A.E. File Copy

Working Papers in
AGRICULTURAL ECONOMICS

March 1992 92-4

DATA REQUIREMENTS FOR ENERGY FORECASTING

Timothy D. Mount

Department of Agricultural Economics
New York State College of Agriculture and Life Sciences
A Statutory College of the State University
Cornell University, Ithaca, New York, 14853-7801

It is the policy of Cornell University actively to support equality of educational and employment opportunity. No person shall be denied admission to any educational program or activity or be denied employment on the basis of any legally prohibited discrimination involving, but not limited to, such factors as race, color, creed, religion, national or ethnic origin, sex, age or handicap. The University is committed to the maintenance of affirmative action programs which will assure the continuation of such equality of opportunity.

DATA REQUIREMENTS FOR ENERGY FORECASTING

Timothy D. Mount
Department Agricultural Economics
Cornell University

MARCH 1992

Abstract

This paper argues that existing models used to forecast energy use can be improved by incorporating more economic and engineering structure. One reason for the lack of structure in models is that existing data sources for energy use and energy prices for major economic sectors are incomplete. In addition, growing public interest in energy efficiency and the environmental effects of energy use will require much better data on how energy is used in contrast to the traditional focus on how much energy is used.

Presented at the Symposium on Short-Term Energy Forecasting, Energy Information Administration, US Department of Energy, November 1991.

1. INTRODUCTION

The basic premise of this paper is that forecasting is not enough. For any government agency, it is inevitable that requests will be made to analyze different policies and evaluate specific scenarios about future events. Consequently, models that simply produce unconditional forecasts will be inadequate for these types of requests.

There are two characteristics that a good forecasting model for energy should have in addition to being accurate. The first is to have a sound structure that is based on economic and engineering principles. This property determines the feasibility of deriving forecasts conditionally on specified assumptions about a policy or scenario. A good structure makes it easier to identify, for example, how to represent the effects of a government intervention in the energy markets.

The second characterisitic of a good model is that it should include variables that are relevant for a given policy or scenario. When the first major price increase for oil occured in 1973, most models of the national economy did not include an explicit variable for the price of imported oil. This made it difficult to forecast the affects of the price increase on the economy. The lesson is that the design of a good model must attempt to anticipate the types of new policy questions that are likely to occur in the near future. While the oil embargo in 1973 was a genuinely unpredictable event, other important changes occur much more slowly. These changes can be identified in time to redesign models and to modify the procedures used to collect data to estimate the models. Examples of changes that have been anticipated and others that have not are discussed in the paper.

The objective of the existing Short-Term Integrated Forecasting System (STIFS) used by the US Energy Information Administration (EIA) is to "generate short-term (up to 8 quarters), monthly forecasts of U.S. supplies, demands, imports, exports, and stocks of various forms of energy" (DOE-STIFS, p. 1). These forecasts are published quarterly in the "Short-Term Energy Outlook" by the Energy Analysis and Forecasting Division of the Office of Energy Markets and End Use. The structure of the STIFS model is predominantly a series of linear regression equations that together form a system of interrelated equations. The selection of the functional form and the estimation technique is done on an equation by equation basis. The main dangers of building a model equation by equation are that structural links across equations are ignored (e.g., cross-equation constraints in a system of demand equations) and statistical properties may be undesirable (e.g., predicted components do not add to the predicted total).

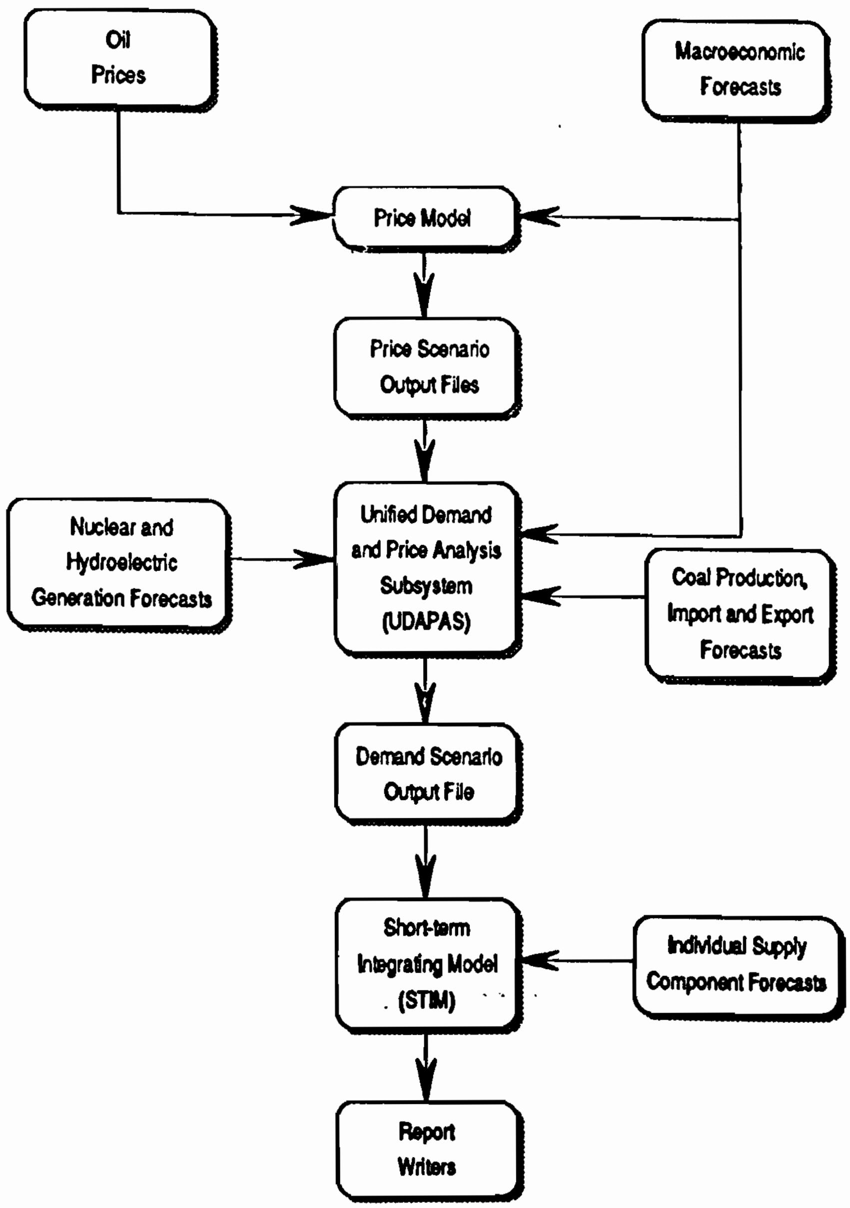
The main conclusion of the paper is that growing public concern about the environment and energy efficiency will lead to new types of questions that can not be answered effectively with current models such as STIFS. A major impediment to redesigning and improving STIFS is that suitable data are not collected on a regular basis. Additional resources should be directed to getting better and more consistent data on how fuels and electricity are used. Although supply issues will still be important for energy markets, greater emphasis should be placed on demand in the future to prepare for the inevitable increase in questions about the use of energy and the effects on environmental quality.

2. DESCRIPTION OF THE STIFS MODEL

The structure of the STIFS model is illustrated in Figure 1. Three basic types of inputs are used to derive forecasts. The first covers measures of economic activity, such as gross national product, personal income and industrial production; the second covers the price of imported crude oil; and the third covers the capacity of nuclear power plants and the amount of hydropower generated. Therefore, forecasts derived from the STIFS model are conditional on these input assumptions, and a wide range of different scenarios can be specified using these inputs. Typically, forecasts are derived for low oil prices, medium oil prices and high oil prices. In addition, other scenarios, such as restricting the quantities of imports of oil, can be evaluated.

There are two major components of the STIFS model. The first is the Unified Demand and Price Analysis Subsystem (UDAPAS), and the second is the Short-Term Integrating Model (STIM). The UDAPAS component determines the demand for different sources of energy by sector, and the STIM component reconciles differences in the forecasts of demand from the UDAPAS with forecasts of supply, including energy conversion processes such as the refining of petroleum products

FIGURE 1
THE STRUCTURE OF THE STIFS MODEL



SOURCE: DOE-STIFS, p. 2.

and the generation of electricity.

A potential problem with the STIFS model is that a variety of different approaches are used to specify the demand for different sources of energy. The demand for gasoline, for example, is represented by a series of equations dealing with miles traveled and the average miles per gallon of vehicles. The demand for other petroleum products is typically handled by fuel, with a distinction between utility and non-utility demand for distillate and residual oil. The demand for electricity and natural gas is specified by sector (residential, commercial, industrial and other).

The types of fuels represented in the STIFS model are summarized in Figure 2. A basic characteristic is the lack of consistency in the coverage by sector (column). For example, electricity and natural gas are represented by sector, but petroleum products are represented for combinations of sectors. These differences reflect the difference between the type of data available from the two regulated industries (electricity and natural gas) compared to the petroleum industry. Nevertheless, the differences in the treatment of distillate oil, for example, compared to electricity and natural gas present a major obstacle to specifying a more integrated economic structure of the demand for fuels and electricity by sector.

In spite of the idiosyncracies that exist in the definitions of fuels in the STIFS model, the objective is to provide a reasonably complete inventory of all sources of energy. In contrast, the representation of prices in the STIFS model does not match the categories of fuels in Figure 2. In addition, the price coverage is only partial (see Figure 3). For example, the price of electricity is considered for the residential sector only. Even though additional data on prices are readily available for electricity. The lack of consistency between quantities and prices represents another major complication in specifying an integrated economic structure of demand by sector.

The objective of any forecasting model is to use available information efficiently. If existing sources of data make it impractical to build an integrated economic model, then a less structured model is inevitable. However, a potentially important cost of using an unstructured approach is that desirable properties of logical consistency, such as market clearing, may be lost. In the STIFS model, the need to reconcile demand and supply forecasts in the STIM component of the model is an indication that the model lacks a formal economic structure.

FIGURE 2
QUANTITIES OF FUELS DEMANDED BY SECTOR

FUEL	SECTOR R	C	I	T	O	U
Petroleum						
Distillate	<------DHFCPUS------>		DSFICPUS	DFACPUS	-	DKEOPUS
Residual	<-------------RFNUPUS---------------->			-	-	RFEOPUS
Petr. Coke	-	-	-	-	-	PCEOPUS
Gasoline	-	-	-	MGTCPUS	-	-
Jet Fuel	-	-	-	JFTCPUS	-	-
Liq. Petr. Gases	<------------------------LXTCPUS------------------------>					
Ethane	<------------------------ETTCPUS------------------------>					
Miscellaneous	<------------------------MSTCPUS------------------------>					
Pentane	<------------------------CPTCPUS------------------------>					
Other Hydrocar.	<------------------------OHRIPUS*------------------------>					
Petr. Feedstock	<------------------------FETCPUS------------------------>					
Natural Gas	NGRCPUS	NGCCPUS	NGINPUS	-	**	NGEOPUS
Coal	<-------CLHCPUS------>		CLXCPUS	-	CCPRPUS	CLEOPUS
Electricity	ESRCPUS	ESCMPUS	ESICPUS	<----------ESOTPUS-------->		-
Hydro	-	-	-	-	-	Exogenous
Nuclear	-	-	-	-	-	NUEOPUS
Wind	-	-	-	-	-	WNEOPUS
Geothermal	-	-	-	-	-	GEEOPUS
Wood and Waste	-	-	-	-	-	WWEOPUS
Imports	-	-	-	-	-	Exogenous
Non-utility	-	-	-	-	-	Trend

FIGURE 2, continued

LEGEND

R	Residential
C	Commercial
I	Industrial
T	Transportation
O	Other
U	Electric Utility

Non-utility demand for distillate (DOC 1, p. 8)
DFNUPUS = DFACPUS + DFHCPUS + DFICPUS

Electricity generation from petroleum (DOC 1, p. 16)
PAEOPUS = DKEOPUS + RFEOPUS + PCEOPUS

* Listed as OLTCPUS (DOC1, p. 17)
** NGLPPUS + NGACPUS (DOC1, pp. 19-20)

FIGURE 3

PRICES OF FUELS BY MARKET LEVEL

FUEL	MARKET LEVEL			
Petroleum	**Average**	**Retail**	**Wholesale**	**Utility**
Crude	Exogenous	-	-	-
Gasoline	-	MGUCUUS	MGWHUUS	-
Distillate	-	D2RCUUS	D2WHUUS	-
Diesel	-	DSTCUUS -	-	-
Residual	-	RFTCUUS	-	RFEUDUS
Jet Fuel	JKTCUUS	-	-	-
Natural Gas				
Wellhead	Trend	-	-	-
Sales	-	NGRCUUS	NGCCUUS* NGICUUS**	NGEUDUS
Coal				
Sales	-	-	-	CLEUDUS
Electricity				
Sales	-	ESRCUUS	-	-

* Price to commercial customers
** Price to industrial customers

The structure of the STIFS model is based on a series of individual demand equations that are estimated one by one. A typical equation uses the quantity of fuel demanded as the dependent variable, and the regressors may include a measure of economic activity (e.g., income), the price of the fuel, weather, other shift variables to reflect unusual events (e.g., strikes), and a lagged dependent variable (or lagged prices) to allow for the dynamic adjustment of demand to changes in the other regressors. Sometimes a price ratio is used to represent the price of the fuel relative to the price of a competing fuel. The existing structure of the STIFS model is not based on the standard restrictions on demand parameters derived from economic theory, and it is not estimated by multivariate techniques. Building more economic structure into the model and using multivariate estimation techniques would improve the efficiency (i.e. accuracy) of the estimators.

3. ALTERNATIVE APPROACHES TO FORECASTING

A. Multivariate Models

In general, the estimation of a system of equations is more efficient if system estimators are used instead of single equation estimators. In some cases, such as vector autoregressive (VAR) models, fewer parameters are needed in a multivariate framework than in estimating the equivalent univariate models. It seems reasonable to expect that the unexplained stochastic residuals of the demand equations for different fuels are related due, for example, to unusual weather conditions. Therefore, the potential for gaining efficiency through using system estimators exists because residuals are correlated across equations and it is also likely that seasonal patterns are related for different fuels. An example of how VAR models perform better than single-equation models is given by Joutz and Trost.

If economic structure is ignored for the present, it is still important to determine whether structural changes have occured. The basic approach is to fit a model for a sample of observations and then to use the estimated model to compute forecasts. If the performance is viewed as inadequate, the model is reestimated, and, if necessary, the form of the model is modified. This general approach to estimating a model is typical of econometric applications. Deciding whether the structure of a model is consistent with new data can be tested formally using, for example, the standard errors of the forecasts derived from the Fair-Parke procedure. If the observed forecasting errors fall outside a computed confidence band, it is an indication that a structural change may have occurred (see Hendry and Ericsson for an example of this type of monitoring of a model's performance).

The underlying logic of a time-series model is to use all the relevant information at a point in time to make a forecast of the future. The natural framework for this type of analysis is to build a recursive structure through time (see Harvey, p. 125). One advantage of this approach is that it provides a natural way to incorporate the effects of autoregressive and moving average residuals on the forecasts. Another advantage is that it is possible to establish efficient methods for updating the forecasting model as new data become available (see Harvey, Chapter 3).

The general conclusion is that there are potential gains in efficiency for forecasting by using multivariate estimators for systems of related equations. In addition, there are efficient methods for updating forecasting models as new data become available. The main advantage of using single-equation estimators is that they are relatively simple to implement, but this simplicity is bought at the expense of efficiency.

B. Economic Structure

An extension to the proposal to use multivariate estimation techniques is to build more structure into the model. Two types of structure could be considered. The first is termed "data admissibility" and it refers to capturing the logical characteristics of the variables through the structure of the model. An example is that component quantities of the total supply of a fuel should be non-negative and add to the total. The second form of structure is economic through, for example, introducing cross-equation constraints in a system of demand equations. These constraints are derived from economic theory and reflect the degree of substitutability between different commodities or factors of production. In both cases, adding structure to a model, if correct, leads to greater efficiency in estimation.

A major obstacle to building economic structure into the demand equations for fuels is the lack of consistency in the available data on demand quantities by sector. An important question is whether it is possible to measure the prices and quantities for electricity and major fuels (natural gas, distillate, residual and coal) by sector, for example, on a monthly basis. The annual model developed by Jorgenson and others is an example of a fully structured economic system for demand and supply that divides the economy into 35 sectors and identifies all major fuels and electricity.

The demand systems in the Jorgenson model are based on linear shares of expenditures on different commodities or input factors derived from translog utility or costs functions. These systems exhibit additivity, implying that the sum of predicted expenditure shares is always one for each system of demand equations. Another fully flexible form of share function, that has the advantage of ensuring non-negative shares as well as additivity, is based on the linear logit model (see Considine and Mount). An application of this model to the mix of fuels used to generate electricity has been developed (Decision Analysis Corporation).

One reason for using a structured economic model is that market clearing can be used to ensure that demand and supply are equated. This is possible if markets are complete (i.e., no prices are omitted). If markets cleared automatically in a model, it would be unnecessary to use the adjustment mechanism embedded in the STIM component of the STIFS model to impose logical consistency between the quantities demanded and the quantities supplied. More importantly, it is possible to use a structured economic model for policy analysis as well as forecasting.

4. ANTICIPATING FUTURE POLICY DIRECTIONS

The current version of the STIFS model was designed to deal with questions of whether supplies of fuels would be adequate to meet demand and to predict price changes. Nevertheless, requirements for policy analyses are changing, and in particular, there is increasing public concern about how fuels are used and the environmental implications of fuel use. As a result of these changes, it is sensible to place more emphasis on the purposes (i.e. end-uses) for using fuels rather than treating the amounts of fuels demanded as the main characteristic of demand. If heating is the end-use, for example, the amount of fuel used depends on the size of buildings and the thermal integrity of their construction. Energy efficiency is determined by the ratio of services delivered to the amount of fuel used.

Monitoring energy efficiency should become a major objective of modeling energy markets. These efforts would compliment the existing objective of monitoring whether supplies of fuels are adequate to meet demand. While efforts have been initiated to gather new data about changes in the supply of energy, there appears to be much less resolve to deal with issues of energy efficiency and demand. A good example of anticipating future policy questions about energy supply is provided by the development of new surveys to measure electricity production by Non-Utility Generators (NUG). The Electric Power Division of EIA has designed surveys to cover this emerging source of electricity generation (see Kimbrough) that has been stimulated by the Public Utilities Regulatory Policies Act (PURPA). NUG sources will become a major component of the future growth of generating capacity in the electric utility industry. In 1986, North American Reliability Council (NERC) anticipated that 9600 MW of NUG capacity would be installed by 1995. However, the current projection is that 18650 MW of NUG capacity will be installed between 1991 and 1995. The new EIA surveys will provide data that will be essential to understanding the changing structure of the electric utility industry and to forming a basis for developing new models of electricity supply.

In contrast to this supply example, perennial questions of demand remain unanswered. Given the importance of gasoline in the US economy, for example, it is remarkable how little is known about how patterns of gasoline use have changed over time. The lack of data on the use of gasoline by different sectors makes it virtually impossible to build structured economic models of demand. To illustrate this point, the demand for gasoline is treated in the STIFS model as a separate sector rather than dividing demand between residential and commercial uses. Furthermore, within residential demand, it would be desirable to distinguish between the use of gasoline for commuting and other types of uses. This information would be valuable for monitoring efficiency gains and also for evaluating policies, such as Regulation XV in Los Angeles, designed to affect communting patterns. Since urban air quality is a major environmental problem throughout the US, better understanding of changing commuting patterns and the implications for fuel use is essential. Periodic surveys such as the Nationwide Personal Transportation Study are simply inadequate, and furthermore, do not provide a basis for building better models of energy use for policy analyses and forecasting.

One serious problem with any attempt to measure the end-use efficiencies of energy is that data needs are truly massive. To keep the costs of data collection within bounds, it will not be possible to collect everything that could be justified on analytical grounds. My proposal would be to establish rolling panels of households and establishments to monitor energy use over time for these panels. Since energy use is determined to a large extent by the existing stock of appliances and equipment, panel data would make it possible to observe and record changes in this stock as well as changes in energy use. These types of rolling panels are already used to measure the cost-of-living in the Consumer Expenditure Survey. Obtaining time-series data on how energy use changes is likely to be more valuable than taking detailed cross-sectional surveys of energy use at infrequent intervals.

5. CONCLUSIONS

The main theme of this paper is that forecasting models of energy use should be based on sound economic and engineering structures. Economic structure should reflect the demand and supply characteristics of energy markets, and the engineering structure should reflect the physical basis for transformation processes such as petroleum refining and the generation of electricity. The rationale for this assertion is that the requirements of federal agencies are not limited to forecasting in the narrow sense of predicting what is going to happen next. Forecasting models will inevitably be used to answer questions about the effects of different policies. It is these situations that make economic structure in a forecasting model valuable because it is generally easier to represent a policy intervention in a structured economic model than it is in an unstructured forecasting model.

A second theme of the paper is that the availability of time-series data limits the types of models that can be developed to answer policy questions. Now that the importance of environmental problems relating to fuel use have been recognized by the public and government, questions about how fuels and electricity are used and the associated energy efficiencies are likely to become more frequent in the future. Consequently, it is no longer sufficient to limit the objectives of a model to forecasting total quantities of fuels demanded. It will be necessary to model how fuels are used, and to provide more insight into changes in the end-uses of fuels and electricity for the residential, commercial and industrial sectors.

A goal for new forecasting models should be to improve the collection of data so that better economic models of the demand for fuels and electricity can be developed. The primary limitation in current sources of data on demand is the lack of information about petroleum products compared to the information that is available about natural gas and electricity. A minimal requirement is that consistent data series on the prices and quantities demanded for each source of energy should be available by sector. At the present time, data sources reflect the supply of energy, and as a result, it is difficult to integrate data on the demand for petroleum products with the demand for other sources of energy. A good example of this problem is the lack of data on gasoline use by sector. Without these data, it is not possible to include gasoline use in models of the residential sector. This in turn makes it difficult to answer questions about commuting patterns.

One approach to improving data collection is to use rolling panels of households and establishments. These panels would provide better time-series data than cross-section surveys. In addition, it would be possible to get information about changes in the stock of appliances and equipment that determine, to a large extent, how efficiently energy is used. In other words, data series would be available that reflect the underlying structure of the demand for fuels and electricity by sector.

One potential danger in gathering better data about energy demand is that these data could be used to enforce regulations about pollution, for example. If this happens, it will be difficult to maintain cooperation in collecting data. It is essential that the identities of people and firms providing data for the EIA are protected, and that a clear separation is made between data collected for information and data collected for the enforcement of government regulations.

Improving the quality of the data collected on energy demand will require time, but there are a number of interim steps that can be taken now to improve existing forecasting models. Many of these steps have already been initiated. First, estimation efficiency can be improved by using multivariate techniques instead of single-equation methods (see Joutz and Trost, November 1991). Second, better methods of evaluating the existing models can be used to measure forecasting accuracy and test for structural changes (see Joutz and Trost, March 1991). Third, better methods of updating forecasting models are available to take advantage of new data (see Harvey). These updating procedures are closely related to testing for structural change. Finally, opportunities exist with existing sources of data, particularly on the supply side, to improve the economic and engineering structure of forecasting models. Examples have been developed by Decision Analysis Corporation for electricity generation and by Considine for petroleum supply.

6. REFERENCES

Considine, Timothy J. "A short-run model of petroleum product supply", Pennsylvania State University, July 1991.

Considine, Timothy J. "Symmetry constraints and variable returns to scale in logit models", Journal of Business and Economic Statistics, Vol. 8, No. 3, July 1990.

Considine, Timothy J. and Timothy D. Mount. "The use of linar logit models for dynamic input demand systems", Review of Economics and Statistics, Vol. 66, 434-443, 1984.

Decision Analysis Corporation of Virginia. "Estimation of the regional electric utility fuel choice model", Final Report to the EIA, October 1991.

Harvey, Andrew C. Forecasting, Structural Time Series Models and the Kalman Filter, Cambridge University Press, Cambridge, England, 1989.

Hendry, David F. and Neil R. Ericsson. "An eonometric analysis of U.K. money demand", American Economic Review, Vol. 81, No. 1, March 1991.

Jorgenson, Dale W. and Peter J. Wilcoxen. "Environmental regulation and U.S. economic growth", RAND Journal of Economics, Vol. 21, No. 2, Summer 1990.

Joutz, Frederick and Robert Trost. "Comparison of Bayesian VAR and unrestricted VAR forecasting models of motor gasoline and distillate fuel prices and quantities", DOE/EIA, November 1991, (DOE - OUTLOOK).

Joutz, Frederick and Robert Trost. "Getting forecast standard errors of motor gasoline demand with stochastic simulation", Working paper presented to The American Statistical Association Committee on Energy, March 1991.

Kimbrough, Mary D. "EIA's initial nonutility survey: experience gained", Working paper presented to The American Statistical Association Committee on Energy Statistics, October 1991.

North American Electric Reliability Council. "1990 Electricity Supply and Demand for 1990-1999", Princeton, New Jersey, November 1990.

US Department of Energy, Energy Information Administration. Short-Term Integrated Forecasting System: 1990 Model Documentation Report, DOE/EIA-MO41, March 1991, (DOE-STIFS).

US Department of Energy, Energy Information Administration. Short-Term Energy Outlook: Annual Supplement 1990, DOE/EIA-0202(90), February 1991, (DOE-OUTLOOK).

OTHER AGRICULTURAL ECONOMICS WORKING PAPERS

No. 91-16	Rural Household Data Collection in Develop9ing Countries: Designing Instruments and Methods for Collecting Time Allocation Data	Julie P. Leones
No. 91-17	Rural Household Data Collection in Developing Countries: Designing Instruments and Methods for Collecting Farm Production Data	Scott Rozelle
No. 91-18	Rural Household Data Collection in Developing Countries: Designing Instruments and Methods for Collecting Off-Farm Income Data	Julie P. Leones Scott Rozelle
No. 91-19	Rural Household Data Collection in Developing Countries: Preparing the Data for Analysis	Tom Randolph
No. 91-20	Clean Power Supply through Cogeneration	Gary W. Dorris Timothy D. Mount
No. 91-21	A Complete, Three-Level Demand System for Foods in Japan	John Dyck Timothy D. Mount
No. 92-01	A Theory of Competition Between Politicians for Political Support	Jo Swinnen Harry de Gorter
No. 92-02	Negotiating a Response to Climate Change: The Role of Biological Emissions	Thomas E. Drennen Duane Chapman
No. 92-03	Environment, Income, and Development in Southern Africa	Duane Chapman