



Independent Statistics & Analysis

U.S. Energy Information
Administration

Supporting Statement for Petroleum Marketing Program

Part B: Collections of Information Employing Statistical Methods

Form EIA-14, Refiners' Monthly Cost Report

Form EIA-182, Domestic Crude Oil First Purchase Report

Form EIA-782A, Refiners'/Gas Plant Operators' Monthly Petroleum Product Sales Report

Form EIA-782C, Monthly Report of Prime Supplier Sales of Petroleum Products Sold for Local Consumption

Form EIA-821, Annual Fuel Oil and Kerosene Sales Report

Form EIA-856, Monthly Foreign Crude Oil Acquisition Report

Form EIA-863, Petroleum Product Sales Identification Survey

Form EIA-877, Winter Heating Fuels Telephone Survey

Form EIA-878, Motor Gasoline Price Survey

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

OMB No. 1905-0174

July 2019



Independent Statistics & Analysis

www.eia.gov

U.S. Department of Energy

Washington, DC 20585

Table of Contents

Part B: Collections of Information Employing Statistical Methods.....	i
B.1. Respondent Universe.....	1
B.2. Statistical Methods.....	4
B.3. Maximizing Response Rates.....	16
B.4. Test Procedures and Form Consultations.....	16
B.5. Statistical Consultations.....	17

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. This includes firms which purchase domestic or foreign crude oil (Forms EIA-182 and EIA-856); firms which refine crude oil into finished petroleum products (Form EIA-14); and firms which supply and/or sell finished petroleum products to customers (Forms EIA-782A, EIA-782C, and EIA-821). 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 Petroleum Product Sales Identification Survey (Form EIA-863) is used to build and maintain the frame for Forms EIA-821, EIA-877, and EIA-888 and has a target population of all firms who sell finished petroleum products. 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 Petroleum Marketing Frame

The frame for Form EIA-863, "Petroleum Product Sales Identification Survey," is constructed from the 2006 and 2010 data reported in the EIA-863 collection cycles conducted in 2007 and 2011, respectively, and from external source lists of petroleum marketers. The frame is maintained by the Office of Petroleum and Biofuels Statistics (PBS). Starting in 2019, PBS will supplement this information with outside source lists of petroleum marketers from State Energy and Tax Offices, Petroleum Marketing Associations, administrative records from governmental sources (e.g., U.S. Bureau of Labor Statistics' (BLS) Quarterly Census of Employment and Wages (QCEW), Secretary of States' business registries), and third-party private data sources (e.g., Oil Price Information Service and the Equifax database). The target population for Form EIA-863 is estimated to be approximately 60,000 companies. This is an increase from the 22,000 companies from 2010 due to possible undercoverage of individually- owned motor gasoline stations from the main source list used in previous cycles of the EIA-863 and the methodology used to compile this information. In the event that the selection of a new sample from the EIA-863 frame results in a disconnect in the data series, EIA includes explanatory text referencing changes in the sample design in the survey methodology description that accompanies the data series.

B.1.2 Monthly Crude Oil Surveys Frames and Target Population

Form EIA-14, “Refiners' Monthly Cost Report”

The target population for this survey is all refiners of crude oil. The frame for Form EIA-14 was constructed from a list of 206 refiners obtained from the Oil and Gas Journal in 1983. The frame is updated periodically via information derived from Form EIA-782A, “Refiners’/Gas Plant Operators’ Monthly Petroleum Product Sales Report,” and Form EIA-810, “Monthly Refinery Report.” There are currently 69 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 in a first purchase transaction. The frame for Form EIA-182 was initially compiled from the 1974 Federal Energy Administration (FEA) 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 percent 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 100 firms in 2018.

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 40 respondents.

B.1.3 Monthly and Annual Petroleum Product Frames and Target Populations

The target population for Form EIA-863 is all firms that sell petroleum products. The firms surveyed on Form EIA-863, along with their associated volumetric data and contact information, serve solely or partially as the sampling frame for Forms EIA-821, “Annual Fuel Oil and Kerosene Sales Report,” EIA-877, “Winter Heating Fuels Telephone Survey,” EIA-878, “Weekly Motor Gasoline Price Survey,” EIA-888, “Weekly On-Highway Diesel Price Survey.” The samples for three weekly surveys, EIA-877, EIA-878, and

EIA-888 are selected from an outlet based frame. The sample for the EIA-821 will continue to be selected from the EIA-863 and will benefit from updating the EIA-863 frame file.

Form EIA-821, “Annual Fuel Oil and Kerosene Sales Report”

The frame for this survey is constructed from Form EIA-863 survey results, supplemented by retailers/resellers and importers of residual fuel oil who were not identified by Form EIA-863. Currently, the sampling frame consists of over 20,000 companies.

Form EIA-782A, “Refiners’ Gas Plant Operators’ Monthly Petroleum Product Sales Report”

The target population for this survey includes the universe of refiners and gas plant operators. Firms that own and operate a refinery or gas plant and sell any of the 14 products listed on the form are required to file Form EIA-782A. The original frame was derived from a consolidated list of refiners known to have reported on several EIA surveys, the frame of gas plant operators from Form EIA-64, “Natural Gas Liquids Operations Report,” and companies identified as refiners or gas plant operators on Form EIA-460, “Petroleum Industry Monthly Report for Product Prices.” In 2018 the frame consists of 100 companies.

Form EIA-782C, “Monthly Report of Prime Supplier Sales of Petroleum Products Sold for Local Consumption”

The target population for this survey includes all suppliers who make the first sale of any of the products listed on Form EIA-782C, and deliver that product into a state for consumption in that state. The product slate includes: motor gasoline, No. 1 distillate, kerosene, fuel oil, diesel fuel, aviation gasoline, jet fuel, No. 4 fuel oil, residual fuel oil, and propane. The original frame was constructed from the 1981 respondent frame of the former Form EIA-25, “Prime Supplier’s Monthly Report.” Currently the frame consists of 200 prime suppliers and is updated as needed due to births, deaths, and mergers.

B.1.4. Weekly Petroleum Product Frames and Target Populations

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. These frames were then adjusted for births and deaths that were identified by EIA 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 6,000 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 9,000 outlets in 38 states in the East Coast, Midwest, Gulf Coast, and Rocky Mountain Regions. To ensure accurate coverage, the two frames are updated annually prior to selection of the samples.

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. The gasoline outlet frame was constructed 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 registries. 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 non-overlapping and one or more sampling regions will comprise a publication cell.

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 the States of Alaska and Hawaii were excluded from the target population. The EIA-888 frame was constructed 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 62,000 service stations and 4,000 truck stops. Based on information from other EIA survey data, the four largest sellers of on-highway diesel in the nation were identified. This allowed for classifying the outlets into three categories: service stations, midsized truck stops, and the top four largest sellers of on-highway diesel fuel.

B.2. Statistical Methods

Six of the ten petroleum marketing surveys— Forms EIA-14, EIA-182, EIA-782A, EIA-782C, EIA-856, and EIA-863—are census surveys so no weighting is required. Imputation is used to account for nonresponse on all surveys in the Petroleum Marketing Program except the EIA-863 which is a survey used for maintaining a petroleum frame file. The target population for Form EIA-856 is all companies with over 500,000 barrels of foreign crude oil in the report month for importation into the United States. Monthly estimates of volume weighted prices are generated from the data reported on Forms EIA-14, EIA-182, EIA-782A, 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-821, EIA-877, and EIA-878, and EIA-888, sample weights are calculated 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 corresponding total volume. Subsequently, the data are sorted by crude 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 Monthly Petroleum Product Surveys

Form EIA-782A and EIA-782C are two monthly petroleum product surveys that are a census. The target population for each of these surveys is small. The volume weighted price is calculated from data reported on Form EIA-782A. The total revenue (price * volume) is divided by a corresponding total volume to arrive at a volume weighted average selling price.

- **Form EIA-782A Aggregation:** Data from this survey are used to estimate the national, Petroleum Administration for Defense Districts (PADD), and state weighted average price for each product and sales-type category. The price and volume data for each company are multiplied and then aggregated across all companies for each product and market level to obtain a total revenue figure. This revenue is then divided by the corresponding total volume to arrive at a volume weighted average selling price.
- **Form EIA-782C Aggregation:** Form EIA-782C collects sale volumes of refined petroleum products. The only estimation procedures used are for summing across companies to calculate state, regional, and national volume totals.

B.2.3 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	Sample Size
EIA-877, "Winter Heating Fuels Telephone Survey"	Residential heating oil sellers: Stratified systematic random sample of approximately 1,320 outlets for 21 states and DC in the East Coast and Midwest Regions. Residential propane sellers: Stratified systematic random sample of approximately 1,680 outlets for 38 states in the East Coast, Midwest, Gulf Coast, and Rocky Mountain Regions.	See section 1	3,000
EIA-878, "Motor Gasoline Price Survey"	Stratified systematic simple random sample.	See section 2	1,000
EIA-888, "On-Highway Diesel Fuel Price Survey"	Stratified Probability Proportional to Size sample of retail outlets from 48 continental states and DC.	See section 3	403

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, which is measured in the same units (current dollars per gallon for weekly gasoline or No. 2 heating oil prices) as the estimate, 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. The coefficient of variation, which is

used as a measure of variability for weekly on-highway no. 2 diesel fuel may also be referred to as the relative standard error, and is the standard error expressed as a fraction of the estimate.

Each average gasoline price estimate published by EIA 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 the Motor Gasoline Price 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.

EIA will publish estimated standard errors for the new sample selected to report on Form EIA-877.

- 1) ***EIA-877 Sample and Estimation:*** 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 current 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. Approximately 1,320 outlets are selected from the frame of residential heating oil sellers, and approximately 1,680 outlets are selected

from the frame of residential propane sellers resulting in a total sample size of 3,000 respondents. The sample size for a given state is determined by how it compares to other states in terms of number of outlets represented on the frame, variability in weekly price and annual sales volume based on available data, and nonresponse in the current sample. Each sampled outlet is assigned a sampling weight that is the inverse of its probability of selection in the sample.

The total sample size increased from 1,506 outlets to 3,000 outlets due to changes in the sampling methodology. First, outlets that are identified as out of business, out of scope, or nonrespondents will no longer be replaced with other outlets from the frame that were not selected in the initial sample size. Second, 9 years have passed since the last EIA-863 collection cycle 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 expect more attrition for outlets in the heating oil sample that are out of business or out of scope. The sample size is 18% higher for heating oil compared to proportional allocation based on the expected number of outlets on the heating oil and propane sampling frames. 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.

Weekly price data are collected by State Energy Offices under the U.S. Energy Information Administration (EIA) State Heating Oil and Propane Program (SHOPP). When sample updates are made, EIA 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.

After the initial sample of outlets has been selected, available sources of information on new outlets since the sampling frames were constructed will be analyzed each year until the sample is redesigned, with the goal of selecting an independent birth sample annually to augment the

initial sample. In designing these annual birth samples, geographic regions will be 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.

(2) 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 U.S. 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.

Outlet names, physical addresses, and ZIP codes were obtained 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. 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.

Each year, the sample will be augmented to account for new outlets that are established. 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 will be 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 1,000 selected 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 two different methods based on the level of nonresponse:

1.) Only Regular grade gasoline volume reported:

If a respondent provided only annual sales volumes for regular grade gasoline for a particular outlet, but it is known that the respondent also sold midgrade and/or premium grade gasoline for that survey period at that particular outlet, imputation of annual midgrade and/or premium grade gasoline volumes are based on the proportion of volumes of midgrade and/or premium grade gasoline reported for the most recent complete calendar year on the EIA-782C Monthly Report of Prime Supplier Sales of Petroleum Products Sold for Local Consumption for the U.S. state/district of the respondent's outlet.

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

$$V_i = \frac{V_r}{\left(\frac{P_r}{100}\right)} \times \left(\frac{P_m}{100}\right)$$

Where,

V_i = Annual volume of either midgrade or premium grade gasoline for a given outlet to be imputed.

V_r = Reported annual volume of regular grade gasoline for a given outlet.

P_r = Percentage of annual volume of regular grade gasoline sold with respect to the total annual gasoline reported on the EIA-782C for the outlet's state/district.

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

Imputation example:

- Outlet A in State B reported 10,000 gallons of regular grade gasoline sold in Year C, but did not provide any volumetric data for premium grade gasoline sold 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 premium grade gasoline volume for Outlet A for Year C would be imputed as follows:

$$\frac{10,000 \text{ gal. reg. gas}}{85\% \text{ reg. gas in State B for Year C}} \times \frac{10\% \text{ pre. gas in State B for Year C}}{100} = 1,176 \text{ gal. pre. g}$$

2.) No gasoline volumes reported:

If a respondent provided no volumetric data, imputation is carried out via a donor imputation method based on an outlet's given sample stratum with each of the 71 sample stratum forming individual donor pools. The donor imputation procedure is as follows:

- Unreported grade of gasoline volumes are identified for outlets based on information on whether outlets sell regular grade, midgrade, and/or premium grade gasoline.
- The sample stratum for each unreported volume is identified and treated as a viable pool for random donor imputation with each reported volume by grade being a possible donor for an outlet's unreported grade of gasoline volume.
- A subject matter expert reviews each sample stratum to remove large outlier volumes by grade from donor pool to ensure imputed values do not overly weight a stratum.
- Volume(s) of gasoline by grade to be imputed are randomly selected from viable donor pool and assigned to unreported outlet volume.

The recurring weekly survey collects the cash price (\$/gal.) 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
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

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

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

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

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

Then the volume weighted average price for a region is as follows:

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

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:

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

where

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

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

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

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

The estimated covariance is as follows:

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

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

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:

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

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

Weekly price imputation for a nonrespondent outlet's price by grade of gasoline are based on a volume weighted average of price increases from the prior week to the current week for other stations located in the nonrespondent's outlet sample stratum.

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

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

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) sample stratum.

$P_{j,t}$ = The current week's (t) reported price for the unreported grade of gasoline for another individual outlet (j) within the nonrespondent outlet's (i) sample stratum.

$P_{j,t-1}$ = 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) sample stratum.

\sum_j = Summation over all other stations within the nonrespondent outlet's (i) sample stratum.

A larger sample size of approximately 1,000 outlets was needed because of the higher variances using annual volumes as weights versus the previous methodology that used the number of pumps located at an outlet as a measure of size

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 – i.e., state and sub-PADD formulations. The sample size is approximately 1,000 outlets with outlets added annually as part of supplemental birth samples. The survey is conducted every Monday (Tuesday on Federal holidays), and more frequently during emergency situations. Data are released on [EIA's website](#) around 5 p.m. each Monday (Tuesday on Federal holidays). Data are made available through email notification to those customers who sign up for that service. The U.S., PADD, sub-PADD, state, and city levels 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 \(TWIP\)](#).

(3) Form EIA-888 Sample Design: The respondents reporting to the weekly diesel price survey represent a stratified probability proportional to size (PPS) sample selected from a frame list of retail outlets. The outlet sampling frame was constructed using commercially available lists from several sources in order to provide comprehensive coverage of truck stops and service stations that sell on-highway diesel fuel in the United States. The frame includes about 62,000 service stations and 4,000 truck stops. Due to statistical and operational considerations, outlets in the States of Alaska and Hawaii are excluded from the target population.

The primary publication cells of the survey include PADDs 2-4, three sub-PADDs within PADD 1, and the two subparts of PADD 5 (the State of California and the West Coast region excluding California). The U.S., the East Coast (PADD 1), and the West Coast (PADD 5) are considered secondary publication cells since their prices are aggregated based on the prices from their primary publication cell components. To select the sample, allocations were first assigned to all primary publication cells through a simulation of coefficients of variation of average prices using historical price data. The target coefficient of variation for each primary publication cell was capped at one percent (1%). Allocations were further assigned to the states covered by each primary publication cell. The distribution of allocations was proportional to the annual state total volume of retail on-highway diesel fuel sales. This allocation procedure yielded a total target sample size of 403 retail outlets. The states were treated as sampling strata in the sample design.

It is believed that on-highway diesel fuel sold through all service stations combined only accounts for a small portion of the retail on-highway diesel fuel market. The truck stops on the frame were also classified into two categories, dependent on whether they belong to the nation's four largest on-highway diesel sellers. Based on information from other survey data and industry sources the proportions of total diesel volumes sold by outlets in the three categories - service stations, mid-sized truck stops, and top four - was assumed to be 20, 55, and 25 percent, respectively. These volume proportions, along with the outlet counts for the three categories on the frame, were used to calculate relative size measures for the outlets in each of the three categories. Pareto Sampling, which is a PPS procedure, and the size measures for each outlet were then used to select sampling units from each state.

The survey is conducted every Monday (Tuesday on Federal holidays), and data are released on [EIA's website](#) around 5 p.m. each Monday (Tuesday on Federal holidays). Data are made available

through email notification to those customers who sign up for that service. The U.S., PADD, sub-PADD, and the State of California levels 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: The reported and imputed prices each week are aggregated in multiple steps to obtain price estimates for publication cells. First, state average prices are calculated as simple unweighted averages of reported and imputed prices. Volumes of on-highway diesel sold in the states, as published by the Federal Highway Administration (FHWA), are then used to weight the state average prices and obtain average prices for primary publication cells. These weights are updated annually when FHWA data are available. Average prices for secondary publication cells are weighted averages of primary publication cell prices based on the proportion of diesel volumes attributable to their component primary publication cells.

B.2.4 Statistical Methods for Annual Petroleum Product Survey

Form EIA-821 Sample Design: The target population for Form EIA-821, "Annual Fuel Oil and Kerosene Sales Report" survey is the universe of all active companies that sell distillate fuel oil, residual fuel oil, or kerosene in the 50 states and the District of Columbia.

The 2006 Form EIA-863 database provided a base sampling frame for the Form EIA-821 survey. Form EIA-863 was mailed to approximately 24,000 companies in January 2007 to collect 2006 state-level sales volume data for No. 2 distillate fuel, residual fuel oil, motor gasoline, and propane. Companies also indicated if they sold kerosene. The No. 2 distillate fuel data were further identified by residential No. 2 fuel oil and by nonresidential retail and wholesale for No. 2 fuel oil and No. 2 diesel fuel; the residual fuel oil data were identified by retail and wholesale. In addition, company/state-level volumes for distillate fuel, residual fuel oil, and kerosene from the 2008 Form EIA-821, 2008 Form EIA-782A, and 2008 Form EIA-782B were also merged with the 2006 Form EIA-863 data. The integrated and comprehensive frame was then used to design and select the 2009 Form EIA-821 sample, which the survey is based.

To select a sample for Form EIA-821, subsidiaries and parents of a company were merged by adding the volumes of parents and subsidiaries in a cluster (i.e., parent-subsidiary combination) to represent the company. The sample was drawn from a multi-attribute frame with four target variables of No. 2

residential fuel oil, No. 2 nonresidential fuel oil, No. 2 nonresidential diesel fuel, and No. 2 wholesale distillate fuel.

A company was classified as a certainty company if it met one of the following criteria:

1. The company (or one of its subsidiaries) was a refiner as identified in the 2008 Form EIA-782A survey.
2. The company had residual fuel oil sales.
3. The company sold any Form EIA-821 product in at least five states.
4. The sum of maximum percentage of the four distillate products at the state level across states was five percent or more.
5. The company reported five percent or more of the total weighted volume in any state for any specified product by end-use category in the 2008 Form EIA-821 survey.

A systematic probability proportional to size design (PPS) was used to sample noncertainty companies. Company State Units (CSUs) were the sampling units. A CSU selected by the sampling procedure was referred to as a “basic” CSU. A company was included in the sample if it had at least one “basic” CSU. All non-“basic” CSUs of a sampled company were referred as “volunteer” CSUs.

In each state, the Dalenius-Hodges procedure was used to stratify CSUs, with each of the four target distillate variables, into zero, low, medium, and high volume four strata. Neyman allocation was used to obtain the sample size for each stratum to meet the target coefficient of variation of five percent. The population of CSUs was divided into mutually exclusive cells by crossing the four stratifications such that every CSU in a particular cell was in the same stratum for each of the four stratifications. Each CSU was assigned a probability of selection, which was the largest sample proportion across all four stratifications. All CSUs within a cell had the same probability of selection. A systematic PPS sample of CSUs was then drawn for the state.

This design produced a final sample of approximately 4,000 companies. Selected companies were asked to report sales by end-use categories for distillate fuel, residual fuel oil, and kerosene.

Form EIA-821 Estimation: For obtaining total estimates of volume, the adjusted probability estimator is used. This estimator, the sum of the weighted volumes, is defined as follows:

$V = \sum_h (\sum_i W_{ih} V_{ih})$, where:

V = total estimated volume,

\sum_h = summation over strata,

\sum_i = summation over units within stratum h ,

W_{ih} = weight attached to unit i in stratum h

(the reciprocal of the probability of selection, P_{ih} , for that unit), and

V_{ih} = volume reported or imputed for units i in stratum h .

Survey nonrespondent volumes are also imputed as the mean of their strata.

B.3. Maximizing Response Rates

All the units that are surveyed may not respond (unit non-response) or may not provide all the information requested (item non-response). 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 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. The strategy for the annual survey is to generate a follow-up email or phone call within two weeks of the reporting deadline and on a monthly basis thereafter. Five months after the reporting deadline, a noncompliance letter requesting submission by a specific date is mailed. Table B2 contains the average annual response rate for the weekly, monthly, and annual surveys from January 2018 to December 2018. The EIA-821 has not finalized data collection for 2018 data.

Table B2: 2018 Average Annual Response Rates for PMP Surveys

Survey	EIA-14	EIA-182	EIA-782A	EIA-782C	EIA-821*	EIA-856	EIA-877	EIA-878	EIA-888
Response Rate	99.9%	100%	99.8%	99.8%	90.8%	100%	99.2%	95.9%	97.5%

*EIA-821 response rate based on 2017 data. Response rates are based on the number of respondents included in the data released on EIA website. Response rate = respondents reporting/total respondents. There is no weighting involved. All response rates shown above are unweighted.

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:

- 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 are able to 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; and
- 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 are able to distinguish their contract price information from the residential retail price they report to EIA on Form EIA-877.

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 FY2011 which serves as a basis for future survey changes, and with staff from the Office of Survey Development and Statistical Integration (SDSI) on cognitive testing of the surveys in the PMP.

PMP staff met with numerous internal data users – 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 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: Ms. Tammy Heppner, Supervisory Statistician, Office of Petroleum and Biofuels Statistics (PBS), 202-586-4748.

For information concerning this request for OMB approval, please contact the agency Forms Clearance Officer Faouzi Aloulou, at 202-586-1344, or faouzi.aloulou@eia.gov.