

## B. COLLECTING INFORMATION BY STATISTICAL METHODS

### A. “Refiners’ Monthly Cost Report”: EIA-14

#### 1. Description of the Survey Plan

The target population for the EIA-14 survey is all refiners of crude oil. There were approximately 206 refiners originally identified from the *Oil and Gas Journal* when the survey began in 1983. Of the 206 refiners on the original list, 68 are currently active and filing Form EIA-14. The frame is kept current using information from other EIA surveys.

#### 2. Sampling Methodology and Estimation Procedures

- a. Sampling. There is no sampling for the EIA-14 since the universe is small.
- b. Estimation Procedures. As in all petroleum price surveys, a volume weighted price is used. Total cost (price times volume) is divided by a corresponding total volume to arrive at a volume weighted average cost.

#### 3. Maximizing Response Rates

To encourage maximum response to the EIA-14, alternative reporting methods are provided. Respondents are allowed to report by mail, fax, phone, or electronically through the excel forms available on EIA’s web site. For nonresponse, a nonrespondent listing is generated within five days of the reporting deadline. Nonrespondent firms are telephoned and asked to submit data. If a firm still does not respond, a noncompliance letter requesting submission by a specific date is sent. The average response rate for the EIA-14 for reference months April 2008 thru March 2009 was 100 percent.

#### 4. Tests of Procedures

Procedures for conducting the EIA-14 survey have been successfully employed for more than 5 years.

#### 5. Statistical Consultations

Ms. Paula Mason of the Petroleum Division, Office of Oil and Gas, (202) 586-1262, is responsible for the statistical aspects of this survey. The Project Manager for the EIA-14 survey is Elizabeth Scott who can be contacted at (202) 586-1258. The contractor responsible for collection and processing of the survey data is:

ABACUS Technology Corporation  
5454 Wisconsin Avenue, Suite 1100  
Chevy Chase, MD 20815

**B. “Domestic Crude Oil First Purchase Report”: EIA-182**

1. Description of the Survey Plan

The target population for the EIA-182 survey is all firms that buy domestic crude oil at the lease boundary, acquiring ownership of the crude in a first purchase transaction. The list initially was 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. Currently, the EIA-182 frame list consists of 82 active firms.

2. Sampling Methodology and Estimation Procedures

- a. Sampling. There is no sampling performed for the EIA-182.
- b. Estimation Procedures. Total cost (amount paid times volume purchased) is divided by corresponding total volume to arrive at a national weighted average price.

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 (plus the Outer Continental Shelf regions, Alaska North Slope and Alaska Other).

Imputation procedures are used to account for missing data and outliers as follows:  
Outliers - Imputation is performed when reported data fail standardized edit checks. The data are imputed by obtaining the month-to-month percentage change for the item in question for all respondents excluding the respondent in question and applying that change to the respondent's prior reporting month's value. Imputation can be for both volume and/or cost, and is noted in the processing system.

Non-respondents – Imputation for non-respondents' volumes is performed automatically, each month, by one program in the First Purchase System. Program-imputed volumes that are generated in report format, are used for intermediate crude oil production estimation purposes only. The volumes are not used to generate published prices in the *PMM*. Published prices are derived from either respondent or manually imputed data only. The data are imputed in the same manner as data which failed the edit checks.

Respondents may make revisions to original data. These revisions are posted to the data base. Most revisions are within 30 days of the original submissions. Hence, the data are initially published as “Preliminary” data. The data for the previous month are revised, if

necessary, and published as “Final.” Revisions beyond the previous month are made if resubmissions, corrections, or late filings significantly alter the former final price. These revisions are included in the *Petroleum Marketing Annual (PMA)* and published in the *Petroleum Supply Monthly (PSM)* concurrent with the publication of the *PMA*.

3. Maximizing the Response Rate

To encourage maximum response to the EIA-182, alternative reporting methods are provided. Respondents are allowed to report by mail, fax, phone, or electronically through the excel forms available on EIA’s web site. In addition, the form is mailed out monthly with a business reply return envelope or mailing label included. Use of the postage paid return envelope/ mailing label increases survey response rates and lowers overall survey costs. For nonresponse, a nonrespondent listing is generated within five days of the reporting deadline. Nonrespondent firms are telephoned and requested to submit data. If a firm still does not respond, a noncompliance letter is sent requesting submission by a specific date. The average response rate for the EIA-182 for reference months April 2008 thru March 2009 was 100 percent.

4. Tests of Procedures

Procedures for conducting the EIA-182 survey have been successfully employed for more than 5 years.

5. Statistical Consultations

Ms. Paula Mason of the Petroleum Division, Office of Oil and Gas, is responsible for the statistical aspects of this survey. The Project Manager for the EIA-182 survey is David Gatton who can be contacted at (202) 586-5995. The contractor responsible for collection and processing of the survey data is:

ABACUS Technology Corporation  
5454 Wisconsin Avenue, Suite 1100  
Chevy Chase, MD 20815

C. “Resellers’/Retailers’ Monthly Petroleum Product Sales Report”: EIA-782B

1. Description of the Survey Plan

The EIA-782B survey has a target population of all resellers of motor gasoline, and resellers and retailers of No. 2 distillate, residual fuel oil, and propane. The original EIA-782B used a frame of distillate fuel oil dealers that consisted of respondents to the EIA-402, “Fuel Oil Identification Survey,” a one-time survey implemented in 1979 and mailed to a listing of 30,000 fuel oil related businesses. The frame was updated using data from the EIA-9A, No. 2 Distillate Price Monitoring Report (the predecessor to the EIA-782B), and the EIA-172, Sales of Fuel Oil and Kerosene.

The current EIA-782B survey sample is referred to as sample 15. It was initiated in 2004, using the EIA-863 (reference year 2002), Petroleum Products Sales Identification Survey, as the sampling frame. The EIA-863 survey was mailed to a listing of approximately 25,000 petroleum products sellers. In addition, data from the EIA-821 for reference year 2002 were mapped to the EIA-863 file because EIA-821 respondents were not required to file the EIA-863. Using this frame file, and data from the previous 782B and EIA-821 samples, a national sample of motor gasoline, distillate fuel oil, and residual fuel oil resellers and retailers was designed and selected. The original sample size was 2,067 companies. The reporting sample decreases through time as businesses sell, merge and go out of business. A company is required to report their sales in all states that they sell petroleum products. If a company is reporting for several states on the frame, each state they report for is treated as a separate company/state reporting unit for sampling purposes because sample allocations are defined at the state level. The current stratification table indicating population and initial sample sizes is available upon request.

## 2. Sampling Methodology and Estimation Procedures

- a. Accuracy Criteria. The required level of accuracy for each of ten target variables was defined by a volume coefficient of variation (CV) of 15 percent at the published State level for No. 2 distillate, and 10 percent for motor gasoline, residual fuel oil, and propane. Given monthly operational budget limitations, these levels were chosen because they yielded samples sizes that best met budget requirements and program requirements. A description of these target variables is contained in the sample design description below. Studies on the relationship of volume CV to price CV have shown that this is expected to produce price CV's of less than one percent. The reliability of current month estimates will vary from these goals due to the deterioration of the frame over time and the changing distribution of prices and volumes on a monthly basis.
- b. Sample Design. EIA-782A respondents (refiners) are selected with certainty due to their small number and because of the relative size of their sales volume. The sales volumes obtained from these surveys were used to assign measures of size for sampling. Nonrefiners comprising 5 percent or more of sales in a State were also selected with certainty. The remaining units on the frame were each assigned probability of selection. The EIA-782 sample uses Pareto sampling, a variant of Poisson sampling (a form of Probability Proportional to Size sampling) that is appropriate for use with permanent random numbers (PRNs). Pareto sampling assigns probabilities of selection  $p$  that add up to  $n$  and assigns a random number  $r$  between 0 and 1 to each unit. The  $n$  smallest values of  $(r-rp)/(p-rp)$  were selected. The EIA-782 design also uses a technique called "collocation" as a form of implicit stratification. Collocation is performed within a cell by applying a monotonic transformation that insures that there will be exactly one number within the interval  $[(k-1)/n, k/n]$  for each positive integer  $k$  less than or equal to  $n$ . For the EIA-782 the number is assigned randomly within the interval.

PRNs are used to draw the sample where the relative position or the value of the number determines whether the unit is sampled or not. The sampling scheme may even be changed for a subsequent sample. This is done using the same set of PRNs derived from the first sample and transforming it accordingly to draw the second sample.

The probabilities, based on the proportion of the volumes each company sells for each frame product, geographic area, and type of sale classification relative to the cell total, were calculated using an algorithm developed by Chromy (1987) and implemented in a SAS program by Laura Zayatz and Richard Sigman (1995). A minimum probability .01, was set so as to cap all weights at 100 to partially address frame errors or changes in the population. A fixed number of companies were sampled. Follow up was conducted to determine which (if any) of the original sample companies were no longer in business or no longer in scope for the survey. The number of companies was then allotted to companies in scope, so that additional companies were selected in sequence until the desired number was achieved according to the CV targeted. The first 2,200 companies (inclusive of refiners that file the EIA-782A) in this ordering were then selected for the sample.

The non-certainty companies were then post-stratified within each defined geographic area and sale's type combination by their volume. The sample weights, the inverse of the probabilities, were multiplied by the sample expectation adjustment which was the ratio of the sum of the probabilities of selection for all frame units in the stratum to the actual sample size of the stratum. The geographic areas were defined as (a) the 24 States in which No. 2 distillate was a significant heating source and 50 States and the District of Columbia for residual and motor gasoline, (b) the 25 States in which propane was a significant energy source, or as (c) the PAD Districts for districts where not all State estimates are provided. Four volume-of-sales strata (certainty, zero, low, and high) were defined with volume boundaries differing by State, sale's type, and product. The design of the EIA-782B sample was based on ten target variables: total retail motor gasoline, total wholesale motor gasoline, residential No. 2 fuel oil, other retail No. 2 fuel oil, total wholesale No. 2 fuel oil, residential propane, total other retail propane, wholesale propane, total retail residual fuel oil, and total wholesale residual fuel oil. A sample size of 2,200 was expected to yield a median level of accuracy for each target variable of volume coefficients of variation (CV) of 15 percent for No. 2 distillate and 10 percent for the other products, determined at the publishable State level (24 States for distillate, 25 for propane, 50 States and the District of Columbia for motor gasoline and residual). Studies on the relationship of volume CV to price CV have shown that this will produce price CVs of less than 1 percent. The reliability of current month estimates will vary from these goals due to the deterioration of the frame over time and the changing distributions of price and volume.

Certainty units are re-evaluated each sample selection/rotation and noncertainty units rotated at roughly 50%. For sample rotation, the random numbers used in

the previous cycles (in the sample at the time) were used as initial random numbers in the rotation procedure. New companies (births) were given new uniformly distributed random numbers. The entire set of random numbers was collocated again by home state and Metropolitan Statistical Area (MSA) status; the collocation simply fit the new cases uniformly among the old. A new set of PRNs was defined using the joint formulas:

- 1) If  $x > kp$  then  $x' = x - kp$
- 2) If  $x < kp$  then  $x' = x - kp + 1$

where  $x$  was the original PRN,  $p$  was the probability of selection of the unit and  $k$  was a constant.

The actual draw of sample 15, the current sample, began with the random numbers used to draw the previous sample 14, with new numbers assigned to any companies that were not in the previous frame. The numbers were then collocated within home state in order to spread out the random numbers through cells reflecting the home states of the companies. Hence, two random numbers were used in the process: one to determine the interval in which a company was placed and the other to determine its position in the interval. This procedure yields a geographically representative sample. Finally the rotation was applied to this new set of numbers, using .37 as the  $k$  in the equations above, the same coefficient that was used the previous cycle.

- c. Sample Rotation. To distribute respondent burden in an equitable manner, the EIA-782B sample is rotated as described in the sample design. The current sample was initiated in October 2004 and is referred to as sample 15. Approximately fifty percent of the non-certainty units in the sample were replaced by new ones randomly selected from the frame.
- d. Estimation Procedures. The EIA-782 used three stages for deriving the weights for estimation. The initial weights were simply the inverse of the probability of selection. The second stage was the adjustment that capped the weights. This was done because an examination of the sum of the inverse probabilities indicated that they could yield an overestimate of the number of company clusters in the frame. In the past this number has exceeded the number of companies in the frame. This seems to have been due to a large number of small probability companies making it into the sample. These companies had a probability of selection of .01, and hence a sampling weight of 100. During the previous cycle reducing the maximum weight to 75 brought the estimated number of companies much closer to the actual count. The third stage was the application of post-stratification weights. Each State and product was divided into up to four noncertainty strata (zero, low, medium and high) plus a certainty stratum (frame non-respondents will be combined with zeroes) as described in the sample design. These strata were collapsed to insure at least three sampled companies in each stratum, and then the ratio of the sum of the probabilities for the entire frame to

the number of sampled companies was multiplied by the weight of each company in the stratum. This had the effect of increasing the weights when the sample size for the stratum was smaller than expected and decreasing the weights when the sample size was bigger than expected. This final sampling weight was multiplied by the ratio of the population total to the sum of the weights for a specific product, so that the sum of these adjusted weights equals the population total for the specific product to produce an unbiased estimate.

Missing data (resulting from incomplete reporting, nonresponse, and values that fail editing) are imputed by weighting together the previous month's reported value and the previous month's predicted value to yield a predicted value (the geometric average) for the current month for each company. The sum of the weighted predicted values for nonrespondents in the current month is then multiplied by a chain link multiplier (the ratio of the sum of the weighted, reported values for respondents in the current month to the sum of the weighted, predicted values for respondents in the current month).

Volume estimates are calculated as the sum of the companies' volumes times the sample weights divided by the sum of the companies' weights. Price estimates are calculated as the sum of the product of each company's volume, price, and final adjusted weight divided by the sum of each company's volume times final adjusted weight.

### 3. Maximizing the Response Rate

To encourage maximum response to the EIA-782B, alternative reporting methods are provided. Respondents are allowed to report by mail, fax, phone, or electronically through the excel forms available on EIA's web site or the PC Electronic Data Reporting Option (PEDRO) software. In addition, the form is mailed out monthly with a business reply return envelope or mailing label included. Use of the postage paid return envelope/ mailing label increases survey response rates and lowers overall survey costs. For nonresponse, a nonrespondent listing is generated within five days of the reporting deadline. Nonrespondent firms are telephoned and requested to submit data. If a firm still does not respond, a noncompliance letter is sent requesting submission by a specific date. Additional noncompliance letters are sent as needed. The expected response rate for resellers and retailers, based on a 12 month performance from April 2008 thru March 2009 for the EIA-782B, is approximately 90 percent.

#### 4. Tests of Procedures

Sample design procedures have been modified and updated as sample requirements have changed. New sample rotations are overlapped for two reporting periods to smooth transition of samples and survey respondents. Statistical procedures for imputation/estimation have been in operation on the EIA-782 for more than 5 years. The methodology has been updated, as the industry, and sample and data requirements have changed. A more detailed description of procedures and methodology is available electronically in the “Explanatory Notes” of the *Petroleum Marketing Annual* at [http://www.eia.doe.gov/pub/oil\\_gas/petroleum/data\\_publications/petroleum\\_marketing\\_annual/current/pdf/enote.pdf](http://www.eia.doe.gov/pub/oil_gas/petroleum/data_publications/petroleum_marketing_annual/current/pdf/enote.pdf)

#### 5. Statistical Consultations

The respondent sample for the EIA-782B was designed and selected by:

ICF MACRO  
11785 Beltsville Drive  
Calverton, MD 20705

Ms. Paula Mason of the Petroleum Division, Office of Oil and Gas, (202) 586-1262, was the EIA Senior Statistician and project manager for the sample design, selection, initiation, as well as survey methodology. The Project Manager for the EIA-782 survey is Tammy Heppner who can be contacted at (202) 586-4748. The contractor responsible for collecting and processing the survey data currently under contract is:

ABACUS Technology Corporation  
5454 Wisconsin Avenue, Suite 1100  
Chevy Chase, MD 20815

#### **D. “Refiners’/Gas Plant Operators’ Monthly Petroleum Product Sales Report”: EIA-782A, and “Monthly Report of Prime Supplier Sales of Petroleum Products Sold for Local Consumption”: EIA-782C**

##### 1. Description of the Survey Plan

The target population for the EIA-782A includes the universe of refiners and gas plant operators. 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 frame is kept current using information from other EIA surveys as well as information from industry journals. The EIA-782A frame currently contains 95 entities reporting monthly. The actual response rate is approximately 99 percent. The target population for the EIA-782C includes all suppliers who make the first sale of any of the products listed on the 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 derived from the respondent frame of the former EIA-25, Prime Supplier's Monthly Report. The current frame has been supplemented with firms qualifying as prime suppliers identified from the EIA-863, Petroleum Products Sales Identification Survey, the EIA-782B, and other available sources. The EIA-782C frame is currently composed of 185 prime suppliers.

## 2. Sampling Methodology and Estimation Procedures

- a. Sampling. The EIA-782A and EIA-782C are census surveys and no sampling takes place.
- b. Estimation Procedures. For the EIA-782A the average price is calculated for each product and marketing level. The price and volume data for each company are multiplied and then aggregated across all companies to obtain a total revenue figure. This revenue is then divided by corresponding total volume to arrive at a volume weighted average price.

Because the EIA-782C is a census survey and only totals are published, the only estimation procedures used are for summing across companies. Missing data for the EIA-782A and EIA-782C are imputed using the same methodology as the EIA-782B.

## 3. Maximizing the Response Rate

The response rates for the EIA-782A and C are maximized in a similar manner as the EIA-782B, previously described with the exception that the forms are not mailed out. To minimize costs, the forms are not mailed out by EIA since respondents prefer alternate modes of transmission. The response rates for both the EIA-782A and EIA-782C surveys are approximately 99 percent.

## 4. Tests of Procedures

The procedures used for the EIA-782A and C have been successfully employed for more than 5 years. Further testing and comparison of methodology is an ongoing project. The methodology has been updated, as the industry and the data requirements have changed. A complete history of procedures and methodology is available electronically in the "Explanatory Notes" of the *Petroleum Marketing Annual* at: [http://www.eia.doe.gov/pub/oil\\_gas/petroleum/data\\_publications/petroleum\\_marketing\\_annual/current/pdf/enote.pdf](http://www.eia.doe.gov/pub/oil_gas/petroleum/data_publications/petroleum_marketing_annual/current/pdf/enote.pdf)

## 5. Statistical Consultations

Contractor and government personnel responsible for the EIA-782 survey series are listed in the section describing the EIA-782B.

**E. “Annual Fuel Oil and Kerosene Sales Report”: EIA-821**

1. Description of the Survey Plan

The target population for the EIA-821 includes all companies that deliver or sell fuel oil or kerosene to ultimate consumers (end-users). The survey’s scientifically drawn sample was selected from the EIA-863 sampling frame, and supplemented by retailers/resellers and importers of residual fuel oil who were not identified by the EIA-863 survey. The EIA-863 (2002) survey collected State-level sales information for calendar year 2002, including volumes of No. 2 distillate fuel oil, residual fuel oil, and motor gasoline sold to end-users and resellers. Most companies that sell only kerosene or distillate fuel oil other than No. 2 were not targeted by the EIA-863. Companies on the frame reporting only kerosene, #1, or #4 fuel are eligible for sampling as a zero volume distillate/residual fuel dealer. The sample was last selected for the 2003 reference year. The original sample size was 4,041 companies, but the size has been reduced since selection due to sales, mergers, and companies ceasing operation.

2. Sampling Methodology and Estimation Procedures

a. Accuracy Criteria. For the EIA-821, accuracy is defined and fixed in terms of relative error, and cost is defined, but not fixed, as respondent and government burden. The goal is to minimize burden, or sample size, while designing for a fixed relative error, or coefficient of variation. Five variables, or product sales categories, specify target coefficients of variation. These are:

- Residential No. 2 distillate volume
- Non-Residential No. 2 distillate volume
- Wholesale No. 2 distillate volume
- Retail residual fuel oil volume
- Wholesale residual fuel oil volume

(“Retail” and “Wholesale” are considered synonymous with “For End Use” and “For Resale” for ease of exposition). Coefficients of variation were targeted at 5 percent at the state and other aggregate levels. The sample design was based on the volumetric target primarily because volumes from the EIA-821 are published.

b. Certainty Strata. The following companies were declared certainties:

- a) The company (or one of its subsidiaries) was a refiner
- b) The company sold any EIA-821 products in at least five States
- c) The company’s sum across states of the maximum of the percentages of each of the three distillate products at the State level was five or more percent
- d) The company reported over five percent of the total weighted volume in any state for the specifically targeted product/end use categories (such as distillate vessel bunkering, distillate electric utility use, etc).
- e) The company reported residual fuel oil sales

The selection of these companies enhanced the efficiency of the sampling design, reduced sampling error and assisted in data continuity from year to year. For the 2003 sample, 746 companies were designated as certainty.

- c. Volumetric Stratification. After frame out-of-scope firms, nonrespondents and certainty elements were removed from consideration, the remaining companies in a State were cross-stratified according to volume of residential No. 2 distillate and the maximum of other retail and wholesale No. 2 distillate (i.e., a two dimensional stratification.). Nonrespondents were sampled as a separate stratum.

For the EIA-821, the number of strata, as well as the cutoff points, was specific to the product and the State. For each State and product there was a non-respondent stratum, a certainty stratum, and a zero volume stratum. The other State/product stratifications varied from one to three strata, representing low to high volume.

The Dalenius-Hodges procedure was used to define the boundaries for the strata. This procedure derived by T. Dalenius and J. L. Hodges, Jr. (Minimum Variance Stratification, *Journal of the American Statistical Association*, 54:88-101) was implemented separately for each State.

In particular, using the data from the frame for each State and distillate product the non-respondents, zero volume respondents and the designated certainties were removed and the Dalenius-Hodges procedure was used to define three stratum boundaries for the product for each State. This was done dividing volumes by 100,000 and by 10,000. If the division by 100,000 resulted in an initial number of categories that was too small, then the division by 10,000 was used. Fourteen different stratifications were obtained from the one to four noncertainty levels (possibly including zero). Residential distillate was crossed with the maximum stratum of nonresidential retail distillate and resale distillate, using the same conceptual stratification for nonresidential retail and resale distillate. For example, if for nonresidential retail the low and zero levels were combined and the high and medium were combined, then the same was done for resale distillate.

- d. Sample Allocation. Totals and standard deviations were calculated for each product and stratum. In order to account for the variability of the values over time, inflation factors were used. The factors used were 1.4, 1.2, and 1.0 for small, medium and large volumes respectively (if only two nonzero strata were defined, the inflation factors were 1.3 and 1.0, respectively). Neyman allocations were used for each product, and the maximum allocation was assigned to the cell. Allocations of 100% were used for the certainty stratum and half the sampling fraction was assigned for the combined noncertainty respondent stratum and the nonrespondent stratum when appropriate. A minimum of 3 respondents was selected from each cell. The allocations were designed to obtain a target Coefficient of Variation of 5 percent.

A random variable was used to draw the sample. A company was selected if one of its states was selected (company state unit, CSU). If a CSU was selected it is referred to as part of the “basic” sample. CSU from the same company cluster not selected are referred to as “volunteer”.

CSU weights were obtained dividing the population of the cell by the allocation. A company weight was obtained analytically, since the samples for different States were independently drawn. The formula used was :  $1/W = 1 - (1 - 1/w_1) \times (1 - 1/w_2) \times \dots (1 - 1/w_k)$  where  $W$  is the company weight, and  $w_1$  to  $w_k$  are the CSU weights. Company weights were provided through this procedure, but for an individual product the weights were adjusted at the stratum level so that the sum of the weights of sampled companies will equal the population at the stratum level.

- e. 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:

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

$$\hat{V} = \text{total estimated volume,}$$

$$\sum_h = \text{summation over strata,}$$

$$\sum_i = \text{summation over units within stratum } h,$$

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

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

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

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

### 3. Maximizing the Response Rate

To encourage maximum response to the EIA-821, alternative reporting methods are provided. Respondents are allowed to report by mail, fax, phone, or electronically through the excel forms available on EIA’s web site. In addition, the form is mailed out annually with a business reply envelope included. Use of the postage paid return envelope increases survey response rates and lowers overall survey costs. For nonresponse after due date, second request letters are mailed to all sample companies who have not responded two weeks after the filing deadline. If no response is received to the second request letter, telephone follow-up procedures are then used to solicit responses. If the telephone follow-up procedures are not successful, then data are

imputed for nonrespondents. The response rate for the EIA-821 survey was 87 percent for the reference year 2007.

4. Tests of Procedures

The procedures used for the EIA-821 survey have been successfully employed for more than five years. Minor changes are implemented as necessary to adjust to changes in the industry. Further testing and analysis are part of an ongoing project.

5. Statistical Consultations

The EIA-821 sample design and sample selection were performed under the guidance of Ms. Paula Mason of the Petroleum Division, Office of Oil and Gas, Energy Information Administration. Ms. Mason can be reached at (202) 586-1262. The Project Manager for the EIA-821 survey is Daniel Walzer who can be contacted at (202) 586-3511.

The contractor responsible for collecting and processing the survey data is:

Science Applications International Corporation  
1710 Goodridge Drive  
McLean, VA 22102

**F. “Monthly Foreign Crude Oil Acquisition Report”: EIA-856**

1. Description of the Survey Plan

All companies that were reporting data on the ERA-51, “Transfer Pricing Report,” as of June 1982, are required to prepare and submit an EIA-856 each month, regardless of the total volumes of crude oil that were imported. In addition, all other 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 an EIA-856 for that month. There are currently 42 companies reporting each month.

2. Sampling Methodology and Estimation Procedures

- a. Sampling. There is no sampling for the EIA-856 because of the small population.
- b. Estimation Procedures. Data are aggregated for publication by calculating a volume weighted average price.

3. Maximizing Response Rates

To encourage maximum response to the EIA-856, alternative reporting methods are provided. Respondents are allowed to report by mail, fax, phone, or electronically through the excel forms available on EIA’s web site. For nonresponse, a nonrespondent listing is generated within five days of the reporting deadline. Nonrespondent firms are

telephoned and asked to submit data. If a firm still does not respond, a noncompliance letter requesting submission by a specific date is sent. The average response rate for the EIA-856 for reference months April 2008 thru March 2009 was 100 percent.

4. Tests of Procedures

Procedures for conducting the EIA-856 survey have been successfully employed for more than 5 years.

5. Statistical Consultations

Ms. Paula Mason of the Petroleum Division, Office of Oil and Gas, (202) 586-1262, is responsible for the statistical aspects of this survey. The Project Manager for the EIA-856 survey is Elizabeth Scott who can be contacted at (202) 586-1258. The contractor responsible for collecting and processing the survey data is:

ABACUS Technology Corporation  
5454 Wisconsin Avenue, Suite 1100  
Chevy Chase, MD 20815

**G. “Petroleum Products Sales Identification Survey”: EIA-863**

1. Description of the Survey Plan

The purpose of the EIA-863 is to construct an enumeration of the universe of resellers and retailers of No. 2 distillate, propane, and residual fuel oil, and resellers of motor gasoline. The form also identifies companies that sell other petroleum products such as kerosene, No. 1 distillate, etc. The survey form collects information on annual sales of the major products by State and by sales category. The EIA-863 survey frame was developed from the match/merge of the predecessor survey EIA-863 (2002), the Dun & Bradstreet Market Identifiers File, State energy office lists, and association mailing lists. The frame is 24,400 companies.

2. Sampling Methodology and Estimation Procedures

- a. Sampling. The EIA-863 survey is a census survey. The respondent frame resulting from this survey will be used for sampling purposes by petroleum sample surveys.
- b. Estimation Procedures. The data received are edited using information from the previous EIA-863 survey, information from other EIA surveys, and information obtained from the States or industry organizations. Tabulations are processed for validation purposes but no estimation or imputation is performed. Company-level status and volume information is the end product.

### 3. Maximizing the Response Rate

To encourage maximum response to the EIA-863, alternative reporting methods are provided. Respondents are allowed to report by mail, fax, phone, or electronically through the excel forms available on EIA's web site. The survey form and instructions are mailed out with a business reply return envelope or mailing label included. Use of the postage paid return envelope/ mailing label increases survey response rates and lowers overall survey costs. For nonresponse after due date, second request letters are sent by mail to all companies who have not responded by the filing deadline. If no response is received to the second request letter, telephone follow-up procedures are then used to solicit responses. For post office returns, a concerted effort is made to obtain address corrections for mailing. The response rate for the EIA-863 for reference year 2006 was 88 percent.

### 4. Tests of Procedures

Procedures for conducting the EIA-863 (2010) survey will be similar to the procedures used in the 2006 survey. The procedures used to conduct the quadrennial survey are constantly reviewed for improvement.

### 5. Statistical Consultations

Ms. Paula Mason of the Petroleum Division, Office of Oil and Gas, (202) 586-1262, is responsible for the statistical aspects of this survey and is the project manager. The contractor and subcontractor responsible for collecting and processing the survey data are:

Science Applications International Corporation  
1710 Goodridge Drive  
McLean, VA 22102

ICF MACRO  
11785 Beltsville Drive  
Calverton, MD 20705

## H. “Winter Heating Fuels Telephone Survey”: EIA-877

### 1. Description of the Survey Plan

Approximately 1200 outlets were selected for the EIA-877 survey and were sent an initial letter explaining the survey. The selected companies are telephoned each week during the heating season (October 1 through March 15) to collect data on No. 2 heating oil and residential propane prices and stocks. If an emergency situation arises, the period of the heating season will be expanded. Most of the companies are telephoned by the State Energy Offices and the data are provided electronically to EIA for processing. The states are responsible for most of the data collection activities and submission of the price data

to EIA. EIA aggregates the data and the results are published electronically in the *Weekly Petroleum Status Report* and made available on Petroleum's website through Petroleum Navigator. Information on individual company or outlet sample weights or volumes in the sample is not shared with the states

## 2. Sampling Methodology and Estimation Procedures

For the No. 2 heating oil data, the sampling frame was derived from the current EIA-863, "Petroleum Product Sales Identification Survey" containing sales volume information by selecting active companies that reported sales of residential No. 2 heating oil. A stratified sample design on residential No. 2 heating oil sales volumes by State was used. Certainty stratum for each state was defined as 5% or more of sales volume in the state based on frame sales volumes.

The allocations were determined in two iterations. The first step involved a bootstrap on the previous heating oil sample and reported data to create a finite population based on replicating the company using the weights and then sampling with PPS from that finite population. A new population was sampled for each bootstrap sample. From each stratum, in each bootstrapped frame, a sample of the same size as the actual sample was drawn and an average price estimate was obtained. A standard error was calculated using the root mean square of deviations from the estimate yielded by the actual sample and divided by the mean to obtain the Coefficient of Variation (CV). Allocations were derived by dividing the CV by .01 (the target CV), the ratio was squared, and multiplied by the allocation of the previous 2001 sample. Maximum state level allocations were set to 35, and minimum allocations set to 15 to address budget and quality concerns.

A Neyman allocation algorithm was used to allocate the noncertainty cases to each of the three stratum, with a minimum of two per stratum. Because the EIA-863 volumes do not reflect the exact volumes of the CSUs several years later, we multiplied the standard errors of the non-certainty strata by inflation factors, to account for the likelihood that some companies have grown and others have contracted. The inflation factors used were 1.7 for the low stratum, 1.4 for the medium and 1.1 for the high (these factors were used in the EIA-782 design for a number of years and were found adequate for that survey). The factors reflect a differential likelihood that a CSU will drift out of a stratum or change its volume considerably. The frame was sorted by State and stratum, and then sampled randomly within each stratum, to obtain the stratified random sample. The original residential fuel oil sample selected in 2004 contained 522 companies. Sample weights were calculated as the inverse of the probability of selection ( $N/n$ ). The expected price coefficient of variation is one to two percent.

Unlike the heating oil sample, the propane sample used outlets as the primary selection unit. Similarly, the active companies that reported sales of residential propane on the EIA-863 formed the sampling frame for the propane portion of the EIA-877 survey. A separate sample design on residential propane sales volumes by state was used. A certainty stratum was defined as companies with volumes of 5% or greater in the state based on frame sales volumes. The certainty companies were mapped to a propane outlet level



file. The number of outlets for selection was determined by the multiple of five percent that each certainty company accounted for in the frame volume. For example, a certainty that represented 10% of the state's volume required that two outlets were selected.

In parallel to heating oil, the allocations were conducted in two iterations. The first step was a bootstrap on the previous residential propane sample. However, the bootstrap took into account the fact that a certainty unit could have multiple outlets. From each stratum in each bootstrapped frame, a sample of the same size as the actual sample was drawn and an average price estimate was obtained. A standard error was calculated from the estimate yielded by the actual sample. The standard error was divided by the mean to get the CV. Allocations were derived by dividing the CV by .01 (the target CV), the ratio squared, and multiplied by the allocation of the previous 2001 sample. Maximum state level allocations were set to 30, and minimum allocations set to 15.

A Neyman allocation algorithm was used to allocate the noncertainty cases to each of two stratum, with a minimum of two per stratum. Because the frame volumes did not reflect the exact volumes of the companies several years later, inflation factors were applied to the noncertainty strata standard errors to account for the likelihood that some companies have grown and others have contracted. The inflation factors used were 1.4 for the low stratum and 1.1 for the high based on the use of those factors successfully in the EIA-782 design. The factors reflected a differential likelihood that a CSU would drift out of a stratum or change its volume considerably.

For sample selection, noncertainty outlets were ordered by State and within stratum by zip code. Using a random starting point, outlets were sampled systematically, that is, every  $k$ th outlet was selected, where  $k$  is the inverse of the outlet-level stratum sampling fraction. Sampling weights for noncertainties in each State were assigned by taking the inverse of the probability of selection for that State and stratum, where the probability of selection for each State equaled the total number of outlets selected for the State, divided by the total number of outlets in the State. Volumes for sampled noncertainty outlets were calculated by dividing the total company volume by the number of noncertainty outlets on the frame representing the company. The original propane sample size as selected in 2004 was 643 outlets.

The name and address outlet list was constructed originally by extracting from the EIA-863 survey companies known to sell propane augmented by a list of individual propane outlets provided by industry associations, Dun and Bradstreet file of primary and secondary retail propane dealers, and respondents to other EIA surveys collecting any information on propane. This file has been maintained as births and deaths are reported on EIA surveys, and through intermittent updates through electronic sources.

Volume weighted average prices are estimated each week for residential propane and residential fuel oil by summing the product of each respondent's reported price by the frame volume and the sample weight and dividing the sum of the sample weighted volumes in each state. All companies in the sample are contacted by telephone to obtain a

high response rate. Nonrespondent firms are telephoned up to three (3) times and requested to submit data.

### 3. Maximizing Response Rates

All companies in the sample will be contacted by telephone to obtain the information. Telephone surveys generally have the highest response rates. Nonrespondent firms are telephoned up to three (3) times and requested to submit data. If a firm does not respond after three (3) attempts to obtain the data by telephone, a noncompliance letter requesting submission by a specific date is sent. The average response rate for 2008-2009 heating season was approximately 99 percent.

### 4. Tests of Procedures

These procedures have been followed for a number of years and no significant problems were encountered. The procedures used to conduct the survey are constantly reviewed for improvement.

### 5. Statistical Consultations

Ms. Paula Mason of the Petroleum Division, Office of Oil and Gas, (202) 586-1262, is responsible for the statistical design. The Project Manager for the EIA-877 survey is Marcela Rourk who can be contacted at (202) 586-4412. Each State Energy Office participating in the grants program is responsible for collecting the data from the sample provided by EIA for that particular state and submitting the data to EIA for aggregation.

## I. “Motor Gasoline Price Survey”: EIA-878

### 1. Description of the Survey Plan

Approximately 800 outlets were selected for the EIA-878 survey. The gasoline outlets selected for the EIA-878 survey are first initiated by telephone and confirmed to be in business. In-business outlets are informed of the purpose of the mandatory survey and informed of the confidentiality and protection afforded their data. Each week the individual outlets are called and asked to report the pump price, by grade, of unleaded gasoline. The collection takes place using a computer assisted telephone interview (CATI) with built in editing. Companies who prefer to report through the headquarters on behalf of their selected outlets are allowed to do so. Companies preferring to report by fax or email are also permitted to report by that method. Data obtained through non-phone methods are entered into the CATI system and treated the same as phone collected prices.

### 2. Sampling Methodology and Estimation Procedures

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 a private commercial data source. Additional information was obtained directly from companies selling retail gasoline to supplement information deficient on the commercial 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.

The gasoline outlet sample is an area sample comprised of both an augmentation to, and rotation of the previous sample cycle of the gasoline survey, the EIA-878 in order to insure continuity in the historical data series. 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. 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. A Goodman-Kish PPS sampling method was used to select counties ordering counties within states by number of stations. A constant number of stations per county was assumed and the proper number of stations were randomly selected for the outlet frame file within each selected county. Once this 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.

To estimate average prices, sample weights were constructed based on the sampled outlet's number of pumps as 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 across grades, formulations and geographic areas were derived using volume data from the EIA-782C "Monthly Report of Prime Supplier Sales of Petroleum Products Sold for Local Consumption", and demographic data from the Bureau of the Census and Department of Transportation on population, number of gasoline stations and number of vehicles.

The target coefficient of variation was set for .4 for the United States, .55 for PADDs and U.S. formulations, .70 for sub-PADDs and the PADD formulations, .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 and data are released on EIA's website:

[http://www.eia.doe.gov/oil\\_gas/petroleum/data\\_publications/wrgp/mogas\\_home\\_page.html](http://www.eia.doe.gov/oil_gas/petroleum/data_publications/wrgp/mogas_home_page.html)

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 publication *Weekly Petroleum Status Report*.

### 3. Maximizing Response Rates

All companies in the sample are contacted through the computer assisted telephone interview system. Companies who prefer to report through the headquarters on behalf of their selected outlets are allowed to do so. Companies preferring to report by fax or email are also permitted to report by that method once they have made arrangements with EIA. Nonrespondent firms are telephoned up to three times. The average response rate for June 2008 to May 2009 was approximately 99 percent.

### 4. Tests of Procedures

This survey began on August 15, 1990. Company cooperation has been outstanding, and no significant problems have been encountered. The sample design and procedures used to conduct the survey are constantly reviewed for improvement, and have been updated a number of times to incorporate those improvements.

### 5. Statistical Consultations

Ms. Paula Mason of the Petroleum Division, Office of Oil and Gas, (202) 586-1262, is responsible for the statistical design of this survey and is the project manager. The contractor responsible for collection and processing of the survey data is:

ICF MACRO  
126 College Street, Suite 2A  
Burlington, VT 05401

ICF MACRO  
11785 Beltsville Drive  
Calverton, MD 20705

## J. "On-Highway Diesel Fuel Price Survey": EIA-888

### 1. Description of the Survey Plan

Approximately 350 outlets were selected for the EIA-888 survey. The EIA-888 survey collects the pump price of diesel fuel sold through gasoline stations and truck stops. With this clearance, prices will be collected for two types of diesel, ultra low sulfur and low sulfur in keeping with the industry's implementation of new EPA requirements during the period in which two types are

sold. However, in accordance with EPA regulations low sulfur diesel will be phased out by the end 2010. Companies selected for the EIA-888 survey were sent an initial letter explaining the survey. Replacement respondents are either faxed or mailed the letter at the time of initiation and informed of the confidentiality and protection afforded their data. Each week the individual outlets are called and asked to report the pump price, by type, of on-highway use diesel fuel. The collection takes place using a computer assisted telephone interview (CATI) with built in editing. Companies who prefer to report through the headquarters on behalf of their selected outlets are allowed to do so. Companies preferring to report by fax or email are also permitted to report by that method. Data obtained through non-phone methods are entered into the CATI system and treated the same as phone collected prices. At this time, an outlet level sampling frame has been constructed based on a match/merge of diesel outlet lists. Sample designs, selection and sample replacement though has been delayed due to a shift of priorities and resources between surveys. The design and selection is now targeted for the first half of 2010.

## 2. Sampling Methodology and Estimation Procedures

The sample for the survey was designed to yield price estimates at the Petroleum Administration Defense District (PADD), sub-PADD, national level, and for the State of California. A standard error of one cent was targeted for PADDs 1, 2 and 3, and one and a half cents of PADDs 4, 5, sub-PADDs 1A, 1B, 1C, and the State of California.

To determine the sample allocations across regions, average standard errors across reporting periods for the previous year of weekly diesel fuel survey prices were calculated for each of the cells. An average sample size was first determined using these standard errors. In addition, a second allocation based on proportional representation within the next larger cell (i.e., more aggregated level cell that the original cell would contribute to) was also obtained. The maximum of these two allocations for each sampling cell was then designated as the sampling cell allocation.

The sample design used a two-phase two stage design. The first phase used the EIA-782 B sample from two rotations, samples 10 and 11. The second sampling phase had two stages, selecting first the company and then the actual outlet. The first stage of the second phase of the sample design used annual state sales volumes for two sample cycles from the EIA-782A and EIA-782B surveys divided by the unit's probability of selection in the monthly survey as a measure of size for Probability Proportionate to Size sampling. These size measures from the two cycles were normalized by assigning 1/2 of the allocation necessary to achieve the target errors in the cell to each cycle and multiplying the allocation by the proportion of the total weighted volume in the cell for the 12 month time period of data for the company state unit (CSU). This allocation procedure yielded a targeted second phase size of 350 outlets for the diesel fuel survey.

Units were selected for the second phase of the sample using probability proportional to size. The frame CSUs were sorted by state and randomly ordered within each state. The normalized size measures were then used to define sampling intervals of 1.0. Using the random order, cumulative size measures were determined where a CSU's cumulative size was the sum of the sizes of all CSUs preceding it and including it. A random number

between 0 and 1 was chosen as a seed, and assigned to the first CSU in PADD 1A. The first CSU whose cumulative size exceeded the seed was sampled and 1.0 was added to the seed. If the CSU's cumulative size measure still exceeded the seed plus 1, the CSU was sampled again and 1 was again added. The sampling continued in this manner selecting the next CSU whose size measure exceeded the count plus seed, until the desired outlet sample size was obtained. The second stage of the second phase took place with the initiation of the sampled companies who were contacted and asked to provide outlet telephone numbers and addresses for the number of outlets in each state that the company sampled. If the CSU was sampled more times than the company had outlets in that state, an outlet was counted more than once.

The design resulted in 2207 company state units in the first phase from the two EIA-782 previous rotations, 282 company state units in the second phase first stage (sub-sampling from the two rotations), and 350 outlets in the second stage of sampling from the company state units the specific outlets.

Since allocations were derived at the sampling cell level, sampling cell ultra low sulfur and low sulfur averages were just simple averages of the CSU prices (the weights from the first and second phases cancel). The average ultra low sulfur and low sulfur prices for regions that constitute a combination of sampling cells were weighted averages of the sampling cell averages where the weights were derived by taking the inverse of the probability proportional to the PADD weighted volumes. The average price of all types of diesel is calculated by weighting the average prices of ultra low sulfur and low sulfur according to the number of outlets in each area selling each of the products as reported for that week in the survey.

The survey is conducted every Monday (Tuesday on Federal holidays), and data are released on EIA's website: <http://tonto.eia.doe.gov/oog/info/wohdp/diesel.asp> 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 publication *Weekly Petroleum Status Report*.

The EIA-888 sample is scheduled for redesign and selection late 2009/early 2010. Similar to the EIA-878, an outlet level frame is being constructed which will serve as the sampling frame. Research is currently being conducted to determine the specific outlet level sample design. Stratification and geographic control variables are being examined in addition to the outlet characteristic of truck stop or service station to evaluate price determining factors. The new sample will be smoothed in by methods such as composite estimation with the overlapping old sample or implementation of the new sample over a series of weeks replacing old respondents with paired newly selected respondents.

3. Maximizing Response Rates

All companies in the sample are contacted through the computer assisted telephone interview system. Companies who prefer to report through the headquarters on behalf of their selected outlets are allowed to do so. Companies preferring to report by fax or email are also permitted to report by that method once they have made arrangements with EIA. Nonrespondent firms are telephoned up to three times. The average response rate for June 2008 to May 2009 was 99 percent.

4. Tests of Procedures

This survey began on February 14, 1994. Company cooperation has been outstanding, and no significant problems have been encountered. The procedures used to conduct the survey are constantly reviewed for improvement.

5. Statistical Consultations

Ms. Paula Mason of the Petroleum Division, Office of Oil and Gas, (202) 586-1262, is responsible for the statistical design of this survey and is project manager. The contractor responsible for collection and processing of the survey data is:

ICF MACRO  
126 College Street, Suite 2A  
Burlington, VT 05401

ICF MACRO  
11785 Beltsville Drive  
Calverton, MD 20705