EIA-914 Monthly Crude Oil and Lease Condensate, and Natural Gas Production Report Methodology

November 2018

Table of Contents

Tables

EIA-914 Monthly Crude Oil and Lease Condensate, and Natural Gas Production Report Methodology

Executive summary

The U.S. Energy Information Administration (EIA) currently collects survey data directly from crude oil and natural gas producers in 15 states,[1] the federal Gulf of Mexico (GOM), and Other States,[2] and uses the reported data to estimate total production for these areas, and the United States. This methodology applies to estimates for natural gas and crude oil production beginning with January 2015, in the 15 states, and the federal GOM.

EIA continues to rely on data from state and other federal agencies to estimate crude oil production for the states within the Other States group, which are not separately reported on the EIA-914, using the Average Lagged Ratio (ALR) method.[3] The ALR method applies to crude oil only (including lease condensate) for all states through December 2014, and only to the states included in Other States thereafter.

EIA estimates monthly production for all individually reported states, for both oil and natural gas, by determining the relationship between sampled operators and all operators in the DrillingInfo (DI) data set, a third-party vendor of well-level data collected by state agencies. This relationship is modeled using a Weighted Least Squares (WLS) linear regression. Other methods are used for some states/areas as indicated in the table below.

Percent of U.S. Total Production by Estimation Method:

	Weighted Least Squares	Direct State Report (mostly Alaska)	Average Lagged Ratio (Crude Oil & Condensate only)	Annual Ratio (Natural Gas only)
Crude Oil & Condensate	94%	5%	1%	N/A
Natural Gas Gross Withdrawals	89%	10%	N/A	1%

---

[1] The states directly sampled by Form EIA-914 are Arkansas, California, Colorado, Kansas, Louisiana, Montana, New Mexico, North Dakota, Ohio, Oklahoma, Pennsylvania, Texas, Utah, West Virginia, and Wyoming, and the federal Gulf of Mexico.

[2] The states not directly sampled and included in Other States are Alabama, Arizona, Florida, Illinois, Indiana, Kentucky, Maryland, Michigan, Mississippi, Missouri, Nebraska, Nevada, New York, Oregon, South Dakota, Tennessee, and Virginia, and federal Pacific Offshore.

[3] The ALR methodology used for EIA crude oil production estimation is available at http://www.eia.gov/petroleum/supply/monthly/pdf/crudemeth.pdf, in the "Average lagged-ratio (ALR) estimation" section.

Background

The EIA-914 began collecting natural gas production data in 2005 from Texas, Oklahoma, Louisiana, New Mexico, Wyoming, the federal Gulf of Mexico, and Other States (as a group) excluding Alaska. In 2010, EIA updated and improved the sampling and estimation methodologies. In 2015, EIA improved the EIA-914 form to begin collecting crude oil and lease condensate production,[4] oil and lease condensate sales volumes by API gravity categories,[5] and expanded the collection to 16 individual states/areas adding Arkansas, California, Colorado, Kansas, Montana, North Dakota, Ohio, Pennsylvania, Utah, and West Virginia, while the Other States category was reduced in coverage from 28 states/areas to 18 states/areas.

There are two parts to the EIA-914 process: the sampling and estimation, each described separately below.

Sampling methodology

The EIA-914 report collects crude oil and lease condensate (combined), and natural gas production volume data on a monthly basis by state from a sample of well operators (oil and gas companies). Under the Federal Energy Administration Act of 1974 (Public Law 93-275), operators are required to submit the EIA-914 report. Hereafter crude oil and lease condensate are referred to as oil, and natural gas as gas. In addition to oil production volumes, oil sales volumes for nine API gravity categories are also collected by state/area.

Two samples, one for oil and one for gas, are drawn each month for each state/area (including Other States). The total sample consists of approximately 375 operators out of roughly 15,000 oil and gas operators in the United States. The sampled group of companies can change by as much as a dozen, but usually by only four or five, each month by adding and dropping companies, and accounting for company acquisitions and sales. Monthly sampling in this manner keeps the sample current and avoids a major change in the sample caused by less frequent updating while minimizing sample turnover.

Data preparation

A file containing each company's latest oil and gas production by state is prepared every month for use in the monthly sample selection process. The latest available DrillingInfo (DI) monthly production data are used to create this file (DI is a commercial vendor of state oil and gas production data). DI acquires well or lease level data from state regulatory agencies, then places these data in a database format, which it sells to third parties, such as EIA. An updated DI database is acquired by EIA every month. The DI database is used for both the sampling and the estimation processes. Data for four of the smaller producing states are not available from DI. For Illinois, Indiana, Kentucky, and Tennessee, annual production data from the EIA-23 survey (Annual Survey of Domestic Oil and Gas Reserves) are used to supplement the state data from DI. Hereafter, references to state data from DI in this document include supplemental data from the EIA-23 report for these four states.

Cutoff sample

The oil and gas production cutoff rates for each state/area are determined once per year and then used

---

[4] Although the survey specifies that respondent companies report their production volumes, some companies are only able to report their crude oil and lease condensate sales volumes. EIA expects that the deviation between a company's production and sales will not be a large amount and that sales typically are a reasonable proxy for production.

[5] During the public comment period for the expanded form, some companies reported they can more easily collect and report API gravity information by production rather than sales. As a result, some respondent companies are reporting API gravity for their production volumes rather than their sales volumes.

every month until the next year. The cutoff rates are designed to yield sample coverage of at least 85 percent of the total oil and gas production of each state. Although operators are selected based on their oil or gas production in a particular state/area, all operators added because of their oil production have an additional selection criterion of producing at least 500 barrels per day in at least one state/area.[6] This constraint limits the number of very small operators in the sample. The application of the 500 barrel per day minimum roughly reduces the sample size by one-half, but also yields less than the 85 percent target coverage for oil in a few states. The loss of some state coverage is an acceptable compromise to dramatically reduce the reporting burden on the industry. The sample is based on coverage of 85 percent production volume to guard against model failure, and testing has shown that this sample methodology yields reasonable relative standard errors.[7]

Adding and dropping companies

Each operator's recent monthly oil and gas production data are compared to the sampling cutoff rates for oil and gas in each state (according to the sampling criteria stated above) to determine if a company is in the sample each month. If a company's production is below the cutoff for six consecutive months in every area where they operate they can be dropped from the sample. Companies meeting the criteria to be dropped from the sample are contacted to confirm that their continued production will remain below all cutoff rates in all states for the foreseeable future. Companies providing confirmation are dropped from the sample. Likewise, companies meeting the criteria to be added to the sample are notified of their selection and given instructions on how to proceed with monthly reporting. Companies with production in DI above the cutoff for four consecutive months in any state for oil or gas are added to the sample.

Other ways companies are added or dropped

Mergers and acquisitions, or buying and selling properties can cause a company's production level to move above or below a sample cutoff value. EIA accommodates the larger company changes in the sample as soon as possible after they occur. Large events in terms of overall production, or share of production in one or more states/areas, that involve two companies in the sample or a company currently in the sample and another that is not in the sample are the most important to quickly accommodate in order to minimize estimation errors (see the section "Potential Sources of Errors"). These larger events are usually in news reports,   newsletters, press releases, industry trade journals, and other media outlets. Both companies involved in the transaction may be contacted to assure continued full accounting of production without potential double counting. Events involving small volumes of production are less important, but the companies involved may still be contacted as resources and time allow. Mergers and sales involving only companies in the nonsampled group do not affect the estimation of production. Most smaller mergers and property sales outside the sample are unknown. Information on mergers and property acquisitions is also requested on the EIA-914 form.

Other States group

The Other States group of states/areas accounts for roughly 3 percent of Lower 48 States oil and gas production and includes 17 states and the federal Pacific Offshore . Some of these states are extremely small producers. Other States also includes four states (Illinois, Indiana, Kentucky, and Tennessee) that

---

[6] Many included operators do produce less than 500 barrels per day of oil in at least one state/area, but, if they are selected because of their oil production, they produce at least 500 barrels per day in the state/area for which they are included in the sample. Operators selected because of their natural gas production may produce less than 500 barrels per day of oil in all the states/areas in which they operate, but this does not eliminate them from the sample.

[7] See "Model failure" and "Sampling error" topics in the "Potential Sources of Errors" section below for more details.

are not in the DI database. Some of the 17 states do not require any production reporting for regulatory purposes, or only require annual reporting to the state. The largest producing states within Other States typically dominate the sample and the smallest producing states may be underrepresented by the companies sampled because of their production in Other States. Also, sampling is affected by the absence of a precise and complete list of all producers and their production in Other States. These circumstances make it difficult to quantify sample coverage for Other States.

Estimation methodology

Weighted Least Squares (WLS) is used to estimate oil and gas production for all of the individually reported states/areas in the Monthly Crude Oil and Lease Condensate, and Natural Gas Production Report (EIA-914). The particular model used is a single regressor linear model with the weights equal to the inverse of the regressor, i.e. the Classical Ratio Estimator (CRE).[8] Past work has shown this to be a robust and 'natural' estimator with a wide variety of applications. The model is structured as:

Equation 1

$$y_i = \beta x_i + e_i$$

Where 'y_i' is the survey reported production rate of operator 'i', 'x_i' is the average production rate from the DI dataset of operator 'i' over a six month time period, and e_i is a normal random variable with mean 0 and variance $x_i \sigma^2$. The particular six month period used to determine 'x_i' is described below in the section "Lag Times." If operators i=1 through i=n are sampled, and operators $i = n + 1$ through $i = N$ are unsampled, then the WLS estimate of β is:

$$\beta = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i}$$

And the estimate of the total production for a state is equal to:

$$\hat{T} = \frac{\sum_{i=1}^{N} x_i \sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i}$$

The difference between this estimate of the total and the true total has expected value 0 and variance equal to:

Equation 2[9]

$$V(T - \hat{T}) = \sigma^2 \left(\sum_{i=n+1}^{N} x_i + \frac{\left(\sum_{i=n+1}^{N} x_i\right)^2}{\sum_{i=1}^{N} x_i} \right)$$

The Standard Error described in the section "Potential Sources of Errors" is the square root of this variance. The estimator for σ^2 is:

---

[8] For a review, see "The Classical Ratio Estimator" by James R. Knaub, published on InterStat, 2005.
[9] Found in "Projected Variance for the Model-Based Classical Ratio Estimator" by James R. Knaub, published on InterStat.

U.S. Energy Information Administration | EIA-914 Monthly Crude Oil and Lease Condensate, and Natural Gas Production Report   Methodology

4

$$\widehat{\sigma^2} = \frac{1}{n-1} \sum_{i=1}^{n} \frac{(y_i - \widehat{y_i})^2}{x_i}$$

Each month an estimate for the current month and the previous two months is generated using each month's sample reported volumes and the latest DI database, revising the previous two month estimates.

Lag times

As described above, the 'x_i' used in the estimation routine is a six month average production rate from the DI dataset. The DI dataset contains all the data that DI has been able to acquire from the states and process. The data are not released by the different states on the same schedule. Some states report their data faster than others, and similarly some operators report their data faster. In addition, revisions to the data are common for recently submitted data. As a result of these considerations, EIA calculates a 'most recently complete month' for each state in the DI dataset. This number is calculated by taking the number of months between the report month and the month at which the state total production volume is anticipated to be within half a percent of its final value. The 'x_i' are based on the most recent month with complete data and the five months previous to it. Typical lags are given in Table 1, but may be adjusted occasionally depending on changing circumstances. For states that only report annually it is necessary to adjust their lag each month.

Table 1. State lags, January 2018

State	Oil Lags (number of months)	Gas Lags (number of months)
Arkansas	4	3
California	5	6
Colorado	13	13
Federal Gulf of Mexico	4	4
Kansas	5	5
Louisiana	5	7
Montana	5	6
New Mexico	6	7
North Dakota	6	6
Ohio	13	13
Oklahoma	18	18
Pennsylvania	13	13
Texas	10	7
Utah	4	4
West Virginia	13	13
Wyoming	4	4

Ohio, Pennsylvania, and West Virginia report their data for conventional wells annually. For these states, the lag is set to place the most recently complete month on December of the most recently reported year.

U.S. Energy Information Administration | EIA-914 Monthly Crude Oil and Lease Condensate, and Natural Gas Production Report Methodology

5

Other States ratio

For Natural Gas production estimates for the Other States, the ratio of total state reported data to the EIA-914 reported data is calculated   based on calendar year volumes and is multiplied by the current month's EIA-914 reported volume to   determine the estimate. State production data for the Other States are collected directly from the states to construct the calendar year total,[10] and also may be incomplete in recent months just as the state data from DI often are for individually sampled states. As mentioned earlier, complete production for some of the Other States may not be available. Therefore, the estimate for the Other States may be lower than an estimate based on complete production data for Other States.

Because EIA publishes oil production estimates for all states individually, including states in the Other States category, EIA uses its alternative oil production estimation methodology that is based on the EIA-182 (Domestic Crude Oil First Purchase Report) data and lagged state reported data as the   official Other States oil production estimation method.[11] (This is the ALR method[12] mentioned earlier.)   The gas production estimate procedure described in the paragraph above is also used to generate oil estimates for the Other States group to compare to a sum of the ALR estimates. The Other States category accounts for roughly 1.5 percent of the total Lower 48 states production for both oil and gas.

Estimates of natural gas lease production and oil sales by API gravity

The EIA-914 collects production data for two separate gas volumes: gas gross withdrawals and lease gas production. Natural gas gross withdrawals are generally gas production measured after lease separation. Natural gas lease production is generally gas that comes off the lease or gas sales. The difference between the two accounts for gas that is used on the lease for fuel, vented and flared, injected, and the removal of non-hydrocarbon gases. The WLS model is used to estimate gross withdrawals, and the ratio of estimated total to   sampled gross withdrawals is multiplied by sampled lease production to estimate lease production.

The EIA-914 collects production data for two separate oil volumes: oil production and oil sales by API gravity category.[13] In order to estimate state production by API category, it is assumed that state production has the same proportional distribution of API categories as the reported API category volumes. That is, the reported distribution of API gravity volumes (including the unknown category) is applied to the estimated total production volume to estimate the total API gravity volumes. Although EIA collects API gravity for 10 categories, including "unknown," the categories are collapsed into four categories for the reporting of state-level API gravity estimates. Many categories at the state level have too few respondents to be reported separately because of confidentiality rules. Suppression of these under-reported categories was accomplished by collapsing categories. For example, the 40.1-45.0 category and the 45.1-50.0 category were combined to create a 40.1-50.0 category. Further, the state-level volumes reported in the "unknown" category are allocated to the individual categories rather than reported separately.

---

[10] EIA analysts and contractors visit state websites and correspond with state officials to acquire the most recently available production data of those states.

[11] An asymmetry in the estimation of gas production for Other States exists because no EIA survey collects state-level gas sales volumes aside from Form EIA-914. Alternatively Form EIA-182 collects oil sales volumes, which may be used to estimate Other   State oil production.

[12] The ALR methodology can be found at http://www.eia.gov/petroleum/supply/monthly/pdf/crudemeth.pdf.

[13] Respondent companies are allowed to report production or sales volumes by API gravity category on Form EIA-914. However, it is expected that sales are predominately reported in the API gravity portion of the survey (Part 4).

U.S. Energy Information Administration | EIA-914 Monthly Crude Oil and Lease Condensate, and Natural Gas Production Report Methodology

6

Potential sources of errors

Alignment of survey and DI datasets

Unknown, deficient reporting of, or incorrectly handled mergers and property sales are likely the largest cause of errors. These events occur every month, making the alignment of the survey and DI datasets a continuous and critical task. The company production in the historical DI dataset must be matched to the reported sample data every month. If an operator in one dataset corresponds to different properties than in the other, the modeled relationship between 'y_i' and 'x_i' is invalid. The unknown or missed mergers and sales are usually small and do not contribute to significant large errors, but it is still possible to miss a larger event. In addition to mergers, sales, and acquisitions, the following are examples of items that can contribute to errors in the alignment of survey data with state data from DI: name changes, multiple name spellings, companies that report under multiple names, and lags between the time of a merger and the time of its appearance in the DI dataset. In practice, these misalignments result in a larger estimated RSE (see below), and so are partially accounted for under 'sampling error.'

Frame coverage

If the sampling frame (state data from DI) does not include all of the operators in a state, then part of the population will be missing in both sampling and estimation, and estimates will be low. This is frame coverage error. In some states, such as the previously mentioned Illinois, Indiana, Kentucky, and Tennessee, the state data from DI are known to be missing or incomplete, and so another method is used to estimate these states, described in the Estimation Methodology for Other States above. In states with long time lags (in which the most recent state data from DI may be more than a year old) frame coverage becomes a concern, as well, since new operators may exist that are not in the frame. EIA attempts to identify and resolve such errors through the use of additional data sources such as state reports, industry news letters, and trade journal publications.

Reporting errors

Reporting errors such as incorrect units or incomplete or otherwise incorrect accounting can occur on the EIA-914. The survey instrument itself was carefully developed and includes detailed instructions for filing data, subject to a common set of definitions similar to those already used by the industry. Editing software has been developed to detect different kinds of probable reporting errors and to flag them for resolution by analysts, either through confirmation of the data by the respondent or through submission of amendments to the filed data (see the section "Data Quality Control Checks below").

Model failure

Both experience and experiment show that the model in Equation 1 holds very well. However, very dynamic events can cause the population to behave abnormally and, particularly when combined with long time lags in the state data from DI, the model may deviate from reality. For example, the rapid development of the Haynesville shale in Louisiana caused a change in the State production trend that, in turn, caused the method to overestimate for a short time late in the 2000s. EIA attempts to identify regions that may be likely to behave in such a way, and if the deviations were to become substantial a change to the estimation methodology, such as a stratification, would have to be made. As an additional safeguard against model failure, EIA targets sample coverage of 85 percent in all states, so that the effects of model failure would be minimized.

Sampling error

Sampling error may be defined as the difference between the estimates obtained from a sample and the results that would have been obtained from a complete enumeration of the frame population. The standard error statistic is a measure of this sampling error, and is the square root of the quantity given in

Equation 2. When presented as a percentage of the estimated total, it is called the relative standard error (RSE). The sampling methodology described in this report has yielded RSE's for the first three months of 2018 gas and oil collection shown in tables 2 and 3 below. RSE's for all states collected individually on the EIA-914 report are published monthly.

Table 2. Natural Gas RSEs for first three months of estimates, 2018

State	January	February	March
Arkansas	0.24%	0.23%	0.26%
California	0.81%	0.75%	0.80%
Colorado	0.52%	0.56%	0.55%
Federal Gulf of Mexico	1.19%	1.24%	1.31%
Kansas	0.69%	0.68%	0.71%
Louisiana	0.72%	0.72%	0.75%
Montana	0.55%	0.58%	0.56%
New Mexico	0.70%	0.78%	0.78%
North Dakota	0.64%	0.58%	0.60%
Ohio	0.83%	1.19%	1.32%
Oklahoma	0.67%	0.72%	0.72%
Other States	-	-	-
Pennsylvania	0.29%	0.26%	0.32%
Texas	0.41%	0.46%	0.47%
Utah	0.62%	0.58	0.65%
West Virginia	1.12%	1.15%	1.44%
Wyoming	0.21%	0.26%	0.30%

Table 3. Crude Oil RSEs for first three months of estimates, 2018

State	January	February	March
Arkansas	1.20%	1.77%	1.69%
California	0.27%	0.30%	0.22%
Colorado	0.96%	0.97%	1.15%
Federal Gulf of Mexico	0.62%	0.68%	0.65%
Kansas	0.96%	1.13%	1.18%
Louisiana	0.87%	0.88%	0.82%
Montana	0.87%	0.73%	1.01%
New Mexico	0.75%	0.71%	0.79%
North Dakota	0.76%	0.72%	0.67%
Ohio	3.15%	3.81%	4.64%
Oklahoma	1.43%	1.36%	1.53%
Other States	-	-	-
Pennsylvania	1.86%	1.55%	1.45%
Texas	0.63%	0.65%	0.64%
Utah	0.52%	0.41%	0.68%
West Virginia	1.31%	1.77%	3.04%
Wyoming	0.73%	0.87%	0.79%

Note that RSEs are not calculated for the Other States region, which uses a different estimation methodology.

U.S. Energy Information Administration | EIA-914 Monthly Crude Oil and Lease Condensate, and Natural Gas Production Report Methodology

9

Replacement of estimates with state data

Given the revision schedule of the natural gas and oil production estimates (i.e., each month data are released for the current month and revised for the two previous months), once a year the estimates for the two previous years for natural gas data and ten previous years for oil data are replaced with final state data. The timing of the revisions/replacement of estimates with state data is done in conjunction with the publication of the *Natural Gas Annual* (NGA) for natural gas production estimates and the publication of the *Petroleum Supply Annual* (PSA) for oil production estimates. The NGA is normally published each September while the PSA is normally published each August.

Data quality control checks

EIA employs automated systems to identify suspect data submissions. The initial set of checks tests the submitted data. Once the suspect data submissions are identified many additional efforts are made to further identify and correct errors in submitted data. Follow up with respondents via personal communication and research is carried out when a data discrepancy is identified. Reported production data are compared with state production data from DI for each sampled company to ensure the companies are correctly identified and matched. Company reported data may also be compared with their state-reported data to assure correct reporting. Mergers and property acquisitions reported in trade press and other media are tracked and companies are routinely contacted to ensure complete accounting without double counting or under reporting production. Companies are also contacted if an unusual or large change in a company's production is reported without explanation. Production estimates are continuously compared to state reported production to check for significant differences.

Outlier process

Some responses from respondent companies are far outside the anticipated value. For example, a company may have extreme growth or decline in recent months. Evaluations of these atypical responses are made and, if, in the opinion of the data analysts and the survey manager, these responses are excessively affecting the survey results, then they may be omitted from the normal estimation process, but added-in later in the process.

Imputation

Company nonresponse occurs from time to time, and occasionally, a company may report a value that is out of the expected range. Omitted submissions, incomplete submissions, and unexplained submitted values that considerably deviate from historically submitted values may require an imputation. Typically, efforts to collect accurate data from nonrespondent companies continue until they submit their data. However, if missing data or suspicious data cannot be acquired from the company or satisfactorily explained by the company, data are imputed at the time of estimation. For natural gas production and sales, and oil production, imputed values are derived by using a three-month average of the most recently available data.

If the company has been a respondent company for at least three months, then the company's historic EIA-914 values are used. In the absence of sufficient 914 company data, data from DI are used to calculate the imputed values.

Oil sales volumes (oil volumes by API gravity category) may be imputed if the data are missing, incomplete, or the distribution of submitted data across the API gravity categories deviates considerably from those historically submitted by the respondent company. The imputed volumes are calculated using the weighted-average distribution from the previous three months. Every effort is made to obtain accurate data from the respondent before using an imputed value.

Suppression

EIA employs statistical disclosure limitation techniques to preserve the confidentiality of the information collected on the EIA-914. The p-percent rule is applied to the statistical aggregates.

Coverage and Response Rate

Coverage and Response Rates are published to give some insight to the quality of the data collection and processing. Coverage is determined by dividing the reported values by the estimated values for each state. Response rate is calculated as the volume reported by companies responding to the survey divided by the volume of reporting companies plus the expected volume for non-reporting companies.

U.S. Energy Information Administration | EIA-914 Monthly Crude Oil and Lease Condensate, and Natural Gas Production Report Methodology

11