

Appendix C

Statistical Considerations

The monthly sales (volume and price) and monthly deliveries (volume) of natural gas to residential, commercial, and industrial consumers presented in this report by State are estimated from data reported on the Form EIA-857, "Monthly Report of Natural Gas Purchases and Deliveries to Consumers." Monthly prices in select States are supplemented with data from the Form EIA-910, "Monthly Natural Gas Marketer Survey." (See Appendix B, "Data Sources," for a description of these Forms.) These estimations must be made from the reported data since the Form EIA-857 is a sample survey. A description of the sample design and the estimation procedures is given below.

Sample Design

The Form EIA-857 is a monthly sample survey of companies delivering natural gas to consumers. It includes inter- and intrastate pipeline companies, and producers, as well as local distribution companies. The survey provides data that are used each month to estimate the volume of natural gas delivered and the price for onsystem sales of natural gas by State to three consumer sectors—residential, commercial, and industrial. Monthly deliveries and prices of natural gas to the electric power sector through 2007 are reported on the Form EIA-906, "Power Plant Report"; Form EIA-920, "Combined Heat and Power Plant Report"; Form FERC-423, "Monthly Report of Costs and Quality of Fuels for Electric Plants"; and Form EIA-423, "Monthly Cost and Quality of Fuels for Electric Plant Report." Beginning with 2008 data, the Form EIA-923, "Power Plant Operations Report," replaced these sources.

Sample Universe. The sample in use for 2008 was selected from a universe of 1,517 companies. These companies were respondents to the Form EIA-176, "Annual Report of Natural and Supplemental Gas Supply and Disposition," for reporting year 2006, who reported sales or deliveries to consumers in the residential, commercial, or industrial sectors. (See Appendix B for a description of the Form EIA-176.)

Sampling Plan. The goal was a sample that would provide estimates of monthly natural gas consumption by the three consuming sectors within each State and the District of Columbia. A stratified

sample using a single stage and systematic selection with probability proportional to size was designed.

The measure of size was the volume of natural gas physically delivered in the State to the three consuming sectors by the company in 2006. There were two strata—companies selected with certainty and companies selected under the systematic probability proportional to size design.

Initial calculations showed that an approximately 25 percent sample of companies would yield reasonably accurate estimates. The sample was selected independently in each State, resulting in a national total of 382 respondent companies.

Certainty Stratum. Since estimates were needed for each of the 50 States and the District of Columbia, the strata were established independently within each State. In 16 States and the District of Columbia where sampling was not feasible due to small numbers of companies and/or small volumes of gas deliveries, all companies were selected. The 16 States were: Alaska, Connecticut, Delaware, Hawaii, Idaho, Maine, New Hampshire, New Jersey, Nevada, North Dakota, Oregon, Rhode Island, South Dakota, Utah, Vermont, and Washington.

For each of the remaining States, the total volumes of industrial sales and deliveries and of the combined residential/commercial sales and deliveries were determined. Companies with natural gas deliveries to the industrial sector or to the combined residential/commercial sector above a certain level were selected with certainty. Since a few large companies often account for most of the natural gas delivered within a State, this ensures those companies' inclusion in the sample. The formula for determining certainty was applied independently in the two consumer sectors—the industrial and the combined residential/commercial. These selected companies, together with the companies in the jurisdictions discussed where sampling was not feasible, formed the certainty stratum.

All companies with natural gas deliveries in sector j greater than the cut-off value (C_j) were included in the certainty stratum. The formula for C_j was:

$$C_j = \frac{X_j}{2n} \quad (1)$$

where:

$C_{.j}$ = cutoff value for consumer sector j,

n = target sample size to be selected for the State, 25 percent of the companies in the State,

X_{ij} = the annual volume of natural gas deliveries by company i to customers in consumer sector j,

$X_{i.}$ = the sum within State of annual gas volumes for company i,

$X_{.j}$ = the sum within State of annual gas volumes in consumer sector j,

$X_{..}$ = the sum within State of annual gas volumes in all consumer sectors.

Noncertainty Stratum. All other companies formed the noncertainty stratum. They were systematically sampled with probability proportional to size. The measure of size for each company was the total volume of gas sales to all consumer sectors ($X_{i.}$). The number of companies to be selected from the noncertainty stratum was calculated for each State, with a minimum of 2.

The formula for selecting the number of noncertainty stratum companies was:

$$m = n \frac{X_2}{X_{..}} \quad (2)$$

where:

m = the sample size for the noncertainty stratum within a State,

X_2 = the sum within State of the $X_{i.}$ for all companies in the noncertainty stratum.

Companies were listed in ascending order according to their measure of size and then a cumulative measure of size in the stratum was calculated for each company. The cumulative measure of size was the sum of the measures of size for that company and all preceding companies on the list. An interval of width I for selecting the companies systematically was calculated.

A uniform random number R was selected between zero and $\left(I = \frac{X_2}{m}\right)I$. The first sampled company was

the first company on the list to have a cumulative measure of size greater than R. The second company selected was the first company on the list to have a cumulative measure of size greater than R + I. R + I

was increased again by I to determine the third company to be selected. This procedure was repeated until the entire sample was drawn.

Subgroups. In five States, the noncertainty stratum was divided into subgroups to ensure that gas in each consumer sector could be estimated. The systematic sample with probability proportional to size design described above was applied independently in each subgroup. The methods for determining the subgroup sample size and calculating the subgroup interval for sample selection were the same as the methods described above for the noncertainty stratum, except that X_2 was the sum within State of the $X_{i.}$ for only those companies in the subgroup.

These subgroups were defined only for the purpose of sample selection. They are:

Kansas, Louisiana and Texas: companies delivering gas only to industrial consumers and all other companies.

Pennsylvania: companies having some deliveries of gas to industrial consumers and all other companies.

Estimation Procedures

Estimates of Volumes. A ratio estimator is applied to the volumes reported in each State by the sampled companies to estimate the total gas sales and deliveries for the State. Ratio estimators are calculated for each consumer sector – residential, commercial, and industrial – in each State where companies are sampled. The following annual data are taken from the most recent submissions of Form EIA-176:

The formula for calculating the ratio estimator (E_{vj}) for the volume of gas in consumer sector j is:

$$E_{vj} = \frac{\gamma_{.j}}{\gamma'_{.j}} \quad (3)$$

where:

γ_j = the sum within State of annual gas volumes in consumer sector j for all companies,

$\gamma'_{.j}$ = the sum within State of annual gas volumes in consumer sector j for those companies in the sample.

The ratio estimator is applied as follows:

$$V_{vj} = Y_{.j} \times E_{vj} \quad (4)$$

where:

V_{vj} = the State estimate of monthly gas volumes in consumer sector j ,

$Y_{.j}$ = the sum within State of reported monthly gas volumes in consumer sector j .

Computation of Natural Gas Prices. The natural gas volumes that are included in the computation of prices represent only those volumes associated with natural gas sales by natural gas companies except as explained below.

The price of natural gas for a State within a sector is calculated as follows:

$$P_j = \frac{R_j}{V_j} \quad (5)$$

where:

P_j = the average price for gas sales within the State in consumer sector j ,

R_j = the reported revenue from natural gas sales within the State in consumer sector j ,

V_j = the reported volume of natural gas sales within the State in consumer sector j .

All average prices are weighted by their corresponding sales volume estimates when national average prices are computed.

The monthly average prices of natural gas to residential and commercial consumers in Georgia, Maryland, New York, Ohio, and Pennsylvania are monthly average prices of natural gas, based on total sales (sales by local distribution companies and natural gas marketers). Beginning in January 2005, the EIA-910 is collected in the States of Florida, Illinois, Massachusetts, Michigan, New Jersey, Virginia, West Virginia, and the District of Columbia as well. Residential and commercial prices represent total deliveries of gas sold to customers in those States as the quality of data collected on the EIA-910 becomes acceptable. Volumes of gas delivered for the account of others to these consumer sectors are not included in the State or national average prices except in these States.

The price of natural gas in the residential and commercial sectors where EIA-910 data are used is calculated as follows:

$$P_c = \left[\left(\frac{R_s}{V_s} \right) * \left(\frac{V_s}{V_s + V_t} \right) \right] + \left[\left(\frac{Rm_s}{Vm_s} \right) * \left(\frac{V_t}{V_s + V_t} \right) \right] \quad (6)$$

P_c = the combined average price for gas sales by local distribution companies and marketers within the State in sector s (residential or commercial),

R_s = the reported revenue from natural gas sales by local distribution companies within the State in s (residential or commercial),

V_s = the reported volume of natural gas sales by local distribution companies within the State in s (residential or commercial),

V_t = the reported volume of natural gas transported by local distribution companies for marketers within the State in s (residential or commercial),

Rm_s = the reported revenue from natural gas sales by marketers within the State in s (residential or commercial),

Vm_s = the reported volume of natural gas sales by a marketer within the State in s (residential or commercial).

Table 22 shows the percent of the total State volume that represents volumes from natural gas sales to the commercial and industrial sectors. This table may be helpful in evaluating commercial and industrial price data. All natural gas prices to the residential sector represent onsystem sales volumes only except in States where EIA-910 data are used.

See the section on consumer price calculations in this Appendix for further price information.

Estimation for Nonrespondents and Edit Failures. A volume for each delivered and transported consumer category is imputed for companies that fail to respond in time for inclusion in the published estimates (unit nonresponse) or for which reported volumes have failed the edit and not been confirmed or corrected (item nonresponse). In both instances, the imputation is carried out in the same way.

The imputed volumes are derived through a two-part procedure:

1) Prediction of monthly volumes for the total commercial, industrial, and residential sectors within Census Division. Census Division refers to the nine divisions into which the U.S. Bureau of the Census groups the 50 States and the District of Columbia for reporting and analysis purposes. Alaska and Hawaii, members of the Pacific Division, are handled separately from other States in that division.

Sector volume includes both sales and transportation components.

For the commercial and residential sectors, the predicted division volume for a month depends on the heating degree days reported by the National Oceanic and Atmospheric Administration (NOAA) for that month within the Census Division. It also depends on an adjustment for the particular month being predicted.

The formula for the predicted division volume in the commercial and residential sectors is

$$\hat{Y}_{jt} = b_0 + (h_j * H_{jt}) + \sum_{t=1}^{12} (d_t * D_t) \quad (7)$$

where:

\hat{Y}_{jt} = the predicted j^{th} division volume in month t ,

b_0 = an intercept term,

h_j = the coefficient for the j^{th} Census Division heating degree days,

H_{jt} = the j^{th} Census Division heating degree days for the t^{th} month being imputed,

d_t = the coefficient for the t^{th} monthly dummy variable D_t , and,

D_t = a dummy variable with value = 1 if the t^{th} month is imputed and 0 otherwise—with one exception. In December, all the dummy variables are equal to 0 and there is no coefficient.

For the industrial sector, the predicted division volume for a month depends on the prior month's division volume. The formula for the predicted division volume in the industrial sector is

$$\hat{Y}_{jt} = b_0 + (b_j * X_{j,t-1}) \quad (8)$$

where:

\hat{Y}_{jt} = the predicted total industrial sector volume for the j^{th} Census Division in month t ,

b_0 = an intercept term,

b_j = the coefficient for the industrial sector volume in the j^{th} Census Division, and,

$X_{j,t-1}$ = the total industrial sector volume in the j^{th} Census Division for the month prior to t .

The coefficients are estimated via ordinary least squares multiple linear regression. The source is a database of monthly sector volumes for the 5 years ending December 31 of the immediately prior calendar year. Coefficient estimation is restricted to companies reporting continuously during the 5 years.

(2) Allocating the monthly sector volume for a particular respondent based on the respondent's share of that sector volume in the latest Form EIA-176 survey.

Once the predicted division volume for a sector is obtained, it is multiplied by an allocation factor to obtain the imputed sector volume for a respondent. The allocation factor is the ratio of that respondent's sector volume to the total of all such sector volumes as reported in the latest Form EIA-176 survey.

The formula for allocating is

$$I_{jtk} = \hat{Y}_{jt} * \left(\frac{v_{jk}}{V_j} \right) \quad (9)$$

where:

I_{jtk} = the imputed monthly sector volume for the k^{th} nonresponse case in Census Division j for month t ,

\hat{Y}_{jt} = the predicted monthly sector volume in Census Division j for month t ,

v_{jk} = nonrespondent k 's reported sector volume for Census Division j in the latest Form EIA-176 survey, and,

V_j = the total reported sector volume for all respondents for Census Division j in the latest Form EIA-176 survey.

Estimation of Revenue. The company's previous month's sector-specific price is multiplied by the corresponding sales volume to impute revenue for that sector.

Final Revisions

Adjusting Monthly Data to Annual Data. After the annual data reported on the Form EIA-176 have been submitted, edited, and prepared for publication in the *Natural Gas Annual*, revisions are made to monthly data. The revisions are made to the volumes and prices of natural gas delivered to consumers that have appeared in the *Natural Gas Monthly (NGM)* to match them to the annual values appearing in the *Natural Gas Annual*. The revised monthly estimates allocate the difference between the sum of monthly estimates and the annual reports, according to the distribution of the estimated values across the months.

Before the final revisions are made, changes or additions to submitted data received after publication of the monthly estimate and not sufficiently large to require a revision to be published in the *NGM*, are used to derive an updated estimate of monthly consumption and revenues for each State's residential, commercial, or industrial natural gas consumption.

For each State, two numbers are revised, the estimated consumption and the estimated price per thousand cubic feet.

The formula for revising the estimated consumption is:

$$V_{jm}^* = V_{jm} + \left[(V_{ja} - V'_{jm}) \left(\frac{V_{jm}}{V'_{jm}} \right) \right] \quad (10)$$

where:

V_{jm}^* = the final volume estimate for month m in consumer sector j,

V_{jm} = the estimated volume for month m in consumer sector j,

V_{ja} = the volume for the year reported on Form EIA-176,

V'_{jm} = the annual sum of estimated monthly volumes.

The price is calculated as described above in the Estimation Procedures section, using the final revised consumption estimate and a revised revenue estimate.

The formula for revising the estimated revenue is:

$$R_{jm}^* = R_{jm} + \left[(R_{ja} - R'_{jm}) \left(\frac{R_{jm}}{R'_{jm}} \right) \right] \quad (11)$$

where:

R_{jm}^* = the final revenue estimate for month m in consumer sector j,

R_{jm} = the estimated revenue for month m in consumer sector j,

R_{ja} = the revenue for the year reported on Form EIA-176,

R'_{jm} = the annual sum of estimated monthly revenues.

Revision of Volumes and Prices for Deliveries to Electric Power Sector. Revisions to monthly deliveries to the electric power sector are published throughout the year as they become available.

Reliability of Monthly Data

The monthly data published in this report are subject to two sources of error -- nonsampling error and sampling error. Nonsampling errors occur in the collection and processing of the data. See the discussion of the Form EIA-857 in Appendix B for a description of nonsampling errors for monthly data.

Sampling error may be defined as the difference between the results obtained from a sample and the results that a complete enumeration would provide. The standard error statistic is a measurement of sampling error.

Standard Errors. A standard error of an estimate is a statistical measure that indicates how the estimate from the sample compares to the result from a complete enumeration. Standard errors are calculated based on statistical theory that refers to all possible samples of the same size and design.

The standard errors for monthly natural gas volume estimates by State are given in Table C1. Ninety-five percent of the time, the volume that would have been obtained from a complete enumeration will lie in the range between the estimated volume minus two standard errors and the estimated volume plus two standard errors.

The standard error of the natural gas volume estimate is the square root of the variance of the estimate. The formula for calculating the variance of the volume estimate is:

$$V(\hat{\gamma}) = \sum_{h=1}^H \left[N_h^2 \frac{\left(1 - \frac{n_h}{N_h} \right)}{n_h(n_h - 1)} \left(\sum_{i=1}^{n_h} (y_i - Tx_j)^2 \right) \right] \quad (12)$$

where:

H = the total number of strata,

N_h = the total number of companies in stratum h,

n_h = the sample size in stratum h,

y_i = the reported monthly volume for company i,

x_i = the reported annual volume for company i,

T = the ratio of the sum of the reported monthly volumes for sample companies to the sum of the reported annual volumes for the sample companies.

Table C1. Standard Error for Natural Gas Deliveries and Price by Consumers, by State, December 2008

State	Volume Million Cubic Feet				Price Dollars per Thousand Cubic Feet		
	Residential	Commercial	Industrial	Total	Residential	Commercial	Industrial
Alabama	461	331	1,781	1,869	NA	NA	NA
Alaska	0	0	0	0	--	--	--
Arizona	2	6	0	6	0.02	0.01	--
Arkansas	4	8	3	10	0.01	0.01	0.03
California	804	129	432	921	0.10	0.11	0.07
Colorado	0	0	0	0	--	--	--
Connecticut	0	0	0	0	--	--	--
Delaware	0	0	0	0	--	--	--
District of Columbia	0	0	0	0	--	--	--
Florida	592	210	4,042	4,090	NA	NA	0.87
Georgia	86	400	1,498	1,553	0.59	NA	NA
Hawaii	0	0	0	0	--	--	--
Idaho	0	0	0	0	--	--	--
Illinois	2,206	227	635	2,307	0.35	0.84	NA
Indiana	563	451	8,646	8,676	0.38	0.66	NA
Iowa	14	43	3,255	3,255	0.13	0.13	0.12
Kansas	489	309	463	742	0.31	0.49	NA
Kentucky	216	232	830	888	0.19	0.17	NA
Louisiana	1,246	295	3,614	3,834	NA	NA	--
Maine	0	0	NA	NA	--	--	--
Maryland	4	5	53	54	0.01	0.02	0.65
Massachusetts	354	87	131	388	0.27	0.25	0.20
Michigan	155	106	208	280	0.02	0.16	0.26
Minnesota	1,198	1,457	1,964	2,723	0.32	0.41	0.99
Mississippi	532	764	670	1,147	NA	0.30	NA
Missouri	253	140	421	511	0.10	0.14	NA
Montana	20	26	0	33	0.04	0.09	--
Nebraska	248	NA	152	NA	0.55	NA	NA
Nevada	0	0	0	0	--	--	--
New Hampshire	0	0	0	0	--	--	--
New Jersey	0	0	0	0	--	--	--
New Mexico	201	27	165	261	0.20	0.25	NA
New York	2,131	846	1,103	2,544	0.13	0.13	NA
North Carolina	139	47	116	187	0.13	0.07	0.40
North Dakota	0	0	0	0	--	--	--
Ohio	1,528	1,218	1,487	2,456	0.30	0.10	0.05
Oklahoma	756	348	683	1,077	0.77	NA	NA
Oregon	0	0	0	0	--	--	--
Pennsylvania	57	25	95	114	0.01	0.01	0.03
Rhode Island	0	NA	0	NA	--	NA	--
South Carolina	38	NA	NA	NA	0.33	0.41	0.96
South Dakota	0	0	0	0	--	--	--
Tennessee	257	251	2,518	2,544	0.18	0.36	NA
Texas	6,289	6,528	4,157	9,973	0.63	NA	0.05
Utah	0	0	0	0	--	--	--
Vermont	0	0	0	0	--	--	--
Virginia	100	35	257	278	0.13	0.16	0.50
Washington	0	0	0	0	--	--	--
West Virginia	40	49	0	63	0.23	0.30	--
Wisconsin	604	786	2,894	3,059	0.24	0.45	0.78
Wyoming	NA	NA	297	NA	0.55	0.98	0.03
Total	7,609	7,025	12,740	16,419	0.15	0.25	0.55

-- Not applicable.

NA Not available.

Source: Energy Information Administration (EIA): Form EIA-857, "Monthly Report of Natural Gas Purchases and Deliveries to Consumers."