Appendix D
Statistical Methodology of Estimating Petroleum Exports Using Data from U.S. Customs and Border Protection

Updated August 16, 2017

Introduction

Beginning August 31, 2016, EIA took its first step to incorporate export data collected by the U.S. Customs Border Protection (CBP) through the Automated Commercial Environment (ACE) system in estimating U.S. petroleum exports used in the *Weekly Petroleum Status Report* (WPSR). The updated WPSR methodology provides a more accurate estimate of U.S. petroleum exports to better inform EIA's weekly estimates of domestic quantity demanded for crude oil, petroleum products, and biofuels.

Exports are an important component for the calculation of EIA weekly estimates of quantity demanded for petroleum products. EIA calculates domestic quantity demanded (measured as product supplied) for petroleum products as:

Domestic quantity demanded = Production + Imports - Stock change - Exports.

EIA conducts weekly surveys to collect data that are used to estimate production, imports and stocks, but not exports. Export data are published on a monthly basis by the U.S. Census Bureau (USCB). To estimate weekly exports, EIA previously used forecasts based on an autoregressive integrated moving-average (ARIMA) procedure to model weekly export statistics using the latest USCB monthly data. EIA obtained official USCB export data for crude oil, petroleum products, and biofuels on a monthly basis approximately six weeks after the close of the reporting month and published them in the *Petroleum Supply Monthly* (PSM). These data provided the inputs to the weekly export ARIMA model. However, as the volume and variability of U.S. exports increased, significant discrepancies were sometimes observed between modeled weekly estimates and final monthly exports data provided by the USCB. Because of the growing impact of exports on EIA's weekly estimates of U.S. petroleum quantity demanded, EIA recognized the need to improve the weekly export estimates used in the WPSR.

Beginning in 2013, EIA investigated data sources and methodologies to improve its weekly estimates of petroleum and biofuels exports. EIA found that administrative data collected by CBP through the ACE system, which cover U.S. exports except for most of those going to Canada, could be used to improve EIA's weekly export estimates. To obtain these data, EIA established a Memorandum of Understanding with CBP in September 2014 through the International Trade Data System (ITDS) process. In August 2016, EIA obtained a letter from Office of Management and Budget (OMB) granting EIA an exception to the provisions of OMB's *Statistical Policy Directive No. 3 on the Compilation, Release, and Evaluation of Principal Federal Economic Indicators.* This exception permits EIA to publish weekly estimates of export volumes of crude oil and six refined petroleum products and biofuel categories in advance of the release of official U.S. trade data from USCB.

Since August 31, 2016, EIA has monitored and continuously analyzed the estimates derived from the recently acquired near real-time CBP export data. After reviewing the results to date, EIA further revised its estimation methodology for crude oil and finished gasoline exports in two phases during the Spring of 2017, as described in the next section.

U.S. Energy Information Administration | Statistical Methodology of Estimating Petroleum Exports Using Data from U.S. Customs and Border Protection

1

Updates to WPSR's Methodology for Estimating Exports of Crude Oil and Finished Gasoline

Phase 1: Implemented starting with the WPSR published on March 8, 2017

- Updated the regression model that is used to estimate total exports from edited CBP data. The regression model uses CBP data as an input and accounts for U.S. exports to Canada that are missing from CBP data. The updated model, referred to as Model 2 (M2) in this document, includes an intercept term and uses daily CBP data that were edited to identify and impute questionable quantities and then aggregated to monthly totals as input. The prior model, referred to as Model 1 (M1) in this document, used monthly USCB data that were interpolated to a weekly series, while M2 uses daily edited CBP data that are aggregated to a monthly series.

  Began using M2 to estimate finished motor gasoline exports. The prior method had been to report finished motor gasoline exports equal to the aggregation of CBP data after editing and imputation.

Phase 2: Implemented starting with the WPSR published on June 7, 2017

- Began using M2 to estimate crude oil exports, replacing the use of the Unobserved Components Model (UCM).

For the previous methodology descriptions, please refer to the archive to find the WPSR for the week of interest.

The following is a summary of the updated statistical methodology for estimating crude oil and finished gasoline exports in the WPSR. The new methodology provides more accurate estimates of crude oil and finished gasoline than previous methods. The estimation methods for distillate fuel, kerosene-type jet fuel, propane, residual fuel oil, and other oils remain unchanged from the methodology presented with the WPSR published on August 31, 2016.

Trade Data Collection and Dissemination

To derive weekly estimates of domestic petroleum product supplies, the WPSR relies on estimates of weekly U.S. exports of crude oil, finished motor gasoline, kerosene-type jet fuel, distillate fuel oil, residual fuel oil, propane, and other oils including biofuels, asphalt, aviation gasoline, butanes, ethane, lubricants, gasoline blending components, miscellaneous products, non-biofuel oxygenates, petroleum coke, unfinished oils, and wax. The export data that EIA uses to develop these weekly estimates are collected by CBP.

In general, entities exporting products from the United States are required to file export transaction data with CBP. This information is then included in CBP's ACE system. As a result of an agreement between the U.S. Customs and Border Protection Agency, EIA now receives these unedited transactional ACE data files from CBP and uses them to develop aggregate estimates of weekly exports that are used in the WPSR. The data from CBP include an export date that makes it possible to group transactions into

WPSR weekly reporting periods. EIA has found that data received from CBP require application of edits and imputation to avoid using questionable data values for analysis, modeling, and aggregation. Edit and imputation methods applied by EIA are described later in this document. EIA is unable to follow up with exporters to resolve questions about the CBP data. USCB may follow up with exporters to address questionable data and, if necessary, obtain corrections during their preparation of official U.S. trade statistics.

Per U.S.-Canadian agreements, entities exporting to Canada are not required to report to CBP, so these files generally exclude Canadian transactions. Products that require an export license are an exception, and export transactions involving licensed products must be reported to CBP regardless of the destination country. Crude oil was a licensed export product until December 2015. USCB has an agreement with Statistics Canada to obtain monthly Canadian imports data from the United States, which USCB then uses to produce monthly estimates of U.S. exports in official U.S. trade statistics. However, EIA does not currently have early access to Canadian import data. To address these issues, EIA developed methodologies to edit the raw CBP data and to estimate weekly exports to Canada.

The CBP data EIA uses to estimate weekly exports begin with the same data that the USCB uses as the initial input for preparation of the monthly *International Trade in Goods and Services* report (FT900), a principal economic indicator. The FT900 has a six-week publication lag, while WPSR data are released each Wednesday for the weekly report period ending the previous Friday. Because the monthly publication cycles are more closely aligned, the FT900 data are used by EIA to report monthly export volumes that are included in EIA's *Petroleum Supply Monthly* (PSM) and *Petroleum Supply Annual* (PSA).

Estimation Methodology

Prior to producing the published estimates for the WPSR, EIA performs the following edit steps after receipt of CBP data for all products:

- Remove duplicate records

- Exclude records with entry dates beyond the reporting period

- Exclude products that are out of scope of the WPSR

- Convert to barrels exported quantities reported in other units

- For each product, calculate the dollar value per barrel from CBP and compare against daily spot prices from Thomson-Reuters to identify extreme values. Use a procedure based on shipment or commodity weight to replace these values, if necessary. Shipment weight is the combined weight of the commodity being shipped and its packaging. Based on EIA's research, the procedure of imputing for extreme exported quantities based on commodity weight typically appears to be slightly more accurate than a procedure based on shipment weight, and much more accurate than a procedure based on shipment or commodity value.

See Attachment A for a more complete list of edit rules that EIA applies to the CBP data.

Following editing, EIA's estimation method consists of two components:

1. Estimates based on aggregations of edited CBP data

   For those products that have accurate coverage in the CBP data, EIA reports the sum of the quantities exported using the edited data. These products include distillate fuel, kerosene-type jet fuel, and propane.

2. Estimates based on post-editing statistical models

   For those products that appear to have coverage differences in the CBP data based on EIA's comparisons to the USCB monthly volumes that are published in the PSM, EIA uses post-editing statistical model to correct for the coverage differences. These products include crude oil, finished motor gasoline, residual fuel oil, and other oils.

Estimation Methodology and Results for Specific Products

Estimates based on aggregations of edited CBP data

EIA's analysis of the CBP data has shown that, once edited, these data can be used to accurately estimate weekly export volumes for propane. For distillates and kero-type jet fuel, EIA is currently using edited CBP data and examining other methods to improve estimates. EIA will continue to monitor weekly export estimates for these products and may use some form of a post-editing statistical model if the weekly estimates consistently diverge noticeably from USCB's monthly exports data.

For the months from September 2016 to May 2017, Figures 1–3 show the aggregated weekly estimates for distillate fuel, kerosene-type jet fuel, and propane produced by using edited CBP data (blue line) compared to USCB/PSM monthly volumes (red line). In Figures 1 and 2, the dotted green line shows what the estimates would have been using the updated regression model (M2). It is not clear from these figures that M2 generally improves the accuracy of the estimates of distillate fuel oil or kerosene-type jet fuel exports, and EIA plans to continue to research possible ways to improve the model. In Figure 2, the shaded gray area shows the portion of the PSM volumes that consists of kerosene-type jet fuel exports to Canada, which is typically no more than 50,000 barrels per day. As can be seen in this figure, the differences between the edited CBP and USCB/PSM series are typically lower than 50,000 barrels per day in absolute value.

Figure 1. Monthly distillate fuel oil exports for September 2016–May 2017
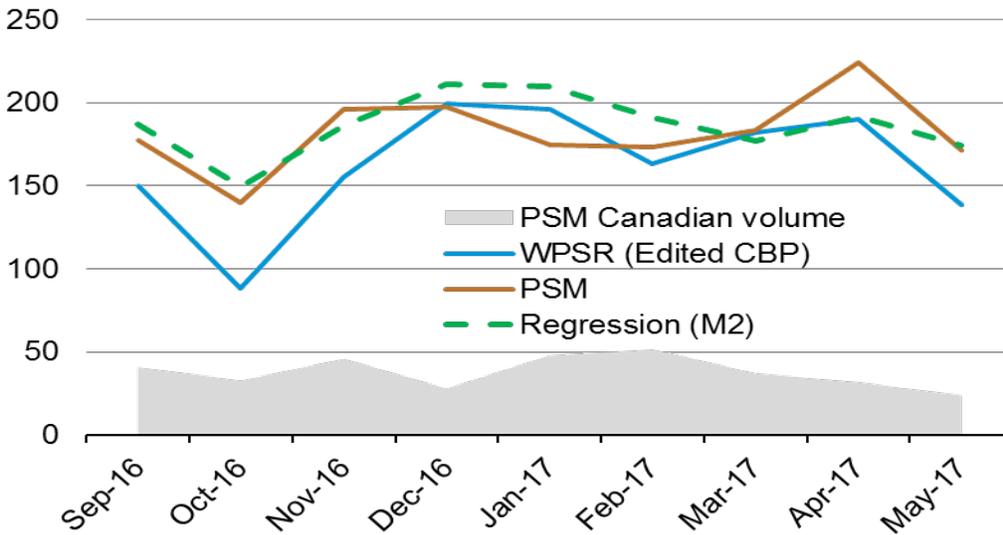
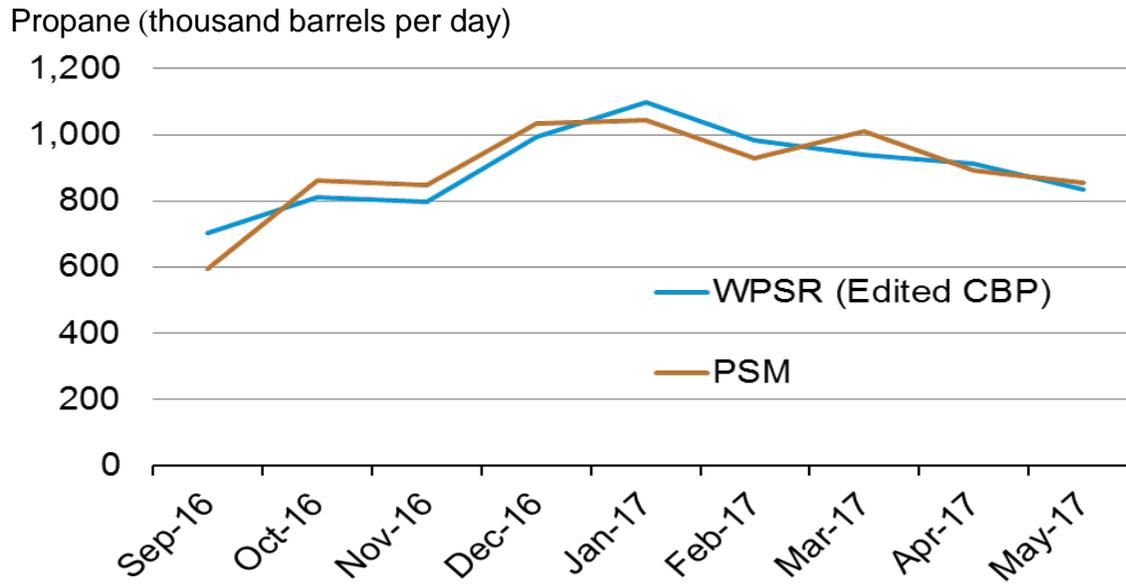Distillate fuel oil (thousand barrels per day)
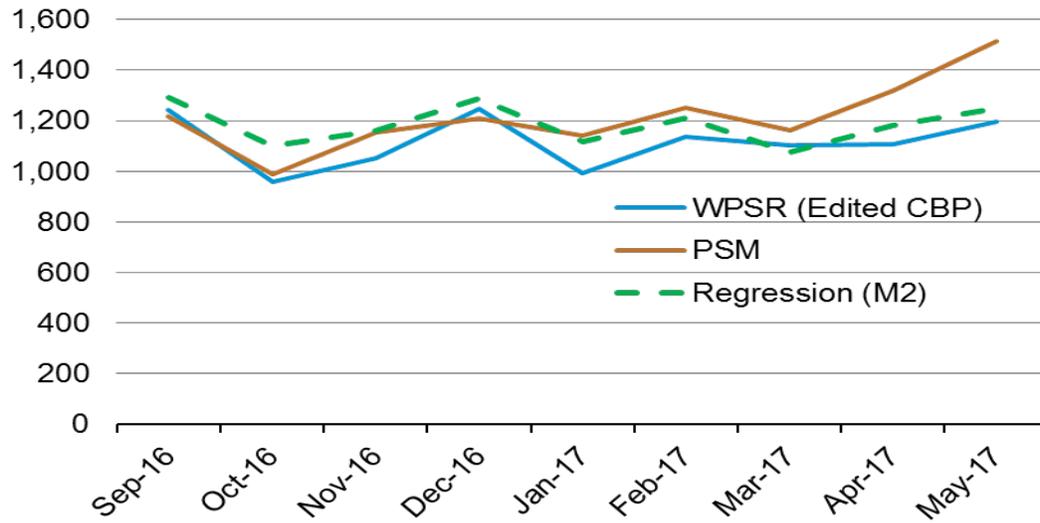
Source: U.S. Energy Information Administration

Figure 2. Monthly kerosene-type jet fuel exports for September 2016–May 2017

Kerosene-type jet fuel (thousand barrels per day)

Source: U.S. Energy Information Administration

Figure 3. Monthly propane exports for September 2016–May 2017

Propane (thousand barrels per day)

Source: U.S. Energy Information Administration

Estimates based on post-editing statistical models

Based on PSM data from September 2016 to March 2017, U.S. exports to Canada account for roughly 40% of total crude oil exports, 12% of total residual fuel oil exports, and 25% of total other oils exports, most of which are not captured in CBP's weekly export files. To address coverage differences for each of these fuels, EIA applies a linear regression model to more accurately estimate the weekly export volumes. In addition, although only about 4% of U.S. finished motor gasoline exports are sent to Canada, EIA found inconsistent reporting in the CBP files that typically led to overestimation when compared to the USCB volumes. This inconsistency necessitated applying a linear regression model to generate more accurate estimates of finished motor gasoline exports beginning March 2017.

The formula for the updated regression model (M2) is:

$$Y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$$

where *t* denotes a month and:

- Y_t is the export of crude oil, finished gasoline, residual fuel oil, or other oils in barrels/day as published in the PSM for month *t*
- x_t is edited CBP export data in barrels/day aggregated for month *t* to be used as the regressor
- β_0 is the intercept
- β_1 is the regression coefficient for the regressor x_t
- ε_t is the error term

The estimated regression coefficients are then applied to the edited CBP export data on a weekly basis to produce the published weekly exports data. This method assumes a similar linear relationship between the PSM and CBP data on a weekly basis for the weeks in a given month. However, if the estimate of weekly exports based on the regression model is less than the reported CBP exports, regression is not applied and instead the reported data (with edits) are published. EIA continues to assess the results of this approach and continues to identify alternative methods to develop weekly export volume estimates, if needed.

For crude oil, Figure 4 demonstrates the typical underestimation of the previous Unobserved Components Model (blue line) compared to M2 (green line) for months from September 2016 to May 2017. In weeks where the edited CBP export volumes exceeded the estimate from the Unobserved Components Model, EIA used the edited CBP exports for WPSR. The shaded gray area shows the portion of the PSM volumes that consists of crude oil exports to Canada, which typically ranges between 200,000 b/d and 400,000 b/d. Although M2 appears to generally result in more accurate estimates for crude oil, EIA continues to research ways to improve the model, particularly when the estimate based on the model is lower than the edited CBP exports.

Figure 4. Comparing the Unobserved Components Model and regression model of crude oil exports for September 2016–May 2017

Crude oil exports (thousand barrels per day)
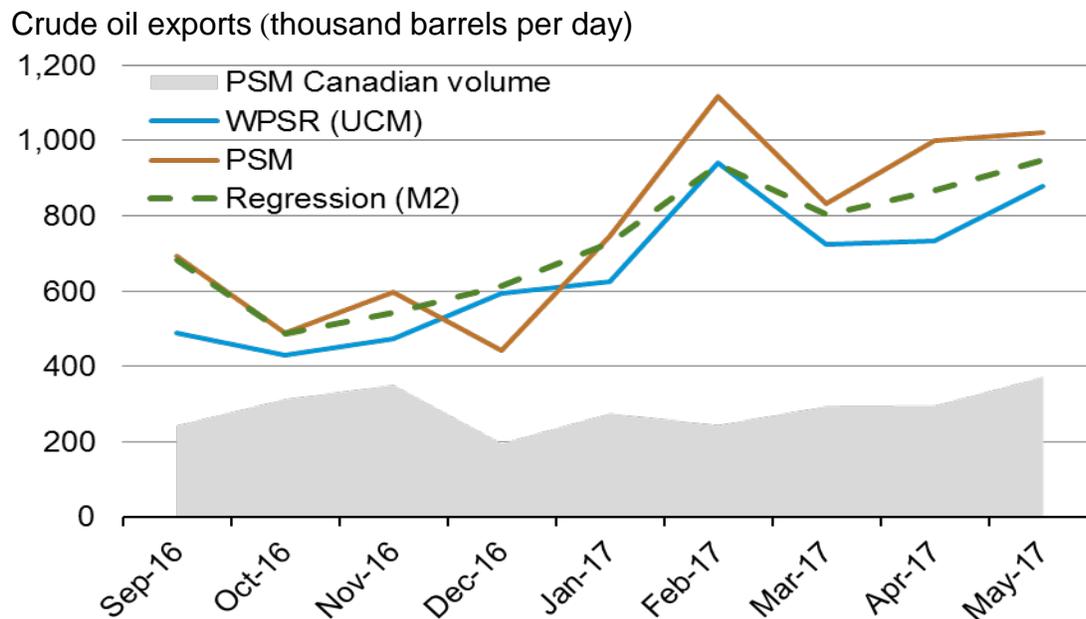
Source: U.S. Energy Information Administration

Figures 5–7 show monthly estimates of other oils, residual fuel oil, and finished motor gasoline exports produced by the regression models (blue and dotted green lines) compared to USCB/PSM volumes (red line). The exports used in the WPSR are shown by the blue line, and the dotted green line shows what the estimates would have been using M2 prior to March 2017. The dotted blue line indicates when the updates to the regression model were made for the WPSR in March 2017. In Figure 5, the shaded gray area shows the portion of the PSM volumes that consists of other oils exports to Canada, which typically

U.S. Energy Information Administration | Statistical Methodology of Estimating Petroleum Exports Using Data from U.S. Customs and Border Protection

7

is about 400,000 b/d. Figure 7 shows the typical overestimation of the edited CBP data when compared with the USCB/PSM volumes.

Figure 5. Monthly other oils exports for September 2016–May 2017

Total other oils (thousand barrels per day)
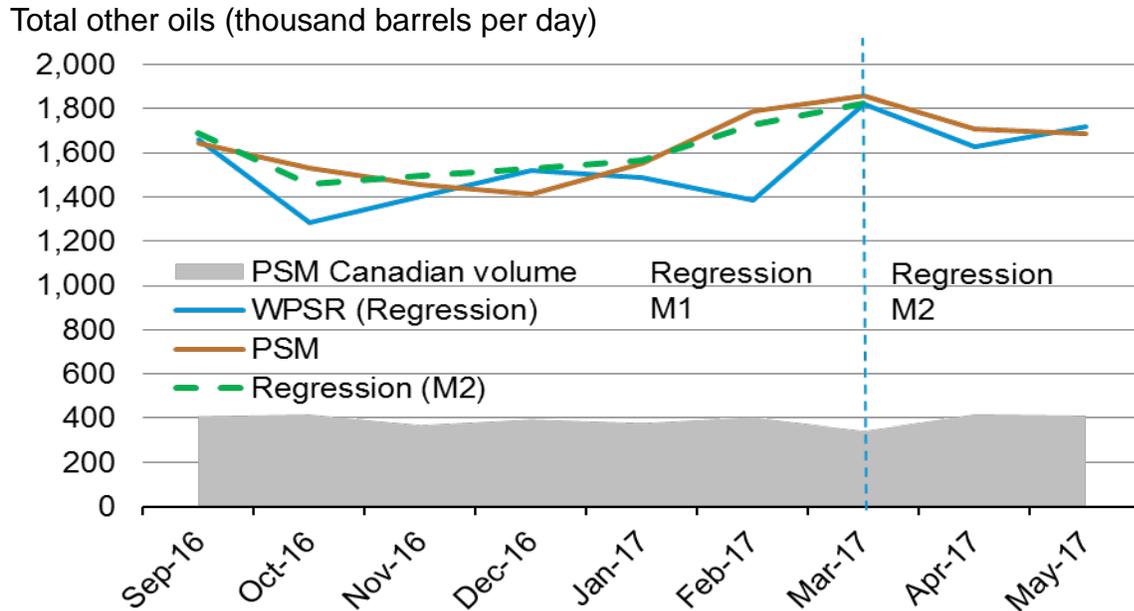
Source: U.S. Energy Information Administration

Figure 6. Monthly residual fuel oil exports for September 2016–May 2017

Residual fuel oil (thousand barrels per day)
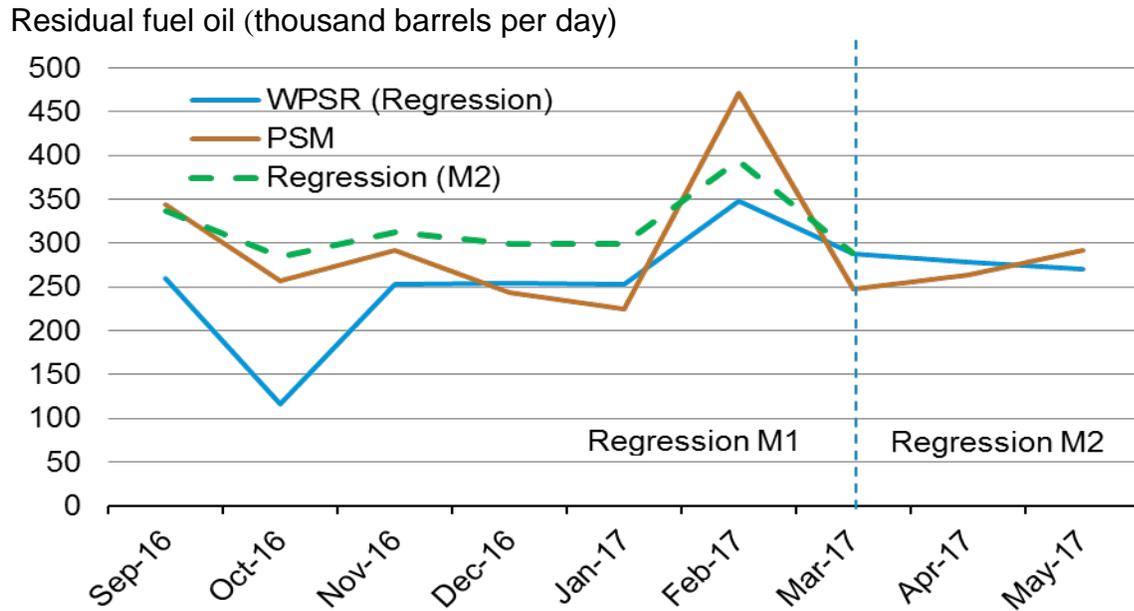
Source: U.S. Energy Information Administration

Figure 7. Monthly finished motor gasoline exports for September 2016–May 2017

Finished motor gasoline (thousand barrels per day)
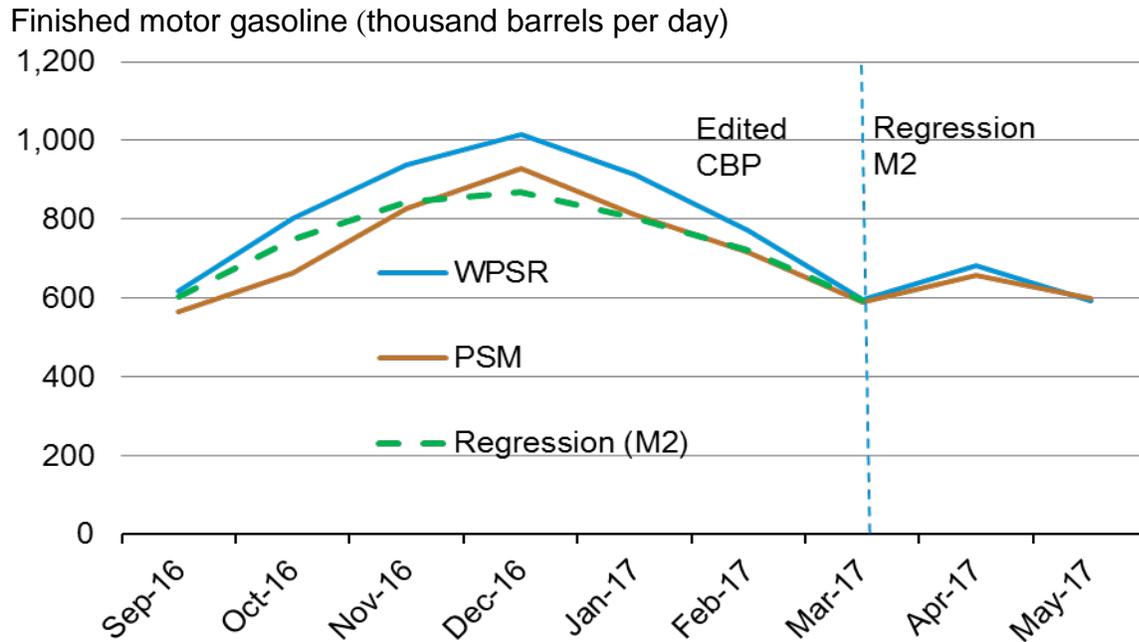
Source: U.S. Energy Information Administration

For September 2016 to May 2017, Figure 8 shows total monthly estimates of exports used for the WPSR compared with USCB/PSM volumes. The differences between these estimates are typically within 10% of the PSM export volumes in absolute value.

Figure 8. Monthly crude oil, petroleum products and biofuels exports for September 2016–May 2017

Total exports (thousand barrels per day)

Source: U.S. Energy Information Administration

U.S. Energy Information Administration | Statistical Methodology of Estimating Petroleum Exports Using Data from U.S. Customs and Border Protection

9
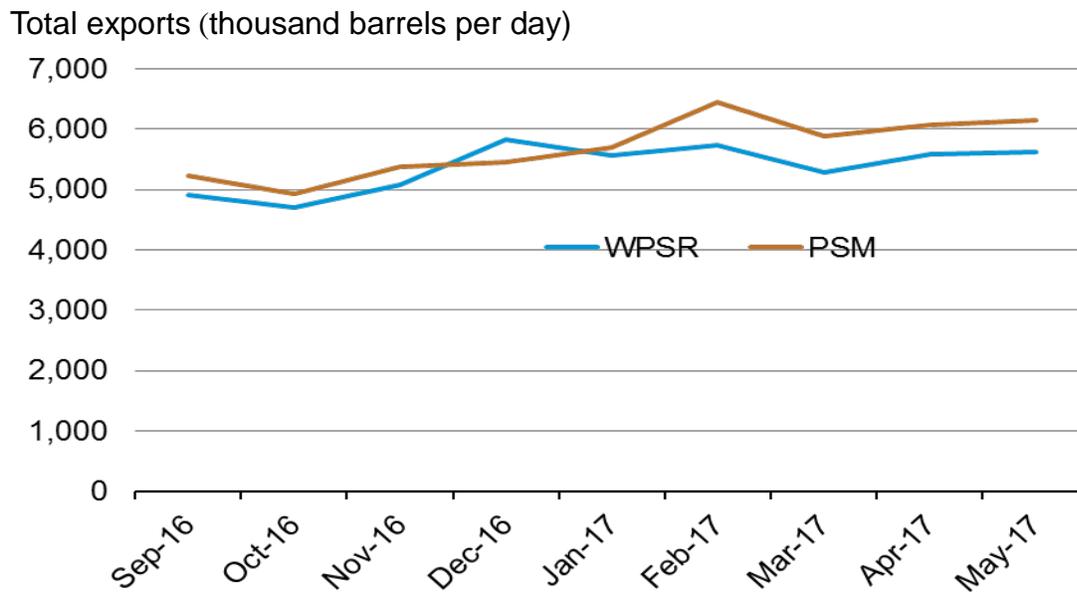
Summary

By making use of near real-time petroleum export data from CBP, published monthly trade data from the USCB, and revised estimation methodologies, EIA's WPSR reports estimates of U.S. weekly exports with improved accuracy. Table 1 below summarizes the estimation methodology used in WPSR for a given product or its component since EIA began incorporating data collected by CBP through the ACE system. EIA will continue to monitor the performance of these methods, and as with crude oil and finished gasoline, EIA will modify them as needed to further improve their accuracy.

Table 1. Summary of Estimation Methodology in WPSR by Product or Component

Product/Component	2016				2017						
	S	O	N	D	J	F	M	A	M	J	J
Crude oil	UCM	UCM	UCM	UCM	UCM	UCM	UCM	UCM	UCM	Model 2	Model 2
Finished motor gasoline	Edited CBP	Edited CBP	Edited CBP	Edited CBP	Edited CBP	Edited CBP	Model 2	Model 2	Model 2	Model 2	Model 2
Kerosene type jet fuel	Edited CBP	Edited CBP	Edited CBP	Edited CBP	Edited CBP	Edited CBP	Edited CBP	Edited CBP	Edited CBP	Edited CBP	Edited CBP
Distillate fuel oil	Edited CBP	Edited CBP	Edited CBP	Edited CBP	Edited CBP	Edited CBP	Edited CBP	Edited CBP	Edited CBP	Edited CBP	Edited CBP
Residual fuel oil	Model 1	Model 1	Model 1	Model 1	Model 1	Model 1	Model 2	Model 2	Model 2	Model 2	Model 2
Propane	Edited CBP	Edited CBP	Edited CBP	Edited CBP	Edited CBP	Edited CBP	Model 2	Model 2	Model 2	Model 2	Model 2
Mogas blending components	Model 1	Model 1	Model 1	Model 1	Model 1	Model 1	Model 2	Model 2	Model 2	Model 2	Model 2
Fuel ethanol	Model 1	Model 1	Model 1	Model 1	Model 1	Model 1	Model 2	Model 2	Model 2	Model 2	Model 2
Other	Model 1	Model 1	Model 1	Model 1	Model 1	Model 1	Model 2	Model 2	Model 2	Model 2	Model 2

Edited CBP	(red)
UCM	(gray)
Model 1 Regression	(yellow)
Model 2 Regression	(green)

Attachment A: Edit Rules

Below are the calculations and edit rules in place to prepare the CBP data for estimates:

1. Identify and remove duplicate records.

2. Identify and remove records that are outside the reporting period.

3. Exclude exports originating from U.S. territories (e.g., Puerto Rico and U.S. Virgin Islands).

4. Map Schedule B/Harmonized Tariff Schedule codes to WPSR products.

5. Keep only products that are within scope of WPSR products.

6. Convert units of quantity not already in barrels (e.g., kg, liters, and tons) to barrels.

7. Based on value (in dollars) divided by quantity of commodity exported, calculate a unit price (dollars/barrel).

8. Compare the unit price with spot prices to determine outliers based on an acceptable range around spot prices.

9. Flag extreme values. Then use the following procedure:

    a. Recalculate weight of the commodity exported based on quantity (in barrels) and a conversion factor.

    b. Compare the recalculated weight to alternative measures of quantity such as shipment or commodity weight that are reported in the record.

    c. If the recalculated weight is within an acceptable range around shipment or commodity weight, then use the original reported quantity in barrels.

    d. If the recalculated weight is not within an acceptable range around the shipment or commodity weight, then convert the shipment or commodity weight to quantity in barrels using a conversion factor and use this as recalculated quantity.