

AN APPROACH TO EVALUATING THE  
ACCURACY OF ENERGY DATA SERIES

Arthur R. Silverberg,  
Food and Drug Administration

INTRODUCTION

This paper describes an approach used in assessing the quality of the Energy Information Administration's (EIA) principal motor gasoline data series. The approach combines two general techniques: an internal assessment, and external comparisons.

We define an internal assessment of a data series to be an evaluation of the design, collection, processing, and publishing of a data series. An internal assessment will attempt to describe and quantify the various components of error in the data system under investigation. The internal assessment tends to be exploratory in nature. The final summary of an internal assessment is an error profile. No attempt will be made to "add up" the individual components of error to arrive at the total survey error.

An external comparison examines the reference data series with other comparable estimates. Comparisons between two or more independently derived estimates are useful in that a difference between estimates of the same quantity may be due to an error in one of the data sources although the comparison itself does not tell us which series is in error. It is also possible that what we believed to be comparable estimates are not. The possibility that the comparison may be at fault is not necessarily unfortunate in that when we find a difference between two "comparable" data series, we are forced to study the phenomenon as well as the data series in greater depth. Comparative analysis tends to be confirmatory in nature.

DESCRIPTION OF THE MOTOR GASOLINE INDUSTRY

The supply of motor gasoline is comprised of marketed production from refineries, petro-chemical plants, gas plants, plus imports. Petro-chemical plants produce motor gasoline by chemically re-processing oils that have been produced by either gas plants or refineries. Refineries, gas plants, petro-chemical plants and importers are the suppliers of motor gas. Once an oil has been produced (not necessarily motor gasoline), the oil may be transported by pipeline, barge, ship, truck, or other means of conveyance. This oil may be re-processed into motor gasoline at a refinery or at a petro-chemical plant. Oils may also be stored at a bulk terminal. Oils can also be mechanically blended into motor gasoline at a blending station. Blending stations differ from petro-chemical plants in that no chemical re-processing is performed at a blending station.

After the motor gasoline is produced or blended, the gasoline may be exported. Exports account for only a very small amount of the total amount of motor gasoline produced in the

United States. During 1980, exports accounted for only .02 percent<sup>1</sup> of the supplied for domestic use value.

Motor gasoline may be sold to wholesalers and jobbers. Wholesalers and jobbers in turn sell the motor gasoline to retail gasoline stations, and other wholesale dealers. Wholesalers and jobbers are not the only types of enterprises that may sell to retailers. Any of the suppliers of motor gasoline may sell directly to retailers. Retailers include gasoline stations, other retailers, and such enterprises as convenience stores and department stores.

Motor gasoline is consumed by bulk customers and retail customers who buy at retail stations. Examples of a bulk consumer include a farmer with a large tank on his farm, or a fleet owner such as a rent-a-car dealer. Bulk consumers may obtain their motor gasoline from wholesalers or jobbers, or any of the suppliers of motor gasoline.

TAXES

When considering retail sales of gasoline, it makes sense to consider the retail price including taxes since this is the price that the consumer faces. On the other hand when considering wholesale or dealer-tank-wagon prices, the price excluding tax is considered, since all taxes have not yet been paid on motor gasoline at this level.

The Federal Government imposes a 4 cents per gallon "manufactures excise tax" on motor gasoline. This tax is imposed on the "manufacturer" of the motor gasoline. Commonly, the refiner, gas plant or petro-chemical plant is considered the manufacturer. The manufacturer will then sell the motor gasoline for a price including Federal Excise tax. The consumer ultimately pays the tax. The consumer reimburses the retailer, who has reimbursed the wholesaler, who has reimbursed the refiner/manufacturer.

All Fifty States, the District of Columbia and Puerto Rico have their own state excise tax on motor gasoline. Licensed distributors pay the excise tax to the states. The licensed distributors are then reimbursed ultimately by the consumer in the same manner as the Federal Excise Tax.

State and local sales taxes may also be imposed. Several, but not all states have a sales tax. This tax may apply to the entire state, or just several counties as in California.

EIA GASOLINE PRICE DATA SERIES

The principal EIA data collection form for motor gasoline is the "Petroleum Industry Monthly Report for Product Prices," Form EIA-460. This form collects respondent's volume sold during the reporting month as well as the respondent's weighted average selling price. Form EIA-460 collects motor gasoline data by grade, premium, regular and no-lead, and by type of transaction, wholesale, dealer-tank-wagon, and retail. (Wholesale

sales differ from dealer-tank-wagon sales in that wholesale sales are sales to suppliers whereas dealer-tank-wagon sales are sales to retail dealers.) Therefore, there are nine grade-by-type distinctions that are collected on the EIA-460. All data are collected excluding taxes, and therefore in many cases adjustments for taxes will be required before external comparisons can be made. Before September 1980, the EIA-460 was known as the FEA-P302-M-1. There is no difference between these two forms as far as motor gasoline is concerned.

The EIA-460 attempts to be a census of all refiners, and resellers having more than \$50,000,000 in annual sales or revenues. Therefore, there is no sampling error.

#### INTERNAL ASSESSMENT

The first question that should be asked is, "What needs to be estimated?" The purpose of Form EIA-460 is to collect price estimates at the refinery level so that the EIA may monitor prices, and perform timely policy and economic analyses. The EIA publishes a national monthly volume weighted average price. No measure of price spread is given.

Exhibits 1 and 2 show the spread by company of reported volume weighted prices for wholesale regular and wholesale unleaded motor gasoline respectively, as a percentage of the national reported volume weighted price on a monthly basis. The top and bottom lines show respectively the maximum and minimum reported average prices. The cross-hatched areas correspond to the inter-decile range. Remember, each firm reports only one volume weighted average price. These Exhibits show the price spread for company average selling prices. The true price spread would be wider. Of these two Exhibits, only Exhibit 2 shows the extremely large price spread during the 1979 Iranian crisis.

Specification error is the difference between the quantity intended to be estimated, and the quantity that the respondents are asked to report. Specification error can be caused by forms and instructions that do not reflect the realities of the phenomena being estimated, or the form and instructions do not ask the correct question.

The grades of motor gasoline used on Form EIA-460 are not consistent with the current grades being sold. The common grades of motor gasoline are premium unleaded, premium leaded, regular unleaded, and regular leaded. The EIA-460 uses the grades premium, regular, and no-lead. It is not clear from the directions to the form as to where premium unleaded should be reported. Based on EIA's Residential Consumption Survey, approximately 41 percent of all motor gasoline sold to households is unleaded regular, 4 percent unleaded premium, 50 percent leaded regular, and 4 percent leaded premium, in 1980.

If we wish to estimate refiner-marketer retail prices, there is no other specification error. There is specification error if we wish to estimate retail prices without regard to the

type of marketer, since independent-marketers are not covered in the EIA-460 frame, and do not have the same average selling price as refiner-marketers. In 1980, annual average retail price charged by refiner-marketers was 117.3 cents per gallon, whereas for independent-marketers the average retail price charged was 120.1 cents per gallon.<sup>2</sup> Refiner-marketers accounted for only 19.2 percent of retail sales.<sup>2</sup>

Selection error is caused by collecting data from frames having duplicates, or from incomplete frames. The intended universe for the EIA-460 motor gasoline prices are refiners and large resellers having over \$50,000,000 or more in annual sales or revenues. As part of a validation study of the EIA-460 data collection system, Transportation and Economic Research Associates, Inc. (TERA) performed an extensive analysis of frame undercoverage.<sup>3</sup> TERA uncovered 30 missing refiners, accounting for approximately 1.5 percent of total U.S. refining capacity. No reliable estimates of the percentage of undercoverage for large resellers are available.

In order to bound the selection error, TERA performed three scenarios. The selling price for the firms that were not in the EIA-460 frame were assumed to be selling at either the highest or lowest selling price of the reporting firms in the missing firms stratum (refiner or reseller). Computations were performed for December 1980 prices of dealer-tank-wagon regular motor gasoline, wholesale unleaded motor gasoline, and retail unleaded motor gasoline. The selection error was bounded at 0.56 percent in the case of wholesale unleaded, 0.18 percent for dealer-tank-wagon regular, and 0.00 percent for retail unleaded.

As part of the EIA-460 validation, the accounting firm of Alexander Grant & Company audited the records of 49 EIA-460 respondents. Of all the audited firms, 26.5 percent (not volume weighted) did not report accurate volume weighted prices. In addition, the auditors were unable to trace the price of 12.3 percent (not volume weighted) of the audited enterprises.

Despite these large percentages of inaccurate or untraceable data, the national volume weighted price was found to be quite accurate. Respondent errors were classified as to either computation errors, transcription errors, timing errors, misinterpretation of instructions, or different accounting conventions. Data for December 1980 for wholesale unleaded gasoline, dealer-tank-wagon regular, and regular unleaded gasoline were audited. The only error type that affected the overall national volume weighted price by more than 0.005 percent is misinterpretation of instructions for unleaded retail gasoline. The error of this type is 0.03 percent, but the relative standard error for the estimate is 0.02 percent. This is despite the fact that about 80 percent of the audited enterprises included out-of-period adjustments in their sales transaction data.

Unit non-response is the failure to obtain information, for any reason, from entities in a sample survey. If no estimates for the non-respondents are made, implicitly a zero is assumed for the non-respondents value, and a population total would be low biased. If an estimation procedure is used to account for the non-response, this may introduce bias. Estimates of averages may also be biased since respondents and non-respondents as a group may have differing characteristics. When a respondent omits one or more items on a survey this is item non-response. Item non-response may cause a bias in the data just as unit non-response.

The estimation of non-response is quite difficult in the case of the EIA-460. Enterprises changing their names, enterprises merging, and enterprises going out of business make an estimate of this error difficult since EIA does not archive the list of those enterprises receiving the EIA-460 each month. When price controls were in effect, files were kept on each enterprise stating which months the enterprise did not file. No files were kept by month. The TERA study discovered that the EIA April 1981 name/address list did not contain 11 refiners and 5 resellers who did actually respond during 1980. Three enterprises listed on the April 1981 name/address list did not respond during 1980. TERA estimates the frame for the EIA-460 contained 186 refiners, and 77 large resellers.

The project manager for the EIA-460 does not believe that non-response is a major problem for the EIA-460. He estimates that 95 percent of all enterprises report before the revised values are published. This 95 percent of all enterprises accounts for about 99 percent of the total volume. All 15 major refiners must report before the final average price is published. Implicitly a zero volume is imputed for non-respondents.

In order to obtain a feeling of how non-response might affect the estimate of volume weighted price, we have plotted volume versus price for each grade of motor gasoline for all three types of sales (retail, dealer-tank-wagon, and wholesale), for January 1977 through June 1981. We have reproduced only two plots in this report. Exhibits 3 and 4 are scatter plots for regular wholesale motor gasoline for June 1977 and June 1979 respectively. All respondents having over 4 percent of the total reported volume are placed at the 5 percent line to protect confidentiality. No respondent reported more than 6 percent of the total reported volume.

Non-response should have little effect in 1977 where the prices tend to cluster around the average price. If the large respondents report, the lack of a response from an enterprise having small volumes should make little difference. A non-respondent can make a large difference in 1979 where the distribution of prices is bi-modal. Large respondents report at two quite different price ranges, one price range about 90 percent of the average price, and one range about 150 percent of the average price. Since the 15 largest refiners

must report before the average price is published, we feel that non-response error is small for volume weighted average price, less than 1 percent.

Inadvertent processing error refers to errors introduced by EIA into the estimates by losing submissions, incorrectly transcribing or coding the data, or making arithmetic errors. This type of non-sampling error is usually small as compared to other types of non-sampling errors.

To check for incorrect transcription errors, the TERA validation of the EIA-460 compared 49 hardcopy December 1980 enterprise submissions with the EIA computer file as of May 1981. A total of 343 data elements were compared. Three differences were found, corresponding to either a revision or a late submission by the enterprise.

TERA also performed two deterministic edit checks. First the volume reported for total motor gasoline was compared to the sum of the volumes reported for the various categories of motor gasoline. Only 0.02 percent of the January-December 1980 data failed this simple test. Also, the reported volume weighted price for total motor gasoline was compared to the volume weighted price of each each of the categories of motor gasoline. TERA found that 0.55 percent of all submissions failed this edit check.

Additionally, we have performed the same test on the archived data and have found similar results. Also we have found that one enterprise reports a volume but no price. This enterprise claims that it does not know the volume weighted selling price of its motor gasoline. The only out-of-range product codes found on the EIA-460 computer database have zero volume and no price.

The September 1977 wholesale no-lead minimum price is too low by a factor of ten, likely a slipped decimal point. Correcting this value increases the volume weighted average price by only 0.08 percent. The June and July 1979 maximum values are suspect. For wholesale regular we see quite a few very low values in 1977 and one in September 1980. There is a consistency between monthly values reported by a particular enterprise. The lowest few prices are shared among only a few enterprises. The highest few prices are shared among another group of enterprises. For some of the EIA-460 product prices, we have noticed that the highest reported prices are from enterprises that have been cited for price overcharges. The high values for 1979 are also likely to be correct since the 90th percentile is much greater than the average price.

To verify the computation that produces the published volume weighted average price, we developed our own program and ran it against the EIA-460 database as of February 10, 1982. The EIA-460 database is continuously updated when late submissions or corrections arrive. Usually, our computed price when rounded to the nearest one-tenth of a cent agrees with the published price. For wholesale regular motor gasoline, there is only one large difference between the two estimates. In September 1979,

our computed price exceeds the published price by four cents per gallon. Upon looking at the data file, we discovered that the enterprise having the highest price, has a volume associated with it that was too large by a factor of 1,000. When we recomputed this price we found that the recomputed price differed from the published price by only 0.4 cents per gallon. The only other large difference that was found between the published price and our computed price was for the sale of wholesale no-lead in September 1979. Here, the same respondent again reported a volume too large by a factor of 1,000.

No judgemental adjustments are made on the EIA-460 data to produce EIA publication, "Monthly Petroleum Product Price Report." The EIA-460 data are collected excluding taxes and the national price estimates are published excluding taxes. The EIA-460 retail gasoline prices are published with taxes in publication, "Petroleum Market Shares, Report on Sales of Retail Gasoline." Since the data are collected excluding taxes, EIA must make an estimate as to the tax rate.

The computer program that calculates the estimates for the "Petroleum Market Shares, Report on Sales of Retail Gasoline" adds a constant value of 13.8 cents per gallon to the reported volume weighted average price. This computer program was written in December 1978 and there is no documentation that indicates that any revisions have been performed on the program.

Using data that was not available to the program office at the time the price estimates were published<sup>4</sup>, we have computed an estimate of the tax rate for 1979 through 1981. We computed a total tax rate of 13.28 cents per gallon for 1979, 13.99 cents per gallon for 1980, and 14.33 cents per gallon for 1981.

Auditor errors are those errors made by the auditors while checking the accuracy of the enterprise responses. Validator errors are those errors made by the auditors while checking the accuracy of the enterprise responses. Validator errors are those errors made by someone such as this author while attempting to understand the system under question. These errors do not affect the accuracy of the original estimates, but affect the perceived accuracy of the original estimates.

#### EXTERNAL COMPARISONS

According to estimates from the Residential Consumption Survey, urban retail prices are higher than rural prices. Retail prices from the Consumer Price Index published by the Bureau of Labor Statistics (BLS), show a trend with larger cities having higher prices. Similar trends are seen in estimates in the "Oil and Gas Journal," although in both the BLS prices and the "Oil and Gas Journal" prices the trend accounts for a very small portion of the total variation between cities. We have no proof that there is an urban effect, but the data we have looked at are consistent with this hypothesis. If this hypothesis is true, the

hypothesis would explain why the city average retail prices reported by BLS and Lundberg, Inc., tend to be higher than the other price data series.

When comparing EIA-460 estimates of wholesale and dealer-tank-wagon sales with those of the Bureau of Labor Statistics, we find sizable differences in price only for wholesale unleaded and regular. We recomputed the EIA-460 estimated prices as a price index based upon 1977 annual sales. Sizable differences between the index price and the volume weighted price were found only for wholesale unleaded and premium. The directions of the differences are the same as in the comparisons of the BLS to EIA-460 volume weighted comparison. It is possible that the differences that we see between the BLS estimates and the EIA published estimates may be caused by BLS using a price index while EIA uses a volume weighted average price.

To confirm the wide spread in prices we have looked at other data sources. The average retail price of motor gasoline varies widely between cities in the BLS retail price series. Bulk terminal prices have a large price spread as confirmed by Platt's Oil Price Handbook and Oilmanac.

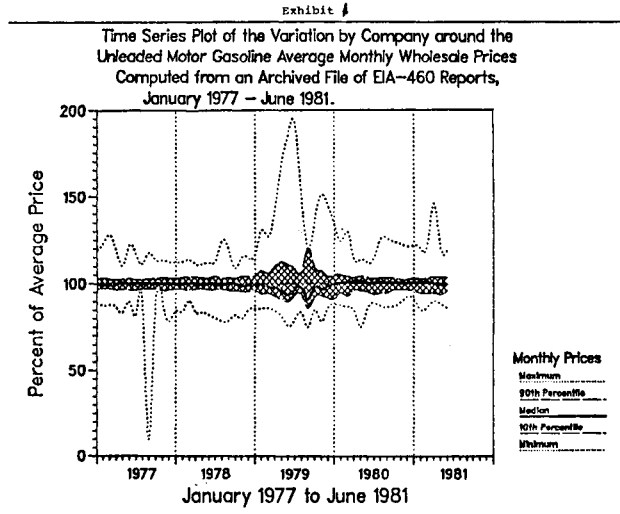
Many different estimates of retail gasoline prices exist. Comparisons to the EIA-460 by grade are not possible because of the specification error, therefore all comparisons involving the EIA-460 must be performed on total motor gasoline. The two household surveys, Auto-Facts and the EIA-141, were compared to each other by grade and were found to be very consistent. The Lundberg and BLS estimates exceeded those published by EIA in "Petroleum Market Shares." All estimates are usually within 5 percent of the EIA "Petroleum Market Shares" estimate of retail motor gasoline prices as illustrated in Exhibit 5. The EIA-460 portion of this published estimate is lower than the overall motor gasoline retail price.

#### CONCLUSION

Readers of average prices need to be warned that the average price may not be meaningful when prices are rapidly changing. I would suggest that in addition to the volume weighted average price, EIA should also publish the 10th and 90th percentiles. At the 10th percentile, 10 percent of the reported volumes are sold by enterprises having a lower monthly volume weighted average price.

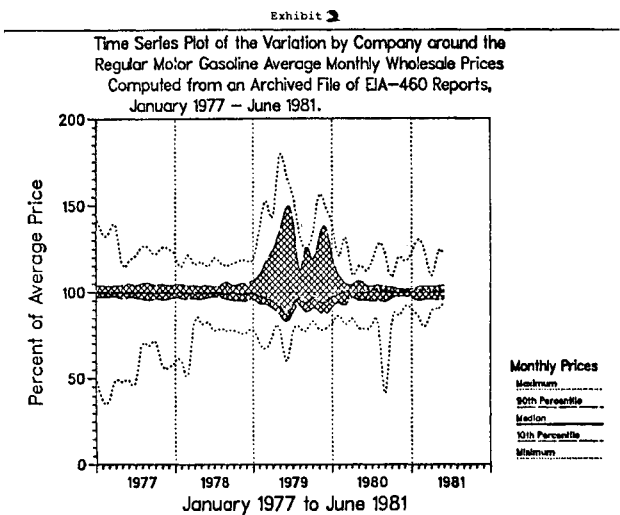
- (1) Energy Information Administration, Crude Petroleum, Petroleum Products and Natural Gas Liquids: 1980 (Final Summary), Energy Data Reports, "Petroleum Statement, Annual," Table 2, (Washington, D.C.: December 4, 1981).
- (2) U.S. Department of Energy, "Petroleum Market Shares, Report on Sales of Retail Gasoline, December 1980," (Washington, D.C., March 31, 1981).

- (3) TERA, Incorporated, "Validation of Selected Petroleum Product Energy Information Systems," (Falls Church, VA September 30, 1981).
- (4) U.S. Department of Transportation, "Highway Statistics 1979," "Highway Statistics 1980," (Washington, D.C., undated).



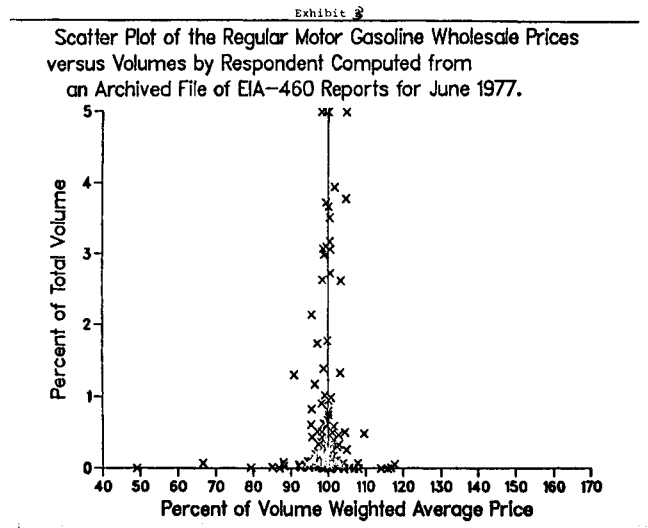
Note: Vertical lines are drawn through January data values. We suspect that the minimum value is incorrect by a factor of 10 in September 1977. The volume associated with the highest price in September 1979 may be too large by a factor of 1,000.

Source: Archived EIA computer file "CN6329.PRJ.SOD82.EIA460.FEB1982.SAS," archived February 10, 1982.



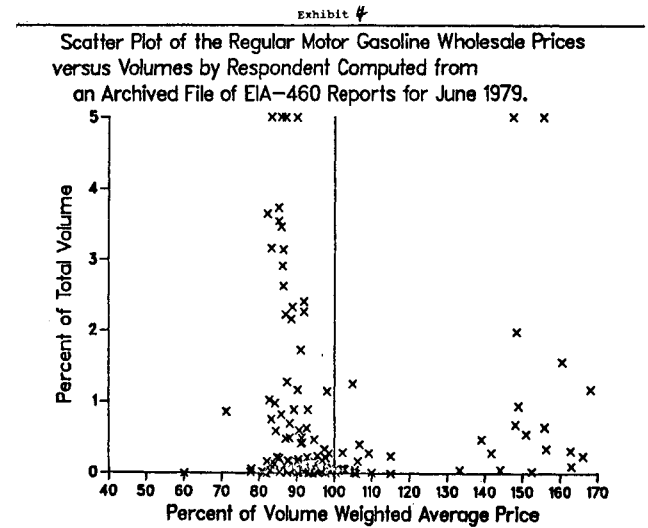
Note: Vertical lines are drawn through January data values.

Source: Archived EIA computer file "CN6329.PRJ.SOD82.EIA460.FEB1982.SAS," archived February 10, 1982.



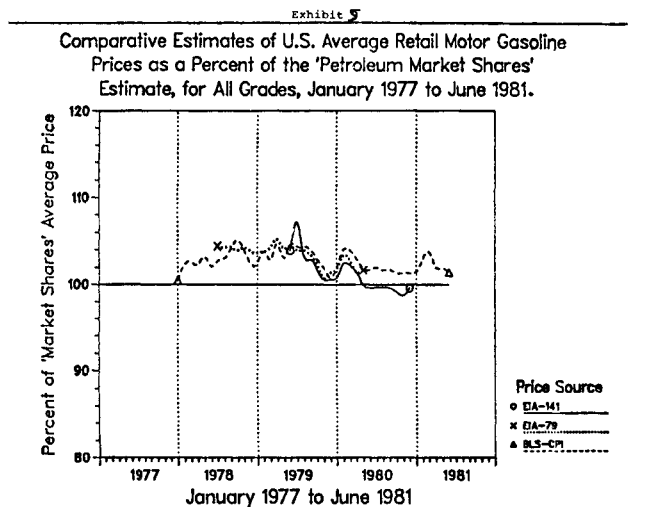
Note: All respondents having 4 percent or more of the total reported sales are plotted at a volume of 5 percent to protect confidentiality. No respondent has a volume over 6 percent of the total reported volume.

Source: Archived EIA computer file "CN6329.PRJ.SOD82.EIA460.FEB1982.SAS," archived February 10, 1982.



Note: All respondents having 4 percent or more of the total reported sales are plotted at a volume of 5 percent to protect confidentiality. No respondent has a volume over 6 percent of the total reported volume.

Source: Archived EIA computer file "CN6329.PRJ.SOD82.EIA460.FEB1982.SAS," archived February 10, 1982.



Note: Vertical lines are drawn through January data values.

Source: For "Petroleum Market Shares," "Petroleum Market Shares, Report on Sales of Retail Gasoline." For EIA-141, Household Transportation Panel, Residential Consumption Survey. For EIA-79, "Monthly Energy Review." For BLS-CPI, "Consumer Prices: Energy and Food" (title varies by month).