Methods of Detecting Sources of Error and Differences in Energy Data Series

Renee Miller, Energy Information Administration
and
Turkan K. Gardenier, TKG Consultants, Ltd.

The Energy Information Administration (EIA) assesses its principal data series on an on-going basis. The results of these assessments are presented in EIA publications; the most recent is entitled, An Assessment of the Quality of Principal Data Series of the Energy Information Administration.[1] These reports 1) describe and document what is known about a data series to avoid misinterpretation and misuse, 2) compare EIA estimates with other series that may present similar information to explain differences between the sources if any are found, and 3) document any discontinuities in a series. For the 1982 assessment [1], we obtained information on the accuracy of a portion of EIA estimates of volumes of petroleum products and natural gas, and estimates of prices of petroleum products by examining both monthly and annual data. The time period covered for monthly estimates was from 1977 to mid-1981. For annual estimates, the time period covered was from 1977 to 1980.

To assess accuracy we used several statistical techniques, e.g., regression, nonparametric tests, and Cumulative Sum charts that are generally used in other contexts. In addition, we used a variety of graphs. In this paper we present examples of how these techniques and graphs can be used to assess data series.

Assessing Accuracy

Accuracy, as used in our assessments, denotes the closeness of a data series to the quantity intended to be measured, i.e., the true value. If the true value were known or if a reliable benchmark were available, accuracy could be quantified easily. In the absence of such information, an alternative is to examine the systematic and random components of error.

The surveys on which our reference estimates (the series that we have chosen to assess) are based are complete enumerations, so there is no sampling error. We therefore checked for the existence of nonsampling error, for example, selection error, nonresponse and processing error.

Nonsampling error was examined by 1) reviewing validation studies that were conducted in the petroleum and natural gas areas, 2) using graphs to identify outlying observations, and 3) performing consistency checks.

Further research was conducted to determine whether these outlying observations were the result of nonsampling error. With respect to the consistency checks, following is an example in the price area.

Wholesale prices should be lower than dealer tankwagon prices, which in turn should be lower than retail prices. We inspected the data to determine whether that was, in fact, the case. If the data were not in the expected order, it could be an indication of an error.

In addition to checking for the existence of nonsampling error, we compared the reference estimate with any similar estimate. Comparisons do not address the issue of accuracy directly in the sense that when there are differences the question of which is the accurate series remains. However, the approach raised questions about specific features of each data collection system. Also, if two independent series are found to correspond closely, it increases confidence in the accuracy of each. Furthermore, comparisons can indicate the range of estimates from different measurement approaches. In some cases, the various estimates bound the target values. Comparisons can also indicate when the relationship among series has changed.

Comparisons were made through the use of descriptive statistics, by fitting regression lines to data from different series and by computing correlation coefficients. We used regression and correlation to get an overall measure of the correspondence between the series. In addition, we used cumulative sums (CUSUMS) and nonparametric tests such as the sign, rank and runs tests. These methods were used for demonstration purposes in the 1982 assessment. Prices of low-sulfur residual fuel oil from the FPC-423 ("Monthly Report of Cost and Quality of Fuels for Electric Plants") and the EIA-460 ("Petroleum Industry Monthly Report for Product Prices") surveys were used as sample data.

Use of Validation Studies

In examining annual data on estimates of additions to and withdrawals from storage for natural gas, we noted some discrepancies between the reference and comparative estimates (Table 1). The reference estimate is based on the "Supply and Disposition of Natural Gas" survey (Form

Table 1. Estimates of Underground Storage of Natural Gas in the United States, 1977-1980 (Billion cubic feet, 14.73 psia at 60 degrees Fahrenheit)

Estimate and Source	1980	1979	1978	1977
Addition to Storage				
Reference Estimate	1,896	2,295	2,278	2,307
EIA-191/FPC-8 Estimate	2,048	2,361	2,328	2,396
AGA Estimate	2,057	2,285	2,271	2,303
Withdrawals from Storage				
Reference Estimate	1,910	2,047	2,158	1,750
EIA-191/FPC-8 Estimate	2,087	2,036	2,180	1,775
AGA Estimate	2,114	2,057	2,151	1,736

EIA-176). The first report year for that form was 1980 and that was the only year for which data were available when we were doing our assessment. In previous years EIA used data from the American Gas Association (AGA) to develop estimates. EIA adjusted the AGA estimates before publication, which explains why the pre-1980 AGA and EIA estimates are similar but not identical. In addition to the AGA estimates, there were comparative estimates available from another set of EIA surveys, Form EIA-191 and FPC Form 8, which are called the Underground Natural Gas Storage Report. As shown in Table 1, the 1980 reference estimates are considerably lower than the comparative estimates. In addition, the reference estimate showed a decrease in withdrawals between 1979 and 1980 while the comparative estimates showed an increase.

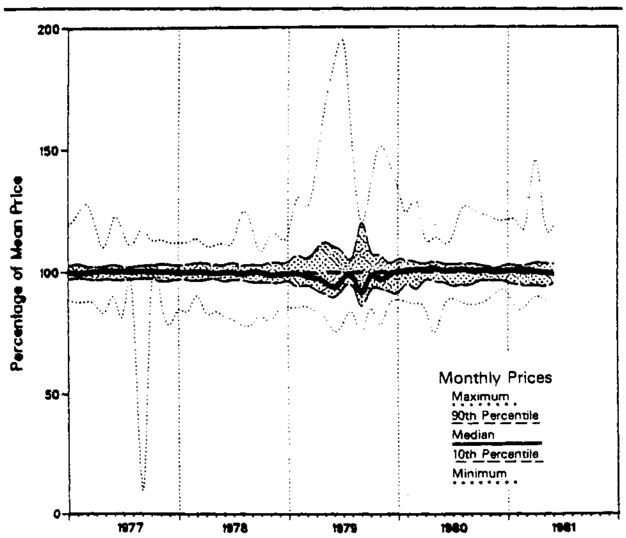
Validation studies were conducted for both the EIA reference and comparative surveys. These studies noted that there were reporting or processing errors in both surveys, and so it was difficult to pinpoint the exact source of the discrepancy between the reference and comparative estimates. In the case of the reference survey, the validation study noted several types of errors made by respondents. The most important were misreporting of state border transactions and the confusion between equity and custody. Respondents are required to report on a custody rather than an equity (ownership) basis on Form EIA-176. However, not all respondents did so. The study noted a source of the problem. Accounting departments keep their records on an equity basis while the gas control departments keep records on a custody basis. These departments may be in separate locations. If the form is filed in the accounting department, it is often difficult for a respondent to provide the data on a custody basis.

In the case of the Underground Natural Gas Storage Report (the EIA comparative estimate), respondents are required to file biweekly during the heating season and monthly during the remainder of the year. Because of the short reporting period during the heating season, respondents frequently submit estimates and then file revisions. Problems have been found in processing the revisions. On a more positive note, the figures for additions to and withdrawals from underground storage are closer for report year 1981 for the reference and comparative estimates than for report year 1980. In addition, Form EIA-176 has been revised for report year 1982, and the instructions have been expanded to clarify and emphasize the reporting on a custody basis.

Graphs

We plotted maximum and minimum and 10th and 90th percentile values as a percentage of the mean for several time periods to determine whether the relationship had changed. A change could be indicative of an aberration. Consider the graph shown in Figure 1 for wholesale no-lead gasoline as an example. Note that the shape of the distribution changed in September 1977. The minimum value appears too low by a factor of ten, probably the result of a slipped decimal. Correcting this value increased the average by only 0.08 percent. As the June and July 1979 maximum wholesale no-lead prices were about 200 percent of the weighted average prices, we initially thought these values were processing errors. However, we found that these values did, in fact, correspond to what the respondents reported. For some of the EIA-460 prices, the highest prices were obtained from respondents that were cited for price overcharges. In addition, we found that there was consistency between monthly values reported by a particular respondent; i.e., the lowest prices are shared among only a few respondents and the highest prices are shared among another group of respondents.

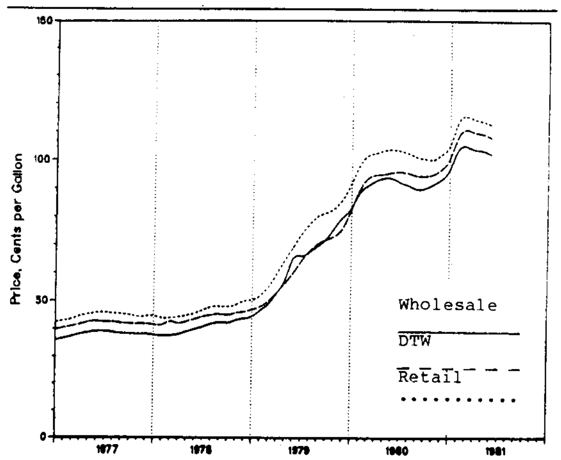
Figure 1. Variations in Company-Based Monthly Wholesale Prices of No-Lead Motor Gasoline, January 1977-June 1981 (Updated File of EIA-460 Submissions)



Consistency Checks

Inspection of motor gasoline data on wholesale, dealer tankwagon (DTW) and retail prices from the EIA-460 survey showed that the prices by grade were generally in the expected order. Unexpected relationships occurred when prices were increasing rapidly in 1979. For several months in 1979, DTW prices were lower than wholesale prices for regular gasoline. In addition, DTW prices were lower than wholesale prices for no-lead gasoline for two months in 1979. Prices by type of sales for regular gasoline are shown in Figure 2 as an example.

Figure 2. Average Monthly Prices, by Type of Sale, of Regular Motor Gasoline, January 1977-June 1981 (Monthly Petroleum Products Price Report)



Regression

We analyzed monthly data by using ordinary linear least squares regression:

$$\hat{Y}_i = a + bX_i + e_i$$

where

$\hat{Y}_i$ = estimated value of Y_i based on the regression equation,

Y_i = estimate of specific process, (for example, reference estimate of volume of motor gasoline),

X_i = alternate estimate of the process, (for example, FHWA estimate of volume of motor gasoline),

e_i = imprecision or unreliability in the measurement,

a,b = two constants in the linear regression line, determined by the method of least squares.

A separate regression line was fit for each comparative estimate. We used the comparative estimates as the independent variables and the reference estimate as the dependent variable. A summary of the results for motor gasoline is presented in Table 2. The reference estimate in this example was from the Joint Petroleum Reporting System (JPRS). We were particularly interested in the magnitude of the "a" and "b" coefficients. The "a" coefficient represents a relatively constant bias in estimation throughout the range of observations. That is, one of the measurements may be consistently lower or higher than the other.

The correlation coefficients between the reference estimate and the comparison estimate exceeded 0.88. A standardized residual of greater than 2 was observed for November 1980 in the Federal Highway Administration (FHWA) comparison. By 1980, the gap between the FHWA and EIA reference estimates had widened. A study of gasoline data systems concluded that the EIA reference estimate missed certain secondary sources of gasoline supply. Beginning in January 1981, blending stations were added as respondents to Form EIA-87 (EIA reference estimate) and definitions were changed to reflect better the flow of products at refineries.

The Durbin-Watson (D-W) test statistic was computed to determine if the residuals correlated over time. The D-W test statistic for the FHWA and American Petroleum Institute (API) residuals was slightly over 2, indicating that the residuals are uncorrelated, while the D-W test statistics for EIA-25 and the P-306 residuals were 0.546 and 0.848 respectively, indicating correlation over time ($\alpha = 0.05$, two-sided test, 54 observations). A cyclical pattern is also evident in the residuals for the P-306 regression, showing that the regression fit is not constant over time.

Despite the problem of time dependence, we found regression and correlation to be appealing methods because they provide an overall measure of correspondence between the series. It would be preferable to submit first order differences between present and previous values to correlation and regression analyses rather than the datapoints themselves. When we re-analyzed the EIA-25 monthly data in this way, the problem of serial correlation was eliminated. The correlation coefficient in this situation was only slightly lower than the one shown on Table 2 (0.87 compared to 0.89).

Table 2. Summary Results of Regression Analysis for Motor Gasoline Using JPRS Volume Estimates (Thousand of Barrels/Day) As Dependent Variables

Independent Variables	Constant Term (a)	Slope Coeff. (b)	Standard Error of (b)	Correlation Coeff. (r)	Standard Error of Regression (As percent of JPRS mean)	Months With Standard Residual >2, <2	Durbin Watson Test
EIA-25	1,266	0.784	0.057	0.887	2.8		0.546
FHWA	497	0.896	0.052	0.923	2.3	11/80	2.193
API (Old Basis)	712	0.899	0.051	0.925	2.3		2.375
P-306	1,114	0.814	0.045	0.929	2.3		0.848

Nonparametric Tests

From one point of view, the most desirable tests for randomness of difference and existence of trend in the differences are those that do not rely on prior specification of the distribution of the series. These nonparametric tests are especially important for energy data because there have been dramatic changes in the shape of the price distributions, particularly when prices are changing rapidly. We used the sign test to test for trends in the difference between two price series: estimates of low-sulfur No. 6 residual fuel oil prices from the EIA-460 and FPC-423 surveys. These series are not strictly comparable, however, because the FPC-423 survey collects data on the cost of receipts from electric utilities only, whereas the EIA-460 survey collects data on the retail selling prices from refiners, large resellers and retailers.

The idea behind the sign test for trend is to pair earlier observations with later observations and to test whether the first member of the resulting pairs tends to be larger (smaller) than the second member. If there are 2N observations, the procedure is to split the series in half and to subtract the N + i^{th} observation from the i^{th}, i ranging from 1 to N. The sign of the difference is recorded and the Binomial test is then applied to this series.

The pairs formed from the residual fuel oil price differences are shown in Table 3. There are nine negative and six positive signs. This is not sufficiently different from the expectation under an equally likely random assignment of signs to indicate a trend at the 10-percent level.

While the sign test is easy to apply, it is not particularly powerful. It more frequently accepts the maintained hypothesis of no trend when there is a trend than do competing tests. The rank test, which is based on Spearman's rank correlation coefficient, is more powerful than the sign test but requires more computation and is less generally applicable.

The rank test operates on the rank order of the data. An increasing trend should produce large values late in the series and a decreasing trend should produce large values early. The rank of the i^{th} observation, R(i), is subtracted from its chronological order i and the result is squared. When two or more observations are the same, each is assigned the average rank. The smallest observation would have rank 1 and the largest would have rank N. The test statistic is based on the sum of the squares. If the differences between two data series increased monotonically throughout the time period, the sum of squares would be zero. If they decreased throughout, the sum of squares would achieve its maximum value of $(1/3)N(N^2-1)$. Thus, very small or very large values indicate a trend. When there are no ties in the ranking this amounts to using Spearman's rho correlation coefficient as the test statistic.

The ranking of the price data differences is shown in Table 4. The sum of the squares of 30 price differences is 3,236. According to tables in Conover [2], this value is neither large enough nor small enough to suggest a trend at the 10-percent significance level.

Table 3. Illustration of the Sign Test for Trend

First 15 Observations	Second 15 Observations	Sign
-0.41	0.83	-
-1.00	-2.29	+
-1.02	-1.02	+
-1.65	0.71	-
-1.73	-0.16	-
-1.42	0.47	-
-3.40	1.97	-
-1.97	-0.47	-
-1.42	-1.43	+
-0.91	-1.34	+
-2.23	-1.21	-
-1.87	-2.05	+
-1.79	-0.35	-
-0.72	0.71	-
0.23	-0.63	+

Table 4. Illustration of the Rank Test for Trend

Price Difference	Magnitude Order[1] R_i	Time Order i	$R_i - i$	$(R_i-i)^2$
-0.41	23.00	1	22.00	484.00
-1.00	18.00	2	16.00	256.00
-1.02	17.00	3	14.00	196.00
-1.65	9.00	4	5.00	25.00
-1.73	8.00	5	3.00	9.00
-1.42	12.50	6	6.50	42.25
-3.40	1.00	7	-6.00	36.00
-1.97	5.00	8	-3.00	9.00
-1.42	12.50	9	3.50	12.25
-0.91	19.00	10	9.00	81.00
-2.23	3.00	11	-8.00	64.00
-1.87	6.00	12	-6.00	36.00
-1.79	7.00	13	-6.00	36.00
-0.72	20.00	14	6.00	36.00
0.23	25.00	15	10.00	100.00
0.83	29.00	16	13.00	169.00
-2.29	2.00	17	-15.00	225.00
-1.62	10.00	18	-8.00	64.00
0.71	27.50	19	8.50	72.25
-0.16	24.00	20	4.00	16.00
0.47	26.00	21	5.00	25.00
1.97	30.00	22	8.00	64.00
-0.47	22.00	23	-1.00	1.00
-1.43	11.00	24	-13.00	169.00
-1.34	15.00	25	-10.00	100.00
-1.19	16.00	26	-10.00	100.00
-2.05	4.00	27	-23.00	529.00
-1.35	14.00	28	-14.00	196.00
0.71	27.50	29	-1.50	2.25
-0.63	21.00	30	-9.00	81.00

[1]Note that the tied observations (numbers 19 and 29) are assigned the average rank.

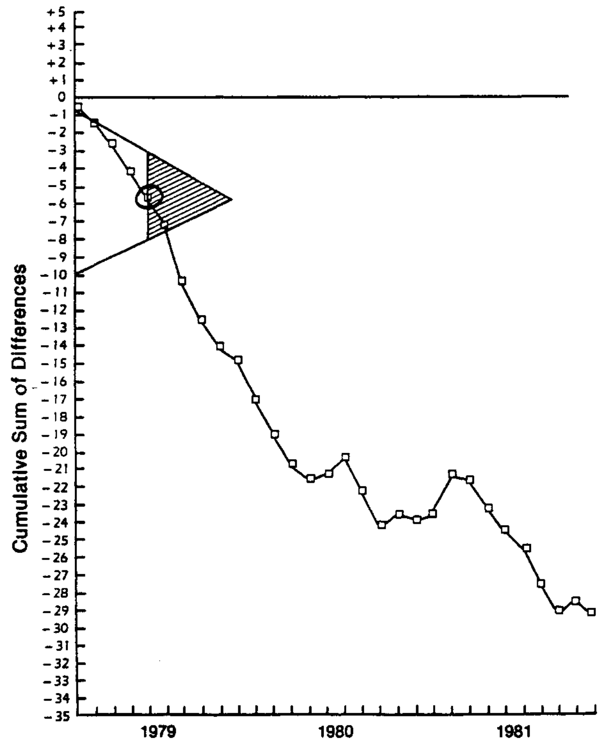
Cumulative Sum (CUSUM) Charts

Cumulative Sums (CUSUMs) have been particularly successful in industry in detecting process drift due to shifts in calibration, tool wear and similar forms of deterioration. Analogous problems in energy data collection, such as frame deterioration and unanticipated supplies from outside the sample universe, may be detectable by CUSUM techniques.

The CUSUM is based on ideas put forward by E.S Page in 1954. It is well adapted to detecting abrupt changes in a parameter value (mean, proportion defective). As described in [3], the CUSUM control chart can be regarded as a sequential sampling procedure "in reverse." We used CUSUM charts to detect changes in relationships between price estimates of low-sulfur No. 6 residual fuel oil from the EIA-460 and FPC-423 surveys (Figure 3). CUSUMS are generally calculated by taking the difference between an observation and a target value. In our situation, we considered the difference between the two series to be our observation with zero being the target value. Figure 3, therefore, shows the cumulative sum of the differences between the two series.

To determine whether the process is out of control, a chart called a V-mask (horizontal V) is superimposed successively over the most recent cumulative sum. If either of the legs of the V-mask is cut by the CUSUM, then the process is out of control. In Figure 3, the V-mask is superimposed at the observation at which the process is first found out of control (marked with a circle). Construction of the V-mask requires explicit decisions on acceptable levels of both false alarms and undetected process changes. It also uses criteria relating to possible bias, i.e., expected magnitude of the change in process.

Figure 3. Cumulative Sum of Differences Between FPC-423 and EIA-460 Surveys of Monthly Prices of No. 6 Residual Fuel Oil (Less Than 0.3 Percent Sulfur) January 1979 and June 1981



V-mask is superimposed on the observations successively. In the diagram it is on the observation marked with a circle which is the point where the process is first found out of control because the legs intersect the trace of the cumulative sums.

In our example, the α - level, the probability of a false alarm was set at approximately 0.1.

The parameters necessary to construct the V-mask, Θ and d can be obtained from the following relationships specified in [3].

$$\tan \Theta = \frac{1}{2} g$$

$$d = \frac{2}{-g^2} \ln \alpha$$

where

d is the distance from the value being evaluated to the apex of the "V,"

Θ is half of the angle formed by the "V," and g is the change in the parameter that we want the CUSUM to detect with fair certainty (we specify g). In this example, g = 1.

An uncertainty parameter, denoted as k, can also be built into the CUSUM computations. This parameter, subtracted from the ongoing Cumulative Sum at each observation point, absorbs random fluctuations.[4]

Some Limitations of Our Approaches

First, fully comparable systems seldom exist. Second, the comparative series may not have been validated. Therefore, discrepancies between series do not necessarily represent inadequacies in a series, but rather the need for further research to explain and/or resolve the differences.

Some of our methods (e.g., regression and CUSUMS) assume independence in either the residuals or the successive measurements. Data indexed over time do not always exhibit i.i.d. (independently and identically distributed) characteristics. Johnson and Bagshaw [5] found that the time required to detect a change using the CUSUM test is shorter with positive (+) first order autocorrelation, i.e., dependence between successive observations. Time needed to detect a change is longer with negative (-) first order autocorrelation. We have observed a positive first order autocorrelation in several of the series. In addition, there are questions about how robust CUSUMs are to misestimates of the standard deviation. Nevertheless, CUSUMs have proven to be very useful in a variety of quality control applications.

Recommendations for Further Research

There are primarily two areas that we believe need to be explored further. The first is the way in which monitoring of differences between energy data series and testing for statistical significance are conducted in the presence of time dependence or autocorrelation within each series. The second area that calls for further research deals with methods to compare more than two data series.

Correction for autocorrelation in applications of the Cumulative Sum charts may be accomplished by first fitting a time-series model, such as Box-Jenkins [6], to reduce or eliminate the autocorrelation effect, then by using CUSUM charts on the residuals from the model [7]. Box-Jenkins models use a set of linear filters, autoregressive (p), differencing (d) and moving average (q) which transform the data series so that successive data are not autocorrelated. We have demonstrated the applicability of time-series methods in [1]; their use for monitoring with CUSUMS is a future recommended application.

In our applications so far we have dealt with comparing only two series with each other. In reality, for most energy data there are more than two sources, within or external to EIA. Comparing more than two series demands an approach to isolating concordance in the multiple time series data. Methods associated with analysis of variance (ANOVA), for example, may be explored to address this issue.

Summary

In summary, we assessed a portion of EIA's estimates of petroleum volumes, petroleum prices and natural gas volumes by examining sources of nonsampling error and by comparisons with other series. In addition, we investigated statistical techniques that could be used for comparing past energy series and for monitoring ongoing series. These techniques include regression, nonparametric tests, and cumulative sum charts. We also presented suggestions and recommendations for further research and applications of these methods to energy data series.

References

1. Energy Information Administration, Office of Statistical Standards, An Assessment of the Quality of Principal Data Series of the Energy Information Administration, Washington, D.C., June 1983.

2. Conover, W. J., Practical Nonparametric Statistics. 2nd Ed., New York: Wiley, 1980.

3. Johnson, N.L., and Leone, F.C., "Cumulative Sum Control Charts, Mathematical Principles Applied to Their Construction and Use: Part I." Industrial Quality Control 18 (12), June 1962.

4. Johnson, N. L., "A Simple Theoretical Approach to Cumulative Sum Control Charts." J. Amer. Stat. Assn., 56, (1961), 835-840.

5. Johnson, R. A., and Bagshaw, M., "The Effect of Serial Correlation on the Performance of CUSUM Tests." Technometrics 16(1), (1975), 103-112.

6. Box, G. E. P., and Jenkins, G. M., Time Series Analysis Forecasting and Control. San Francisco: Holden Day, 1976.

7. Gardenier, T. G., "Moving Averages in Environmental Monitoring." Simulation, 39(2), (1982), 50-58.