

National Institute for Statistical Science: Project Descriptions ¹
Ruey-Pyng Lu, Statistics and Methods Group
202-586-5985, ruey-pyng.lu@eia.doe.gov

Abstract

The Energy Information Administration (EIA) has established an Applied Methodology Student Research Program to be conducted jointly with the National Institute of Statistical Sciences (NISS). This program is designed to engage students in graduate programs in collaborative, interdisciplinary research on topics of interest to EIA. Research projects undertaken by students and their technical mentors at their home universities and their research sponsors at EIA will advance methodology and address specific research questions that are important to EIA programs. Students may apply either for full-time summer support or for research support during the academic year as either a full or partial stipend or as a cooperative arrangement that would provide research access to sensitive microdata. Ordinarily academic year research will be conducted at the student's university; summer research may be conducted either on-site at EIA or at the student's university with visits to EIA as necessary to consult with the EIA project sponsor and to access and utilize EIA data, as required by the particular project. Graduate students in statistical science disciplines (statistics, operations research, economics, etc) together with their faculty mentors are encouraged to identify a research area proposed by EIA and to define research objectives in conjunction with EIA staff preparatory to submitting a full proposal. Awards are made based on the strength of the applications. These are renewable based on accomplishments and approved future plans. Citizenship is not required; women and members of minority groups are especially encouraged to apply.

We are in the process of developing specific project statements, in the hope of encouraging students to work on specific projects of use to EIA. We know it is important to provide enough detail so the researcher will have an idea as to the methodology and to data and its availability. So far we have developed draft descriptions of four research projects, and they are included below for your review and discussion.

¹This is a working document prepared by the Energy Information Administration (EIA) in order to solicit advice and comment on statistical matters from the American Statistical Association Committee on Energy Statistics. This topic will be discussed at EIA Spring 2007 meeting with the Committee to be held April 19 and 20, 2007.

Questions to ASA Energy Committee:

1. What can EIA do to attract the interest of students and new researchers in EIA projects?
 - a. How should these project descriptions be improved?
 - b. How should we advertise the program?

Research Project 1

Project Name: Development and testing of alternative editing and imputation methods for Energy Information Administration establishment surveys.

Research Question: Can new model-based or other approaches to micro data editing and imputation improve the accuracy of EIA's micro data and/or simplify EIA's data processing systems?

Background: For some of EIA's monthly surveys, regression, time series and other model-based imputation techniques have been used to impute individual respondent data that are missing due either to non-selection in the sample, non-response, or edit failures. Many other editing and imputation techniques are used within the Federal Statistical System. Alternative model-based or non-model-based approaches to imputation should be tested with the goals of improving accuracy and/or simplifying the imputation processes.

Short description: EIA will provide monthly data at respondent level from the largest (sampled) companies, and annual data at respondent level from the census of all companies. The research project has two parts:

- 1) Study the data sets provided and identify a minimum of three alternative imputation techniques that may be appropriate for imputing certain key variables. Some of the variables of interest will have seasonal patterns. All the proposed imputation methods should be appropriate for implementation in the context of a large-scale statistical production system.
- 2) Empirically test the proposed methods and evaluate their accuracy relative to each other and to the methods currently used by EIA. Prepare a detailed report describing the research and the results.

Data: The EIA data sets will most likely contain data collected from establishments in the electrical power, oil, and/or natural gas industries. In addition to the datasets that EIA prepares specifically for this project, the researcher may identify and use other publicly available data, as appropriate.

Methodology: A wide variety of statistical editing and imputation techniques are discussed in the statistics literature. Many involve regression and/or time series modeling techniques. Others simpler methods (e.g., hot deck) are commonly used by statistical agencies but may require creative modification to be suitable for new applications. The researcher should be familiar with standard imputation methods, survey estimation techniques, and techniques for testing alternative imputation procedures.

Selected References:

1. Rubin, Donald B., **Multiple Imputation for Nonresponse in Surveys**, Wiley, 1987.
2. Kish, Leslie, **Survey Sampling**, Wiley, 1965
3. Kirkendall, N. and Sedransk, J., "Data Analysis for the EIA-826: Test results," presented at the 2005 Spring meeting of the American Statistical Association.

Research Project 2

Project Name: Combined Heat and Power Plant Fuel Allocation Methodology

Research Question: How to partition fuel used at Combined Heat and Power plants (CHP's) into fuel used for generating electricity and fuel used for other purposes (such as process heat)? The issue arises because the cogeneration environments used in industry or for some commercial applications produce both electricity and heat. EIA wishes to capture the total amount of fuel used for electricity generation alone. Companies can and do report good data for total amounts of electricity generated and for total fuel used.

Background: Initially, EIA collected data from CHP's concerning total fuel use, and "useful thermal output" measured in BTU. The data were used in a model to estimate the fuel used to produce the useful thermal output. In about 2002, EIA modelers stated that the data were not accurate, and that they understated the fuel used to generate electricity. In 2003, EIA changed its survey forms to ask respondents to estimate the percent of fuel that is used for electricity generation. In 2005, EIA modelers observed that data collected in this new way tended to overstate the amount of fuel used to generate electricity. EIA is in need of a better way to estimate the amount of fuel used to generate electricity. What data should be collected, and what model should be used?

Short Description: The researcher will develop, test, and evaluate a new methodology for allocating the reported fuel consumption by CHP plants between fuel for electricity and fuel for useful thermal output. The new methodology must be designed to substantially reduce or eliminate the effects of the persistent data collection problems EIA has encountered in this area. Specific tasks will include:

- Review current and past Electric Power Division (EPD) approaches taken to this problem.
- Become familiar with EPD data collection instruments and systems.
- Perform a literature review on allocation methods and issues.
- Recommend a new methodology for performing the allocation, as well as respondent level data that would be needed.

Data: Form EIA-920 - Monthly Nonutility Power Plant Report, which contains monthly data on electricity generation and fuel use by type of fuel. Also available is information about the plant equipment and boiler units. What data would we give them? Could take a few observations that illustrated some things in Kaplan's talk.

Methodology: The new proposed methodology should address the problem at both the data collection and the data processing/estimation stages. The researcher should be familiar

with data collection, data editing, and estimation techniques used in establishment surveys.

Selected References:

1. Aviel Verbruggen “**Combined Heat and Power (CHP) essentials,**” *International Journal of Energy Technology and Policy* 2007 – Vol. 5, No.1 pp. 1 – 16.
2. Stan Kaplan, briefing to the ASA Committee on Energy Statistics, “Making Adjustments to Survey Data When Collected Data Do Not Meet Expectations,” April 2006.
http://www.eia.doe.gov/smg/asa_meeting_2006/spring/files/adjsurveyd.ppt

Research Project 3

Project Name: U.S. Biodiesel Production and Distribution

Research Questions: How can EIA improve its analysis of the current status of and near-term outlook for the U.S. biodiesel production and distribution industries?

Background: Section 503 of the Energy Policy Act of 1992 requires EIA to provide estimates of replacement fuel consumption. Although biodiesel is a replacement fuel, EIA does not presently collect data on biodiesel prices or production. In addition to fulfilling EIA’s general mandate for timely, reliable, and relevant information in the alternate fuels area, this project should provide at least anecdotal but reliable price and quantity information on biodiesel in partial fulfillment of Section 503.

Short Description: This project includes, but is not limited to:

- Briefly summarize the history of biodiesel, how it is made, and its characteristics. Refer to existing EIA documents wherever possible, e.g., Biodiesel Performance, Costs, and Use
- Biodiesel production
 - Describe biodiesel producers.
 - Describe how biodiesel is currently produced, e.g., using what feedstocks/methods?
 - List the major producers; provide a brief description of their operations. Provide a comprehensive list of all the known biodiesel producers and their locations/capacities and quantities produced.
 - Discuss fuel quality and handling issues producers face, if any
 - Discuss biodiesel blends produced shipped. Do biodiesel producers only ship pure biodiesel, or do they blend it sometimes/always?
- Biodiesel distribution
 - Describe biodiesel distributors.
 - Who purchases biodiesel from producers?
 - How does biodiesel enter the petroleum supply chain?
 - How much biodiesel is sold outside the petroleum supply chain, to whom is it sold, and how does this biodiesel become a final product?

- What biodiesel blends do blenders/wholesalers distribute?
- Provide documentable sample prices for biodiesel over the past 24 months.

The project's goals are (1) to provide general information on biodiesel market activity and (2) to create a current, comprehensive database of information on biodiesel producers, prices, consumption levels, and market activity. The project's deliverables include a database containing biodiesel information and a final report prepared for distribution on EIA's website. The report should include appropriate tables and graphics presenting information from the database.

Data: The researcher should perform a comprehensive search for publicly available data on biodiesel production, blending, distribution, and use.

Methodology: This project is primarily a literature/web search and documentation/database creation project. The researcher should be skilled in the use of library resources and web search engines as well as database design and creation.

Selected References:

EIA has received a very large number of customer requests for biodiesel information during the past year. This project would help fulfill a number of these requests.

Research Project 4

Project Name: Potential use of cointegration analysis in Energy Information Administration projection models.

Research Question: Are there cointegration relationships between energy price series (oil, gas, coal, electricity, etc.). If there are such relationships, what are they and how should (can) the Energy Information Administration adapt its short and/or medium term models to accommodate these relationships?

Background: West Texas Intermediate (WTI) oil spot prices and Henry Hub (HH) natural gas spot prices appear to have moved together in the past 15 years for substantial amounts of time; they have also moved in opposite directions for other 3-6 month periods. Some energy experts argue that oil and gas prices must be related in the longer term and that they tend to converge toward some long term Btu price relationship which is confused or obfuscated by the variability of weather or other unusual shocks affecting supply and demand. Others argue that the relationship is far more complex and involves all energy forms and possibly technological alternatives that may convert one form of energy to another (like coal to liquids for transportation or coal to electricity in generation).

Previous EIA analysis [7] suggested that oil and gas prices were cointegrated. Two major concerns about the original EIA analysis were the need to use (a) a trend term and (b) many seasonal dummy variables to smooth out some of the data points. Although a trend term is common in most time-series data, it is unsatisfying simply to say that a series will

continue increasing exponentially over 20 years. One possible explanation in this case is that there may be missing critical variables, like the price of coal. Finally, the number of deleted data points, treated by using dummy variables for them, seemed uncharacteristically large.

Short description: This project has two parts: 1) fitting cointegration models to energy, and possibly related non-energy, price series to establish long term and short term price relationships, and to determine whether these have changed over time; 2) advising EIA concerning how identified relationships should (could) be used in EIA's short and midterm models to enhance their projection capabilities. Part 1 involves considering frequency of data (daily, weekly, monthly, annual), and modeling details such as advantages and disadvantages of including trend terms and other economic variables or commodities. Part 2 could involve work with appropriate EIA staff, possibly on location, to develop sufficient understanding of EIA models to recommend how the findings could be incorporated into the short term and midterm models.

Data: EIA can provide (monthly?) time series for xxx that have been used in previous analysis, and the daily data that were used to construct them. Researcher will identify and use other publicly available time series, as appropriate.

Methodology: Cointegration analysis is a method of analysis for non-stationary time series that identifies short and long term relationships between series. The second part of the project will require reviewing EIA's short and long term modeling documentation to assess how the cointegration results might be used.

Selected References:

1. Alexander, Carol, **Market Models, A Guide to Financial Data Analysis**, Wiley, 2001
2. Chan, Ngai Hang, **Time Series Applications to Finance**, Wiley, 2002
3. Hendry, David and Juselius, Katarina, **The Energy Journal**, "Explaining Cointegration Analysis: Part I, Volume 21, Number 1, 2000, pages 1 -42.
4. Hendry, David and Juselius, Katarina, **The Energy Journal**, "Explaining Cointegration Analysis: Part II, Volume 22, Number 1, 2001, pages 75-120.
5. EIA Documentation for mid-term model, xxx module of the NEMS (web link)
6. EIA Documentation for the Short Term Energy Model, (web link)
7. Villar, Jose and Joutz, Fredrick, "The Relationship between Crude Oil and natural gas prices", feature article in EIA's Natural Gas Monthly, 2006, http://www.eia.doe.gov/pub/oil_gas/natural_gas/feature_articles/2006/reloilgaspri/reloilgaspri.pdf. Submitted for publication.

EIA contact: Dr. Andy S. Kydes