RICE UNIVERSITY

Impact of News on Crude Oil Futures

by

Yu Han

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE:

Katherine B. Enor
Professor of Statistics

Barbara Ostdiek
Associate Professor of Finance and Statistics,
Senior Associate Dean of Jones Graduate School
of Business

David W. Scott
Noah Harding Professor of Statistics

Stuart M. Turnbull
Professor of Finance, Bauer College of
Business, University of Houston

Houston, Texas
April, 2017
ABSTRACT

Impact of News on Crude Oil Futures

by

Yu Han

Crude oil futures are the most actively traded commodity futures in the world, with more than 3 billion barrels per year in open interest. In part one of this thesis we use related news information to model the price dynamics of oil futures. We first examine the empirical patterns of oil market news data processed by Thompson Reuters News Analytics and of the intra day trading data of the WTI futures price traded on NYMEX. We find that news has significant impact on the returns and negative news has higher impact on returns and volatility clustering than does positive news. Motivated by these findings, we build a six-factor stochastic model for prices on the entire oil futures curve using spot prices, interest rates, convenience yields, stochastic volatilities, and positive and negative news events. The Kalman filter is applied to obtain quasi-maximum likelihood estimators. The estimation results show that news can significantly explain price movements, volatility clustering, skewness and kurtosis, and that negative news has a higher explanatory power for price dynamics than does positive news.

In the second part of this thesis, we develop a bivariate-EVT framework for the natural gas market. In the United States, spot natural gas is traded at more than one hundred hubs with different prices. All prices are based on a spread relative to the Henry Hub spot price. Yet due to the limited liquidity and data availability at most hubs, latest volatility models use realized variance calculated from intra day data. We adapt a modified realized beta GARCH model framework, using the Henry
Hub futures as the market factor, to model the price dynamics of individual spot trading hubs. We find that this model can better explain the price dynamics of individual hubs than does a standard GARCH model. We then develop an associated bivariate-EVT model for the tail risks, derive and calculate the expected shortfall measures. We compare our model to standard GARCH model, we find the new Realized Beta GARCH-Bivariate EVT model can better model tail risks.
# Contents

Abstract v
List of Illustrations vii
List of Tables ix

1 **Background** 1

2 **Thomson Reuters News Analytics** 7
   2.1 News Arrival Timestamps 7
      2.1.1 News Arrival Date 7
      2.1.2 News Arrival Time 11
   2.2 Relevance Fields 12
      2.2.1 RELEVANCE 12
      2.2.2 SENT_WORDS and TOT_WORDS 13
      2.2.3 MENTION_1 and TOT_SENTS 14
   2.3 Sentiment Fields 14
   2.4 Novelty Fields 16
   2.5 Relationship between Relevance, Sentiment and Novelty 16
   2.6 Newsscope Metadata: ITEM_TYPE 17

3 **Modeling the Price of Crude Oil Futures** 30
   3.1 Crude Oil Futures Price and Relationship with News 30
      3.1.1 Understanding the Crude Oil Futures 30
      3.1.2 Filter for Major News 33
      3.1.3 News and Futures Price Dynamics 35
   3.2 Our Model 48
3.2.1 Crude Oil Spot Price ................................................. 48
3.2.2 Interest Rate ......................................................... 49
3.2.3 Convenience Yield .................................................. 49
3.2.4 Stochastic Volatility ................................................. 50
3.2.5 News ................................................................. 50
3.2.6 Predicting Futures Price and Variance ......................... 51
3.2.7 Reduced Models without News ................................. 52
3.3 Estimation .............................................................. 53
   3.3.1 Estimating the News Arrival Process ......................... 53
   3.3.2 Estimating the Interest Race Parameters $a, m$ and $\sigma_r$ ... 55
   3.3.3 Use Kalman Filter to Estimate the Remaining Parameters ... 57
3.4 Fitting Result ......................................................... 61
   3.4.1 Parameter Estimation and Properties ......................... 61
   3.4.2 Price Fitting ...................................................... 62
   3.4.3 Conditional Variance ........................................... 62
3.5 Robustness ............................................................ 62
   3.5.1 Estimation on Sub-Samples ................................... 62
   3.5.2 Estimation Results for Different Sentiment Probability Thresholds ... 65
   3.5.3 Out of Sample Estimation .................................... 66
3.6 Conclusion ............................................................ 70

4 A RealGARCH and bivariate EVT Framework for Tail Risk Estimation ............................................. 71
   4.1 Background .......................................................... 71
   4.2 Fiting a modified Realized Beta GARCH Model .................. 72
   4.3 Maximum Likelihood Estimation for the Modified Realized Beta GARCH Model ............ 74
      4.3.1 Likelihood for Model of Market Factor ...................... 74
      4.3.2 Likelihood for Model of Individual Hubs .................. 75
   4.4 Data ................................................................. 76
   4.5 Realized Beta GARCH Model Fitting Result ..................... 79
4.6 Bivariate Extreme Value Theory and Expected Shortfall ...................... 83
  4.6.1 Derive the Distribution of Excess Residuals ......................... 83
  4.6.2 Derive Expected Shortfall for Returns at level $\pi$ .................... 85

4.7 EVT Result ........................................................................... 91
  4.7.1 Rolling Estimation ......................................................... 92
  4.7.2 Expected Shortfall ......................................................... 95

4.8 Conclusions .......................................................................... 96

Appendices .................................................................................. 98

A Futures Price ........................................................................... 99

B Higher Moments of Return ...................................................... 110

C Futures Realized Volatility ...................................................... 118
Illustrations

2.1 Daily Number of News Counting: 2004-2012 ........................................... 8
2.2 Distribution of Number of News Arrived within a Week: 2004-2012 .......... 9
2.3 Weekly News Arrival Counting: 2004-2012 ............................................. 10
2.4 Monthly Pattern of News Arrival Process: 2004-2012 ............................. 11
2.5 Daily News Arrival Count vs. Days Left to Maturity and Days Left to Trading Termination: 2004-2012 .......................................................... 19
2.6 Hourly News Arrival Counting Distribution in a Day: 2004-2012 ............ 20
2.7 News Relevance Distribution ................................................................. 21
2.8 News SENT_WORDS and TOT_WORDS Distribution ............................. 21
2.9 News RELEVANCE, SENT_WORDS and TOT_WORDS ....................... 22
2.10 News MENTION_1 and TOT_SENTS Distribution ................................ 22
2.11 News MENTION_1, TOT_SENTS and RELEVANCE .............................. 23
2.12 News Sentiment Distribution ............................................................. 24
2.13 Daily News Sentiment Probability Distribution ..................................... 25
2.14 Positive and Negative News Time Series vs. Oil Price .......................... 26
2.15 Number of Linked News Items in the 5 History Periods ............................ 27
2.16 News Sentiment Probabilities vs. Relevance ....................................... 28
2.17 News Novelty vs. Average Relevance .................................................. 28
2.18 Item Type Distribution ........................................................................ 29

3.1 Price, Return and Realized Variance of First WTI Futures Contract (F1) and Three-Month Treasury Rate ................................................................. 37
3.2 Extreme Returns and Associated Major News Event of First Crude Oil Futures Contract (F1) During the Oil Bubble peak: December 3, 2007 to June 30, 2009 ........................................... 38
3.3 Extreme Realized Variance Daily Changes and Associated Major News Event of First Crude Oil Futures Contract (F1) During the Oil Bubble peak: December 3, 2007 to June 30, 2009 ........................................... 39
3.4 Daily Net News Quantiles vs Contemporaneous Daily Return and Realized Volatility ................................................................. 43
3.5 Paired Plot of Daily Conditional Variance from GARCH process and Total Conditional Variance, for Model with and without News ................. 65

4.1 Comparing Henry Hub Futures price with Spot price ...................... 77
4.2 Henry Hub Futures Daily Price, Return and Realized Variance ............ 79
4.3 Correlation between Henry Hub Futures Returns and Spot Returns ........ 80
4.4 Comparing the Tail Distribution of Standardized Residuals ................. 83
4.5 Expected Shortfall for $\alpha = 0.95, 0.99$ and 0.995 ......................... 96
Tables

2.1 Summary of Daily Number of News Counting: 2004-2012 .................. 9
2.2 Summary of Relevance Scores: 2004-2012 ........................................ 13
2.3 Relationship between Sentiment and Relevance: 2004-2012 .............. 17
2.4 Item Types vs. Sentiment & Relevance ............................................. 18

3.1 Summary Statistics for Crude Oil Futures Data .................................. 31
3.2 Effect of 1% Log Return Outliers on Summary Statistics .................. 32
3.3 Impact of Applying Relevance Filter on News Data ........................... 40
3.4 Impact of Applying Sentiment Probability Filter on News Data ............ 41
3.5 Summary Statistics for Major News .................................................. 42
3.6 Bootstrapping Test Result on Correlations ........................................ 45
3.7 Analysis of News Explaining Clustering of Realized Variance ............... 46
3.8 Jump Intensity Parameters for News ................................................ 55
3.9 Summary Statistics And Parameters Estimation for Interest Rate Process .... 56
3.10 Estimation Results for Full Sample Period ........................................ 63
3.11 Summary Statistics for Returns of Filtered Futures Price .................... 64
3.12 Robustness: Comparing Results for Sub-samples ............................. 67
3.13 Robustness: Estimation Results for Different Sentiment Probability Thresholds 68
3.14 Robustness: Estimation Results Using Different Contracts .................. 69

4.1 Summary Statistics for Henry Hub Futures and Spot Returns ............... 78
4.2 Realized Beta GARCH Model Full Sample Fitting Result .................... 81
4.3 Comparing Residuals of Realized Beta GARCH Model with Standard GARCH Model ................................................................. 82
4.4 RealGARCH and GPD Parameters from Oct 5, 2012 to Aug 22, 2016 92
4.5 Parameter Estimations for Realized GARCH Model with Rolling Time Window 94
4.6 GPD Parameter Estimations for Realized GARCH Model Residuals with Rolling Time Window 95
4.7 Summary Statistics for Expected Shortfall on Individual Hub Model 97
Chapter 1

Background

There have been many studies that investigate the relationship between market information and asset prices. Engle and Ng (1993) [21] define a news impact curve to measure how new information is incorporated into volatility. They combine aggregated news into ARCH-family volatility models, and apply the models to Japanese equities market. They find that nonparametric ARCH models with news components perform better during extreme market conditions. Cutler, Poterba and Summers (1998) [16] look at seven microeconomic indicators including dividend, industrial production, M1, long and short interest rates, CPI and volatility. They find that these factors can explain 20% of stock price movements, using a vector autoregressive (VAR) model. Bessembinder, Chan and Seguin (2004) [7] use factors extracted from markets as proxies for information: absolute stock returns as information flow proxy; open interest of S&P500 index futures as divergence of opinion proxy; and S&P 500 index futures daily volume or stock trading volume as trading volume proxy. They build a linear model using these factors and find that trading volume is positively related to open interest changes, plus that the impact of market information is higher on equity index trading, as compared to the impact of firm-specific information on individual stock trading. Antweiler and Frank (2004) [2] examine the internet stock message boards, and its relationship with stock market returns and volatilities. They aggregate information on message board into 3 factors: the number of messages, bullishness of the messages, and disagreement of opinions. They discover strong evidence that these factors help predict volatility, yet for returns they only find that a higher number of messages predicts negative subsequent returns. Tetlock (2007) [40] extracts daily content from Wall Street Journal columns and process the articles with General Inquirer (GI). He uses the principal components to predict stock market returns and obtains statistically signifi-
cant and robust results. He finds that high media pessimism has negative correlation with prices, and usually high or low pessimism is related to high trading volume. Garcia (2013) [22] builds a similar model using New York Times financial news columns and extracts positive and negative sentiment proxies to study DJIA index prices. He uses a VAR model as well and finds that one-period lagged sentiment parameters are highly significant, especially during periods of recession. Bollen, Mao and Zeng (2011) [8] also try to predict the DJIA index, but they use information extracted from twitter feeds and processed by OpinionFinder and Google-Profile of Mood States. Boudoukh, Feldman, Kogan and Richardson (2013) [10] use news from Dow Jones Newswire tagged for the S&P 500 companies and classify days into 3 categories: no news, un-identified news and identified news. They use a linear regression method and find that if news information can be identified and tone can be determined, it will increase the $R^2$ statistic of the model.

Recent literature uses machine learning techniques to process news and build sentiment indicators. Groß-Klußmann and Hautsch (2011)[26] choose pre-processed news data from Thompson Reuters News Analytics (TRNA) and use Name Entity Recognition and Part-of-Speech Tagging algorithm to calculate the probability of news item tones. They apply a VAR model to predict market return and volatility and find that more positive tones can predict higher positive returns and lower volatility. Heston and Sinha (2014) [32] extract 900, 754 news articles and process them with Harvard General Inquirer Psychosocial Dictionary and Thompson Reuters sentiment Engine. They find that daily news aggregation predicts stock returns for the next 1 to 2 days and that weekly news aggregation can predict quarterly returns. In addition, they document that market reacts to positive news faster than to negative news. Da, Engelberg and Gao (2015) [17] construct a Financial and Economic Attitudes Revealed by Search (FEARS) index using searching query volume for certain key words, and use it to predict S&P 500 index returns and mutual fund outflow, and obtain significant parameters when predicting contemporaneous, $t + 1$ and $t + 2$ returns, and daily mutual fund outflow. Loughran and McDonald (2011) [36] argue that only using the count of negative words in
textual analyses will mis-classify common words which do not have a tone in financial statements. They find that 73.8% of the negative words defined by the popular Harvard Psycho-sociological Dictionary are not negative in financial news context, such as cost, liability and foreign.

Weather related news plays a crucial role in some markets. Boudoukh, Richardson, Shen and Whitelaw (2007) [11] examine the impact of weather news of related areas in US on the Frozen Concentrated Orange Juice (FCOJ) market. They construct a general model that captures the relationship between FCOJ futures returns and the daily low temperature in Orland, FL, and find that the model can explain 50% of the variations in returns for days with a minimum temperature below the freezing point.

There are also studies on how market participants react to news. Harris and Raviv (1993) [30] believe that traders receive the same information and the difference is in how they interpret the news. They build a trade model to simulate market participants’ behavior and find that absolute changes in traders’ expected returns are positively correlated with trading volume. Baker and Wurgler (2006) [3] study how investor sentiments affect the cross-section of stock returns. They build an investor sentiment index and find that when sentiments are estimated to be high, arbitrageurs-prone securities, such as small stocks, non-dividend stocks or high volatility stocks, tend to have lower subsequent returns. Barnerjee and Kremer (2010) [4] build a Bayesian model to investigate the relationship between investor disagreement and trading volume. They show that major disagreements among investors will lead to high trading volume and high volatility, and increase the autocorrelation of volume and volatility time series. Borovkova and Lammiman (2011) [9] adopt an event study approach and find that negative events are accompanied by losses of greater absolute magnitude as compared to that of gains from positive events.

Different investor groups tend to react differently to information. Barber and Odean (2008) [5]
show that individual investors are net buyers of stocks in the news or stocks with unusually high volume or volatility. Griffin, Hirschey and Kelly (2011) [25] show that developed and emerging markets also react to news differently: developed markets react to public news considerably more strongly than do emerging markets.

Some news are anticipated, like scheduled job market reports or the Energy Information Agency petroleum storage report; others are not, like the Swiss Franc un-pegging from the Euro. Engle, Hensen and Lunde (2011) [20] study news on 29 large US companies collected by Dow Jones Factiva find that a large proportion of changes in asset prices can be explained by unanticipated news. Bauwens, Omrane and Giot (2005) [6] investigate the volatility of the EUR/USD market following news announcement, using ARCH-family models. They find that scheduled news announcements tend to have more pre-announcement volatility rise, but the unscheduled announcements do not (except for central bank movement rumors); they find no evidence of major volatility change following news announcements, indicating that private information plays an important role in the FOREX market. Leon and Sebestyen (2008) [34] try to model Euro area interest rates and find that the most significant factors are ECB monetary policy surprises and employment reports from the United States.

How to aggregate news matters. Dzielinski and Hasseltoft (2013) [18] build two daily measures from Thomson Reuters News Analytics data: news tone is defined to be the sum of positive and negative news weighted by sentiment probabilities; news dispersion is the standard deviation of news tone. They find that news tone can predict stock returns (positively) and volatility (negatively), and that news dispersion predicts investor disagreement (positively), turnover,(positively) stock returns(negatively) and volatility (positively).

Novelty of news is another important component. Wu and Huberman (2007) [43] establish a
novelty factor by analyzing the dynamics of collective attention among one million users on an interactive website. Later, Wu and Huberman (2008) [44] find evidence between novelty and news article popularity. Hafez (2011) [27] builds a trailing sentiment index that combines news sentiments and novelty and tests the return predictability of this index on a simple long-short trading strategy. This sentiment index produces a cumulative return of 1.2 from May 2005 to Dec 2009, while without the novelty component the cumulative return shrinks to 0.5.

There are times when markets experience extreme returns but no public news is released regarding the underlying assets. Chan (2003) [13] finds that for stocks with extreme returns, if there is public news associated with them then the returns tend to keep their momentum, otherwise the returns are likely to reverse after the news announcement. These extreme returns do not seem to be related to private information, as Vega (2006) [41] shows by using a measure called probability of private information-based trading (PIN), first developed by Easley and O’Hara (1992) [19]. Hillert, Jacobs and Mullter (2014) [33] find similar results as in Chan (2003) [13] showing that stocks with higher levels of media coverage tend to exhibit stronger return momentum. Their findings are based on 2.2 million news articles collected from newspapers.

standard deviation. Christoffersen, Jacobs and Li (2016) [14] develop a discrete-time jump-family model with endogenous jumps having either constant or dynamic intensities. They find evidence that jumps are important for modeling crude oil futures, especially during crisis periods.
Chapter 2

Thomson Reuters News Analytics

We use news data provided by Thomson Reuters News Analytics (TRNA) tagged for the crude oil market, from Jan 8, 2004 to Dec 28, 2012. Thomson Reuters collected all news items reported by its news service and tagged them for related assets. There are 2261 days and 808,958 news items during this sample period. Each news item represents a piece of news information published by Thomson Reuters, processed and scored in terms of relevance, sentiment, novelty and other fields. Each item contains a publication timestamp, type (whether this is an article, alert, append or overwrite), and many characteristic scores useful for analysis. A few important fields are discussed in the following sections.

2.1 News Arrival Timestamps

2.1.1 News Arrival Date

Daily Number of News Distribution

Each news item comes with a timestamp that marks the date and time of release, allowing us to investigate patterns in news arrival process. Figure 2.1.1 panel A plots the time series of the daily total number of news \( n_t \) arrived on day \( t \), and panel B is the histogram to show \( n_t \)'s distribution. Also table 2.1 gives some summary statistics for the daily number of news for our sampled years. We have the following observations:

- Figure 2.1.1 panel A shows that there exists a large variation in the daily number of news, ranging from less than 10 a day to almost 800. We also observe more news arriving during the oil bubble period around 2008.
- Figure 2.1.1 panel B shows there are two clusters in the distribution of the daily number of news. This is due to the differences in the news arrival process between weekdays and weekends: the first cluster is associated with daily news count for weekends, during which we receive on average less than 50 news items per day; the second cluster is for weekdays, on which we tend to get around 300 items.

- From table 2.1 we confirmed the pattern of the larger number of daily news during the oil bubble period (2007-2009), and we can also identify a few days when the maximum number of news arrived—for example July 8, 2008 is about the beginning of the bubble-bursting sell-off.

Figure 2.1 : **Daily Number of News Counting: 2004-2012**

Panel A: Time series of daily total news arrival counting over time; Panel B: Histogram of daily total news arrival counting. Sample period is from Jan 8, 2004 to Dec 28, 2012.

**Seasonality Within a Week**

Figure 2.2 shows the distribution of the number of news arrived within a week. We can see that the number of news arrivals peaks on Wednesdays, which might be related to the fact that the
<table>
<thead>
<tr>
<th>Year</th>
<th>Avg Daily#</th>
<th>Median Daily#</th>
<th>StDev of Daily#</th>
<th>Max Daily#</th>
<th>Max Daily# Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004</td>
<td>205.8</td>
<td>251.5</td>
<td>126.4</td>
<td>498</td>
<td>2004-10-07</td>
</tr>
<tr>
<td>2005</td>
<td>234.3</td>
<td>254.0</td>
<td>156.4</td>
<td>678</td>
<td>2005-05-27</td>
</tr>
<tr>
<td>2006</td>
<td>298.3</td>
<td>355.0</td>
<td>183.9</td>
<td>743</td>
<td>2006-05-04</td>
</tr>
<tr>
<td>2007</td>
<td>293.6</td>
<td>350.0</td>
<td>191.3</td>
<td>716</td>
<td>2007-11-07</td>
</tr>
<tr>
<td>2008</td>
<td>313.5</td>
<td>368.5</td>
<td>188.9</td>
<td>776</td>
<td>2008-07-08</td>
</tr>
<tr>
<td>2009</td>
<td>259.9</td>
<td>302.0</td>
<td>162.4</td>
<td>674</td>
<td>2009-07-30</td>
</tr>
<tr>
<td>2010</td>
<td>226.5</td>
<td>261.0</td>
<td>130.8</td>
<td>517</td>
<td>2010-05-26</td>
</tr>
<tr>
<td>2011</td>
<td>267.9</td>
<td>314.0</td>
<td>148.7</td>
<td>621</td>
<td>2011-03-11</td>
</tr>
<tr>
<td>2012</td>
<td>225.2</td>
<td>270.0</td>
<td>124.7</td>
<td>484</td>
<td>2012-04-17</td>
</tr>
</tbody>
</table>

Table 2.1: **Summary of Daily Number of News Counting: 2004-2012**

Energy Information Agency Weekly Petroleum Report—one of the most important reports in the industry—is released every Wednesday. Also there are significantly fewer news arriving during weekends, explaining the two-cluster bi-model distribution in figure 2.1.1 panel B. Figure 2.3 plots the distribution of the number of news counting by week. From panel A we can see during times we get about 1500 ~ 2500 news on crude oil every week; in panel B we now have a slightly positively skewed single peak distribution, with a skewness of 0.097. Panel B only has one cluster, which confirms that the two-cluster pattern in the daily news count distribution in figure 2.1.1 panel
B comes from weekday seasonality.

Panel A: Weekly Number of News Counting over Time: 2004-2012
Panel B: Histogram of Weekly Number of News Counting: 2004-2012

Figure 2.3: **Weekly News Arrival Counting: 2004-2012**

Panel A: Weekly total news arrival counting over time; Panel B: Histogram of weekly total news arrival counting. Sample period is from Jan 8, 2004 to Dec 28, 2012.

### Monthly Distribution

Figure 2.4 shows the monthly pattern of the news arrival process: panel A plots the monthly distribution of daily number of news; and panel B shows the distribution of monthly total number of news. We do not observe a strong seasonal pattern in news related to the crude oil futures market, unlike in other markets such as agricultural commodities.

### Distribution with respect to Days Left to Maturity

Next in figure 2.5 we investigate how the daily number of news evolves over the trading and settling process: Panel A shows the distribution of the daily number of news versus the nearest maturity date (first calendar day of next month); and Panel B shows the distribution of daily number of news
Panel A plots the monthly distribution of the daily number of news. Panel B shows the distribution of monthly total number of news. Sample period is from Jan 8, 2004 to Dec 28, 2012.

versus the nearest trading termination day, as the crude oil futures maturing next month will stop trading on the third business day before the 25th of current month. Both plots are relatively flat, indicating that we do not observe significant seasonal patterns of news during different phases of the futures trading process.

2.1.2 News Arrival Time

Thomson Reuters News Analytics also records the exact time when each news item was released. Figure 2.6 Panel A shows the hourly news arrival count distribution in a day. Here we see a clear clustering around the active trading hours (EST 9:00 AM to 2:30 PM), and the two peaks happen around the market opening and close. Panel B of Figure 2.6 shows the evolution of the average hourly number of news for individual years. We confirmed that the clustering pattern around the active hours persists over the years and found that the magnitude of this clustering is much higher
during the oil bubble period around 2008.

2.2 Relevance Fields

Relevance fields describe how relevant the news item is to the crude oil futures market. For example, consider two news items: (A) “PetroChina steps up LNG spot imports to meet winter demand” (Dec 31, 2012); (B) “Russia to ship extra 400,000 T of oil to Belarus in Q1” (Dec 29, 2012). Both items will have an impact on the crude oil futures market, but clearly oil investors have more concerns on news B than on news A. These relevance fields are designed to give a quantitative description of the relevance levels of a news item.

2.2.1 RELEVANCE

The "RELEVANCE" field is a real valued number between 0 and 1 which measures the relevance between news and the underlying asset. For the example regarding the two news items in the previous section, news item A has a RELEVANCE score of 0.088, while news item B has a score of 1.0. The score is calculated by comparing the relative number of occurrences of the asset with the number of occurrences of other assets within the text of the news item to other words: if the asset is exclusively mentioned in the item, relevance is set to 1. In addition, if the asset is mentioned in the headline, the relevance is also 1. For items mentioning multiple assets, the asset with the most mentions will have the highest relevance; and an asset with a lower amount of mentions will have a lower relevance score.

Figure 2.7 shows the distribution of relevance scores: Panel A shows the distribution of the whole sample period. We can see that the majority(≈ 65%) of news items are of relevance 1, and other scores distribute relatively evenly; Panel B plots the relevance score percentage distribution of each year, with the detailed numbers shown in table 2.2. Although the percentage distributions are rel-
atively consistent, we do see a lower level of relevance during the oil bubble period (2007-2009).

<table>
<thead>
<tr>
<th>Year</th>
<th>1</th>
<th>0.9-1</th>
<th>0.8-0.9</th>
<th>0.7-0.8</th>
<th>0.6-0.7</th>
<th>0.5-0.6</th>
<th>0.4-0.5</th>
<th>0.3-0.4</th>
<th>0.2-0.3</th>
<th>0.1-0.2</th>
<th>0-0.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004</td>
<td>75.39%</td>
<td>3.44%</td>
<td>4.54%</td>
<td>6.28%</td>
<td>1.56%</td>
<td>2.15%</td>
<td>1.58%</td>
<td>1.59%</td>
<td>1.72%</td>
<td>1.27%</td>
<td>0.55%</td>
</tr>
<tr>
<td>2005</td>
<td>63.48%</td>
<td>2.78%</td>
<td>5.17%</td>
<td>7.03%</td>
<td>3.04%</td>
<td>3.66%</td>
<td>2.21%</td>
<td>2.50%</td>
<td>4.18%</td>
<td>3.14%</td>
<td>2.82%</td>
</tr>
<tr>
<td>2006</td>
<td>55.80%</td>
<td>2.39%</td>
<td>4.69%</td>
<td>7.19%</td>
<td>2.87%</td>
<td>4.51%</td>
<td>2.80%</td>
<td>3.09%</td>
<td>6.25%</td>
<td>5.30%</td>
<td>5.11%</td>
</tr>
<tr>
<td>2007</td>
<td>56.89%</td>
<td>1.77%</td>
<td>4.16%</td>
<td>7.03%</td>
<td>2.59%</td>
<td>4.13%</td>
<td>2.49%</td>
<td>3.24%</td>
<td>5.54%</td>
<td>5.90%</td>
<td>6.28%</td>
</tr>
<tr>
<td>2008</td>
<td>59.99%</td>
<td>1.73%</td>
<td>4.24%</td>
<td>6.71%</td>
<td>2.41%</td>
<td>4.01%</td>
<td>2.61%</td>
<td>2.95%</td>
<td>4.45%</td>
<td>4.85%</td>
<td>6.04%</td>
</tr>
<tr>
<td>2009</td>
<td>60.66%</td>
<td>1.69%</td>
<td>4.13%</td>
<td>6.47%</td>
<td>2.73%</td>
<td>4.41%</td>
<td>2.59%</td>
<td>3.16%</td>
<td>4.43%</td>
<td>4.68%</td>
<td>5.05%</td>
</tr>
<tr>
<td>2010</td>
<td>71.95%</td>
<td>2.30%</td>
<td>4.42%</td>
<td>5.30%</td>
<td>2.13%</td>
<td>3.56%</td>
<td>1.95%</td>
<td>2.26%</td>
<td>2.99%</td>
<td>2.22%</td>
<td>0.92%</td>
</tr>
<tr>
<td>2011</td>
<td>76.11%</td>
<td>1.54%</td>
<td>3.16%</td>
<td>4.93%</td>
<td>1.62%</td>
<td>3.69%</td>
<td>1.89%</td>
<td>2.11%</td>
<td>2.68%</td>
<td>1.58%</td>
<td>0.69%</td>
</tr>
<tr>
<td>2012</td>
<td>70.65%</td>
<td>1.98%</td>
<td>4.46%</td>
<td>6.08%</td>
<td>2.56%</td>
<td>3.87%</td>
<td>2.30%</td>
<td>2.32%</td>
<td>2.64%</td>
<td>2.16%</td>
<td>0.96%</td>
</tr>
</tbody>
</table>

Table 2.2: Summary of Relevance Scores: 2004-2012

This table shows detailed distribution of relevance score, by calculating the percentage of relevance scores falling into each bin over the years, plus an average relevance score at the last column.

2.2.2 SENT_WORDS and TOT_WORDS

According to Thomson Reuters News Analytics [35], SENT_WORDS refers to the number of lexical tokens (words and punctuation) in the sections of the item text that are deemed relevant to the asset. TOT_WORDS is the total number of lexical tokens (words and punctuation) in the item.

Figure 2.8 shows the distribution of SENT_WORDS and TOT_WORDS. We see an interesting multiple peak distribution here, which is due to news item categorization: the first peak corresponds to these short alert news items; while the others correspond to longer news articles.

Figure 2.9 examines the relationship between SENT_WORDS, TOT_WORDS and RELEVANCE, by plotting SENT_WORDS against TOT_WORDS, using color to indicate relevance scores. We can see that if a higher percentage of TOT_WORDS are determined to be SENT_WORDS (closer to the 45 degree line), then it’s more likely to have a higher RELEVANCE score. It’s interesting that we have some news items with low SENT_WORDS to TOT_WORDS ratio but have RELEVANCE score equal to 1, especially in the long articles area (large TOT_WORDS).
2.2.3 MENTION_1 and TOT.SENTS

According to Thomson Reuters News Analytics [35], MENTION_1 refers to the position of the first sentence in which the scored asset (crude oil in our case) is mentioned. If the value of this field is 0, this means the asset was never mentioned in the news (think of the news on Iranian nuclear talk progress: this may not mention oil directly, but any progress makes Iran one step closer to starting exporting its huge oil reserve again); if the value is 1, then the asset is mentioned in the headline, which will give the news a RELEVANCE score of 1; if value is greater than one, say $n$, then oil is not mentioned in the headline but at the $n - 1^{th}$ sentence of the news item text body. For our data during the whole sample period, 38.4% of news mentioned oil in the headline, 31.2% mentioned in the text body, and 30.4% never explicitly mentioned oil.

TOT.SENT is the total number of sentences in the news item, together with MENTION_1 we can determine the location of the first mentioning in the news.

Figure 2.10 shows the distribution of MENTION_1 and TOT.SENTS, for news that mentioned oil in the text body only. We see that most news items mentioned oil in the first 10 sentences, if ever mentioned. We also see the multiple-peak distribution in the TOT.SENTS again, which is due to the same news type categorization reason.

Figure 2.11 investigated the relationship between MENTION_1, TOT.SENTS and RELEVANCE, for news items that mentioned oil only in the text body. We can see that if oil was mentioned early (closer to x-axis), it’s more likely to have a higher RELEVANCE score.

2.3 Sentiment Fields

In the Thomson Reuters News Analytics data, each news item is scored with three sentiment probabilities of it having a positive($p^+$), neutral($p^0$) or negative($p^-$) impact on the market, which
will add up to 1. A sentiment flag is selected according to the highest sentiment probability: if being positive is most probable, this news item is marked with a sentiment flag equal to 1; if neutral then marked with 0; and negative marked with −1.

Figure 2.12 Panel A shows the distribution of news items in the three categories according to this definition, we see that positive and negative are dominating sentiment while neutral only consists about 17%; and Panel B shows the percentage distribution of the three sentiments over years.

Figure 2.13 shows the sentiment probability distributions. The negative sentiment probability is plotted against the positive sentiment probability, and the color indicates sentiment classification: red represents negative news, green positive news, and gray neutral news. We can see some internal structure in the distribution even within the same sentiment category—this is due to Thomson Reuters modifying algorithms for news type categorization.

If we compare the number of positive and negative news arrived each day with the oil price (Figure 2.14), we find some interesting patterns. The upper panel plots the number of positive news (left, green) and negative news (right, red) over time, and we have put the price of the most active crude oil futures contract on the second Y-axis. The lower panel plots the percentage of daily news arrivals being positive (left, green) and being negative (right, red), with the same oil futures price on the second Y-axis. We can see that the number of positive (negative) news tends to move together with (against) the oil price. This trend is more significant in the daily percentage plots: for periods with more news being positive, prices tend to go up, and vice versa. We will check this pattern more systematically in the next chapter.


2.4 Novelty Fields

The novelty of news is measured by counting the number of linked news items in the past. More specifically, Thomson Reuters News Analytics defined five history periods: 12 hours, 24 hours, 3 days, 5 days and 7 days. If a news item has more linked items in the history periods, it has a lower novelty. If a news item has no linked item in the past 7 days we define its novelty score equal to 1.

Figure 2.15 shows the distribution of the number of linked items in the above 5 history periods, we can see that the patterns after the past 3 days is barely changing, indicating that most of the time the reporting life of new information is within 3 days. If there is no linked files in the past 7 days, we believe it is safe to say that information contained in this news item is new to the market.

2.5 Relationship between Relevance, Sentiment and Novelty

Having discussed the news relevance, sentiment and novelty, we now take a look at the relationship between these fields. Figure 2.16 illustrates the distribution of sentiment probabilities within relevance ranges, and table 2.3 presents the average relevance for sentiment probability ranges. Below are a few observations:

- For items marked as positive news, the positive sentiment probability level is relatively independent of the relevance levels; this is also true for neutral news.

- For items marked as negative news, the relevance level is positively correlated with the negative sentiment probability level.

Figure 2.17 shows the relationship between novelty and relevance. The x-axis indicates the number of linked items, and the y-axis is the average relevance scores. We can see from this plot that the more linked items a news item has in the past, the more likely it has a higher relevance level.
Table 2.3: Relationship between Sentiment and Relevance: 2004-2012
This table presents the average relevance for sentiment probability ranges. Note that the positive news column’s average relevance only includes news that is marked as positive; the other averages are defined similarly.

<table>
<thead>
<tr>
<th>Sentiment Probability</th>
<th>Average Relevance</th>
<th>Sentiment Probability</th>
<th>Average Relevance</th>
<th>Sentiment Probability</th>
<th>Average Relevance</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3~0.35</td>
<td>0.79</td>
<td>0.3~0.35</td>
<td>0.79</td>
<td>0.3~0.35</td>
<td>0.79</td>
</tr>
<tr>
<td>0.35~0.4</td>
<td>0.83</td>
<td>0.35~0.4</td>
<td>0.79</td>
<td>0.35~0.4</td>
<td>0.83</td>
</tr>
<tr>
<td>0.4~0.45</td>
<td>0.80</td>
<td>0.4~0.45</td>
<td>0.83</td>
<td>0.4~0.45</td>
<td>0.81</td>
</tr>
<tr>
<td>0.45~0.5</td>
<td>0.81</td>
<td>0.45~0.5</td>
<td>0.85</td>
<td>0.45~0.5</td>
<td>0.82</td>
</tr>
<tr>
<td>0.5~0.55</td>
<td>0.83</td>
<td>0.5~0.55</td>
<td>0.83</td>
<td>0.5~0.55</td>
<td>0.84</td>
</tr>
<tr>
<td>0.55~0.6</td>
<td>0.79</td>
<td>0.55~0.6</td>
<td>0.84</td>
<td>0.55~0.6</td>
<td>0.84</td>
</tr>
<tr>
<td>0.6~0.65</td>
<td>0.79</td>
<td>0.6~0.65</td>
<td>0.86</td>
<td>0.6~0.65</td>
<td>0.85</td>
</tr>
<tr>
<td>0.65~0.7</td>
<td>0.78</td>
<td>0.65~0.7</td>
<td>0.87</td>
<td>0.65~0.7</td>
<td>0.88</td>
</tr>
<tr>
<td>0.7~0.75</td>
<td>0.79</td>
<td>0.7~0.75</td>
<td>0.87</td>
<td>0.7~0.75</td>
<td>0.89</td>
</tr>
<tr>
<td>0.75~0.8</td>
<td>0.80</td>
<td>0.75~0.8</td>
<td>0.87</td>
<td>0.75~0.8</td>
<td>0.88</td>
</tr>
<tr>
<td>0.8~0.85</td>
<td>*</td>
<td>0.8~0.85</td>
<td>0.85</td>
<td>0.8~0.85</td>
<td>0.87</td>
</tr>
<tr>
<td>0.85~0.9</td>
<td>*</td>
<td>0.85~0.9</td>
<td>0.84</td>
<td>0.85~0.9</td>
<td>*</td>
</tr>
<tr>
<td>0.9~0.95</td>
<td>*</td>
<td>0.9~0.95</td>
<td>*</td>
<td>0.9~0.95</td>
<td>*</td>
</tr>
<tr>
<td>0.95~1</td>
<td>*</td>
<td>0.95~1</td>
<td>*</td>
<td>0.95~1</td>
<td>*</td>
</tr>
</tbody>
</table>

2.6 Newsscope Metadata: ITEM_TYPE

News items in Thomson Reuters News Analytics are classified into the following 4 types:

1. **Alert**: The news item was an alert in the system, which is usually just one sentence.

2. **Article**: The news item was a fresh story, usually longer with a body text.

3. **Append**: The news item was generated to append texts to an existing story.

4. **Overwrite**: The news item was generated to replace an existing story.

Figure 2.18 investigates the distribution of item types. Panel A is the pie chart of all items grouped by item types. We can see that the majority of the items are news articles (69%), with fewer news alerts (19%) and news appends (12%), and very few news overwrites (0.42%). Panel B shows the same points as in figure 2.13 but colored by item types, where we can see that the internal structures in the plots are indeed due to the item types.

Table 2.4 summarizes the distribution of sentiment and relevance within each item type. We can see that alerts have a much higher average relevance score and neutral news percentage than do other item types.
Table 2.4: **Item Types vs. Sentiment & Relevance**

This table reports the following statistics for each item type: percentage of items with negative, neutral or positive sentiment, average relevance score.

<table>
<thead>
<tr>
<th>Item Type</th>
<th>Negative</th>
<th>Neutral</th>
<th>Positive</th>
<th>Average Relevance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alert</td>
<td>34.70%</td>
<td>28.39%</td>
<td>36.91%</td>
<td>1.00</td>
</tr>
<tr>
<td>Append</td>
<td>37.16%</td>
<td>17.33%</td>
<td>45.51%</td>
<td>0.69</td>
</tr>
<tr>
<td>Article</td>
<td>38.42%</td>
<td>13.33%</td>
<td>48.25%</td>
<td>0.80</td>
</tr>
<tr>
<td>Overwrite</td>
<td>39.51%</td>
<td>12.93%</td>
<td>47.56%</td>
<td>0.85</td>
</tr>
</tbody>
</table>

In this chapter we examined the Thomson Reuters News Analytics data tagged for crude oil futures market. We checked the distribution of the news arrival process at weekly, daily and hourly level, and checked for seasonality. We looked scores including sentiment words, total words, position of sentence of first mention, and novelty. We carefully investigate how relevance and sentiment is defined for news item, and checked it’s distribution over time, item type and underlying asset price movements. In the next chapter we will use this dataset to model the price dynamics of crude oil futures. We will first check the relationship in a model free environment, then propose and implement a stochastic for futures prices using news as a factor.
Figure 2.5: Daily News Arrival Count vs. Days Left to Maturity and Days Left to Trading Termination: 2004-2012

Panel A shows the distribution of daily number of news versus nearest maturity date (first calendar day of next month); and Panel B shows the distribution of daily number of news versus the nearest trading termination day, as the crude oil futures maturing next month will stop trading on the third business day before the 25th of current month. Sample period is from Jan 8, 2004 to Dec 28, 2012.
Panel A shows the hourly news arrival count distribution in a day; Panel B shows the evolution of the average hourly number of news for individual years. Sample period is from Jan 8, 2004 to Dec 28, 2012.

Figure 2.6: **Hourly News Arrival Counting Distribution in a Day: 2004-2012**
Figure 2.7: **News Relevance Distribution**

Panel A: Relevance score distribution for the full sample period; Panel B: Relevance score percentage distribution over years, from 2004 to 2012.

Figure 2.8: **News SENT_WORDS and TOT_WORDS Distribution**

Left: SENT_WORDS distribution for the full sample period; Right: TOT_WORDS distribution over years, from 2003 to 2012.
Figure 2.9: News RELEVANCE, SENT_WORDS and TOT_WORDS
This figure shows the relationship between SENT_WORDS, TOT_WORDS and RELEVANCE, by plotting SENT_WORDS against TOT_WORDS, using color to indicate relevance scores.

Figure 2.10: News MENTION_1 and TOT_SENTS Distribution
Left: MENTION_1 distribution for the full sample period; Right: TOT_SENTS distribution over years, from Jan 8, 2004 to Dec 28, 2012.
Figure 2.11: News MENTION, TOT_SENTS and RELEVANCE
If oil was mentioned early (closer to x-axis), then it’s more likely to have a higher RELEVANCE score.
Figure 2.12: News Sentiment Distribution

Panel A: News sentiment distribution using THE original TRNA rule, which categorizes news using the largest sentiment probability; Panel B: News sentiment percentage distribution split into years.
Figure 2.13: Daily News Sentiment Probability Distribution

We plot the negative sentiment probability against the positive sentiment probability, and the color indicates sentiment classification: red represents negative news, green positive news, and gray neutral news. Sample period is from Jan 8, 2004 to Dec 28, 2012.
Figure 2.14: Positive and Negative News Time Series vs. Oil Price

The upper panel plots the number of positive news (left, green) and negative news (right, red) over time, and we have put the price of the most active crude oil futures contract on the second Y-axis. The lower panel plots the percentage of daily news arrivals being positive (left, green) and being negative (right, red), with the same oil futures price on the second Y-axis.
Figure 2.15: **Number of Linked News Items in the 5 History Periods**

These 5 pie charts show the distribution of the number of linked news items of the following 5 history periods: past 12 hours, past 1 day, past 3 days, past 5 days and past 7 days.
Figure 2.16: News Sentiment Probabilities vs. Relevance
Panel A: Positive sentiment probability vs. Relevance; Panel B: Negative sentiment probability vs. Relevance. From this plot we can see that positive sentiment probabilities are negatively proportional to relevance, while negative sentiment probabilities are positively proportional to relevance.

Figure 2.17: News Novelty vs. Average Relevance
Average relevance versus number of linked news items in the 5 history periods, respectively. This plot shows that relevance is negatively related to news novelty.
Figure 2.18: **Item Type Distribution**

Panel A: Pie chart of Item Type Distribution from the full sample; Panel B: Negative vs. Positive sentiment Probability Classified by Item Type. We can see that the internal structures in figure 2.13 are indeed due to different item types, possibly because TRNA applied different algorithms on scoring different item types.
Chapter 3

Modeling the Price of Crude Oil Futures

3.1 Crude Oil Futures Price and Relationship with News

3.1.1 Understanding the Crude Oil Futures

We used data of Light Sweet Crude Oil (WTI) futures traded on the Chicago Mercantile Exchange, with a sample period from January 8, 2004 to December 28, 2012, including 2261 trading days. They are the world’s most actively traded energy futures contracts, with more than 3 billion barrels per year in open interest. Maturities of monthly contracts listed can go as far as 10 years in the future, with trading activity peaks at around one month from maturity dates.

Following Gorton, Hayashi and Rouwenhorst (2013) [24], we roll commodity futures before they are less than one month (around 22 trading days) to maturity. We look at the first 12 contracts $F_1, F_2, \ldots, F_{12}$, where $F_i$ denotes the contracts matures in $i$ months. Figure 3.1 plots the daily settlement price (Panel A), log returns(Panel B), $F_6/F_1$ price ratio (Panel C) and realized volatility (Panel C) of the $F_1$ contract; where the realized volatility measures calculated from intra-day data, using method described in Appendix C. We observe significant clustering of extreme returns and high realized volatility around the oil bubble period, which is from December 1, 2007 to January 30, 2009. Panel C indicates that futures term structure is in contango most of the times, although there are periods in the first few years and during the first half of the oil bubble that it is in backwardation.

Table 3.1 shows the summary statistics of these futures contracts. Panel A shows the mean, vari-
ance, skewness and excess kurtosis of the daily log returns for futures at different curve positions, also included in Panel A is trading volume and open interest. We observe consistent decreasing of the variance, skewness and kurtosis along the futures curve. Panel B examined the same set of summary statistics for $F_1$ on different years, and we can see that there exist large variations over time, for example the variance statistics is significantly higher around the oil bubble period.

Table 3.2 compares the summary statistics of the log returns before and after a two tailed 1% win-

**Table 3.1: Summary Statistics for Crude Oil Futures Data**

**Panel A: Summary Statistics on F1-F12**

<table>
<thead>
<tr>
<th>$F_n$</th>
<th>$\hat{\mu}$</th>
<th>$\hat{\sigma}^2$</th>
<th>$\hat{\gamma}$</th>
<th>$\hat{\kappa}$</th>
<th>Volume</th>
<th>Open Interest</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>4.41e-04</td>
<td>5.32e-04</td>
<td>-0.056</td>
<td>3.188</td>
<td>56048</td>
<td>114751</td>
</tr>
<tr>
<td>F2</td>
<td>4.50e-04</td>
<td>4.85e-04</td>
<td>-0.125</td>
<td>2.940</td>
<td>26227</td>
<td>72855</td>
</tr>
<tr>
<td>F3</td>
<td>4.61e-04</td>
<td>4.53e-04</td>
<td>-0.143</td>
<td>2.839</td>
<td>14970</td>
<td>55654</td>
</tr>
<tr>
<td>F4</td>
<td>4.71e-04</td>
<td>4.30e-04</td>
<td>-0.155</td>
<td>2.807</td>
<td>10321</td>
<td>47289</td>
</tr>
<tr>
<td>F5</td>
<td>4.80e-04</td>
<td>4.10e-04</td>
<td>-0.154</td>
<td>2.778</td>
<td>7814</td>
<td>41068</td>
</tr>
<tr>
<td>F6</td>
<td>4.89e-04</td>
<td>3.93e-04</td>
<td>-0.158</td>
<td>2.767</td>
<td>6062</td>
<td>36993</td>
</tr>
<tr>
<td>F7</td>
<td>4.95e-04</td>
<td>3.79e-04</td>
<td>-0.157</td>
<td>2.767</td>
<td>5033</td>
<td>34414</td>
</tr>
<tr>
<td>F8</td>
<td>5.01e-04</td>
<td>3.66e-04</td>
<td>-0.161</td>
<td>2.769</td>
<td>4190</td>
<td>30590</td>
</tr>
<tr>
<td>F9</td>
<td>5.05e-04</td>
<td>3.54e-04</td>
<td>-0.166</td>
<td>2.763</td>
<td>3278</td>
<td>27763</td>
</tr>
<tr>
<td>F10</td>
<td>5.08e-04</td>
<td>3.43e-04</td>
<td>-0.170</td>
<td>2.753</td>
<td>2725</td>
<td>26619</td>
</tr>
<tr>
<td>F11</td>
<td>5.12e-04</td>
<td>3.33e-04</td>
<td>-0.173</td>
<td>2.757</td>
<td>2538</td>
<td>24633</td>
</tr>
<tr>
<td>F12</td>
<td>5.15e-04</td>
<td>3.25e-04</td>
<td>-0.176</td>
<td>2.731</td>
<td>2305</td>
<td>22333</td>
</tr>
<tr>
<td>Average</td>
<td>4.85e-04</td>
<td>4.00e-04</td>
<td>-0.145</td>
<td>2.987</td>
<td>11793</td>
<td>44580</td>
</tr>
</tbody>
</table>

**Panel B: Summary Statistics for F1**

<table>
<thead>
<tr>
<th>Year</th>
<th>$\hat{\mu}$</th>
<th>$\hat{\sigma}^2$</th>
<th>$\hat{\gamma}$</th>
<th>$\hat{\kappa}$</th>
<th>Volume</th>
<th>Open Interest</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004</td>
<td>1.05e-03</td>
<td>4.92e-04</td>
<td>-0.288</td>
<td>0.562</td>
<td>27395</td>
<td>68714</td>
</tr>
<tr>
<td>2005</td>
<td>1.46e-03</td>
<td>3.81e-04</td>
<td>0.242</td>
<td>0.201</td>
<td>35339</td>
<td>85504</td>
</tr>
<tr>
<td>2006</td>
<td>-1.89e-04</td>
<td>2.69e-04</td>
<td>-0.008</td>
<td>-0.309</td>
<td>39147</td>
<td>100305</td>
</tr>
<tr>
<td>2007</td>
<td>1.92e-03</td>
<td>3.32e-04</td>
<td>-0.111</td>
<td>-0.052</td>
<td>63686</td>
<td>125572</td>
</tr>
<tr>
<td>2008</td>
<td>-3.19e-03</td>
<td>1.23e-03</td>
<td>0.097</td>
<td>1.458</td>
<td>61253</td>
<td>112487</td>
</tr>
<tr>
<td>2009</td>
<td>1.82e-03</td>
<td>9.78e-04</td>
<td>-0.154</td>
<td>2.054</td>
<td>65012</td>
<td>124421</td>
</tr>
<tr>
<td>2010</td>
<td>4.26e-04</td>
<td>3.07e-04</td>
<td>-0.127</td>
<td>-0.077</td>
<td>80131</td>
<td>140599</td>
</tr>
<tr>
<td>2011</td>
<td>2.87e-04</td>
<td>4.59e-04</td>
<td>-0.550</td>
<td>1.770</td>
<td>76387</td>
<td>138995</td>
</tr>
<tr>
<td>2012</td>
<td>-5.10e-04</td>
<td>2.61e-04</td>
<td>0.529</td>
<td>4.021</td>
<td>55534</td>
<td>135521</td>
</tr>
</tbody>
</table>

This tables presents summary statistics for crude oil futures trading and pricing dynamics, we reported the mean, annualized standard deviation, skewness and excess kurtosis of the daily close-to-close returns, the average daily trading volume and open interest from F1 to F12, where $F_n$ refers to futures contract with expiration in $n$ months. Assume $R_{n,t}$ is the return on day $t$, $t \in \{1, \ldots, T\}$ for $F_n$, it is calculated using $\ln(F_{t,n}) - \ln(F_{t-1,n})$, where $F_{t,n}$ refers to the close price for $F_{t,n}$ at time $t$. We estimate the mean ($\hat{\mu}$), variance ($\hat{\sigma}^2$), normalized skewness $\hat{\gamma}$, and excess kurtosis $\hat{\kappa}$, using the definitions recommended by Joanes and Gill (1998) for skewed distributions. The reporting period of this table contains 2261 trading days from January 5, 2004 to December 28, 2012.

This 100% columns present summary statistics of the complete sample; 98% columns present summary statistics of the sample after removed the top and bottom 1% extreme log returns. We
observe significant differences in standard deviation, Skewness and Kurtosis statistics before and after the Winsorizing treatment, indicating that outliers have significant impact on the data. In order to check this more closely, we first examine the possible causes of some the most extreme returns.

<table>
<thead>
<tr>
<th>Table 3.2 : Effect of 1% Log Return Outliers on Summary Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>100%</td>
</tr>
<tr>
<td>F1</td>
</tr>
<tr>
<td>F2</td>
</tr>
<tr>
<td>F3</td>
</tr>
<tr>
<td>F4</td>
</tr>
<tr>
<td>F5</td>
</tr>
<tr>
<td>F6</td>
</tr>
<tr>
<td>F7</td>
</tr>
<tr>
<td>F8</td>
</tr>
<tr>
<td>F9</td>
</tr>
<tr>
<td>F10</td>
</tr>
<tr>
<td>F11</td>
</tr>
<tr>
<td>F12</td>
</tr>
<tr>
<td>Average</td>
</tr>
</tbody>
</table>

This table compares the summary statistics of the log returns before and after a two-tailed 1% Winsorizing: 100% columns present summary statistics of the complete sample; 98% columns present summary statistics of the sample after removed the top and bottom 1% extreme log returns.

Figure 3.2 panel A plots the daily prices and log returns of $F1$ from December 3, 2007 to June 30, 2009, around the peak of oil bubble, plus the days with largest 5 positive returns and 5 negative returns marketed by $P1, \ldots, P5$ and $N1, \ldots, N5$, respectively; in panel B we dig into the news archive and found the following events that caused these huge price movements. These are all major news influencing either the future demand or supply of crude oil, events that do not happen everyday.

The pattern we find in Figure 3.2 seems to indicate three things: first, large price movements are likely caused by important supply or demand information hitting the market; second, the influence of information on price is directional, that news predicting higher future demand or lower supply will drive the price higher, and news predicting lower future demand or higher supply will drive
price lower; third, not all news are equal and we should find a way to filter for major news events. As a result of these observations, when we incorporate news events into futures pricing dynamics, we should split the process positive information and negative information and treat them differently.

Similar pattern of information shock can be found regarding the daily changes in realized variance of the futures price, which we can see from figure 3.3. Note that for these volatility jumps, the events we can identify are all negative news, which again confirmed an asymmetrical affect of news with different sentiment on the futures price dynamics.

### 3.1.2 Filter for Major News

As discussed before, not all news items are created equal: some are highly relevant to the underlying asset, some are not; some contain extreme sentiment that will shake the market, while some might not even create a minor ripple in the price. To filter for those major news events we first start with the relevance criterion. Table 3.3 analyzed the impact of news time series if applied a criterion of using only the news items with relevance score equal to 1. In Panel A we checked the summary statistics before and after filter using relevance: the raw data has on average 358 items per day, with 164 (45%) of the news being positive and 134 (38%) being negative. If we only keep the news items with relevance score equal to 1, we end up with 230 items in total per day, with 96 (42%) being positive and 91 (40%) being negative.

Panel B shows the time series regression result, following Engle, Hansen and Lunde (2012) [20]. $R^2_A$ stands for the $R^2$ measure of regression of $z_t$ only on $z_{t-1}$, $z_t$ can be either daily returns or log changes in realized variance. Then $R^2_{AN}$ is the $R^2$ for $z_t$ regressing on $z_{t-1}$, $n_t^+$ and $n_t^-$, the latter two variables being the contemporaneous daily count of positive and negative news. We observe
significant improvement for return regression when added contemporaneous news variable, and minor improvement for log changes in realized variance, and these changes are statistically significant using the Wald test \((W_{+N})\) with a \(F_{2,N_{obs}−3}\) statistic. Also we found that using news with relevance equal to 1 provides more \(R^2\) increase, for both returns and log changes in realized variance. We also checked the explanatory power of using lagged news on the same regression model, and result shows they have little impact on \(R^2_{AL}\), and the Wald statistics \((W_{+L})\) are not significant.

Next we examine the effect of adding filters on sentiment probabilities, as we want to eliminate those news items that only have marginal impacts and are mostly just noises. We propose to set a threshold \((\pi)\) for sentiment probabilities, and if a news item has positive sentiment probability \(p(+) \geq \pi\), it is a major positive news; if it has negative sentiment probability \(p(−) \geq \pi\), it is major negative news; otherwise it will be categorized into neutral news. Table 3.4 examines the impact of applying such a threshold for different \(\pi\) values.

Panel A presents the daily average number of news under different thresholds, plus the Spearman correlation between major positive news and major negative news. If we don’t apply any filter on sentiment probability (first row), we have on average 96 positive news, 43 neutral news and 91 negative news per day, with a correlation between positive and negative news at 0.30 and it is statistically significant. This significant correlation only disappears for threshold higher than 0.65. For example when we have \(\pi = 0.75\), we end up with 31 major positive news (32% of number of news without sentiment filter) and 33 major negative news (36%), with close to zero correlation.

Panel B and C used a similar time series regression method as in table 3.3 to analyze the impact of different sentiment probability thresholds, for daily returns (Panel B) and log changes in realized variance (Panel C). We find that for \(R^2_{+N}\) with contemporaneous news time series, a higher threshold generally adds more to \(R^2\) for both returns and log changes of realized variance. This indicates that if by increasing the threshold, we actually filter out noise and keep the information
that has the most significant impact on the price dynamics. We also checked the impact of adding lagged news, and as in table 3.3 the lagged news has little impact on the $R_{+L}^2$.

We decide to use threshold value $\pi = 0.75$, as it will remove the significant correlation between positive and negative news and produce high $R_{+N}^2$. We will check the robustness of the model using data with $\pi = 0.65$ in the results.

Hence we rebuild the news time series using the following filters:

- We only use news items that have relevance score equal to 1.
- We define positive news as items with positive sentiment probability greater or equal to 0.75; define negative news as items with negative sentiment probability greater or equal to 0.75; the rest of the items are categorized as neutral news.

Table 3.5 shows summary statistics of news under these criterion. To summarize, the unfiltered news has 808,631 items in total (358 per day), with 370,901 (164 per day) positive news and 303,519 (134 per day) negative news; after filtering we have 69,173 (31 per day) positive news and 73,821 (33 per day) negative news.

### 3.1.3 News and Futures Price Dynamics

Equipped with the news times series generated from the new definitions, we can take a closer look at the relationship between news and futures price dynamics. We define net news meaning the number of positive news minus the number of negative news items:

$$NetNews_t = PositiveNews_t - NegativeNews_t$$

(3.1)

Figure 3.4 plot the distribution of daily returns versus net news quantiles. We find a positive relationship between net news and returns, and a negative relationship between net news and realized
variance. To further investigate this pattern, we perform a bootstrapping-based test in table 3.6.
Figure 3.1: Price, Return and Realized Variance of First WTI Futures Contract (F1) and Three-Month Treasury Rate

This figure presents information of the West Texas Intermediate (WTI) front contract futures contract (F1), from January 5, 2004 to December 28, 2012. This period spans 2261 trading days, first contract (F1) means always using the futures contract maturing next month. Panel A plots the daily settlement price; Panel B plots the daily returns calculated from daily settlement price; Panel C plots the daily price ratio between F6 and F1, with the solid horizontal line indicating one (price equivalence), and the ratio greater than one implies an upward sloping term structure; Panel D plots realized variance for the nearest to maturity futures contract (F1). We obtained daily settlement price from Commodity Research Bureau, and intraday data for calculating realized volatility from TickData.
Figure 3.2: Extreme Returns and Associated Major News Event of First Crude Oil Futures Contract (F1) During the Oil Bubble peak: December 3, 2007 to June 30, 2009

The upper plot panel A shows F1 price and panel B shows daily return after the oil bubble burst, from December 1st, 2007 to June 30, 2009, with days having top five positive and top five negative returns marked on the plot. Here IEA refers to International Energy Agency, EIA refers to the U.S. Energy Information Administration. The lower table shows the returns $R_t$ on these extreme days along with the associated major news event identified, plus returns of previous three days ($R_{t-3}, R_{t-2}, R_{t-1}$) for reference.
Panel A: F1 Realized Variance During Oil Bubble

Panel B: F1 Realized Variance Daily Change During Oil Bubble

<table>
<thead>
<tr>
<th>Label</th>
<th>Date</th>
<th>$\Delta_{\sigma^2}$</th>
<th>$\sigma^2_0$</th>
<th>$\sigma^2_{t-1}$</th>
<th>$\sigma^2_t$</th>
<th>News</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>2008-10-10</td>
<td>-0.23</td>
<td>0.70</td>
<td>0.36</td>
<td>0.13</td>
<td>0.83 global equity market crashed with Dow plunged 697 points; concern about weakening global demand for fuel;</td>
</tr>
<tr>
<td>P2</td>
<td>2008-10-16</td>
<td>-0.14</td>
<td><strong>0.39</strong></td>
<td>0.28</td>
<td>0.15</td>
<td>0.54 EIA reported crude stocks up 5.6 mil BBLS vs forecast of 1.9 mil.</td>
</tr>
<tr>
<td>P3</td>
<td>2008-11-13</td>
<td>-0.05</td>
<td><strong>0.27</strong></td>
<td>0.25</td>
<td>0.20</td>
<td>0.47 IEA cuts world oil demand forecast on weak economy.</td>
</tr>
<tr>
<td>P4</td>
<td>2008-12-10</td>
<td>-0.05</td>
<td><strong>0.47</strong></td>
<td>0.32</td>
<td>0.27</td>
<td>0.74 EIA predicted world oil demand to fall for first time in decades</td>
</tr>
<tr>
<td>P5</td>
<td>2008-12-31</td>
<td>-0.14</td>
<td><strong>0.47</strong></td>
<td>0.37</td>
<td>0.23</td>
<td>0.70 EIA reported crude stock up 500,000 BBLS, vs forecast of 1.5 mil BBLS.</td>
</tr>
</tbody>
</table>

Figure 3.3: Extreme Realized Variance Daily Changes and Associated Major News Event of First Crude Oil Futures Contract (F1) During the Oil Bubble peak: December 3, 2007 to June 30, 2009

In the upper plots, panel A shows daily realized variance ($\sigma^2_t$), panel B shows daily realized variance change ($\Delta_{\sigma^2} = \sigma^2_t - \sigma^2_{t-1}$) after the oil bubble burst, from December 3, 2007 to June 30, 2009, with days having top five positive realized variance changes ($\Delta_{\sigma^2}$) marked on the panel B. Here IEA refers to International Energy Agency, EIA refers to the U.S. Energy Information Administration.

The lower table shows the realized variance changes $\Delta_{\sigma^2}$ on these extreme days along with the associated major news event or reasons identified, plus realized variance change of the previous day ($\Delta_{\sigma^2_{t-1}} = \sigma^2_{t-1} - \sigma^2_{t-2}$) and realized variance of previous two days ($\sigma^2_{t-2}$, $\sigma^2_{t-1}$, $\sigma^2_t$) for reference.
Table 3.3: Impact of Applying Relevance Filter on News Data

Panel A: Impact of Applying Relevance Filter on Summary Statistics For Daily News Data

<table>
<thead>
<tr>
<th>Year</th>
<th>Unfiltered News</th>
<th>News with Relevance=1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total (+)</td>
<td>(n)</td>
</tr>
<tr>
<td>2004</td>
<td>277</td>
<td>127(46%)</td>
</tr>
<tr>
<td></td>
<td>40(14%)</td>
<td>111(40%)</td>
</tr>
<tr>
<td>2005</td>
<td>319</td>
<td>159(50%)</td>
</tr>
<tr>
<td></td>
<td>49(15%)</td>
<td>110(35%)</td>
</tr>
<tr>
<td>2006</td>
<td>416</td>
<td>216(52%)</td>
</tr>
<tr>
<td></td>
<td>61(15%)</td>
<td>133(32%)</td>
</tr>
<tr>
<td>2007</td>
<td>409</td>
<td>214(53%)</td>
</tr>
<tr>
<td></td>
<td>61(15%)</td>
<td>133(32%)</td>
</tr>
<tr>
<td>2008</td>
<td>430</td>
<td>213(49%)</td>
</tr>
<tr>
<td></td>
<td>59(14%)</td>
<td>158(37%)</td>
</tr>
<tr>
<td>2009</td>
<td>363</td>
<td>174(47%)</td>
</tr>
<tr>
<td></td>
<td>57(16%)</td>
<td>133(37%)</td>
</tr>
<tr>
<td>2010</td>
<td>316</td>
<td>125(40%)</td>
</tr>
<tr>
<td></td>
<td>62(19%)</td>
<td>130(41%)</td>
</tr>
<tr>
<td>2011</td>
<td>373</td>
<td>139(37%)</td>
</tr>
<tr>
<td></td>
<td>74(20%)</td>
<td>159(43%)</td>
</tr>
<tr>
<td>2012</td>
<td>315</td>
<td>110(35%)</td>
</tr>
<tr>
<td></td>
<td>70(22%)</td>
<td>135(43%)</td>
</tr>
<tr>
<td>All</td>
<td>358</td>
<td>164(45%)</td>
</tr>
<tr>
<td></td>
<td>59(17%)</td>
<td>134(38%)</td>
</tr>
</tbody>
</table>

Panel B: Impact of Applying Relevance Filter on Time Series Regression Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Unfiltered News</th>
<th>News with Relevance=1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Return</td>
<td>∆ln(Realized Variance)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R²A (%)</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>∆₁N (R²)</td>
<td>6.56</td>
</tr>
<tr>
<td></td>
<td>W₁N</td>
<td>79.04</td>
</tr>
<tr>
<td></td>
<td>R²AN (%)</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>∆₁L (R²)</td>
<td>-0.12</td>
</tr>
<tr>
<td></td>
<td>W₁L</td>
<td>0.18</td>
</tr>
</tbody>
</table>

This table examines the impact of applying a relevance filter to the news data by only keeping news items with relevance score equal to 1. Results are produced using data from January 5, 2004 to December 28, 2012.

Panel A provides daily average summary statistics on daily news data for individual years. We compare the distribution of unfiltered news with news filtered for relevance. The numbers in parenthesis indicate the percentage of news items fall into each category out of total under the current filtering condition.

Panel B performed a time series regression analysis comparison between unfiltered news and news filtered for relevance, regarding their explanatory power towards F1 daily price movements. The column names are the dependent variables(y), either daily returns or daily log change in realized variance measures; the row names indicate the adjusted R² and diagnostic statistics with different explanatory variables structures: R²A is the R² measure of the basic AR(1) model:

\[ z_t = \alpha_0 + \alpha_1 z_{t-1} + \epsilon_t \]

R²AN is the adjusted R² after added contemporaneous news:

\[ z_t = \alpha_0 + \alpha_1 z_{t-1} + b^{(+)} n_t^{(+)} + b^{(-)} n_t^{(-)} + \epsilon_t \]

and ∆₁N is the percentage change in R² with the contemporaneous news variables, an *** sign is appended if this change is significant at 0.001 level (*** will be attached if significant at 0.01 level, * for 0.05 level); and W₁N is the Wald test F statistic for adding the new factors, with dfnum = 2 and dfdenom = 2257. Similarly R²AL indicate the adjusted R² with the lagged news, i.e z_t = \alpha_0 + \alpha_1 z_{t-1} + b^{(+)} n_{t-1}^{(+)} + b^{(-)} n_{t-1}^{(-)} + \epsilon_t, and Wald statistic and R² compared to the basic AR(1) model are defined similarly.
Table 3.4: Impact of Applying Sentiment Probability Filter on News Data

Panel A: Impact on Summary Statistics

<table>
<thead>
<tr>
<th>Threshold</th>
<th>Positive</th>
<th>Neutral</th>
<th>Negative</th>
<th>Corr(+/-)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA</td>
<td>96</td>
<td>43</td>
<td>91</td>
<td>0.30(0.04)</td>
</tr>
<tr>
<td>0.45</td>
<td>89(93%)</td>
<td>57(133%)</td>
<td>84(92%)</td>
<td>0.24(0.04)</td>
</tr>
<tr>
<td>0.55</td>
<td>68(72%)</td>
<td>96(224%)</td>
<td>65(72%)</td>
<td>0.09(0.04)</td>
</tr>
<tr>
<td>0.65</td>
<td>54(56%)</td>
<td>130(302%)</td>
<td>47(51%)</td>
<td>0.01(0.04)</td>
</tr>
<tr>
<td>0.75</td>
<td>31(32%)</td>
<td>167(387%)</td>
<td>33(36%)</td>
<td>0.00(0.04)</td>
</tr>
</tbody>
</table>

Panel B: Impact on Return Regression Results

<table>
<thead>
<tr>
<th>Threshold</th>
<th>(R^2_{+N}(%))</th>
<th>(\Delta_{+N}(R^2))</th>
<th>(W_{+N})</th>
<th>(R^2_{+L}(%))</th>
<th>(\Delta_{+L}(R^2))</th>
<th>(W_{+L})</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA</td>
<td>0.14</td>
<td>0.14</td>
<td></td>
<td>0.14</td>
<td>0.14</td>
<td></td>
</tr>
<tr>
<td>0.33</td>
<td>9.34</td>
<td>9.20***</td>
<td>116.2</td>
<td>0.06</td>
<td>-0.08</td>
<td>0.5</td>
</tr>
<tr>
<td>0.45</td>
<td>9.77</td>
<td>9.63***</td>
<td>122.1</td>
<td>0.06</td>
<td>-0.08</td>
<td>0.6</td>
</tr>
<tr>
<td>0.55</td>
<td>9.71</td>
<td>9.57***</td>
<td>121.3</td>
<td>0.07</td>
<td>-0.07</td>
<td>0.7</td>
</tr>
<tr>
<td>0.65</td>
<td>10.00</td>
<td>9.86***</td>
<td>125.2</td>
<td>0.09</td>
<td>-0.05</td>
<td>1.0</td>
</tr>
<tr>
<td>0.75</td>
<td>10.57</td>
<td>10.43***</td>
<td>133.2</td>
<td>0.14</td>
<td>0.00</td>
<td>1.5</td>
</tr>
</tbody>
</table>

Panel C: Impact on \(\Delta \ln(\text{Realized Variance})\) Regression Results

<table>
<thead>
<tr>
<th>Threshold</th>
<th>(R^2_{+N}(%))</th>
<th>(\Delta_{+N}(R^2))</th>
<th>(W_{+N})</th>
<th>(R^2_{+L}(%))</th>
<th>(\Delta_{+L}(R^2))</th>
<th>(W_{+L})</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA</td>
<td>19.99</td>
<td></td>
<td></td>
<td>19.99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.33</td>
<td>20.42</td>
<td>0.44***</td>
<td>7.7</td>
<td>19.96</td>
<td>-0.11</td>
<td>1.1</td>
</tr>
<tr>
<td>0.45</td>
<td>20.48</td>
<td>0.49***</td>
<td>8.4</td>
<td>19.96</td>
<td>-0.11</td>
<td>1.0</td>
</tr>
<tr>
<td>0.55</td>
<td>20.61</td>
<td>0.62***</td>
<td>10.3</td>
<td>19.96</td>
<td>-0.11</td>
<td>1.1</td>
</tr>
<tr>
<td>0.65</td>
<td>20.99</td>
<td>1.00***</td>
<td>15.9</td>
<td>19.93</td>
<td>-0.13</td>
<td>0.7</td>
</tr>
<tr>
<td>0.75</td>
<td>20.84</td>
<td>0.85***</td>
<td>13.7</td>
<td>19.97</td>
<td>-0.10</td>
<td>1.2</td>
</tr>
</tbody>
</table>

This table evaluates how news sentiment probability filtering will affect its explanatory power on returns and realized variances, where for threshold value \(\tau\), we re-define positive news as items with positive sentiment probability \(p^+(\tau)\), and re-define negative news as items with \(p^-(\tau)\), the rest are re-categorized as neutral news. Results are produced using data from January 5, 2004 to December 28, 2012. Panel A provides summary statistics under different filtering thresholds for sentiment probabilities. Positive/Neutral/Negative indicate the daily average number of news in each category with certain sentiment probability thresholds, the number in parenthesis gives the percentage of news that remains out of the original population before filtering for sentiment probabilities. Corr(+/−) gives the Spearman correlation between the positive and negative news time series.

Panel B investigates the impact of filtering for sentiment probabilities using a regression analysis approach. This presents a similar regression test as in Panel B of Table 7: adjusted \(R^2\), \(\Delta_{+N}\) as the percentage changes in \(R^2\) after added contemporaneous news time series obtained under given sentiment probability thresholds, an *** sign is appended if this change is significant at 0.001 level (** will be attached if significant at 0.01 level, * for 0.05 level). \(W_{+N}\) is the Wald test F statistic for adding the new factors, with \(df_{num} = 2\) and \(df_{denom} = 2257\). For the row names, \(NA\) indicate no news involved, hence its a simple \(zt = a_0 + a_1 z_{t-1} + \varepsilon_t\) model. Then we present regression results with model:

\[zt = a_0 + a_1 z_{t-1} + \delta^+(n^{+})_{n,t} + \delta^-(n^{-})_{n,t} + \varepsilon_t\]

\(n^{(+/-)}_{t}\) being news time series under threshold \(\tau\). Similar analysis have been performed using lagged news (lag 1), presented in columns marked with \(+L\).
Table 3.5: Summary Statistics for Major News

<table>
<thead>
<tr>
<th></th>
<th>Average (+)</th>
<th>StDev (+)</th>
<th>25% Quantile (+)</th>
<th>50% Quantile (+)</th>
<th>75% Quantile (+)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n)</td>
<td>(n)</td>
<td>(n)</td>
<td>(n)</td>
<td>(n)</td>
</tr>
<tr>
<td>2004</td>
<td>28</td>
<td>153</td>
<td>15</td>
<td>18</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>15</td>
<td>13</td>
<td>13</td>
<td>15</td>
</tr>
<tr>
<td>2005</td>
<td>31</td>
<td>144</td>
<td>14</td>
<td>21</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>53</td>
<td>13</td>
<td>16</td>
<td>22</td>
</tr>
<tr>
<td>2006</td>
<td>35</td>
<td>163</td>
<td>13</td>
<td>26</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>38</td>
<td>13</td>
<td>22</td>
<td>30</td>
</tr>
<tr>
<td>2007</td>
<td>35</td>
<td>166</td>
<td>12</td>
<td>27</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>41</td>
<td>12</td>
<td>12</td>
<td>20</td>
<td>16</td>
</tr>
<tr>
<td>2008</td>
<td>37</td>
<td>178</td>
<td>13</td>
<td>29</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>19</td>
<td>19</td>
<td>28</td>
<td>36</td>
</tr>
<tr>
<td>2009</td>
<td>32</td>
<td>158</td>
<td>11</td>
<td>25</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>12</td>
<td>12</td>
<td>20</td>
<td>28</td>
</tr>
<tr>
<td>2010</td>
<td>28</td>
<td>166</td>
<td>10</td>
<td>20</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>45</td>
<td>19</td>
<td>19</td>
<td>20</td>
<td>28</td>
</tr>
<tr>
<td>2011</td>
<td>30</td>
<td>207</td>
<td>12</td>
<td>22</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>58</td>
<td>21</td>
<td>21</td>
<td>32</td>
<td>42</td>
</tr>
<tr>
<td>2012</td>
<td>21</td>
<td>165</td>
<td>8</td>
<td>15</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>41</td>
<td>13</td>
<td>13</td>
<td>27</td>
<td>34</td>
</tr>
<tr>
<td>All</td>
<td>31</td>
<td>167</td>
<td>13</td>
<td>21</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>17</td>
<td>17</td>
<td>22</td>
<td>30</td>
</tr>
</tbody>
</table>

This table presents summary statistics and arrival process parameter estimations for the major news time series, major news refers to positive or negative news when a threshold of 0.75 is applied. We provide the average number, the standard deviation and the 25%, 50%, 75% quantile of the time series constructed by daily number of major news of each sentiment category, acquired from the news time series using a sentiment probability threshold $\tau = 0.75$. These summary statistics are reported for each year as well as for the whole sample period. The full sample period includes 2261 trading days from January 5, 2004 to December 28, 2012.
Panel A1, A2 and A3 plots the distribution of daily return versus F1 daily net news quantiles for different sample periods, daily returns defined as $R_{n,t} = \ln(F_{t,n}) - \ln(F_{t-1,n})$. For each box, the top and bottom of the box represent the 25th (Q1) and 75th (Q3) percentile, respectively; and and upper and lower end of the whiskers represent $\min(\max(\text{data}), Q3 + 1.5IQR)$ and $\max(\min(\text{data}), Q1 - 1.5IQR)$, respectively. $IQR = Q3 - Q1$ is the inter quantile range. Boxes on the left represent days with more negative net news, while boxes on the right represent days with more positive net news.

We observe a trend of positive contemporaneous correlation between net news and daily returns.

Panel B1, B2 and B3 plots the distribution of daily realized volatility versus F1 daily net news quantiles, with boxes on the left represent more negative net news days and boxes on the right represent more positive net news days. Full sample period is from January 5, 2004 to December 28, 2012. Bubble period is from January 5, 2004 to June 30, 2009; Post Bubble period is from July 1, 2009 to December 28, 2012.
Table 3.6 calculated the Spearman correlation between news of different sentiment and futures returns of different curve positions, plus the correlation between news with different sentiments. A few observations:

1. We find significant positive correlation between positive news and return, with a decreasing pattern over the futures curve;

2. We find significant positive correlation between positive news and squared returns, with a decreasing pattern over the futures curve;

3. We do not observe any significant correlation between neutral news and return;

4. We do observe a positive correlation between neutral news and squared return, but standard errors are high compared to the correlation measure;

5. We observe significant negative correlation between negative news and return, with a decreasing pattern over the futures curve;

6. We observe significant positive correlation between negative news and squared returns, with a decreasing pattern over the futures curve;

7. Correlation between negative news and returns (squared returns) consistently has a higher magnitude than those with positive news.

8. We do not observe significant correlation between the number of positive and negative news items, however neutral news is significantly correlated with both positive and negative news.

We have shown that information from news can provide significant improvement in time series regression models of returns and log changes of realized variances in table 3.4. Next in table 3.7 we examine if news can also explain the clustering behavior of realized variance. In this table
Table 3.6: Bootstrapping Test Result on Correlations

Panel A: Correlations between the Major News and Futures Price Movements

<table>
<thead>
<tr>
<th>Time Series</th>
<th>Positive News</th>
<th>Neutral News</th>
<th>Negative News</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>vs Return</td>
<td>vs Squared Return</td>
<td>vs Return</td>
</tr>
<tr>
<td>F1</td>
<td>0.137(0.026)</td>
<td>0.069(0.028)</td>
<td>0.037(0.019)</td>
</tr>
<tr>
<td>F2</td>
<td>0.136(0.027)</td>
<td>0.066(0.027)</td>
<td>0.036(0.019)</td>
</tr>
<tr>
<td>F3</td>
<td>0.134(0.027)</td>
<td>0.066(0.027)</td>
<td>0.034(0.019)</td>
</tr>
<tr>
<td>F4</td>
<td>0.133(0.026)</td>
<td>0.067(0.026)</td>
<td>0.034(0.019)</td>
</tr>
<tr>
<td>F5</td>
<td>0.133(0.027)</td>
<td>0.066(0.027)</td>
<td>0.035(0.019)</td>
</tr>
<tr>
<td>F6</td>
<td>0.130(0.026)</td>
<td>0.067(0.027)</td>
<td>0.034(0.019)</td>
</tr>
<tr>
<td>F7</td>
<td>0.130(0.027)</td>
<td>0.068(0.027)</td>
<td>0.036(0.019)</td>
</tr>
<tr>
<td>F8</td>
<td>0.129(0.026)</td>
<td>0.067(0.028)</td>
<td>0.036(0.020)</td>
</tr>
<tr>
<td>F9</td>
<td>0.129(0.026)</td>
<td>0.065(0.027)</td>
<td>0.036(0.020)</td>
</tr>
<tr>
<td>F10</td>
<td>0.129(0.026)</td>
<td>0.065(0.028)</td>
<td>0.037(0.020)</td>
</tr>
<tr>
<td>F11</td>
<td>0.128(0.026)</td>
<td>0.065(0.028)</td>
<td>0.038(0.020)</td>
</tr>
<tr>
<td>F12</td>
<td>0.128(0.026)</td>
<td>0.064(0.028)</td>
<td>0.038(0.020)</td>
</tr>
</tbody>
</table>

Panel B: Correlations within Major News of Different Sentiments

<table>
<thead>
<tr>
<th>Sentiments</th>
<th>Positive News</th>
<th>Neutral News</th>
<th>Negative News</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Positive News</td>
<td>Neutral News</td>
<td>Negative News</td>
</tr>
<tr>
<td></td>
<td>vs Return</td>
<td>vs Squared Return</td>
<td>vs Return</td>
</tr>
</tbody>
</table>

This table presents bootstrapping test results for the Pearson correlations, for sample period from January 5, 2004 to December 28, 2012.

Panel A is for contemporaneous correlations between major news time series of different sentiments and futures returns/squared returns for contracts F1 to F12, with standard error from bootstrapping presented in the parenthesis. We used squared returns as a proxy for realized variance, since the liquidity for further contracts won’t provide reliable estimation for realized variance. Panel B tested the contemporaneous correlations within news time series between different sentiments, with standard error presented in the parenthesis.

we depict the effect of adding news factors to the half life of the impulse response function from autoregressive models of realized variances. Half life is a measure of the persistence of volatility shocks, which is calculated as the number of time units (days) it takes for half of the cumulative impulse to pass. A larger half life means a shock takes longer to pass, indicating a stronger clustering pattern in the model. Our analysis shows that adding contemporaneous news will significantly reduce the half life, meaning that it can help explain part of the volatility clustering embedded in the autoregressive structure of realized variance, which is consistent with findings by Engle, Hanse and Lunde (2012) [20].

Another interesting finding is that when we add the positive and negative news respectively, we observe that negative news provides much greater improvement in reducing the half life, increas-
ing the log likelihood of the model and decreasing the corrected AIC measure. Consistent with previous findings, this indicates an asymmetric pattern of how news with different sentiment affects the volatility, a pattern that also shows up in the result of our model in the later part of this paper. We have shown in a model free environment that the news time series we defined is related to the crude oil futures prices in many ways: we find that net news has a positive relationship with returns and a negative one with realized variances; we find in a time series regression frame work, news can help explain returns and log changes of realized variance, and negative news contribute more to this explanatory power than does positive news; news can also explain volatility clustering (mainly negative news); news affects the whole futures curve, although at different magnitude for different curve positions. In the next section we build a five-factor stochastic model for this rela-

Table 3.7 : Analysis of News Explaining Clustering of Realized Variance

<table>
<thead>
<tr>
<th></th>
<th>AR(5)</th>
<th>AR(5) plus Contemporaneous News</th>
<th>AR(5) plus Lagged News</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$A_r$</td>
<td>$A_r^{(+)}$</td>
<td>$A_r^{(-)}$</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>66.8</td>
<td>66.8</td>
<td>66.8</td>
</tr>
<tr>
<td>Half Life</td>
<td>36</td>
<td>36</td>
<td>36</td>
</tr>
<tr>
<td>$% \Delta H_{HF}$</td>
<td>*</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>3838.4</td>
<td>3838.5</td>
<td>3838.5</td>
</tr>
<tr>
<td>Corrected AIC</td>
<td>-7664.7</td>
<td>-7663.0</td>
<td>-7663.0</td>
</tr>
<tr>
<td>LR Test p value</td>
<td>*</td>
<td>0.80</td>
<td>0.93</td>
</tr>
</tbody>
</table>

denoted as $AR(5)$. We calculated the adjusted $R^2$, the Half Life, the log likelihood and corrected AIC for the model. Half Life here means the the time it takes for the cumulative impulse response to diminish, for example you have a unit shock $\epsilon_1 = 1$ and then no shock afterwards $\epsilon_i = 0$, $i = 2, 3, \ldots$, then the half life is the time that the 50% impulse created by the $\epsilon_1$ to pass, and here we can expect that $a_i$ with larger magnitude tend to have longer half life, as the initial shock will be more persistent. Then $A_r^{(+)}$ will indicate a model with the original autoregressive structure, but added a contemporaneous positive news factor:

$$RV_t = \sum_{i=1}^{5} a_i RV_{t-i} + \epsilon_t$$

denoted as $AR(5)$. Then $A_r^{(-)}$ is the $AR(5)$ model plus contemporaneous negative news; and $A_r^{(+)} A_r^{(-)}$ means added both news time series. $A_r^{(+)}$ and $A_r^{(-)}$ are defined similarly but using news factors at lag 1 ($n_{t-1}$).

A reduction of half life of impulse response function when new variables are added will indicate that these variables can help explain the clustering in the realized variance, a larger reduction is associated with higher explanatory power.
tionship, and the estimation results confirmed the above findings via parameter values or properties summary.
3.2 Our Model

We propose a five factor model for the crude oil futures prices, including:

1. **Crude oil spot price.** By assumption, futures price should equal to the spot price at maturity. Current futures price is spot price plus the cost (or loss) added by holding the physical asset until maturity. We denote the spot price at time $t$ as $S_t$, and the log spot price as $X_t = \log(S_t)$.

2. **Interest rate.** Interest rate directly affects the cost of holding the asset, we denote it as $r_t$.

3. **Convenience yield.** Convenience yield is the premium for holding the underlying asset. We denote it as $\delta_t$.

4. **Stochastic volatility.** This term is described by a $GARCH(1,1)$ process following Heston and Nandi (2000) [31], denoted as $h_t$.

5. **News.** We only consider positive and negative news. The number of positive news items arriving on day $t$ is denoted as $n(t^+)$, the number of negative news items is denoted as $n(t^-)$.

We first discuss the details of how to model these factors, and then derive a predictive model for futures prices in the following sections.

3.2.1 Crude Oil Spot Price

We describe the dynamics of log spot price under P measure by:

$$X_{t+\Delta} - X_t = (r_t - \delta_t)\Delta - t_t^Q + e_{t+1}^{(X)}$$

(3.2)

Here $\Delta$ is the length of unit time interval. $t_t^Q$ is given by:

$$\exp(t_t^Q) = E_t^Q[\exp(e_{t+1}^{(X)})]$$

(3.3)
where this expectation is calculated under the risk neutral measure (Q measure).

We split the random term into two parts: a volatility component and a jump component:

\[ e_t^{(X)} = e_t^{(V)} + e_t^{(J)} \]  

(3.4)

We will discuss how to construct these error components later.

### 3.2.2 Interest Rate

The interest rate process is modeled by a Gaussian mean reverting process, given by:

\[ r_{t+1} - r_t = a(m - r_t) + e_t^{(r)} \]  

(3.5)

where

\[ e_t^{(r)} \sim N(0, \Delta \sigma_r^2) \]  

(3.6)

### 3.2.3 Convenience Yield

We model the convenience yield \( \sigma_t^\delta \) by:

\[ \delta_{t+1} - \delta_t = [\kappa(\alpha - \delta_t) + \psi_\delta X_t + \psi_\delta h_t] \Delta + e_{t+1}^{\delta} \]  

(3.7)

where

\[ e_t^{(\delta)} = \beta^{(\delta)} e_t^{(V)} + \beta^{(\delta)} e_t^{(J)} + \bar{e}_t^{(\delta)} \]  

(3.8)

\[ \bar{e}_t^{(\delta)} \sim N(0, \Delta \sigma_r^2) \]  

(3.9)
3.2.4 Stochastic Volatility

We assume the stochastic volatility term has a \(GARCH(1,1)\) process, following the work of Heston and Nandi (2000). In other words we propose the model:

\[
\begin{align*}
    h_t &= \omega_h + b_h h_{t-1} + a_h(\bar{e}_t^{(V)} - c_h\sqrt{h_{t-1}}) \\
    \bar{e}_t^{(V)} &\sim N(0, 1)
\end{align*}
\]

(3.10)  (3.11)

Also the \(e_t^{(V)}\) term in equation 3.4 and 3.8 is:

\[
    e_t^{(V)} = \sqrt{h_{t-1}\Delta \bar{e}_t^{(V)}}
\]

(3.12)

3.2.5 News

We incorporate positive and negative news into modeling of the futures price. Denote the number of positive news arrived at time \(t\) by \(n_t^{(+)}\), and number of negative news arrived by \(n_t^{(-)}\). We assume each Poisson random variable with intensity \(\lambda_t^{(+)}\) and \(\lambda_t^{(-)}\), respectively. Taking positive news as an example, the intensity can be modeled using (negative process are modeled similarly):

\[
    \lambda_t^{(+)} = y_0^{(+)} + y_1^{(+)}n_{t-1}^{(+)}
\]

subject to the conditions \(y_0^{(+)} \geq 0\) and \(y_1^{(+)} \geq 0\). Negative news are modeled similarly.

We assume each positive and negative news will create a jump in the futures price, with jump sizes being \(\theta_j^{(+)}\) and \(\theta_j^{(-)}\), respectively. The jump sizes are assumed to have a normal distribution with jump mean size \(\mu_{\theta}^{(\pm)}\) and jump standard deviation \(\sigma_{\theta}^{(\pm)}\), i.e.:

\[
\begin{align*}
    \theta_j^{(+)} &\sim N(\mu_{\theta}^{(+)}, \sigma_{\theta}^{2(+)}), \\
    \theta_j^{(-)} &\sim N(\mu_{\theta}^{(-)}, \sigma_{\theta}^{2(-)})
\end{align*}
\]

(3.14)  (3.15)
The error term from jump component $e_t^{(J)}$ in equation 3.4 and 3.8 is then constructed as:

$$
e_t^{(J)} = \sum_{j=1}^{n_t^{(+)}} \theta_j^{(+)} + \sum_{j=1}^{n_t^{(-)}} \theta_j^{(-)}$$

(3.16)

### 3.2.6 Predicting Futures Price and Variance

In order to derive the futures price under the risk neutral measure (Q measure), we need the Radon-Nykodym derivative:

$$\frac{\Delta Q}{\Delta P} = \frac{\exp(-\Lambda \nu_{t+1})}{L(\Lambda; \nu_t)}$$

(3.17)

where $L(\Lambda; \nu_t)$ is an expectation under the pricing measure, or more precisely:

$$L(\Lambda; \nu_t) = E^P_t[\exp(-\Lambda \nu_{t+1})]$$

(3.18)

$$\nu_t = (e_t^{(V)}, e_t^{(\delta)}, e_t^{(r)}, \sum_{j=1}^{n_t^{(+)}} \theta_j^{(+)}, \sum_{j=1}^{n_t^{(-)}} \theta_j^{(-)})$$

(3.19)

$$\Lambda = (\Lambda^{(h)}, \Lambda^{(\delta)}, \Lambda^{(r)}, \Lambda^{(+)}, \Lambda^{(-)})$$

(3.20)

The vector $\nu_t$ represents the random terms from different factors, and $\Lambda$ vector represents the associated market prices of risk.

Assume a futures contract with maturity $T$, with $n$ intervals until maturity. The futures price at time $t$ is then given by:

$$F(t, T) = \exp(B_n + D_n^{(X)} X_t + D_n^{(r)} r_t - D_n^{(\delta)} \delta_t + D_n^{(+)} n_t^{(+)} + D_n^{(-)} n_t^{(-)} + G_n^{(h)} h_t)$$

(3.21)
where the coefficients are calculated in a recursive way, please see Appendix A for detailed derivations. The variance of futures price is given by:

$$\text{var}_t(R_{t+1}) = 2(G_{n-1}^{(h)}a_h)^2 + h_t(2G_{n-1}^{(h)}a_h c_h - u_{n-1}^{(V)} \sqrt{\Delta})^2$$

$$+ \sum_{j \in A} \lambda_t^{(j)} \Delta[(D_{n-1}^{(j)} + \mu_j u_{n-1}^{(j)})^2 + (\sigma_j u_{n-1}^{(j)})^2]$$

$$+(D_{n-1}^{(r)})^2 \sigma_t^2 \Delta + (D_{n-1}^{(\delta)})^2 \sigma_{(\delta)}^2 \Delta$$

(3.22) (3.23) (3.24)

where $u_{n-1}^{(V)} = D_{n-1}^{(X)} - \beta_D^{(V)} D_{n-1}^{(\delta)}$ and $u_{n-1}^{(J)} = D_{n-1}^{(X)} - \beta_D^{(J)} D_{n-1}^{(\delta)}$. 

3.2.7 Reduced Models without News

We can derive a sub-model by removing the news component, and end up with a four-factors model with a GARCH(1,1) process, motivated by [15]. Such a model has the following form:

$$X_{t+1} - X_t = (r_t - \delta_t) \Delta - l_t^{(X)} + \epsilon_{t+1}^{(X)}$$

$$\delta_{t+1} - \delta_t = [\kappa(\alpha - \delta_t) + \psi_D X_t + \psi_h h_t] \Delta + \epsilon_{t+1}^{(\delta)}$$

$$r_{t+1} - r_t = a(m - r_t) \Delta + \epsilon_{t+1}^{(r)}$$

$$h_t = w_h + b_h h_{t-1} + a_h (\epsilon_t^{(V)} - c_h \sqrt{h_{t-1}})^2$$

where:

$$l_t^{(X)} = \left(\frac{1}{2} - \Lambda_h\right) h_t \Delta$$

$$e_{t}^{(r)} = e_{t}^{(r)}$$

$$e_{t}^{(\delta)} = \beta_D^{(V)} e_{t}^{(V)} + \epsilon_{t}^{(\delta)}$$

$$e_{t}^{(X)} = e_{t}^{(V)}$$

(3.25) (3.26) (3.27) (3.28) (3.29) (3.30) (3.31) (3.32)
We will denote this sub-model as the "No News Model", and use it as a benchmark to compare against our full model.

### 3.3 Estimation

Our full model includes twenty eight parameters, summarizing them here:

1. Four interest rate parameters $a, m, \sigma_r$ and $\Lambda^{(r)}$.

2. Four GARCH parameters $\omega_h, a_h, b_h$ and $c_h$.

3. Seven convenience yield parameters $\kappa, \alpha, \sigma_\delta, \psi_\delta, \psi_h, \beta^V$ and $\beta^J$.

4. Eight jump parameters: $y_0^{(+/−)}, y_1^{(+/−)}, \mu^{(+/−)}$ and $\sigma^{(+/−)}$.

5. Eight market price or risk parameters: $\Lambda^{(h)}, \Lambda^{(r)}, \Lambda^{(δ)}$ and $\Lambda^{(+/−)}$.

Our estimation process proceeds as follows: we will first estimate the jump parameters $y_0^{(+/−)}, y_1^{(+/−)}$ using a maximum likelihood approach from the news time series, second estimate the interest rate parameters $a, m, \sigma_r$ from the three month Treasury data, and finally estimate the remaining 21 parameters using quasi-maximum likelihood approach and the Kalman filtering.

#### 3.3.1 Estimating the News Arrival Process

We assume that the news arrival process $n_t$ follows a Poisson process with intensity $\lambda_t$. For time interval $\Delta_t$, the number of news arrivals should follow a Poisson distribution, in order to obtain a maximum likelihood estimator for $\lambda_t$, the likelihood function is as follows: (assume there are in total $n_t$ news arrived at time $t$, and our sample period runs from $t = 1$ to $T$)

$$
\bar{L} = \prod_{t=1}^{T} \frac{\lambda^{n_t}}{n_t!} \exp(-\lambda)
$$

(3.33)
The logarithm of the likelihood function is

\[ L = \sum_{t=1}^{T} n_t \ln(\lambda) - \lambda - \ln(n_t!) \]  

(3.34)

Therefore,

\[ \frac{\partial L}{\partial \lambda} = \sum_{t=1}^{T} n_t \frac{1}{\lambda} - 1 \]  

(3.35)

so that

\[ \frac{1}{\lambda} \sum_{t=1}^{T} n_t = T \]  

(3.36)

which can be written

\[ \frac{1}{T} \sum_{t=1}^{T} n_t = \bar{\lambda} \]  

(3.37)

Hence

\[ \hat{\lambda} = \frac{m}{\sum_{i=1}^{m} X_i} = \frac{1}{T_n} \sum_{t=1}^{T} n_t = \bar{\lambda} \]  

(3.38)

We use one day as our unit time interval, and hence our daily intensity is:

\[ \hat{\lambda}^{(+/-)} = \sum_{i=1}^{n^{(+/-)}} I_{X_{t,i}}^{(+/-)} \]  

(3.39)

\[ = n_t \]  

(3.40)

Hence the arrival intensity of news during a time interval \( \Delta \) is \( \hat{\lambda}_{t}^{(+/-)} \Delta \), and it is modeled by:

\[ \lambda_{t}^{(+/-)} \Delta = y_{0}^{(+/-)} + y_{1}^{(+/-)} n_{t}^{(+/-)} \]  

(3.41)

where \( n_{t}^{(+)} \) represents the number of positive (negative) news that arrived in the prior period until \( t \). Since we estimate our model on a daily basis, equation 3.41 can be re-written as:

\[ \lambda_{t}^{(+/-)} = y_{0}^{(+/-)} + y_{1}^{(+/-)} n_{t-1}^{(+/-)} \]  

(3.42)
here $\lambda_t^{(+/-)}$ is the positive (negative) news arrival intensity on day $t$. We maximize the following log likelihood to estimate $y_{0}^{(+/-)}$ and $y_{1}^{(+/-)}$:

$$L^{(+/-)} = \sum_{t=1}^{T} n_{t}^{(+/-)} \ln(\lambda_{0}^{(+/-)} + \lambda_{1}^{(+/-)} n_{t-1}^{(+/-)}) - (\lambda_{0}^{(+/-)} + \lambda_{1}^{(+/-)} n_{t-1}) - \ln(n_{t}^{(+/-)}!)$$ (3.43)

The estimation results are presented in Table 3.8. Apart from the full sample period from January 8, 2004 to December 28, we also estimate the parameters for two sub-periods: the oil bubble period until June 30, 2009, and post bubble period after that.

### Table 3.8: Jump Intensity Parameters for News

<table>
<thead>
<tr>
<th></th>
<th>Bubble Period</th>
<th>PostBubble Period</th>
<th>Full Sample Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_{0}^{(+)}$</td>
<td>24.171 (0.413)</td>
<td>17.531 (0.446)</td>
<td>20.225 (0.296)</td>
</tr>
<tr>
<td>$y_{1}^{(+)}$</td>
<td>0.271 (0.012)</td>
<td>0.339 (0.016)</td>
<td>0.339 (0.009)</td>
</tr>
<tr>
<td>$y_{0}^{(-)}$</td>
<td>18.573 (0.335)</td>
<td>21.210 (0.471)</td>
<td>18.801 (0.269)</td>
</tr>
<tr>
<td>$y_{1}^{(-)}$</td>
<td>0.384 (0.011)</td>
<td>0.420 (0.013)</td>
<td>0.424 (0.008)</td>
</tr>
</tbody>
</table>

This table presents parameter separately estimated just using this filtered news data, for model

$$\lambda_t^{(j)} = y_{0}^{(j)} + y_{1}^{(j)} n_{t}^{(j)}$$

where $j \in (+, -)$ and $\lambda_t^{(j)}$ is the Poisson process intensity at time $t$.

#### 3.3.2 Estimating the Interest Rate Parameters $a$, $m$ and $\sigma_r$

We estimate $a$, $m$ and $\sigma_r$ separately under our mean-reverting assumption of the interest rate process, using the 3 month treasury rate data we obtained from Federal Reserve Bank of St. Louis. Note that although we estimate these parameters using the treasury data separately, in the Kalman filtering process the interest rate process is still treated as a latent variable, just with given parameters estimated from this step.

Table 3.9 panel A presents summary statistics of the daily treasury rate, from which we can observe
that the rate has two phases: during the Oil Bubble period (January 5, 2004 to June 30, 2009) the rate was volatile and at a relatively high level; after the bubble (July 1, 2009 to December 29, 2012) the rate basically become flat with low volatility level. Panel B shows parameters for the Bubble period, Post Bubble and the full sample period, respectively, and these parameters are consistent with the pattern we observed in panel A: higher reverting speed \((a)\), larger reverting mean \((m)\) and greater variance term \(\sigma_r\).

Table 3.9: Summary Statistics And Parameters Estimation for Interest Rate Process

### Panel A: Summary Statistics for Three-Month Treasury Rate Data

<table>
<thead>
<tr>
<th>Year</th>
<th>Mean</th>
<th>Median</th>
<th>StDev</th>
<th>Max</th>
<th>Min</th>
<th>ACF(1)</th>
<th>ACF(3)</th>
<th>ACF(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004</td>
<td>0.0140</td>
<td>0.0134</td>
<td>0.0045</td>
<td>0.0226</td>
<td>0.0087</td>
<td>0.26</td>
<td>0.04</td>
<td>0.15</td>
</tr>
<tr>
<td>2005</td>
<td>0.0322</td>
<td>0.0315</td>
<td>0.0052</td>
<td>0.0408</td>
<td>0.0231</td>
<td>0.23</td>
<td>0.14</td>
<td>0.26</td>
</tr>
<tr>
<td>2006</td>
<td>0.0485</td>
<td>0.0492</td>
<td>0.0024</td>
<td>0.0513</td>
<td>0.0416</td>
<td>0.20</td>
<td>0.10</td>
<td>0.23</td>
</tr>
<tr>
<td>2007</td>
<td>0.0447</td>
<td>0.0484</td>
<td>0.0071</td>
<td>0.0519</td>
<td>0.0287</td>
<td>0.10</td>
<td>0.16</td>
<td>0.22</td>
</tr>
<tr>
<td>2008</td>
<td>0.0139</td>
<td>0.0155</td>
<td>0.0081</td>
<td>0.0327</td>
<td>0.0000</td>
<td>0.26</td>
<td>0.16</td>
<td>0.32</td>
</tr>
<tr>
<td>2009</td>
<td>0.0015</td>
<td>0.0016</td>
<td>0.0007</td>
<td>0.0032</td>
<td>0.0002</td>
<td>0.11</td>
<td>0.14</td>
<td>0.23</td>
</tr>
<tr>
<td>2010</td>
<td>0.0014</td>
<td>0.0015</td>
<td>0.0003</td>
<td>0.0018</td>
<td>0.0004</td>
<td>0.21</td>
<td>0.07</td>
<td>0.05</td>
</tr>
<tr>
<td>2011</td>
<td>0.0005</td>
<td>0.0003</td>
<td>0.0005</td>
<td>0.0016</td>
<td>0.0000</td>
<td>0.31</td>
<td>0.22</td>
<td>0.19</td>
</tr>
<tr>
<td>2012</td>
<td>0.0009</td>
<td>0.0009</td>
<td>0.0002</td>
<td>0.0014</td>
<td>0.0001</td>
<td>0.20</td>
<td>0.08</td>
<td>0.07</td>
</tr>
</tbody>
</table>

### Panel B: Interest Rate Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Oil Bubble Period</th>
<th>Post Bubble Period</th>
<th>Full Sample Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>1.6062 (0.3180)</td>
<td>0.5824 (0.5004)</td>
<td>1.2127 (0.1789)</td>
</tr>
<tr>
<td>(m)</td>
<td>0.0275 (0.0033)</td>
<td>0.0016 (0.0010)</td>
<td>0.0167 (0.0027)</td>
</tr>
<tr>
<td>(\sigma_r)</td>
<td>0.0124 (0.0002)</td>
<td>0.0019 (0.0000)</td>
<td>0.0099 (0.0001)</td>
</tr>
</tbody>
</table>

Panel A provided summary statistics of the Three Month Treasury rate, obtained from Federal Reserve Bank of St. Louis, for a sample period from January 5, 2004 to December 28, 2012. Panel B presents parameter separately estimated just using this interest rate data, for a mean-reverting process:

\[
r_{t+1} - r_t = a(m - r_t)\Delta + e_{t+1}^{(r)}
\]

where \(e_{t+1}^{(r)} \sim N(0, \Delta \sigma_r^2)\). Full sample period is from January 5, 2004 to December 28, 2012. Bubble period is from January 5, 2004 to June 30, 2009; Post Bubble period is from July 1, 2009 to December 28, 2012.
3.3.3 Use Kalman Filter to Estimate the Remaining Parameters

For Kalman filtering, the measurement equation is

\[ y_t = c_t + Z_t \ast \alpha_t + G_t \ast \epsilon_t \] (3.44)

where \( \epsilon_t \sim N(0, I_n) \).

\[ y_t = \begin{pmatrix} ln(F(t, T_1)) \\ \vdots \\ ln(F(t, T_n)) \end{pmatrix} \] (3.45)

and

\[ \alpha_t = \begin{pmatrix} ln(S_t) \\ \delta_t \\ r_t \end{pmatrix} \] (3.46)

Where equation 3.45 is the vector of futures prices, and equation 3.46 is the vector of the state variables.

The transition equation is:

\[ \alpha_{t+1} = d_t + T_t \ast \alpha_t + H_{t+1} \ast \eta_{t+1} \] (3.47)

where \( \eta_t \sim N(0, I_3) \).

For the state variables we have: log spot price, convenience yield and interest rate, which are latent
variables in the model. For the spot price process, with \( X(t) = \ln(S(t)) \), we have:

\[
X(t + 1) - X(t) = (r_t - \delta_t)\Delta - l_t^Q + e_t^{(X)}
\]

\[
= r_t \Delta - \delta_t \Delta - l_t^Q + e_t^{(V)} + e_t^{(J)} + e_t^{(V)}
\]

\[
= [-l_t^Q + \sum_{j=1}^{n_t^{(+)}} \theta_j^{(+)} + \sum_{j=1}^{n_t^{(-)}} \theta_j^{(-)}] + r_t \Delta - \delta_t \Delta + \sqrt{h_t}e_t^{(V)}
\]

\[
+ \sum_{j=1}^{n_t^{(+)}} \sigma_j^{(+)} e_t^{(+)} + \sum_{j=1}^{n_t^{(-)}} \sigma_j^{(-)} e_t^{(-)} + \sqrt{h_t} \Delta e_t^{(V)}
\]

For the convenience yield process, we have:

\[
\delta(t + 1) - \delta(t) = [\kappa(\alpha - \delta_t) + \psi_\delta X_t + \psi_h h_t] \Delta + e_{t+1}^\delta
\]

\[
= [\kappa(\alpha - \delta_t) + \psi_\delta X_t + \psi_h h_t] \Delta
\]

\[
+ \beta_\delta^{(V)} e_t^{(V)} + \beta_\delta^{(J)} e_t^{(J)} + \sqrt{\Delta} \delta e_t^{(\delta)}
\]

\[
= (\kappa \alpha + \psi_h h_t) \Delta + \beta_\delta^{(V)} \left( \sum_{j=1}^{n_t^{(+)}} \theta_j^{(+)} + \sum_{j=1}^{n_t^{(-)}} \theta_j^{(-)} \right)
\]

\[
- \kappa \Delta \delta_t + \psi_\delta \Delta X_t + \sqrt{\Delta} \delta e_t^{(\delta)} + \beta_\delta^{(V)} e_t^{(V)}
\]

\[
= [(\kappa \alpha + \psi_h h_t) \Delta + \beta_\delta^{(V)} \left( \sum_{j=1}^{n_t^{(+)}} \theta_j^{(+)} + \sum_{j=1}^{n_t^{(-)}} \theta_j^{(-)} \right)]
\]

\[
- \kappa \Delta \delta_t + \psi_\delta \Delta X_t + \sqrt{\Delta} \delta e_t^{(\delta)}
\]

\[
+ \beta_\delta^{(V)} \sqrt{h_t} \Delta e_t^{(V)} + \beta_\delta^{(J)} \left( \sum_{j=1}^{n_t^{(+)}} \sigma_j^{(+)} e_t^{(+)} + \sum_{j=1}^{n_t^{(-)}} \sigma_j^{(-)} e_t^{(-)} \right)
\]

For the interest rate process, we have:

\[
r(t + 1) - r(t) = a \Delta (m - r_t) + e_{t+1}^r
\]

\[
= a * m \Delta - a \Delta r_t + \sigma_r \Delta e_{t+1}^{(r)}
\]
Hence equation 3.44 becomes

\[
\begin{pmatrix}
X_{t+1} \\
\delta_{t+1} \\
r_{t+1}
\end{pmatrix} =
\begin{pmatrix}
\frac{d_t^{(X)}}{d_t^{(\delta)}} \\
a * m \Delta
\end{pmatrix} +
\begin{pmatrix}
1 & -\Delta & \Delta \\
\psi(\delta) \Delta & 1 - \kappa \Delta & 0 \\
0 & 0 & 1 - a \Delta
\end{pmatrix}
\begin{pmatrix}
X_t \\
\delta_t \\
r_t
\end{pmatrix} + H_t
\begin{pmatrix}
\frac{\epsilon_t^{(X)}}{\epsilon_t^{(\delta)}}
\end{pmatrix}
\]

where

\[
d_t^{(X)} = r_t \Delta - \nu Q_t + \sum_{j=1}^{n_t^{(+)}} \mu_j^{(+)} + \sum_{j=1}^{n_t^{(-)}} \mu_j^{(-)}
\]

\[
d_t^{(\delta)} = (\kappa \alpha + \psi_h r_t) \Delta + \beta_t^{(\delta)} \left( \sum_{j=1}^{n_t^{(+)}} \mu_j^{(+)} + \sum_{j=1}^{n_t^{(-)}} \mu_j^{(-)} \right)
\]

and

\[
H_t^2 =
\begin{bmatrix}
\sigma_{(X)}^2 & \sigma_{(X,\delta)} & 0 \\
\sigma_{(X,\delta)} & \sigma_{(\delta)}^2 & 0 \\
0 & 0 & \sigma_r^2 \Delta
\end{bmatrix}
\]

where

\[
\sigma_{(X)}^2 = n_t^{(+)}(\sigma_j^{(+)} )^2 + n_t^{(-)}(\sigma_j^{(-)} )^2 + h_t \Delta
\]

\[
\sigma_{(\delta)}^2 = (\beta_t^{(\delta)} )^2 h_t \Delta + (\beta_t^{(\delta)} )^2 [n_t^{(+)}(\sigma_j^{(+)} )^2 + n_t^{(-)}(\sigma_j^{(-)} )^2] + \sigma_r^2 \Delta
\]

\[
\sigma_{(X,\delta)} = (\beta_t^{(\delta)} )h_t \Delta + (\beta_t^{(\delta)} )[n_t^{(+)}(\sigma_j^{(+)} )^2 + n_t^{(-)}(\sigma_j^{(-)} )^2]
\]

It follows that futures price is given by:

\[
E_t^Q[S(T)] = \exp(B_t + D_t^{(X)} X_t + D_t^{(\delta)} r_t - D_t^{(\delta)} \delta_t + D_t^{(+)} n_t^{(+)}) + D_t^{(-)} n_t^{(-)} + G_t^{(h)} h_t)
\]
Equation 3.47 can be written as:

\[
\begin{pmatrix}
\ln(F(t, T_1)) \\
\vdots \\
\ln(F(t, T_n))
\end{pmatrix} =
\begin{pmatrix}
c_{T_1} \\
\vdots \\
c_{T_n}
\end{pmatrix} +
\begin{pmatrix}
D^{(X)}_{nT_1} - D^{(\delta)}_{nT_1} & D^{(r)}_{nT_1} \\
\vdots & \vdots \\
D^{(X)}_{nT_n} - D^{(\delta)}_{nT_n} & D^{(r)}_{nT_n}
\end{pmatrix} \begin{pmatrix}
X_t \\
\delta_t \\
r_t \\
\epsilon_t^{(T_1)}
\end{pmatrix} +
\begin{pmatrix}
G_t \\
\vdots
\end{pmatrix}
\]

where

\[
c_{T_i} = B^{(r)}_{nT_i} r_t + D^{(\delta)}_{nT_i} n_t^{(+)} + D^{(-)}_{nT_i} n_t^{(-)} + G^{(h)}_{nT_i} h_t
\]

\[
G_t^2 = \text{Diag}[(\sigma_t^{T_1})^2, \ldots, (\sigma_t^{T_n})^2]
\]

\[
(\sigma_t^{T_i})^2 = 2(G^{(h)}_{nT_i} a_h)^2 + h_{t-1} (2G^{(h)}_{nT_i} a_h c_h - u^{(V)}_{nT_i} \sqrt{\Delta})^2
\]

\[
+ \sum_{j \in A} \lambda_{j-1}^{(j)} \Delta [(D^{(j)}_{nT_i} + \mu_j u^{(J)}_{nT_i})^2 + (\sigma^{(j)} u^{(J)}_{nT_i})^2]
\]

\[
(D^{(r)}_{nT_i})^2 \sigma_t^2 \Delta + (D^{(\delta)}_{nT_i})^2 \sigma_t^{(\delta)} \Delta
\]

We perform quasi-maximum likelihood optimization with Kalman filter to obtain the parameter estimations, we used the FKF R package to do the filtering.

Note that although the GARCH process \( h_t \) is assumed to be a latent variable, it is not a linear process and also does not have constant variance. We simplified the estimation on the GARCH process while still keeping it as a latent factor, by assuming the variance of the front contract (F1) is measured without error, and is equal to the realized variance measure on that day, i.e \( \text{Var}(\ln(F(t, T_1))) = \sigma^2_{t,F_1} \) and \( \sigma^2_{t,F_1} \) is the realized variance for F1 at time t. To do that, we
need to set:

\[ \text{var}_t(R_t) = \sigma^2_{t,F_1} \]

\[ = 2(G_n^{(h)}a_h)^2 + h_{t-1}(2G_n^{(h)}a_h c_h - u_n^{(V)} \Delta )^2 \]  \hspace{1cm} (3.69)

\[ + \sum_{j \in A} \lambda^{(j)}_{t-1} \Delta [(D_n^{(j)} + \mu_j u_n^{(j)})^2 + (\sigma_{(j)} u_n^{(j)})^2] \]  \hspace{1cm} (3.70)

\[ (D_n^{(r)})^2 \sigma^2_r \Delta + (D_n^{(\delta)})^2 \sigma^2_{(\delta)} \Delta \]  \hspace{1cm} (3.71)

\[ (D_n^{(r)})^2 \sigma^2_r \Delta + (D_n^{(\delta)})^2 \sigma^2_{(\delta)} \Delta \]  \hspace{1cm} (3.72)

Hence given the parameter values and \( \lambda_{t-1} \), we can solve for \( h_{t-1} \) as a latent variable. We then use this latent variable along with others \((X_t, \delta_t \text{ and } r_t)\) to calculate the likelihood for the given parameters.

### 3.4 Fitting Result

#### 3.4.1 Parameter Estimation and Properties

Table 3.10 reported the parameter estimate properties on the full sample period, for the No News model and With News model. Overall we observe a higher likelihood and lower corrected AIC for the full model over the GARCH model. Also we observed that on average 39.42% of the total conditional variance is explained by news, where 30.76% explained by negative news and 8.66% explained by positive news. This asymmetrical explanatory power is consistent with our findings in the empirical checks. In terms of root means squared errors (RMSE), the With News model performs better for contracts with closer maturities, while the No News model is better for far contracts, but overall the RMSE is close.

Taking a closer look at the parameters, most of them are significant, and the ones that are not significant relate to the market prices of risks. For the GARCH parameters, we observe lower \( w_h, a_h \) and \( b_h \) after news was added to the model, indicating a smaller intercept for stochastic volatility process and weaker autoregressive behavior, confirming with previous preliminary checks where we find news can help explain volatility clustering. Regarding the jump parameters, the jump mean
for positive news ($\mu^{(+)}$) is smaller in magnitude compared to that of negative news ($\mu^{(-)}$), so are the jump standard deviations. This suggests that negative news have higher per item impact on prices than positive news, consistent with our previous findings.

### 3.4.2 Price Fitting

Table 3.11 reported summary statistics of the fitted futures price, constructed very similarly to table 3.1. By comparing Panel A of these two tables, we find that we can fit the first two moments rather well, while for skewness and kurtosis we tend to over-estimate for closer contracts and under-estimate for far contracts. For Panel B we can reproduce the same panel from table 3.1 well, with a tendency to slightly over-estimate the variance.

### 3.4.3 Conditional Variance

Figure 3.5 plots the log of conditional variance for model with news against log of conditional variance for model without news. Panel A is the paired plot for conditional variance just from the GARCH process, we find that without news the model tend to over-estimate the conditional variance, which is consistent with findings by Engle, Hansen and Lunde (2012) [20]. Panel B compares the total conditional variance, the No News model under-estimate for low volatility days and over-estimate for high-volatility days.

### 3.5 Robustness

#### 3.5.1 Estimation on Sub-Samples

From Figure 3.1 Panel A we find that the crude oil price undergo a very volatile period for our sample: it first climbed up to $147 per barrel, then drop to under $40 quickly, after that the price become more flat. Similarly, interest rate also have very different behavior for the oil bubble period and post bubble period, as shown in table 3.9: rate increased from 0.014 in 2004 to 0.0447
Table 3.10: Estimation Results for Full Sample Period

<table>
<thead>
<tr>
<th></th>
<th>No News</th>
<th>With News</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GARCH Parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\omega_h$</td>
<td>7.55e-03(1.69e-03)***</td>
<td>4.97e-03(1.89e-03)***</td>
</tr>
<tr>
<td>$\alpha_h$</td>
<td>2.93e-03(1.26e-06)***</td>
<td>2.06e-03(6.88e-06)***</td>
</tr>
<tr>
<td>$\gamma_h$</td>
<td>0.81(4.18e-06)***</td>
<td>0.76(1.34e-05)***</td>
</tr>
<tr>
<td>$\delta_h$</td>
<td>7.70(1.39e-03)***</td>
<td>9.90(2.06e-03)***</td>
</tr>
<tr>
<td><strong>Convenience Yield Parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\kappa$</td>
<td>1.83(3.06e-02)***</td>
<td>1.67(2.61e-02)***</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.34(1.63e-02)***</td>
<td>0.32(3.55e-02)***</td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.17(2.38e-03)***</td>
<td>3.04e-02(1.05e-02)***</td>
</tr>
<tr>
<td>$\Psi_d$</td>
<td>-6.99e-02(4.75e-03)***</td>
<td>-9.58e-02(4.62e-03)***</td>
</tr>
<tr>
<td>$\Psi_h$</td>
<td>-2.75e-01(2.08e-02)***</td>
<td>-9.62e-01(0.09)***</td>
</tr>
<tr>
<td>$\beta_V$</td>
<td>1.81(2.11e-02)***</td>
<td>1.79(1.87e-02)***</td>
</tr>
<tr>
<td>$\beta_J$</td>
<td>-8.3e+00(0.06)***</td>
<td>-8.3e+00(0.06)***</td>
</tr>
<tr>
<td><strong>Jumps Size Parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu^{(+)}$</td>
<td>6.98e-05(3.32e-06)***</td>
<td>5e-05(8.49e-06)***</td>
</tr>
<tr>
<td>$\sigma^{(+)}$</td>
<td>-7.82e-05(3.43e-06)***</td>
<td>-7.82e-05(3.43e-06)***</td>
</tr>
<tr>
<td>$\mu^{(-)}$</td>
<td>1.38e-04(5.06e-06)***</td>
<td>1.38e-04(5.06e-06)***</td>
</tr>
<tr>
<td>$\sigma^{(-)}$</td>
<td>-3.44e+00(6.11)</td>
<td>-3.44e+00(6.11)</td>
</tr>
<tr>
<td><strong>Market Price of Risks Parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Lambda^p$</td>
<td>0.98(2.94e-02)***</td>
<td>0.49(4.72e-02)***</td>
</tr>
<tr>
<td>$\Lambda^d$</td>
<td>1.49(2.55e-03)***</td>
<td>2.90(&gt; 10)</td>
</tr>
<tr>
<td>$\Lambda^r$</td>
<td>2.90(&gt; 10)</td>
<td>2.76(&gt; 10)</td>
</tr>
<tr>
<td>$\Lambda^{(+)}$</td>
<td>-3.44e+00(6.11)</td>
<td>1.17(3.57)</td>
</tr>
<tr>
<td>$\Lambda^{(-)}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Properties</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>92556.9</td>
<td>92814.1</td>
</tr>
<tr>
<td>Corrected AIC</td>
<td>-185095.6</td>
<td>-185587.5</td>
</tr>
<tr>
<td>LR Test p Value</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Average RMSE</td>
<td>0.00554</td>
<td>0.00577</td>
</tr>
<tr>
<td>F1 RMSE</td>
<td>0.01336</td>
<td>0.01247</td>
</tr>
<tr>
<td>F3 RMSE</td>
<td>0.00447</td>
<td>0.00422</td>
</tr>
<tr>
<td>F6 RMSE</td>
<td>0.0034</td>
<td>0.00408</td>
</tr>
<tr>
<td>% Var from News</td>
<td>39.42</td>
<td></td>
</tr>
<tr>
<td>% Var from (+) News</td>
<td>8.66</td>
<td></td>
</tr>
<tr>
<td>% Var from (−) News</td>
<td>30.76</td>
<td></td>
</tr>
</tbody>
</table>

This table compares parameter estimations and properties between the model with news and model without news, for the full sample period from January 5, 2004 to December 28, 2012. Standard error of parameters are shown in the parenthesis following the estimation value, with significance marker attached: a * indicate significance at 0.05 level, a ** indicate significance at 0.01 level, a *** indicate significance at 0.001 level. For the properties panel we report the converged log likelihood from the maximum likelihood estimation, the corrected AIC, the p value from a likelihood ratio test between these two nested models, and the average conditional variance percentage explained by news, also by positive and negative news, respectively.
Table 3.11: Summary Statistics for Returns of Filtered Futures Price

<table>
<thead>
<tr>
<th>Panel A: Statistics for Filtered F1-F12 Price</th>
<th>$\hat{\mu}$</th>
<th>$\hat{\sigma}^2$</th>
<th>$\hat{\gamma}$</th>
<th>$\hat{\kappa}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>4.49e-04</td>
<td>7.25e-04</td>
<td>0.078</td>
<td>6.409</td>
</tr>
<tr>
<td>F2</td>
<td>4.59e-04</td>
<td>6.07e-04</td>
<td>0.037</td>
<td>5.857</td>
</tr>
<tr>
<td>F3</td>
<td>4.67e-04</td>
<td>5.14e-04</td>
<td>-0.018</td>
<td>5.214</td>
</tr>
<tr>
<td>F4</td>
<td>4.75e-04</td>
<td>4.41e-04</td>
<td>-0.072</td>
<td>4.578</td>
</tr>
<tr>
<td>F5</td>
<td>4.82e-04</td>
<td>3.84e-04</td>
<td>-0.115</td>
<td>4.028</td>
</tr>
<tr>
<td>F6</td>
<td>4.88e-04</td>
<td>3.40e-04</td>
<td>-0.155</td>
<td>3.530</td>
</tr>
<tr>
<td>F7</td>
<td>4.94e-04</td>
<td>3.05e-04</td>
<td>-0.187</td>
<td>3.109</td>
</tr>
<tr>
<td>F8</td>
<td>4.99e-04</td>
<td>2.78e-04</td>
<td>-0.214</td>
<td>2.754</td>
</tr>
<tr>
<td>F9</td>
<td>5.03e-04</td>
<td>2.57e-04</td>
<td>-0.237</td>
<td>2.453</td>
</tr>
<tr>
<td>F10</td>
<td>5.08e-04</td>
<td>2.40e-04</td>
<td>-0.251</td>
<td>2.220</td>
</tr>
<tr>
<td>F11</td>
<td>5.12e-04</td>
<td>2.27e-04</td>
<td>-0.262</td>
<td>2.035</td>
</tr>
<tr>
<td>F12</td>
<td>5.16e-04</td>
<td>2.17e-04</td>
<td>-0.271</td>
<td>1.875</td>
</tr>
<tr>
<td>Average</td>
<td>4.88e-04</td>
<td>3.78e-04</td>
<td>-0.073</td>
<td>6.203</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: Statistics for Filtered F1 Price</th>
<th>$\hat{\mu}$</th>
<th>$\hat{\sigma}^2$</th>
<th>$\hat{\gamma}$</th>
<th>$\hat{\kappa}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004</td>
<td>1.16e-03</td>
<td>6.62e-04</td>
<td>-0.330</td>
<td>1.221</td>
</tr>
<tr>
<td>2005</td>
<td>1.30e-03</td>
<td>4.35e-04</td>
<td>0.282</td>
<td>0.220</td>
</tr>
<tr>
<td>2006</td>
<td>-6.32e-06</td>
<td>2.80e-04</td>
<td>0.047</td>
<td>-0.296</td>
</tr>
<tr>
<td>2007</td>
<td>1.84e-03</td>
<td>3.64e-04</td>
<td>-0.234</td>
<td>0.299</td>
</tr>
<tr>
<td>2008</td>
<td>-2.68e-03</td>
<td>2.26e-03</td>
<td>0.389</td>
<td>2.693</td>
</tr>
<tr>
<td>2009</td>
<td>1.86e-03</td>
<td>1.29e-03</td>
<td>0.004</td>
<td>2.818</td>
</tr>
<tr>
<td>2010</td>
<td>6.13e-04</td>
<td>3.46e-04</td>
<td>-0.330</td>
<td>0.434</td>
</tr>
<tr>
<td>2011</td>
<td>3.30e-04</td>
<td>5.91e-04</td>
<td>-0.858</td>
<td>4.053</td>
</tr>
<tr>
<td>2012</td>
<td>-3.56e-04</td>
<td>2.87e-04</td>
<td>0.704</td>
<td>5.281</td>
</tr>
</tbody>
</table>

This table presents summary statistics for returns of filtered futures price. We reported the mean, annualized standard deviation, skewness and excess kurtosis of the returns, defined similarly as in table 3.1.

in 2007, and then collapsed after the financial crisis. Hence we divide our sample period into two sub periods: the oil bubble period from January 8, 2004 to June 30, 2009; the post bubble period from July 1, 2009 to December 28, 2012. We estimated our model on these two sub periods, and reported the results in table 3.12.

Most of the parameters are relatively stable for the two periods. The GARCH process has a higher constant intercept ($w_h$) term during the oil bubble, as the market is highly volatile for this period. The jump parameters are relatively stable during the two periods, consider that we have more news during the oil bubble period, news actually contributes more to pricing. We observe higher RMSE levels for the oil bubble period than the post bubble period, as the volatile market condition is
making it more difficult to fit the model. We observe a higher percentage of conditional variance explained by news: 42.39% for the bubble period compared to 37.86% for the post bubble period.

3.5.2 Estimation Results for Different Sentiment Probability Thresholds

In table 3.4 we examined the impact of applying different sentiment probability thresholds on the news data, and we find that we need $\pi \geq 0.65$ to remove noisy items and correlation between positive and negative news. We used $\pi = 0.75$ in our model estimation, and here we check the result with $\pi = 0.65$. Table 3.13 compares the estimation results for the With News model using data obtained using $\pi = 0.65$ versus using $\pi = 0.75$.

The news arrival intensity parameters have increased for the $\pi = 0.65$ scenario, as we have more news item under this threshold. The Jump size parameters have decreased, indicating smaller per
item impact from news, which is as expected since we added more noisy to the news time series with $\pi = 0.65$. Changes in other parameters are mostly small, and the properties of these two sets of results are very close as well, with $\pi = 0.75$ providing slightly higher likelihood and smaller RMSE statistics.

### 3.5.3 Out of Sample Estimation

In this section we estimate the model only using the odd contracts (F1, F3, F5, F7, F9, F11), and see if they can predict the prices for the even contracts. Table 3.14 reported the estimation results, and for the properties we first estimate the parameters using either all contracts (first column) or odd contracts (second column), and then use these parameters to calculate the log likelihood and RMSEs for all the contracts. The differences in parameters are generally very small, and using all contracts provides a higher likelihood than just using odd contracts. The RMSEs are similar as well: the RMSEs first decrease (F1 to F3), and then increase afterwards (F4 to F12). In the parenthesis we used bootstrapping method to calculate the 99% confidence interval of the RMSEs, and we found that after the differences in RMSE are not significant at this confidence level.
### Table 3.12: Robustness: Comparing Results for Sub-samples

<table>
<thead>
<tr>
<th></th>
<th>Oil Bubble Period</th>
<th>Post Bubble Period</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No News</td>
<td>With News</td>
</tr>
<tr>
<td><strong>GARCH Parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$w_h$</td>
<td>4.88e-02(4.88e-03)***</td>
<td>2.74e-02(5.27e-03)***</td>
</tr>
<tr>
<td>$a_h$</td>
<td>3.29e-06(8.94e-07)***</td>
<td>1.43e-05(1.6e-06)***</td>
</tr>
<tr>
<td>$b_h$</td>
<td>0.91(4.54e-06)***</td>
<td>0.90(6.95e-05)***</td>
</tr>
<tr>
<td>$c_h$</td>
<td>10.37(&gt; 10)</td>
<td>10.61(6.62)</td>
</tr>
<tr>
<td><strong>Convenience Yield Parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\kappa$</td>
<td>1.83(2.56e-02)***</td>
<td>1.62(2.35e-02)***</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.26(4.51e-02)***</td>
<td>0.29(1.08e-03)***</td>
</tr>
<tr>
<td>$\sigma_\delta$</td>
<td>0.18(2.88e-03)***</td>
<td>3.03e-02(1.49e-02)**</td>
</tr>
<tr>
<td>$\Psi^\delta$</td>
<td>-1.1e-01(5.92e-03)***</td>
<td>-9.42e-02(6.86e-03)***</td>
</tr>
<tr>
<td>$\Psi^h$</td>
<td>-3.94e-01(0.10)***</td>
<td>3.46e-02(0.08)</td>
</tr>
<tr>
<td>$\beta^V$</td>
<td>1.87(1.93e-02)***</td>
<td>1.84(1.96e-02)***</td>
</tr>
<tr>
<td>$\beta^J$</td>
<td>-6.64e+00(1.04)***</td>
<td></td>
</tr>
<tr>
<td><strong>Jumps Size Parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu^{(+)}$</td>
<td>6.22e-05(4.98e-06)***</td>
<td></td>
</tr>
<tr>
<td>$\sigma^{(+)}$</td>
<td>5.01e-05(1.61e-05)***</td>
<td></td>
</tr>
<tr>
<td>$\mu^{(-)}$</td>
<td>-7.58e-05(6.42e-06)***</td>
<td></td>
</tr>
<tr>
<td>$\sigma^{(-)}$</td>
<td>2.18e-04(2.33e-05)***</td>
<td></td>
</tr>
<tr>
<td><strong>Market Price of Risks Parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Lambda^{(+)}$</td>
<td>0.14(2.01e-02)***</td>
<td>0.11(0.05)***</td>
</tr>
<tr>
<td>$\Lambda^{(-)}$</td>
<td>3.45(&gt; 10)</td>
<td>1.17(&gt; 10)</td>
</tr>
<tr>
<td>$\Lambda^{(0)}$</td>
<td>-3.84e+00(&gt; 10)</td>
<td>3.88(&gt; 10)</td>
</tr>
<tr>
<td>$\Lambda^{(-)}$</td>
<td>1.74(2.69)</td>
<td></td>
</tr>
<tr>
<td><strong>Properties</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>54980.8</td>
<td>55135.9</td>
</tr>
<tr>
<td>Corrected AIC</td>
<td>-109943.3</td>
<td>-110230.9</td>
</tr>
<tr>
<td>LR Test p Value</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Average RMSE</td>
<td>0.00612</td>
<td>0.00571</td>
</tr>
<tr>
<td>F1 RMSE</td>
<td>0.0146</td>
<td>0.01318</td>
</tr>
<tr>
<td>F3 RMSE</td>
<td>0.00485</td>
<td>0.00403</td>
</tr>
<tr>
<td>F6 RMSE</td>
<td>0.00388</td>
<td>0.00392</td>
</tr>
<tr>
<td>% Var from News</td>
<td>42.39</td>
<td></td>
</tr>
<tr>
<td>% Var from (+) News</td>
<td>5.06</td>
<td></td>
</tr>
<tr>
<td>% Var from (-) News</td>
<td>37.33</td>
<td></td>
</tr>
</tbody>
</table>

This table test the robustness of our estimation by comparing result on the Oil Bubble period (January 5, 2004 to June 30, 2009) and Post Bubble period (July 1, 2009 to December 28, 2012). Standard error of parameters are shown in the parenthesis following the estimation value, with significance marker attached: a * indicate significance at 0.05 level, a ** indicate significance at 0.01 level, a *** indicate significance at 0.001 level.

For the properties panel, we report the converged log likelihood of the ML estimators, the corrected AIC, the p value from a likelihood ratio test between these two nested models, and the average conditional variance percentage explained by news, also by positive and negative news, respectively.
Table 3.13: Robustness: Estimation Results for Different Sentiment Probability Thresholds

<table>
<thead>
<tr>
<th>Parameters</th>
<th>$\pi = 0.65$</th>
<th>$\pi = 0.75$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GARCH Parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$w_h$</td>
<td>7.07e-03(2.18e-03)***</td>
<td>4.97e-03(1.89e-03)***</td>
</tr>
<tr>
<td>$q_h$</td>
<td>8.27e-05(3.58e-05)***</td>
<td>2.06e-03(6.88e-06)***</td>
</tr>
<tr>
<td>$b_h$</td>
<td>0.90(2.18e-03)***</td>
<td>0.76(1.34e-05)***</td>
</tr>
<tr>
<td>$c_h$</td>
<td>11.44(7.59)</td>
<td>9.90(2.06e-03)***</td>
</tr>
<tr>
<td><strong>Convenience Yield Parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\kappa$</td>
<td>1.52(2.68e-02)***</td>
<td>1.67(2.61e-02)***</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.13(4.54e-02)***</td>
<td>0.32(3.55e-02)***</td>
</tr>
<tr>
<td>$\sigma_\delta$</td>
<td>3.06e-02(1.28e-02)***</td>
<td>3.04e-02(1.05e-02)***</td>
</tr>
<tr>
<td>$\Psi_d$</td>
<td>-2.89e-02(4.33e-03)***</td>
<td>-9.58e-02(4.62e-03)***</td>
</tr>
<tr>
<td>$\Psi_h$</td>
<td>-1.49e-01(1.65e-02)***</td>
<td>-9.62e-01(0.09)***</td>
</tr>
<tr>
<td>$\beta_{V}$</td>
<td>1.72(1.92e-02)***</td>
<td>1.79(1.87e-02)***</td>
</tr>
<tr>
<td>$\beta_{I}$</td>
<td>-9.43e+00(0.51)***</td>
<td>-8.3e+00(0.06)***</td>
</tr>
<tr>
<td><strong>Jumps Size Parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu^{(+)}$</td>
<td>5.57e-05(2.79e-06)***</td>
<td>6.98e-05(3.32e-06)***</td>
</tr>
<tr>
<td>$\sigma^{(+)}$</td>
<td>3.41e-05(1.07e-05)***</td>
<td>5e-05(8.49e-06)***</td>
</tr>
<tr>
<td>$\mu^{(-)}$</td>
<td>-6.79e-05(3.74e-06)***</td>
<td>-7.82e-05(3.43e-06)***</td>
</tr>
<tr>
<td>$\sigma^{(-)}$</td>
<td>8.76e-05(7.64e-06)***</td>
<td>1.38e-04(5.06e-06)***</td>
</tr>
<tr>
<td><strong>News Arrival Intensity Parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$y_{t}^{(+)}$</td>
<td>32.06(0.43)***</td>
<td>20.23(0.30)***</td>
</tr>
<tr>
<td>$y_{t}^{(-)}$</td>
<td>0.40(7.87e-03)***</td>
<td>0.34(9.42e-03)***</td>
</tr>
<tr>
<td>$y_{t}^{(+)}$</td>
<td>26.42(0.36)***</td>
<td>18.80(0.27)***</td>
</tr>
<tr>
<td>$y_{t}^{(-)}$</td>
<td>0.43(7.59e-03)***</td>
<td>0.42(8.1e-03)***</td>
</tr>
<tr>
<td><strong>Market Price of Risks Parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Lambda^{(s)}$</td>
<td>1.05(0.12)***</td>
<td>0.49(4.72e-02)***</td>
</tr>
<tr>
<td>$\Lambda^{(r)}$</td>
<td>1.73(&gt; 10)</td>
<td>2.90(&gt; 10)</td>
</tr>
<tr>
<td>$\Lambda^{(+)}$</td>
<td>1.68(&gt; 10)</td>
<td>2.76(&gt; 10)</td>
</tr>
<tr>
<td>$\Lambda^{(-)}$</td>
<td>-1.3e+00(0.56)**</td>
<td>-3.44e+00(6.11)</td>
</tr>
<tr>
<td><strong>Properties</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>92682.4</td>
<td>92814.1</td>
</tr>
<tr>
<td>Corrected AIC</td>
<td>-185323.8</td>
<td>-185587.5</td>
</tr>
<tr>
<td>LR Test p Value</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Average RMSE</td>
<td>0.00596</td>
<td>0.00577</td>
</tr>
<tr>
<td>F1 RMSE</td>
<td>0.0123</td>
<td>0.01247</td>
</tr>
<tr>
<td>F3 RMSE</td>
<td>0.00408</td>
<td>0.00422</td>
</tr>
<tr>
<td>F6 RMSE</td>
<td>0.00449</td>
<td>0.00408</td>
</tr>
<tr>
<td>F12 RMSE</td>
<td>0.00608</td>
<td>0.00578</td>
</tr>
<tr>
<td>% Var from News</td>
<td>38.97</td>
<td>39.42</td>
</tr>
<tr>
<td>% Var from (+) News</td>
<td>11.26</td>
<td>8.66</td>
</tr>
<tr>
<td>% Var from (−) News</td>
<td>27.71</td>
<td>30.76</td>
</tr>
</tbody>
</table>

This table tests the robustness of our estimation by comparing results on full sample periods (January 5, 2004 to December 28, 2012) using news data under a threshold (π) equal to 0.65 and 0.75, respectively. Standard error of parameters are shown in the parenthesis following the estimation value, with significance marker attached: a * indicate significance at 0.05 level, a ** indicate significance at 0.01 level, a *** indicate significance at 0.001 level. For the properties panel, we report the converged log likelihood of the ML estimators, the corrected AIC, the p value from a likelihood ratio test between these two nested models, and the average conditional variance percentage explained by news, also by positive and negative news, respectively.
Table 3.14: Robustness: Estimation Results Using Different Contracts

<table>
<thead>
<tr>
<th></th>
<th>Using All Contracts</th>
<th>Using Odd Contracts</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GARCH Parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$w_h$</td>
<td>4.97e-03(1.89e-03)**</td>
<td>4.98e-03(6.91e-04)**</td>
</tr>
<tr>
<td>$a_h$</td>
<td>2.06e-03(6.88e-06)**</td>
<td>1.94e-03(6.32e-07)**</td>
</tr>
<tr>
<td>$b_h$</td>
<td>0.76(1.34e-05)**</td>
<td>0.76(3.29e-06)**</td>
</tr>
<tr>
<td>$c_h$</td>
<td>9.90(2.06e-03)**</td>
<td>10.44(3.85e-04)**</td>
</tr>
<tr>
<td><strong>Convenience Yield Parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\kappa$</td>
<td>1.67(2.51e-02)**</td>
<td>1.45(3.68e-02)**</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.32(3.55e-02)**</td>
<td>0.14(4.31e-02)**</td>
</tr>
<tr>
<td>$\sigma_\delta$</td>
<td>3.04e-02(1.05e-02)**</td>
<td>4.05e-02(2.41e-02)*</td>
</tr>
<tr>
<td>$\Psi_\delta$</td>
<td>-9.58e-02(4.62e-03)**</td>
<td>-9.6e-02(7.4e-03)**</td>
</tr>
<tr>
<td>$\Psi_\beta$</td>
<td>-9.62e-01(0.09)**</td>
<td>-6.59e-01(0.06)**</td>
</tr>
<tr>
<td>$\beta^V$</td>
<td>1.79(1.87e-02)**</td>
<td>1.36(2.48e-02)**</td>
</tr>
<tr>
<td>$\beta^J$</td>
<td>-8.3e+00(0.06)**</td>
<td>-7.27e+00(3.19e-02)**</td>
</tr>
<tr>
<td><strong>Jumps Size Parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu^{(+)}$</td>
<td>6.98e-05(3.32e-06)**</td>
<td>6.34e-05(5.27e-06)**</td>
</tr>
<tr>
<td>$\sigma^{(+)}$</td>
<td>5.05(8.49e-06)**</td>
<td>5.1e-05(2.78e-05)*</td>
</tr>
<tr>
<td>$\mu^{(-)}$</td>
<td>-7.82e-05(3.43e-06)**</td>
<td>-8.18e-05(5.67e-06)**</td>
</tr>
<tr>
<td>$\sigma^{(-)}$</td>
<td>1.38e-04(5.06e-06)**</td>
<td>1.39e-04(9.52e-06)**</td>
</tr>
<tr>
<td><strong>Market Price of Risks Parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Lambda^{E}$</td>
<td>0.49(4.72e-02)**</td>
<td>0.71(0.07)*****</td>
</tr>
<tr>
<td>$\Lambda^{(i)}$</td>
<td>2.90(&gt; 10)</td>
<td>2.19(&gt; 10)</td>
</tr>
<tr>
<td>$\Lambda^{r}$</td>
<td>2.76(&gt; 10)</td>
<td>2.53(&gt; 10)</td>
</tr>
<tr>
<td>$\Lambda^{(+)}$</td>
<td>-3.44e+00(6.11)</td>
<td>-3.09e+00(&gt; 10)</td>
</tr>
<tr>
<td>$\Lambda^{(-)}$</td>
<td>1.17(3.57)</td>
<td>2.20(9.21)</td>
</tr>
<tr>
<td><strong>Properties</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>92814.1</td>
<td>91960.9</td>
</tr>
<tr>
<td>F1 RMSE</td>
<td>0.0125(0.0116,0.0133)</td>
<td>0.0114(0.0107,0.0121)</td>
</tr>
<tr>
<td>F2 RMSE</td>
<td>0.00688(0.00634,0.00743)</td>
<td>0.00616(0.00569,0.00664)</td>
</tr>
<tr>
<td>F3 RMSE</td>
<td>0.00422(0.00381,0.00462)</td>
<td>0.00402(0.00366,0.00439)</td>
</tr>
<tr>
<td>F4 RMSE</td>
<td>0.00326(0.00303,0.00353)</td>
<td>0.00341(0.00318,0.00365)</td>
</tr>
<tr>
<td>F5 RMSE</td>
<td>0.00356(0.00338,0.00373)</td>
<td>0.0037(0.00354,0.00385)</td>
</tr>
<tr>
<td>F6 RMSE</td>
<td>0.00408(0.00392,0.00424)</td>
<td>0.00414(0.00398,0.00431)</td>
</tr>
<tr>
<td>F7 RMSE</td>
<td>0.00447(0.00427,0.00467)</td>
<td>0.00455(0.00434,0.00475)</td>
</tr>
<tr>
<td>F8 RMSE</td>
<td>0.00459(0.00435,0.00484)</td>
<td>0.00476(0.00451,0.00502)</td>
</tr>
<tr>
<td>F9 RMSE</td>
<td>0.00454(0.00428,0.00481)</td>
<td>0.00484(0.00455,0.00513)</td>
</tr>
<tr>
<td>F10 RMSE</td>
<td>0.00454(0.00427,0.00484)</td>
<td>0.00493(0.00464,0.00523)</td>
</tr>
<tr>
<td>F11 RMSE</td>
<td>0.0049(0.00464,0.00516)</td>
<td>0.0053(0.005,0.00561)</td>
</tr>
<tr>
<td>F12 RMSE</td>
<td>0.00578(0.00549,0.00607)</td>
<td>0.00609(0.00576,0.00642)</td>
</tr>
</tbody>
</table>

This table tests the robustness of our estimation by comparing results on full sample periods (January 5, 2004 to Dec 28, 2012) using futures data of all 12 contracts and just the odd contracts, respectively. Standard errors of parameters are shown in the parenthesis following the estimation value, with significance marker attached: a * indicate significance at 0.05 level, a ** indicate significance at 0.01 level, a *** indicate significance at 0.001 level. For the properties panel, we report the converged log likelihood of the ML estimators on all 12 contracts, just the odd contracts, plus just the even contracts, respectively.
3.6 Conclusion

In the project we study the relationship between news and crude oil futures prices. In a model-free environment, we have shown that news has significant correlation with crude oil futures return and variance. This effect is persistent on all curve positions. It is also asymmetrical: negative news has higher correlation with returns and variance than does positive news, and negative news can explain more volatility clustering than does positive news.

We build a five factor stochastic model to incorporate these preliminary results, in which we explicitly model the impact of news on futures price dynamics, with the jump intensity being an exogenous factor extracted from news data. We also separate the positive and negative news processes, hence allowing news with different sentiments to have asymmetric impact on futures pricing.

Our model can fit the futures price and term structure well. And by comparing to models without news, we demonstrate that our model with news provides a better description of the price dynamics. We show that news can explain a significant amount of the conditional variance of futures prices. We also confirm the asymmetrical impact between positive and negative news by showing: negative news has higher jump mean and standard deviation than does positive news, and negative news contributes more to the conditional variance than does positive news.
Chapter 4

A RealGARCH and bivariate EVT Framework for Tail Risk Estimation

4.1 Background

Natural gas is another important and actively traded commodity, with a unique market structure. The natural gas market in the US has adapted a central pricing benchmark system, and the most important benchmark is the futures contracts deliverable at Henry Hub, actively traded on NYMEX. Prices at other market hubs are quoted as the spread relative to the pricing benchmark. These hubs are usually intersections of pipelines, or interfaces of pipelines with local distribution systems (Kaminski (2012)). The central pricing benchmarks are very liquid, with intra prices available, while trades at market hubs are much less liquid and only daily prices are available. This structure makes it difficult to study the pricing dynamics at these individual hubs, motivating us to ask the question: Can we model price dynamics at individual hubs with the benchmark?

Hansen, Huang and Shek (2012)[28] first introduce a realized GARCH framework to jointly model the returns and realized volatility, by using a realized measure of volatility to model the conditional variance of returns. They show that adding the realized volatility component can provide significant improvement over a standard GARCH model. Later in Hansen, Lunde and Voev (2014) [29], they extend the model to a multivariate framework, modeling market factors and individual assets together. In the first part of this chapter, we use a modified realized beta GARCH model where the pricing benchmark (Henry Hub NG futures) is the market factor and returns and realized volatility observations are utilized. For the individual assets we use the spot prices at individual trading hubs, where only daily returns are available. We modify the realized beta GARCH framework to suit our
After obtaining realized beta GARCH model results, we examine the behavior of the residuals and model the tail risk using a bivariate-EVT model. There has been significant research on modeling the tail risk for the past decades, most of which uses a Generalized Pareto distribution framework. McNeil and Frey (2000) [37] propose an approach to estimate the tail risk and expected shortfall, using the EVT theory on residuals of a standard GARCH model. Allen, Singh and Powell (2013) [1] apply the EVT theory to market indices and compare the Value-at-Risk measures with expected shortfall, and investigate hedging strategies in extreme conditions. Wang et al. (2010)[42] use a GARCH-EVT-Copula framework to model foreign exchange portfolios’ tail risk, calculating the Value-at-Risk and Expected Shortfall. They found that t-Copula and Clayton-Copula have better performance than a Normal-Copula. In the second part of this chapter, we propose a bivariate-EVT method for modeling the tail behavior of residuals from a realized beta GARCH model, using this framework to derive and calculate the expected shortfall.

## 4.2 Fitting a modified Realized Beta GARCH Model

Following Hansen, Lunde and Voev (2014) [29], we first build a realied beta GARCH model for the market factor (Henry Hub NG1 futures) and individual hub (spot market). We model the return
\((r_{0,t})\) of market factor as:

\[
r_{0,t} = \mu_0 + \sqrt{h_{0,t}} z_{0,t} \tag{4.1}
\]

\[
\log h_{0,t} = a_0 + b_0 \log(h_{0,t-1}) + c_0 \log(x_{0,t-1}) + \tau_0(z_{0,t-1}) \tag{4.2}
\]

\[
\log x_{0,t} = \xi_0 + \varphi_0 \log(h_{0,t}) + \delta_0(z_{0,t}) + u_{0,t} \tag{4.3}
\]

\[
\tau_0(z) = \tau_1 z + \tau_2 (z^2 - 1) \tag{4.4}
\]

\[
\delta_0(z) = \delta_{0,1} z + \delta_{0,2} (z^2 - 1) \tag{4.5}
\]

\[
z_{0,t} \sim i.i.d.N(0,1) \tag{4.6}
\]

\[
u_{0,t} \sim i.i.d.N(0,\sigma^2_{u0}) \tag{4.7}
\]

For return \((r_{i,t})\) at hub \(i\):

\[
r_{i,t} = \mu_i + \sqrt{h_{i,t}} z_{i,t} \tag{4.8}
\]

\[
z_{i,t} = \rho_{i,t} z_{0,t} + \sqrt{1 - \rho_{i,t}^2} w_{i,t} \tag{4.9}
\]

\[
w_{i,t} = \frac{z_{i,t} - \rho_{i,t} z_{0,t}}{1 - \rho_{i,t}^2} \sim N(0,1) \tag{4.10}
\]

\[
\log(h_{i,t}) = a_i + b_i \log(h_{i,t-1}) + d_i \log(h_{0,t}) + \tau_i(z_{i,t-1}) \tag{4.11}
\]

And for the correlation term \(\rho_{i,t}\), we use a Fisher transformation \(\rho \mapsto F(\rho) \equiv \frac{1}{2} \log \frac{1+\rho}{1-\rho}\), where:

\[
F(\rho_{i,t}) = a_{i0} + b_{i0} F(\rho_{i,t-1}) + c_{i0} F(y_{i,t-1}) \tag{4.12}
\]

\[
F(y_{i,t}) = \xi_{i0} + \psi_{i0} F(\rho_{i,t}) + v_{i,t} \tag{4.13}
\]
The measurement errors $u_{0,t}, v_{i,t}$ are correlated with the covariance matrix:

$$\Sigma = \begin{bmatrix} \sigma^2_{u_0} & \sigma_{u_0,v_i} \\ \sigma_{u_0,v_i} & \sigma^2_{v_i} \end{bmatrix}$$

### 4.3 Maximum Likelihood Estimation for the Modified Realized Beta GARCH Model

Because the model we are trying to estimate consists of a large number of individual hubs, we will adopt a hierarchical approach in estimation. First we estimate the parameters and filter for innovations of the market factor model. Second, we use the result of the market factor model to estimate and filter for individual hubs, as recommended by Hanse, Lunde and Voev (2014) [29] in their Appendix B.2.

#### 4.3.1 Likelihood for Model of Market Factor

The market factor model components have the following conditional density function:

$$f(r_{0,t}, x_{0,t} | \mathbf{\tilde{s}}_{t-1}) = f_{r_0}(r_{0,t} | \mathbf{\tilde{s}}_{t-1}) f_{x_0}(x_{0,t} | r_{0,t}, \mathbf{\tilde{s}}_{t-1})$$

where:

$$r_{0,t} \sim N(\mu_{0,t}, h_{0,t})$$

$$u_{0,t} \sim N(0, \sigma^2_{u_0})$$
Then for the marginal model of the market factor, after filtering for the latent states of $h_{0,t}$, we can obtain the log likelihood function following Hansen, Lunde and Voev (2014) [29]:

\begin{align}
\mathcal{L}_0 &= \mathcal{L}_{z0} + \mathcal{L}_{u0} \\
\mathcal{L}_{z0} &= \sum_{t=1}^{T} \log(h_{0,t}) + \frac{(r_{0,t} - \mu_0)^2}{h_{0,t}} \\
\mathcal{L}_{u0} &= \sum_{t=1}^{T} \log(\sigma_{u0}^2) + \frac{(\log(x_{0,t}) - \xi_0 - \varphi_0 \log h_{0,t} - \delta_0(z_{0,t}))^2}{\sigma_{u0}^2}
\end{align}

We maximize $\mathcal{L}_0$ to obtain parameter estimators for the market factor model.

### 4.3.2 Likelihood for Model of Individual Hubs

For our modified model of individual hubs, we have the following conditional density function:

\begin{equation}
\begin{aligned}
f(r_{i,t}, y_{i,t} | \mathbf{\theta}_{t-1}) &= f_{r_{i,t}}(r_{i,t} | r_{i,t}^0, x_{0,t}, \mathbf{\theta}_{t-1}) \ast f_{y_{i,t}}(y_{i,t} | r_{i,t}, r_{0,t}^0, x_{0,t}, \mathbf{\theta}_{t-1}) \\
\text{where we have:}
\end{aligned}
\end{equation}

\begin{align}
r_{i,t} &\sim N(\mu_i + \rho_{i,t}\sqrt{h_{i,t}}z_{0,t}, (1 - \rho_{i,t}^2)h_{i,t}) \\
v_{i,t} &\sim N\left(\frac{\sigma_{v_{i,u0}}}{\sigma_{u0}^2}; \frac{\sigma_{v_i}^2}{\sigma_{u0}^2} - \frac{\sigma_{v_{i,u0}}^2}{\sigma_{u0}^2}\right)
\end{align}

We can show $\sigma_{v_i}^2 - \frac{\sigma_{v_{i,u0}}^2}{\sigma_{u0}^2} > 0$ using Cauchy-Schwartz inequality.

After we filtered for the latent states of $h_{i,t}$ and $\rho_{i,t}$, we have the following log likelihood func-
We maximize $\mathcal{L}_i$ to obtain parameter estimators for the model of individual hubs.

4.4 Data

In North America there are more than one hundred hubs that have natural gas spot trading, all priced at a spread added to the Henry Hub price. Because the realized beta GARCH model estimates parameters for each hub separately, in this paper we only show the result of one pair as an example. We use the NYMEX Henry Hub futures prices as the market factor, and the Henry Hub spot prices as the individual hub factor. Results for other hubs can be derived in the same way once data become available.

We have obtained the daily settlement price of Henry Hub spot price from the Energy Information Agency, from January 8, 1998 to August 22, 2016, for a total of 4733 trading days. We also acquired daily and intra-day trading data for NYMEX Henry Hub natural gas futures (NG) from TickData, for the same sample period. In figure 4.1 panel A we compare the futures price and spot price, and in panel B we plot the spread of these two. We observe that most of the times the spread is distributed around zero, but occasionally with significant deviations.

Table 4.1 presents summary statistics for the log return of the futures ($r_{0,t}$) and spot ($r_{i,t}$) prices. We observe large variations in the moments for different years, as well as the Spearman correlation between these two returns series. Overall we see that the spot returns have higher variance.
Figure 4.1: Comparing Henry Hub Futures price with Spot price

Panel A we compared the futures price and spot price: the black solid line is the futures price and red dashed line is the spot price. Panel B we plot the spread of these two, by subtracting spot price from the futures price. The sample period is from January 8, 1998 to August 22, 2016

and larger skewness and kurtosis than the futures returns.

With the intra-day data of Henry Hub future, we can calculate realized variance, using the method describe in Zhang et al (2005) [45]. Figure 4.2 depicts the realized variance along with futures price and return. From the figure we can see that clustering of realized variance also correspond to periods with extreme returns and price slides. This realized variance time series is $x_{0,t}$ in equation 4.4.
Table 4.1: Summary Statistics for Henry Hub Futures and Spot Returns

<table>
<thead>
<tr>
<th>Year</th>
<th>Futures Returns</th>
<th></th>
<th></th>
<th>Spot Returns</th>
<th>Correlation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \hat{\mu} )</td>
<td>( \hat{\sigma}^2 )</td>
<td>( \hat{\gamma} )</td>
<td>( \hat{\kappa} )</td>
<td>( \hat{\mu} )</td>
<td>( \hat{\sigma}^2 )</td>
</tr>
<tr>
<td>1998</td>
<td>-5.80e-04</td>
<td>1.21e-03</td>
<td>0.439</td>
<td>1.656</td>
<td>-5.41e-04</td>
<td>2.59e-03</td>
</tr>
<tr>
<td>1999</td>
<td>5.05e-04</td>
<td>9.36e-04</td>
<td>0.299</td>
<td>1.110</td>
<td>4.60e-04</td>
<td>9.51e-04</td>
</tr>
<tr>
<td>2000</td>
<td>5.57e-03</td>
<td>1.16e-03</td>
<td>-0.287</td>
<td>2.166</td>
<td>5.92e-03</td>
<td>1.50e-03</td>
</tr>
<tr>
<td>2001</td>
<td>-5.56e-03</td>
<td>1.97e-03</td>
<td>0.439</td>
<td>1.656</td>
<td>-5.41e-04</td>
<td>2.59e-03</td>
</tr>
<tr>
<td>2002</td>
<td>2.39e-03</td>
<td>1.26e-03</td>
<td>0.202</td>
<td>0.243</td>
<td>2.49e-03</td>
<td>1.55e-03</td>
</tr>
<tr>
<td>2003</td>
<td>1.03e-03</td>
<td>1.87e-03</td>
<td>-0.439</td>
<td>1.656</td>
<td>9.08e-04</td>
<td>5.48e-03</td>
</tr>
<tr>
<td>2004</td>
<td>-6.82e-05</td>
<td>1.26e-03</td>
<td>0.202</td>
<td>0.243</td>
<td>-6.18e-03</td>
<td>3.33e-03</td>
</tr>
<tr>
<td>2005</td>
<td>2.31e-03</td>
<td>9.65e-04</td>
<td>0.586</td>
<td>1.110</td>
<td>4.60e-04</td>
<td>9.51e-04</td>
</tr>
<tr>
<td>2006</td>
<td>-6.82e-05</td>
<td>1.26e-03</td>
<td>0.202</td>
<td>0.243</td>
<td>-6.18e-03</td>
<td>3.33e-03</td>
</tr>
<tr>
<td>2007</td>
<td>2.31e-03</td>
<td>9.65e-04</td>
<td>0.586</td>
<td>1.110</td>
<td>4.60e-04</td>
<td>9.51e-04</td>
</tr>
<tr>
<td>2008</td>
<td>-6.82e-05</td>
<td>1.26e-03</td>
<td>0.202</td>
<td>0.243</td>
<td>-6.18e-03</td>
<td>3.33e-03</td>
</tr>
<tr>
<td>2009</td>
<td>2.31e-03</td>
<td>9.65e-04</td>
<td>0.586</td>
<td>1.110</td>
<td>4.60e-04</td>
<td>9.51e-04</td>
</tr>
<tr>
<td>2010</td>
<td>-6.82e-05</td>
<td>1.26e-03</td>
<td>0.202</td>
<td>0.243</td>
<td>-6.18e-03</td>
<td>3.33e-03</td>
</tr>
<tr>
<td>2011</td>
<td>2.31e-03</td>
<td>9.65e-04</td>
<td>0.586</td>
<td>1.110</td>
<td>4.60e-04</td>
<td>9.51e-04</td>
</tr>
<tr>
<td>2012</td>
<td>-6.82e-05</td>
<td>1.26e-03</td>
<td>0.202</td>
<td>0.243</td>
<td>-6.18e-03</td>
<td>3.33e-03</td>
</tr>
<tr>
<td>2013</td>
<td>2.31e-03</td>
<td>9.65e-04</td>
<td>0.586</td>
<td>1.110</td>
<td>4.60e-04</td>
<td>9.51e-04</td>
</tr>
<tr>
<td>2014</td>
<td>-6.82e-05</td>
<td>1.26e-03</td>
<td>0.202</td>
<td>0.243</td>
<td>-6.18e-03</td>
<td>3.33e-03</td>
</tr>
<tr>
<td>2015</td>
<td>2.31e-03</td>
<td>9.65e-04</td>
<td>0.586</td>
<td>1.110</td>
<td>4.60e-04</td>
<td>9.51e-04</td>
</tr>
<tr>
<td>2016</td>
<td>-6.82e-05</td>
<td>1.26e-03</td>
<td>0.202</td>
<td>0.243</td>
<td>-6.18e-03</td>
<td>3.33e-03</td>
</tr>
<tr>
<td>Average</td>
<td>-2.87e-05</td>
<td>1.13e-03</td>
<td>0.346</td>
<td>5.910</td>
<td>-1.39e-04</td>
<td>2.00e-03</td>
</tr>
</tbody>
</table>

This table presents summary statistics for Henry Hub futures and spot prices, we reported the mean, annualized standard deviation, skewness and excess kurtosis of the daily close-to-close returns. Assume \( R_t \) is the return on day \( t \), \( t = 1, \ldots, T \), it is calculated using \( \ln(F_t) - \ln(F_{t-1}) \), where \( F_t \) refers to the close price at time \( t \). We estimate the mean (\( \hat{\mu} \)), variance (\( \hat{\sigma}^2 \)), normalized skewness (\( \hat{\gamma} \)), and excess kurtosis (\( \hat{\kappa} \)), using the definitions recommended by Joanes and Gill (1998) for skewed distributions. In the last column we also reported the Spearman correlation between the futures and spot returns. The reporting period of this table contains 4733 trading days from January 8, 1998 to August 22, 2016.

We further calculate the correlation between futures returns and spot returns:

\[
y_{i,t} = \text{corr}(r_{i,t}, r_{0,t} | \delta_{t-1}) \tag{4.26}
\]

\[
y_{i,t} = \text{CorrSpearman}(r_{0,t-252}, \ldots, r_{0,t-1} | r_{i,t-252}, \ldots, r_{i,t-1}) \tag{4.27}
\]

using a 252 trading days rolling time window. This series will be used as a proxy for \( y_{i,t} \) in equation 4.13, which is the realized measure of correlation between market factor and individual asset. Hence for any given time \( t \), the correlation is obtained using returns from the previous trading year. Figure 4.3 plots this correlation measure. We observe quite significant movements during our sample period, especially around the 2008 financial crisis.

Note that because we are fitting the realized beta GARCH model for EVT purpose, we are more interested in the extreme negative returns (extreme losses) than in positive returns (extreme gains).
Figure 4.2: Henry Hub Futures Daily Price, Return and Realized Variance

Panel A plots the Henry Hub front futures daily price, front futures being the futures contracts with the shortest maturity, rolled when it’s around one month from delivery month. Panel B plots the daily returns and panel C shows the daily realized variance, calculated from intraday data using the Zhang et al (2005) method.

Hence we fit our model using the negative returns of Henry Hub futures and spot returns, the realized variances and correlation measure remains the same. In the next section we will present the fitting results and also compare to results using a standard GARCH model.

4.5 Realized Beta GARCH Model Fitting Result

Table 4.2 presents the estimation results for the realized Beta GARCH model for Henry Hub futures and spot prices. Most of the parameters are significant except the two mean terms and the covariance term, which fit our expectation as the full sample standard deviation for returns in Table 4.1 is much larger than the mean, for both futures and spot returns. Table 4.3 compares the distribution of residuals of Realized Beta GARCH model (RealGARCH) to those of standard GARCH model.
Figure 4.3: Correlation between Henry Hub Futures Returns and Spot Returns
This plot shows the Spearman correlation measure between the Henry Hub futures returns and spot returns, using a rolling windows of 252 trading day, from January 8, 1998 to August 22, 2016. The blue solid line near 0.3 represent the average correlation.

(sGARCH), both for the market model and the individual hubs model. We reported the parameter estimations and standard errors, standard errors are obtained from the numerical Hessian matrix. We choose a standard GARCH model as benchmark because it is the default choice for modeling the price dynamics of individual hubs, in the absence of realized variance measures. We estimate the sGARCH model and calculated the following summary statistics on the residuals: mean, standard deviation, skewness and excess kurtosis. We also reported the same set of summary statistics for the market factor standardized residuals ($z_{0,t}$ series in equation 4.7) of Realized Beta GARCH model, and the standardized residuals for the individual hub model ($w_{i,t}$ series of equation 4.10). Note these two standardized residuals are by definition independent. Below are a few observations:

- Residual mean values are consistently smaller for RealGARCH model than sGARCH model,
This table presents fitting results of models for the market factor and Henry HubSpot factor (denoted as hub \(i\)) on the full sample period, including 4733 trading days from January 8, 1998 to August 22, 2016.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>S.E.</th>
<th>Parameter</th>
<th>Value</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\mu )</td>
<td>-9.91e-05</td>
<td>2.87e-04</td>
<td>(\mu_i)</td>
<td>1.16e-04</td>
<td>3.2e-04</td>
</tr>
<tr>
<td>(a_0)</td>
<td>0.090</td>
<td>0.031</td>
<td>(a_i)</td>
<td>0.382</td>
<td>0.089</td>
</tr>
<tr>
<td>(b_0)</td>
<td>0.820</td>
<td>0.010</td>
<td>(b_i)</td>
<td>0.790</td>
<td>0.077</td>
</tr>
<tr>
<td>(c_0)</td>
<td>0.175</td>
<td>0.009</td>
<td>(d_i)</td>
<td>0.255</td>
<td>0.074</td>
</tr>
<tr>
<td>(\tau_1)</td>
<td>-0.026</td>
<td>0.003</td>
<td>(\tau_1)</td>
<td>-0.021</td>
<td>0.011</td>
</tr>
<tr>
<td>(\tau_2)</td>
<td>0.023</td>
<td>0.002</td>
<td>(\tau_2)</td>
<td>0.094</td>
<td>0.005</td>
</tr>
<tr>
<td>(\xi_0)</td>
<td>-1.182</td>
<td>0.136</td>
<td>(a_{i0})</td>
<td>0.034</td>
<td>0.005</td>
</tr>
<tr>
<td>(\phi_0)</td>
<td>0.935</td>
<td>0.019</td>
<td>(b_{i0})</td>
<td>0.133</td>
<td>0.009</td>
</tr>
<tr>
<td>(\sigma_{0.1})</td>
<td>-0.007</td>
<td>0.006</td>
<td>(c_{i0})</td>
<td>0.780</td>
<td>0.005</td>
</tr>
<tr>
<td>(\sigma_{0.2})</td>
<td>0.074</td>
<td>0.003</td>
<td>(\xi_{i0})</td>
<td>-0.043</td>
<td>0.009</td>
</tr>
<tr>
<td>(\sigma_{a0})</td>
<td>0.540</td>
<td>0.004</td>
<td>(\phi_{i0})</td>
<td>1.110</td>
<td>4.78e-05</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(\sigma_{vi})</td>
<td>0.006</td>
<td>3.54e-05</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(\sigma_{u_0,v_i})</td>
<td>2.84e-09</td>
<td>1.88e-04</td>
</tr>
</tbody>
</table>

- For the market model the residual standard deviation for RealGARCH model is closer to 1 compared to that from sGARCH model.

- The magnitude of skewness and excess kurtosis are lower for RealGARCH model than sGARCH model.

This shows that residuals of our modified Realized Beta GARCH model behave better than do those from standard GARCH models. We also notice that all the residuals are leptokurtic, with excess kurtosis greater than zero, and this has motivated us to develop a better way for tail risk estimation. Next we take a closer look at distributions of the residuals that are far from zero. Figure 4.4 compares the standardized residuals from the realized beta GARCH model \((w_i,t)\) as in equation 4.10) with that from a standard GARCH model, plus with a normal distribution benchmark, for the individual hub model. In this figure we calculate the percentage of residuals that is
Table 4.3: **Comparing Residuals of Realized Beta GARCH Model with Standard GARCH Model**

<table>
<thead>
<tr>
<th>Residuals From</th>
<th>Market Model</th>
<th>Individual Hub Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$sGARCH(1,1)$</td>
<td>RealGARCH</td>
</tr>
<tr>
<td>Mean</td>
<td>0.003</td>
<td>0.002</td>
</tr>
<tr>
<td>StDev</td>
<td>1.171</td>
<td>1.001</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.857</td>
<td>-0.296</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>13.003</td>
<td>2.149</td>
</tr>
</tbody>
</table>

This table compares the distribution of residuals of Realized Beta GARCH model to those of standard GARCH model, both for market model and individual hubs model. We estimated a standard GARCH model and calculated the following summary statistics on the residuals: mean, standard deviation, skewness and excess kurtosis. We also reported the same set of summary statistics for the $z_{0,t}$ series of the market model of Realized Beta GARCH model, and the $w_{i,t}$ series of the individual hub model. The sample period includes 4733 trading days from January 8, 1998 to August 22, 2016.

greater than a given value between 2 and 5. For example for the normal benchmark curve, when the x axis (distance from zero) is at 2.33, the y axis (percentage of residuals that is greater than 2.33 out of all residuals) is at 0.005, which means 0.5% of all residuals will be greater than 2.33 if they perfectly follow a normal distribution.

In figure 4.4, the black solid curve shows the distribution of residuals from the realized beta GARCH model, and the blue dashed curve is for standard GARCH model. Both curves are above the normal benchmark curve (grey dashed line), meaning they both exhibit heavier tails than normal distribution. Overall the Realized Beta GARCH curve is below the standard GARCH curve, indicating that the our model performs better in predicting large returns in Henry Hub spot prices, compared to the standard GARCH model which doesn’t incorporate a market factor. The normal distribution clearly underestimates large losses, and we need to make extra assumptions to appropriately model the tail behavior of the natural gas spot market returns. In the next section we will model the tail behavior of the standardized residuals, and propose a bivariate EVT approach to calculate the expected shortfalls.
4.6 Bivariate Extreme Value Theory and Expected Shortfall

4.6.1 Derive the Distribution of Excess Residuals

After fitting a modified Beta GARCH model, we obtain the standardized and independent residuals $z_{0,t}$ and $w_{i,t}$. The next step is to check the distribution of these residuals in order to determine the threshold for the Peak-over-Threshold (PoT) method. We focus on one tail (the tail with extreme negative returns) and choose threshold $\pi_i$ for hub $i$. Then we assume all residuals $w_{i,t}$ above $\pi_i$ are i.i.d. and follow a Generalized Pareto Distribution. Denote $y_{i,t} = w_{i,t} - \pi_i$, then the C.D.F. of $y_{i,t}$
is given by:

\[ F_{\pi_i}(y_{i,t}) = P(X - \pi_i \leq y_{i,t} | X > \pi_i) \]

\[ = \frac{F(y_{i,t} + \pi_i) - F(\pi_i)}{1 - F(\pi_i)} \]  

\[ \sim G_{\xi,\sigma}(y_{i,t}) \]  

(4.28)  

(4.29)  

(4.30)

with:

\[ G_{\xi,\beta}(y_{i,t}) = \begin{cases} 
1 - (1 + \frac{\xi w_i}{\beta})^{-1/\xi}, & \text{if } \xi_i \neq 0 \\
1 - \exp(-\frac{w_i}{\beta}), & \text{if } \xi_i = 0 
\end{cases} \]  

(4.31)

where \( \beta > 0 \). Also we should have:

\[ 1 - F(w_{i,t}) = P(X > \pi_i)P(X > w_{i,t} | X > \pi_i) \]

\[ = (1 - F(\pi_i))P(X - \pi_i > w_{i,t} - \pi_i | X > \pi_i) \]  

\[ = (1 - F(\pi_i))(1 - F_{\pi_i}(w_{i,t} - \pi_i)) \]  

\[ = (1 - F(\pi_i))(1 - F_{\pi_i}(y_{i,t})) \]  

(4.32)  

(4.33)  

(4.34)  

(4.35)

The first term \((1 - F(\pi_i))\) can be estimated by:

\[ (1 - F(\pi_i)) \approx \frac{N_{\pi_i}}{N} \]  

(4.36)

where \( N_{\pi_i} \) is the number of residuals over the threshold, \( N \) is the the total number of residuals. Then the second term \((1 - F_{\pi_i}(y_{i,t}))\) can be modeled as a GPD as we discussed. Hence we have:

\[ 1 - F(w_{i,t}) \approx \frac{N_{\pi_i}}{N}(1 + \frac{\xi_i(w_{i,t} - \pi_i)}{\beta_i})^{-1/\xi_i} \]  

(4.37)

given \( \pi_i \). In practice we follow the approach used by McNeil and Frey (2000) and fix \( N_{\pi_i} = k \) conditioning on \( k \ll N \). For our sample of 4733 trading days, we used a \( k = 100 \) and hence
the quantile \( q_i \) is \( 1 - k/N \). This makes \( \pi_i \) equal to the \((k + 1)\)th element in the ordered series \( w_{i,(1)} \geq w_{i,(2)} \geq \cdots \geq w_{i,(N)} \).

Then the cumulative density function becomes:

\[
F(w_{i,t}) \approx 1 - \frac{k}{N} (1 + \frac{\xi_i (w_{i,t} - \pi_i)}{\beta_i})^{-1/\xi_i}
\]  

(4.38)

and the quantile can be obtained by inverting the cumulative density function:

\[
z_{q_i} \approx \pi_i + \frac{\beta_i}{\xi_i} (\frac{1 - q_i}{k/N})^{-\xi_i} - 1
\]  

(4.39)

### 4.6.2 Derive Expected Shortfall for Returns at level \( \pi \)

**Derive Expected Shortfall for Market Factor Return \( r_{0,t} \)**

For market factor return we have:

\[
r_{0,t} = \mu_0 + \sqrt{h_{0,t}} z_{0,t}
\]  

(4.40)

\[
z_{0,t} \sim i.i.d. N(0, 1)
\]  

(4.41)

Let \( F_{r_0}(x) \) denote marginal distribution of \( r_{0,t} \), and \( F_{r_{0,t+1}+\cdots+r_{0,t+h}}(x) \) denote the predictive return distribution over the next \( h \) time units, given information up to \( t \). Then the unconditional quantile \( x_\alpha \) and expected shortfall \( S_\alpha \) at level \( \alpha \) are:

\[
x_\alpha = \inf \{ x \in \mathfrak{R} : F_{r_0}(x) \geq \alpha \}
\]  

(4.42)

\[
S_\alpha = E[X \mid X > x_\alpha]
\]  

(4.43)
The conditional quantile for the next $h$ time units $x^{t}_{\alpha}(h)$ and conditional expected shortfall are given by:

$$x^{t}_{\alpha}(h) = \inf \{ x \in \mathcal{R} : F_{r_{0,t+1}+\cdots+r_{0,t+h}}(x) \geq \alpha \}$$

(4.44)

$$S^{t}_{\alpha}(h) = E[\sum_{j=1}^{h} X_{t+j} | \sum_{j=1}^{h} X_{t+j} > x^{t}_{\alpha}(h), \mathcal{F}_{t}]$$

(4.45)

If we focus on the conditional expected shortfall of a one-step predictions, we have:

$$F_{r_{0,t+1}|\mathcal{F}_{t}}(x) = P\{ \mu_{0} + \sqrt{h_{0,t+1}}z_{0,t+1} \leq x | \mathcal{F}_{t} \}$$

(4.46)

$$= P( z_{0,t+1} \leq \frac{x - \mu_{0}}{\sqrt{h_{0,t+1}}} | \mathcal{F}_{t} )$$

(4.47)

$$= F_{Z}( \frac{x - \mu_{0}}{\sqrt{h_{0,t+1}}} )$$

(4.48)

Hence:

$$S^{t}_{\alpha}(1) = \mu_{0} + \sqrt{h_{0,t+1}}E[Z | Z > z_{0,\alpha}]$$

(4.49)

Assume the threshold we picked for $z_{0,t+1}$ is $\pi_{0}$, note that we have:

$$y_{0,t+1} = z_{0,t+1} - \pi_{0}$$

(4.50)

$$\sim G_{\xi_{0},\sigma_{0}}(y_{0,t+1})$$

(4.51)

by making $z_{0,\alpha} > \pi_{0}$, we have:

$$Z - z_{0,\alpha} | Z > z_{0,\alpha} = (Z - \pi_{0}) - (z_{0,\alpha} - \pi_{0}) | (Z - \pi_{0}) > (z_{0,\alpha} - \pi_{0})$$

(4.52)

$$\sim G_{\xi_{0},\sigma_{0}+\xi_{0}(z_{0,\alpha} - \pi_{0})}(Z - z_{0,\alpha})$$

(4.53)
Hence we have:

\[
E[Z|Z > z_{0,\alpha}] = z_{0,\alpha} \left( \frac{1}{1 - \xi_0} + \frac{\sigma_0 - \xi_0 \pi_0}{(1 - \xi_0) z_{0,\alpha}} \right) \tag{4.54}
\]

and

\[
S^*_\alpha(1) = \mu_0 + \sqrt{h_{0,t+1} z_{0,\alpha}} \left( \frac{1}{1 - \xi_0} + \frac{\sigma_0 - \xi_0 \pi_0}{(1 - \xi_0) z_{0,\alpha}} \right). \tag{4.55}
\]

We can derive the distribution of the other side of the tail similarly. Although it may not be of interest here, it can be used in the next section when deriving the expected shortfall of individual hubs. We can simplify the estimation by just performing the same process on the actual log returns (instead of the negative log returns), and we denote the tail distribution on the other tail is:

\[
Z' - z'_{0,\alpha} \mid Z' > z'_{0,\alpha} \sim G_{\xi'_0, \sigma'_0 + \xi'_0 (z'_{0,\alpha} - \pi'_0)}(Z' - z'_{0,\alpha}) \tag{4.56}
\]

which will be used for the case when \( \rho_{i,t} \) is negative.

**Derive Expected Shortfall for Individual Hubs Return** \( r_{i,t} \)

For individual hubs return we have:

\[
r_{i,t} = \mu_i + \sqrt{h_{i,t}} z_{i,t} \tag{4.57}
\]

\[
z_{i,t} = \rho_{i,t} z_{0,t} + \sqrt{1 - \rho_{i,t}^2} w_{i,t} \tag{4.58}
\]

Hence

\[
r_{i,t} = \mu_i + \sqrt{h_{i,t}} \rho_{i,t} z_{0,t} + \sqrt{h_{i,t}(1 - \rho_{i,t}^2)} w_{i,t} \tag{4.59}
\]
which can be written as:

\[
    r'_{i,t} = \frac{r_{i,t} - \mu_i}{\sqrt{h_i(t)(1 - \rho_{i,t}^2)}} - \frac{\rho_{i,t}}{\sqrt{(1 - \rho_{i,t}^2)}} z_{0,\alpha} - z_{i,\alpha}
\]

(4.60)

\[
    = \frac{\rho_{i,t}}{\sqrt{(1 - \rho_{i,t}^2)}} (z_{0,t} - z_{0,\alpha}) + (w_{i,t} - z_{i,\alpha})
\]

(4.61)

Denote \( \lambda_{i,t} = \frac{\rho_{i,t}}{\sqrt{(1 - \rho_{i,t}^2)}} \) and we get:

\[
    r'_{i,t} = \lambda_{i,t} (z_{0,t} - z_{0,\alpha}) + (w_{i,t} - z_{i,\alpha})
\]

(4.62)

\[
    = \lambda_{i,t} y_{0,t} + y_{i,t}
\]

(4.63)

where \( z_{0,t} \) and \( w_{i,t} \) are independent and are \( i.i.d. N(0, 1) \).

**For the case when \( \lambda_{i,t} \geq 0 \):**

The tail distributions of \( y_{0,t} \) and \( y_{i,t} \) are:

\[
    (y_{0,t} = z_{0,t} - z_{0,\alpha}) | z_{0,t} > z_{0,\alpha} \sim G_{\xi_0, \sigma_0 + \xi_0(z_{0,\alpha} - \pi_0)} (z_{0,t} - z_{0,\alpha})
\]

(4.64)

\[
    (y_{i,t} = w_{i,t} - z_{i,\alpha}) | w_{i,t} > z_{i,\alpha} \sim G_{\xi_i, \sigma_i + \xi_i(z_{i,\alpha} - \pi_i)} (w_{i,t} - z_{i,\alpha})
\]

(4.65)

Assume \( \sigma'_i = \sigma_i + \xi_i(z_{i,\alpha} - \pi_i) \), \( i \in 0, 1, \ldots, N \), for \( \xi_i \neq 0 \) we have:

\[
    f(y_{i,t} | y_{i,t} > 0) = \frac{1}{\sigma'_i} \left( 1 + \frac{\xi_i y_{i,t}}{\sigma'_i} \right)^{-1/\xi_i - 1}
\]

(4.66)

\[
    f(y'_{0,t} = \lambda_{i,t} y_{0,t} | \lambda_{i,t} y_{0,t} > 0) = \frac{1}{\lambda_{i,t} \sigma'_0} \left( 1 + \frac{\xi_0 y'_{0,t}}{\lambda_{i,t} \sigma'_0} \right)^{-1/\xi_0 - 1}
\]

(4.67)

\[
    \sim G_{\xi_0, \lambda_{i,t} \sigma'_0} (y'_{0,t})
\]

(4.68)
As \( y_{0,t} \) and \( y_{i,t} \) are independent, we need to find the joint conditional density of \( f(y_{i,t} = y_{0,t} + y_{i,t} | y_{0,t} > 0, y_{i,t} > 0) \). The characteristic function for \( \xi \neq 0 \) a GPD is:

\[
\phi_X(t) = \sum_{j=0}^{\infty} \left[ \frac{(it\sigma)^j}{\sum_{k=0}^{j} (1 - k\xi)} \right] \quad j = 0, 1, \ldots \tag{4.69}
\]

\[
= 1 + \frac{it\sigma}{1 - \xi} - \frac{t^2\sigma^2}{(1 - \xi)(1 - 2\xi)} - \ldots \tag{4.70}
\]

Hence we have:

\[
\phi_{y_{i,t}}(t) = 1 + \frac{it\sigma'}{1 - \xi_i} - \ldots \tag{4.71}
\]

\[
\phi_{y_{0,t}'}(t) = 1 + \frac{it\sigma'0\lambda_{i,t}}{1 - \xi_0} - \ldots \tag{4.72}
\]

\[
\phi_{y_{i,t}+y_{0,t}'}(t) = \phi_{y_{i,t}}(t) * \phi_{y_{0,t}'}(t) \tag{4.73}
\]

\[
= 1 + it\frac{\sigma'_i(1 - \xi_0) + \sigma'_0\lambda_{i,t}(1 - \xi_i)}{1 - (\xi_0 + \xi_i - \xi_0\xi_i)} + \ldots \tag{4.74}
\]

\[
(4.75)
\]

Hence the distribution of \( y_{i,t} = y_{0,t} + y_{i,t} \) is approximately:

\[
y_{i,t} | y_{0,t} > 0, y_{i,t} > 0 \sim G_{\xi_0+\xi_i,\xi_0\xi_i,\sigma'_i(1-\xi_0)+\sigma'_0\lambda_{i,t}(1-\xi_i)}(y_{i,t}) \tag{4.76}
\]

Therefore the one day ahead expected shortfall can be approximated by:

\[
S_{\alpha}^t(1) = \mu_i + \sqrt{h_{i,t}(1 - \rho^2_{i,t})}z_{Y,\alpha} \left( \frac{1}{1 - \xi_Y} + \frac{\sigma_Y - \xi_Y\pi_Y}{(1 - \xi_Y)z_{Y,\alpha}} \right) \tag{4.77}
\]
where:

\[
\begin{align*}
\xi_Y &= \xi_0 + \xi_i - \xi_0 \xi_i \\
\sigma_Y &= \sigma_i'(1 - \xi_0) + \sigma'_0 \lambda_{i,t}(1 - \xi_i) \\
z_{Y,\alpha} &= \lambda_{i,t} z_{0,\alpha} + z_{i,\alpha} \\
\pi_Y &= \lambda_{i,t} \pi_0 + \pi_i \\
z_{i,\alpha} &> \pi_i, \quad i \in (0, 1, \ldots, N)
\end{align*}
\] (4.78)

For the case when \( \lambda_{i,t} < 0 \):

The case when \( \lambda_{i,t} < 0 \) (in other words \( \rho_{i,t} \) is negative) can be derived similarly. Equation 4.62 can be written as:

\[
\begin{align*}
\dot{r}_{i,t} &= \lambda_{i,t}(z_{0,t} - z_{0,\alpha}) + (w_{i,t} - z_{i,\alpha}) \\
&= -\lambda_{i,t}(-z_{0,\alpha} - (-z_{0,t})) + (w_{i,t} - z_{i,\alpha}) \\
&= \tilde{\lambda}_{i,t}(\tilde{z}_{0,t} - \tilde{z}_{0,\alpha}) + (w_{i,t} - z_{i,\alpha}) \\
&= \tilde{\lambda}_{i,t}\tilde{y}_{0,t} + y_{i,t}
\end{align*}
\] (4.83)

The difference is that for the market model we estimated the generalized Pareto distribution on the extreme gains instead of losses. Hence the market factor is estimated on the positive return, not the negative return. Assume the extreme gains has the following distribution:

\[
(\tilde{y}_{0,t} = \tilde{z}_{0,t} - \tilde{z}_{0,\alpha}) | \tilde{z}_{0,\alpha} > \tilde{z}_{0,\alpha} \sim G_{\tilde{\xi}_0, \tilde{\sigma}_0 + \tilde{\xi}_0(\tilde{z}_{0,\alpha} - \tilde{y}_0)}(\tilde{z}_{0,t} - \tilde{z}_{0,\alpha}).
\] (4.87)

Then the distribution of \( \dot{y}_{i,t} = \dot{y}_{0,t} + y_{i,t} \) becomes:

\[
\dot{y}_{i,t} | \dot{y}_{0,t} > 0, y_{i,t} > 0 \sim G_{\dot{\xi}_0 + \dot{\xi}_i - \dot{\xi}_0 \xi_i, \sigma_i'(1 - \dot{\xi}_0) + \sigma'_0 \lambda_{i,t}(1 - \xi_i)}(\dot{y}_{i,t})
\] (4.88)
It then follows that the one day ahead expected shortfall is

\[ S^e(t) = \mu_i + \sqrt{h_{i,t}(1 - \rho_{i,t}^2)} \tilde{z}_{Y,\alpha} \left( \frac{1}{1 - \xi_Y} + \frac{\tilde{\sigma}_Y - \tilde{\xi}_Y \tilde{\pi}_Y}{(1 - \xi_Y) \tilde{z}_{Y,\alpha}} \right) \]  \hspace{1cm} (4.89)

where:

\[ \tilde{\xi}_Y = \tilde{\xi}_0 + \xi_i - \tilde{\xi}_0 \xi_i \]  \hspace{1cm} (4.90)

\[ \tilde{\sigma}_Y = \sigma'_i(1 - \tilde{\xi}_0) + \sigma'_0 \lambda_i, t(1 - \xi_i) \]  \hspace{1cm} (4.91)

\[ \tilde{z}_{Y,\alpha} = \tilde{\lambda}_{i,t} \tilde{z}_{\alpha,0} + z_{i,\alpha} \]  \hspace{1cm} (4.92)

\[ \tilde{\pi}_Y = \tilde{\lambda}_{i,t} \tilde{\pi}_0 + \pi_i \]  \hspace{1cm} (4.93)

\[ z_{i,\alpha} > \pi_i, \quad i \in (0, 1, \ldots, N) \]  \hspace{1cm} (4.94)

4.7 EVT Result

Following McNeil and Frey (2000) [37], we pick \( k = 100 \) and fit the model on the most recent 1000 trading days. We choose this instead of fitting on the full sample period as we want to perform back-testing with a 1000-day rolling window in the next section. This setup means the ordered residual \( w^{(1)} \geq w^{(2)} \geq \ldots w^{(N)}, w^{(1)}, w^{(2)}, \ldots, w^{(100)} \) will be modeled using the Generalized Pareto Distribution, and parameters are fitted to \( w^{(1)} - w^{(101)}, w^{(2)} - w^{(101)}, \ldots, w^{(100)} - w^{(101)} \). In Table 4.4 we report the estimation results for extreme gains and losses, for both the market model and individual hub model, during the sample period from October 5, 2012 to August 22, 2016 (1000 trading days). Panel A reports the Realized Beta GARCH model parameters and standard errors, and Panel B shows threshold value \( z_{(k+1)} \) and the Generalized Pareto Distribution parameter estimates and diagnosis, using maximum likelihood estimation. Here we take \( k = 100 \), and standard errors are calculated from the numerical Hessian matrix.
Table 4.4: RealGARCH and GPD Parameters from Oct 5, 2012 to Aug 22, 2016

Panel A: Realized Beta GARCH Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Market Model</th>
<th>Individual Hub Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_0$</td>
<td>-2.74e-04</td>
<td>$\mu_i$ 2.10e-04</td>
</tr>
<tr>
<td>$a_0$</td>
<td>0.582</td>
<td>$a_i$ 0.652</td>
</tr>
<tr>
<td>$b_0$</td>
<td>0.842</td>
<td>$b_i$ 0.726</td>
</tr>
<tr>
<td>$c_0$</td>
<td>0.210</td>
<td>$d_i$ 0.349</td>
</tr>
<tr>
<td>$\tau_1$</td>
<td>-0.045</td>
<td>$\tau_1$ 0.009</td>
</tr>
<tr>
<td>$\tau_2$</td>
<td>0.031</td>
<td>$\tau_2$ 0.092</td>
</tr>
<tr>
<td>$\xi_0$</td>
<td>-3.826</td>
<td>$a_{0,0}$ -0.048</td>
</tr>
<tr>
<td>$\phi_0$</td>
<td>0.608</td>
<td>$b_{0,0}$ 0.271</td>
</tr>
<tr>
<td>$\sigma_{0,1}$</td>
<td>-0.030</td>
<td>$c_{0,0}$ 0.995</td>
</tr>
<tr>
<td>$\sigma_{0,2}$</td>
<td>0.099</td>
<td>$\xi_{0,0}$ 0.050</td>
</tr>
<tr>
<td>$\sigma_{u0,0}$</td>
<td>0.587</td>
<td>$\phi_{0,0}$ 0.727</td>
</tr>
</tbody>
</table>

Panel B: GPD Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$z_{0,t}$</th>
<th>$w_{i,t}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gains</td>
<td>Losses</td>
<td>Gains</td>
</tr>
<tr>
<td>$z(k+1)$</td>
<td>1.2735</td>
<td>1.2240</td>
</tr>
<tr>
<td>$\xi$</td>
<td>-0.1868</td>
<td>0.0153</td>
</tr>
<tr>
<td>$(s.e)_\xi$</td>
<td>0.0949</td>
<td>0.0941</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.6982</td>
<td>0.5290</td>
</tr>
<tr>
<td>$(s.e)_\beta$</td>
<td>0.0954</td>
<td>0.0726</td>
</tr>
</tbody>
</table>

In this table we report the estimation results for extreme gains and losses, for both market model and individual hub model, during the sample period from Oct 5, 2012 to Aug 22, 2016 (1000 trading days). Panel A reports the Realized Beta GARCH model parameters and standard errors, and Panel B shows threshold value $z(k+1)$ and the Generalized Pareto Distribution parameter estimates and diagnosis, using maximum likelihood estimation. Here we take $k = 100$, and standard errors are calculated from the numerical Hessian matrix.

4.7.1 Rolling Estimation

In this section we use a rolling time window of 1000 days to fit the model, and then compare the parameter estimates. We fit our model using data $r_{t-999}, r_{t-998}, \ldots, r_{t-1}, r_t$, where $r_i$ is the negative of returns on day $i$. We then present summary statistics for estimated parameters in table 4.5, for the Realized Beta GARCH model for market (Panel A) and individual hub (Panel B). We reported the average and median of parameters, plus the 6 smallest estimates, the 1%, 5%, 95%
and 99% quantiles, plus the 6 largest estimates.
### Table 4.5: Parameter Estimations for Realized GARCH Model with Rolling Time Window

#### Panel A: Parameter Estimations for Market Model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Median</th>
<th>Min</th>
<th>Min(1)</th>
<th>Min(2)</th>
<th>Min(3)</th>
<th>Min(4)</th>
<th>Min(5)</th>
<th>1%</th>
<th>5%</th>
<th>95%</th>
<th>99%</th>
<th>Max(-5)</th>
<th>Max(-4)</th>
<th>Max(-3)</th>
<th>Max(-2)</th>
<th>Max(-1)</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_0$</td>
<td>-1.417e-04</td>
<td>-2.131e-04</td>
<td>-0.011</td>
<td>-0.010</td>
<td>-0.009</td>
<td>-0.009</td>
<td>-0.008</td>
<td>-0.008</td>
<td>-0.005</td>
<td>-0.003</td>
<td>-0.002</td>
<td>0.002</td>
<td>0.005</td>
<td>0.008</td>
<td>0.008</td>
<td>0.009</td>
<td>0.011</td>
<td></td>
</tr>
<tr>
<td>$a_0$</td>
<td>0.278</td>
<td>0.262</td>
<td>0.957</td>
<td>3.487</td>
<td>5.047</td>
<td>5.065</td>
<td>0.001</td>
<td>0.007</td>
<td>0.012</td>
<td>0.010</td>
<td>0.011</td>
<td>0.012</td>
<td>0.012</td>
<td>0.008</td>
<td>0.007</td>
<td>0.009</td>
<td>0.011</td>
<td></td>
</tr>
<tr>
<td>$b_0$</td>
<td>0.792</td>
<td>0.782</td>
<td>0.434</td>
<td>0.493</td>
<td>0.523</td>
<td>0.524</td>
<td>0.055</td>
<td>0.153</td>
<td>0.641</td>
<td>0.916</td>
<td>0.903</td>
<td>0.905</td>
<td>0.907</td>
<td>0.917</td>
<td>0.920</td>
<td>0.994</td>
<td>0.999</td>
<td></td>
</tr>
<tr>
<td>$c_0$</td>
<td>0.225</td>
<td>0.229</td>
<td>0.082</td>
<td>0.084</td>
<td>0.088</td>
<td>0.088</td>
<td>0.121</td>
<td>0.159</td>
<td>0.106</td>
<td>0.462</td>
<td>0.891</td>
<td>0.903</td>
<td>0.916</td>
<td>0.917</td>
<td>0.920</td>
<td>0.994</td>
<td>0.999</td>
<td></td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>-0.029</td>
<td>-0.025</td>
<td>-0.112</td>
<td>-0.110</td>
<td>-0.102</td>
<td>-0.100</td>
<td>-0.078</td>
<td>-0.072</td>
<td>-0.086</td>
<td>-0.032</td>
<td>-0.022</td>
<td>-0.044</td>
<td>-0.009</td>
<td>-0.019</td>
<td>-0.026</td>
<td>-0.027</td>
<td>-0.037</td>
<td></td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0.038</td>
<td>0.032</td>
<td>0.004</td>
<td>0.009</td>
<td>0.015</td>
<td>0.013</td>
<td>0.022</td>
<td>0.014</td>
<td>0.020</td>
<td>0.011</td>
<td>0.023</td>
<td>0.170</td>
<td>0.040</td>
<td>0.019</td>
<td>0.034</td>
<td>0.041</td>
<td>0.120</td>
<td></td>
</tr>
<tr>
<td>$\delta_1$</td>
<td>0.770</td>
<td>0.776</td>
<td>0.571</td>
<td>0.578</td>
<td>0.580</td>
<td>0.594</td>
<td>0.678</td>
<td>0.687</td>
<td>0.636</td>
<td>0.770</td>
<td>0.776</td>
<td>0.844</td>
<td>0.872</td>
<td>0.900</td>
<td>0.905</td>
<td>0.920</td>
<td>0.922</td>
<td></td>
</tr>
<tr>
<td>$\delta_2$</td>
<td>0.004</td>
<td>0.004</td>
<td>-0.076</td>
<td>-0.058</td>
<td>-0.057</td>
<td>-0.056</td>
<td>-0.035</td>
<td>-0.035</td>
<td>-0.053</td>
<td>-0.040</td>
<td>-0.040</td>
<td>0.035</td>
<td>0.051</td>
<td>0.073</td>
<td>0.081</td>
<td>0.096</td>
<td>0.127</td>
<td></td>
</tr>
<tr>
<td>$\sigma_{12}$</td>
<td>0.540</td>
<td>0.528</td>
<td>0.029</td>
<td>0.033</td>
<td>0.045</td>
<td>0.046</td>
<td>0.060</td>
<td>0.113</td>
<td>0.054</td>
<td>0.056</td>
<td>0.050</td>
<td>0.151</td>
<td>0.131</td>
<td>0.148</td>
<td>0.149</td>
<td>0.155</td>
<td>0.158</td>
<td></td>
</tr>
</tbody>
</table>

This table presents summary statistics for estimate parameters using a 1000 day rolling window, for the Reaized Beta GARCH model for market (Panel A) and individual hub (Panel B). We reported the average and median of parameters, plus the 6 smallest estimates, the 1%, 5%, 95% and 99% quantiles, plus the 6 largest estimates.
Table 4.6: GPD Parameter Estimations for Realized GARCH Model Residuals with Rolling Time Window

<table>
<thead>
<tr>
<th></th>
<th>$\xi_0$</th>
<th>$\beta_0$</th>
<th>$\mu_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: GPD Parameters for $z_{0,t}$</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>-0.011</td>
<td>0.509</td>
<td>1.210</td>
</tr>
<tr>
<td>Median</td>
<td>-0.018</td>
<td>0.504</td>
<td>1.217</td>
</tr>
<tr>
<td>Min</td>
<td>-0.162</td>
<td>0.367</td>
<td>0.830</td>
</tr>
<tr>
<td>Min(1)</td>
<td>-0.161</td>
<td>0.373</td>
<td>0.835</td>
</tr>
<tr>
<td>Min(2)</td>
<td>-0.160</td>
<td>0.374</td>
<td>0.841</td>
</tr>
<tr>
<td>Min(3)</td>
<td>-0.158</td>
<td>0.377</td>
<td>0.848</td>
</tr>
<tr>
<td>Min(4)</td>
<td>-0.157</td>
<td>0.379</td>
<td>0.855</td>
</tr>
<tr>
<td>Min(5)</td>
<td>-0.156</td>
<td>0.381</td>
<td>0.855</td>
</tr>
<tr>
<td>1%</td>
<td>-0.138</td>
<td>0.406</td>
<td>0.978</td>
</tr>
<tr>
<td>5%</td>
<td>-0.113</td>
<td>0.427</td>
<td>1.166</td>
</tr>
<tr>
<td>95%</td>
<td>0.105</td>
<td>0.600</td>
<td>1.250</td>
</tr>
<tr>
<td>99%</td>
<td>0.184</td>
<td>0.626</td>
<td>1.300</td>
</tr>
<tr>
<td>Max(-5)</td>
<td>0.237</td>
<td>0.643</td>
<td>1.350</td>
</tr>
<tr>
<td>Max(-4)</td>
<td>0.246</td>
<td>0.644</td>
<td>1.354</td>
</tr>
<tr>
<td>Max(-3)</td>
<td>0.249</td>
<td>0.644</td>
<td>1.362</td>
</tr>
<tr>
<td>Max(-2)</td>
<td>0.249</td>
<td>0.647</td>
<td>1.387</td>
</tr>
<tr>
<td>Max(-1)</td>
<td>0.250</td>
<td>0.647</td>
<td>1.393</td>
</tr>
<tr>
<td>Max</td>
<td>0.274</td>
<td>0.654</td>
<td>1.402</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$\xi_0$</th>
<th>$\beta_0$</th>
<th>$\mu_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel B: GPD Parameters for $w_{i,t}$</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.032</td>
<td>0.560</td>
<td>1.148</td>
</tr>
<tr>
<td>Median</td>
<td>0.082</td>
<td>0.553</td>
<td>1.155</td>
</tr>
<tr>
<td>Min</td>
<td>-0.345</td>
<td>0.373</td>
<td>0.684</td>
</tr>
<tr>
<td>Min(1)</td>
<td>-0.342</td>
<td>0.382</td>
<td>0.829</td>
</tr>
<tr>
<td>Min(2)</td>
<td>-0.307</td>
<td>0.393</td>
<td>0.854</td>
</tr>
<tr>
<td>Min(3)</td>
<td>-0.300</td>
<td>0.395</td>
<td>0.856</td>
</tr>
<tr>
<td>Min(4)</td>
<td>-0.285</td>
<td>0.397</td>
<td>0.874</td>
</tr>
<tr>
<td>Min(5)</td>
<td>-0.282</td>
<td>0.399</td>
<td>0.897</td>
</tr>
<tr>
<td>1%</td>
<td>-0.229</td>
<td>0.415</td>
<td>0.995</td>
</tr>
<tr>
<td>5%</td>
<td>-0.151</td>
<td>0.439</td>
<td>1.034</td>
</tr>
<tr>
<td>95%</td>
<td>0.189</td>
<td>0.680</td>
<td>1.224</td>
</tr>
<tr>
<td>99%</td>
<td>0.252</td>
<td>0.751</td>
<td>1.239</td>
</tr>
<tr>
<td>Max(-5)</td>
<td>0.285</td>
<td>0.700</td>
<td>1.255</td>
</tr>
<tr>
<td>Max(-4)</td>
<td>0.287</td>
<td>0.806</td>
<td>1.256</td>
</tr>
<tr>
<td>Max(-3)</td>
<td>0.301</td>
<td>0.808</td>
<td>1.293</td>
</tr>
<tr>
<td>Max(-2)</td>
<td>0.303</td>
<td>0.814</td>
<td>1.294</td>
</tr>
<tr>
<td>Max(-1)</td>
<td>0.338</td>
<td>0.822</td>
<td>1.348</td>
</tr>
<tr>
<td>Max</td>
<td>0.340</td>
<td>0.829</td>
<td>1.355</td>
</tr>
</tbody>
</table>

This table presents summary statistics for estimate parameters using a 1000 day rolling window, for the residuals Realized Beta GARCH model for market ($z_{0,t}$, Panel A) and individual hub ($w_{i,t}$, Panel B). We reported the average and median of parameters, plus the 6 smallest estimates, the 1%, 5%, 95% and 99% quantiles, plus the 6 largest estimates.

4.7.2 Expected Shortfall

For the past two decades there have been studies criticizing the VaR measure as a risk management tool, mainly based on two reasons: VaR numbers are not necessarily sub-additive, making it very difficult to apply on portfolio level analysis; VaR only reports a loss given a confidence parameter,
it doesn’t tell people the potential size of the loss once the situation exceeds normal conditions. In this paper, we adapted an expected shortfall approach, and derived and calculated it for our bivariate EVT model. Figure 4.5 plots the expected shortfall calculated using equation 4.77, for \( \alpha = 0.95 \) (red), 0.99 (blue) and 0.995 (green), respectively. We also plotted the negative returns of Henry Hub spot returns in the plot (black vertical bars). We see the expected shortfall generally tracks the extreme returns.

Figure 4.5: Expected Shortfall for \( \alpha = 0.95, 0.99 \) and 0.995

In this plot we present the expected shortfall calculated using equation 4.77, for \( \alpha = 0.95 \) (red dashed curve), 0.99 (blue dashed curve) and 0.995 (green dashed curve), respectively. We also plotted the negative returns of Henry Hub spot returns in the plot (black vertical bars)

4.8 Conclusions

We developed a bivariate-EVT framework for the natural gas market, by adapting a modified realized beta GARCH model framework, using the Henry Hub futures as the market factor to model the price dynamics of individual spot trading hubs. We find that this model can better explain the
Table 4.7: Summary Statistics for Expected Shortfall on Individual Hub Model

<table>
<thead>
<tr>
<th></th>
<th>( \alpha = 95% )</th>
<th>( \alpha = 99% )</th>
<th>( \alpha = 99.5% )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.209</td>
<td>0.265</td>
<td>0.289</td>
</tr>
<tr>
<td>Median</td>
<td>0.257</td>
<td>0.319</td>
<td>0.345</td>
</tr>
<tr>
<td>Min</td>
<td>0.035</td>
<td>0.039</td>
<td>0.040</td>
</tr>
<tr>
<td>Min(1)</td>
<td>0.035</td>
<td>0.039</td>
<td>0.041</td>
</tr>
<tr>
<td>Min(2)</td>
<td>0.035</td>
<td>0.039</td>
<td>0.041</td>
</tr>
<tr>
<td>Min(3)</td>
<td>0.035</td>
<td>0.039</td>
<td>0.041</td>
</tr>
<tr>
<td>Min(4)</td>
<td>0.035</td>
<td>0.039</td>
<td>0.041</td>
</tr>
<tr>
<td>Min(5)</td>
<td>0.035</td>
<td>0.040</td>
<td>0.041</td>
</tr>
<tr>
<td>1%</td>
<td>0.036</td>
<td>0.040</td>
<td>0.042</td>
</tr>
<tr>
<td>5%</td>
<td>0.037</td>
<td>0.043</td>
<td>0.046</td>
</tr>
<tr>
<td>95%</td>
<td>0.419</td>
<td>0.521</td>
<td>0.568</td>
</tr>
<tr>
<td>99%</td>
<td>0.497</td>
<td>0.618</td>
<td>0.670</td>
</tr>
<tr>
<td>Max(-5)</td>
<td>0.695</td>
<td>0.853</td>
<td>0.938</td>
</tr>
<tr>
<td>Max(-4)</td>
<td>0.727</td>
<td>0.894</td>
<td>0.966</td>
</tr>
<tr>
<td>Max(-3)</td>
<td>0.792</td>
<td>1.005</td>
<td>1.097</td>
</tr>
<tr>
<td>Max(-2)</td>
<td>0.953</td>
<td>1.224</td>
<td>1.341</td>
</tr>
<tr>
<td>Max(-1)</td>
<td>1.164</td>
<td>1.488</td>
<td>1.627</td>
</tr>
<tr>
<td>Max</td>
<td>2.436</td>
<td>3.101</td>
<td>3.387</td>
</tr>
</tbody>
</table>

This table presents summary statistics for expected shortfall, applied on the residuals Realized Beta GARCH model for individual hub. We reported the average and median of parameters, plus the 6 smallest estimates, the 1%, 5%, 95% and 99% quantiles, plus the 6 largest estimates.

price dynamics of individual hubs over a standard GARCH model. We then develop an associated bivariate-EVT model for the tail risks and expected shortfall, and compare the results to extreme returns. We find the expected shortfall measures generally capture extreme returns.
Appendices
Appendix A

Futures Price

The spot commodity price is denoted by $S_t$. Let $S_t = \exp(X_t)$.

\[
X_{t+1} - X_t = (r_t - \delta_t)\Delta_t - l_t + \epsilon_{t+1}^{(X)}
\]  
(A.1)

where $E_t^Q[\exp(\epsilon_{t+1}^{(X)})] = \exp(l_t^Q)$. Therefore

\[
E_t^Q[\exp(X_{t+1} - X_t)] = \exp((r_t - \delta_t)\Delta_t)
\]  
(A.2)

which can be rewritten in the form

\[
E_t^Q[S_{t+1}] = S_t \exp((r_t - \delta_t)\Delta_t)
\]

This is the discrete time equivalent of the continuous time expressions (4) and (36) used by Casassus and Collin-Dufresne (2005).

We need to derive results under the pricing measure $Q$. The Radon-Nykodym derivative is

\[
\frac{\Delta Q}{\Delta P} = \frac{\exp(-\Lambda'\nu_{t+1})}{L(\Lambda; \nu_t)}
\]  
(A.3)

where $L(\Lambda; \nu_t) = E_t^P[\exp(-\Lambda'\nu_{t+1})]; \nu_t = (e_t^{(V)}, e_{t-1}^{(r)}, \hat{e}_t^{(\delta)}, U_t^+, U_t^-, U_t)$ is a $(6, 1)$ vector of error terms and $\Lambda$ a $(6, 1)$ vector of prices of risk, assumed constant.
The Laplace transform of the jump size is given by

\[ E_P^t[\exp(-u\theta^{(j)}_k)] = \exp(-u\mu_{(j)} + \frac{1}{2}u^2\sigma_{(j)}^2) \equiv \phi^{(j)}(u) \]  
(A.4)

where \( u \) is a real constant and \( j \in A \). Let \( U^{(j)}_{t+1} = \sum_{k=1}^{N^{(j)}_t} \theta^{(j)}_k \), so that the Laplace transform

\[ E_P^t[\exp(-uU^{(j)}_{t+1})] = \exp[-(1 - \phi^{(j)}(u))T\lambda^{(j)}_t] \]  
(A.5)

where \( \phi^{(j)}(u) \) is defined by expression (A.4). Under the \( Q \) measure, the Laplace transform of \( U^{(j)}_{t+1} \) is given by

\[ E_Q^t[\exp(-uU^{(j)}_{t+1})] = \frac{\exp[-(1 - \phi^{(j)}(\Lambda^{(j)}_\lambda + u))T\lambda^{(j)}_t\Delta]}{\exp[-(1 - \phi^{(j)}(\Lambda^{(j)}_\lambda))T\lambda^{(j)}_t\Delta]} \]

\[ = \exp[\phi^{(j)}(\Lambda^{(j)}_\lambda + u) - \phi^{(j)}(\Lambda^{(j)}_\lambda)T\lambda^{(j)}_t\Delta] \]  
(A.6)

where \( j \in A \). For \( e^{(V)}_{t+1} \), under the \( Q \) measure,

\[ E_Q^t[\exp(-ue^{(V)}_{t+1})] = \frac{E_Q^t[\exp(-(\Lambda_V + u)e^{(V)}_{t+1})]}{E_Q^t[\exp(-\Lambda_V e^{(V)}_{t+1})]} \]

\[ = \exp[\frac{1}{2}u^2 + u\Lambda_V]hT\Delta \]

Therefore, given \( \exp(l_{T-1}) = E_Q^{(X)_T}[\exp(e^{(X)_T}_{T-1})] \), then

\[ l_{T-1} = \left( \frac{1}{2} + \Lambda_V \right)hT-1\Delta \]

\[ + \sum_{j \in A}(\phi^{(j)}(\Lambda^{(j)}_\lambda - 1) - \phi^{(j)}(\Lambda^{(j)}_\lambda))T^{(j)}_T\lambda^{(j)}_{T-1}\Delta \]
At time $T - 1$, using (A.2), we have

$$E^Q_{T-1}[S_T] = S_{T-1} \exp[(r_{T-1} - \delta_{T-1})\Delta]$$

$$= \exp[D^{(X)}_1 X_{T-1} + D^{(r)}_1 r_{T-1} - D^{(\delta)}_1 \delta_{T-1}]$$

where*

$$B_1 = 0$$

$$D^{(X)}_1 = 1$$

$$D^{(r)}_1 = \Delta$$

$$D^{(\delta)}_1 = \Delta$$

$$D^{(j)}_1 = 0, \ j \in A$$

$$G^{(h)}_1 = 0$$

Now

$$D^{(X)}_1 X_{T-1} + D^{(r)}_1 r_{T-1} - D^{(\delta)}_1 \delta_{T-1}$$

$$= D^{(X)}_1 [X_{T-2} + (r_{T-2} - \delta_{T-2})\Delta - t^Q_{T-2} + e^{(X)}_{T-1}]$$

$$+ D^{(r)}_1 [r_{T-2} + a(m - r_{T-2})\Delta + e^{(r)}_{T-1}]$$

$$- D^{(\delta)}_1 [\delta_{T-2} + h_{T-2} \psi_{\delta} \Delta + h_{T-2} \psi_{h} \Delta + e^{(\delta)}_{T-1}]$$

$$= X_{T-2}(D^{(X)}_1 - D^{(\delta)}_1 \psi_{\delta} \Delta) + r_{T-2}(D^{(X)}_1 \Delta + D^{(r)}_1 - D^{(\delta)}_1 a \Delta)$$

$$- \delta_{T-2}(D^{(X)}_1 \Delta + D^{(\delta)}_1 \kappa \Delta) - h_{T-2}(D^{(\delta)}_1 \psi_{h} \Delta)$$

$$- D^{(X)}_1 t^Q_{T-2} + D^{(r)}_1 am \Delta - D^{(\delta)}_1 \kappa \alpha \Delta$$

$$+ D^{(X)}_1 e^{(X)}_{T-1} + D^{(r)}_1 e^{(r)}_{T-1} - D^{(\delta)}_1 e^{(\delta)}_{T-1}$$

*The terms $B_1$, $D^{(j)}_1$ and $G^{(h)}_1$ are added to be consistent with the tables presented later.
where \( E_{T-2}^Q[\exp(e_{T-1}^{(X)})] = \exp(l_{T-2}^Q) \). Now using (??)

\[
D_1^{(X)} e_{T-1}^{(X)} + D_1^{(r)} e_{T-1}^{(r)} - D_1^{(\delta)} e_{T-1}^{(\delta)} = u_1^{(V)} e_{T-1}^{(V)} + u_1^{(J)} e_{T-1}^{(J)} + D_1^{(r)} e_{T-1}^{(r)} - D_1^{(\delta)} e_{T-1}^{(\delta)}
\]

where

\[
u_1^{(V)} \equiv D_1^{(X)} - D_1^{(\delta)} \beta_\delta^{(V)} \quad \text{and} \quad u_1^{(J)} \equiv D_1^{(X)} - D_1^{(\delta)} \beta_\delta^{(J)}
\]

First consider

\[
E_{T-2}^P[\exp(D_1^{(r)} e_{T-1}^{(r)})] = \exp\left(\frac{1}{2}(D_1^{(r)})^2 \Delta \sigma_r^2\right)
\]

so that

\[
E_{T-2}^Q[\exp(D_1^{(r)} e_{T-1}^{(r)})] = \frac{E_{T-2}^P[\exp(-\Lambda_r + D_1^{(r)} \bar{e}_{T-1}^{(r)})]}{E_{T-2}^P[\exp(-\Lambda_r \bar{e}_{T-1}^{(r)})]} = \exp\left(\frac{1}{2}((D_1^{(r)})^2 - 2D_1^{(r)} \Lambda_r \Delta \sigma_r^2)\right)
\]

and second

\[
E_{T-2}^Q[\exp(-D_1^{(\delta)} \bar{e}_{T-1}^{(\delta)})] = \exp\left(\frac{1}{2}((D_1^{(\delta)})^2 + 2D_1^{(\delta)} \Lambda_\delta \Delta \sigma_\delta^2)\right)
\]

Finally we need to consider

\[
E_{T-2}^Q[\exp(u_1 e_{T-1}^{(X)})] = E_{T-2}^Q[\exp(u_1 e_{T-1}^{(V)})] E_{T-2}^Q[\exp(u_1 e_{T-1}^{(J)})]
\]

given the assumption of independence. Evaluating these terms, we have

\[
E_{T-2}^Q[\exp(u_1 e_{T-1}^{(V)})] = \exp\left(\frac{1}{2}u_1^2 - u_1 \Lambda_h \Delta_{T-2}\right)
\]
and from (A.6)

\[ E_{T-2}^Q[\exp(u_1 U_{T-1}^{(j)})] = \exp[(\phi^{(j)}(\Lambda_\lambda^{(j)}) - u_1) - \phi^{(j)}(\Lambda_\lambda^{(j)}) )\lambda_T^{(j)}]\Delta \]

\[ j \in A, \text{ implying} \]

\[ l_{T-2}^Q = (\frac{1}{2} - \Lambda_h)h_{T-2}\Delta + \sum_{j \in A} (\phi^{(j)}(\Lambda_\lambda^{(j)}) - 1) - \phi^{(j)}(\Lambda_\lambda^{(j)}) )\lambda_T^{(j)}\Delta \]

Therefore, after substituting for \( l_{T-2}^Q \), we have

\[
E_{T-2}^Q[S_T] = \exp[X_{T-2}D_2^{(X)} + r_{T-2}D_2^{(r)} - \delta_{T-2}D_2^{(\delta)}] \\
- D_1^{(X)}(\frac{1}{2} - \Lambda_h)h_{T-2}\Delta - D_1^{(X)}\sum_{j \in A} (\phi^{(j)}(\Lambda_\lambda^{(j)}) - 1) - \phi^{(j)}(\Lambda_\lambda^{(j)}) )\lambda_T^{(j)}\Delta \\
+ D_1^{(r)}am\Delta - D_1^{(\delta)}K\alpha\Delta \\
+ \frac{1}{2}(u_1^{(V)})^2 - u_1^{(V)}\Lambda_h]h_{T-2}\Delta + \sum_{j \in A} [\phi^{(j)}(\Lambda_\lambda^{(j)}) - u_1^{(j)}) - \phi^{(j)}(\Lambda_\lambda^{(j)}) )\lambda_T^{(j)}\Delta \\
- (\psi hD_1^{(\delta)}\Delta)h_{T-2} + \frac{1}{2}[(D_1^{(r)})^2 - 2D_1^{(r)}\Lambda_r]\sigma_r^2\Delta \\
+ \frac{1}{2}[(D_1^{(\delta)})^2 + 2D_1^{(\delta)}\Lambda_\delta]\sigma_\delta^2\Delta + B_1]
\]

Remembering that \( \lambda_T^{(j)} = y_0^{(j)} + y_1^{(j)} n_{T-2}^{(j)} \), then the above expression can be written

\[
E_{T-2}^Q[S_T] = \exp(B_2 + D_2^{(X)} X_{T-2} + D_2^{(r)} r_{T-2} - D_2^{(\delta)}\delta_{T-2} + G_2^{(h)} h_{T-2} + \sum_{j \in A} D_2^{(j)} u_{T-2}^{(j)})
\]
\[
B_1 + D_1^{(r)} am\Delta + \frac{1}{2}(D_1^{(r)})^2 - 2D_1^{(r)} \Lambda_{\tau} \sigma_{\tau}^2 \Delta \\
B_2 = -D_1^{(d)} \kappa \alpha \Delta + \frac{1}{2}(D_1^{(d)})^2 + 2D_1^{(d)} \Lambda_{\delta} \sigma_{\delta}^2 \Delta \\
+ \sum_{j \in A} [\delta^{(j)}(\Lambda^{(j)}_\lambda - u_1^{(j)}) - D_1^{(X)} \delta^{(j)}(\Lambda^{(j)}_\lambda - 1) + (D_1^{(X)} - 1) \delta^{(j)}(\Lambda^{(j)}_\lambda)]y_0^{(j)} \Delta \\
D_2^{(X)} = D_1^{(X)} - D_1^{(d)} \psi_{\delta} \Delta \\
D_2^{(r)} = D_1^{(X)} \Delta + D_1^{(r)} (1 - \alpha \Delta) \\
D_2^{(d)} = D_1^{(X)} \Delta + D_1^{(d)} (1 - \kappa \Delta) \\
G_2^{(h)} = (\frac{1}{2}(u_1^{(V)})^2 - u_1^{(V)} \Lambda_{h}) \Delta - \psi_{h} D_1^{(d)} \Delta - (\frac{1}{2} - \Lambda_{h}) D_1^{(X)} \Delta \\
[ \delta^{(j)}(\Lambda^{(j)}_\lambda - u_1^{(j)}) - D_1^{(X)} \delta^{(j)}(\Lambda^{(j)}_\lambda - 1) + (D_1^{(X)} - 1) \delta^{(j)}(\Lambda^{(j)}_\lambda)]y_1^{(j)} \Delta \\
\quad j \in A \\
u_1^{(V)} = D_1^{(X)} - D_1^{(d)} \beta_{\delta}^{(V)} \\
u_1^{(J)} = D_1^{(X)} - D_1^{(d)} \beta_{\delta}^{(J)}
At time $T-3$, consider

$$
D_2^{(X)} X_{T-2} + D_2^{(r)} r_{T-2} - D_2^{(\delta)} \delta_{T-2} + G_2^{(h)} h_{T-2}
$$

$$
= D_2^{(X)} [X_{T-3} + (r_{T-3} - \delta_{T-3}) \Delta - \mu_{T-3}^{(Q)} + e_{T-2}^{(X)}]
+ D_2^{(r)} [r_{T-3} + a(m - r_{T-3}) \Delta + e_{T-2}^{(r)}]
- D_2^{(\delta)} [\delta_{T-3} + \kappa(\alpha - \delta_{T-3}) \Delta + X_{T-3} \psi h \Delta + e_{T-2}^{(\delta)}]
+ G_2^{(h)} [w_h + b_h h_{T-3} + a_h (e_{T-2}^{(V)} - c_h \sqrt{h_{T-3}})^2]
$$

$$
= -D_2^{(X)} l_{T-3}^{Q} + D_2^{(r)} a m \Delta - D_2^{(\delta)} \kappa \alpha \Delta + G_2^{(h)} w_h
$$

$$(D_2^{(X)} - D_2^{(\delta)} \psi h \Delta) X_{T-3}
+ r_{T-3} [D_2^{(X)} \Delta + D_2^{(r)} a \Delta]
- \delta_{T-3} [D_2^{(X)} \Delta + D_2^{(\delta)} \kappa \Delta]
+ G_2^{(h)} [b_h h_{T-3} + a_h (e_{T-2}^{(V)} - c_h \sqrt{h_{T-3}})^2] - h_{T-3} D_2^{(\delta)} \psi h \Delta
+ D_2^{(X)} e_{T-3}^{(X)} + D_2^{(r)} e_{T-2}^{(r)} - D_2^{(\delta)} e_{T-2}^{(\delta)}
$$

where $E_{T-3}^{Q} [\exp(e_{T-2}^{(X)})] = \exp(l_{T-3}^{Q})$. Now using (??)

$$
D_2^{(X)} e_{T-2}^{(X)} + D_2^{(r)} e_{T-2}^{(r)} - D_2^{(\delta)} e_{T-2}^{(\delta)}
$$

$$
= e_{T-2}^{(V)} u_2^{(V)} + e_{T-2}^{(J)} u_2^{(J)} + D_2^{(r)} e_{T-1}^{(r)} - D_2^{(\delta)} e_{T-1}^{(\delta)}
$$

where

$$
u_2^{(V)} \equiv D_2^{(X)} - D_2^{(\delta)} \beta_2^{(V)} \quad \text{and} \quad u_2^{(J)} \equiv D_2^{(X)} - D_2^{(\delta)} \beta_2^{(J)}
$$

To evaluate the expectation under the $Q$ measure for the volatility terms, consider

$$
-\Lambda_h \sqrt{h_{T-3}} \Delta e_{T-2}^{(V)} + u_2^{(V)} \sqrt{h_{T-3}} \Delta e_{T-2}^{(V)} + G_2^{(h)} a_h (e_{T-2}^{(V)} - c_h \sqrt{h_{T-3}})^2
$$

$$
= G_2^{(h)} a_h [e_{T-2}^{(V)} - (c_h - \bar{u}_2/(2G_2^{(h)} a_h)) \sqrt{h_{T-3}}]^2 + (c_h \bar{u}_2 - \bar{u}_2^2/(4G_2^{(h)} a_h)) h_{T-3}
$$
where $\bar{u}_2 = (-\Lambda_h + u_2^{(V)})\sqrt{\Delta}$ and the term $\Lambda_h\sqrt{h_{T-3}}\Delta e_{T-2}^{(V)}$ comes from the numerator in the change of measure. Let $\gamma = (c_h - \bar{u}_2/(2G_2^{(h)}a_h))\sqrt{h_{T-3}}$, so that

$$
\frac{E_{T-3}^{(P)}\{\exp[G_2^{(h)}a_h(e_{T-2}^{(V)} - \gamma)^2]\}}{E_{T-3}^{(P)}[\exp(-\Lambda_h\sqrt{h_{T-3}}\Delta e_{T-2}^{(V)})]} = \exp[-\frac{1}{2}\ln(1 - 2G_2^{(h)}a_h) + G_2^{(h)}a_h\gamma^2/(1 - 2G_2^{(h)}a_h) - \frac{1}{2}\Lambda_h^2h_{T-3}\Delta]
$$

Note we require that $1 - 2G_2^{(h)}a_h > 0$.

Let $Z_{T-2}^{(j)} = D_2^{(j)}n_{T-2}^{(j)} + u_2^{(j)}U_{T-2}^{(j)}$, so that the Laplace transform is

$$
E_{T-3}^{P}[\exp(-uZ_{T-2}^{(j)})] = \exp[-(1 - \hat{\phi}^{(j)}(u))\lambda_{T-3}\Delta]
$$

where $\hat{\phi}_2^{(j)}(u) = \exp(-uD_2^{(j)})\phi^{(j)}(uu_2^{(j)})$.

$$
E_{T-3}^{Q}[\exp(-uZ_{T-2}^{(j)})] = \frac{E_{T-3}^{P}[\exp(-(\Lambda_\lambda^{(j)} + u)Z_{T-2}^{(j)})]}{E_{T-3}^{P}[\exp(-\Lambda_\lambda^{(j)}Z_{T-2}^{(j)})]}
= \exp[(\hat{\phi}_2^{(j)}(u + \Lambda_\lambda^{(j)}) - \hat{\phi}_2^{(j)}(\Lambda_\lambda^{(j)}))\lambda_{T-3}\Delta]
$$

Setting $u = -1$,

$$
E_{T-3}^{Q}[\exp(Z_{T-2}^{(j)})] = \exp[(\hat{\phi}_2^{(j)}(-1 + \Lambda_\lambda^{(j)}) - \hat{\phi}_2^{(j)}(\Lambda_\lambda^{(j)}))\lambda_{T-3}\Delta]
$$

We need to consider

$$
E_{T-3}^{Q}[\exp(D_2^{(r)}e_{T-2}^{(r)})] = \exp\left(\frac{1}{2}[(D_2^{(r)})^2 - 2D_2^{(r)}\Lambda_r]\sigma_r^2\Delta\right)
$$

and second

$$
E_{T-3}^{Q}[\exp(D_2^{(\delta)}e_{T-2}^{(\delta)})] = \exp\left(\frac{1}{2}[(D_2^{(\delta)})^2 + 2D_2^{(\delta)}\Lambda_\delta]\sigma_\delta^2\Delta\right)
$$
Finally we need to compute

\[ \exp(t_{T-3}^Q) = E_{T-3}^Q[\exp(e_{T-2}^{(X)})] \]

Given the assumption of independence.

\[ E_{T-3}^Q[\exp(u_2 e_{T-2}^{(X)})] = E_{T-3}^Q[\exp(u_2 e_{T-2}^{(V)})] E_{T-2}^Q[\exp(u_2 U_{T-2})] \]

First we have

\[ E_{T-3}^Q[\exp(u_2 e_{T-2}^{(V)})] = \exp\left(\frac{1}{2}(u_2)^2 - u_2 \Lambda_h h_{T-3} \Delta\right) \]

and from (A.6)

\[ E_{T-3}^Q[\exp(u_2 U_{T-2}^{(j)})] = \exp\left(\phi^{(j)}(\Lambda_{\lambda}^{(j)} - u_2) - \phi^{(j)}(\Lambda_{\lambda}^{(j)})\lambda_{T-3}^{(j)} \Delta\right) \]

Therefore

\[ t_{T-3}^Q = \left(\frac{1}{2} - \Lambda_h \right) h_{T-3} \Delta + \sum_{j \in A} \left(\phi^{(j)}(\Lambda_{\lambda}^{(j)} - 1) - \phi^{(j)}(\Lambda_{\lambda}^{(j)})\lambda_{T-3}^{(j)} \Delta\right) \]

Substituting these terms and simplifying, gives

\[ E_{T-3}^Q[S_T] = \exp(B_3 + D_3^{(X)} X_{T-3} + D_3^{(r)} r_{T-3} - D_3^{(\delta)} \delta_{T-3} + G_3^{(h)} h_{T-3} + \sum_{j \in A} D_3^{(j)} n_{T-3}^{(j)}) \]

where
\[ B_2 + D_2^{(r)} a m \Delta + \frac{1}{2}[(D_2^{(r)})^2 - 2D_2^{(r)} \Lambda \sigma^2 \Delta \]

\[ -D_2^{(d)} k \alpha \Delta + \frac{1}{2}[(D_2^{(d)})^2 + 2D_2^{(d)} \Lambda \sigma^2 \Delta \]

\[ \sum_{j\in A}[\hat{\phi}_2^{(j)}(-1 + \Lambda^{(j)}_\lambda) - \hat{\phi}_2^{(j)}(\Lambda^{(j)}_\lambda) - D_2^{(x)}[\phi^{(j)}(\Lambda^{(j)}_\lambda - 1) - \phi^{(j)}(\Lambda^{(j)}_\lambda)]]y_0^{(j)} \Delta \]

\[ + G_2^{(h)} w_h - \frac{1}{2} \ln(1 - 2G_2^{(h)} a_h) \]

\[ \mathbf{D}_{3}^{(x)} = D_2^{(x)} - D_2^{(d)} \psi_\beta \Delta \]

\[ \mathbf{D}_{3}^{(r)} = D_2^{(x)} \Delta + D_2^{(r)} (1 - a \Delta) \]

\[ \mathbf{D}_{3}^{(d)} = D_2^{(x)} \Delta + D_2^{(d)} (1 - \kappa \Delta) \]

\[ G_3^{(h)} = \mathbf{G}_2^{(h)} \mathbf{b}_\mathbf{h} + \mathbf{G}_2^{(h) a_h} (c_h - \bar{u}_2/(2 \mathbf{G}_2^{(h) a_h}))^2 - \frac{1}{2} \Lambda^2 \Delta \]

\[ + (c_h \bar{u}_2 - \bar{u}_2^2/(4 \mathbf{G}_2^{(h) a_h})) - D_2^{(x)}(1 - \Lambda_h) \Delta - D_2^{(d)} \psi_\beta \Delta \]

\[ \mathbf{D}_{3}^{(j)} = \hat{\phi}_2^{(j)}(-1 + \Lambda^{(j)}_\lambda) - \hat{\phi}_2^{(j)}(\Lambda^{(j)}_\lambda) - D_2^{(x)}[\phi^{(j)}(\Lambda^{(j)}_\lambda - 1) - \phi^{(j)}(\Lambda^{(j)}_\lambda)]]y_1^{(j)} \Delta \]

\[ \hat{\phi}_2^{(j)}(u) = \exp(-uD_2^{(j)}) \phi^{(j)}(u \bar{u}_2^{(j)}) \]

\[ u_2^{(v)} = D_2^{(x)} - D_2^{(d)} \beta_\delta^{(v)} \]

\[ u_2^{(j)} = D_2^{(x)} - D_2^{(d)} \beta_\delta^{(j)} \]

\[ \bar{u}_2 = (-\Lambda_h + u_2^{(v)}) \sqrt{\Delta} \]

After this point the analysis repeats itself. Therefore at time \( t \), interval \( n \), the futures price is given by

\[ F(t, T) = E_t^Q[S_T] = \exp(B_n + D_n^{(x)} X_t + D_n^{(r)} r_t - D_n^{(d)} \delta_t + G_n^{(h)} h_t + \sum_{j\in A} D_n^{(j)} n_t^{(j)}) \]
where

\[ B_{n-1} + D_{n-1}^{(r)} a_m \Delta + \frac{1}{2} [(D_{n-1}^{(r)})^2 - 2D_{n-1}^{(r)} \sigma^2 r] \Delta \]

\[ B_n = -D_{n-1}^{(d)} \kappa \alpha \Delta + \frac{1}{2} [(D_{n-1}^{(d)})^2 + 2D_{n-1}^{(d)} \sigma^2 d] \Delta \]

\[ \sum_{j \in \mathcal{A}} \hat{\phi}_{n-1}^{(j)}(-1 + \Lambda_{\lambda}^{(j)}) - \hat{\phi}_{n-1}^{(j)}(\Lambda_{\lambda}^{(j)}) - D_{n-1}^{(X)} [\phi^{(j)}(\Lambda_{\lambda}^{(j)} - 1) - \phi^{(j)}(\Lambda_{\lambda}^{(j)})] y_{\theta}^{(j)} \Delta \]

\[ + G_{n-1}^{(h)} w_h - \frac{1}{2} \ln(1 - 2G_{n-1}^{(h)} a_h) \]

\[ D_n^{(X)} = D_{n-1}^{(X)} \Delta + D_{n-1}^{(r)} (1 - a \Delta) \]

\[ D_n^{(d)} = D_{n-1}^{(X)} \Delta + D_{n-1}^{(d)} (1 - \kappa \Delta) \]

\[ C_{n}^{(h)} = G_{n-1}^{(h)} b_h + \frac{G_{n-1}^{(h)} a_h}{(1 - 2G_{n-1}^{(h)} a_h)} (c_h - \bar{u}_{n-1}/(2G_{n-1}^{(h)} a_h))^2 - \frac{1}{2} \Lambda_h^2 \Delta \]

\[ + (c_h \bar{u}_{n-1} - \bar{u}_{n-1}^2 / (4G_{n-1}^{(h)} a_h)) - D_{n-1}^{(X)} (\frac{1}{2} - \Lambda_h) \Delta - D_{n-1}^{(d)} \psi_h \Delta \]

\[ D_n^{(j)} = [\hat{\phi}_{n-1}^{(j)}(-1 + \Lambda_{\lambda}^{(j)}) - \hat{\phi}_{n-1}^{(j)}(\Lambda_{\lambda}^{(j)}) - D_{n-1}^{(X)} [\phi^{(j)}(\Lambda_{\lambda}^{(j)} - 1) - \phi^{(j)}(\Lambda_{\lambda}^{(j)})] y_{\theta}^{(j)} \Delta \]

\[ \hat{\phi}_{n-1}^{(j)}(u) = \frac{\exp(-uD_{n-1}^{(j)}) \phi^{(j)}(uu_{n-1}^{(j)})}{D_{n-1}^{(X)} - D_{n-1}^{(d)} \beta^{(V)}_{n-1}} \]

\[ u_{n-1}^{(V)} = D_{n-1}^{(X)} - D_{n-1}^{(d)} \beta^{(V)}_{n-1} \]

\[ u_{n-1}^{(j)} = D_{n-1}^{(X)} - D_{n-1}^{(d)} \beta^{(j)}_{n-1} \]

\[ \bar{u}_{n-1} = (-\Lambda_h + u_{n-1}^{(V)}) \sqrt{\Delta} \]

Casassus and Collin-Dufresne (2005) describe a three jump model (see Appendix F). Direct comparison of the solutions is difficult due to different assumptions.
Appendix B

Higher Moments of Return

Let $R_{t+1} = \ln(F(t+1, T)/F(t, T))$, we want to derive the variance, skewness and kurtosis of $R_{t+1}$ at time $t$, and the error sources of the return are:

$$
Error(R_{t+1}) = D_{n-1}^{(X)} e_{t+1}^{(X)} + D_{n-1}^{(r)} e_{t+1}^{(r)} - D_{n-1}^{(\delta)} e_{t+1}^{(\delta)} + G_{n-1}^{(h)} Error(h_{t+1}) + \sum_{j \in A} D_{n-1}^{(j)} Error(n_{t}^{(j)})
$$

$$
= D_{n-1}^{(X)} (e_{t+1}^{(V)} + e_{t+1}^{(J)}) + D_{n-1}^{(r)} (e_{t+1}^{(r)} - e_{t+1}^{(J)}) - D_{n-1}^{(\delta)} (e_{t+1}^{(V)} + e_{t+1}^{(J)}) + G_{n-1}^{(h)} a_{h} (\bar{e}_{t}^{2(V)} - 2c_{h} \sqrt{\bar{h}_{t} e_{t}^{(V)}}) + \sum_{j \in A} D_{n-1}^{(j)} Error(n_{t}^{(j)})
$$

$$
= (D_{n-1}^{(X)} \bar{h}_{t} \Delta e_{t+1}^{(V)} - D_{n-1}^{(\delta)} \beta_{V} \sqrt{h_{t} \Delta e_{t+1}^{(V)}} + G_{n-1}^{(h)} a_{h} ((\bar{e}_{t}^{(V)})^2 - 2c_{h} \sqrt{\bar{h}_{t} e_{t}^{(V)}}))
$$

$$
+ (D_{n-1}^{(X)} (e_{t+1}^{(J)} - e_{t+1}^{(\delta)})) + \sum_{j \in A} D_{n-1}^{(j)} Error(n_{t}^{(j)}) + D_{n-1}^{(r)} \bar{e}_{t+1}^{(r)} - D_{n-1}^{(\delta)} \bar{e}_{t+1}^{(\delta)}
$$

$$
= (u_{n-1}^{(V)} \bar{h}_{t} \Delta - 2G_{n-1}^{(h)} a_{h} c_{h} \sqrt{h_{t}} e_{t+1}^{(V)} + G_{n-1}^{(h)} a_{h} (\bar{e}_{t}^{(V)})^2)
$$

$$
+ u_{n-1}^{(J)} e_{t+1}^{(J)} + \sum_{j \in A} D_{n-1}^{(j)} Error(n_{t}^{(j)}) + D_{n-1}^{(r)} \bar{e}_{t+1}^{(r)} - D_{n-1}^{(\delta)} \bar{e}_{t+1}^{(\delta)}
$$
We can split the error sources into the following components and identify their distributions, respectively:

\[ A = (u_{n-1}^{(V)} \sqrt{h_t} - 2G_{n-1}^{(h)} a_h c_h \sqrt{h_t})^2 + G_{n-1}^{(h)} a_h (\varepsilon_t^{(V)})^2 \]
\[ = G_{n-1}^{(h)} a_h [(\varepsilon_t^{(V)})^2 - (c_h - \frac{u_{n-1}^{(V)} \sqrt{\Delta}}{2G_{n-1}^{(h)} a_h}) \sqrt{h_t})^2] \]
\[ \sim G_{n-1}^{(h)} a_h \chi^2_{\text{NonCentral}}(1, (c_h - \frac{u_{n-1}^{(V)} \sqrt{\Delta}}{2G_{n-1}^{(h)} a_h}) \sqrt{h_t})^2) \]
\[ B = u_{n-1}^{(J)} e_{t+1}^{(J)} \]
\[ = u_{n-1}^{(J)} \sum_{j=1}^{N_j^+} \psi_j^{(+)} + \sum_{j=1}^{N_j^-} \psi_j^{(-)} \]
\[ \sim u_{n-1}^{(J)} [\text{CompoundPoisson}(N_{t+1}^{(+)} \Delta, N(\mu^{(+)}, \sigma^{(+)})) + \text{CompoundPoisson}(N_{t+1}^{(-)} \Delta, N(\mu^{(-)}, \sigma^{(-)}))] \]
\[ C = \sum_{j \in A} D_{n-1}^{(j)} \text{Error}(n_t^{(j)}) \]
\[ \sim D_{n-1}^{(+)} \text{Poisson}(N_{t+1}^{(+)} \Delta) + D_{n-1}^{(-)} \text{Poisson}(N_{t+1}^{(-)} \Delta) \]
\[ D = D_{n-1}^{(r)} e_{t+1}^{(r)} - D_{n-1}^{(\delta)} e_{t+1}^{(\delta)} \]
\[ \sim N(0, (D_{n-1}^{(r)} \sigma_r)^2 \Delta + (D_{n-1}^{(\delta)} \sigma_\delta)^2 \Delta) \]

But B and C are not independent and can be combined into one Compound Poisson Process:

\[ B + C = BC \]
\[ \sim u_{n-1}^{(J)} [\text{CompoundPoisson}(N_{t+1}^{(+)} \Delta, N(\mu^{(+)} + \frac{D_{n-1}^{(+)}(J)}{u_{n-1}^{(J)}}, \sigma^{(+)})) \]
\[ + \text{CompoundPoisson}(N_{t+1}^{(-)} \Delta, N(\mu^{(-)} + \frac{D_{n-1}^{(-)}(J)}{u_{n-1}^{(J)}}, \sigma^{(-)})) \]

Next we will calculate the cumulants \( \kappa_n, n = (1, 2, 3, 4) \) for the three processes, and then obtain the higher moments from cumulants.
Cumulants

Cumulant of Non-Central Chi Square Distribution (A)

In this case we have:

\[
A = \left( G_{(n-1)a_h}^{(h)} \right) \ast A'
\]

\[
A' \sim \chi^2(1, \lambda_A)
\]

\[
\lambda_A = (c_h - \frac{u_{n-1}^{(V)} \sqrt{\Delta}}{2G_{n-1}^{(h)}a_h})^2 h_t
\]

the moment generating function is:

\[
m_X(t) = (1 - 2t)^{-\frac{1}{2}} \exp\left( \frac{\lambda_A t}{1 - 2t} \right)
\]

And the cumulant is:

\[
K_X(t) = \log(m_X(t))
\]

\[
= -\frac{1}{2} \log(1 - 2t) + \frac{\lambda_A t}{1 - 2t}
\]

Perform Taylor expansion and combine terms then we get:

\[
K_X(t) = \left( \frac{2}{2} + \frac{2}{2} \lambda_A \right) t + \left( \frac{2^2}{4} + \frac{2^2}{2} \lambda_A \right) t^2
\]

\[
+ \left( \frac{2^3}{2 \ast 3} + \frac{2^3}{2} \lambda_A \right) t^3 + \left( \frac{2^4}{2 \ast 3} + \frac{2^4}{2} \lambda_A \right) t^4 + \ldots
\]

\[
= \sum_{m=1}^{\infty} \kappa_m \frac{t^2}{m!}
\]
Hence we have:

\[
\begin{align*}
\kappa_2^{A'} &= 2(1 + 2\lambda_A) \\
\kappa_3^{A'} &= 8(1 + 3\lambda_A) \\
\kappa_4^{A'} &= 48(1 + 4\lambda_A)
\end{align*}
\]

Since \( \kappa_m(cX) = c^m \kappa_m(X) \), after adding the constant we have:

\[
\begin{align*}
\kappa_2^A &= 2(1 + 2\lambda_A)(G^{(h)}_{(n-1)} a_h)^2 \\
\kappa_3^A &= 8(1 + 3\lambda_A)(G^{(h)}_{(n-1)} a_h)^3 \\
\kappa_4^A &= 48(1 + 4\lambda_A)(G^{(h)}_{(n-1)} a_h)^4
\end{align*}
\]

**Cumulant of Compound Poisson Distribution (BC)**

In this case we have:

\[
\begin{align*}
BC &= (u_{n-1}^{(j)})^* (BC^{(+)}) + BC^{(-)} \\
BC^{(+)} &= \text{CompoundPoisson}(N_{t+1}^{(+)} \Delta, N(\mu^{(+)}, \frac{D^{(+)\ast}_{n-1}}{u_{n-1}^{(j)}}, \sigma^{(+)}) \\
BC^{(-)} &= \text{CompoundPoisson}(N_{t+1}^{(-)} \Delta, N(\mu^{(-)}, \frac{D^{(-\ast)}_{n-1}}{u_{n-1}^{(j)}}, \sigma^{(-)})
\end{align*}
\]
We will derive the cumulant for the positive process, and negative process can be derived similarly.

The cumulant for the Poisson process is:

\[ K^{(+)}_{\lambda^{(+)}_t}(t) = \lambda^{(+)}_t \Delta(e^t - 1) = \lambda^{(+)}_t \Delta \left( \sum_{m=1}^{\infty} \frac{t^m}{m!} \right) \]

And the cumulant for the Normal process is:

\[ K^{(+)}_{\mu^{(+)}}(t) = \mu^{(+)} + \frac{D^{(+)}_{(n-1)}}{u^{(j)}_{n-1}} + \frac{\sigma^{(+)}_t^2}{2} \]

Hence the cumulant for the Compound Poisson process is:

\[ K^{(+)}_{\lambda^{(+)}_t}(K^{(+)}_{\mu^{(+)}}(t)) = \lambda^{(+)}_t \Delta \left( \sum_{m=1}^{\infty} \frac{(\mu^{(+)} + \frac{D^{(+)}_{(n-1)}}{u^{(j)}_{n-1}} + \frac{\sigma^{(+)}_t^2}{2})^m}{m!} \right) \]

\[ = \lambda^{(+)}_t \Delta \left( \mu^{(+)} + \frac{D^{(+)}_{(n-1)}}{u^{(j)}_{n-1}} + \frac{\sigma^{(+)}_t^2}{2} + (\mu^{(+)} + \frac{D^{(+)}_{(n-1)}}{u^{(j)}_{n-1}} + \frac{\sigma^{(+)}_t^2}{2})^2 \right) \]

\[ + \lambda^{(+)}_t \Delta \left( \frac{1}{2} ((\mu^{(+)} + \frac{D^{(+)}_{(n-1)}}{u^{(j)}_{n-1}})(\sigma^{(+)}_t)^2 + \frac{1}{6} (\mu^{(+)} + \frac{D^{(+)}_{(n-1)}}{u^{(j)}_{n-1}})^3) + \frac{1}{8} (\sigma^{(+)}_t)^4 + \frac{1}{4} (\mu^{(+)} + \frac{D^{(+)}_{(n-1)}}{u^{(j)}_{n-1}})^2 (\sigma^{(+)}_t)^2 + \frac{1}{4} (\mu^{(+)} + \frac{D^{(+)}_{(n-1)}}{u^{(j)}_{n-1}})^4 t^4 \right) \]

\[ + \ldots \]
Hence we have:

\[
\begin{align*}
\kappa_2^{BC(+)} &= \lambda_t^{(+)} \Delta [(\mu^{(+)}) + \frac{D_{n-1}^{(+)}}{u_{n-1}^{(+)}})^2 + (\sigma^{(+)})^2] \\
\kappa_3^{BC(+)} &= \lambda_t^{(+)} \Delta [3((\mu^{(+)}) + \frac{D_{n-1}^{(+)}}{u_{n-1}^{(+)}}))(\sigma^{(+)})^2 + (\mu^{(+)}) + \frac{D_{n-1}^{(+)}}{u_{n-1}^{(+)}})] \\
\kappa_4^{BC(+)} &= \lambda_t^{(+)} \Delta [3(\sigma^{(+)})^4 + 6(\mu^{(+)}) + \frac{D_{n-1}^{(+)}}{u_{n-1}^{(+)}}(\sigma^{(+)})^2 + 6(\mu^{(+)}) + \frac{D_{n-1}^{(+)}}{u_{n-1}^{(+)}})] \nonumber
\end{align*}
\]

And similarly:

\[
\begin{align*}
\kappa_2^{BC(-)} &= \lambda_t^{(+)} \Delta [(\mu^{(-)}) + \frac{D_{n-1}^{(-)}}{u_{n-1}^{(-)}})^2 + (\sigma^{(-)})^2] \\
\kappa_3^{BC(-)} &= \lambda_t^{(+)} \Delta [3((\mu^{(-)}) + \frac{D_{n-1}^{(-)}}{u_{n-1}^{(-)}}))(\sigma^{(-)})^2 + (\mu^{(-)}) + \frac{D_{n-1}^{(-)}}{u_{n-1}^{(-)}})] \\
\kappa_4^{BC(-)} &= \lambda_t^{(+)} \Delta [3(\sigma^{(-)})^4 + 6(\mu^{(-)}) + \frac{D_{n-1}^{(-)}}{u_{n-1}^{(-)}}(\sigma^{(-)})^2 + 6(\mu^{(-)}) + \frac{D_{n-1}^{(-)}}{u_{n-1}^{(-)}})] \nonumber
\end{align*}
\]

Hence the cumulants for the Compound Poisson process is:

\[
\begin{align*}
\kappa_2^B &= (u_{n-1}^{(+)})^2 (\kappa_2^{BC(+)} + \kappa_2^{BC(-)}) \\
\kappa_3^B &= (u_{n-1}^{(+)})^3 (\kappa_3^{BC(+)} + \kappa_3^{BC(-)}) \\
\kappa_4^B &= (u_{n-1}^{(+)})^4 (\kappa_4^{BC(+)} + \kappa_4^{BC(-)}) \nonumber
\end{align*}
\]

**Cumulant of Normal Distribution (D)**

As \( D \sim N(0, (D_{n-1}^{(r)} \sigma_r)^2 \Delta + (D_{n-1}^{(\delta)} \sigma_\delta)^2 \Delta) \) we should have:

\[
\begin{align*}
\kappa_2^D &= (D_{n-1}^{(r)} \sigma_r)^2 \Delta + (D_{n-1}^{(\delta)} \sigma_\delta)^2 \Delta \\
\kappa_3^D &= 0 \\
\kappa_4^D &= 0 \nonumber
\end{align*}
\]
Derive Higher Moments

Variance

Variance is also the second cumulant $\kappa_2$, where $\kappa_2 = \kappa_2^A + \kappa_2^{BC} + \kappa_2^D$, hence we have:

\[
\text{var}_t(R_{t+1}) = \kappa_2 = 2(1 + 2(c_h - \frac{u_{n-1}^{(V)} \sqrt{\Delta}}{2G_{n-1}^{(h)} a_h})) (G_{(n-1)}^{(h)} a_h)^2 \\
+ (u_{n-1}^{(J)})^2 \lambda_t^{(+)} \Delta [(\mu^{(+)} + \frac{D_{n-1}^{(+)}}{u_{n-1}^{(J)}})^2 + (\sigma^{(+)})^2] \\
+ (u_{n-1}^{(J)})^2 \lambda_t^{(-)} \Delta [(\mu^{(-)} + \frac{D_{n-1}^{(-)}}{u_{n-1}^{(J)}})^2 + (\sigma^{(-)})^2] \\
+ (D_{n-1}^{(r)} \sigma_r)^2 \Delta + (D_{n-1}^{(d)} \sigma_d)^2 \Delta
\]

Skewness

Skewness is defined by:

\[
\text{Skew}_t(R_{t+1}) = \frac{\kappa_3}{(\kappa_2)^2}
\]

where $\kappa_2$ is already given above, $\kappa_3$ is:

\[
\kappa_3 = 8(1 + 3(c_h - \frac{u_{n-1}^{(V)} \sqrt{\Delta}}{2G_{n-1}^{(h)} a_h})) (G_{(n-1)}^{(h)} a_h)^3 \\
+ (u_{n-1}^{(J)})^3 \lambda_t^{(+)} \Delta [3((\mu^{(+)} + \frac{D_{n-1}^{(+)}}{u_{n-1}^{(J)}})) (\sigma^{(+)})^2 + (\mu^{(+)} + \frac{D_{n-1}^{(+)}}{u_{n-1}^{(J)}})^3] \\
+ (u_{n-1}^{(J)})^3 \lambda_t^{(-)} \Delta [3((\mu^{(-)} + \frac{D_{n-1}^{(-)}}{u_{n-1}^{(J)}})) (\sigma^{(-)})^2 + (\mu^{(-)} + \frac{D_{n-1}^{(-)}}{u_{n-1}^{(J)}})^3]
\]
Excess Kurtosis

Excess Kurtosis is defined by:

\[ \text{ExcessKurt}_t(R_{t+1}) = \frac{\kappa_4}{\kappa_2^2} \]

\( \kappa_2 \) is given above, \( \kappa_4 \) is:

\[ \kappa_4 = 48(1 + 4(c_h - \frac{u_{n-1}^{(V)} \sqrt{\Delta}}{2G_n^{(h)} a_h})^2 h_t))(G_{(n-1)}^{(h)} a_h)^4 \]

\[ + (u_{n-1}^{(J)})^4 \lambda_t^{(+)} \Delta^2(\sigma^{(+)}^4 + 6(\mu^{(+)}^2 + \frac{D_{n-1}^{(+)}}{u_{n-1}^{(V)}}(\sigma^{(+)}^2 + 6(\mu^{(+)} + \frac{D_{n-1}^{(+)}}{u_{n-1}^{(V)}})^4) \]

\[ + (u_{n-1}^{(J)})^4 \lambda_t^{(+)} \Delta^2(\sigma^{(-)}^4 + 6(\mu^{(-)}^2 + \frac{D_{n-1}^{(-)}}{u_{n-1}^{(V)}}(\sigma^{(-)}^2 + 6(\mu^{(-)} + \frac{D_{n-1}^{(-)}}{u_{n-1}^{(V)}})^4) \]
Appendix C

Futures Realized Volatility

We used the ZMA (Zhang, Mykland and Ait-Sahalia, 2005) method to estimate daily realized covariance and used sub-sampling method to correct the bias, steps shown below:

1. Create a daily grid:
   (a) We include all trading information between 9:00 a.m to 2:30 pm (ET), which is 5 hours and 30 minutes, hence 19800 seconds. We should have a 19800 by 1 vector to represent the price grid for each day, since we assume that price is refreshed every second.
   (b) Refresh price every second: current price at a specific time means the nearest transaction price of all previous transactions (including price at that specific moment).
   (c) The price before the first transaction happens was just using the price of the first transaction on that day, to make calculation easier. Since there are usually at most a few seconds (out of 19800 seconds) before the first transaction, this will not make much difference.

2. Calculate the realized volatility estimator using complete data:

\[
RV_t^{(alt)} = \sum_{t_i \in G} (P_{t_i,+} - P_{t_i})^2
\]  

(C.1)

3. Sub-sampling method:
   (a) Partition the grid into J non-overlapping sub grids, each contains equally spaced observations. For each grid we have the realized volatility estimator calculated in the same
way as shown in equation C.1, but only on the sub-sampled grid. Terms from a-th sub-sample is denoted as \( RV_t^{(k,a)} \), \( a \in 1, 2, \ldots, J \).

(b) We use five-minute frequency data, which means we sample every 5 minutes, with different starting time. Since there are 300 seconds in 5 minutes, we are going to have 300 sub samples for each day (J=300). And we have \( RV^{\text{ave}} \) as follows (a is the index for sub-samples, \( a \in 1, 2, 3, \ldots, J \)), for both diagonal and off-diagonal terms:

\[
RV_t^{\text{ave}} = \frac{1}{J} \sum_{a=1}^{J} RV_t^{(a)} \tag{C.2}
\]

(c) the bias corrected volatility terms should be:

\[
RV_t = RV_t^{(\text{ave})} - \frac{\bar{m}}{m} RV_t^{(\text{all})} \tag{C.3}
\]

For both diagonal and off-diagonal terms, where \( \bar{m} = \frac{m - J + 1}{J} \) is the average number of G(j) sub grid elements.

In this way we have obtained realized volatility estimations for each trading day.
Bibliography


