Revisions in Natural Gas Monthly Consumption and Price Data, 2004 – 2007

This report summarizes the method in which natural gas consumption data is collected and processed for publication and details the most notable revisions in natural gas consumption data for the period 2004 to 2007. It is intended to assist data users in evaluating the quality of the monthly consumption and price data for residential, commercial, and industrial consumers of natural gas. Questions or comments on the contents of this report should be directed to Joe Conklin at joseph.conklin@eia.doe.gov or 202-586-6664 or to Amy Sweeney at amy.sweeney@eia.doe.gov or 202-586-2627.

The U.S. Energy Information Administration publishes monthly data pertaining to the natural gas industry in the *Natural Gas Monthly* (*NGM*). A sample survey, Form EIA-857, entitled "Monthly Report of Natural Gas Purchases and Deliveries to Consumers," is used to obtain preliminary estimates for volume and price data for natural gas delivered to residential, commercial, and industrial consumers. Preliminary estimates are published in the *NGM* with a 2-month lag. For example, estimates for January 2007 were published in the March 2007 *NGM*.

The primary findings of this report concerning revisions made to the consumption and price estimates from the first preliminary estimates based on the EIA-857 to the final adjusted figures reflecting the annual EIA-176, "Annual Report of Natural and Supplemental Gas Supply and Disposition" data over the period from 2004 to 2007 include:

- At the national level, the mean of monthly revision errors for 2007 is lower than that of 2004.
- The industrial consumption and price series exhibit the largest 4-year (2004-2007) mean revision error.
- Delaware and Florida exhibit the largest 4-year mean revision errors for the residential consumption and price series.
- Alaska and Massachusetts exhibit the largest 4-year mean revision errors for the industrial consumption series as well as the commercial price series.
- New York exhibits one of the largest 4-year revision errors means for the commercial consumption and industrial price series.
- Overall, the lack of a discernible pattern in 4-year and annual mean revision errors suggests that a variety of factors are responsible for the revision errors from 2004 to 2007. These factors include the series' sensitivity to market conditions such as seasonality, the price of substitutes for natural gas, and economic variability, as well as State- and respondent-specific issues.

Survey Design and Statistical Considerations

The monthly sales (volume and price) and monthly deliveries (volume) of natural gas to residential, commercial, and industrial consumers presented in the *NGM* are estimated from data reported on the Form EIA-857. These estimations must be made from the reported data since the Form EIA-857 is a monthly sample survey. The Form EIA-857 includes inter- and intrastate pipeline companies, and local distribution companies.

The first stage of the Form-857 process is to determine who will be included in the sample. The sample is selected from a universe of companies that reported sales or deliveries to consumers in the residential, commercial, or industrial sectors on the Form EIA-176 for the reporting year two years prior. For example, the sample selected for 2009 was from a universe of all the companies that were respondents to the Form EIA-176 for reporting year 2007.

The sample is stratified using single-stage and systematic selection with probability proportional to size. The measure of size is the volume of natural gas physically delivered in the State to the three consuming sectors by the company in the reporting year two years prior. There are two strata—companies selected with certainty and companies selected under the systematic probability proportional to size design. The sample is selected independently in each State.

The formula for determining certainty is applied independently in the two consumer sectors—the industrial and the combined residential/commercial. The certainty stratum includes all companies in States where sampling is not feasible because of small numbers of companies and/or small volumes of gas deliveries. The certainty stratum also includes companies with natural gas deliveries above a certain level. Since a few large companies often account for most of the natural gas delivered within a State, this method ensures those companies' inclusion in the sample.

All other companies form the noncertainty stratum. They are systematically sampled with probability proportional to size. The measure of size for each company is the total volume of gas sales to all consumer sectors. The number of companies to be selected from the noncertainty stratum is calculated for each State, with a minimum of two.

Companies are listed in ascending order according to their measure of size, and a cumulative measure of size in the stratum is calculated for each company. In some States, the noncertainty stratum is divided into subgroups to ensure that natural gas in each consumer sector could be estimated. After the sample has been drawn, separate methods are used to estimate volumes and prices. To estimate the total gas sales and deliveries for the State, a ratio estimator is applied to the volumes reported by the sampled companies in each State. Ratio estimators are calculated for each consumer sector — residential, commercial, and industrial — in each State where companies are sampled. Annual data from the most recent submissions of Form EIA-176 are used in the formula for the ratio estimator.

Each month about half the responses are received by the due date, although response rates by first publication of the relevant month are approximately 95 percent. When a response is extremely late, volumes are imputed. When the company's submission is eventually received, the submitted data are entered into the data system and used for subsequent processing and revisions. 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 derived using the same two-part procedure.

The first part of the derivation of imputed volumes (including both sales and transportation components) involves the prediction of monthly volumes for the total commercial, industrial, and

residential sectors within Census Division. Alaska and Hawaii, members of the Pacific Division, are handled separately from other States in that division.

The second part of the derivation of imputed volumes involves 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.

Monthly data are revised after the annual data reported on the Form EIA-176 have been submitted, edited, and prepared for publication in the *Natural Gas Annual*. Form EIA-176 is mailed to all identified interstate and intrastate natural gas pipeline companies; investor- and municipally-owned natural gas distributors; underground natural gas storage operators; synthetic natural gas plant operators; field, well, or processing plant operators that deliver natural gas directly to consumers, including their own industrial facilities; andcompanies that transport gas across a State border through field or gathering facilities. Each company is required to file if it meets any of the survey specifications. The final nonresponse rates have averaged less than 5 percent in recent years.

Revisions are made to the volumes and prices of natural gas delivered to consumers that have appeared in the *NGM* to conform 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 price is calculated as described above, using the final revised consumption estimate and a revised revenue estimate.

The monthly data published in the *Natural Gas Monthly* are subject to two sources of error –- nonsampling error and sampling error. Nonsampling errors occur in the collection and processing of the data and have not been measured. 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. The standard error statistics are provided by State and sector in every edition of the *NGM*.

Methodology

This report analyzes the differences between data collected from the EIA-857 and EIA-176 that appear in the *NGM*, or revision errors. Throughout this report, data obtained from EIA-857 are referred to as initial estimates, while data from EIA-176 are referred to as revised or annually adjusted values. The analysis focuses on data series for consumption and price for residential, commercial, and industrial consumers, as these series are common to both the EIA-857 and EIA-176 surveys. Revision errors are a percentage term which can be defined as:

Revision error = [(revised value – initial estimate)/revised value]*100

A positive revision error indicates an underestimation, whereas a negative revision error indicates an overestimation. Revision errors were calculated for each month in each year in the series, 2004 to 2007. These monthly revision errors were averaged (simple unweighted average) to obtain an annual average revision error. These annual means were totaled and divided by four to obtain the 4-year mean. All tables in this report provide the 4-year statistics. For tables providing statistics for individual years for each series, see Appendix A of this report.

At the national level, the industrial consumption and price series exhibit the largest 4-year mean revision error.

Although the commercial consumption series has the largest positive and negative annual mean revision errors, the 4-year mean revision errors for each customer class within the consumption series are lower than those of the corresponding customer classes within price series. Overall, the industrial series exhibit the largest 4-year means within both the consumption and price series of approximately 0.6 percent and 1.3 percent, respectively (Table 1).

Table 1: 4-Year Mean Revision Errors of Monthly U.S. Natural Gas Data (in percent)

Series	Volume Delivered to Consumers	Average Price
Residential	-0.15	0.30
Commercial	0.25	0.29
Industrial	0.59	1.28

Source: Energy Information Administration, Natural Gas Division

Higher revision errors among the industrial series are attributable to a number of factors. The industrial series are more sensitive to seasonality, the price of substitutes for natural gas, and economic conditions compared with the commercial and residential series. Furthermore, large commercial and industrial sales are indexed to spot market prices, subjecting data within the industrial price series to a higher degree of volatility. In addition, industrial prices are more sensitive to supply disruptions, weather variation, and demand levels than residential and commercial prices.

An examination of the mean annual revision error for each customer class within both series provides some explanation of the discrepancies among 4-year means.

At the national level, revision errors in 2007 were on average smaller than in 2004.
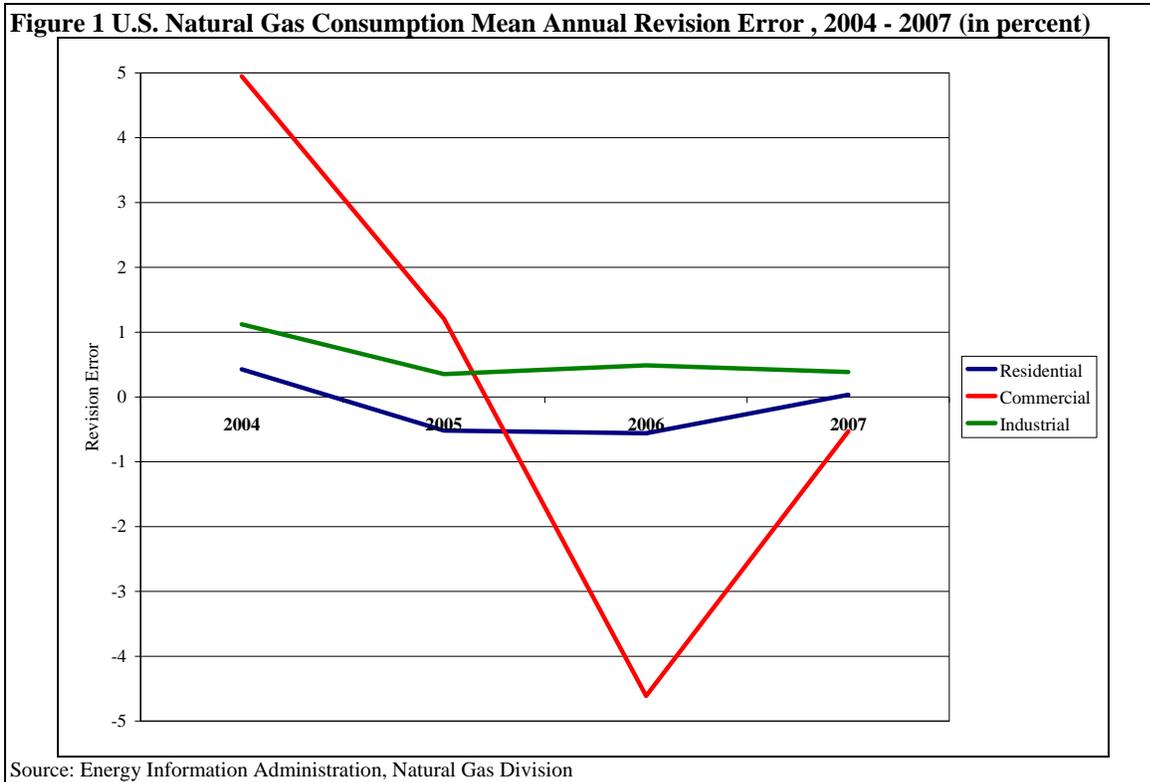
This section examines the revision error annual means for the residential, commercial, and industrial consumption and price series from 2004 to 2007. Some customer classes' annual means exhibit non-monotonic curves, or curves which oscillate between positive and negative values. The existence of non-monotonic curves obscures the formulation of definitive conclusions regarding overall trends. However, a comparison of 2004 and 2007 annual means in combination with analyses of the ranges and signs of each customer class' annual means provides some insight regarding the data.

Overall, the 2007 revision error annual mean for each customer class within the consumption series is lower in absolute terms relative to their respective 2004 annual means. The residential

series' annual means range between -0.6 percent and 0.4 percent, while the industrial series' annual means range between 0.4 percent and 1.1 percent. By comparison, the annual means for the commercial series range between -4.6 and 5 percent. Notably, the lower value of the commercial series' range is for 2006, during which a reporting error occurred. For the months of August through October 2006, a California natural gas distributor initially reported electric consumption of natural gas as commercial consumption. Annually adjusted values for 2006 reflect the correction of this error.

Only the 2004 revision error annual mean for the residential consumption series is positive, indicating that initial 2004 monthly values for residential consumption were typically underestimates. In 2005 and 2006, initially published values for residential consumption were typically overestimates, as these years' negative annual means indicate. The 2007 annual mean is negligible (0.03 percent), indicating that initially published values for residential consumption usually did not require further revision. Notably, the annual means for the industrial series are positive from 2004 to 2007, indicating that initially published values were usually underestimates throughout the study period.

Conversely, revision error annual means for the commercial consumption series indicate that initially published values were typically underestimates in 2004 and 2005, but overestimates in 2006 and 2007 (Figure 1). Notably, the commercial series exhibits larger annual means than the industrial series, but the industrial series has a higher 4-year mean. The commercial series annual means' fluctuations between negative and positive values produced a deceivingly low 4-year mean.

Figure 1 U.S. Natural Gas Consumption Mean Annual Revision Error , 2004 - 2007 (in percent)

Source: Energy Information Administration, Natural Gas Division

Like the consumption series, the 2004 revision error annual means for each customer class within the price series are lower relative to 2007 annual means. However, the respective ranges of the annual means for each customer class within the price series are smaller than those of the consumption series. The annual means for the commercial and industrial price series exhibit relatively wider ranges of -0.9 percent and 1.9 percent and 0.1 percent and 2.6 percent, respectively. The annual means for the residential series range between 0.1 and 0.5 percent.

The signs of the revision error annual means for the residential and industrial series indicate that initially published values were typically underestimates in every year of the study period. The initially published values of the commercial series were also usually underestimates in 2004 and 2005, but overestimates in 2006 and 2007 (Figure 2). Similar to the commercial consumption series, the commercial price series annual means' fluctuations between negative and positive values produced a deceivingly low 4-year mean.

Figure 2 U.S. Natural Gas Price Mean Annual Revision Error, 2004 – 2007 (in percent)

Source: Energy Information Administration, Natural Gas Division

These findings provide several conclusions regarding the data. Primarily, the 2004 revision error annual means are lower in absolute terms than the 2007 revision error annual means for every customer class within both series, suggesting that fewer initial monthly estimates required further revision in 2007 than in 2004. In addition, the initially published values for the commercial consumption and price series were typically underestimates in 2004 and 2005, but overestimates in 2006 and 2007. The initially published values for the industrial consumption and price series were underestimates for every year in the study period. The residential consumption and price series typically exhibited the smallest year-to-year fluctuations in revision error annual means. Overall, these conclusions emphasize these series' sensitivity to market conditions and other

external factors. See Appendix B of this report for the monthly revision errors of individual series by year.

The high 4-year mean revision errors of several States reflect the impact of State- and respondent-specific factors.

Reflecting 4-year mean revision errors at the national level, the industrial consumption and price series generally exhibit the largest 4-year means, while the residential series exhibit the smallest. Notably, only five States have one of the highest 4-year mean for multiple series (Table 2).

Table 2 Top Three States with Highest 4-Year Mean Revision Errors for the Consumption and Price Series

Volume Delivered to Consumers		Average Price	
Customer Class	Mean	Customer Class	Mean
Residential		**Residential**	
Delaware	2.23	New Jersey	2.82
New Hampshire	-2.12	Delaware	-2.56
Florida	-1.97	Florida	-2.49
Commercial		**Commercial**	
Texas	-16.68	Alaska	-11.49
New York	11.77	Mississippi	2.78
North Carolina	3.05	Massachusetts	2.19
Industrial		**Industrial**	
Alaska	-26.47	Oklahoma	-9.16
Massachusetts	-21.97	Tennessee	7.43
Connecticut	-8.24	New York	-7.14

Source: Energy Information Administration, Natural Gas Division

Regarding residential consumption and price, Delaware and Florida have some of the highest 4-year means for both series. In absolute terms, these States' 4-year means for the consumption series are approximately 2 percent, but closer to 3 percent for the price series. On the 2006 EIA-176, a large respondent in Delaware made significant upward revisions to the data it reported on the EIA-857 for residential consumption and price, likely affecting this State's 4-year mean for these series.

New York exhibits one of the highest 4-year means for the commercial consumption and industrial price series. This State's 4-year means indicate that on average, initial values for the commercial consumption series were underestimates, while initial values for the industrial price series were usually overestimates during the study period. In 2004, annually adjusted values for New York's commercial consumption reflect the inclusion of a large respondent's sales to itself in end-use commercial consumption, resulting in low initial values for this series. In addition, a large respondent in New York's 2007 commercial consumption data are still in question and may be revised in future natural gas publications. Regarding industrial price data, a large respondent in New York made significant revisions to the 2005 and 2006 EIA-857 data for industrial price on the EIA-176, likely resulting in overestimated initial values for this series.

Massachusetts and Alaska exhibit some of the highest 4-year means for the industrial consumption and commercial price series. Massachusetts' 4-year means indicate that initial

values for the industrial consumption series were typically overestimates, but initial values for the commercial price series were usually underestimates. In 2004, one of the largest respondents to the EIA-857 in Massachusetts experienced a substantial decrease in industrial deliveries, likely contributing to this State's negative 4-year mean for this series.

Alaska's 4-year means indicate that on average, initial values for the industrial consumption series and commercial price series were overestimates. Alaska's high 4-year mean for these series may be attributable to the presence of one company that necessitated a large revision combined with the presence of a low number of natural gas companies reporting data. For example, in 2005 only five Alaskan companies reported commercial prices, while in 2004 only three companies reported industrial sales. The low number of natural gas companies operating in Alaska during these years necessitated the use of a total universe in the EIA-857, the survey which typically uses a sample to determine initial estimates. Furthermore, one company reporting data in both 2004 and 2005, Enstar, had over half of the commercial and industrial market shares in Alaska during those years.

State- and respondent-specific issues contributed to the high 4-year means of other States. For example, Connecticut exhibits a high negative 4-year mean for the industrial consumption series, indicating initial values for this series were typically overestimates. In 2005, an unusually mild winter in Connecticut caused lower-than-expected consumption levels, resulting in data reported on the EIA-176 to be lower than data reported on the EIA-857. In another instance, a large respondent in Oklahoma was unable to report accurate industrial price data until the end of the year in 2006 and 2007, resulting in this State's high 4-year mean.

Changes in data collection also affected the 4-year means of some States. For example, in 2001 EIA began collecting marketer data, which affected the residential price of States with a higher number of alternative suppliers. Data from natural gas marketers serving residential customers began being collected and published in New Jersey in 2005 and the State exhibits the highest 4-year mean relative to other States in the residential price series.

Conclusion

Overall, the industrial consumption and price series exhibit the largest 4-year mean revision errors. However, as the commercial consumption and price series' annual means oscillate between positive and negative values, the 4-year means for these series are deceptively low. Examination of the top three States with the highest 4-year means illustrates the impact of State- and respondent-specific factors on revision errors.

Overall, the lack of a discernible pattern in 4-year and annual mean revision errors suggests a variety of factors are responsible for the revision errors from 2004 to 2007. These factors may include the series' sensitivity to market conditions such as seasonality, the price of substitutes for natural gas, and economic variability, as well as State- and respondent-specific issues.

Appendix A: Revision Errors at the National and State level by Year

Table A1 Revision Error Means and Actual Change Means in Monthly U.S. Natural Gas Data, 2004

Series	Volume Delivered to Consumers			Average Price		
Customer Class	Residential	Commercial	Industrial	Residential	Commercial	Industrial
Revision Error	0.42	4.95	1.12	0.54	1.87	2.56
Actual Change	0.42	11.92	6.50	0.06	0.18	0.17

Note: For the mean revision error, all units are percentages. For the actual change means, units for volume delivered to consumers are MMcf, while units for average price are dollars per Mcf.
Source: Energy Information Administration, Natural Gas Division

Table A2 Revision Error Means and Actual Change Means in Monthly U.S. Natural Gas Data, 2005

Series	Volume Delivered to Consumers			Average Price		
Customer Class	Residential	Commercial	Industrial	Residential	Commercial	Industrial
Revision Error	-0.52	1.21	0.35	0.45	0.30	1.95
Actual Change	-1.25	3.58	1.75	0.06	0.03	0.17

Note: For the mean revision error, all units are percentages. For the actual change means, units for volume delivered to consumers are MMcf, while units for average price are dollars per Mcf.
Source: Energy Information Administration, Natural Gas Division

Table A3 Revision Error Means and Actual Change Means in Monthly U.S. Natural Gas Data, 2006

Series	Volume Delivered to Consumers			Average Price		
Customer Class	Residential	Commercial	Industrial	Residential	Commercial	Industrial
Revision Error	-0.56	-4.61	0.49	0.10	-0.90	0.07
Actual Change	-0.92	-8.25	2.67	0.02	-0.10	0.00

Note: For the mean revision error, all units are percentages. For the actual change means, units for volume delivered to consumers are MMcf, while units for average price are dollars per Mcf.
Source: Energy Information Administration, Natural Gas Division

Table A4 Revision Error Means and Actual Change Means in Monthly U.S. Natural Gas Data, 2007

Series	Volume Delivered to Consumers			Average Price		
Customer Class	Residential	Commercial	Industrial	Residential	Commercial	Industrial
Revision Error	0.03	-0.53	0.39	0.12	-0.11	0.56
Actual Change	4.67	-0.50	2.33	0.02	-0.01	0.04

Note: For the mean revision error, all units are percentages. For the actual change means, units for volume delivered to consumers are MMcf, while units for average price are dollars per Mcf.
Source: Energy Information Administration, Natural Gas Division

Table A5 Top Three Largest State Revision Errors for U.S. Natural Gas Consumption, 2004

Residential	Revision Error Mean (Percentage)	Actual Change Mean (MMcf)
New Hampshire	-7.12	-27.25
Massachusetts	-4.65	-391.82
Nebraska	-4.12	-81.33
Commercial		
New York	28.59	6,772.33
Maryland	-10.01	-487.25
Pennsylvania	-6.76	-679.75
Industrial		
Massachusetts	-53.77	-1,017.25
Alaska	-50.05	-1,906.67
Maryland	23.96	453.25

Source: Energy Information Administration, Natural Gas Division

Table A6 Top Three Largest State Revision Errors for U.S. Natural Gas Consumption, 2005

Residential	Revision Error Mean (Percentage)	Actual Change Mean (MMcf)
Utah	-5.60	39.58
North Carolina	4.36	54.17
Massachusetts	-3.31	-252.27
Commercial		
Texas	-29.17	-3,114.18
New York	23.54	6,582.58
Wisconsin	6.36	251.58
Industrial		
Massachusetts	-36.68	-1,388.33
Alaska	-36.63	-1,650.08
Connecticut	-14.16	-244.58

Source: Energy Information Administration, Natural Gas Division

Table A7 Top Three Largest State Revision Errors for U.S. Natural Gas Consumption, 2006

Residential	Revision Error Mean (Percentage)	Actual Change Mean (MMcf)
Delaware	7.34	48.92
Washington	-3.65	-138.42
Massachusetts	3.13	242.58
Commercial		
Texas	-41.06	-4,655.75
North Carolina	8.01	274.08
Maine	-4.89	-15.50
Industrial		
West Virginia	13.15	351.75
Colorado	-11.07	-910.17
Arizona	8.06	127.92

Source: Energy Information Administration, Natural Gas Division

Table A8 Top Three Largest State Revision Errors for U.S. Natural Gas Consumption, 2007

Residential	Revision Error Mean (Percentage)	Actual Change Mean (MMcf)
West Virginia	-8.14	-164.00
Montana	-5.84	-53.42
Nebraska	3.37	46.00
Commercial		
West Virginia	-13.38	-260.33
Delaware	-11.58	-70.42
Massachusetts	7.26	385.55
Industrial		
Nebraska	22.78	1,097.36
Alaska	-19.44	-272.33
Kansas	12.45	1,314.92

Source: Energy Information Administration, Natural Gas Division

Table A9 Top Three Largest State Revision Errors for U.S. Natural Gas Prices, 2004

Residential	Revision Error Mean (Percentage)	Actual Change Mean (dollars per Mcf)
New Hampshire	6.86	1.07
New Jersey	6.10	0.67
Florida	-3.00	-0.60
Commercial		
Alaska	-12.63	-0.51
Georgia	6.46	0.82
Washington	5.92	0.57
Industrial		
Alaska	29.84	1.11
Montana	-21.51	-1.39
Tennessee	18.82	1.39

Source: Energy Information Administration, Natural Gas Division

Table A10 Top Three Largest State Revision Errors for U.S. Natural Gas Prices, 2005

Residential	Revision Error Mean (Percentage)	Actual Change Mean (dollars per Mcf)
New Jersey	5.80	0.83
Florida	-2.69	-0.62
Georgia	-2.53	-0.49
Commercial		
Alaska	-6.40	-0.30
New York	-2.48	-0.33
South Carolina	1.94	0.26
Industrial		
New York	-11.11	-1.13
Colorado	-10.13	-0.88
Georgia	-7.59	-0.77

Source: Energy Information Administration, Natural Gas Division

Table A11 Top Three Largest State Revision Errors for U.S. Natural Gas Prices, 2006

Residential	Revision Error Mean (Percentage)	Actual Change Mean (dollars per Mcf)
Delaware	-7.81	-1.51
Colorado	3.78	0.43
Louisiana	3.31	0.57
Commercial		
Alaska	-26.17	-1.19
New York	-8.48	-0.95
Maine	-5.76	-0.65
Industrial		
Montana	23.11	2.80
Oklahoma	-12.34	-1.06
New York	-9.60	-0.89

Source: Energy Information Administration, Natural Gas Division

Table A12 Top Three Largest State Revision Errors for U.S. Natural Gas Prices, 2007

Residential	Revision Error Mean (Percentage)	Actual Change Mean (dollars per Mcf)
Tennessee	-4.24	-0.63
Georgia	-2.99	-0.61
South Carolina	-2.95	-0.65
Commercial		
Tennessee	-6.10	-0.73
Louisiana	-4.89	-0.54
Massachusetts	2.48	0.38
Industrial		
Oklahoma	-17.10	-1.52
Illinois	-16.16	-1.47
Pennsylvania	-13.40	-1.39

Source: Energy Information Administration, Natural Gas Division

Appendix B: National Monthly Revision Errors by Series

Figure B1 U.S. Natural Gas Residential Consumption Revision Errors, 2004 - 2007 (in percent)
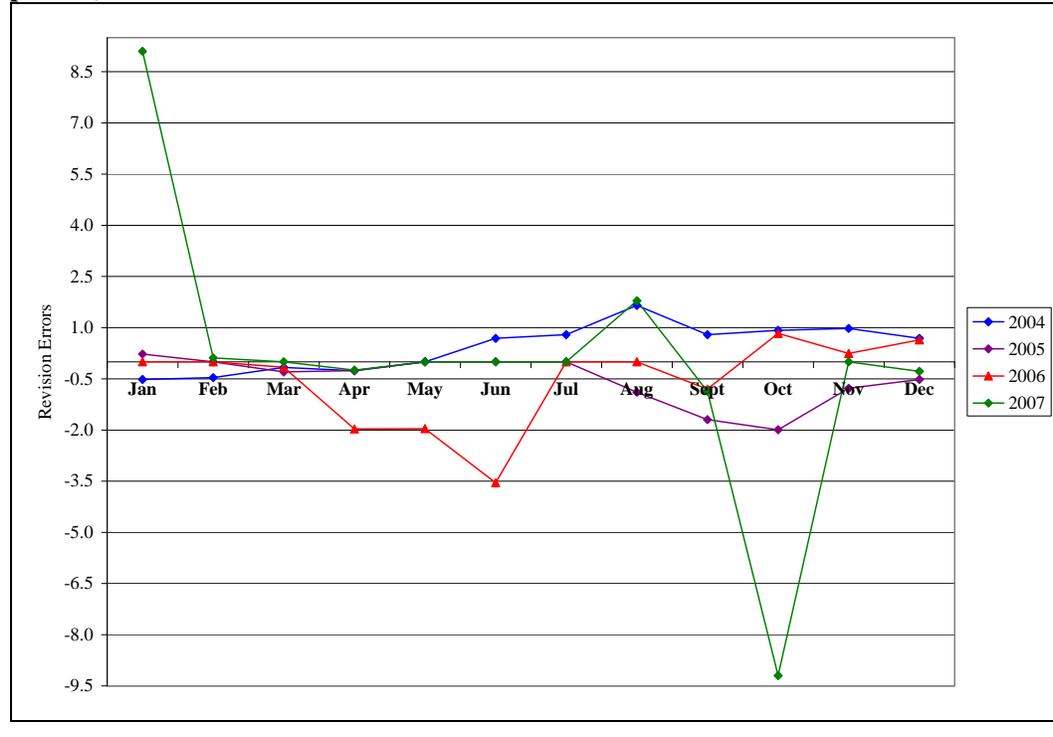
Figure B2 U.S. Natural Gas Commercial Consumption Revision Errors, 2004 – 2007 (in percent)
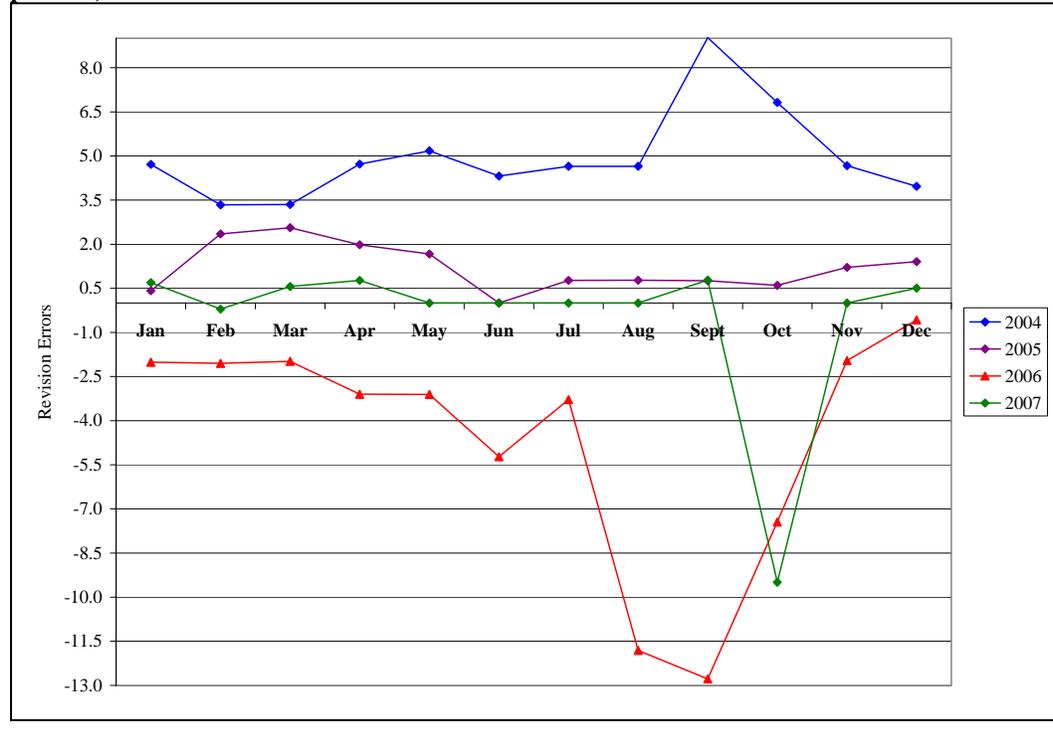

13

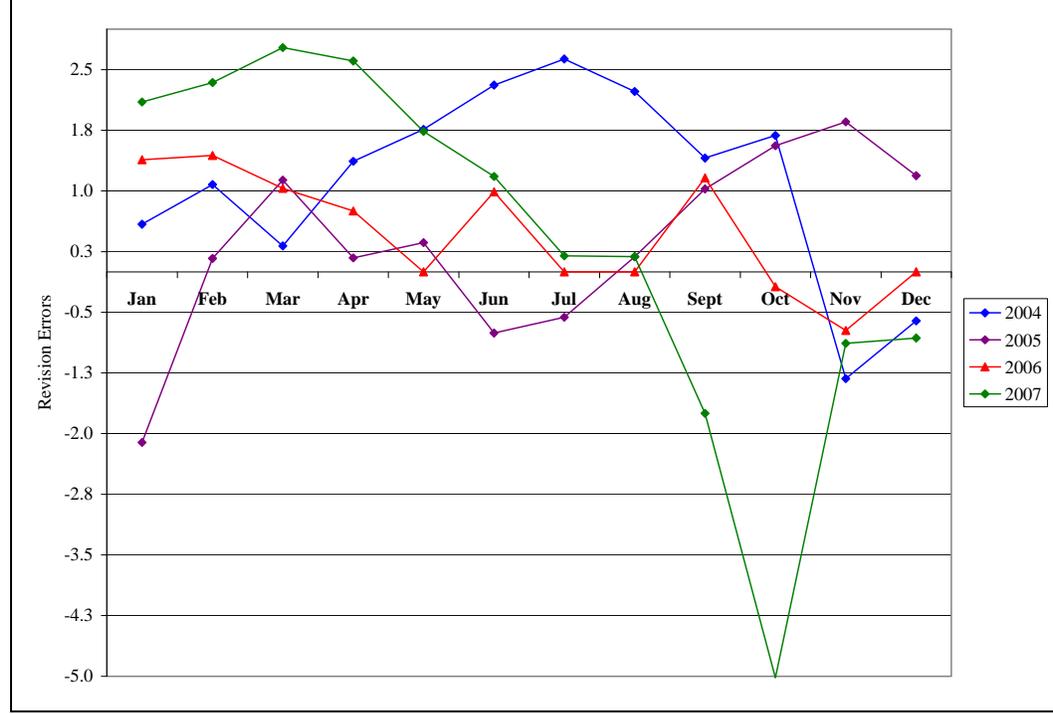**Figure B3 U.S. Natural Gas Industrial Consumption Revision Errors, 2004 – 2007 (in percent)**

Figure B4 U.S. Natural Gas Residential Price Revision Errors, 2004 – 2007 (in percent)
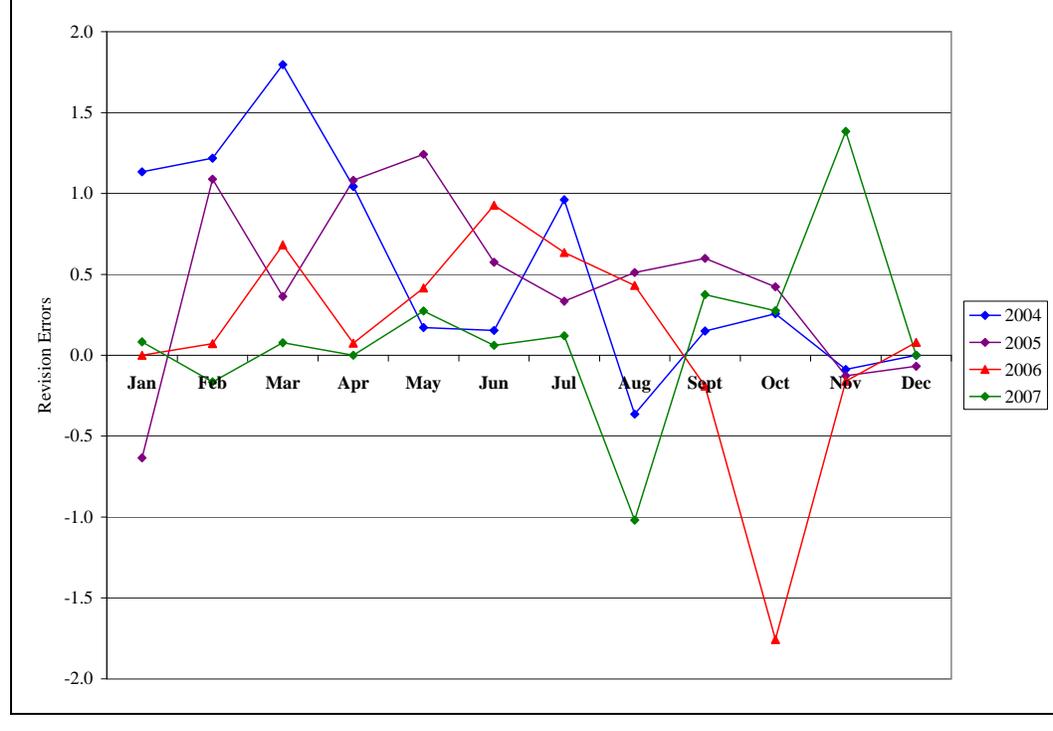
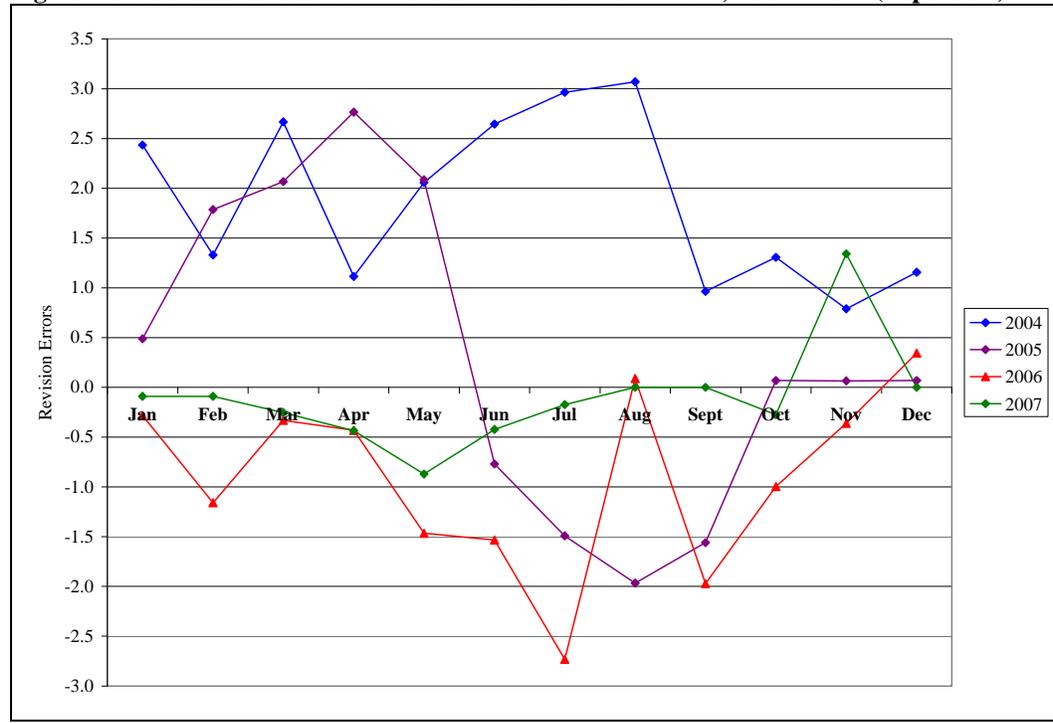**Figure B5 U.S. Natural Gas Commercial Price Revision Errors, 2004 – 2007 (in percent)**

Figure B6 U.S. Natural Gas Industrial Price Revision Errors, 2004 – 2007 (in percent)