Parameter Variation and the Components of Natural Gas Price Volatility

Dr. Matthew Brigida

Department of Finance
College of Business Administration
Clarion University of Pennsylvania
Clarion, PA 16214
mbrigida@clarion.edu

October 8, 2014

Introduction

Estimating a static coefficient for a deseasoned gas storage or weather variable implicitly assumes that market participants react identically throughout the year (and over each year) to that variable. This means they assume:

- the variable is no more important in winter.
- market participants do not adapt.

These are unrealistic assumptions of economic behavior.

Further, natural gas uncertainty is likely not simply attributable to the regression error term, but rather also due to changes in how market participants link prices to storage and weather (the coefficients), and the uncertainty in these parameter estimates.

Introduction

In this analysis we will model natural gas price returns as a linear function of gas storage and weather variables, and we will allow the coefficients of this function to vary continuously over time (TVP model).

- TVP model will allow market participants to adapt their reactions to information contained in variables, and also afford an estimate of conditional heteroskedasticity due to both parameter uncertainty and a standard error term.
- We will use this to also estimate a time series of the proportion of total volatility attributable to each independent variable.
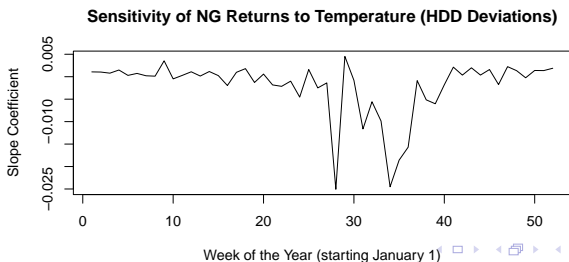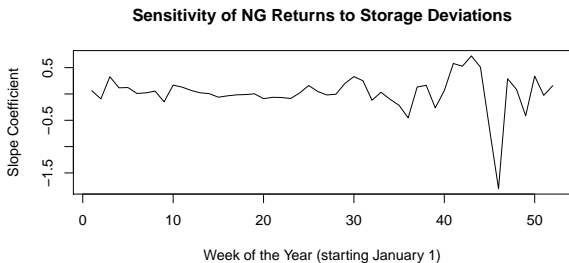
Preliminary Evidence

We first use a series of OLS regressions to investigate whether:

- the sensitivity of natural gas prices to changes in storage and the weather (HDD) is time-varying.
- the proportion of natural gas volatility attributable to these variables time-varying.

To do so we first calculate the average natural gas price return, storage deviation, and HDD deviation for each week of the year. We then run 52 (one for each week) OLS regressions,
$ng_{w,i} = \beta_0 + \beta_1 Stor_{w,i} + \beta_2 HDD_{w,i} + e_i$ where w denotes a particular week of the year, and i ranges over the 14 years in our sample.

Figure: Sensitivity of natural gas returns to deviations in storage and HDD. The sensitivities were estimated as slope coefficients from linear regressions $ng_{w,i} = \beta_0 + \beta_1 Stor_{w,i} + \beta_2 HDD_{w,i} + e_i$ where w denotes a particular week of the year, and i ranges over the 14 years in our sample.

Sensitivity of NG Returns to Storage Deviations

Slope Coefficient

Week of the Year (starting January 1)

Sensitivity of NG Returns to Temperature (HDD Deviations)

Slope Coefficient

Week of the Year (starting January 1)

Figure: Proportion of natural gas return volatility attributable to storage and HDD. The proportions were estimated from linear regressions $ng_{w,i} = \beta_0 + \beta_1 Stor_{w,i} + \beta_2 HDD_{w,i} + e_i$ where w denotes a particular week of the year, and i ranges over the 14 years in our sample.

Proportion of Volatility from Storage and Temperature
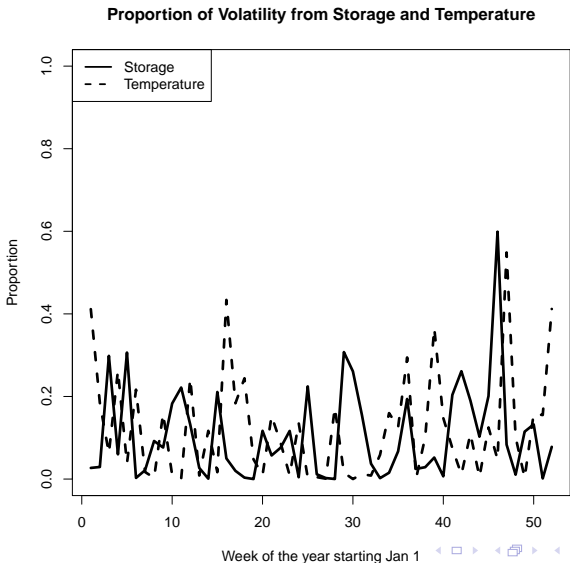
Parameter Stability Tests

We use a test to detect departures from constancy in time-series regression relationships proposed by Brown, Durban, and Evans (1975). The specific test is referred to by the authors as the 'homogeneity test'[1].

- Null hypothesis of the test is that the regression parameters are equal at each time point.

---

[1]found in the section: *2.5. Moving Regressions*

Parameter Stability Tests

The sample period is split into nonoverlapping intervals of arbitrary length n, and the 'between group over within groups' ratio of mean sum of squares is calculated as the test statistic.

- Under H_0 the test statistic is distributed as $F(kp - k, T - kp)$ where k is the number of regressors, p the number of intervals, and T is the number of observations.

Applying this test to $ng_t = \beta_0 + \beta_1 Stor_t + \beta_2 HDD_t + e_t$, for values of n ranging from 20 to 50, we are able to reject the null at the 5% level of significance for all n

Result

We reject the null, which is stable regression coefficients.

Engle and Watson (1985) suggest a random walk in cases where market participants adjust their estimate of the state only on the arrival of new information.

- In natural gas markets, however, participants will also likely adjust parameter estimates based on season.

Structural form of the parameters

To test whether the coefficients are random walks we will estimate $ng_t = \beta_0 + \beta_1 Stor_t + \beta_2 HDD_t + e_t$ using OLS. We also estimate separate equations for each independent variable. We then use that one-half times the regression sum of squares of $\frac{\hat{e}_t^2}{\sigma_e^2} = \gamma_0 + \gamma_1 t(x_t^2) + \mu_t$ where x is the vector of independent variables, is distributed $\chi^2(k)$ under the null of stable coefficients.

- The alternative hypothesis is that the OLS regression exhibits heteroskedasticity consistent with random walk coefficients.

Structural form of the parameters

Using this test we are able to reject the null for the HDD coefficient at the 5% level. We do not reject the null for the storage coefficient ($Stor$). Considering this with the results of the previous section implies the $Stor$ coefficient is time-varying though not in a random walk fashion.

- Jointly testing all regression parameters, we also do not reject the null.

TVP Model

Measurement equation:

$$ng_t = \beta_{0,t} + \beta_{1,t}Stor + \beta_{2,t}HDD + e_t, \quad e_t \sim N(0, \sigma_e^2) \qquad (1)$$

and for coefficient n we let the transition equation take the form of a random walk:

$$\beta_{n,t} = \beta_{n,t-1} + \xi_{n,t}, \quad \xi_t \sim N(0, \sigma_{\xi_n}^2) \qquad (2)$$

Note ng_t denotes log returns in natural gas futures prices, and *Stor* and *HDD* represent deviation from normal storage, and heating degree days respectively.

Estimation of the model is done using the Kalman Filter and Prediction Error Decomposition. The likelihood function was maximized using the *optim* function in the *R* (2014) programming language.

Model Results

Estimation of the model affords:

- Time-varying parameters: $\beta_{n,t|t-1}$ and $\beta_{n,t|t}$
- The conditional variance: $H_{t|t-1} = x_{t-1} P_{t|t-1} x'_{t-1} + \sigma_e^2$
  where x_{t-1} is the vector of explanatory variables (*Stor* and
  HDD), $P_{t|t-1}$ is the variance-covariance matrix of the time
  varying regression coefficients ($\beta_{t|t-1}$), and σ_e^2 is the variance of
  the disturbance term.

Components of Volatility

To calculate the proportion of natural gas volatility attributable to a particular variable (*Stor* or *HDD*) we:

- zero out the variable in x_{t-1} and any row or column in $P_{t|t-1}$ which involves that variable.
- We then recalculate $H_{t|t-1}$, which affords the conditional variance of natural gas prices without that variable.
- The difference between the full conditional variance and the conditional variance without the variable, affords the conditional variance attributable to that variable.

Serial Dependence:

- We test the heteroscedasticity-adjusted one-period-ahead forecast errors, $H_{t|t-1}^{-\frac{1}{2}}\eta_{t|t-1}$, for serial dependence using the Box-Pierce and Ljung-Box tests. We run tests for lags from 1 to 52 weeks. Both tests do not reject the null of no serial dependence for all lags.

ARCH Effects:

- we use both the Lagrange Multiplier test of Engel (1982) and the Ljung-Box test on the squares of the heteroscedasticity-adjusted one-period-ahead forecast errors $\left(H_{t|t-1}^{-\frac{1}{2}} \eta_{t|t-1} \right)^2$.

- The tests disagree:
  - Ljung-Box test failing to reject the null of no ARCH effects
 - the Lagrange Multiplier rejecting the null

We therefore conclude there is evidence of ARCH effects.

Results Summary

The model was estimated using varying initial parameters and the maximum log-likelihood over the many estimations was 1650.

- The standard deviation of the error term in the measurement equation (σ_e) is 5.49%.
- The standard deviations of the error terms in the transition equations (σ_{ξ_n}):
  - intercept: 0.0015
  - *Stor*: 0.4811
  - *HDD*: 0.0001

Figure: Below are plots of the Kalman filtered estimated coefficients. The plots are over the full sample period.
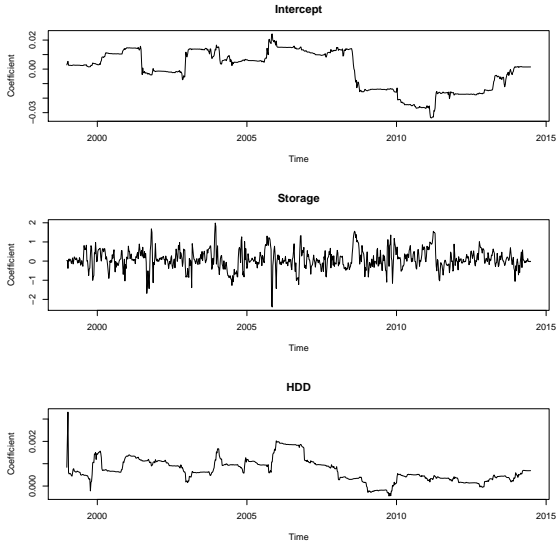
Results: Coefficients

- The storage coefficient appears to be a stationary series, and has a mean of 0.08. The range of variation through the seasons is from -2.40 to 1.99.
- The weather coefficient shows some seasonal variation, however the mean of this coefficient seems to vary with time (i.e. nonstationary).
  - For the period 1999 to 2007, the mean was 0.0010. For the period 2008 to 2014 the mean dropped to 0.0002. This 80% drop is evidence that market participants vary their reaction to underlying variables over multi-year periods, as well as throughout the year.
- Using the ADF test, the intercept and weather coefficients contain a unit root, however a unit root is rejected in the storage coefficient.

The mean forecast uncertainty from the TVP model is 10.60%, whereas the mean absolute value of natural gas returns is 5.31%. The unconditional standard deviation of returns from the GARCH(1,1) model is 8.45%.

- This shows, on average, there is more forecast uncertainty in natural gas returns than would be implied by the error term alone. That is, parameter uncertainty plays an important role in natural gas return volatility.

Figure: Weekly volatility measures. Below are measures of weekly volatility in natural gas returns, estimated over the full sample period.

Weekly Volatility
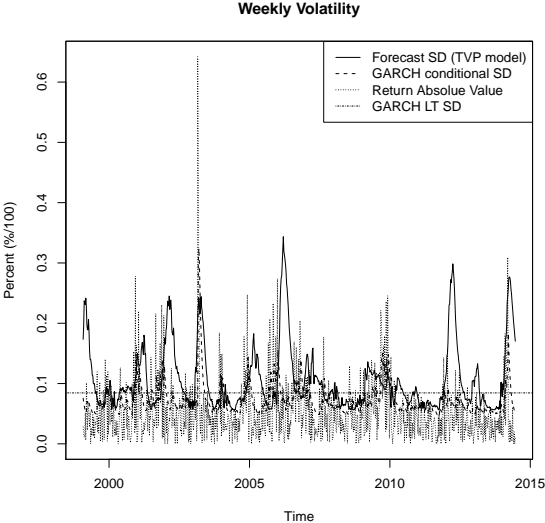
Results: Time-Varying Components of Volatility

Summer:

- Storage accounts for 50% of volatility with approximately 25% of the volatility coming from weather and the intercept term.

Winter:

- The proportion of forecast volatility due to weather often becomes the prime component of volatility (often accounting for 40% of the total) and the portion attributable to storage drops to around 30%.
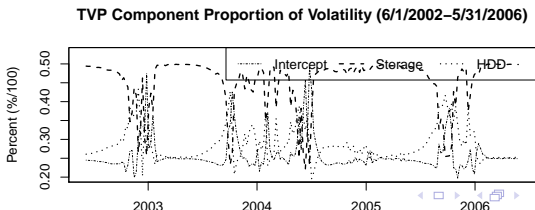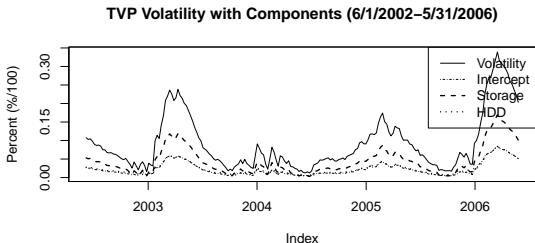
These results are consistent with common accounts of traders focusing on storage amounts during the summer injection season, as this is an indicator of whether there will be enough working gas in storage to meet winter demand.

Figure: Below are plots of total volatility (forecast uncertainty) with the proportions of that volatility attributable to each variable (top frame), and the proportion of total volatility attributable to each factor (bottom frame). The plots are over the full sample period.

TVP Volatility with Components: Full Sample

TVP Component Proportion of Volatility: Full Sample
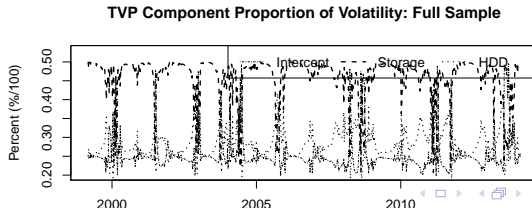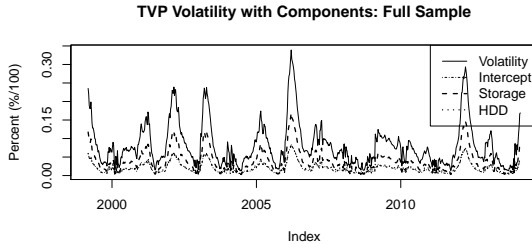
Figure: Below are plots of total volatility (forecast uncertainty) with the proportions of that volatility attributable to each variable (top frame), and the proportion of total volatility attributable to each factor (bottom frame). The plots are over a representative subsample period.

TVP Volatility with Components (6/1/2002−5/31/2006)

TVP Component Proportion of Volatility (6/1/2002−5/31/2006)

This analysis models how, in aggregate, market participants' parameters linking storage and weather evolve over time. This can help a market participants to understand their own adaptation process.

- The filtering algorithm affords forecasts of next week's coefficients, and coefficient uncertainly, as well as next week's natural gas return, given this week's data.

Application: Hedge Ratios

Power producers to hedge input prices by buying natural gas futures, and hedge demand risk by buying/selling weather derivatives (heating and cooling degree days).

- Say a company buys gas, and sells heating degree day futures. Let h_G and h_H denote the hedge ratio for the gas and HDD futures contracts respectively. To calculate the optimal hedge ratios the company will seek to minimize the variance of the combined position:

$$Var((h_G \Delta F_G - \Delta S_G) + (\Delta S_H - h_H \Delta F_H)) \qquad (3)$$

Proposition 1

The optimal hedge ratios which solve this minimization problem are (proof in the appendix):

$$h_G^* = \frac{\sigma_{F_H}^2 (Cov(\Delta F_G, \Delta S_H) - Cov(\Delta F_G, \Delta S_G)) + Cov(\Delta F_G, \Delta F_H)(Cov(\Delta F_H, \Delta S_G) - Cov(\Delta F_H, \Delta S_H))}{(Cov(\Delta F_G, \Delta F_H))^2 - \sigma_{F_G}^2 \sigma_{F_H}^2}$$

and

$$h_H^* = \frac{\sigma_{F_G}^2 (Cov(\Delta F_H, \Delta S_G) - Cov(\Delta F_H, \Delta S_H)) + Cov(\Delta F_G, \Delta F_H)(Cov(\Delta F_G, \Delta S_H) - Cov(\Delta F_G, \Delta S_G))}{(Cov(\Delta F_G, \Delta F_H))^2 - \sigma_{F_G}^2 \sigma_{F_H}^2}$$

We can see the optimal hedge ratios are functions of the variances and covariances of the changes in spot and futures prices.

- To estimate the effect of the proportion of natural gas uncertainty due to weather, we have estimated:

$$Cov(\Delta F_{G_t}, \Delta S_{H_t}) = \beta_0 + \beta_1 PropHDD_t + \mu_t$$

where $Cov(\Delta F_{G_t}, \Delta S_{H_t})$ is estimated from time $t-2$ to $t+2$ and $PropHDD_t$ is the average proportion from $t-2$ to $t+2$.

Application: Hedge Ratios

- The resulting estimate of the slope coefficient (β_1) is 5.59 and is significant at less than the 0.1% level.
- The proportion of natural gas uncertainty from HDD explains about 10% of the variation in $Cov(\Delta F_{G_t}, \Delta S_{H_t})$.

The proportion of natural gas price uncertainty from HDD has a strong positive effect on $Cov(\Delta F_{G_t}, \Delta S_{H_t})$, and thereby affects the hedge ratio.

This is evidence that the optimal hedge ratio will be affected over time by the proportion of volatility from the weather.

Conclusion

- In this analysis we have modeled natural gas returns explicitly allowing for market participants to learn over time, and to react differently to present changes in economic variables.
- we found the time series of the Kalman filtered estimates of the *Stor* coefficient did not contain a unit root. This implies that we can make inferences about future coefficient values.
- We also found evidence that the weather (*HDD*) coefficient did contain a unit root, and weather became a less important determinant of natural gas returns in 2007.

Conclusion

In an original application of the TVP model, we decomposed conditional volatility into a time series of each contributing factor to that volatility.

- This showed that storage is the dominant component of natural gas volatility throughout the year, with weather being the largest contributing factor only during periods in the winter.

Lastly, we showed that results of this analysis have particular applications to hedging and trading in natural gas markets.

Questions/Comments?

Appendix

Proof of Proposition 1: The variance of the combined position (V) is:

$$V = Var\left(h_G \Delta F_G - h_H \Delta F_H + \Delta S_H - \Delta S_G\right) =$$

$$Var\left(h_G \Delta F_G - h_H \Delta F_H\right) + Var\left(\Delta S_H - \Delta S_G\right) + 2Cov\left(h_G \Delta F_G - h_H \Delta F_H, \Delta S_H - \Delta S_G\right) =$$

$$= h_G^2 \sigma_{F_G}^2 + h_H^2 \sigma_{F_H}^2 - 2h_G h_H Cov\left(\Delta F_G \Delta F_H\right) + Var\left(\Delta S_H - \Delta S_G\right) +$$

$$+2\left(h_g Cov(\Delta F_G, \Delta S_H) - h_G Cov(\Delta F_G, \Delta S_G) - h_H Cov(\Delta F_H, \Delta S_H) + h_H Cov(F_H, S_G)\right)$$

Taking the partial derivatives of the variance of the combined position with respect to h_G and h_H and setting them equal to zero gives:

$$\frac{\partial V}{\partial h_G} = h_G \sigma_{F_G}^2 - h_H Cov(\Delta F_G, \Delta F_H) + Cov(\Delta F_G, \Delta S_H) - Cov(\Delta F_G, \Delta S_G) = 0$$

$$\frac{\partial V}{\partial h_H} = h_H \sigma_{F_H}^2 - h_G Cov(\Delta F_G, \Delta F_H) + Cov(\Delta F_H, \Delta S_G) - Cov(\Delta F_H, \Delta S_H) = 0$$

We then solve this system of equations for h_G and h_H. To do so write the system of equations as:

$$h_G A - h_H B + Q = 0$$

$$-h_G B + h_H I + R = 0$$

$$\Rightarrow h_H = \frac{h_G A + Q}{B}$$

$$\Rightarrow h_G = \frac{h_H I + R}{B}$$

Plugging h_H into h_G and solving for h_G:

$$h_G = \frac{\left(\frac{h_G A + Q}{B}\right) I + R}{B} \Rightarrow h_G^* = \frac{QI + RB}{B^2 - AI}$$

Plugging h_G back into h_H and solving for h_H:

$$h_H = \frac{\left(\frac{QI + RB}{B^2 - AI}\right) A + Q}{B} \Rightarrow h_H^* = \frac{QB + RA}{B^2 - AI}$$

Plugging back in for A, B, I, Q, and R affords the solution.