Parameter Variation & the Components of Natural Gas Price Volatility

Matthew Brigida*

September 14, 2014

Abstract

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. In this analysis we model natural gas returns as a linear function of gas storage and weather variables, and we allow the coefficients of this function to vary continuously over time. This formulation takes into account that market participants continuously try to improve their forecasts of market prices, and this likely means they continuously change the scale of their reaction to changes in underlying variables. We use this model to also calculate conditional natural gas volatility and the proportion of volatility attributable to each factor. We find that return volatility is higher in the winter, and this increase is attributable to increases in the proportion of volatility due to weather and natural gas storage. We provide time series estimates of the changing proportion of volatility attributable to each factor, which is useful for hedging and derivatives trading in natural gas markets.

*Associate Professor of Finance, Clarion University of Pennsylvania, 840 Wood St. Clarion PA 16214. matt@complete-markets.com

1 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. It assumes market participants find the variable no more meaningful in winter, and that they do not adapt to what has occurred in the market. These are unrealistic assumptions of economic behavior. Market participants continuously try to improve their forecasts of market prices, and this likely means they continuously change the scale of their reaction to changes in underlying variables. Moreover, 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.

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. This may be referred to as a time-varying-parameter (TVP) model, and it can be estimated using the Kalman filter. The 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[1]. We will use this to also estimate a time series of the

[1] Whereas an ARCH framework doesn't specify the source of the conditional heteroskedasticity

proportion of total volatility attributable to each independent variable. Such a procedure, as far as we know, is unique in the literature. Our results are useful for hedging and derivatives trading in natural gas markets.

Our procedure differs from previous work on volatility spillovers (Diebold and Yilmaz 2009 & 2012) in several and important ways. Firstly, they consider how total volatility in one variable spills over into another, whereas we measure the effect of volatility induced by uncertainty in the parameters of the function linking the variables. Second, to estimate a time series of volatility spillovers, they must use a rolling window approach. Because out method is based on the Kalman filter, time-varying estimates of parameter uncertainty are built in to the model.

The paper is organized as follows. Section 1.1 describes the data. Section 2 provides preliminary evidence for our TVP model. Section 3 tests for parameter instability in our equation, and the structural form of time-varying parameters. Section 4 introduces our main model and the calculation of conditional variances. Section 5 summarizes the results and performs diagnostic tests. Section 6 highlights applications and section 7 concludes.

1.1 Data

Natural gas price and storage data are from the U.S. Energy Information Administration and are available through the Administration's application programming interface. Heating Degree Day (HDD) data are from the U.S. National Oceanic and Atmospheric Administration (NOAA). We use HDD deviations from the norm to measure temperature fluctuations. This measures both warm ($HDD < 0$) and cold ($HDD > 0$) temperature[2]. All data are sampled at the weekly frequency.

2 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, and also whether 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. From these regressions we extract a set of coefficients for each week, as well as proportion of total natural gas volatility attributable to each independent variable.

A plot of the regression coefficients over each week of the year is in figure 1 below. Figure 2 shows the proportion of volatility from storage and weather for each week. What can be easily seen is that the sensitivity of natural gas returns to storage and the weather varies throughout the year. Further, the proportion of volatility from storage and the weather is generally higher

[2] We don't include deviations in cooling degree days because, in general, a large positive deviaton from the norm in CDD tends to coincide with a large negative deviation from the norm in HDD (the variables have a less than -0.50 correlation coefficient).

after the 40th week of the year. These initial results are consistent with our hypothesis, and motivate a more thorough examination of these time-varying relationships.

[INSERT FIGURES 1 AND 2 HERE]

3 Parameter Stability Tests

Prior to implementing the time-varying parameter model, we test firstly for parameter instability. If we can reject stability in the parameters, this is evidence in favor of time-varying parameters. We then investigate, the functional form of the state equation.

Similar to Kim and Nelson (1989), to test for parameter instability 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'[3]. The null hypothesis of the test is that the regression parameters are equal at each time point (stable regression coefficients).

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.

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

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 (most p-values are below 0.1%). We therefore reject the stability of the regression coefficients.

3.1 Structural form of the parameters

Also of interest, is the structural form of the time-varying 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.

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. This result is from Breusch and Pagan (1979), and this test is also employed in Kim and Nelson (1989).

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. This is consistent with our intuition that market participants are adjusting the coefficient based on season. Jointly testing all regression parameters, we also do not reject the null.

4 Time-Varying Parameter Model and Conditional Variance

The preliminary regression results and parameter stability tests are evidence in favor of parameters which vary with time, however it is likely that market participants (as rational economic agents) use past and present information when deciding on the appropriate present and future coefficients. This motivates a model where the coefficients are allowed to be updated in a Bayesian fashion when new information arrives, much like the views of market participants. An appropriate specification in this case is a time-varying-parameter model where the parameters are updated using the Kalman filter.

For the main model of or analysis, the measurement equation is:

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

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

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

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.

This time-varying-parameter model will estimate varying regression coefficients, but also affords an estimate of conditional volatility through the conditional variance of forecast errors from the Kalman filter (see Kim and Nelson (1989)). The present analysis further decomposes this conditional volatility into the contribution from each factor.

From the Kalman filter we estimate the conditional variance as $H_{t|t-1} = x_{t-1}P_{t|t-1}x'_{t-1}+\sigma_e^2$ 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.

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.

5 Results

5.1 Diagnostic Tests

We first 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.

To test for 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, however, with the Ljung-Box test failing to reject the null of no ARCH effects, and the Lagrange Multiplier rejecting the null. We therefore conclude there is evidence of ARCH effects[4]. This analysis will focus on the components of volatility from TVP, and leave the addition of ARCH effects, if any, to further analyses.

5.2 Results Summary

The model was estimated using varying initial parameters and the maximum log-likelihood over the many estimations was 1650. A chart of the Kalman filtered estimates of the time-varying regression coefficients is in figure 3 below. The standard deviation of the error terms in the measurement equation (σ_e) is 5.49%. The standard deviations of the error terms in the intercept,

[4] This may be a suitable avenue of further research

Stor, and HDD transition equations (σ_{ξ_n}) are 0.0015, 0.4811, and 0.0001 respectively (these reflect the units of independent variables).

Figure 3 shows the time series of each estimated slope coefficient of our model. While each coefficient shows substantial variation, the storage coefficient exhibits marked seasonal variation. The storage coefficient appears to be a stationary series, and has a mean of 0.08 and a 0.51. 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. 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. Further research may be able to determine why market participants became less sensitive to weather after 2007, perhaps it was a shift in overall winter weather severity.

To confirm the above we tested for a unit root in each of the coefficient series. Both the intercept and the weather coefficient contained a unit root. However, the storage coefficient rejected a unit root at the 1% level of significance. The augmented Dickey-Fuller test for a unit root was employed.

Figure 4 below shows a time series of weekly natural gas volatility using TVP and GARCH(1,1) models. 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.

In figures 5 and 6 we see the time varying components of natural gas forecast volatility. In the bottom panes we generally see that, in the summer, storage accounts for 50% of volatility with approximately 25% of the volatility coming from weather and the intercept term.

However in the 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%.

This shows the marked effect of winter on the drivers of natural gas volatility. Moreover, it is 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.

6 Applications

6.1 Application to Hedge Ratios

It is common for 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:

$$Var((h_G \Delta F_G - \Delta S_G) + (\Delta S_H - h_H \Delta F_H)) \tag{3}$$

Proposition 1 The optimal hedge ratios which solve this minimization problem are (proof in the appendix):

$$h_G^* = \frac{\sigma_{F_H}^2 \left(Cov(\Delta F_G, \Delta S_H) - Cov(\Delta F_G, \Delta S_G)\right) + Cov(\Delta F_G, \Delta F_H) \left(Cov(\Delta F_H, \Delta S_G) - Cov(\Delta F_H, \Delta S_H)\right)}{(Cov(\Delta F_G, \Delta F_H))^2 - \sigma_{F_G}^2 \sigma_{F_H}^2} \tag{4}$$

and

$$h_H^* = \frac{\sigma_{F_G}^2 \left(Cov(\Delta F_H, \Delta S_G) - Cov(\Delta F_H, \Delta S_H)\right) + Cov(\Delta F_G, \Delta F_H) \left(Cov(\Delta F_G, \Delta S_H) - Cov(\Delta F_G, \Delta S_G)\right)}{(Cov(\Delta F_G, \Delta F_H))^2 - \sigma_{F_G}^2 \sigma_{F_H}^2} \tag{5}$$

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 regressed the covariance of the log changes in natural gas futures prices with the changes in HDD on the proportion of natural gas price uncertainty from HDD, $Cov(F_{G_t}, S_{H_t}) = \beta_0 + \beta_1 PropHDD_t + \mu$, where $Cov(F_{G_t}, S_{H_t})$ is estimated from time $t-2$ to $t+2$ and $PropHDD_t$ is the average proportion from $t-2$ to $t+2$.

The resulting estimate of the slope coefficient (β_1) is 5.59 and is significant at less than the 0.1% level. From the regression we can see that the proportion of natural gas uncertainty from HDD explains about 10% of the variation in $Cov(F_{G_t}, S_{H_t})$. As we would expect, the proportion of natural gas price uncertainty from HDD has a strong positive effect on $Cov(F_{G_t}, S_{H_t})$, 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.

Importantly, this time-varying effect on the optimal hedge ratio cannot be captured simply by using seasonal covariances. This is because the proportion of uncertainty from the weather can fluctuate greatly from one year's winter to another.

6.2 Natural Gas Trading

The results of this analysis are useful to natural gas traders and other market participants. In fact, 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.

Importantly, 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. This also has clear implications for market participants, however more out-of-sample work in this area is needed prior to implementation in a trading environment.

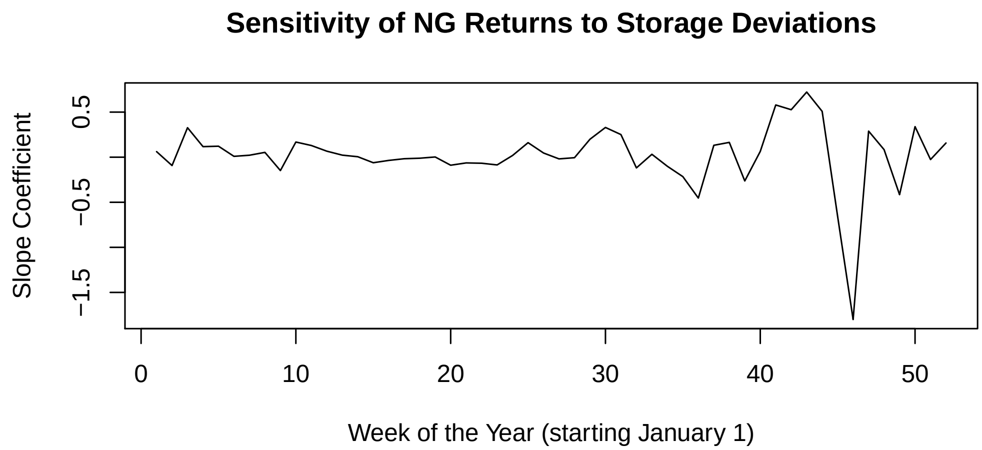
7 Conclusion

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. This learning and adaptation, and the attendant parameter uncertainty, constitutes another source of time varying conditional volatility.

In so doing we have found evidence of significant variation in the coefficients linking natural gas returns and its underlying fundamental factors. Further, 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. The modeling and out-of-sample prediction of future *Stor* coefficient values would be useful to include within future research. 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.

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.

Figure 1: 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.



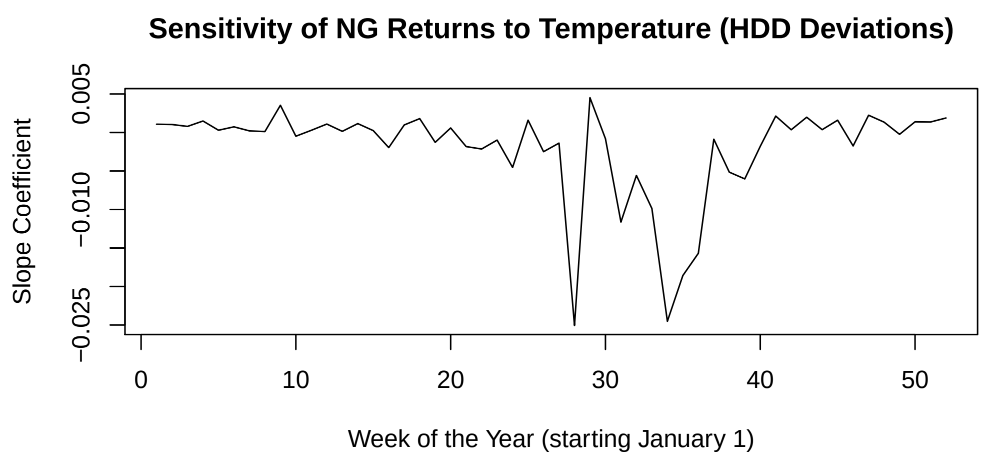


Figure 2: Proportion of natural gas return volatility attributable to storage and HDD. The proportions were estimated from linear regressions $ng_{w,i} = \beta_0 + \beta_1 Stor_{w,i} + \beta_2 HDD_{w,i} + e_i$ where w denotes a particular week of the year, and i ranges over the 14 years in our sample.

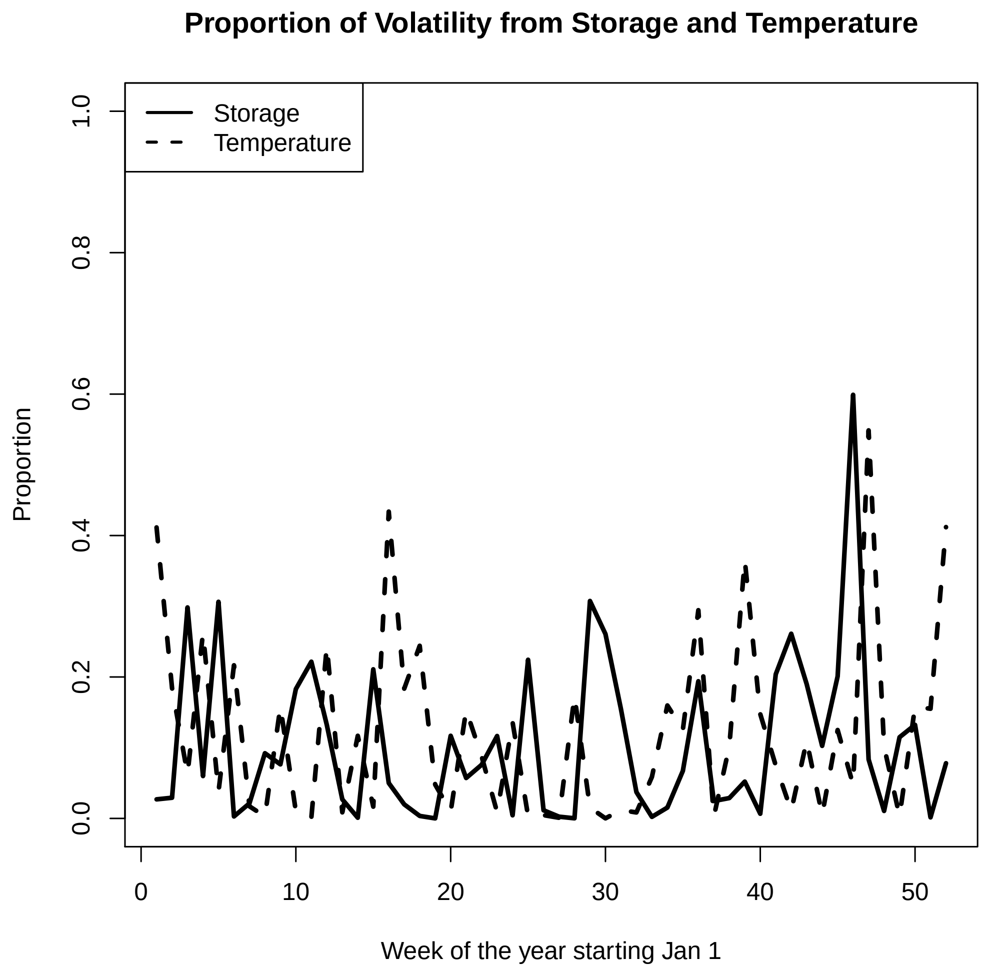


Figure 3: Below are plots of the Kalman filtered estimated coefficients. The plots are over the full sample period.

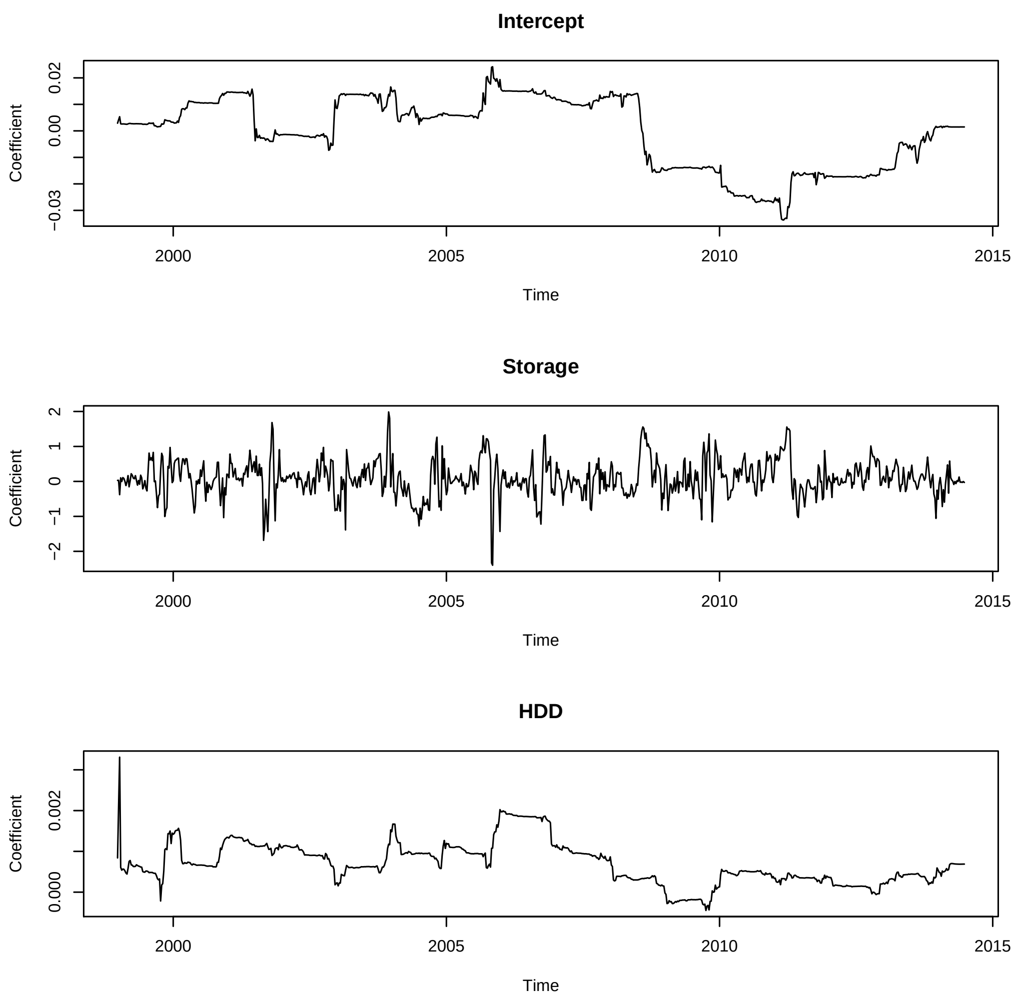


Figure 4: Weekly volatility measures. Below are measures of weekly volatility in natural gas returns, estimated over the full sample period.

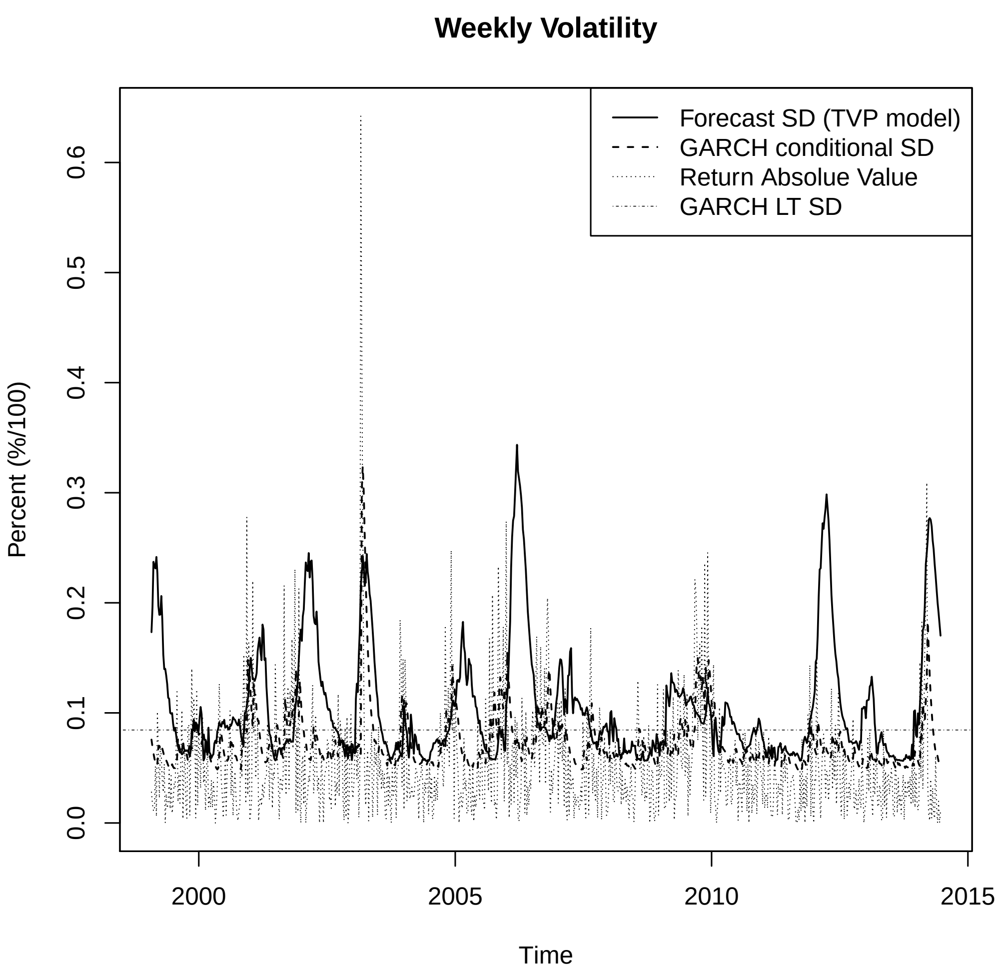
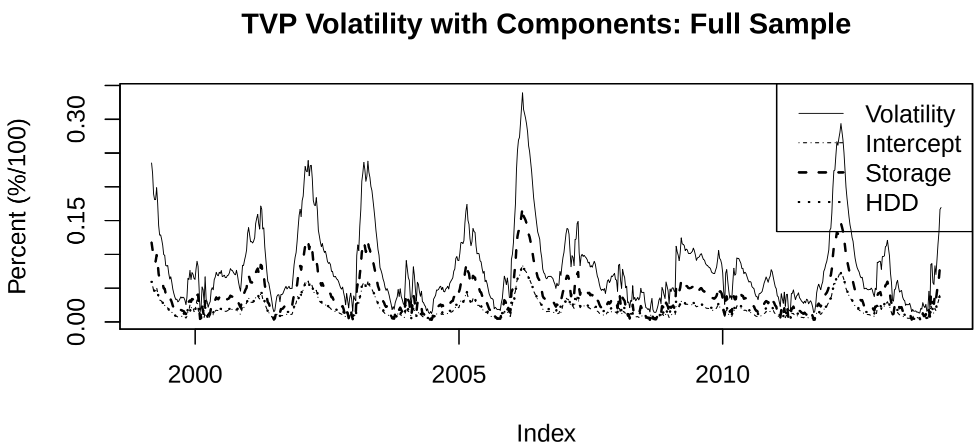


Figure 5: Below are plots of total volatility (forecast uncertainty) with the proportions of that volatility attributable to each variable (top frame), and the proportion of total volatility attributable to each factor (bottom frame). The plots are over the full sample period.



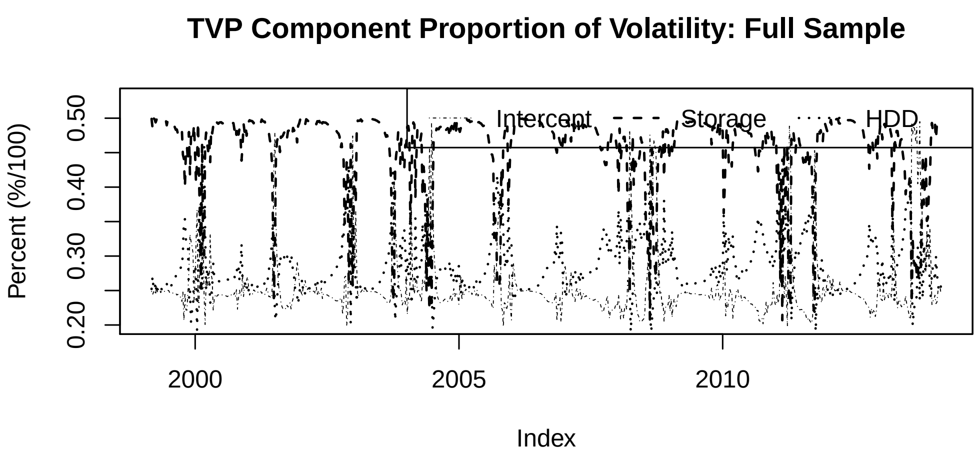
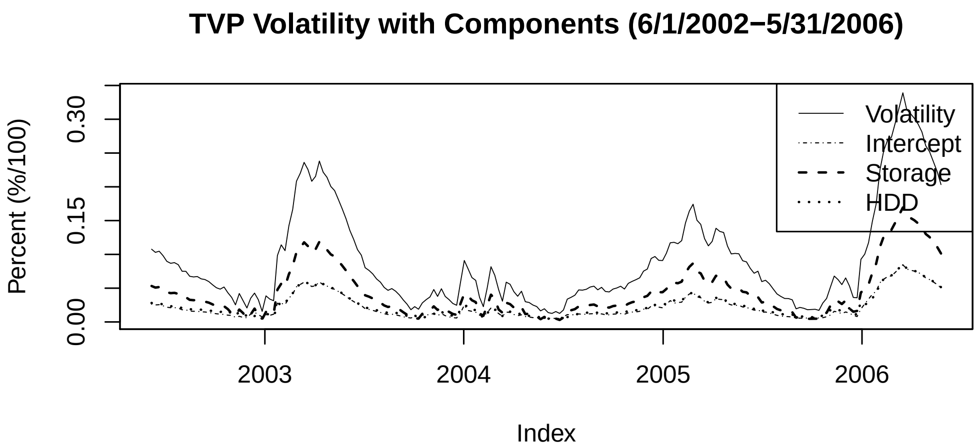
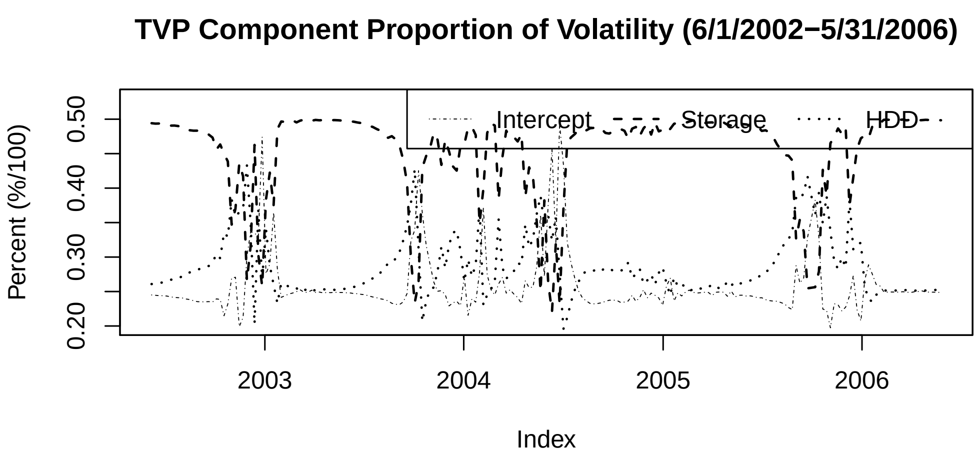


Figure 6: Below are plots of total volatility (forecast uncertainty) with the proportions of that volatility attributable to each variable (top frame), and the proportion of total volatility attributable to each factor (bottom frame). The plots are over a representative subsample period.





8 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.

References

Brown, R.L., Durbin, J., and Evans, J.M. (1975) "Techniques for Testing Constancy of Regression Relationships over Time," *Journal of the Royal Statistical Society*, Ser. B, 37, 149–192.

Diebold, F.X. and K. Yilmaz (2009). "Measuring financial asset return and volatility spillovers, with application to global equity markets". *The Economic Journal* 119(534), 158–171.

Diebold, F.X. and K. Yilmaz (2012). "Better to give than to receive: Predictive directional measurement of volatility spillovers" *International Journal of Forecasting* 28(1), 57–66.

Engle, R.F. and Watson, M.W. (1985) "Application of Kalman Filtering in Econometrics," unpublished report, World Congress of the Econometric Society, Cambridge, MA.

Kim, C.J., and C.R. Nelson (1989) "The time-varying-parameter model for modeling changing conditional variance: The case of the Lucas Hypothesis," *Journal of Business and Economic Statistics* 7, 433–440.

R Core Team (2014). "R: A language and environment for statistical computing. R Foundation for Statistical Computing," Vienna, Austria. URL http://www.R-project.org/.