Market Fundamentals and the Dynamics of Natural Gas Futures Volatility: An Augmented GARCH Approach

Ibrahim Ergen, Federal Reserve Bank of Richmond, Baltimore, MD
Islam Ridvanoglu, Zirve University, Gaziantep, Turkey

July 9, 2014

Abstract

We investigated the determinants of daily volatility for natural gas nearby-month futures traded in NYMEX within a GARCH framework augmented with market fundamentals. Consistent with the previous literature, we found that volatility is much higher on the storage level announcement days, on Mondays and during winters defined as December, January and February. We also confirmed the previous finding that high volatility is associated with weather shocks in excess of seasonal norms. Samuelson’s hypothesis is also investigated and a significant time to maturity effect is detected. The impact of the (storage x season) interaction is investigated for the first time in this paper. Inclusion of this interaction term in our regressions produced very interesting results: The mainstream finding in the literature that lower storage levels result in higher volatility is valid only during winter. At other times, it is actually higher storage levels that results in higher volatility. Economic intuition behind this novel finding is discussed in detail. This study also fills the gap in the literature regarding the out-of-sample forecasting accuracy of GARCH models augmented with the market fundamentals. We found that augmentation with market fundamentals improves the forecast accuracy evidenced by reduced mean absolute error and mean squared error compared to standard GARCH models with no market fundamentals.
1 Introduction

Natural gas futures contracts began trading on the New York Mercantile Exchange (NYMEX) in April 1990. One contract is written on 10,000 MMBTU of natural gas to be delivered to Henry Hub.\footnote{MMBTU= 1 Million British Thermal Units (BTU). A BTU is a unit used to describe the heat value (energy content) of fuels. A BTU is defined as the amount of heat required to raise the temperature of one pound of liquid water by one degree from 60°F to 61°F at a constant pressure of one atmosphere. (www.wikipedia.org)}\footnote{Henry Hub is a point on the natural gas pipeline system in Erath, Louisiana. It interconnects with nine interstate and four intrastate pipelines. Spot and future prices set at Henry Hub are denominated in $/MMBTU and are generally regarded as the primary price set for the North American natural gas market. (www.wikipedia.org)} Contracts for delivery in each month and for up to six years out are traded at any point in time. Trading in a given contract ends three business days before the first calendar day of the delivery month.

Trading in natural gas futures has skyrocketed in recent years. Open interest in Nymex natural gas futures grew at a rate of 15.2 percent per year during the last decade (Chiou-Wei, Linn and Zhu, 2007). Daily volume is in the order of 60,000 to 100,000 contracts for the nearby month futures and 20,000 to 60,000 contracts for the second nearby futures.\footnote{Nearby Future Contract is the earliest maturing contract. This corresponds to the next month delivery contract for natural gas futures.} \footnote{Second Nearby Contract is the second earliest maturing contract In natural gas markets, it corresponds to the contract for delivery on the month after the next month.}

The volatility of natural gas prices has received increasing attention in recent years. The extreme fluctuations in both spot and futures prices caused researchers and market practitioners to focus on the sources of this high volatility. Whether news about natural gas market fundamentals, or excessive speculation and irrational investor behavior is responsible for the high volatility is an ongoing debate. In this paper, empirical evidence is provided that natural gas futures price volatility is driven by market fundamentals within a GARCH type of dynamic volatility framework.

The response of prices to shifts in supply and demand depends on price elasticity of the commodity. In general, natural gas markets are highly inelastic in both the supply and demand side; hence, the price is very responsive to short-term changes in both, which results in high volatility. Two key fundamental pieces of information affecting the natural gas markets are the level of working gas in storage facilities
and weather changes. These are generally regarded as proxies for supply and demand.

For supply conditions, the storage report is perceived as the most important piece of information by natural gas market participants. The report is currently prepared by the Energy Information Administration (EIA) of the Department of Energy (DOE). Anecdotal evidence on the effect of the natural gas storage report is abundant in the financial press. The following appeared in Communications, Energy and Paperworkers Union of Canada (CEP) News on Thursday, October 30, 2008:

*Underground natural gas storage in the U.S. increased 46 billion cubic feet (Bcf) in the week ending Oct. 24, according to the Energy Information Administration (EIA)'s weekly report on Thursday. Expectations had been for a 41 Bcf increase. Following the report, natural gas prices bottomed out, and are now at new session lows of 6.528 from pre-report levels of 6.811.***

The report provides the level of total underground working gas and the historical average of this quantity for the equivalent time periods of last five years. Although the effects of storage surprises on short-term volatility have been emphasized in many studies (Gregoire and Boucher 2008; Mu 2007; Linn and Zhu (2004)), we did not identify any research considering the effects of the interaction between the storage surprises and seasonality. This paper addresses this issue. In the winter, natural gas demand spikes, and the supply is unable to react quickly since the production of natural gas is uniform across seasons. When this happens, low storage levels, as compared to historical averages may be regarded as tight supply situations and put pressure on gas prices, which results in high volatility. Conversely, during the other seasons, higher storage levels than historical averages may increase concerns regarding the capacity of storage facilities and result in highly volatile natural gas prices. In this paper, supporting evidence is presented for this hypothesis, which implies asymmetric effect of storage surprises in different seasons.

For demand conditions, the key information for short-term volatility dynamics is a change in the weather. Upon the observation of unexpected cold weather during the winter months, the extra demand

---

*Working gas in storage is the volume of gas in the reservoir that is in addition to the cushion or base gas. Base gas is the volume of gas needed as a permanent inventory to maintain adequate reservoir pressures. (www.doe.eia.gov)*
for heating pushes the prices up causing higher volatility since the supply cannot adjust to such changes in the short run. In recent years, power generation plants have used more natural-gas-fueled technology. Natural gas usage in electricity generation rose from 12 percent to 17 percent between 1990 and 2006 (Hartley et al., 2007.a). As a result, hotter than expected temperatures during the summer months increase the demand for cooling and might have a similar effect on volatility. Empirical evidence is presented that weather anomalies, measured as the degree days in excess of seasonal norms, result in increased short-term volatility.

In addition to the supply-and-demand driven volatility, the previous literature found significantly higher volatility on Mondays (Mu 2007; Murry and Zhu 2004). The study of Fleming et al. (2004) explains this effect based on the continuous weather information flow during the long nontrading weekend period.

Additionally, time to maturity may be another determinant of futures market volatility. Samuelson (1965) was the first to claim that the volatility of futures prices increases as the contract maturity gets closer. Using a very large futures dataset on 6,805 contracts, Daal et al. (2006) found that the maturity effect was much stronger for agricultural and energy commodity futures than it was for financial futures. Using extreme value method, to measure the daily volatility of natural gas futures from daily low and daily high prices, Gregorio and Boucher (2008) found that the maturity effect was significant even after controlling for storage surprises. On the other hand, Mu (2007) tested for the maturity effect by fitting separate GARCH models to the nearby futures contracts and the second nearby contracts and comparing the fitted daily volatilities. In this paper, we present empirical evidence of the maturity effect by directly including a time to maturity variable in the conditional variance equation of the GARCH model, a substantially different approach from past research.

In this study, GARCH models are used as an econometric tool in order to account for the dynamic nature of short-term market volatility. Some of the more recent studies on natural gas volatility augment the GARCH models with market fundamentals in order to focus on the determinants of volatility (Ates and Wang 2008; Mu 2007; Pyndick 2004; Murry and Zhu 2004). However, the literature does not address the out-of-sample forecasting accuracy of GARCH models augmented with market fundamentals. In this paper, we also study the out-of-sample forecasting accuracy of several simple GARCH models together
with augmented GARCH models using a constant-size sliding-sample methodology.

The remaining sections of this paper are organized as follows: In section 2, a statistical analysis of the nearby month futures returns is presented, and natural gas volatility with respect to the market fundamentals is analyzed without imposing any econometric structure. In section 3 the econometric model is introduced and estimation results are presented. Out-of-sample volatility forecasting accuracy of augmented GARCH models is tested in section 4. The Usefulness of augmented GARCH models for natural gas price risk measurement is evaluated in section ???. Lastly, in section 5, we discuss the conclusions.

2 Data and Statistical Analysis

Natural gas futures price data from February 2001 to May 2008 were obtained from NYMEX contracts. Contract-by-contract price data are available from DataStream. The return series for nearby month futures are constructed in two steps. First, returns for individual contracts $i$ are calculated by

$$ r_{t,i} = \ln\left( \frac{F_{t,i}}{F_{t-1,i}} \right) $$

where $F_{t,i}$ is the price of the futures contract $i$ at time $t$. Then, the nearby month contract return for time $t$ is obtained as,

$$ r_{t,nb} = r_{t,j} $$

where $j$ is the earliest maturing contract. In other words, day $t$ is in month $j - 1$. By first obtaining the returns and then rolling over the contracts, constructing price series from different contracts is avoided, which may distort the data. Therefore, all nearby futures returns $r_{nb,t}$ are tradable and realizable. At the end of this procedure, the nearby month futures return data are obtained that run from January 4, 2001 to April 23, 2008, a total of $T = 1,823$ daily observations.

Summary statistics for nearby month futures returns are presented in Table 1. The numbers in parenthesis are the probability values for the associated tests. The returns are right skewed for this sample period and exhibit excess kurtosis. Consequently, the Jarque-Bera test rejects the null hypothesis of normal distribution. These statistics imply that natural gas futures returns are not normally distributed.
However, a comparison with the summary statistics of emerging market stock returns given in Table ?? reveals that natural gas futures returns are much closer to being normal despite their higher standard deviation.

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>LBQ(5)</th>
<th>LBQ(10)</th>
<th>LBQ²(5)</th>
<th>LBQ²(10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>-0.09538</td>
<td>5.31</td>
<td>7.526</td>
<td>61.81</td>
<td>74.195</td>
</tr>
<tr>
<td>Median</td>
<td>-0.09935</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variance</td>
<td>12.7719</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard deviation</td>
<td>3.5738</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skewness</td>
<td>0.3779</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kurtosis</td>
<td>4.9046</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jarque-Bera</td>
<td>1877.15</td>
<td>(0.00)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Summary Statistics for Nearby Month Futures Returns

The Ljung-Box test at lags 5 and 10 does not reject the null hypothesis of no autocorrelation in raw returns. However, for squared returns, the null hypothesis of no autocorrelation is rejected strongly. This suggests that there is strong volatility persistence in the data. Therefore, the LM-ARCH test (Engle, 1982) is administered and the null hypothesis of no ARCH effects is strongly rejected. An augmented Dickey Fuller test (Dickey and Fuller, 1979) is administered with a constant and 12 lags in the unit root regression without a time trend. The null hypothesis of non-stationarity is rejected, so non-stationarity is not a problem in the analysis.

The gas storage report was being announced by the American Gas Association (AGA) until May 2002 on Wednesdays at 2:00 pm. Since then, the report is released every Thursday at 10:30 am by the Energy Information Administration (EIA). The report provides information on storage levels the Friday before, net weekly changes in storage levels, and the storage levels one-year before. In addition, the five-year historical average for the equivalent time period and the difference between the current level and the five-year average are reported. The deviation from the historical average is the key variable in this study.

The storage data are publicly available from the EIA website. From all the above-mentioned variables, the downloadable data only includes the storage levels. we followed a two-step procedure to construct the deviations from historical levels. First, the weekly data are

---

6The data are downloadable from http://tonto.eia.doe.gov/dnav/ng/ngstor_wkly_s1_w.htm
interpolated to obtain a daily storage level data.\textsuperscript{7} This is needed so that storage data for the same day in each of the previous five years are available. In the second step, the storage deviation variable $SD_t$ is constructed as

\begin{equation}
SD_t = S_{t,s} - \frac{1}{5} \sum_{i=1}^{5} S_{t,s-i}
\end{equation}

where $S_{t,s}$ is the level of storage on day $t$ in year $s$. This two-step procedure is the same methodology followed by the EIA to construct the five-year averages and the deviations from historical averages while preparing the storage report.\textsuperscript{8}

A time series plot of the working gas in underground storage is presented in Figure 1 panel A.

The demand is highly seasonal, and the supply is uniform across seasons in the natural gas market. As such, storage levels also exhibit very strong seasonality because it balances the difference between the supply and demand providing a buffer to the market. In Figure 1 panel B, the time series plot of the storage deviation from the five-year historical average is provided. Note that during the winters of 2001 and 2003 the storage levels were significantly lower than their five-year historical averages, and these periods also coincide with high volatility in natural gas markets.

we obtained the daily realized temperature data running from January 1960 to April 2008 from the trading floor of a very active natural gas trading firm. The dataset includes daily minimum and daily maximum temperatures for seven locations: Atlanta Hartsfield Airport, Chicago Midway Airport, Chicago O’Hare Airport, Dallas Fort Worth Airport, New York Central Park, New York JFK Airport, and New York LaGuardia Airport. The discussion of the weather modeling is left to Section 3.2.

\subsection*{2.1 A First Look at the Natural Gas Volatility}

In this section, we analyze the nearby month futures volatility with respect to natural gas market fundamentals without imposing any econometric structure. The previous literature emphasizes higher volatility

\textsuperscript{7}Simple linear interpolation is used here.

\textsuperscript{8}A complete documentation of the EIA methodology can be found at the link: http://www.eia.doe.gov/oil_gas/natural_gas/ngs/methodology.html.
Figure 1: Working Gas In Underground Storage and Its Deviation From Five Year Historical Mean

on Mondays and storage report announcement days. In some studies, winter was found not to be associated with higher volatility after
controlling for other factors, we believe this result is driven by using a broad definition for winter: November to March. This coincides with the period when withdrawal from storage is higher than the injection to storage; hence, known as the withdrawal season. Here, we restrict the definition of winter to include only December, January, and February.

In Table 2, the standard deviations of nearby month futures returns are presented for several subgroups based on certain characteristics. There are several important patterns in this table. In panel A, standard deviations are calculated for Mondays, storage report announcement days (SDDAYs), and all other days along with winter and non-winter days.

<table>
<thead>
<tr>
<th>Panel-A: By Mondays SDDAYs and Winter</th>
<th>Winter</th>
<th>Non-Winter</th>
<th>All Seasons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monday</td>
<td>6.30</td>
<td>4.24</td>
<td>4.79</td>
</tr>
<tr>
<td>SDDAY</td>
<td>4.37</td>
<td>3.66</td>
<td>3.85</td>
</tr>
<tr>
<td>Other Days</td>
<td>3.43</td>
<td>2.79</td>
<td>2.96</td>
</tr>
<tr>
<td>All Days</td>
<td>4.27</td>
<td>3.31</td>
<td>3.57</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: By Winter and Bidweek</th>
<th>Bidweek</th>
<th>Non-Bidweek</th>
<th>All Days</th>
</tr>
</thead>
<tbody>
<tr>
<td>Winter</td>
<td>5.57</td>
<td>4.00</td>
<td>4.27</td>
</tr>
<tr>
<td>Non-Winter</td>
<td>3.34</td>
<td>3.31</td>
<td>3.31</td>
</tr>
<tr>
<td>All Seasons</td>
<td>4.03</td>
<td>3.49</td>
<td>3.57</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel C: By SD and Winter</th>
<th>$SD &gt; 0$</th>
<th>$SD &lt; 0$</th>
<th>All Days</th>
</tr>
</thead>
<tbody>
<tr>
<td>Winter</td>
<td>3.73</td>
<td>5.66</td>
<td>4.27</td>
</tr>
<tr>
<td>Non-Winter</td>
<td>3.42</td>
<td>2.84</td>
<td>3.31</td>
</tr>
<tr>
<td>All Seasons</td>
<td>3.51</td>
<td>3.80</td>
<td>3.57</td>
</tr>
</tbody>
</table>

Table 2: Standard Deviation of Daily Returns Broken into Groups

The standard deviation is highest for Monday returns, 4.79. Although lower than Mondays, the standard deviation on storage report announcement days, 3.85, is higher than the standard deviation for all other days, 2.96. The winter effect is also evident. The return volatility for winter days, 4.27, is higher than the return volatility for
non-winter days, 3.31. The reason for higher volatility on storage report announcement days is obvious. The storage report is regarded as the most important piece of information by market practitioners, and it is priced as soon as it becomes available. Using intraday data, Linn and Zhu (2004) showed that the new information is absorbed into the prices within minutes. The volatility of the 10:30-10:35 am interval is much higher than any other five-minute interval during the trading day. However, when Thursdays were excluded from their sample, this effect is completely gone. Fleming et al. (2004) explain the Monday effect with the continuous flow of weather information. They find that the variance ratios of trading to nontrading periods are significantly lower for weather sensitive markets compared to equity markets. They attribute the difference to the continuing flow of weather information over the nontrading period, whereas the information flow for equity markets is reduced in the nontrading period. Additionally, the ratios get even lower over the weekend compared to weekdays, which supports their hypothesis further.

The volatilities for subgroups generated by the Cartesian product of days and seasons also make complete sense. Standard deviations for different days have the same order both in and outside of winter,
and winter volatility is consistently higher than non-winter volatility in any subgroup of days. A bar plot of the statistics presented in Table 2 panel A is presented in Figure 2(a).

Panel B of Table 2 analyzes the standard deviations of the returns in the same way, but now the sample is split as the bidweek and non-bidweek days across winter and non-winter days. In the natural gas market, the largest volume of spot trading occurs in the last five business days of every month known as “bidweek.” This is when producers are trying to sell their core production and consumers are trying to buy for their core natural gas needs for the upcoming month (see www.naturalgas.org). The average prices set during bidweek are commonly the prices used in physical contracts over the next month. Since the trading in futures contracts terminates on the third business day before the first business day of the next month and bidweek is the last five business days of the month, the last three business days of trading for the nearby month contract coincides with the bidweek. In analyzing the maturity effect first proposed by Samuelson (1965), we use bidweek as a natural cutoff point. In panel B of Table 2, the standard deviation of the returns on the bidweek days is 4.03, whereas the standard deviation of returns not on the bidweek days is 3.49. The maturity effect is particularly strong during winter. The standard deviation of bidweek days in winter is 5.57, but the standard deviation of non-bidweek days in winter is 4. On the other hand, there is only a marginal difference between the standard deviations outside of winter, 3.34 for bidweek days and 3.31 for non-bidweek days. A bar plot of the statistics presented in Table 2 panel B is provided in Figure 2(b).

In panel C, $SD$ is the storage deviation variable constructed by (3). The short-term volatility studies in the literature found that lower than expected storage levels results in increased volatility because this signals a tight supply situation to the natural gas market (Mu 2007). However, to my knowledge, there are no studies analyzing the effects of the interaction of storage levels with seasonality. In panel C, the finding of the previous literature is first confirmed and then challenged. The standard deviation of the periods in which $SD < 0$; that is, storage level is lower than five-year historical average, is 3.8, whereas the standard deviation of the periods in which $SD > 0$ is 3.51. This confirms the previous literature. However, a more careful examination of the table reveals that this relationship is valid only during the winter months when supply tightness is really a big problem. During the
winter, those periods with \( SD < 0 \) have a standard deviation of 5.66, whereas those periods with \( SD > 0 \) have a standard deviation of 3.73. During the non-winter months, the effect is just the opposite: Returns of those periods with \( SD > 0 \) have a standard deviation of 3.42, whereas the returns of those periods with \( SD < 0 \) have a standard deviation of 2.84. This should be because of the concerns regarding the storage capacities. Very high storage levels during non-winter months when demand is minimal increase the concerns about whether there will be enough storage space to store the production for winter demand. This puts a pressure on the price of storage space, which naturally spills over to natural gas prices, causing excessive volatility. Lee Van Atta (2008) cites the excess volatility as one of the most important reasons leading to excessive storage construction over the past few years. This view is consistent with the finding here. A bar plot of the statistics presented in Table 2 panel C is presented in Figure 2(c).

### 2.2 Levene Tests for Variance Equality

We formally test for the equality of variances among some subsamples of the data in this section. The robust Brown-Forsythe (1974) type Levene (1960) test statistics and associated probability values are presented in Table 3. In each row of the table the null hypothesis of equal variances across the \( J \) groups in the second column is tested. In the first row, the null hypothesis of equal variances for winter and non-winter returns is rejected. In the second row, equality of variances for Mondays, storage report announcement days, and all other days is rejected. In the third row, six groups are constructed as the Cartesian product of the winter groups in the first row and days in second row. The equality of variances across these groups is rejected.
### Table 3: Brown-Forsythe Type Levene Tests for Equality of Variances

<table>
<thead>
<tr>
<th>Groups</th>
<th>J</th>
<th>BFL Test Statistic</th>
<th>Distribution under $H_0$</th>
<th>p.Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) W, NW</td>
<td>2</td>
<td>17.596</td>
<td>$F_{1,1821}$</td>
<td>2.863e-05</td>
</tr>
<tr>
<td>(2) MD, STD, OD</td>
<td>3</td>
<td>39.896</td>
<td>$F_{2,1820}$</td>
<td>0.00</td>
</tr>
<tr>
<td>(3) W&amp;MD, W&amp;STD, W&amp;OD NW&amp;MD, NW&amp;STD, NW&amp;OD</td>
<td>6</td>
<td>20.656</td>
<td>$F_{5,1817}$</td>
<td>0.00</td>
</tr>
<tr>
<td>(4) BW, NBW</td>
<td>2</td>
<td>0.2512</td>
<td>$F_{1,1821}$</td>
<td>0.6163</td>
</tr>
<tr>
<td>(5) W&amp;BW, W&amp;NBW, NW&amp;BW, NW&amp;NBW</td>
<td>4</td>
<td>6.2663</td>
<td>$F_{3,1819}$</td>
<td>0.00031</td>
</tr>
<tr>
<td>(6) W&amp;BW, W&amp;NBW</td>
<td>2</td>
<td>1.2409</td>
<td>$F_{1,425}$</td>
<td>0.2659</td>
</tr>
<tr>
<td>(7) SD &gt; 0, SD &lt; 0</td>
<td>2</td>
<td>0.8743</td>
<td>$F_{1,1821}$</td>
<td>0.3499</td>
</tr>
<tr>
<td>(8) W&amp;SD &gt; 0, W&amp;SD &lt; 0 NW&amp;SD &gt; 0, NW&amp;SD &lt; 0</td>
<td>4</td>
<td>12.978</td>
<td>$F_{3,1819}$</td>
<td>2.18e-08</td>
</tr>
<tr>
<td>(9) W&amp;SD &gt; 0, W&amp;SD &lt; 0</td>
<td>2</td>
<td>7.4212</td>
<td>$F_{1,425}$</td>
<td>0.0067</td>
</tr>
</tbody>
</table>

Notes:
1. The abbreviations in the group names are as follows: W for winter, NW for non-winter, MD for monday, STD for storage report announcement day, OD for other days, BW for bidweek, NBW for non-bidweek.
2. The Brown-Forsythe-Levene test statistic is computed as

$$F = \frac{\sum_{j=1}^{J} n_j (D_{tj} - \hat{M}_j)^2}{\sum_{j=1}^{J} \sum_{t=1}^{n_j} (D_{tj} - \hat{D}_j)^2} \frac{N - J}{J - 1}$$

where $D_{tj} = |r_{tj} - \hat{M}_j|$ and $r_{tj}$ is the return for day $t$ in group $j$; $\hat{M}_j$ is the sample median of the $n_j$ returns in group $j$; $\hat{D}_j = \sum_{t=1}^{n_j} (D_{tj}/n_j)$ is the mean absolute deviation from the median $M_j$ in group $j$; and $\hat{D}_j = \sum_{j=1}^{J} \sum_{t=1}^{n_j} (D_{tj}/N)$ is the grand mean where $N = \sum_{j=1}^{J} n_j$. The test statistic is distributed as $F_{J-1, N-J}$ under the null hypothesis of equality of variances across the $J$ groups.
3. The original Levene test uses the mean instead of the median. The optimal choice depends on the underlying distribution. However, Brown-Forsythe type test based on the median is recommended since it provides good robustness against non-normal data while retaining good statistical power.
These results are consistent with those presented in Table 2 and Figure 2. The result in row four—equal variances for bidweek days and non-bidweek days cannot be rejected—is somewhat surprising. In row five, four groups are produced from the Cartesian product of winter groups in row one and bidweek groups in row four. The equality of variances is rejected in this case. However, this may be because unequal variances of winter and non-winter dominating the analysis. To get rid of that effect, we test the equality of variances for those bidweek and non-bidweek days only in winter. The results in row six still cannot reject the equality of variances, although the probability value gets much smaller compared to that in row four. Therefore, there is not strong evidence for unequal variances for bidweek and non-bidweek days. In row seven, the equality of variances for periods with positive storage deviation and negative storage deviation is tested, and the equal variance hypothesis can not be rejected. Constructing four groups based on the Cartesian product of winter and the sign of storage deviation, the equal variance hypothesis is rejected. In the last row, to control for the winter effect, we constructed two groups with positive and negative storage deviations only from winter returns. Now, the null hypothesis of equal variances is rejected. This is consistent with my hypothesis that storage deviation has asymmetric effects during winter and outside of winter.

3 Empirical Model and Estimation Results

The Ljung-Box test statistics for squared returns in Table 1 suggest that there is strong volatility persistence for natural gas nearby month futures returns. Consequently, the LM-ARCH tests confirmed the existence of ARCH effects. In order to take this persistence into account, a GARCH volatility model is adopted as the econometric tool in this section. The focus of this study is completely on the estimation and out-of-sample prediction of daily volatility. Therefore, no structure is specified for the mean equation of the GARCH model. Instead, zero expected return is assumed. Since the day-ahead return is very difficult to forecast, this approach is common for volatility forecasting studies. Consequently, the specification of the empirical model is as
follows:

\[ r_t = \sigma_t z_t \]
\[ \sigma^2_t = \omega + \alpha r^2_{t-1} + \beta \sigma^2_{t-1} + \gamma X_t \]  

(4)

where \( r_t \) is the nearby month futures return on day \( t \) given by (1) and (2), \( \sigma_t \) is the conditional volatility, \( z_t \) is the shocks to the data generating process with \( E[z_t] = 0, E[z^2_t] = 1 \). Lastly, \( X_t \) is a vector of exogenous variables capturing the dynamics of the natural gas market volatility. The parameters of the model are obtained by maximizing the following log–likelihood function:

\[ \log L(\omega, \alpha, \beta, \gamma) \propto \sum_{t=1}^{n} \left( \log \sigma^2_t(\omega, \alpha, \beta, \gamma) - \frac{r^2_t}{\sigma^2_t(\omega, \alpha, \beta, \gamma)} \right). \]  

(5)

This likelihood function assumes that the shocks \( z_t \) are normally distributed.

### 3.1 Day-of-the-Week, Seasonality and Maturity Modeling

Inspired by the statistical analysis presented in panel A of Table 2, the following variables are included in the model.

- \( SDDAY_t \): A dummy variable for the storage report announcement days
- \( MON_t \): A dummy variable for Mondays
- \( WIN_t \): A dummy variable for winter days, with the winter defined as December, January, and February

Additionally, panel B of Table 2 presents preliminary evidence regarding the maturity effect on futures volatility. However, the Brown-Forsythe type Levene test does not confirm the unequality of the variances for the bidweek and non-bidweek days. Therefore, using a dummy variable for bidweek is not justified. Instead, we construct more general variables to capture the maturity effect:

- \( TTM \): The number of business days to the maturity of nearby month futures contract
- \( TTMWIN \): Time to maturity variable on winter days. It is constructed as \( TTMWIN_t = TTM_t \times WIN_t \) to model the asymmetric maturity effect across seasons.

The results of the estimations including these first set of exogenous variables are presented in Table 4. The first estimation is for a
simple GARCH(1,1) model. Starting with the second estimation, one more variable is included in the model at each time. The estimation results are consistent with the previous data analysis. Volatility is significantly higher on storage report announcement days, Mondays, and winter days. In estimation-5, the time to maturity (TTM) variable is significant and has the correct negative sign. So, the volatility of futures returns increases as the maturity gets closer. However, it lost its significance in estimation-6 after TTMWIN is also included in the estimation. This is consistent with the previous idea that the maturity effect is present only during winter months. More formally, the coefficient of time to maturity variable during non-winter months is $\gamma_4$. On the other hand, during the winter months, it is $\gamma_4 + \gamma_5$. Therefore, testing for the conditional hypothesis that the maturity effect is present only during the winter requires a statistical test of the null hypothesis $H_0: (\gamma_4 = 0 \text{ and } \gamma_4 + \gamma_5 < 0)$. The $t$ statistic for $\gamma_4$ is -0.15. Also, the $t$ statistic for $\gamma_4 + \gamma_5$ is calculated using the variance-covariance matrix of estimated parameters and reported in Table 4 as -8.416. This confirms the null hypothesis that the maturity effect is present only in winter. In the final estimation, the non-significant TTM variable is dropped from the estimation.
<table>
<thead>
<tr>
<th>Model</th>
<th>$\omega$</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$\gamma_1$</th>
<th>$\gamma_2$</th>
<th>$\gamma_3$</th>
<th>$\gamma_4$</th>
<th>$\gamma_5$</th>
<th>H-L</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>0.255</td>
<td>0.077</td>
<td>0.908</td>
<td>4.211</td>
<td>3.69</td>
<td>0.341</td>
<td>-0.047</td>
<td>-0.217</td>
<td>10.65</td>
</tr>
<tr>
<td></td>
<td>(3.53***</td>
<td>(8.62***</td>
<td>(3.99***</td>
<td>(6.00***</td>
<td>(6.13***</td>
<td>(6.98***</td>
<td>(1.645*</td>
<td>(6.73***</td>
<td></td>
</tr>
<tr>
<td>(2)</td>
<td>0.500</td>
<td>0.073</td>
<td>0.902</td>
<td>4.09</td>
<td>4.47</td>
<td>0.162</td>
<td>-0.047</td>
<td>-0.217</td>
<td>12.02</td>
</tr>
<tr>
<td></td>
<td>(-2.30**</td>
<td>(7.02***</td>
<td>(6.29***</td>
<td>(4.46***</td>
<td>(7.73***</td>
<td>(6.81***</td>
<td>(1.645*</td>
<td>(6.73***</td>
<td></td>
</tr>
<tr>
<td>(3)</td>
<td>-0.997</td>
<td>0.080</td>
<td>0.873</td>
<td>4.31</td>
<td>4.01</td>
<td>0.162</td>
<td>-0.047</td>
<td>-0.217</td>
<td>10.48</td>
</tr>
<tr>
<td></td>
<td>(-0.997)</td>
<td>(9.39***</td>
<td>(6.60***</td>
<td>(4.96***</td>
<td>(6.81***</td>
<td>(1.645*</td>
<td>(1.645*</td>
<td>(6.73***</td>
<td></td>
</tr>
<tr>
<td>(4)</td>
<td>-1.02</td>
<td>0.078</td>
<td>0.859</td>
<td>4.09</td>
<td>4.47</td>
<td>0.341</td>
<td>-0.047</td>
<td>-0.217</td>
<td>10.83</td>
</tr>
<tr>
<td></td>
<td>(-5.53***</td>
<td>(6.00***</td>
<td>(4.73***</td>
<td>(7.73***</td>
<td>(4.35***</td>
<td>(4.35***</td>
<td>(1.645*</td>
<td>(6.73***</td>
<td></td>
</tr>
<tr>
<td>(5)</td>
<td>-0.503</td>
<td>0.070</td>
<td>0.874</td>
<td>4.31</td>
<td>4.01</td>
<td>0.162</td>
<td>-0.047</td>
<td>-0.217</td>
<td>12.02</td>
</tr>
<tr>
<td></td>
<td>(-1.66*)</td>
<td>(5.61***</td>
<td>(4.96***</td>
<td>(6.81***</td>
<td>(1.645*</td>
<td>(1.645*</td>
<td>(1.645*</td>
<td>(6.73***</td>
<td></td>
</tr>
<tr>
<td>(6)</td>
<td>-1.06</td>
<td>0.073</td>
<td>0.863</td>
<td>4.69</td>
<td>4.44</td>
<td>2.40</td>
<td>-0.0029</td>
<td>-0.215</td>
<td>10.83</td>
</tr>
<tr>
<td></td>
<td>(-3.11***</td>
<td>(5.68***</td>
<td>(5.38***</td>
<td>(6.98***</td>
<td>(7.18***</td>
<td>(7.18***</td>
<td>(6.73***</td>
<td>(8.29***</td>
<td></td>
</tr>
<tr>
<td>(7)</td>
<td>-1.10</td>
<td>0.073</td>
<td>0.865</td>
<td>4.67</td>
<td>4.41</td>
<td>2.38</td>
<td>-0.0029</td>
<td>-0.215</td>
<td>10.83</td>
</tr>
<tr>
<td></td>
<td>(-5.50***</td>
<td>(5.68***</td>
<td>(5.35***</td>
<td>(7.30***</td>
<td>(9.38***</td>
<td>(9.38***</td>
<td>(8.29***</td>
<td>(8.29***</td>
<td></td>
</tr>
</tbody>
</table>

$t.stat(\hat{\gamma}_4 + \hat{\gamma}_5) = -8.416$ (In Estimation-6)
Significance Codes: * 10%, ** 5%, *** 1%

$$r_t = \sigma_t z_t$$
$$\sigma_t^2 = \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2 + \gamma_1 SDDAY_t + \gamma_2 MON_t + \gamma_3 WIN_t + \gamma_4 TTM_t + \gamma_5 TTMWIN_t$$

Table 4: GARCH Estimation Results For Nearby Month Futures Returns
A high level of persistence in natural gas futures volatility, as measured by the sum $\alpha + \beta$, is evident in these estimations. The volatility literature suggests that the persistence in volatility might be the result of driving exogenous variables that are persistent themselves. Therefore, such variables should reduce the level of volatility persistence once they are included in the conditional variance equation of GARCH models. This kind of behavior is observed in the parameter estimates with exogenous variables in the volatility equation. While $\alpha + \beta = 0.985$ in the simple GARCH estimation, it reduced to 0.938 in the final estimation. The half-life of a volatility shock is defined as the time it takes for half of the shock to vanish and is given by:

$$\text{Half Life} = \frac{\log(0.5)}{\log(\alpha + \beta)}$$

The half-lives estimated for all models are presented in the final column of Table 4. After including the persistent covariates, the half-life of a volatility shock decreased to 10.83 days from the 45.86 days in the simple GARCH model.

### 3.2 Storage and Weather Modeling

In order to incorporate the asymmetric effects of storage levels during different seasons, two additional variables are constructed: $SD_t$ and $SDWIN_t$. The first variable $SD_t$ is the same variable constructed in Section 2. It is the deviation of storage level from its five-year historical average. Once $SD_t$ is calculated for a storage report announcement day, the same number is used on the following days until the next storage report announcement. This is different from the previous literature that included the storage surprise variable only for the announcement days (Mu 2007; Gregoire and Boucher 2008). This is because we regard this variable not only as a proxy for storage surprise but as a proxy for supply tightness during winter and as a proxy for tightness in storage space supply in other seasons. The second variable $SDWIN_t$ is a proxy for supply tightness during winter. It is constructed as $SDWIN_t = SD_t \times WIN_t$. This variable enables us to model the asymmetric effect of storage levels for different seasons. The expectation is a positive coefficient for $SD_t$ and a higher negative coefficient for $SDWIN_t$ to confirm the hypothesis that low storage levels increase the volatility in winter, whereas high storage levels increase the volatility at other times.
The weather modeling is accomplished by the well-known degree day variables. These are quantitative indices used to reflect the demand for energy. Experience shows that there is no need for heating or cooling if the outside temperature is $65^\circ F$. Consequently, Heating Degree Days and Cooling Degree Days variables are defined as:

$$
HDD_t = \max(0, 65 - T_{ave,t})
$$
$$
CDD_t = \max(0, T_{ave,t} - 65),
$$

where $T_{ave,t}$ is the average of the maximum and minimum observed temperature on day $t$. There are two common ways of modeling weather shocks, either with ex-post forecast errors or with temperature anomalies, defined as the deviation of degree days variables from their seasonal norms (Mu 2007). Here, we follow the second approach because the forecast data are not available. The following weather shock variables are constructed: $HDD.Shock_t$: This is defined as the deviation of Heating Degree Days from the seasonal norms over the forecasting horizon. Following Mu (2007), the forecasting horizon is chosen to be seven days since the weather forecasts from the public media are typically broadcast for seven days ahead.\(^9\)

$$
HDD.Shock_t = \sum_{i=t+1}^{t+7} (HDD_i - HDD.Norm_i),
$$

where $HDD.Norm_t$ is the historical 30-year average of HDD on day $t$. The historical 30-year average is the definition of the National Weather Service (NWS) for the seasonal norm. Since we do not have the actual forecast data, the realized $HDD$ is used in creating this variable. $CDD.Shock_t$: This is defined as the deviation of CDD from the seasonal norms and calculated in the same way as:

$$
CDD.Shock_t = \sum_{i=t+1}^{t+7} (CDD_i - CDD.Norm_i),
$$

where $CDD.Norm_t$ is the 30-year historical average of CDD on day $t$.

Both weather variables are constructed for Chicago, New York, Atlanta, and Dallas. Then, the natural gas consumption weighted average of these locations is calculated.\(^{10}\) These national averages for

\(^9\)Results are robust to the choice of a forecasting horizon as 8, 9, or 10 days.

\(^{10}\)We used the same weights used in Mu (2007): 0.42 for Chicago, 0.28 for New York, 0.17 for Atlanta, and 0.13 for Dallas.
weather shock variables are plotted in Figure 3. Heating degree day shocks are closer to zero in summer months, whereas cooling degree day shocks are closer to zero in winter months since both the thirty-year historical averages and the actual realizations get closer to zero.

Estimation results including the storage and weather variables are presented in Table 5. In estimation-8, only the two storage variables are added to the last estimation in Table 4. Both variables have the correct sign and are significant at all conventional levels. The positive sign for the coefficient of $SD_t$ suggests that high storage levels increase the short-term volatility during the non-winter period. Also, the negative coefficient for $SDWIN_t$ is greater than the positive coefficient of the $SD_t$, which suggests that it is the low storage levels resulting in high volatility during winter months. This asymmetric effect is tested more formally later in estimation-11.
Figure 3: National Weather Shock Variables
\[
t_{\text{stat}}(\hat{\gamma}_5 + \hat{\gamma}_6) = -3.094 \quad \text{(In estimation-11)}
\]

Significance Codes: * 10%, ** 5%, *** 1%

\[
\begin{align*}
    r_t &= \sigma_t z_t \\
    \sigma_t^2 &= \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2 + \gamma_1 SDDAY_t + \gamma_2 MON_t + \gamma_3 WIN_t + \gamma_4 TTMWIN_t + \gamma_5 SD_t + \gamma_6 SDWIN_t + \gamma_7 HDD.Shock_t + \gamma_8 CDD.Shock_t
\end{align*}
\]

Table 5: GARCH Estimation Results with Storage and Weather Variables For Nearby Month Futures Returns
In estimation-9, only the two weather variables are added to the last estimation in Table 4. They are both significant at the 10 percent confidence level with a correct positive sign. Higher degree days than the seasonal norms increase short-term volatility. One unexpected result is that the significance of $CDD.Shock$ is stronger than the significance of $HDD.Shock$. This might be due to the other variables accounting for higher volatility in winter. If the winter dummy variable and its interaction term with the maturity were excluded from the model as in estimation-10, the significance of $HDD.Shock$ becomes stronger than the significance of $CDD.Shock$. In this estimation, HDD is significant at the 1 percent level and CDD is significant at the 5 percent level. Lastly, in estimation-11, we include the storage and weather variables together. The inference for the storage variables remains the same. The asymmetric effect of storage deviation variable $SD_t$ across seasons can be formally tested with the null hypothesis of $H_0 : (\gamma_5 > 0 \text{ and } \gamma_5 + \gamma_6 < 0)$. The $t$ statistic for $\gamma_5$ is 3.34. Also, the $t$ statistic for $\gamma_5 + \gamma_6$ is calculated using the variance-covariance matrix of estimated parameters and reported in Table 5 as -3.094. This confirms the asymmetric effect of the storage variable across seasons. As for weather variables, they still have the correct positive signs, but they lost their significance after the addition of the storage variables. Note that the storage variables were included in the model as a proxy for supply, and the weather shocks were included as a proxy for demand in the natural gas market. However, storage levels could be thought of as the result of the combination of supply and demand forces, thereby providing an explanation for the reduction in the significance of weather variables. The winter dummy is another factor that reduces the significance of weather variables due to nonorthogonality as discussed before.

4 Out-of-Sample Forecast Accuracy

Recent papers employing GARCH models to investigate the effects of natural gas market fundamentals on the volatility dynamics of natural gas futures report results only for in-sample estimations. These estimation results provide valuable information for understanding the volatility dynamics of natural gas futures. However, the out-of-sample predictive power of these augmented GARCH models has not been tested. In this section, day-ahead volatility predictions are made for
natural gas nearby month futures using simple GARCH models, as well as their augmented counterparts, and the accuracy of these forecasts are compared.

The empirical methodology followed here is known as a sliding window scheme. To make a prediction for day \( t \) where \( t \in \{501, 502, \ldots, T\} \), only the returns \( \{r_{t-1}, r_{t-2}, \ldots r_{t-500}\} \) are used. So, the length of the sliding window is chosen as 500 observations. That is, returns 1 through 500 are used to predict the volatility for day 501; returns 2 through 501 are used to predict the volatility for day 502, and so on. Since \( T = 1,823 \), there are \( T - 500 = 1,323 \) volatility predictions. One problem in out-of-sample forecasting is that the variables \( \text{HDD.Shock}_t \) and \( \text{CDD.Shock}_t \) are using information from the future. On day \( t \), the volatility for day \( t + 1 \) is being forecasted, but at that time these variables are not available yet in the information set. To solve this problem, first weather shock forecasts are obtained by fitting an ARIMA(1,2,1) model to the last 500 calendar days of temperature data for all four cities. The ARIMA(1,2,1) is chosen based on the Schwarz information criterion (SIC). The natural gas consumption weighted average of weather shock variables across the four cities is calculated as the final weather shock forecasts. Then, the weather shock forecasts are used in augmented GARCH models to forecast the volatility. A simpler approach is to use the appropriate lags of weather shock variables. This can be regarded as forecasting the next seven days’ weather shocks as being equal to the last seven days’ weather shocks. The presented results are from the ARIMA forecasting approach, but using a simpler lag approach provides the same results.

4.1 Other Simple Models for Forecasting

Random Walk Model: 
With a random walk assumption, the volatility forecast for day \( t \) is the realized volatility on day \( t - 1 \). It is used as the benchmark model.

\[
\hat{\sigma}_t = |r_{t-1}| .
\] (10)

Moving Average Model: 
Moving average models are widely used by natural gas market practitioners. In this paper, 20-day and 60-day moving averages are used
that correspond to one-month and three-month trading days.

\[
\hat{\sigma}_t = \sqrt{\frac{1}{m} \sum_{i=1}^{m} r_{t-i}^2}.
\]  

(11)

Implied Volatility:

Annualized implied volatilities for the closest-to-the-money call options are obtained from the trading floor of a very active natural gas trading firm. Dividing annualized implied volatility by \(\sqrt{250}\), the daily implied volatilities are obtained. Forecasting can be done as follows:

\[
\hat{\sigma}_t = \frac{\sigma_{imp,t}}{\sqrt{250}}.
\]

(12)

4.2 Forecast Accuracy Results

After making the forecast and observing the realized volatility the next day, the volatility forecast error can be defined as

\[
FE = \sigma_t - \hat{\sigma}_t
\]

\[
FE = |r_t| - \hat{\sigma}_t,
\]

where \(\hat{\sigma}_t\) is the forecasted volatility, and \(\sigma_t\) is the realized volatility for day \(t\). The latter equality follows because, in the absence of intraday data, the most common approach in the literature is to use the absolute value of return as the realized volatility. Three statistical measures are used for measuring forecast accuracy. These are mean absolute error (MAE), mean squared error (MSE), and Theil’s U statistic.

\[
MAE = \frac{1}{n} \sum_{t=1}^{n} |r_t| - \hat{\sigma}_t|,
\]

(13)

\[
MSE = \frac{1}{n} \sum_{t=1}^{n} (|r_t| - \hat{\sigma}_t)^2,
\]

(14)

\[
Theil's\ U = \sqrt{\frac{\sum_{t=1}^{n} (|r_t| - \hat{\sigma}_t)}{\sum_{t=1}^{n} (|r_t| - |r_{t-1}|)}}.
\]

(15)

Theil’s U statistic can be thought as a relative accuracy measure. It is the ratio of the root mean squared error of the chosen model to
the root mean squared error of the random walk model. Out-of-sample forecasting accuracy measures for the simple and augmented GARCH models as well as other simple forecasting methods are presented in Table 6.

<table>
<thead>
<tr>
<th>Model</th>
<th>MAE</th>
<th>Rank (MAE)</th>
<th>MSE</th>
<th>Rank (MSE)</th>
<th>Theil’s U</th>
<th>Rank (Theil’s U)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Garch(1,1)</td>
<td>1.984</td>
<td>6</td>
<td>6.284</td>
<td>5</td>
<td>0.7908</td>
<td>5</td>
</tr>
<tr>
<td>Model (7)</td>
<td>2.005</td>
<td>8</td>
<td>6.366</td>
<td>6</td>
<td>0.7959</td>
<td>6</td>
</tr>
<tr>
<td>Model (8)</td>
<td>1.931</td>
<td>3</td>
<td>6.016</td>
<td>2</td>
<td>0.7737</td>
<td>2</td>
</tr>
<tr>
<td>Model (9)</td>
<td>1.970</td>
<td>5</td>
<td>6.271</td>
<td>4</td>
<td>0.7899</td>
<td>4</td>
</tr>
<tr>
<td>Model (11)</td>
<td>1.8913</td>
<td>2</td>
<td>5.974</td>
<td>1</td>
<td>0.7710</td>
<td>1</td>
</tr>
<tr>
<td>MA(20) Model</td>
<td>1.8730</td>
<td>1</td>
<td>6.195</td>
<td>3</td>
<td>0.7852</td>
<td>3</td>
</tr>
<tr>
<td>MA(60) Model</td>
<td>1.963</td>
<td>4</td>
<td>6.459</td>
<td>8</td>
<td>0.8017</td>
<td>8</td>
</tr>
<tr>
<td>Implied Vol</td>
<td>2.004</td>
<td>7</td>
<td>6.411</td>
<td>7</td>
<td>0.7987</td>
<td>7</td>
</tr>
<tr>
<td>Random Walk</td>
<td>2.239</td>
<td>9</td>
<td>10.049</td>
<td>9</td>
<td>1</td>
<td>9</td>
</tr>
</tbody>
</table>

Table 6: Accuracy Statistics for Out-of-Sample Forecasting

Model 7 is based on the estimation-7 in Table 4. The forecasting accuracy of this model is slightly worse than the simple GARCH model giving marginally higher MAE, MSE, and Theil’s U statistics. Model 8 and 9 are based on estimations 8 and 9 in Table 5. Both models perform better than the simple GARCH model. Moreover, storage variables in model 8 improve the performance much more than the weather variables in model 9. Lastly, model 11 includes all variables and among all the GARCH models, it provides the lowest MAE, MSE, and Theil’s U statistics, making a great improvement on the simple GARCH model.

So, the evidence is not conclusive that any augmentation of the GARCH models would increase the accuracy in out-of-sample forecasting as in the case of model 7. However, with carefully chosen variables to account for natural gas market fundamentals, it is possible to increase the forecast accuracy as in models 8, 9, and 11.

The simple forecasting schemes generally perform very poorly with the exception of the 20-day moving average. MA(20) method ranks first with a very small margin in terms of MAE and third in terms of MSE and Theil’s U. The MSE measure penalizes models with the square of their errors, and the MA(20) does not perform as good as with the linear penalty function used in MAE calculation. Therefore,
this simple model is producing very good forecasts in general, but once it is wrong it is way of the target. On the other hand, model 11 that includes all fundamental variables within a GARCH framework ranks second in terms of MAE and first in terms of MSE and Theil’s U, thereby consistently providing good forecast accuracy.

5 Conclusion

Recently, a new literature emerged on modeling short-term volatility dynamics of natural gas futures. This research focuses on the augmentation of GARCH models and its variants with the natural gas market fundamentals in order to understand the sources of high volatility in natural gas prices. In this paper, several new findings contributing to this literature have been presented, and more importantly forecasting the accuracy of these models is analyzed for the first time.

First of all, the effect of storage levels on short-term volatility is asymmetric across the seasons. During the winter months, lower storage levels than the five-year historical average were found to be increasing the short-term volatility. In contrast, it is the high levels of storage causing excess volatility in other seasons. This can be attributed to the changing concerns of market players at different seasons. In the winter, low storage levels are perceived as a tight supply situation causing excess volatility. At other times, the market is mainly concerned about the storage space supply. Therefore, high levels of storage cause excess volatility.

Secondly, the maturity effect for natural gas nearby month futures is found to be a significant determinant of volatility only in the winter months. This result is confirmed by both data analysis and econometric estimation of the GARCH models, including the maturity variable and its interaction with the seasonality in the volatility equation. Since winter is the season when demand is highly inelastic, traders might be overreacting to new information arrival closer to the maturity and this causes excess volatility.

In addition to these new findings, this study confirms some of the previous results in the literature. Higher volatility is observed on Mondays possibly due to the accumulation of weather information over the non-trading weekend. Storage report announcement days also exhibit higher volatility than other days since new arriving information is priced very fast in this case. Volatility on winter days,
defined as December, January, and February, is found to be higher than other seasons. Lastly, weather shocks in excess of the seasonal norms increase the short-term volatility. However, when storage variables and the winter dummy is included in the model, they take the significance of weather variables. This might be because the storage variables are taking care of both supply and demand dynamics and/or non-ortgonality with the winter dummy.

As for forecasting accuracy, the augmented GARCH models with carefully chosen fundamental variables have the potential to decrease MAE, MSE, and Theil’s U statistics. The model using storage and weather variables in addition to other variables capturing the Monday, storage report announcement, winter, and maturity effects provides the best forecasting accuracy, thereby greatly improving on the simple GARCH model forecasts. This is a very important finding because better volatility forecasts can be used in option pricing and hedging natural gas exposures. Fleming et al. (2001) suggests focusing on the economic significance of time varying predictable volatility instead of evaluating the statistical performance of volatility models. Future research can be conducted in such applications of augmented GARCH models for natural gas volatility.

Lastly, we found that a simple GARCH model provides very good backtesting result in risk estimation, and there is no room for improvement by augmenting the model with market fundamentals. This is because natural gas futures return distribution does not exhibit very fat tails. Overall, the results suggest that the volatility forecasting performance can be increased by augmentation of the GARCH models, whereas a simple GARCH model can be preferable for the risk measurement of linear portfolios considering the simplicity advantage.
References


