Methodology for EIA Weekly Retail Price Estimates of Winter Fuels

April 2024

Table of Contents

Overview of Methodology Changes

On October 7, 2019, we implemented new samples based on methodological changes to improve the accuracy of our estimated weekly retail prices of residential No. 2 heating oil and propane. These methodological changes include:

- Updated sampling frames based exclusively at the outlet level
- New sampling and estimation methodologies
- The publication of standard errors for the weekly price estimates

We will monitor the results from these samples and the performance of the methods.

As a result of these methodology changes, differences between the published price estimates for the 2018–19 heating season and 2019–20 heating season reflect changes in both the market and the samples that we use to estimate the weekly prices. Documentation on the previous sampling methodology we used before these changes is archived on our website. The sections below discuss our methodology for the new samples. We updated these sections to reflect the new monthly data collection from these samples during the non-heating season (from April through September) beginning with 2024.

Weekly Prices of Winter Fuels and Annual Sales Volume Data Collection

We collect information on retail prices of residential No. 2 heating oil and propane from two independent samples (one for each of the products) of retail outlets across the United States using Form EIA-877, *Winter Heating Fuels Telephone Survey*. During the heating season (from October through March), this survey collects prices every Monday for these fuels. The weekly survey also collects data by state on the price of these fuels offered to residential consumers, excluding taxes and discounts such as those given to cash or high-volume purchasers. The data collected represent the price as of 7:00 a.m. local time on Monday even during the weeks when data collection may be delayed a day or two because of a holiday. During the non-heating season (from April through September), the monthly survey similarly collects prices for these fuels on the second Monday of the month for use during the subsequent heating season.

During the heating season, State Energy Offices and EIA collect data as a joint effort as part of the State Heating Oil and Propane Program (SHOPP). During the non-heating season, EIA collects the data. We collect residential No. 2 heating oil prices for 21 states in the East Coast and Midwest regions and the District of Columbia (DC). We collect residential propane prices for 38 states in the East Coast, Midwest, Gulf Coast, and Rocky Mountain regions.

We collect the weekly or monthly prices by state via telephone or email from the two samples of outlets. All collected prices go through automated error checks during data collection and data processing. We verify any data flagged for potential errors with the respondents. We use imputation (a statistical replacement process) to estimate prices for outlets we cannot reach during the data collection or validation process.

We use price data from the samples to calculate volume-weighted average price estimates for the fuels at the national, regional, and state levels. The volumes are based on the most recently available annual sales volume data by state obtained from the sampled outlets using Form EIA-877. The volume data are used only for statistical purposes and are not published.

During the heating season, we publish the average price estimates for residential No. 2 heating oil and propane at about 1:00 p.m. eastern time every Wednesday, except when a government holiday falls on or before the Wednesday of that week. In this case, we release the data (still representing Monday's prices) on Thursday or Friday, depending on the day of the week when the holiday occurs. Similarly, during the non-heating season, we publish the average price estimates for those fuels at about 1:00 p.m. every Wednesday of the week following the second Monday of the month, except when a government holiday falls on or before the Wednesday of that week. Note, published data reflect prices collected on the second Monday. For more information, see the *Holiday Release Schedule for the Heating Oil and Propane Update.*

Sampling

The two target populations, corresponding to the two fuel types (No. 2 heating oil and propane), are all active retail outlets located in the states covered by SHOPP that sell residential No. 2 heating oil or propane to residences in the United States. These target populations do not include outlets in states that are out of scope to SHOPP and sell these fuels to residences in states that are in scope to SHOPP. However, we believe that these outlets would have little effect on the published weekly or monthly price estimates. These outlets typically sell smaller volumes to states that are in scope to SHOPP, compared with outlets that are in these states.

To design the new samples, we constructed frames of about 5,000 retail No. 2 heating oil outlets and 7,000 retail propane outlets located in states that are in scope to SHOPP. We updated the original frames that were based on companies identified in the 2006 and 2010 data collection cycles of Form EIA-863, *Petroleum Product Sales Identification Survey*, to create the new outlet frames. We identified births and deaths by using outlet names, physical and mailing addresses, and ZIP codes from the most recent SHOPP samples, as well as administrative records and third-party data sources. These data sources included State Heating Energy Assistance Program Office lists, the Bureau of Labor Statistics' *Quarterly Census of Employment and Wages* (QCEW), Equifax database, Data.com, and the National Propane Gas Association's membership list. We assigned each outlet on a frame to a county based on its available address information.

We used similar sampling procedures for estimating weekly prices of residential No. 2 heating oil and propane. The new outlet samples are stratified systematic samples of about 1,000 retail No. 2 heating oil outlets in 21 states in the East Coast and Midwest regions and DC and of about 1,600 retail propane outlets in 38 states in the East Coast, Midwest, Gulf Coast, and Rocky Mountain regions. For each frame, we assign an outlet to a primary sampling stratum defined by which state the outlet is located in. These primary sampling strata are nonoverlapping, and one or more primary sampling strata may be combined to correspond to a publication cell.

The primary sampling strata are further sub-stratified into secondary strata based on the average sales volume by outlet. We based these volumes on the most recently available company sales data collected from Form EIA-863, *Petroleum Product Sales Identification Survey*, and Form EIA-821, *Annual Fuel Oil and Kerosene Sales Report*, as well as annual sales volumes collected from propane outlets in the previous SHOPP sample. For a given primary stratum, outlets with the largest average sales volumes that meet specified size criteria are assigned to the *certainty* substratum and are given a sampling weight of 1. The remaining outlets in the primary stratum are assigned to one of two noncertainty substrata based on average company sales volumes—the *large* substratum and the *other* substratum. The *large* substratum combined with the *certainty* substratum accounts for about 75% of the aggregate of the primary stratum's most recently available company sales volumes. The *other* substratum includes all other outlets in the primary stratum, including outlets with small average company sales volumes and outlets with no available company sales volume data.

To perform the sampling for most of the substrata, we sort the outlets by county and ZIP code and select an independent systematic random sample without replacement. For a small number of propane substrata in which the sample size was more than half of the size of the substratum, we selected a simple random sample to avoid selection bias. In this case, the bias results from selecting only the outlets at the top of the sorted list. The sample size for a primary stratum is determined by how the state compares with other states in terms of:

- Number of outlets represented on the frame
- Variability in weekly price data during the heating season from the previous SHOPP sample
- Attrition in the previous SHOPP sample as a result of outlets that were identified as being nonrespondents, out of business, or out of scope of the survey

Until we redesign the samples, we annually assess available sources of information on new outlets that were not in the initial construction of the frames. As needed to ensure accurate coverage, we select independent birth samples to augment the initial samples. In designing these birth samples, geographic regions will be oversampled where relatively higher rates of sample attrition occur.

Imputation and Estimation

We calculate the survey response rate separately by fuel type (residential No. 2 heating oil and propane) based on the annual volumes represented by the reporting outlets in the sample. For each of the two samples, the volumes represented by the reporting outlets in the weekly or monthly survey (in terms of total weighted annual sales volume) account for at least 80% at the U.S. level.

We handle item and unit nonresponse to weekly or monthly winter fuel prices and annual sales volumes at the outlet level by imputation, using previous survey data reported by the outlet and survey data reported from other outlets in the sample.

The estimation for weekly and monthly prices uses two sources of data from the *Winter Heating Fuels Telephone Survey*: annual sales volumes for each outlet in the samples and weekly or monthly price data for those outlets. Before implementing the new samples, we began collecting annual sales volumes for the retail outlets in the samples from owners of the outlets.

The sampling weight for a given sampled outlet is the reciprocal of the outlet's probability of selection in the sample. Using the annual sales volume data to estimate average prices, we constructed the volume weight for a given sampled outlet by multiplying its sampling weight by its annual sales volume. These volume weights are applied each week during the heating season to the reported or imputed outlet winter fuel prices to obtain weighted average price estimates for the geographic areas that we publish. We use the same estimation procedure during the non-heating season to produce the monthly weighted average price estimates that we publish.

For quality assurance purposes, we withhold an average price estimate from publication if the outlets that reported price data contribute less than half of the total weighted annual sales volume for all active, in-scope outlets in the sample that could have reported price data.

In addition to publishing retail price estimates, we publish wholesale price estimates for No. 2 heating oil and propane. Wholesale heating oil prices are calculated for selected states from the Oil Price Information Service's (OPIS) city-level prices for states where sufficient data are available. Wholesale propane prices are provided to EIA directly from OPIS. We calculate most regional and U.S. average price estimates as weighted averages of the state-level price estimates using as weights the most recently published annually aggregated state-level prime supplier volumes collected in Form EIA-782C, *Monthly Report of Prime Supplier Sales of Petroleum Products Sold for Local Consumption*. The exceptions are Kansas and Texas, where the most recent state residential volumes from the petroleum marketing surveys are used in the weighting because of the influence of the Conway (Kansas) and Mt. Belvieu (Texas) propane storage facilities on prime supplier volumes for those states.

Sampling Error, Measures of Sampling Variability, and Confidence Intervals

Sampling error is a statistical term for the error caused by observing a sample instead of the entire sampling frame. Statistics based on a sample, such as averages, generally differ from statistics for the entire frame because the sample includes only a subset of the frame.

Statisticians use measures of sampling variability, such as the standard error and the coefficient of variation, to measure the sampling error. These measures of sampling variability are typically estimated from the sample that was selected. The standard error is a measure of the sampling variability of the estimate based on all possible samples that could have been selected using the chosen sample design. It is measured in the same units as the estimate, in this case, dollars per gallon for weekly or monthly winter fuel prices. The coefficient of variation, which may also be referred to as the relative standard error, is the standard error expressed as a fraction of the estimate.

Each average price estimate published by EIA has a corresponding estimated standard error published in the *Detailed Price and Standard Error Report*. For quality assurance purposes, average price estimates are flagged if the corresponding estimated coefficient of variation is more than 5%.

Data users can use the estimated standard error to compute a confidence interval centered about the corresponding published average price estimate with a desired level of confidence. We selected only one of many possible samples for the *Winter Heating Fuels Telephone Survey.* If it were possible to

construct a confidence interval for each of these possible samples, the percentage of confidence intervals containing the census value (if we had surveyed the entire sampling frame) would be expected to equal the level of confidence. For example, if it were possible to construct a 95% confidence interval for each possible sample, then one would expect that 95% of these confidence intervals would contain the value obtained from taking a census of the sampling frame.

To determine the width of the confidence interval for a published average price estimate, users can compute the margin of error (MOE) using the estimated standard error. The MOE is defined as the estimated standard error of the estimate multiplied by the standard normal percentile for the level of confidence, rounded up to the nearest unit used in publishing the corresponding estimate. The lower bound of the confidence interval is the estimate minus the MOE, and the upper bound of the confidence interval is the estimate plus the MOE. For the standard normal percentile, 1.645 is used for a 90% confidence interval, and 1.96 is used for a 95% confidence interval.

For example, suppose an average price estimate of $1.670 has an estimated standard error of $0.0230482625709464 in the *Detailed Price and Standard Error Report*. The 95% margin of error would be 1.96 multiplied by $0.0230482625709464, which rounds up to $0.046. The 95% confidence interval would then be $1.670 +/- $0.046, or $1.624 to $1.716.

Non-Sampling Errors

Potential errors unrelated to sampling, called non-sampling errors, include various response and operational errors, such as those related to data collection, respondent reporting, transcription, and nonresponse. All these types of errors could also occur even if every known outlet had been surveyed under the same conditions as the sample survey. Although non-sampling error is not measured directly, we employ quality control procedures throughout the survey process.

During the heating season, preliminary price estimates for the previous week may be revised in the publication for the current week based on data that were received late or were revised. Revision error is the difference between published preliminary and final estimates for the previous week. During the non-heating season, the price estimates published in the previous publication will not be revised as a result of the additional week allotted for data processing and validation.

Confidentiality of Information

Beginning with the 2019–20 heating season, the information collected on Form EIA-877 is protected under the Confidential Information Protection and Statistical Efficiency Act of 2018 (CIPSEA). As a result, the information collected on Form EIA-877 is used for statistical purposes only and is confidential by law. Under CIPSEA and other applicable federal laws, responses will not be disclosed in identifiable form without consent. Under the Federal Cybersecurity Enhancement Act of 2015, federal information systems are protected from malicious activities through cybersecurity screening of transmitted data. Every EIA employee, as well as every agent, is subject to a jail term, a fine, or both if they make public any identifiable information reported on Form EIA-877.