Accuracy of Petroleum Supply Data

by Tammy G. Heppner and Carol L. French

Overview

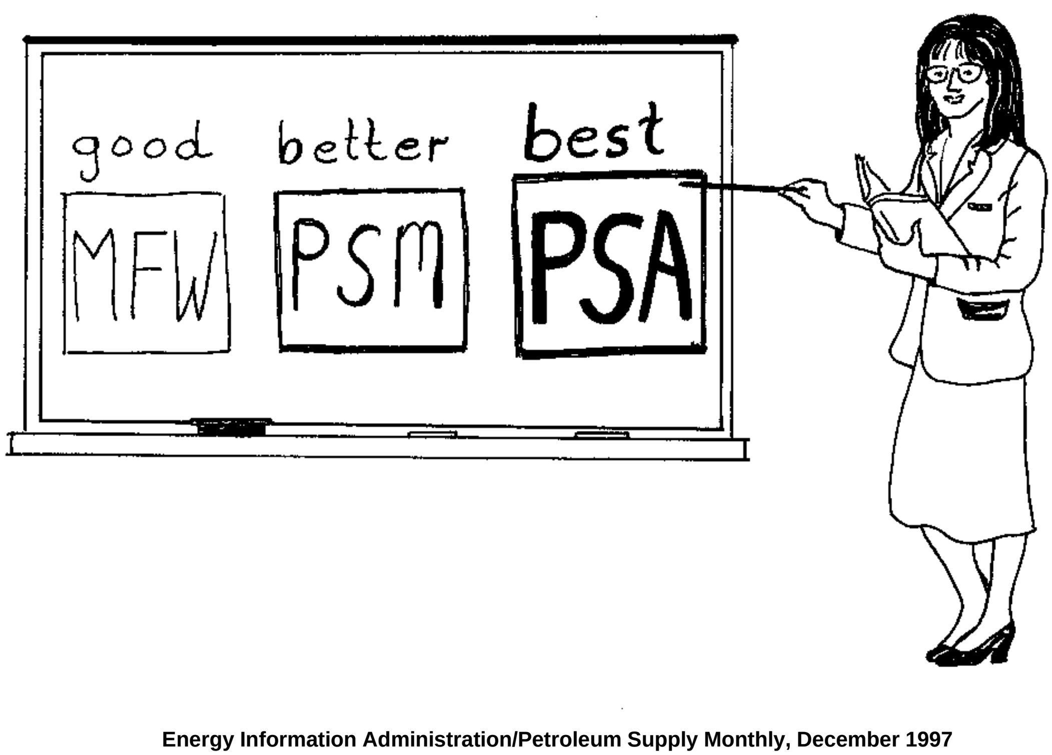
Data collected by the Petroleum Supply Division (PSD) of the Energy Information Administration (EIA) provided an accurate picture of petroleum supply in 1996. These data were presented in a series of PSD publications: the *Weekly Petroleum Status Report* (*WPSR*), the *Winter Fuels Report* (*WFR*), the *Petroleum Supply Monthly* (*PSM*), and the *Petroleum Supply Annual* (*PSA*). Weekly estimates in the *WPSR* and *WFR* were the first values available.

Figure FE1 illustrates the improvement in accuracy from the weekly estimates to the interim monthly values to the final petroleum supply values. The monthly-from-weekly (MFW) data are the least accurate but "good." The *PSM* data are more accurate or "better" and the *PSA* data are the most accurate or "best." For 1996, 66 petroleum supply data series were analyzed to determine how close the *PSM* values were to the final *PSA* values. For these series, 47 out of the 66 were within 1 percent of the *PSA* values in terms of mean absolute percent error. Sixty-one petroleum supply data series were analyzed to see how close the MFW estimates were to the final *PSA* values. For these 61 series, 26 were within 2 percent of the *PSA* values in terms of mean absolute percent error and, of those, 11 were within 1 percent.

Two major factors that contribute to the *PSM* values being more accurate than the MFW estimates are: (1) the greater length of time between the close of the reference period and the publication date of the *PSM*; and, (2) some MFW values are estimates whereas many *PSM* respondents extract their actual data from automated accounting systems. The greater length of time allows more in-depth review of the data by the respondents and EIA. Within 2 months of the close of a reference month, interim values are published in the *PSM*. The weekly data are more quickly available. The *WPSR* is available electronically 5 days after and in hardcopy 7 days after the close of the reference week (excluding holiday weeks). Propane data are available electronically and in the *WPSR*. Additionally, they were published until February 1996 in hardcopy in the *WFR*. About 5 months after the end of the reference year, final monthly values, reflecting any resubmissions, are published in the *PSA*.

Historically, the weekly publications (the *WPSR* and *WFR*) and the monthly (*PSM)* provided volumes of crude oil and

Figure FE1. Grading the Accuracy of 1996 PSD Data



petroleum products data at relatively increasing levels of accuracy. This article provides petroleum analysts with a measure of the degree to which, on average, estimates and interim values vary from their final values.

The Petroleum Supply Reporting System

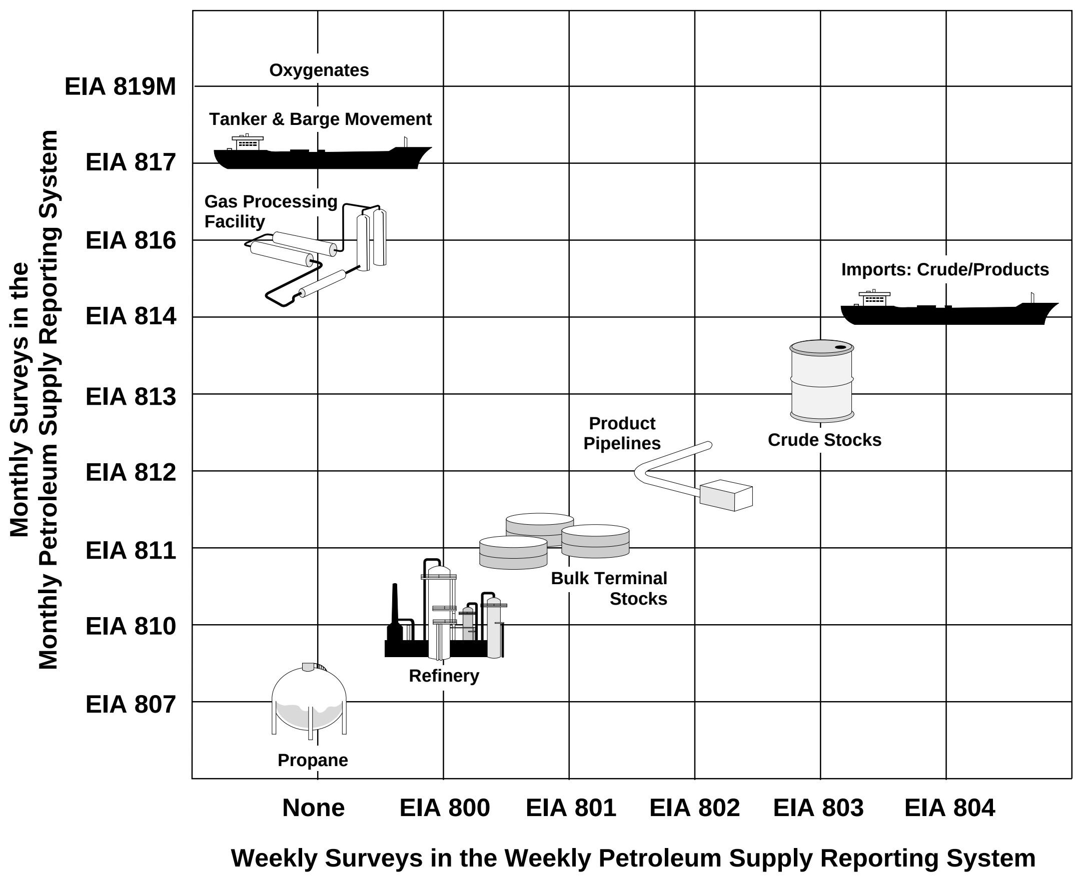
The 15 surveys in the Petroleum Supply Reporting System (PSRS) track the supply and disposition of crude oil, petroleum products, and natural gas liquids in the United States. To maintain a database with historically accurate observations and current estimates from the petroleum industry, EIA administers three survey series: weekly, monthly, and biennial (every other year).

The PSRS is organized into two data collection subsystems, the Weekly Petroleum Supply Reporting System (WPSRS) and the Monthly Petroleum Supply Reporting System (MPSRS). The WPSRS processes data from the five weekly surveys. The MPSRS includes eight monthly surveys, one biennial survey, and the Form EIA-807, which collects propane data monthly from April through September and weekly from October through March.

Figure FE2 displays the petroleum supply and distribution system and indicates the points at which petroleum supply data are collected. Both weekly and monthly surveys are administered at five key points along the petroleum production and supply path: (1) refineries, (2) bulk terminals, (3) product pipelines, (4) crude oil stock holders, and (5) importers of crude oil and products.

Due to 1996 budget reductions, EIA has eliminated and/or changed the collection and publication of two data series. The two data series affected by this decision are U.S. refinery capacity, collected on the Form EIA-820, "Biennial Refinery Report," and U.S. oxygenate production capacity, collected on the Form EIA-819A, "Annual Oxygenate Capacity Report." Annual U.S. refinery capacity data collection and publication normally presented each year in Volume 1 of the *PSA* have been moved to a biennial schedule. These data were collected for 1996. The annual U.S. oxygenate production capacity data normally presented each year in Volume 1 of the *PSA* have been eliminated. Also, since February 1996, publication of the *Winter Fuels Report* has been eliminated, but the data are still available electronically.

Figure FE2. Petroleum Supply Reporting System: Surveys and Subsystems



Source: Energy Information Administration, Petroleum Supply Reporting System.

The Weekly Petroleum Supply Reporting System

The WPSRS contains the data collected from the five weekly surveys. Each weekly survey is distributed to a sample of the corresponding monthly survey's universe. In Figure FE2, the icons represent the target population of the monthly and weekly surveys of the PSRS. For example, the target population for the survey Forms EIA-801 and EIA-811 are bulk terminal stocks. Thus, the respondents to the Form EIA-801 are a sample of the respondents who report to Form EIA-811. For the weekly surveys, EIA aims for a minimum 90-percent multi-attribute-cutoff sample from the respondents to the corresponding monthly survey. In choosing the sample for each product, companies are ranked in descending order by volume. Respondents are chosen in order, down the list until the sample includes those companies contributing at least 90 percent of a variable's total volume. For example, for distillate fuel oil stocks, the weekly sample includes those respondents whose combined volumes of stocks for distillate fuel oil from refineries, bulk terminals, and pipelines constitute at least 90 percent of the total volume of distillate fuel oil stocks as reported in the corresponding monthly surveys.

With these weekly surveys, EIA can provide timely, relatively accurate snapshots of the U.S. petroleum industry every week. The weekly surveys collect information on the supply and disposition of selected petroleum products and crude oil. The reference period for each weekly survey begins at 7:01 a.m. each Friday and ends at 7:00 a.m. the following Friday. Respondents report their data via telephone, facsimile, or EIA's electronic data collection software package, the Personal Computer Electronic Data Reporting Option (PEDRO). All respondents must submit their data by 5:00 p.m. on the Monday following the end of the reference period. During 2 working days, quality control procedures are executed. Cell values determined to be unusual or inconsistent with other cell values are flagged. The validity of the value of each flagged cell is investigated. Some flagged values are verified by the respondent to be correct; other flagged cells are corrected; and the remaining flagged values are referred to as unresolved. Nonrespondent and unresolved flagged data are imputed using an exponentially smoothed mean of the respondents' historical data.

Within 7 days of the close of the reference week, data are made available to the public in three forms: through the EIA electronic publishing system (EPUB), hardcopy (through the *WPSR*), and EIA's internet web site. Except when holidays delay data processing schedules, values for the weekly variables are available via EPUB at 9:00 a.m. on the Wednesday following the close of the reference week. The hardcopy *WPSR* is distributed on the Friday morning following the close of the reference week. Beginning in September 1995, the weekly data were made available on the internet (http://www.eia.doe.gov) and are on the same schedule as EPUB.

The Monthly Petroleum Supply Reporting System

The reference period for the monthly surveys starts on the first day of the month at 12:01 a.m. and ends on the last day of the month at midnight. Except for the Form EIA-819M, the deadline for filing monthly surveys is the 20th calendar day following the end of the report month. Data collection for the Form EIA-819M begins on the seventh working day of the month. Form EIA-819M data are solicited by telephone or received by facsimile. Data for the other monthly surveys are reported via telephone, facsimile, or PEDRO.

During the period of data editing, either the respondent or EIA staff may identify an error. If the respondent discovers an error, the EIA representative for a particular survey is notified and the value is corrected. If EIA's edits diagnose an unusual value, an EIA representative will determine if the value is correct or incorrect by calling the company and/or reviewing historical data.

Within 60 days of the close of the reference month, all of the interim monthly data are published in the *PSM*. However, customer satisfaction surveys conducted by EIA during 1995 showed a need for faster release of available monthly data. In response to this need, beginning in November 1995, EIA implemented a plan for early release of monthly petroleum statistics approximately 45 days after the end of the report period. The preliminary data are presented in four tables: "U.S. Daily Average Supply and Disposition of Crude Oil and Petroleum Products," "Imports of Crude Oil and Petroleum Products into the United States by Country of Origin," "Stocks of Crude Oil and Petroleum Products by Petroleum Administration for Defense (PAD) District," and "Refinery, Bulk Terminal, and Natural Gas Plant Stocks of Selected Petroleum Products by PAD District and State". These preliminary tables are available on the internet and EPUB approximately on the 13th of each month. After incorporation of petroleum exports and crude oil production, these tables are replaced with final tables between the 20th and the 23rd of each month. In addition to the internet, beginning in March 1996, monthly data became available on EIA's CD-ROM called the Energy InfoDisc, which is released quarterly.

Throughout the year, EIA accepts data revisions of monthly data. If a revision is made after the *PSM* has been published, it is referred to as a resubmission. Resubmissions for earlier months are published in Appendix C of the *PSM* and are reflected in the *PSA*. Beginning with the February 1994 *PSM*, a new table (Table H1, Petroleum Supply Summary) was included to show early estimates of monthly data. The current-month values in Table H1 are preliminary estimates based on weekly submissions. These monthly-from-weekly estimates become available in the *WPSR* and on EPUB on the Wednesday following the first Friday of each month.

Table FE1. Average Coverage for Weekly Surveys, 1996 and 1995 (Percent of Final Monthly Volumes Included in Monthly-from-Weekly Sample)

Product	Stocks						Production		Imports	
	Refinery		Bulk Terminal		Pipeline					
	1996	1995	1996	1995	1996	1995	1996	1995	1996	1995
Total Motor Gasoline	98	97	93	92	97	97	99	98	96	98
Jet Fuel	98	98	94	94	99	99	99	98	98	80
Distillate Fuel Oil	97	96	88	87	98	98	97	97	89	87
Residual Fuel Oil	94	94	91	90	--	--	94	94	94	96
Crude Oil	96	95	--	--	--	--	--	--	95	96

-- = Not Applicable.
Source: Energy Information Administration, Petroleum Supply Reporting System.

Within 5 months of the end of the calendar year, the final monthly values for the previous year are published in the *PSA*. These values reflect all *PSM* resubmissions and other data corrections. The values contained in the *PSA* are EIA's most accurate measures of petroleum supply industry activity.

Factors Affecting Data Accuracy

Maintaining an accurate database is a major goal of EIA. The quality of the data drives the quality of all qualitative and quantitative analyses conducted using these data. Accuracy and timeliness are primary attributes of high quality data. Accuracy of survey data is measured as the closeness of the published values to the true values (i.e., those values that would be obtained if the target population had been correctly surveyed and all the data had been precisely recorded).

Respondents to the monthly surveys have more time to file than the weekly respondents, enabling them to collect, review, and revise their monthly data more carefully than they do their weekly data. Additionally, EIA has more time to edit the monthly data. Also, some weekly respondents report estimates while many monthly respondents extract actual data from accounting systems. Thus, the monthly data are more accurate.

Some mechanisms introducing error, such as nonresponse, are not totally preventable. Other sources for errors, such as sampling errors, are unique to a particular type of survey. One situation where sampling error occurs is if the group of sampled respondents is dissimilar to the full population. Within the PSRS, only weekly surveys, the monthly oxygenate survey, Form EIA-819M, and the propane survey, Form EIA-807, are at risk of having sampling errors. However, all surveys in the PSRS are at risk for nonsampling errors, such as: (1) insufficient coverage of respondents (the survey frame does not include all members of the target population); (2) nonresponse; (3) response error; and (4) internal processing errors such as incorrect data entry. A detailed discussion of factors influencing data accuracy and how they are minimized in the PSRS follows.

Samples and Sampling Error

A sample is a subsection of a universe identifying members of a target population. The weekly surveys are administered to samples of the monthly populations to reduce respondent burden and to expedite the turnaround of data from survey respondents to the public. As with any sample, the values obtained are different from those obtained if the full universe had been surveyed. Sampling error is the difference between a sample estimate and a population value.

There are five samples, one for each weekly petroleum supply survey, in the WPSRS. For these surveys, the sampling error is minimized by using a minimum 90-percent multi-attribute-cutoff sample from the corresponding monthly survey's frame. At the end of each month, updates are made to the samples and survey frames if a 90-percent coverage was not obtained.

For the weekly surveys, better coverage will most likely reduce sampling error. As shown in Table FE1, 1996 coverage was comparable to 1995. Nineteen of the 21 product and supply type combinations had coverage above 90 percent in 1996. For 14 of the 21 combinations, 1996 coverage increased from 1995. Tabulations were done before rounding of the coverage values. The average coverage of jet fuel refinery imports increased by 18 percentage points, from 80 to 98 percent.

Nonsampling Error

Unlike sampling errors, all survey data, even those from a census survey, are at risk of incurring nonsampling errors. There are two categories of nonsampling errors, random and systematic. With random error, on average, and over time, values will be overestimated by the same amount they are underestimated. Therefore, over time, random errors do not bias the data, but they will give an inaccurate portrayal at any point in time. On the other hand, systematic error is a source of bias in the data, since these patterns of errors are made repeatedly. The following is a discussion of how the

Table FE2. Average Response Rates for Monthly and Weekly Surveys, 1996

Survey Site	Respondents to Monthly Surveys			Respondents to Weekly Surveys		
	Average Universe Size	Average Number of Respondents	Percent [1]	Average Weekly Sample Size	Average Number of Respondents	Percent [2]
Refinery	264	257	97.4	188	184	97.9
Bulk Terminal	318	313	98.4	75	72	96.2
Pipeline	82	82	99.8	43	42	97.6
Crude Oil Stocks	175	174	99.5	85	83	98.6

[1] The average response rates for monthly surveys are calculated by summing the individual monthly response rates and dividing by 12.
[2] The average response rates for weekly surveys are calculated by summing the individual weekly response rates and dividing by 52.
Note: Percents are calculated before rounding.
Source: Energy Information Administration, Petroleum Supply Reporting System.

four most frequently occurring types of nonsampling error are minimized within the PSRS.

Frame Updates

The list of all companies identified as members of the target population is called a frame. If members of the target population are not included in the frame, there is an undercount of the aggregate data. To diminish the chance of undercounting, the PSRS frames are continually updated. New companies are identified through continual review of petroleum industry periodicals, newspaper articles, and correspondence from respondents. During the frames update, each frame is scrutinized to assure completeness.

Maintaining a Low Nonresponse

Survey respondents are required by law to report to EIA (see Explanatory Note 6 of the *PSM* for a description of action for chronic nonresponse). The 1996 response rates for the weekly surveys and their complementary monthly surveys are enumerated in Table FE2. The 1996 average response rate for each of the EIA weekly surveys was over 96 percent (same as 1995). The corresponding monthly surveys had slightly better response rates of over 97 percent (an increase of 1 percent from 1995).

To mitigate the effect of nonresponse, imputed values are calculated for all unreported values except monthly imports. Weekly imputed values are the exponentially smoothed mean of that respondent's historical values for that variable. Monthly imputed values are the previous month's value for the particular respondent and variable. For imports, however, there is a great deal of fluctuation from one reference period to another, with respondents frequently having no imports of a particular product. As a result, zero is the value imputed for nonreported cells on the monthly survey. In addition, the monthly imports are collected at a much greater level of detail than the weekly imports, which makes imputation difficult.

Reducing Response Error

Over the past 5 years, many structural and procedural improvements to the PSRS system have been made in order to reduce the problem of nonsampling errors. One such improvement has been the PEDRO system, which permits all weekly and monthly survey data except the Form EIA-819M and Form EIA-807 to be submitted to EIA electronically. A respondent entering values via PEDRO may execute edit routines prior to transmission of the survey responses. These routines include consistency and outlier (extreme value) checks of the data. Unusual or nonreported cells are flagged and, prior to transmission of the data, a representative of the company is able to review and verify or correct data in the flagged cells.

Even with sophisticated edit checks, response error (the difference between the reported value and the actual value) remains the most likely cause of data inaccuracy. The weekly surveys are more susceptible to response error since some of their values are estimates. Many monthly respondents abstract their actual data from accounting systems and thus are generally more accurate.

Maintaining accurate accounting records, however, does not ensure against response error. For example, numbers can be transposed within the correct cell; an otherwise correct value may be entered in the wrong cell; a respondent may misinterpret the intent of a question; or the wrong units may be used.

Survey Clarity

The terms, layout, and definitions on all survey forms are periodically reviewed for completeness, clarity, and consistency across surveys. At regular intervals, survey intent, as well as what data are collected, are subject to industry and government review. To the extent possible, industry changes in terminology and practice are incorporated into the PSRS on an ongoing basis.

Data Assessment

Each of the variables included in these analyses is of current and historical interest. Of the 66 variables for which both *PSM* and *PSA* values were published, only 61 of them were published weekly throughout 1996. For each variable, six measures of accuracy were calculated to compare the differences between the MFW and *PSM* values relative to the *PSA* values.

- **Error** is the difference between the estimate or interim value and the final value for a given month. For inputs, production, stock change, imports, exports, and product supplied, values are expressed in units of thousands of barrels per day. For stocks, values are expressed in units of thousands of barrels.

  MFW Error = MFW Volume - *PSA* Volume

  PSM Error = *PSM* Volume - *PSA* Volume

- **Percent Error** is the error for a given month divided by the final value for a given month, and multiplied by 100.

$$\text{MFW Percent Error} = \frac{\text{MFW Error}}{PSA\text{ Volume}} \times 100$$

$$PSM\text{ Percent Error} = \frac{PSM\text{ Error}}{PSA\text{ Volume}} \times 100$$

- **Mean absolute error** is the weighted average over the 12 months of the year of the absolute values of the errors for each month. The mean absolute error measures the average magnitude of the revisions that took place over a year. Outliers increase the mean absolute error. The number of days in the month is used for weighting all product categories except stocks. Stocks are weighted equally for each of the 12 months.

- **Mean absolute percent error** is the weighted average over the 12 months of the year of the absolute values of the percent errors. It provides a measure of the average magnitude of the revisions relative to final values. The mean absolute percent error has an inverse relationship with data accuracy; i.e., the smaller the mean absolute error, the closer the interim data are to the final data; conversely, the larger the mean absolute percent error, the greater the difference in the interim value and the final value. Outliers inflate the mean absolute percent error.

- **Range** is the difference between the smallest and largest percent errors. The range shows the dispersion of the percent differences between interim and final values.

- **Median** of the percent errors is the point at which half the values are higher and half are lower. Unlike the mean, the median is not affected by an outlier. In these analyses, each distribution has 12 observations. The median is the average of the sixth and seventh ordered observation.

The average final absolute volumes and the mean absolute percent error for MFW estimates and *PSM* interim values for 1996 and 1995 are presented in Table FE3. The average final absolute volumes are presented to give the reader an idea of the magnitude of these volumes. Variables with very small volumes are prone to larger percent changes because a modest volume change is being compared to a small final volume. The mean absolute error and the size of the volumes involved must both be included in the interpretation of data accuracy.

The 1996 MFW mean absolute percent errors which were within 2 percent of their respective *PSA* values (26 of the 61 MFW series), and the 1996 *PSM* mean absolute percent errors which were within 1 percent of their *PSA* values (47 of the 66 *PSM* series), are distinguished by a single asterisk. Mean absolute percent errors that were greater than 10 percent are marked by a double asterisk. There were 15 such MFW series and 1 *PSM* series.

For 1996, 12 of the 14 production series have a single asterisk in the *PSM* column, indicating a mean absolute percent error of less than 1 percent from the *PSA*. For distillate fuel oil, two relatively new subcategories are low sulfur (0.05 percent sulfur and under) and high sulfur (over 0.05 percent sulfur). In prior years, respondents had trouble classifying distillate fuel oil into these subcategories. This has been resolved as the table shows the decrease in mean absolute percent error from 1995 to 1996 for each of the products. The increase in the MFW mean absolute percent error for oxygenated motor gasoline production is partially a result of a decrease in the average absolute volume. Another factor is that weekly fuel ethanol supply and disposition data are not available; therefore, the weekly oxygenated motor gasoline field production is based on the latest available monthly value.

The single asterisks in Table FE3 by the stock series show that, as in prior years, the stock values for both MFW estimates and *PSM* interim values are very close to the final *PSA* values. A major exception is the double asterisk shown by the MFW percent error for oxygenated motor gasoline stocks. Similar to the percent error for production of this product, the increase is related to the average absolute volume. Reformulated gasoline was first reported on the weekly system for the week ending September 23, 1994; therefore, now there are two full years of data to compare. Fuel ethanol and methyl tertiary butyl ether stocks are not collected weekly, but are collected on the Form EIA-819M, "Monthly Oxygenate Telephone Report." The survey provides production data and preliminary stock data from a sample of respondents reporting on the monthly surveys and from the universe of oxygenate producers. These data are displayed in Appendix D of the *PSM*. Interim data are collected later on the monthly surveys and published in the *PSM*.

Stock change is the difference between stocks at the beginning of the month and stocks at the end of the month. Since the monthly change in stock levels is small compared to the stock levels themselves, a large percent error in stock change can occur when the percent errors in stock levels are small.

Crude oil stock change is one of the components in the calculation of unaccounted for crude oil (calculated disposition minus calculated supply of crude oil). For both the MFW and the *PSM* numbers, the volume of the unaccounted for crude oil may be increased by a combination of factors including an

Table FE3. Summary Statistics for Differences Between Interim and Final Data, 1996 and 1995

Variable	*PSA* Average Absolute Volumes		Monthly-from-Weekly Mean Absolute Percent Error		*PSM* Mean Absolute Percent Error	
	1996	1995	1996	1995	1996	1995
Crude Oil Production (thousand barrels/day)	6,465	6,560	* 0.79	0.83	* 0.50	0.52
Refinery Operations						
Refinery Crude Oil Inputs (thousand barrels/day)	14,195	13,973	* 0.32	0.45	* 0.19	0.08
Operable Utilization Rate (percent)	94	92	* 1.20	1.95	* 0.70	1.50
Production (thousand barrels/day)						
Total Production	18,467	18,060	--	--	* 0.25	0.26
Refinery Production	16,324	15,994	* 1.44	1.45	* 0.25	0.29
Finished Motor Gasoline	7,647	7,589	* 0.96	0.89	* 0.70	0.21
Reformulated Motor Gasoline	2,221	1,884	2.97	2.44	1.83	1.39
Oxygenated Motor Gasoline	454	876	** 56.41	22.39	3.09	3.19
Other Motor Gasoline	4,972	4,828	4.00	2.97	* 0.60	0.48
Jet Fuel	1,515	1,416	* 1.27	0.91	* 0.11	0.18
Distillate Fuel Oil	3,316	3,155	* 0.42	1.26	* 0.30	0.14
Low Sulfur Distillate Fuel Oil	2,084	1,944	* 1.72	2.63	* 0.25	1.92
High Sulfur Distillate Fuel Oil	1,232	1,211	2.52	2.85	* 0.77	2.89
Residual Fuel Oil	726	788	4.17	2.69	* 0.80	0.40
Other Products	5,264	5,112	--	--	* 0.59	0.77
Propane	1,044	1,022	--	--	* 0.18	0.32
Other Products Refinery Production	3,204	3,176	7.18	5.86	* 0.58	0.48
Stocks (thousand barrels)						
Total Stocks	1,525,640	1,608,840	* 0.57	0.32	* 0.07	0.11
Total Stocks, excl. SPR	944,599	1,017,175	* 0.95	0.50	* 0.12	0.17
Total Crude Stocks	884,400	913,192	* 0.27	0.47	* 0.11	0.17
Crude Oil Stocks, excl. SPR	303,359	321,526	* 0.76	1.34	* 0.31	0.49
SPR Stocks	581,041	591,666	* 0.04	0.00	* 0.00	0.00
Refined Products Stocks	641,240	695,648	* 1.43	0.90	* 0.10	0.07
Total Motor Gasoline Stocks	200,699	206,435	* 1.03	0.96	* 0.22	0.15
Reformulated Motor Gasoline Stocks	39,413	37,658	2.18	2.71	1.24	0.42
Oxygenated Motor Gasoline Stocks	1,522	3,326	** 22.23	13.70	8.92	44.06
Other Motor Gasoline Stocks	118,855	124,030	* 1.56	0.88	* 0.20	0.42
Jet Fuel Stocks	38,246	40,600	* 1.68	1.68	* 0.20	0.31
Distillate Fuel Oil Stocks	106,959	125,824	* 1.22	1.57	* 0.19	0.13
Low Sulfur Distillate Fuel Oil Stocks	59,291	63,007	* 1.67	1.62	* 0.28	0.61
High Sulfur Distillate Fuel Oil Stocks	47,668	62,817	2.48	1.98	* 0.33	0.63
Residual Fuel Oil Stocks	36,367	38,041	2.69	1.80	* 0.27	0.41
Other Products Stocks	258,969	284,748	3.30	1.76	* 0.18	0.17
Propane Stocks	38,386	42,670	4.89	1.69	* 0.48	0.20
Fuel Ethanol Stocks	1,425	3,282	5.62	2.92	1.39	0.78
Methyl Tertiary Butyl Ether Stocks	9,718	10,344	6.26	6.43	* 0.62	0.74
Stock Change (thousand barrels/day)						
Total Stock Change	563	422	** 84.30	259.63	7.56	16.10
Total Crude Stock Change	211	220	**1,258.53	86.77	** 220.89	9.30
Refined Products Stock Change	535	456	** 71.92	114.79	8.08	11.44
Imports (thousand barrels/day)						
Total Imports	9,478	8,835	2.68	1.87	* 0.84	0.25
Total Crude Imports	7,508	7,230	* 1.75	1.91	* 0.39	0.28
Crude Oil Imports, excl. SPR	7,508	7,230	* 1.75	1.91	* 0.39	0.28
SPR Imports	0	0	* 0.00	0.00	* 0.00	0.00
Refined Products Imports	1,971	1,605	** 10.37	8.41	2.74	0.97
Finished Motor Gasoline Imports	336	265	** 13.68	9.77	4.16	2.91
Reformulated Motor Gasoline Imports	174	117	** 15.10	8.44	6.27	3.38
Oxygenated Motor Gasoline Imports	0	7	* 0.00	38.76	* 0.00	8.49
Other Motor Gasoline Imports	163	142	** 22.45	14.37	3.61	3.84
Jet Fuel Imports	111	106	** 10.75	15.80	2.87	0.83

See footnotes at end of table.

Table FE3. Summary Statistics for Differences Between Interim and Final Data, 1996 and 1995 (Continued)

Variable	*PSA* Average Absolute Volumes		Monthly-from-Weekly Mean Absolute Percent Error		*PSM* Mean Absolute Percent Error	
	1996	1995	1996	1995	1996	1995
Distillate Fuel Oil Imports	230	193	** 11.31	12.91	2.45	1.30
Low Sulfur Distillate Fuel Oil Imports	112	77	9.92	23.83	2.84	0.11
High Sulfur Distillate Fuel Oil Imports	118	116	** 19.54	10.82	2.29	2.14
Residual Fuel Oil Imports	248	187	** 13.03	9.73	* 0.79	0.69
Other Products Imports	1,045	863	** 15.95	12.55	5.63	1.91
Propane Imports	119	102	--	--	2.22	0.37
Exports (thousand barrels/day)						
Total Exports	981	949	6.31	8.19	* 0.00	0.01
Crude Oil Exports	110	95	** 41.76	28.12	* 0.00	0.18
Refined Products Exports	871	855	7.13	8.54	* 0.00	0.03
Total Net Imports (thousand barrels/day)	8,498	7,886	3.03	10.40	* 0.94	0.28
Products Supplied (thousand barrels/day)						
Total Products Supplied	18,309	17,725	* 1.49	1.51	* 0.41	0.23
Finished Motor Gasoline Supplied	7,891	7,789	* 1.43	1.25	* 0.52	0.22
Jet Fuel Supplied	1,578	1,514	3.07	2.30	* 0.43	0.34
Distillate Fuel Oil Supplied	3,365	3,207	* 1.88	4.09	* 0.30	0.41
Residual Fuel Oil Supplied	848	852	6.65	5.07	1.10	0.84
Other Products Supplied	4,627	4,363	3.98	3.81	* 0.73	0.49
Propane Supplied	1,136	1,096	--	--	* 0.47	0.49

-- = Not Applicable.
* = For MFW values, mean absolute percent error less than or equal to 2; for *PSM* values, mean absolute percent error less than or equal to 1.
** = Mean absolute percent error greater than or equal to 10.
SPR = Strategic Petroleum Reserve
Notes: •Error is the difference between Monthly-from-Weekly estimates or interim monthly data published in the *Petroleum Supply Monthly* and the final value as published in the *Petroleum Supply Annual.* Percent error is the error multiplied by 100 and divided by the final published value. Mean absolute error is the weighted average of the absolute errors. Mean absolute percent error is the weighted average of the absolute percent errors. The number of days in the month is used for weighting all product categories except stocks. Stocks are weighted equally for each of the 12 months. •Totals may not equal sum of components due to independent rounding.
Source: Energy Information Administration, Petroleum Supply Reporting System and the U.S. Bureau of the Census.

understatement of imports, an overstatement of exports, an understatement of crude oil production, an understatement of stock withdrawals, and an overstatement of crude oil inputs. The overstatement of crude oil inputs can be caused by injections along crude oil pipelines of natural gas liquids. When refiners receive this mixture, they process it as crude oil. As seen in Table FE3, the production, imports, and refinery inputs of crude oil have a small mean absolute percent error relative to crude oil stock change.

For petroleum products, stock change is a component in the calculation of product supplied (representing the consumption of petroleum products). Unlike the other variables, stock change values can be negative. Stock change thus has an added dimension by which to evaluate accuracy; this is the correctness of the direction of the change. Table FE4 provides a measure of accuracy of the direction of MFW and *PSM* stock change values for 1996 and 1995. The 1996 direction of stock change was similar to the 1995 direction.

Table FE4. Number of Months In Which the Direction of MFW and *PSM* Stock Change Values Differed From *PSA*

	Number of Months	
	1996	1995
Total Stock Change		
MFW and *PSA* Values	4	5
PSM and *PSA* Values	0	0
Total Crude Stock Change		
MFW and *PSA* Values	1	1
PSM and *PSA* Values	2	0
Refined Products Stock Change		
MFW and *PSA* Values	2	2
PSM and *PSA* Values	0	0

Source: Energy Information Administration, Petroleum Supply Reporting System.

For imports, one reason for the large mean absolute percent errors in the MFW values is that shipments do not always arrive during the week in which they were expected. This has a greater impact when the end of the month occurs in the middle of the week. Eight of the 15 MFW import series in Table FE3 showed an increase in mean absolute percent error from 1995 to 1996. For the *PSM*, 13 of the 16 import series increased.

With the exception of refinery receipts in the Virgin Islands, EIA does not collect export data. They are gathered by the U.S. Customs Service on a monthly basis and are compiled by the U.S. Bureau of the Census. They are received by EIA on a monthly basis approximately 7 weeks after the close of the reporting month. The weekly estimates for exports are projections based on past monthly data. Because the export data are highly variable, it is difficult to obtain estimates of comparable quality to domestic estimates.

Product supplied is the calculation of field production, plus refinery production, plus imports, plus unaccounted for crude oil, minus stock change, minus crude oil losses, minus refinery inputs, minus exports. Therefore, the accuracy of product supplied is affected by the individual components.

Box and Whisker Plots

Example 1 in the shaded box titled "Structure of Box and Whisker Plots," is a simplified illustration of the box and whisker plots that follow. The box and whisker plots map the 5-year trends in historical accuracy of weekly estimates and monthly interim values. The details provided by the box and whisker plots include: historical trends, the range of monthly percent errors, direction of the error (i.e., overestimation or underestimation), and the identification of unusual values.

Each box and whisker plot is placed on a graph, where the horizontal axis represents the year and the vertical axis represents the percent error. The center horizontal line for all the box and whisker plots is zero percent error. For each variable studied, a pair of charts, each containing five box and whisker plots (one for each year, from 1992 through 1996), are presented side-by-side; the chart on the left contains the percent errors for the MFW estimates, and the chart on the right contains the percent errors for the *PSM* values. To facilitate the comparison of MFW percent errors and the *PSM* percent errors, the plots have the same scale.

The position of the box along the y-axis denotes whether the MFW or *PSM* values are predominantly overestimates or underestimates of the *PSA* values. For example, if the majority of the MFW values were over estimates, more than half of the box would be above the zero percent error line.

Crude Oil Production and Crude Oil Inputs

Crude oil production data are not collected through any of EIA's surveys. EIA's Dallas Field Office assembles data collected from State agencies responsible for measuring crude oil production. Based on historical trends and data reported on Form EIA-182, "Domestic Crude Oil First Purchase Report," EIA estimates weekly and monthly production. Final estimates based on revised Form EIA-182 data are published in the *PSA*. Figure FE3 presents errors of MFW and *PSM* values relative to *PSA* values for crude oil production and inputs. Over the last 5 years, both MFW and *PSM* crude oil production values have been quite close to the *PSA* values. The 1996 MFW percent error distribution has values ranging from -1.63 to 1.16 percent. Similarly, there is a tight distribution for the percent errors of crude oil *PSM* values, ranging between -0.89 and 0.77 percent. All of the months are within 1 percent of the final *PSA* values. In 1996, it was the first time in the past five years for the MFW and *PSM* percent errors to have a positive median. The small percent errors of both MFW and *PSM* crude oil values demonstrate the consistency and precision of EIA's estimation procedures for weekly and monthly crude oil production.

Refinery crude oil inputs had the smallest range (1.31) of all the other MFW plots analyzed. Most of the MFW values were within 0.50 percent of the final values. Historically, the *PSM* refinery crude oil inputs have been extremely close to their final values. In 1996, all of the *PSM* values were within 0.28 percent of the final values, except one outlier in November (0.51). This was due to resubmissions. The majority of *PSM* values were underestimates.

Product Production

As expected, *PSM* interim values for production of each of the four major petroleum products were superior to their comparable MFW estimates. Figures FE4 and FE5 contain the box and whisker plots for motor gasoline and distillate fuel oil production, and residual fuel oil and jet fuel production, respectively.

In contrast to last year, Figure FE4 illustrates that the 1996 MFW estimates of motor gasoline production were primarily underestimates of the final values. Over the 5-year period, 1996 had the smallest range (2.53) and the largest absolute median (-0.80). With the exception of 1996, the *PSM* interim motor gasoline production values have historically been excellent. The 1996 *PSM* percent error distribution has values ranging from -1.43 to -0.28. This is the largest range over the 5-year period, with July 1996 (-1.43) having the largest *PSM* absolute percent error over the 60-month period. During the annual processing of 1996 data, there were additional submissions that increased the *PSA* volumes of motor gasoline production. Therefore, the MFW and *PSM* underestimated motor gasoline production.

Structure of Box and Whisker Plots

All box and whisker plots discussed in this article are the visual presentation of a variable's distribution of 12 values of percent errors for either MFW or *PSM* values relative to *PSA* values for a given year. In general, box and whisker plots group data, ordered from smallest to largest, into four areas of equal frequency, quartiles, and show the range and dispersion of data within the quartiles. Sometimes the values of quartiles must be interpolated, i.e., if there are two values that meet the criteria of a quartile, then the average of the two must be taken. Presented below is a discussion of components of box and whisker plots and how they apply to the 12-value distribution illustrated in Example 1: -35, -20, -11, -9, 0, 0, 0, 0, 4.5, 5.5, 15, and 20.

- **First Quartile**

Twenty-five percent of the values are equal to or below the first quartile. In Example 1, the first quartile is the average of the third and fourth ordered observations, i.e., (-11+(-9))/2=-10. The first quartile demarcates the lower boundary of the box.

- **Second Quartile**

The second quartile is the median, and it intersects the box. Fifty percent of the observations are equal to or below the median; in our example, the values of these six observations are: 0, 0, -9, -11, -20, and -35. Also, for this example, the median is the average of the sixth and seventh value, 0, i.e., (0+0)/2. The plot provides the value of the median (the second quartile) as well as information on how the median compares in magnitude to the rest of the observations. Outliers distort the magnitude of the mean, whereas a median is not distorted since it is the actual value that falls in the middle of the distribution. Since outliers have occurred in the distributions of values of PSRS variables, a median is preferred to a mean when assessing accuracy.

- **Third Quartile**

Seventy-five percent of the observations (9 in this case) have values equal to or below the third quartile. In Example 1, the third quartile is 5, i.e., (4.5+5.5)/2. The third quartile demarcates the upper boundary of the box.

- **Box**

The box contains half of all the values. In Example 1, as well as in each box found in Figures FE3-FE11, a minimum of six values are contained within the box. The interquartile range is the length of the box, the difference between the first and third quartiles. The interquartile range for Example 1 is 15, i.e., 5-(-10).

- **Whiskers**

Each whisker extends out from the box, one from the first quartile and the other from the third quartile, to the most extreme value that still falls within 1.5 times the interquartile range. In Example 1, a whisker extends from the third quartile, 5, to 20, which is the maximum value and is within 1.5 interquartile ranges of 5 (as it is less than 5+(1.5*15)=27.5). Also in Example 1, the lower whisker extends from the first quartile -10, to -20, which is the lowest value of the distribution within 1.5 interquartile ranges of the first quartile.

- **Fourth Quartile**

The fourth quartile is the maximum value of the distribution. In Example 1, the fourth quartile, 20, also demarcates the upper value of the top whisker as it is within 1.5 interquartile ranges of the third quartile.

- **Outlier**

An outlier, identified as an asterisk, is an observation that is more than 1.5 interquartile ranges greater than the third quartile, or more than 1.5 interquartile ranges less than the first quartile. In Example 1, there is one outlier, -35. It is less than the lower whisker's threshold value, which is -32.5 (-10-(1.5*15)). The importance of the occurrence of an outlier depends on the distribution of the variable. If the interquartile range is very tight and the outlier is in close proximity, then there is little concern about the occurrence of that outlier. (See Figure FE3, MFW vs *PSA* of Crude Oil Production for 1992.)

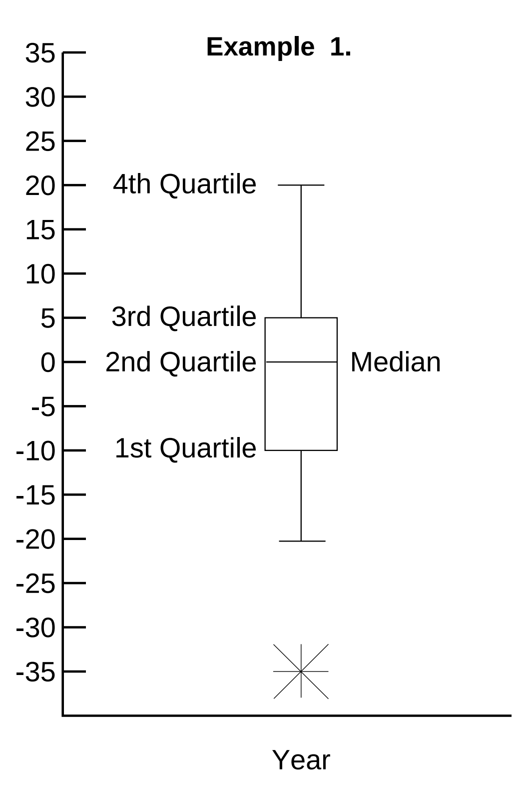
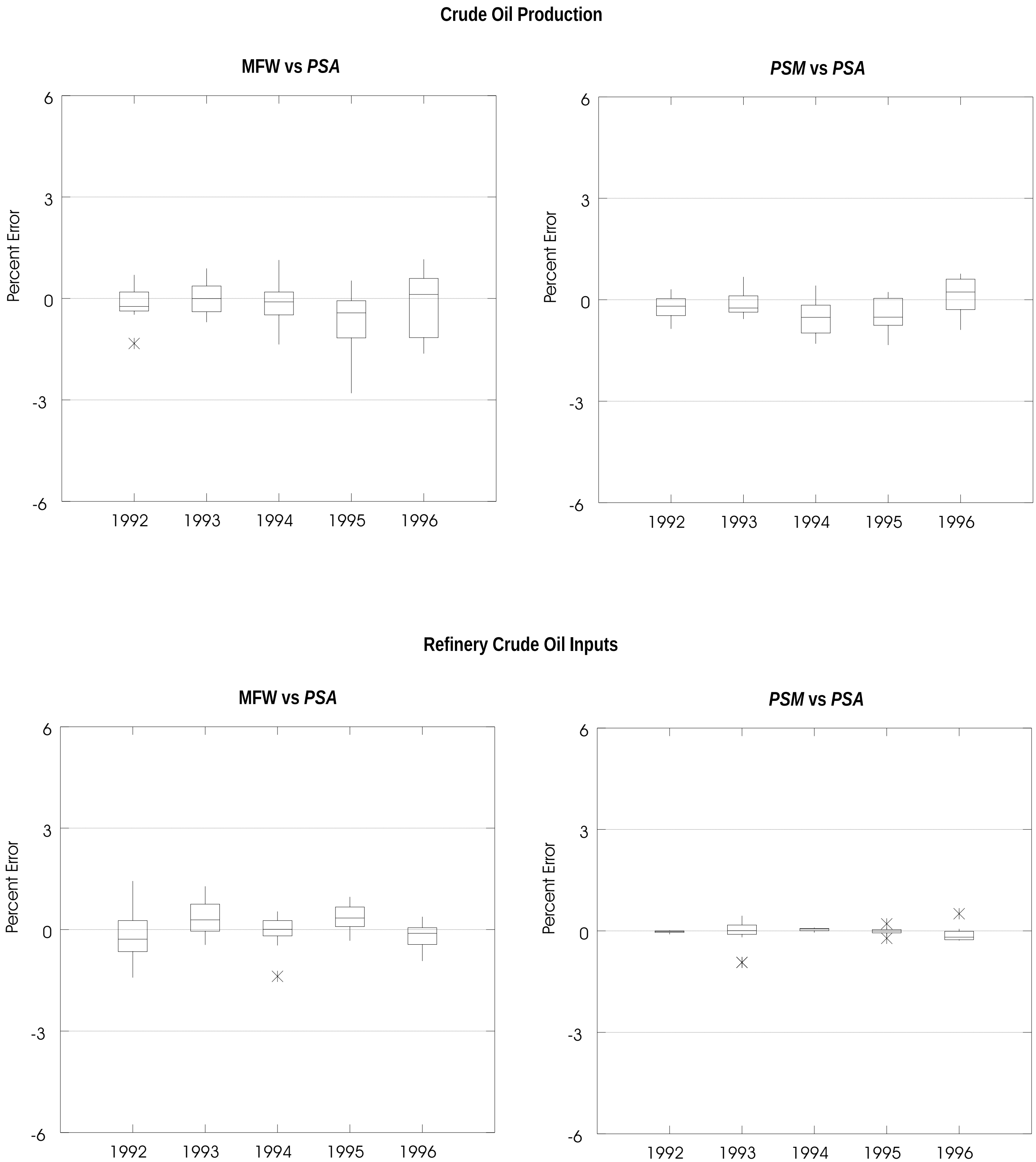
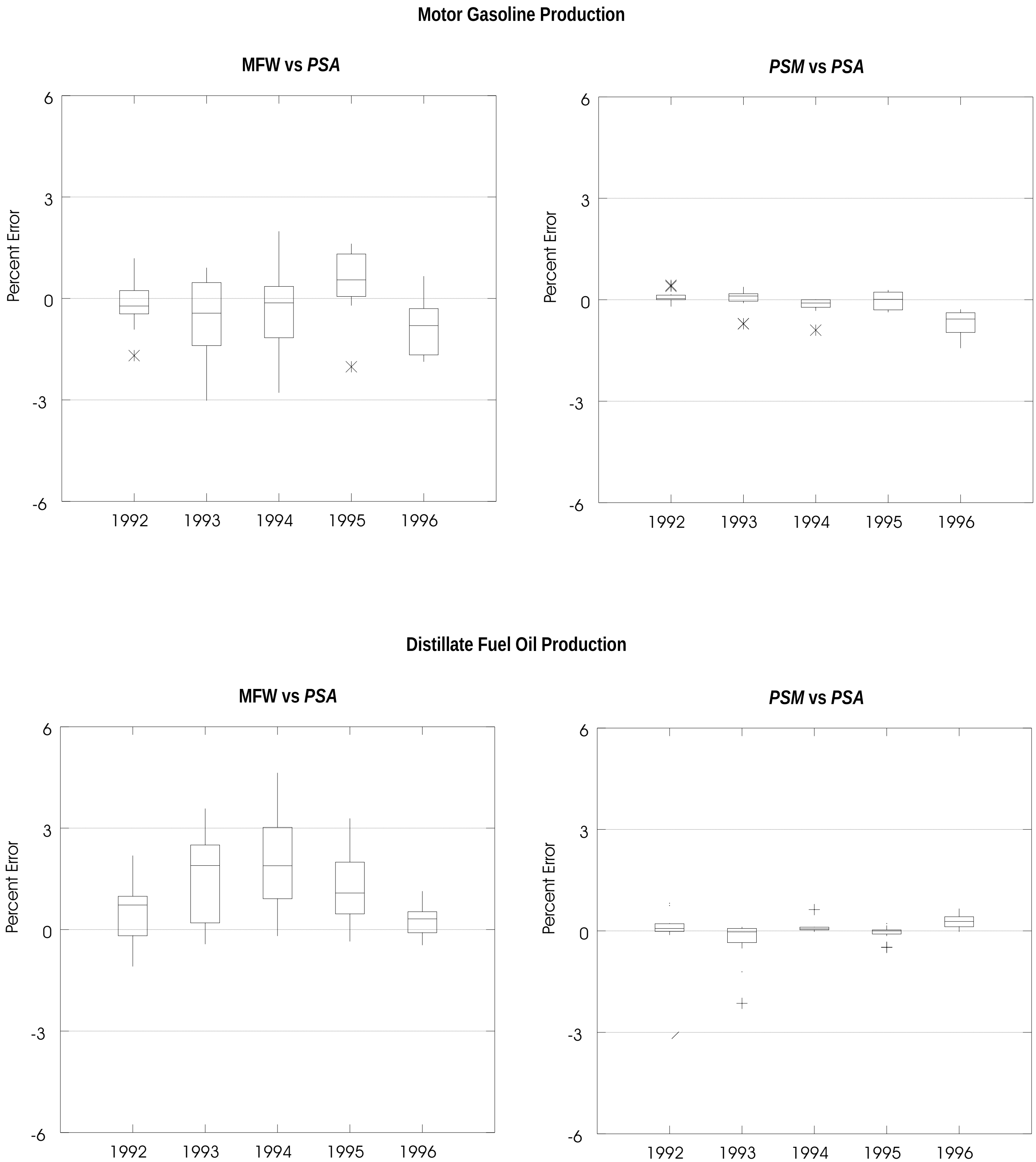


Figure FE3. Range of Percent Errors for MFW and *PSM* Crude Oil Production and Refinery Crude Oil Inputs Data, 1992 - 1996



Source: Energy Information Administration, Petroleum Supply Reporting System.

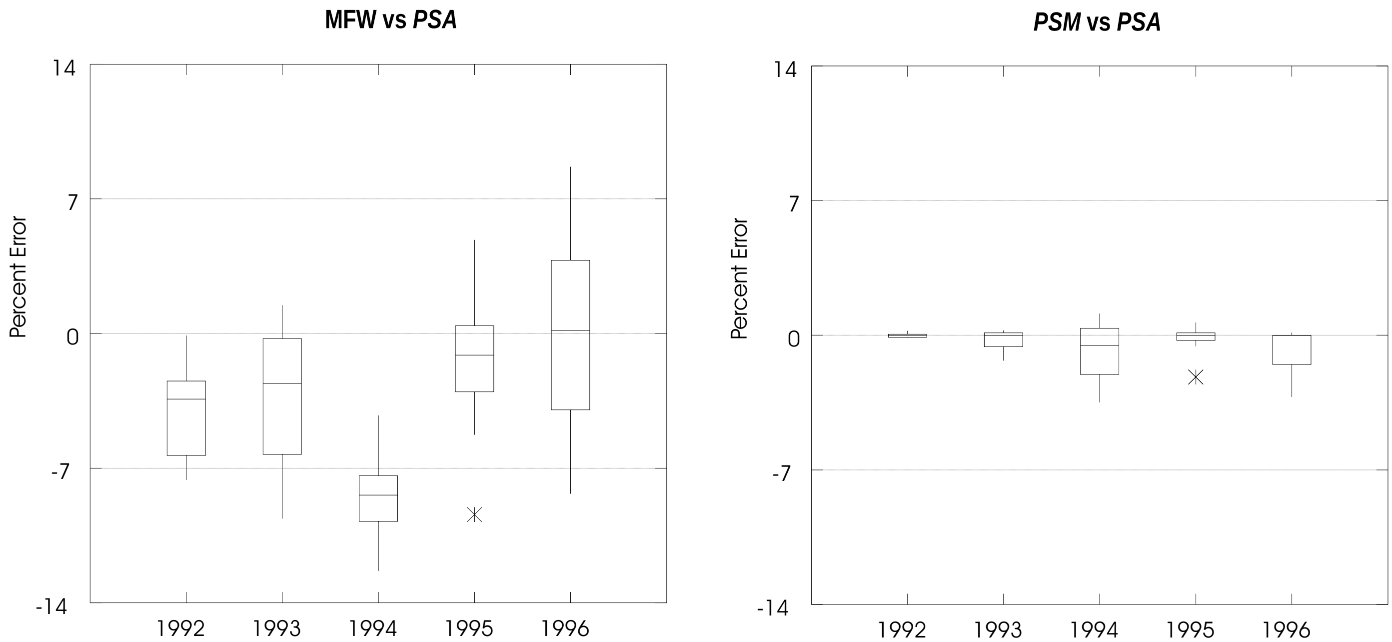
Figure FE4. Range of Percent Errors for MFW and *PSM* Motor Gasoline and Distillate Fuel Oil Production Data, 1992 - 1996



Source: Energy Information Administration, Petroleum Supply Reporting System.

Figure FE5. Range of Percent Errors for MFW and *PSM* Residual Fuel Oil and Jet Fuel Production Data, 1992 - 1996

Residual Fuel Oil Production



Jet Fuel Production

Source: Energy Information Administration, Petroleum Supply Reporting System.

For distillate fuel oil production, 1996 MFW percent errors had the smallest range over the 5-year period, ranging from -0.46 to 1.14 percent. The problem in prior years of respondents misclassifying residual fuel oil as distillate fuel oil on the weekly surveys has been resolved. As in prior years, *PSM* interim values for distillate fuel oil production are close to final values. All of the percent errors for 1996 were within 0.66 percent; and in 58 of the last 60 months, *PSM* percent errors have been within 1 percent of the final values. Distillate fuel oil production had the smallest range (0.69) of all the other *PSM* plots analyzed.

Figure FE5 shows the box and whisker plots for residual fuel oil production and jet fuel production. For MFW percent errors of residual fuel oil production, the largest range over the 5-year period was observed in 1996 (17.0). These values were not primarily underestimates as in the prior 4 years. The 1996 distribution of *PSM* percent errors for residual fuel oil ranged from -3.21 to 0.14 percent.

In 1996, the range of percent errors for the MFW and *PSM* values of jet fuel production was similar to the previous 4 years. The MFW percent errors, ranging from -1.80 to 4.58 percent, had the largest range over the 5-year period although the 1996 median was the closest to zero for that period. The outlier in October (4.58) due to company misreporting contributed to the large range. Similarly, the range for the *PSM* percent errors was the largest over the 5-year period, due to an outlier in November (0.93). This outlier was the largest percent error over the 60-month period. Also, there was an outlier in December (0.19). These outliers resulted from company resubmissions.

Stocks

Figures FE6, FE7, and FE8 show the yearly distribution of percent errors for stocks of crude oil, motor gasoline, distillate fuel oil, residual fuel oil, jet fuel, and propane. Figure FE6 shows the box and whisker plots for crude oil stocks and motor gasoline stocks. In contrast to 1995, the 1996 MFW estimates for crude oil did not predominately underestimate the final values. Over the 5-year period, the median for the 1996 percent errors was the closest to zero. In contrast to the last 2 years, the 1996 *PSM* interim values were mostly overestimates and were within 0.72 percent of the *PSA*.

The range of 1996 MFW percent errors for motor gasoline stocks, from -2.57 to 2.07 percent, was similar to the previous 4 years. The median of 0.34 percent is the closest to zero for the 5 years studied. The *PSM* percent errors were within 0.37 percent and the median was close to zero.

Figure FE7 shows box and whisker plots for distillate and residual fuel oil stocks. Similar to prior years, most of the MFW estimates for 1996 distillate fuel oil stocks were underestimates but the range of percent errors was the smallest over the 5-year period. The *PSM* percent errors were tightly grouped around the median of -0.09 percent and were mostly underestimates of the final *PSA* values.

Residual fuel oil typically has larger percent errors than other stock series. Most of the MFW values for 1996 were underestimates, ranging from -5.21 to 1.45 percent error. The 1996 range of the *PSM* percent errors (1.10) was the smallest over the 5-year period with most of the values within 0.50 percent.

The box and whisker plots for jet fuel stocks and propane stocks are shown in Figure FE8. In contrast to prior years, most of the 1996 MFW jet fuel stocks underestimated the final values. The percent errors ranged from -4.77 to 0.86 percent with a median of -1.28 percent. As in prior years, the 1996 *PSM* percent errors were tightly grouped about the median. In 1996, the range (0.76) and median (-0.01) of percent errors for jet fuel stocks were the smallest over the 5-year period.

Most of the 1996 MFW propane stocks underestimated the *PSA* values. The range for the percent errors was the largest over the 5-year period, with March 1996 having the largest absolute percent error over the 60-month period. For 1996, all of the *PSM* interim values were underestimates and all but two were within 1 percent of the final *PSA* values.

Imports

Figures FE9, FE10, and FE11 show the yearly distributions of percent errors for the imports of crude oil and four products: motor gasoline, distillate fuel oil, residual fuel oil, and jet fuel. Because of the irregularity of imports for crude oil and petroleum products, the magnitude and range of percent errors for both the MFW and the *PSM* imports numbers can be expected to be much larger and wider than for production and stocks.

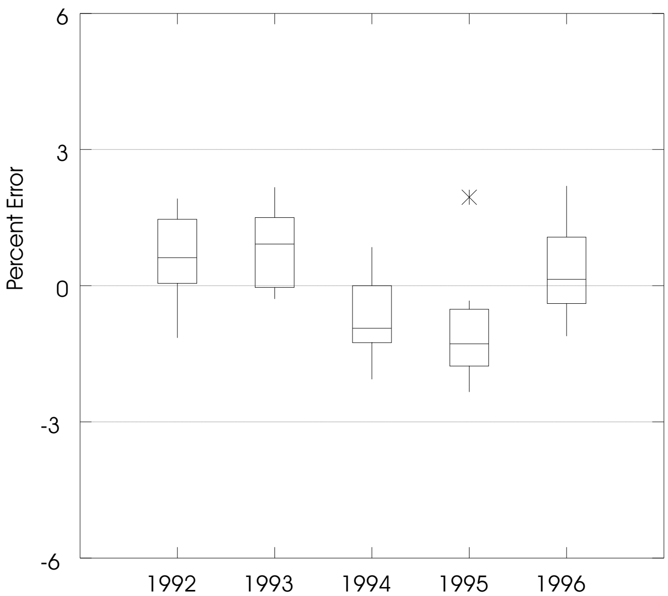
Figure FE9 shows that the range of the 1996 MFW percent errors of imports of crude oil is similar to prior years, ranging from -4.27 to 2.69 percent. All but one of the *PSM* percent errors were within 1 percent of the final values.

The distributions of percent errors of the MFW estimates and *PSM* interim values for 1992 through 1996 of motor gasoline and distillate fuel oil imports are shown in Figure FE10. In 1996, the MFW percent errors for motor gasoline imports ranged from -19.66 to 34.32 percent, with January having the largest percent error over the 60 months studied. This range is the largest for all the MFW plots analyzed for 1996. For all months in 1996, *PSM* interim values were overestimated and January had the largest percent error (13.20) over the 60-month period. The 1996 median percent error (3.29) was the largest compared to the prior 4 years.

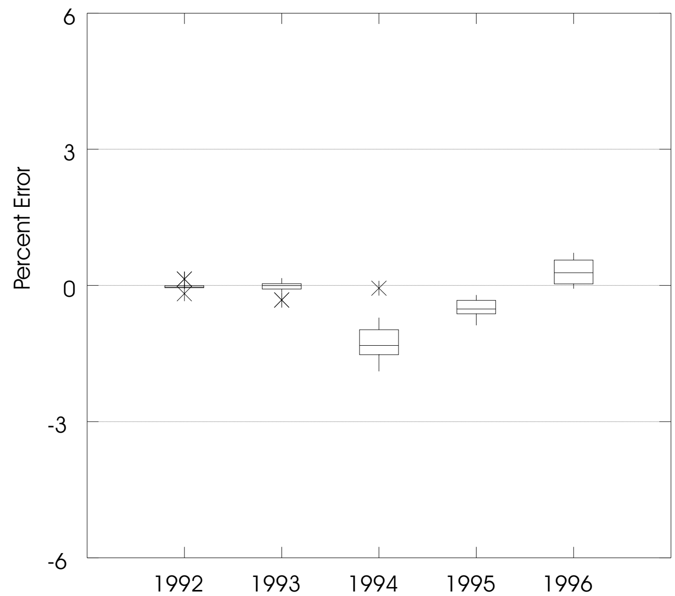
Figure FE6. Range of Percent Errors for MFW and *PSM* Crude Oil Stocks Excluding SPR and Motor Gasoline Stocks Data, 1992 -1996

Crude Oil Stocks Excluding SPR

MFW vs *PSA*

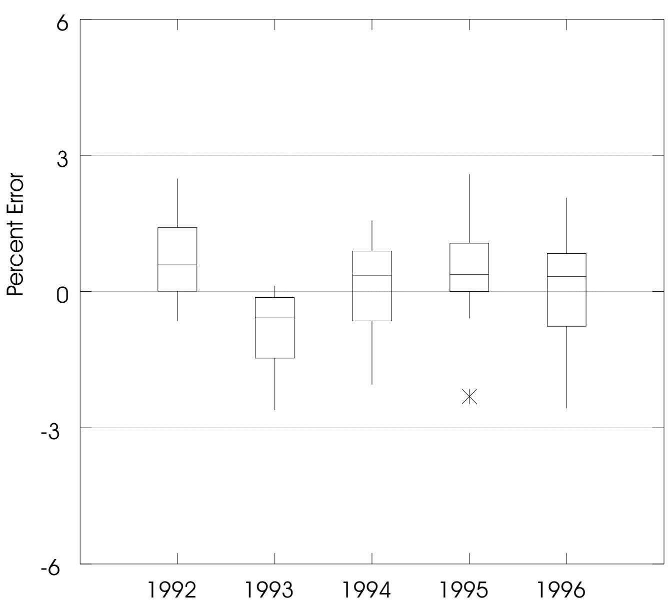


PSM* vs *PSA

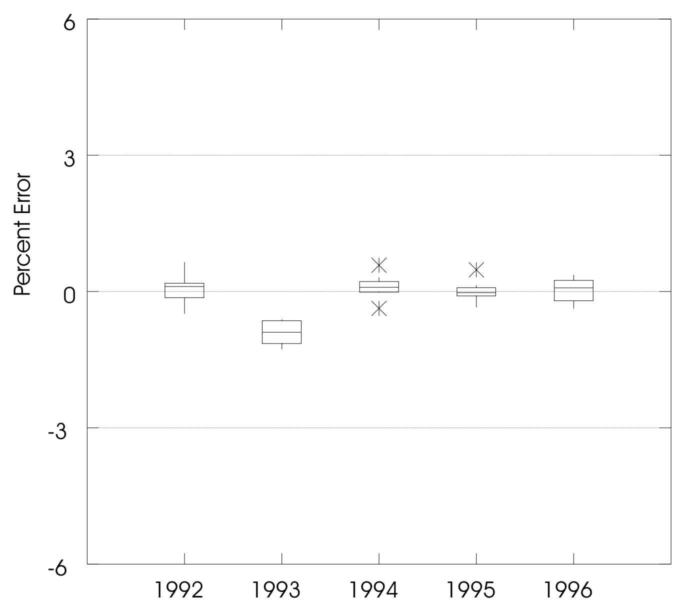


Motor Gasoline Stocks

MFW vs *PSA*

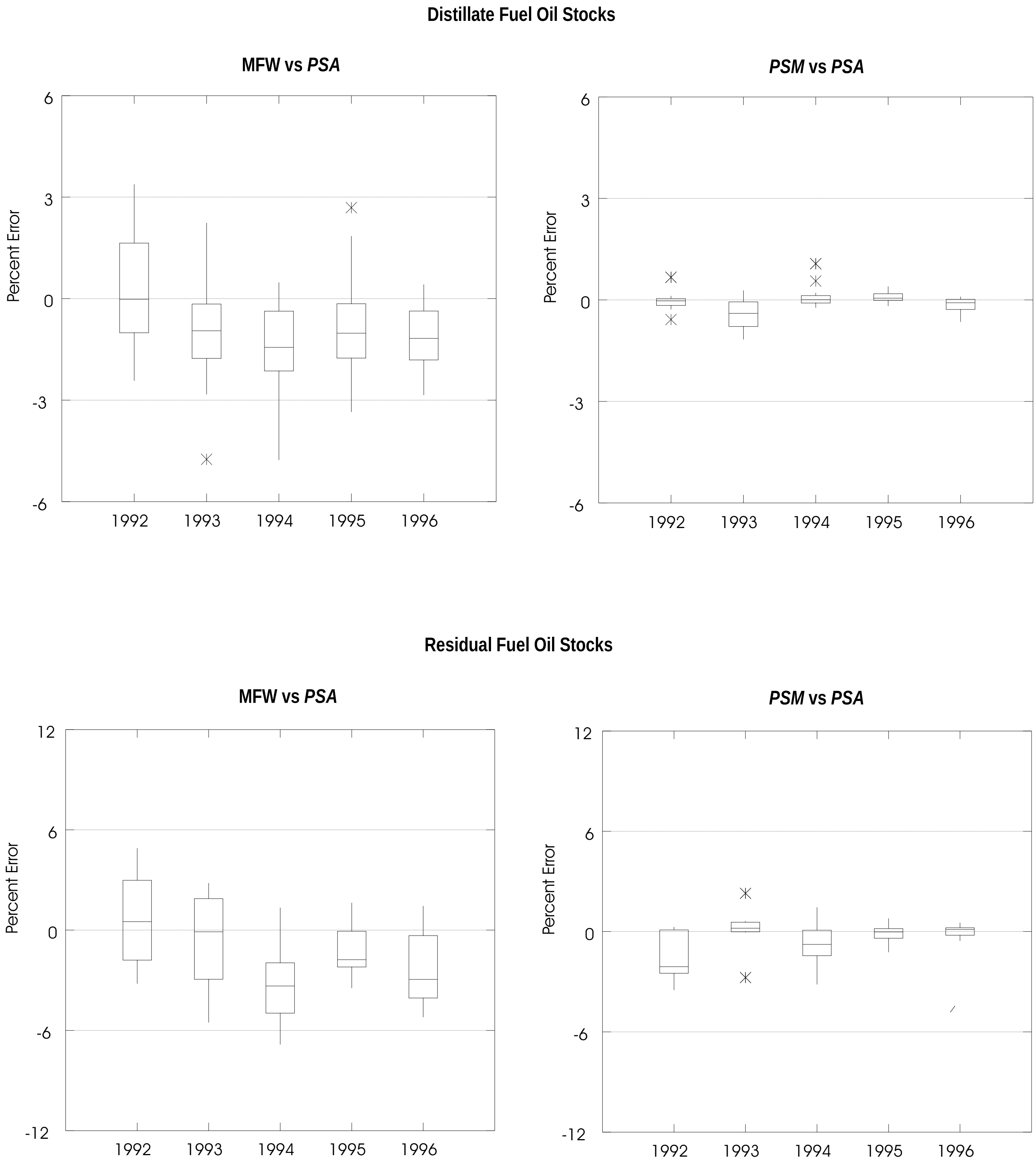


PSM* vs *PSA



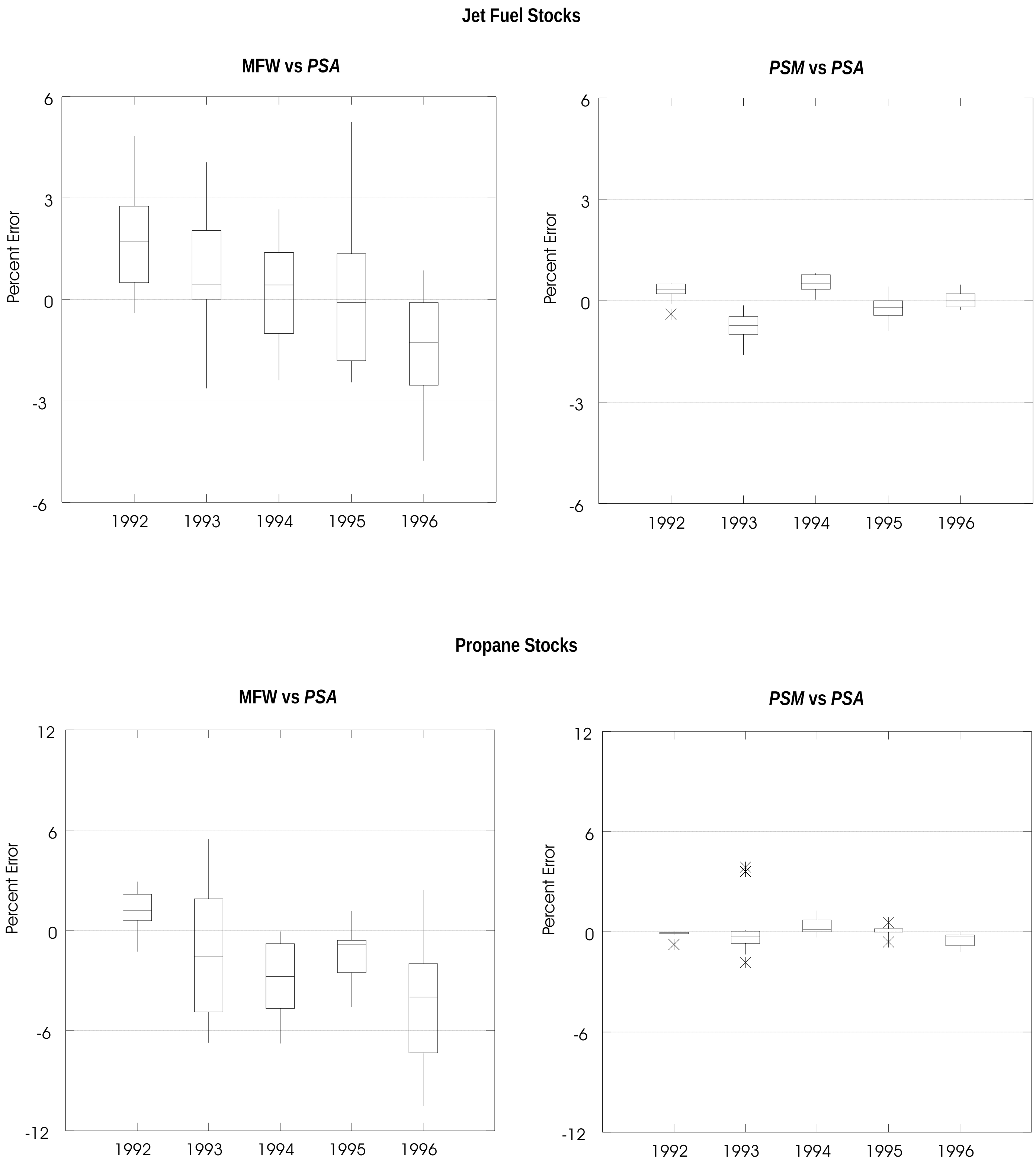
Source: Energy Information Administration, Petroleum Supply Reporting System.

Figure FE7. Range of Percent Errors for MFW and *PSM* Distillate Fuel Oil and Residual Fuel Oil Stocks Data, 1992 - 1996



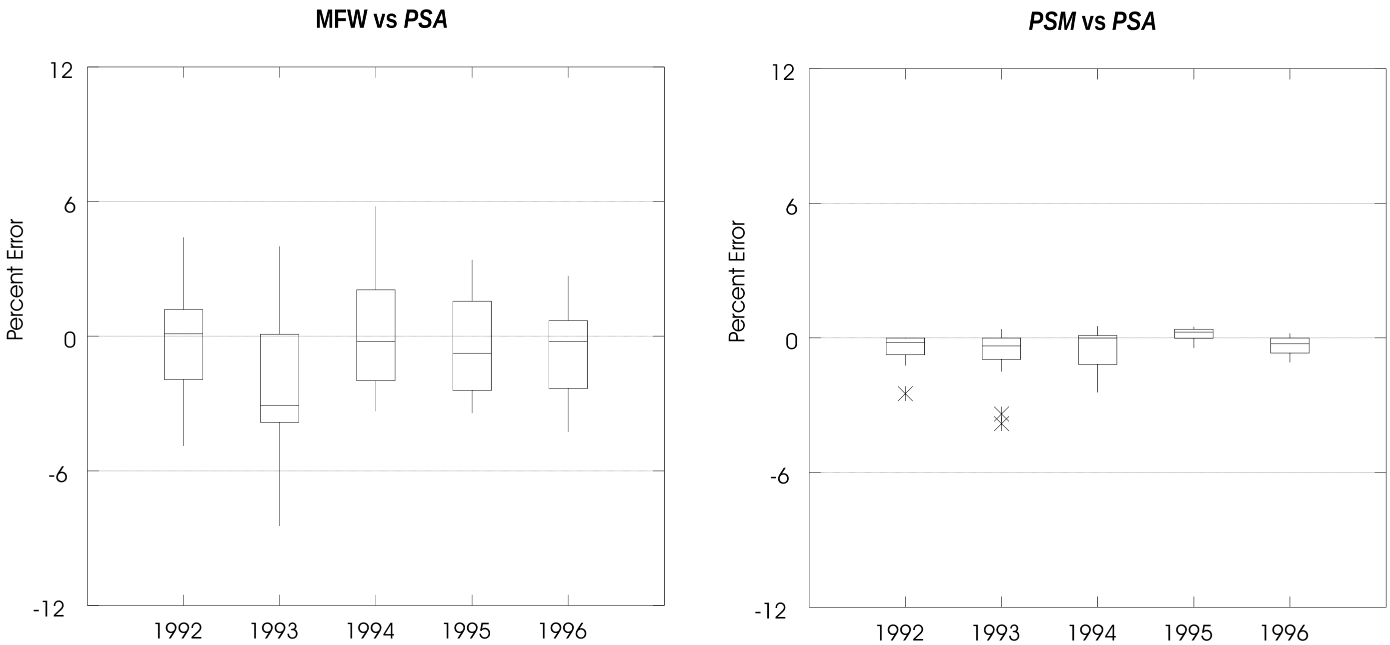
Source: Energy Information Administration, Petroleum Supply Reporting System.

Figure FE8. Range of Percent Errors for MFW and *PSM* Jet Fuel Stocks and Propane Stocks Data, 1992 - 1996



Source: Energy Information Administration, Petroleum Supply Reporting System.

Figure FE9. Range of Percent Errors for MFW and *PSM* Crude Oil Imports Excluding Strategic Petroleum Reserve Data, 1992 - 1996



Source: Energy Information Administration, Petroleum Supply Reporting System.

All of the 1996 MFW estimates for distillate fuel oil imports were understated except for the 2 outliers in February (12.90) and March (13.67). These were caused by the misreporting of high sulfur distillate. The median absolute percent error was the largest for the 5-year period. In contrast to prior years, there were no outliers for the 1996 *PSM* percent errors but there were more resubmissions.

Figure FE11 shows the box and whisker plots for residual fuel oil imports and jet fuel imports. As in prior years, most of the 1996 MFW estimates for residual fuel oil imports were underestimates. Over the 5-year period, 1996 had the largest range (45.85). Three outliers occurred in the 1996 *PSM* percent errors: June (3.57), August (-4.41), and November (-1.48). These were the only resubmissions for the year.

The range of percent errors for MFW estimates of jet fuel imports for 1996 was the smallest over the 5-year period, ranging from -23.60 to 8.99 percent. As in prior years, there was an outlier in *PSM* percent errors. This occurred in February (8.00) due to resubmissions. The range of 1996 percent errors was the largest of all *PSM* plots analyzed.

Conclusion

In summary, similar to previous years, the interim *PSM* data were closer in value to the final *PSA* volumes than the MFW estimates. This is largely a result of the longer time period provided to process the monthly data and monthly respondents' accounting systems.

In 1996, 47 of 66 interim values were within 1 percent (mean absolute percent error) of the final values; 26 of 61 MFW estimates were within 2 percent (mean absolute percent error) of the final values; and 11 of those 26 were within 1 percent. As in previous years, the accuracy of 1996 preliminary and interim values varied by product and by petroleum supply type. As a group, stocks continued to have the most accurate MFW estimates and *PSM* interim values.

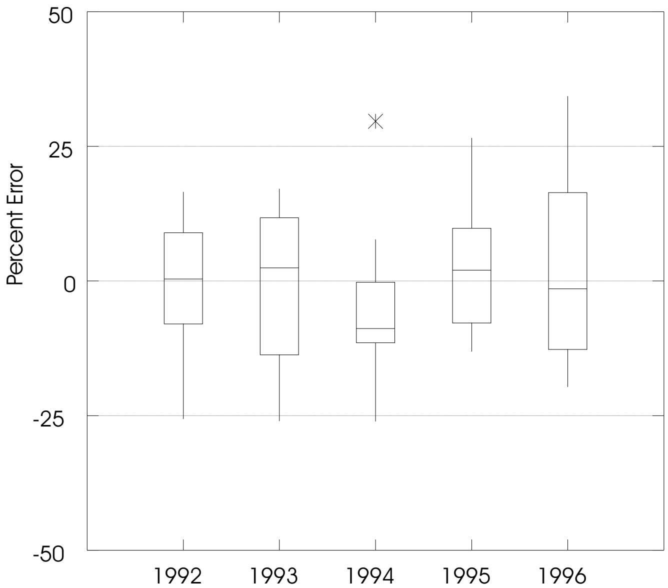
The good coverage for weekly surveys across petroleum supply type and product combinations has contributed to the accuracy of weekly estimates. In 1996, for 19 of the 21 categories, coverage was above 90 percent. The consistently high response rate, above 96 percent for the weekly surveys and above 97 percent for the monthly surveys, contributes to the high level of accuracy of these data.

To successfully maintain and improve the accuracy of these data, PSD has made progress in its business re-engineering effort. The organization has been restructured to facilitate major processes including “getting data in the door,” survey management, editing, statistical methodology, publication, and customer outreach. Some of the techniques being researched and developed are graphical data validation and optical scanning. Improvements are being made in survey design, sampling, editing procedures, and edit parameters. The result of these efforts should enable the PSD to provide more accurate weekly and monthly data estimates.

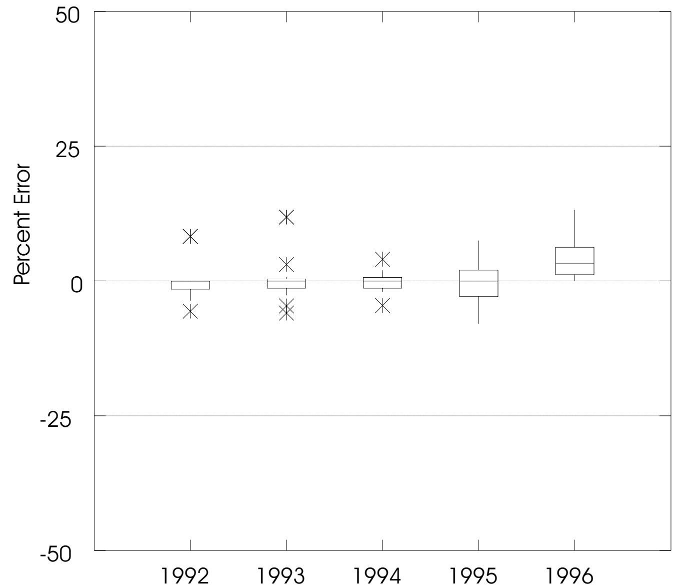
Figure FE10. Range of Percent Errors for MFW and *PSM* Motor Gasoline and Distillate Fuel Oil Imports Data, 1992 - 1996

Motor Gasoline Imports

MFW vs *PSA*

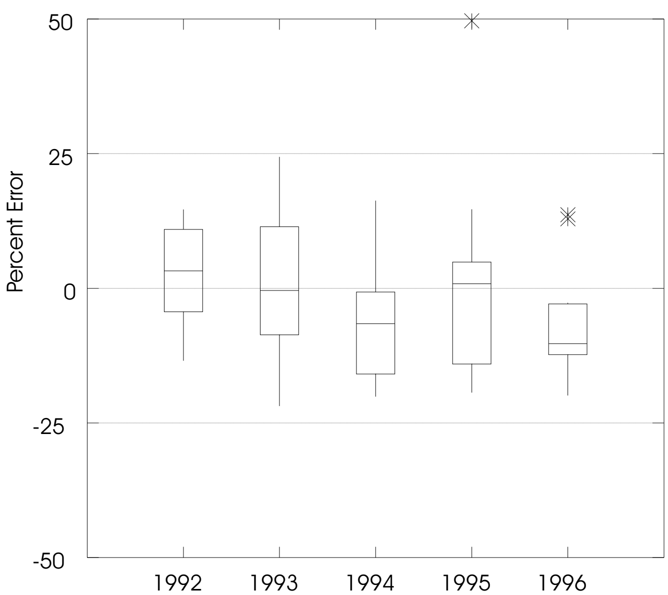


PSM* vs *PSA

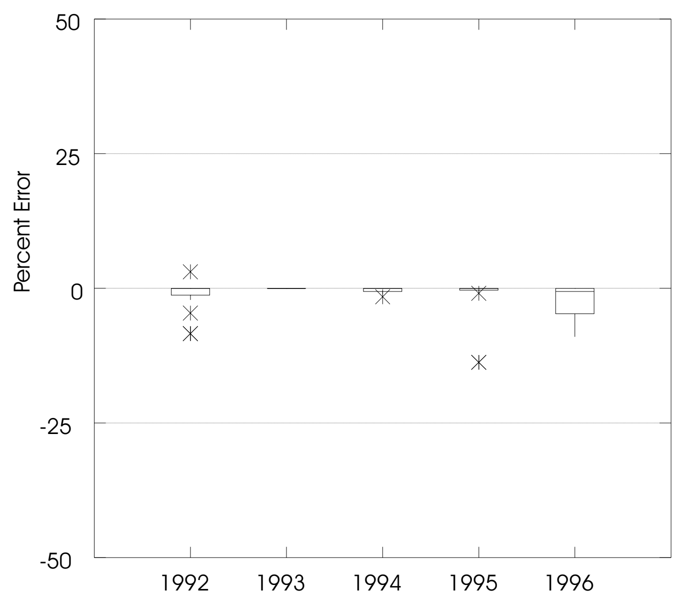


Distillate Fuel Oil Imports

MFW vs *PSA*

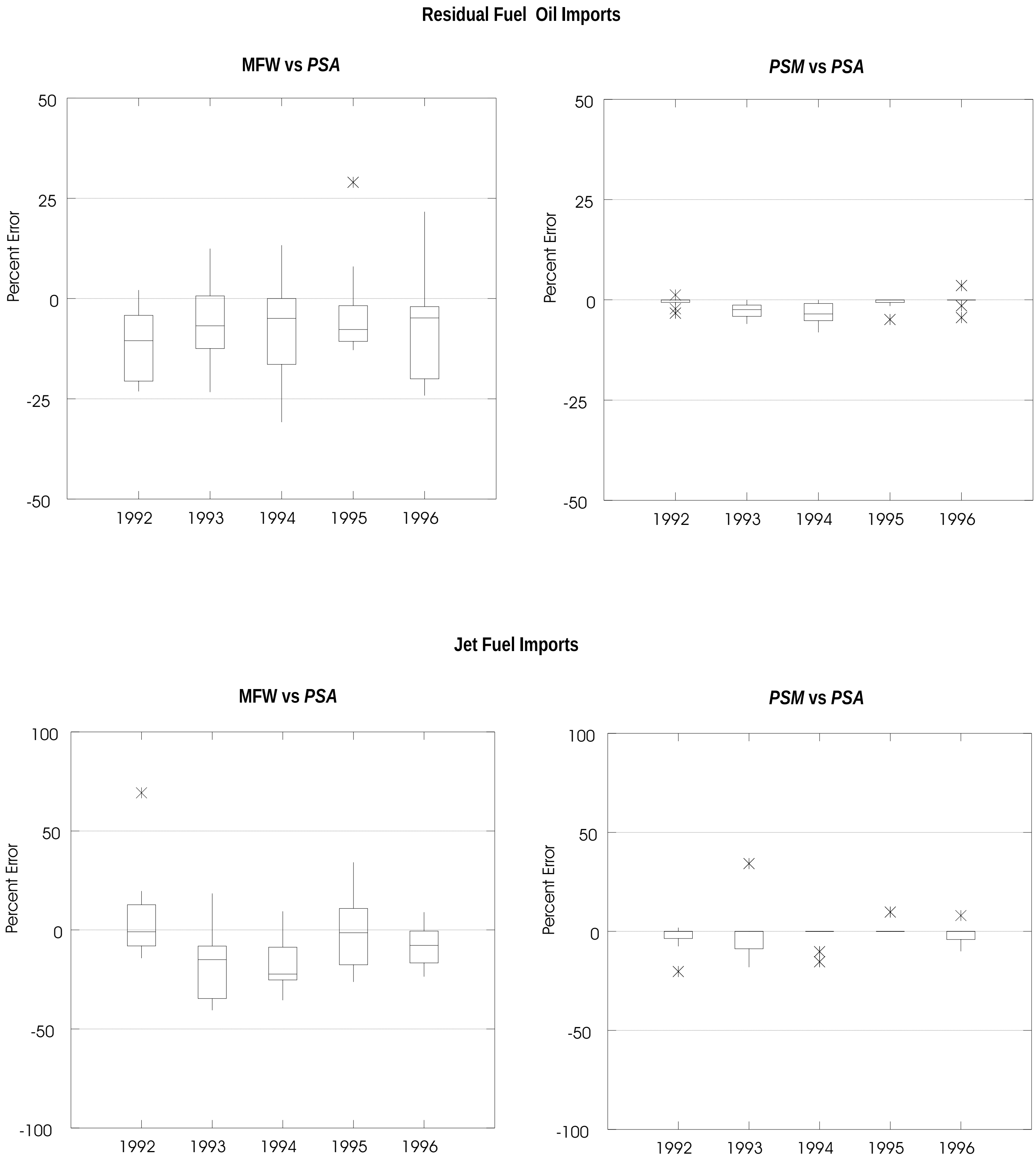


PSM* vs *PSA



Source: Energy Information Administration, Petroleum Supply Reporting System.

Figure FE11. Range of Percent Errors for MFW and *PSM* Residual Fuel Oil and Jet Fuel Imports Data, 1992 - 1996



Source: Energy Information Administration, Petroleum Supply Reporting System.