Supporting Statement for Petroleum Marketing Program

Part B: Collections of Information Employing Statistical Methods

OMB No. 1905-0174

Form EIA-14, Refiners' Monthly Cost Report
Form EIA-182, Domestic Crude Oil First Purchase Report
Form EIA-856, Monthly Foreign Crude Oil Acquisition Report
Form EIA-877, Winter Heating Fuels Telephone Survey
Form EIA-878, Motor Gasoline Price Survey
Form EIA-888, On-Highway Diesel Fuel Price Survey

November 2024

The U.S. Energy Information Administration (EIA), the statistical and analytical agency within the U.S. Department of Energy (DOE), prepared this report. By law, our data, analyses, and forecasts are independent of approval by any other officer or employee of the U.S. Government. The views in this report do not represent those of DOE or any other federal agencies.

Table of Contents

Table of Tables

B.1. Respondent Universe

The Petroleum Marketing Program collects data from two types of entities: firms and outlets. In terms of the flow of petroleum products from the upstream to downstream markets, the data system begins by collecting data from firms, generally parent companies that have complex structures with multiple offices, locations, subsidiaries, etc. The target population for the surveys is the firms, which are defined in terms of their oil market activities. The target population includes firms that purchase domestic or foreign crude oil (Forms EIA-182 and EIA-856) and firms that refine crude oil into finished petroleum products (Form EIA-14). The remaining surveys focus on individual retail outlets selling the products to consumers (Forms EIA-877, EIA-878, and EIA-888) and collect price data on a weekly basis. The frames for the monthly surveys are kept current using information from other surveys as well as information from industry journals, administrative records, and other sources.

B.1.1. Monthly crude oil survey frames and target population

Form EIA-14, **Refiners' Monthly Cost Report**
The target population for this survey is all refiners of crude oil. EIA constructed the frame for Form EIA-14 from a list of 206 refiners obtained from the *Oil & Gas Journal* in 1983. The frame is updated periodically using information derived from Form EIA-810, *Monthly Refinery Report*. There are currently 62 active respondents filing Form EIA-14.

Form EIA-182, **Domestic Crude Oil First Purchase Report**
The target population for this survey is all firms that buy domestic crude oil at the lease boundary, acquiring ownership of the crude oil in a first purchase transaction. EIA initially compiled the frame for Form EIA-182 from the 1974 Federal Energy Administration's (which later became EIA) *Oil and Gas Survey of Producers and Operators*. Collection of data from first purchasers began in February 1976. By 1978, the frame consisted of 340 respondents. Of these respondents, 198 purchased more than 150,000 barrels per year, which represented 99.9% of the total reported volume. Following Executive Order 12287—Decontrol of Crude Oil and Refined Petroleum Products in January 1981, many small firms went out of business or were absorbed by larger companies. By January 1986, the frame had been reduced to 170 respondents. Over the years, adjustments to the frame have mostly been deaths, with relatively few births. The size of the frame has declined from 340 firms in 1978 to 91 firms in 2023.

Form EIA-856, **Monthly Foreign Crude Oil Acquisition Report**
All companies acquiring more than 500,000 barrels of foreign crude oil in the report month for importation into the United States are required to submit this form monthly. The frame is updated periodically from information reported on Form EIA-814, *Monthly Imports Report*. Currently, the frame consists of 38 respondents.

B.1.2. Weekly petroleum product frames and target population

Form EIA-877, **Winter Heating Fuels Telephone Survey**
Constructed separately for residential propane sellers and residential heating oil sellers in selected states, the two outlet-level sampling frames for this survey consist of outlets originally identified from the Form EIA-863 in 2006 and 2010 data collection cycles. EIA then adjusted these frames for births and

deaths that identified using multiple administrative records and third-party data sources including State Energy Office lists, Bureau of Labor Statistics' *Quarterly Census of Employment and Wages* (QCEW), the Equifax database, Dun and Bradstreet, and the National Propane Gas Association business list. The sampling frame for outlets that sell residential heating oil consists of about 5,000 outlets in 21 states and the District of Columbia in the East Coast and Midwest regions, while the sampling frame for outlets that sell residential propane consists of about 7,000 outlets in 38 states in the East Coast, Midwest, Gulf Coast, and Rocky Mountain regions. To ensure accurate coverage, the two frames are assessed annually and updated, if needed, prior to selection of samples of newly identified outlets (births). Prior to the 2020–2021 heating season, independent birth samples of 7 heating oil outlets and 29 propane outlets were selected. Prior to the 2022–2023 heating season, independent birth samples of 9 heating oil outlets and 23 propane outlets were selected.

Form EIA-878, Motor Gasoline Price Survey

The target population is all active retail gasoline outlets in the United States for a given week. The population includes two types of outlets—big-box and non-big-box outlets. Big-box outlets typically sell large volumes of gasoline at discounted prices.

The sample for Form EIA-878 was drawn from a frame of approximately 130,000 retail gasoline outlets in the United States that were active in 2016. EIA constructed the gasoline outlet frame by combining outlet level name and address information purchased from the Oil Price Information Service (OPIS) with information from other sources including the Yellow Pages and Secretary of States' business registers. The individual outlets in the frame were assigned to counties after converting the physical addresses to geographic coordinates. The outlets were then assigned either as reformulated or conventional gasoline areas based on the published geographic areas as defined by the U.S. Environmental Protection Agency program and some state-defined reformulated gasoline program areas. The outlets were then further assigned to city areas based on the geographic areas as defined by EIA. These regions are nonoverlapping, and one or more sampling regions will comprise a publication cell. To ensure accurate coverage, the frame is assessed annually and updated, if needed, prior to selection of a sample of newly identified outlets (births). In 2018 and 2020, birth samples of size 145 and 187, respectively, were selected.

Form EIA-888, On-Highway Diesel Fuel Price Survey

The target population is all active retail diesel outlets in the contiguous United States for a given week. The population includes two types of outlets—truck stops and service stations that sell on-highway diesel fuel. Due to statistical and operational considerations, outlets in Alaska and Hawaii were excluded from the target population. EIA constructed the EIA-888 frame using commercially available lists from several sources. These sources were used to provide a comprehensive coverage of truck stops and service stations that sell on-highway diesel fuel in the contiguous United States. The frame includes around 73,000 service stations and 9,500 truck stops. To ensure accurate coverage, the frame is assessed annually and updated, if needed, prior to selection of a sample of newly identified outlets (births).

B.2. Statistical Methods

Three of the six petroleum marketing surveys—Forms EIA-14, EIA-182, and EIA-856—are census surveys. Imputation is used to account for nonresponse on all surveys in the Petroleum Marketing Program. The target population for Form EIA-856 is all companies importing over 500,000 barrels of foreign crude oil into the United States in the report month. Monthly estimates of volume-weighted prices are generated from the data reported on Forms EIA-14, EIA-182, and EIA-856. The total cost or revenue (price * volume) is divided by a corresponding total volume to calculate the volume-weighted average price.

For EIA-877, EIA-878, and EIA-888, EIA calculates sample weights as the inverse of the probability of selection of the sampled company or outlet. The price surveys create volume-weights for use in estimation, which are the product of the sampling weight multiplied by a sales volume measure for the outlet.

B.2.1 Statistical methods for monthly crude oil surveys

Form EIA-14 is a census. A volume-weighted price at the national and regional levels is calculated by dividing the total cost (price * volume) by the corresponding total volume.

Form EIA-182 is a census. To obtain a volume-weighted first purchase average price at the national and state level, the total cost (price per barrel of crude oil * volume purchased) is divided by the corresponding total volume. Subsequently, the data are sorted by crude oil stream within each state. These data are aggregated across all companies reporting purchases from a given state. Weighted average prices for crude oil are then derived for each producing state and for the Outer Continental Shelf regions, Alaska North Slope, and Alaska Other.

Form EIA-856 is a census of those companies importing over 500,000 barrels of foreign crude oil into the United States in the report month.

B.2.2 Statistical methods for weekly petroleum product surveys

All three of the following surveys collect weekly prices for petroleum products, including motor gasoline, on-highway diesel, heating fuel, and propane. Table B1 contains a summary of the sample design for the weekly surveys (Forms EIA-877, EIA-878, and EIA-888) in the Petroleum Marketing Program.

Table B1. Summary of sample design

Survey name	Sample design	Selection procedure	Current sample size	Proposed sample size
EIA-877, *Winter Heating Fuels Telephone Survey*	Residential heating oil sellers: Stratified systematic random sample of approximately 1,200 outlets for 21 states and DC in the East Coast and Midwest regions Residential propane sellers: Stratified systematic random sample of approximately 1,800 outlets for 38 states in the East Coast, Midwest, Gulf Coast, and Rocky Mountain regions	See Form EIA-877 sample design	2,663	3,000
EIA-878, *Motor*	Stratified systematic random sample of retail outlets for 50	See Form EIA-	1,165	1,300

Gasoline Price Survey	states and DC	878 sample design		
EIA-888, *On-Highway Diesel Fuel Price Survey*	Stratified systematic random sample of retail outlets from 48 continental states and DC	See Form EIA-888 sample design	590	650

- EIA uses measures of sampling variability, such as the standard error and the coefficient of variation, to measure the sampling error in the three weekly price surveys referenced above. These measures of sampling variability are 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, and it is measured in the same units as the estimate (current dollars per gallon [gal] for weekly gasoline prices, on-highway diesel prices, or No. 2 heating oil prices). The coefficient of variation, which is often referred to as the relative standard error, is the standard error expressed as a fraction of the estimate.

- Each weekly average price estimate EIA publishes has a corresponding estimated standard error published with the weekly price estimates. 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. For example, EIA selected only one of many possible samples for a given weekly survey. If a confidence interval were constructed 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 one could construct a 95% confidence interval for each possible sample that could be selected, 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 given 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.

Form EIA-877 sample design

Similar sampling and estimation procedures are used for estimating weekly prices of residential No. 2 heating oil and propane. For the No. 2 heating oil and propane sampling frames, primary strata are

defined based on the state of an outlet's location, which is the most detailed geographic level used for published estimates. Secondary strata within a primary (state) stratum are based on the relative sales volume of the companies that own the outlets and annual sales volumes collected from propane outlets in the previous sample.

Each secondary stratum is sorted by county and ZIP code. From each sampling frame, a systematic random sample is selected from each secondary stratum. Sorting in this way imposes an implicit stratification so that we prevent selecting the sample for a given substratum consisting of outlets in only a certain part of the state. From the frame of residential heating oil sellers, 1,024 outlets were selected, and 1,631 outlets were selected from the frame of residential propane sellers, resulting in a total sample size of 2,655 outlets. The sample size for a given state is determined by how it compares with other states in terms of number of outlets represented on the frame, variability in weekly price based on available data, and sample attrition in the previous sample due to outlets that were nonrespondents, out of business, or out of scope. Each sampled outlet is assigned a sampling weight that is the inverse of its probability of selection in the sample. Constraints on the minimum sample size for a given state and the maximum sampling weight were used.

The total sample size increased from 1,506 outlets to 2,655 outlets due to changes in the sampling and estimation methodologies. First, outlets that are identified as out of business, out of scope, or nonrespondents are no longer replaced with other outlets from the frame that were not selected in the initial sample size. Second, the EIA-863 collection cycle was last conducted in 2011 for the 2010 survey year so the sampling frames for heating oil and propane were updated using various administrative and third-party data sources that have varying degrees of accuracy. We expected more attrition for outlets in the heating oil sample that are out of business or out of scope. Third, while creating the sampling frames at the outlet level allows more control in the sample design in terms of certain characteristics of the outlets such as their geography and company ownership, sales volume information is limited, and the increased sample size reflects this lack of accurate information on the measure size for the outlets in the sample design. Due to changes in the population over time, we expect the total sample size for the next new sample to be approximately 3,000 outlets, which comprises approximately 1,200 residential heating oil sellers and approximately 1,800 residential propane sellers.

Form EIA-877 estimation
Weekly price data are collected by State Energy Offices as part of the U.S. Energy Information Administration (EIA) State Heating Oil and Propane Program (SHOPP). When EIA updates the sample, it collects recent annual sales volume data from respondents. The volume weight for a given sampled outlet is then constructed by multiplying its sampling weight by its reported or imputed annual sales volume. These volume weights are applied each week to the reported or imputed outlet prices to obtain weighted average price estimates for the geographic areas that EIA publishes. Item and unit nonresponse to weekly prices and annual sales volumes are handled at the outlet level by imputation using prior survey data reported by the outlet and survey data reported from other outlets in the sample.

The method used to produce weighted average price estimates from the heating oil and propane samples is detailed below.

The following notation in calculating weighted average price estimates and their measures of sampling variability is used:

k = type of outlet (secondary stratum based on size category for heating oil or propane);

j = sampling region (primary stratum based on state of the outlet's location);

i = outlet;

N_{jk} = population size (number of outlets);

n_{jk} = sample size in sampling region j and type of outlet k;

P = price;

V = volume;

W_{jk} = sampling weights = inverses of probability of selection; and

f_{jk} = denotes the sampling fraction in sampling region j and type of outlet k, for certainty strata f_{jk} = 1.

By definition,

$$W_{jki} = \frac{N_{jk}}{n_{jk}}$$

Define
$$x_{jki} = W_{jki} V_{jki}$$

$$y_{jki} = W_{jki} V_{jki} P_{jki}$$

$$x_j = \sum_{i,k} x_{jki}$$

$$y_j = \sum_{i,k} y_{jki}$$

Then the volume-weighted average price for a region is as follows: $\hat{P}_j = \dfrac{y_j}{x_j}$

Relvariance (relative variance) is calculated by dividing the estimated variance of a statistic by the square of the statistic.

The relvariance of $\hat{P}_j$ is as follows:

$$V^2{}_{P_j} = V^2{}_{x_j} + V^2{}_{y_j} - 2 V_{x_j y_j}$$

where,

$$V^2_{x_j} = \frac{\sigma^2_{x_j}}{x^2_j} \quad \text{and} \quad V_{x_j y_j} = \frac{\sigma_{x_j y_j}}{(x_j)(y_j)}$$

The estimated variance for x_j is as follows (the variance of y_j is defined similarly):

$$\hat{\sigma}^2_{x_j} = \sum_k (1 - f_{jk}) \frac{n_{jk}}{n_{jk} - 1} \sum_i (x_{jki} - \bar{x}_{jk})^2$$

where,

$$\bar{x}_{jk} = \frac{\sum_i x_{jki}}{n_{jk}}$$

The estimated covariance is as follows:

$$\hat{\sigma}_{x_j y_j} = \sum_k (1 - f_{jk}) \frac{n_{jk}}{n_{jk} - 1} \sum_i (x_{jki} - \bar{x}_{jk})(y_{jki} - \bar{y}_{jk})$$

where,

$$\bar{y}_{jk} = \frac{\sum_i y_{jki}}{n_{jk}}$$

The volume-weighted average price for a publication region, which is made up of one or more sampling regions j, is as follows:

$$\hat{P} = \frac{\sum_j y_j}{\sum_j x_j}$$

The estimated relvariance of p is given as follows:

$$\hat{V}^2_P = \frac{\sum_j \hat{\sigma}^2_{x_j}}{\left(\sum_j x_j\right)^2} + \frac{\sum_j \hat{\sigma}^2_{y_j}}{\left(\sum_j y_j\right)^2} - \frac{2\sum_j \hat{\sigma}^2_{x_j y_j}}{\left(\sum_j y_j\right)\left(\sum_j x_j\right)}$$

where,

RSE(P) = $\left(\hat{V}^2 ¿¿ P\right)¿^{(1/2)}$

After the initial sample of outlets has been selected, available sources of information on new outlets since the sampling frames were constructed are analyzed each year until the sample is redesigned to determine whether it is necessary to update the sampling frames and select an independent birth sample of newly identified outlets to augment the initial sample. In designing these birth samples, geographic regions are oversampled where there are relatively higher rates of sample attrition because of outlets that were selected in the initial sample and subsequently went out of business.

Form EIA-878 sample design

The sample for the *Motor Gasoline Price Survey* was drawn from a frame of approximately 130,000 retail gasoline outlets in the United States that were active in 2016. The gasoline outlet frame was constructed by combining U.S. Energy Information Administration outlet information from a private commercial source with information contained on existing EIA petroleum product frames and surveys, federal and state administrative records, and other publicly available sources.

EIA obtained outlet names, physical addresses, and ZIP codes from the private commercial data source. The individual outlets in the frame were assigned to counties after converting the physical addresses to geographic coordinates. The outlets were then assigned either as reformulated or conventional gasoline areas based on the published geographic areas as defined by the U.S. Environmental Protection Agency program and some state-defined reformulated gasoline program areas. The outlets were then further assigned to city areas based on the geographic areas as defined by EIA.

The new gasoline outlet sample is a stratified systematic sample with a total size of 1,000 retail outlets. We expect the sample size for the next new sample to be approximately 1,300 retail outlets due to an increase in the number of big-box outlets over time and changes in the definitions of the reformulated gasoline program areas. Retail gasoline outlets are assigned to primary sampling strata based on physical address. 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 then substratified by retail gasoline outlet type (big-box or non-big-box). The total sample size is allocated to the sampling substrata in proportion to the number of outlets in the cell after weighting the big-box substrata in recognition of larger annual sales volume per outlet compared with non-big-box substrata.

Sampling within each sampling substratum is performed by ordering the outlets by county and ZIP code and selecting an independent systematic random sample without replacement. This procedure results in adequate sample representation by ZIP code within a given substratum.

In 2018 and 2020, birth samples were selected and added to the sample in 2019 and 2020, respectively, to account for new outlets that were identified. Also, each year, some geographic regions may experience relatively higher annual rates of outlets going out of business. Those geographic regions with relatively higher rates of sample attrition are oversampled to account for this impact.

Form EIA-878 estimation

In the first phase of this survey, a measure of size—annual volumetric sales by grade of gasoline—was collected from the sample of outlet owners with additional collection annually for new birth samples.

These volumes were applied each week to the reported outlet gasoline prices to get volume-weighted average prices and measures of uncertainty.

Imputation of annual volumes for nonrespondents relies on three different methods based on the level of nonresponse:

1. Only total gasoline volume for all three grades of gasoline (that is, regular, midgrade, and premium combined) is reported:

   If a respondent provided only total annual sales volume for all three grades of gasoline for a particular outlet, but it is known that the respondent also sold all three grades of gasoline for that survey period at that particular outlet, imputation of annual gasoline volumes by grade are based on the proportion of historical reported volumes by grade from the Form EIA-782C, *Monthly Report of Prime Supplier Sales of Petroleum Products Sold for Local Consumption* for the U.S. state or district of the respondent's outlet.

   The following is the formula for imputation for respondents reporting only total annual gasoline sales volume for all three grades of gasoline:

   $$V_i = T_i \times \left(\frac{P_V}{100} \right)$$

   where,

   V_i = annual volume of either regular, midgrade, or premium grade gasoline for a given outlet *i* to be imputed;

   T_i = reported total annual volume for all three grades of gasoline for a given outlet *i*; and

   P_V = percentage of annual volume of either regular, midgrade, or premium grade gasoline sold with respect to the total annual gasoline reported on the EIA-782C for the outlet's state or district.

   Imputation example:

   - Outlet A in State B reported 20,000 gallons of total gasoline for all three grades sold in Year C but did not provide any sales volumes by grade of gasoline despite selling such a product in Year C.

   - On EIA-782C collected data, State B had proportions for all gasoline sold by prime suppliers by grades of 85% regular, 5% midgrade, and 10% premium reported for the most recent complete calendar year near Year C.

   - The gasoline volume by grade for Outlet A for Year C would be imputed as follows:

$$20{,}000 \text{ gal gasoline} \times \frac{85\% \ \text{reg. gas in State B for Year C}}{100} = 17{,}000 \text{ gal reg. gas}$$

$$20{,}000 \text{ gal gasoline} \times \frac{5\% \ \text{mid. gas in State B for Year C}}{100} = 1{,}000 \text{ gal mid. gas}$$

$$20{,}000 \text{ gal gasoline} \times \frac{10\% \ \text{prem. gas in State B for Year C}}{100} = 2{,}000 \text{ gal prem. gas}$$

2. Only regular grade gasoline volume is reported:

If a respondent provided only annual sales volume for regular grade gasoline for a particular outlet but also reports weekly prices for midgrade and/or premium grade gasoline, then imputation of annual midgrade and/or premium grade gasoline volumes are based on the proportion of weighted volumes of midgrade and/or premium grade gasoline to regular grade gasoline as reported by EIA-878 respondents, calculated separately by type of outlet (big box or other).

The following is the formula for imputation for respondents only reporting annual regular grade gasoline volume sales:

$$V_i = G_i \times \frac{\sum_{j \epsilon R} W_j V_j}{\sum_{j \epsilon R} W_j G_j}$$

where,

V_i = annual volume of either midgrade or premium grade gasoline for a given outlet *i* to be imputed;

G_i = reported annual volume of regular grade gasoline for a given outlet *i*;

R = the set of EIA-878 respondents who reported volumes for all three grades of gasoline for outlets of the same type (big box or other) as outlet *i*; and

W_i = sampling weight for outlet *i* equal to the inverse of the outlet's probability of selection.

Imputation example:

Outlet A is a big-box outlet that reported 10,000 gallons of regular grade gasoline but did not provide any volumetric data for premium grade gasoline sold despite selling such a product.

Using data reported in EIA-878 by big-box outlets, the ratios of weighted volumes for medium grade and premium grade to regular grade are 30% and 10%, respectively.

The premium grade gasoline volume for Outlet A would be imputed as follows:

$$10,000 \text{ gal regular gas} \times 0.1 = 1,000 \text{ gal premium gas}$$

3. No gasoline volumes reported:

If a respondent provided no volumetric data, imputation is carried out via a donor imputation method with the pool of potential donors for the outlet based on the type of the outlet (big box or other) and region (the Petroleum Administration for Defense District [PADD] or sub-PADD) in which the outlet is located. The donor imputation procedure is as follows:

The type of outlet and region are identified and treated as a viable pool for random donor imputation with each reported volume by grade being a potential donor for an outlet's unreported grade of gasoline volume.

A subject matter expert reviews the potential donors and may remove large outlier volumes by grade from the donor pool to ensure that imputed values do not overly weight the prices for the outlet with unreported volumes.

Volumes of gasoline by grade to be imputed are randomly selected from the viable donor pool, incorporating the sampling weights of the potential donors, and assigned to unreported outlet volume.

The recurring weekly survey collects the cash price (dollars per gallon) information by gasoline grade from each respondent and uses a weighted price estimation method as detailed below.

For the weekly price estimation of means and variances, the following notation is used:

k = type of outlet (1 = big box, 2 = other);

j = sampling region;

i = outlet;

N_{jk} = population size (number of outlets);

n_{jk} = sample size in sampling region j and type of outlet k;

P = price;

V = volume;

W_{jk} = sampling weights = inverses of probability of selection; and

f_{jk} = denotes the sampling fraction in region j and type of outlet k, for certainty strata f_{jk} = 1.

By definition,

$$W_{jki} = \frac{N_{jk}}{n_{jk}}$$

Define
$$x_{jki} = W_{jki} V_{jki}$$

$$y_{jki} = W_{jki} V_{jki} P_{jki}$$

$$x_j = \sum_{i,k}^{\square} x_{jki}$$

$$y_j = \sum_{i,k}^{\square} y_{jki}$$

Then the volume-weighted average price for a region is as follows: $\hat{P}_j = \dfrac{y_j}{x_j}$

Relvariance (relative variance) is calculated by dividing the estimated variance of a statistic by the square of the statistic.

The relvariance of $\hat{P}_j$ is as follows:

$$V^2_{P_j} = V^2_{x_j} + V^2_{y_j} - 2 V_{x_j y_j}$$

where,

$$V^2_{x_j} = \frac{\sigma^2_{x_j}}{x^2_j} \quad \text{and} \quad V_{x_j y_j} = \frac{\sigma_{x_j y_j}}{(x_j)(y_j)}$$

The estimated variance for x_j is as follows (the variance of y_j is defined similarly):

$$\hat{\sigma}^2_{x_j} = \sum_k^{\square} (1 - f_{jk}) \frac{n_{jk}}{n_{jk} - 1} \sum_i^{\square} (x_{jki} - \bar{x}_{jk})^2$$

where,

$$\bar{x}_{jk} = \frac{\sum_i^{\square} x_{jki}}{n_{jk}}$$

The estimated covariance is as follows:

$$\hat{\sigma}_{x_j y_j} = \sum_k^{\square} (1 - f_{jk}) \frac{n_{jk}}{n_{jk} - 1} \sum_i^{\square} \left(x_{jki} - \overline{x}_{jk} \right) \left(y_{jki} - \overline{y}_{jk} \right)$$

where,

$$\overline{y}_{jk} = \frac{\sum_i^{\square} y_{jki}}{n_{jk}}$$

The volume-weighted average price for a publication region, which is made up of one or more sampling regions j, is as follows:

$$\hat{P} = \frac{\sum_j^{\square} y_j}{\sum_j^{\square} x_j}$$

The estimated relvariance of p is given as follows:

$$\hat{V}^2_P = \frac{\sum_j^{\square} \hat{\sigma}^2_{x_j}}{\left(\sum_j^{\square} x_j \right)^2} + \frac{\sum_j^{\square} \hat{\sigma}^2_{y_j}}{\left(\sum_j^{\square} y_j \right)^2} - \frac{2 \sum_j^{\square} \hat{\sigma}^2_{x_j y_j}}{\left(\sum_j^{\square} y_j \right) \left(\sum_j^{\square} x_j \right)}$$

where,

$$\text{RSE(P)} = \left(\hat{V}^2 ¿¿ P \right) ¿^{(1/2)}$$

Weekly price imputation for a nonrespondent outlet's price by grade of gasoline is based on a volume-weighted average of price change from the prior week to the current week for other stations of the same outlet type (big box or other) and the region (PADD or sub-PADD) in which the outlet is located.

The weekly price imputation formula for prices by nonrespondent outlet is as follows:

$$\hat{P}_{i,t} = P_{i,t-1} + \frac{\sum_j^{\square} V_j \left(P_{j,t} - P_{j,t-1} \right)}{\sum_j^{\square} V_j}$$

where,

$\hat{P}_{i,t}$ = imputation price estimate for nonrespondent price for a particular grade of gasoline for a nonrespondent outlet (*i*) during current week (*t*);

$P_{i,t-1}$ = the outlet's (*i*) prior week's (*t-1*) reported price for the unreported grade of gasoline;

V_j = the annual volume reported of the unreported grade of gasoline for another individual outlet (j) within the nonrespondent outlet's (i) outlet type and region;

$P_{j,t}$ = the current week's (t) reported price for the unreported grade of gasoline for another individual outlet (j) within the nonrespondent outlet's (i) outlet type and region; and

$P_{j,t-1}$ = the prior week's (t-1) reported price for the unreported grade of gasoline for another individual outlet (j) within the nonrespondent outlet's (i) outlet type and region.

A larger sample size of approximately 1,000 outlets was needed due to changes in the sampling and estimation methodologies. First, outlets that are identified as out of business, out of scope, or nonrespondents are no longer replaced with other outlets from the frame that were not selected in the initial sample. Second, the EIA-863 collection cycle was last conducted in 2011 for the 2010 survey year, and the sampling frame for motor gasoline was updated using various administrative and third-party data sources that have varying degrees of accuracy. Third, although creating the sampling frames at the outlet level allows more control in the sample design in terms of certain characteristics of the outlets such as their geography and company ownership, sales volume information is limited, and the increased sample size reflects this lack of accurate information on the measure size for the outlets in the sample design.

The target coefficient of variation was set at 0.4 for the United States, 0.55 for PADDs and U.S. formulations, 0.70 for sub-PADDS and the PADD formulations, 0.85 for cities and states, and 1.0 for the remaining cells—that is, state and sub-PADD formulations. The sample size is approximately 1,000 outlets, with outlets added as part of supplemental birth samples to augment the sample when necessary based on annual assessments. The EIA-878 survey every Monday (Tuesday on federal holidays) and more frequently during emergency situations. Data are released on EIA's website around 5:00 p.m. each Monday (Tuesday on federal holidays). Customers can also sign up to receive the data via email. The U.S., PADD, sub-PADD, state, and city level regular gasoline average prices are made available on EIA's prerecorded telephone hotline at 202-586-6966 and in the publications *Weekly Petroleum Status Report* and *This Week in Petroleum* (TWIP).

Form EIA-888 sample design

The respondents reporting to the weekly diesel price survey were selected using a stratified systematic random sample selected from a frame list of retail outlets. The outlet sampling frame was constructed using commercially available lists from several sources to provide comprehensive coverage of truck stops and service stations that sell on-highway diesel fuel in the United States. The frame includes about 73,000 service stations and 9,500 truck stops. Due to statistical and operational considerations, outlets in Alaska and Hawaii are excluded from the target population. For the purposes of this frame, a 'truck stop' is any on-highway diesel retail outlet that was classified as a truck stop by one or more of the third-party data sources that were used to construct the frame. EIA developed a model using historical sales data from Form EIA-821 to estimate sales volumes at the outlet level. This model helped inform stratification of the truck stops into four size categories based on various auxiliary variables found in several of the third-party data sets used. Variables used for stratification were truck diesel lane counts,

traffic volumes on nearby roadways, truck parking availability at the outlet, and sales of Diesel Exhaust Fluid (DEF).

The primary strata for the survey include PADDs 2–4, three sub-PADDs within PADD 1, and the two subparts of PADD 5 (California and the West Coast region excluding California), for a total of eight regions. There were a total of 38 sampling strata that were formed by substratifying the primary strata using a service-station category and the four truck stop size categories.

EIA designed the new EIA-888 sample based on a sample size of 590 outlets, which is an increase of 187 outlets from the previous sample size of 403 outlets. The larger sample size is needed due to changes in the sampling and estimation methodologies. First, outlets that are identified as out of business, out of scope, or nonrespondents are no longer replaced with other outlets from the frame that were not selected in the initial sample. Based on the previous EIA-888 sample, EIA expects an attrition rate of about 10%, but the attrition rate may vary among the strata. Second, the EIA-863 collection cycle was last conducted in 2011 for the 2010 survey year, and the sampling frame of outlets for on-highway diesel was updated using various administrative and third-party data sources that have varying degrees of accuracy. Third, although creating the sampling frames at the outlet level allows more control in the sample design in terms of certain characteristics of the outlets such as their geography and company ownership, sales volume information is limited, and the increased sample size reflects this lack of accurate information on the measure size for the outlets in the sample design. Constraints on the minimum sample size for a given stratum and the maximum sampling weight were used in allocating the sample to the strata. Due to changes in the population over time, we expect the total sample size for the next new sample to be approximately 650 outlets. Birth samples of newly identified outlets will be selected when needed, based on annual assessments of the frame.

The EIA-888 survey is conducted every Monday (Tuesday on federal holidays), and data are released on EIA's website around 5:00 p.m. each Monday (Tuesday on federal holidays). Customers can also sign up to receive the data via email. The U.S., PADD, sub-PADD, and California level retail on-highway diesel average prices are made available on EIA's prerecorded telephone hotline at 202-586-6966 and in the publications *Weekly Petroleum Status Report* and *This Week in Petroleum* (TWIP).

Form EIA-888 estimation

In the first phase of this survey, a measure of size—annual sales of on-highway diesel—will be collected from all respondents in the sample with additional collection for new birth samples. These volumes will be multiplied each week by the reported outlet diesel prices to get volume-weighted average prices and measures of sampling variability.

Imputation for volume nonrespondents will be carried out via donor imputation. The donor pool of potential donors for a given nonrespondent will be based on respondents having similar characteristics as the nonrespondent with respect to location, type of outlet, and size category, preferably in the same stratum. Expert judgment may be used to exclude outliers that are not deemed sufficiently similar. Volumes of on-highway diesel to be imputed are randomly selected from the viable donor pool, incorporating the sampling weights of the potential donors, and assigned to unreported outlet volume. For outlets that sell on-highway diesel at different prices for heavy trucks versus cars and light trucks but

respondents are unable to split their sales volume to correspond with the two prices, the truck volume will similarly be imputed via donor imputation.

To estimate weekly prices, first, any price nonrespondents will be imputed using the following weekly price imputation formula for prices by nonrespondent outlet as follows:

$$\hat{P}_{i,t} = P_{i,t-1} + \frac{\sum_j V_j \left(P_{j,t} - P_{j,t-1}\right)}{\sum_j V_j}$$

where,

$\hat{P}_{i,t}$ = imputation price estimate for nonrespondent price for on-highway diesel for a nonrespondent outlet (*i*) during current week (*t*);

$P_{i,t-1}$ = the outlet's (*i*) prior week's (*t-1*) reported price for on-highway diesel;

V_j = the annual volume reported of on-highway diesel for another individual outlet (*j*) within the nonrespondent outlet's (*i*) sample stratum;

$P_{j,t}$ = the current week's (*t*) reported price for on-highway diesel for another individual outlet (*j*) within the nonrespondent outlet's (*i*) sample stratum;

$P_{j,t-1}$ = the prior week's (*t-1*) reported price for on-highway diesel for another individual outlet (*j*) within the nonrespondent outlet's (*i*) sample stratum; and

$\Sigma_{j=}$ summation over all other stations within the nonrespondent outlet's (*i*) sample stratum.

Then, the reported and imputed prices will be multiplied by the corresponding sales volumes. The following notation, similar to the EIA-878, is used here:

k = type of outlet (secondary stratum based on truck-stop size or service station);

j = sampling region (primary stratum based on PADD or sub-PADD);

i = outlet;

N_{jk} = population size (number of outlets in sampling region *j* and type of outlet *k*);

n_{jk} = sample size in sampling region *j* and type of outlet *k*;

P = price;

V = volume;

W_{jk} = sampling weights = inverses of probability of selection; and

f_{jk} = denotes the sampling fraction in region *j* and type of outlet *k*, for certainty strata f_{jk} = 1.

By definition,

$$W_{jki} = \frac{N_{jk}}{n_{jk}}$$

Define $\quad x_{jki} = W_{jki} V_{jki}$

$$y_{jki} = W_{jki} V_{jki} P_{jki}$$

$$x_j = \sum_{i,k}^{\square} x_{jki}$$

$$y_j = \sum_{i,k}^{\square} y_{jki}$$

Then the volume-weighted average price for a region is as follows: $\hat{P}_j = \dfrac{y_j}{x_j}$

Relvariance (relative variance) is calculated by dividing the estimated variance of a statistic by the square of the statistic.

The relvariance of $\hat{P}_j$ is as follows:

$$V^2_{P_j} = V^2_{x_j} + V^2_{y_j} - 2 V_{x_j y_j}$$

where,

$$V^2_{x_j} = \frac{\sigma^2_{x_j}}{x^2_j} \quad \text{and} \quad V_{x_j y_j} = \frac{\sigma_{x_j y_j}}{(x_j)(y_j)}$$

The estimated variance for x_j is as follows (the variance of y_j is defined similarly):

$$\hat{\sigma}^2_{x_j} = \sum_k^{\square} (1 - f_{jk}) \frac{n_{jk}}{n_{jk} - 1} \sum_i^{\square} (x_{jki} - \bar{x}_{jk})^2$$

where,

$$\bar{x}_{jk} = \frac{\sum_i^{\square} x_{jki}}{n_{jk}}$$

The estimated covariance is as follows:

$$\hat{\sigma}_{x_j y_j} = \sum_k^{\square} \left(1 - f_{jk}\right) \frac{n_{jk}}{n_{jk} - 1} \sum_i^{\square} \left(x_{jki} - \overline{x}_{jk}\right)\left(y_{jki} - \overline{y}_{jk}\right)$$

where,

$$\overline{y}_{jk} = \frac{\sum_i^{\square} y_{jki}}{n_{jk}}$$

The volume-weighted average price for a publication region, which is made up of one or more sampling regions *j*, is as follows:

$$\hat{P} = \frac{\sum_j^{\square} y_j}{\sum_j^{\square} x_j}$$

The estimated relvariance of *p* is given as follows:

$$\hat{V}^2_P = \frac{\sum_j^{\square} \hat{\sigma}^2_{x_j}}{\left(\sum_j^{\square} x_j\right)^2} + \frac{\sum_j^{\square} \hat{\sigma}^2_{y_j}}{\left(\sum_j^{\square} y_j\right)^2} - \frac{2\sum_j^{\square} \hat{\sigma}^2_{x_j y_j}}{\left(\sum_j^{\square} y_j\right)\left(\sum_j^{\square} x_j\right)}$$

where,

RSE(P) = $\left(\hat{V}^2 ¿¿ P\right)¿^{(1/2)}$

B.3. Maximizing Response Rates

All the units that are surveyed may not respond (unit nonresponse) or may not provide all the information requested (item nonresponse). Alternative modes of data collection and follow-up are employed to encourage maximum response to the surveys in the Petroleum Marketing Program (PMP). Respondents are allowed to report by mail, fax, phone, or electronically using Excel forms or a fillable PDF available from the survey directory on EIA's website.

The nonresponse strategy for each of the monthly surveys is to generate a follow-up email or phone call within five days of the reporting deadline. Late respondents on both the weekly and monthly surveys are emailed or called and asked to submit their data. If a weekly respondent still fails to respond after our initial reminder during data collection, secondary contacts are emailed or called for data. If a firm repeatedly fails to respond, a noncompliance letter requesting submission by a specific date is sent. Table B2 contains the average annual response rate for the weekly and monthly surveys from January 2023 to December 2023. For EIA-877, the weeks covered are from the most recent heating season, which took place from October 2023 to March 2024.

Table B2. 2023 average annual response rates° for PMP surveys

Survey	EIA-14	EIA-182	EIA-856	EIA-877^±	EIA-878#	EIA-888*
Response rate	99.5%	100%	99.3%	94.4%	89.5%	93.3%

° Response rates are based on the number of respondents included in the data released on EIA's website. Response rate = respondents reporting/total eligible respondents. No weighting is involved. All response rates shown above are unweighted.
^EIA-877 response rate is based on the October 2023 to March 2024 heating season for both heating oil and propane.
± For EIA-877, weighted item response rates based on annual sales for heating oil and propane are 94.6% and 97.4%, respectively.
#For EIA-878, the weighted item response rate based on annual sales volume for all formulations of regular motor gasoline at the U.S. level is 91.4%.
*For EIA-888, the weighted item response rate based on annual sales volume for all formulations of diesel fuel at the U.S. level is 94.2%.

B.4. Test Procedures and Form Consultations

In August 2018, EIA conducted research on the ability of EIA survey respondents to report retail price information for No. 2 heating oil and propane. Specific objectives of this research project were:

- To determine how respondents are reporting the retail price for No. 2 heating oil and propane
- To better understand the price tiers offered by publicly traded companies that make retail sales of No. 2 heating oil and/or propane
- To determine if respondents are reporting the retail price for residential customers only
- To assess whether respondents can separate residential customers from their commercial and/or industrial customers
- To assess whether respondents are including any type of discount in the residential retail price they report every Monday
- To determine how companies use contract prices and how these differ from the retail price they report
- To check the burden per response on the form including how much time it takes respondents to report annual volumes

The findings from this research project show that most respondents to this survey, in general, correctly report their residential retail price offered to residential customers of No. 2 heating oil and propane every week exclusive of all discounts. The interviews also revealed that many respondents offer a contract price to their customers that is different from the retail price reported weekly on Form EIA-877. The contract price offered to residential customers is a predetermined price a customer pays for agreeing to purchase a certain quantity of heating fuel over a certain time period at a specific price. Participants indicated that these contracts are common in the industry for managing price risk and are a significant business element. As such, establishments that offer these contracts keep detailed records on the terms and price of their contracts and can distinguish their contract price information from the residential retail price they report to EIA on Form EIA-877.

EIA-878 Schedule B and EIA-888 Schedule B cognitive research study 2021
A cognitive research study was implemented to test both survey forms EIA-878 Schedule B and EIA-888 Schedule B to ensure that respondents would be able to accurately complete each form. The Survey

Development Team contacted 318 organizations for cognitive interviews. The invitations to participate in the cognitive interviews resulted in 10 interviews, one email response to EIA-878 Schedule B questions, and two no-shows. The response rate was approximately 3.45%. The EIA-878 cognitive results showed that the Instructions and Part 1: Identification Information sections had no comprehension problems by the majority of respondents. A major recommendation that was brought up by several respondents during the Part 1 section was to develop a multi-station form. In Part 3, an obstacle for respondents was retrieving annual motor gasoline volumes. It is important to note, that no one stated that it was impossible to retrieve this information, however, it appears to be time consuming for many of the current contacts reporting this information to EIA. Based on the responses, corporate-level employees may gather this type of information quicker and more efficiently than for station-level employees. The results for the EIA-888 cognitive study showed that most respondents had comprehension issues regarding the purpose of the survey form. When asked to paraphrase and define in their own words the purpose in the Instructions section, they were unable to accurately define the section. Overall, respondents did not have any issues with Part 1 of the survey form. There were no issues with the comprehension of diesel bays or heavy-duty trucks. However, there were some comprehension issues with diesel products like off-road diesel, No.1 diesel, and biodiesel. A retrieval issue found in the research was the inability for many respondents to split the diesel volumes between truck diesel and total diesel. Based on the feedback and recommendations received from the cognitive study, the program office was able to implement various changes that further improved Form EIA-878 Schedule B and Form EIA-888 Schedule B.

B.5. Statistical Consultations

Publicly available studies and research papers prepared by EIA statisticians and contractors regarding surveys in the PMP are available upon request. This list includes only publicly available reports. In addition, staff worked with contractor Z, Inc. to conduct a quality assessment in fiscal year 2011 that serves as a basis for future survey changes, and with staff from the Office of Statistical Methods & Research (SMR) on cognitive testing and usability studies for the surveys in the PMP.

PMP staff met with numerous internal data users—staff that work on the *Annual Energy Outlook* (AEO), *International Energy Outlook* (IEO*), Short-Term Energy Outlook* (STEO), and State Energy Database System (SEDS)—to consider their needs. In addition, staff also gave presentations at the following conferences to obtain feedback from external data users:

- American Statistical Association (ASA) Conference (2009 and 2010)
- State Energy Data Needs Workshop (2009)
- Energy Markets and Financial Initiative (2010)
- EIA's Annual Energy Conference (2011)
- Kauffman Foundation Forum on Establishment Surveys (2011)
- Federal Committee for Statistics and Methodology (FCSM) (2012)
- FCSM Statistical Policy Conference (2016)
- FCSM Research and Policy Conference (2018)

- Joint Statistical Meetings—American Statistical Association (2018)

Contact for the Petroleum Marketing Program: Rosalyn Berry, Supervisory Survey Statistician, Office of Energy Production, Conversion & Delivery (EPCD), petroleummarketingprogram@eia.gov.

For information concerning this request for OMB approval, please contact the agency Forms Clearance Officer Gerson Morales at 202-586-7077 or Gerson.Morales@eia.gov.