into winter and summer forecasts.

Halvorsen (1978) discusses an econometric method of forecasting residential energy demand. According to Halvorsen, residential electricity demand can be expected to be a function of both economic and non-economic variables. These economic and non-economic variables include: the price of electricity, prices of substitutes and complements, the level of income,

28

climate variables, demographic variables, and the characteristics of housing stock are all included in the model.

Halvorsen's residential electricity demand equation is:

$$LnQ_R = \alpha_Q + \alpha_E LnP_R + \alpha_Y LnY + \alpha_G LnG_R + \alpha_A LnA + \alpha_D LnD + \alpha_J LnJ + \alpha_U LnU + \alpha_M LnM$$
$$+ \alpha_H LnH + \alpha_T T + \mu_R$$

Where:

Q_R represents the average annual residential electricity sales per customer, in thousands

of kWh

P_R represents the average real price of residential electricity, in cents per kWh

Y represents the average real income per capital, in thousands of dollars

G_R represents the average real gas price for all types of gas, in cents per therm

A represents the index of real wholesale prices of selected electric appliances

D represents heating degree days

J represents average July temperature, in degrees Fahrenheit

U represents percentage of population living in rural areas

M represents percentage of housing units in multiunit structures

H represents average size of households

T represents time

u represents a disturbance term

Crow *et al.* and Halvorsen both include the concept of rural versus urban energy usage in their model. Crow *et al.* does it by including the number of rooms in a typical residence in the model. The large rural population in West Virginia means that many people live in homes rather than apartments, and this variable captures this concept. Halvorsen captures the concept by including a variable representing the percentage of the population living in rural areas. The

percentage of people living in rural areas is included in the model, because those living in rural and urban areas usually have different availability of substitutes, and alternative energy sources. Also, if a majority of the population live in rural areas, then it would be logical to assume that a large portion of the population live in single-unit structures rather than multi-unit structures as apartments are usually an urban phenomenon. Those living in apartments are expected to use less electricity than those living in large single family homes. This variable is interesting for West Virginia, because the majority of the population live in rural areas.

Cowing and McFadden (1984) discuss the Oak Ridge National Laboratory (ORNL) model of residential energy demand forecasting. The ORNL disaggregate forecast is based upon housing and urban differences, family size differences, and income. As discussed earlier, interesting population and demographic differences are considered with this model. However, the forecast model requires data that are not available for the southern portion of West Virginia.

The ORNL model is a complex model whose purpose is to satisfy "the most pressing requirements for regional forecasts of the impacts of heterogeneous policies" (p. 45). The creators of the model realized that the end-use of electricity is disaggregated consumers who use it for appliances, and they deal with this problem by permitting some demographic segmentation. Components of the ORNL model are outlined in Table 1.

Table 1: ORNL Model Components

Base-year data	New equipment market shares, base year
	Equipment market shares, base year
	Annual average energy use, new equipment, 1970
	New equipment prices, base year
	Interest rate
	Ratio of 1970 to 1970 usage factors
	New equipment installations, base year
	Fuel prices plus income, base year
Technological parameters	Air conditioning-space heat load reduction ratio
	New equipment technological parameter
	New construction technological parameters
	Retrofit technological parameters
	Average equipment lifetimes
	Lifetime of investments in thermal shell
Behavioral parameters	Appliance market share elasticity with respect to operating coast
	Usage elasticity with respect to operating cost
	Interest rates for PV cost minimization
	Market share equation slope coefficient
	Ratio of short- to long-run usage elasticities
	A market penetration rate parameter
	A horizon after which life-cycle cost is minimum
	Maximum saturation
	Average equipment lifetimes
Exogenous forecasts	New equipment prices, relative
	Real prices for fuels plus income
	Average size of existing housing units
	Average annual energy use, new equipment, before adjustment
	Thermal integrity for retrofit homes
	Number of homes which are retrofit
	Total number of occupied housing units
	Total number of new housing units
	Fractions of new homes with room – central air conditioning
	Status quo new equipment energy use
	Size of new housing units
	Average thermal integrity, new structures

(Source : Cowing and McFadden, p.46)

Not all of the variables listed in Table 1 above are used because some of the variables

measure the same concept, are believed to be zero, or "have common values or are obtained by a

relatively simple interpolation as part of the preprocessing of the input files" (p. 46). The

problem with the model is that many of the variables are based upon the forecasters' judgement,

which leaves the forecast vulnerable to error. Also, much of the data are difficult to obtain because it is not available on the specific level necessary.

The design of the ORNL model is a block recursive structure. The model has a housing submodel which "forecasts the number of households in a region, additions to housing stocks required to accommodate these households, and the average size of new dwellings" (p. 49). An appliance saturation model is also a part of the ORNL model which is an "econometrically estimated model of equipment ownership, classified by type of equipment and fuel type, as a function of equipment and flue prices" (p. 57).

The ORNL model accounts for changing demographics, economic conditions, and new technologies. Therefore, "the ORNL model is designed to forecast annual residential energy consumption, classified by five fuel types, three dwelling types, and eight end-uses for up to thirty years" (p. 48).

Cowing and McFadden (1984) and Koomey, Brown, Richey, Johnson, Sanstad, and Shown (1995) discuss the Residential End-Use Energy Planning System (REEPS). REEPS was developed by the Electric Power Research institute and is designed to "provide end-use specific forecasts for energy consumption at the household level" (Cowing and McFadden, p. 61) and to "allow users to define customized models for various energy end-uses in the residential sector, including appliances and heating , ventilation, and air conditioning (HVAC) equipment" (Koomey *et al.*, p. 1).   What makes the model work is a generated sample of simulated households. REEPS "predicts four categories of energy-related decisions: appliance investment decisions in new dwellings, retrofit appliance investments in existing dwellings, appliance operating efficiencies, and annual energy consumption (Cowing and McFadden, p. 64). The most common application of REEPS can be used to "forecast the long-term impacts of energy policies

on a population whose characteristics change over time, reflecting projected demographic and economic growth" (Cowing and McFadden, p. 66).

The REEPS modeling system has four steps. The first step is to formulate a sample of "households with information on their socioeconomic attributes, appliance holdings and characteristics of the area in which they reside" (Cowing and McFadden, P. 64) or as Koomey *et al.* say that "the first step is "accounting for changes in stock based on equipment decay" (p. 1). Step two is the simulated model which forecasts purchases and installations of space heating, water heating, central air conditioners, room air conditioners, stoves, and dishwashers. It is divided into three categories: 1) physically attached appliances in new housing construction, 2) attached appliances in existing housing, and 3) other appliances. The third step consists of "calculating annual energy consumption by fuel type given the appliance portfolio and the household, dwelling, and geographical characteristics" (Cowing and McFadden, p. 65). The fourth step is "after the appliance and energy usage decisions are simulated for the current period, a new sample is created for the subsequent period simulation" (Cowing and McFadden, p. 65-66).

There are three inputs for the REEPS model. These inputs are exogenous inputs that are "used to forecast the general macroeconomic circumstances under which the energy-and technology-specific projections occur (Koomey *et al.*, p. 6)." The input elements of the REEPS model are shown in Table 2.

Table 2: REEPS Model Exogenous Inputs

Exogenous Inputs	Fuel prices
	Legal availability
	Market availability
	Household income
	Household size
	Income
End-use Technology Inputs	Ownership
	Equipment efficiency
	Equipment size/capacity
	Equipment UEC
	Price
	Vintage blocks
	Min/max lifetimes
Thermal shell Inputs	Heat gain/loss
	Floor area
	Thermal shell
	Purchases
	Equipment size/capacity
	Equipment efficiency
	Equipment UEC
	Stock
	Ownership
	Site energy consumption

(Source: Koomey *et al.*, p. 6-8)

The method of decision making in the REEPS model is decision models which are used "to forecast the effect of different decisions made in the course of owning and operating residential appliances and HVAC equipment" (Koomey *et al.*, p. 9). These decision models determine "the value of four key variables for appliances and HVAC equipment throughout the course of the forecast: ownership, efficiency, usage, and size/capacity" (Koomey *et al.*, p. 9).

Neither the ORNL model nor the REEPS model are applicable for model development for the southern part of West Virginia. Both models are so complex that it would be difficult to implement either one for such a small area. Also, the data necessary for each model are not available on a regional basis. Therefore, the model would not be an accurate representation of residential energy demand for southern West Virginia.

Houthakker (1980) addresses the issue of marginal price versus average price. Houthakker suggests that "quite apart from theoretical considerations, the marginal price differed considerably from the average price in its behavior over time. This divergence is presumably related to the effect of higher fuel prices on the marginal electricity price" (p. 31).

Houthakker's model follows:

$$\log q_{it} = a_i + \beta \log q_{it-1} + \gamma \log y_{it} + \delta \log p_{it} + \xi_1 \log\left(h_{it}\middle/h_{it-1}\right) + \xi_2 \log\left(c_{it}\middle/c_{it-1}\right) + \xi_3 \log z_{it} + \mu_{it}$$

Where:

i represents the ith state

t represents the thth year

q represents consumption in kWh per capital per year

y represents adjusted disposable income per capital (deflated by the regional CPI)

p represents the marginal price electricity (actually the mean of the marginal prices at the

　　beginning and end of year t, deflated by the regional CPI)

h represents heating degree days

c represents cooling degree days

z represents the average price of gas (i.e., the revenue from residential gas customers

　　divided by the quantity, in Btu, sold to these customers), deflated as before

μ represents the error term

Houthakker discusses the concept of marginal pricing for electric price versus average pricing. The marginal price is the incremental price or the price of the next unit of energy consumed or produced. Average prices are commonly used in energy forecasts, because the determination of the marginal cost is difficult.

Houthakker also used the Consumer Price Index (CPI) deflator to calculate the constant dollar values for the income and price variables to account for changes in inflation over time. This converted the price and income variables into constant rather than current dollar values.

Houthakker's model could be used to produce a statistical forecast for total energy demand including both summer and winter since the cooling degree-day and heating degree-day variables are both included. Other than the variations mentioned above Houthakker's model is similar to those discussed earlier and the same discussion of the temperature and economic variables would hold.

3.3 Energy Demand Models that use the Probabilistic Method

Sam O. Sugiyama (1992) describes the criterion used by the Bonneville Power Administration to forecast energy use. An aggregate econometric energy model was developed by Bonneville to forecast monthly electricity sales for its public utility customers. The independent variables in the model are weather (in terms of heating and cooling degree-days), economic activity (regional non-agricultural employment) and electric prices. Dummy variables were used to account for "shoulder months." Therefore, Bonneville Power Administration's model is written as:

$$MW_t^R = MW_t^{EMP} + MW_t^{DD} + MW_t^{Oth}$$

Where:

MW^{EMP} represents the electric energy attributable to employment,

MW^{DD} represents electric energy attributable to degree-days

Sugiyama used probabilistic forecasting rather than the more traditional methods of energy forecasting such as the statistical method. The stochastic system for regression

estimation is an alternative because this type of analysis recognizes that uncertainty exists with regards to the data. It is easy to calibrate detailed baseline forecasts, and the model gives the opportunity to examine the effects of major events overall as long as they fall within the current scenario ranges without re-running the detailed models.

Bonneville's model has all of the traditional components of an electricity demand model—an economic variable, a weather variable and a variable representing electricity price. What makes it interesting is that inputs are supplied in terms of probabilistic distributions, and the probabilistic method is used to analyze the impact of these probabilistic variables upon energy demand which gives a measure of the risk involved in the forecasts. This measure of risk allows the risk involved in the project to be analyzed and considered in the forecast. The concept of using probabilistic analysis to forecast energy demand is a concept discussed later in this thesis.

4.0   Theory

This section describes the theory involved in comparing the statistical and probabilistic methods of analysis for a long-term forecast for Appalachian Power Company's service area in West Virginia. Included in the theory section are discussions of regression analysis, risk analysis, probability and probabilistic analysis, sensitivity analysis and scenario analysis, Monte Carlo Simulation, preference theory and certainty equivalence, forecasting, predictive versus explanatory models, functional form, and finally the Consumer Price Index.

4.1   Theoretical Concepts Used in Investigation

Regression analysis is a means of analyzing the independent factors involved in a particular question and determining which one(s) impact(s) the dependent variable. Therefore, regression analysis is used to address the question of what factor(s) impact(s) the summer peak, and how each factor considered singularly or in combination contributes to demand. Decision analysis can be used to determine the risk involved in the final model decided upon through regression analysis.

Monte Carlo Simulation is a method that uses probabilistic distribution functions of the input variables in conjunction with computer simulation to arrive at a probabilistic distribution of the dependent variable.

Certainty equivalence is a means of setting the level of acceptable risk, whether the decision-maker is risk seeking or risk averse. The level of acceptable risk then allows the decision-maker to decide upon which scenario suits their level of risk.

4.2   Risk

The future is uncertain, and no one knows exactly what tomorrow will bring. We do not know how the action we take today will affect some unknown future event. "Risk implies that a given action has more than one possible outcome" (Palisade Corporation, 1996, p. 2-5). If a situation is determined to be risky, then risk must be factored into the decision process and this is when risk analysis takes over. Both regression analysis and Monte Carlo Simulation incorporate risk into the analysis but the methods do so in different ways.

4.2.1 Risk Analysis

Risk analysis is a method of assessing the impacts of risk on decision situations. It accounts for uncertainty. Risk is the evaluation of a situation using a known mechanism that incorporates the probabilities of occurrence for success and failure and/or different outcomes for the situation. The use of quantitative approaches to incorporate risk and uncertainty into analysis results may allow the decision-maker to be more successful in achieving the objective in the long run. The goal is to help the decision-maker choose a course of action, given better understanding of the possible outcomes that could occur. Risk is often portrayed in an analysis by the inclusion of a probability density function about a single input or output value or as confidence limits about a regression estimate. The theories behind all derived probabilities and probability density functions have common roots.

4.2.2 Probability

Probabilistic analysis is based on the concept of probability. An example of probability as described by Howard (1968) is: "Let A be an event and S be the state of information on

which the probability of the occurrence of A is to be assigned. Then {A/S} is the symbol for the probability of A given S. If x is the random variable, then the probability density or mass function of x assigned on the basis of S is {x/S}" (p. 212).

Probability analysis is a method of determining how much risk is involved in selecting a particular scenario. Probabilistic analysis is the study of the uncertainty of events. "The distribution and relative possibilities of values are assigned to a given variable will remain characteristic of that value if factors affecting the variable will remain constant" (Stermole and Stermole, 1993, p. 250).

There are five steps that should be followed when doing probabilistic analysis:

1. Develop a model by defining the problem or situation in a spreadsheet program. In other words, develop the forecast model in the spreadsheet program.

2. Identify uncertainty in variables in the spreadsheet worksheet and specifying the possible values of these variables with probability distributions and identifying the uncertain worksheet results that are unwanted and change them for each range of values.

3. Estimate the possible correlation among the input variables and make the appropriate entries in the probabilistic program.

4. Analyze the model using simulation to determine the range and probabilities of all possible outcomes for the results of the worksheet.

5. Describe the risk involved in the decision and then make a decision based on the results and personal preferences.

Sensitivity analysis, scenario analysis, and Monte Carlo simulation are all methods of risk analysis. Each has its own impact upon the analysis.

4.2.3 Determining Probabilities

Three methods exist for determining probabilities: the classical method, the relative frequency method, and the subjective method. Regardless of the method used there are two basic requirements that must be met: (Anderson, 1978, p. 26)

1) The probability values assigned to each sample point must be between 0 and 1. That is, if we let Ei indicate the sample point and P(Ei) indicate the probability of the sample point, we must have

$$0 \leq P(E_i) \leq 1, \text{ for all } i$$

2) The sum of all the sample point probabilities must be 1. For example, if a sample space has k sample points we have

$$P(E_1) + P(E_2) + \ldots + P(E_k) =$$

$$\sum_{i=1}^{k} P(E_i) = 1$$

The classical method is used when two outcomes are equally likely to occur, in other words a fifty-fifty chance of occurring. An example of the classical method is a coin toss – it has a fifty-fifty chance of being heads or of being tails.

The relative frequency method is a method of determining probabilities. It is used when the two possible outcomes do not have an equal possibility of occurring. An example of the relative frequency method is to "suppose that in a test market evaluation of the product 400 potential customers were actually contacted: 100 actually purchased the product, while 300 did

41

not" (Anderson, 1978, p. 27). The way of determining the probability is by dividing 100 by 400 or 25% of those surveyed actually purchased the product while 75% did not.

The subjective method is used when the "experimental outcomes are not equally likely and where experimental or historical relative frequency data are unavailable" (Anderson, 1978, p. 28). In other words, the probabilities are based upon the knowledge of the past history and one's own personal knowledge of the event. An example is a sports team. When determining the probability the team will win one must delve into their knowledge of the team's players, injuries, coaching, etc. and compare it to what is known of their opponent and then make a judgement as to the probability that the team will win. But Anderson says that "after considering all available information, you must specify a probability value that expresses your degree of belief that the team will win its next game" (Anderson, 1978, p. 28).

4.3   Regression Analysis

Regression analysis is a statistical tool that compares the relationship between two or more variables. It is the identification of a relationship between a dependent variable and one or more independent variables. Regression analysis measures how much of the movement in the dependent variable is explained by the independent variables. "Econometricians use regression analysis to make quantitative estimates of economic relationships that previously have been completely theoretical in nature" (Studenmund, 1992, p. 5). It is a statistical tool that attempts to explain changes in the dependent variable as a function of the independent variable(s) through the "quantification of a single equation" (Studenmund, 1992, p. 5)

Definitions of some of the regression statistics used to observe the outcome follow:

- **Population** is the group from which the sample is being drawn, while the **sample** comes from the population. According to Studenmund (1992) when the sample size nears the population size, the unbiased estimate coefficient nears the value of the true population.

- **Mean** is the sample average. "The mean is the most common measure for the center or location of data on an economic variable" (Griffiths, Hill, and Judge, 1992, p. 22). The mean is represented by the equation:

$$\overline{X} = \frac{\sum_{t=1}^{T} X_t}{T} = \frac{X_1 + X_2 + ...+ X_T}{T}$$

Where: T represents the number of observations

- **Regression Coefficients** measures the contribution of the independent variable to the prediction. Each coefficient multiplies the corresponding variable in forming the best prediction of the dependent variable. The C or constant coefficient is the base level of the prediction when all other independent variables are zero.

- **Regression Estimation** is a statistically estimated curve or function in which a dependent variable, such as sales or peak loads, is explained given the values of the independent variables, such as the electricity price, temperature, and income.

- **Least Squares** is an estimations criterion ("estimation") used in regression analysis, which minimizes the sum of squared deviations between the actual observation and the regression. Least squares, when properly applied yields unbiased estimates of minimum variance.

- **Residuals** are the difference between the actual and fitted values of the dependent variable. "They give an indication of the likely errors that the regression would make in a forecasting application" (Quantitative Micro Software, 1995, p. 162). To obtain the residuals, the estimated values are subtracted from the actual values. The smaller the residual, the better the fit and the closer the y hats will be to the ys. A residual is a real-world value that is calculated for each observation every time a regression is run.

$$u = y - y^f$$

Where:

y^f represents the fitted values. Fitted values are the predicted values from a regression computed by applying the regression coefficients to the independent variables. The fitted values are represented by the equation:

$$y^f = Xb$$

b is the least squares coefficient and is represented by the equation:

$$b = (X'X)^{-1}X'y$$

y represents the dependent variable

- **Standard Errors** measure the statistical reliability of the regression coefficients where the larger the standard error, the more statistical noise effects the coefficient. It measures the strength of the residuals. "About two-thirds of the residuals will lie in a range from minus one standard error to plus one standard error, and 95% of the residuals will be in a range from minus two to plus two

44

standard errors"(Quantitative Micro Software, 1995, p. 160). The standard error

is represented by the equation:

$$s = \sqrt{\frac{1}{T-k} u'u}$$

Where:

T represents the number of observations

k represents the number of coefficients

u represents the residuals

- **Sample Variance** is a measure of dispersion. In other words, it shows how the

    sample data varies around the mean. It is represented by the equation:

$$s_x^2 = \frac{\sum_{t=1}^{T} (x_t - \bar{x})^2}{T-1}$$

Where:

x represents the variable

T represents the observations on the economic variable x

x bar is the sample mean

In other words, the sample variance is "the average squared distance between x_t,

the observed values of x and the sample arithmetic sample mean x bar" (Griffiths

et al., 1993, p. 23).

- **Standard Deviation** is a vital measure of a normal distribution. It is the square

    root of the variance. It is a measure of the dispersion in the

    sampling distribution of β_k. The standard deviation is represented by the

    equation:

$$s_x = \sqrt{s_x^2}$$

Where:

s_x^2 represents the sample variation

"Large values of s_x and s_x^2 indicate more dispersion of the values of x about the arithmetic sample mean, and small values of s_x and s_x^2 indicate that the values of x are clustered around the sample mean" (Griffiths *et al.*, 1993, p. 23).

- **Degrees of Freedom** are the number of observations minus the number of coefficients estimated. The lower the degrees of freedom, the less reliable the estimate is likely to be. "Thus the increase in the quality of the fit cause by the addition of a variable needs to be compared to the decrease in the degrees of freedom before a decision can be made with respect to the statistical impact of the added variable" (Studenmund, 1992, p. 50).

$$n - k - 1 > 0$$

If n-k-1 ≤ 0, then "the equation cannot be estimated, and if the degrees of freedom are low, precision is low" (Studenmund, 1992, p. 410). When this happens the best course of action is to include more observations in the model.

- **The Error Term** is the difference between the observed y and the true regression equation or the expected value of y. It is a theoretical value that can never be observed. The error term is derived from the omitted explanatory variables. The variation described by the error term is probably the result of "omitted influences, measurement error, incorrect functional form, or purely random and totally

unpredictable occurrences" (Studenmund, 1992, p. 10). The equation describing

the error term is:

$$\varepsilon_i = Y_i - E\left(Y_i \middle/ X_i\right)$$

Where:

Y_i represents the observed value of Y

Y_i/X_i represents the expected value of Y

The difference between the residual and the error term is that the error term can

never be observed while the residual is a real world value that is calculated for

each observation every time a regression is run (Studenmund, 1992, p. 16). This

is so because "the residual is the difference between the observed y and the

estimated regression line (y hat), while the error term is the difference between

the observed y and the true regression equation (the expected value of y)"

(Studenmund, 1992, p. 16).

- **R^2** measures the "goodness of fit." In other words, it measures the success of the

  regression in predicting the values of the dependent variable within the sample. It

  is the fraction of the variance of the dependent variable explained by the

  dependent variable. In other words, it is the proportion of the variance in the

  dependent variable explained by variation in the independent variables. The

  higher the R^2 value, or the closer to one it is, the better the estimated regression

  equation fits the sample data, and the closer the R^2 value is to zero the worse the

  fit of the regression equation with the sample data. The equation which represents

  R^2 is:

$$R^2 = 1 - \frac{u'u}{(y - \bar{y})'(y - \bar{y})}$$

Where:

u represents the residuals

y represents the independent variable

y bar represents the mean of the dependent variable

There is a problem with R^2, in that adding another independent variable to the equation can never decrease R^2. Therefore adjusted R^2 is needed.

- **Adjusted R^2** is R^2 adjusted for degrees of freedom. A slightly different variance is used with this measure than with R^2. Adjusted R^2 will increase or decrease depending on whether the improvement in fit caused by the addition of the new variable outweighs the loss of the degree of freedom. It is usually less than R^2 if there is more than one independent variable. Adjusted R^2 is represented by the equation:

$$\bar{R}^2 = 1 - (1 - R^2) \bullet \frac{T - 1}{T - k}$$

Where:

u represents the residuals

T represents the number of observations

k represents the number of regressors

- **F-statistic** is a method of testing the null hypothesis that includes more than one coefficient. It works by determining whether the overall fit of an equation is significantly reduced by constraining the equation to fit the null hypothesis. It is a test of the hypothesis that all of the coefficients in a regression are zero except the

intercept or constant. If the F-statistics exceeds a critical level, at least one of the coefficients is probably non-zero.

$$F = \frac{\sum (\hat{y} - \bar{y})^2 / k}{\sum e_i^2 / (n - k - 1)}$$

Where:

n represents the number of observations

k represents the number of regressors

- **Multicollinearity** occurs when the independent variable is a perfect linear function of one or more of the independent variables. The consequences of multicollinearity are (Studenmund, 1995, p. 264-267):

  - The estimates will remain unbiased

  - The variances of the estimates will increase

  - The computed t-scores will fall

  - Estimates will become very sensitive to changes in specification

  - The overall fit of the equation will be largely unaffected

  - The estimation of nonmulticollinear (orthogonal) variables will remain unaffected

  - The severity of multicollinearity worsens its consequences.

- **Correlation Coefficient** measures the strength and direction of the linear relationship between two variables (X_1 and X_2). The value lies between -1 and 1. "The correlation between X_1 and X_2 is 1 or –1 if X_1 is a perfect positive or negative linear function of and X_2" (Griffiths *et al.*, 1993, p.44). The larger the

correlation coefficient the larger the linear relationship between the two variables. The relationship is represented by the following equation:

$$\rho = \frac{\text{cov}(X_1, X_2)}{\sqrt{\text{var}(X_1)}\sqrt{\text{var}(X_2)}}$$

- **The Null Hypothesis** is the actual hypothesis being tested and the **Alternative Hypothesis** is the logical alternative that will be accepted if the hypothesis is rejected. An example of a null hypothesis (H_0) and an alternative hypothesis (H_1) is:

$$H_{0:} \quad \beta = \beta_0$$
$$H_1: \quad \beta \neq \beta_0$$

"A statistical test of the null hypothesis is carried out by examining whether or not the information about the value of β contained in a sample of the data supports, or appears to be consistent with, the null hypothesis" (Griffith *et al.*, 1992, p.133). To test a true null hypothesis a standardized normal random variable is constructed (also known as the t-statistic):

$$z = \frac{(b - \beta_0)}{\sqrt{\delta^2/T}}$$

z has the true standard normal distribution $N(0,1)$ if the null hypothesis is true. Griffiths *et al.* discuss three steps for testing a hypothesis about the mean of a normal population when δ^2 is known. The steps are:

(a)   Use the least squares estimate b of β to calculate the value of z

(b)   Reject the null hypothesis H_0: $\beta = \beta_0 \, |z| \geq z_c$ or

(c)   Do not reject the null hypothesis if $|z| < z_c$

- **Level of Significance** indicates the probability of observing an estimated t-value greater than the critical t-value. The level of significance is usually between 0.01 and 0.05.

  It is represented by the equation:

  $$P[b \geq \beta_U] = P[b \leq \beta_L] = \alpha/2$$

  "If $b \geq \beta_L$ or $b \leq \beta_U$ we reject the null hypothesis and if $\beta_L \leq b \leq \beta_U$ we do not reject the null hypothesis. When a null hypothesis is rejected it means that the sample information does not support the null hypothesis and thus we conclude that it is unlikely to be true" (Griffiths *et al.*, 1993, p.135).

- **Probability** shows the likelihood of rejecting the null hypothesis with the t-statistic value that was produced. Usually, a probability lower than 0.05 is strong evidence that the null hypothesis will be rejected.

- **Confidence Intervals** can be calculated about the forecasted value. This confidence interval is calculated using the following equation:

  $$s_F^2 = s^2 \left[1 + \frac{1}{n} + \frac{(X_F - \bar{X})^2}{\sum (X_i - \bar{X})^2} \right]$$

4.4   Sensitivity Analysis

Sensitivity analysis is a way of varying one or more factors to determine what impact the variance has upon the forecast. In other words, when running the simulation using risk analysis software, only one of the variables is altered in the equation while the others remain the same to determine what impact the change has upon the forecast. Each time a simulation is run another variable is altered and the results from all of the simulations are compared. Sensitivity analysis thereby identifies those factors that have the greatest impact upon the forecast which "allows evaluators to gather additional data in a more efficient manner" (Torries, 1998, p. 55). Sensitivity analysis is frequently used to "determine how much change in a variable would be necessary to reverse the decision based on average-value or best-guess estimate" (Stermole and Stermole, p. 246) or to determine the impact of price changes upon the forecast.

4.5   Scenario Analysis

Scenario analysis is a means by which "decision makers understand the uncertainty created by multiple combinations of input factor values; as a result, they sometimes investigate the results of scenarios in which combinations of variables are changed" (Torries, 1998, p. 56). "The scenario analysis identifies of combinations of inputs which leads to output target values" (Palisade Corporation, 1995, p. 3-7). Scenario analysis uses groupings to determine which causes a particular output value. While there are many different types of scenarios in scenario analysis, a common arrangement is to compare the base case or the expected case, the optimistic case, and the pessimistic case. The base or expected case is developed from the "best" estimates from the forecast. The base or expected case is not a true expected value, because "the expected case can only be determined through probabilistic analysis as the distribution of the input factors

has not been considered in the determination of the base case" (Torries, 1998, p. 57). The pessimistic case shows the probable outcome when nothing goes as expected, while the optimistic case shows the results when everything goes better than expected.

The three cases can be used to estimate the distribution of the input values. A basic concept of probability theory is that a range of values and probabilities are used, because a single input cannot reflect the risk associated with the variables. Using the least, most likely, and greatest values of a variable allows for a more accurate quantification of the risks included in the variable than one single value. Crow *et al.* (1981) suggests that: "the future is ultimately unknowable. Therefore, it is reasonable to select not only the best guess about the future, but also maximum and minimum plausible values so that some range of likely electricity consumption can be forecast" (p. 49-50).

This arrangement of the three cases can be portrayed as a triangular distribution as shown in Figure 3. If a triangular distribution is symmetrical it has the same property as a normal distribution in that the value that occurs the most is also the mean value (Torries, 1998, p. 60). However, the distribution of triangular distributions may also be skewed.
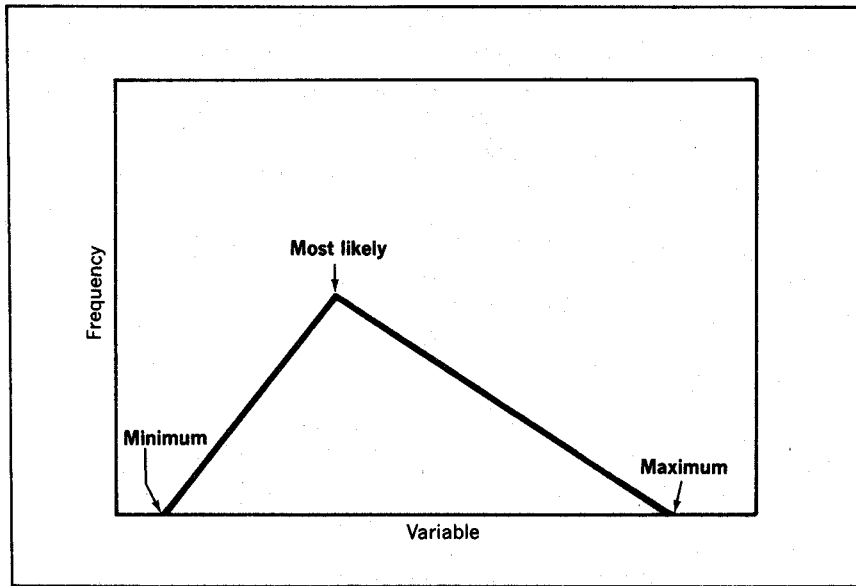
Figure 3:   Triangular Distribution

"The lowest and highest values are considered to have low probabilities (say, less than 5%) and the base values the highest probabilities of occurrence" (Torries, 1998, p. 59). The area under the triangle represents the frequency of all the events that can occur. Thus, greater frequencies have greater area, and smaller frequencies have a smaller area. The use of probabilistic inputs for the variables accounts for the risk involved with uncertain values of these variables.

4.6   Monte Carlo Simulation

Monte Carlo Simulation is a technique used to determine the probabilistic distribution of an outcome that relies on all probable scenarios. "Monte Carlo sampling refers to the traditional technique for using random or pseudo-random numbers to sample from a probability

distribution" (Palisade Corporation, 1995, p. B-5). It is a method of recognizing and determining risk. Monte Carlo Simulation produces not only one answer, but rather a series of answers or a range over which the results vary as a function of probability of occurrence and also a most expected result. In other words, Monte Carlo Simulation generates hundreds of alternative (scenarios) for a project. The answer may fall anywhere within the range of the results produced.

All inputs are independent of one another. Thus, when the variables are correlated a correction to the Monte Carlo Simulation is required. To account for correlation, a correlation matrix in the Monte Carlo Simulation is used. The correlation value for each set of variables is entered in the matrix, and when the simulation is run, correlation among the variables is accounted for.

Monte Carlo simulation is done using risk analysis software. The results of the analysis may be viewed differently by different people. This variety of opinions is because of the differences in the level of acceptable risk among individuals.

4.7   Preference Theory & Certainty Equivalence

Walls and Torries (1998) discuss the concept of preference theory and certainty equivalence. With preference theory a utility function is produced which is representative of the individual's or business' willingness to accept or reject risk. The utility function is represented by the equation:

$$u(x) = -e^{-x/RT}$$

Where:

RT represents the risk tolerance coefficient

x represents the monetary variable

e represents exponential constant

Certainty equivalence is used to determine the individual's preference for what type of risk they are: risk seeking or risk averse. The equation representing certainty equivalence is:

$$C_x = -RT \ln \left(\sum_{i=1}^{n} p_i e^{-x_i/RT} \right)$$

Where:

p_i represents the probability of outcome I

x_i represents the value of outcome I

n represents the total number of possible outcomes

Certainty equivalence makes a distinction between the risks involved between two different outcomes. "The Cx valuation measures the trade-offs between potential and uncertain upside gains versus downside losses with respect to the investor's risk propensity" (p. 125).

4.8   Forecasting

Forecasting is a means of predicting what is going to happen in the future—next month, year, decade, etc. Accurate forecasting requires high quality data, application of the appropriate forecasting technique, and knowledgeable interpretation. "The accuracy of such forecasts depends in large measure on the degree to which the past is a good guide to the future" (Studenmund, 1995, p. 4).

Prior to building a forecast model, the first step is to clearly understand the problem in order to establish the forecast range and objectives (Al-Alawi, 1996, p. 142). Therefore, the four steps of forecasting are:

1. Collect and analyze the data through graphs to determine if any of the numbers seem inconsistent with the others.

2. Make assumptions consistent with the data

3. Test the relationships using statistical analysis

4. Feed back the results—iterate, con problems until get reliable and reasonable results (Burns, 1984, p. 1439).

A forecasting model for electricity demand includes:

- Historical, current, and predicted values of demand-determining variables

- Identification of the equations and parameters that relate those variables to electricity demand

- A method for solving the electricity demand model for each forecast year and summarizing the results. (Mitchell, Park, & Labrune, 1986, p.4)

Burns (1984) suggests that "the time period should be long enough to permit a trend to develop and so short as to depend solely on the outcome of particular events" (p. 1439). Forecasting is the estimation of the expected value of a dependent variable for observations, which are not in the sample data set, but rather at some point in the future (Studenmund, 1995, p. 574). "Forecasts are generated (via regressions) by estimating an equation for the dependent variable to be forecasted, and substituting values for each of the independent variables (for the observations to be forecasted) into the equation" (Studenmund, 1995, p. 595).

A problem that may arise is that just because an equation is an excellent fit with the sample data, does not mean that the equation will forecast well outside of the time period of the data set. Another problem is the need to anticipate the impact of future technology (Burns, 1984). Yet another problem is that the farther into the future one goes, the more likely errors are

going to occur in the forecast (Granger, 1986, p. 7). In other words, no forecast beyond the short-term can be assumed to be highly accurate, but the more accurate the forecast, the better the decisions that are made from it.

Electricity load forecasts are inherently uncertain. Therefore, the forecaster needs to understand the uncertainty and include it in the analysis. "The forecaster needs to present quantitative measures of forecast uncertainty and to understand the sources of forecast error as they relate to key factors that influence electricity demand" (Charles River Associates, 1995, p. 8). A method of forecasting electricity demand that produces accuracy for one to three years is not usually the best method for forecasting long-term demand. "Disaggregating forecast models by end use, timing of loads, geography, or other facts can improve forecast accuracy or the usefulness of forecasts in planning, but it increases model complexity and makes models more difficult to obtain" (Charles River Associates, 1995, p. 9).

Because forecasts are inherently uncertain, uncertainty exists about the future direction of explanatory variables. Uncertainty also exists in relationships or "causality" among different dependent and independent variables such as changes in consumer behavior in response to higher or lower prices. Also, uncertainty exists in the observed explanatory variables due to presence of "noise" or errors of estimation and measurement plus other factors. Thus, there is no guarantee that the past is an accurate representation of the future. Yet, another forecasting problem is the need to anticipate future technology (Burns, 1984).

Another problem with forecasting is that of inconsistencies in the data. The forecast is intended only for Appalachian Power Company's service area in West Virginia. In a perfect world all of the data should be based upon that service area. The problem arises when not all of the data are available on such a specific level. It is only available for the state or not a national

level in some instances. While these values can be used as proxies for Appalachian Power Company's service area, the forecast is not the most accurate for that area.

Another problem is that of geographic regions. This problem is in conjunction with the inconsistency of data. As said previously, not all of the data were available for only Appalachian Power Company's service area. Therefore, some of the data are only available on a state or national basis. The geography is different in Appalachian Power Company's service area from that of rest of the state. Therefore, data that do not recognize these differences may not be appropriately used.

4.9   Predictive versus Explanatory Models

Predictive models attempt to determine what is going to happen in the future, while explanatory models attempt to explain a situation based upon empirical evidence. Where predictive models are used to "predict" future events, explanatory models are used to "explain" why an event took place.

Predictive models are used to predict what the future outcome may be. Predictive models extrapolate the trend to determine what the future will be—future growth, decline, or remain the same. Predictive models are used for the planning of future projects, acquisitions, or long-term need. Predictive models are usually time-series in nature as time-series is a mean of predicting future outcomes based upon past trends. Trend extrapolation (see section 2.3.1) is also used to determine future trends.

Explanatory models attempt to explain why something is the way it is based upon some theory. Therefore, explanatory models must be based on theory—if there is no theory, then the

situation cannot be explained. Explanatory models show if the theory tested makes sense. It is a test of whether or not the empirical evidence is consistent with the explanation.

4.10 Functional Form

The functional form is the form that the electricity demand equation takes, e.g. linear, logarithmic, or exponential. The choice of a functional form is a critical one. To choose a functional form, one should determine the logical relationship between the independent and dependent variables and compare them with the properties of various functional forms. The one that comes closest to agreeing with the underlying theory should be chosen. As Griffith *et al.* (1993) suggests: "The first consideration in this regard should always be to choose a functional form whose characteristics reflect the economic nature of the relationship. Should the elasticities be constant over the whole range of the explanatory variables? Should the marginal responses (first order partial derivatives) be constant over the whole range of explanatory variables" (p. 344)?

The functional form chosen to represent the residential summer peak forecast for Appalachian Power Company is the semi-log form. The semi-log functional form is used when economic theory shows that some but not all of the variables are expressed in terms of their logs as the following equation depicts (Studenmund, 1995, p. 220):

$$Y_i = \beta_0 + \beta_1 \ln X_{1i} + \beta_2 \ln X_{2i} + \varepsilon_i$$

The elasticity for the semi-log equation can be expressed as (Studenmund, 1995, p. 220):

$$\eta_{Y,X_1} = \frac{\Delta Y}{\Delta X_1} \bullet \frac{X_1}{Y} = \frac{\beta_1}{Y}$$

The equation shows that β_1/y decreases as y increases.

The traditional method of energy demand analysis use either linear or double-log forms for determining energy demand. According to Plourde and Ryan (1985), the double log form is a reasonable choice of a functional form for an equation to determine energy demand. The reasoning behind this is that the estimated coefficients are estimates of the elasticities, and the standard errors provide measures of the variability of the estimated elasticities. Yet, there are problems with the double log form and forecasting energy demand. One problem is that because the elasticities are constant they do not vary according to income level or price. "An additional problem with the use of double-log forms in energy demand analysis is the relationship of the form to the economic theory of consumer (or producer) behavior" (p. 107). In other words, it is also important to determine whether energy demand functions which use the double-log form can be "obtained from a less restrictive representation of consumer preferences" (p. 107).

Harvey (1983) highlights the idea that when there is an economic time series the logarithm should be used, because if a series contains seasonal components, e.g. temperature, these components tend to combine multiplicatively. In other words, a seasonal pattern appears more stable after a logarithm is taken.

Munley, Taylor, and Formbly (1990) dismiss the idea of traditional linear and double log forms of energy modeling. Instead, they show that the method that produces the best results is when the demand equation is raised to approximately the one-half power. Munley *et al.* tested the linear functional form, the inverse functional form where kWh was raised to the one-half power, and the double log functional form to determine which method is most appropriate. What is interesting about the results is that they showed that the price and income elasticities of electricity demand are almost identical for all three functional forms tested (p. 187). Also, the

"hypothesis tests about the significance of individual independent variables would lead to the same general conclusions under any of the three specifications" (p.187). Although Munley *et al.* admit that the hypothesis tests for the three functional forms are not statistically different, they still conclude that the linear or logarithmic forms should be rejected and the inverse functional form where kWh is raised to the one-half power is the correct functional form.

Nan and Murry (1992) determine that the flexible double-log is the proper functional form for an energy demand equation. This functional form is developed to meet the assumptions of consumer behavior. "The traditional log-linear functional form has some short comings of constant elasticities. The regression equations in this study, with varied estimated elasticities, have overcome some of these short comings" (p. 153). The problem with this method is that of the constant elasticities.

4.11 Consumer Price Index

The consumer Price Index (CPI) adjusts dollar figures for inflation and is produced by the Bureau of Labor Statistics (BLS). The CPI is a statistical measure of change, over time, of the prices of goods and services in major expenditure groups—food, housing, apparel, transportation, and medical care typically purchased by urban customers. "The CPI measures changes in prices only of goods and services that families purchase for consumption, which account for about sixty-five percent of total output. The remaining thirty-five percent goes to business investment, government services, and the goods and services that we buy from and sell to foreigners" (Miller, 1994, p. 153).

A price index compares the cost of a sample "market basket" of goods and services in a specific month relative to the cost of the same "market basket" in an earlier reference period. A

price index is calculated starting at some base year which is the reference year for comparison of prices in other years. It is defined by the equation:

$$\text{Price Index} = \frac{\text{Cost of Market Basket}}{\text{Cost of Market Basket in Base year}} * 100$$

The CPI has some problems associated with it. One is that it is a fixed-quantity price index which means "that each month the BLS samples only prices rather than relative quantities purchased by consumers" (Miller, 1994, p. 135). The reason that this is a problem is that of the demand curve and the substitution effect. As prices increase consumers substitute in favor of a lower price good, yet when prices decrease, the consumer purchases more of the first product. Another problem is that the BLS only obtains list prices, it does not account for free warranties, sales, etc. Also, the BLS does not account for quality changes. Yet another problem is that the BLS does not account for improvements on a product whose price remains the same despite the improvements.

5.0   Methodology

This section describes the methodology used to forecast summer peak demand for Appalachian Power Company's service area in West Virginia. Included in the discussion are the variables that potentially impact summer peak electricity demand for this service area. Also discussed are the steps taken to analyze the model using regression analysis and Monte Carlo Simulation.

5.1   Model Development

Residential electricity demand is affected by many different factors. Halvorsen suggests that "residential energy demand is assumed to be a function of the number of customers, electricity and gas prices, per capital income, and several noneconomic variables" (Halvorsen, 1970, p. 612). Therefore, when forecasting a residential energy demand model, one must account for many different variables that combine to form the basis of a residential energy demand model. The forecaster needs to account for the consumers' buying power. The majority of residential energy demand models include variables that measure economic capacity such as income, temperature or cooling/heating degree-days, price of a substitute such as natural gas, and the price of electricity. Some models also include a variable for the number of residential customers.

Heat waves, humidity, and soaring temperatures cause problems during the summer when temperatures reach uncomfortable levels. When the temperature rises, the demand for electricity increases due to the increased demand for cooling devices. In other words, everyone turns up their air conditioning. Thus, the summer peak usually occurs on one of the warmest day of the

summer. A temperature variable such as maximum temperature or cooling degree-days measures the relationship between temperature and electricity demand.

When electricity prices are high the consumer tends to use less electricity and when the prices are low the consumer tends to use more. Also, if a consumer has a choice between two comparable goods, the consumer will usually choose the less expensive good. In the case of electricity consumption, a substitute could be natural gas.

Total income is included in an electricity model to measure the consumer's buying power. It is used to determine how much the consumer can afford to pay for electricity. When income is higher a consumer can afford to spend more on electricity which increases demand. The consumer whose income is low must watch and use only the electricity they can afford. Therefore, electricity consumption is affected by a consumer's income. Also, those with higher incomes tend to have larger homes than those with lower incomes, and larger homes use more electricity to cool the home.

Without consumers there is no need to forecast residential energy demand. Therefore, a variable measuring the number of customers should be included in the model. One method of measuring the number of customers is to actually measure the total number of residential customers in Appalachian Power Company's service area. Another method is to use total population in the service area. By measuring the total population, the researcher is able to include some of the consumers of electricity that may not actually be included in the number of residential customers.

5.2   Description of the Data

All data are yearly from 1980 to 1995 (shown in Appendix A in Tables 12-14). The data series starts with 1980 because of the difficulties in using data prior to the oil crisis of the 1970's. If the data started with the 1970's the results would be inordinately skewed because of the higher fuel prices resulting from the crisis. This situation does not exist in the United States today for today's fuel prices are relatively stable and are expected to remain stable over the forecast time period. Thus, to forecast the demand for electricity accurately it is important to use the scenario that exists in today's world, rather than that of the 1970's.

The data are also at the county-level where county-level data were available. If county-level data were unavailable, then state-level data were used. County-level data were desired because Appalachian Power Company only services certain counties in West Virginia. These counties are Boone, Cabell, Fayette, Jackson, Kanawha, Lincoln, Logan, McDowell, Mason, Mercer, Mingo, Nicholas, Putnam, Raleigh, Wayne, and Wyoming counties. The county data also included data for West Virginia Power Company, a small electric company supplied by Appalachian Power Company. The counties in the West Virginia Power Company service area are Greenbriar, Monroe, and Summers counties. These counties are highlighted in Figure 4 below.
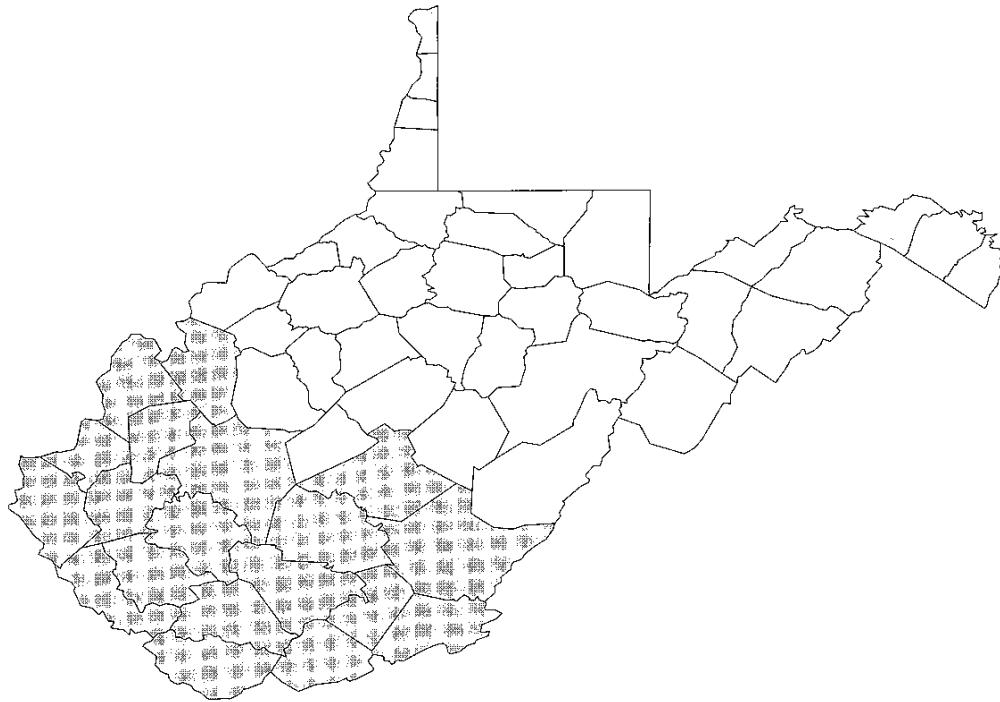
Figure 4:   Map of Appalachian Power Company's Service Area in West Virginia

When available, monthly data for the month of the summer peak were used. When unavailable quarterly data were used for the quarter when the summer peak occurred.

5.3   Variables

The variables represent the numerous factors that influence summer peak residential demand. While the results may show that all of the variables do not have an impact upon summer peak electricity demand in West Virginia, they may have an impact upon electricity demand in other states or on the national-level.

5.3.1 Peak Summer kWh

Total energy demand is the amount of electricity consumed by Appalachian Power Company's residential customers on the day of the summer peak. Appalachian Power Company provided the data for energy demand.

5.3.2 Trend Variable

A trend variable is simply a number one through fourteen. It measures the trend involved in the forecast. A trend variable helps to determine if there is a trend in the data.

5.3.3 Maximum Temperature

Temperature is a variable that measures the highest temperature on the day of the summer peak. The theory behind this is that the hotter the temperature, the more electricity needed due to increased use of air conditioning, fans and other cooling devices. The problem with the temperature variable is that it does not account for other factors that affect temperature: solar radiation, wind, humidity, or other variables.   The temperature data were taken from the West Virginia Climatological Journal for each month, day, and year of the peak day. Since demand for home cooling will be positively related to summer temperature, the demand and temperature should be positively related.

5.3.4 Cooling Degree-Days

Cooling degree-days are obtained by adding the high and low temperature for the day and then subtracting that average from 65, the base number. Cooling degree-days relate the day's temperature to the energy demands of air conditioning. Cooling degree-days can be used to

relate how much more or less one might spend on air conditioning in one part of the country than someone in another part of the country. Crow *et al.* (1981) suggests that one problem with using cooling degree-days is that they do not take into account all weather factors, e.g. solar radiation, wind, and humidity. Another problem discussed by Crow *et al.* is the unreliability of the benchmark temperature. It may not be the correct choice for all areas or even for all utilities. The temperature data used to calculate cooling degree-days were found in the <u>West Virginia Climatological Journal</u> for the peak month, day, and year. Cooling degree-days and the demand for electricity should be positively related.

5.3.5 The Number of Residential Customers

Since this is a study of a residential summer peak model, there needs to be a variable that measures the total number of residential customers. Electricity use by residential customers tends to peak in the morning when the household is getting up and getting ready for work and school and again at night when they return home and cook dinner, watch television, etc. Therefore, the residential customer variable is one of the most important when modeling residential energy demand. The data for total number of residential customers were found in <u>Appalachian Power Company's Annual Statistical Report</u> to the West Virginia Public Service Commission for each year of the study. Since residential demand depends on the number of residential customers, the relationship between residential demand and electricity is expected to be positive.

5.3.6 Total Population

Total population is a measure of the total population in Appalachian Power Company's service area in West Virginia. A more exact measure is the total number of residential customers, but total population is an attempt to measure the differences in the population, e.g., urban and rural differences and different demographics in the population. The total population also accounts for all people that use electricity where total residential customers only measures the number of people that purchase electricity from the power company. Total population figures for West Virginia were found on the REIS CD-ROM 1969-1995. The relationship between population and electricity demand is expected to be positive, because the greater the population the more demand upon electricity.

5.3.7 Percent Rural Population

Urban and rural customers have different needs when it comes to residential electricity consumption. Urban consumers tend to use less electricity than rural consumers, because a greater number of urban consumers live in apartments and other multi-unit structures, while rural consumers tend to live in homes. Multi-unit structures take less electricity to heat and light than homes. Therefore, rural consumers tend to use more electricity than urban consumers. Since data were unavailable for total urban and rural residential customers, percent rural population was used as a proxy. Data for total urban and rural population were found in the 1980 Census of Population, General Population Characteristics for West Virginia, and the 1990 Census of Population and Housing Summary Tape File 3A CD-ROM.

5.3.8 Price of Electricity

The price of electricity is important, because it shows an individual's willingness to pay. Le theorizes that "since income is limited, the consumer will react to price and income changes by changing demand to satisfy budget constraints as the price of electricity increases, the price and income effects show that with the higher prices the consumer may conserve or use a substitute good" (Le, 1983, p. 221). The only problem with this variable is that the price of electricity is so low in West Virginia that it may not have a recognizable impact upon demand. Energy prices from 1983-1987 are from Appalachian Power Company's Annual Statistical Report to the West Virginia Public Service Commission, and energy prices from 1988-1995 are from EIA-861 ("Annual Electric Utility Report") published in Electric Sales and Revenue. The price of electricity and the demand for electricity should be negatively related.

5.3.9 Price of Natural Gas

When developing an economic model one always has to consider the effects of substitute goods on the demand for the good in question. When modeling the demand for electricity the natural substitute is natural gas. If given a choice, the consumer will chose the one with the lower cost to heat or cool their home. The problem with this theory is that many people are not given the choice unless they build their own home or include fuel choice when buying a home. Therefore, consumer choice may be obscured by the costs of home construction.

Natural gas prices for West Virginia were found on Energy Information Administration's home page. The price series for natural gas was selected for the years 1984 to 1995. Since the natural gas industry experienced deregulation in the years prior to 1984, natural gas prices would

not be an accurate reflection of today's demand. The price of natural gas and the demand for electricity are negatively related, because natural gas is a substitute for electricity.

5.3.10 Per Capita Income

Total personal income is a measurement of the economic strength of the consumer. In other words, as personal income increases, the consumption in the household sector is expected to increase. Total income was found on the REIS-CD ROM 1969-1995. The demand for electricity and total income should be positively related to summer peak, because the higher an individual's income the greater their demand for electricity.

5.3.11 Age

Age is an interesting variable, because the older one is the less electricity they are likely to use. Younger people tend to use more electricity than the elderly, because they have families. A family uses more electricity than just one or two people. Also, younger people tend to have more appliances that use electricity, e.g. computers. Also, for most elderly people, their income decreases when they reach retirement age. Therefore, they will lessen their electricity use to meet their current income standards. Therefore, the age differences within a community may have an impact upon residential energy use. Age was obtained from the 1980 Census of Population, Characteristics of the Population, General Social and Economic Characteristics for West Virginia, and the 1990 Census of Population, Social and Economic Characteristics for West Virginia.

5.3.12   Total Number of Households

The total number of households variable is used because the household is who is obtaining the energy for residential use. Therefore, the more households there are the greater the demand for energy. The total number of households data were obtained from was obtained from the 1980 Census of Population and Housing, Summary Characteristics for Government Units and Standard Metropolitan Statistical Areas for West Virginia, and the 1990 Census of Population and Housing, Summary Population and Housing Characteristics for West Virginia.

5.4   Method of Analysis

The two types of analysis of the Appalachian Power Company's summer peak demand model for its service area in West Virginia were completed to better understand the risk and uncertainty involved in the forecast. A conventional regression analysis was completed and the results analyzed. The results of the regression analysis were then used to complete a probabilistic forecast using Monte Carlo simulation.

The first step of the statistical analysis is to develop a model with summer peak energy demand in Appalachian Power Company's service area in West Virginia as the dependent variable and various combinations of the variables discussed in Section 5.3 above as the independent variables. The next step is to run a regression analysis on the equation to determine which variables have an impact upon the dependent variable and which do not. Those independent variables that do not appear to have an impact upon the dependent variable are removed from the equation and a regression is run on the new model. This process is repeated until the results appear to make sense by choosing the most statistically significant results. Any problems such as multicollinearity among the independent variables are expected to be corrected

in this process. However, multicollinearity could not be removed from the forecast results because of unresolved problems with the data set. These problems are discussed in the next section, Section 6.0.

The first step in Monte Carlo Simulation is to use the statistical model as the basis of the probabilistic model. Next the data set for the variables in the statistical model are forecasted until the year 2005. This forecast is accomplished by regressing each independent variable against a trend variable. The results of this regression equation are used to produce a forecasted value for the independent variables. The results of the forecasted values of the independent variables are analyzed to determine if they are logical. If they are not logical as determined by analyzing a graph of the forecasted values and comparing future values to past values, then the forecast is accomplished by using the subjective method of analysis of trend extrapolation.

A distribution about the forecasted values is calculated using the equation for calculating a confidence interval. If the resulting confidence intervals of the input values do not appear to be logical based upon an analyses of the graphs of the confidence intervals and the input values, the distributions for these values are also determined using trend extrapolation. A distribution about the estimated coefficients was also calculated using the standard error of the estimated coefficients.

The next step is to develop a probabilistic equation based upon the statistical equation using the distributions about the coefficients, the coefficients, the distributions about the forecasted variables and the forecasted variables. The Monte Carlo simulation on the model is accomplished by using a commercial risk analysis software. The final step is to analyze the results of the Monte Carlo simulation to determine if they are logical based upon an analysis of the results for statistical significance.

Once the statistical analysis and probabilistic analysis are completed a forecast is developed using each method. Although the data are insufficient, which makes the results unreliable, a comparison of the two methods is done to analyze their differences and similarities.

6.0   Discussion of Regression Results

Numerous regressions were conducted to determine which factors influence summer peak residential energy demand for Appalachian Power Company's service area in West Virginia. These results were then compared to the results of the probabilistic analysis. The four regressions in Table 3 below will be discussed in detail. These results are not the best results as defined by a subjective choice based upon many different factors, such as the expected sign, the standard error of the estimated, the R-squared value, etc., but because they produced the most logical results. All other regressions may be viewed in Appendix B in Tables 18-20.

Table 3: Results of Regression Analysis

	Regression 1	*Regression 2*	*Regression 3*	*Regression 4*
Intercept	1944.215 (0.1029)[*]	-7144.578 (-0.9489)	4.627 (4.195)	-186296.355 (-1.053)
Trend		1.883 (0.05)		-4.311 (-0.112)
Electricity Price	-22390.546 (-0.0168)	-420504.007 (-0.295)	-203.87 (-2.482)	-447279.569 (-0.313)
Temperature	22.727 (2.877)	21.825 (2.742)	0.000627 (1.606)	21.851 (2.752)
Residential Customers		0.034 (1.068)		
Log Residential Customers				33623.535 (1.051)
Log Per Capita Income	1591.848 (2.435)	-1247.583 (-0.537)	0.0325 (0.319)	
Per Capita Income				-0.038 (-0.569)
Log Population	-1455.71 (-0.505)		-0.575 (-3.868)	
Log Percent Rural			-1.284 (-2.473)	
R-Squared	0.8347	0.8621	0.9887	0.8627
Adjusted R-Squared	0.7612	0.7759	0.9816	0.777
Standard Error	81.134	78.611	0.00398	78.418
F-Statistic	11.362	10.000	140.063	10.057

*The values in ( ) are the t-statistics

The most logical results were chosen first on the basis of expected sign, and then other statistical factors were considered. Expected signs are what sign the variables are expected to have prior to running a regression. If a sign is not as expected it could be an indication that there is a problem with the results or that there is some matter the researcher failed to consider. Although the sign may be unexpected, it may still be statistically significant. Therefore, the significance of the variable should also be considered.

In Regression 1, temperature and per capita income have expected signs, but electricity price and log population do not. In Regression 2, price, temperature, and per capita income all have expected signs, but the variable for the number of residential customers does not. In Regression 3, price temperature, log per capita income, and log percent rural all have expected signs, but log population does not. In regression 4, electricity price, temperature, per capita income, and log residential customers all have the expected sign.

The Adjusted R-squared values for all four regressions are within a reasonable range. A high R-squared value is desired, but depending on the situation, one that is very high can indicate problems. Therefore, the high R-squared values combined with the low calculated t-scores are the first indication that there is a problem with the results.

The value of the critical t-value for Regression 1 is ±1.812. The t-statistic calculated for the intercept, electricity price, and log population all fall within the acceptable range for the calculated t-statistic value, while temperature and log per capita income do not. The critical t-value for Regression 2 is also ±1.83. The variables whose calculated t-statistics fall within the acceptable range of ±1.83 are the intercept, the trend variable, electricity price, residential customers, and log per capita income and the variable that does not is temperature. The critical t-value for regression 3 is again ±1.83. For this regression the calculated t-statistics for

temperature, and log per capita income all fall within the acceptable range of ±1.83, but the intercept, electricity price log population, and log percent rural do not. The critical t-value for regression 4 is also ±1.83. The calculated t-statistics for the intercept, trend, electricity price, per capita income, and log residential customers all fall within the acceptable range of ±1.83, while the calculated t-statistic for temperature does not.

The correlation statistic gives an indication of when two variables are closely related to one another. (Refer to Table 17 in Appendix B for the correlation matrix.) Per capita income is highly correlated with age at 0.92, residential customers at 0.99, percent rural at -0.93, and natural gas price at 0.95. Age is highly correlated with natural gas price at 0.89, and number of residential customers at 0.95. Percent rural is highly correlated with natural gas price at –0.90 and residential customers at –0.91. Natural gas price and residential customers are highly correlated at 0.946. Electricity price is not highly correlated with any other variable.

Many problems exist within the data. With such a small sample size, the dispersion in the sample is not large enough to produce results that make sense. This problem compounds the problem of multicollinearity. Auxiliary regressions—regression run where one of the dependent variables is regressed against all of the others to determine where the problems lie--were run to determine what variables were affected by the multicollinearity. The results can be viewed in Tables 21-26 in Appendix B. These results show that multicollinearity exists, because the R-squared values are high while the t-scores are low. Another indication of multicollinearity is the high R-squared values combined with low t-scores in the regression results calculated to forecast energy demand. The high standard error of the estimated coefficients is another indication of multicollinearity. One way to solve the problem of multicollinearity is to drop one or more of the variables from the equation. This was done, but did not solve the problem. Another way to

solve the problem is to increase the sample size. This solution was considered, but not done because increasing the sample size would include data from the 1970's which was unacceptable due to the high-energy prices resulting from the energy crisis.   Therefore with the insufficient data that exists for regional energy forecasting, the problem of multicollinearity could not be solved.

Thus, the results of the regression analysis show that none of the regressions produce results that are acceptable. Without accurate data the results cannot be reliable. To confirm the results and to compare the two techniques a probabilistic analysis was done.

6.1   Discussion of Probabilistic Simulation Results

A probabilistic analysis was conducted on Regression 2 and Regression 4. These regressions were chosen because the variables for per capita income have a negative sign. Although not expected, there is some theory behind why the sign may be negative whereas the unexpected negative signs in Regression 1 and Regression 3 for the variable representing the number of residential customers has no theory as to why it may be negative. One reason per capita income may be unexpectedly negative is that those with higher incomes may have better and more efficient housing than those with lower incomes. Another reason is that those with higher incomes are less likely to use inefficient post construction cooling units than those with lower incomes.

Each variable used in Regression 2 and Regression 4 was forecasted until the year 2005 by regressing each variable against time. The results of those equations can be viewed in Tables 27-29 in Appendix B. The resulting equation was then used to forecast the value for each variable. The results of this forecast were unacceptable. When graphed they did not appear to

follow the trend of the variable (see Figures 23-25 in Appendix B for graphs). The reason why the forecast did not follow the trend is that the early values for the variables are different than the latter values which results in a forecast trend that is either lower or higher than the trend. Therefore, trend extrapolation was used to forecast the future values of the variables. See Figure 5, Figure 6, and Figure 7 below for graphs of the extrapolation of future values.

The equation for calculating a confidence interval about the mean value was used to calculate a confidence interval about the forecasted values for the input variables and coefficients. The confidence intervals can be viewed in Tables 30-31 and in Figures 26-29 in Appendix B. The confidence intervals calculated in this manner do not make sense for they are too large to be realistic. The reason they are so large is that the standard errors of the estimated coefficients are used to calculate them and the large standard errors indicate that there is a large variance about the variables. Therefore, the confidence intervals were determine by using subjective trend extrapolation. A graph of each variable was studied and a distribution was drawn about the trend as seen in Figures 5, 6, and 7:
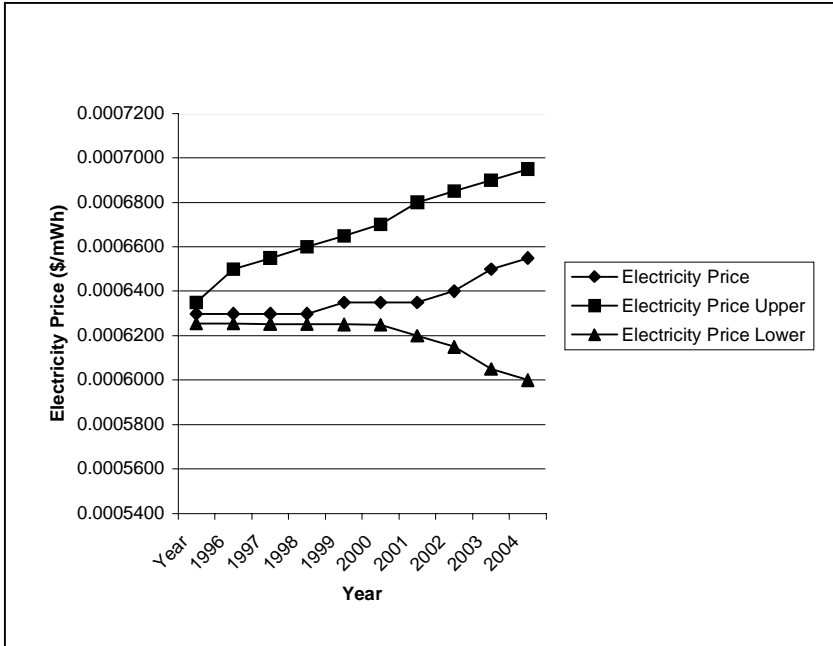
Figure 5: Distribution for Appalachian Power Company's Residential Electricity Price ($/mWh) for West Virginia using Trend Extrapolation
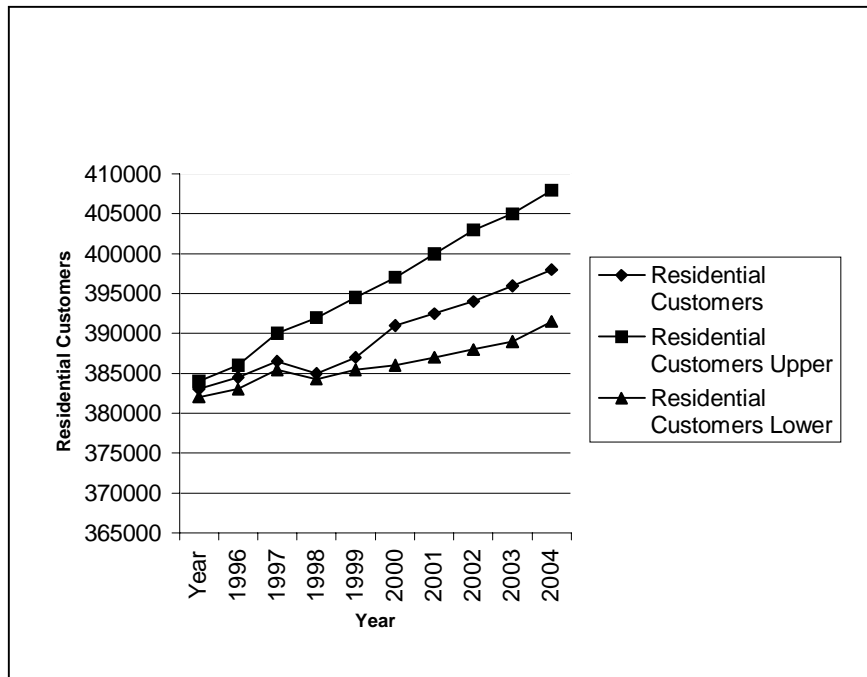
Figure 6: Distribution for the Number of Residential Customers for Appalachian Power Company's West Virginia Service Area using Trend Extrapolation
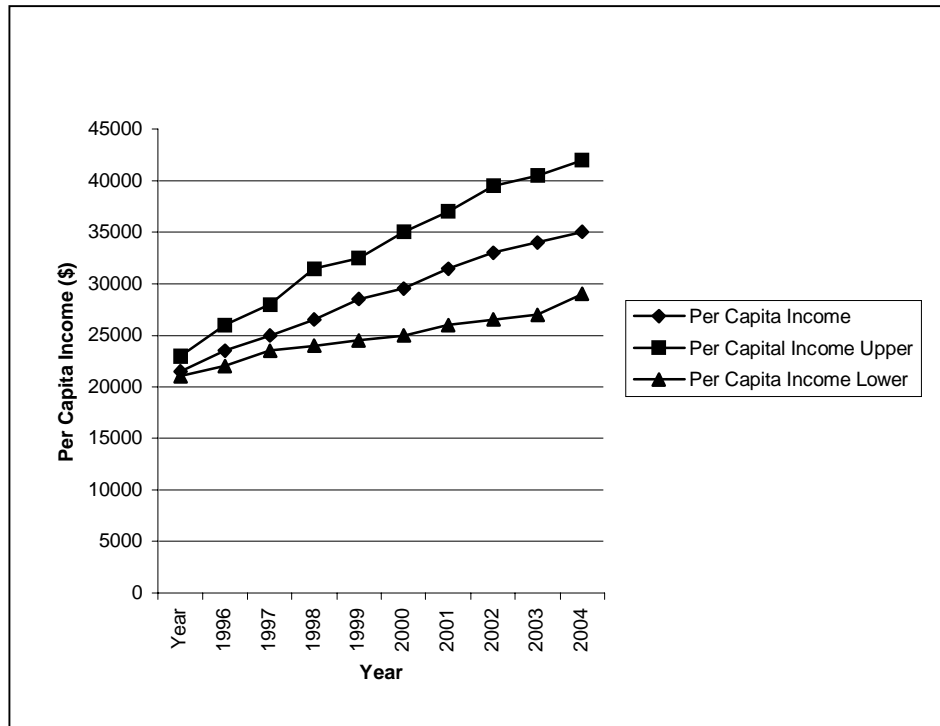
Figure 7: Distribution for Per Capita Income for Appalachian Power Company's West Virginia Service Area using Trend Extrapolation

The distribution about electricity price in Figure 5 is skewed to the high side because prices have been historically stable in West Virginia. However, with deregulation taking place it is more likely that residential prices will rise over the next few years rather than decrease because of the probability of eliminating the residential subsidy that appears to be supplied by higher industrial rates. The distribution increases as time goes on because the future is uncertain and with competition coming due to deregulation, prices could moderate at some point and even begin to decrease. This uncertainty is reflected by a larger distribution about the expected value.

The distribution for the number of residential customers as shown in Figure 6 is skewed toward the high end. It is more likely that the number of residential customers in southern West Virginia will rise rather than fall. Therefore, the distribution for the maximum needs to be larger to account for this anticipated trend.

Figure 7 shows that the distribution about per capita income is relatively unchanged as a percent of the expected value. The distribution range is not skewed in either direction. Per capita income in West Virginia has been increasing steadily over time and it is unlikely that this will change. It is also likely to continue to follow the present trend

The confidence interval about the coefficients was calculated by adding or subtracting the standard error for each coefficient from the estimated coefficient which are shown in Table 4 for Regression 4 and Table 5 for Regression 2.

Table 4: Confidence Intervals about the Estimated Coefficients for Regression 4

	Mean	Upper	Lower
Intercept	-186296	-9447	-363144
Trend	-4.31	34.16	-42.79
Electricity Price	-447279.57	979852.84	-1874411.97
Temperature	21.85	29.79	13.91
Log Residential Customers	33623.54	65621.66	1625.41
Per Capita Income	-0.04	0.03	-0.10

Table 5: Confidence Intervals about the Estimated Coefficients for Regression 2

	Mean	Upper	Lower
Intercept	-7144.578	387.9	-14677.06
Trend	1.883	39.48	-35.72
Electricity Price	-420504.007	1006186.31	-1847194.33
Temperature	21.825	29.79	13.87
Residential Customers	0.034	0.066	0.002
Log Per Capita Income	-1247.583	1077.45	-3572.62

In Table 4 and Table 5 the confidence intervals about the estimated coefficients are large because of large standard errors about the estimated coefficients. The large standard errors are an indication of a large variance about the estimated coefficients, and, as discussed earlier, an indication of multicollinearity among the variables.

The maximum and minimum values of the range of values chosen must have the same probability of occurrence in order to be used in the Monte Carlo Simulation. It is assumed that this probability of occurrence is small. In other words, if there are two triangular distributions, the maximum and minimum values of each distribution have the same probability of occurrence. The maximum value of one distribution must have the same probability of occurring as the maximum value of the other distribution, and the minimum values for each distribution have the same probability of occurrence as all other minimum values.

Summaries of the variables used in the Monte Carlo Simulation for Regression 2 and Regression 4 are given in Tables 6 and 7 respectively. The "most likely" value represents the maximum value in an assumed triangular distribution. Minimum and maximum values about the "most likely" values complete the definition of the triangular distribution for each variable.

Table 6: Variables for Monte Carlo Simulation for Regression 4

Variable
Intercept
Trend
Trend Coefficient
Electricity price coefficient
Electricity price variable
Temperature coefficient
Temperature variable
Log Number of residential customers coefficient
Log Number of residential customers variable
Per capita income coefficient
Per capita income variable

Table 7: Variables for Monte Carlo Simulation for Regression 2

Variable
Intercept
Trend
Trend Coefficient
Electricity price coefficient
Electricity price variable
Temperature coefficient
Temperature variable
Number of residential customers coefficient
Number of residential customers variable
Log per capita income coefficient
Log per capita income variable

Having the high, "most likely," and low values for each variable and coefficient allows

for the risk to be incorporated into the analysis.

Commercial risk analysis software was used to conduct the probabilistic analysis using

triangular distributions. Unlike the statistical analysis, the probabilistic analysis gives a range of

values as answers. The range of values in Table 8 presents the forecast for Appalachian Power

Company's summer peak energy demand for Regression 4 using 1000 iterations.

Table 8: Results of Monte Carlo Simulation for Regression 4

Name	Minimum	Mean	Maximum
1996 / MW/HR	-346321	3484	355696
1997 / MW/HR	-346465	3478	355971
1998 / MW/HR	-346652	3528	356354
1999 / MW/HR	-346751	3476	356628
2000 / MW/HR	-346843	3505	356873
2001 / MW/HR	-346936	3555	357158
2002 / MW/HR	-347069	3570	357467
2003 / MW/HR	-347153	3597	357783
2004 / MW/HR	-347226	3630	357988
2005 / MW/HR	-347453	3661	358275

The problem with the results of the probabilistic analysis for Regression 4 is that electricity demand can never be a negative number as the low value shows, and the high value is so high that it is unrealistic. The mean value is within a reasonable range. The results are shown in Figure 8, which shows the probability of each value occurring for the year 2000, and Figure 9, which show the same statistics for the year 2005. The plots illustrate the unreasonable results from the Monte Carlo Simulation.
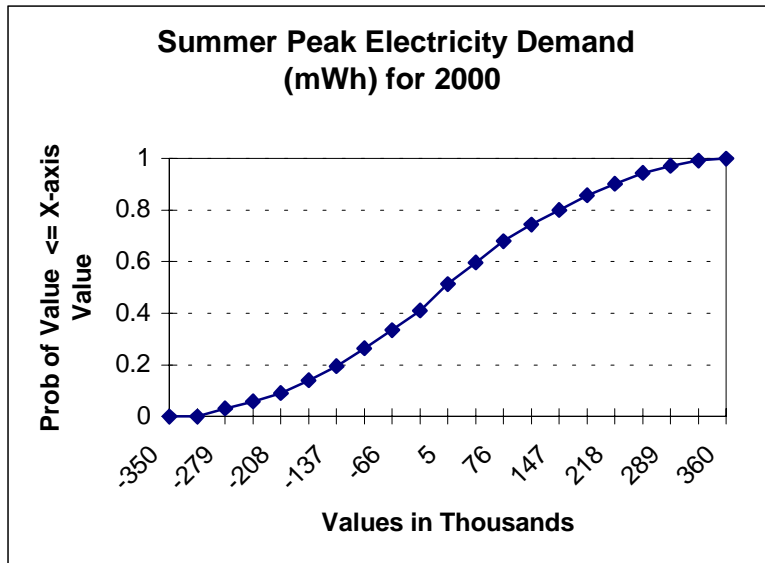
Figure 8: Regression 4: West Virginia's Summer Peak Residential Electricity Demand (mWh) for Appalachian Power Company for 2000
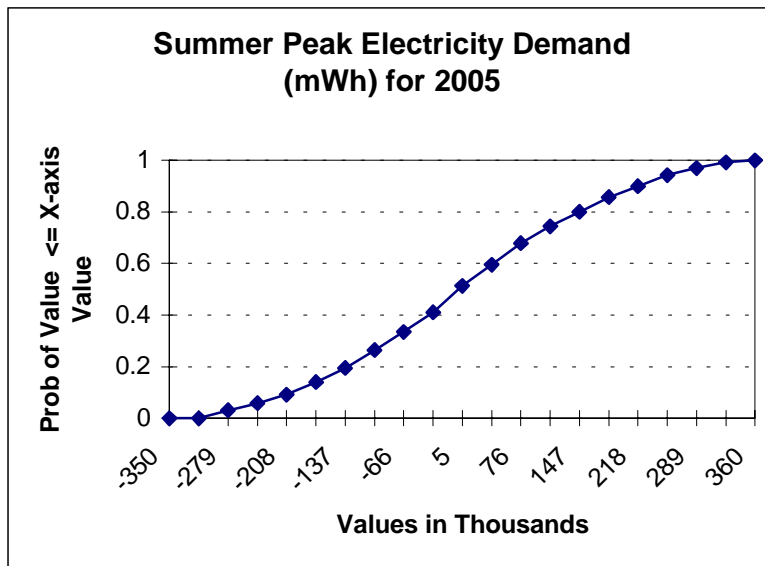
Figure 9: Regression 4: West Virginia's Summer Peak Residential Electricity Demand (mWh) for Appalachian Power Company for 2005

The summary results of the Monte Carlo Simulation for Regression 2 in Table 9 show the forecast for Appalachian Power Company's residential summer peak energy demand for 1000 iterations.

Table 9: Results of Monte Carlo Simulation for Regression 2

Cell	Name	Minimum	Mean	Maximum
B100	1996 / MW/HR	-21443	2386	27568
B101	1997 / MW/HR	-19658	2524	24121
B102	1998 / MW/HR	-20308	2414	25399
B103	1999 / MW/HR	-16550	2219	26616
B104	2000 / MW/HR	-18414	1968	23601
B105	2001 / MW/HR	-19333	2401	25181
B106	2002 / MW/HR	-18204	2446	26915
B107	2003 / MW/HR	-17590	1997	27886
B108	2004 / MW/HR	-21750	2349	28700
B109	2005 / MW/HR	-18056	2254	24257

As with the results for Regression 4, the minimum value is negative which is impossible for electricity demand, while the maximum value is so high that it is unrealistic. The mean value is within a reasonable range. Figure 10 and Figure 11 highlight the probability of the value of electricity demand at various percentages for 2000 and 2005.

The Monte Carlo simulation for both Regression 4 and Regression 2 show uneven growth in electricity demand from one year to another. One explanation for this decline is that the demand for summer peak energy is dependent upon the temperature. One summer may be warmer than the next, resulting in a fluctuation in summer peak demand.
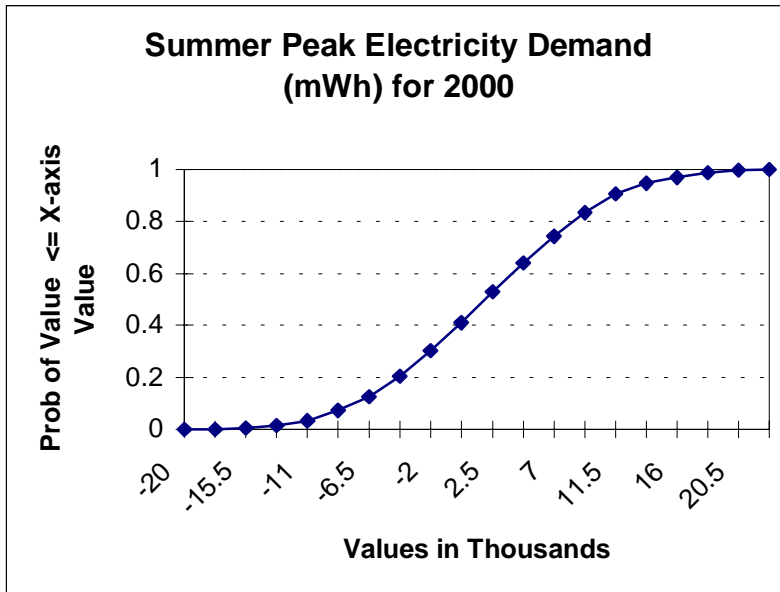
Figure 10: Regression 2: West Virginia's Summer Peak Residential Electricity Demand (mWh) for Appalachian Power Company for 2000
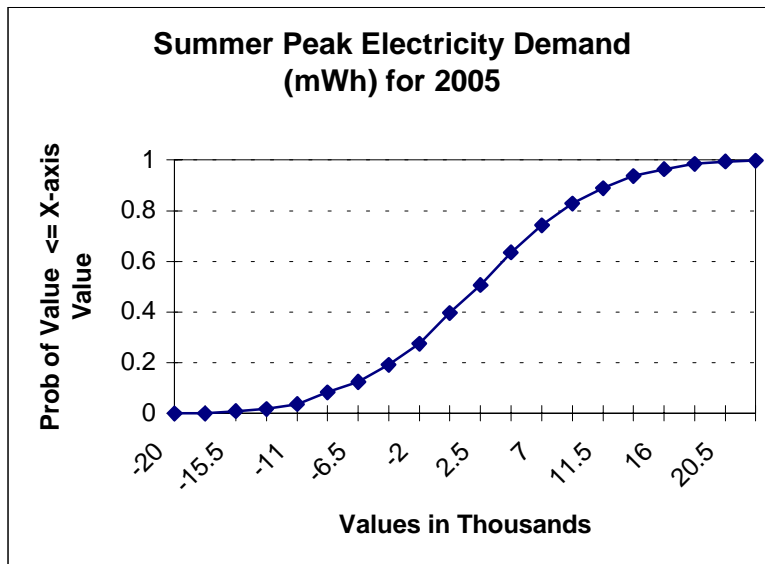
Figure 11: Regression 2: West Virginia's Summer Peak Residential Electricity Demand (mWh) for Appalachian Power Company for 2005

The results of the probabilistic analysis could be created using the statistical method by using the standard errors to calculate the maximum and minimum values around the forecasted values. Probabilistic analysis was used, because it computes different scenarios based upon the confidence interval. Although similar conclusions may be reached using statistical analysis by looking at the t-statistic, probabilistic analysis gives a visual portrayal of the input and output distributions. This difference may be important if the input distributions are not normally distributed. To compare the forecasting using the statistical and probabilistic methods, a forecast was developed with each technique. The forecast developed using statistical analysis for Regression 4 and Regression 2 are shown in Tables 10 and 11 respectively.

Table 10: Forecast for Regression 4 using Statistical Method

	Minimum	Mean	Maximum	Standard Error Of Forecast
1996	3819	3963	4107	78.418
1997	3948	4092	4235	78.418
1998	4076	4220	4364	78.418
1999	4072	4216	4360	78.418
2000	4217	4361	4505	78.418
2001	4401	4545	4689	78.418
2002	4529	4673	4816	78.418
2003	4635	4779	4923	78.418
2004	4738	4882	5026	78.418
2005	4843	4987	5131	78.418

Table 11: Forecast for Regression 2 using Statistical Method

	Minimum	Mean	Maximum	Standard Error of Forecast
1996	2436	2293	2436	78.611
1997	2435	2291	2435	78.611
1998	2469	2326	2469	78.611
1999	2386	2243	2386	78.611
2000	2415	2271	2415	78.611
2001	2532	2388	2532	78.611
2002	2545	2401	2545	78.611
2003	2570	2427	2570	78.611
2004	2622	2479	2622	78.611
2005	2674	2531	2674	78.611

The results show that the statistical method forecasts summer peak electricity demand at a much higher level for Regression 4 than does the statistical method. The results also show that the statistical method for Regression 2 produces results that are higher than the probabilistic method, but the difference is not as great as between the results for Regression 4. These differences are cause by the negatively skewed distributions of the input variables used in the Monte Carlo Simulation compared to the normal distributions assumed in the statistical analysis.

As discussed earlier, the mean for both probabilistic forecasts fluctuates, while the mean for both statistical forecasts show a steady increase. A reason for this is that Monte Carlo Simulation accounts for the random changes in temperature from year to year, while regression analysis does not.

Preference theories involving certainty equivalence could have been used to evaluate the results, as shown by the probability distribution of energy demand. Certainty equivalence identifies the risk preference of the decision-maker given two different outcomes.   The reasons that certainty equivalence is not used are: the results are not significant, and no information exists on the risk preference of the user of results in the data.

7.0   Summary and Conclusions

The summary and conclusions section discusses the results of the study, the implications of those results, the studies that could be conducted in the future, and the limitations of the study.

7.1   Summary

Regression analysis and Monte Carlo Simulation were compared in the development of a summer peak residential energy demand model for Appalachian Power Company's service area in West Virginia. Regression analysis is used to develop and test the validity of a summer peak forecasting model and determine if a more precise model could be developed. Risk is incorporated into the analysis, but the statistical output needs to be analyzed to determine the inherent risk. The regression models that had the most statistically significant and logical results were then used for the Monte Carlo Simulation. A commercial risk analysis software was used to conduct the Monte Carlo Simulation and to produce a range of results given the probabilistic distribution of inputs. Both regression analysis and Monte Carlo Simulation use the concept of risk in the analysis, but do so in a different manner.

During the 1980's the use of air conditioning increased in the West Virginia residential sector, and the demand for electricity in the summer kept pace. Therefore, a summer peak residential energy demand model is needed to forecast growing demand.

A summer peak demand model developed for Appalachian Power Company's service area in southern West Virginia was constructed. The model considered the following variables: maximum temperature on the day of Appalachian Power Company's summer peak in West Virginia, yearly average electricity price for Appalachian Power Company's service area in West Virginia, per capita income for Appalachian Power Company's service area in West Virginia, the total number of residential customers in Appalachian Power Company's Service area, the total

93

population in Appalachian Power Company's service area in West Virginia, natural gas price in West Virginia, cooling degree days on the day of Appalachian Power Company's summer peak in West Virginia, the percentage of rural population in Appalachian Power Company's service area in West Virginia, the average age in Appalachian Power Company's service area in West Virginia, and the total number of households in Appalachian Power Company's service area in West Virginia. The data were analyzed by regression analysis to determine what impact each variable has upon electricity demand. The identified regression equation was then used to do a forecast. The results of the regression analysis were then used to formulate the inputs for a forecast based on Monte Carlo Simulation.

An analysis of the statistical output shows that multicollinearity exists among the variables, which results in unreliable forecasts. The Monte Carlo Simulation confirms the problems identified with the statistical analysis. The forecast resulted in negative minimum demand values and high values so high to be unrealistic. Therefore, the probabilistic forecast is also unreliable.

The results of the probabilistic forecast can be computed by the statistical method using the calculated t-statistics and the standard errors, but the probabilistic method produces a visual picture of the results. A forecast developed using regression analysis and Monte Carlo Simulation shows that the statistical method produces a forecast of steadily increasing demand, while Monte Carlo Simulation produces a forecast that varies throughout the forecast time period. This variation is the result of the statistical analysis not adapting to changing conditions, such as temperature fluctuations, while the Monte Carlo Simulation does adapt. Neither method produced results that were acceptable. Therefore, both methods produce unacceptable results when the data are unreliable.

7.2 Implications of the Analysis

Although neither forecasting method produced reliable forecasts of electricity demand, the exercise did result in a number of implications. The most obvious implication is the need for adequate data. The data set must be large enough so that the past can accurately reflect the future. Without a sufficient data set, problems arise such as the ones discussed in the Results Section. Without an adequate data set, electricity demand may be over or underestimated.

Both regression analysis and Monte Carlo Simulation indicate that there are problems with the data. The regression analysis indicates that multicollinearity exists, while the Monte Carlo Simulation produces minimum and maximum values that are unrealistic. Also, the standard error of the forecast calculated using regression analysis can be used to calculate the distribution about the forecasted values as Monte Carlo Simulation does.

When the results of the forecast are not logical, it may be advantageous to use subjective inputs. Subjective inputs rely upon the interpretation of the trend of the data to forecast future values. Therefore, when the results of the resulting confidence intervals are not logical, it is necessary to revert back to subjective methods of trend extrapolation.

Regression analysis and Monte Carlo Simulation both incorporate risk into the analysis, but in different ways. Monte Carlo Simulation gives a visual representation of risk in the form of the probability distribution graph, while with regression analysis the statistical results need to be analyzed to determine the risk.

7.3 Limitations and Further Research

The most glaring limitation of the study is the small sample size. A small sample size does not give much of an indication into the past as to what will happen in the future. Thus, the small sample size in this study does not give a great historical indication of what past demand

was and the impact the factors affecting demand had upon it. Another problem with the small sample size is that multicollinearity exists. When multicollinearity exists it has the effect of making the effective sample size even smaller. Thus, with a small sample size and multicollinearity co-existing the sample size is even smaller which eliminates the ability to accurately predict the future from the past. A way to solve this problem would be to include data from the 1970's but use a dummy variable to account for the data from the abnormal energy market conditions in the 1970's. Another way to solve this problem would be to use a simple extrapolation.

An annual event limitation exists. The summer peak occurs only once every summer. This annual summer peak limits the rest of the data obtained because it should be on a daily basis as well. The problem arises from the fact that it is difficult to obtain data this precise. Data are usually available on only a monthly, quarterly, or yearly basis, but rarely on a daily basis.

Regional limitations on the data produce problems. The study is only for Appalachian Power Company's service area in southern West Virginia. Therefore, the data should be for this area only, but most data are unavailable on such a specific level. Data are usually available on only a state or national-level which is not precise enough for the forecast.

The structural change limitation is another limitation placed upon the data. The data begins with the 1980's, because that is when residential air conditioning use became wide spread. Since the study is for summer peak residential energy demand and the consumers did not begin using large levels of energy during the summer until the use for residential air conditioning increased, the study did not go beyond this time period. Another reason the data begins with the 1980's is that the oil crisis existed in the 1970's. Because the energy industry today is different than the one that existed in the 1970's, the data does not begin until the 1980's. Beginning with

the 1970's would skew the results. However, a dummy variable could be used to differentiate data from the 1970's and data from the 1980's to correct for these problems.

A future study could correct some of the problems in the forecast and perhaps develop a more reliable summer peak model for the state of West Virginia. Although, the types of variables impacting demand would likely be similar, some variables may differ. For example, a variable representing total population variable rather than one representing the number of residential customers variable might be used. The impact of deregulation upon electricity demand in West Virginia could be included in another study.

One concept to consider for a future study is the impact age has upon residential demand. West Virginia is a state with an aging population, and younger and older consumers use electricity differently. The elderly do not use as much electricity as younger individuals and with an aging population in West Virginia the demands placed upon the system peak may not be as great as expected. A study that addresses this question could provide answers that might help the elderly save money on their electricity bill by instituting conservation measures. It would also identify the impact a large elderly population has upon a region's energy demand.

An interesting idea would be to study the impact school age children have upon summer peak demand. Children are home from school all day during the summer. This means that electrical appliances such as televisions, computers, and radios are on all day, and rather than turning the air conditioning down when someone is not home, as may people do, the air conditioning would have to remain on at a high-level to remain comfortable for the children. Therefore, with children are home during the summer for the hottest portion of the day the residential summer peak may be affected.

Another study could determine what impact the sharp rise in per capita income in West Virginia has upon summer peak electricity demand (see Figure 17 in Appendix A). Normally the higher the income the greater the demand for energy. Therefore, it would be interesting to see what impact this increase in income in West Virginia has upon energy demand.

A future study could analyze the impact of the migration from the urban to rural areas that is taking place in West Virginia. As Figure 20 in Appendix A shows West Virginia's rural population is decreasing in Appalachian Power Company's service area. Therefore, it would be interesting to study what impact this has upon energy demand in West Virginia. Rural consumers tend to use more electricity than urban consumers do because rural consumers tend to live in homes while more urban consumers live in apartments and other multi-unit structures. Therefore, studying the impact of the migration from rural to urban may help forecasters determine what can be done to meet the changing needs of West Virginia's consumers.

List of References

Ackerman, Gary. (1985). Short-Term Load Prediction for Electric-Utility Control of Generating Units. In D.W. Bunn and E.D. Farmer (Eds.), <u>Comparative Models for Electrical Load Forecasting</u>. (pp. 33-42). New York: John Wiley & Sons Ltd..

Al-Alawi, Saleh M & Syed M. Islam. (1996). Principles of Electricity Demand Forecasting: Part 1 Methodologies. <u>Power Engineering Journal,</u>10 (3), 139-143.

Al-Garni, Ahmed Z., Syed M. Zubair & Javeed S. Nizami. (1994). A Regression Model for Electric-Energy-Consumption Forecasting in Eastern Saudi Arabia. <u>Energy, 19</u> (10), 1043-1049.

Anderson, David R., Dennis J. Sweeney, & Thomas A. Williams. (1978). <u>Essentials of Management Science: Applications to Decision Making</u>. St. Paul, Minn: West Publishing Co.

Bartels , Robert & Denzil G. Fiebig. (1996). Metering and Modelling Residential End-Use Electricity Load Curves. <u>Journal of Forecasting, 15,</u> 415-426.

Baum, Vladimir. (1993). Glowing in the Gloom. <u>Petroleum Economist, 60,</u> 6-9.

Bureau of Business and Economic Research. Consumer Price Index-All Urban Consumers, West Virginia, (1980-1995), wvbeis.be.wvu.edu/public/cat/cpi80p.txt.

Burns, Thomas G. (1984). The Art and Science of Energy Forecasting. <u>Journal of Petroleum Technology, 36,</u> 1437-1442.

Charles River Associates. (1986). A <u>Guide to Electricity Forecasting Methodology</u>. Washington D.C.: Edison Electric Institute.

Cowing, Thomas G. & Daniel L. McFadden. (1984). <u>Microeconomic Modeling and Policy Analysis: Studies in Residential Energy Demand</u>. Orlando, Fla.: Academic Press, Inc., Harcourt Brace, Jovanovich Publishers.

Crow, Robert T., Michael Robinson, and Raymond L. Squitieri. (1981). <u>Forecasting Electricity Sales and Loads: A Handbook for Small Utilities</u>.

Donnelly, William A. (1987). <u>The Econometrics of Energy Demand: A Survey of Applications</u>. New York: Praeger.

Granger, C.W.J. & Paul Newbold. (1986). <u>Forecasting Economic Time Series</u> (2nd ed.). Orlando, Fla.: Academic Press, Harcourt Brace Jonavich Publishers.

Griffiths, William E., R. Carter Hill, & George G. Judge. (1993). <u>Learning and Practicing Econometrics.</u> New York: John Wiley & Sons, Inc.

Gupta, P.C. (1985). Adaptive Short-term Forecasting of Hourly Loads Using Weather Information. In D.W. Bunn and E.D. Farmer, Comparative Models for Electrical Load Forecasting. (pp. 43-56) New York: John Wiley & Sons Ltd.

Halvorsen, Robert. (1970). Demand for Electric Energy in the United States. Southern Economic Journal, 42, 610-625.

Halvorsen, Robert. (1978). Econometric Models of U.S. Energy Demand. Lexington, Mass.: Lexington Books, D.C. Heath and Company.

Harvey, Andrew C. (1983). Time Series Models (2nd ed.). Cambridge, Mass.: The MIT Press.

Houthakker, Hendrikk S. (1980). Residential Electricity Revisited. The Energy Journal, 1 (1), 29-41.

Howard, Ronald A. (1968). The Foundations of Decision Analysis. IEEE Transactions on Systems Science and Cybernetics, 4 (3), 211-219.

Kher, Lov Kumar, Fereidoon P. Sioshansi, & Soroosh Sorooshian. (1987). Energy Demand Modeling with Noisy Input-Output Variables. The Energy Journal, 8 (4), 33-49.

Koomey, Jonathan G., Richard E. Brown, Robert Richey, Francis X. Johnson, Alan H. Sanstad, & Leslie Shown. (1995). Residential Sector End-Use Forecasting with EPRI-REEPS 2.1: Summary Input Assumptions and Results. Energy Analysis Program, Energy and Environment Division, Ernest Orlando Lawrence Berkley National Laboratory, University of California, Berkeley, CA.

Labys, Walter C. (1999). Modeling Mineral and Energy Markets. Boston, Mass.: Kluwer Academic Publishers.

Lakshminarayana, G.S. (1991). Importance of Long-Range Energy Demand Forecasting in Power System Planning. Irrigation and Power, 48 (1), 29-37.

Le, S.V. (1983). An Econometric Approach to Peak Load Forecasting. In S. Nakamura (Ed.), Proceedings of Electric Load Forecasting and Generating Capacity Expansion. Columbus, Ohio: National Regulatory Research Institute.

Leung, Ping Sun, and Walter Miklius. (1994). Accuracy of Electric Power Consumption Forecasts Generated by Alternative Methods: The Case of Hawaii. Energy Sources, 16, 289-299.

Megill, Robert E. (1984). An Introduction to Risk Analysis (2nd ed.). Tulsa, Okla.: PennWell Publishing Company.

Miller, Roger LeRoy. (1994). Economics Today: The Macro View (8th ed.). New York: Harper Collins College Publishers.

Mitchell, Bridger M., Rolla Edward Park, & Francis Labrune. (1986). <u>Projecting the Demand for Electricity: A Survey and Forecast.</u> Santa Monica, Calif.: Rand.

Munley, Vincent G., Larry W. Taylor, & John P. Formby. (1990). Electricity Demand in Multi-Family, Renter-Occupied Residences. <u>Southern Economic Journal, 57</u> (1), 178-194.

Nan, Gehuang D. & Donald A. Murry. (1992). Energy Demand with the Flexible Double-Logarithmic Functional Form. <u>The Energy Journal, 13</u> (4), 149-159.

Palisade Corporation. <u>Risk Analysis and Simulation Add-In for Microsoft Excel or Lotus 1-2-3.</u> (1996). Windows version, Newfield, Release 3.0 User's Guide. NY: Palisade Corporation.

Plourde, Andre & David Ryan. (1985). On the Use of Double-Log Forms in Energy Demand Analysis. <u>The Energy Journal, 6</u> (4), 105-113.

Quantitative Micro Software. (1995). <u>E-Views Manual</u>. Irvine, Calif.: Quantitative Micro Software.

<u>REIS-CD ROM</u>, 1969-1995.

Stermole, Franklin J. & John M. Stermole. (1993) Uncertainty and Risk Analysis. <u>Economic Evaluation and Investment Decision Methods</u>. Golden, Colo: Investment Evaluations Corporation.

Studenmund, A.H. with the assistance of Henry J. Cassidy. (1992). <u>Using Econometrics: A Practical Guide</u>. New York: Harper Collins Publishers.

Sugiyama, Sam O. (1992.) <u>Simple Model Representations of Detailed Load Forecasting Models</u>. Proceedings: Eighth Electric Utility Forecasting Symposium, Baltimore, Maryland.

Torries, Thomas F. (1998). <u>Evaluating Mineral Projects: Applications and Misconceptions</u>. Littleton, Colo.: Society for Mining, Metallurgy, and Exploration, Inc.

U.S. Department of Commerce, Economics and Statistics Administration, Bureau of the Census. (1982). <u>1980 Census of Population, General Population Characteristics, West Virginia</u>. Washington, DC: U.S. Government Printing Office.

U.S. Department of Commerce, Economics and Statistics Administration, Bureau of the Census. (1983). <u>1980 Census of the Population, Characteristics of the Population, General Social and Economic Characteristics, West Virginia</u>. Washington, DC: U.S. Government Printing Office.

U.S. Department of Commerce, Economics and Statistics Administration, Bureau of the Census. (1993). <u>1990 Census of Population, Social and Economic Characteristics, West Virginia</u>. Washington, DC: U.S. Government Printing Office.

U.S. Department of Commerce, Economics and Statistics Administration, Bureau of the Census. (1982). <u>1980 Census of Population and Housing, Summary Characteristics for Government Units and Standard Metropolitan Statistical Areas, West Virginia</u>. Washington, DC: U.S. Government Printing Office.

U.S. Department of Commerce, Economics and Statistics Administration, Bureau of the Census. (1991). <u>1990 Census of Population and Housing, Summary Population and Housing Characteristics, West Virginia</u>. Washington, DC: U.S. Government Printing Office.

U.S. Department of Commerce, Bureau of the Census. (1992). <u>1990 Census of Population and Housing Summary Tape File 3A</u>.

U.S. Department of Commerce, National Oceanic and Atmospheric Administration. (1982). <u>West Virginia Climatological Data</u>. (vol. 90, no. 7). Asheville, North Carolina: National Climatic Data Center.

U.S. Department of Commerce, National Oceanic and Atmospheric Administration. (1983). <u>West Virginia Climatological Data</u>. (vol. 91, no. 8). Asheville, North Carolina: National Climatic Data Center.

U.S. Department of Commerce, National Oceanic and Atmospheric Administration. (1984). <u>West Virginia Climatological Data</u>. (vol. 92, no. 8). Asheville, North Carolina: National Climatic Data Center.

U.S. Department of Commerce, National Oceanic and Atmospheric Administration. (1985). <u>West Virginia Climatological Data</u>. (vol. 93, no. 8). Asheville, North Carolina: National Climatic Data Center.

U.S. Department of Commerce, National Oceanic and Atmospheric Administration. (1986). <u>West Virginia Climatological Data</u>. (vol. 94, no. 7). Asheville, North Carolina: National Climatic Data Center.

U.S. Department of Commerce, National Oceanic and Atmospheric Administration. (1987). <u>West Virginia Climatological Data</u>. (vol. 95, no. 7). Asheville, North Carolina: National Climatic Data Center.

U.S. Department of Commerce, National Oceanic and Atmospheric Administration. (1988). <u>West Virginia Climatological Data</u>. (vol. 96, no. 8 Asheville, North Carolina: National Climatic Data Center.

U.S. Department of Commerce, National Oceanic and Atmospheric Administration. (1989).
West Virginia Climatological Data. (vol. 97, no. 7). Asheville, North Carolina: National
Climatic Data Center.

U.S. Department of Commerce, National Oceanic and Atmospheric Administration. (1990).
West Virginia Climatological Data. (vol. 98, no. 7). Asheville, North Carolina: National
Climatic Data Center.

U.S. Department of Commerce, National Oceanic and Atmospheric Administration. (1991).
West Virginia Climatological Data. (vol. 99, no. 7). Asheville, North Carolina: National
Climatic Data Center.

U.S. Department of Commerce, National Oceanic and Atmospheric Administration. (1992).
West Virginia Climatological Data. (vol. 100, no. 7). Washington, DC: U.S.
Government Printing Office.

U.S. Department of Commerce, National Oceanic and Atmospheric Administration. (1993).
West Virginia Climatological Data. (vol. 101, no. 8). Asheville, North Carolina:
National Climatic Data Center.

U.S. Department of Commerce, National Oceanic and Atmospheric Administration. (1994).
West Virginia Climatological Data. (vol. 102, no. 6). Asheville, North Carolina:
National Climatic Data Center.

U.S. Department of Commerce, National Oceanic and Atmospheric Administration. (1995).
West Virginia Climatological Data. (vol. 103, no. 8). Asheville, North Carolina:
National Climatic Data Center.

U.S. Department of Energy, Energy Information Administration. (1988-1995). Annual Electric
Utility Report. Electric Sales and Revenue.

U.S. Department of Energy, Energy Information Administration. Historical Gas Annual 1930-
1996. (DOE/EIA-0131(96)). www.eia.doe.gov.

U.S. Department of Labor, Bureau of Labor Statistics. Consumer Price Index, All Urban
Consumers, U.S. City Average Electricity, Base Period, 1982-84=100,
http://stats.bls.gov/.

U.S. Department of Labor, Bureau of Labor Statistics. Consumer Price Index, All Urban
Consumers, U.S. City Average Utility Gas Service, Base Period, 1982-84=100,
http://stats.bls.gov/.

Walls, Michael R., and Thomas F. Torries. (1998). A Review of Risk Tolerance and Certainty
Equivalence. Evaluating Mineral Projects: Applications and Misconceptions (pp. 123-
130). Littleton, Colo.: Society for Mining, Metallurgy, and Exploration, Inc.

West Virginia Public Service Commission. (1982-1996). <u>Appalachian Power Company's Annual Statistical Report</u>.

Other Sources

Anderson, Kent P. (1973). <u>Residential Energy Use: An Econometric Analysis</u>. Prepared for the National Science Foundation, R-1297-NSF, Santa Monica, CA: Rand.

Arsenault, E., J-T. Bernard, C.W. Carr, & E. Genest-Laplante. (1995). A Total Energy Demand Model of Quebec. <u>Energy Economics, 17</u> (2), 163-171.

Bentzen, Jan and Tom Engsted. (1993). Short-and Long-Run Elasticities in Energy Demand: A Cointegration Approach. <u>Energy Economics, 15,</u> 9-16.

Bernard, Jean-Thomas, Denis Bolduc, & Donald Be`langer. (1996). Quebec Residential Energy Demand: A Microeconomic Approach. <u>Canadian Journal of Economics, 29,</u> 92-113.

Bodger, P.S. (1987). Logistic and Energy Substitution Models for Electricity Forecasting: A Comparison Using New Zealand Consumption Data. <u>Technological Forecasting and Social Change, 31,</u> 27-48.

Bunn, D.W. & E.D. Farmer. (1985a). Economic and Operational Context of Electric Load Prediction. In D.W. Bunn and E.D. Farmer (eds.) <u>Comparative Models for Electrical Load Forecasting</u> (p.p. 3-11). New York: John Wiley & Sons Ltd.

Bunn, D.W. & E.D. Farmer. (1985b). Review of Short-term Forecasting Methods in the Electric Power Industry. In D.W. Bunn and E.D. Farmer (eds.) <u>Comparative Models for Electrical Load Forecasting</u> (pp. 13-30). New York: John Wiley & Sons Ltd.

Canova, Fabio, and Bruce E. Hansen. (1995). Are Seasonal Patterns Constant Over Time? A Test for Seasonal Stability. <u>Journal of Business & Economic Statistics, 13</u> (3), 237-253.

Chan, Hing Lin and Shu Kam Lee. (1996). Forecasting the Demand for Energy in China. <u>The Energy Journal, 17</u> (1), 19-30.

Cromwell, Jeff B., Walter C. Labys, & Michel Terraza. (1994). <u>Univariate Tests for Time Series Models</u>. (Michael S. Lewis-Beck, eds.), Sara Miller McCune, A SAGE University Paper.

Enders, Walter. (1995). <u>Applied Econometric Time Series</u>. New York: John Wiley & Sons, Inc.

Engle, Robert. F., Chowdhury Mustafa, & John Rice. (1992). Modelling Peak Electricity Demand. <u>Journal of Forecasting, 11,</u> 241-252.

Halvorsen, Robert. (1975). Residential Demand for Electric Energy. <u>The Review of Economics & Statistics, 57,</u> 12-18.

Hawkins, R.G. (1975). The Demand for Electricity: A Cross-Section Study of New South Wales and the Australian Capital Territory. The Economic Record, 51, 1-18.

U.S. Department of Energy, Energy Information Administration. Historical Monthly Energy Review 1973-1992. www.eia.doe.gov

Irisarri, G.D., S.E. Widergren, & P.D. Yehsakul. (1982). On-line Load Forecasting for Energy Control Center Application. IEEE Transactions on Power Apparatus and Systems, 1 (101) 71-78.

Irish, William F. (1992). Using Multiple Models in Producing Long-term Energy Forecasts. Proceedings: Eighth Electric Utility Forecasting Symposium, Baltimore, Maryland.

Kahn, Edward. (1991). Electric Utility Planning and Deregulation. American Council for an Energy Efficient Economy, Washington, D.C. and Berkeley, CA.

Kennedy, Peter. (1998). A Guide to Econometrics (4th ed.). Cambridge, Mass: The MIT Press.

Kras, Michael. (1987). Energy Forecasting: The Epistemological Context. Futures: The Journal of Forecasting and Planning, 19 (3), 254-275.

Labys, Walter C. (1982). Measuring the Validity and Performance of Energy Models. Energy Economics, 4, 159-167.

Lee, Ray-Shine & Nirvikar Singh. (1994). Patterns in Residential Gas and Electricity Consumption: An Econometric Analysis. Journal of Business & Economic Statistics, 12 (2), 233-241.

Mitchell, Bridger M., Judith Wilson Ross, & Rolla Edward Park. (1986). A Short Guide to Electric Load Forecasting. Santa Monica, Calif: Rand.

Mountain, Dean C & Evelyn Lawson. (1992). A Disaggregated Nonhomothetic Modeling of Responsiveness to Residential Time-of-Use Electricity Rates. International Economic Review, 33 (1), 181-207.

Murphy, Frederic H. & Allen L. Soyster. (1983). Economic Behavior of Electric Utilities. Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

Pierce, William Spangar. (1996). Economics of the Energy Industry (2nd ed.). West Port, Connecticut: Praegar Publishers

Ramanathan, Ramu, Clive Granger, & Robert Engle. (1985). Two Step Modelling for Short-Term Forecasting. In D.W. Bunn and E.D. Farmer (eds.) Comparative Models for Electrical Load Forecasting. (pp. 131-157). New York: John Wiley & Sons.

Regulation Energy. http://www.kemijoki.fi/saatoe.htm#SAATO.

Ryan, David L., Yu Wang, & Andre` Plourde. (Special Issue, April 1996). Assymmetric Price Responses of Residential Energy Demand in Ontario." Canadian Journal of Economics, XXIX, 317-323.

Sarma, U.K., S. Sinha, & T.K. Basu. (1987). Medium Range Forecasting of Power System Load (Energy) Demand. International Journal of Systems Science, 18 (9), 1691-1702.

Schneider, Alan M., Tadashi Takenawa, & Dean A. Schiffman. (1985). 24- Hour Electricity Utility Load Forecasting. In D.W. Bunn and E.D. Farmer (eds.) Comparative Models for Electrical Load Forecasting (pp. 87-108). New York: John Wiley & Sons Ltd.

Stanton, K. Neil. and Pradeep C. Gupta. (1970). Forecasting Annual or Seasonal Peak Demand in Electric Utility Systems. IEEE Transactions on Power Apparatus and Systems, 89 (5/6), 951-959.

Sun, J.W. (1996). Real Rural Residential Energy Consumption in China, 1990. Energy Policy, 24 (9), 827-839.

Taylor, Lester D. (1975). The Demand for Electricity: A Survey. Bell Journal of Economics & Management Science, 6, 74-109.

Tuan, Nguyen Anh & Thierry Lefevre. Analysis of Household Energy Demand in Vietnam. Energy Policy, 24 (12), 1089-1099.

Uri, Noel, D. (1979). A Mixed Time-Series/Econometric Approach to Forecasting Peak System Load. Journal of Econometrics, 9, 155-171.

Uri, Noel. D. & Saad A. Hassanein. (1986). A Note on Modeling the High-Frequency Demand for Energy. Energy Sources, 8 (2/3), 87-98.

Yamasaki, Eiji & Norio Tominaga. (1997). Evolution of an Aging Society and Effect on Residential Energy Demand. Energy Policy, 25 11, 903-912.

Appendix A: The Data

Table 12: The Data

Year	Appalachian Power Company's West Virginia Residential Summer Peak (mWh)[1]	Trend	Appalachian Power Company's Residential Electricity Price ($/mWh) for West Virginia (1990 Constant dollars)[2]	Maximum Temperature (Degrees F) on Day of Appalachian Power Company's West Virginia Summer Residential Peak[3]	Per Capita Income (1990 Constant $) for Appalachian Power Company's West Virginia Service Area[4]
1982	1786	1	0.000637	92	12849
1983	1969	2	0.000641	98	12512
1984	1878	3	0.000664	91	12906
1985	1887	4	0.000641	96	13057
1986	1792	5	0.000606	92	13397
1987	1828	6	0.000586	95	13358
1988	2151	7	0.000580	99	13663
1989	2017	8	0.000563	91	13879
1990	2072	9	0.000550	94	15013
1991	2129	10	0.000571	98	15619
1992	2072	11	0.000592	92	17047
1993	2268	12	0.000625	98	18188
1994	2138	13	0.000620	93	19443
1995	2274	14	0.000630	92	20998

1. Source: Appalachian Power Company

2. Source: 1983-1997, Appalachian Power Company's Annual Statistical Report to the West Virginia Public Service Commission
    1988-1995, EIA-861 ("Annual Electric Utility Report") published in Electric Sales and Revenue

3. Source: West Virginia Climatological Journal, 1982-1995

4. Source: REIS-CD ROM 1969-1995

Table 13: The Data Continued

Year	Total Population for Appalachian Power Company's West Virginia Service Area[5]	Appalachian Power Company's Residential Customers in West Virginia[6]	% Rural Population for Appalachian Power Company's Service Area in West Virginia[7]	Total Households for Appalachian Power Company's West Virginia Service Area[8]
1982	1013534	359720	0.63	375420
1983	1064309	361430	0.62	372147
1984	998333	362437	0.62	368875
1985	985439	362868	0.62	365602
1986	970674	363288	0.62	362329
1987	953193	363338	0.62	359057
1988	932400	363820	0.63	355784
1989	915416	366183	0.63	352512
1990	904834	368054	0.63	349239
1991	906332	369625	0.61	345966
1992	909145	371818	0.60	342694
1993	913738	374874	0.59	339421
1994	913568	377225	0.58	336149
1995	913775	381678	0.57	332876

5. Source: <u>1990 Census of Population and Housing Summary Tape File 3A</u> CD ROM

6. Source: <u>Appalachian Power Company's Annual Statistical Report</u> to the West Virginia Public Service Commission, 1982-1995.

7. Source: <u>1980 Census of Population, General Population Characteristics, West Virginia</u>, and the <u>1990 Census of Population and Housing Summary Tape File 3A</u> CD ROM

8. Source: <u>1980 Census of Population and Housing, Summary Characteristics for Government Units and Standard Metropolitan Statistical Areas for West Virginia</u>, and the <u>1990 Census of Population and Housing, Summary Population and Housing Characteristics for West Virginia</u>

Table 14: The Data Continued

Year	Residential Natural Gas Price ($/Btu) in West Virginia (1990 Constant $)[9]	Cooling Degree-Day on for Appalachian Power Company's West Virginia Residential Summer Peak[10]	Average Age for Appalachian Power Company's West Virginia Service Area[11]
1982	5.64	15.0	30.84
1983	5.57	20.0	31.40
1984	5.64	16.5	31.95
1985	5.93	17.5	32.51
1986	6.19	17.5	33.07
1987	6.12	17.0	33.62
1988	5.66	20.0	34.18
1989	5.76	14.5	34.73
1990	6.46	18.0	35.29
1991	6.58	19.0	35.85
1992	6.50	19.5	36.40
1993	7.06	18.0	36.96
1994	7.43	15.0	37.51
1995	7.46	17.0	38.07

9. Source: Energy Information Administration Homepage, http://www.eia.doe.gov/

10. Source: West Virginia Climatological Journal, 1982-1995

11. Source: 1980 Census of Population, General Social and Economic Characteristics for West Virginia, and the 1990 Census of Population, Social and Economic Characteristics for West Virginia

Table 15:

Electricity Price and Natural Gas Price Index Percentages (Base year = 1990)		
	Electricity	Natural Gas
1982	81.64%	92.23%
1983	84.29%	107.64%
1984	89.65%	108.42%
1985	92.87%	107.70%
1986	94.06%	102.46%
1987	93.82%	97.78%
1988	94.89%	97.18%
1989	97.72%	99.86%
1990	100.00%	100.00%
1991	103.78%	101.25%
1992	105.79%	103.09%
1993	107.91%	109.41%
1994	107.96%	111.52%
1995	110.39%	105.79%
Source: Bureau of Labor Statistics Data Homepage		

Table 16

West Virginia Consumer Price Index Percentages for Income (Base year = 1990)	
1982	73.83%
1983	76.21%
1984	79.50%
1985	82.33%
1986	83.86%
1987	86.92%
1988	90.51%
1989	94.87%
1990	100.00%
1991	104.21%
1992	107.35%
1993	110.56%
1994	113.39%
1995	116.60%
Source: Bureau of Business & Economic Research homepage, wvbeis.be.wvu.edu/public/cat/cpi80p.txt	
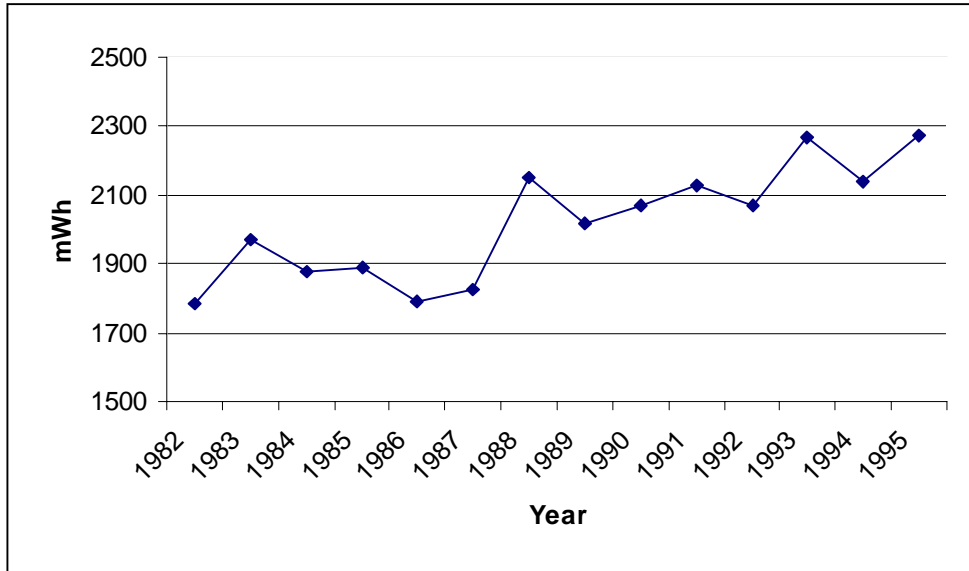
(Source: Table 12)

**Figure 12: Appalachian Power Company's
West Virginia Residential Summer Peak (mWh)**
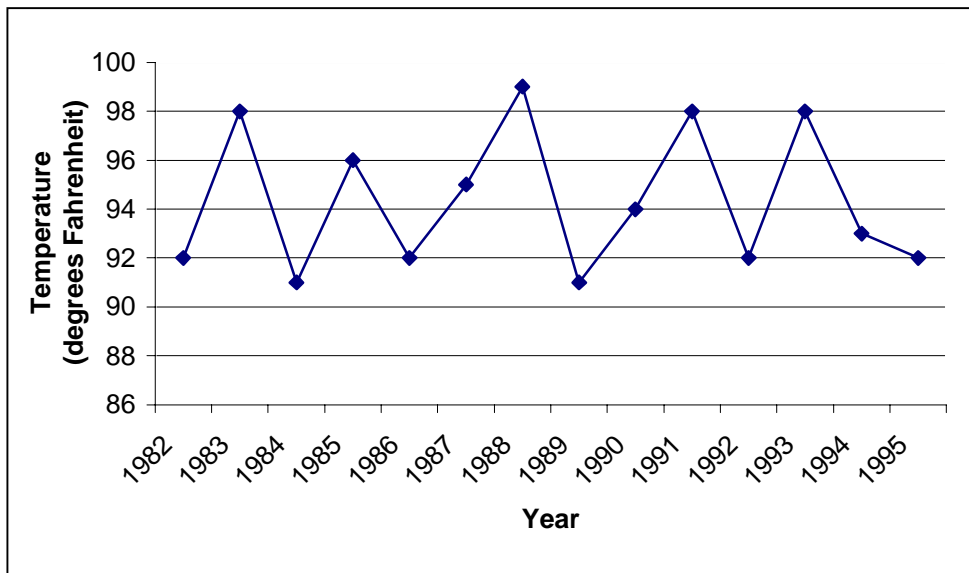
(Source: Table 12)

**Figure 13: Maximum Temperature (degrees Fahrenheit) on Day of
Appalachian Power Company's West Virginia Residential Summer Peak**
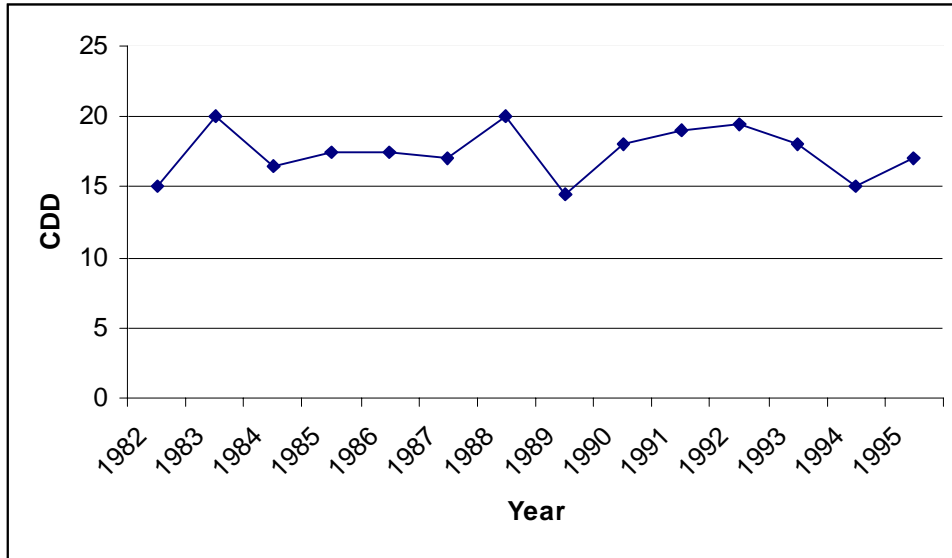
(Source: Table 14)

**Figure 14: Cooling Degree-Days on Day of Appalachian
Power Company's West Virginia Residential Summer Peak**
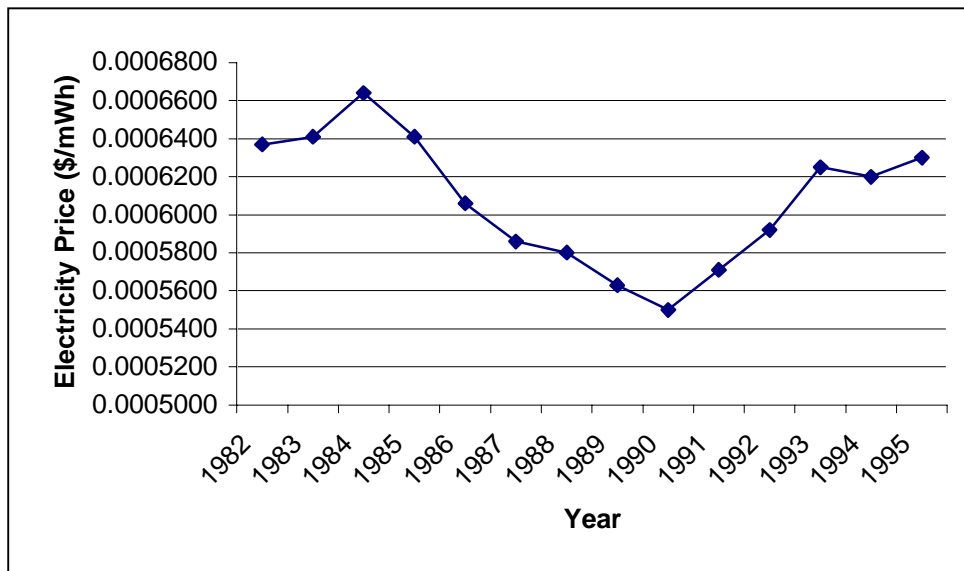
(Source: Table 12)

**Figure 15: Appalachian Power Company's Average Residential
Electricity Price ($/mWh) for West Virginia in 1990 Constant Dollars**
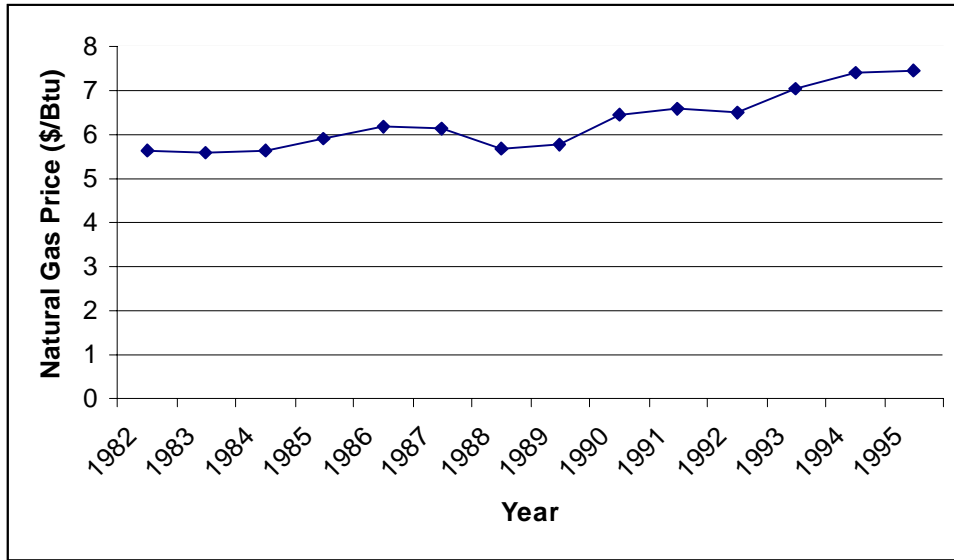
(Source: Table 14)

**Figure 16: Residential Natural Gas Price ($/Btu)
for West Virginia in 1990 Constant Dollars**
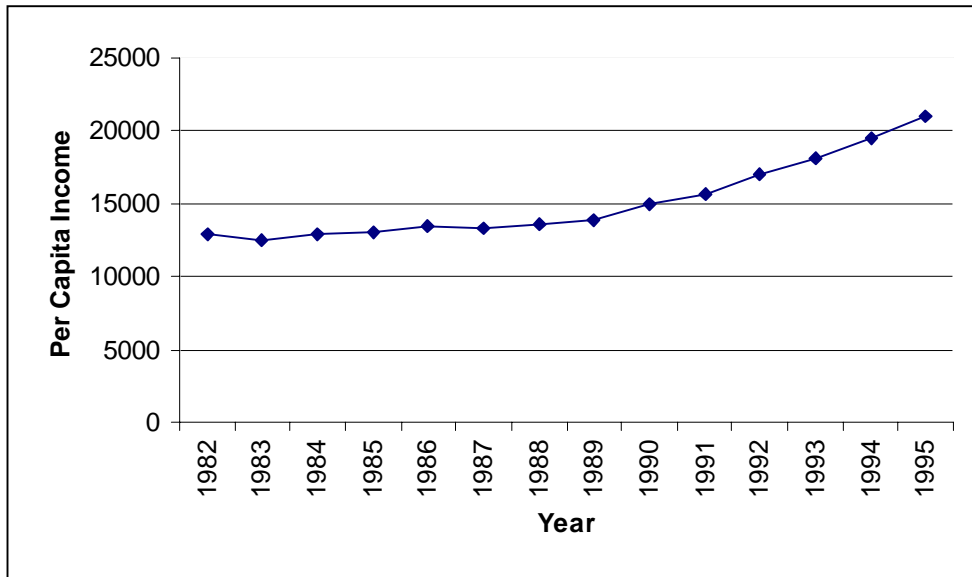
(Source: Table 12)

**Figure 17: Per Capita Income for Appalachian Power Company's
West Virginia Service Area in 1990 Constant Dollars**
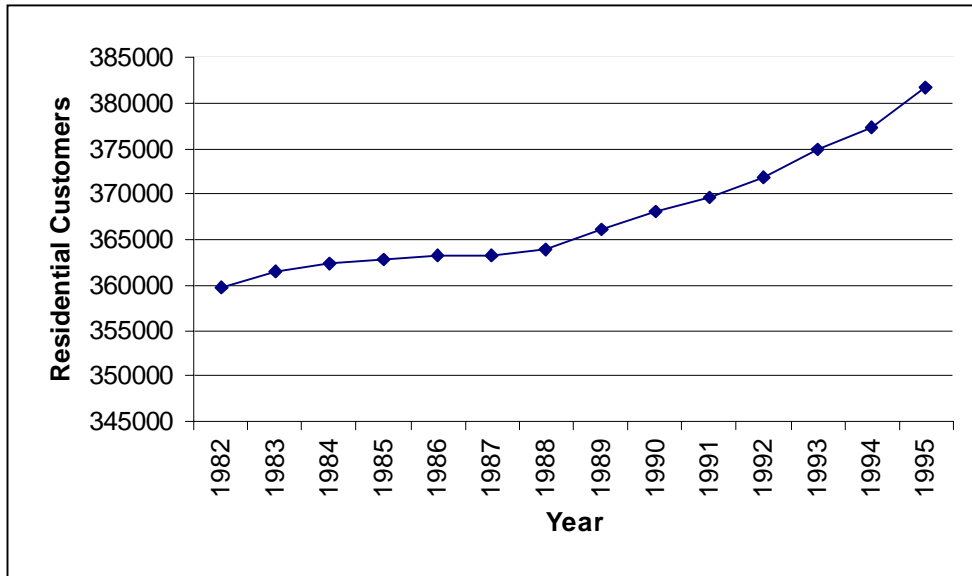
(Source: Table 13)

**Figure 18: Appalachian Power Company's
Residential Customers in West Virginia**
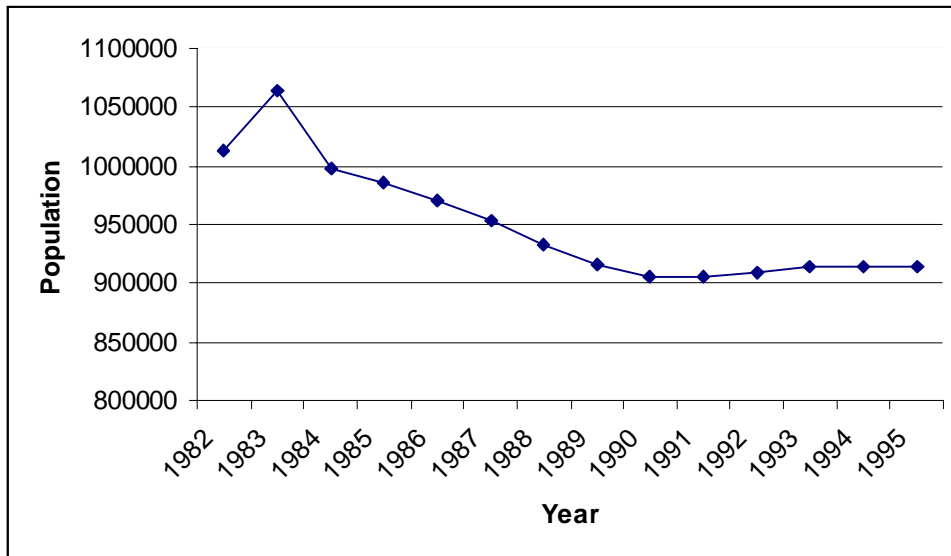
(Source: Table 13)

**Figure 19: Total Population for Appalachian
Power Company's Service Area in West Virginia**
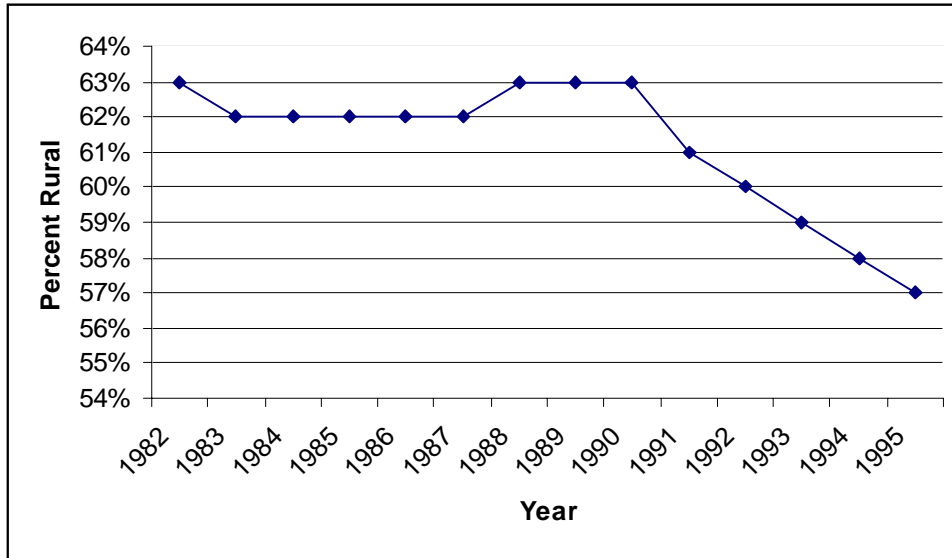
Figure 20: Percent Rural Population for Appalachian Power Company's Service Area in West Virginia
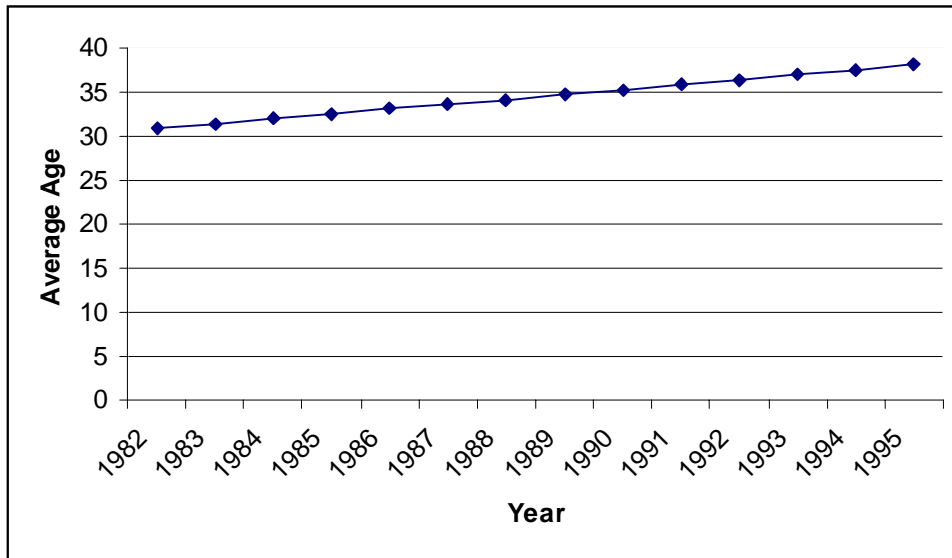
Figure 21: Average Age for Appalachian Power Company's Service Area in West Virginia
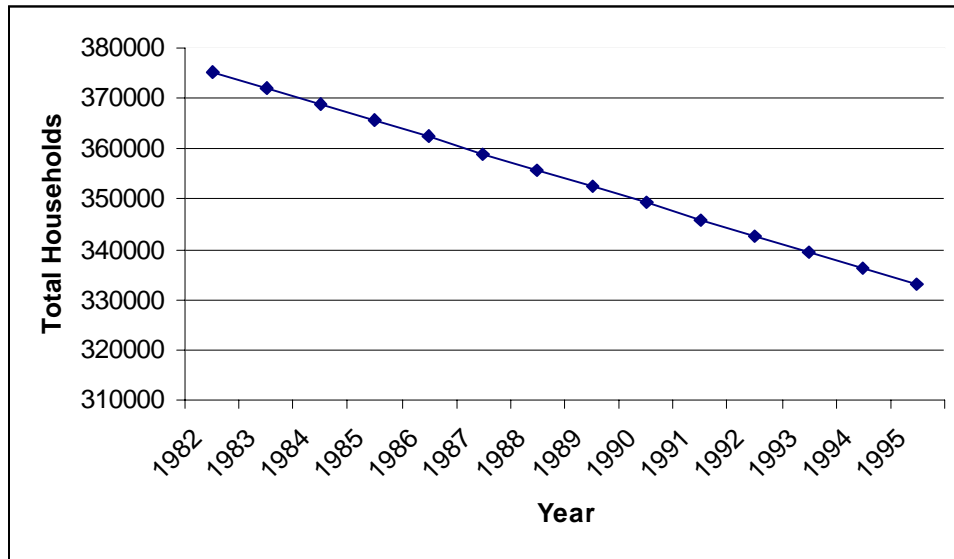
(Source: Table 13)

Figure 22: Total Households for Appalachian Power Company's Service Area in West Virginia

Appendix B: The Results

Table 17: Correlation Matrix

	Electricity Price	Temperature	Per Capita Income	Population	Residential Customers	Percent Rural	Total Households	Natural Gas Price	CDD	Average Age
Electricity Price	1									
Temperature	-0.117	1								
Per Capita Income	-0.012	-0.106	1							
Population	0.650	0.056	-0.662	1						
Residential Customers	-0.076	-0.088	0.991	-0.705	1					
Percent Rural	-0.281	0.079	-0.929	0.411	-0.909	1				
Total Households	0.352	0.003	-0.920	0.874	-0.948	0.771	1			
Natural Gas Price	-0.057	-0.061	0.951	-0.662	0.946	-0.898	-0.894	1		
CDD	-0.148	0.681	-0.083	0.050	-0.061	0.057	0.018	-0.100	1	
Average Age	-0.352	-0.002	0.920	-0.874	0.948	-0.771	-1.000	0.894	0.019	1

* Unexpected Sign

These results generally make sense.

Table 18: The Statistical Results

	Regression 5*	Regression 6	Regression 7	Regression 8
Intercept	-13596.727 (-1.093)	-160284.29 (-1.0849)	-16429.12 (-3.418)	-8548.768 (-1.032)
Trend	-2.789 (-0.075)	0.285 (0.007)	-4.66 (-0.218)	16.891 (0.436)
Electricity Price	-427499.045 (-0.302)	-43.94 (-0.306)	-634079.268 (-0.782)	52702.86 (0.036)
Temperature	21.839 (2.754)	21.839 (2.743)	23.952 (5.27)	
CDD				23.267 (1.734)
Residential Customers	0.0392 (1.058)		0.038 (2.07)	0.027 (0.795)
Log Residential Customers		29764.389 (1.069)		
Log Per Capita Income		-1231.639 (-0.532)	1071.813 (0.75)	365.64 (0.138)
Per Capita Income	-0.0392 (-0.586)			
Log Population				
Log Percent Rural				
Natural Gas Price			-267.193 (-4.224)	-208.494 (-1.758)
R-Squared	0.863	0.862	0.96	0.865
Adjusted R-Squared	0.777	0.776	0.93	0.749
Standard Error	78.35	78.607	44.607	83.18
F-Statistic	10.08	10.002	28.86	7.467

 * Regressions 1-4 are shown in Table 3

Table 19: The Statistical Results Continued

	Regression 9	Regression 10	Regression 11	Regression 12
Intercept	-48858.431	-16450.767	107594.776	-833962.565
	(-2.003)	(-3.423)	(2.117)	(-0.0004)
Trend	96.529	-4.559	675.304	-5145.176
	(2.248)	(-0.213)	(3.042)	(-0.0002)
Electricity Price	3759001.47	-623279.377	2432703.408	3382590.986
	(2.109)	(-0.769)	(1.581)	(1.962)
Temperature	15.745	23.971	19.003	
	(2.161)	(5.277)	(3.33)	
CDD				-0.614
				(-0.034)
Log Per Capita Income	2893.689	1066.407		2165.708
	(1.691)	(0.747)		(1.299)
Log Population	6739.089		2535.243	
	(1.685)		(0.771)	
Residential Customers		0.038		
		(2.078)		
Population				0.003
				(1.91)
Log Percent Rural	27975.75		3845.979	
	(2.529)		(1.429)	
Natural Gas Price		-268.803		-178.5
		(-4.229)		(-1.42)
Log Average Age			-81001.99	
			(-2.665)	
Average Age				15032.636
				(1.383)
Log Natural Gas Price			-2087.155	
			(-1.849)	
Percent Rural				13938.377
				(2.256)
Total Households				0.946
				(0.0002)
R-Squared	0.918	0.961	0.96	0.949
Adjusted R-Squared	0.847	0.928	0.91	0.835
Standard Error	64.939	44.575	49.46	67.436
F-Statistic	12.999	28.899	20.07	8.313

Table 20: The Statistical Results Continued

	Regression 13	Regression 14	Regression 15
Intercept	-286687.246 (0.0002)	41967.2 (0.271)	-189267.578 (-1.853)
Trend	-2069.341 (-0.0002)	391.283 (0.584)	-3.616 (-0.163)
Electricity Price	2599341.454 (2.326)	2248950.992 (1.322)	-617583.459 (-0.748)
Temperature	15.232 (2.514)	19.156 (3.122)	23.847 (5.17)
Log Per Capita Income	2938.818 (2.719)	1761.253 (0.454)	
Per Capita Income			0.019 (0.476)
Log Population		3119.25 (0.83)	
Log Residential Customers			34280.396 (1.855)
Population	0.002 (2.043)		
Log Percent Rural		14874.47 (1.397)	
Natural Gas Price	-178.54 (-2.532)		-260.106 (-4.12)
Log Average Age		-43562.22 (-0.491)	
Average Age	5429.793 (0.914)		
Log Natural Gas Price		-2448.503 (-1.688)	
Percent Rural	11457.082 (2.885)		
Total Households	0.265 (7.8657×10^{-5})		
R-Squared	0.98	0.96	0.96
Adjusted R-Squared	0.92	0.898	0.93
Standard Error	41.99	53.105	45.3
F-Statistic	22.14	15.263	27.946

Table 21: Auxiliary Regression with Appalachian Power Company's Residential Electricity Price for West Virginia as Dependent Variable

Regression Statistics

Multiple R	0.96
R Square	0.91
Adjusted R Square	0.86
Standard Error	0.00001
Observations	14

ANOVA

	Df	SS	MS	F	Significance F
Regression	5	1.4E-08	2.8E-09	16.89	0.0005
Residual	8	1.3E-09	1.7E-10		
Total	13	1.5E-08			

	Coefficients	Standard Error	t Stat	P-value
Intercept	0.008011	0.005720	1.400375	0.198971
Temperature	0.000001	0.000001	0.500949	0.629897
log percapinc	-0.000198	0.000324	-0.611855	0.557618
log pop	-0.000737	0.000772	-0.955539	0.367284
log age	-0.002134	0.000860	-2.481842	0.038002
log perrural	-0.004886	0.001411	-3.463475	0.008524

Table 22: Auxiliary Regression with Maximum Temperature on Day of Appalachian Power Company's West Virginia Residential Summer Peak as the Dependent Variable

Regression Statistics

Multiple R	0.54
R Square	0.29
Adjusted R Square	-0.15
Standard Error	3.14
Observations	14

ANOVA

	df	SS	MS	F	Significance F
Regression	5	32.45	6.49	0.66	0.66
Residual	8	78.76	9.84		
Total	13	111.21			

	Coefficients	Standard Error	t Stat	P-value
Intercept	-1886.59	1403.68	-1.34	0.22
Electricity Price	42460.51	84760.07	0.50	0.63
Log percapinc	-27.22	80.21	-0.34	0.74
Log pop	259.96	175.75	1.48	0.18
Log age	389.01	242.15	1.61	0.15
Log perrural	387.01	525.48	0.74	0.48

Table 23: Auxiliary Regression with Per Capita Income for Appalachian Power Company's West Virginia Service Area as the Dependent Variable

Regression Statistics	
Multiple R	0.99
R Square	0.98
Adjusted R Square	0.97
Standard Error	0.01
Observations	14

ANOVA

	df	SS	MS	F	Significance F
Regression	5	0.069	0.014	73.42	1.82E-06
Residual	8	0.002	2E-04		
Total	13	0.071			

	Coefficients	Standard Error	t Stat	P-value
Intercept	6.29	6.43	0.98	0.36
Electricity Price	-225.28	368.19	-0.61	0.56
Temperature	-0.001	0.002	-0.34	0.74
Log pop	-0.58	0.84	-0.68	0.51
Log age	0.39	1.21	0.32	0.76
Log perrural	-4.28	1.83	-2.33	0.05

Table 24: Auxiliary Regression with Total Population for Appalachian Power Company's West Virginia Service Area as the Dependent Variable

Regression Statistics	
Multiple R	0.98
R Square	0.96
Adjusted R Square	0.94
Standard Error	0.01
Observations	14

ANOVA

	Df	SS	MS	F	Significance F
Regression	5	0.0062	0.001237811	39.56	1.96E-05
Residual	8	0.0003	3.12856E-05		
Total	13	0.0064			

	Coefficients	Standard Error	t Stat	P-value
Intercept	7.73	0.45	17.30712973	1.26532E-07
Electricity Price	-138.91	145.38	-0.955538735	0.367
Temperature	0.00083	0.00	1.479129651	0.177
Log percapinc	-0.095	0.14	-0.681446532	0.515
Log age	-1.133	0.29	-3.867718586	0.005
Log perrural	-1.82	0.72	-2.52858909	0.035

Table 25: Auxiliary Regression with Percent Rural Population for Appalachian Power Company's West Virginia Service Area as the Dependent Variable

Regression Statistics	
Multiple R	0.99
R Square	0.99
Adjusted R Square	0.98
Standard Error	0.002
Observations	14

ANOVA

	df	SS	MS	F	Significance F
Regression	5	0.002	0.0004	106.04	4.33E-07
Residual	8	3.34E-05	4.17E-06		
Total	13	0.002			

	Coefficients	Standard Error	t Stat	P-value
Intercept	2.21	0.64	3.45	0.01
Electricity Price	-122.78	35.45	-3.46	0.01
Temperature	0.0002	0.0002	0.7365	0.4825
Log percapinc	-0.09	0.04	-2.33	0.05
Log pop	-0.24	0.10	-2.53	0.04
Log age	-0.34	0.14	-2.47	0.04

Table 26: Auxiliary Regression with Average Age for Appalachian Power Company's West Virginia Service Area as the Dependent Variable

Regression Statistics	
Multiple R	0.99
R Square	0.99
Adjusted R Square	0.98
Standard Error	0.004
Observations	14

ANOVA

	df	SS	MS	F	Significance F
Regression	5	0.01	0.002	140.06	1.45E-07
Residual	8	0.0001	1.59E-05		
Total	13	0.01			

	Coefficients	Standard Error	t Stat	P-value
Intercept	4.63	1.10	4.20	0.00
Electricity Price	-203.87	82.14	-2.48	0.04
Temperature	0.00	0.00	1.61	0.15
Log percapinc	0.03	0.10	0.32	0.76
Log pop	-0.57	0.15	-3.87	0.00
Log perrural	-1.28	0.52	-2.47	0.04

Table 27: Regression Equation used to Forecast Appalachian Power Company's Residential Electricity Price for West Virginia

Electricity price

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.352049897
R Square	0.12393913
Adjusted R Square	0.050934058
Standard Error	3.34817E-05
Observations	14

ANOVA

	df	SS	MS	F	Significance F
Regression	1	1.93E-09	1.90E-09	1.698	0.217
Residual	12	1.35E-08	1.12-09		
Total	13	1.54E-08			

	Coefficients	Standard Error	t Stat	P-value
Intercept	0.00063	1.89E-05	33.29	3.42E-13
Trend	-2.89E-06	2.22E-06	-1.30	0.22

Table 28: Regression Equation used to Forecast Per Capita Income for Appalachian Power Company's West Virginia Service Area

Per Capita Income
SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.92
R Square	0.85
Adjusted R Square	0.83
Standard Error	1118.49
Observations	14

ANOVA

	df	SS	MS	F	Significance F
Regression	1	82468517.8	82468517.8	65.92	3.23E-06
Residual	12	15012203.47	1251016.96		
Total	13	97480721.26			

	Coefficients	Standard Error	t Stat	P-value
Intercept	10622.31	631.41	16.82	1.04E-09
Trend	602.08	74.16	8.12	3.23E-06

Table 29: Regression Equation used to Forecast Log Residential Customers for Appalachian Power Company's West Virginia Service Area

Log Residential Customers
SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.95
R Square	0.90
Adjusted R Square	0.90
Standard Error	0.0025
Observations	14

ANOVA

	df	SS	MS	F	Significance F
Regression	1	0.0007	0.0007	112.51	1.88796E-07
Residual	12	7.53846E-05	6.28205E-06		
Total	13	0.0008			

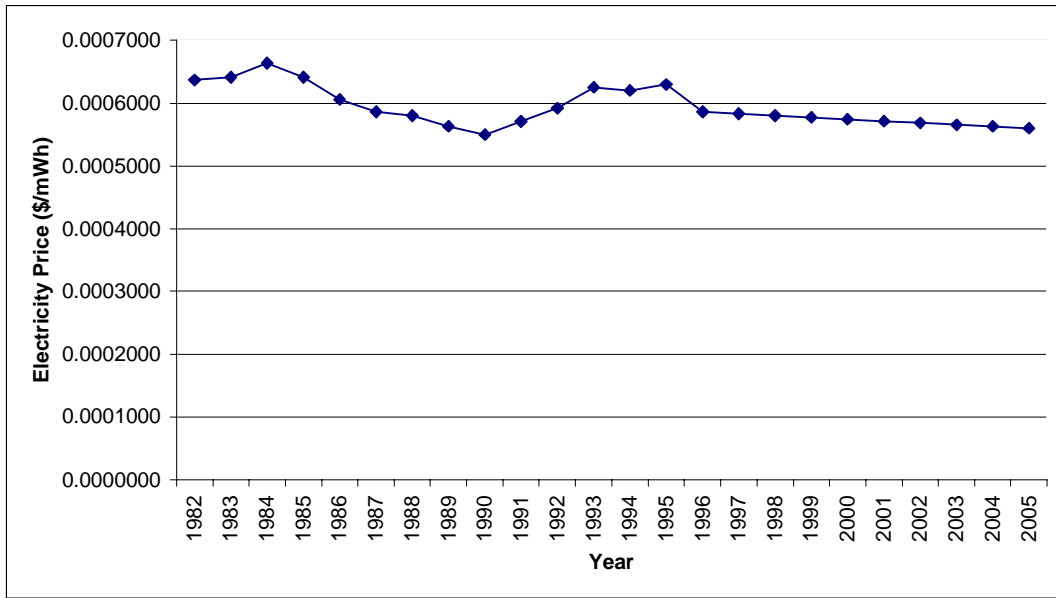|	Coefficients	Standard Error	t Stat	P-value
Intercept	5.55	0.0014	3,923.99	5.05438E-38
Trend	0.0018	0.0002	10.61	1.88796E-07

Figure 23: Graph of Appalachian Power Company's Residential Electricity Price ($/mWh) for West Virginia using Forecast Equation
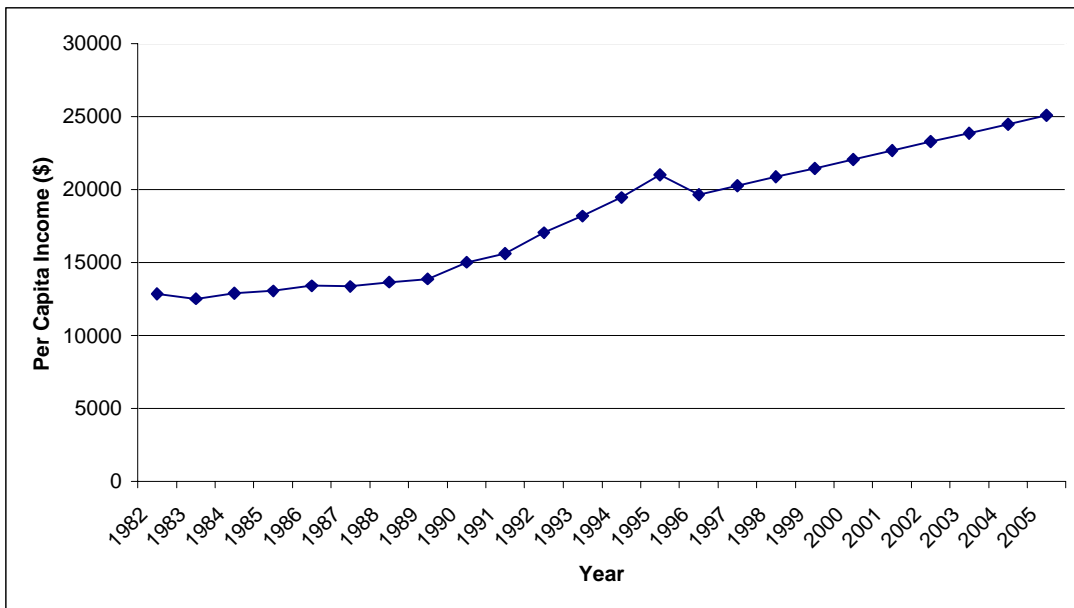

Figure 24: Graph of Per Capita Income for Appalachian Power Company's West Virginia Service Area using Forecast Equation
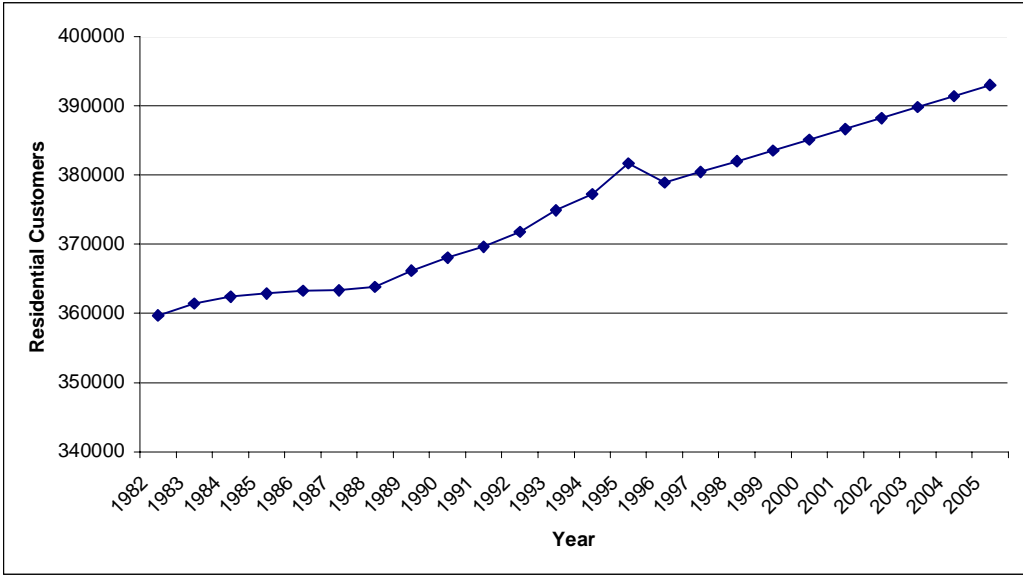
Figure 25: Graph of Appalachian Power Company's Residential Customers in West Virginia using Forecast Equation

Table 30: Confidence Intervals of Forecast Values of Exogenous Variables using Confidence Interval Equation ($\alpha = 0.05$)

Year	Electricity Price	Electricity Price Upper	Electricity Price Lower
1996	0.0005859	720737.385	-720737.384
1997	0.0005830	726437.896	-726437.894
1998	0.0005801	732850.970	-732850.969
1999	0.0005772	739976.609	-739976.607
2000	0.0005743	747814.811	-747814.810
2001	0.0005714	756365.577	-756365.576
2002	0.0005685	765628.907	-765628.906
2003	0.0005656	775604.801	-775604.799
2004	0.0005627	786293.258	-786293.257
2005	0.0005598	797694.280	-797694.278

Year	Temperature	Temperature Upper	Temperature Lower
1996	94.357	99	91
1997	94.357	99	91
1998	94.357	99	91
1999	94.357	99	91
2000	94.357	99	91
2001	94.357	99	91
2002	94.357	99	91
2003	94.357	99	91
2004	94.357	99	91
2005	94.357	99	91

Year	Per Capita Income	Per Capital Income Upper	Per Capita Income Lower
1996	19653.49254	19653.531	-19653.531
1997	20255.57142	20255.612	-20255.612
1998	20857.6503	20857.693	-20857.693
1999	21459.72918	21459.774	-21459.774
2000	22061.80806	22061.855	-22061.855
2001	22663.88695	22663.937	-22663.937
2002	23265.96583	23266.019	-23266.019
2003	23868.04471	23868.101	-23868.101
2004	24470.12359	24470.183	-24470.183
2005	25072.20247	25072.265	-25072.265

Table 31: Confidence Intervals of Forecast Values of Exogenous Variables using Confidence Interval Equation ($\alpha = 0.05$) Continued

Year	Residential Customers	Residential Customers Upper	Residential Customers Lower
1996	378902	379062	-379062
1997	380443	380650	-380650
1998	381990	382250	-382250
1999	383543	383863	-383863
2000	385103	385488	-385488
2001	386669	387126	-387126
2002	388242	388777	-388777
2003	389821	390441	-390441
2004	391406	392118	-392118
2005	392998	393808	-393808

Table 32: Confidence Intervals using Simple Trend Extrapolation

Year	Electricity Price	Electricity Price Upper	Electricity Price Lower
1996	0.0006300	0.000635	0.0006255
1997	0.0006300	0.000650	0.0006254
1998	0.0006300	0.000655	0.0006253
1999	0.0006300	0.000660	0.0006252
2000	0.0006350	0.000665	0.0006251
2001	0.0006350	0.000670	0.0006250
2002	0.0006350	0.000680	0.0006200
2003	0.0006400	0.000685	0.0006150
2004	0.0006500	0.000690	0.0006050
2005	0.0006550	0.000695	0.0006000

Year	Temperature	Temperature Upper	Temperature Lower
1996	94	99	91
1997	94	99	91
1998	94	99	91
1999	94	99	91
2000	94	99	91
2001	94	99	91
2002	94	99	91
2003	94	99	91
2004	94	99	91
2005	94	99	91

Year	Per Capita Income	Per Capital Income Upper	Per Capita Income Lower
1996	21500	23000	21000
1997	23500	26000	22000
1998	25000	28000	23500
1999	26500	31500	24000
2000	28500	32500	24500
2001	29500	35000	25000
2002	31500	37000	26000
2003	33000	39500	26500
2004	34000	40500	27000
2005	35000	42000	29000

Table 33: Confidence Intervals using Simple
Trend Extrapolation (Continued)

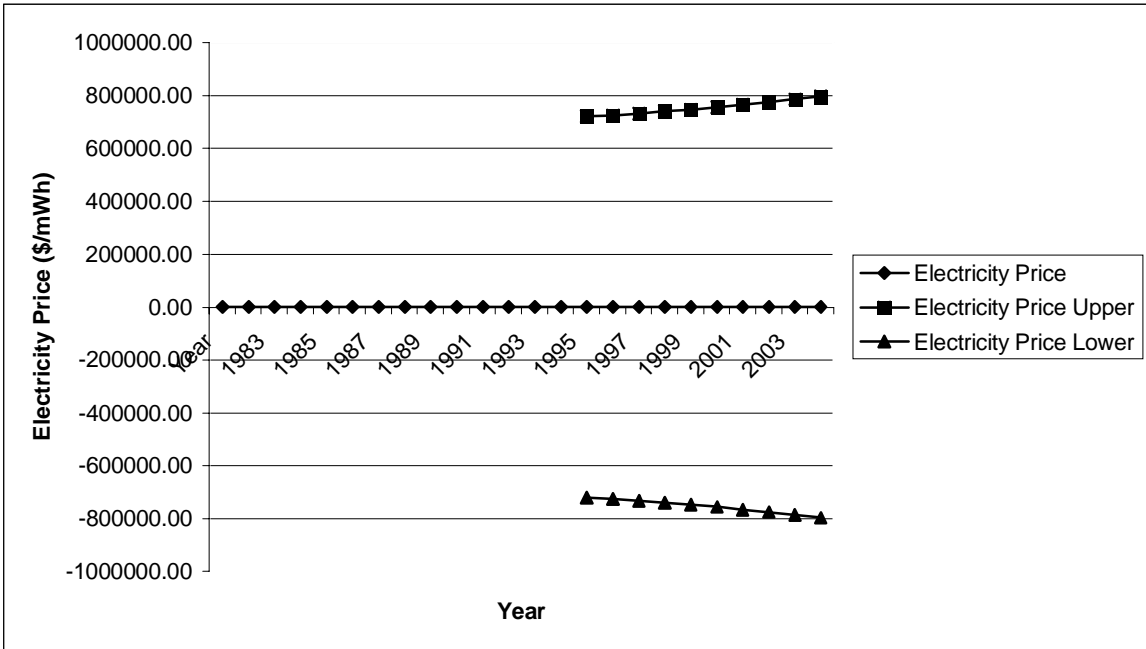| Year	Residential Customers	Residential Customers Upper	Residential Customers Lower
1996	383000	384000	382000
1997	384500	386000	383000
1998	386500	390000	385500
1999	385000	392000	384250
2000	387000	394500	385500
2001	391000	397000	386000
2002	392500	400000	387000
2003	394000	403000	388000
2004	396000	405000	389000
2005	398000	408000	391500

Figure 26: Graph of Appalachian Power Company's Residential Electricity Price ($/mWh) for West Virginia Confidence Interval using Confidence Interval Equation
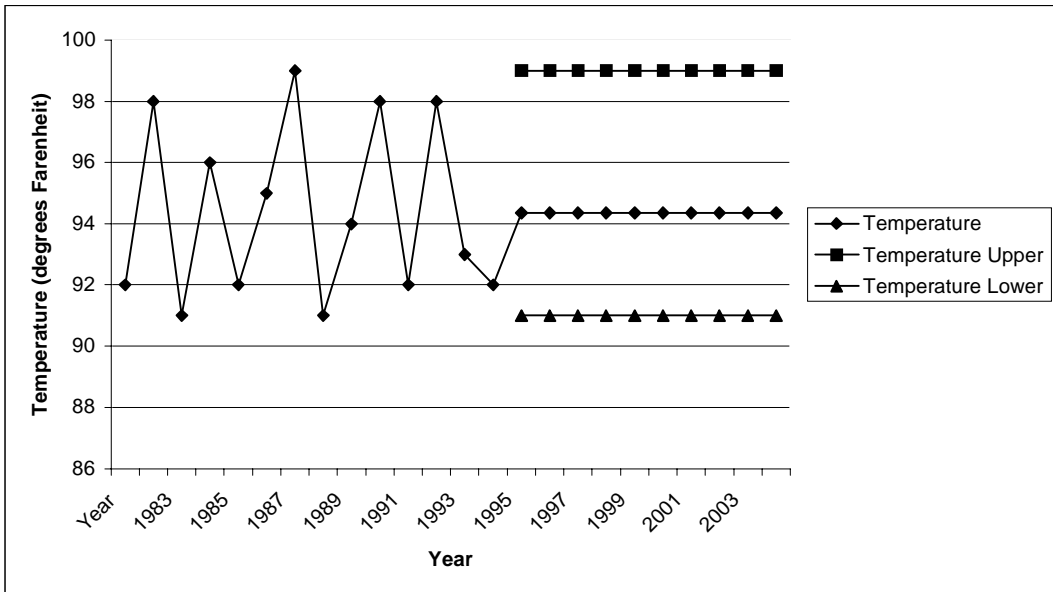

Figure 27: Graph of Confidence Interval for Maximum Temperature on Day of Appalachian Power Company's West Virginia Residential Summer Peak
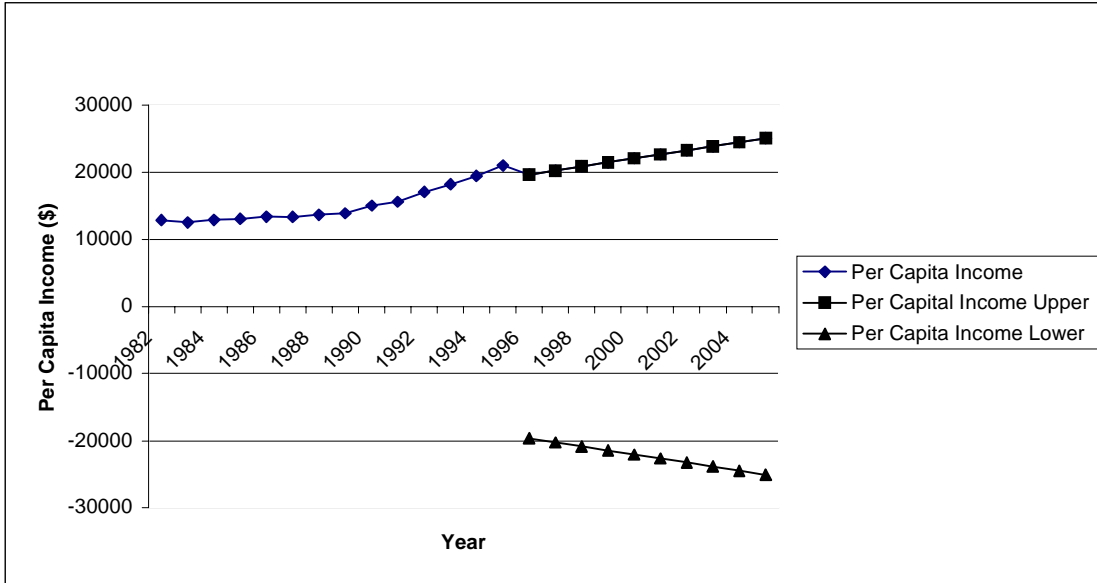
**Figure 28: Graph of Confidence Interval for Per Capita Income for
Appalachian Power Company's West Virginia Service Area
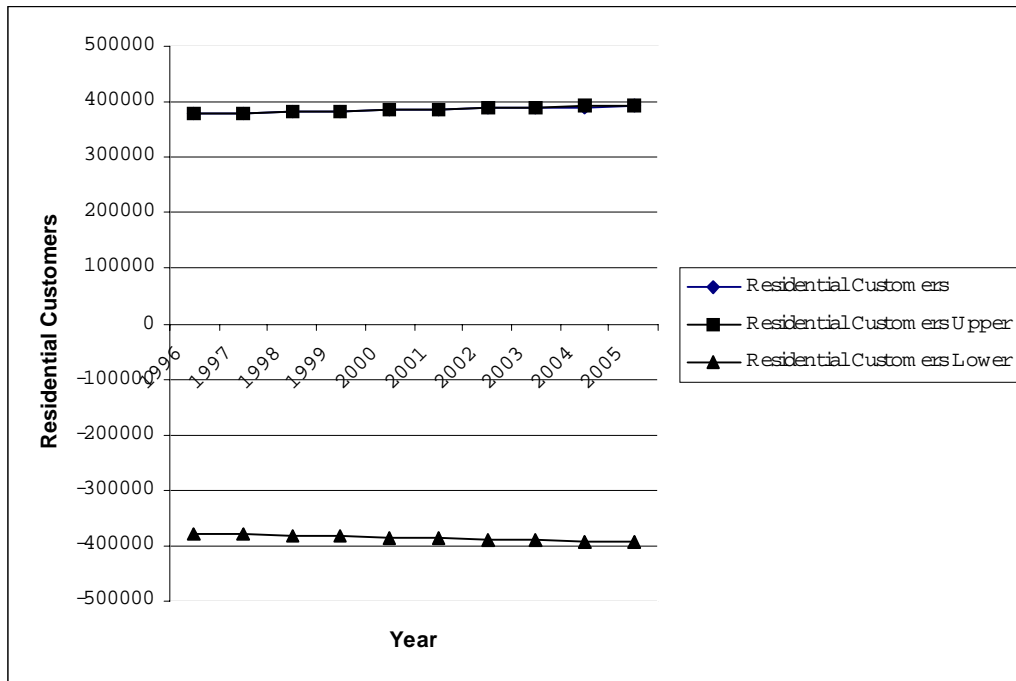using the Confidence Interval Equation**

**Figure 29: Graph of Appalachian Power Company's Residential
Customers in West Virginia using the Confidence Interval Equation**

Kathleen Ann Cullen

Education:

West Virginia University
Masters of Science: August 1999
Field: Agricultural & Natural Resource Economics

Saint Vincent College
Bachelor of Arts: May 1995
Major: Political Science

Employment History

March 1997 – Present
West Virginia University National Research Center for Coal & Energy, Morgantown WV
Administrative Assistant

Duties: Assist with project management on projects such as the Electric Industry Research Group, Industries of the Future – West Virginia, and Electronics Recycling. Job includes assisting with proposal preparation, budget tracking, conference planning, and meeting coordination

August 1995 – March 1997
West Virginia University
College of Agriculture, Forestry, & Consumer Sciences, Morgantown, WV
Graduate Research Assistant

Duties: Literature searches, data searches, data analysis, and aiding with research and paper preparation

June – August 1996
West Virginia Public Service Commission
Summer Intern

Duties: Literature searches, data searches, aided with the development of a residential summer energy demand model for Appalachian Power Company's service area in West Virginia.

August 1994 – May 1995
Saint Vincent College, Latrobe, PA
Tutor for History Department and Interdisciplinary writing/computer lab

Duties: Aiding students with questions in history, aiding in writing and editing of papers, and aiding with computer problems.

May – August 1993
Fayette County District Attorney's Office, Uniontown, PA
Summer Intern

Duties: Record checks on defendants, aiding with basic office support, and watching court cases.