

# PETROLEUM MARKETING PROGRAM SUPPORTING STATEMENT

EIA's Supporting Statement for 2013 Petroleum Marketing Surveys: OMB Number 1905-0174



## Introduction

To help accomplish EIA's overall mission to provide high quality, informative energy data and analyses, the Office of Petroleum and Biofuels will employ ten statistical surveys to gather relevant, accurate and timely information from a variety of business establishments that participate in the production and distribution of crude oil and various refined petroleum products in the United States and its Territories. Within this context, the standard practices applicable and relevant to the design and operation of these statistical surveys are identified below:

- *Section 1: Frame and Target Population.* Reporting thresholds can be used to determine a survey frame and thereby exclude smaller companies from reporting. Exclusions result in a frame that does not completely cover the target population of companies of all sizes. While many of the frames are stable and have small populations, other surveys have large frames which are very dynamic over time.
- *Section 2: Sampling Methodology and Estimation Procedures.* A data collection may be based on a sample of the population rather than on a complete enumeration. Variation occurs by chance because a particular sample is surveyed. This variation decreases with increased sample size. A preferred sample size is determined in order to keep key statistics under a specified threshold of sampling error, but sometimes the actual sample size is solely determined by cost. The estimation methodology and variance of the population also impact reliability of published data.
- *Section 3: Maximizing Response Rate.* All the units that are surveyed may not respond (unit non-response) or may not provide all the information requested (item non-response). The strategy to minimize nonresponse is discussed in Section 3.
- *Section 4: Tests of Procedures.* Numerous test procedures and evaluations were conducted to enhance the information quality of the disseminated data and to minimize any potential for measurement error.
- *Section 5: Statistical Consultation.* An internal peer review and cognitive testing yielded revisions to survey forms and instructions. The purpose of these initiatives was to ensure consistency across forms, especially with regard to data collected regarding petroleum products. Testing of the revised

survey forms and instructions with respondents provided feedback used to redesign the forms and instructions. These revisions seek to ensure consistency of the data collected from respondents.

Numerous staff, contractors, respondents, and data users were involved in the redesign of the Petroleum Marketing Program. This included consultation with both internal and external data users as indicated in Appendix L.

Additional information regarding survey and statistical methods is available in Technical Notes ([PDF](#)) of the Petroleum Marketing Monthly.

## 1. Frame and Target Population

The Petroleum Marketing Program collects data from two types of entities in general. In terms of the market chain, 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 (EIA-182, EIA-856); firms which refine crude oil into finished petroleum products (EIA-14); and firms which supply and/or sell finished petroleum products to customers (EIA-782A; EIA-782C; EIA-821, EIA-877). The remaining surveys focus on individual outlets selling the products to consumers (EIA-878, EIA-888) and collect price data on a weekly basis. The Petroleum Product Sales Identification Survey (EIA-863) is used to build and maintain the frame for several of these surveys. The frames for the monthly surveys are kept current using information from other surveys as well as information from industry journals and other sources.

### A. Petroleum Marketing Frame:

Frame building for the Petroleum Product Sales Identification Survey (EIA-863) begins with the EIA-863 frame constructed from the previous cycle of the EIA-863, relevant data records from an internal frame system maintained by the Office of Petroleum and Biofuels Statistics, and outside source lists of petroleum marketers from State Energy Offices and Petroleum

Marketing Associations. The target population for the EIA-863 includes over 22,000 companies.

B. Monthly Crude Oil Surveys Frames and Target Population

- EIA-14, “Refiners' Monthly Cost Report” The target population for this survey is all refiners of crude oil. The frame for the Refiners' Monthly Cost Report (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 the EIA-782A (Refiners'/Gas Plant Operators' Monthly Petroleum Product Sales Report) and the EIA-810 (Monthly Refinery Report). There are currently 72 active respondents filing the Form EIA-14.
- 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 the Domestic Crude Oil First Purchase Report (EIA-182) was compiled from the 1974 Federal Energy Administration (FEA) Oil and Gas Survey of Producers and Operators. The size of the frame has declined from 340 firms in 1978 to 82 firms in 2012. The original frame was derived from a consolidated list of refiners known to have reported on several EIA surveys; and the frame of gas plant operators from the EIA-64, “Natural Gas Liquids Operations Report.”

The list frame 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, 198 purchased more than 150,000 barrels per year and together represented 99.9 percent of the total reported volume. Following decontrol 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.

- 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 for the Monthly Foreign Crude Oil Acquisition Report (EIA-856) is composed of all companies reporting data in June 1982 on the Transfer Pricing Report (ERA-51) regardless of the total volumes of crude oil that were imported and all companies acquiring more than 500,000 barrels of oil in the reference month for importation into the United States. Currently the frame consists of 43 respondents.

### C. Monthly and Annual Petroleum Product Frames and Target Populations

The target population for the EIA-863 is all firms which sell petroleum products. The firms surveyed on the EIA-863, along with their associated volumetric data and tracking information, serve as the sampling frame for Forms EIA-821, “Annual Fuel Oil and Kerosene Sales Report,” EIA-877, “Winter Heating Fuels Telephone Survey,” EIA-878, Daily Motor Gasoline Price Survey,” EIA-888 “Weekly On-Highway Diesel Price Survey,” and other ad hoc surveys, such as the National Petroleum Council Surveys.

- EIA-821, “Annual Fuel Oil and Kerosene Sales Report.” The frame for this survey is constructed from the EIA-863 survey results, supplemented by retailers/resellers and importers of residual fuel oil who were not identified by the EIA-863 survey. Currently, the sampling frame consists of over 22,000 companies.
- 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. The frame for this survey was derived from a list of refiners that reported on past EIA surveys and a list of gas plant operators derived from the Natural Gas Liquids Operations Report. In 2012 the frame consists of 95 companies. The original frame was derived from a

consolidated list of refiners known to have reported on several EIA surveys; and the frame of gas plant operators from the EIA-64, Natural Gas Liquids Operations Report.

- 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, residual fuel oil, and propane. The original frame was constructed from the respondent frame of the former 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.

#### D. Weekly Petroleum Product Frames and Target Population

- EIA-877, “Winter Heating Fuels Telephone Survey.” The company/state level frame for this survey is constructed from company-state units identified in the Petroleum Product Sales Identification Survey (EIA-863). The current sampling frame for units which sell either residential heating oil or propane or heating oil consists of 14,000 companies in 24 states. For residential propane sellers, an outlet level frame is used for the second phase selection of individual outlets to represent the companies selected in the first phase. This frame currently includes about 4,200 active outlets.
- EIA-878, “Motor Gasoline Price Survey.” The sample for the Motor Gasoline Price Survey was drawn from a frame of approximately 115,000 retail gasoline outlets. The gasoline outlet frame was constructed by combining outlet level name and address information purchased from the Oil Price Information Service (OPIS) with information contained on the EIA-782A and B surveys from sellers of motor gasoline through retail outlets. Additional information was obtained directly from companies selling retail gasoline to supplement information deficient on the OPIS list. The individual frame outlets were mapped to counties using their zip codes. The outlets were then assigned to the published geographic areas as defined by the EPA program area, or for conventional gasoline areas, as defined by the

Census Bureau's Standard Metropolitan Statistical Areas (SMSA) using their county assignment.

- EIA-888, "On-Highway Diesel Fuel Price Survey." 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.

## 2. Sampling Methodology and Estimation Procedures

Six of the ten petroleum marketing surveys—the EIA-14, the EIA-182, the EIA-782A, the EIA-782C, the EIA-856, and the EIA-863—are censuses so no weighting is required. The EIA-856 is a cut-off sample and the target population is all companies with over 500,000 barrels of foreign crude oil in the report month for importation into the United States. One respondent with less than 500,000 barrels is required to report because they were originally on the frame in 1982. The volume weighted price is used for estimation in all of these surveys. The total cost, price times volume, is divided by a corresponding total volume to arrive at a volume weighted average cost.

For the remaining surveys, sample weights are typically 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 times a sales volume measure for the outlet.

### A. Sampling Methodology and Statistical Procedures for Monthly Crude Oil Surveys

- 1) The EIA-14 is a census.
- 2) The 182 is a census. To obtain a national weighted average price, the total cost (amount paid times 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.

3) The EIA-856 is a cut-off sample, with a census of those companies importing over 500,000 barrels of foreign crude oil in the report month for importation into the United States for one respondent who is required to report that was on the frame in 1982.

**B. Sampling Methodology and Statistical Procedures for Monthly Petroleum Product Surveys**

There are two monthly petroleum product surveys and both are a census since the target population for each of these surveys is small. The volume weighted price is used for estimation in all of these surveys. The total cost, price times volume, is divided by a corresponding total volume to arrive at a volume weighted average cost.

1. EIA-782A Aggregation: Data from this survey is used to estimate the national average price for each product and market level - e.g., U.S., PADD, and State. 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 corresponding total volume to arrive at a volume weighted average price.
2. EIA-782C Aggregation: Because the EIA-782C is a census survey and only totals are published, the only estimation procedures used are for summing across companies.

**C. Sampling Methodology and Estimation Procedures for Weekly Petroleum Product Surveys**

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

<b>Table B1: Summary of Sample Design</b>			
<b>Survey Name</b>	<b>Sample Design</b>	<b>Selection Procedure</b>	<b>Sample Size</b>
EIA-877, "Winter Heating Fuels Telephone Survey"	Stratified random sample with certainty for 21 States and DC in the East Coast and Midwest Regions	See section A	527 outlets



		below	
EIA-877, "Winter Heating Fuels Telephone Survey"	Two-stage sample for 24 States in the East Coast and Midwest Region. Stratified Probability Proportional to Size (PPS) sample of Company-State Units in the first stage and simple random sample of outlets in the second stage.	See section A below	720 outlets
EIA-878. "Motor Gasoline Price Survey"	Stratified simple random sample.	See section B below	801
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 C below	403

- 1) *EIA-877 Sample and Estimation:* Separate sampling and estimates procedures are used for residential no. 2 heating oil and propane. These procedures are described below.

*Sample and Estimation Procedures for No. 2 Heating Oil Prices on EIA-877:* The No.2 heating oil price data are reported by a statistical sample. According to the requirement of the State Heating Oil and Propane Program (SHOPP) program, 21 States and the District of Columbia (DC) in the East Coast and Midwest regions participate in the No.2 heating oil price survey. The sampling frame used was a list of all Company State Units (CSUs) in those 21 States and DC that reported residential No. 2 heating oil sales on the 2006 Form EIA-863, "Petroleum Product Sales Identification Survey." CSUs that sold at least five percent (5%) of the residential No.2 heating oil in a State, as reported on the frame survey EIA-863, were automatically included in the sample and are referred to as certainty units. The remaining CSUs, referred to as non-certainty units, were stratified into three groups by their residential No.2 heating oil sales volumes in each State, as reported on the EIA-863. Strata boundaries were determined using the Dalenius-Hodges procedure. The sample allocations were designed generally to yield average price coefficients of variation (CV) of one percent (1%), but, due to budget constraints, individual State sample sizes were capped at 35 even if the target CV was not met. In addition, a minimum size of fifteen was required for each of the 21 States. The sample of

CSUs within each stratum was a simple random sample. The residential No.2 heating oil sample size inclusive of certainty and non-certainty units is 527 CSUs.

The residential No. 2 heating oil prices (excluding taxes) for a given State are based on the results of a telephone survey of marketers and refiners, one for each of the two products. Data are collected by State Energy Offices under the Energy Information Administration (EIA) State Heating Oil and Propane Program (SHOPP).

To estimate the average residential No. 2 heating oil price data for a State, the sample and volume weights are applied to the reported price, summed and divided by the sum of the weighted volume:

$$\frac{\sum_{j=1}^s \sum_{i=1}^{n_j} w_{ij} v_{ij} p_{ij}}{\sum_{j=1}^s \sum_{i=1}^{n_j} w_{ij} v_{ij}}$$

where  $w_{ij}$  = sample weight

$v_{ij}$  - volume weight

$p_{ij}$  = price,  
 $i$  = respondent

$n_j$  = sample size of stratum  $j$ , and

$s$  = number of strata

The sample weights  $w_{ij}$  were calculated as ratios of population number of CSUs to the sampled number of CSUs in each stratum. Volume weights  $v_{ij}$  were assigned using the data reported in the frame survey.

State level residential No. 2 heating oil average prices are then aggregated into regional and overall averages with State level total residential No. 2 heating oil volumes as weights.

These fixed volume weights indicate the relative importance of the individual companies according to the size of their sales at the time of the frame. Therefore, changes in the average price across time reflect only the change in the price being offered by the company, and not the change in the amounts sold. Price indexes constructed using fixed volumes, such as these annual sales, are known as Laspeyres Indexes. One alternative method of weighting, used in Paasche Indexes, uses current weights. This method would require each company to report the number of gallons sold at the reported price each pricing period and would be more burdensome on the companies. Both methods of weighting are correct but provide different averages particularly when volumes are changing. It has been argued in the literature that during periods of change, the Laspeyres method has a tendency to overestimate price changes, while the Paasche method tends to underestimate price changes.

In this survey, it is expected that the change in volumes weekly during the heating season is small. Residential sales are not bulk in nature and do not tend to reflect discounts on price for large volume purchases. Absolute changes in volume with a year's time would more likely reflect demand and be consistent across companies within a geographic area. Therefore, even though the volume weights used in the calculation of average prices in the SHOPP tend to lag behind the actual volumes sold in the reference period, fixed volumes are used to reduce company burden and enable timely release of average prices.

*Sample and Estimation Procedures for Propane Prices on EIA-877:* Propane price data are collected from a sample of companies for 24 States in the East Coast and Midwest regions and the District of Columbia. The sampling frame used was a list of all Company State Units (CSUs) in those 24 States that reported residential propane sales on the 2006 Form EIA-863, "Petroleum Product Sales Identification Survey." The population of the survey was first stratified by State, which is the publication cell. Due to high residential propane price variation and budget constraints, sample sizes of all strata were limited to 30 even though the target CV of one percent (1%) was not met in many States. To select the sample, the CSUs in each State were ordered by zip code in order to control for the geographic location of the companies. A Probability Proportional to Size (PPS) Systematic Sample with Probability Minimum Replacement, using the propane volumes reported on the frame survey EIA-863 as a measure of sampling unit size, was then selected from this ordering. With the

targeted maximum sample size of 30 CSUs in each State, any CSU that sold more than 3.3 percent (1/30) of the residential propane in a State was selected at least once. Within each sampled CSU, a simple random sample of residential propane outlets was drawn by using an outlet address listing EIA developed with information provided by the industry and State energy officials. The number of outlets selected from each CSU was the same as the number of times that CSU was selected in the PPS sample. The resulting total number of outlets selected to report on the propane sample was 720.

In cases where there were fewer outlets in a sampled CSU than the number of times that CSU was sampled in the PPS sample, all outlets for that CSU were selected and their weights were adjusted to  $n'_i/n_i$ , where  $n'_i$  is the number of times that CSU was sampled and  $n_i$  is the number of outlets sampled. This was also the practice if a CSU preferred to report on the survey by providing the average of all its outlets in a given State. Therefore, the actual number responding each month may deviate from the 720 outlets sampled.

To estimate the average residential propane price data for a State, a simple average of the prices from each sampled outlet yields a valid estimate as a result of the cancellation of sample weights of the PPS sample design and volume weights in the estimate:

$$\frac{1}{n} \sum_{i=1}^n \frac{n'_i}{n_i} p_i$$

where  $p_i$  = price,

$i$  = outlet respondent,

$n$  = sample size, and

$n'_i/n_i$  = weight adjustment as described in the previous paragraph

Regional and overall averages are calculated as averages of the State level residential propane prices with State level total residential propane volumes as weights.

- 2) *EIA-878 Sample*: The sample for the Motor Gasoline Price Survey was drawn from a frame of approximately 115,000 retail gasoline outlets. The gasoline outlet frame was constructed by

combining information purchased from a private commercial source with information contained on existing EIA petroleum product frames and surveys. Outlet names and zip codes were obtained from the private commercial data source. Additional information was obtained directly from companies selling retail gasoline to supplement information on the frame. The individual frame outlets were mapped to counties using their zip codes. The outlets were then assigned to the published geographic areas as defined by the EPA program area, or for conventional gasoline areas, as defined by the Census Bureau's Standard Metropolitan Statistical Areas (SMSA) by using their county assignment.

The gasoline outlet sample is an area sample that is comprised of both an augmentation to, and rotation of the previous sample cycle of the gasoline survey, the EIA-878. The augmentation outlets were obtained by first, sampling counties, and then, sampling the outlets from the gasoline outlet frame within those counties within each sampling cell<sup>[1]</sup>. Every county in the U.S. was assigned to the corresponding sampling cell as defined. After the counties were assigned, the standard deviations of gasoline prices for these sampling cells were estimated using the prices from the previous sample of the gasoline survey. These deviations and the number of stations from the Census Bureau's County Business Patterns (CBP) were used to determine the required number of outlets to be sampled. The statistical technique used was the Chromy allocation algorithm, an iterative procedure to determine the number of units required for each sampling cell. To select the sample of outlets, counties within each sampling cell were sorted within states, and the required number of outlets selected from the outlet frame randomly. Once the augmentation portion of the sample was obtained, standard deviations were re-estimated, combining the previous gasoline sample outlets and newly sampled outlets. The Chromy algorithm was applied again to determine the revised sample cell requirements. The previous sample's outlets were then sub-sampled to insure a self-weighting sample within each stratum, and allocations satisfied by sampling half from each of the self-weighting sub-sample and the old sample.

Prior to the development of the outlet frame, only company level data were available. Therefore, the previous samples required two-phase sample design to select the outlets. In the first phase of the design, retail gasoline companies were selected with probability proportional to the total

volume in each state they sold gasoline, as reported in the EIA Monthly Petroleum Product Sales Report survey. The second phase, the selection of individual outlets from the selected companies, was performed using information obtained directly from the sampled companies during the sample initiation. This design permitted the use of a simple average for estimating average prices for city and state gasoline prices, but required volume weighted prices for more aggregated published areas with respect to geography, formulation, and grade. However, with the publication of additional city and state average prices, this design approach was insufficient and required redesign with the increase of the geographic detail to include nine states and ten cities.

*EIA-878 Estimation:* To estimate average prices, sample weights were constructed based on the sampled outlet's number of pumps, a proxy for sales volume. These weights are applied each week to the reported outlet gasoline prices to obtain averages for the specific formulations, grades and geographic areas. Weights used in aggregating grades, formulations and geographic areas were derived using volume data from the EIA "Monthly Report of Prime Suppliers Sales of Petroleum Products Sold for Local Consumption", and demographic data from the U.S. Census Bureau and the Department of Transportation on population, number of gasoline stations and number of vehicles.

The target coefficient of variation was set for 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 - e.g. state and sub-PADD formulations. The sample size is approximately 800 outlets. The survey is conducted every Monday (Tuesday on Federal holidays), and more frequently during emergency situations. Data are released on [EIA's website](#) by 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) *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 Petroleum Administration for Defense Districts (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](#) by 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 \(TWIP\)](#).

*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 in 2010, as published by the Federal Highway Administration, are then used to weight the State average prices and obtain average prices for primary publication cells. 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.

#### 4) Sampling Methodology and Estimation Procedures for Annual Petroleum Product Survey

*EIA-821 Sample Design:* The target population for the 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 EIA-863 database provided a base sampling frame for the EIA-821 survey. The EIA-863, "Petroleum Product Sales Identification Survey," 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 EIA-821 survey, 2008 EIA-782A survey ("Refiners'/Gas Plant Operators'



Monthly Petroleum Product Sales Report”), and 2008 EIA-782B survey (“Resellers'/Retailers' Monthly Petroleum Product Sales Report”) were also merged with the 2006 EIA-863 data. The integrated and comprehensive frame was then used to design and select the 2009 EIA-821 sample, which the 2010 survey is based.

To select a sample for the EIA-821 survey, 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:

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

*EIA-821 Estimation Procedures:* 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.

### 3. Maximizing Response Rates

Alternative modes of data collection and follow-up are employed to encourage maximum response to the survey in the Petroleum Marketing Program (PMP). Respondents are allowed to report by mail, fax, phone, or electronically using excel forms available from the [survey directory](#) on EIA's web site.

The nonresponse strategy for each of the monthly surveys is to generate a follow-up letter or email 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 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 monthly and weekly surveys from January 2011 to December 2011.

**Table B2: 2011 Average Annual Response Rate for Weekly and Monthly PMP Surveys**

Survey	EIA-14	EIA-182	EIA-782A	EIA-782C	EIA-856	EIA-877	EIA-878	EIA-888
Response Rate	99.7%	100%	99.0%	98.8%	100%	98.8%	98.9%	99.6%

---

## 4. Tests of Procedures

In preparation for clearance of the 2013 Petroleum Marketing Program by the Office of Management and Budget (OMB), the Survey Development Team (SDT) conducted an assessment of the current forms and proposed recommendations that would improve the forms for future iterations. The evaluation focused primarily on efficient form design and valid data collections in order to improve data quality, reduce respondent burden, and improve process efficiency at EIA.

The first step of the process was to conduct internal early-stage scoping interviews with the OPBS managers and analysts. Early-stage scoping is a method that uses semi-structured interviews to understand substantive issues from the subject's point of view using the subject's language. The primary objective of this method is to systematically gather and document common form design issues, unclear instructions, and problematic data requests across the petroleum forms. Based on the findings from the early-stage scoping interviews, the SDT may recommend record-keeping studies, cognitive interviews, and/or usability testing with external respondents.

The team conducted internal early-stage scoping interviews. In-person interviews were conducted with survey managers and analysts and typically lasted about one hour. None of the interviewers were audio recorded. Individuals that participated in the interviews were selected based upon recommendations from the Office of Petroleum and Biofuels Statistics.

The Survey Development Team conducted seven interviews for the EIA-782C during the winter of 2011. Respondents were selected for testing based on recommendation by PBS' team leads and included larger sized companies. All interviews were conducted over the telephone and usually lasted about 45 minutes. Usually, SDT conducts early-stage scoping interviews in-person because face-to-face interviews provide richer information, and allow us to see respondent's records. However due to resource constraints this was not possible.

A summary of additional pre-clearance activities, involving meeting with stakeholders – both internal and external - to enhance information quality is available in the following 2012 Federal Committee for Statistics and Methodology (FCSM) paper, [Achieving Information Quality via Continuous Quality Improvement](#).

## 5. Statistical Consultations

Publicly available studies and research papers prepared by EIA statisticians and contractors regarding surveys in the Petroleum Marketing Program are available in Appendix M. This list includes only publicly available reports. In addition, staff worked with contractor, Z, Inc. to conduct a quality assessment in FY2011 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 – AER, AEO, IEO, SEDS, and STEO - 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 (2012)

Contacts for the Petroleum Marketing Program include:

- Ms. Shawna Waugh, Mathematical Statistician, Office of Petroleum and Biofuels Statistics (PBS), (202) 586-6484
- Mr. Jason Worrall, title, Office of Survey Design and Statistical Development (SDSI), (202) 586-6075