Essays on Volatility Estimation and Forecasting of Crude Oil Futures

YANG, Xiaoran

A thesis submitted for the degree of Doctor of Philosophy in Finance

Essex Business School

University of Essex

October, 2016 (Submitted)

April, 2017 (Revised)

Colchester, Essex, the United Kingdom

Acknowledgments

First and foremost, I would like to express my deepest and most sincere gratitude to my supervisors Professor Neil Kellard and Dr Nikolaos Vlastakis for their excellent guidance, patience and providing me with an excellent atmosphere for doing research. My thesis could not have been accomplished without their assistance and dedicated involvement in every step through the process. I would like to thank you both very much for your support and understandings over these past four years. I have been extremely lucky to have Professor Neil Kellard and Dr Nikolaos Vlastakis as my supervisors.

I would also like to acknowledge the academic and administrative support of the University of Essex and the Essex Business School.

Last, none of this could have happened without my family. I would like to thank my parents YANG Ning and LIU Xiaohong for giving birth to me in the first place and their unconditional support, both financially and emotionally, throughout my life.

For any errors or inadequacies that may remain in this work, the responsibility is entirely my own.

CONTENTS

Introduction 1

Chapter 1. Forecasting Crude Oil Market Volatility by using GARCH models: Evidence of Using High Frequency Data and Daily Data 7

Abstract 7

1. Introduction 8

2. Literature Review 11

2.1. Forecasting by using high-frequency data 11

2.2 Forecast the crude oil volatility with daily data 14

3. Data and methodology 18

3.1. Data and data properties 18

3.2. Model estimation 29

3.3. Forecast and SPA test 34

4. Estimation results for different volatility models 38

5. Forecast comparison 43

6. Conclusion 52

References: 54

Chapter 2. Forecasting Crude Oil Market Volatility by using HAR-RV models: Evidence of Using High Frequency Data 59

Abstract 59

1. Introduction 60

2. Literature review 62

2.1. Forecasting the volatility of crude oil 62

2.2. Forecasting volatility by using realised volatility 64

3. Volatility estimation, jump specification and volatility modelling 65
3.1. Volatility estimation by using intraday data 65
3.2. Volatility model specification 68
4. Data description 72
5. Model Estimation 80
6. Forecast evaluation 91
6.1. Diebold-Mariano test 91
6.2. Superior Predictive Ability (SPA) test 95
6.3. The comparison of forecasting performance between HAR models and GARCH models 103
7. Conclusion 113
References 114
Chapter 3. Co-movement Estimation and Volatility Forecasting of Crude Oil Market and US Stock Market: Evidence of MGARCH, Wavelet and High Frequency Data 119
Abstract 119
1. Introduction 119
2. Literature Review 122
2.1. Cross market co-movements of crude oil market and stock markets 122
2. 2. Volatility forecast on financial assets. 128
3. Data Description 130
4. Methodology 143
4.1. Modelling dynamic conditional correlation 143
4.2 Wavelet method 145
4.3. Forecast 147
5. Empirical Finds and Analysis 150
5.1. The Empirical Findings of Wavelet analysis 150

5.2. The Empirical Findings of DCC-GARCH model 155

6. Forecast evaluation 169
7. Conclusion 173
References 175
Conclusion 179

Introduction

Volatility estimation and forecasting of financial assets, especially commodity assets such as crude oil, has been the focus of research in areas such as investment analysis, derivative securities pricing and risk management. Poon and Granger (2003) suggest that volatility forecasts can play the role of a "barometer for the vulnerability of financial markets and the economy". In this thesis, I estimate volatility of crude oil futures and evaluate the volatility forecasting performances of alternative models for crude oil futures by employing high-frequency data in Chapter 1 and Chapter 2. In Chapter 3, I link the volatility of crude oil market with that of the US stock market, study the co-movements of the most traded commodity and the stock market of the largest capitalisation by employing Multi-GARCH model and wavelet method and evaluate the forecasting performance of Multi-GARCH model on the two financial assets.

Comparatively, high frequency data/ intraday data contain more information than daily data on daily transactions and provide more accuracy on volatility estimation and forecast evaluation (Andersen & Bollerslev, 1998). Many studies advocate high frequency data (Koopman, Jungbacker & Hol, 2005; Marlik, 2005) and many studies evaluate the performance of different models on volatility forecasting (Andersen & Bollerslev, 1998; ABDL, 2001, 2003; Corsi, 2009; Engle & Gallo, 2006; Shephard & Sheppard 2010; Celik & Ergin 2014; Sevi, 2014).

The literature on volatility forecasting by using high-frequency data covers 4 main aspects: 1. assessments of the standard volatility model at high frequencies, 2. model comparisons by using between high-frequency and daily data, 3.studies of the realised volatility, 4. data properties of specific assets/series.

For the first aspect, there is still no consensus on whether other traditional time series models are able to capture the properties of high-frequency data or fit the intraday data. Researches supporting that the traditional time series models are able to fit the intraday data include Rahman & Ang (2002); Pong et al. (2004); Chortareas et al. (2014) but some other studies document opposite evidence (Jones, 2003; Baillie et al., 2004).

The second aspect of the volatility literature studies the virtues and drawbacks of using high-frequency data and compares volatility forecast evaluation by between using intraday data and using daily data. Beltratti & Morana (1999) show that at half-hour frequency the coefficients of the GARCH volatility model are not very different from those estimated on the basis of an IGARCH model. Hol and Koopman (2002) indicate that an ARFIMA model fitted to the realised volatility outperforms other alternative models. Martens and Zein (2004) find that high-frequency data improve both the measurement accuracy and the forecasting performance and they show that long memory models improve the forecasting performance. Pong et al. (2004) find that the most accurate volatility forecasts are generated using high frequency returns rather than a long memory model specification.

Many researches focus on realised volatility measure and its application. Since Andersen and Bollerslev (1998) demonstrate a dramatic improvement in the volatility forecasting performance of a daily GARCH model by using 5 min data as a volatility measure proxy, a great number of studies have focused on realised volatility forecasting and its properties. Andersen, Bollerslev, Diebold, and Labys (ABDL, 1999 and 2001) recommend forecasting the realised volatility by using the ARFIMA model and show that the realised volatility is a consistent estimator of the integrated volatility. The findings make contribution to the empirical basis of using the realised volatility in volatility forecasting directly. Tseng et al. (2009) find that realised range-based bi-power variation (RBV), a replacement of realised variance which is immune to jumps, is a better independent variable for future volatility prediction and the jump components of realised-range variance have little predictive power for oil futures contracts. Sevi (2014) studies the crude oil market with Heterogeneous Auto-Regressive model (HAR) and its variants of realised volatility and compare their performance in light with Diebold-Mariano test.

For the fourth part in the literature, many studies focus on the properties of high-frequency data for some specific financial assets. First order negative autocorrelation, non-normal distributions, an increasing fat tail with an increasing frequency, and periodicity are documented as stylised properties in the literature (Dacorogna et al. 2001). Microstructure noise and optimal sampling frequency (Hansen & Lunde (2006), Bandi & Russel (2005)) are well discussed as a technical topic for high-frequency data as well.

In this thesis, Chapter 1 assesses the standard volatility model at intraday frequency and makes model comparisons by using between high-frequency and daily data. Chapter 2 studies the realised volatility and compares the forecasting performance of realised volatility model and GARCH series model. The data properties of crude oil futures are determined in both chapters.

Chapter 1 fills the gap in the literature by modelling and forecasting crude oil volatility at both daily and intraday frequencies. I use a number of GARCH-class models to describe several facts on volatility based on the work of Kang et al. (2009) and Wei et al. (2010). I also adopt several loss functions including SPA test (Hansen, 2005) to evaluate the forecasting performance among different models. I discuss whether high frequency data of crude oil futures fit GARCH family models in the last. I find that none of the GARCH-class models outperforms the others at intraday data frequency. Our finding is against the results in ABDL (2001), Corsi (2009), Martens and Zein (2004) and Chortareas et al. (2011) which all document that long memory specification in high-frequency data can improve the forecasting power and accuracy significantly. EGARCH model is superior to other models when it comes to daily data and it is different from the finding of Kang et al. (2009) in which FIGARCH performs well.

My findings suggests that the traditional time series models are not good to fit intraday data. Therefore, new efforts should be made to find new models to forecast volatility in a high-frequency framework. I also find that the intraday crude oil returns are consistent with the stylised properties of other financial series such as stock market indices and exchange rates at high frequencies in many respects. It might reflect general features which all intraday data share.

Since the univariate GARCH models are documented as not fit for intraday data in Chapter 1, in Chapter 2 I assess the performance of Heterogeneous Autoregressive model of Realised volatility (HAR-RV) on crude oil futures with the same data set as in Chapter 1. Corsi (2009) proposes HAR-RV model and therefore introduces a way to specify and forecast volatility with the information of high-frequency data or intraday data in spite of the model's simple structure. Sevi (2014) expands the HAR-RV model by decomposing volatility into continuous and jump components, positive and negative semi-variance and considering leverage effect. His analysis suggests the decomposition of realised variance improves the in-sample fit but fails to improve the out-of-sample forecast performance. Following Sevi (2014) I specify and forecast volatility of the most traded commodity in the world by using front-month WTI futures contract. Moreover, I compare the forecasting performance among HAR-RV series models and GARCH series models which are studied in Chapter 1. It is valuable to compare HAR-RV models with GARCH and FIGARCH models because HAR-RV model is not able to depict the long memory property of volatility due to its simplicity while FIGARCH model considers the long memory character by using fractional integration.

In Chapter 2, I find that the decomposition of continuous components and signed jumps do not help to improve the in-sample fit. The in-sample fit of complicated HAR-RV models are as good as the simple HAR-RV model proposed by Corsi (2009). Second, the information of in-sample fit of semi-variance decomposition is mixed. Third, the complicated model containing all the decomposed components outperforms simple models or is as good as models without decomposed components at worst for prediction comparison. Last, the comparison between HAR series models and GARCH series models is inconclusive, which is against Andersen, Bollerslev, Christoffersen, and Diebold (2006, chap. 15), who find that even based on simple autoregressive structures such as the HAR provide much better results than GARCH-type models.

After adding findings to the literature on volatility forecasting by using high-frequency data of one single asset-crude oil in terms of the four aspects mentioned above, I extend the study of volatility forecasting of crude oil futures, a single financial asset to multi-asset background. Studying relationship between the crude oil market and stock markets is an ongoing issue in the finance literature recently. A large group of researchers are working on the strength of cross market relationship. Recent studies concentrating on the linkage between the oil market and the US stock market include Hammoudeh et al. (2004), Kilian and Park (2009), Balcilar and Ozdemir (2012), Elyasiani et al. (2012), Fan and Jahan-Parvar (2012), Alsalman and Herrera (2013), Mollick and Assefa (2013), Conrad et al. (2014), Kang et al. (2014), Khalfaoui et al. (2015) and Salisu and Oloko (2015). Since the introduction of the wavelet method, wavelet tool has become a small branch of finance research. In Chapter 3, I use the DCC-GARCH and wavelet-based measures of co-movements to find out the relationship between the two financial assets in time and frequency domain features of the data and make forecasting evaluation of DCC-GARCH model under different time frequencies. To the knowledge of mine, there is no empirical paper studying the linkage between crude oil and stock market with high frequency data or intraday data. Chapter 3 fills the gap in the existing literature.

In Chapter 3, I find that wavelet method helps to identify the long/short term investment behaviours at daily data frequency and that intraday data improve the forecast performance of traditional time series method. The findings of Chapter 3 have empirical implications in asset allocation and risk management for investment decisions such as the construction of dynamic optimal portfolio diversification strategies and dynamic value-at-risk methodologies.

Chapter 1. Forecasting Crude Oil Market Volatility by using GARCH models: Evidence of Using High Frequency Data and Daily Data

Abstract

We evaluate the performance of volatility estimation and forecast of West Texas Intermediate (WTI) crude oil futures based on intraday data and daily by employing a number of linear and nonlinear generalised autoregressive conditional heteroskedasticity (GARCH) class models. We assess the one-step out-of-sample volatility forecasts of the GARCH-class models by using different loss functions and the superior predictive ability (SPA) test for intraday data and daily data respectively. Our results indicate that the majority of GARCH series models except FIAPARCH model cannot provide satisfactory forecasting result of the volatility of WTI crude oil futures by using intraday data while EGARCH model for daily return data outperforms other models for WTI crude oil futures.

1. Introduction

Volatility forecasting of financial assets including commodity is one of the heated topics in finance research. Poon and Granger (2003) suggest that volatility forecasts can play the role of a "barometer for the vulnerability of financial markets and the economy". On the other hand, Modelling and forecasting crude oil volatility are important inputs into econometric models, portfolio selection models, and option pricing formulas. The access to high frequency data opens a new stage to volatility modelling and forecasting of returns of financial assets. In this paper, we assess the volatility forecasting performances of a number of GARCH class models for NYMEX WTI light crude oil futures by using high-frequency data and daily data respectively.

Compared with traditional daily data—daily returns or daily volatility, high frequency data contain more information on daily transactions and provide more accuracy on volatility estimation and forecast evaluation (Andersen & Bollerslev, 1998). Many studies advocate high frequency data (Koopman, Jungbacker & Hol, 2005; Marlik, 2005) and a number of studies evaluate the performance of different models on volatility forecasting (Andersen & Bollerslev, 1998; ABDL, 2001, 2003; Corsi, 2009; Engle & Gallo, 2006; Shephard & Sheppard 2010; Celik & Ergin 2014, Sevi, 2014).

A lot of studies are conducted on foreign exchange volatility forecasting (ABDL, 2001, 2003; Martens, 2001; Chortareas et al. 2011) and the volatility forecasting on stock markets (Chernov et al. 2003; Celik & Ergin 2014) by employing high frequency or intraday data, but limited research has been done on forecasting the volatility of crude oil by employing high frequency data/ intraday data (Sevi 2014) to the best of our knowledge.

Our study fills the gap in the literature by modelling and forecasting crude oil volatility at both daily and intraday frequencies. My work extends the previous research in three different ways.

First, based on the work of Kang et al. (2009) and Wei et al. (2010), I use a number of GARCH-class models to describe several facts about volatility. Second, I adopt several loss functions including SPA test (Hansen, 2005) to evaluate the forecasting performance among different models. Third, we discuss whether the employment of high frequency data of crude oil futures fits GARCH family models.

We find that most of the GARCH-class models cannot outperform the others when it comes to intraday data except FIAPARCH model. FIAPARCH model's performance is in line with some research papers in the literature ABDL (2001), Corsi (2009), Martens and Zein (2004) and Chortareas et al. (2011) which all document that long memory specification in high-frequency data can improve the forecasting power and accuracy significantly. The different results for other complicated GARCH models stem from the more up-to-date data sample period used in this study. EGARCH model is superior to other models when it comes to daily data and it is different from the finding of Kang et al. (2009) in which FIGARCH performs well.

Our findings provides a solid piece of evidence to the cons part in the discussion that whether the traditional time series models are good to fit intraday data. We find that the traditional volatility model cannot fit the data when we employ intraday data. After de-seasonalising the raw returns of the crude oil futures and putting in GARCH family models, it emerges that no GARCH model can produce satisfactory forecast results except FIAPARCH model. Thus, the new efforts should be made to find new models to forecast volatility in a high-frequency framework.

We find that the intraday crude oil returns are consistent with the stylised properties of other financial series such as stock market indices and exchange rates at high frequencies in many respects. This becomes a piece of evidence that these properties are not limit to certain kinds of high-frequency data. It might reflect some general features which all intraday data share.

The paper proceeds as follows. Section 2 reviews some of the main findings in the volatility forecasting literature. Section 3 discusses the data and methodology I use. Section 4 introduces estimation results. Section 5 compares the out-of-sample forecast performance of alternative models. Section 6 concludes.

2. Literature Review

2.1. Forecasting by using high-frequency data

The literature on volatility forecasting by using high-frequency data covers 4 aspects mainly: 1.studies of the realised volatility, 2. model comparisons by using between high-frequency and daily data, 3. assessments of the standard volatility model at high frequencies, and 4. data properties of specific assets/series.

Since the true volatility is unobservable, daily squared returns are often used as a proxy measure of volatility. By using 5 min data as a new volatility measure, Andersen and Bollerslev (1998) demonstrate a dramatic improvement in the volatility forecasting performance of a daily GARCH model (foreign exchange). Since then, a great number of studies have focused on realised volatility forecasting and its properties. Andersen, Bollerslev, Diebold, and Labys (ABDL, 1999 and 2001) recommend forecasting the realised volatility by using the ARFIMA model and show that the realised volatility is a consistent estimator of the integrated volatility. ABDL (2001) show that if realised volatility is modelled directly by a parametric model rather than simply being used in the evaluation of other models' forecasting behaviours, the realised volatility can improve forecasting when it comes to the ARFIMA model on foreign exchange rates. The findings above make contribution to the empirical basis of using the realised volatility in volatility forecasting directly but it is limited to foreign exchange rate.

The second aspect of the volatility literature studies the virtues and drawbacks of using high-frequency data and compares volatility forecast evaluation by between using intraday data and using daily data. Beltratti & Morana (1999) estimate volatility models on the basis of high frequency (half-hour) data for the Deutsche mark–US dollar exchange rate and compare the results to those obtained from volatility models estimated on the basis of daily data. Their high frequency data cover 1996 (from January 1, 1996 to December 31,1996, excluding week-ends and holidays), containing 12576 observations excluding week-ends while the daily data they use start with December 31, 1972 and end with January 31, 1997, corresponding to 6545 observations. They apply MA(1)-GARCH(1,1), MA(1)-GARCH(2,1) and MA(1)-FIGARCH(1,d,1) models to two sets of data. They categorise high-frequency data into three kinds: raw returns, deterministically filtered returns and stochastically filtered returns and they apply GARCH model and FIGARCH model to the three kinds of returns respectively. They show that even at the high (half-hour) frequency the coefficients of the GARCH volatility model are not very different from those estimated on the basis of an IGARCH model. Marlik (2005) studies the foreign exchange volatility by using hourly data of the British pound and the euro vis-a-vis the U.S. dollar. The period to which the data correspond starts in December 2001 and ends in March 2002 and is approximately the same for both currencies. Put it in another way, the author uses hourly data covering four months. The author applies GARCH model, FIGARCH, EGARCH, FIEGARCH and SV models to the two currencies. Moreover the author just employs raw return of hourly data instead of filtered returns. They find that euro is considerably more volatile when compared to British pound. Martens (2001) studies volatility forecast of foreign exchange by using half-hour returns of several major exchange rates: the spot rate between the Deutsche mark and the US dollar (DEM/USD) and that of the Japanese yen and the US dollar (YEN/USD) for all of 1996. The author excludes the returns from Friday 21:00 GMT through to Sunday 21:00 GMT thus leaves 261 days each with 48 half-hour returns in his research. The author sets July 1 through to December 31, 1996 as out-of-sample forecast period for the daily volatility forecasts for the DEM/USD and YEN/USD exchange. GARCH models are applied to de-seasonalised returns and raw returns respectively. Martens and Zein (2004) find that high-frequency data improve both the measurement accuracy and the forecasting performance and they show that long memory models improve the forecasting performance. Hol and Koopman (2002) use S&P 100 stock index to compare the predictive powers of realised volatility models and daily time-varying volatility models and their out-of-sample evaluation result indicate that an ARFIMA model fitted to the realised volatility outperforms other alternative models. Pong et al. (2004) compare exchange rate volatility forecasts obtained from an option implied volatility model, a short memory model (ARMA), a long memory model (ARFIMA) and a daily GARCH model. They find that the most accurate volatility forecasts are generated using high frequency returns rather than a long memory specification.

It is proved that the realised volatility model is able to fit the intraday data and has a good performance, however, there is still no consensus on whether other traditional time series models are able to capture the properties of high-frequency data or fit the intraday data. Rahman & Ang (2002) study the intra-day return volatility process by employing NASDAQ stock data. Their data set consists of transaction prices, bid-ask spread, and trading volumes from January 1, 1999 to March 31, 1999, for a subset of thirty stocks from NASDAQ 100 Index. They calculate 5 minute returns for this sample period. They add trading volume to the regression of conditional variance equation of GARCH model and they find that a standard GARCH (1, 1) is able to describe the intraday volatility. Chortareas et al. (2014) find that the traditional volatility model could also be an alternative for volatility forecasting in a high-frequency framework and should be considered along with the newer models but some other research document opposite evidence (Jones, 2003). Baillie et al. (2004) use three spot exchange rates: the British pound (BP), Swiss franc (SF) and the Deutsche mark (DM) vis-a-vis the US dollar ($) to measuring non-linearity, long memory and self-similarity. They use two datasets from quite distinct periods where the underlying institutional dynamics are different, to see if the apparent data generating process remains stable. The first dataset they use are recorded every hour from 0.00 a.m. (2 January 1986) through 11:00 a.m. (15 July 1986) at Greenwich Mean Time (GMT). The second dataset contains every 30 min spot price for the complete 1996 calendar year for the DM–$, $–BP and SF–$ exchange rates. The sample period is from 00:30 GMT (1 January 1996) through 00:00GMT (1 January 1997). They filter the return series with two methods: non-linear deterministic method and stochastic methodology and they apply MA-FIGARCH model to the two filtered return series. They find that the estimates of the long memory parameter are remarkably consistent across time aggregations and currencies and are suggestive of self-similarity but it is found to be too weak to be exploitable for forecasting purposes.

For the fourth part, many studies focus on the properties of high-frequency data for some specific financial assets. First order negative autocorrelation, non-normal distributions, an increasing fat tail with an increasing frequency, and periodicity are documented as stylised properties in the literature (Dacorogna et al. 2001). Microstructure noise and optimal sampling frequency (Hansen & Lunde (2006), Bandi & Russel (2005)) are well discussed as a technical topic for high-frequency data as well.

2.2 Forecast the crude oil volatility with daily data

Agnolucci (2009) compares the predictive ability of two approaches which can be used to forecast volatility: GARCH-type models where forecasts are obtained after estimating time series models and an implied volatility model where forecasts are obtained by inverting one of the models used to price options. He has estimated GARCH models by using daily returns from the generic light sweet crude oil future based on the West Texas Intermediate (WTI) traded at the NYMEX. Data on the price of the contract have been sourced from the Bloomberg database. The collected sample goes from 31/12/1991 to 02/05/2005. The WTI future contract quoted at the NYMEX is the most actively traded instrument in the energy sector. He evaluates which model produces the best forecast of volatility for the WTI future contract, evaluated according to statistical and regression-based criteria, and also investigates whether volatility of the oil futures are affected by asymmetric effects, whether parameters of the GARCH models are influenced by the distribution of the errors and whether allowing for a time-varying long run mean in the volatility produces any improvement on the forecast obtained from GARCH models.

Kang et al. (2009) investigate the efficacy of volatility models for three crude oil markets — Brent, Dubai, and West Texas Intermediate (WTI) — with regard to its ability to forecast and identify volatility stylized facts, in particular volatility persistence or long memory. The data they use are three crude oil spot prices (in US dollars per barrel) obtained from the Bloomberg databases. The datasets consist of daily closing prices over the period from January 6, 1992 to December 29, 2006, and the last one year's data are used to evaluate out-of-sample volatility forecasts. They assess persistence in the volatility of the three crude oil prices using conditional volatility models. The CGARCH and FIGARCH models are better equipped to capture persistence than are the GARCH and IGARCH models. The CGARCH and FIGARCH models also provide superior performance in out-of-sample volatility forecasts. They conclude that the CGARCH and FIGARCH models are useful for modelling and forecasting persistence in the volatility of crude oil prices. Wei et al. (2010) extend the work of Kang et al. (2009). They use a number of linear and nonlinear GARCH models to capture the volatility features of two crude oil markets: Brent and WTI. They also carry out superior predictive ability test (SPA test) and other loss functions to evaluate the forecasting power of different models. They use daily price data (in US dollars per barrel) of Brent and WTI from 6/1/1992 to 31/12/2009.

Mohammadi and Su (2010) examine the usefulness of several ARIMA-GARCH models for modelling and forecasting the conditional mean and volatility of weekly crude oil spot prices in eleven international markets over the 1/2/1997–10/3/2009 period with weekly data. In particular, they investigate the out-of-sample forecasting performance of four volatility models — GARCH, EGARCH and APARCH and FIGARCH over January 2009 to October 2009. Forecasting results are somewhat mixed, but in most cases, the APARCH model outperforms the others. Also, conditional standard deviation captures the volatility in oil returns better than the traditional conditional variance. Finally, shocks to conditional volatility dissipate at an exponential rate, which is consistent with the covariance-stationary GARCH models than the slow hyperbolic rate implied by the FIGARCH alternative.

Hou and Suardi (2012) consider an alternative approach involving nonparametric method to model and forecast oil price return volatility considering the use of parametric GARCH models to characterise crude oil price volatility is widely observed in the empirical literature. Focusing on two crude oil markets, Brent and West Texas Intermediate (WTI), they show that the out-of-sample volatility forecast of the nonparametric GARCH model yields superior performance relative to an extensive class of parametric GARCH models. The data which are sampled from 6 January 1992 to 30 July 2010 are obtained from DataStream database service. The improvement in forecasting accuracy of oil price return volatility based on the nonparametric GARCH model suggests that this method offers an attractive and viable alternative to the commonly used parametric GARCH models.

Though crude oil plays a vital role in commodity market and global economy, few research focus on forecasting the crude oil volatility based on high-frequency data and on how alternative models outperform others. Corsi (2009) and Sevi (2014) study the volatility estimation and forecasting of crude oil futures with intraday data with HAR-type model. This paper focuses on crude oil volatility forecasting at high frequencies and the comparison of alternative GARCH-series models' forecasting performance and thus, fills the gap in the existing literature.

3. Data and methodology

3.1. Data and data properties

The original data we obtain are 15 min price data of the NYMEX light, sweet (low-sulphur) crude oil futures contract provide by Tick Data. Crude oil futures is the world's most actively traded commodity, and the NYMEX light, sweet (low-sulphur) crude oil (WTI) futures contract is the world's most liquid crude oil futures, as well as the world's largest-volume futures contract trading on a physical commodity. The data I use span the period from 25th March 2009 to 25th March 2013, containing 1033 trading days.

High frequency data contain more information on financial assets. Theoretically, the higher the frequency of the data, the more accurate the volatility estimation will be. While on the other hand, microstructure frictions, such as price discreteness and measurement errors may affect the effectiveness of high frequency data (ABDL, 1999; Bandi & Russell, 2005). I employ 15 minute data in this paper in order to mitigate microstructure effects of high frequency data, which is consistent with ABDE (2001).

NYMEX light, sweet (low-sulphur) crude oil futures has open outcry trading from 9:00 to 14:30 EST on weekdays. Investors can also trade oil futures via NYMEX electronic trading platform from 17:00 on Sunday to 17:15 the next day and from 18:00 to 17:15 (New York Time) on weekdays. The trading volumes on weekends are rather small therefore we remove weekend returns from the sample following the common practice in the literature (Chortareas et al. 2011; Celik & Ergin 2014). I obtain 89732 observations in total after the data is cleared. The daily data is used as a comparison.

The intraday return series $r_{t,m}$ is given as follow:

$$r_{t,m} = \ln\left(P_{t,m}\right) - \ln\left(P_{t,m-1}\right) \quad (1)$$

Where $P_{t,m}$ is the close-mid price at the mth time stamp on day t. Figure 1 shows the intraday prices of crude oil futures.

The daily return r_t is given as follows:

$$r_t = \ln(P_t) - \ln(P_{t-1}) \tag{2}$$

Figure 2 shows the comparison between the intraday returns of NYMEX light, sweet (low-sulphur) crude oil futures return series and those of the daily returns. Figure 3 indicates the comparison between the realised volatility and the daily volatility. Figure 4 shows the distribution of the 15 min returns and daily returns. Table 1 represents the descriptive statistics of the two intraday/daily return series.

Figure 1. Plots of 15 minute price series.

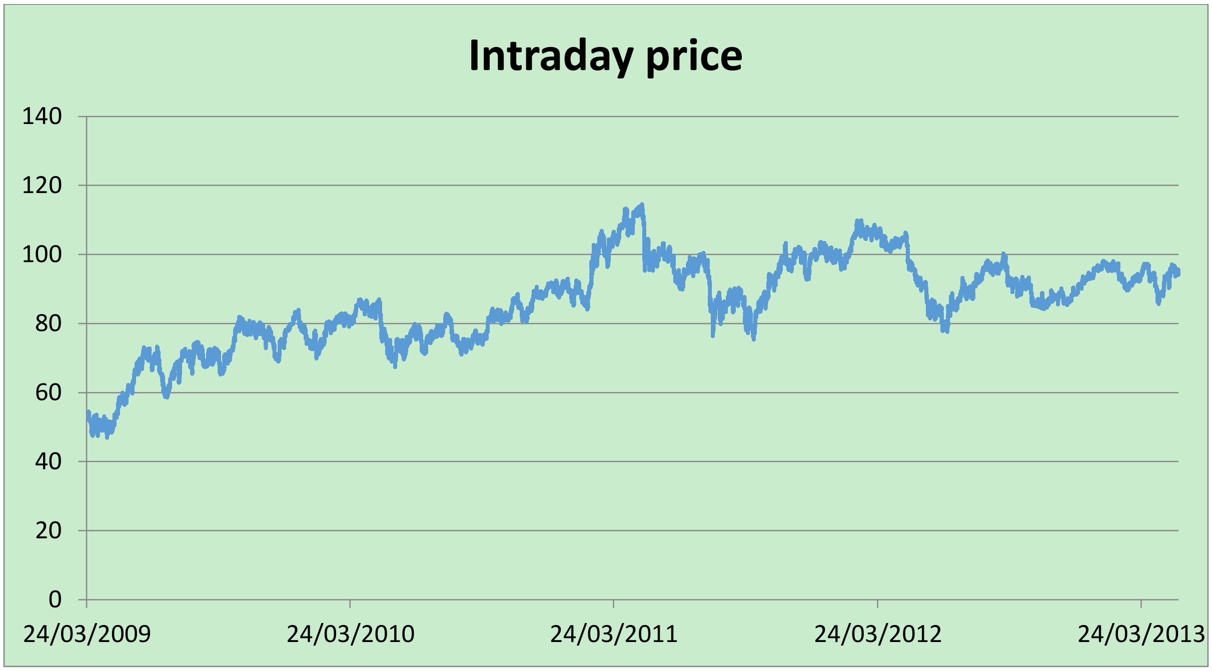


Figure 2. Plots of 15 minute return series and daily return series.

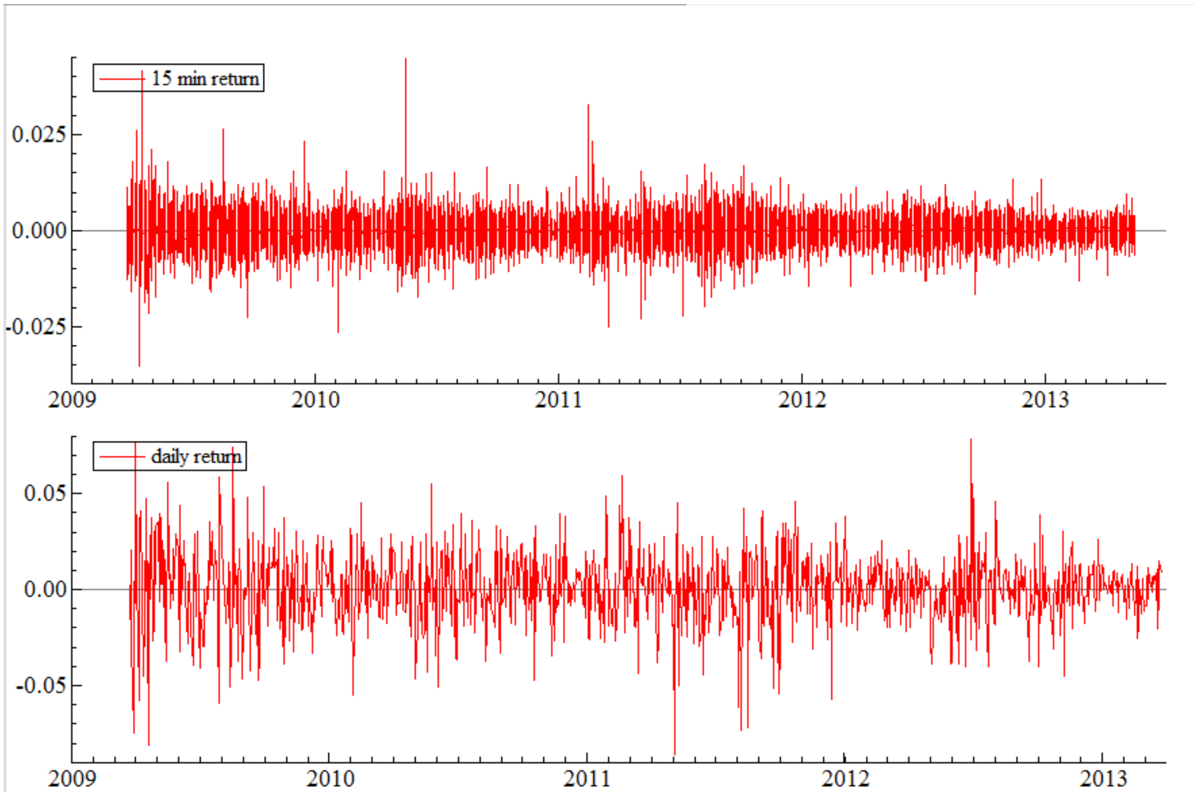


Figure 3. Plots of realised volatility and daily volatility.

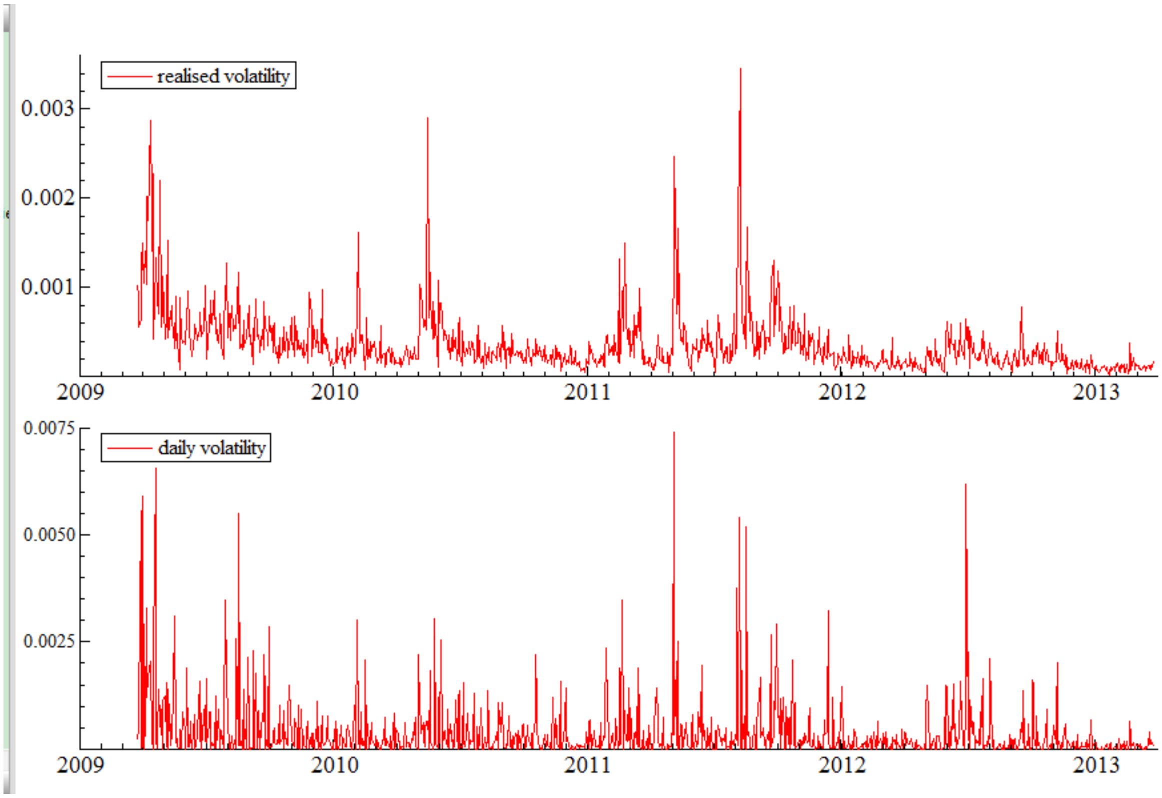


Figure 4. The distribution of 15 min return data and the daily return data

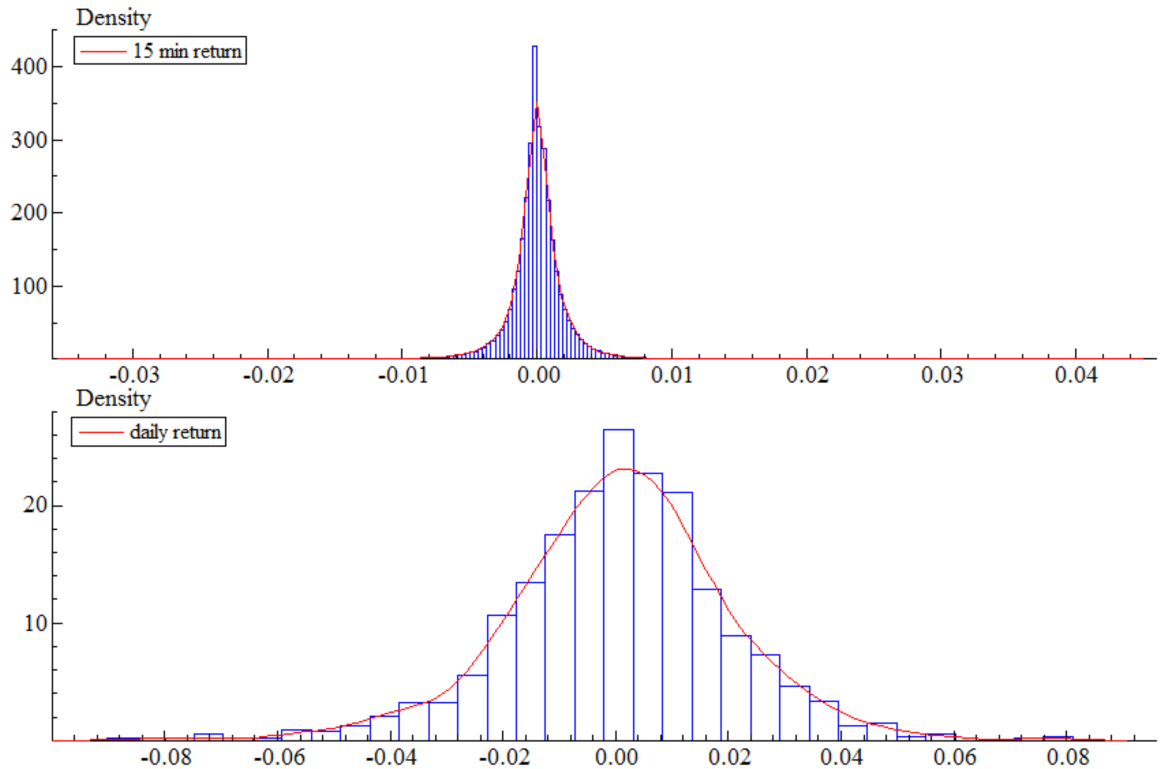


Table 1. Summary statistics of 15 minute returns series and daily return series.

	Mean ($\times 10^{-6}$)	S.D ($\times 10^{-3}$)	Skewness	Kurtosis	ADF	GPH
15min return	6.21	2.046	0.070065	19.07676	-303.574	-0.005 (0.003)
Daily return	550	19.646	-0.22522	4.674699	-34.0487	-0.056 (0.029)

Notes: The table shows the descriptive statistics of the 15 min returns and daily returns of the crude oil futures. Both series are skewed and fat tailed distributed. The sample period is from 25th March 2009 to 25th March 2013, containing 1033 trading days. The standard errors are in the parentheses in the last column.

Figure 2 shows that the movements of the 15 min returns and the daily returns are not consistent. High-frequency data carry more information thus several jumps in the daily returns are smoothed out in the 15 min returns. Figure 3 also indicates the inconsistence between the realised volatility which is constructed from the squared intraday returns and daily volatility which is equal to the squared daily returns. The movements of the two volatility proxies are not synchronised and the scalars of the two volatilities on the Y-axis are not the same. It is shown that the values of the realised volatility are much smaller than the values of the daily volatility. The distributions of the 15 min returns present that the 15 min returns are much more leptokurtic than the daily returns.

Numbers in Table 1 indicate features of 15 minute returns of crude oil and these of daily returns. The crude oil shares some stylised properties of high-frequency returns of other financial assets in the literature. The mean value of crude oil returns is approximately zero, which is common among financial assets. The skewness of crude oil intraday return is 0.07, suggesting the distribution leans leftward. The kurtosis is way larger than 3, indicating the distribution is fat tailed. The augmented Dickey-Fuller unit root test supports the rejection of the null hypothesis of a unit root at the 1% significance level, implying the return series is stationary. The p-value of the GPH test on the 15 min returns is 0.0833, implying the non-rejection of the null hypothesis that the long memory parameter is zero. Meanwhile the statistics of the daily returns are different from the intraday returns. The mean and standard deviation are much larger than those of the 15 min returns and the skewness is negative rather than positive compared to the skewness of the 15 min returns. The negative skewness indicates the distribution of daily returns is rightward rather than leftward which is a feather of the 15 min returns. The negative value of the ADF test statistics implies the daily returns are stationary and the GPH test result indicates the long memory parameter is zero.

Dacorogna et al. (2001) find that a well-documented stylised fact of high-frequency returns which is the negative first order autocorrelation in the return. Figure 5 indicates the autocorrelation function of the 15 min return series of crude oil. The first order autocorrelation of the 15 min returns of crude oil is negative, which is consistent with the literature (Goodhart, 1989; Goodhart and Figliuoli, 1992; Goodhart et al. 1995). Literature documents that a large negative autocorrelation is followed by rather small autocorrelations in the subsequent lags which is caused by the bounce between the bid and ask prices. However, for the crude oil return, the first order autocorrelation is just -0.012, which is not large enough to dominate the subsequent lags. The coefficients of autocorrelations in the subsequent lags are close to zero and the P-values of the Q-stat are almost zero for the following 12 lags thus the null hypothesis of no autocorrelation for 12 lags cannot be rejected. However, considering the small amount of the first order autocorrelation, we will not take moving average into consideration when we construct the mean equation of the regression in the following parts of this paper.

Figure 5. The autocorrelation function of the 15 minute returns (12 lags)

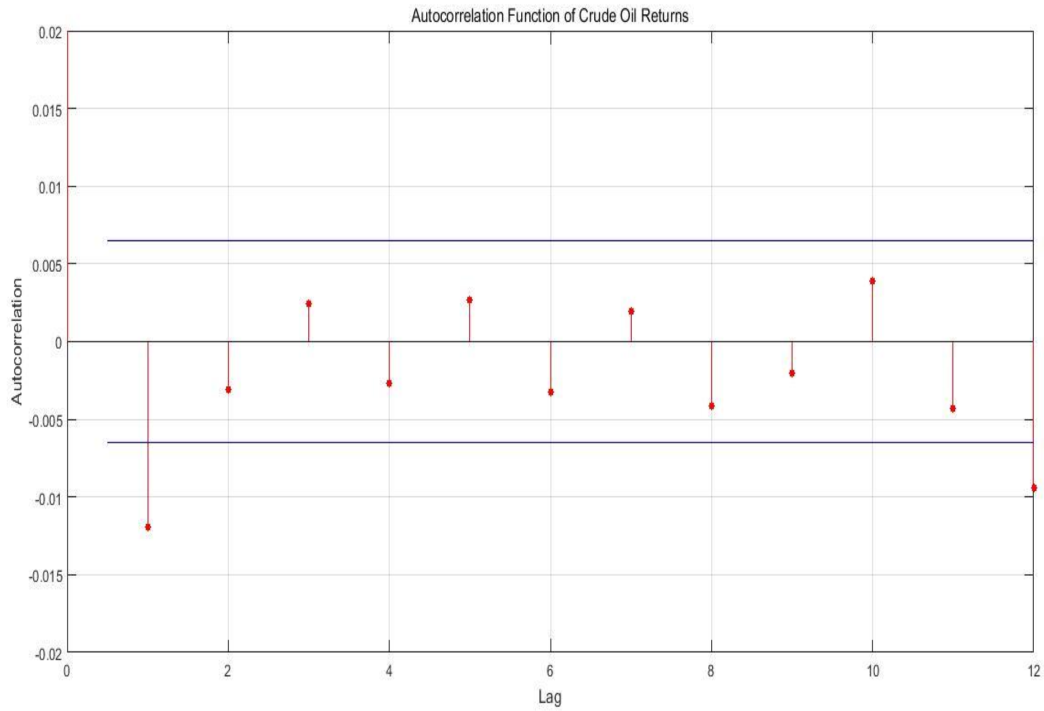
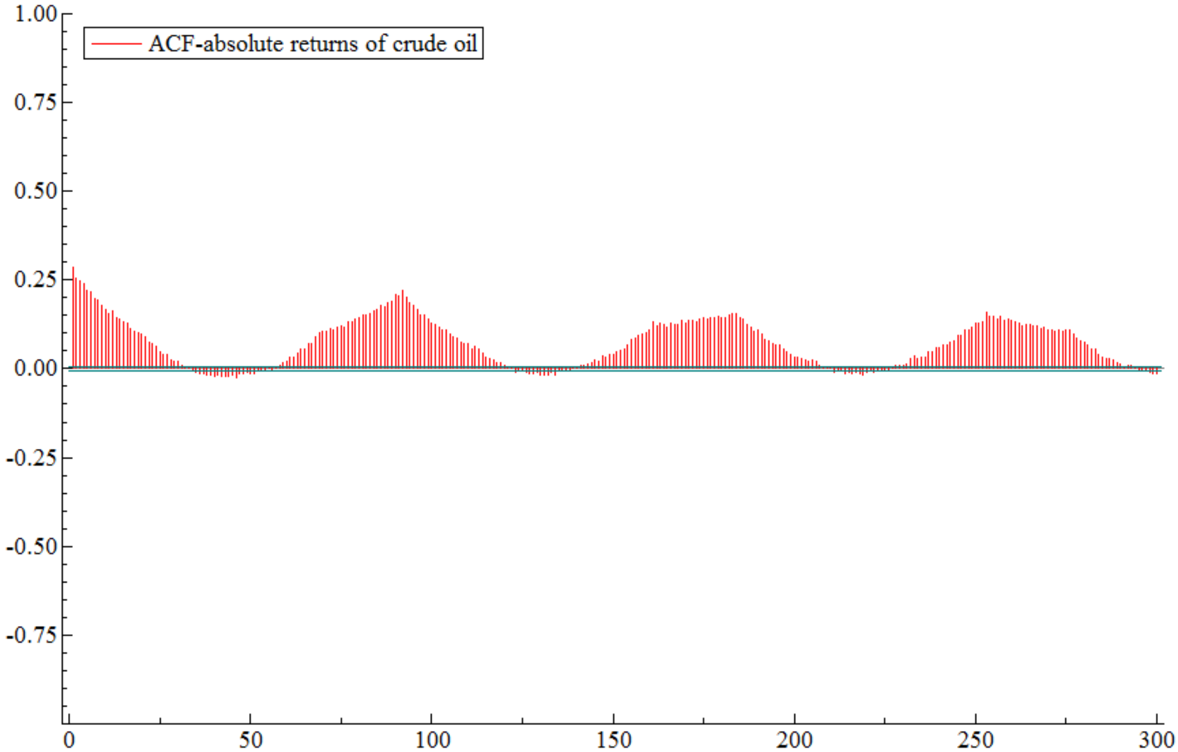


Figure 6. The autocorrelation function of absolute 15 min returns for crude oil futures for 300 lags.



Periodicity is another stylised fact of intraday volatility series. Figure 6 shows the autocorrelation function of absolute returns for crude oil futures. The U-shaped plot reveals the periodicity in a trading day. Crude oil is traded from Sunday to Friday 6:00 p.m. - 5:15 p.m. New York time/ET with a 45-minute break each day beginning at 5:15 p.m. thus there are 278 observations for each 24 hours. One can observe that the U pattern recurs approximately at 92 lags, suggesting periodicity within one day. The autocorrelation peaks at the beginning and the end of the 24 hour grids and it bottoms in the midday. This finding is consistent with those of other studies (Andersen and Bollerslev, 1997; Barbosa, 2002; Dacorogna et al. 2001). There is no sign of disappearance of autocorrelation in the absolute returns in Figure 6.

In brief, the return series of the 15 min crude oil in my study shares the stylised facts of high frequency financial returns well documented in the literature. It has a zero mean while it is fat tailed and marginally positive skewed. The return series exhibits small negative first order autocorrelation and it reveals that periodicity pattern exists in intraday volatility.

3.2. Model estimation

The volatilities of intraday returns have a strong periodicity in 1-day interval, which is demonstrated in the previous section. Martens et al. (2002) suggest that intraday periodic patterns do not fit the traditional time series models, (e.g., GARCH-type models) directly because the GARCH-type model are easily distorted by the pattern. Thus, we use the de-seasonalised filtered returns to estimate GARCH-type models instead of the original returns directly. According to Taylor and Xu (1997), we have

$$\tilde{r}_{t,n} = \frac{r_{t,n}}{S_{t,n}} \quad (n = 1,2, \dots, N) \tag{3}$$

where $r_{t,n}$ is the nth intraday return on day t and $S_{t,n}$ is the corresponding seasonality term, for N intraday periods. $S_{t,n}$ is equal to the averaging the squared returns for each intraday period:

$$S_{t,n}^2 = \frac{1}{T}\sum_{t=1}^{T} r_{t,n}^2 \quad (n = 1,2, \dots, N) \tag{4}$$

where T is the number of days in the sample. It's an effective method to smooth the seasonality feature so we use the de-seasonalised returns in the following part of the paper.

The intraday return series is nearly symmetric and has a high kurtosis thus I assume the returns series follows the symmetric student T distribution while for the symmetric student T distribution,

$$E\left|z_{t,n-1}\right| = 2\frac{\Gamma\left(\frac{1+v}{2}\right)\sqrt{v-2}}{\sqrt{\pi}\Gamma(v/2)} \tag{5}$$

where v indicates the degree of freedom of the student T distribution and $\Gamma(.)$ is the Gama function.

We employ a series of GARCH family models for two different time frequencies for volatility forecasting. Bollerslev (1986) proposes the GARCH model and Sadorsky (2006) demonstrates that the GARCH (1, 1) model works well for crude oil volatility. The standard GARCH (1, 1) model for intraday data is given by:

$$\tilde{r}_{t,n} = \mu + \varepsilon_{t,n}\,, \quad \varepsilon_{t,n}|\Omega_{t,n-1} \sim T_v(0, h_{t,n})$$

$$h_{t,n} = \omega + \alpha\varepsilon_{t,n-1}^2 + \beta h_{t,n-1} \tag{6}$$

where μ denotes the conditional mean, ω, α and β are the parameters of the variance equation with parameter restrictions $\omega > 0$, $\alpha > 0$, $\beta > 0$ and $\alpha + \beta < 1$. The error term $\varepsilon_{t,n}$ based on the information set $\Omega_{t,n-1}$ follows a student's T distribution T_v with zero mean, variance $h_{t,n}$ and degree of freedom v. Considering the expected return of the intraday price is almost zero, the conditional mean μ will not be reported in the following parts of the paper while it is still in the regression. The daily GARCH model is given as follows:

$$r_t = \mu + \varepsilon_t\,, \quad \varepsilon_t|\Omega_{t,n-1} \sim T_v(0, h_t)$$

$$h_t = \omega + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1}\,, \tag{7}$$

The restrictions on parameters of the daily GARCH model are the same as these of the intraday

GARCH model. The error term of the daily GARCH model also follows a student's T distribution T_v with zero mean, variance $h_{t,n}$ and degree of freedom v.

Engle and Bollerslev (1986) introduced IGARCH model which captures infinite persistence in the conditional variance. The model setting of IGARCH model is similar to that of the GARCH model but with the parameter restriction $\alpha + \beta = 1$. We also apply IGARCH model to both intraday returns and daily returns. Thus for intraday returns, the IGARCH model is given as follows:

$$\tilde{r}_{t,n} = \mu + \varepsilon_{t,n}\ ,\ \ \varepsilon_{t,n}|\Omega_{t,n-1} \sim T_v(0, h_{t,n})$$

$$h_{t,n} = \omega + \alpha\varepsilon_{t,n-1}^2 + \beta h_{t,n-1} \tag{8}$$

$$s.t. \alpha + \beta = 1$$

And the daily IGARCH model is expressed as:

$$r_t = \mu + \varepsilon_t\ ,\ \ \varepsilon_t|\Omega_{t,n-1} \sim T_v(0, h_t)$$

$$h_t = \omega + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1}\ , \tag{9}$$

$$s.t. \alpha + \beta = 1$$

Cont (2001) presents the stylised facts of financial assets such as long memory volatility effect and asymmetric leverage effect and others. Many GARCH family models are developed to capture these stylised features of the financial assets. We will apply the following GARCH family models to estimate and forecast the volatility of crude oil futures to capture long memory volatility effect and asymmetric leverage effect.

Glosten et al. (1993) construct the GJR model to capture the asymmetric leverage volatility effect, i.e., the negative shocks will have larger impact on the volatility of the time series. The GJR model for intraday returns is given as follows:

$$\tilde{r}_{t,n} = \mu + \varepsilon_{t,n}\ ,\ \ \varepsilon_{t,n}|\Omega_{t,n-1} \sim T_v(0, h_{t,n})$$

$$h_{t,n} = \omega + [\alpha + \gamma I(\varepsilon_{t,n-1} < 0)]\varepsilon_{t,n-1}^2 + \beta h_{t,n-1}, \tag{10}$$

where $I(.)$ is an indicator function. If $\varepsilon_{t,n-1}$ is negative, then $I(.) = 1$ and $I(.) = 0$ if $\varepsilon_{t,n-1}$ is not negative. γ is the asymmetric leverage coefficient and it captures the leverage effect of the volatility.

The GJR model setting for the daily returns is given as follows:

$$r_t = \mu + \varepsilon_t \, , \;\; \varepsilon_t|\Omega_{t,n-1} \sim T_v(0, h_t)$$

$$h_t = \omega + [\alpha + \gamma I(\varepsilon_{t-1} < 0)]\varepsilon_{t-1}^2 + \beta h_{t-1} \, , \tag{11}$$

EGARCH model (Nelson, 1990) is another GARCH family model which captures the volatility leverage effect. Nelson argues that the nonnegative constraints in the linear GARCH model are too restrictive. To loosen the nonnegative constraints on parameters α and β of GARCH model, Nelson proposes the EGARCH model where no restrictions are placed on these parameters in the EGARCH model. The specification of EGARCH model for the intraday returns is

$$\tilde{r}_{t,n} = \mu + \varepsilon_{t,n} \, , \;\; \varepsilon_{t,n}|\Omega_{t,n-1} \sim T_v(0, h_{t,n})$$

$$\log(h_{t,n}) = \omega + \alpha z_{t,n-1} + \gamma(|z_{t,n-1}| - E|z_{t,n-1}|) + \beta \log(h_{t,n-1}), \tag{12}$$

Where $z_{t,n-1}$ depends on the assumption made on the unconditional density of $z_{t,n-1}$ and γ is the asymmetric leverage coefficient to capture the volatility leverage effect.

The EGARCH model for daily return is given as:

$$\tilde{r}_{t,n} = \mu + \varepsilon_t \, , \;\; \varepsilon_t|\Omega_{t,n-1} \sim T_v(0, h_t)$$

$$\log(h_t) = \omega + \alpha z_{t-1} + \gamma(|z_{t-1}| - E|z_{t-1}|) + \beta \log(h_{t-1}), \tag{13}$$

GARCH models above capture short-term volatility features while fractionally integrated GARCH (FIGARCH) model (Baillie et al., 1996, 2004; Andersen and Bollerslev, 1997) captures the long memory properties of the volatility. The FIGARCH model assumes the finite persistence of volatility shocks (no such persistence exists in the GARCH framework), i.e., long-memory behaviour and a slow rate of decay after a volatility shock. Comparatively, an IGARCH model implies the complete persistence of a shock, and apparently quickly fell out of favour. The FIGARCH(1,d,1) is reduced to a GARCH(1,1) if the fractional integration parameter d is 0 and it is reduced to an IGARCH(1,1) if d is 1. The FIGARCH (1, d, 1) model for intraday returns can be written as follows:

$$\tilde{r}_{t,n} = \mu + \varepsilon_{t,n}\,, \quad \varepsilon_{t,n}|\Omega_{t,n-1} \sim T_v(0, h_{t,n})$$

$$h_{t,n} = \omega + \beta h_{t,n-1} + [1 - (1 - \beta L)^{-1}(1 - \varphi L)(1 - L)^d]\varepsilon_{t,n}^2\,, \tag{14}$$

where $0 \leq d \leq 1$, $\omega > 0$, $\varphi, \beta < 1$. d is the fractional integration parameter and L is the lag operator. The fractional integration parameter d allows autocorrelations to decay at a slow hyperbolic rate which characterises the long-memory feature. If d is set between zero and one, FIGARCH model is able to describe intermediate ranges of persistence since it lies within d=1 representing the complete integrated persistence of volatility shocks and d=0 representing the geometric decay.

The FIGARCH specification for the daily return is given as follows:

$$r_t = \mu + \varepsilon_t\,, \quad \varepsilon_t|\Omega_{t,n-1} \sim T_v(0, h_t)$$

$$h_t = \omega + \beta h_{t-1} + [1 - (1 - \beta L)^{-1}(1 - \varphi L)(1 - L)^d]\varepsilon_t^2 \tag{15}$$

Based on FIGARCH, Tse (1998) introduces the fractionally integrated asymmetric power ARCH (FIAPARCH) model to capture long memory and asymmetry in volatility simultaneously. The FIAPARCH (1, d, 1) model for intraday returns is written as follows:

$$\tilde{r}_{t,n} = \mu + \varepsilon_{t,n}\,, \quad \varepsilon_{t,n}|\Omega_{t,n-1} \sim T_v(0, h_{t,n})$$

$$h_{t,n} = \omega(1 - \beta)^{-1} + [1 - (1 - \beta L)^{-1}(1 - \varphi L)(1 - L)^d](|\varepsilon_{t,n}| - \gamma\varepsilon_{t,n})^\delta\,, \tag{16}$$

where $0 \leq d \leq 1$, $\omega, \delta > 0$, φ, $\beta < 1$ and $-1 < \gamma < 1$. FIAPARCH model is reduced to FIGARCH model if $\gamma = 0$ and $\delta = 2$.

FIAPARCH model for daily return is given as follows:

$$r_t = \mu + \varepsilon_t\ ,\ \ \varepsilon_t|\Omega_{t,n-1} \sim T_v(0, h_t)$$

$$h_t = \omega(1-\beta)^{-1} + [1-(1-\beta L)^{-1}(1-\varphi L)(1-L)^d](|\varepsilon_t| - \gamma\varepsilon_t\ )^\delta \tag{17}$$

Davidson (2004) proposed the hyperbolic GARCH (HYGARCH) model, which nests both the GARCH and FIGARCH models as special cases. The HYGARCH model is covariance stationarity and it obeys hyperbolically decaying impulse response coefficients just like the FIGARCH model. The HYGARCH (1, d, 1) model for intraday returns is determined as follows:

$$\tilde{r}_{t,n} = \mu + \varepsilon_{t,n}\ ,\ \ \varepsilon_{t,n}|\Omega_{t,n-1} \sim T_v(0, h_{t,n})$$

$$h_{t,n} = \omega + \{1 - [1-\beta L]^{-1}\varphi L\{1 + k[(1-L)^d - 1]\}\}\varepsilon_{t,n}^2 \tag{18}$$

where $0 \leq d \leq 1, \omega > 0, k \geq 0, \varphi, \beta < 1$ and L is the lag operator.

The HYGARCH (1, d, 1) model for daily returns is defined as follows:

$$r_t = \mu + \varepsilon_t\ ,\ \ \varepsilon_t|\Omega_{t,n-1} \sim T_v(0, h_t)$$

$$h_t = \omega + \{1 - [1-\beta L]^{-1}\varphi L\{1 + k[(1-L)^d - 1]\}\}\varepsilon_t^2 \tag{19}$$

In summary, we employ 7 GARCH family models to describe and forecast the volatility of the WTI crude oil futures by using intraday 15 min return series and daily return series respectively.

3.3. Forecast and SPA test

The crude oil observations are from 25th March 2009 to 25th March 2013 and we divide the whole sample into two subgroups: the in-sample data for volatility modelling covering from 25th March 2009, to 1nd November 2012, and the out-of-sample data for model evaluation is from 2nd November 2012, to 25th March 2013, covering 100 trading days and containing 8595 observations. We use a rolling window method and produce one-step ahead volatility forecasts for intraday and daily model therefore, each step is one-day for daily data while it is 15 min each step for our high frequency data. This procedure is repeated 100 times in order to produce 100 daily volatility forecasts for daily out-of-sample evaluation and 8595 times to yield intraday volatility forecasts for intraday out-of-sample evaluation. The rolling window estimation requires adding one new observation and dropping the most distant one therefore the sample size employed in estimating the models remains fixed and the forecasts do not overlap.

Actual volatility (variance) is assessed using the squared returns and denoted as σ_t^2. The volatility forecast obtained using a GARCH-class model is indicated by $\hat{\sigma}_t^2$. Various forecasting criteria or loss functions can be considered to assess the predictive accuracy of a volatility model. However it is not obvious which loss function is more appropriate for the evaluation of volatility models. Hence, rather than making a single choice we use the following 9 different loss functions as forecasting criteria:

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(\sigma_t^2 - \hat{\sigma}_t^2)^2 \qquad (20)$$

$$\text{MedSE} = Median(\sigma_t^2 - \hat{\sigma}_t^2)^2 \qquad (21)$$

$$ME = \frac{1}{n}\sum_{t=1}^{n}(\sigma_t^2 - \hat{\sigma}_t^2) \qquad (22)$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|\sigma_t^2 - \hat{\sigma}_t^2| \qquad (23)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(\sigma_t^2 - \hat{\sigma}_t^2)^2} \qquad (24)$$

$$HMAE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{\sigma_t^2 - \hat{\sigma}_t^2}{\sigma_t^2}\right| \qquad (25)$$

$$AMAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{\sigma_t^2 - \hat{\sigma}_t^2}{(\sigma_t^2 + \hat{\sigma}_t^2)/2}\right| \qquad (26)$$

$$U = \frac{\sqrt{\frac{1}{n}\sum_{t=1}^{n}(\sigma_t^2 - \hat{\sigma}_t^2)^2}}{\sqrt{\frac{1}{n}\sum_{t=1}^{n}(\sigma_t^2)} + \sqrt{\frac{1}{n}\sum_{t=1}^{n}(\hat{\sigma}_t^2)}} \qquad (27)$$

$$logloss = -\frac{1}{n}\sum_{t=1}^{n}(\sigma_t^2 \log(\hat{\sigma}_t^2) + (1 - \sigma_t^2)\log(1 - \hat{\sigma}_t^2)) \qquad (28)$$

where n is the number of forecasting data. In the forecasting comparison part, the subscript indicating the observation number within a day is omitted because we do not make cross comparison between same models in different time frequencies. The 9 loss functions are Mean Squared Error (MSE), Median Squared Error (MedSE), Mean Error (ME), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Heteroskedastic Mean Squared Error (HMSE), Mean Absolute Percentage Error (MAPE), Adjusted Mean Absolute Percentage Error (AMAPE), Theil Inequality Coefficient (THEIL) and Logarithmic Loss Function (LL) respectively. Additional discussion of these criteria can be found in Brooks, Burke, and Persand (1997) for more details about these measures.

When we use a particular loss function to compare two models, we cannot clearly conclude that the forecasting performance of model A is superior to that of model B. Such a conclusion cannot be made on the basis of just one loss function and just one sample. Recent research has focused on a testing framework for determining whether a particular model is outperformed by another one (e.g., Diebold and Mariano, 1995; White, 2000). Hansen (2005) extends the White framework known as the superior predictive ability (SPA) test. The SPA test has been shown to have good power properties and to be more robust than previous approaches.

The SPA test can be used to compare the performance of two or more forecasting models at a time. Forecasts are evaluated using a pre-specified loss function and the "best" forecasting model is the one that produces the smallest expected loss. In a SPA test, the loss function relative to the benchmark model is defined as $X_{t,l}^{(0,i)} = L_{t,l}^{(0)} - L_{t,l}^{(i)}$, where $L_{t,l}^{(0)}$ is the value of the loss function l at time t for a benchmark model M_0 and $L_{t,l}^{(i)}$ is the value of the loss function l at time t for another competitive model M_i for $i = 1, \dots, K$. The SPA test is used to compare the forecasting performance of a benchmark model against its K competitors. The null hypothesis that the benchmark or base model is not outperformed by any of the other competitive models is expressed as $H_0\colon \max_{i=1,\dots,K} E(X_{t,l}^{(0,i)}) \leq 0$. It is tested with the statistic $T_l^{SPA} = \max_{i=1,\dots,K}(\sqrt{n}\bar{X}_{i,l}/$

$\sqrt{\lim_{n\to\infty} var(\sqrt{n}\bar{X}_{i,l})}$), where n is is the number of forecast data points and $\bar{X}_{i,l} = \frac{1}{n}\sum_{t=1}^{n} X_{t,l}^{(0,i)}$. $\lim_{n\to\infty} var(\sqrt{n}\bar{X}_{i,l})$ and the p-value of the T_l^{SPA} are obtained by using the stationary bootstrap procedure discussed by Politis and Romano (1994). Hansen (2005) summarises that the p-value of a SPA test indicates the relative performance of a base model M_0 in comparison with alternative models M_i . A high p-value indicates that we are not able to reject the null hypothesis that "the base model is not outperformed".

4. Estimation results for different volatility models

Table 2 and table 3 present the in-sample estimation results for the alternative volatility models presented in model framework section for two time frequencies. For each table, the upper part shows the values and standard errors of each parameter and the lower part presents the diagnostic results of the standardised residuals.

After reading table 1, I conclude that βs in all the models are significant at 1% level. For IGARCH and EGARCH model, βs are much close to 1 (larger than 0.9) and βs in GARCH model and GJR model are also close to 1 (larger than 0.8). The large βs suggest the high persistence of volatility in the intraday data. The asymmetric leverage coefficients γs for intraday regression are significant in GJR, EGARCH and FIAPARCH models, indicating the leverage effect exists. The power coefficient δ in FIAPARCH model is close to 2 and it is significantly different from zero and I cannot reject the hypothesis that δ is 2 at 5% significance level while I reject the hypothesis that δ is 1 at 1% level. That δ is close to 2 indicates that conditional variance is more fit for the intraday data than conditional standard deviation. The fractional difference parameter ds in FIGARCH, FIAPARCH and HYGARCH are all significant and the value is from 0.45 to 0.4725, suggesting a large degree of long-memory volatility in intraday returns. The value of degree of freedom of the student's T distribution ranges from 5.99 to 6.09 and are all significant in all GARCH family models, suggesting the kurtosis of the returns.

The lower part of Table 2 provides the diagnostic tests of the corresponding GARCH family models for 15 min intraday data. The log likelihood function values and AIC values are close to each other for alternative GARCH family models except EGARCH model. The log likelihood function value and the value of AIC of EGARCH are much lower than those of other GARCH family models. The Ljung-Box Q tests and ARCH tests results are quite mixed for intraday data.

The Ljung-Box Q-statistics of lag order 20 of the standardized residuals are all significant at 1% level in each model except IGARCH, rejecting the null hypothesis that there is no serial correlation in the standardized residuals; while the Ljung-Box Q-statistics of lag order 20 of the squared standardized residuals is not significant for FIGARCH model only. ARCH test results show that the standardized residuals still have heteroskedasticity feature except FIGARCH model and HYGARCH model.

The daily return regression output and diagnostic tests are given in Table 2. Similar to the output of GARCH, IGARCH, GJR and EGARCH model output for intraday returns, βs in these models are very close to 1 and are significant at 1% level, indicating the volatility of daily data is persistent in WTI market. The asymmetric leverage coefficients γs in GJR and EGARCH model is significant, suggesting the negative shocks will have a larger impact on the volatility than positive shocks. While γ in FIAPARCH is not significant. This result is consistent with Cheong (2009) and Wei et al. (2010). The value of the power coefficient δ in FIAPARCH model employing daily data is 1.997, which is very close to 2 and I do not reject the hypothesis that δ is 2 at the 5 % level. This result is similar to the FIAPARCH output of the intraday return, which present that conditional variance is more fit to the crude oil return than conditional standard deviation. The fractional difference parameter ds in FIGARCH and FIAPARCH are significant and the values are 0.258 and 0.184 respectively. The results indicate the volatility of the crude oil contains long-memory character. All the parameters of HYGARCH model are not significant except the degree of freedom of the student’s T distribution thus the performance of HYGARCH is not fit for crude oil returns.

The lower part of Table 3 provides the diagnostic tests of the corresponding GARCH family models for daily data. The log (L) and AIC values are much close to each other under the alternative GARCH family models. For GARCH family model employing daily data, The Ljung-Box Q-statistics of lag order 20 of the squared standardized residuals and ARCH tests indicate FIGARCH, FIAPARCH and HYGARCH outperform the other 4 models while the Ljung-Box Q-statistics of lag order 20 of the standardized residuals tell an opposite story. All the Q-statistics of the standardized residuals and the ARCH statistics except the ARCH statistics under EGARCH are not significant at 5% level, which indicates that the residuals have no autocorrelation and ARCH effect.

Swanson et al. (2006) argue that we are supposed to choose a preferred model based on its forecasting performance rather than their in-sample fit. Therefore I carry out out-of-sample forecasting performance to evaluate alternative GARCH family models.

Table 2. Estimation results of different volatility models for intraday returns

	GARCH	IGARCH	GJR	EGARCH	FIGARCH	FIAPARCH	HYGARCH
ω x 10^6	0.01221*** (0.0028)	0.02762 (0.0016)	0.0122*** (0.0028)	0.0000 (0.0166)	0.0468*** (0.0086)	0.0128*** (0.0025)	0.0172 (0.0147)
A	0.1001*** (0.0010)	0.078083*** (0.0017381)	0.100111*** (0.0010350)	0.271113*** (0.0068354)			
B	0.800025*** (0.0021910)	0.921917*** (0.000286)	0.800025*** (0.0021917)	0.955319*** (0.00024038)	0.452940*** (0.013664)	0.400140*** (0.015277)	0.448520*** (0.022339)
d.o.f	6.011470*** (0.015824)	6.026217*** (0.14406)	6.011470*** (0.015394)	5.999317*** (0.11790)	6.089591*** (0.060163)	6.012063*** (0.024139)	5.997117*** (0.15620)
γ			0.010122*** (0.0030080)	-0.078280*** (0.0029402) 0.270658*** (0.00024756)		0.010863*** (0.0019776)	
Log Alpha (HY)							0.016572 (0.0090933)
δ						2.000181*** (0.0053816)	
φ					0.130278*** (0.0092180)	0.099942*** (0.011534)	0.126694 (0.015074)
d					0.472533*** (0.0071312)	0.450144*** (0.0053950)	0.464303*** (0.014638)
Diagnostic							
Log(L)	335108.544	401539.058	335278.276	114588.408	328694.918	352379.885	393581.536
AIC	-8.260191	-9.897705	-8.264350	-2.824394	-9.862134	-8.685849	-9.701481
Q(20)	494.876*** [0.0000000]	16.2711 [0.6996701]	537.457*** [0.0000000]	55.5864*** [0.0000335]	67.4981*** [0.0000005]	491.552*** [0.0000000]	215.758*** [0.0000000]
$Q^2(20)$	277.088*** [0.0000000]	151.098*** [0.0000000]	282.397*** [0.0000000]	91.5607*** [0.0000000]	6.35074 [0.9945546]	217.559*** [0.0000000]	12.5546 [0.8173234]
ARCH(20)	17.410*** [0.0000]	6.8890*** [0.0000]	17.805*** [0.0000]	11.552*** [0.0000]	0.31674 [0.9984]	12.386*** [0.0000]	0.63793 [0.8875]

Notes: the numbers in parentheses are standard errors of the estimations. Log(L) is the logarithm maximum likelihood function value. AIC is the average Akaike information criterion. Q(20) and $Q^2(20)$ are the Ljung–Box Q-statistic of lag order 20 computed on the standardized residuals and squared standardized residuals, respectively. ARCH(20) is the non-heteroskedasticity statistic of order 20. P-values of the statistics are reported in square brackets. ** and *** denote significance at the 5% and 1% levels, respectively.

Table 3. Estimation results of different volatility models for daily returns

	GARCH	IGARCH	GJR	EGARCH	FIGARCH	FIAPARCH	HYGARCH
ω x 10^4	0.135486 (0.075531)	0.034278 (0.039289)	0.102000 (0.055122)	0.000544 (12.998)	0.535345 (0.46157)	0.485799 (1.9011)	0.055273 (0.93261)
α	0.065141** (0.026221)	0.071372** (0.043119)	0.008735 (0.015840)	0.020320 (0.15456)			
β	0.901656*** (0.037753)	0.928628*** (0.008606)	0.919959*** (0.028861)	0.999308*** (0.0012490)	0.192791 (0.52391)	-0.161725 (0.54603)	0.148453 (0.69051)
d.o.f	8.406655*** (2.0608)	7.003380*** (1.6289)	9.408019*** (2.5921)	6.759639*** (1.8483)	8.372224*** (2.0506)	9.539912*** (2.5541)	8.206247*** (2.0179)
γ			0.089790*** (0.033702)	-0.068631 (0.036998) 0.4110*** (0.071263)		0.454404 (0.34889)	
HY							0.360136 (0.71845)
δ						1.997314*** (0.61248)	
φ					0.000000 (0.56190)	-0.255096 (0.52410)	0.000000 (0.79986)
d					0.258486*** (0.062712)	0.183622** (0.074691)	0.151379 (0.14814)
Diagnostic							
Log(L)	2350.947	2347.775	2356.222	2307.596	2352.048	2357.519	2352.235
AIC	-5.028825	-5.024169	-5.037989	-4.931610	-5.029042	-5.036483	-5.0273
Q(20)	27.9886 [0.1096686]	25.7596 [0.1738983]	28.2193 [0.1043095]	22.1826 [0.3306860]	28.5784 [0.0963982]	29.4656 [0.0789886]	28.3319 [0.1017727]
$Q^2(20)$	17.7095 [0.4749414]	19.9536 [0.3354371]	20.0119 [0.3321486]	33.9349** [0.0128306]	14.2030 [0.7157638]	17.1048 [0.5159099]	14.5209 [0.6945593]
ARCH(20)	1.0760 [0.3695]	1.1882 [0.2562]	1.1667 [0.2760]	1.7437** [0.0226]	0.81558 [0.6962]	0.94017 [0.5352]	0.83414 [0.6727]

Notes: the numbers in parentheses are standard errors of the estimations. Log(L) is the logarithm maximum likelihood function value. AIC is the average Akaike information criterion. Q(20) and $Q^2(20)$ are the Ljung–Box Q-statistic of lag order 20 computed on the standardized residuals and squared standardized residuals, respectively. ARCH(20) is the non-heteroskedasticity statistic of order 20. P-values of the statistics are reported in square brackets. ** and *** denote significance at the 5% and 1% levels, respectively.

5. Forecast comparison

Table 4 produces the one-step out-of-sample volatility forecasts valuation of alternative GARCH family models by employing intraday data. The out-of-sample period is from 2nd November 2012 to 25th March 2013, covering 100 trading days and containing 8595 observations. There are 9 different forecast evaluations in table 1 and the performance of alternative models is different under different valuation criteria. FIGARCH performs best when it comes to mean squared error (MSE), mean error (ME) or root mean squared error (RMSE) while GARCH model outperforms other models if we stick to median squared error (MedSE), mean absolute error (MAE) or mean absolute percentage error (MAPE). FIAPARCH is the best under the criterion of adjusted mean absolute percentage Error (AMAPE). A look at Theil inequality coefficient (TIC) tells that Fractional GARCH models such as FIGARCH, FIAPARCH and HYGARCH outperform GARCH, IGARCH, GJR and EGARCH models and GARCH, IGARCH, GJR models are almost naïve guess considering their TIC values are close to 1. The TIC value of EGARCH is 1, which suggests that the forecast of EGARCH model is just naïve guesswork. To sum up, GARCH model performs well in terms of two criteria: mean absolute error and mean absolute percentage error; FIGARCH also performs well according to three criteria: mean squared error, mean error and root mean squared error. GJR performs the best under median squared error and logarithmic loss function, FIAPACH and HYGARCH perform well in adjusted mean absolute percentage error and Theil inequality coefficient respectively. The performance of EGARCH model is the worst among the models being compared.

Table 4. Forecast valuation of one-step out-of-sample volatility forecasts of alternative GARCH models of intraday data

	GARCH	IGARCH	GJR	EGARCH	FIGARCH	FIAPARCH	HYGARCH
MSE	3.256e-011 (5)	1.621e-008 (6)	3.254e-011 (4)	0.9929 (7)	2.951e-011 (1)	2.966e-011 (2)	3.02e-011 (3)
MedSE	2.438e-014 (2)	1.22e-008 (6)	2.241e-014 (1)	1 (7)	2.588e-012 (4)	3.132e-013 (3)	4.529e-012 (5)
ME	1.395e-006 (5)	-0.0001099 (6)	1.388e-006 (4)	-0.9946 (7)	-2.33e-007 (2)	4.383e-007 (3)	-8.104e-007 (1)
MAE	1.462e-006 (2)	0.0001101 (6)	1.46e-006 (1)	0.9946 (7)	2.063e-006 (4)	1.698e-006 (3)	2.463e-006 (5)
RMSE	5.706e-006 (5)	0.0001273 (6)	5.704e-006 (4)	0.9964 (7)	5.432e-006 (1)	5.446e-006 (2)	5.495e-006 (3)
MAPE	243.5 (1)	2.166e+005 (6)	255.8 (2)	1.846e+009 (7)	3231 (4)	1739 (3)	4331 (5)
AMAPE	0.6258 (3)	0.9519 (6)	0.6242 (2)	1 (7)	0.6685 (4)	0.6191 (1)	0.6962 (5)
TIC	0.9712 (6)	0.9497 (4)	0.9699 (5)	1 (7)	0.7371 (2)	0.7687 (3)	0.6913 (1)
LL	8.35 (2)	48.05 (6)	8.318 (1)	251.6 (7)	13.23 (4)	10.85 (3)	15.25 (5)

Notes: Numbers in brackets indicate the performance ranking of alternative models under each loss function.

Table 5 presents the one-step out-of-sample volatility forecasts valuation of alternative GARCH family models by employing daily data. Contrary to the findings of alternative GARCH models employing intraday data, EGARCH model of daily data outperforms other models in terms of the most criteria. The Theil inequality coefficient of FIAPARCH model is less than that of EGARCH, which is the only loss function indicating daily EGARCH is outperformed by any other daily GARCH type model.

The discussion above provide the performance of different models according to different criteria. To check the reliability and robustness of the forecasts, we refer to SPA test for more information.

Table 5. Forecast valuation of one-step out-of-sample volatility forecasts of alternative GARCH models of daily data

	GARCH	IGARCH	GJR	EGARCH	FIGARCH	FIAPARCH	HYGARCH
MSE	1.283e-007 (5)	1.687e-007 (7)	1.193e-007 (3)	7.732e-008 (1)	1.541e-007 (6)	1.038e-007 (2)	1.264e-007 (4)
MedSE	1.005e-007 (5)	1.344e-007 (7)	8.977e-008 (3)	3.08e-008 (1)	1.311e-007 (6)	7.374e-008 (2)	9.773e-008 (4)
ME	-0.0002361 (5)	-0.0002889 (7)	-0.0002258 (3)	-9.15e-005 (1)	-0.0002782 (6)	-0.0001867 (2)	-0.0002305 (4)
MAE	0.0003113 (5)	0.0003627 (7)	0.0002996 (3)	0.0001929 (1)	0.0003502 (6)	0.000269 (2)	0.0003071 (4)
RMSE	0.0003582 (5)	0.0004108 (7)	0.0003455 (3)	0.0002781 (1)	0.0003926 (6)	0.0003223 (2)	0.0003555 (4)
MAPE	292.2 (5)	297.5 (6)	286.5 (3)	163 (1)	327.3 (7)	262.2 (2)	287.4 (4)
AMAPE	0.6887 (5)	0.7088 (7)	0.6834 (3)	0.6029 (1)	0.7075 (6)	0.6671 (2)	0.6865 (4)
TIC	0.553 (4)	0.5787 (7)	0.5432 (2)	0.5518 (3)	0.5681 (6)	0.54 (1)	0.5535 (5)
LL	10.51 (5)	11.14 (7)	10.32 (3)	8.258 (1)	11.07 (6)	9.803 (2)	10.44 (4)

Notes: Numbers in brackets indicate the performance ranking of alternative models under each loss function.

Table 6. SPA test results evaluated by the MAE and MSE for intraday GARCH model

	MAE	MSE	MAE	MSE
	Models		t-statistics	
Benchmark	Intraday GARCH	Intraday GARCH	-	-
Most Significant	GJR	GJR	5.87510	7.91513
Best model	GJR	GJR	5.87510	7.91513
Model_25%	FIGARCH	FIGARCH	-3.64346	5.70474
Median_50%	HYGARCH	HYGARCH	-5.64952	5.13410
Model_75%	FIAPARCH	FIAPARCH	-11.38561	2.82256
Worst model	IGARCH	IGARCH	-20.01088	-9.61660
SPA test p-value	MAE	MSE		
	0.00000	0.00270		

Notes: Table 6 shows the SPA test results for different models. The benchmark model selected is the intraday GARCH model. The null hypothesis of the test is that the benchmark model is not inferior to the other candidate models. The test chooses the most significant model, the best model, models with performances of 75%, 50% and 25% relative to the benchmark model, and the worst model. P-values are reported in the last row.

Table 7. SPA test results evaluated by the MAE and MSE for intraday FIAPARCH model

	MAE	MSE	MAE	MSE
	Models		t-statistics	
Benchmark	Intraday FIAPARCH	Intraday FIAPARCH	-	-
Most Significant	FIGARCH	HYGARCH	15.46191	0.60762
Best model	FIGARCH	HYGARCH	15.46191	0.60762
Model_25%	HYGARCH	FIGARCH	14.90305	-0.14373
Median_50%	GJR	GJR	11.42375	-2.81174
Model_75%	GARCH	GARCH	11.38561	-2.82256
Worst model	IGARCH	IGARCH	-17.79533	-11.18634
SPA test p-value	MAE	MSE		
	0.0000	0.32920		

Notes: Table 7 shows the SPA test results for different models. The benchmark model selected is the intraday FIAPARCH model. The null hypothesis of the test is that the benchmark model is not inferior to the other candidate models. The test chooses the most significant model, the best model, models with performances of 75%, 50% and 25% relative to the benchmark model, and the worst model. P-values are reported in the last row.

Table 8. SPA test results evaluated by the MAE and MSE for daily FIGARCH model

	MAE	MSE	MAE	MSE
	Models		t-statistics	
Benchmark	Daily FIGARCH	Daily FIGARCH	-	-
Most Significant	GJR	HYGARCH	3.69650	2.83204
Best model	GJR	HYGARCH	3.69650	2.83204
Model_25%	GARCH	FIAPARCH	3.64346	0.14373
Median_50%	HYGARCH	GJR	-13.02806	-5.69430
Model_75%	FIAPARCH	GARCH	-15.46191	-5.70474
Worst model	IGARCH	IGARCH	-20.68853	-10.68270
SPA test p-value	MAE	MSE		
	0.00000	0.00000		

Notes: Table 8 shows the SPA test results for different models. The benchmark model selected is the daily FIGARCH model. The null hypothesis of the test is that the benchmark model is not inferior to the other candidate models. The test chooses the most significant model, the best model, models with performances of 75%, 50% and 25% relative to the benchmark model, and the worst model. P-values are reported in the last row.

Table 9. SPA test results evaluated by the MAE and MSE for daily EGARCH model

	MAE	MSE	MAE	MSE
	Models		t-statistics	
Benchmark	Daily EGARCH	Daily EGARCH	-	-
Most Significant	FIAPARCH	FIAPARCH	-10.10507	-4.69190
Best model	FIAPARCH	FIAPARCH	-10.10507	-4.69190
Model_25%	GJR	GJR	-11.52714	-5.02383
Median_50%	IGARCH	HYGARCH	-11.57345	-6.62125
Model_75%	HYGARCH	FIGARCH	-12.10224	-7.26162
Worst model	FIGARCH	IGARCH	-13.39979	-7.89615
SPA test p-value	MAE	MSE		
	0.49060	0.46160		

Notes: Table 9 shows the SPA test results for different models. The benchmark model selected is the daily EGARCH model. The null hypothesis of the test is that the benchmark model is not inferior to the other candidate models. The test chooses the most significant model, the best model, models with performances of 75%, 50% and 25% relative to the benchmark model, and the worst model. P-values are reported in the last row.

Selected SPA test results are illustrated in tables 6 to 9. Table 6 and Table 7 present the SPA results for intraday GARCH and FIAPARCH models respectively and Table 8 and Table 9 present the SPA results for the daily FIGARCH and EGARCH models respectively. We carry out SPA test for different benchmark models and compare the volatility models under two pre-determined loss function, Mean Squared Error (MSE) and Mean Absolute Error (MAE). I rank the models according to their performance against that of the benchmark model from the best to the worst by reading the t-statistics. P-value of SPA test is based on 10000 bootstrap samples in the empirical test. A high P-value suggests it is less likely to reject the null hypothesis that the base model is not outperformed by all of the other models.

P-values From Table 6 and Table 7 are almost close to zero under either MSE or MAE, which present none of the volatility models outperforms the other models. The only exception is intraday FIAPARCH model. The P-value of SPA test under the loss function MSE is 0.3292, suggesting FIAPARCH is not outperformed by the other models. Comparatively, the SPA test results of daily models indicate EGARCH model outperforms the other volatility model. The P values from the SPA test results for other models as benchmark are close to zero, which is similar to those in the results in Table 6 and Table 8. Those results are not presented in this paper to avoid repeat results. Considering EGARCH model outperforms other models under different loss functions, we conclude daily EGARCH model is superior to other models. However, for volatility models employing intraday data, none of the GARCH-type models are superior to the others except FIAPARCH.

6. Conclusion

In this article, we employ a greater number of GARCH-class models and many loss functions and carry out the superior predictive ability (SPA) test to estimate and compare the forecasting performance on the basis of intraday data and daily data. Several GARCH family models such as GJR, EGARCH, APARCH, FIGARCH, FIAPARCH, and HYGARCH capture long-memory volatility and/or the asymmetry leverage effect in volatility. We find that intraday FIAPARCH model is not outperformed by other models while other complicated GARCH series models do not pass the SPA test so we conclude that none of the GARCH-class models outperforms the others when it comes to intraday data except FIAPARCH model. The FIAPARCH model result is in line with some research papers in the literature such as ABDL (2001), Corsi (2009), Martens and Zein (2004) and Chortareas et al. (2011) which all document that long memory specification in high-frequency data can improve the forecasting power and accuracy significantly however other complicated GARCH models' forecast performance are against the exiting literature. The data we use in this paper covers the post-crisis period (2009-2013) which are more up-to-date than the data covered in the existing literature and this is the root of the difference between this paper and the existing literature since we use the same or similar methodology and models as in the literature. EGARCH model is superior to other model when it comes to daily data and it is different from the finding of Kang et al. (2009) in which FIGARCH performs well. The difference of daily data model between this paper and Kang et al. lies in the different data sample as well since same/similar models are used in the two studies. Our findings provide a solid piece of evidence to the cons part in the discussion that whether the traditional time series models are good to fit intraday data. We find that the traditional volatility model cannot fit the data when we employ intraday data. After de-seasonalising the raw returns of the crude oil futures and putting in GARCH family models, it emerges that the majority of GARCH models cannot produce satisfactory forecast results except FIAPARCH model. Thus, the new efforts should be made to find new models to forecast volatility in a high-frequency framework.

We find that the intraday crude oil returns are consistent with the stylised properties of other financial series such as stock market indices and exchange rates at high frequencies in many respects. This becomes a piece of evidence that these properties are not limit to certain kinds of high-frequency data. It might reflect some general features which all intraday data share.

Agnolucci (2009) proposes the question "whether the comparison of volatility forecasting models is influenced by the criterion used in the exercise." Our findings indicate that the rankings of the performance of volatility models are different when different criteria are applied. The results of our paper suggest that economists and financial practitioners should not arbitrarily choose a volatility forecasting model by referring to the existing research. Which model can be trusted depends on not only the given data sample but also the correspondence of the particular forecasting purpose with the loss function considered.

References:

Andersen, T. G., & Bollerslev, T. (1997). Intraday periodicity and volatility persistence in financial markets. Journal of Empirical Finance, 4, 115–158.

Andersen, T. G., & Bollerslev, T. (1998). Answering the skeptics: yes, standard volatility models do provide accurate forecasts. International Economic Review, 39, 885–905.

Andersen, T.G., Bollerslev, T., Diebold, F. X., & Elbens, H. (2001). The distribution of stock return volatility. Journal of Financial Economics, 61, 43-76.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (1999). (Understanding, optimizing, using and forecasting) realised volatility and correlation. Manuscript, Northwestern University, Duke University and University of Pennsylvania. Published in revised form as "Great realizations". Risk. 105–108.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2001). The distribution of exchange rate volatility. Journal of the American Statistical Association, 96, 42–55.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modelling and forecasting realised volatility. Econometrica, 71, 579–625.

Agnolucci, P. (2009). Volatility in crude oil futures: a comparison of the predictive ability of GARCH and implied volatility models. Energy Economics, 31, 316–321.

Barbosa, P. A. (2002). Investigating high frequency exchange rate from the Brazilian Sisbex market. Working paper. http://ssrn.com/abstract=428581.

Baillie, R.T., Bollerslev, T., & Mikkelsen, H.O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 74, 3–30.

Baillie, R.T., Cecen, A.A, Erkal, C., & Han, Y. (2004). Measuring non-linearity, long memory and self-similarity in high-frequency European exchange rates. International Financial Markets, Institutions and Money, 14, 401–418.

Bandi, F. M., & Russell, J. R. (2005). Separating microstructure noise from volatility. Journal of Financial Economics, 79, 655–692.

Beltratti, A., & Morana, C. (1999). Computing value at risk with high frequency data. Journal of Empirical Finance, 6, 431–455.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31, 307–327.

Brooks, C., Burke,S. P., & Persand, G. (1997): "Linear and Non-Linear (Non-)Forecastability of High-Frequency Exchange Rates," Journal of Forecasting, 16, 125–145.

Celik, S., & Ergin, H. (2014). Volatility forecasting using high frequency data: Evidence from stock markets. Economic Modelling, 36, 176–190.

Chernov, M., Gallant, R., Ghysels, E., & Tauchen, G. (2003). Alternative models for stock price dynamics. Journal of Econometrics, 116, 225–257.

Chortareas, G., Jiang, Y., & Nankervis, J. (2011). Forecasting exchange rate volatility using high-frequency data: Is the euro different? International Journal of Forecasting, 27, 1089-1107.

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. Quantitative Finance, 1, 223–236.

Corsi, F. (2009). A simple approximate long-memory model of realised volatility. Journal of Financial Econometrics, 7(2), 174–196.

Dacorogna, M. M., Gencay, R., Muller, U. A., Olsen, R. B., & Pictec, O. V. (2001). An introduction to high-frequency finance. San Diego: Academic Press.

Davidson, J. (2004). Moment and memory properties of linear conditional heteroscedasticity models, and a new model. Journal of Business & Economic Statistics, 22, 16–29.

Diebold, F.X., & Mariano, R.S. (1995). Comparing predictive accuracy. Journal of Business & Economic Statistics, 13, 253–263.

Engle, R.F., & Bollerslev, T. (1986). Modelling the persistence of conditional variances. Econometric Reviews, 5, 1–50.

Engle,R. & Gallo, G. (2006). A multiple indicators model for volatility using intra-daily data. Journal of Econometrics, 131, 3–27.

Glosten, L.R., Jagannathan, R., & Runkle, D.E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. Journal of Finance, 48, 1779–1801.

Goodhart, C.A.E. (1989). News and foreign exchange market. In Proceedings of the Manchester Statistical Society, 1-79.

Goodhart, C.A.E., & Figliuoli, L. (1992). The geographical location of the foreign exchange market: A test of an "islands" hypothesis. Journal of International and Comparative Economics, 1, 13-27.

Goodhart, C.A.E., Ito, T., & Payne, R.G. (1995). One Day in June 1993: A study of the working of Reuters 2000-2 electronic foreign exchange trading system. The Microstructure of Foreign Exchange Markets. University of Chicago Press, 107-182.

Hansen, P.R. (2005). A test for superior predictive ability. Journal of Business & Economic Statistics, 23, 365–380.

Hansen, P.R., & Lunde, A. (2006). Realised variance and market microstructure noise. Journal of Business and Economic Statistics, 24, 127-218.

Hol, E., & Koopman, S. J. (2002). Stock index volatility forecasting with high frequency data. Tinbergen Institute Discussion Paper No. 2002-068/4.

Hou, A. & Suardi, S. (2012). A nonparametric GARCH model of crude oil price return volatility. Energy Economics, 34(2), 618–626.

Jones, B. (2003). Is ARCH useful in high frequency foreign exchange applications? Research paper No. 24. Applied Finance Centre, Macquarie University.

Kang, S.H., Kang, S.M., & Yoon, S.M. (2009). Forecasting volatility of crude oil markets. Energy Economics, 31, 119–125.

Koopman, S. J., Jungbacker, B., & Hol, E. (2005). Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements. Journal of Empirical Finance, 12(3), 445–475.

Marlik, A. K. (2005). European exchange rate volatility dynamics: an empirical investigation. Journal of Empirical Finance, 12, 187–215.

Martens, M. (2001). Forecasting daily exchange rate volatility using intraday returns. Journal of International Money and Finance, 20(1), 1-23.

Martens, M., Chang, Y. C., & Taylor, S. (2002). A comparison of seasonal adjustment methods when forecasting intraday volatility. Journal of Financial Research, 25(2), 283–299.

Martens, M., & Zein, J. (2004). Predicting financial volatility: high-frequency time-series forecasts vis-`a-vis implied volatility. Journal of Futures Markets, 24, 1005–1028.

Mohammadi, H., & Su, L. (2010). International evidence on crude oil price dynamics: applications of ARIMA–GARCH models. Energy Economics, 32 (5), 1001–1008.

Nelson, D.B. (1991). Conditional heteroskedasticity in asset returns: a new approach. Econometrica, 59, 347–370.

Politis, D.N., & Romano, J.P. (1994). The stationary bootstrap. Journal of the American Statistical Association, 89, 1303–1313.

Pong, S., Shackleton, M., Taylor, S. J., & Xu, X. (2004). Forecasting currency volatility: a comparison of implied volatilities and AR(FI)MA models. Journal of Banking and Finance, 28(9), 2541–2563.

Poon, S. H., & Granger, C. (2003). Forecasting financial market volatility: a review. Journal of Economic Literature, 41(2), 478–539.

Rahman, S., & Ang, K. P. (2002). Intraday return volatility process: evidence from NASDAQ stocks. Review of Quantitative Finance and Accounting, 19, 155–180.

Sadorsky, P. (2006). Modelling and forecasting petroleum futures volatility. Energy Economics, 28, 467–488.

Sevi, B. (2014). Forecasting the volatility of crude oil futures using intraday data. European Journal of Operational Research, 235, 643–659.

Shephard, N., & Sheppard, K. (2010). Realising the future: forecasting with high frequency based volatility (HEAVY) models. Journal of Applied Econometrics, 25, 197-231.

Swanson, N.R., Elliott, G., Ghysels, E., & Gonzalo, J. (2006). Predictive methodology and application in economics and finance: volume in honor of the accomplishments of Clive W.J. Granger. Journal of Econometrics, 135, 1–9.

Taylor, S., & Xu, X. (1997). The incremental volatility information in one million foreign exchange quotations. Journal of Empirical Finance, 4, 317–340.

Tse, Y.K. (1998). The conditional heteroscedasticity of the Yen–Dollar exchange rate. Journal of Applied Econometrics, 13, 49–55.

Wei, Y., Wang, Y., & Huang, D. (2010). Forecasting crude oil market volatility: Further evidence using GARCH-class models. Energy Economics, 32. 1477–1484.

White, H. (2000). A reality check for data snooping. Econometrica, 68, 1097–1126.

Chapter 2. Forecasting Crude Oil Market Volatility by using HAR-RV models: Evidence of Using High Frequency Data

Abstract

The increasing availability of high-frequency data in recent years help scholars to have wider fields to specify and forecast volatility in financial markets. Realised volatility becomes a widely used tool due to the easy access of high-frequency data and the decomposition of jump and continuous components, the decomposition of positive jumps and negative jumps and the discussion of leverage effect may describe volatility better and produce more accurate volatility forecast. Based on Corsi (2009) and Sevi (2014) this study compares the in-sample specification and out-of-sample performance among a series of Heterogeneous Autoregressive (HAR) models from 25th March 2009 to 25th March 2013 by using front-month WTI futures contract. This study finds opposite results to Sevi (2014) and it indicates that the decomposition between jumps and the continuous components and negative and positive realised semi-variances improves the in-sample fit and it improves the out-of-sample forecasting as well.

1. Introduction

It is widely known that volatility specification and forecast is a barometer for the vulnerability of financial markets and the economy. Recent literature indicates that models employing high-frequency data are able to provide more accurate prediction of volatility. Corsi (2009) proposes Heterogeneous Autoregressive model of Realised volatility (HAR-RV) and therefore introduces a way to specify and forecast volatility with the information of high-frequency data or intraday data in spite of the model's simple structure and the absence of long-memory properties of volatility. Following Corsi (2009), Sevi (2014) expands the HAR-RV model by decomposing volatility into continuous and jump components, positive and negative semi-variance and considering leverage effect. His analysis suggests the decomposition of realised variance improves the in-sample fit while the out-of-sample forecast performance is not improved.

This paper extends Sevi's (2014) work and there are several motivations to support the study. In the first place, following Sevi (2014) we specify and forecast volatility of the most traded commodity in the world by using front-month WTI futures contract. Second, we decompose volatility into continuous components following Andersen, Bollerslev and Diebold (2007) and jump components. Third, we compare the forecasting performance among HAR-RV series models and GARCH series models. HAR-RV model is not able to depict the long memory property of volatility due to its simplicity regardless of the decomposition mentioned in the paper while FIGARCH model considers the long memory character by using fractional integration. It is valuable to compare the simple and complicated HAR-RV models with GARCH and FIGARCH models and to rank their forecasting performance. Last, SPA test is utilised to compare the performance of models mentioned above.

Our main findings are as follows: first, the decomposition of continuous components and signed jumps do not help to improve the in-sample fit. The in-sample fit of complicated HAR-

RV models are as good as the genuine HAR-RV model by Corsi (2009). Second, the information of in-sample fit of the decomposition of variance into semi-variance is mixed. Third, when it comes to comparing the prediction ability, the complicated model containing all the decomposed components outperforms simple models or is as good as models without decomposed components at worst. Last, the results of forecasting performance between HAR-RV models and GARCH-type models are quite mixed. It indicates that the forecasting performance of GARCH model and FIGARCH model is better than HAR-RV models when it comes to DM test while the complicated HAR-CSJd outperforms GARCH models in SPA test. The layout of the paper is as follows: Section 2 reviews the literature and Section 3 provides data description. Section 4 presents HAR-RV series models and GARCH models we estimate in the paper and the decomposed components in HAR-RV models. Section 5 and 6 illustrates the in-sample fit and out-of-sample performance of HAR-RV series models respectively while Section 7 mainly compares the forecasting performance between some certain HAR-RV models and GARCH series models. Section 8 concludes the paper.

2. Literature review

2.1. Forecasting the volatility of crude oil

Since the introduction of GARCH model (Bollerslev, 1986), GARCH model and its variants have become the main stream models to specify the volatility of crude oil futures with daily data. Much research has been done to evaluate the forecasting performance among GARCH family models and other alternative models. Literature documents that volatility in oil returns is clustering and persistent and GARCH family models fit the data well (e.g. Adrangi et al., 2001; Agnolucci, 2009; Aloui and Mabrouk, 2010; Cabedo and Moya, 2003; Charles and Darne, 2014; Chkili et al., 2014; Fong and See, 2002; Giot and Laurent, 2003; Hou and Suardi, 2012; Kang et al., 2009; Kang and Yoon, 2013; Mohammadi and Su, 2010; Morana, 2001; Narayan and Narayan, 2007; Sadeghi and Shavvalpour, 2006; Sadorsky, 2006; Wei, et al., 2010).

Sadorsky (2006) uses several different univariate and multivariate statistical models to estimate forecasts of daily volatility in a series of petroleum futures price returns. He finds the TGARCH model fits well for heating oil and natural gas volatility and the GARCH model fits well for crude oil and unleaded gasoline volatility. Models like state space model, vector autoregressive model and bivariate GARCH do not outperform the single equation GARCH model. Most models outperform a random walk.

Mohammadi and Su (2010) investigate the forecasting performance of four GARCH family models — GARCH, EGARCH and APARCH and FIGARCH over January 2009 to October 2009 by employing weekly crude oil spot prices in eleven international markets. They find that the APARCH model outperforms the other models and shocks to conditional volatility dissipate at an exponential rate, which is consistent with the covariance-stationary GARCH models rather than at the slow hyperbolic rate implied by the FIGARCH alternative. This evidence is in line with Sadorsky and McKenzie (2008) but it is contrast to Kang et al (2009). Kang et al. e investigates the a series of volatility models for three crude oil markets and compare the ability to forecast and identify volatility stylized facts, in particular volatility persistence and long memory. They find that The CGARCH and FIGARCH models fit the data better than the GARCH and IGARCH models when it comes to volatility persistence fitting and CGARCH and FIGARCH models also provide superior performance in out-of-sample volatility forecasts considering DM test results. This kind of contradiction may be due to different data frequency (weekly data vs. daily data) and different number of markets.

Wei et al. (2010), Kang and Yoon (2013) extend the work of Kang et al (2009) respectively. Wei et al. employ more GARCH variants models to fit the data and compare their out-of-sample performance by using SPA test proposed by Hansen (2005). They find that no model can outperform all of the other models for either the Brent or the WTI market considering different loss functions. But APARCH GARCH or FIGARCH model which capture long-memory and/or asymmetric volatility, exhibit greater forecasting accuracy than standard GARCH model, especially in volatility forecasting over longer time horizons. Kang and Yoon (2013) focus on the long-memory properties of the same financial assets studied by Sadorsky (2006) by using a batch of long-memory models. They discover that their volatility models fit the daily data well but none of them is outperforming the others based on Diebold-Mariano test. Fong and See (2002) show the regime shift model outperforms GARCH model in-sample but the out-of-sample performance is inconclusive.

2.2. Forecasting volatility by using realised volatility

The existence of intraday/ high-frequency data has drawn researchers' attention and high frequency finance has become a fast-growing field in the past few years. Volatility forecasting is one of the hot-discussed topics in high frequency finance due to the availability of high frequency data.

Many researchers model and forecast volatility with intraday data (Andersen and Bollerslev, 1997; Andersen, Bollerslev, Diebold and Labys, 2001; Koopman et al., 2005; Bollerslev, et al, 2009; Chortareas et al., 2011; Sevi, 2014). A large branch of the literature is comparing volatility forecast by using high frequency data with that by daily data and some study also consider the option implied volatility (Martens, 2011; Koopman, 2002; Martens and Zen, 2004; Pong et al., 2004; Chortareas et al., 2011). Pong et al. (2004) compare the exchange rate volatility forecasts among an ARMA model, an ARFIMA model, a daily GARCH model and option implied volatility model. They find that the most accurate volatility forecasts are the model using high frequency returns. Martens and Zein (2004) document that high frequency data improve the measurement accuracy and the forecasting performance. High frequency data contains more information than daily data and it reflects more reality.

Some scholars use intraday futures commodity data to study the property of commodity markets. Sevi (2014) studies the crude oil market with HAR and its variants of realised volatility and compare their performance in light with Diebold-Mariano test. Tseng et al. (2009) fit the HAR-CJ model using the realised-range proxy as an independent variable to replace realised variance and find that realised range-based bi-power variation (RBV), which is immune to jumps, is a better independent variable for future volatility prediction. Similar to the findings for financial markets, they also find that the jump components of realised-range variance have little predictive power for oil futures contracts.

3. Volatility estimation, jump specification and volatility modelling

3.1. Volatility estimation by using intraday data

We focus on realised volatility computed by employing intraday returns. The standard choice for estimating realised volatility introduced by Andersen et al. (2001) is as follows:

$$RV_{t,M} = \sum_{i=1}^{M} r_{t,i}^2 \tag{1}$$

It shows that for a given day t, the realized variance is computed as the sum of squared intraday returns $r_{t,i}$ at a given sampling frequency $1/M$. The sampling frequency is a key parameter to realised volatility. Using data with the highest possible frequency theoretically optimises the accuracy of the daily volatility estimation but a generally accepted practice is to consider intervals between 5 and 30 minutes (e.g., ABDL, 2003; Hol & Koopman, 2002; Martens, 2001). Therefore the sampling frequency in this paper is set as 15 minutes following the general principle.

Theoretically, realised volatility has the following asymptotic character if the frequency goes to infinity:

$$RV_{t,M} \to \int_{t-1}^{t} \sigma_s^2 ds + \sum_{i=1}^{J(t)} K^2(t_i) \quad in\ probability \tag{2}$$

This equation illustrates that volatility tends to the sum of a continuous and a jump component. Ideally, we would like to get the best estimate of the integrated volatility and leave out the discontinuous component, but the RV unavoidably aggregates both types of risk in the presence of jumps. Barndorff-Nielsen and Shephard (2004) have shown how to disentangle the continuous component and proposed a broader class of realized measures based on bipower variation which allows to estimate $\int_{t-1}^{t} \sigma_s^2 ds$ robustly to jumps.

$$BPV_{t,M} = \xi_1 \sum_{i=1}^{M-1} |r_{t,i}||r_{t,i+1}| \tag{3}$$

where $\xi_1 = \sqrt{2/\pi} \approx 0.79788$. The BPV is a consistent estimator of integrated volatility, and it decomposes the realized variance into its diffusive and non-diffusive parts.

Although very popular, the BPV_t is affected by small sample (upward) bias in the presence of discontinuities. If a given return contains a jump, then the jump effect will not vanish when multiplied by the preceding and following absolute return, and the estimator will not converge to the integrated volatility. To avoid this drawback, Andersen, Dobrev, and Schaumburg (2012) propose the median realized variance (MedRV) which has better properties in realistic settings. The MedRV estimator is given as follows:

$$MedRV_{t,M} = \frac{\pi}{6-4\sqrt{3}+\pi}\left(\frac{M}{M-2}\right)\sum_{i=2}^{M-1} med(|r_{t,i-1}|, |r_{t,i}|, |r_{t,i+1}|)^2 \quad (4)$$

This estimator is more robust to the occurrence of zero-returns. We decide to consider the MedRV as an alternative for our analysis and provide a comparison between the BPV and the MedRV estimators. Both BPV_t and $MedRV_t$ feasibly estimate the $\int_{t-1}^{t} \sigma_s^2 ds$ in the presence of jumps. However, the evidence of the impact of discontinuities on volatility forecasts is contrasting. Andersen et al. (2007) find that they do not contribute to future volatility while Corsi et al. (2010) document an expected positive influence and attributes the puzzle to the small sample bias of BPV_t. One of the explanations to the puzzle is that positive jumps are averaged out by negative ones, therefore Barndorff-Nielsen et al. (2010) further decompose the realised volatility in two complementary components to capture and separate the sign effect of returns

$$RSV_{t,M}^{-} = \sum_{i=1}^{M} r_{t,i}^2 \times I(r_{t,i} < 0) \quad (5)$$

$$RSV_{t,M}^{+} = \sum_{i=1}^{M} r_{t,i}^2 \times I(r_{t,i} > 0) \quad (6)$$

Patton and Sheppard (2015) define signed jumps as the difference between positive and negative realized semi-variances based on the two estimators:

$$\Delta J_{t,M} = RSV_{t,M}^{+} - RSV_{t,M}^{-} \quad (7)$$

Apart from semi-variance decomposition, we need to consider the jump components of the volatility $\sum_{i=1}^{J(t)} K^2(t_i)$ if they are statistically significant. We follow Huang and Tauchen (2005) to detect the volatility jump diffusion. The test statistic at day t is as follows:

$$ZJ_{BPV}(t,M) = \frac{\sqrt{M}(RV_{t,M} - BPV_{t,M})RV_{t,M}^{-1}}{((\xi_1^{-4} + 2\xi_1^{-2} - 5)\max\{1, TQ_{t,M}RV_{t,M}^{-2}\})^2} \quad (8)$$

where $TQ_{t,M}$ is the realized tri-power quarticity.

$TQ_{t,M} = M\xi_{4/3}^{-3}\sum_{i=1}^{M-1}|r_{t,i}|^{4/3}|r_{t,i+1}|^{4/3}|r_{t,i+2}|^{4/3}$ and the estimator converges to integrated quarticity in probability. ZJ_{BPV} statistic follows the standard normal distribution.

A similar test applies to MedRV estimator as well. The MedRV jump test statistic is given as follows:

$$ZJ_{MedRV}(t,M) = \frac{\sqrt{M}(RV_{t,M} - MedRV_{t,M})RV_{t,M}^{-1}}{(0.96\max\{1, MedRQ_{t,M}MedRV_{t,M}^{-2}\})^{1/2}} \quad (9)$$

where $MedRQ_{t,M}$ is an estimator of integrated quarticity given as follows:

$$MedRQ_{t,M} = \frac{3\pi}{72-52\sqrt{3}+9\pi}\left(\frac{M}{M-2}\right)\sum_{i=2}^{M-1} med(|r_{t,i-1}|, |r_{t,i}|, |r_{t,i+1}|)^4 \quad (10)$$

We compare the performance of the two statistic to detect which provides an improvement in forecasting volatility.

The jump component in the integrated volatility cannot be neglected if the ZJ test is significant and it is necessary to be taken into consideration in volatility modelling. The jump component for the BPV estimator is defined as follows:

$$J_{t,\alpha}(M) = [RV_{t,M} - BPV_{t,M}] \times I(ZJ_{BPV}(t,M) > \Phi_\alpha) \quad (11)$$

where I[.] is an indicator function taking value of 1 if the jump test statistic is larger than a critical value of standard normal distribution. The significance level is set as 95% in this study.

Another jump component is taken into consideration in this paper as well which stands as:

$$C_{t,\alpha}(M) = BPV_{t,M} \times I(ZJ_{BPV}(t,M) > \Phi_\alpha) + RV_{t,M} \times I(ZJ_{BPV}(t,M) \le \Phi_\alpha) \quad (12)$$

$C_{t,\alpha}(M)$ plays as a "switch" of $BPV_{t,M}$ and $RV_{t,M}$ and ensures that the sum of the squared jump component and the continuous component equals the RV whenever jump components are significant or not.

3.2. Volatility model specification

Our empirical study is based on HAR model proposed by Corsi (2009). This model can be estimated by using standard ordinary least square (OLS) but still capture long memory feature of volatility. Corsi's (2009) HAR-RV model is specified as:

Model 1. HAR-RV

$$RV_{t+1,t+h} = \beta_0 + \beta_1 RV_t + \beta_5 RV_{t-1,t-4} + \beta_{22} RV_{t-5,t-21} + \varepsilon_t \qquad (13)$$

where $RV_{t+1,t+h}$ indicates the average realised variance over the period [t+1, t+h], i.e.

$$RV_{t+1,t+h} = \frac{\sum_{i=1}^{h} RV_{t+i}}{h} \qquad (14)$$

and ε_t is the error term.

HAR-RV model forecasts the realised volatility over the period [t+1, t+h] by using the one-day, one-week and one-month lagged averaged realised volatility.

Apart from the original Corsi's (2009) model, we employ a series of alternatives of HAR model to carry out the model specification and volatility forecast comparison afterwards.

Model 2. HAR-RV-J

The HAR-RV-J is introduced by Andersen, Bollerslev, and Diebold (2007). This model composes a jump component using the one-day lagged squared jump. ABD find that the coefficient of the jump component is negative and significant. The model is specified as follows:

$$RV_{t+1,t+h} = \beta_0 + \beta_1 RV_t + \beta_5 RV_{t-1,t-4} + \beta_{22} RV_{t-5,t-21} + \beta_{SQJ} J_t + \varepsilon_t \qquad (15)$$

Model 3. HAR-CJ

HAR-CJ model is specified in ABD (2007) as well and continuous and squared jumps components are separated at different horizons in this model.

$$RV_{t+1,t+h} = \beta_0 + \beta_{C1} C_t + \beta_{SQJ1} J_t + \beta_{C5} C_{t-1,t-4} + \beta_{SQJ5} J_{t-1,t-4} + \beta_{C22} C_{t-5,t-21} +$$
$$\beta_{SQJ22} J_{t-5,t-21} + \varepsilon_t \qquad (16)$$

Model 4. PS

The PS model is developed by Patton and Sheppard (2015) and this model decomposes the one-day lagged realised volatility into a positive and negative component.

$$RV_{t+1,t+h} = \beta_0 + \beta_1^+ RSV_t^+ + \beta_1^- RSV_t^- + \beta_5 RV_{t-1,t-4} + \beta_{22} RV_{t-5,t-21} + \varepsilon_t \quad (17)$$

Model 5. PSlev

PSlev model is derived from PS model and it captures the leverage effect. PSlev model is developed by Patton and Sheppard (2011) as well

$$RV_{t+1,t+h} = \beta_0 + \beta_1^+ RSV_t^+ + \beta_1^- RSV_t^- + \gamma RV_t I_{[r_t<0]} + \beta_5 RV_{t-1,t-4} + \beta_{22} RV_{t-5,t-21} + \varepsilon_t \quad (18)$$

Model 6. HAR-RSV

Patton and Sheppard (2011) argue that positive and negative realized semi-variances can have different forecasting power at different lags.

$$RV_{t+1,t+h} = \beta_0 + \beta_1^+ RSV_t^+ + \beta_1^- RSV_t^- + \beta_5^+ RSV_{t-1,t-4}^+ + \beta_5^- RSV_{t-1,t-4}^- + \beta_{22}^+ RSV_{t-5,t-21}^+ + \beta_{22}^- RSV_{t-5,t-21}^- + \varepsilon_t \quad (19)$$

Model 7. CG model

Chen and Ghysels (2011) propose an alternative of the HAR-RSV model that includes the one-day lagged squared jump component. It is specified as:

$$RV_{t+1,t+h} = \beta_0 + \beta_1^+ RSV_t^+ + \beta_1^- RSV_t^- + \beta_5^+ RSV_{t-1,t-4}^+ + \beta_5^- RSV_{t-1,t-4}^- + \beta_{22}^+ RSV_{t-5,t-21}^+ + \beta_{22}^- RSV_{t-5,t-21}^- + \beta_{SQJ} J_t + \varepsilon_t \quad (20)$$

Model 8. HAR-RV-SJ

Patton and Sheppard (2011) develop this model in favour of the HAR-CJ model.

$$RV_{t+1,t+h} = \beta_0 + \beta_J \Delta J_t + \beta_C C_t + \beta_5 RV_{t-1,t-4} + \beta_{22} RV_{t-5,t-21} + \varepsilon_t \quad (21)$$

Model 9. HAR-CSJ

HAR-CSJ model is developed by Sevi (2014). He follows Patton and Sheppard (2011) and consider jumps over short period of time and the signs of the jumps.

$$RV_{t+1,t+h} = \beta_0 + \beta_{1J} \Delta J_t + \beta_{1C} C_t + \beta_{5J} \Delta J_{t-1,t-4} + \beta_{5C} C_{t-1,t-4} + \beta_{J22} \Delta J_{t-5,t-21} +$$

$\beta_{22C}C_{t-5,t-21} + \varepsilon_t$ (22)

Model 10. HAR-RV-SJd

HAR-RV-SJd is developed by Patton and Sheppard (2011). This model disentangles positive and negative signed jumps

$$RV_{t+1,t+h} = \beta_0 + \beta_J^{+}\Delta J_t I_{[\Delta J_t>0]} + \beta_J^{-}\Delta J_t I_{[\Delta J_t<0]} + \beta_C C_t + \beta_5 RV_{t-1,t-4} + \beta_{22} RV_{t-5,t-21} + \varepsilon_t \quad (23)$$

Model 11. HAR-CSJd

Sevi (2014) expands HAR-RV-SJd model and creates HAR-CSJd which disentangle between positive and negative signed jumps at various horizons:

$$RV_{t+1,t+h} = \beta_0 + \beta_{1J}^{+}\Delta J_t I_{[\Delta J_t>0]} + \beta_{1J}^{-}\Delta J_t I_{[\Delta J_t<0]} + \beta_{1C} C_t + \beta_{5J}^{+}\Delta J_{t-1,t-4} I_{[\Delta J_{t-1,t-4}>0]} +$$

$$\beta_{5J}^{-}\Delta J_{t-1,t-4} I_{[\Delta J_{t-1,t-4}<0]} + \beta_{5C} C_{t-1,t-4} + \beta_{22J}^{+}\Delta J_{t-5,t-21} I_{[\Delta J_{t-5,t-21}>0]} +$$

$$\beta_{22J}^{-}\Delta J_{t-5,t-21} I_{[\Delta J_{t-5,t-21}<0]} + \beta_{22C} C_{t-5,t-21} + \varepsilon_t \quad (24)$$

In this model, $\Delta J_{t-1,t-4}$ is not calculated as the difference between positive RSV and negative RSV over 4 days but as the sum of signed jump over the 4 days.

For the purpose of forecasting comparison, we also illustrate GARCH model and FIGARCH model in this section. Bollerslev (1986) proposes the GARCH model and Sadorsky (2006) demonstrates that the GARCH (1, 1) model works well for crude oil volatility. The standard GARCH (1, 1) model is given by:

$$r_t = \mu + \varepsilon_t\ ,\ \ \varepsilon_t|\Omega_{t,n-1} \sim T_v(0, h_t)$$

$$h_t = \omega + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1}\ , \quad (25)$$

where μ denotes the conditional mean, ω, α and β are the parameters of the variance equation with parameter restrictions $\omega > 0$, $\alpha > 0$, $\beta > 0$ and $\alpha + \beta < 1$. The error term $\varepsilon_{t,n}$ based on the information set $\Omega_{t,n-1}$ follows a student's T distribution T_v with zero mean, variance $h_{t,n}$ and degree of freedom v.

FIGARCH model (Baillie et al., 1996, 2004; Andersen and Bollerslev, 1997) captures the long memory properties of the volatility. The FIGARCH model assumes the finite persistence of volatility shocks (no such persistence exists in the GARCH framework), i.e., long-memory behaviour and a slow rate of decay after a volatility shock. The FIGARCH (1, d, 1) model can be written as follows:

$$r_t = \mu + \varepsilon_t \ , \ \ \varepsilon_t|\Omega_{t,n-1} \sim T_v(0, h_t)$$

$$h_t = \omega + \beta h_{t-1} + [1 - (1 - \beta L)^{-1}(1 - \varphi L)(1 - L)^d]\varepsilon_t^2 \qquad (26)$$

where $0 \leq d \leq 1$, $\omega > 0$, $\varphi, \beta < 1$. d is the fractional integration parameter and L is the lag operator. The fractional integration parameter d allows autocorrelations to decay at a slow hyperbolic rate which characterises the long-memory feature. If d is set between zero and one, FIGARCH model is able to describe intermediate ranges of persistence since it lies within d=1 representing the complete integrated persistence of volatility shocks and d=0 representing the geometric decay.

4. Data description

The original data I obtain are 15 min price data of the NYMEX light, sweet (low-sulphur) crude oil futures contract provide by Tick Data. Crude oil futures is the world's most actively traded commodity, and the NYMEX light, sweet (low-sulphur) crude oil (WTI) futures contract is the world's most liquid crude oil futures, as well as the world's largest-volume futures contract trading on a physical commodity. The data I use span the period from 25th March 2009 to 25th March 2013, containing 1033 trading days.

High frequency data contain more information on financial assets. Theoretically, the higher the frequency of the data, the more accurate the volatility estimation will be. While on the other hand, microstructure frictions, such as price discreteness and measurement errors may affect the effectiveness of high frequency data (ABDL, 1999; Bandi & Russell, 2005). I employ 15 minute data in this paper in order to mitigate microstructure effects of high frequency data, which is consistent with ABDE (2001).

NYMEX light, sweet (low-sulphur) crude oil futures has open outcry trading from 9:00 to 14:30 EST on weekdays. Investors can also trade oil futures via NYMEX electronic trading platform from 17:00 on Sunday to 17:15 the next day and from 18:00 to 17:15 (New York Time) on weekdays. The trading volumes on weekends are rather small and I remove weekend returns from the sample following the common practice in the literature (Chortareas et al. 2011; Celik & Ergin 2014). I obtain 89732 observations in total after I clear the data.

The intraday return series $r_{t,m}$ is given as follow:

$$r_{t,m} = \ln(P_{t,m}) - \ln(P_{t,m-1}) \qquad (27)$$

Where $P_{t,m}$ is the close-mid price at the nth time stamp on day t. Figure 1 shows the intraday prices of crude oil futures.

Figure 1. Plots of 15 minute price series.

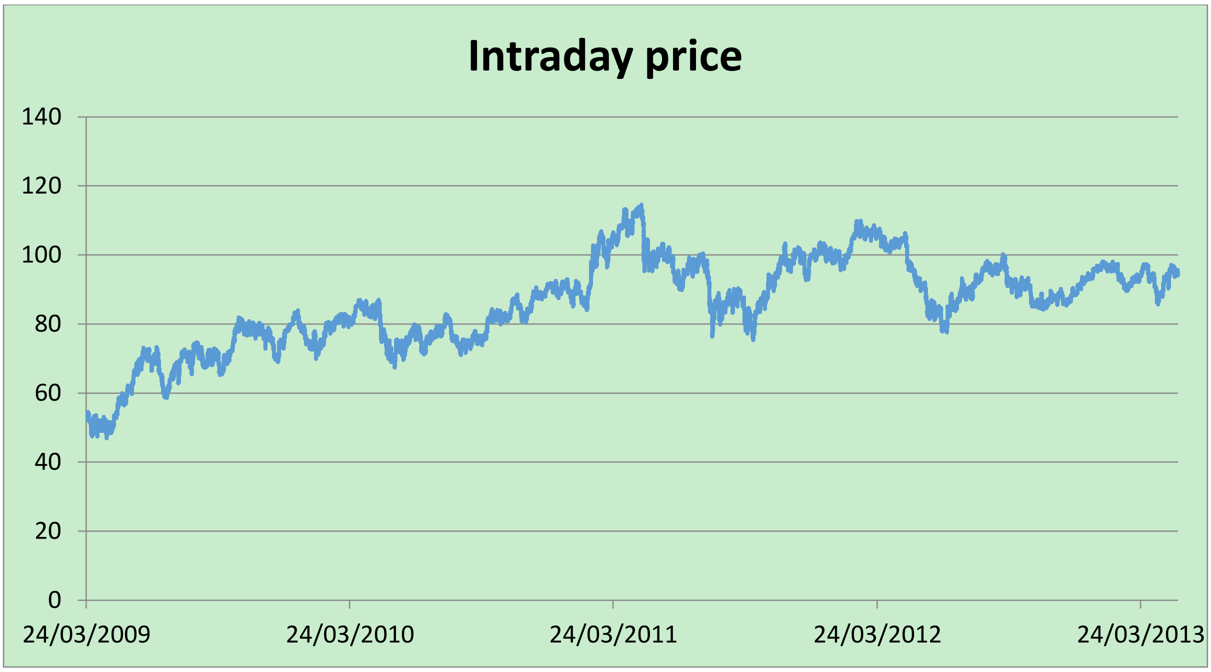


Figure 2 shows the comparison between the intraday returns of NYMEX light, sweet (low-sulphur) crude oil futures return series and those of the daily returns. Figure 3 indicates the comparison between the realised volatility and the daily volatility. Figure 4 shows the distribution of the 15 min returns and daily returns. Figure 5 illustrates autocorrelation function for the realised volatility and realised semi-variances. Table 1 represents the descriptive statistics of 15min return series.

Figure 2. Plots of 15 minute return series and daily return series.

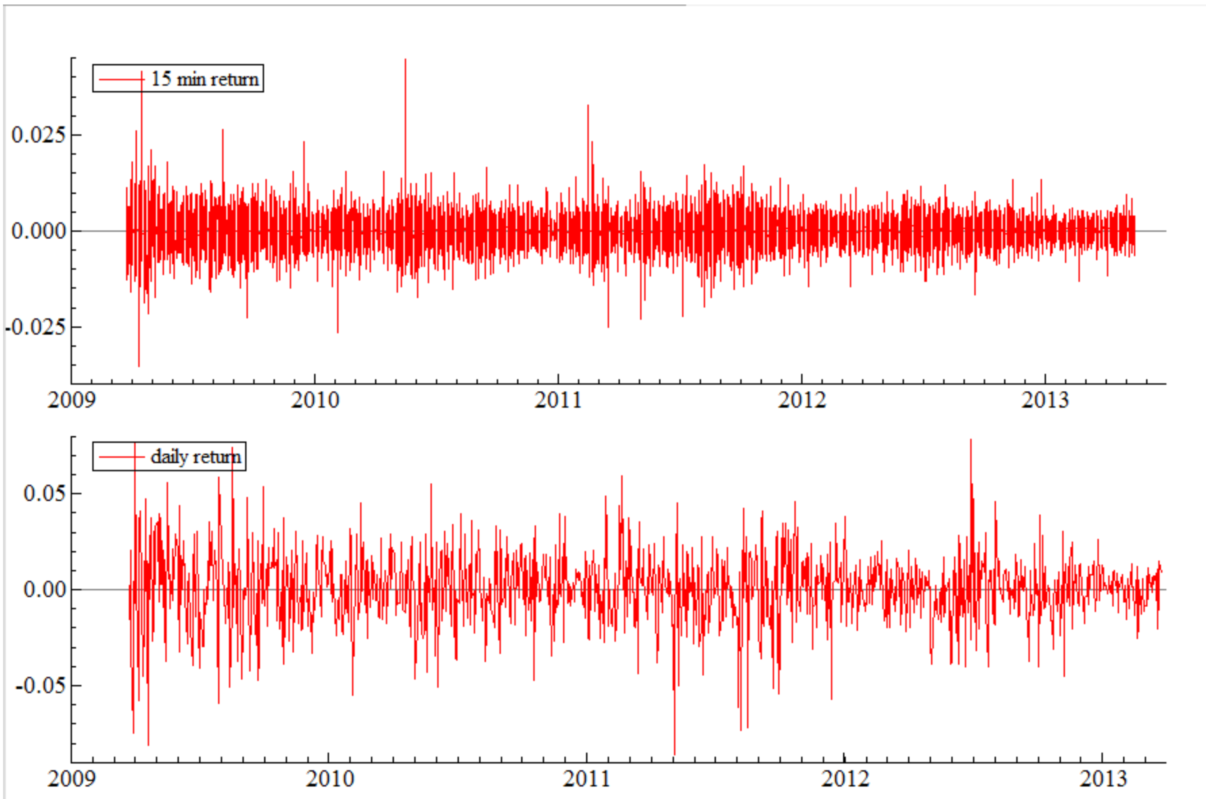


Figure 3. Plots of realised volatility.

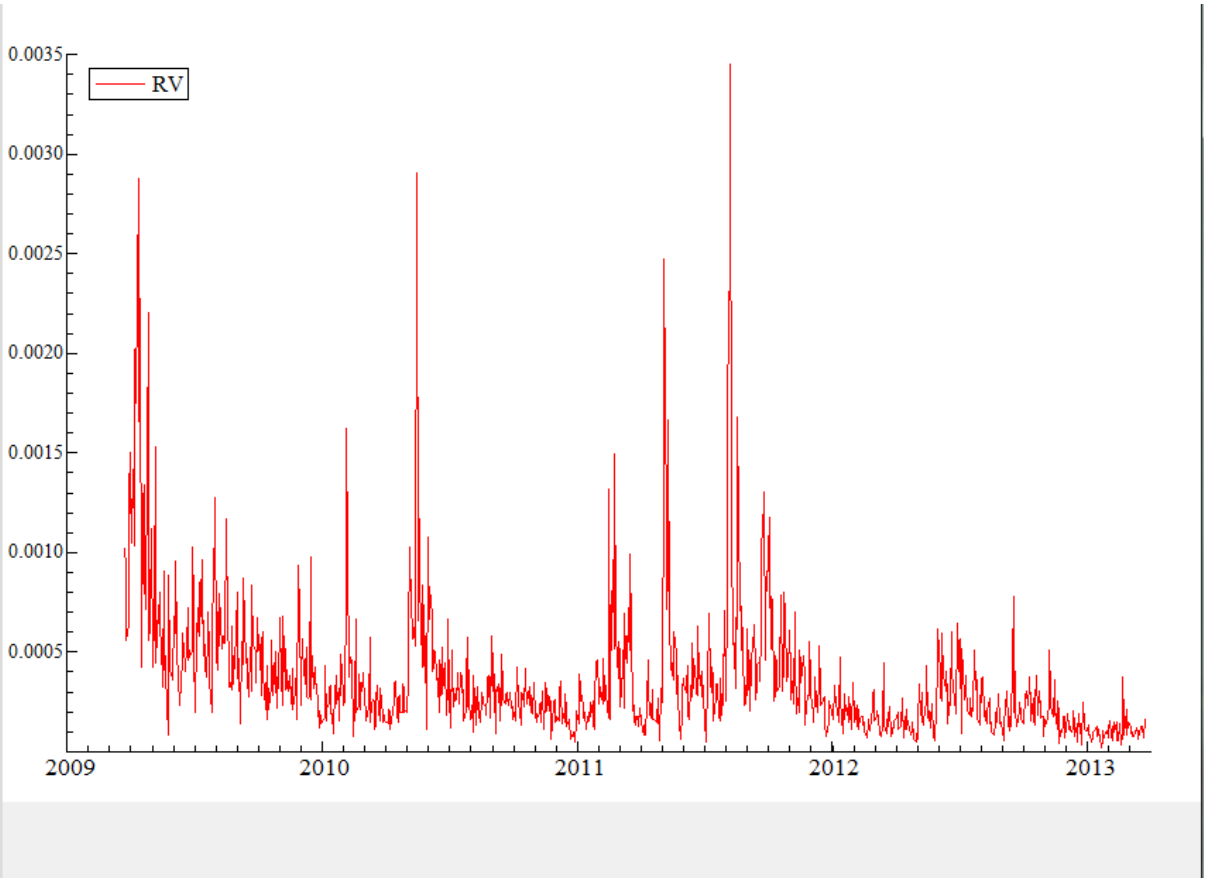


Figure 4. The distribution of 15 min return data and the daily return data

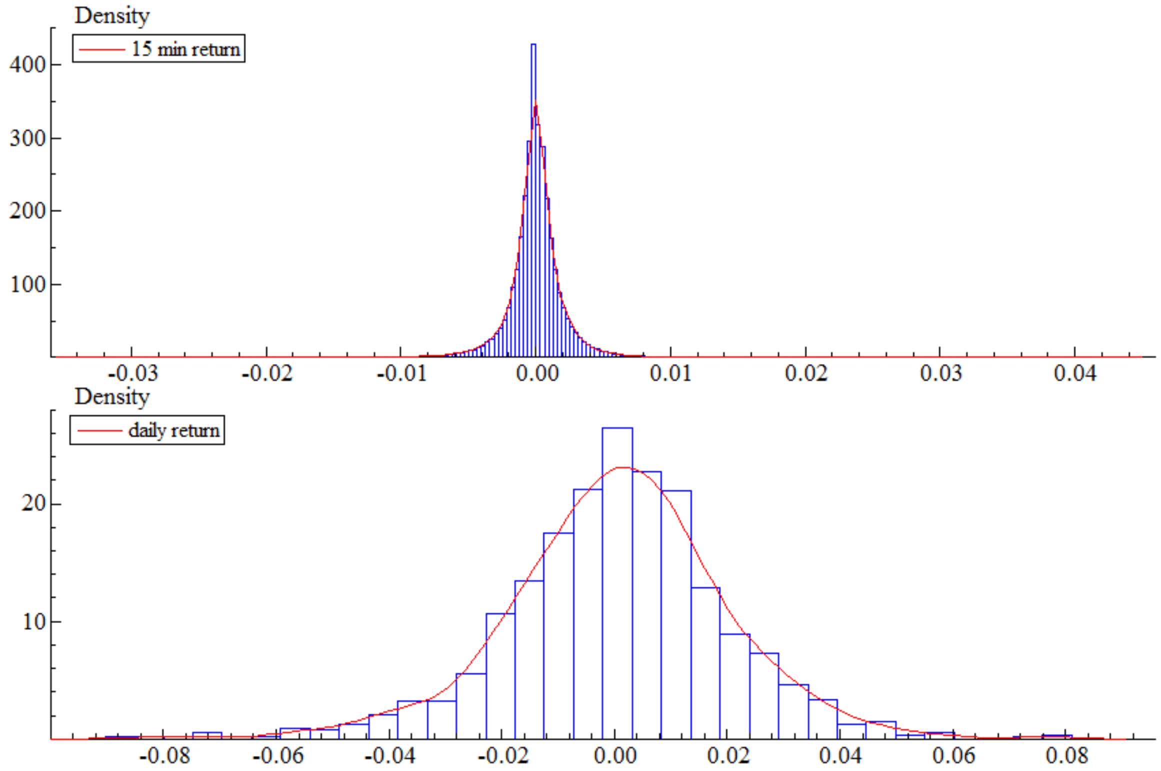
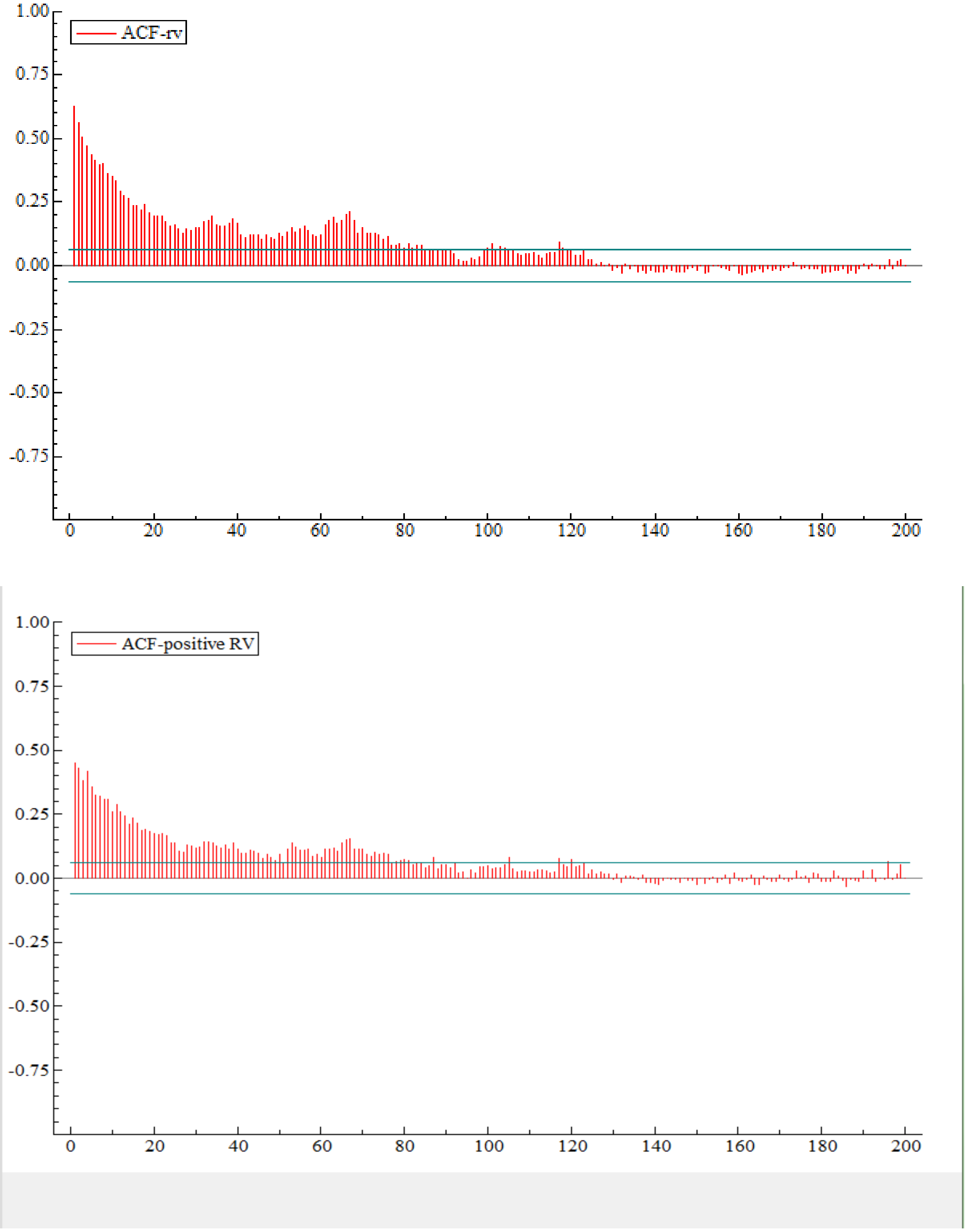
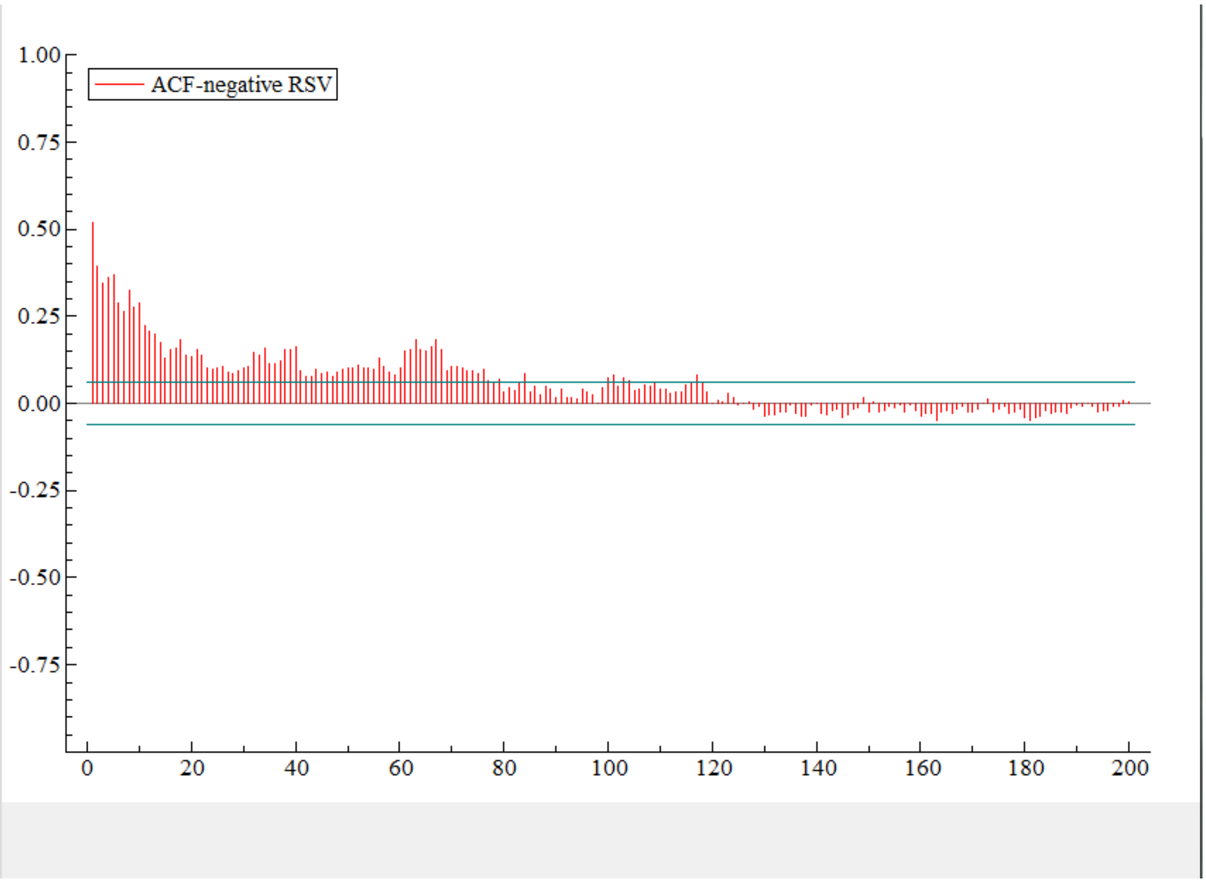


Figure 5. Sample autocorrelation for crude oil futures RV (top), positive RV (middle) and negative RV (bottom)





ACF-negative RSV
1.00
0.75
0.50
0.25
0.00
-0.25
-0.50
-0.75
0
20
40
60
80
100
120
140
160
180
200

Table 1. Summary statistics of 5 minute returns series.

	Mean ($\times 10^{-6}$)	S.D ($\times 10^{-3}$)	Skewness	Kurtosis	ADF	GPH
15 min return	6.21	2.046	0.070065	19.07676	-303.574	-0.00545191 (0.00314829)

Notes: The standard errors are in the parentheses in the last column.

Figure 2 shows that the movements of the 15 min returns are not consistent. High-frequency data carry more information thus several jumps in the daily returns are smoothed out in the 15 min returns. Figure 3 also indicates the inconsistence between the realised volatility which is constructed from the squared intraday returns. The distributions of the 15 min returns presents that the 15 min returns are much more leptokurtic than the daily returns.

Figures 1 indicate features of 15 minute returns of crude oil and these of daily returns. The crude oil shares some stylized properties of high-frequency returns of other financial assets in the literature. The mean value of crude oil returns is approximately zero, which is common among financial assets. The skewness of crude oil is 0.07, suggesting the distribution leans leftward. The kurtosis is way larger than 3, indicating the distribution is fat tailed. The augmented Dickey-Fuller unit root test supports the rejection of the null hypothesis of a unit root at the 1% significance level, implying the return series is stationary. The p-value of the GPH test on the 15 min returns is 0.0833, implying the non-rejection of the null hypothesis that the long memory parameter is zero. Meanwhile the statistics of the daily returns are different from the intraday returns. The mean and standard deviation are much larger those of the 15 min returns and the skewness is negative rather than positive compared to the skewness of the 15 min returns. The negative skewness indicates the distribution of daily returns is rightward rather than leftward which is a feather of the 15 min returns. The negative value of the ADF test statistics implies the daily returns are stationary and the GPH test result indicates the long memory parameter is zero.

5. Model Estimation

In this part, I estimate models for time horizon h=1 (daily horizon) and h=5 (weekly horizon). I follow ABD (2007) considering the authors estimate their models for h=1, h=5 and h=22.

I first compare the R squared for the same model but for different horizon. Generally, R squared are larger when time horizon becomes longer except for HAR-RV-SJd model containing MedRV jump component whose R squared for h=5 is less that for h=1. However, the R-squareds for different models and for the two horizons are not dramatically different from each other: all of them lie in the interval from 0.3 to 0.5. The highest R squared I obtain is from model HAR-CSJd containing MedRV jump component for time horizon h=5.

Table 2 to Table 5 report the parameters of all models. Table 2 and Table 3 report models containing MedRV jumps and BPV jumps with h=5 respectively. Table 4 and Table 5 report models containing MedRV jump and BPV jumps with h=1 apiece. Jumps, signed jumps and signed semi-variances are the main features of this paper so their values and significance are discussed here. Moreover, BPV is down biased (Corsi et al. (2010)) therefore comparing the in-sample fitness between BPV models and MedRV models is also discussed here.

The parameters of HAR-RV model are all significant at 5% level which is consistent to the empirical literature while the R squared of HAR-RV model here is much higher than that in ABD. For HAR-RV-J model, The R squared of HAR-RV-J model is only marginally higher than that of HAR-RV model but the parameters of the jump component is negative and highly significant regardless of MedRV jump or BPV jump. The sign of the jump component HAR-RV-J model is negative and significant which is consistent with the finding in ABD and Sevi (2014). Therefore, squared jumps in the volatility will reduce the influence of previous and current volatility on future's volatility and this kind of offsetting effect is significant.

Significance of variables in HAR-CJ model is quite mixed. For HAR-CJ model containing MedRV jump component, the current continuous jump is highly significant but the squared jump component is not. Some of the lagged jump components are not significant except the 1 month lag continuous jump. For model with MedRV jump for h=1, the R squared does not increase a lot compared to the simply HAR-RV model and it is even smaller than that of HAR-RV-J model. For model with BPV jump, continuous jumps are all significant while the squared jump components are not significant in most cases. This finding is consistent with the literature that the continuous component has most information about future volatility and the squared jump component does not have. Compare the performance of HAR-CJ model with MedRV and HAR-CJ with BPV, the R squared of the former one is higher than that of the latter for h=1 and h=5.

The R squared of PS model is only slightly higher than the simple HAR-RV model therefore the decomposition between positive and negative semi-variance does not improve the regression fitness. The positive semi-variance for h=1 is not significant which is contrary to the finding in Sevi (2014) where all the semi-variances are highly significant but the finding is in favour of Patton and Sheppard (2011). The leverage effect component in PSlev model is not significant either suggesting the leverage effect does not impact future's volatility.

Coefficients in HAR-RSV and CG model are quite like those in HAR-CJ model. There exist some significant variables but the majority of the variables are insignificant. This piece of evidence contrasts with Sevi (2014) and it indicates the decomposition of the variance does not make contribution to the predicting of future's volatility. The comparison of squared jump components in MedRV CG and BPV CG model is interesting: squared jump in CG with BPV is not significant while squared jump in CG with MedRV for h=5 is significant under 10% and squared jump in CG with MedRV for h=1 is significant under 5%. Significant as MedRV jump components are, the R squared of CG with the two jump components are almost the same. This piece of evidence suggests MedRV is superior to BPV but they two hardly contribute to explaining future's volatility in oil market which is contrary to Sevi (2014).

The signed jump component is a main feature in HAR-RV-SJ model introduced by Patton and Sheppard (2011) but its performance in oil market is mixed. The signed jump component in HAR-RV-SJ model for h=1 is significant under 5% level but it is insignificant in HAR-RV-SJ model for h=5. Comparing the R squared from HAR-RV-SJ model with that from HAR-RV model, I conclude the signed jump component does not help forecast volatility.

HAR-RV-SJd and HAR-CSJd contain the decomposed signed jumps and their corresponding weekly and monthly lags. The current negative signed jump are all significant while positive signed jump are insignificant in most cases. This is contrary to Sevi (2014). The significance of one-week lagged and one month lagged signed jumps are quite mixed indicating the noise it maintains. The R squared from HAR-CSJd is higher than that from other models but it may because the increased number of the explanatory variables or the new information the significant variables bring in.

I conclude the in-sample fitness performance as follows. There is no outperforming model in terms of the explanatory power i.e. R squared. Squared jumps help to reduce future's volatility to some extent. MedRV jump is more significant than BPV jump component but their contribution to volatility explanation is limited. The information of the decomposition of variance into semi-variance is mixed which is against Sevi's (2014) finding that considering independently the squared jump component, the continuous component, signed jumps and realised semi-variances of both signs significantly help to improve the fit of the predictive regressions.

Table 2. Parameters of the 11 models. Jump components here indicate MedRV jumps and time span h is 5.

	HAR-RV	HAR-RV-J	HAR-CJ	PS	PSlev	HAR-RSV	CG	HAR-RV-SJ	HAR-CSJ	HAR-RV-SJd	HAR-CSJd
β_0	1.15E-04*** (6.122)	1.13E-04*** (6.019)	5.35E-05*** (2.623)	1.11E-04*** (6.108)	1.10E-04*** (6.051)	1.09E-04*** (5.904)	1.10E-04*** (5.861)	1.13E-04*** (6.073)	5.74E-05*** (2.575)	1.06E-04*** (6.140)	3.91E-0.5** (2.029)
β_1	0.278*** (5.416)	0.328*** (5.853)									
β_5	0.261*** (3.193)	0.230*** (2.937)		0.266*** (3.460)	0.274*** (3.351)			0.237*** (3.159)		0.250*** (3.201)	
β_{22}	0.166** (2.231)	0.172** (2.303)		0.177** (2.384)	0.179** (2.424)			0.183** (2.367)		0.178** (2.442)	
β_{SQJ1}		-0.246*** (-2.922)	0.069 (1.612)				-0.156* (-1.758)				
β_{C1}			0.365*** (8.693)					0.318*** (6.076)	0.367*** (8.718)	0.266*** (4.629)	0.247*** (6.138)
β_{SQJ5}			0.152 (0.848)								
β_{C5}			0.0003 (0.007)						0.0009 (0.020)		-0.004 (-0.101)
β_{SQJ22}			0.357 (1.160)								
β_{C22}			0.609*** (4.462)						0.648*** (4.522)		0.588*** (4.547)
β_1^+				0.124*** (2.975)	0.126*** (2.731)	0.112** (2.241)	0.189** (2.506)				
β_1^-				0.423*** (4.335)	0.358*** (2.730)	0.422*** (4.423)	0.413*** (4.629)				
β_5^+						0.128 (0.490)	0.128 (0.492)				
β_5^-						0.420** (2.412)	0.379** (0.028)				
β_{22}^+						0.291 (0.979)	0.268 (0.894)				

β_{22}^{-}						0.096 (0.275)	0.118 (0.336)				
γ					0.047 (1.233)						
$\Delta J1$								-0.065 (-1.082)	-0.019 (-0.334)		
$\Delta J1^{-}$										-0.336*** (-3.041)	-0.296*** (-2.736)
$\Delta J1^{+}$										0.123** (2.485)	0.111** (2.277)
$\Delta J5$									-0.098 (-0.583)		
$\Delta J5^{-}$											-1.022*** (-3.949)
$\Delta J5^{+}$											0.638*** (3.407)
$\Delta J22$									-0.09 (-0.272)		
$\Delta J22^{-}$											-0.873 (-1.364)
$\Delta J22^{+}$											0.793 (1.461)
R^2	0.381	0.391	0.428	0.391	0.392	0.393	0.396	0.391	0.424	0.403	0.466

Notes: Estimation is by OLS. Newey-West adjusted t-statistics are given in brackets.

* Statistical significance at 10% level.

** Statistical significance at 5% level.

*** Statistical significance at 1% level.

Table 3. Parameters of the 11 models. Jump components here indicate BPV jumps and time span h is 5.

	HAR-RV	HAR-RV-J	HAR-CJ	PS	PSlev	HAR-RSV	CG	HAR-RV-SJ	HAR-CSJ	HAR-RV-SJd	HAR-CSJd
β_0	1.15E-04*** (6.122)	1.14E-04*** (6.098)	1.11E-04*** (5.296)	1.11E-04*** (6.108)	1.10E-04*** (6.051)	1.09E-04*** (5.904)	1.09E-04*** (5.906)	1.12E-04*** (6.119)	1.08E-04*** (5.671)	1.07E-04*** (6.176)	9.85E-05*** (5.615)
β_1	0.278*** (5.416)	0.292*** (5.444)									
β_5	0.261*** (3.193)	0.254*** (3.182)		0.266*** (3.460)	0274*** (3.351)			0.263*** (3.411)		0.278*** (3.449)	
β_{22}	0.166** (2.231)	0.162** (2.150)		0.177** (2.384)	0.179** (2.424)			0.169** (2.203)		0.169** (2.298)	
β_{SQJ1}		-0.157** (-2.378)	0.134** (8.674)				0.019 (0.306)				
β_{C1}			0.291*** (5.430)					0.283*** (5.706)	0.281*** (6.087)	0.232 (3.925)	0.219*** (4.042)
β_{SQJ5}			0.271** (2.275)								
β_{C5}			0.253*** (3.007)						0.260*** (3.022)		-0.214** (2.098)
β_{SQJ22}			-0.098 (-0.208)								
β_{C22}			0.176** (2.044)						0.191** (2.109)		0.163 (1.266)
β_1^+				0.124*** (2.975)	0.126*** (2.731)	0.112** (2.241)	0.107 (1.585)				
β_1^-				0.423*** (4.335)	0.358*** (2.730)	0.422*** (4.423)	0.424*** (4.390)				
β_5^+						0.128 (0.490)	0.129 (0.492)				
β_5^-						0.420** (2.412)	0.422** (2.388)				
β_{22}^+						0.291 (0.979)	0.294 (0.996)				

β_{22}^{-}						0.096 (0.275)	0.094 (0.270)				
γ					0.047 (1.233)						
$\Delta J1$								-0.099 (-1.600)	-0.106 (-1.630)		
$\Delta J1^{-}$										-0.311*** (-2.635)	-0.329*** (-2.679)
$\Delta J1^{+}$										0.050 (0.686)	0.044 (0.590)
$\Delta J5$									-0.117 (-0.646)		
$\Delta J5^{-}$											-0.450 (-1.449)
$\Delta J5^{+}$											0.091 (0.428)
$\Delta J22$									0.040 (0.121)		
$\Delta J22^{-}$											-0.121 (-0.206)
$\Delta J22^{+}$											0.228 (0.607)
R^2	0.381	0.383	0.383	0.391	0.392	0.393	0.393	0.385	0.386	0.391	0.396

Notes: Estimation is by OLS. Newey-West adjusted t-statistics are given in brackets.

* Statistical significance at 10% level.

** Statistical significance at 5% level.

*** Statistical significance at 1% level.

Table 4. Parameters of the 11 models. Jump components here indicate MedRV jumps and time span h is 1.

	HAR-RV	HAR-RV-J	HAR-CJ	PS	PSlev	HAR-RSV	CG	HAR-RV-SJ	HAR-CSJ	HAR-RV-SJd	HAR-CSJd
β_0	7.41E-05*** (5.227164)	7.09E-05*** (4.986)	4.35E-05** (2.368)	6.53E-05*** (4.863)	6.43E-05*** (4.733)	6.21E-05*** (3.621)	6.27E-05*** (4.271)	6.75E-05*** (4.780)	4.87E-05** (2.454)	5.57E-05*** (3.325)	2.60E-0.5 (1.423)
β_1	0.395*** (3.669)	0.494*** (4.780)									
β_5	0.267*** (2.711)	0.205** (2.221)		0.278*** (3.178)	0.287*** (3.212)			0.229*** (2.625)		0.251*** (2.798)	
β_{22}	0.157** (2.279)	0.171** (2.520)		0.181*** (2.743)	0.183*** (2.752)			0.187*** (2.754)		0.179 *** (2.851)	
β_{SQJ1}		-0.491*** (-3.641)	-0.004312 (0.9538)				-0.302** (-2.226)				
β_{C1}			0.549*** (7.727)					0.462*** (4.722)	0.538*** (16.692)	0.374*** (3.505)	0.389*** (5.036)
β_{SQJ5}			0.128 (0.787)								
β_{C5}			0.016 (0.391)						0.0195 (0.437)		0.006 (0.167)
β_{SQJ22}			0.621** (2.073)								
β_{C22}			0.398*** (3.719)						0.458*** (6.040)		0.390*** (4.055)
β_1^+				0.067 (0.877)	0.069 (0.806)	0.048 (0.787)	0.196* (1.722)				
β_1^-				0.704*** (3.778)	0.188*** (3.354)	0.702*** (12.087)	0.685*** (4.087)				
β_5^+						0.016 (0.108)	0.015 (0.063)				
β_5^-						0.525** (4.633)	0.447*** (2.957)				
β_{22}^+						0.245	0.200				

						(0.858)	(0.596)				
β_{22}^{-}						0.177 (0.576)	0.219 (0.591)				
γ					0.054 (0.995)						
$\Delta J1$								-0.195** (-2.157)	-0.142*** (-2.785)		
$\Delta J1^{-}$										-0.649*** (-3.228)	-0.606*** (-2.906)
$\Delta J1^{+}$										0.121 (1.446)	0.121** (1.252)
$\Delta J5$									-0.140 (-1.319)		
$\Delta J5^{-}$											-0.930*** (-3.312)
$\Delta J5^{+}$											0.467*** (2.496)
$\Delta J22$									-0.073 (-0.253)		
$\Delta J22^{-}$											-1.359* (-1.847)
$\Delta J22^{+}$											1.347** (2.159)
R^2	0.351	0.375	0.374	0.378	0.379	0.382	0.389	0.385	0.376	0.405	0.412

Notes: Estimation is by OLS. Newey-West adjusted t-statistics are given in brackets.

* Statistical significance at 10% level.

** Statistical significance at 5% level.

*** Statistical significance at 1% level.

Table 5. Parameters of the 11 models. Jump components here indicate BPV jumps and time span h is 1.

	HAR-RV	HAR-RV-J	HAR-CJ	PS	PSlev	HAR-RSV	CG	HAR-RV-SJ	HAR-CSJ	HAR-RV-SJd	HAR-CSJd
β_0	7.41E-05*** (5.227164)	7.14E-05*** (5.158)	7.01E-05*** (4.675)	6.53E-05*** (4.863)	6.43E-05*** (4.733)	6.21E-05*** (4.228)	6.20E-05*** (4.223)	6.63E-05*** (4.931)	6.07E-05** (4.119)	5.85E-05*** (4.472)	4.81E-0.5*** (3.382)
β_1	0.395*** (3.669)	0.432*** (4.027)									
β_5	0.267*** (2.711)	0.248*** (2.611)		0.278*** (3.178)	0.287*** (3.212)			0.269*** (3.032)		0.298*** (3.207)	
β_{22}	0.157** (2.279)	0.147** (2.124)		0.181*** (2.743)	0.183*** (2.752)			0.168** (2.470)		0.169 ** (2.548)	
β_{SQJ1}		-0.491*** (-3.287)	0.012 (0.549)				-0.078** (-0.732)				
β_{C1}			0.432*** (4.020)					0.405*** (4.070)	0.399*** (4.174)	0.313*** (2.670)	0.295** (2.564)
β_{SQJ5}			0.253 (1.420)								
β_{C5}			0.248** (2.523)						0.262*** (2.707)		0.227** (2.316)
β_{SQJ22}			0.038 (0.088)								
β_{C22}			0.153** (2.030)						0.204** (2.466)		0.144 (1.177)
β_1^+				0.067 (0.877)	0.069 (0.806)	0.048 (0.604)	0.069 (0.679)				
β_1^-				0.704*** (3.778)	0.188*** (3.354)	0.702*** (3.824)	0.695*** (3.837)				
β_5^+						0.016 (0.068)	0.015 (0.062)				
β_5^-						0.525** (3.418)	0.518*** (3.356)				
β_{22}^+						0.245	0.230				

						(0.755)	(0.695)				
β_{22}^{-}						0.177 (0.491)	0.186 (0.513)				
γ					0.054 (0.995)						
$\Delta J1$								-0.245** (-2.509)	-0.256*** (-2.588)		
$\Delta J1^{-}$										-0.633*** (-2.924)	-0.656*** (-2.980)
$\Delta J1^{+}$										0.027 (0.256)	0.017 (0.162)
$\Delta J5$									-0.2098 (-1.364)		
$\Delta J5^{-}$											-0.558** (-2.046)
$\Delta J5^{+}$											-0.002 (-0.012)
$\Delta J22$									-0.0387 (-0.105)		
$\Delta J22^{-}$											-0.333 (-0.539)
$\Delta J22^{+}$											0.274 (0.734)
R^2	0.351	0.359	0.359	0.378	0.379	0.382	0.382	0.375	0.377	0.388	0.393

Notes: Estimation is by OLS. Newey-West adjusted t-statistics are given in brackets.

* Statistical significance at 10% level.

** Statistical significance at 5% level.

*** Statistical significance at 1% level.

6. Forecast evaluation

6.1. Diebold-Mariano test

Diebold and Mariano (1995) and West (1996) develop the Diebold–Mariano–White (DMW) statistic which compares the forecast ability of two competing models, requires a loss function that is a measure of the difference between the realised value and the forecast in a pseudo out-of-sample forecasting environment. the loss function relative to the benchmark model is defined as $X_{t;l}^{(A,B)} = L_{t;l}^{(A)} - L_{t;l}^{(B)}$, where $L_{t,l}^{(A)}$ is the value of the loss function l at time t for a benchmark model M_A and $L_{t,l}^{(B)}$ is the value of the loss function l at time t for the competitive model M_B .

Then, a DMW test of equal predictive accuracy is a simple Wald test that the expected value of this difference is zero. The DMW statistic is then given by:

$$DM = \bar{X}_{t=1,2,\dots,\tau;l}^{(A,B)} \Big/ \hat{\Sigma}_\tau / \sqrt{\tau} \tag{28}$$

where $\hat{\Sigma}_\tau$ is an estimator of the asymptotic standard deviation of $\Sigma_\tau = \sqrt{var[\sqrt{\tau}\bar{X}_{t=1,2,\dots,\tau;l}^{(A,B)}]}$ and τ is the number of predictions for each forecast horizon. The statistic follows standard normal distribution and permits an easy comparison of pairs of models at each horizon.

I compare the out-of-sample performances of the eleven models. I use Diebold-Mariano (DM) statistics to compare the forecast predictability of two competing models. The loss function I choose is mean squared error.

I report DM statistics for the models in table 6 and 7. In table 6, the models containing jumps are MedRV jump models while models comprising jumps in Table 7 are BPV jump models. Table 6 and 7 are read in the following way: the statistic number compares the model whose name is in the headline with the model whose name is in the head column. Number in brackets indicates the P-value of the DM statistic. A negative statistic indicates the model in the head column outperforms the one in the headline and the corresponding P-value indicates if the null hypothesis that the two forecasts have the same accuracy should be rejected or not.

Table 6 indicates that HAR-RV model is significantly inferior to 5 models while HAR-CJ model significantly outperforms the most of the other model except HAR-CSJ and HAR-CSJd. HAR-CSJ model is superior to the most of the other model except HAR-CJ and HAR-CSJd. HAR-CSJd model's performance is good as well because it outperforms the most of the model except HAR-CS and HAR-CSJ. Therefore, the best performing models in Table 6 are HAR-CS, HAR-CSJ and HAR-CSJd.

Table 7 tells a similar story when models comprising the BPV jumps. The simple HAR-RV is outperformed by many other models such as HAR-CJ, HAR-RSV, HAR-CSJ and HAR-CSJd. The performance of HAR-CJ model is no longer as good as the one comprising MedRV jumps while HAR-CSJ still outperforms the most of the alternative models. The best performing model in table 7 is HAR-CSJd model since its statistics compared by other models are all significantly positive.

To summarise, whatever the jump component is BPV or MedRV, the HAR-CSJd outperforms the other models in most cases while the simple HAR-RV model's forecast performance is inferior to many other models. HAR-CJ model containing MedRV jump performs well in it comes to forecasting while HAR-CJ model with BPV jump is at best as good as other alternative models. This finding is contrary to Sevi's (2014) finding in that HAR-CSJd model in Sevi's paper is not superior and HAR-RV model for long time horizon outperforms other models.

Table 6. DM statistic of the mean squared error of the forecast from alternative models

	HAR-RV-J	HAR-CJ	PS	Pslev	HAR-RSV	CG	HAR-RV-SJ	HAR-CSJ	HAR-SJd	HAR-CSJd
HAR-RV	1.495751 (0.134719)	14.14831 (1.91E-45)***	0.882018 (0.377767)	1.369795 (0.170751)	1.694479 (0.090174)*	1.910731 (0.056039)*	0.701995 (0.482682)	13.95058 (3.12E-44)***	1.362478 (0.173047)	11.08730 (1.45E-28)***
HAR-RV-J		13.90131 (6.22E-44)***	-0.280774 (0.778884)	0.442793 (0.657916)	0.780069 (0.435350)	1.531923 (0.125541)	-0.578522 (0.562912)	14.09908 (3.85E-45)***	1.014237 (0.310470)	11.08342 (1.51E-28)***
HAR-CJ			-13.99564 (1.66E-44)***	-13.76507 (4.13E-43)***	-14.20714 (8.27E-46)***	-13.81877 (1.96E-43)***	-13.34241 (1.31E-40)***	-0.588258 (0.556359)	-10.55009 (5.07E-26)***	0.578479 (0.562940)
PS				2.546691 (0.010875)**	1.982784 (0.047392)**	2.291278 (0.021947)**	-0.090465 (0.927918)	14.34648 (1.12E-46)	1.497145 (0.134356)	11.98837 (4.09E-33)***
PSlev					0.549769 (0.582478)	1.032055 (0.302046)	-0.868289 (0.385236)	14.10215 (3.68E-45)***	0.970639 (0.331728)	11.94907 (6.57E-33)***
HAR-RSV						0.924649 (0.355148)	-1.155670 (0.247816)	14.78945 (1.71E-49)***	0.638495 (0.523152)	12.21870 (2.47E-34)***
CG							-2.492339 (0.012690)**	14.47823 (1.66E-47)***	0.434673 (0.663800)	11.79875 (3.96E-32)***
HAR-RV-SJ								13.75686 (4.63E-43)***	1.488015 (0.136747)	11.24251 (2.52E-29)***
HAR-CSJ									-11.04847 (2.23E-28)***	0.779782 (0.435519)
HAR-RV-SJd										10.34229 (4.54E-25)***

Notes: A positive test statistic indicates the model in the head line outperforms the one in the head column. P values are given in brackets. Models comprising jumps mentioned in this table are based on MedRV jump detection. Statistical significance at 10%, 5% and 1% are highlighted by *, ** and *** respectively.

Table 7. DM statistic of the mean squared error of the forecast from alternative models.

	HAR-RV-J	HAR-CJ	PS	Pslev	HAR-RSV	CG	HAR-RV-SJ	HAR-CSJ	HAR-SJd	HAR-CSJd
HAR-RV	3.611071 (0.000305)***	13.73897 (5.93E-43)***	0.882018 (0.377767)	1.369795 (0.170751)	1.694479 (0.090174)*	1.641236 (0.100748)	0.921840 (0.356612)	3.041721 (0.002352)***	1.143385 (0.252879)	2.826803 (0.004702)***
HAR-RV-J		17.48959 (1.72E-68)***	0.569771 (0.568833)	1.094689 (0.273653)	1.411304 (0.158155)	1.360301 (0.173735)	0.461938 (0.644126)	2.721476 (0.006499)***	0.944735 (0.344794)	2.705716 (0.006816)***
HAR-CJ			-0.867077 (0.385900)	-0.224705 (0.822209)	0.038122 (0.969590)	0.009350 (0.992540)	-1.679437 (0.093067)*	0.823914 (0.409989)	-0.102989 (0.917972)	1.761832 (0.078098)*
PS				2.546691 (0.01088)**	1.982784 (0.047392)**	1.909945 (0.05614)*	-0.788254 (0.430548)	3.161925 (0.001567)***	0.980827 (0.326678)	4.211609 (2.54E-05)***
PSlev					0.549769 (0.582478)	0.493379 (0.621745)	-1.922767 (0.054509)*	1.529238 (0.126205)	0.131664 (0.895250)	3.368621 (0.000755)***
HAR-RSV						-1.504959 (0.132335)	-2.217286 (0.026604)**	1.833882 (0.066672)*	-0.180615 (0.856670)	2.956664 (0.003110)***
CG							-2.110406 (0.034823)**	1.823154 (0.068280)*	-0.150667 (0.880238)	2.996223 (0.002733)***
HAR-RV-SJ								5.296440 (1.18E-07)***	1.122830 (0.261510)	3.790724 (0.000150)***
HAR-CSJ									-0.740147 (0.459211)	2.092357 (0.036407)**
HAR-RV-SJd										4.359302 (1.30E-05)***

Notes: A positive test statistic indicates the model in the headline outperforms the one in the head-column. P values are given in brackets. Models comprising jumps mentioned in this table are based on BPV jump detection. Statistical significance at 10%, 5% and 1% are highlighted by *, ** and *** respectively.

6.2. Superior Predictive Ability (SPA) test

Apart from DM test, I carry out SPA test to detect the forecast superiority of the RV and RV jump models. The SPA test can be used to compare the performance of two or more forecasting models at a time. Forecasts are evaluated using a pre-specified loss function and the "best" forecasting model is the one that produces the smallest expected loss. In a SPA test, the loss function relative to the benchmark model is defined as $X_{t,l}^{(0,i)} = L_{t,l}^{(0)} - L_{t,l}^{(i)}$, where $L_{t,l}^{(0)}$ is the value of the loss function l at time t for a benchmark model M_0 and $L_{t,l}^{(i)}$ is the value of the loss function l at time t for another competitive model M_i for $i = 1, \ldots, K$. The SPA test is used to compare the forecasting performance of a benchmark model against its K competitors. The null hypothesis that the benchmark or base model is not outperformed by any of the other competitive models is expressed as $H_0\colon \max_{i=1,\ldots,K} E(X_{t,l}^{(0,i)}) \leq 0$. It is tested with the statistic $T_l^{SPA} = \max_{i=1,\ldots,K} (\sqrt{n}\bar{X}_{i,l} / \sqrt{\lim_{n\to\infty} var(\sqrt{n}\bar{X}_{i,l})}\ )$, where n is the number of forecast data points and $\bar{X}_{i,l} = \frac{1}{n}\sum_{t=1}^{n} X_{t,l}^{(0,i)}$. $\lim_{n\to\infty} var(\sqrt{n}\bar{X}_{i,l})$ and the p-value of the T_l^{SPA} are obtained by using the stationary bootstrap procedure discussed by Politis and Romano (1994). Hansen (2005) summarises that the p-value of a SPA test indicates the relative performance of a base model M_0 in comparison with alternative models M_i . A high p-value indicates that we are not able to reject the null hypothesis that "the base model is not outperformed".

SPA test selects six models out of a number of alternative models and the six models are the most significant model, the best model, models with performance of 75%, 50% and 25% relative to the benchmark model and the worst performance model. Each model I employ will be regarded as a benchmark model so that the null hypothesis of the SPA test that the benchmark model is not inferior to other models can be tested.

The following tables indicate the SPA test results for selected models. Models comprising jumps in Table 8 to Table 10 are MedRV jump models while models containing jumps in Table 11 to Table 13 are BPV jump models. The loss functions I choose are MSE and MAE and the P-value of the test is produced at the bottom of each table.

Table 8. SPA test results evaluated by the MAE and MSE for HAR-RV model with MedRV jump component

	MAE	MSE	MAE	MSE
	Models		t-statistics	
Benchmark	HAR-RV	HAR-RV	-	-
Most Significant	HAR-CSJd	CG	12.28756	11.71140
Best model	HAR-CSJd	HAR-CSJd	12.28756	11.18509
Model_25%	HAR-CSJ	HAR-CSJ	4.41251	2.24168
Median_50%	HAR-RSV	PSlev	2.76050	3.29203
Model_75%	HAR-RV-J	HAR-RV-J	1.47870	2.58759
Worst model	HAR-RV-SJ	HAR-RV-SJ	-12.00076	-11.02174
SPA test p-value	MAE	MSE		
	0.0000	0.0000		

Notes: Table 8 shows the SPA test results for different models. The benchmark model selected is HAR-RV model with MedRV jump component. The null hypothesis of the test is that the benchmark model is not inferior to the other candidate models. The test chooses the most significant model, the best model, models with performances of 75%, 50% and 25% relative to the benchmark model, and the worst model. P-values are reported in the last row.

Table 9. SPA test results evaluated by the MAE and MSE for HAR-RV-J model with MedRV jump component

	MAE	MSE	MAE	MSE
	Models		t-statistics	
Benchmark	HAR-RV-J	HAR-RV-J	-	-
Most Significant	CG	CG	13.20796	12.08039
Best model	HAR-RV-SJd	HAR-CSJd	12.22289	11.03887
Model_25%	HAR-CSJ	HAR-CSJ	4.14038	1.81780
Median_50%	HAR-RSV	PSlev	2.10960	2.56997
Model_75%	HAR-RV	HAR-RV	-2.47174	-2.58759
Worst model	HAR-RV-SJ	HAR-RV-SJ	-13.28610	-10.58156
SPA test p-value	MAE	MSE		
	0.0000	0.0000		

Notes: Table 9 shows the SPA test results for different models. The benchmark model selected is HAR-RV-J model with MedRV jump component. The null hypothesis of the test is that the benchmark model is not inferior to the other candidate models. The test chooses the most significant model, the best model, models with performances of 75%, 50% and 25% relative to the benchmark model, and the worst model. P-values are reported in the last row.

Table 10. SPA test results evaluated by the MAE and MSE for HAR-CSJd model with MedRV jump component

	MAE	MSE	MAE	MSE
	Models		t-statistics	
Benchmark	HAR-CSJd	HAR-CSJd	-	-
Most Significant	HAR-RV-J	HAR-RV-J	-6.24035	-4.40787
Best model	HAR- RV-SJd	HAR-RV-SJd	-7.56712	-5.07985
Model_25%	HAR-RV-SJ	HAR-RV-SJ	-8.57026	-5.86700
Median_50%	PSlev	PS	-12.50230	-10.39378
Model_75%	HAR-RV	HAR-RV	-12.22289	-11.03887
Worst model	CG	CG	-17.01871	-14.00429
SPA test p-value	MAE	MSE		
	0.50650	0.52640		

Notes: Table 10 shows the SPA test results for different models. The benchmark model selected is HAR-CSJd model with MedRV jump component. The null hypothesis of the test is that the benchmark model is not inferior to the other candidate models. The test chooses the most significant model, the best model, models with performances of 75%, 50% and 25% relative to the benchmark model, and the worst model. P-values are reported in the last row.

Table 11. SPA test results evaluated by the MAE and MSE for HAR-RV model with BPV jump component

	MAE	MSE	MAE	MSE
	Models		t-statistics	
Benchmark	HAR-RV	HAR-RV	-	-
Most Significant	HAR-CJ	HAR-CJ	22.96357	17.33296
Best model	HAR-CSJd	HAR-CSJd	6.47564	6.07192
Model_25%	HAR-CSJ	HAR-CSJ	4.71027	5.13676
Median_50%	HAR-RSV	HAR-RSV	2.83639	3.15563
Model_75%	HAR-CJ	HAR-CJ	22.96357	17.33296
Worst model	HAR-RV-J	HAR-RV-J	11.74953	8.89234
SPA test p-value	MAE	MSE		
	0.0000	0.0000		

Notes: Table 11 shows the SPA test results for different models. The benchmark model selected is HAR-RV model with BPV jump component. The null hypothesis of the test is that the benchmark model is not inferior to the other candidate models. The test chooses the most significant model, the best model, models with performances of 75%, 50% and 25% relative to the benchmark model, and the worst model. P-values are reported in the last row.

Table 12. SPA test results evaluated by the MAE and MSE for HAR-CJ model with BPV jump component

	MAE	MSE	MAE	MSE
	Models		t-statistics	
Benchmark	HAR-CJ	HAR-CJ	-	-
Most Significant	HAR-CSJd	HAR-CSJd	4.80494	4.48272
Best model	HAR-CSJd	HAR-CSJd	4.80494	4.48272
Model_25%	HAR-CSJ	HAR-CSJ	1.62682	1.98999
Median_50%	HAR-RSV	HAR-RSV	0.56885	0.76741
Model_75%	HAR-RV-J	HAR-RV-J	-2.45372	-2.04172
Worst model	HAR-RV	HAR-RV	-14.21593	-12.48918
SPA test p-value	MAE	MSE		
	0.0000	0.0003		

Notes: Table 11 shows the SPA test results for different models. The benchmark model selected is HAR-CJ model with BPV jump component. The null hypothesis of the test is that the benchmark model is not inferior to the other candidate models. The test chooses the most significant model, the best model, models with performances of 75%, 50% and 25% relative to the benchmark model, and the worst model. P-values are reported in the last row.

Table 13. SPA test results evaluated by the MAE and MSE for HAR- CSJd model with BPV jump component

	MAE	MSE	MAE	MSE
	Models		t-statistics	
Benchmark	HAR-CSJd	HAR-CSJd	-	-
Most Significant	HAR-CSJ	HAR-CSJ	-3.32102	-2.94373
Best model	HAR-RV-SJd	HAR-RV-SJd	-4.56521	-3.14259
Model_25%	HAR-RV-SJ	HAR-RV-SJ	-6.34987	-5.81179
Median_50%	PSlev	PSlev	-7.43376	-7.42629
Model_75%	HAR-RV-J	HAR-RV-J	-4.80494	-4.48272
Worst model	HAR-RV	HAR-RV	-5.91385	-5.58171
SPA test p-value	MAE	MSE		
	0.4799	0.48510		

Notes: Table 13 shows the SPA test results for different models. The benchmark model selected is HAR-CSJd model with BPV jump component. The null hypothesis of the test is that the benchmark model is not inferior to the other candidate models. The test chooses the most significant model, the best model, models with performances of 75%, 50% and 25% relative to the benchmark model, and the worst model. P-values are reported in the last row.

Ranking information in Table. 8, 9, 11 and 12 is quite consistent. For each benchmark model, the ranking information from the best model to the worst model is almost the same in spite of some small changes. For MedRV models the worst performing model is HAR-RV-SJ in most cases and the worst performing model for BPV models is HAR-RV model. In most cases, HAR-CSJd is the best performing model compared to other alternative models. The P-value of the SPA test support the null hypothesis that the HAR-CSJd model is not inferior to other model considering. For other benchmark models, the SPA test P-value is close to zero, indicating that the null hypothesis that the benchmark model is not inferior to other models is rejected. Those results are not presented in this paper to avoid repeat results. It illustrates that apart from HAR-CSJd, other models are at least inferior to one of the competing models. The finding of SPA test is contrary to Sevi (2014) as well. Sevi's finding suggests that the decompositions between jumps and the continuous components and negative and positive realised semi-variances do not improve the forecast performance for crude oil asset. However, HAR-CSJd mode, which is the most complicated model I employ at best outperforms simple model such as HAR-RV and other decomposition models or is as good as them at worst. Sevi (2014) studies data covering from January 1987 to December 2010 which is long and slightly dated. The sample in this study covers from 25th March 2009 to 25th March 2013. Though data in the two studies have overlapping part, the non-overlapping part outweigh the overlapping part. Since same HAR series models are utilised in the two studies, this contradiction stems from the difference of the two data samples.

6.3. The comparison of forecasting performance between HAR models and GARCH models

I choose two batches of models to compare their forecasting performance. One batch contains base HAR-RV model and the good performing model: HAR-CJ and HAR-CSJd model according to DM test and SPA test. The other batch has GARCH model and FIGARCH model which takes long memory property into consideration. The forecast period is still from 3rd Nov. 2012 to 25th Mar. 2013. To make the comparison applicable, I sum up the 5 min volatility within the same day to construct the realised volatility during this period and carry on the comparison the realised volatility from HAR models and GARCH family models.

I report DM statistics for the models in Table 14 and 15. In table 14, the models containing jumps are BPV jump models while models comprising jumps in Table 15 are MedRV jump models. Table 14 and 15 are read in the following way: the statistic number compares the model whose name is in the headline with the model whose name is in the head column. Number in brackets indicates the P-value of the DM statistic. A negative statistic indicates the model in the head column outperforms the one in the headline and the corresponding P-value indicates if the null hypothesis that the two forecasts have the same accuracy should be rejected or not.

Table 14. DM statistic of the mean squared error of the forecast from alternative models (I).

	HAR-CJ	HAR-CSJd	GARCH	FIGARCH
HAR-RV	13.73897 (5.93E-43)***	2.826803 (0.004702)***	27.3795 (4.812E-165)***	12.5386 (4.59E-36)***
HAR-CJ		1.761832 (0.078098)*	26.3004 (1.897E-152)***	11.7888 (4.458E-32)***
HAR-CSJd			20.9017 (5.167E-97)***	8.9729 (2.888E-19)***
GARCH				-45.2982 (0.0000) ***

Notes: Models comprising jumps mentioned in this table are based on BPV jump detection. A positive test statistic indicates the model in the head line outperforms the one in the head column. P values are given in brackets. Statistical significance at 10%, 5% and 1% are highlighted by *, ** and *** respectively.

Table 15. DM statistic of the mean squared error of the forecast from alternative models (II).

	HAR-CJ	HAR-CSJd	GARCH	FIGARCH
HAR-RV	14.14831 (1.91E-45)***	11.08730 (1.45E-28)***	27.3795 (4.812E-165)***	12.5386 (4.59E-36)***
HAR-CJ		0.578479 (0.562940)	33.6754 (1.325E-248) ***	5.7032 (1.17579E-08) ***
HAR-CSJd			13.8765 (8.793E-44) ***	0.2667 (0.789700154)
GARCH				-45.2982 (0.0000) ***

Notes: Models comprising jumps mentioned in this table are based on MedRV jump detection. A positive test statistic indicates the model in the headline outperforms the one in the head-column. P values are given in brackets. Statistical significance at 10%, 5% and 1% are highlighted by *, ** and *** respectively

The comparison among HAR-RV, HAR-CJ and HAR-CSJd model has been discussed and the attention is given to the comparison between HAR-RV models and GARCH models. The HAR-RV family models are not performing well when the GARCH model and FIGARCH model are added to the comparison batch. GARCH model and FIGARCH model outperform HAR-RV models according to DM test. GARCH model significantly outperforms HAR-RV, HAR-CJ, HAR-CSJd and FIGARCH model pairwise while FIGARCH model is only inferior to GARCH model while it still outperforms the representatives of HAR-RV family models. This is an interesting piece of evidence in that Andersen, Bollerslev, Christoffersen and Diebold (2006) find that the simple HAR model provide much better results than GARCH-type model. They say that the GARCH models only use daily data while HAR models employ more information contained in intraday day. However, the regression of GARCH model and FIGARCH model in this paper is set with intraday return (5 min return) therefore the GARCH model and FIGARCH model also reflect the information which HAR-RV models have.

The following tables present the selected SPA test results. Models comprising jumps in Table 16 and Table 17 are BPV jump models while models containing jumps in Table 18 to Table 19 are MedRV jump models. The loss functions I choose are MSE and MAE. The null hypothesis of SPA test is that the benchmark model is not inferior to other alternative models. The test produces the most significant model, the best model, models with performances of 75%, 50% and 25% relative to the benchmark model and the worst performing model. P-values of the test are illustrated at the bottom of each table.

Table 16. SPA test results evaluated by the MAE and MSE for HAR- CSJd model with BPV jump component

	MAE	MSE	MAE	MSE
	Models		t-statistics	
Benchmark	HAR-CSJd	HAR-CSJd	-	-
Most Significant	FIGARCH	FIGARCH	2.87360	0.45580
Best model	FIGARCH	FIGARCH	2.87360	0.45580
Model_25%	HAR-CJ	HAR-CJ	-4.80494	-4.48272
Median_50%	HAR-RV	HAR-RV	-6.47564	-6.07192
Model_75%	HAR-RV	HAR-RV	-6.47564	-6.07192
Worst model	GARCH	GARCH	-4.14656	-3.34757
SPA test p-value	MAE	MSE		
	0.0000	0.35940		

Notes: Table 16 shows the SPA test results for different models. The benchmark model selected is HAR-CSJd model with BPV jump component. The null hypothesis of the test is that the benchmark model is not inferior to the other candidate models. The test chooses the most significant model, the best model, models with performances of 75%, 50% and 25% relative to the benchmark model, and the worst model. P-values are reported in the last row.

Table 17. SPA test results evaluated by the MAE and MSE for FIGARCH model

	MAE	MSE	MAE	MSE
	Models		t-statistics	
Benchmark	FIGARCH	FIGARCH	-	-
Most Significant	HAR-CJ	HAR-CJ	-4.26269	-1.26024
Best model	GARCH	GARCH	-3.19604	1.90754
Model_25%	HAR-CJ	HAR-CJ	-4.26269	-1.26024
Median_50%	HAR-RV	HAR-RV	-4.67949	-1.57134
Model_75%	HAR-RV	HAR-RV	-4.67949	-1.57134
Worst model	HAR-CSJd	HAR-CSJd	-7.21755	-5.07617
SPA test p-value	MAE	MSE		
	0.47520	0.91600		

Notes: Table 17 shows the SPA test results for different models. The benchmark model selected is FIGARCH model. Candidate HAR series models are with BPV jump component. The null hypothesis of the test is that the benchmark model is not inferior to the other candidate models. The test chooses the most significant model, the best model, models with performances of 75%, 50% and 25% relative to the benchmark model, and the worst model. P-values are reported in the last row.

Table 18. SPA test results evaluated by the MAE and MSE for HAR- CSJd model with MedRV jump component

	MAE	MSE	MAE	MSE
	Models		t-statistics	
Benchmark	HAR-CSJd	HAR-CSJd	-	-
Most Significant	FIGARCH	FIGARCH	-4.65927	-2.75550
Best model	HAR-CJ	HAR-CJ	-3.51678	0.537890
Model_25%	FIGARCH	FIGARCH	-4.65927	-2.75550
Median_50%	HAR-RV	HAR-RV	-6.24035	-4.40787
Model_75%	HAR-RV	HAR-RV	-6.24035	-4.40787
Worst model	GARCH	GARCH	-8.95971	-4.69843
SPA test p-value	MAE	MSE		
	0.51920	0.54940		

Notes: Table 18 shows the SPA test results for different models. The benchmark model selected is HAR-CSJd model with MedRV jump component. The null hypothesis of the test is that the benchmark model is not inferior to the other candidate models. The test chooses the most significant model, the best model, models with performances of 75%, 50% and 25% relative to the benchmark model, and the worst model. P-values are reported in the last row.

Table 19 SPA test results evaluated by the MAE and MSE for GARCH model

	MAE	MSE	MAE	MSE
	Models		t-statistics	
Benchmark	GARCH	GARCH	-	-
Most Significant	HAR-CJ	FIGARCH	8.95971	5.02232
Best model	HAR-CJ	HAR-CJ	8.95971	4.69843
Model_25%	FIGARCH	FIGARCH	7.14033	5.02232
Median_50%	HAR-RV	HAR-RV	5.36427	4.06353
Model_75%	HAR-RV	HAR-RV	5.36427	4.06353
Worst model	HAR-CSJd	HAR-CSJd	3.25576	2.19376
SPA test p-value	MAE	MSE		
	0.00000	0.00000		

Notes: Table 19 shows the SPA test results for different models. The benchmark model selected is GARCH model. Candidate HAR series models are with MedRV jump component. The null hypothesis of the test is that the benchmark model is not inferior to the other candidate models. The test chooses the most significant model, the best model, models with performances of 75%, 50% and 25% relative to the benchmark model, and the worst model. P-values are reported in the last row.

SPA test results indicate that HAR-CSJd is not inferior to other models and FIGARCH model is not inferior to HAR-RV models if the jump components of HAR-RV models are bi-power variation. However, the SPA test performance of GARCH model is not as good as the DM test performance of GARCH model. The null hypothesis that GARCH model is not inferior to other models is rejected due to its low P value regardless of the jump component of HAR-RV models. In a nutshell, the results of forecasting performance between HAR-RV models and GARCH-type models are quite mixed. It indicates that the forecasting performance of GARCH model and FIGARCH model is better than HAR-RV models when it comes to DM test while SPA test results are the other way round.

7. Conclusion

This paper provides the comparison within empirical performance of a series HAR-type models and several GARCH-type models. In-sample analysis indicates that there is no outperforming model while squared jumps help to reduce future's volatility to some extent. MedRV jump is more significant than BPV jump component but their contribution to volatility explanation is limited. The information of the decomposition of variance into semi-variance is mixed. The out-of-sample performance comparison presents the most complicated HAR-type model outperforms other simple HAR-type models while the comparison between GARCH-type models and HAR-type models is inconclusive, which is against Andersen, Bollerslev, Christoffersen, and Diebold (2006, chap. 15), who find that even based on simple autoregressive structures such as the HAR provide much better results than GARCH-type models. The forecast performance contradiction stems from the different data sample periods: the data sample span from March 2009 to March 2013, which is more up-to-date than data utilised the existing literature.

One limit of our study is the comparison criteria we employ are not voluminous. The forecasting performance tests are limited to DM test and SPA test. Stepwise SPA test (Hsu et al., 2010), an improvement on the conservation of SPA test has already been introduced to the literature before the writing of the paper. Stepwise SPA (SSPA) test is not adopted in the thesis, which is a limitation in the paper.

A potential extension of the current study is to study the linkage among different markets and assets based on HAR-type model or apply intraday data to multivariable GARCH models such as DCC model (Engle, 2002) or Correlated ARCH (Christodoulakis & Satchell, 2002).

References

Adrangi, B., Chatrath, A., Dhana, K., & Raffiee, K. (2001). Chaos in oil prices? Evidence from futures markets. Energy Economics, 23(4), 405–425.

Aloui, C., & Marbrouk, S. (2010). Value-at-risk estimations of energy commodities via long-memory, asymmetry and fat-tailed GARCH models. Energy Policy, 38(5), 2326–2339. Greater China Energy: Special Section with regular papers.

Agnolucci, P. (2009). Volatility in crude oil futures: A comparison of the predictive ability of GARCH and implied volatility models. Energy Economics, 31, 316–321.

Andersen, T. G., & Bollerslev, T. (1997). Intraday periodicity and volatility persistence in financial markets. Journal of Empirical Finance, 4, 115–158.

Andersen, T. G., Bollerslev, T., Christoffersen, P. F., & Diebold, F. X. (2006). Volatility and correlation forecasting. In G. Elliott, C. W. J. Granger, & A. Timmermann (Eds.). Handbook of economic forecasting (Vol. 2). Elsevier.

Andersen, T.G., Bollerslev, T., Diebold, F. X., & Elbens, H. (2001). The distribution of stock return volatility. Journal of Financial Economics, 61, 43-76.

Andersen, T. G., Bollerslev, T., & Diebold, F. X. (2007). Roughing it up: Including jump components in the measurement, modeling and forecasting of return volatility. Review of Economics and Statistics, 89, 701–720.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2001). The distribution of exchange rate volatility. Journal of the American Statistical Association, 96, 42–55.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (1999). (Understanding, optimizing, using and forecasting) realized volatility and correlation. Manuscript, Northwestern University, Duke University and University of Pennsylvania. Published in revised form as “Great realizations”. Risk. 105–108.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modelling and forecasting realized volatility. Econometrica, 71, 579–625.

Andersen, T. G., Dobrev, D., & Schaumburg, E. (2012). Jump-robust volatility estimation using nearest neighbor truncation. Journal of Econometrics, 169, 75–93.

Baillie, R.T., Bollerslev, T., & Mikkelsen, H.O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 74, 3–30.

Baillie, R.T., Cecen, A.A, Erkal, C., & Han, Y. (2004). Measuring non-linearity, long memory and self-similarity in high-frequency European exchange rates. International Financial Markets, Institutions and Money, 14, 401–418.

Bandi, F. M., & Russell, J. R. (2005). Separating microstructure noise from volatility. Journal of Financial Economics, 79, 655–692.

Barndorff-Nielsen, O., & Shephard, N. (2004). Power and bipower variation with stochastic volatility and jumps. Journal of Financial Econometrics, 2, 1–37.

Barndorff-Nielsen, O., Kinnebrock, S., & Shephard, N. (2010). Measuring downside risk – Realised semivariance. In T. Bollerslev, J. Russell, & M. Watson (Eds.), Volatility and time series econometrics: Essays in honor of Robert F. Engle. Oxford University Press.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. Journal of Econometrics. 31. 307-327.

Bollerslev, T., & Wright, J. H. (2001). High-frequency data, frequency domain inference, and volatility forecasting. Review of Economics and Statistics, 83, 596–602.

Cabedo, J. D., & Moya, I. (2003). Estimating oil price 'value at risk' using the historical simulation approach. Energy Economics, 25, 239–253.

Charles, A., & Darne O. (2014). Volatility persistence in crude oil markets. Energy Policy. 65, 729–742.

Chen, X., & Ghysels, E. (2011). News – good or bad – and its impact on volatility predictions over multiple horizons. Review of Financial Studies, 24, 46–81.

Chkili, W., Hammoudeh. S., Nguyen, D.K. (2014). Volatility forecasting and risk management for commodity markets in the presence of asymmetry and long memory. Energy Economics, 41, 1–18.

Chortareas, G., Jiang, Y., & Nankervis, J. (2011). Forecasting exchange rate volatility using high-frequency data: Is the euro different? International Journal of Forecasting, 27, 1089-1107.

Christodoulakis, G.A., & Satchell, S.E. (2002). Correlated ARCH (CorrARCH): Modelling the time-varying conditional correlation between financial asset returns. European Journal of Operational Research, 139(2), 351–370.

Corsi, F. (2009). A simple approximate long memory model of realized volatility. Journal of Financial Econometrics, 7, 174–196.

Corsi, F., Pirino, D., & Renò, R. (2010). Threshold bipower variation and the impact of jumps on volatility forecasting. Journal of Econometrics, 159, 276–288.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. Journal of Business & Economic Statistics, 13, 253–263.

Engle, R.F. (2002). Dynamic Conditional Correlation: a New Simple Class of Multivariate GARCH Models, Journal of Business and Economic Statistics, 20, 339-350.

Fong, W. M., & See K. H. (2002). A Markov switching model of the conditional volatility of crude oil futures prices. Energy Economics, 24(1), 71–95.

Hansen, P.R. (2005). A test for superior predictive ability. Journal of Business & Economic Statistics 23, 365–380.

Hol, E., & Koopman, S. J. (2002). Stock index volatility forecasting with high frequency data. Tinbergen Institute Discussion Paper No. 2002-068/4.

Hou, A., & Suardi, S. (2012). A nonparametric GARCH model of crude oil price return volatility. Energy Economics, 34(2), 618–626.

Hsu, P., Hsu, Y., & Kuang, C. (2010). Testing the predictive ability of technical analysis using a new stepwise test without data snooping bias. Journal of Empirical Finance, 17, 471-484.

Huang, X., & Tauchen, G. (2005). The relative price contribution of jumps to total price variance. Journal of Financial Econometrics, 3, 456–499.

Giot, P., & Laurent, S. (2003). Market risk in commodity markets: A VaR approach. Energy Economics, 25, 435–457.

Kang, S.H., Kang, S.M., & Yoon, S.M., (2009). Forecasting volatility of crude oil markets. Energy Economics, 31, 119–125.

Kang, S.H., & Yoon, S.M. (2013). Modelling and forecasting the volatility of petroleum futures prices. Energy Economics, 36, 354–362.

Koopman, S. J., Jungbacker, B., & Hol, E. (2005). Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements. Journal of Empirical Finance, 12, 445–475.

Martens, M. (2001). Forecasting daily exchange rate volatility using intraday returns. Journal of International Money and Finance, 20(1), 1-23.

Martens, M., & Zein, J. (2004). Predicting financial volatility: high-frequency time-series forecasts vis-`a-vis implied volatility. Journal of Futures Markets, 24, 1005–1028.

Mohammadi, H., & Su, L. (2010). International evidence on crude oil price dynamics: Applications of ARIMA-GARCH models. Energy Economics, 32(5), 1001–1008.

Morana, C. (2001). A semiparametric approach to short-term oil price forecasting. Energy Economics, 23(3), 325–338.

Patton, A. J., & Sheppard, K. (2015). Good volatility, bad volatility: Signed jumps and the persistence of volatility. The Review of Economics and Statistics, 97(3), 683-697.

Politis, D.N., & Romano, J.P., (1994). The stationary bootstrap. Journal of the American Statistical Association, 89, 1303–1313.

Pong, S., Shackleton, M., Taylor, S. J., & Xu, X. (2004). Forecasting currency volatility: a comparison of implied volatilities and AR(FI)MA models. Journal of Banking and Finance, 28(9), 2541–2563.

Sadorsky, P. (2006). Modeling and forecasting petroleum futures volatility. Energy Economics, 28, 467–488.

Sadorsky, P., & McKenzie, M. D. (2008). Power transformation models and volatility forecasting. Journal of Forecasting, 27, 587–606.

Sevi, B. (2014). Forecasting the volatility of crude oil futures using intraday data. European Journal of Operational Research, 235, 643–659.

Tseng, T.C., Chung, H., & Huang, C.S. (2009). Modeling jump and continuous components in the volatility of oil futures. Studies in Nonlinear Dynamics & Econometrics, 13(3), 5 (Article 5).

Wei, Y., Wang, Y., & Huang, D. (2010). Forecasting crude oil market volatility: Further evidence using GARCH-class models. Energy Economics, 32, 1477–1484.

West, K. D. (1996). Asymptotic inference about predictive ability. Econometrica, 64, 1067–1084.

Chapter 3. Co-movement Estimation and Volatility Forecasting of Crude Oil Market and US Stock Market: Evidence of MGARCH, Wavelet and High Frequency Data

Abstract

The study of cross market linkage between the crude oil markets and stock markets plays a fundamental role in modern finance background. We examine the relationship between WTI crude oil futures and S&P 500 stock index in the time–frequency space in this paper. The sample period in this paper starts from 8 Oct 2001 9:30 to 30 Oct 2015 16:00. We use the DCC-GARCH and wavelet-based measures of co-movements to find out the relationship between the two financial assets in time and frequency domain features of the data. In the time series domain, intraday data are employed and the performances of intraday data and daily data are compared. A rolling window analysis is utilised to construct out-of-sample one-day-ahead forecast of dynamic conditional expected returns and variances. We find that wavelet method is instrumental to identify the long/short term investment behaviours with the help of daily data and intraday data improve the forecast performance of traditional time series method. The findings of this paper have empirical implications in asset allocation and risk management for investment decisions.

1. Introduction

Studying relationship between different markets, especially between the crude oil market and stock markets is an ongoing issue in the finance literature recently. Tang and Xiong (2012) find that the price co-movements between various commodities after 2004 have greatly increased and that the prices of non-energy commodities have become increasingly correlated with oil prices. Most of the recent work analysing cross markets co-movements has been based on time domain aspect of analysis and ignored frequency domain. Rua and Nunes (2009) claim that the higher strength in the co-movements of stock returns at lower frequencies suggesting higher return from international diversification in the short-term relative to the long-term after studying stock markets. Cross markets movements are complex since it involves different investors with different term objectives. Therefore, the standard time series econometric method which considers the time components only usually loses one side of information. To be more specific, studies based on time series analysis lose frequency aspect information and studies based on frequency domain lose the time aspect information. Compared to previous contribution on cross market co-movements, we follow Uddin et al. (2013) and employ a balance between time and frequency aspect of the data. To be specific, we employ the wavelet approach, which allows us to study the frequency components of time series without losing the time information. This method helps us to discover cross market interactions which remain hidden in econometric methods. Another advantage is that the wavelet analysis approach is model-free. It makes wavelet method a powerful tool to compare with other time series or frequency based estimation methods which are based on estimation methods. Thus the wavelet application in the cross market co-movement can provide insights into changing patterns of cross market co-movements and it enables simultaneous assessment of short term and long term cross market co-movements and detects change in market linkages over time.

The aim of this paper is to examine the strength of the co-movement between crude oil market and the US stock in the time and frequency space by resorting to wavelet analysis using daily data and intraday data from Oct 8th 2001 9:30 to Oct 30 2015 16:00 and make forecasting evaluation of DCC-GARCH model under different time frequencies. The wavelet method results are relatively easy to interpret and offer considerable amount of information on co-movements and lead-lag relationships of the two markets containing time as well as frequency domain information. Moreover, the wavelet approach allows to evaluate co-movements across different investment horizons and to distinguish between short term and long term investors.

To relate our findings to the standard econometric literature, we connect our approach with the standard econometric approach of (Engle, 2002) dynamic conditional correlations from a multivariate DCC-GARCH model in order to measure the co-movements between the crude oil market and the US stock market. The research makes contribution to the literature in three aspects by detecting 1. Whether wavelet method fits the intraday data; 2. Whether traditional time series method fits intraday data well; 3. Do high-frequency data improve the forecasting performance of traditional time series method? We make contribution to the literature by answering the three questions above.

The paper proceeds as follows. Section 2 gives the literature review on 1. Cross market co-movements of crude oil market and stock markets and 2. Volatility forecast on financial assets. Section 3 documents the data. Section 4 explains the methodology, while Section 5 discusses empirical results. The final section discusses the key findings and concludes.

2. Literature Review

2.1. Cross market co-movements of crude oil market and stock markets

The discussion and analyse of the relationship between oil market and US stock market have been well documented in the literature. Hamilton (1983, 1985, 2003) in his seminal articles illustrates exogenous oil supply shocks may be a reason for recessions and periods of low economic growth. Several studies have analysed the relationship between oil and stock market since then. Huang et al. (1996) and Jones and Kaul (1996) are pioneers to explore the relationship by using empirical methods. Huang et al. (1996) investigate the dynamic interactions between futures prices of crude oil traded in the New York Mercantile Exchange (NYMEX) and US stock prices and they find that the return volatility spill-over from oil futures to stocks is very weak. On the contrary, Jones and Kaul (1996) find that US stock prices react significantly to oil shocks.

Present researchers use different methodologies, different data frequencies and different proxies for oil market and US stock market to detect the relationship. Recent studies concentrating on the linkage between the oil market and the US stock market include Hammoudeh et al. (2004), Kilian and Park (2009), Balcilar and Ozdemir (2012), Elyasiani et al. (2012), Fan and Jahan-Parvar (2012), Alsalman and Herrera (2013), Mollick and Assefa (2013), Conrad et al. (2014), Kang et al. (2014), Khalfaoui et al. (2015) and Salisu and Oloko (2015).

The main methodologies used in the literature are VAR type models and GARCH type models. A bunch of the researches model the linkage between oil market and the US stock market by employing VAR type model. (Kilian and Park (2009), Balcilar and Ozdemir (2012), Fan and Jahan-Parvar (2012), Alsalman and Herrera (2013), Kang et al. (2014)). Kilian and Park (2009) use a structural vector auto regression (SVAR) model on monthly data covering from January 1973 to December 2006. The aggregate US stock return they utilise is constructed from monthly returns on the Centre for Research in Security Prices (CRSP) value-weighted market portfolio and the oil price is based on the US refiner's acquisition cost of crude oil, as reported by the US Department of Energy. These variables are employed in real terms by deflating them with the US consumer price index (CPI). They find that the response of aggregate US real stock returns differ greatly depending on whether the increase in the price of crude oil is driven by demand or supply shocks in the crude oil market. They show that positive shocks to the global demand for industrial commodities cause both higher real oil prices and higher stock prices, which helps explain the resilience of the US stock market to the recent surge in the price of oil. They also find that oil demand and oil supply shocks combined account for 22% of the long-run variation in US real stock returns. Following the work of Kilian and Park (2009), Kang et al. (2014) also utilize an SVAR model to investigate how the demand and supply shocks driving the global crude oil market affect US bond market returns and they use monthly data covering the period from January 1982 to December 2011. They follow Kilian and Park (2009) method to compute the real oil price but they use bond market instead of stock market. the US bond return were constructed from an index of US aggregate bond holdings and the real aggregate US bond return was measured by deflating its nominal term by the US CPI. Contrary to the findings of Kilian and Park (2009), they find that a positive oil market-specific demand shock is associated with significant decreases in US bond returns. In addition, their evidence shows that the demand and supply shocks driving the global crude oil market jointly account for 30.6% of the long run variation in US real bond returns. Balcilar and Ozdemir (2013) consider monthly data from February 1990 to July 2011 and they employ a Markov switching vector autoregressive (MS–VAR) model. They divide S&P500 index into different sub-groups such as Industry, Energy, Energy Equipment & Services, Oil and Gas and Consumable fuels, Oil and Gas Exploration and Production, Oil and Gas Storage and Transportation indexes. The oil futures price is used as a proxy for oil price. They do not find any lead–lag Granger causality, but the results based on the MS–VAR model clearly show that oil futures price has strong regime prediction power for a sub-grouping of S&P 500 stock index during various sub-periods in the sample, while there is a weak evidence for the regime prediction power of a sub-grouping of S&P 500 stock indexes for oil futures. Fan and Jahan-Parvar (2012) employ WTI spot and NYMEX light sweet crude futures prices for oil price while the US stock returns were computed from average monthly value weighted returns on forty nine US industry level portfolios composing of NYSE, AMEX, and NASDAQ stocks. They employ both linear regression model and vector autoregressive (VAR) model with monthly data from January 1979 to January 2009. They find that oil–price predictability is concentrated in relatively small number of industries. Alsalman and Herrera (2013) estimate a simultaneous equation model, which is a VAR model in essence, comprising of symmetric and asymmetric responses of stock returns to positive and negative oil price shocks by using monthly data from January 1973 to December 2009. Excess returns of all NYSE, AMEX, and NASDAQ stocks are used as proxies of US stock market and US composite refiners' acquisition cost as the proxy of crude oil market. Their in-sample evidence suggests that the increase of oil price helps to forecast aggregate US stock returns as well as industry-level returns one-year ahead.

The studies using GARCH-type model to illustrate the connection between oil market and the US stock market include Hammoudeh et al. (2004), Elyasiani et al. (2011), Mollick and Assefa (2013), Conrad et al. (2014), Salisu and Oloko (2015). Hammoudeh et al. (2004) use two US markets of oil prices: the WTI spot and 1- to 4-month NYMEX futures prices and the proxies for the US Stocks are the S&P oil sector stock indices which include Oil Exploration and Production, Oil & Gas Refining & Marketing, Oil-Domestic Integrated, Oil-International Integrated, and the overall Oil Composite. They employ both univariate and multivariate ARCH/GARCH models with daily data for the period July 17, 1995 to October 10, 2001. They find that there are bi-directional interactions between the US oil stock returns and the spot return and the futures return of crude oil. Elyasiani et al. (2011) employ daily data from 11 December 1998 to 29 December 2006. The data they use are NYMEX crude oil futures and thirteen industry sectors market portfolio of NYSE, AMEX, and NASDAQ stocks. Using the ARCH and GARCH models, they find strong evidence to support the idea that oil price volatility contributes to a systematic risk at the industry level as nine of the thirteen sectors in question show significant relationships between oil-futures return distribution and industry excess return. These industries are affected by oil futures returns or oil futures return volatility, either or both. Mollick and Assefa (2013) employ the GARCH and MGARCH–DCC models using daily data from January 1999 to December 2011. They use S&P 500, Dow Jones, NASDAQ, and Russell 2000 indexes returns as proxies for US stock returns and WTI for oil price. They find that US stock returns are slightly and negatively affected by oil prices and by the exchange rate (USD/Euro) before the financial crisis. However, from mid- 2009 onwards, the stock returns are documented to be positively affected by oil prices and a weaker USD/Euro. Conrad et al. (2014) use a modified Dynamic Conditional Correlations–Mixed Data Sampling (DCC–MIDAS) specification proposed in Colacito et al. (2011) and further extended by Engle et al. (2013) to explore the relationship between the US stock market and crude oil market. They employ the daily returns of the CRSP value-weighted portfolio, which is based on all NYSE, AMEX and NASDAQ stocks and WTI oil data covering from January 1993 to November 2011. They find that variables that contain information on current and future economic activity are able to predictors of changes in the oil–US stock correlation. Salisu and Oloko (2015) use ARMA (1, 1)-BEKK-AGARCH (1, 1) model to model the relationship between crude oil market and US stock market. They use Daily data of Brent and WTI crude oil price and S&P 500 stock from 1 Feb. 2002 to 4 Apr. 2014. Their empirical evidence suggests a significant positive return spillover from US stock market to oil crude market and bi-directional shock spillovers between the two markets. Both markets illustrate asymmetric volatility effect and volatility spillover from oil market to stock market become more pronounced after a structural break which coincides with the time of global economic slowdown.

New method is introduced into the literature on the linkage of oil market and stock market recently apart from the two mainstream methodologies mentioned above. Khalfaoui et al. (2015) introduce a new approach incorporating both multivariate GARCH models and wavelet analysis: wavelet-based MGARCH approach. By using daily oil price and daily stock market indices of G7 countries spanning from 2 June, 2003 to 7 February, 2012, they investigate the spill-over effects of volatility and shocks between oil prices and the G-7 stock markets. Equipped with a wavelet-based GARCH–BEKK approach, they find strong evidence of time-varying volatility in all markets. Oil price and stock market prices are directly affected by their own news and volatilities and indirectly affected by the volatilities of other prices and wavelet scale. The results show also, that mean and volatility spillover effects were decomposed into many sub-spillovers on different time scales according to heterogeneous investors and market participants. Moreover, hedging ratios vary across scales. Recent papers introduce wavelet approaches to identify the relationships between stock markets and oil markets. Reboredo and River-Castro (2014) examine the relationship between oil and stock markets in Europe and the USA at the aggregate and sectoral levels using wavelet multi-resolution analysis. They find evidence of contagion and positive interdependence between these markets after 2008. Martín-Barragán et al. (2015) investigate the impact of oil shocks and stock market crashes on correlations between stock and oil markets and they also find evidence of contagion, in particular during the 2008 and 2011 stock market falls which supports the results from Reboredo and River-Castro (2014). Madaleno and Pinho (2014) find the relationship between oil prices and sector stock returns is ambiguous and that that long run market dynamics are more uncertain.

Other studies investigating the crude oil market and non-US stock market include Park and Ratti (2008), Arouri et al. (2011, 2012) and Wang et al. (2013).

Like the studies on the US, different methodologies such as vector autoregressive (VAR) model, vector error-correction model (VECM), univariate and multivariate GARCH-type models including the BEKK (Baba, Engle, Kraft and Kroner over parameterization), CCC (Constant Conditional Correlation) and DCC (Dynamic Conditional Correlation) are applied to non-US cases. Arouri et al. (2011) employ VAR (1)–GARCH(1,1) for stock markets in the Gulf Cooperation Council (GCC) countries and Arouri et al. (2012) employ the same model for the stock markets in Europe. Wang et al. (2013) use structural VAR model examine the relationship between oil prices and stock Oil price shocks and stock market activities between oil-importing and oil-exporting countries.

In a nutshell, various empirical studies suggest that the choice of methodology, proxies of variables and country characters may affect the linkage between crude oil and stock market (Kilian and Park (2009), Balcilar and Ozdemir (2013), Fan and Jahan-Parvar (2012), Alsalman and Herrera (2013), Mollick and Assefa (2013), Conrad et al. (2014), Kang et al. (2014)) or daily data (Hammoudeh et al. (2004), Elyasiani et al. (2011), Salisu and Oloko (2015), Khalfaoui et al. (2015)). To the knowledge of the author, there is no empirical paper studying the linkage between crude oil and stock market with high frequency data or intraday data. This study fills the gap in the existing literature.

2. 2. Volatility forecast on financial assets.

Since the true volatility is unobservable, daily squared returns are often used as a proxy measure of volatility. By using 5 min data as a new volatility measure, Andersen and Bollerslev (1998) demonstrate a dramatic improvement in the volatility forecasting performance of a daily GARCH model (foreign exchange). Since then, a great number of studies have focused on realized volatility forecasting and its properties. Andersen, Bollerslev, Diebold, and Labys (ABDL, 1999 and 2001) recommend forecasting the realised volatility by using the ARFIMA model and show that the realised volatility is a consistent estimator of the integrated volatility. ABDL (2001) show that if realised volatility is modelled directly by a parametric model rather than simply being used in the evaluation of other models' forecasting behaviours, the realised volatility can improve forecasting when it comes to the ARFIMA model on foreign exchange rates. The findings above make contribution to the empirical basis of using the realised volatility in volatility forecasting directly.

Kang et al. (2009) investigate the efficacy of volatility models for three crude oil markets — Brent, Dubai, and West Texas Intermediate (WTI) — with regard to its ability to forecast and identify volatility stylized facts, in particular volatility persistence or long memory. The data they use are three crude oil spot prices (in US dollars per barrel) obtained from the Bloomberg databases. The datasets consist of daily closing prices over the period from January 6, 1992 to December 29, 2006, and the last one year's data are used to evaluate out-of-sample volatility forecasts. They assess persistence in the volatility of the three crude oil prices using conditional volatility models. The CGARCH and FIGARCH models are better equipped to capture persistence than are the GARCH and IGARCH models. The CGARCH and FIGARCH models also provide superior performance in out-of-sample volatility forecasts. They conclude that the CGARCH and FIGARCH models are useful for modelling and forecasting persistence in the volatility of crude oil prices. Wei et al. (2010) extend the work of Kang et al. (2009). They use a number of linear and nonlinear GARCH models to capture the volatility features of two crude oil markets: Brent and WTI. They also carry out superior predictive ability test (SPA test) and other loss functions to evaluate the forecasting power of different models. They use daily price data (in US dollars per barrel) of Brent and WTI from 6/1/1992 to 31/12/2009.

Mohammadi and Su (2010) examine the usefulness of several ARIMA-GARCH models for modeling and forecasting the conditional mean and volatility of weekly crude oil spot prices in eleven international markets over the 1/2/1997–10/3/2009 period with weekly data. In particular, they investigate the out-of-sample forecasting performance of four volatility models — GARCH, EGARCH and APARCH and FIGARCH over January 2009 to October 2009. Forecasting results are somewhat mixed, but in most cases, the APARCH model outperforms the others. Also, conditional standard deviation captures the volatility in oil returns better than the traditional conditional variance. Finally, shocks to conditional volatility dissipate at an exponential rate, which is consistent with the covariance-stationary GARCH models than the slow hyperbolic rate implied by the FIGARCH alternative.

3. Data Description

The original data we obtain are 5 min price data of the crude oil futures and S&P 500 index. The NYMEX light, sweet crude oil futures contract data is provided by Tick Data and the intraday data of S&P 500 by Pi Trading. Crude oil futures is the world's most actively traded commodity, and the NYMEX light, sweet (low-sulphur) crude oil (WTI) futures contract is the world's most liquid crude oil futures, as well as the world's largest-volume futures contract trading on a physical commodity. S&P 500 index is one of the most commonly followed equity indices, and many consider it one of the best representations of the U.S. stock market, and a barometer for the U.S. economy. The time span in this study is from 8th Oct 2001 9:30 to 30 Oct 2015 16:00, containing 3524 trading days.

High frequency data contain more information on financial assets. Theoretically, the higher the frequency of the data, the more accurate the volatility estimation will be. While on the other hand, microstructure frictions, such as price discreteness and measurement errors may affect the effectiveness of high frequency data. We follow Andersen and Bollerslev (1998) who demonstrate a dramatic improvement in the volatility forecasting performance of a daily GARCH model. Bandi and Russell also (2005) propose a rule for the calculation of the optimal sampling frequency for the realised volatility. They suggest that for the US stock market, the optimal frequencies vary between 0.4 and 13.8 min. We employ 5 minute data in this paper which lies in the optimal frequency interval.

NYMEX light, sweet (low-sulphur) crude oil futures has open outcry trading from 9:00 to 14:30 EST on weekdays. Investors can also trade oil futures via NYMEX electronic trading platform from 17:00 on Sunday to 17:15 the next day and from 18:00 to 17:15 (New York Time) on weekdays. The trading volumes on weekends are rather small therefore we remove weekend returns from the sample following the common practice in the literature (Chortareas et al. 2011; Celik & Ergin 2014). I obtain 264878 observations in total after the data is cleared. The daily data is used as a comparison.

The intraday return series $r_{t,m}$ is given as follow:

$$r_{t,m} = \ln(P_{t,m}) - \ln(P_{t,m-1}) \quad (1)$$

Where $P_{t,m}$ is the close-mid price at the mth time stamp on day t. Figure 1 and 2 show the intraday prices of crude oil futures and S&P 500 index respectively.

The daily return r_t is given as follows:

$$r_t = \ln(P_t) - \ln(P_{t-1}) \quad (2)$$

Figure 1 and 2 illustrate the prices of crude oil futures and S&P 500 index respectively. Figure 3 and 4 indicate the return series of intraday crude oil and stock market respectively. Figure 5 and 6 show the return series of daily crude oil and stock market respectively.

Figure 1. Time series plot of crude oil futures

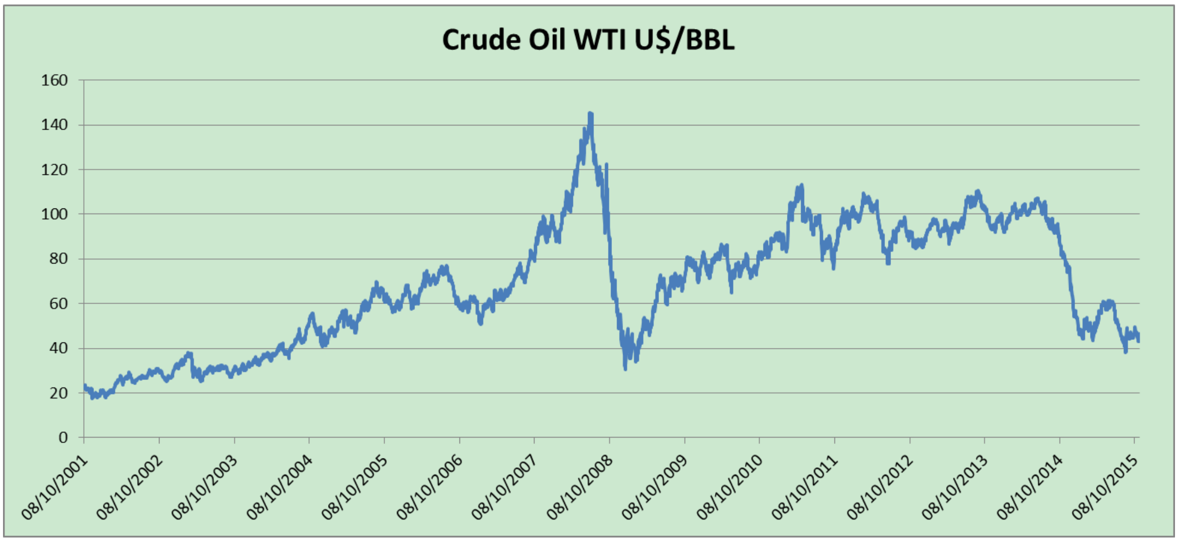


Figure 2. Time series plot of S&P 500 index

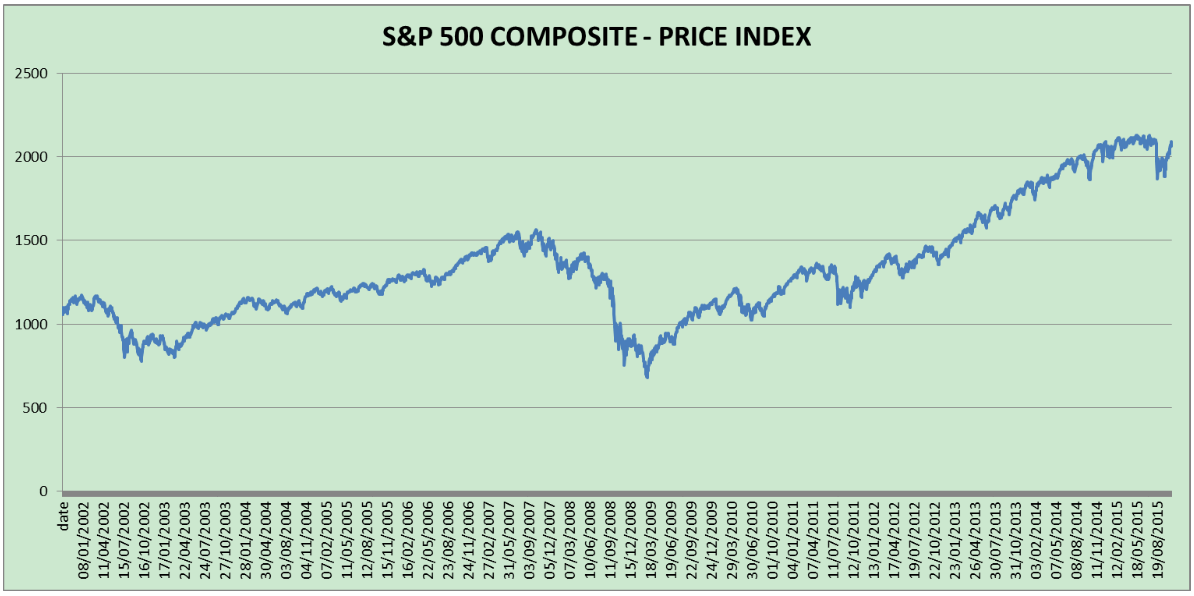


Figure 3. Graph of intraday returns of crude oil

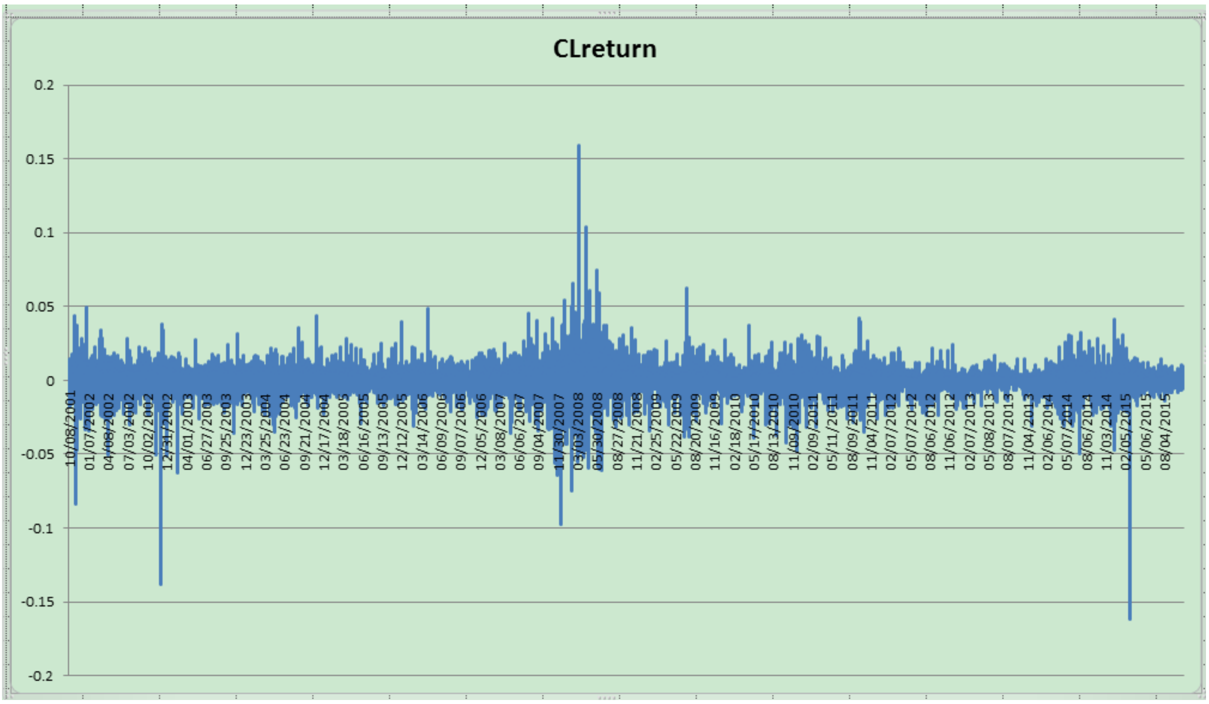


Figure 4. Graph of intraday returns of stock market

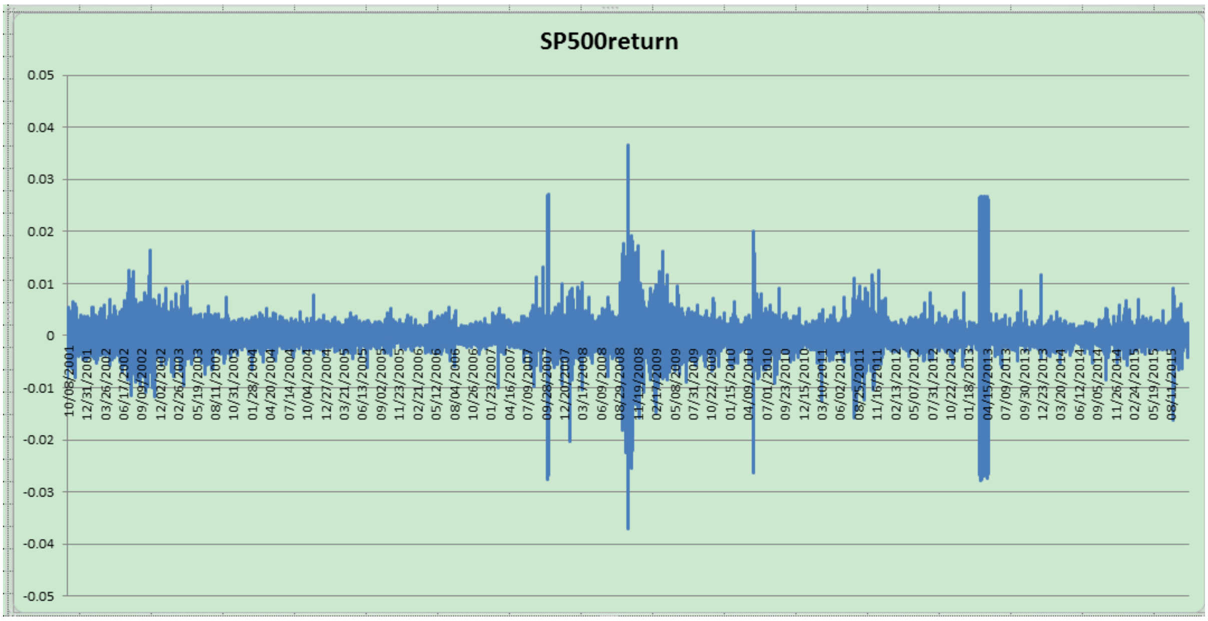


Figure 5. Graph of daily returns of crude oil

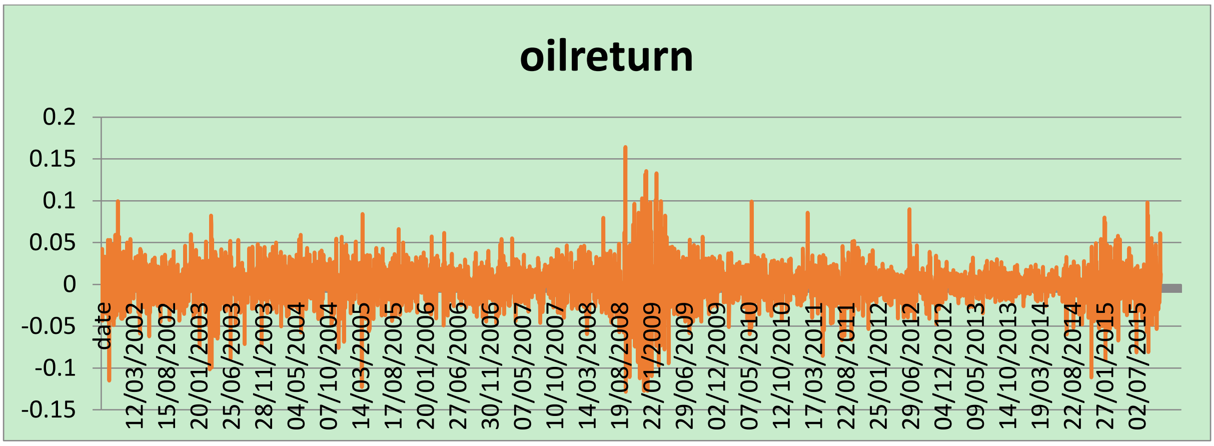


Figure 6. Graph of daily returns of stock market

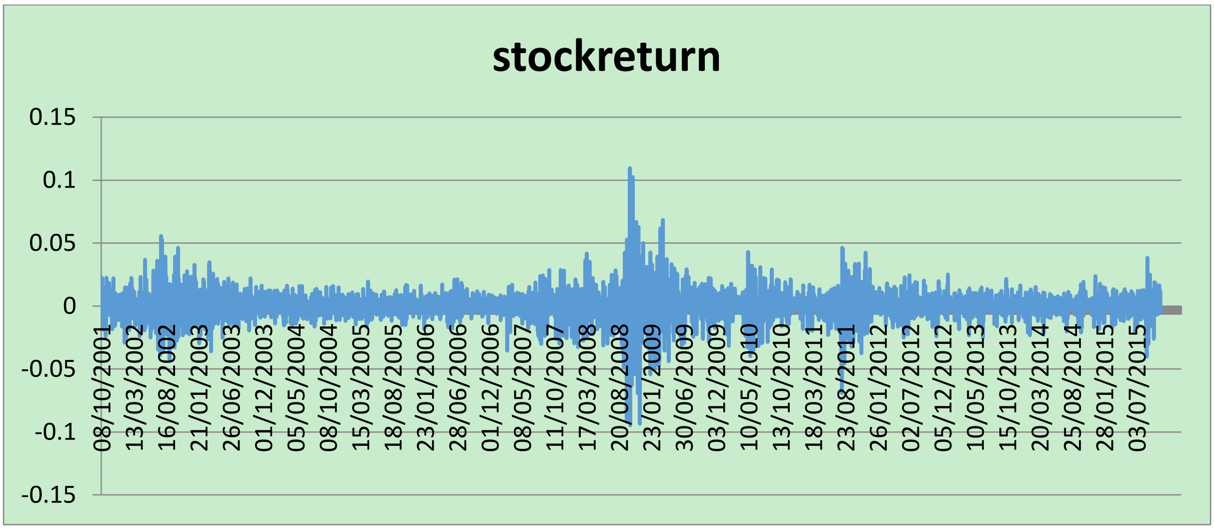


Table 1. Descriptive statistics of intraday return series

	Crude Oil	S&P 500
Mean	0.00000267	0.00000168
Median	0.00000000	0.00000759
S.D	0.002586	0.00129
Skewness	-1.070631	0.007958
Kurtosis	213.85	119.1156
Jarque- Bera	50700000	15400000
Jarque- Bera Probability	0.0000	0.0000
Observations	273591	273591
ADF	-305.476	-273.95

Table 2. Unconditional correlation for intraday returns

correlation	Crude Oil	S&P 500
Crude Oil	1	
S&P 500	-0.000276 (-0.144296)	1

Notes: t ratio is reported in the round brackets.

Table 3. Descriptive statistics of daily returns

	Crude Oil	S&P 500
Mean	0.0002	0.000181
Median	0.0000	0.000348
S.D	0.0236	0.01228
Skewness	-0.0975	-0.221472
Kurtosis	7.4910	12.7275
Jarque- Bera	3089.97	14499.59
Jarque- Bera Probability	0.0000	0.0000
Observations	3523	3523
ADF	-33.087	-31.194

Table 4. Unconditional correlation for daily returns

correlation	Crude Oil	S&P 500
Crude Oil	1	
S&P 500	0.233198 (14.52387)	1

Notes: t ratio is reported in the round brackets.

Figure 7. The autocorrelation function of the 5 minute returns of crude oil (15 lags)

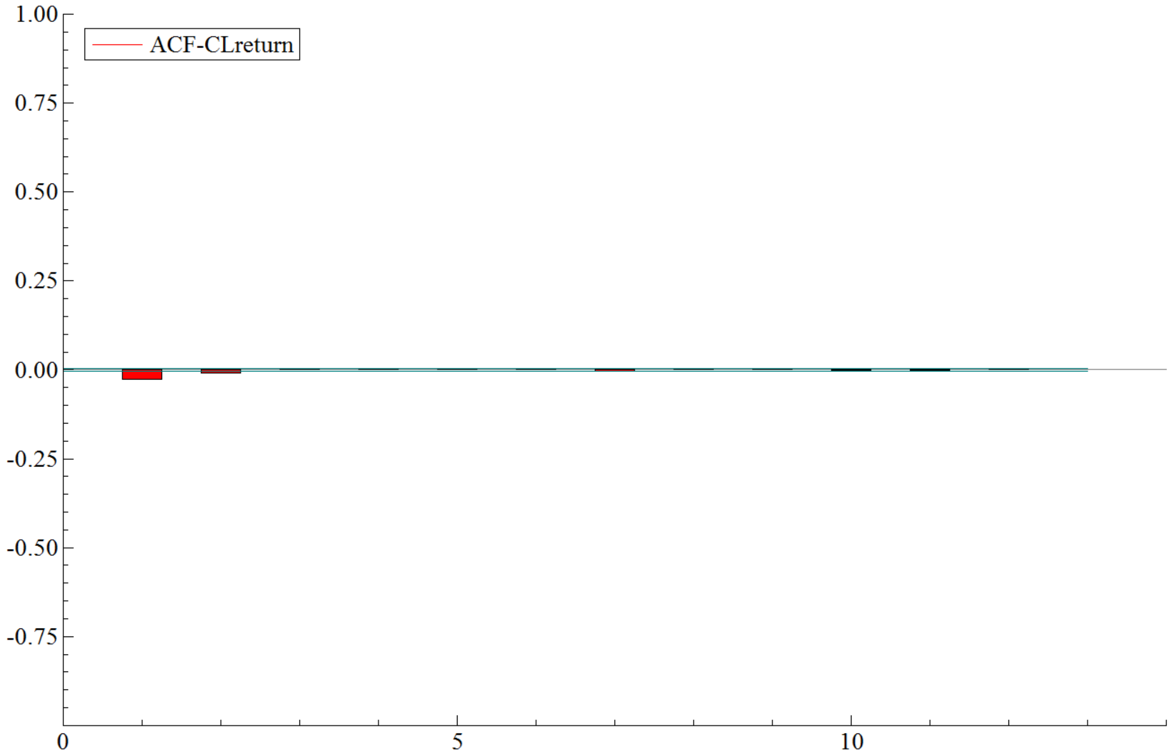


Figure 8. The autocorrelation function of absolute 5 min returns for crude oil futures for 300 lags.

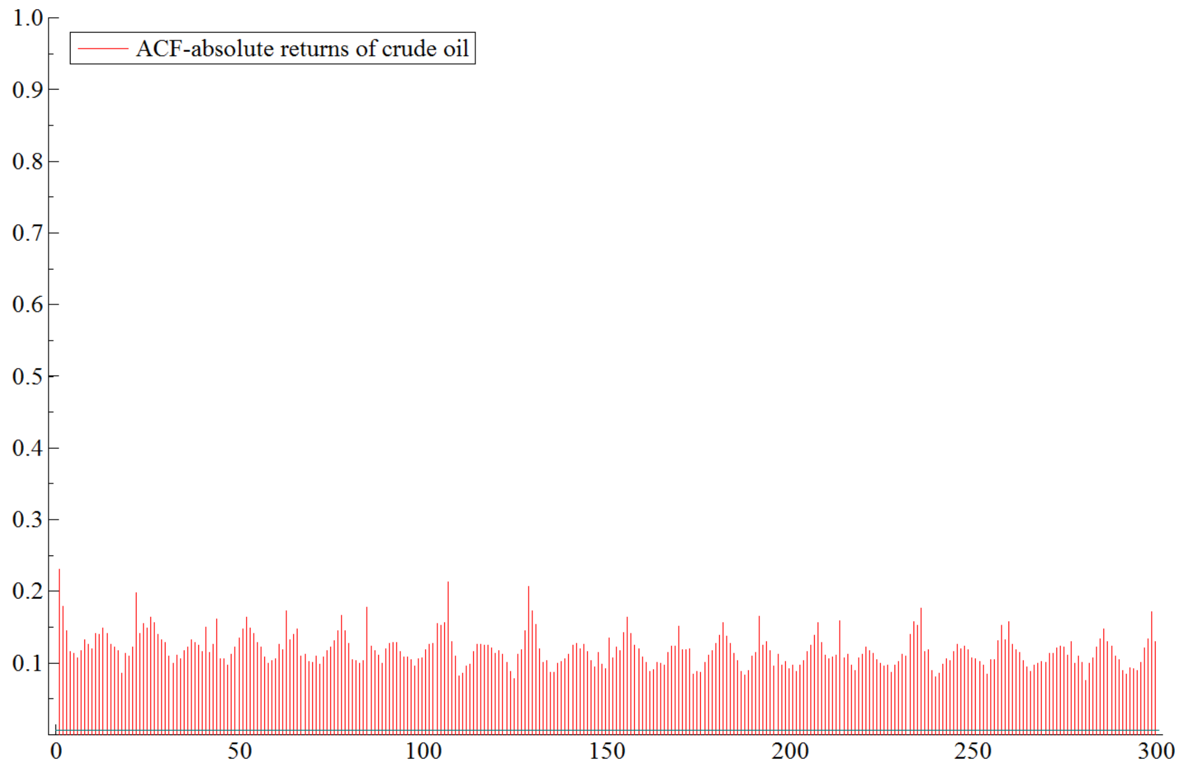


Figure 9. The autocorrelation function of the 5 minute returns of S&P 500 index (20 lags)

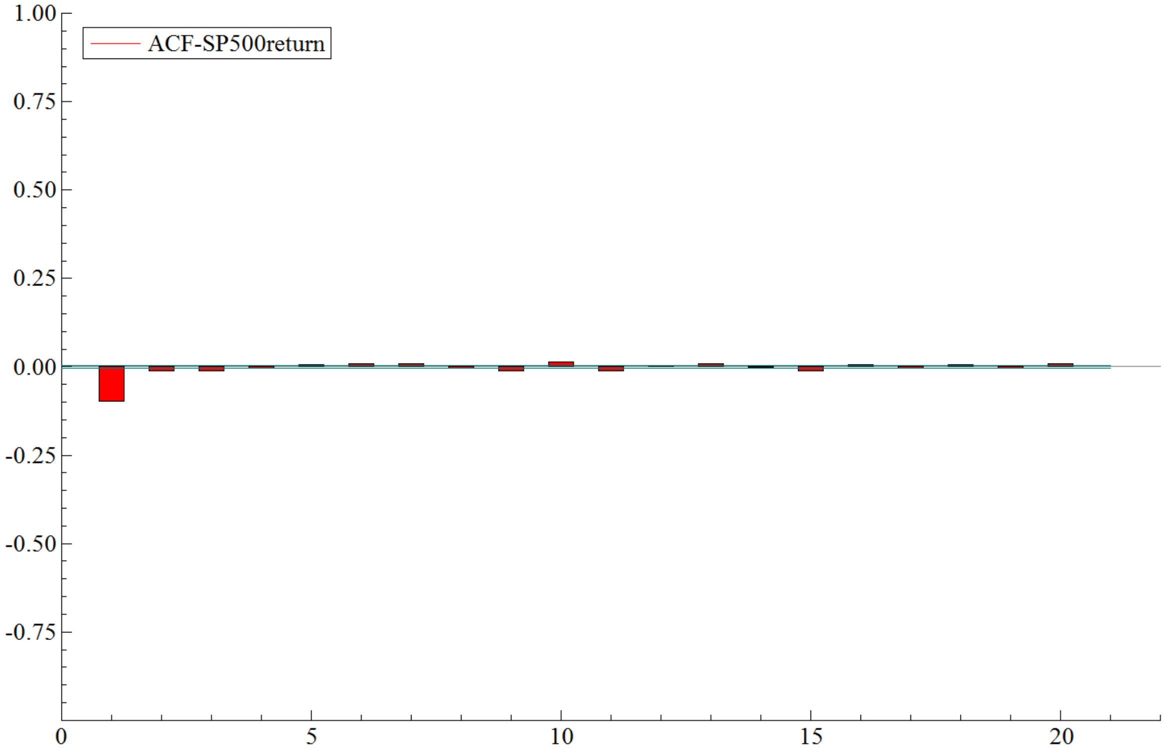
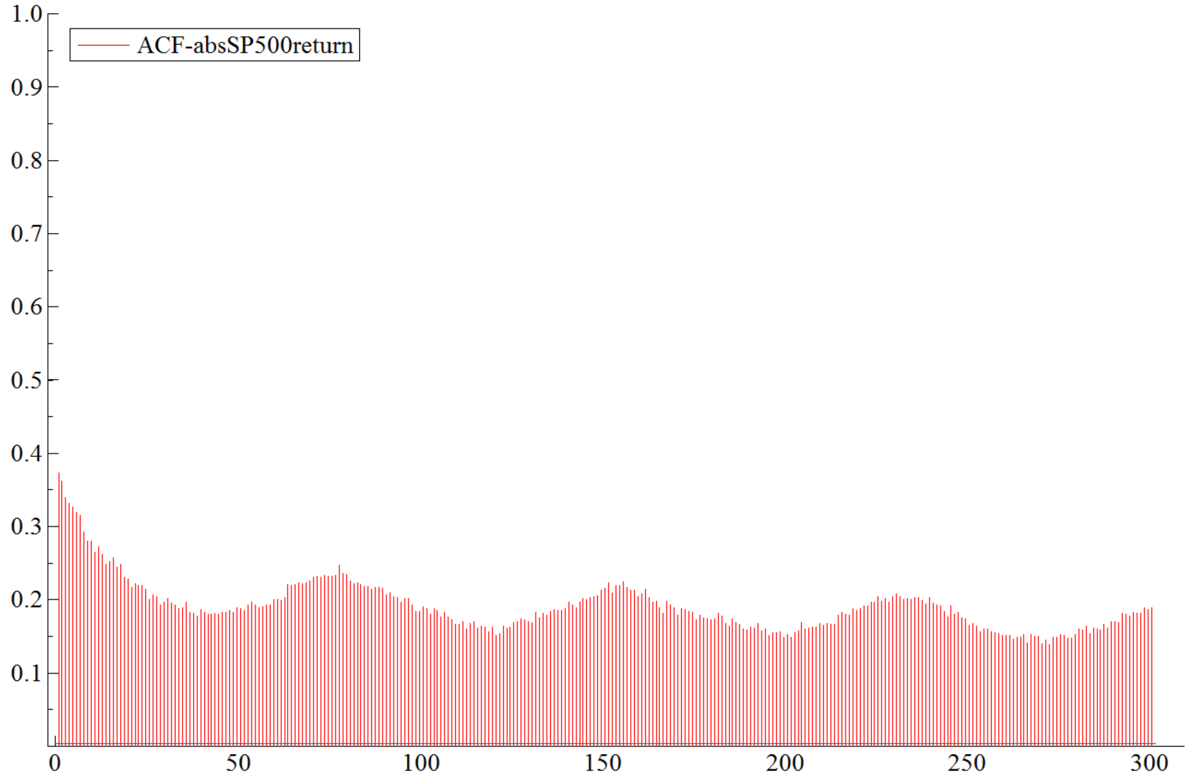


Figure 10. The autocorrelation function of absolute 5 min returns for S&P 500 futures for 300 lags.



From Table 1 we can summarise that intraday returns document extremely high kurtosis. The skewness of crude oil return is slightly negative while the stock market are positively skewed. The Jarque-Bera tests statistic on the two markets strongly reject the normal distribution hypothesis. The descriptive statistics of the daily data are different from that of intraday data. The kurtosis values of daily data from the two markets are comparatively lower than those from the intraday data of the two markets respectively and the two markets are negatively skewed under daily observation. The Jarque-Bera tests statistic on the two markets strongly reject the normal distribution hypothesis under daily observation as well. The distinct of unconditional correlations between 5 min observations (Table 2) and daily observations (Table 4) is also large. A small negative unconditional correlation (-0.000276) is observed under 5 min observation while a positive unconditional correlation (0.233198) is documented under daily observation.

The extremely large negative values of ADF test results indicate that all the return series do not reject the null hypothesis that there is no unit root in the series.

Figure 3 to 6 show that the movements of the 5 min returns and the daily returns are not consistent. High-frequency data carry more information thus several jumps in the daily returns are smoothed out in the 5 min returns.

Dacorogna et al. (2001) find that a well-documented stylised fact of high-frequency returns which is the negative first order autocorrelation in the return. Figure 7 and 9 indicate the autocorrelation function of the 5 min return series of crude oil and S&P 500 index respectively. The first order autocorrelation of the 5 min returns of crude oil and S&P 500 index are negative, which is consistent with the literature (Goodhart, 1989; Goodhart and Figliuoli, 1992; Goodhart et al. 1995). Literature tells that a large negative autocorrelation is followed by rather small autocorrelations in the subsequent lags which is caused by the bounce between the bid and ask prices. However, for the crude oil return, the first order autocorrelation is not large enough to dominate the subsequent lags. The coefficients of autocorrelations in the subsequent lags are close to zero and the P-values of the Q-stat are almost zero for the following 12 lags thus the null hypothesis of no autocorrelation for 12 lags cannot be rejected. For the stock market return, the first order autocorrelation is large enough to dominate the subsequent lags.

Periodicity is another stylised fact of intraday volatility series. Figure 8 and 10 show the autocorrelation function of absolute returns for crude oil futures and stock market respectively.

The U-shaped plot reveals the periodicity in a trading day. One could clearly read that crude oil has more U-turns than stock market for the same time lag number. There is no sign of disappearance of autocorrelation in the absolute returns in Figure 8 and 10.

In brief, the return series of the 5 min crude oil and stock market in my study share the stylised facts of high frequency financial returns well documented in the literature. It has a zero mean while it is fat tailed and marginally skewed. The return series of two assets exhibit negative first order autocorrelation and it reveals that periodicity pattern exists in intraday volatility of two assets.

4. Methodology

In this section we discuss the econometric methodology we will employ in our research work. First approach is DCC GARCH model. The second approach we use is wavelet analysis. In wavelet analysis we will rely on wavelet power spectrum, cross-wavelet analysis, wavelet coherency and phase differences. The wavelet power spectrum demonstrates the volatilities and spikes in the data series; cross-wavelet analysis can be interpreted as co-variance of time series analysis; wavelet coherency can be interpreted as correlation in the time series analysis; and phase difference provide the evidence of lead-lag relationship. Our both approaches has similarity in the sense that they show time-varying correlation over period of time. DCC GARCH approach shows time-varying correlation over period of time in two dimensions while cross-wavelet approach shows the same in three dimensions. The difference of the two approaches lies that DCC GARCH approach provides a single correlation coefficient for a point of time while wavelet coherency approach computes several correlation coefficients for a point of time by varying frequencies.

4.1. Modelling dynamic conditional correlation

The volatilities of intraday returns have a strong periodicity in 1-day interval, which is demonstrated in the previous section. Martens et al. (2002) suggest that intraday periodic patterns do not fit the traditional time series models, (e.g., GARCH-type models) directly because the GARCH-type model are easily distorted by the pattern. Thus, we use the de-seasonalised filtered returns to estimate GARCH-type models instead of the original returns directly. According to Taylor and Xu (1997), we have

$$\tilde{r}_{t,n} = \frac{r_{t,n}}{S_{t,n}} \quad (n = 1,2, \dots, N) \tag{3}$$

where $r_{t,n}$ is the nth intraday return on day t and $S_{t,n}$ is the corresponding seasonality term, for N intraday periods. $S_{t,n}$ is equal to the averaging the squared returns for each intraday period:

$$S_{t,n}^2 = \frac{1}{T}\sum_{t=1}^{T} r_{t,n}^2 \quad (n = 1,2,\dots,N) \tag{4}$$

where T is the number of days in the sample. It's an effective method to smooth the seasonality feature so we use the de-seasonalised returns in the following part of the paper.

The intraday return series is nearly symmetric and has a high kurtosis thus I assume the returns series follows the symmetric student T distribution while for the symmetric student T distribution,

$$E\left|z_{t,n-1}\right| = 2\frac{\Gamma\left(\frac{1+v}{2}\right)\sqrt{v-2}}{\sqrt{\pi}\Gamma(v/2)} \tag{5}$$

where v indicates the degree of freedom of the student T distribution and $\Gamma(.)$ is the Gama function.

Most previous works assessing cross-market time-varying correlation employ the multivariate DCC model developed by Engle (2002). This model is suitable to assess co-movements between the markets we study because it allows us to infer the cross-market conditional correlations straightforwardly.

Assume that stock market returns from the k series are multivariate student T distributed with zero mean and conditional variance-covariance matrix H_t, our multivariate DCC-GARCH model for intraday data can be presented as follows:

$$\begin{cases} \tilde{r}_{t,n} = \mu_t + \varepsilon_t,\ \varepsilon_t | I_{t-1} \sim T_v(0, H_t) \\ H_t = D_t R_t D_t \end{cases} \tag{6}$$

The DCC-GARCH model for daily data is expressed as:

$$\begin{cases} r_t = \mu_t + \varepsilon_t,\ \varepsilon_t | I_{t-1} \sim T_v(0, H_t) \\ H_t = D_t R_t D_t \end{cases} \tag{7}$$

where r_t and $\tilde{r}_{t,n}$ are the $(k \times 1)$ vector of the returns; ε_t is a $(k \times 1)$ vector of zero mean return innovations conditional on the information available at time $t-1$; $\mu_{i,t} = \beta_{i0} + \beta_{i1} r_{i,t-1}$ for market i, D_t is a $(k \times k)$ diagonal matrix with elements on its main diagonal being the conditional standard deviations of the returns on each market in the sample and $R_t$is the $(k \times$

k) conditional correlation matrix. D_t and R_t are defined as follows:

$$D_t = diag(h_{11,t}^{1/2}, \dots, h_{kk,t}^{1/2}) \quad (8)$$

Where $h_{ii,t}$ is chosen to be a univariate GARCH (1, 1) process;

$$R_t = (diagQ_t)^{-1/2} Q_t (diagQ_t)^{-1/2} \quad (9)$$

Where $Q_t = (1-\alpha-\beta)\bar{Q} + \alpha u_{t-1}u'_{t-1} + \beta Q_{t-1}$ indicates a $(k \times k)$ symmetric positive definite matrix with $u_{i,t} = \varepsilon_{i,t}/\sqrt{h_{ii,t}}$, $\bar{Q}$ is the $(k \times k)$ unconditional variance matrix of u_t and α and β are nonnegative scalar parameters satisfying $\alpha + \beta < 1$.

The equation of the conditional correlation coefficient ρ_{ij} between two markets i and j is given as follows:

$$\rho_{ij} = \frac{(1-\alpha-\beta)\bar{q}_{ij} + \alpha u_{i,t-1}u_{j,t-1} + \beta q_{ij,t-1}}{\sqrt{(1-\alpha-\beta)\bar{q}_{ij} + \alpha u_{i,t-1}^2 + \beta q_{ii,t-1}}\sqrt{(1-\alpha-\beta)\bar{q}_{ij} + \alpha u_{j,t-1}^2 + \beta q_{jj,t-1}}} \quad (10)$$

q_{ij} indicates the element located in the ith row and jth column of the symmetric positive definite matrix Q_t. The two-stage procedure is employed to estimate the regression output of the DCC-GARCH model. Univariate GARCH (1, 1) model is estimated for each market in the first stage and the standardised residuals obtained from the first stage are used to estimate the conditional correlations.

The log-likelihood function is expressed as follows:

$$L = -\frac{1}{2}\sum_{t=1}^{T}[nlog(2\pi) + 2log|D_t| + \log(R_t) + u'_t R_t^{-1} u_t] \quad (11)$$

4.2 Wavelet method

A wavelet is a function with zero mean and that is localised in both frequency and time. A wavelet can be characterised by how localised it is in time Δt and frequency $\Delta\omega$ or the bandwidth). One particular wavelet, the Morlet, is defined as follows:

$$\psi_0(\eta) = \pi^{-1/4} e^{i\omega_0\eta} e^{-\eta^2/2} \quad (12)$$

Where ω_0 and η are dimensionless frequency and time respectively. To maintain a good balance e between time and frequency localisation, the frequency parameter ω_0 is set to be 6 (see Foufoula-Georgiou, 1995; Grinsted, 2004; Rua and Nunes, 2009).

The wavelet function is applied to be a bandpass filter to the time series. The continuous wavelet transform of a time series x_n, $(n = 1,2,\ldots,N)$ with uniform time steps δt is defined as the convolution of x_n with the scaled and normalized wavelet (see Grinsted, 2004). The equation is written as follows:

$$W_n^X(s) = \sqrt{\frac{\delta t}{s}} \sum_{n'=1}^{N} x_{n'} \psi_0 [(n' - n)\frac{\delta t}{s}] \tag{13}$$

The term $|W_n^X(s)|^2$ is defined as wavelet power. The complex argument of $W_n^X(s)$ is interpreted as the local phase.

The wavelet coherency measure (WTC) is used to reveal how coherent the cross wavelet transform is in time frequency space and measure the extent of synchronisation of a pair of time series. Following Torrence and Webster (1998) we define the wavelet coherence of two time series as

$$R_n^2(s) = \frac{\left|S(s^{-1}W_n^{XY}(s))\right|^2}{S(s^{-1}\left|W_n^X(s)\right|^2)\cdot S(s^{-1}\left|W_n^Y(s)\right|^2)} \tag{14}$$

Where S is a smoothing operator written as:

$$S(W) = S_{scale}(S_{time}(W_s(s))) \tag{15}$$

where S_{scale} denotes smoothing along the wavelet scale axis and S_{time} smoothing in time. It is natural to design the smoothing operator so that it has a similar footprint as the wavelet used. For the Morlet wavelet a suitable smoothing operator is given by Torrence and Webster (1998)

$$S_{time}(W)|_s = \left(W_s(s) \cdot c_1^{\frac{-t^2}{2s^2}}\right)|_s, \tag{16}$$

$$S_{time}(W)|_s = (W_s(s) \cdot c_2 \Pi(0.6s))|_s \tag{17}$$

where c_1 and c_2 are normalization constants and Π is the rectangle function. The factor of 0.6 is the empirically determined scale decorrelation length for the Morlet wavelet (Torrence and

Compo, 1998). The numerator is the absolute value squared of the smoothed cross-wavelet specturum and denominator represents the smoothed wavelet power spectra (Torrence and Webster 1999; Rua and Nunes 2009). The definition of $R_n^2(s)$ closely resembles that of a traditional correlation coefficient, and it is useful to think of the wavelet coherency as a localised correlation coefficient in time frequency space. The value of $R_n^2(s)$ gives a quantity between 0 and unity, and the higher the value indicates higher co-movement between two markets. The robustness of this approach is that it enables us to identify area of co-movement between two series in the time frequency space and significance of the wavelet coherence measured by the Monte Carlo simulation methods (Torrence and Compo (1998). In this paper, we will employ the Wavelet Coherency measure, instead of the Wavelet Cross Spectrum employed by Aguiar-Conraria et al. (2008).

4.3. Forecast

We employ 15 min data to detect the forecasting performance of the two assets. The whole sample data observations are from 8 Oct 2001 9:30 to 30 Oct 2015 16:00 and we divide the whole sample into two subgroups: the in-sample data for volatility modelling covering from 8 Oct 2001, to 4 Jun 2015, and the out-of-sample data for model evaluation is from 5 Jun 2015, to 30 Oct 2015, covering 100 trading days and containing 3200 observations (There are 32 observations within a day for 15 min data). We use a rolling window method and produce one-step ahead daily volatility forecasts for daily models and 32-step-ahead intraday volatility forecasts for intraday models. This procedure is repeated 100 times in order to produce 100 daily volatility forecasts for evaluation out-of-sample. The rolling window estimation requires adding one new observation and dropping the most distant one therefore the sample size employed in estimating the models remains fixed and the forecasts do not overlap.

Considering the true volatility is unobservable, we follow Andersen and Bollerslev (1998) and use a realised volatility series constructed from 5 min returns which we use in the model estimation part as a proxy for the true volatility, i.e.,

$$\sigma_{rv,m}^2 = \sum_{m=1}^{M} r_{t,m}^2 \tag{18}$$

where $r_{t,n}^2$ are 5 min interval squared returns and $\sigma_{rv,t}^2$ is the realised variance on day t.

Actual volatility (variance) is assessed using the squared returns and denoted as σ_t^2. The volatility forecast obtained by using a GARCH-class model is indicated by $\hat{\sigma}_t^2$. Various forecasting criteria or loss functions can be considered to assess the predictive accuracy of a volatility model. However it is not obvious which loss function is more appropriate for the evaluation of volatility models. Hence, rather than making a single choice we use the following 9 different loss functions as forecasting criteria:

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(\sigma_t^2 - \hat{\sigma}_t^2)^2 \tag{19}$$

$$\text{MedSE} = Median(\sigma_t^2 - \hat{\sigma}_t^2)^2 \tag{20}$$

$$ME = \frac{1}{n}\sum_{t=1}^{n}(\sigma_t^2 - \hat{\sigma}_t^2) \tag{21}$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|\sigma_t^2 - \hat{\sigma}_t^2| \tag{22}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(\sigma_t^2 - \hat{\sigma}_t^2)^2} \tag{23}$$

where n is the number of forecasting data. The 5 loss functions are Mean Squared Error (MSE), Median Squared Error (MedSE), Mean Error (ME), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE. Additional discussion of these criteria can be found in Brooks, Burke, and Persand (1997) for more details about these measures.

The actual volatility σ_t^2 is set to be realised volatility $r_{t,n}^2$ from 5 min data and volatility forecast obtained $\hat{\sigma}_t^2$ for a single day is the realised volatility obtained from the 15 min data. Volatility forecast from daily data is conditional volatility obtained from daily DCC-GARCH model. We also compare the mean forecast. For the mean forecast performance of 15 min data, the actual mean series are the 15 min returns we employ in the paper and the forecasted values are obtained from the rolling window estimation procedure. We also compare the real daily returns and the forecasted values from the one-day ahead rolling-window procedure.

5. Empirical Finds and Analysis

5.1. The Empirical Findings of Wavelet analysis

We present results of wavelet coherency and phase-relationship between crude oil market and the US stock market in this part. To report the results of the dynamics of cross-market return co-movement obtained by applying the cross-wavelet coherency approach we refer to multi-colour graphs. The vertical axis represents the frequency and for intraday data, the unit is 5 minutes. The time is depicted in horizontal axis (5 min as well for intraday return). For the daily return output, the vertical and horizontal axes are still indicating frequency and time respectively but the unit is changed to 1 day. The following figures present the estimated wavelet coherency and phase relationship between crude oil and the US stock market. The significance is obtained via Monte Carlo simulations. Contours denote wavelet-squared coherency, the thick black contour is the 5% significance level and outside of the thin line is the boundary affected zone.

Figure 11. Wavelet coherency of crude oil market and the US stock market by intraday data

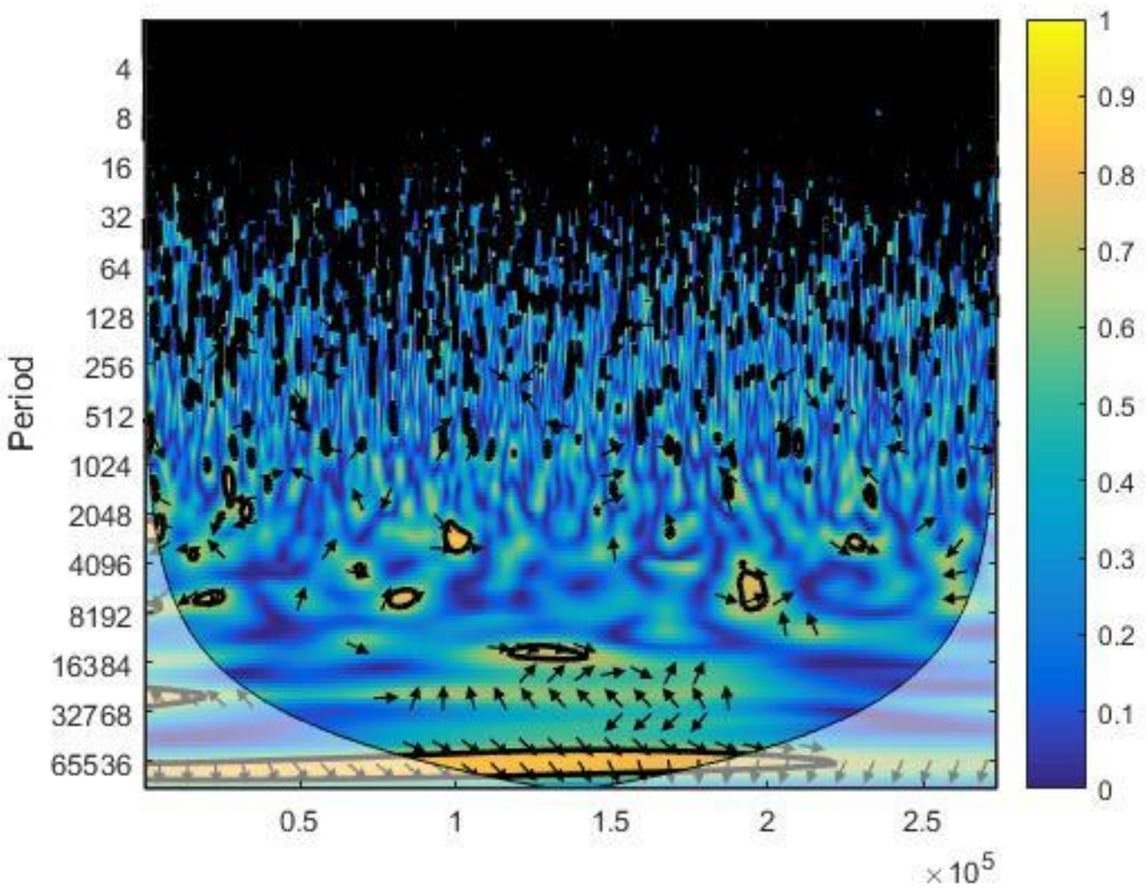


Figure 12. Wavelet coherency of crude oil market and the US stock market by daily data

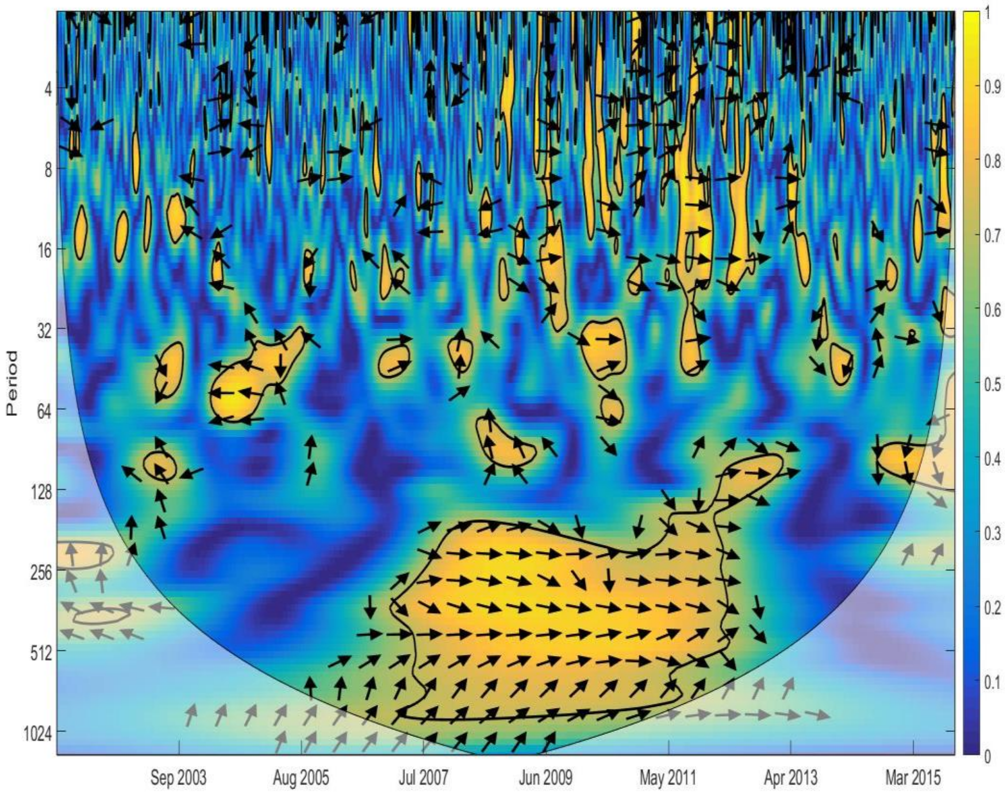


Figure 11. presents the Cross-wavelet coherency of co-movement between the crude oil market and S&P 500 stock index via intraday data and Figure 12. Documents the Cross-wavelet coherency of co-movement between the crude oil market and S&P 500 stock index via daily data. The thick blue contour designates the 95% confidence level estimated from Monte Carlo simulations using phase randomised surrogate series. The (downward pointing) cone of influence indicates the region affected by the edge effect due to finite-length time series (See Torrence and Compo "A Practical Guide to Wavelet Analysis"). The colour code for power ranges from blue (low correlation in volatility) to yellow (high correlation in volatility). The phase difference between the two series is indicated by arrows. Arrows pointing to the right mean that the returns of the two markets are in phase. Arrows pointing to the right –down indicate that crude oil is leading the co-movement towards the stock market and the arrows pointing to the right –up indicate that the stock market is leading the co-movement. Arrows pointing to the left illustrate that the two variables are out of phase. Arrows pointing to the left-up indicate that the crude oil market is leading co-movement towards the stock market and the arrows pointing to the left- down indicate that the co-movement of the stock market is leading. The in-phase situation suggests that the two variables are having cyclical effect on each other and an out-of-phase situation shows that variable having anti-cyclical effect on each other. The following table provides a concise reading of arrows mentioned above.

Table 5. The reading of arrows in cross-wavelet coherency figures

Arrow direction	Interpretation
Arrow pointing to the right	In-phase (cyclical effect) in two markets
Arrows pointing to the right –down	The crude oil is leading the S&P 500 stock
arrow pointing to the right – up	The S&P 500 is leading the crude oil
Arrow pointing to the left	Out-of-phase (anti-cyclical effect) in two markets
Arrow pointing to the left-up	The crude oil is leading the S&P 500 stock
arrow pointing to the left-down	The S&P 500 is leading the crude oil

For intraday data, we could read that the wavelet coherency is only large at highest period (a yellow tape-shape at the bottom of Figure 11.) For the significant part within the cone, all the arrows are pointing down-right, indicating that the crude oil is leading the S&P 500 stock. However, since only a small tape-shape exists, the result implies that the wavelet coherency may not fit intraday data.

Information from Figure 12 on the phases shows us that the relationship between the two markets is not homogeneous across scales/periods because it clearly documents that arrows are pointing left and right, up and down. Moreover, the cross wavelet coherency is high at low frequencies/large periods and coherence is not statistically significant at the highest scale/smallest period. The multi-colour settings of cross wavelet coherency provide us a method to detect areas of varying co-movement among return series over time across frequencies. Strong co-movement in the time-frequency space suggests the fail of diversification.

5.2. The Empirical Findings of DCC-GARCH model

Table 6. illustrate the regression results for 5 min data on the full-sample data. The diagnostic tests on the standardised residuals of DCC-GARCH models are displayed in Table 7.

Table 6. The DCC-GARCH regression estimation results on intraday data

	Crude oil	S&P 500
β_0	0.0000 (0.1511)	0.0000 (1.698)
β_1	-0.01596*** (-5.718)	0.01214*** (3.031)
ω_0	2.06*10^(-10) (1.347)	8.519*10^(-9)*** (3.151)
ω_1	0.00126*** (16.97)	0.11544*** (7.885)
ω_2	0.99873*** (8790)	0.889122*** (67.38)
α		0.001269 (0.9817)
β		0.859617*** (13.04)
ρ		0.003258 (1.31)
D.o.f		3.625*** (349.9)

Notes: P values are given in brackets. Statistical significance at 10%, 5% and 1% are highlighted by *, ** and *** respectively.

Table 7. Tests on the standardised residuals (intraday)

	Skewness	Kurtosis	Jarque-Bera	Q(50)	$Q^2(50)$
Crude Oil	-0.99598 (28.763)	43.589 (586.20)	34411 (1702)	46.5694 [0.6118]	7.08599 [0.9999]
S&P 500	-0.31274 (9.0316)	12.224 (176.55)	31219 (2796)	31.9065 [0.9783]	25.9035 [0.9981]
AIC	-20.523520				
H(50)	191.684 [0.6510783]				
$H^2(50)$	183.630 [0.7601109]				
Li(50)	191.729 [0.6502248]				
$Li^2(50)$	183.738 [0.7583382]				

Notes: T ratio are given in round brackets. Ljung-Box test for autocorrelation of order of 50 on standardised residuals and squared standardised residuals are reported as $Q(50)$ and $Q^2(50)$ respectively. Akaike Information Criteria is reported as AIC. Hosking's multivariate Portmanteau statistics on standardised residuals and squared standardised residuals with the order of 50 are reported as $H(50)$ and $H^2(50)$ respectively. Li and McLeod's multivariate Portmanteau statistics on standardised residuals and squared standardised residuals with the order of 50 are reported as $Li(50)$ and $Li(50)$ respectively. P values of the corresponding tests are documented in square brackets.

Table 6. documents the parameter estimation of DCC-GARCH model on intraday data. The values of the coefficients prior to the lagged return terms in the mean equations are small but the coefficients are significant. All the coefficients in the variance equations but the constant term for crude oil are significant. The coefficients ω_1 representing the ARCH effect in the two equations respectively are negligible while coefficients ω_2 representing GARCH effect are quite large. The summation of ARCH and GARCH coefficients α and β is less than 1 for the variance and covariance equations, meeting the stationary conditions for the MGARCH model. The average time-varying conditional correlation ρ is close to zero and insignificant while the degree of freedom on the T distribution in the regression is significant.

Table 7. contains the results of tests on the standardised residuals from DCC-GARCH model on intraday data. The skewness of the standardised residuals from the model are slightly negative and the kurtosis of the standardised residuals are way less than that of the return series of the two financial assets. The Jarque-Bera test result on the standardised residuals indicates that the unconditional distribution of the standardised residuals do not pass the test. Lung-Box statistics are calculated to test the autocorrelation issue on the standardised residuals and squared standardised residuals with the lag of 50. The high values of probability in the parenthesis do not reject the no-autocorrealtion hypothesis. The robust tests on the standardised residuals suggest that the DCC-GARCH model specification in the paper is able to describe the dynamics of the conditional covariance matrix.

Table 8. The regression results on daily data

	Crude oil	S&P 500
β_0	0.0005 (1.487)	0.0005*** (4.076)
β_1	-0.0496*** (-2.664)	-0.0605*** (-3.649)
ω_0	3.287*10^(-6) (2.064)	1.7007*10^(-6)*** (3.706)
ω_1	0.0528*** (4.359)	0.089199*** (7.620)
ω_2	0.9423*** (68.96)	0.895797*** (70.55)
α		0.025019*** (2.869)
β		0.970737*** (86.61)
ρ		0.215843*** (2.239)
D.o.f		7.04*** (13.82)

Notes: P values are given in brackets. Statistical significance at 10%, 5% and 1% are highlighted by *, ** and *** respectively.

Table 9. Robust tests on the standardised residuals

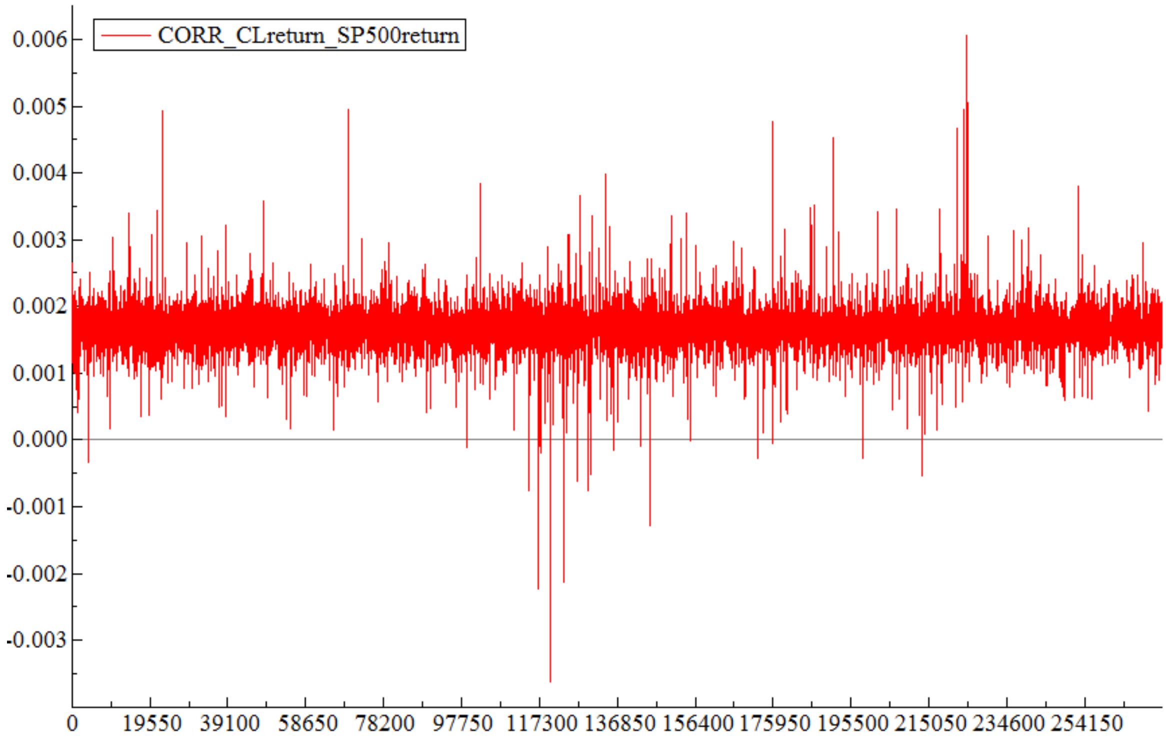
	Skewness	Kurtosis	Jarque-Bera	Q(50)	$Q^2(50)$
Crude Oil	-0.14623 (3.6179)	4.8424 (22.798)	532.12 (2453)	42.4854 [0.7658350]	70.8812 [0.0276024]
S&P 500	-0.36899 (9.1295)	4.2991 (16.076)	341.36 (5478)	64.7970 [0.0778162]	42.5710 [0.7628986]
AIC	-11.489503				
H(50)	215.983 [0.2084192]				
$H^2(50)$	227.995 [0.0707293]				
Li(50)	215.943 [0.2089626]				
$Li^2(50)$	228.042 [0.0704391]				

Notes: T ratio are given in round brackets. Ljung-Box test for autocorrelation of order of 50 on standardised residuals and squared standardised residuals are reported as $Q(50)$ and $Q^2(50)$ respectively. Akaike Information Criteria is reported as AIC. Hosking's multivariate Portmanteau statistics on standardised residuals and squared standardised residuals with the order of 50 are reported as $H(50)$ and $H^2(50)$ respectively. Li and McLeod's multivariate Portmanteau statistics on standardised residuals and squared standardised residuals with the order of 50 are reported as $Li(50)$ and $Li(50)$ respectively. P values of the corresponding tests are documented in square brackets.

Table 8. documents the parameter estimation of DCC-GARCH model on daily data. Similar to the results on intraday data, the values of the coefficients prior to the lagged return terms in the mean equations are small but the coefficients are significant. All the coefficients in the variance equations but the constant term for crude oil are significant. The coefficients ω_1 representing the ARCH effect in the two equations respectively are negligible while coefficients ω_2 representing GARCH effect are quite large. The summation of ARCH and GARCH coefficients α and β is less than 1 for the variance and covariance equations, meeting the stationary conditions for the MGARCH model. What is contrary to the regression on the intraday data is the conditional correlation. The average time-varying conditional correlation for daily data is 0.22 and significant while that for intraday data is 0.003 and insignificant. The existing literature does not have a clear explanation for the correlation difference between the performance of DCC-GARCH model on intraday data and daily data.

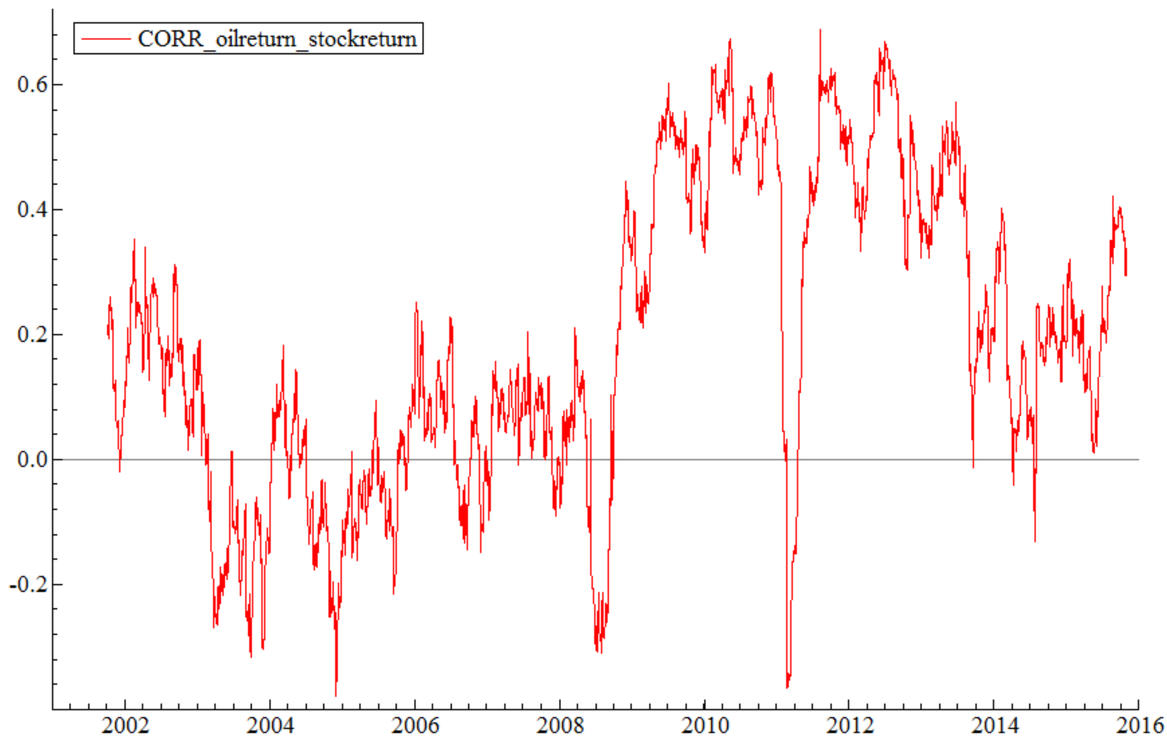
Table 9. contains the results of tests on the standardised residuals from DCC-GARCH model on daily data. The skewness of the standardised residuals from the model are slightly negative and the kurtosis of the standardised residuals are less than that of the return series of the two financial assets. The Jarque-Bera test result on the standardised residuals indicates that the unconditional distribution of the standardised residuals do not pass the test. Lung-Box statistics are calculated to test the autocorrelation issue on the standardised residuals and squared standardised residuals with the lag of 50. The high values of probability in the parenthesis do not reject the no-autocorrelation hypothesis. The robust tests on the standardised residuals suggest that the DCC-GARCH model specification in the paper is able to describe the dynamics of the conditional covariance matrix.

Figure 13. Dynamic conditional correlations of the two series via intraday returns



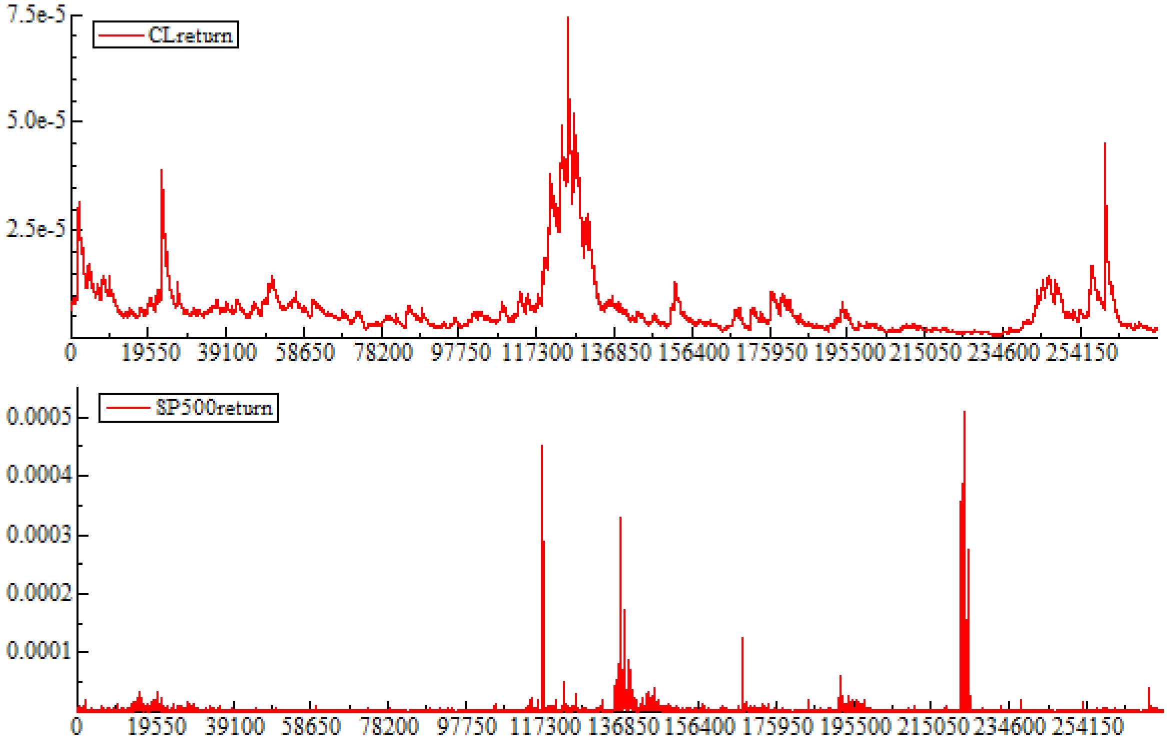
Notes: This figure plots the dynamic conditional correlation between crude oil market and stock market. The horizontal axis indicates the observation number.

Figure 14. Dynamic conditional correlations of the two series via daily returns



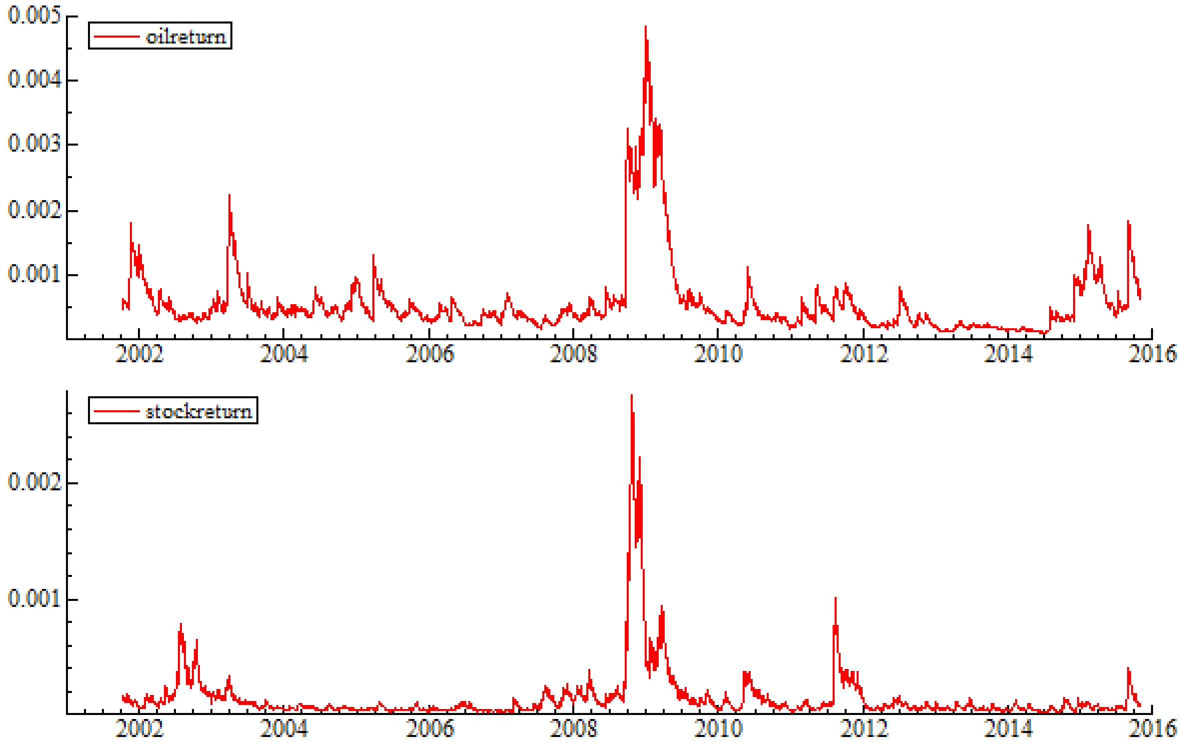
Notes: This figure plots the dynamic conditional correlation between crude oil market and stock market. The horizontal axis indicates the year span.

Figure 15. Dynamic conditional volatility of intraday crude oil and stock market



Notes: This figure plots the volatilities of crude oil market and stock market respectively. The horizontal axis indicates the observation number.

Figure 16. Dynamic conditional volatility of daily crude oil and stock market



Notes: This figure plots the volatilities of crude oil market and stock market respectively. The horizontal axis indicates the year span.

Figure 13 and 14 illustrate the visual information on dynamic conditional correlations and dynamic conditional volatilities of crude oil market and stock market. From the figures one could read the difference of correlation between intraday data and daily data which is similar to the table output in Table 6 and 8. The dynamic conditional correlation for the intraday data is near zero and the peak value is less than 0.006 and the largest negative value is larger than -0.004. For intraday data, the dynamic correlation does not move dramatically. On the contrary, the dynamic conditional correlation obtained from daily data is more volatile than that obtained from intraday data. The two markets indicated by the daily data co-move in the same direction for the most of the time while the two markets move in a different direction in the year 2003, 2005 and 2008. The correlation for the two markets are negative for a short time in the year 2011 and 2014 respectively. The findings here are similar to the findings obtained by using wavelet coherency method. After employing the wavelet coherency method on intraday data, we are able to read that the wavelet coherency is not large or significant for short periods and it is only significant and large for some large periods. The overall performance of wavelet coherency method on intraday data is not significant and the values from wavelet coherency method are small which becomes the mirror image of correlation analysis. The daily data performances for correlation analysis and wavelet coherency are consistent as well. The correlation obtained from DCC-GARCH model is significant and the average value of the time-varying correlation cannot be ignored. The wavelet coherency result on daily data indicates that the coherency is large and significant for large period from July 2007 to 2012. The only difference is that in the wavelet coherency analysis, all the arrows are pointing to the right (right-up and right-down) in the large contour close to the cone at the bottom which suggests that the two financial assets move "in phase", i.e. in the same direction. But we are able to read from the dynamic correlation analysis from the DCC-GARCH model that the correlation is negative for the two financial assets in the year of 2008 and in the first quarter of year 2011 dramatically. It is due to the fact that correlation is a 2-dimension measurement measuring how strong two variables are connected in the time domain while the wavelet coherency analysis is a 3-dimension measurement measuring the relationship of two variables not only in time domain but frequency domain. In the figure depicting the wavelet coherency relationship between stock market and crude oil market in daily data frequency, at the top of the figure there are arrows pointing left at short periods while at the same time there exist arrows pointing right at the bottom for long periods in the same figure. This is the advantage of using wavelet coherency analysis because it distinguish the long period feature and short period feature of the connections of two variables.

Figure 15 and Figure 16 depict the dynamic conditional volatilities of the two financial assets under two different time frequencies. Figure 15 documents the volatilities of crude oil and stock market under 5-min data and Figure 16 documents those of crude oil and stock market under daily data. By reading the two figures we are able to find that the volatility trends of crude oil are similar but the peaks do not occur at the same time. The volatility peaks occurring in the crude oil intraday data prior to the peaks in the crude oil daily data from 2002 to 2013. One example is that the largest volatility peak of crude oil intraday data occurs at 14 Jun 2008 while the extreme volatility peak of crude oil daily data occurs at 4 Jan 2009. We are able to conclude a rule of thumb that from 2002 to 2013, the volatility peaks of crude oil would be documented in intraday data 6-8 months ahead of the peaks in daily data. However, this rule of thumb is no longer instrumental after 2014 when the volatility graphs are not alike for intraday data and daily data. Moreover, the volatility scale of intraday data is way less than that of daily data. It is clear to explain because we divide the daily time span into several 5 min span therefore for each time observation in intraday data, the volatility scale is less than the volatility scale for each time observation in daily data.

The volatility figures for stock market in two different time frequencies tell different stories in

Figure 15 and 16. For intraday data, the volatility is quite stable and close to zero for each time stamp and it has 5 extreme peaks in the time span we study. Daily volatility is more volatile than the intraday counterpart and the peaks in daily volatility are not synchronised with intraday volatility peaks except Sep 2008 and Jul 2011. For other cases, there are no clear lead-lag or synchronisation phenomenon for the stock market volatility for two different time frequencies for the same sample length.

6. Forecast evaluation

We evaluate the forecasting performance of DCC-GARCH model in this part. We follow Chortareas et al. (2011), employ 15 min data to detect the forecasting performance of the two assets. The whole sample data observations are from 8 Oct 2001 9:30 to 30 Oct 2015 16:00 and we divide the whole sample into two subgroups: the in-sample data for volatility modelling covering from 8 Oct 2001, to 4 Jun 2015, and the out-of-sample data for model evaluation is from 5 Jun 2015, to 30 Oct 2015, covering 100 trading days and containing 3200 observations (There are 32 observations within a day for 15 min data). The rolling window method is employed and we produce one-step ahead daily volatility forecasts for daily models and 32-step-ahead intraday volatility forecasts for intraday models. This procedure is repeated 100 times in order to produce 100 daily volatility forecasts for evaluation out-of-sample.

For DCC-GARCH with intraday data, the actual volatility σ_t^2 is set to be realised volatility $r_{t,n}^2$ from 5 min data and volatility forecast obtained $\hat{\sigma}_t^2$ for a single day is the realised volatility obtained from the 15 min data. Volatility forecast from daily data is conditional volatility obtained from daily DCC-GARCH model. We also compare the mean forecast. For the mean forecast performance of 15 min data, the actual mean series are the 15 min returns we employ in the paper and the forecasted values are obtained from the rolling window estimation procedure. We also compare the real daily returns and the forecasted values from the one-day ahead rolling-window procedure. Therefore, by employing the realised volatility from 15 min data and daily volatility from daily data, we are able to compare the volatility forecast performance of DCC-GARCH model among different data frequencies since the comparison criteria are set to be the same among different data frequencies.

Table 10 and Table 11 document the forecasting result of the DCC-GARCH model for 15 min data and daily data respectively. There are 5 loss functions mentioned in the methodology part to indicate the performance of the DCC-GARCH model. Within each table, we are able to compare the forecast accuracy of mean equations of crude oil and S&P 500 index and we can make cross-table comparison of the volatility forecast performance. Put it another way, we compare the loss functions measuring the volatility of each financial asset calculated from different time frequencies.

Table 10. Forecast valuation of one-day out-of-sample volatility forecasts of DCC-GARCH model of intraday data

	Oil Mean	Stock Mean	Oil Vol	Stock Vol
MSE	1.62E-05	0.000141	2.18E-09	6.37E-08
MedianSE	3.52E-06	4.11E-07	4.27E-11	5.47E-09
ME	-3.5E-05	-0.00013	-7.3E-06	2.2E-07
MAE	0.002763	0.00111	1.72E-05	0.000141
RMSE	0.004026	0.011884	4.67E-05	0.000252

Notes: The value of each loss function for the forecast valuation of the first and the second moment of oil and stock is documented in every cell.

Table 11. Forecast valuation of one-day out-of-sample volatility forecasts of DCC-GARCH model of daily data

	Oil Mean	Stock Mean	Oil Vol	Stock Vol
MSE	0.000801	0.000133	2.43E-06	0.006466
MedianSE	0.000244	2.64E-05	2.43E-07	8.38E-13
ME	0.002462	0.00071	-1.1E-05	0.001079
MAE	0.020488	0.008029	0.000901	0.001081
RMSE	0.028299	0.011551	0.00156	0.080409

Notes: The value of each loss function for the forecast valuation of the first and the second moment of oil and stock is documented in every cell.

After reading the numbers in Table 10 we can conclude that DCC-GARCH provides a more accurate forecast for oil than stock market in terms of mean equations. Crude oil outperforms stock market in all loss functions in terms of mean equations. For daily data, it is the other way round. For all loss functions, the forecast results of the mean of stock market are more accurate than those of the mean of crude oil market.

We also make cross-table comparison of the volatility forecast performance. We are able to read that the values of all loss functions of financial assets in intraday frequencies are less than those in daily frequency except the median squared error for stock market volatility. By using the loss functions in this study, we can conclude that the using of intraday data improves the forecast ability of DCC-GARCH model. Our finding is in line with Pong et al. (2004) and Chortareas et al. (2011).

7. Conclusion

The co-movements of crude oil/stock returns and volatilities are important in asset allocation and risk management. In this paper, we employ continuous wavelet analysis and traditional time series model DCC-GARCH model to assess the relationship between S&P 500 stock market and crude oil market.

Wavelet method allows for the examination of the time-and frequency varying co-movements of financial assets within a unified framework. Wavelet analysis is a model-free approach which can distinguish between short and long run relations for a single time series or for two series to detect the relationship between the two financial assets. We make contribution to the literature by extending wavelet analysis framework into intraday data. Unfortunately, the wavelet approach fails to detect the relationship between the crude oil market and stock market due to the large number of observations involved. However, wavelet analysis does fit daily data and it is able to distinguish between short term investment behaviours and long term investment behaviours. Madaleno and Pinho (2014) find that the relationship between oil prices and sector stock returns is ambiguous because phase and anti-phase relationships exist for different horizons at the same observation time. Our daily data findings are in line with their results.

Apart from using wavelet method, we also apply DCC-GARCH model to estimate and forecast the return and volatility of crude oil market and stock market. We are among the pioneers to identify the relationship between the two vital financial assets with the help of intraday data. We find that intraday data and daily data fit DCC-GARCH model well. We also find that by using daily data, the two financial assets in questions have similar volatility figures and we are able to conclude a rule of thumb that from 2002 to 2013, the volatility peaks of crude oil would be documented in intraday data 6-8 months ahead of the peaks in daily data.

By measuring the forecasting performance of intraday DCC-GARCH model and daily DCC-GARCH model, we can conclude that the using of intraday data improves the forecast ability of DCC-GARCH model. Our finding support the theory saying that the employment of high frequency data is instrumental for improving the forecasting performance of traditional time series method.

One limit of our study is that we are not able to forecast the wavelet approach results due to its model-free nature. The in-sample comparison between DCC-GARCH model and wavelet approach is valuable while the out-of-sample comparison is also a key interesting point in later research.

A potential extension of the current study is to study the contagion of the stock and oil markets (Reboredo and River-Castro, 2014; Martín-Barragán et al., 2015). Also, to develop the current wavelet analysis approach and to make the forecast feature available in a way are promising for future's research.

The analysis conducted has a number of practical implications to practitioners and policy makers. Findings in the paper can be applied to the construction of dynamic optimal portfolio diversification strategies and value-at-risk methodologies since changes in correlation/wavelet coherency impacts portfolio weights.

References

Aguiar-Conraria, L., Azevedo, N., & Soares, M. J. (2008). Using wavelets to decompose the time–frequency effects of monetary policy. Physica A: Statistical mechanics and its Applications, 387(12), 2863-2878.

Alsalman, Z., & Herrera, A. (2013). Oil Price Shocks and the U.S. Stock Market: Do Sign and Size Matter? Energy Journal, 36(3), 171-188.

Andersen, T. G., & Bollerslev, T. (1998). Answering the skeptics: yes, standard volatility models do provide accurate forecasts. International Economic Review, 39, 885–905.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (1999). (Understanding, optimizing, using and forecasting) realized volatility and correlation. Manuscript, Northwestern University, Duke University and University of Pennsylvania. Published in revised form as "Great realizations". Risk. 105–108.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2001). The distribution of exchange rate volatility. Journal of the American Statistical Association, 96, 42–55.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modelling and forecasting realized volatility. Econometrica, 71, 579–625.

Arouri, M.H., Jouini, J., & Nguyen, D. K. (2011). Volatility spillovers between oil prices and stock sector returns: Implications for portfolio management. Journal of International Money and Finance, 30(7), 1387–1405.

Balclar, M., & Ozdemir. Z. (2013). The causal nexus between oil prices and equity market in the U.S.: A regime switching model. Energy Economics, 39, 271–282.

Bandi, F. M., & Russell, J. R. (2005). Separating microstructure noise from volatility. Journal of Financial Economics, 79, 655–692.

Brooks, C., S. P. Burke, & Persand, G. (1997). Linear and Non-Linear (Non-)Forecastability of High-Frequency Exchange Rates. Journal of Forecasting, 16, 125–145.

Chortareas, G., Jiang, Y., & Nankervis, J. (2011). Forecasting exchange rate volatility using high-frequency data: Is the euro different? International Journal of Forecasting, 27, 1089-1107.

Colacito, R., Engle, R., & Ghysels, E. (2011). A component model for dynamic correlations. Journal of Econometrics, 164(1), 45–59. Annals Issue on Forecasting.

Dacorogna, M. M., Gencay, R., Muller, U. A., Olsen, R. B., & Pictec, O. V. (2001). An introduction to high-frequency finance. San Diego: Academic Press.

Elyasiani, E., Mansur, I., & Odusami, B. (2011). Oil price shocks and industry stock returns. Energy Economics, 33, 966–974.

Engle, R.F. (2002). Dynamic Conditional Correlation: a New Simple Class of Multivariate GARCH Models, Journal of Business and Economic Statistics, 20, 339-350.

Fan, Q., & Jahan-Parvar, M. (2012). U.S. industry-level returns and oil prices. International Review of Economics & Finance, 22(1), 112–128.

Goodhart, C.A.E. (1989). News and foreign exchange market. In Proceedings of the Manchester Statistical Society, 1-79.

Goodhart, C.A.E., & Figliuoli, L. (1992). The geographical location of the foreign exchange market: A test of an "islands" hypothesis, Journal of International and Comparative Economics, 1, 13-27.

Goodhart, C.A.E., Ito, T., & Payne, R.G. (1995). A study of the working of Reuters 2000-2 electronic foreign exchange trading system, NBER Technical Working Paper No: 179.

Grinsted, A., Moore, J.C., & Jevrejeva, S. (2004). Application of the cross wavelet transform and wavelet coherence to geophysical time series. Nonlinear Processes in Geophysics, 11,561–566.

Hamilton, J.D. (1983). Oil and the macroeconomy since World War II. Journal of Political Economy, 91, 228–248.

Hamilton, J.D. (1985). Historical causes of postwar oil shocks and recessions. Energy Journal, 6, 97–116.

Hamilton, J. D. (2003). What is an oil shock? Journal of Econometrics, 113(2), 363–398.

Hammoudeh, S., Dibooglub, S., & Aleisa, E. (2004). International Review of Economics & Finance, 13(4), 427–453.

Huang, R.D., Masulis, R.W., & Stoll, H.R. (1996). Energy shocks and financial markets Journal of Futures Market, 16 (1), 1–27.

Jones, C.M., & Kaul, G. (1996). Oil and the stock markets Journal of Finance, 51 (2), 463–491.

Kang, S.H., Kang, S.M., & Yoon, S.M. (2009). Forecasting volatility of crude oil markets. Energy Economics, 31, 119–125.

Kang, W., Ratti, R. A., & Yoon, K. H. (2014). The impact of oil price shocks on U.S. bond market returns. Energy Economics, 44, 248–258.

Khalfaouia, R., Boutahara, M., & Boubaker, H. (2015). Analyzing volatility spillovers and hedging between oil and stock markets: Evidence from wavelet analysis. Energy Economics, 49, 540–549.

Kilian, L., & Park, C. (2009). The impact of oil price shocks on the U.S. stock market. International Economic Review, 50(4), 1267–1287.

Madaleno, M., & Pinho, C. (2014). Wavelet dynamics for oil-stock world interactions. Energy Economics, 45, 120–133.

Martín-Barragán, B., Ramos, S.B., & Veiga, H. (2015). Correlations between oil and stock markets: A wavelet-based approach. Economic Modelling, 50, 212–227.

Martens, M., Chang, Y. C., & Taylor, S. (2002). A comparison of seasonal adjustment methods when forecasting intraday volatility. Journal of Financial Research, 25(2), 283–299.

Mohammadi, H., & Su, L. (2010). International evidence on crude oil price dynamics: applications of ARIMA–GARCH models. Energy Economics, 32 (5), 1001–1008.

Mollick, A., & Assefa, T. (2013). U.S. stock returns and oil prices: The tale from daily data and the 2008–2009 financial crisis. Energy Economics, 36, 1–18.

Park, J., & Ratti, R. A. (2008). Oil price shocks and stock markets in the U.S. and 13 European countries. Energy Economics, 30(5), 2587–2608.

Reboredo, J.C., & Rivera-Castro, M. A. (2014). Wavelet-based evidence of the impact of oil prices on stock returns. International Review of Economics and Finance, 29, 145–176.

Rua, A. (2010). Measuring comovement in time–frequency space. Journal of Macroeconomics, 32, 685–691.

Rua, A., & Nunes, L.C. (2005) Coincident and leading indicators for the euro area: a frequency band approach, International Journal of Forecasting, 21, 503–23.

Rua, A., & Nunes, L. C. (2012). A wavelet-based assessment of market risk: the emerging markets case, Quarterly Review of Economics and Finance, 52, 84–92.

Rua, A., & Nunes, L. C. (2009). International comovement of stock returns: A wavelet analysis. Journal of Empirical Finance, 11, 632–639.

Salisu, A., & Oloko, T. (2015). Modeling oil price–US stock nexus: A VARMA–BEKK–AGARCH approach. Energy Economics, 50, 1–12.

Tang, K., & Xiong, W. (2012). Index Investment and the Financialization of Commodities. Financial Analysts Journal, 68, 54-74.

Torrence, C., & Compo, G.P., (1998). A practical guide to wavelet analysis. Bulletin of the American Meteorological Society, 79, 605–618.

Uddin, G., Arouri, M., & Tiwari, A. (2013). Co-Movements between Germany and International Stock Markets: Some New Evidence from DCC-GARCH and Wavelet Approaches. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2311724

Wei, Y., Wang, Y., & Huang, D. (2010). Forecasting crude oil market volatility: Further evidence using GARCH-class models. Energy Economics, 32. 1477–1484.

Conclusion

This thesis studies an array of volatility estimation models and evaluates the forecasting performance of those estimation models on high-frequency data/intraday data and daily data of WTI crude oil futures in the first two chapters. I also study the linkage of crude oil futures and the US stock market and evaluate the forecasting results of multi-variable GARCH model in Chapter 3.

In Chapter 1, I employ a greater number of GARCH-class models and many loss functions and carry out the superior predictive ability (SPA) test to estimate and compare the forecasting performance on the basis of intraday data and daily data. Several GARCH family models such as GJR, EGARCH, APARCH, FIGARCH, FIAPARCH, HYGARCH capture long-memory volatility and/or the asymmetry leverage effect in volatility. None of the GARCH-class models outperforms the others when it comes to intraday data. Our finding is against some research papers in the literature such as ABDL (2001), Corsi (2009), Martens and Zein (2004) and Chortareas et al. (2011) which all document that long memory specification in high-frequency data can improve the forecasting power and accuracy significantly. EGARCH model is superior to other model when it comes to daily data and it is different from the finding of Kang et al. (2009) in which FIGARCH performs well.

Our findings provides a solid piece of evidence to the cons part in the discussion that whether the traditional time series models are good to fit intraday data. We find that the traditional volatility model cannot fit the data when we employ intraday data. After de-seasonalising the raw returns of the crude oil futures and putting in GARCH family models, it emerges that no GARCH model can produce satisfactory forecast results.

We find that the intraday crude oil returns are consistent with the stylised properties of other financial series such as stock market indices and exchange rates at high frequencies in many respects. This becomes a piece of evidence that these properties are not limit to certain kinds of high-frequency data. It might reflect some general features which all intraday data share. Agnolucci (2009) proposes the question “whether the comparison of volatility forecasting models is influenced by the criterion used in the exercise.” Our findings indicate that the rankings of the performance of volatility models are different when different criteria are applied to.

The results of Chapter 1 suggest that economists and financial practitioners should not arbitrarily choose a volatility forecasting model by referring to the existing research. Which model can be trusted depends on not only the given data sample but also the correspondence of the particular forecasting purpose with the loss function considered.

In Chapter 2, I present results from an empirical analysis of a batch of predictive HAR-type time-series models whose aim is to forecast realised volatility. For the in-sample fitness performance, there is no outperforming model in terms of the explanatory power i.e. R squared. Squared jumps help to reduce future’s volatility to some extent. MedRV jump is more significant than BPV jump component but their contribution to volatility explanation is limited. The information of the decomposition of variance into semi-variance is mixed which is against Sevi’s (2014) finding that considering independently the squared jump component, the continuous component, signed jumps and realised semi-variances of both signs significantly help to improve the fit of the predictive regression.

The out-of-sample performance comparison presents the most complicated HAR-type model outperforms other simple HAR-type models. The comparison between GARCH-type models and HAR-type models is inconclusive. This finding is against Andersen, Bollerslev, Christoffersen, and Diebold (2006, chap. 15), who find that even based on simple autoregressive structures such as the HAR provide much better results than GARCH-type models.

In Chapter 3, the results show that the wavelet approach fails to detect the relationship between the crude oil market and stock market due to the large number of observations involved. However, wavelet analysis does fit daily data and it is able to distinguish between short term investment behaviours and long term investment behaviours.

By utilising DCC-GARCH model, I find that intraday data and daily data fit DCC-GARCH model well. I also find that by using daily data, the two financial assets in questions have similar volatility figures and I am able to conclude a rule of thumb that from 2002 to 2013, the volatility peaks of crude oil would be documented in intraday data 6-8 months ahead of the peaks in daily data. The forecasting evaluation shows that the using of intraday data improves the forecast ability of DCC-GARCH model.

There are still limitations of the thesis. First, the time series models in the thesis are GARCH series models and HAR series models. Though the comparison of the forecast performance of different models is the highlight of the thesis, the in-sample specification and modelling are the cornerstone of forecasting performance. For the in-sample specification, regime-switching models and stochastic volatility models are not mentioned in this thesis. Second, the forecasting performance tests are limited to DM test and SPA test. Stepwise SPA test (Hsu et al., 2010), an improvement on the conservation of SPA test has already been introduced to the literature before the writing of the thesis. Stepwise SPA (SSPA) test is not adopted in the thesis, which is also a limitation in the thesis. Third, the wavelet approach results utilised in the thesis are not able to detect evidence of contagion in the two markets (see Reboredo and River-Castro, 2014; Martín-Barragán et al., 2015). Future research would extend the thesis based on the limitations mentioned above.

The analysis conducted has a number of practical implications to practitioners and policy makers. Results of Chapter 1 and 2 suggest that economists and financial practitioners should not arbitrarily choose a volatility forecasting model by referring to the existing research. Which model can be trusted depends on not only the given data sample but also the correspondence of the particular forecasting purpose with the loss function considered. Findings in Chapter 3 can be applied to the construction of dynamic optimal portfolio diversification strategies and value-at-risk methodologies since changes in correlation/wavelet coherency impacts portfolio weights.