

Part B

Collection of Information Employing Statistical Methods

B.1. Respondent¹ Universe

The electric power surveys collectively cover the entire range of companies involved in the generation, transmission, distribution, and sales of electricity. Of the six surveys in this package, three surveys are of the entire universe (or nearly the entire universe) based on more exacting filing requirements given in those surveys; a fourth survey is for additional information as it becomes available. A fifth survey has both an annual census and a monthly sample component; and a sixth survey is a monthly sample survey corresponding to one of the annual census surveys. The respondent frame for each survey is:

- **Form EIA-411** – The target population for this annual census comprises all electricity generators and electric utilities in the United States. The eight Regions of the North American Electric Reliability Corporation (NERC) collect the data from the target population units. Each Region assembles the required information using input from the member electricity generators and electric utilities in its geographic area. The Regions submit the compiled data to the NERC headquarters, where it is consolidated and forwarded to the EIA.
- **Form EIA-826** – The target population for this monthly survey comprises all U.S. electric utilities, electric service providers, and distribution companies. Cutoff sampling is used to select the sample for the Form EIA-826, which includes most of the investor-owned utilities (188), 4 Federal utilities, all electric service providers (92), all distribution companies, and a sample of approximately 164 municipal, cooperative, State and political subdivision utilities that have sales to end-use customers.
- **Form EIA-860** – The target population for this annual census comprises all existing and proposed (for operation within 5 years) electric power plants that have a total generator nameplate capacity of 1 megawatt or greater. Companies complete the form for all the plants they operate. There are approximately 2,700 entities that operate and/or propose to operate about 5,800 facilities, containing over 19,000 generators, who are required to file the Form EIA-860. The respondents to this survey form the basis of the EIA electric power entity frame, from which samples for other surveys are drawn.
- **Form EIA-860M** – The target population for this monthly census comprises power plants within the EIA-860 target population that have either (a) a new generator scheduled to begin commercial operations within the next 12 months, or

¹ Respondents refer to entities in a survey frame.

- (b) an existing generator scheduled for retirement within the next 12 months, or
(c) an existing generator undergoing modifications resulting in changes in capacity or other major modifications that are scheduled to be completed within 1 month. Respondents are the operators of the power plants where these new generators and existing generators are located. Based on the number of plants putting new generators into service in 2008 and 2009, the EIA estimates that in a typical month the Form EIA-860M will be used to collect data from approximately 124 respondent entities.
- **Form EIA-861** – The target population for this annual census comprises participants in the electric power industry involved in the generation, transmission, or distribution of electricity in the United States and its territories. Target population members include electric utilities, wholesale power marketers (registered with the Federal Energy Regulatory Commission), energy service providers (registered with the States), and electric power producers. There are approximately 3,300 entities in the United States involved in the generation, transmission, and distribution of electric energy. This survey serves as the universe from which the sample for the Form EIA-826 is drawn.
 - **Form EIA-923** – The target population for this annual census comprises all electric plants in the United States that are connected to the electric power grid and have a generating capacity of 1 megawatt or greater. While the target population is defined in terms of plants, the respondents for the Form EIA-923 are companies, which report data for the eligible plants they operate. There are approximately 5,573 operating power plants (being reported by 2,800 respondents) for which data will be collected through Form EIA-923. Data will be reported monthly for a sample of approximately 1,781 plants, although this may be adjusted as the data are evaluated. Monthly respondents will report on Schedules 1, 3, 4, and 5, plus Schedule 2 if they have a fossil-fueled capacity of 50 megawatts or greater. At the end of the year, the monthly respondents will report on Schedules 6 and 7, plus Schedule 8, if they have a steam-electric organic-fueled capacity of 10 megawatts or greater. Those respondents who are not in the monthly frame will file annually. They will file Schedules 1, 3, 4, 5, 6, and 7, plus Schedule 2 if they have a fossil-fueled capacity of 50 megawatts or greater and Schedule 8, if they have a steam-electric organic-fueled capacity of 10 megawatts or greater.

B.2. Statistical Methodology

To limit the burden on industry respondents, two of the monthly surveys, the Form EIA-826 and the Form EIA-923, are sent to only a sample of units in the target populations. The samples are *cutoff* samples, i.e., they are basically comprised of all units with measures of size larger than a predefined threshold. This is complicated by compromises due to the need for data on multiple variables of interest. The cutoff sampling eliminates the monthly reporting burden for smaller industry participants. Because smaller units

have, in the past, been responsible for a high percentage of reporting errors, the cutoff sampling may also reduce the levels of non-sampling error affecting the published estimates. (See Knaub (2007, 2008) on cutoff sampling in general, Royall (1970) on model variance, and Knaub (2001) on model bias and variance.) The remainder of this section provides detail on the sampling and estimation methods used for the two sample surveys.

Form EIA-826 Sampling

For the Form EIA-826, the sample is composed of those utilities that typically sell most of the electricity in each category (or end-use sector) in each State. The sample is made up of:

- All investor-owned utilities (IOUs), except for a few small IOUs in Alaska
- All energy service providers
- All Federal utilities
- All entities selling in the public transportation sector
- A sample of the municipal and cooperative utilities.

The frames for Schedule B (energy service providers) and Schedule C (distribution companies) are not always complete, as information from the States on these entities is not always available in a timely manner. In these cases, the two types of respondents are reconciled at the State level and added to the State totals. (Classical ratio estimation can be used for variance estimation. See Knaub (1991), pages 776 and 777, “Incompletely Specified Auxiliary Data.”) A zero-intercept, ratio model (see Royall and Cumberland, 1978) is used to estimate total sales and revenue by end-use sector and State. The sample eliminates the smaller respondents, thus reducing burden and reducing the source of non-sampling errors.

The Form EIA-826 sample design and estimation procedures employ a linear regression model to represent the relationship between the respondent’s annual data value (e.g., sales) from the prior year and the corresponding monthly value for the current month. The prior year’s annual data come from the Form EIA-861. Data values for units not in the sample are estimated from the prior year’s annual data and the estimated parameters of the regression model. Data from sample units for which there is no historical Form EIA-861 data (e.g., units new to the target population) are not used to estimate the relationship between the prior year’s annual value and the current monthly value. The reported current monthly data are, however, used in estimating totals for publication groups. (See Knaub (2002).) If a sample unit’s annual data are deemed reliable, and its Form EIA-826 (monthly) data are considered unreliable, the annual data are used (as for the non-sampled units) to impute the monthly Form EIA-826 data. As mentioned above, a census is performed within the Form EIA-826 for the power marketers or energy service providers (ESP) data, and their totals are added to the estimated (imputed) entities to obtain the estimates for the entire universe.

Form EIA-826 Monthly Sample Selection from the Form EIA-861 Annual Frame

The monthly cutoff sample thresholds for the Form EIA-826 were originally selected based on the criterion of having estimated relative standard error (RSE) values less than 1 percent for all publication groups. The RSE is a percentage measure of the precision of a survey statistic and is used in part as one way to measure sampling error induced by sampling. RSEs are estimated to account for using model-based predicted monthly values in place of missing and non-sampled data for the quantities of interest (revenues, sales, etc.), based on monthly sampled data from the Form EIA-826, and the corresponding annual (Form EIA-861) data for the units not in the monthly sample. Threshold values for the cutoff sampling have been adjusted over time to maintain low RSEs for the published estimates.

Form EIA-923 Sampling

One of the goals of the original Form EIA-923 sample selection process was to reduce the sample size from the separate predecessor Forms EIA-906 and EIA-920. Not only did this reduce respondent burden, but it also allowed the EIA survey staff to focus its resources on a smaller sample to ensure a higher quality of data. A reduction in sample size was deemed especially important in the commercial and industrial sectors due to sometimes questionable data quality and the difficulty in collecting data from many of the smaller facilities. This original goal continues to be an important focus of attention of the ongoing Form EIA-923 sampling strategy (See Douglas (2007)).

The cutoff sampling process for the Form EIA-923 sample is similar to the one described above for the monthly Form EIA-826 sample. Since the original Form EIA-923 sample was established with the 2008 data collection cycle, it is estimated that 261 plants will have been added to the monthly sample to maintain the targeted sample coverage ratios by the inception of the 2011 data collection cycle. These sample additions are deemed necessary as large currently planned and under construction plants become operational. Ongoing sample validation studies may produce other necessary sample additions as needed.

Gross generation was the main focus of the original sample selection process and its high correlation with other data elements on the Form EIA-923 ensured good coverage results for other reported values. Since then, sample validation studies were conducted on fuel consumption, receipts, costs and stocks and the sample was adjusted accordingly.

Sampling parameters are assigned to each sampling stratum. The strata are defined by facility type, energy source, and geographic region. (See "publication groups" in Knaub (1999).) For instance, one stratum is identified as electric utilities burning coal in the South Atlantic Census Division. The types of stratification groups are briefly described below.

Facility Type Classification for Form EIA-923

The four facility type categories comprise seven sectors for which data are collected. These four categories, which correspond to the facility type classifications published in the Electric Power Monthly (EPM), are (1) electric utilities, (2) independent power producers, (3)

commercial facilities, and (4) industrial facilities. Table 8 below shows the seven sectors. (Combined Heat and Power Plant is abbreviated CHP.)

Table 8. Facility Types

Sector Classification Number	Sector Classification Description	Facility Type Stratification Group
1	Regulated Electric Utility	Electric Utilities
2	IPP (Non-CHP)	Independent Power Producers
3	IPP (CHP)	Independent Power Producers
4	Commercial (Non-CHP)	Commercial Facilities
5	Commercial (CHP)	Commercial Facilities
6	Industrial (Non-CHP)	Industrial Facilities
7	Industrial (CHP)	Industrial Facilities

Energy Source Classification for Form EIA-923

The 14 energy source categories, which correspond to the energy source classifications published in the EPM, are aggregations of the 36 different fuel types for which data are collected on the survey. Table 9 gives the 14 energy source categories and the corresponding stratification categories. The energy source codes are defined in the instructions for completing Form EIA-923. (See Appendix C.)

Table 9. Energy Source Aggregations

Reported Energy Source Code	Energy Source Stratification Group
NG	Natural Gas
NUC	Nuclear
HPS ²	Pumped Storage
WAT	Conventional Hydroelectric
PC	Petroleum Coke
GEO	Geothermal
SUN	Solar
WND	Wind
BFG, OG, PG	Other Gas
WDL, WDS, BLQ	Wood
OTH, MSN, TDF, PUR	Other Sources
BIT, LIG, SC, SUB, WC	Coal
RFO, DFO, JF, KER, OO, WO	Petroleum
AB, LFG, MSB, OBG, OBL, OBS, SLW	Waste

² Pumped Storage facilities do not actually report energy source code HPS, rather they report energy source code WAT combined with a prime mover code of PS to differentiate them from conventional hydroelectric facilities. The energy source is renamed to HPS for simplicity sake only.

Geographic Regions Classification for Form EIA-923

The 10 geographic sampling groups correspond to 10 modified Census division regions published in the EPM. The States assigned to each division are shown in Table 10.

Table 10. State/Census Division Aggregations

States	Modified Census Divisions
AK, HI	Pacific Non-Contiguous
NJ, NY, PA	Mid-Atlantic
CA, OR, WA	Pacific Contiguous
AL, KY, MS, TN	East Central
AR, LA, OK, TX	West Central
IL, IN, MI, OH, WI	East North Central
CT, ME, MA, NH, RI, VT	New England
IA, KS, MN, MO, NE, SD, ND	West North Central
AZ, CO, ID, NT, NV, NM, UT, WY	Mountain Region
DE, DC, FL, GA, MD, NC, SC, VA, WV	South Atlantic

Original Sample Selection Criteria for Form EIA-923

The Form EIA-923 sample was chosen to provide reasonably accurate results for multiple attributes (published aggregate numbers) while minimizing the burden on the industry and the Federal government. The following five steps were used in selecting plants for the monthly sample:

1. Select preliminary cutoff samples based on nameplate capacity values
2. Add sample units, where necessary, based on generation, consumption and stocks
3. Add sample units, where necessary, to provide adequate sample counts for estimation groups
4. Add sample units, where necessary, to reduce relative standard errors (RSEs) of key estimates to acceptable levels
5. Add other facilities, based on special-case criteria.

The first three steps were designed to ensure adequate coverage of the target population by including all of the largest contributors to key data elements. The fourth step helps ensure that the published estimates meet reasonable reliability standards, which is the key goal, given acceptable resource expenditure. The final criterion covers special cases, as described below.

Facilities in the target population that meet any one of the sample selection criteria applied at any of the five steps were included in the final sample. Further, any additional prime movers and energy sources used by a sample facility were also included in the sample even if

individually they did not meet any of the sample selection criteria. Each sample facility reports data for all combinations of prime mover and fuel source each month. All nuclear and pumped storage facilities are included in the monthly sample. The remainder of this section provides further detail on the sampling steps.

Step 1: Select Cutoff Samples Based on Nameplate Capacity. Initially, pre-determined capacity coverage percentages were tested to ensure a certain proportion of operational Form EIA-860 capacity is covered within each sampling group. Stand-by and back-up generators were not included in the operational capacity totals when data were aggregated to the level of prime mover, and only the largest consumed fuel source for each generator were used in identifying the sample groupings. Different target coverage percentages were selected for each facility classification, and were applied to all regions and energy sources within each classification. When the capacity cutoff percentage yields a capacity cutoff of less than 25 megawatts, then a default value of 25 megawatts was used instead. Otherwise, the percentages of capacity included in the sample are listed below.

1. Electric utilities – 70 percent
2. Independent power producers – 70 percent
3. Commercial facilities – 50 percent
4. Industrial facilities – 50 percent.

Step 2: Add Units Based on Generation, Consumption, and Stocks. Facilities accounting for large percentages of actual past reported gross generation, fuel consumption, or fuel stocks, were added to the sample, even if their nameplate capacities fell below the capacity coverage percentage cutoff.

Step 3: Add Units to Ensure Adequate Sample in Estimation Groups. Estimation strata identical to those employed in the Form EIA-923 regression imputation system were examined. Units below the threshold value were added to any group with fewer than 10 usable observations, until the usable count was brought up to 10.

Step 4: Add Sample to Meet Reliability Standards. Weighted multiple regressions, identical to those currently employed in the Form EIA-923 imputation system, were run, and relative standard error (RSE or CV) estimates were calculated for each publication group by month. An additional diagnostic measure, the relative standard error for a superpopulation (RSESP) was calculated to indicate the adequacy of the regression model fit. Limits for both measures (RSE and RSESP) were set individually for each facility classification and applied to all energy sources for the U.S. total for each classification.

If one or both of the error measures fell outside of the limits, the next largest facilities, ranked by gross generation, were included until the RSE/RSESP's were brought into a reasonable range. It is important to note that if only the RSESP estimate was out of range, then it was difficult to lower the estimate of RSESP based on sampling alone. In these cases, a change in modeling may be necessary. The RSE/RSESP data quality limits are outlined below.

1. Electric utilities – RSE less than 5 percent and RSESP less than 20 percent
2. Independent power producers – RSE less than 5 percent and RSESP less than 20 percent
3. Commercial facilities – RSE less than 10 percent and RSESP less than 30 percent
4. Industrial facilities – RSE less than 10 percent and RSESP less than 30 percent.

Step 5: Add Special Cases. Finally, additional facilities were added to the sample as necessary. These include storage-only facilities (used in estimating stocks); new facilities for which the EIA has no prior-year’s annual data for use in regression imputation; and any large, easy to survey facilities which the survey staff identified as being desirable in the sample.

Table 11 shows the sample coverage by facility type and Table 12 shows the sample coverage by energy source.

Table 11. Form EIA-923 Sample Coverage by Facility Type

		2011 Sample		
Facility Type	Total Count	Count	Percent By Count	Percent By Volume
Electric Utilities	2,649	854	32	97
Independent Power Producers	2,170	790	36	93
Industrial Facilities	545	112	21	66
Commercial Facilities	209	25	12	52
Total	5,573	1,781	32	94

Table 12. Form EIA-923 Sample Coverage by Energy Source

Energy Source Grouping	2011 Sample Coverage (percent by volume)
Coal	97
Geothermal	67
Hydroelectric	77
Natural Gas	92
Nuclear	100
Other Gas	81
Other Sources	62
Petroleum	92
Petroleum Coke	87
Pumped Storage	100
Solar	95
Waste	43
Wind	80
Wood	65
Total	94

REFERENCES:

The regression estimation/imputation procedures used for the Form EIA-826 and Form EIA-923 are documented and discussed in the on-line statistics journal, *InterStat*, in the following articles:

- “Using Prediction-Oriented Software for Survey Estimation,” at the following URL: <http://interstat.stat.vt.edu/interstat/articles/1999/abstracts/g99001.html-ssi>
- “Using Prediction-Oriented Software for Survey Estimation - Part II: Ratios of Totals,” at the following URL: <http://interstat.stat.vt.edu/interstat/articles/2000/abstracts/u00002.html-ssi>
- “Using Prediction-Oriented Software for Survey Estimation - Part III: Full Scale Study of Variance and Bias,” at the following URL: <http://interstat.stat.vt.edu/interstat/articles/2001/abstracts/u01001.html-ssi>.

The method described in these articles is generally useful for both small area estimation and imputation, with adjustments as described in those documents. Additional documentation and references include:

- (1) “Model-Based Sampling, Inference and Imputation,” available on the EIA Web site at: <http://www.eia.doe.gov/cneaf/electricity/forms/eiawebme.pdf>
- (2) “Weighting in Regression for Use in Survey Methodology,” *InterStat*, available at:

<http://interstat.stat.vt.edu/InterStat/ARTICLES/1997/abstracts/A97001.html--ssi>.

(3) "Some Applications of Model Sampling to Electric Power Data," *ASA Proceedings of the Survey Research Methods Section*, available at:

www.amstat.org/sections/SRMS/proceedings/papers/1991_133.pdf

(4) Royall, R.M., and W.G. Cumberland (1978), "Variance Estimation in Finite Population Sampling," *Journal of the American Statistical Association*, 73, 351-358

(5) "The Classical Ratio Estimator," *InterStat*, available at:

<http://interstat.statjournals.net/YEAR/2005/abstracts/0510004.php>

(6) "Cutoff Sampling and Inference," *InterStat*, available at:

<http://interstat.statjournals.net/YEAR/2007/abstracts/0704006.php>.

(7) "Cutoff vs. Design-Based Sampling and Inference For Establishment Surveys," *InterStat*, available at:

<http://interstat.statjournals.net/YEAR/2008/abstracts/0806005.php?Name=806005>

(8) Douglas, Joel R.(2007), "Model-Based Sampling Methodology for the new EIA-923," Presented to the American Statistical Association and EIA's Joint Meeting on Energy Statistics, October 18, 2007,

http://www.eia.doe.gov/smg/asa_meeting_2007/fall/files/modeleia923.ppt.

B.3. Methods to Maximize Response Rates

For all of the EIA electric power respondents, the response rates are close to or equal to 100 percent. For 2008 annual data, all 11,117 annual respondents (aggregated across all surveys) submitted their data and typically only about 3-7 out of 2,252 monthly 2009 data respondents did not submit their data in any given month. To maximize response rates, the EIA forms have been designed and the instructions have been written to be clear and concise to help the respondent complete the forms. Data that are not expected to change from year-to-year or month-to-month are pre-populated on the forms. Forms and/or notifications are mailed or emailed early to maximize the time that respondents have to complete the surveys. As noted, the EIA Internet Data Collection (IDC) System makes forms available on-line as soon as respondents obtain a secure ID and password. Given the high IDC use rate in 2009 (approximately 95 percent of the monthly reports and an estimated 90 percent of the annual reports), most of those respondents will merely log on in the next data collection period and access their required forms. Form(s) due dates are the same each period so that respondents can schedule their completion activities. The notification and due dates for each survey are provided in Table 6.

The non-respondents are contacted by email, telephone, and letter to request data submission until an insignificant non-response rate is obtained. Follow-up letters and emails citing failure to file the required form are mailed to all non-respondents. If no response occurs as a result of the letters, additional correspondence, requesting

immediate submission of the appropriate data, is sent to the supervisor of the primary contact and, if necessary to higher-level management officials at the non-respondent entity. These letters are sent from the Office Director or, if necessary, from the EIA Administrator. Statistical imputation fills any gaps created by the small amount of non-response.

Respondents who file via the IDC System are given the opportunity to either correct or explain unusual data during their submission. The explanations are reviewed by the EIA staff. Respondents are called if further clarification is needed. For those respondents that do not file via the IDC, but rather on a hardcopy of the form, telephone calls are made to confirm corrections or clarifications of any unusual data.

In addition, the EIA has recently developed an improved centralized frame system which affords all survey staff almost immediate knowledge of changes in plant ownership and/or contacts; such changes contributed to non-response in the past. The new system is integrated with the IDC System so that access can be given to new owners and/or contacts quickly.

B.4. Tests of Procedures

The electric power surveys are established continuing surveys and testing was done at the time they were being established. It is the Electric Power Division's policy to test in several phases. First, the proposed forms are reviewed by internal EIA subject matter and survey methodology experts. The second phase of the testing involves sending draft forms to representatives of the major segments of the electric power industry. Finally, the proposed forms are tested with actual volunteer survey respondents. These respondents are asked to review the forms, and then they are debriefed by EIA to make sure they understand the concepts being measured, can successfully navigate the forms, and have the data in their business records. Changes are made at all stages of testing to incorporate feedback.

B.5. Forms Consultation

During 2009, the Electric Power Division met with a variety of stakeholders to make them aware of the general proposals for form changes and to elicit their suggestions, concerns, and needs. The following is a list of some of the organizations with whom the EIA met.

- American Council for an Energy Efficient Economy
- American Public Power Association
- American Statistic Association
- American Wind Energy Association

- DOE, Office of Electricity Delivery and Energy Reliability
- DOE, Office of Fossil Energy
- Edison Electric Institute
- Electricity Consumers Resource Council
- Electricity Storage Association
- Electric Power Supply Association
- Federal Energy Regulatory Commission
- National Association of Regulatory Utility Commissioners
- National Association of State Utility Consumer Advocates
- National Hydropower Association
- National Mining Association
- National Rural Electric Cooperative Association
- Natural Resources Defense Council
- North American Electric Reliability Corporation
- Ocean Renewable Power Company
- Ozone Transportation Commission
- Platts
- Science and Technology Policy Institute
- Solar Energy Industries Association
- U. S. Environmental Protection Agency.

For additional information concerning these surveys, please contact Rebecca A. Peterson at 202-586-4509 or at rebecca.peterson@eia.doe.gov. For information concerning this request for OMB approval, please contact the agency Clearance Officer, Jason Worrall, at 202-586-6075 or at jason.worrall@eia.doe.gov.