

Asyl Bakanova

# The futures price volatility in the crude oil market

Submitted for the degree of Ph.D. in Economics at

*Faculty of Economics  
University of Lugano  
Lugano, Switzerland*

*Thesis Committee:*

Prof. G. Barone-Adesi, advisor, University of Lugano

Prof. P. Gagliardini, University of Lugano

Prof. F. X. Diebold, University of Pennsylvania

11<sup>th</sup> January 2011

©2011 by Asyl Bakanova



# Acknowledgments

This PhD dissertation would have never been completed without the help, support, and guidance of numerous people, all of whom deserve special thanks and appreciation.

First, I am thankful to my supervisor, Giovanni Barone-Adesi, for supporting and guiding me throughout my research process, his valuable suggestions and feedback on my thesis. I would also like to thank my thesis reviewers - Patrick Gagliardini and Francis X. Diebold - for their thorough and valuable comments, which have significantly improved this thesis.

My special gratitude goes to Francis X. Diebold for hosting me at the University of Pennsylvania, and thus, giving me an opportunity to continue my research in one of the world's best institutions.

I also want to thank my wonderful family - my parents, brother - for their endless support. They have always had faith and confidence in me. Last, but not least, I thank my husband Attilio for his love, support, patience and encouragement. And, most importantly, I thank my precious daughter Arianna for bringing so much joy and happiness and giving a new meaning of life. This work is dedicated to them.

I acknowledge the financial support from the Swiss Finance Institute, University of Lugano, NCCR-FINRISK and from the Swiss National Science Foundation.



# Contents

<b>1</b>	<b>Modeling volatility in the crude oil futures market with time-series models</b>	<b>1</b>
1.1	Introduction . . . . .	2
1.2	Literature review . . . . .	4
1.3	Volatility measures . . . . .	7
1.4	Methodology . . . . .	9
1.4.1	Time-varying volatility models . . . . .	9
1.4.2	Forecasting methodology . . . . .	13
1.5	Data description and summary statistics . . . . .	14
1.6	Empirical results . . . . .	22
1.6.1	Volatility measures . . . . .	22
1.6.2	In-sample estimation results . . . . .	25
1.6.3	Out-of-sample forecasting results . . . . .	27
1.7	Conclusion . . . . .	29
<b>2</b>	<b>Information content of implied volatility in the crude oil futures market</b>	<b>34</b>
2.1	Introduction . . . . .	35
2.2	Literature review . . . . .	38
2.3	Volatility measures . . . . .	42
2.3.1	Black-Scholes implied volatility . . . . .	42
2.3.2	Model-free implied volatility . . . . .	43
2.3.3	Realized volatility . . . . .	45
2.4	Methodology . . . . .	46
2.4.1	In-sample estimation . . . . .	46
2.4.2	Out-of-sample forecasting . . . . .	48
2.5	Data description and summary statistics . . . . .	51

2.6	Empirical results . . . . .	54
2.6.1	In-sample estimation results . . . . .	54
2.6.2	Out-of-sample forecasting results . . . . .	57
2.7	Robustness analysis . . . . .	65
2.7.1	Error-in-variables . . . . .	66
2.7.2	Subsample analysis . . . . .	68
2.8	Conclusion . . . . .	69
<b>3</b>	<b>Variance risk premium in the crude oil futures</b>	<b>72</b>
3.1	Introduction . . . . .	73
3.2	Literature review . . . . .	74
3.3	Variance Risk Premium: Definition and Measurement . . . . .	77
3.4	Methodology . . . . .	80
3.5	Empirical results . . . . .	82
3.6	Conclusion . . . . .	85

# List of Tables

1.1	Contract specifications: light, sweet crude oil futures traded at NYMEX.	15
1.2	Descriptive statistics of daily crude oil futures returns. . . . .	17
1.3	Autocorrelations of crude oil futures squared returns. . . . .	20
1.4	Descriptive statistics of log range and log absolute returns. . . . .	22
1.5	Correlation between volatility estimators. . . . .	24
1.6	Descriptive statistics of the volatility measured by different estimators. .	25
1.7	Efficiency of the range-based volatility estimators. . . . .	26
1.8	In-sample estimation results of GARCH-type volatility models. . . . .	31
1.9	In-sample estimation results of HAR model. . . . .	32
1.10	Out-of-sample forecasting performance. . . . .	32
1.11	Diebold-Mariano test results. . . . .	33
2.1	Contract specifications: light, sweet crude oil futures options traded at NYMEX. . . . .	52
2.2	Descriptive statistics: realized volatilities and implied volatilities. . . . .	56
2.3	In-sample estimation results of GARCH models. . . . .	57
2.4	In-sample estimation results of EGARCH models. . . . .	58
2.5	Forecast regression of model-free implied volatility. . . . .	59
2.6	Forecast regression of historical volatility. . . . .	60
2.7	Forecast regression of GARCH volatility forecasts. . . . .	61
2.8	Forecast regression of Black-Sholes implied volatility. . . . .	62
2.9	Information content of model-free implied volatility versus historical volatility and the GARCH(1,1) volatility forecast. . . . .	63
2.10	Information content of Black-Scholes implied volatility versus historical volatility and the GARCH(1,1) volatility forecast. . . . .	64

2.11	Information content of Black-Scholes implied volatility versus model-free implied volatility. . . . .	65
2.12	Information content of model-free implied volatility versus historical volatility using overlapping samples. . . . .	66
2.13	Information content of Black-Scholes implied volatility versus historical volatility using overlapping samples . . . . .	67
2.14	Information content of implied volatility: instrumental variable estimates	68
2.15	Descriptive statistics for realized model-free implied volatilities: Subsample period analysis. . . . .	69
2.16	Information content of implied volatility: subperiod analysis. . . . .	70
3.1	Descriptive statistics: realized variance, implied variance, and variance risk premium. . . . .	82
3.2	Time-variation in variance risk premium. . . . .	85

# List of Figures

1.1	Dynamics of daily prices of WTI crude oil futures market (US Dollar/Barrel) - January 3, 1991 to December 14, 2006. . . . .	16
1.2	Crude oil futures return distribution. . . . .	18
1.3	QQ plot of the crude oil futures returns. . . . .	19
1.4	Dynamics of daily returns of WTI crude oil futures market (US Dol- lar/Barrel) - January 2, 1991 to December 14, 2006. . . . .	21
1.5	Range-based vs. traditional close-to-close volatility estimators in the crude oil futures market. . . . .	23
2.1	Time-series plot of the model-free implied volatility. . . . .	53
2.2	First differences of model-free implied volatility. . . . .	54
2.3	Time series plot of the realized volatility measures. . . . .	55
3.1	Time series plot of the implied, realized variances, and variance risk premium.	83



# Abstract

The main goal of this thesis is to present and evaluate different procedures for modeling and forecasting volatility, and examine the relative accuracy of these forecasts using data from the light, sweet crude oil futures market traded at New York Mercantile Exchange (NYMEX). First, we consider various volatility models and find that the models, which account for long memory, has the best forecasting performance over the longer horizons. We use the range-based volatility estimators, based on high, low, opening, and closing prices, as volatility proxies. Next, we apply the model-free methodology to extract implied volatility from prices of options on light, sweet crude oil futures and analyze its information content. Our results show that model-free implied volatility, although biased, has predictive power and it is an efficient measure of future realized volatility. We find that forecasts based on historical prices do not contain information that was not impounded in implied volatility. Finally, we estimate the variance risk premium in the crude oil futures market and analyze whether this risk premium is priced, and thus, causes implied volatility to be a biased forecast of future realized volatility. In line with previous evidence on equity indices and currencies, we show that the volatility risk premium in the crude oil futures market has been negative and time-varying, in particular, for the period of high volatility between September 2001 and December 2006.



# Introduction

Reliable and accurate forecasts of volatility are important inputs for various finance applications, including derivatives pricing, risk management, portfolio selection and asset allocation. Since asset price volatility is not directly observable, much effort has been devoted to extracting volatility from other observable variables. As a result, a range of volatility models has been developed to measure and forecast volatility and incorporate it into both theoretical and applied research.

One common approach for forecasting volatility involves building time series models for volatility, using past asset prices in combination with other relevant historical information. These models fall into one of two categories, the autoregressive conditional heteroskedastic (ARCH) family, first introduced by Engle (1982), and the stochastic volatility (SV) family, which traces its roots to Clark (1973). SV models explicitly include an unobserved shock to the return variance into representation of the volatility dynamics, which makes them more difficult to estimate.<sup>1</sup> In contrast, an advantage of the ARCH models is that they treat the conditional variance given past information as observable, and maximum-likelihood methods can be used to estimate the model parameters. Indeed, researchers have increasingly used the ARCH model and its various extensions, as they have proven to capture the time-varying volatility observed in financial data.

Despite the empirical success of ARCH models in modeling the in-sample volatility of asset prices, some studies have reported that these models provide poor out-of-sample forecasts and explain little of the variability of *ex post* realized volatility (see, e.g., Figlewski (1997)). Andersen and Bollerslev (1998) argue that the poor performance of the ARCH models to provide good out-of-sample forecasts is not a failure of the models per se, but a failure to specify correctly the volatility proxy, on the basis of which the

---

<sup>1</sup>The SV models have been extensively surveyed in Ghysels, Harvey and Renault (1996) and Shephard (1996), among others.

forecasting performance is evaluated. The log absolute or squared returns, which had been primarily used as volatility proxies, are very noisy due to the measurement errors. Thus, they do not provide reliable and efficient inferences regarding the underlying latent volatility and its dynamics.

Consequently, Andersen and Bollerslev (1998) suggest an alternative measure for ex post latent volatility estimated from high-frequency intraday returns. They argue that this measure, under suitable conditions, is an unbiased and highly efficient estimator of return volatility, and thus, it allows more accurate volatility forecast evaluation.<sup>2</sup> However, the high-frequency volatility has been criticized for its susceptibility to microstructure effects as a result of nonsynchronous trading, discrete price observations, intraday periodic volatility patterns and bid-ask bounce.<sup>3</sup> In addition, the reliable high-frequency data is not readily available for all financial markets.

Alternative estimators of latent volatility, which have been known for a long time, are the extreme-value, or range-based, estimator.<sup>4</sup> The advantage of using range-based, or extreme-value, estimators is that they require the daily opening, closing, high and low prices data, which are widely available for stock markets, commodities, and currencies over long historical spans. In addition, Alizadeh, Brandt, and Diebold (2002), using both analytical and numerical methods, show that the log range, defined as the difference between the logarithms of the daily high and low prices, is a highly efficient volatility estimator. It is approximately Gaussian, much less noisy than alternative volatility proxies such as log absolute or squared returns, and robust to bid-ask bounce and related microstructure noise.

Alternative approach to forecasting volatility, which has recently gained the popularity, is based on the information contained in option prices. Traditionally, options have been used for managing the price risk due to their ability to hedge against unfavorable market moves. But in addition to their hedging function, options are also useful to infer information about market's assessment of the underlying asset's future volatility. This can be attributed to the underlying forward-looking approach in the formation of volatility expectations, which is able to capture information beyond that contained in historical

---

<sup>2</sup>See also Andersen, Bollerslev, Diebold and Ebens (2001), Andersen, Bollerslev, Diebold and Labys (2003), Barndorff-Nielsen and Shephard (2002), among others.

<sup>3</sup>See, e.g., Aït-Sahalia, Mykland and Zhang (2005), Bandi and Russell (2004), Hansen and Lunde (2006).

<sup>4</sup>See, e.g., Parkinson (1980), Garman and Klass (1980), Beckers (1983), Ball and Torous (1984), Rogers and Satchell (1991), Yang and Zhang (2000).

returns.

The ability of the volatility forecast implied by option prices to predict future volatility is considered a measure of the information content of option prices. It has been frequently tested in the literature and the results have been somewhat mixed. In general, the results of the studies on information content of implied volatility provide sufficient evidence of its efficiency and superior predictive ability. However, most of these studies also find that implied volatility is an upward-biased predictor of future realized volatility, such that the implied volatility is on average higher than the realized volatility. One of the potential explanations for the bias in implied volatility is the presence of a volatility risk premium.<sup>5</sup> This is because options are used to hedge against adverse volatility shocks and hence investors are willing to pay a premium to hedge against market volatility going up.

Another reason for the observed bias in option implied volatility as a forecast of future realized volatility is the measurement error and/or model misspecification. For example, most of the previous research on the information content of implied volatility focuses on the at-the-money Black-Scholes implied volatility. However, one of the main assumptions of the Black-Scholes model is constant volatility. Since a constant volatility model is used to analyze options, whose prices have been established under conditions of stochastic volatility in the market, this results in measurement errors due to model misspecification. In addition, by concentrating on at-the-money options only, these studies ignore information contained in other options.

Beginning with Breeden and Litzenberger (1978), there has been considerable effort in deriving probability densities that are consistent with all option prices observed at the same point in time. Based on these studies, Britten-Jones and Neuberger (2000) derive a model-free implied volatility measure that incorporates the whole cross-section of option prices, not only at-the-money prices. Their methodology does not rely on any option pricing formula and no assumption is made regarding the underlying stochastic process. Therefore, the measurement error from model misspecification is reduced.

The main goal of this thesis is to present and evaluate different procedures for modeling and forecasting volatility, and examine the relative accuracy of these predictions using data from the light, sweet crude oil futures market traded at New York Mercantile Exchange (NYMEX). Most of the research on volatility modeling and forecasting so far

---

<sup>5</sup>See, e.g., Rosenberg and Engle (2002), Bakshi and Madan (2006), Bollerslev, Gibson, and Zhou (2008).

has focused on equity and foreign exchange markets. Much less attention has been paid to commodities markets, which are becoming increasingly attractive to investors as an alternative asset class. Among commodity markets, crude oil market is one of the biggest and most actively traded markets in the world.

According to NYMEX, the light, sweet crude oil futures contract is the world's largest by volume trading and the most liquid contract on physical commodity. Because of the higher liquidity and low transaction costs on the exchange, this market has attracted a wide range of participants and has evolved from a primarily physical activity into one of the most complex financial markets with different financial instruments and much longer trading horizons. As a result, estimating and forecasting volatility for pricing and hedging purposes has become even more challenging. Given the extremely high volatility that characterizes this market, it is a good candidate for testing volatility forecasting models.

The outline of the thesis is as follows.

In Chapter 1, we estimate various GARCH specifications, including GARCH(1,1), Exponential GARCH (EGARCH), GJR-GARCH, and Asymmetric Power ARCH (APARCH). In order to reproduce the long memory dependence observed in the data, we also estimate Fractionally Integrated GARCH (FIGARCH) and Heterogeneous Autoregressive (HAR) models. To improve the accuracy of the inference regarding predictive performance, we rely on range-based volatility measures as a proxy for the unobservable volatility process. Unlike in previous research conducted in other markets, we do not find a significant negative and asymmetric contemporaneous relationship between changes in the volatility and the returns. For the case of one-step ahead forecasts, the GARCH(1,1) model has the best performance, whereas GJR-GARCH model that allows for asymmetries seems to be the worst model in terms of predictive ability. Our results also provide evidence that the HAR model is able to capture the long memory observed in the data. In general, for the longer horizons, models which account for long memory have superior out-of-sample forecasting performance compared to that of the GARCH and EGARCH specifications.

In Chapter 2, we apply the model-free methodology to extract implied volatility from prices of options on light, sweet crude oil futures and analyze its information content. Next, we evaluate the information content and forecasting accuracy of the model-free implied volatility relative to the information contained in alternative volatility forecasts in the crude oil futures market. In-sample tests are conducted by embedding implied volatility as an additional explanatory variable in GARCH specifications. The out-of-

sample forecasts are evaluated using both univariate and encompassing regressions. The results show that model-free implied volatility has predictive power, and therefore, it is a useful measure of expected future volatility, although option prices provide biased predictions of future realized volatility. Out-of-sample forecasting results also provide evidence that implied volatility subsumes historical volatility, GARCH-type volatility forecast, and Black-Scholes implied volatility. We also find that the implied volatility has better forecasting performance in the period after September 11, 2001, which corresponds to a higher volatility in the crude oil futures market.

In Chapter 3, we estimate the variance risk premium in the crude oil futures market as the difference between the realized variance and the synthetic variance swap rate. We also analyze whether this risk premium is priced and thus causes implied volatility to be a biased forecast of future realized volatility. The historical behavior of the variance risk premium in this market is also examined. In line with previous evidence focusing mainly on equity indices and currencies, we show that volatility risk premium has been negative and time-varying. This can be interpreted as the willingness of investors to accept a significantly negative average return to long variance swaps on the crude oil futures to hedge against market volatility going up, as market volatility shocks are considered unfavorable and investors demand high compensation for bearing such shocks.



# Chapter 1

## Modeling volatility in the crude oil futures market with time-series models

## 1.1 Introduction

Volatility has been one of the most investigated subjects in both empirical and theoretical financial economics. This is because of its critical role in applications, such as risk management, portfolio allocation, derivatives pricing and hedging. A number of methods has been proposed for volatility modeling and forecasting. Yet, how to model and estimate volatility still remains a challenging task, mainly because asset price volatility is not directly observable in the market, and thus, it has to be extracted.

The autoregressive conditional heteroskedasticity (ARCH) family of models, proposed by Engle (1982) and generalized by Bollerslev (1986), has been one of the most widely used models. Most of the studies have documented that ARCH models tend to have a relatively good in-sample forecasting performance and are successful in capturing the time dependence in volatility. This has led to the development of a large number of ARCH models and their extensions to incorporate various stylized facts of volatility and the time series properties of financial returns. Despite the success of the ARCH models and their extensions in modeling the in-sample volatility of asset prices, however, numerous studies report that these models provide less satisfactory out-of-sample forecasts (e.g., Figlewski (1997), Poon and Granger (2003)).

Andersen and Bollerslev (1998) argue that the failure of the GARCH models to provide good out-of-sample forecasts is not a failure of the GARCH model per se, but a failure to specify correctly the ex post volatility measure, based on which the forecasting performance is measured. They show that the standard approach of using daily squared returns based on daily closing prices as the measure for ex post volatility is flawed, mainly because this measure provides a very noisy proxy for the "true" volatility process. In addition, by using only closing prices to compute daily volatility measure, only two observations are employed, and the path of the price inside the reference period is ignored.

Consequently, Andersen and Bollerslev (1998) suggest an alternative measure of ex post volatility, known as realized volatility, which is computed from high-frequency data as the sum of intraday squared returns over a given sampling period. A number of studies has shown that under the assumption that the logarithmic asset price process is a semimartingale, this measure converges uniformly in probability to the quadratic variation of the process (see, e.g., Andersen, Bollerslev, Diebold, and Labys (2001, 2003); Barndorff-Nielsen and Shephard (2002), Meddahi (2002)). Thus, the realized volatility

measure based on high-frequency intraday data offers a more accurate ex-post observation of the true unobserved return variance than the traditional estimator based on daily squared returns.

However, the realized volatility obtained from high-frequency data has been criticized for its susceptibility to microstructure effects as a result of bid-ask bounce, discrete price observations and non-synchronous trading, which contaminate intraday prices and distort the realized volatility measure (see, e.g., Aït-Sahalia, Mykland and Zhang (2005), Bandi and Russell (2006), Hansen and Lunde (2006)). In addition, reliable high-frequency data is still not readily available for all financial markets.

Alternative estimators of latent volatility that have been known for a long time are extreme-value, or range-based, estimators (see, e.g., Parkinson (1980), Garman and Klass (1980), Rogers and Satchell (1991)). Knowing extreme values allows to get close to the real underlying process, even if the whole path of asset prices is unknown. Since the price range utilizes more information from the sample path in contrast to estimators based on closing prices only, it is expected to be more efficient.<sup>1</sup> Moreover, Alizadeh, Brandt, and Diebold (2002) showed that range-based volatility is more robust to microstructure noise.

Most of the earlier studies on volatility modeling have focused on equity, bond and exchange rate markets. Relatively less attention has been paid to commodities markets, which are becoming increasingly popular among investors. Investing in commodities does not only provide hedge against inflation, but it also reduces the risk of unfavorable correlations with traditional investments through portfolio diversification. One of the world's biggest and most widely traded commodity markets is the crude oil market. Because of its importance for the economies in general, large fluctuations in oil prices can have adverse impacts not only for macroeconomy, but for financial markets as well (Kroner et al. (1995)).

In this study, we consider volatility modeling and forecasting in the light, sweet crude oil futures market. According to the New York Mercantile Exchange (NYMEX), this futures contract has become the world's largest-volume and most liquid futures contract trading on a physical commodity. In particular, we estimate various GARCH specifications, including GARCH(1,1), Exponential GARCH (EGARCH), GJR-GARCH, and Asymmetric Power ARCH (APARCH). In order to reproduce the long memory dependence observed in the data, we also estimate Fractionally Integrated GARCH (FI-

---

<sup>1</sup> Andersen and Bollerslev (1998) show that, under certain assumptions, the daily range is as efficient as the realized volatility based on returns sampled every four hours.

GARCH) and Heterogeneous Autoregressive (HAR) models. One of the main issues in evaluating the predictive ability of volatility models is that the true underlying volatility process is not observed. To improve the accuracy of the inference regarding predictive performance, we rely on range-based volatility measures as proxies for the unobservable volatility process.

Our results can be summarized as follows. First, unlike in previous research conducted for other markets, we do not find a significant negative and asymmetric contemporaneous relationship between changes in the volatility and the returns. For the case of one-step-ahead forecasts, the GARCH(1,1) model has the best performance, whereas GJR-GARCH model that allows for asymmetries seems to be the worst model in terms of predictive ability. Our results also provide evidence that the HAR model is able to capture the long memory observed in the data and has a superior out-of-sample performance. In general, for the longer horizons, models which account for long memory, have superior out-of-sample forecasting performance compared to that of the GARCH and EGARCH specifications.

The rest of this chapter is organized as follows. Section 1.2 provides a summary of the relevant literature. Section 1.3 discusses different volatility measures used in the study. Section 1.4 outlines the adopted methodology. Section 1.5 describes the data which is used for this study. The empirical findings are reported in Section 1.6. Finally, Section 1.7 concludes and suggests the directions for future research.

## 1.2 Literature review

GARCH models have been extensively applied for modeling the return volatility across different markets.<sup>2</sup> In the crude oil market, much research has been done to evaluate the forecasting performance of different volatility models in the spot markets. For example, Kang et al. (2009) compare the out-of-sample forecasting ability of different GARCH models, including the standard GARCH, FIGARCH, CGARCH, and IGARCH. They find that the CGARCH and FIGARCH models could capture the long-memory volatility of crude oil markets and obtain superior performance compared to that of the GARCH and IGARCH specifications. In a study following that of Kang et al. (2009), Cheong (2009) investigate the out-of-sample forecasting performance of four GARCH-class mod-

---

<sup>2</sup>Reviews can be found, among others, in Bollerslev, Chou and Kroner (1992), Diebold (2004), Engle and Patton (2001), Shephard (1996), Poon and Granger (2003).

els (GARCH, APARCH, FIGARCH, and FIAPARCH) under three loss functions in the crude oil spot markets. They find that the simplest and most parsimonious GARCH model fits better than the other models in the Brent crude oil market, while the FIAPARCH model outperforms other models in the WTI crude oil market.

Similarly, Mohammadi and Su (2010) compared the forecasting accuracy of four GARCH-class models (GARCH, EGARCH, APARCH, and FIGARCH) under two loss functions in eleven international crude oil spot markets. In contrast to Kang et al. (2009), they do not find strong support for the FIGARCH model, whereas out-of-sample forecasting evaluations of the conditional mean and variance favor EGARCH(1,1) and APARCH(1,1) models. Recently, Wei et al. (2010) extend the previous work on modeling volatility in the crude oil spot markets by including a greater number of linear and nonlinear GARCH-class models and using more loss functions in the framework of the superior predictive ability (SPA) test. They conclude that no single model outperforms all of the other models across the different loss functions. However, the nonlinear models do perform better than the linear ones in the long-run volatility forecasting of crude oil prices.

Several studies compare the volatility forecasting also in the crude oil futures markets. Sadorsky (2006) uses different univariate and multivariate models to estimate forecasts of daily volatility in petroleum futures price returns over the period from 1988 through 2003. They find that the GARCH model fits well for crude oil futures volatility. Agnolucci (2009) also estimates different GARCH models for the crude oil futures market, including CGARCH and TGARCH models with normal,  $t$ -student and GED error distributions, and finds that models with errors following GED distribution performs the best, while more complicated GARCH specifications do not outperform a simple GARCH(1,1) model.

Most of these studies use daily squared returns as volatility proxy, which is considered to be a very noisy estimator. On the other hand, the range has been shown to be a more efficient estimator of volatility than the variance estimator based on closing prices only. The idea of using information on the daily high and low prices, as well as the opening and closing prices, goes back to Parkinson (1980) and Garman and Klass (1980) with further contributions by Ball and Torous (1984), Rogers and Satchell (1991), Kunitomo (1992), and Yang and Zhang (2000). The advantage of the range data is that it is available for most financial assets over long time periods. Thus, many studies have examined the role of the range variable in a variety of contexts.

Schwert (1990) applies the range to two simple forecasting frameworks and finds that the range data do not help predict stock returns but do add significant information in predicting volatility. Gallant, Hsu and Tauchen (1999) prove the usefulness of range data in a stochastic volatility framework. Bali (2003) shows empirically that the value-at-risk (VaR) estimation using the extreme-value approach is superior to that of the standard approach. And Bali and Weinbaum (2005) demonstrate, at the daily frequency, extreme value estimators are less biased than the traditional estimator and they are also significantly more efficient.

Alizadeh, Brandt and Diebold (2002) extensively examine the properties of the log range volatility proxy. They show that the measurement error associated with the log range is about one quarter of the error of the standard volatility proxies, such as the absolute or squared open-to-close returns, and hence, the log range is much more efficient. In addition, the range is more robust to bid-ask bounce and related microstructure noise and it is approximately Gaussian. Thus, range-based volatility estimators are expected to provide more accurate volatility forecast evaluation.

By adding microstructure noise to the Monte Carlo simulation, Shu and Zhang (2006) support the finding of Alizadeh, Brandt and Diebold (2002) that range estimators are fairly robust towards microstructure effects. Similarly, Brandt and Diebold (2006) develop a range-based covariance measure and show that this measure is highly efficient and also robust to microstructure noise. Martens and van Dijk (2007) also indicate that in a continuous trading model, the realized range volatility estimate is five times more efficient than the traditional measure of realized variance.

In the context of volatility forecasting, Engle and Gallo (2006) confirm that the spread between the highest and lowest points of a daily price series has predictive power for realized volatility and can provide improved volatility estimates. Vipul and Jacob (2007) also examine the forecasting performance of extreme-value volatility estimators and find that most of the volatility models, including GARCH model, show a better performance with the use of extreme-value estimators, compared to the traditional volatility estimator.

The value of the range-based estimators has also been shown in studies of more complex volatility models. For example, Chou (2005) constructs the conditional autoregressive range (CARR) model and finds that the CARR model does provide more accurate volatility forecasts than the GARCH model. Brandt and Jones (2006) use the daily range to improve the EGARCH model, showing that volatility is predictable at horizons as long as one year, in part because of the improvement introduced by the range-based volatility

measure.

More recently, Chou and Liu (2010) examine the economic value of volatility timing for the range-based volatility model and conclude that the range-based volatility model has more significant economic value than the one based on returns. He, Kwok, Wan (2010) show that a modeling framework that exploits the interaction between the daily high price, low price and the range can significantly improve forecasts based on the linear ARIMA and naïve models. And Brownlees and Gallo (2010) evaluate the value-at-risk (VaR) forecast performance against some alternatives based on daily returns and find that the daily range performs remarkably well.

### 1.3 Volatility measures

The main issue in evaluating the predictive ability of volatility models is that since the true underlying volatility process is not observed, a good proxy is needed. Various volatility estimators have been suggested in the literature. Traditionally, the sample standard deviation of close-to-close returns computed over the relevant horizon has been used as a proxy for the unobservable volatility process. It is given as follows:

$$\hat{\sigma}_{C,t}^2 = \left( \ln \frac{C_t}{C_{t-1}} \right)^2 \quad (1.1)$$

where  $C_t$  is the closing price of the trading day  $t$ . However, if the previous trading day is quite volatile and the closing price happens to be the same as the opening price, then the lagged daily squared return would be zero. Thus, this estimator does not take into account the information given by the path of the price inside the period of reference. In addition, Andersen and Bollerslev (1998) show that due to the measurement error, this volatility measure is very noisy.

Recently, the sum of squared high-frequency intraday returns has been suggested as a more precise measure of volatility. When intraday data are unavailable or unreliable, an alternative approach is to use the information contained in the highest and lowest prices observed during the trading day. A number of these extreme-value volatility estimators have been proposed, which are, at least in theory, significantly more efficient than traditional estimator based on closing prices only.

Assuming an underlying geometric Brownian motion with no drift and no opening jumps, Parkinson (1980) developed an estimator which uses the daily high and low prices

of the asset for estimating its volatility. This estimator can be represented as follows:

$$\hat{\sigma}_{P,t}^2 = \frac{1}{4\ln(2)} \left( \ln \frac{H_t}{L_t} \right)^2 \quad (1.2)$$

where  $H_t$  and  $L_t$  are the high and the low prices on day  $t$ , respectively. Parkinson (1980) showed that this estimator is about five times more efficient than the traditional one based on closing prices only (i.e., its sampling variance is about five times lower).

Since the price path is not observable when the market is closed, Garman and Klass (1980) suggest a method to mitigate the effect of discontinuous observations by including opening and closing prices along with the high and low prices. The simple version of Garman and Klass estimator is:

$$\hat{\sigma}_{GK,t}^2 = 0.511 \left( \ln \frac{H_t}{L_t} \right)^2 - 0.019 \left[ \ln \left( \frac{C_t}{O_t} \right) \ln \left( \frac{H_t L_t}{O_t^2} \right) - 2 \ln \left( \frac{H_t}{O_t} \right) \ln \left( \frac{L_t}{O_t} \right) \right] - 0.383 \left[ \ln \frac{C_t}{O_t} \right]^2 \quad (1.3)$$

where  $O_t$  is the opening price on day  $t$ . Garman and Klass (1980) also showed that their estimator is theoretically more efficient than the traditional estimator.

The derivation of both Parkinson and Garman and Klass estimators depends on the assumption of continuous price paths. Moreover, these estimators are based on the assumption of driftless geometric Brownian motion process, which could lead to an overestimation of volatility. Rogers and Satchell (1991) relaxed the assumption of no drift and proposed an estimator which is given by:

$$\hat{\sigma}_{RS,t}^2 = \ln \left( \frac{H_t}{C_t} \right) \ln \left( \frac{H_t}{O_t} \right) + \ln \left( \frac{L_t}{C_t} \right) \ln \left( \frac{L_t}{O_t} \right) \quad (1.4)$$

In case if the security price has a drift and the price movement is largely explained by the drift rather than the volatility, the Rogers and Satchell estimator should be more efficient than the previous estimators.

Finally, Yang and Zhang (2000) propose an estimator that is independent of any drift and consistent in the presence of opening price jumps. The Yang and Zhang estimator uses the opening, high, low and closing prices of multiple periods to increase the efficiency. However, this estimator does not give the estimate of volatility based on the data of a single period. Therefore, we will not use it here.<sup>3</sup>

---

<sup>3</sup>We will also ignore the estimator proposed by Kunitomo (1992) as it requires tick data.

## 1.4 Methodology

### 1.4.1 Time-varying volatility models

Many studies have showed that the GARCH(1,1) specification provides a good first approximation to the statistical features of financial return series at the daily, weekly or monthly horizons. It is very parsimonious and is given by the following equations:

$$\begin{aligned} r_t &= \mu_t + \varepsilon_t, \varepsilon_t = z_t \sigma_t, z_t \sim NID(0, 1) \\ \hat{\sigma}_t^2 &= \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \end{aligned} \quad (1.5)$$

where  $r_t$  is the daily return,  $\mu_t$  is the conditional mean of the daily return,  $\sigma_t^2$  is the conditional variance of returns in period  $t$ ,  $\varepsilon_{t-1}$  is the last period's shock, and  $\sigma_{t-1}^2$  is the conditional variance of returns in period  $t-1$ . The parameter restrictions are  $\omega > 0$ ,  $\alpha > 0$ ,  $\beta > 0$ , and  $\alpha + \beta < 1$ . The latter constraint is necessary to ensure that the unconditional variance is finite, since the conditional variance evolves over time. The model can be extended to higher order GARCH( $p, q$ ) models by including additional lagged squared innovations and/or conditional variances. However, while many studies consider extensions, the GARCH(1,1) still remains very robust and is preferred in many cases.

In equilibrium, investors taking on additional risk should be compensated through higher expected return, which implies a positive coefficient in the risk-return relation. Time-varying risk premia can be modeled by including some function of the variance as an additional regressor in the conditional mean equation (1.5). The GARCH-in-mean (GARCH-M) model proposed by Engle et al. (1987) allows for the direct effect of volatility changes on asset prices in a short memory GARCH-type model, by introducing the conditional variance into the mean return equation. It is given by the following specification:

$$\begin{aligned} r_t &= \mu_t + \lambda \sigma_{t-1} + \varepsilon_t \\ \sigma_t^2 &= \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \end{aligned} \quad (1.6)$$

If  $\lambda$  is positive and statistically significant, then increased risk, given by an increase in the conditional variance, leads to a rise in the mean return. Thus  $\lambda$  can be interpreted as a risk premium.

The early generation of GARCH models, such as ARCH( $p$ ), GARCH( $p, q$ ) and their in-mean generalization have the ability of reproducing volatility clustering. However, they

assume that positive and negative shocks of the same absolute magnitude will have the identical effect on the conditional variance. In contrast, the volatility has been shown to respond asymmetrically to past negative and positive innovations, with negative returns resulting in higher volatilities.<sup>4</sup> Therefore, a symmetry constraint on the conditional variance function in past shocks is inappropriate.

This limitation has led to the development of more flexible volatility specifications, which would allow positive and negative shocks to have a different impact on volatility. One such model that tries to accomodate asymmetric effect is Exponential GARCH (EGARCH) model proposed by Nelson (1990), who argued that the nonnegative constraints in the linear GARCH model are too restrictive. Thus, in EGARCH model specificataion there are no restrictions on parameters  $\alpha$  and  $\beta$ . It is given as:

$$\log(\sigma_t^2) = \omega + \alpha(|z_{t-1}| - E(|z_{t-1}|)) + \gamma z_{t-1} + \beta \log(\sigma_{t-1}^2) \quad (1.7)$$

where  $z_t = \sigma_t^{-1} \varepsilon_t$ . The conditional variance given by Equation (1.7) is a function of the conditional variance of returns in period  $t - 1$ ,  $\sigma_{t-1}^2$ , last period's shock,  $\varepsilon_{t-1}$ , which has been standardized to have unit variance, and the deviation of its absolute value from the mean absolute value. The estimated conditional variance is strictly positive and it does not require non-negativity constraints used in the estimation of GARCH models. Including both the standardized value and its absolute value allows the variance equation to capture any asymmetry in the relation between market returns and conditional volatility.

The GJR-GARCH model by Glosten, Jagannathan and Runkle (1993) is another model that accounts for asymmetric effects of positive and negative innovations in the GARCH process. The GJR-GARCH(1,1) model can be represented as follows:

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \gamma \varepsilon_{t-1}^2 I(\varepsilon_{t-1} < 0) + \beta \sigma_{t-1}^2 \quad (1.8)$$

where  $I(\cdot)$  is an indicator function, such that when the condition in  $(\cdot)$  is met,  $I(\cdot) = 1$ , and 0 otherwise. Past negative return shocks will have a larger impact on the future conditional variance when  $\gamma > 0$ . The non-negativity condition is satisfied provided that  $\omega > 0$ ,  $\alpha > 0$ ,  $\beta \geq 0$ , and  $\alpha + \gamma \geq 0$ . If  $\gamma < 0$ , the model is still admissible, provided that  $\alpha + \gamma \geq 0$ .

---

<sup>4</sup>This asymmetric effect was first discovered in the stock market by Black (1976), and confirmed by the findings of French, Schwert and Stambaugh (1987), Nelson (1990), and Schwert (1990), among others.

Finally, the asymmetric power ARCH (APARCH) model of Ding et al. (1993) is one of the most promising ARCH-type models. This model nests several ARCH-type models and has been found to be particularly relevant in many recent applications. The APARCH(1,1) model is defined as follows:

$$\hat{\sigma}_t^\delta = \omega + \alpha(|\varepsilon_{t-1}| - \gamma\varepsilon_{t-1})^\delta + \beta\sigma_{t-1}^\delta \quad (1.9)$$

where parameter  $\delta$  ( $\delta > 0$ ) plays the role of a Box-Cox transformation of the conditional standard deviation  $\sigma_t$ , and  $\gamma$  reflects the asymmetric effect. The APARCH model nests several ARCH extensions as special cases, including the GARCH(1,1) model when  $\delta = 2$  and  $\gamma = 0$ , and the GJR-GARCH(1,1) model when  $\delta = 2$ .

Despite the success of the standard GARCH models in capturing time dependence in volatility, they still fail to capture strong volatility persistence observed in financial data. The long-memory property of asset market volatility has been the subject of extensive research and a number of empirical studies have argued for the importance of modeling the long-memory dynamic dependence in financial market volatility. Several different parametric ARCH and stochastic volatility models have been proposed in the literature for capturing this phenomenon (e.g., Andersen and Bollerslev (1997), Dacorogna et al. (2001)). For example, Andersen et al. (2003) adopt the class of autoregressive fractionally integrated moving average (ARFIMA) processes, introduced by Granger and Joyeux (1980) and Hosking (1981).

In this study, the first model that we use to account for volatility persistence is the Fractionally Integrated ARCH (FIGARCH) model introduced by Baillie, Bollerslev and Mikkelsen (1996). The model implies the finite persistence of volatility shocks, i.e., long-memory behavior and a slow rate of decay after a volatility shock. The FIGARCH is given by:

$$\varphi(L)(1-L)^d\varepsilon_t^2 = \omega + (1-\beta(L))v_t \quad (1.10)$$

where  $v_t = \varepsilon_t^2 - \sigma_t^2$ ,  $\varphi(L)$  is an infinite order polynomial for  $0 < d < 1$ ,  $\varphi(L) = [1 - \beta(L) - \alpha(L)]$ , and all the roots of  $\varphi(L)$  and  $[1 - \beta(L)]$  are all outside the unit circle.

Rearranging the terms in (1.10), we obtain:

$$\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + [1 - \beta L - (1 - \alpha L - \beta L)(1-L)^d]\varepsilon_t^2 \quad (1.11)$$

The parameter  $d$  characterizes the long-memory property of hyperbolic decay in volatility

because it allows autocorrelations to decay at a slow hypoerbolic rate. The advantage of the FIGARCH process is that for  $0 < d < 1$ , it is flexible enough to allow for ranges of persistence between the geometric decay associated with  $d = 0$  and the complete and integrated persistence of volatility shocks associated with  $d = 1$ . Thus, the FIGARCH model nests several GARCH models. When  $d = 0$ , FIGARCH is reduced to GARCH model, and when  $d = 1$  FIGARCH becomes integrated GARCH (IGARCH).

The problem with the FIGARCH model is that it still fails to reproduce the multiscaling behavior observed in financial data. The second model we consider is the heterogeneous autoregressive (HAR) model, which was proposed by Corsi (2009) and motivated by the Heterogeneous Market Hypothesis of Müller et al. (1997). In this model, the conditional variance is estimated as simple linear autoregressive model with volatilities realized over different time horizons. Thus, the HAR model represents an additive cascade of unobserved partial volatilities generated by different market components. At each cascade level, the corresponding partial volatility process is a function not only of its past value, but also the expected values of the longer-term partial volatilities. Although the HAR representation does not possess long-memory, the mixing of several volatility components is able of reproducing slow volatility autocorrelation decay.

To define the HAR model, let the multi-period realized variance, calculated as the sum of the corresponding one-period measures, be denoted by:

$$\sigma_{t,t+h} = \frac{\sigma_{t+1} + \sigma_{t+2} + \dots + \sigma_{t+h}}{h} \quad (1.12)$$

where  $h = 1, 2, \dots$  and  $\sigma_{t,t+1} = \sigma_{t+1}$ , by definition. Provided that the expectations exist and under stationarity condition,  $E(\sigma_{t,t+h}) = E(\sigma_{t+1})$  for all  $h$ . We will refer to these normalized measures for  $h = 1$ ,  $h = 5$  and  $h = 22$  as the daily, weekly and monthly volatilities, respectively.

Then the daily HAR model of Corsi (2009) may be expressed as:

$$\sigma_{t+1} = \beta_0 + \beta_D \sigma_t + \beta_W \sigma_{t-5,t} + \beta_M \sigma_{t-22,t} + \varepsilon_{t+1} \quad (1.13)$$

where  $t = 1, 2, \dots, T$ .<sup>5</sup>

In summary, five GARCH-class models, i.e. GARCH, GARCH-M, GJR-GARCH, EGARCH, and APARCH models, are used to describe and forecast the short-term volatil-

---

<sup>5</sup>Volatilities over other horizons could be included as additional explanatory variables, but in practice, daily, weekly and monthly measures are capable of capturing the long-memory feature.

ity of the crude oil futures prices. In addition, to capture the volatility persistence, we also apply long-memory FIGARCH and HAR models.

### 1.4.2 Forecasting methodology

In order to examine the appropriateness of the volatility models, we test their performance in predicting the volatility of crude oil futures prices. First, we estimate each model with data from January 03, 1991 to December 30, 2001, retaining the last five years of the sample from December 1, 2001 to December 14, 2006 for out-of-sample forecast comparisons. To obtain 1-, 5-, and 22-step-ahead forecasts on day  $t + 1$ ,  $t + 5$ , and  $t + 22$ , respectively, the estimation period is rolled forward by adding one, five, and twenty two new day(s) and dropping the most distant day(s). In this way the sample size used in estimating the models stays at a fixed length and the forecasts do not overlap. With longer forecast horizons, the error term will be serially correlated. Thus, we rely on Newey-West heteroskedasticity consistent covariance matrix estimator with 5, 10, and 44 lags for the daily, weekly, and monthly regression estimates, respectively.

In the context of volatility forecasting, several statistical procedures have been used for assessing the quality of competing forecasts. In our study, to compare the accuracy of out-of-sample volatility forecasts from different models against the realized ex post volatility obtained using range-based estimators, we use the mean squared error (MSE) and the mean absolute error (MAE). Their loss functions are defined as:

$$MSE = \frac{1}{T} \sum_{t=1}^T (\sigma_{a,t}^2 - \sigma_{f,t}^2)^2 \quad (1.14)$$

$$MAE = \frac{1}{T} \sum_{t=1}^T |\sigma_{a,t}^2 - \sigma_{f,t}^2| \quad (1.15)$$

where  $T$  is the number of out-of-sample forecasts,  $\sigma_{a,t}^2$  is the realized ex post volatility measured using range-based estimators, and  $\sigma_{f,t}^2$  is the one-step-ahead forecast obtained from either GARCH, EGARCH, FIGARCH or HAR models.

The MSE and MAE statistics allow to sort among models based on their out-of-sample forecasting accuracy. However, they do not formally test whether improvements in forecast accuracy between two models are statistically significant. Several procedures to formally test for the statistical significance of the observed differences in the MSE and

MAE criteria and the predictive ability of the underlying forecasting models have been proposed in the literature. One of them is the Diebold and Mariano (1995) test, which involves a pairwise comparison of the forecasts from each model.

Diebold and Mariano (1995) proposed a test of equal predictive ability of two competing models based on the null hypothesis of no difference in the predictive accuracy of the two forecasts. Let  $e_{1t}$  and  $e_{2t}$  denote forecast errors from two competing models and  $g(e_{1t})$  and  $g(e_{2t})$  represent their associated loss functions. The loss differential is then:  $d_t = g(e_{1t}) - g(e_{2t})$ . The Diebold-Mariano (DM) test statistic is given by:

$$DM = \frac{\bar{d}}{\sqrt{\hat{V}(\bar{d})}} \quad (1.16)$$

where  $\bar{d} = \frac{1}{T} \sum_{t=1}^T d_t$  is the sample mean loss differential and  $\hat{V}(\bar{d})$  is the asymptotic variance of  $\bar{d}$ . Assuming the sequence  $\{d_t\}_{t=1}^T$  is covariance stationary, under the null hypothesis of  $E(d_t) = 0$ , the asymptotic distribution of the sample mean loss differential  $\bar{d}$  is  $\sqrt{T}(\bar{d} - \mu) \rightarrow N(0, V(\bar{d}))$ . The null hypothesis, that the benchmark model 1 has equal predictive ability with model 2, is investigated against the alternative hypothesis that the benchmark model has superior predictive ability. Thus, it is rejected if the test statistic is negative and statistically significant.

## 1.5 Data description and summary statistics

The dataset for this study consists of observations on the daily closing, opening, high and low prices of West Texas Intermediate (WTI) crude oil futures. These futures contracts are traded on New York Mercantile Exchange (NYMEX), a subsidiary of the Chicago Mercantile Exchange (CME). Contract specifications and trading details, which are available from the NYMEX website, are provided in Table 1.1. The dataset covers the period from January 3, 1991 through December 14, 2006. We split the full sample period into an in-sample estimation period from January, 1991 through December, 2001, and an out-of-sample forecast evaluation period from December, 2001 through December 2006. The time period covered by our sample makes the crude oil futures market an ideal setting for testing alternative models for volatility, given the shocks that occurred during this period (e.g., Asian financial crisis, Russian default, Long Term Capital Management, terrorist attacks on the World Trade Center in New York, invasion in Iraq, etc.).

Table 1.1: Contract specifications: light, sweet crude oil futures traded at NYMEX.

Product Symbol	CL
Contract Unit	1,000 barrels
Price Quotation	U.S. dollars and cents per barrel
Minimum fluctuation	\$0.01 per barrel
Expiration of Trading	Trading in the current delivery month ends on the third business day prior to the 25th calendar day of the month preceding the delivery month. If the 25th calendar day of the month is a non-business day, trading ends on the third business day prior to the last business day preceding the 25th calendar day.
Listed Contracts	Crude oil futures are listed 9 years forward as follows: consecutive months are listed for the current year and the next five years; in addition, the June and December contract months are listed beyond the sixth year.
Settlement Type	Physical

From the dataset we construct a daily price series of the nearby futures contract, which is typically the most actively traded. The nearby contract is the futures contract that is closest to maturity. We roll over the contract month one week before the contract expiration day. To eliminate any effect at the time of rollover, we compute daily futures returns using the price data from the identical contract. Therefore, on the day of rollover, we gather futures prices for both the nearby and the first-deferred contracts, so that the daily return on the day after rollover is measured with the same contract month.<sup>6</sup>

There are several advantages to using futures prices that have been shown in the literature (see, e.g., Fleming et al. (1996), Alizadeh et al. (2002)). First, since they result from open outcry, the transactions are open to all the participants in the market and orders are executed at the best price. Second, the closing, or settlement, prices are very scrutinized as they are used for marking to market the account balances. Finally, the lower transaction costs in the futures markets allow faster and easier arbitrage.

Figure 1.1 displays the evolution of prices for the NYMEX WTI crude oil futures. The impact of several economic and geopolitical events is evident in this graph. From 1991 until 1995 the oil prices were relatively stable. In the period from 1997 through 1998, there is a downward trend due to tension in the Middle East and the currency crisis in East Asian countries. However, in early 1999 OPEC cut down production and prices

<sup>6</sup>The estimation results show that the choice of timing for rolling over futures contracts has insignificant effects.

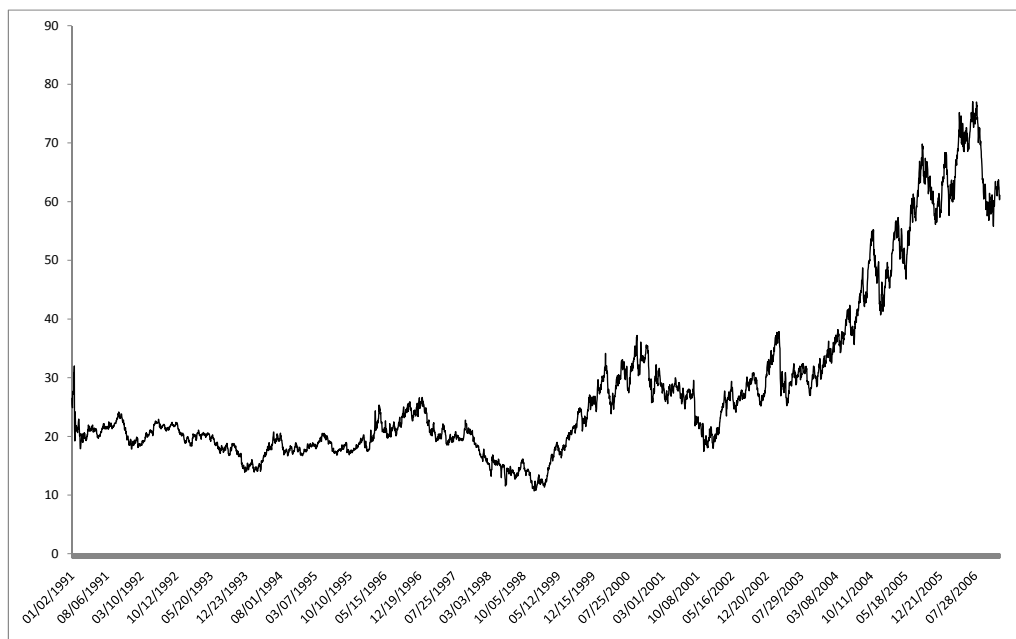


Figure 1.1: Dynamics of daily prices of WTI crude oil futures market (US Dollar/Barrel)  
- January 3, 1991 to December 14, 2006.

started to increase again. In 2001 there is again a downward trend, and starting from September 2001, following the terrorist attacks, the prices have been going up. The US military action in Iraq in 2003 contributed further to this increase in oil prices.

Next, all daily prices are converted into daily percentage return series in the following way. Let  $C_t$  denote the closing price of the crude oil futures on day  $t$ . The return at time  $t$  is then  $r_t = 100 \times \ln(C_t/C_{t-1})$  for  $t = 1, 2, \dots, T$ . We first discuss the properties of daily returns and then investigate the volatility clustering. If volatility clustering is confirmed, then the use of the GARCH family models is justified.

Table 1.2: Descriptive statistics of daily crude oil futures returns. The table presents descriptive statistics of the crude oil futures returns: sample average, standard deviation, skewness, kurtosis, minimum value, maximum value. JB is the Jarque and Bera (1980) test statistic for the null hypothesis of normality in the sample returns distribution. ADF and PP are the statistics of the augmented Dickey-Fuller and Philips-Perron unit root tests, respectively. ARCH(5) is the Engle (1982) test for 5th order ARCH effects. p-values are shown in parentheses. The sample period is from January 3, 1991 to December 14, 2006.

Statistic	Returns
Mean	0.0177
Std. Dev.	1.0325
Skewness	-0.2560
Kurtosis	6.0309
Maximum	6.1804
Minimum	-7.1852
JB stat	2590
p-value	(0.000)
ADF stat	-46.871
p-value	(0.000)
PP stat	-62.899
p-value	(0.000)
ARCH(5)	26.214
p-value	(0.000)

Basic descriptive statistics for returns series is presented in the Table 1.2. The daily mean return (0.0177%) is very small in comparison to its standard deviation (1.0325%). Crude oil futures returns conform to several stylized facts, which have been extensively documented in the literature for many financial markets. The coefficient of kurtosis,

which measures the peak of a distribution relative to the normal one, indicates that returns are characterized by fat tails. This can also be seen from graphs of the density functions in Figure 1.2.

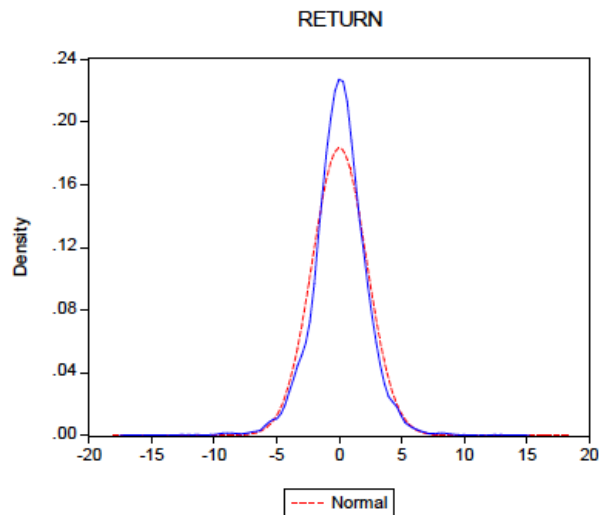


Figure 1.2: Crude oil futures return distributions. The figure shows kernel estimates of the density of daily crude oil futures returns. The sample period is from January 3, 1991 through December 14, 2006. The solid line is the estimated density of raw returns and the dashed line is a  $N(0, 1)$  density for visual reference.

Coefficient of skewness is different from zero and negative for returns indicating that there is an asymmetry of the probability distribution, such that the returns series is skewed to the left. Thus, both skewness and kurtosis coefficients of the series indicate that the distributions of daily returns are non-Gaussian.<sup>7</sup> This is also confirmed by the Jarque-Bera (1980) test statistics, which rejects the null hypothesis that the sample returns come from a normal distribution. QQ plot of crude oil returns in Figure 1.3

<sup>7</sup>For comparison, the skewness coefficient is equal to zero and the kurtosis coefficient is equal to three for a normally distributed random variable.

indicates that both large positive and large negative shocks are responsible for the non-normality of the series.

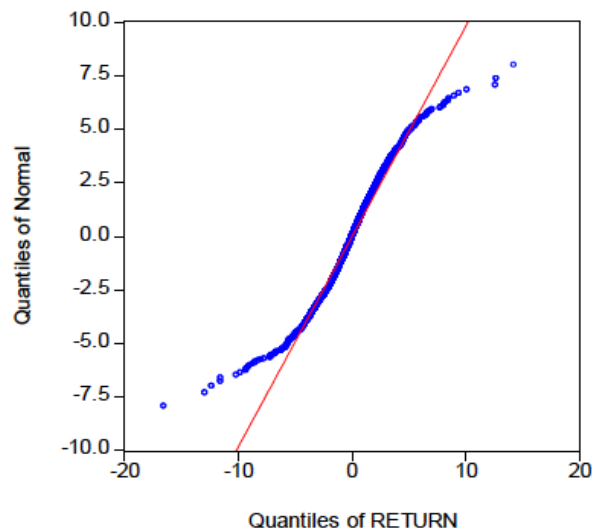


Figure 1.3: QQ plot of the crude oil futures returns. This graph shows the QQ plot, which is a graph of the quantiles of the standardized residual distribution against the corresponding quantiles of a normal distribution. If the returns are normally distributed, their QQ plot is a straight line with a unit slope.

Table 1.2 also presents two unit root tests for returns series - the Augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) statistics. These two statistics test for a unit root in the univariate representation of a time series and are calculated with an intercept in the test regression. The null hypothesis is a non-stationary time series and the alternative hypothesis is a stationary time series. Both the ADF and PP test statistics reported in Table 1.2 reject the hypothesis of a unit root in the return series at the 1% significance level, implying that the return series are stationary, and thus, may be modeled directly without further transformations.

In order to test for the presence of the ARCH effect and justify the appropriateness of

GARCH-type models, we compute and report the statistics associated with the ARCH-LM test in Table 1.2. To test the null hypothesis that there is no ARCH effect, we run the regression of the squared residuals on a constant and lagged squared residuals up to order five. We get significant results for the returns series, which confirms the presence of conditional heteroskedasticity and indicates the use of GARCH-type models.

Table 1.3: Autocorrelations of crude oil futures squared returns.  $\rho$  is the autocorrelation for up to the 10th order serial correlation computed on the standardized returns. Q-stat is the Ljung-Box test statistics for testing the joint significance of autocorrelations of squared returns. The sample period includes daily data from January 3, 1991 to December 14, 2006.

Lags	$\rho$	Q-stat	p-value
1	0.148	87.453	0.000
2	0.078	111.80	0.000
3	0.121	169.96	0.000
4	0.060	184.54	0.000
5	0.077	208.37	0.000
6	0.067	226.45	0.000
7	0.043	233.92	0.000
8	0.033	238.35	0.000
9	0.024	240.61	0.000
10	0.040	246.97	0.000

The existence of volatility clustering in daily, weekly, or monthly returns has been extensively documented in the literature. This feature is also confirmed in Figure 1.4 which shows the time series of daily crude oil futures returns. The graph of the returns series provides an evidence of volatility clustering, when periods of high volatility are followed by high volatility, while periods of low volatility are followed by low volatility. In other words, the variance of the forecast errors depends on the size of the preceding disturbance. This behavior is known to frequently occur in financial markets.

Statistically, volatility clustering implies a strong autocorrelation in squared returns; thus, a simple method for detecting volatility clustering is to calculate the first-order autocorrelation coefficient in squared returns. To test the joint hypothesis that sample autocorrelations are simultaneously equal to zero, the Ljung and Box (1978) portmanteau tests for up to tenth-order serial correlation in the squared returns is used. The values of Q-test statistic (reported in Table 1.3) show that the null hypothesis of no autocorrelation

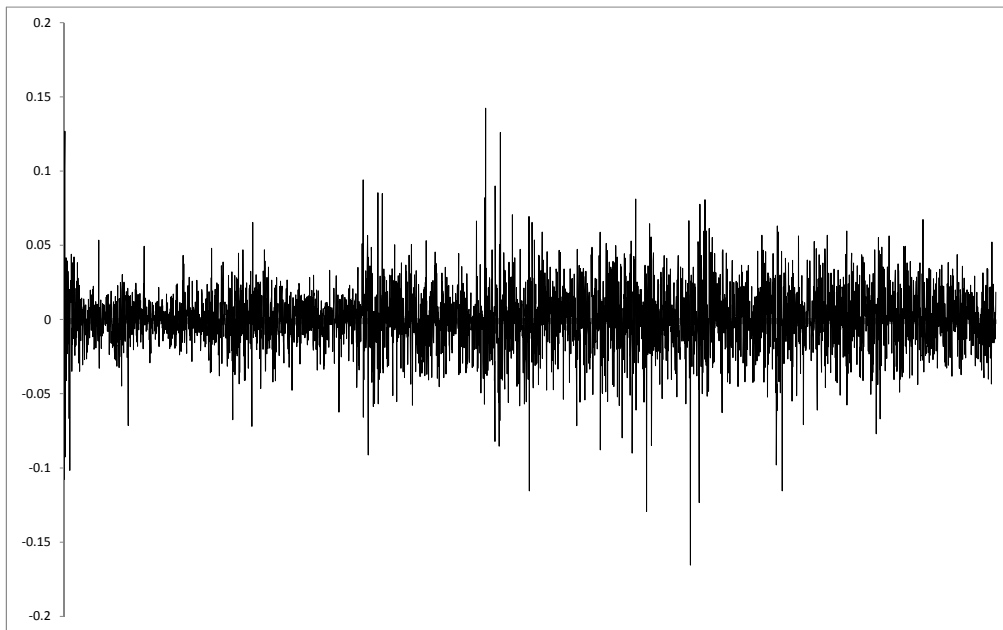


Figure 1.4: Dynamics of daily returns of WTI crude oil futures market (US Dollar/Barrel)  
- January 2, 1991 to December 14, 2006.

for up to the 10th order is rejected, indicating volatility persistence in the crude oil futures returns, which can be further captured by the long-memory models.

## 1.6 Empirical results

### 1.6.1 Volatility measures

Table 1.4 presents descriptive statistics for the log absolute returns and the log range, defined as the difference between the highest and the lowest prices in logarithms. First, the log range is preferable in terms of its smaller standard deviation. Second, the skewness and kurtosis of the log range are 0.243 and 3.589, respectively, and are closer to the corresponding values of 0 and 3 for a normal random variable compared to those for log absolute returns. This conclusion is confirmed by checking the Jarque-Bera statistic, which is also much smaller for the log range.

Table 1.4: Descriptive statistics of log range and log absolute returns. The table presents the descriptive statistics of the log absolute returns and the log range, defined as the difference between the highest and the lowest prices in logarithms. JB is the Jarque-Bera test statistic for the null hypothesis of normality. Q(20) is the Ljung-Box (1978) Q statistics for 20th order sample autocorrelation in the residuals. p-values are given in parentheses. The sample period is from January 1, 1991 through December 14, 2006.

Statistic	Log range	Log absolute returns
Mean	-3.516	-4.443
Std dev	0.421	1.043
Skewness	0.243	-0.848
Kurtosis	3.589	3.780
JB t-stat	66.610 (0.000)	394.993 (0.000)
Q(20)	755.43 (0.000)	20.710 (0.001)

Next, we estimate the daily volatility estimates from the closing prices and then from high, low, opening and closing prices using Parkinson, Garman-Klass, and Rogers-Satchell volatility estimators. Figure 1.5 exhibits these volatility estimates over the period from January 1991 through December 2006. The volatility estimates based on closing

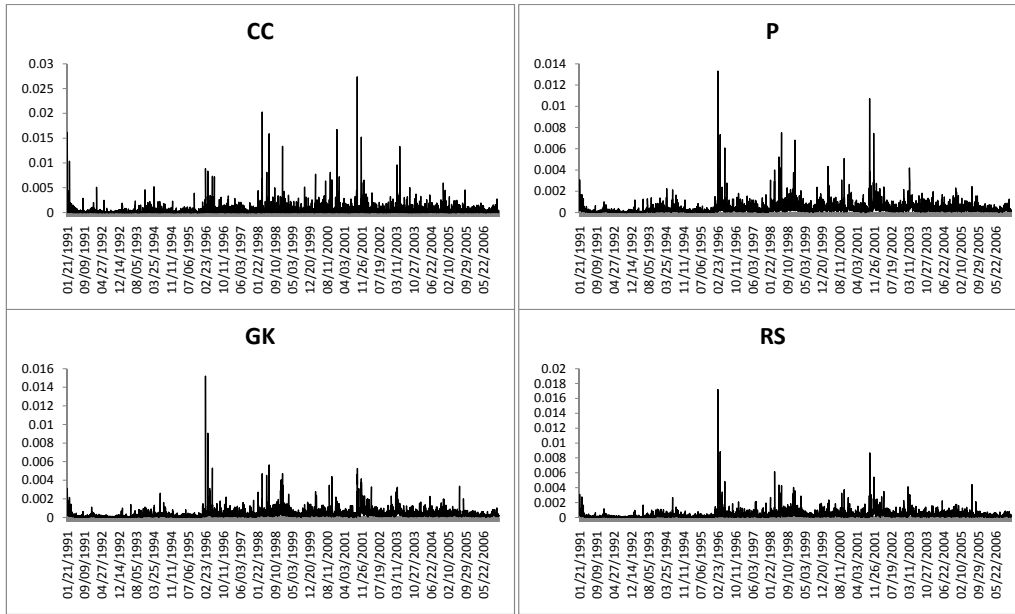


Figure 1.5: Range-based vs. traditional close-to-close volatility estimators in the crude oil futures market. The figure presents different volatility estimators: traditional close-to-close (CC), Parkinson (P), Rogers-Satchell (RS) and Garman-Klass (GK). The sample period is from January 3, 1991 to December 14, 2006.

prices only are far more variable than the other estimates, which can be attributable to the smaller efficiency of this estimator. All estimates are highly correlated and appear to move in a similar way through time, as well as to cluster in different periods at the same time.

Table 1.5: Correlation between volatility estimators. This table presents pairwise correlations between close-to-close (C), Parkinson (P), Garman-Klass (GK), and Rogers-Satchell (RS) estimators. The sample period is from January 3, 1991 to December 14, 2006.

	Correlation coefficients			
	C	P	GK	RS
C	1	0.732	0.531	0.382
P		1	0.928	0.816
GK			1	0.965
RS				1

Table 1.5 reports the pairwise correlation coefficients between various volatility estimators. The correlation coefficients provide strong evidence that all range-based volatility estimates are highly correlated with each other, which suggests that range-based measures provide similar views of volatility patterns. The correlations between traditional estimator and range-based volatilities are relatively weaker, which is consistent with higher noise contained in standard deviations.

Next, we present in Table 1.6 descriptive statistics for the range-based volatility estimates. First, the volatility measures are all skewed and exhibit excess kurtosis. The Ljung-Box test statistics show that there is substantial serial correlation among realized volatility estimates, with the magnitudes of the autocorrelations from one to five lags all highly significant. Garman and Klass estimator has the smallest standard deviation, whereas Parkinson estimator has the smallest coefficients of skewness and kurtosis. Compared with the small autocorrelations of squared returns due to measurement errors, the large and slowly decaying autocorrelations of the range-based estimates provide evidence of strong volatility persistence.

Finally, Table 1.7 reports the mean squared errors (MSE) of the various volatility estimators. We consider the estimator with the lowest MSE as the most efficient. In general, each of the three range-based estimators outperform the traditional log of the squared returns measure. Our analysis suggests that the Garman and Klass estimator

Table 1.6: Descriptive statistics of the volatility measured by different estimators. The table presents sample average, standard deviation, coefficients of skewness and kurtosis, and autocorrelations up to the fifth lag for the following estimators: traditional close-to-close estimator (C), Parkinson (P), Rogers-Satchell (RS), and Garman-Klass (GK). The sample period is from January 1, 1991 through December 14, 2006.

	Realized volatility			
	C	P	GK	RS
<i>Statistic</i>				
Mean	0.000474	0.000383	0.000378	0.000374
Std. Dev.	0.001152	0.000557	0.000523	0.000555
Skewness	9.599212	8.616663	9.644215	10.99818
Kurtosis	149.0755	138.2569	198.7595	252.3184
<i>Autocorrelation</i>				
lag(1)	0.149	0.278	0.289	0.248
lag(2)	0.080	0.273	0.273	0.213
lag(3)	0.121	0.233	0.233	0.192
lag(4)	0.060	0.176	0.195	0.143
lag(5)	0.077	0.190	0.194	0.169

is consistently the most efficient range-based volatility measure in the crude oil futures market.<sup>8</sup>

## 1.6.2 In-sample estimation results

Table 1.8 presents the in-sample results of estimation of the GARCH(1,1), GARCH-in-mean, EGARCH, GJR-GARCH, APARCH and FIGARCH models for the crude oil futures market using range-based estimates as volatility proxy. We estimate the selected models using the quasi-maximum likelihood estimation technique and assuming normally distributed errors (see Bollerslev and Wooldridge (1992)). Values in parentheses are standard errors of corresponding parameter estimates.  $\text{Log}(L)$  is the value of maximized Gaussian log likelihood function. Additional indicators and diagnostic tests on standardized residuals are computed in order to detect the best time-series model for

<sup>8</sup>This result is in contrast to the studies of Brown (1990) and Alizadeh (1998), where they argue that opening and closing prices are highly influenced by microstructure effects and there is little efficiency gain from including them in volatility estimation. However, our results are in line with Bali and Weinbaum (2005), who for S&P 500 futures market find that the Garman and Klass estimator is the single best-performing estimator.

Table 1.7: Efficiency of the range-based volatility estimators. The table reports the mean squared error (MSE) of different volatility estimators.  $\hat{\sigma}_C$  is the close-to-close estimator,  $\hat{\sigma}_P$  is the Parkinson estimator,  $\hat{\sigma}_{GK}$  is the Garman and Klass estimator, and  $\hat{\sigma}_{RS}$  is the Rogers and Satchell estimator.

Volatility	MSE
$\hat{\sigma}_C$	0.0081
$\hat{\sigma}_P$	0.0041
$\hat{\sigma}_{GK}$	0.0039
$\hat{\sigma}_{RS}$	0.0056

a given data set.

The coefficient  $\alpha$  captures influence of new shocks on volatility. Estimates of this parameter are statistically significant and positive for all five models. The parameter  $\beta$ , which measures persistence of volatility shocks, is positive, close to 1 and statistically significant at 1% level for all model specifications. The sum of the coefficients on the lagged squared error and lagged conditional variance is close to one. This implies that shocks to the conditional variance are highly persistent. In case of GARCH-M specification, the estimated parameter on the mean equation has a negative sign, but it is not statistically significant. The coefficient of the variance in the mean equation is not statistically significant either. We would thus conclude that in the crude oil futures market, there is no feedback from the conditional variance to the conditional mean.

The asymmetry coefficient  $\gamma$  is not statistically significant for either EGARCH or GJR-GARCH specifications. It is not statistically significant in the APARCH model either. Therefore, there is no strong evidence that shocks have asymmetric effects on the volatility of crude oil futures prices. Moreover, in EGARCH model, the coefficient estimate is negative, suggesting that positive shocks imply a higher next period conditional variance than negative shocks of the same sign. This is in contrast to what is observed in the stock markets.

The value of the power coefficient  $\delta$  is 1.892 for the APARCH model. The null hypothesis is rejected for  $\delta = 1$ , but not for  $\delta = 2$  at the 5% significance level. These results indicate that the crude oil futures series are better modeled with conditional variance rather than conditional standard deviation. Finally, the fractional difference parameter  $d$  reported in Table 1.8 is significant in the FIGARCH model and equals to

0.392, implying a significant degree of long-memory volatility in the crude oil futures market. Both AIC and BIC statistics are the smallest for FIGARCH model.

In order to test whether each model has adequately captured the persistence in volatility and there is no ARCH effect left in the residuals from the models, ARCH-LM test was conducted for lags up to five. The tests indicate that the standardized residuals do not exhibit any ARCH effects at 1% significance level for GARCH(1,1), GARCH(1,1)-M, GJR-GARCH, and APARCH models and at 10% for EGARCH and FIGARCH models. Also Ljung-Box test on squared residuals indicates that squared residuals are independent, which is a necessary condition to conclude that the models are adequate.

Next, Table 1.9 reports the results of the estimation of the HAR model of Equation (1.13) for the crude oil futures market. Here we consider four proxies of volatility: close-to-close (C), Parkinson (P), Garman-Klass (GK) and Rogers-Satchell (RS). From the values of the  $t$ -statistics, it is clear that the realized volatility aggregated over the three different horizons is highly significant except for the coefficient of daily realized volatility obtained using closing prices only. According to Corsi (2009), the noisier estimation of the daily realized volatility induces a lack of significance of the daily volatility component, while weekly and monthly realized volatilities, being averages over longer periods, contain less noise and more information in the volatility process.

As can be seen from Table 1.9, the estimates for  $\beta_D$ ,  $\beta_W$ , and  $\beta_M$  confirm the existence of highly persistent volatility dependence. The relative importance of the daily volatility component increases from the daily to the weekly to the monthly regressions, whereas the monthly volatility component tends to be relatively more important. Thus, the HAR model is capable of capturing the long-memory component.

### 1.6.3 Out-of-sample forecasting results

To assess the accuracy of the model forecasts used in this study, we compare the predictions to the actual realized volatility measures; i.e., for the one-day horizon forecasts. In addition to the one-day forecasts, we also calculate one-week and one-month forecasts defined by the average of the forecasts from 1 to 5, and 1 to 22 days-ahead, respectively.

Here we consider the out-of-sample forecasting performance of the GARCH(1,1), EGARCH, FIGARCH, and HAR models. Table 1.10 reports the Mean Square Error (MSE) and Mean Absolute Error (MAE) for the forecasts from each of the four different models based on the data over the out-of-sample period and for 1-, 5-, and 22-day

forecast horizons. The out-of-sample results essentially confirm the earlier in-sample findings. For one-step-ahead predictions, the superiority of GARCH(1,1) model holds across both MAE and MSE criteria. The worst model is the EGARCH model, which is dominated by all other models across all horizons.<sup>9</sup> In the multistep ahead case, long-memory models tend to outperform GARCH models. The results suggest that the HAR model has superior forecasting ability than the other three models at 5- and 22-day forecast horizons.

We now proceed by performing pairwise comparison of models with the Diebold-Mariano (DM) test statistic. Since there is no evidence of asymmetric effect in the crude oil futures market and, as a result, the EGARCH model performs the worst across all horizons, we will not consider it. Thus, we will compare the predictive ability of a simple GARCH, FIGARCH and HAR model. In Table 1.11, DM test results with forecast errors measured by MAE and MSE and across all horizons are reported. The null hypothesis is that of equal predictive accuracy of the two models - model 1/model 2. The statistical significance of the difference is indicated by the number of asterisks attached to the statistic. A significantly positive (negative)  $t$ -statistic indicates that the model 2 is dominated (dominates) the competitor model 1.

First, Table 1.11 shows that the MAE superiority of the HAR forecasts over the GARCH(1,1) and FIGARCH forecasts is significant at the 1% level for longer horizons. For one-step ahead forecasts, we reject the null of equal forecast accuracy between the GARCH model and both HAR and FIGARCH models implying that the GARCH is significantly superior to HAR and FIGARCH one-step ahead forecasts. The above conclusions hold also in terms of MSE comparisons. For relatively short forecast horizons, the GARCH(1,1) and HAR yield equally good forecasts in MSE terms, but as  $h$  increases to 22, it is apparent that the HAR is the superior choice of the two. These findings suggest that the HAR forecasts outperform those of FIGARCH and GARCH forecasts for multiple-step ahead forecasts.

In summary, the GARCH models are useful for short-run (one-day) volatility forecasting. However, the models, which can capture long-memory volatility, exhibit greater forecasting accuracy than the linear ones, especially in volatility forecasting over longer horizons, such as weekly or monthly.

---

<sup>9</sup>This result is consistent with Day and Lewis (1993), who also found that EGARCH does not outperform the simple GARCH model in forecasting volatility in the crude oil futures market.

## 1.7 Conclusion

One of the most frequent questions that is investigated in the finance literature is which model to be used for volatility forecasting. In this study, we have evaluated the relative forecasting performance of different volatility models, at various horizons, for the light, sweet crude oil futures market. As a volatility proxy, based on which the forecasting performance was examined, we use range-based volatility estimators based on high, low, opening, and closing prices, which seem to outperform the traditional log of squared returns volatility measure. The out-of-sample forecasting performance of these models is compared using standard forecast summary statistics such as mean squared error (MSE), mean absolute error (MAE), and Diebold and Mariano (1995) test statistics. We estimate most commonly used GARCH specifications, including GARCH(1,1), GARCH-M, EGARCH, GJR-GARCH, and APARCH models. In addition, to account for volatility persistence, we estimate FIGARCH and HAR models.

Our findings can be summarized as follows. In line with previous studies, we find that volatility is a highly persistent process for the crude oil futures market. However, counter to a number of previous studies, our results show that the asymmetric effect is not strong in the crude oil futures market. Also we do not find any evidence of the risk-return tradeoff. For the case of one-step ahead forecasts, a simple GARCH(1,1) performs at least as well as the models that take into account the long-memory component. Such a finding is robust to the choice of the loss function. However, the results suggest that the HAR model has superior forecasting ability than the other models at longer forecast horizons.

Several important questions remain. First, although the statistical evidence in favor of long memory appear quite strong, there are alternative hypotheses that must be addressed. Long memory also has potential links to regimes and structural breaks in volatility. In particular, it may be the case that the volatility process is subject to occasional structural breaks, which are known to induce spurious indications of strong persistence in the data.

Second, the high-frequency intradaily returns contain more information about financial market volatility beyond that available from daily or lower frequency observations. Thus, the high-frequency returns may be used to construct improved ex-post interdaily return volatility measurements, which may further improve the ability to separate short-run and long-run volatility components and allow for improved estimates of the apparent

long-memory features in return volatility.

Third, return volatility is often modeled as driven by the random arrival of new information within so-called mixture-of-distribution characterization of the price process (e.g., Clark (1973), Harris (1987), Andersen (1996)). Thus, it would also be interesting to examine whether other variables, such as trading volume, open interest or market liquidity are related to the return volatility process in the crude oil futures market, as for example, in the work by Gallant et al. (1992).

Finally, it would be interesting in the future research to explore modeling and forecasting volatility and correlation between energy markets, as well as energy and other financial markets. All these issues await future research.

Table 1.8: In-sample estimation results of GARCH-type volatility models. The table presents the parameter estimates for the GARCH(1,1), GARCH-M, GJR-GARCH, EGARCH, APARCH, and FIGARCH models. Values in parentheses are standard errors of corresponding parameter estimates. Log(L) is the logarithm of the maximum likelihood function value. AIC is the average Akaike information criterion. BIC is the Bayesian information criterion. Q(10), Q(15), and Q(20) are the LjungBox Q-statistic computed on the standardized residuals of order 10, 15, and 20, respectively. ARCH(5) is the non-heteroskedasticity statistic of order 5.  $p$ -values of the statistics are reported in brackets. \*\* and \* denote significance at the 5% and 1% levels, respectively.

	GARCH	GARCH-M	GJR-GARCH	EGARCH	APARCH	FIGARCH
$\lambda$	-	1.155	-	-	-	-
	-	(1.132)	-	-	-	-
$\omega$	0.048*	0.047*	0.047*	-0.206*	0.038*	0.245*
	(0.000)	(0.000)	(0.000)	(0.034)	(0.011)	(0.075)
$\alpha$	0.062*	0.062*	0.058*	0.127*	0.065*	-
	(0.009)	(0.009)	(0.014)	(0.020)	(0.008)	-
$\beta$	0.929*	0.930*	0.930*	0.986*	0.930*	0.513*
	(0.008)	(0.008)	(0.009)	(0.003)	(0.009)	(0.073)
$\gamma$	-	-	0.007	-0.019	0.024	0.012
	-	-	(0.016)	(0.013)	(0.042)	(0.024)
$\delta$	-	-	-	-	1.892*	
	-	-	-	-	(0.228)	
$\varphi$	-	-	-	-	-	0.235
	-	-	-	-	-	(0.0062)
$d$						0.422
	-	-	-	-	-	(0.055)
Log(L)	9858	9859	9858	9868	9860	9856
Q(10)	13.829	13.784	13.839	14.852	13.840	12.967
Q(15)	21.627	21.577	21.729	22.607	21.739	20.984
Q(20)	23.283	23.249	23.401	24.396	23.411	23.298
AIC	-4.946	-4.946	-4.945	-4.951	-4.946	-4.938
BIC	-4.939	-4.938	-4.937	-4.943	-4.939	-4.929
ARCH(5)	0.826	0.826	0.800	2.047	0.815	0.679

Table 1.9: In-sample estimation results of HAR model. The table presents the parameter estimates for the HAR model, using traditional close-to-close (C), Parkinson (P), Garman and Klass (GK), and Rogers and Satchell (RS) volatility estimators. Values in parentheses are the standard errors of corresponding parameter estimates. AIC is the average Akaike information criterion. BIC is the Bayesian information criterion. \*\*\*, \*\* and \* denote significance at the 10%, 5% and 1% levels, respectively.

	CC	P	GK	RS
Intercept	0.0002* (0.000)	$8.99 \times 10^{-5}$ * (0.000)	$7.66 \times 10^{-5}$ * (0.000)	$8.23 \times 10^{-5}$ * (0.000)
$\beta^d$	0.021 (0.018)	0.087** (0.019)	0.091** (0.019)	0.078** (0.018)
$\beta^w$	0.175*** (0.041)	0.279** (0.037)	0.241** (0.038)	0.192** (0.039)
$\beta^m$	0.341* (0.059)	0.396* (0.045)	0.463* (0.044)	0.507* (0.047)
AIC	-4.946	-4.946	-4.951	-4.945
BIC	-4.939	-4.938	-4.943	-4.937
$R^2$	0.037	0.151	0.164	0.134

Table 1.10: Out-of-sample forecasting performance. This table reports the Mean Square Error (MSE) and Mean Absolute Error (MAE) for the 1-, 5-, and 22-day forecasts obtained from GARCH, EGARCH, FIGARCH and HAR models.

Horizon	Loss function	GARCH	EGARCH	FIGARCH	HAR
1	MAE	0.01379	0.01533	0.1469	0.01429
	MSE	0.00032	0.00050	0.00042	0.00034
5	MAE	0.01748	0.01885	0.01537	0.01302
	MSE	0.00063	0.00072	0.00054	0.00042
22	MAE	0.2018	0.2231	0.1847	0.1724
	MSE	0.00081	0.00092	0.00078	0.00075

Table 1.11: Diebold-Mariano test results. This table reports Diebold-Mariano (DM) test statistics with forecast errors measured by MAE and MSE for 1-, 5-, and 22-day out-of-sample forecasts. \*\*\*, \*\* and \* denote significance at the 10%, 5% and 1% levels, respectively.

MAE	1	5	22
GARCH/HAR	-6.12*	-5.08*	-6.96*
HAR/FIGARCH	-7.10*	-6.12*	-8.21*
FIGARCH/GARCH	-7.55*	-6.44*	-7.92*
MSE	1	5	22
GARCH/HAR	-3.08*	-2.12*	-3.15*
HAR/FIGARCH	-4.23*	-3.02*	-3.42*
FIGARCH/GARCH	-4.76*	-3.92*	-4.12*



## Chapter 2

# Information content of implied volatility in the crude oil futures market

## 2.1 Introduction

Volatility is an important input in many finance applications, such as asset pricing, portfolio selection, and risk management, and it is hardly surprising that great effort has been made to obtain reliable volatility forecasts. Recently, there has been much interest in obtaining improved volatility estimates from observed derivatives prices. This is because the option's current market price in an informationally efficient market should accurately reflect all currently available information beyond that contained in asset's price history. And if the correct option pricing model is used, then implied volatility should provide an optimal forecast of future volatility of underlying asset through option expiration and subsume any forecast generated by time-series models based on historical information.

Assessing the information content and forecasting performance of implied volatility has been an ongoing research issue and prior empirical studies provide mixed evidence. In general, earlier studies document that implied volatility is not an efficient predictor of future volatility (e.g., Day and Lewis (1992), Lamoureux and Lastrapes (1993), Canina and Figlewski (1993)). However, more recent studies suggest that econometric problems may compromise inference regarding the forecast quality of implied volatility. Overcoming these problems, they conclude that implied volatility still outperforms other volatility measures in forecasting future volatility (e.g., Jorion (1995), Christensen and Prabhala (1998)). Although the results of these studies are contradictory, most authors find that option implied volatilities provide efficient, but biased, forecasts of future volatility.

Traditionally, implied volatility has been backed out from Black and Scholes (1973) option pricing model, which is based on the assumptions that are in contrast with actual market observations.<sup>1</sup> Thus, the implied volatility computed from Black-Scholes model does not reflect the true expectation of future realized volatility, and tests, based on the Black-Scholes implied volatility, are rather joint tests of market efficiency and the model misspecification. Due to these shortcomings of the Black-Scholes model, an enormous research concentrates on developing option pricing models that could account for stochastic volatility and thus reflect market prices of options more accurately.

One alternative that has been suggested in the literature is the model-free implied volatility, that does not depend upon any model and parametric assumptions about un-

---

<sup>1</sup>For example, one of the assumptions of the Black-Scholes model is constant volatility. However, it is well known, that options of the same time to maturity, but different strikes, appear to produce different implied volatility estimates for the same underlying asset, so called volatility smile.

derlying volatility dynamics.<sup>2</sup> Therefore, the measurement errors resulting from model misspecification are reduced. In addition, the model-free implied volatility gathers information contained not only in near-the-money options, but from a cross-section of put and call options with strikes spanning the full range of possible values for the underlying asset. In fact, this methodology with some modifications has been adopted by the Chicago Board Options Exchange (CBOE) in constructing the monthly volatility index, named VIX, from S&P 500 index option prices.<sup>3</sup> The VIX has been widely followed by financial market participants and is considered not only to be the market's expectation of the volatility in the S&P500 index over the next month, but also to reflect investor sentiment and risk aversion.

Most of the studies on information content of the implied volatility have focused on stock, bond or foreign exchange markets, in which very often the options and underlying assets do not trade on the same exchange and their closing prices are not synchronous. In addition, transaction costs for index options or individual stock options are non negligible and these transaction costs may allow market prices to diverge significantly from theoretical prices (Canina and Figlewski (1993)). The goal of this study is to assess the predictive ability of implied volatility extracted from the light, sweet crude oil futures options traded at the New York Mercantile Exchange (NYMEX), where options on futures and underlying futures contracts are traded side by side on the same floor and their prices are observed simultaneously, which reduces the measurement errors.

Our focus is on the light, sweet crude oil futures market, because it deserves attention for several reasons. According to the NYMEX, this futures contract has become the world's largest-volume and most liquid futures contract trading on a physical commodity. Since one of the characteristics of prices in the oil markets is volatility, which is both relatively high and variable over time, it provides a good opportunity to evaluate the speed with which new information is incorporated in implied volatility relative to other forecasts of volatility. In addition, since they are futures options, they carry very low transaction costs for arbitrage activity.

We estimate the implied volatility using model-free methodology, which aggregates information on market volatility in the term structure of option prices. To the best of our knowledge, this is one of the first studies to test the efficiency of the model-free implied

---

<sup>2</sup>See, e.g., Carr and Madan (1998) and Britten-Jones and Neuberger (2000).

<sup>3</sup>Until 2003, the VIX had been based on S&P100 options, rather than S&P500, and calculated using Black-Scholes formula, rather than model-free approach.

volatility in the crude oil futures market. We evaluate the information content and forecasting accuracy of model-free implied volatility relative to the information contained in alternative volatility forecasts in the crude oil futures market. The in-sample information content of implied volatility is examined by adding the implied volatility from the options as an exogenous variable in GARCH specifications. Then, the out-of-sample forecasting power of implied volatility is compared to forecasts based on past historical information and Black-Scholes implied volatility.

We summarize our findings as follows. First, for the crude oil futures market, the model-free implied volatility contains important information for predicting future volatility. We find that the model-free implied volatility significantly improves the in-sample forecasting power of GARCH-type models. Out-of-sample forecasting results provide evidence that implied volatility subsumes historical volatility, GARCH-type volatility forecast, and Black-Scholes implied volatility, although option prices provide biased predictions of future realized volatility. Overall, our findings are consistent with the hypothesis that the crude oil futures options markets are efficient in that implied volatility can forecast future volatility and that volatility forecasts based on past returns provide no incremental information.

The results from robustness analysis reinforce the conclusions. We check the robustness of our results from two aspects. First, we evaluate the error-in-variable issue in univariate and encompassing regressions. We find that measurement errors in the model-free implied volatility do bias our slope estimates toward zero, and once we correct for the bias, there is stronger evidence that implied volatility is a good forecast of future realized volatility. Then we divide our data into two subsample periods, before September 11, 2001 terrorist attacks and after. We find that the implied volatility predicts future realized volatility much better in the second period, which corresponds to a higher volatility in the crude oil futures market.

The outline of the chapter is the following. In Section 2.2 we present the literature review on information content of implied volatility forecasts. Section 2.3 describes different measures of volatility employed in this study. In Section 2.4 we describe the methodological approach for studying information content and forecast ability of implied volatility relative to historically-based volatility forecasts. In Section 2.5 we describe the data and present summary statistics. The empirical within-sample estimation results followed by out-of-sample forecasting results are presented in Section 2.6. In Section 2.7 we present the results of the robustness analysis. The chapter concludes with a summary

and suggestions for future research.

## 2.2 Literature review

The hypothesis that implied volatility is a rational forecast of subsequently realized volatility has been frequently tested in the literature.<sup>4</sup> Empirical research across countries and markets so far has failed to provide a definitive answer, as the prior studies provide mixed evidence.

Early research on the predictive content of implied volatility found that it explains variation in future volatilities better than historical volatility. Latane and Rendleman (1976) were the first to find that realized volatility of 24 stocks calculated from in-sample period and extended into the future is more closely related to implied volatility. Also Chiras and Manaster (1978) and Beckers (1981) using stock options and the basic Black-Scholes option pricing model find that implied volatility forecasts explain a large amount of the cross-sectional variations of individual stock volatilities. Moreover, Schmalensee and Trippi (1978) showed that implied volatility rises when stock price falls and that implied volatilities of different stocks tend to move together.

In subsequent research from a time-series perspective, Day and Lewis (1992) develop a new methodology based on ARCH models. They examine the information content of implied volatility, obtained from S&P 100 index options from 1985 through 1989, relative to ARCH estimates of conditional volatility by adding the implied volatility to ARCH models as an exogenous variable. They find that both volatilities contain predictive power about future volatility, but conditional volatility based on ARCH modeling contains information beyond that reflected in implied volatility. Lamoureux and Lastrapes (1993) reach a similar conclusion by examining implied volatility from options on ten individual stocks over the period between April 1982 and March 1984. They suggest that both GARCH forecasts and implied volatility may contain valuable information and given the implied volatility and GARCH forecasts together, there is no further relevant information about future volatility to be extracted from the historical volatility.

In contrast, the empirical results from Canina and Figlewski (1993), who study the daily closing prices of call options on the S&P 100 index from March 1983 through March 1987, suggest that implied volatility is a poor forecast of subsequent realized volatility on the underlying index. Based on an encompassing regression analysis, they find that

---

<sup>4</sup>Poon and Granger (2003) provide an extensive review of the literature on volatility forecasting.

compared to historical volatility, implied volatility has virtually no correlation with future return volatility and does not appear to incorporate information contained in historical return volatility.

Such contradictory evidence regarding the efficiency of the implied volatility may be explained by several methodological issues, such as error-in-variables, overlapping data, and/or missing variables, which potentially contaminate the results of the above studies. For example, Day and Lewis (1992) use implied volatilities obtained from options with shortest times to maturity ranging from 7 to 36 days, which is then compared to one-week-ahead future volatility. Similarly, Lamoureux and Lastrapes (1993) examine implied volatilities based on stock options with maturities of between 64 and 129 trading days relative to the one-day-ahead volatility, resulting in a maturity mismatch problem. In addition, most of these papers construct their data on a daily basis, resulting in an overlap of consecutive observations in the time series of historical and future volatility causing serial correlation.

Overcoming these problems, more recent papers find evidence that implied volatility embedded in option prices is informationally efficient in forecasting future volatility. These studies (e.g., Moraux et al. (1999), Bates (2000), Simon (2003)) consider longer time series, possible regime shift around the October 1987 crash and the use of non-overlapping samples, and confirm that implied volatility still outperforms other volatility measures in forecasting future volatility. In particular, Christensen and Prabhala (1998), using a longer time series and monthly non-overlapping sample, find that implied volatility of at-the-money one-month S&P 100 options outperforms historical volatility in forecasting future volatility and subsumes all information in historical volatility. Moreover, they show that implied volatility is an unbiased forecast of ex-post realized volatility after the 1987 stock market crash.

Fleming (1998) argue that the use of Black-Scholes model to compute American option leads to a measurement error as a result of model misspecification and a better forecast ability of implied volatility can be achieved by improving the measurement of implied volatility. Their results show that the implied volatility from S&P 100 index options performs better in predicting future realized volatility of S&P 100 returns than historical volatility. However, the implied volatility is still an upward biased forecast of realized volatility. Ederington and Guan (2002) study the Black-Scholes implied volatility of futures options on the S&P 500 index and also conclude that the apparent bias and inefficiency of implied volatility from early studies are due to measurement errors. After

correcting for these measurement errors, they find that implied volatility is an efficient forecast for future volatility.

To address the problems associated with the model misspecification in obtaining implied volatility, model-free methods of calculating implied volatilities directly from option prices have emerged in recent years. Jiang and Tian (2005) show that the model-free implied volatility is more highly correlated with future realized volatility than either at-the-money Black-Scholes volatility or the realized volatility calculated from high-frequency data. Unlike the traditional concept of implied volatility, the model-free methodology is not based on any particular option pricing model, and therefore, the measurement error from model misspecification is minimal. Moreover, this model-free implied volatility measure incorporates the whole cross-section of option prices, not only at-the-money prices.

Most notably, in 2003 the Chicago Board of Exchange (CBOE) has adopted a model-free methodology to calculate the volatility index VIX, which is a measure of the implied volatility of S&P 500 index options. Due to the popularity of the VIX index, numerous studies have examined whether the VIX index forecasts future volatility better than volatility obtained from historical information. For example, Blair et al. (2001) conclude that volatility forecasts provided by the VIX index are more accurate for out-of-sample forecasting than forecasts based on intra-day returns or GARCH models. Corrado and Miller (2005) argue that the VIX yields forecasts that are biased upward, but they are more efficient than forecasts based on historical volatility. However, Becker et al. (2006) provide evidence that other historical volatility estimates can improve volatility forecasts based on the VIX alone. Giot and Laurent (2007) estimate a model for S&P 500 realized volatility with lagged values of the VIX index as explanatory variables and conclude that it is nearly as good as a model with jump and continuous components of historical volatility.

The information content of implied volatility for other assets has been also been studied. For example, Jorion (1995), using data on currency options, finds that implied volatility dominates the moving average and GARCH models in forecasting future volatility, although the implied volatility does appear to be a biased volatility forecast. Similarly, Taylor and Xu (1997) compare the forecast ability of implied volatility and historical volatility using options on spot currencies traded at the Philadelphia Stock Exchange. They find that when realized volatility and historical volatility are constructed from daily data, the implied volatility dominates historical volatility in forecasting re-

alized volatility. Amin and Ng (1997) study the information content of the options on short term forward interest rates, known as Eurodollar options, which are traded on the Chicago Mercantile Exchange (CME). They show that implied volatilities contain more information about future volatility than statistical time series models, however, the implied volatility forecasts are improved by the use of historical information.

In the case of commodities, Kroner et al. (1995) find that volatility forecasts combining implied volatility and GARCH-based estimates tend to perform better than either of them alone. Szakmary et al. (2003) also find that for a large majority of the 35 futures options markets, implied volatility, though not a completely unbiased predictor of future volatility, outperforms historical volatility as a predictor of future volatility. They also find that historical volatility is subsumed by implied volatility for most of the 35 markets examined.

Implied volatilities from options written on crude oil futures were also examined in the literature. Day and Lewis (1993) compare Black-Scholes implied volatility from call options on crude oil futures to a simple historical volatility and to out-of-sample forecasts from GARCH and EGARCH models. Their results indicate much better forecasting performance for implied volatility than that of the GARCH and historical models in this market. However, in a more recent study, Agnolucci (2009) concludes that GARCH-type models seem to perform better than the implied volatility obtained from inverting the Black equation.

The existing literature shows that the relation between implied and realized volatility tends to be affected by the measurement of implied volatility, the measurement of realized volatility, the econometric models and even the data set used. Although the results are somewhat mixed, the overall opinion seems to be that implied volatility is an efficient forecast of future volatility since the information contained in implied volatility subsumes all other available information, and therefore, it is a useful measure of expected future volatility. However, in almost all studies it appears to be a biased forecast. Given the equivocal results and conclusions across different asset classes, it is clear that further research on the predictive power of implied volatility is needed.

## 2.3 Volatility measures

### 2.3.1 Black-Scholes implied volatility

Implied volatility refers to the value of the volatility parameter that would set the theoretical option value equal to its current market price. Implied volatility can be computed by using a number of approaches, depending on the models used to price options. One of the most common approach for extracting implied volatility, that many studies of information content rely on, is the Black and Scholes (1973) options pricing model. Here we use the pricing formula from Black (1976), which has a future contract, rather than a stock, as underlying. The option pricing formula is represented as follows:

$$c(S, K, T, r, \sigma) = [F_t N(d_1) - K N(d_2)] e^{-r(T+\Delta)} \quad (2.1)$$

where  $c$  is the price of the call option,  $F_t$  is the price of underlying futures contract with delivery date  $\Delta$  periods after option expiration,  $T$  is the time to expiration of the option,  $K$  is the exercise or strike price of the option,  $r$  is the riskless interest rate.  $N$  indicates the value of the standard cumulative normal distribution evaluated at  $d_1 = \frac{\ln(F_t/K) + (r + \sigma_t^2/2)T}{\sigma_t \sqrt{T}}$  and  $d_2 = d_1 - \sigma_t \sqrt{T}$ . Thus, given observations on the option price  $c$  and other observed parameter values, an implied volatility estimate,  $\sigma$ , can be backed out by numerical inversion. We estimate implied volatility using the nearest-to-the-money options (both call and put) and then computing the arithmetic average from the value of implied volatility obtained from the put option and the value obtained from the call option.

The main problem with Black-Scholes option pricing model is that it is based on the assumptions, which are in contrast with actual market observations. For example, the Black-Scholes model assumes that the volatility will remain constant over the life of the option. In practice, however, volatility has been shown to be time-varying. Also, the logarithmic returns of the underlying asset are assumed to follow a normal distribution, whereas financial market returns exhibit both skewness and excess kurtosis. In addition, by focusing only on at-the-money options, the information contained in other options is ignored. However, the Black-Scholes implied volatility of a short-term option still remains a good proxy for the expected realized volatility of the underlying price process.

### 2.3.2 Model-free implied volatility

Beginning with Breeden and Litzenberger (1978), there has been considerable effort in deriving probability densities that incorporate the whole cross-section of option prices, not only at-the-money prices. An enormous number of different techniques have been suggested for the estimation of the risk-neutral density. Most of the approaches aim to find a parametric or nonparametric functional form that interpolates between available strike prices and extrapolates outside their range. Derman and Kani (1994), Dupire (1994), Rubinstein (1994, 1998), Carr and Madan (1998), Demeterfi et al. (1999) were among the first studies to derive tree models that include the complete set of observed options. Extending their work, Britten-Jones and Neuberger (2000) propose an alternative implied volatility measure based on the concept of the fair value of future variance that was further improved by various researchers (e.g., Carr and Wu (2006), Jiang and Tian (2005)).

In particular, let  $IV_{t,t+\Delta}$  denote the time  $t$  implied volatility measure computed as a weighted average, or integral, of a continuum of  $\Delta$ -maturity options. Under the assumptions that the underlying asset does not make dividend payments and the risk-free rate is zero, the risk-neutral expected sum of squared returns between two dates  $(t, t + \Delta)$  is

$$IV_{t,t+\Delta} = E^Q \left[ \int_t^{t+\Delta} \left( \frac{dS_t}{S_t} \right)^2 \right] = 2 \int_0^\infty \frac{C(t + \Delta, K) - C(t, K)}{K^2} dK \quad (2.2)$$

where  $E^Q[\dots]$  refers to expectation under the risk-neutral measure  $Q$ ,  $C(t, K)$  is an observed call price maturing at time  $t$  and strike price  $K$ , and  $S_t$  is the asset price at time  $t$ . Since no model for the underlying asset price is required in the derivation of (2.2), it is referred to as the model-free measure of variance. It only requires two cross sections of call prices with strikes spanning from zero to infinity, one with time to maturity  $t$  and the other with time to maturity  $t + \Delta$ .

Jiang and Tian (2005) show how to implement model-free implied volatility in practice by relaxing the assumptions of no dividends and a zero risk-free rate, which requires several modifications to (2.2). The zero interest rate assumption is relaxed by utilizing the relationship between options on the underlying asset and options on a related asset

with zero drift

$$IV_{t,t+\Delta} = E^Q \left[ \int_t^{t+\Delta} \left( \frac{dS_t}{S_t} \right)^2 \right] = 2 \int_0^\infty \frac{C(t+\Delta, Ke^{r(t+\Delta)}) - C(t, Ke^{rt})}{K^2} dK \quad (2.3)$$

where  $r$  is the risk-free rate.

Next, in order to produce a volatility measure that is comparable to implied volatility, the asset return variance between the current date and some later date  $T$  must be obtained. This is done by setting  $t = 0$  and  $t + \Delta = T$ , where  $T$  is the maturity to be evaluated. In this case the model-free variance formula simplifies to

$$E^Q \left[ \int_0^T \left( \frac{dS_t}{S_t} \right)^2 \right] = 2 \int_0^\infty \frac{C(T, Ke^{rT}) - \max(S_0 - K, 0)}{K^2} dK \quad (2.4)$$

where  $\max(S_0 - K, 0)$  is time-zero value of a call with strike price  $K$ .

Now only a single cross-section of option prices is needed to compute the model-free variance in (2.4). However, there are two problems that arise when computing model-free implied variance with (2.4), discretization of strike prices and truncation of the integration domain. The right-hand side of (2.4) specifies an integral over a continuum of strike prices ranging from 0 to  $\infty$ . Strike prices, however, are available only over a subset of this range, which implies that the integration domain will be truncated. Moreover, a continuum of strike prices is sampled by a discrete set of strike prices resulting in discretization problem.<sup>5</sup>

Therefore, the following equation provides approximation of (2.4)

$$2 \int_0^\infty \frac{C(T, Ke^{rT}) - \max(S_0 - K, 0)}{K^2} dK \approx 2 \sum_{K=K_L}^{K_H} \frac{C(T, Ke^{rT}) - \max(S_0 - K, 0)}{K^2} \Delta K \quad (2.5)$$

where  $K_L$  is the lowest available strike price,  $K_H$  is the highest available strike price, and  $\Delta K$  is the difference between adjacent strike prices.

---

<sup>5</sup>Jiang and Tian (2005) show that under reasonable assumptions, these approximation errors are negligible.

### 2.3.3 Realized volatility

As volatility is a latent variable, a proxy for ex post realized volatility needs to be computed. Here five different volatility estimators are used. The first measure of realized volatility, which is widely used in the volatility literature, is the daily squared returns obtained from closing prices:

$$\sigma_{C,t}^2 = \left( \ln \frac{C_t}{C_{t-1}} \right)^2 \quad (2.6)$$

where  $C_t$  is the closing futures price on day  $t$ , and  $C_{t-1}$  is the closing futures price on day  $t - 1$ .

The second volatility estimator is the Parkinson (1980) extreme value estimator that uses daily high and low prices. Parkinson (1980) estimator for a security following driftless geometric Brownian motion is given by

$$\sigma_{P,t}^2 = \frac{1}{4\ln(2)} \left( \ln \frac{H_t}{L_t} \right)^2 \quad (2.7)$$

where  $H_t$  is the high and  $L_t$  is the low prices, respectively, on a trading day  $t$ .

The third volatility estimator is the Garman and Klass (1980) estimator. Garman and Klass (1980) estimator assumes Brownian motion with zero drift and no opening jumps and is defined by:

$$\sigma_{GK,t}^2 = 0.511 \left( \ln \frac{H_t}{L_t} \right)^2 - 0.019 \left[ \ln \left( \frac{C_t}{O_t} \right) \ln \left( \frac{H_t L_t}{O_t^2} \right) - 2 \ln \left( \frac{H_t}{O_t} \right) \ln \left( \frac{L_t}{O_t} \right) \right] - 0.383 \left[ \ln \frac{C_t}{O_t} \right]^2 \quad (2.8)$$

where  $O_t$  is the opening prices on a trading day  $t$ .

The next estimator we use is the Rogers and Satchell (1991) estimator, which allows for drift and it is given by:

$$\sigma_{RS,t}^2 = \ln \left( \frac{H_t}{C_t} \right) \ln \left( \frac{H_t}{O_t} \right) + \ln \left( \frac{L_t}{C_t} \right) \ln \left( \frac{L_t}{O_t} \right) \quad (2.9)$$

.

All of the above measures are based on daily futures prices, and hence, they are one-period measures. In order to match the exact period covered by each of the implied volatilities in the construction of our realized volatility series and time-series forecasts, we calculate the multiperiod measures by summing the daily realized volatilities of the

trading days in the given period until the option expiration. Assume there are  $T$  trading days ( $t = 1, \dots, T$ ), then:

$$\sigma_t^T = \sqrt{\frac{252}{T} \sum_{t=1}^T \sigma_{i,t}^2}$$

where  $i = C, P, GK, RS$ .

And, finally, we also use the Yang and Zhang (2000) volatility estimator, which is independent of any drift and consistent in the presence of opening price jumps. It is calculated as a weighted average of the Rogers-Satchell estimator, the close-open volatility and the open-close volatility with the weights chosen to minimize the variance of estimator<sup>6</sup>:

$$\sigma_{YZ,t} = \sqrt{\frac{1}{N} \sum_{t=1}^N \left[ \ln \frac{O_t}{C_{t-1}} - \overline{\ln \frac{O_t}{C_{t-1}}} \right]^2 + \frac{\kappa}{N} \sum_{t=1}^N \left[ \ln \frac{C_t}{O_t} - \overline{\ln \frac{C_t}{O_t}} \right]^2 + (1 - \kappa) \sigma_{RS,t}^2} \quad (2.10)$$

with:

$$\kappa = \frac{0.34}{1.34 + \frac{N+1}{N-1}}$$

where  $\overline{\ln \frac{O_t}{C_{t-1}}} = 1/N \sum_{t=1}^N \ln \frac{O_t}{C_{t-1}}$ ,  $\overline{\ln \frac{C_t}{O_t}} = 1/N \sum_{t=1}^N \ln \frac{C_t}{O_t}$ , and  $\sigma_{RS,t}$  is the Rogers-Satchell (1991) estimator.

## 2.4 Methodology

### 2.4.1 In-sample estimation

Following Day and Lewis (1993), we first examine the in-sample performance of the implied volatility series, by fitting GARCH models to the daily returns of the crude oil futures, which have proven to capture the persistence in the conditional variance. The most widely used time-varying volatility model is the GARCH(1,1) specification, which is specified as follows:

$$r_t = \mu + \varepsilon_t, \varepsilon_t \sim N(0, h_t) \quad (2.11)$$

$$h_{t+1} = \alpha_0 + \alpha_1 \varepsilon_t^2 + \beta_1 h_t \quad (2.12)$$

---

<sup>6</sup>Since the opening jump is usually associated with the unexpected information coming during non-trading time, the incorporation of opening jump may lead to a better measurement of the realized volatility.

where  $r_t$  is the continuously compounded return of the underlying futures computed by taking the logarithm of the ratio of two consecutive days' closing prices,  $\mu_t$  is the mean daily return,  $h_{t+1}$  is the conditional variance of returns in period  $t + 1$ ,  $\varepsilon_t$  is the current period's shock, and  $h_t$  is the conditional variance of returns in period  $t$ . The parameter restrictions are  $\alpha_0 > 0$ ,  $\alpha_1 \geq 0$ ,  $\beta_1 \geq 0$ ,  $\alpha_1 + \beta_1 < 1$ .

Several studies have found that stock return volatility increases more after negative shocks than after positive shocks. To account for this asymmetric effect, we also consider the exponential GARCH (EGARCH) model proposed by Nelson (1990), in which conditional variance is an asymmetric function of the residuals  $\varepsilon_t$ :

$$\ln(h_{t+1}) = \alpha_0 + \beta_1 \ln(h_t) + \gamma \frac{\varepsilon_t}{\sqrt{h_t}} + \alpha_1 \left[ \frac{|\varepsilon_t|}{\sqrt{h_t}} - \sqrt{\frac{2}{\pi}} \right] \quad (2.13)$$

The variance specifications in Equations (2.12) and (2.13) are then tested against models, which include implied volatility forecast,  $IV$ , as an exogenous variable:

$$h_{t+1} = \alpha_0 + \alpha_1 \varepsilon_t^2 + \beta_1 h_t + \delta IV_t \quad (2.14)$$

and

$$\ln(h_{t+1}^2) = \alpha_0 + \beta_1 \ln(h_t^2) + \gamma \frac{\varepsilon_t}{\sqrt{h_t^2}} + \theta \left[ \frac{|\varepsilon_t|}{\sqrt{h_t}} - \sqrt{\frac{2}{\pi}} \right] + \delta \ln(IV_t) \quad (2.15)$$

The coefficient  $\delta$  can be interpreted as a measure of how much incremental information implied volatility contributes to the model. The null hypothesis is given by  $H_0 : \delta = 0$  in (2.14) and (2.15) and is tested using a likelihood ratio test. If the null hypothesis cannot be rejected, the conclusion would be that implied volatility contains no incremental information for explaining volatility beyond that derived from a GARCH model.

The question whether the GARCH(1,1) and EGARCH specifications for volatility contain information that is not impounded in implied volatilities is also examined by estimating Equations (2.14) and (2.15) under the constraint that the parameters  $\alpha_1$  and  $\beta_1$  in Equation (2.14) and  $\theta$ ,  $\gamma$  and  $\beta_1$  in Equation (2.15) are all equal to zero. These sets of restrictions on (2.14) and (2.15) test whether the lagged squared error and lagged conditional variance from a GARCH model contain any additional explanatory power once implied volatility is included in the specification. In this case the equations

describing the conditional volatility dynamics are:

$$h_{t+1} = \alpha_0 + \delta IV_t \quad (2.16)$$

and

$$\ln(h_{t+1}) = \alpha_0 + \delta \ln(IV_t) \quad (2.17)$$

The parameter estimation for Equations (2.16) and (2.17) also can be used to test whether implied volatilities are unbiased predictors of within-sample volatility. If implied volatilities are unbiased estimates of future realized volatility, the parameter estimates of  $\alpha_0$  and  $\delta$  will be close to zero and one, respectively. All of these restrictions are tested using a likelihood ratio test.

## 2.4.2 Out-of-sample forecasting

Although the previous tests allow us to assess the relative information content of time series models and implied volatilities, they do not represent a true test of relative forecasting power, since the entire sample period for which implied volatilities are available is used to both estimate and test the models. In addition, the GARCH model has a forecast horizon of one day, while the implied volatility forecasts volatility over the time until option expiration (up to one month). Hence we also conduct an out-of-sample forecasting tests.

Parameters for the GARCH-type models used in the forecast have been estimated on a 5-year rolling sample. Specifically, we use the data for the period from January, 2, 1991 through December 30, 1995, to estimate the models, and then make a one-step-ahead forecast of the following day's volatility. Then the sample is rolled forward one observation at a time, constructing a new one-step-ahead forecast at each stage. To control for a maturity mismatch problem, we use a procedure suggested by Day and Lewis (1992) to transform the daily GARCH forecasts into a forecast of volatility over a horizon comparable to the forecast horizon for the implied volatility.

Following Szakmary et al. (2003) and Covrig and Low (2003), among others, we examine the following three hypotheses to test whether the implied volatility contains a significant amount of information beyond that contained in alternative volatility forecasts:

- H1. Implied volatility is an unbiased estimator of future realized volatility.
- H2. Implied volatility has more explanatory power than historical volatility (or the

GARCH volatility forecast) in forecasting future realized volatility.

H3. Implied volatility incorporates all information regarding future volatility; historical volatility (or the GARCH volatility forecast) contains no information beyond what is already contained in implied volatility.

To test the above hypotheses, we employ a simple and intuitive approach commonly used in the literature for testing the information content - the regression framework. This methodology is still by far the dominant approach within the literature addressing the efficiency and bias issue of implied volatility forecasts. First, we run Mincer and Zarnowitz (1969) univariate predictive regressions, in which realized volatility is regressed on the corresponding implied, and possibly alternative, volatility forecasts as follows:

$$\sigma_{t,T}^2 = b_0 + b_i \hat{\sigma}_{t,T}^2 + \varepsilon_t \quad (2.18)$$

where  $\sigma_{t,T}^2$  is the *ex post* average realized volatility of the underlying asset realized over the period  $t$  to  $t + T$  until option expiration,  $\hat{\sigma}_{t,T}^2$  is a forecast of future volatility over the period  $t$  to  $t + T$  based on the information available at the end of period  $t$ , and  $\varepsilon_t$  represents the forecast error. In particular, in our study we compare four volatility forecasts: GARCH volatility forecasts, the model-free implied volatility, the Black-Scholes implied volatility and the naïve historical volatility. Thus, we will consider the following regressions:

$$\sigma_t^2 = b_0 + b_1 \hat{\sigma}_{MF,t}^2 + \varepsilon_t \quad (2.19)$$

$$\sigma_{RV,t}^2 = b_0 + b_2 \hat{\sigma}_{G,t}^2 + \varepsilon_t \quad (2.20)$$

$$\sigma_{RV,t}^2 = b_0 + b_3 \hat{\sigma}_{HV,t}^2 + \varepsilon_t \quad (2.21)$$

$$\sigma_{RV,t}^2 = b_0 + b_4 \hat{\sigma}_{BS,t}^2 + \varepsilon_t \quad (2.22)$$

where  $\hat{\sigma}_{MF,t}^2$  is the model-free implied volatility at time  $t$  until option expiration,  $\hat{\sigma}_{G,t}^2$  is the average step-ahead GARCH(1,1) forecast,  $\hat{\sigma}_{HV,t}^2$  is the naïve forecast based on historic volatility over the previous days, and  $\hat{\sigma}_{BS,t}^2$  is the Black-Scholes implied volatility at time  $t$  until option expiration.

If H1 is true - that model-free implied volatility is an unbiased predictor of the future realized volatility - we would expect  $b_0 = 0$  and  $b_1 = 1$  in Equation (2.19). Moreover, if implied volatility is efficient, the residuals  $\varepsilon_t$  from regression (2.19) should be white noise and uncorrelated with any variable in the market's information set. If H2 is true - that

implied volatility contains more current market information than other volatility measures - we would expect a higher adjusted  $R^2$  from regression (2.19) than from regressions (2.20-2.22).

Next, the *ex post* realized volatility is regressed on several forecasts generated by different models in encompassing regressions. If the forecast series have significant coefficients, we conclude that it encompasses the forecasts whose coefficients are not significant. The test regression is of the form:

$$\sigma_{RV,t}^2 = b_0 + b_1 \hat{\sigma}_{MF,t}^2 + b_2 \hat{\sigma}_{G,t}^2 + b_3 \hat{\sigma}_{HV,t}^2 + \varepsilon_t \quad (2.23)$$

If H3 is correct and implied volatility forecast contains independent information that is useful in predicting future volatility, the regression coefficient of that forecast ( $b_1$ ) should be significantly greater than zero. And if implied volatility efficiently incorporates all information regarding future volatility, we would expect  $b_2 = 0$  and  $b_3 = 0$  in (2.23) since other volatility measures should have no explanatory power beyond that already contained in implied volatility.

In addition, since we estimate implied volatility using both model-free methodology and Black-Scholes option pricing model, we also test which one is more efficient in forecasting future realized volatility. If the model-free implied volatility is more efficient than the Black-Scholes, regressions with model-free implied volatility should result in higher adjusted  $R^2$ . Furthermore, as a more formal test, we run the following regression with both model-free implied volatility and the Black-Scholes implied volatility included as explanatory variables:

$$\sigma_{RV,t}^2 = b_0 + b_1 \hat{\sigma}_{MF,t}^2 + b_4 \hat{\sigma}_{BS,t}^2 + \varepsilon_t \quad (2.24)$$

Intuitively, if model-free implied volatility subsumes Black-Scholes implied volatility, we would expect the coefficient  $b_1$  to be significant and the coefficient  $b_4$  to be insignificant.

All the regressions above are estimated with standard OLS. In order to account for the possible presence of serial correlation in the data, the Newey-West covariance correction for serial correlation is used.

## 2.5 Data description and summary statistics

The dataset used in the study contains daily time series of light, sweet crude oil futures and options on these futures traded at the NYMEX for the period from January 03, 1991 through December 14, 2006. We split the full sample into an in-sample period of estimation from 1991 through 1995, and an out-of-sample forecast period from 1996 through 2006. There are a number of practical advantages to using the data on light, sweet crude oil from the NYMEX. First, both futures and options on these futures are traded on the same floor, which facilitates hedging and arbitrage and makes the market more efficient. Second, the trading closes at the same time, therefore, there are no non-synchronicity biases.

NYMEX lists American-style options on the futures with monthly expirations several years into the future with the same maturity date as the futures<sup>7</sup>. Upon exercise, the cash flow to a call option equals the difference between the current futures price and the exercise price. The owner of a put option receives the difference between the exercise price and the current futures price. In Table 2.1 we provide main features and contract specifications obtained from the CME website.

From the dataset we obtain observations on daily settlement prices for call and put options. We consider only options at the nearest maturities. As option prices can be very volatile when approaching the expiration date, we exclude options with less than eight days to expiration and use the next nearest maturities instead. The reason for choosing options that expire the next calendar months is that they are the most liquid ones. We match all puts and calls by trading date, maturity, and strike. For each pair, we drop strikes for which put/call price is less than \$0.01. Generally, a large number of options meet these selection criteria. Options which violate boundary conditions or are either deep in- or out-of-the-money are not considered.

To estimate the ex post realized volatility, we use the data on daily high, low, opening and closing prices for underlying futures contracts also traded at the NYMEX. We use only the futures contracts with the same contract months as options to ensure the best match between the implied volatility and the realized volatility calculated from subse-

---

<sup>7</sup>NYMEX also lists European-style options on light, sweet crude oil. However, the trading history is much shorter and liquidity is much lower than for the American-style options. Since the options we use are American type, their prices could be slightly higher than prices of the corresponding European options due to the early exercise premia. As shown by several studies, the difference, however, is very small for short-term at-the-money futures options that we focus on (e.g., Jorion (1995)).

Table 2.1: Contract specifications: light, sweet crude oil futures options traded at NYMEX.

Underlying Futures	Light Sweet Crude Oil Futures (CL)
Product Symbol	LO
Price Quotation	U.S. dollars and cents per barrel.
Option Style	American
Minimum fluctuation	\$0.01 per barrel
Expiration of Trading	Trading ends three business days before the termination of trading in the underlying futures contract.
Listed Contracts	Crude oil options are listed nine years forward as follows: consecutive months are listed for the current year and the next five years; in addition, the June and December contract months are listed beyond the sixth year.
Strike Prices	Twenty strike prices in increments of \$0.50 per barrel above and below the at-the-money strike price, and the next 10 strike prices in increments of \$2.50 above the highest and below the lowest existing strike prices for a total of at least 61 strike prices.
Settlement Type	Exercise into Futures

quent futures prices. At the rollover date, the prices of the new contract on the previous day before the rollover is used. We further construct the time series of non-overlapping monthly implied volatilities, since previous work has shown that the use of overlapping data leads to problems in the statistical analysis. To overcome the problems associated with overlapping sampling, we follow recent literature and consider a non-overlapping sample on a monthly basis. Because the implied volatility we use is of a constant maturity (one month) in our non-overlapping samples, we are able to compute the future volatility over the exact remaining life of the option, eliminating the maturity mismatch problem.

The risk-free rate of interest for each option expiration series is computed for each day using the average of the bid and ask discounts for the U.S. Treasury bill, whose maturity is closest to the option expiration date.

Figure ?? shows the time-series of the model-free implied volatility for the period from January 1996 through December 2006. As we can see, it was relatively stable in the early 1996, but it becomes more volatile starting from 1998. The spikes in volatility coincide with the events, such as Gulf War of late 1997, Asian financial crisis of 1997, the Russian and LTCM crisis of 1998, the 9/11 terrorist attacks and military operations in

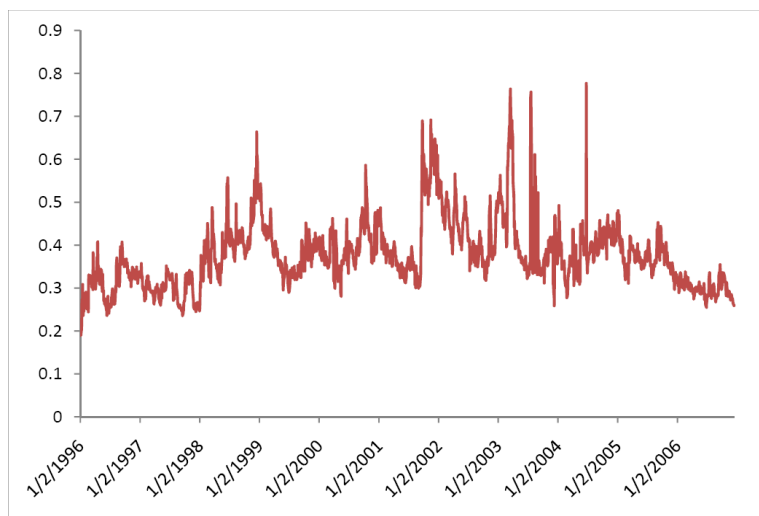


Figure 2.1: Time-series plot of the model-free implied volatility. The figure presents the time series of the 30-day model-free implied volatility over the sample period from January 1996 through December 2006.

Iraq in 2003. Figure 2.2 shows the first differences of the implied volatility for the entire sample and points to heteroskedasticity in the data.

In Figure ?? we present the time-series of the various estimates of daily realized volatility. The peaks of these estimates are approximately synchronous, however, estimators based on the range data are seem to be less volatile than the classical estimator based on closing prices only.

Summary statistics for the levels and logarithms of both realized and implied volatilities are provided in Table 2.2 for the entire sample. If implied volatility were given by the conditional expectation of future realized volatility, we would expect that realized volatility and implied volatility had equal unconditional means. In the crude oil futures market, both average implied volatility and average log implied volatility exceed the means of the corresponding realized volatility, possibly reflecting a negative price of volatility risk (Bollerslev and Zhou (2006)). Thus, the results suggest that implied volatility is unlikely to be an unbiased estimator of realized volatility, which will be formally tested in the next few sections. The bias is similar regardless of whether the implied volatility is estimated using model-free methodology or Black-Scholes option pricing model.

The implied volatility is skewed to the right and displays excess kurtosis. The co-

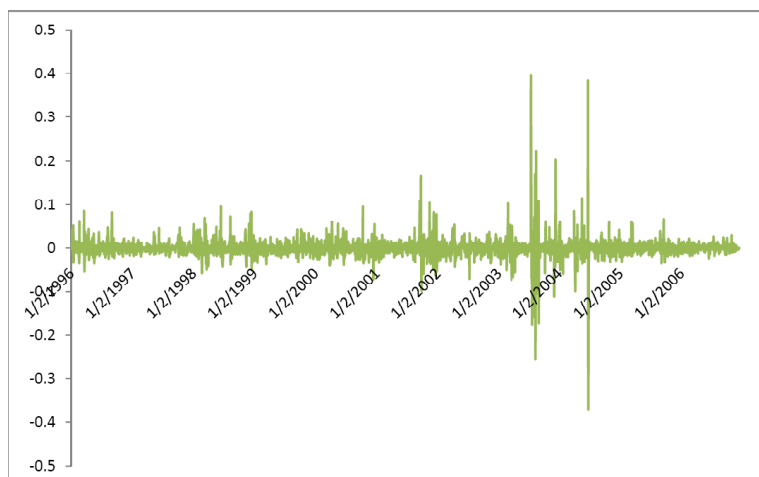


Figure 2.2: First differences of model-free implied volatility.

efficients of skewness of all measures of realized volatilities are positive; the mass of probability in the right side of the distribution appears slightly larger than on the left side. All the realized volatility measures appear to be leptokurtic, while the distributions of the log-volatility series are less so. Both realized and implied volatilities exhibit pronounced serial correlation with extremely slow decay in their autocorrelations. However, an Augmented DickeyFuller test cannot reject the presence of a unit root at the 5% level.

## 2.6 Empirical results

### 2.6.1 In-sample estimation results

Tables 2.3 and 2.4 present the parameter estimates for the GARCH(1,1) and EGARCH specifications for the crude oil futures market volatility. All models are estimated by quasi-maximum likelihood. We provide robust standard errors (Bollerslev and Wooldridge (1992)) in parentheses. The standard GARCH(1,1) reported in the first row confirms the common finding of strong persistence in volatility ( $\alpha = 0.891$ ).

We also provide the coefficient estimates for the Equation (2.14), which includes options market volatility information as an exogenous variable in the GARCH(1,1) model for the conditional volatility. It appears from the coefficient estimates and their standard

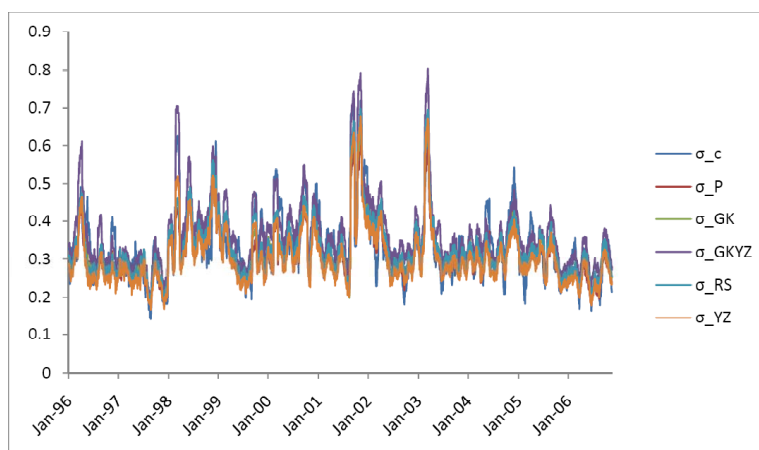


Figure 2.3: Time series plot of the realized volatility measures. The figure presents the time series of realized volatility measures: squared returns (C), Parkinson (P), Rogers-Satchell (RS), Garman-Klass (GK), and Yang-Zhang (YZ). The sample period is from January 1996 to December 2006.

errors under the GARCH(1,1) specification that the implied volatility term ( $\delta$ ) is positive, which suggests that conditional volatility is an increasing function of the observed levels of implied volatility. It is also statistically significant, while the GARCH terms ( $\alpha_1$  and  $\beta_1$ ) are not. This evidence suggests that implied volatility contains some incremental information that is not reflected by historical returns. The maximum likelihood has significantly increased compared with that from the standard GARCH(1,1) model. Likelihood ratio tests of the null hypothesis  $\alpha_1 = \beta_1 = 0$  in Eq. (2.14) can be evaluated by comparing  $LR = 2(L_1 - L_0)$  with  $\chi^2_2$ , where  $L_0$  and  $L_1$  are the maximum log-likelihoods for (2.12) and (2.14). The test values reject the hypothesis that implied volatility has no incremental information at the 0.001% significance level.

In Equation (2.16) we restrict all parameters, except for  $\delta$ , to zero to see how implied volatility describes the volatility process on its own. Comparing the log-likelihood of this model to the standard GARCH(1,1) model reveals that this model performs significantly better, and the size of the coefficient shows that the model-free implied volatility captures the persistence in the volatility series well. We also test the hypothesis that returns contain no volatility information in addition to that already contained in option prices by

Table 2.2: Descriptive statistics for implied and realized volatilities. This table reports the descriptive statistics for the levels and logarithms of different realized and implied volatility measures. The realized volatility measures are: close-to-close ( $\sigma_C$ ), Parkinson ( $\sigma_P$ ), Garman-Klass ( $\sigma_{GK}$ ), Rogers-Satchel ( $\sigma_{RS}$ ), and Yang-Zhang ( $\sigma_{YZ}$ ). The implied volatility measures are: model-free implied volatility ( $\sigma_{IV}^2$ ) and Black-Scholes implied volatility ( $\sigma_{BS}$ ). The sample period is from January 1991 through December 2006.

	$\sigma_C$	$\sigma_P$	$\sigma_{GK}$	$\sigma_{RS}$	$\sigma_{YZ}$	$\sigma_{IV}$	$\sigma_{BS}$
Mean	0.335	0.315	0.328	0.330	0.307	0.372	0.361
St. Dev.	0.090	0.068	0.074	0.078	0.077	0.072	0.075
Kurtosis	6.040	5.988	7.066	8.149	5.574	4.367	4.701
Skewness	1.489	1.335	1.575	1.766	1.898	0.983	1.203
	$\log\sigma_C$	$\log\sigma_P$	$\log\sigma_{GK}$	$\log\sigma_{RS}$	$\log\sigma_{YZ}$	$\log\sigma_{IV}$	$\log\sigma_{BS}$
Mean	-0.488	-0.511	-0.494	-0.492	-0.524	-0.437	-0.452
St. Dev.	0.106	0.088	0.089	0.093	0.097	0.080	0.085
Kurtosis	3.673	3.779	4.170	4.492	4.666	3.126	3.241
Skewness	0.627	0.536	0.683	0.746	0.838	0.412	0.512

comparing maximum log-likelihoods values from Equations (2.16) and (2.14). Comparing the test values with  $\chi^2$  shows that the null hypothesis is rejected and options prices subsume information contained in historical returns for predicting conditional volatility. Finally, we test the hypothesis that implied volatility is an unbiased estimate for one-period-ahead future volatility, which implies that  $\alpha_0 = 0$  and  $\delta = 1$  in Equation (2.16). The results of the likelihood ratio tests cannot reject the hypothesis at the 0.5% significance level.

Tests of the informational efficiency of the options market using the EGARCH specification lead to the similar conclusions as those reported for the GARCH(1,1) specification. Moreover, the reported estimate of  $\theta$  is not significantly different from zero, which indicates that the asymmetry effect is not significant in the crude oil futures market. This is in contrast to the stock market where negative news shocks have a larger impact on volatility than positive shocks. Therefore, the EGARCH model may not be more useful than GARCH(1,1) in forecasting volatility in this market.

The within-sample results demonstrate that implied volatilities contain incremental information regarding the variability of returns, and the information contained in the implied volatility series subsumes the information contained in the time series of returns.

Table 2.3: In-sample estimation results of GARCH models. This table provides the estimation of GARCH models in Eqs. (2.12), (2.14), and (2.16). The numbers in parentheses are standard errors of the coefficients computed using the robust inference procedures by Bollerslev and Wooldridge (1988).  $\text{Log}(L)$  is the logarithm maximum likelihood function value.  $\chi^2$  statistics are for equations (2.12) and (2.16). The sample period is from January 1991 to December 2006. \*\* and \* denote significance at the 5% and 1% levels, respectively.

Variance Specification	$\alpha_0$	$\alpha_1$	$\beta_1$	$\delta$	Log(L)	$\chi^2$
(2.12)	0.058* (0.026)	0.069* (0.011)	0.891* (0.017)	- -	7113.236	69.73
(2.14)	0.000 (0.000)	0.034 (0.032)	0.715 (0.089)	0.459* (0.102)	7214.214	-
(2.16)	0.015 (0.000)	- -	- -	0.711* (0.112)	7172.473	30.27

Hence, in-sample, the model-free implied volatility seems to be a better measure of future market volatility than past volatility. By adding implied volatility, we see that the model significantly improves the standard GARCH(1,1) and EGARCH models. Interestingly, the addition of this exogenous variable reduces the coefficient on the GARCH term, implying that the model-free implied volatility series captures a part of the persistence in the volatility process.

## 2.6.2 Out-of-sample forecasting results

Until now the tests have characterized the within-sample properties of volatility information, since the likelihoods of GARCH models were maximized over the complete sample period. Next, we compare the out-of-sample forecasting ability of historical volatility forecasts, forecasts from standard GARCH models, and implied volatility forecasts.

First, table 2.5 reports the regression results for testing the predictive power of model-free implied volatility using different measurements of realized volatility. The following conclusions can be made from these results. First, the slope coefficient for implied volatility is significantly different from zero at 1% significance level for all realized volatility measurements, indicating that the model-free implied volatility contains information in forecasting future realized volatility. The coefficient estimate ranges from 0.737 for the realized volatility measured by the traditional close-to-close estimator to 0.872 for the

Table 2.4: In-sample estimation results of EGARCH models. This table provides the estimation of EGARCH models in Eqs. (2.13), (2.15), and (2.17). The numbers in parentheses are standard errors of the coefficients computed using the robust inference procedures by Bollerslev and Wooldridge (1988).  $\text{Log}(L)$  is the logarithm maximum likelihood function value.  $\chi^2$  statistics are for equations (2.13) and (2.17). The sample period is from January 1991 to December 2006.\*\* and \* denote significance at the 5% and 1% levels, respectively.

Variance Specification	$\alpha_0$	$\beta_1$	$\theta$	$\gamma$	$\delta$	$\text{Log}(L)$	$\chi^2$
(2.13)	0.315* (0.058)	0.983* (0.006)	-0.015 (0.011)	0.009* 0.012	- -	7110.511	89.32
(2.15)	0.273 (0.210)	0.892 (0.132)	-0.005 (0.000)	0.004 (0.001)	0.437* (0.142)	7198.219	
(2.17)	0.131** (0.047)	- -	- -	- -	0.633* (0.005)	7167.751	34.56

realized volatility measured by the Garman-Klass estimator. The forecasting ability of model-free implied volatility is improved when a better measurement of realized volatility is used. The adjusted  $R^2$  is 0.285 when realized volatility is constructed from closing prices only, but it is 0.431 when realized volatility is constructed from daily range data. The results show that using range data significantly improves the predictability of the asset return volatility. However, although the model-free implied volatility contains significant information regarding future realized volatility, it is not an unbiased estimator. The Wald-statistics for the null hypothesis that jointly  $b_0 = 0$  and  $b_1 = 1$  is rejected at 1% significance level when realized volatility is measured by either of these estimators.

The results for the predictive ability of the historical volatility are similar to those of model-free implied volatility. As we can see, the slope coefficient is significantly different from zero for all measurements of volatility, indicating that the historical volatility does contain information in forecasting next period realized volatility. However, the null hypothesis for the unbiasedness of historical volatility is also rejected at the 1% significance level. All slope coefficients are significantly less than one. That is, historical volatility also contains information concerning future realized volatility; however, it is not an unbiased estimator. The explanatory power of historical volatility is quite low compared to the implied volatility in forecasting future realized volatility. The adjusted  $R^2$  decreases sharply, the highest  $R^2$  is 0.299 when realized volatility is measured by the Parkinson

Table 2.5: Forecast regression of model-free implied volatility. This table reports OLS regression coefficients and standard errors in parentheses for Equation (2.19) based on the monthly non-overlapping sample for the period January 1996 through December 2006. Different measurements of realized volatility are used: close-to-close ( $\sigma_C$ ), Parkinson ( $\sigma_P$ ), Garman-Klass ( $\sigma_{GK}$ ), Rogers-Satchel ( $\sigma_{RS}$ ), and Yang-Zhang ( $\sigma_{YZ}$ ). F is the Wald-statistic tests the joint hypothesis that  $b_0 = 0$  and  $b_1 = 1$ . DW is the Durbin-Watson statistics. \*\*\*, \*\* and \* denote significance at the 1%, 5% and 10% levels, respectively.

Volatility measure	$b_0$	$b_1$	Adj. $R^2$	F	DW
$\sigma_{CC}$	0.112*** (0.041)	0.737*** (0.104)	0.285	22.714***	2.12
$\sigma_P$	0.091*** (0.029)	0.750*** (0.077)	0.431	63.983***	2.08
$\sigma_{GK}$	0.095*** (0.044)	0.872*** (0.085)	0.356	4.715***	2.04
$\sigma_{RS}$	0.080*** (0.035)	0.818*** (0.094)	0.418	29.988***	2.09
$\sigma_{YZ}$	0.069*** (0.035)	0.787*** (0.035)	0.393	62.039***	2.06

estimator, while the regression on implied volatility gives  $R^2$  of 0.431 for the similar measurement of realized volatility. All results support the conclusion that historical volatility has less predictive ability than implied volatility in forecasting future realized volatility.

In Table 2.7 the results of regressing realized volatility on GARCH volatility forecasts are presented. As can be seen, they are similar to the results in the previous Table 2.6. Thus, taken alone, both GARCH and historical volatilities are statistically significant, but their information content is quite inferior to the implied volatility predictive power as judged by adjusted  $R^2$ .

And, finally, we present the regression results for testing the relative predictive power of the Black-Scholes implied volatility in Table 2.8. The adjusted  $R^2$  of Equation (2.22) using Black-Scholes implied volatility as the explanatory variable is substantially higher than those using either historical or the GARCH(1,1) volatility forecast as the explanatory variable. As can be seen from Table 2.8, the Black-Scholes implied volatility also contains useful information in predicting future realized volatility, as indicated by highly significant coefficients. Thus, also the Black-Scholes implied volatility contains more information than either historical volatility or the GARCH(1,1) volatility forecast in

Table 2.6: Forecast regression of historical volatility. This table reports regression coefficients and standard errors in parentheses for Equation (2.21) based on the monthly non-overlapping sample for the period January 1996 through December 2006. Different measurements of realized volatility are used: close-to-close ( $\sigma_C$ ), Parkinson ( $\sigma_P$ ), Garman-Klass ( $\sigma_{GK}$ ), Rogers-Satchel ( $\sigma_{RS}$ ), and Yang-Zhang ( $\sigma_{YZ}$ ). F is the Wald-statistic tests the joint hypothesis that  $b_0 = 0$  and  $b_1 = 1$ . DW is the Durbin-Watson statistics. \*\*\*, \*\* and \* denote significance at the 1%, 5% and 10% levels, respectively.

Volatility measure	$b_0$	$b_3$	Adj. $R^2$	F	DW
$\sigma_{CC}$	0.206*** (0.033)	0.415*** (0.088)	0.177	40.323***	2.12
$\sigma_P$	0.153*** (0.026)	0.581*** (0.074)	0.299	30.426***	2.14
$\sigma_{GK}$	0.224*** (0.034)	0.463*** (0.080)	0.203	44.22***	2.09
$\sigma_{RS}$	0.192*** (0.028)	0.489*** (0.077)	0.224	39.201***	2.15
$\sigma_{YZ}$	0.184*** (0.028)	0.465*** (0.080)	0.205	41.600***	2.16

predicting future realized volatility.

Overall, our finding that implied volatility contains substantial information about future realized volatility in the crude oil futures market is consistent with H1, although it does not appear to be an unbiased estimator. These results are in line with prior studies on information content of the implied volatility. The predictive power of both model-free and Black-Scholes implied volatilities is much superior to that of historical volatility or the GARCH(1,1) volatility forecast in explaining future realized volatility, evidence consistent with H2.

And now we test H3 by fitting regression models that include model-free implied volatility and other volatility forecasts as specified in Equation (2.23). In Table 2.9 the results of encompassing regressions are presented. The slope coefficient for implied volatility decreases slightly compared to its values in the univariate regression, but remains statistically significant. The  $t$ -statistic for the null hypothesis  $b_1 = 0$  is rejected at 1% significance level. Past volatility, taken alone, does have some predictive power for future realized volatility. However, once the implied volatility is added as an explanatory variable, the regression coefficients for either historical volatility or GARCH volatility

Table 2.7: Forecast regression of GARCH volatility forecasts. This table reports regression coefficients and standard errors in parentheses for Equation (2.20) based on the monthly non-overlapping sample for the period January 1996 through December 2006. Different measurements of realized volatility are used: close-to-close ( $\sigma_C$ ), Parkinson ( $\sigma_P$ ), Garman-Klass ( $\sigma_{GK}$ ), Rogers-Satchel ( $\sigma_{RS}$ ), and Yang-Zhang ( $\sigma_{YZ}$ ). F is the Wald-statistic tests the joint hypothesis that  $b_0 = 0$  and  $b_1 = 1$ . DW is the Durbin-Watson statistics. \*\*\*, \*\* and \* denote significance at the 1%, 5% and 10% levels, respectively.

Volatility measure	$b_0$	$b_2$	Adj. $R^2$	F	DW
$\sigma_{CC}$	0.148*** (0.038)	0.547*** (0.099)	0.206	31.215***	2.04
$\sigma_P$	0.140*** (0.029)	0.540*** (0.077)	0.284	74.458***	2.07
$\sigma_{GK}$	0.164*** (0.039)	0.609*** (0.101)	0.229	14.532***	2.09
$\sigma_{RS}$	0.157*** (0.032)	0.535*** (0.086)	0.234	40.731***	1.94
$\sigma_{YZ}$	0.142*** (0.033)	0.516*** (0.085)	0.225	68.813***	2.05

forecasts drop from their values in the univariate regressions for all measurements of realized volatility. In fact, the  $t$ -statistic for the null hypothesis that  $b_2 = 0$  cannot be rejected even at 10% significance level, indicating that historical volatility virtually contains no incremental information in forecasting future realized volatility in addition to the implied volatility. The coefficient for the GARCH volatility forecast is statistically significant, but it has a much smaller effect on future realized volatility compared to implied volatility.

The better performance of model-free implied volatility over other volatility forecasts is also evident from the small incremental adjusted  $R^2$  resulting from adding historical volatility to implied volatility. The increase in the adjusted  $R^2$  is very marginal. These results suggest that including either historical volatility or the GARCH volatility forecast adds very little or no information to that already contained in implied volatility. However, the forecasts cannot be considered to be unbiased since the coefficient  $b_1$  is significantly less than one. The F-test, rejects the joint hypothesis of the unbiasedness of  $b_0 = b_2 = b_3 = 0$  and  $b_1 = 1$ .

Table 2.8: Forecast regression of Black-Scholes implied volatility. This table reports regression coefficients and standard errors in parentheses for Equation (2.22) based on the monthly non-overlapping sample for the period January 1996 through December 2006. Different measurements of realized volatility are used: close-to-close ( $\sigma_C$ ), Parkinson ( $\sigma_P$ ), Garman-Klass ( $\sigma_{GK}$ ), Rogers-Satchel ( $\sigma_{RS}$ ), and Yang-Zhang ( $\sigma_{YZ}$ ). F is the Wald-statistic tests the joint hypothesis that  $b_0 = 0$  and  $b_1 = 1$ . DW is the Durbin-Watson statistics. \*\*\*, \*\* and \* denote significance at the 1%, 5% and 10% levels, respectively.

Volatility measure	$b_0$	$b_4$	Adj. $R^2$	F	DW
$\sigma_{CC}$	0.121*** (0.043)	0.701*** (0.104)	0.263	37.338***	2.18
$\sigma_P$	0.104*** (0.033)	0.696*** (0.082)	0.395	51.180***	2.14
$\sigma_{GK}$	0.132*** (0.042)	0.771*** (0.105)	0.338	19.432***	2.05
$\sigma_{RS}$	0.118*** (0.037)	0.686*** (0.089)	0.391	45.721***	1.99
$\sigma_{YZ}$	0.106*** (0.036)	0.663*** (0.091)	0.368	65.113***	2.08

Next, we show the regression estimates of Equation (2.23) for the Black-Scholes implied volatility. The Black-Scholes implied volatility subsumes historical volatility or the GARCH volatility forecast in predicting future realized volatility, since all the coefficients in the encompassing regressions are close to unity and are statistically significant, whereas those for historically-biased volatility measures are both close to zero and statistically insignificant. Similarly to the model-free implied volatility, the high Wald-statistics strongly reject the unbiasedness hypothesis of  $b_0 = b_2 = b_3 = 0$  and  $b_4 = 1$ .

We compare the goodness of fit in models, where the Black-Scholes implied volatility is the explanatory variable, with that in models, where the model-free implied volatility is the explanatory variable. It is quite intuitive that the implied volatility that yields a higher explanatory power will be more efficient in predicting future volatility. When we compare the model adjusted  $R^2$  using the model-free implied volatility in Table 2.5 and 2.9 to those using the Black-Scholes implied volatility in Table 2.8 and 2.10, the model adjusted  $R^2$ s using the model-free implied volatility are uniformly higher than those using the Black-Scholes implied volatility. These seem to suggest, that the model-free implied volatility is more efficient than the Black-Scholes implied volatility, also as judged by

Table 2.9: Information content of model-free implied volatility versus historical volatility and the GARCH(1,1) volatility forecast. This table reports regression coefficients and standard errors (in parentheses) for Equation (2.23) based on the monthly non-overlapping samples for the period January 1996 through December 2006. F is the Wald-statistic tests the joint hypothesis that  $b_0 = 0$  and  $b_1 = 1$ ,  $b_2 = b_3 = 0$ . DW is the Durbin-Watson statistics. \*\*\*, \*\* and \* denote significance at the 1%, 5% and 10% levels, respectively.

Volatility measure	$b_0$	$b_1$	$b_2$	$b_3$	Adj. $R^2$	DW	F
$\sigma_{CC}$	0.083 * ** (0.038)	0.779 * ** (0.136)	-0.066 (0.152)	0.257* (0.153)	0.298	2.18	34.92***
$\sigma_P$	0.065 * ** (0.026)	0.711*** (0.127)	-0.112 (0.165)	0.254** (0.107)	0.462	2.16	37.46***
$\sigma_{GK}$	0.058 (0.041)	0.844*** (0.182)	-0.135 (0.123)	0.313** (0.139)	0.375	2.15	34.62***
$\sigma_{RS}$	0.053 (0.032)	0.868*** (0.149)	-0.228 (0.161)	0.275** (0.123)	0.444	2.19	38.42***
$\sigma_{YZ}$	0.047 (0.033)	0.804*** (0.146)	-0.176 (0.150)	0.239** (0.119)	0.413	2.14	33.29***

higher coefficients.

Furthermore, we develop a more formal test by running an encompassing regression, in which future volatility is regressed on both the Black-Scholes and model-free implied volatilities as specified in Equation 2.24). Intuitively, the implied volatility that retains its statistical significance in this multivariate setting is more efficient. The encompassing regression results are displayed in Table 2.11.

We can see from Table 2.11 that the Black-Scholes implied volatility loses its significance when model-free implied volatility is included. Therefore, the model-free implied volatility subsumes the Black-Scholes implied volatility in predicting future volatility. Our findings based on the encompassing regressions are consistent with those using the goodness-of-fit analysis earlier. In theory, the model-free implied volatility may be expected to outperform the Black-Scholes implied volatility because the former is much less subject to the various measurement errors, as shown by Jiang and Tian (2005).

To check whether our results are sensitive to the sampling procedure, we also use daily, overlapping samples. As mentioned before, the daily overlapping sampling procedure presents a serious problem of serial correlation, which leads to an underestimated

Table 2.10: Information content of Black-Scholes implied volatility versus historical volatility and the GARCH(1,1) volatility forecast. This table reports regression coefficients and standard errors (in parentheses) for Equation (2.23) based on the monthly non-overlapping samples for the period January 1996 through December 2006. F is the Wald-statistic tests the joint hypothesis that  $b_0 = 0$  and  $b_1 = 1$ ,  $b_2 = b_3 = 0$ . DW is the Durbin-Watson statistics. \*\*\*, \*\* and \* denote significance at the 1%, 5% and 10% levels, respectively.

Volatility measure	$b_0$	$b_1$	$b_2$	$b_3$	Adj. $R^2$	DW	F
$\sigma_{CC}$	0.074 * ** (0.028)	0.562 * ** (0.116)	-0.077 (0.137)	0.179* (0.141)	0.285	2.21	32.98***
$\sigma_P$	0.061 * ** (0.024)	0.675*** (0.123)	-0.127 (0.159)	0.216** (0.112)	0.422	2.17	34.67***
$\sigma_{GK}$	0.054 (0.044)	0.709*** (0.181)	-0.139 (0.126)	0.312** (0.149)	0.363	2.14	37.82***
$\sigma_{RS}$	0.056 (0.030)	0.801*** (0.159)	-0.252 (0.147)	0.276** (0.129)	0.398	2.18	36.57***
$\sigma_{YZ}$	0.051 (0.038)	0.781*** (0.166)	-0.172 (0.159)	0.256** (0.139)	0.404	2.11	33.62***

standard error of the coefficient for historical volatility. To reduce this bias, we compute our  $t$ -statistics using the Newey and West (1987) method, which corrects for both heteroskedasticity and autocorrelation in the residuals, caused by the overlapping sampling procedure.

The regression results using the overlapping data for the model-free implied volatility are reported in Table 2.12. As can be seen, using overlapping samples does not alter our earlier conclusions. In particular, the predictive power of model-free implied volatility remains much superior to that of historical volatility, while it subsumes the information contained in historical volatility in predicting future volatility. The overlapping procedure generally results in a higher adjusted  $R^2$ .

We employ the overlapping sampling procedure and re-run the regressions using daily sample for the Black-Scholes implied volatility as well. The model estimates with their  $t$ -statistics adjusted by the Newey and West (1987) method are displayed in Table 2.13. The adjusted  $R^2$  for Equation (2.22) is much higher than that of Equation (2.21), where the independent variable is historical volatility. In the encompassing regression (2.23), all coefficients for implied volatility are significantly different from zero, whereas those

Table 2.11: Information content of Black-Scholes implied volatility versus model-free implied volatility. This table reports regression coefficients and  $t$ -statistics (in parentheses) for Eq. (2.24). DW is the Durbin-Watson statistic.\*\*\*, \*\* and \* denote significance at the 1%, 5% and 10% levels, respectively.

$b_0$	$b_1$	$b_2$	Adj. $R^2$	DW
0.079** (3.29)	0.737*** (4.62)		0.432	0.10
0.152** (3.78)		0.579*** (5.041)	0.355	0.09
0.058*** (4.03)	0.814*** (4.78)	-0.125 (1.47)	0.427	0.07

for historical volatility are much smaller in magnitude, with most being statistically insignificant. These results suggest that our findings regarding the efficiency of implied volatility versus historical volatility are invariant to the sampling procedure.

Overall, we can summarize the results of the out-of-sample tests as follows. Taken alone, all of the model-free implied volatility, Black-Scholes implied volatility, historical volatility, and GARCH volatility forecast contain information concerning future realized volatility. However, model-free implied volatility provides better out-of-sample forecasts than any of the alternative forecasting models for volatility, suggesting that options traders incorporate information beyond that contained in the past series of futures prices in their implicit predictions of the future volatility. Although the model-free implied volatility contains information regarding future volatility, it is not an unbiased estimator.

## 2.7 Robustness analysis

In this section, we perform several tests to verify the robustness of our results. First, we evaluate the impacts of error-in-variable problems on our regressions where the implied volatility is used as a regressor. Second, we analyze whether the information efficiency of implied volatility varies significantly over different subsample periods.

Table 2.12: Information content of model-free implied volatility versus historical volatility using overlapping samples. This table reports the results of the predictive regressions of realized volatility on historical volatility forecast and model-free implied volatility based on the daily overlapping samples for the period January 1996 through December 2006. The number in parentheses are Newey-West (1987) standard errors of the estimated parameters.

$b_0$	$b_1$	$b_2$	Adj. $R^2$	DW
0.079*** (3.29)	0.537*** (4.62)		0.332	0.10
0.182*** (3.78)		0.379 (5.041)	0.125	0.09
0.058*** (4.03)	0.644*** (4.78)	-0.155 (1.47)	0.327	0.07

### 2.7.1 Error-in-variables

The fact that the slope estimate is significantly below one could be either due to implied volatility being a biased forecast or due to the bias induced by the error-in-variable problem. Such errors-in-variable problems cause correlation between regressor and error terms in the forecasting equations (2.19) and (2.23), so called endogeneity problem. Implied volatility may be measured with error as a result of non-synchronous option and futures prices, misspecification of the option pricing formula, etc.

In fact, Christensen and Prabhala (1998) assume that the error-in-variables problem causes implied volatility to appear both biased and inefficient. In particular, it generates a downward bias for the slope coefficient of implied volatility and an upward bias for the slope coefficient of past volatility, which explains the under-estimation of implied volatility and the over-estimation of past volatility. As a result, the usual OLS will lead to false conclusions concerning implied volatility predictive power.

Consequently, they propose to use instrumental variable framework as a way of correcting error-in-variable problems in implied volatility. Since past volatility is related to future volatility and since implied volatility reflects future volatility information, we should have an implied volatility that depends on past volatility. Therefore, efficient estimation can be performed under an instrumental variable procedure.

The procedure is the following. First, the implied volatility is regressed on an instrument and, following Christensen and Prabhala (1998), we choose  $\sigma_{IV,t-1}$  as a natural

Table 2.13: Information content of model-free implied volatility versus historical volatility using overlapping samples. This table reports regression coefficients and  $t$ -statistics (in parentheses) based on the daily overlapping samples for the period January 1996 through December 2006. Standard errors of the coefficients are adjusted by Newey and West's (1987) consistent covariance estimator for both heteroskedasticity and autocorrelation. DW is the Durbin-Watson statistic.

$b_0$	$b_1$	$b_2$	Adj. $R^2$	DW
0.079*** (3.29)	0.537*** (4.62)		0.332	0.10
0.182*** (3.78)		0.379 (5.041)	0.125	0.09
0.058*** (4.03)	0.644*** (4.78)	-0.155 (1.47)	0.327	0.07

candidate for the instrument. Then, equation (2.19) is estimated using OLS by replacing implied volatility,  $\sigma_{IV,t}$ , with the fitted values from this regression of  $\sigma_{IV,t}$  on  $\sigma_{IV,t-1}$ . Within this framework, the following equation is used:

$$\sigma_{IV,t} = a + c\sigma_{IV,t-1} + \varepsilon_t \quad (2.25)$$

Next, implied volatility is regressed on  $\sigma_{IV,t-1}$  and  $\sigma_{HV,t}$  as follows:

$$\sigma_{IV,t} = a + c\sigma_{IV,t-1} + \sigma_{HV,t} + \varepsilon_t \quad (2.26)$$

Then the fitted values from this regression are used in equation (2.23), which is estimated by OLS method.<sup>8</sup>

Table 2.14 reports the estimates and standard errors of the parameters for the instrumental variable estimation. The slope estimate is 0.931, larger than the least squares estimate reported in Table 2.2. The difference between the two estimates captures the bias induced by the measurement errors in the model-free implied volatility and after correcting for the bias, the slope coefficient is significantly different than the null value of one, suggesting that the implied volatility is still a biased predictor of the future realized

<sup>8</sup>These fitted values can also be used in estimating equation (2.19), which is equivalent to estimating equation (2.23) with two instruments,  $\sigma_{IV,t-1}$  and  $\sigma_{HV,t}$ , instead of using only one instrument. According to Christensen and Prabhala (1998), both approaches give empirically similar results. Thus, we provide the results of only the first one.

Table 2.14: Information content of implied volatility: instrumental variables estimates. The standard errors of the parameters are in parentheses. DW is the Durbin-Watson statistics. The sample period is from January 1996 through December 2006.

Panel A: first stage regressions estimates				
Dependent variable: $\sigma_t^{IV}$				
Intercept	IV	HV	$adj.R^2$	DW
0.058 (0.001)	0.876 (0.005)		0.548	2.04
0.053 (0.002)	0.847 (0.007)	-0.067 (0.123)	0.564	2.17
Panel B: second stage IV estimates				
Dependent variable: $\sigma_t^{RV}$				
Intercept	IV	HV	$adj.R^2$	Durbin-Watson
-0.022 (0.066)	0.831 (0.025)		0.39	2.00
-0.024 (0.083)	0.853 (0.031)	0.06 (0.038)	0.42	2.14

volatility.

### 2.7.2 Subsample analysis

Next, we perform a robustness check to ensure that our results are not driven by a specific sample selection. For this purpose we split our sample into two parts. The crude oil futures market has experienced a much higher volatility following the terrorist attacks of 9/11 in 2001. To study whether the forecasting performance of implied volatility changes in period with more volatility than in period with less volatility, we divide our sample into two subsamples, with September 2001 as the dividing point. The first subsample includes observations before September 2001. The second subsample includes all the observations after this date.

Table 2.15 reports the summary statistics of the realized volatility, as measured by various volatility estimators, and the model-free implied volatility under the two subsample periods. On average, both the realized volatilities and the implied volatility are higher in the second subsample. We also perform the regression (2.19) for both subsamples. The results are reported in Table 2.16. As can be seen from the results, the implied volatility

Table 2.15: Descriptive statistics for realized and model-free implied volatilities: Sub-sample period analysis. This table reports the descriptive statistics for the levels and logarithms of realized and implied volatility measures. The volatility measures are: close-to-close ( $RV^C$ ), Parkinson ( $RV^P$ ), Garman-Klass ( $RV^{GK}$ ), Rogers-Satchel ( $RV^{RS}$ ), Yang-Zhang ( $RV^{YZ}$ ) and model-free implied volatility ( $IV$ ). The Panel A presents the summary statistics for monthly volatilities for the period from January 2001 to September 2001, while Panel B presents the summary statistics for the period September 2001 to December 2006.

	$RV^C$	$RV^P$	$RV^{GK}$	$RV^{GK}$	$RV^{RS}$	$RV^{YZ}$	$IV$
<i>Panel A: Subperiod 01/1996 to 09/2001</i>							
Mean	0.344	0.318	0.328	0.370	0.331	0.309	0.354
St. Dev.	0.092	0.066	0.071	0.096	0.076	0.075	0.059
Kurtosis	5.509	4.165	4.385	7.180	5.065	5.374	3.599
Skewness	1.252	0.852	0.954	1.421	1.071	1.221	0.589
<i>Panel B: Subperiod 10/2001 to 12/2006</i>							
Mean	0.326	0.312	0.326	0.360	0.329	0.306	0.392
St. Dev.	0.088	0.071	0.077	0.094	0.081	0.079	0.0794
Kurtosis	7.463	7.906	9.739	12.873	11.306	9.207	3.807
Skewness	1.849	1.802	2.168	2.726	2.440	2.604	0.951

has a higher slope coefficient in the second sample, also implied volatility explains more variability in the realized volatility as judged by the adjusted  $R^2$ . Moreover, if we apply the instrumental variable approach to both subsamples, we get that the slope coefficient of implied volatility increases substantially and is not statistically different from one. Therefore, controlling for error-in-variable problem, in the second subperiod, the implied volatility becomes an unbiased predictor of future realized volatility.

## 2.8 Conclusion

Using the database on the crude oil futures options traded at NYMEX, we compute the model-free implied volatility and investigate its properties with respect to its forecasting power for future realized volatility both in-sample and out-of-sample. In-sample, we find that volatility forecasts estimated from options prices contain incremental information relative to standard GARCH specifications for conditional volatility, which only use the information in past returns. Furthermore, the hypothesis that past returns have no

Table 2.16: Information content of implied volatility: subperiod analysis. The table presents the ordinary least squares (OLS) and instrumental variable (IV) estimates of specification (2.19). The realized volatility is measured with the Garman-Klass volatility estimator. The results reported in Panel A are based on nonoverlapping monthly volatility observations from January 1996 to August 2001, while those reported in Panel B are based on nonoverlapping monthly volatility observations from October 2001 to December 2006. Number in parentheses denote standard errors of the coefficients.

Panel A: 01/1996 - 08/2001			
Method	Intercept	IV	$adj.R^2$
OLS	0.102 (0.036)	0.690 (0.095)	0.345
IV	0.066 (0.023)	0.758 (0.075)	0.341
Panel B: 10/2001 - 12/2006			
Method	Intercept	IV	$adj.R^2$
OLS	0.041 (0.031)	0.815 (0.065)	0.502
IV	-0.034 (0.023)	0.935 (0.032)	0.393

incremental information content in addition to the information conveyed by the options market cannot be rejected.

The out-of-sample volatility forecasts confirm the above results. Implied volatility measures outperform historical volatility and GARCH(1,1) volatility forecasts in forecasting future realized volatility. Moreover, the model-free implied volatility encompasses all the alternative volatility forecasts, so that adding historical volatility, GARCH volatility or Black-Scholes volatility as explanatory variables does not increase the forecast ability by substantial amount. Furthermore, we employ an overlapping sampling procedure and repeat the analysis for both implied volatilities and reach the same conclusions as above.

Since the transaction costs for arbitrage activity are significantly lower for crude oil futures options than for equity options, this market is more efficient. Therefore, the implied volatility is more likely to reflect the market's ex ante expectation of the future volatility in the crude oil futures option market. After all, implied volatility is a forward-looking forecast, unlike any historically-based volatility. However, consistent with previous studies, and despite its superior performance over the historically-based

volatility measures, implied volatility is not an unbiased estimator for future volatility.

The implied volatility appears to be less biased and more efficient once we account for error-in-variables and apply instrumental variable approach. We also find that in the light, sweet crude oil market, the implied volatility performs better during the more volatile period following September 11 terrorist attacks.

Numerous interesting directions for future research remain.

One of the directions for future research is to understand the sources of the forecasting bias in the crude oil futures options market in order to improve the quality of the option-based forecasts of future volatility. One potential explanation for the observed bias in option implied volatility is the presence of a volatility risk premium, which may induce dependence between the regressor and the residuals and contaminate the inference regarding the efficiency of option markets for predicting realized volatility. Thus, accounting for volatility risk premium may reduce the bias.

A recent literature also documents that use of high-frequency data leads to significant improvements in measuring and modeling volatility. For example, Poteshman (2000) finds that half of the bias in implied volatility forecasts obtained from S&P500 futures options was removed, when actual volatility was estimated with a more efficient volatility estimator based on intraday 5-minute returns. Thus, using realized volatility obtained from high-frequency intraday data may lead to more efficient inference.

Recent studies have also suggested the importance of explicitly allowing for jumps in the estimation of stochastic volatility models and in the pricing of options and other derivatives. Thus, volatility forecast accuracy may be further improved by explicitly modeling the jump component, if present.

Relatively little work has been done on modeling and forecasting implied volatility itself, compared with the extensive literature on volatility modeling and forecasting in general. Thus, it would be interesting to model time series of model-free implied volatility with variables, such as returns, trading volume and open interest, and assess the forecasting performance of these models.

Finally, since the results suggest that implied volatility and time-series forecasts contain independent information about future volatility, a promising and important avenue for future research lies in combining time-series forecasts with implied volatility. We leave further work along these lines for future research.

## Chapter 3

### Variance risk premium in the crude oil futures

## 3.1 Introduction

A number of studies have examined if the implied volatility has any explanatory power regarding the future realized volatility, if it outperforms other volatility forecast measures, and if it is an unbiased estimator of ex post realized volatility.<sup>1</sup> Although the results of these studies provide sufficient evidence of the efficiency of implied volatility, most of these studies also find that implied volatility is an upward-biased predictor, such that the implied volatility is on average higher than the realized volatility. This difference may be caused by measurements errors or model misspecification. However, Poteshman (2000) shows that the bias in the implied volatility does not disappear after measurement errors are corrected by using last-period implied and realized volatilities as instrumental variables.

One potential explanation for the observed bias in option implied volatility as a forecast of future realized volatility is the presence of a volatility risk premium. In fact, when examining variance swap contracts, Carr and Wu (2009) and Bondarenko (2007) find large, negative and statistically significant unconditional variance risk premia for stock indices, as the option implied volatility, on average, exceeds the ex-post realized volatility of the underlying asset by a substantial amount. Moreover, the variance risk premia appear to fluctuate over time, typically increasing with the level of expected volatility itself and showing sensitivity to the business cycle. Thus, a negative market volatility risk premium has the interpretation that investors are willing to pay a premium to hold options in their portfolio.

The variance risk premium is not directly observable. However, an empirical proxy can be constructed from the difference between implied variance and the conditional expectation of realized variance. The use of BlackScholes implied volatilities and/or realized volatilities from daily returns generally results in biased and inefficient estimates of the risk premium parameter, leading to unreliable statistical inference. It was only with the development of model-free implied volatility that it became possible to develop more precise tests for the existence of risk premia in option prices.

The concept of the model-free implied volatility was developed in original work of Dupire (1994) and Neuberger (1994). The concept is referred to as "model-free" because it does not rely on any particular parametric model, unlike the Black-Scholes implied

---

<sup>1</sup>See, e.g., Latane and Rendleman (1976), Day and Lewis (1992), Canina and Figlewski (1993), Lamoureux and Lastrapes (1993), Jorion (1995), Christensen and Prabhala (1998), Ederington and Guan (2002), and Blair, Poon, and Taylor (2001).

volatility. The model-free implied volatility aims to measure the expected return variation evaluated under the risk-neutral, or pricing, measure. Since volatility is stochastic, this measure will typically differ from the expected return variation under the actual, or objective, measure. Thus, in general, implied volatilities will include premiums compensating for the risk associated with the exposure to the volatility.

The goal of this study is to quantify the variance risk premium in the crude oil futures market by synthesizing the variance swap rate from prices of options written on the crude oil futures and analyze its dynamics. Variance swap is defined as an over-the-counter contract that pays the difference between a standard estimate of the realized variance and the fixed variance swap rate. Since the price of such variance swaps is zero at initiation, the variance swap rate represents the risk-neutral expected value of the realized variance over the life of the swap. Thus, the difference between the ex post realized variance and this synthetic variance swap rate can be used to measure the variance risk premium.

Our findings are similar to those reported for equity markets. We find that the average variance risk premium is strongly negative in the crude oil futures market. This indicates that variance buyers are willing to pay a premium to hedge away the risk of an increase in volatility. We also find that the market variance risk premium is time-varying and correlated with the variance swap rate.

The remainder of the chapter is organized as follows. Section 3.2 describes previous literature relevant to this study. Section 3.3 provides the definition of the variance risk premium. Section 3.4 describes the methodology. Section 3.5 presents the empirical results. Section 3.6 concludes.

## 3.2 Literature review

Lamoureux and Lastrapes (1993) were among the first to build a link between the *ex post* realized volatility and at-the-money implied volatility based on the Hull and White (1987) option pricing model in the equity market. Their findings suggest that the bias in the volatility forecasts may be due to the omitted volatility risk premium. Similarly, Jackwerth and Rubinstein (1996) report that at-the-money implied volatilities of call and put options are consistently higher than their realized volatilities, suggesting that a negative volatility premium could be an explanation to this empirical irregularity. Chernov (2001) finds that when implied volatility is discounted by a volatility risk premium and

when the errors-in-variables problems in historical and realized volatility are removed, the unbiasedness of the S&P100 index option implied volatility cannot be rejected over the sample period from 1986 to 2000.

Bakshi and Kapadia (2003) also investigate the pricing of market volatility risk premium in options written on the S&P 500 index by considering the gains on a delta-hedged option portfolio. They find that the gains from this portfolio are significantly negatively correlated with the level of market volatility, which is consistent with the negative volatility risk premium. They also find that the magnitude of the negative premium is higher for out-of-the-money options and therefore dependent on moneyness levels. Similar results were obtained by Carr and Wu (2009). They show that the difference between the ex-post realized variance and the synthetic variance swap rate quantifies the variance risk premium and that this variance risk premium is strongly negative and time-varying.

When examining variance swap contracts, Bondarenko (2007) find large, negative and statistically significant unconditional variance risk premia for stock indices, as the option implied volatility, on average, exceeds the ex-post realized volatility of the underlying asset by a substantial amount. They also find that these premia cannot be explained by commonly used risk factors. Moreover, the variance risk premia appear to fluctuate over time, typically increasing with the level of expected volatility itself.

Bates (1996, 2000, 2003), Benzoni (2002), Chernov and Ghysels (2000), Pan (2002), Eraker (2004), among others, try to explicitly model and estimate variance risk premium in equity markets and find it to be significant based on the information contained in the option prices and the returns on the underlying assets. These studies calibrate structural models to option prices allowing for the existence of variance risk and report a significant volatility risk premium. However, these models usually are based on some strong assumptions and hence, their results are prone to model misspecification. Bondarenko (2004) and Todorov (2010) show that the difference between the model-free implied variance and the realized variance can be used to estimate the variance risk premium. Bollerslev, Gibson, and Zhou (2008) also confirm that using model-free implied volatility from options, the volatility risk premium can be estimated nearly as well as if the actual risk-neutral implied volatility and continuous time integrated volatility were used.

Economic theory provides several reasons why the price of risk of innovations in market volatility should be negative. For example, Campbell (1993, 1996) show that investors hedge against changes in market volatility, because increasing volatility represents a deterioration in investment opportunities. French, Schwert, and Stambaugh

(1987) and Campbell and Hentschel (1992) show that periods of high volatility also tend to coincide with downward market movements. Coval and Shumway (2001) investigate the properties of at-the-money zero-beta straddles on the S&P 100 index and find that their returns are consistently lower than the risk-free rate, suggesting the presence of a negative volatility risk premium in the prices of options. They show that investors are willing to pay a premium to hold options in their portfolio as a hedge against adverse shocks. This makes the option price higher than what it would have been if volatility risk were not priced which causes negative volatility risk premium.

Bakshi and Madan (2006) show that the disparity between two volatilities arises from a non-Gaussian property of returns and from the non-risk-neutrality of investors. More specifically, they find that the difference between implied and realized volatilities, denoted as the volatility spread, is the largest when the physical marginal density function of asset returns, is negatively skewed and fat tailed. Further studies estimate the volatility risk compensation and its relation with macroeconomic and financial determinants. For example, Bollerslev, Tauchen and Zhou (2009) observe that realized volatility is systematically lower than implied volatility and determined by AAA corporate bond spread over Treasuries and the  $P/E$  ratio. Bollerslev and Todorov (2010) suggest that the large premium is tied to time-varying compensation for tail risks. Fornari (2010) show that interest rate volatility has embodied a large and negative compensation for volatility risk, in line with other studies focusing on different asset classes. They also found that the volatility risk compensation is extremely time varying.

The variance risk premiums have been estimated for currencies and equities. However, the evidence has been so far nearly absent for commodity markets. Using Monte Carlo simulation, Doran and Ronn (2008) demonstrate that the volatility risk premium is the only parameter that generates the disparity between implied and realized volatilities in energy markets, even in the presence of jumps and a jump premium. They use a two-step procedure to estimate volatility risk premium based on the Heston's volatility specification. But the limitation of the method is that it places strong restrictions on the data-generating process, e.g., the correlation between the volatility and the asset return is zero. Trolle and Schwartz (2010) estimate the variance risk premium for crude oil and natural gas using the model-free methodology and find that the average variance risk premia are negative for both commodities. Moreover, they find that the energy variance risk premia are time varying.

Thus, the empirical evidence suggests a large, negative and strongly time-varying

variance risk premium that is also negatively correlated with market returns.

### 3.3 Variance Risk Premium: Definition and Measurement

The variance risk premium is defined as the difference between the ex-ante risk neutral expectation of the future return variance and the objective or statistical expectation of the return variance over the  $[t, T]$  time interval,

$$VRP_t \equiv E_t^Q(RV_{t,T}) - E_t^P(RV_{t,T}) \quad (3.1)$$

which is not directly observable in practice. To construct an empirical proxy for such a variance risk premium, one need to estimate various reduced-form counterparts of the risk neutral and physical expectation, i.e.,  $\hat{V}RP_t \equiv \hat{E}_t^Q(RV_{t,T}) - \hat{E}_t^P(RV_{t,T})$ .

The concept of a variance risk premium can be also shown in the context of a variance swap, which is a derivative instrument contingent on volatility itself. A variance swap has zero net market value at entry and it allows investors to trade future realized variance of a given asset against current implied variance. At maturity, the payoff to the long side of the swap is equal to the difference between the realized variance over the life of the swap contract and a variance swap rate:

$$[RV_{t,T} - IV_{t,T}]L \quad (3.2)$$

where  $RV_{t,T}$  is the realized annualized return variance between time  $t$  and  $T$ ,  $IV_{t,T}$  is the fixed variance swap rate that is determined at time  $t$  and paid at time  $T$ , and  $L$  is the notional dollar amount that converts the variance difference into a dollar payoff. In the absence of arbitrage opportunities, the variance swap rate equals the risk-neutral expected value of the realized variance

$$IV_{t,T} = E_t^Q[RV_{t,T}] \quad (3.3)$$

where  $E_t^Q[\cdot]$  is the time- $t$  conditional expectation operator under some risk-neutral measure  $Q$ .

Britten-Jones and Neuberger (2000) show that the risk-neutral expected sum of

squared returns on a financial asset over any interval  $[t, T]$  can be obtained from the set of market prices at time  $t$  for European options expiring on time  $T$ . The derivation assumes a full set of call options with a continuum of strike prices and a continuum of maturities is traded on the asset. Under regularity assumptions, they find that

$$IV_{t,T} \equiv E_t^Q(RV_{t,T}) = \frac{2}{B_t(T)(T-t)} \left( \int_0^{F(t)} \frac{P(t, T, K)}{K^2} dK + \int_{F(t)}^\infty \frac{C(t, T, K)}{K^2} dK \right) \quad (3.4)$$

where  $B_t(T)$  is the time- $t$  price of a bond paying one dollar at time  $T$ ,  $K$  is a strike price,  $C(t, T, K)$  is a call option, when  $K$  is greater than the futures price  $F_t$ , and  $P(t, T, K)$  is the put option, when  $K \leq F_t$ .

This measure can be estimated simply using the current market prices of traded options. It remains valid under very general assumptions on the underlying stochastic process, so that it is a model-free measure. The model-free implied volatility implicit in this identity is an expectation with respect to the risk-neutral distribution. If volatility risk is priced, the expected volatility under the risk-neutral measure and the expected volatility under the physical measure differ by a quantity reflecting the associated risk premium.

The continuum of option prices is needed for estimation of the risk-neutral implied variance, which is not available in the financial markets, but accurate approximations can be obtained through a discretization procedure. Under reasonable assumptions about the underlying asset dynamics,  $IV_{t,T}$  can be constructed on the basis of a finite number of strikes, which is a fairly accurate approximation to the true (unobserved) risk-neutral expectation of the future market variance (Jiang and Tian (2005), Carr and Wu (2008), Bollerslev, Gibson and Zhou (2008)).

In the studies on the volatility risk premium in the equities market, the risk-neutral expectation  $\hat{E}_t^Q(RV_{t,T})$  is typically replaced by the CBOE (Chicago Board of Exchange) implied variance index, so-called VIX, for the S&P 500 equity index<sup>2</sup>. We will also apply the CBOE's methodology for calculating VIX to the crude oil futures market. The formula for calculating the index is:

$$\sigma_{t,T}^2 = \frac{2e^{r_f T}}{T} \sum_{i=I_l}^{I_u} \frac{\Delta K_i}{K_i^2} Q(K_i) - \frac{1}{T} \left[ \frac{F_0}{K_0} - 1 \right]^2 \quad (3.5)$$

<sup>2</sup>See [www.cboe.com/micro/vix/vixwhite.pdf](http://www.cboe.com/micro/vix/vixwhite.pdf) for details on calculating the VIX

where  $K_i$  is the strike price of the  $i$ -th out-of-the-money option (a call if  $K_i > F$  and a put otherwise);  $\Delta K_i$  is the difference between strike prices defined as  $\Delta K_i = \frac{K_{i+1} - K_{i-1}}{2}$ ;  $Q(K_i, T)$  is the settlement price of the option with strike price  $K_i$ ;  $F$  is a forward index level derived from the nearest to the money option prices by using put-call parity. The last term in (3.5) reflects a correction for the discrepancy between  $K_0$  and the forward price which turns all discretely sampled option values into out-of-the-money prices.

The CBOE determines these basic variables from the available options data in three separate steps. First, at a given point in time, and for each relevant maturity separately, they identify the strike price  $K$  for which the distance between the quoted midpoints of the call and put prices is minimal. This strike is then used to compute the implied forward index level according to the formula,  $F = K + e^{rT}(C(K, T) - P(K, T))$ . Finally, the first strike price available below the forward rate  $F$ .

Since only a few options maturity dates are available on any given day, the CBOE applies an additional approximation technique to calculate a constant 30-calendar day bracket of implied volatility. First, the VIX measure for the two expiration dates closest to thirty calendar days,  $T_1$  and  $T_2$ , but excluding options with less than seven calendar days to expiry, is obtained. Then, linear interpolation between  $\sigma_{T_1}^2$  and  $\sigma_{T_2}^2$  to obtain an estimate at 30-day maturity leads to:

$$\sigma_t^{MF} = \sqrt{\frac{365}{30} \times \left\{ T_1 \sigma_{T_1}^2 \left[ \frac{N_{T_2} - 30}{N_{T_2} - N_{T_1}} \right] + T_2 \sigma_{T_2}^2 \left[ \frac{30 - N_{T_1}}{N_{T_2} - N_{T_1}} \right] \right\}}$$

where  $N_{T_1}$  and  $N_{T_2}$  denote the number of actual days to expiration for the two maturities. Thus, the squared value of the VIX can be regarded as the risk-neutral expectation of the annualized 30 calendar day return variance.

Finally, in order to define the measure in quantifying the actual return variation, let  $C_t$  denote the closing futures price,  $O_t$  denote the opening futures price,  $H_t$  denote the high futures price,  $L_t$  denote the low futures price on a trading day  $t$ . The realized variance over the discrete  $t$  to  $T$  time interval can be measured using Yang-Zhang range-based volatility estimator as follows:

$$\sigma_{YZ,t} = \sqrt{\frac{1}{N} \sum_{t=1}^N \left[ \ln \frac{O_t}{C_{t-1}} - \overline{\ln \frac{O_t}{C_{t-1}}} \right]^2 + \frac{\kappa}{N} \sum_{t=1}^N \left[ \ln \frac{C_t}{O_t} - \overline{\ln \frac{C_t}{O_t}} \right]^2 + (1 - \kappa) \sigma_{RS,t}^2} \quad (3.6)$$

with:

$$\kappa = \frac{0.34}{1.34 + \frac{N+1}{N-1}}$$

where  $\overline{\ln \frac{O_t}{C_{t-1}}} = 1/N \sum_{t=1}^N \ln \frac{O_t}{C_{t-1}}$ ,  $\overline{\ln \frac{C_t}{O_t}} = 1/N \sum_{t=1}^N \ln \frac{C_t}{O_t}$ , and  $\sigma_{RS,t} = \ln \left( \frac{H_t}{C_t} \right) \ln \left( \frac{H_t}{O_t} \right) + \ln \left( \frac{L_t}{C_t} \right) \ln \left( \frac{L_t}{O_t} \right)$  is the Rogers-Satchell (1991) estimator.

The dataset used in the study contains daily time series of light, sweet crude oil futures and options on these futures traded at the NYMEX for the period from January 1996 through December 2006. NYMEX lists futures contracts with monthly expirations several years into the future and American-style options on these futures, which expire three business days before the underlying futures contracts.

From the dataset we obtain observations on daily settlement prices for call and put options. We consider only options at the two nearest maturities. When the time to maturity for the nearby options is less than eight days, we switch to the next two maturities. We match all puts and calls by trading date, maturity, and strike. Options which violate boundary conditions are not considered. To estimate the ex post realized volatility, we use the data on daily high, low, opening and closing prices for underlying futures contracts also traded at the NYMEX. We use only the futures contracts with the same contract months as options to ensure the best match between the implied volatility and the realized volatility calculated from subsequent futures prices.

### 3.4 Methodology

As shown in Carr and Wu (2009), using  $P$  to denote the statistical probability measure, the variance swap rate is linked to the realized variance through the following equation:

$$IV_{t,T} = \frac{E_t^P[M_{t,T}RV_{t,T}]}{E_t^P[M_{t,T}]} = E_t^P[m_{t,T}RV_{t,T}] \quad (3.7)$$

where  $M_{t,T}$  is a pricing kernel and  $m_{t,T} = M_{t,T}/E_t^P[M_{t,T}]$  is a normalized pricing kernel. These equivalent pricing equations imply that the expected difference under the physical measure between realized volatility and the swap rate (model-free implied volatility) reflects a variance risk premium,  $-Cov_t^P(m_{t,T}, RV_{t,T})$ , arising from the covariation between the normalized pricing kernel  $m_{t,T}$  and return variance  $RV_{t,T}$ ,

$$E_t^P[RV_{t,T} - IV_{t,T}] = -cov_t^P(m_{t,T}, RV_{t,T})$$

Hence, the conditional expectation  $E_t^P[RV_{t,T} - IV_{t,T}]$  equals the conditional variance risk premium in effect for the period  $[t, T]$ .

Equation (3.7) can be decomposed into the following two terms:

$$IV_{t,T} = E_t^P[m_{t,T}RV_{t,T}] = E_t^P[RV_{t,T}] + Cov_t^P(m_{t,T}, RV_{t,T}) \quad (3.8)$$

The first term  $E_t^P[RV_{t,T}]$  represents the time-series conditional mean of the realized variance. The second term captures the conditional covariance between the normalized pricing kernel and the realized variance. The negative of this covariance defines the return variance risk premium. Thus, a direct estimate of the average variance risk premium is the sample average of the difference between the variance swap rate and the realized variance,  $RP_{t,T} \equiv RV_{t,T} - IV_{t,T}$ .

Finally, dividing both sides of Equation (3.8) by  $IV_{t,T}$ , equation (3.8) can be represented in return terms:

$$1 = E_t^P[m_{t,T} \frac{RV_{t,T}}{IV_{t,T}}] = E_t^P[\frac{RV_{t,T}}{IV_{t,T}}] + Cov_t^P(m_{t,T}, \frac{RV_{t,T}}{IV_{t,T}}) \quad (3.9)$$

If  $IV_{t,T}$  is the forward cost of a variance swap investment, then  $(RV_{t,T}/IV_{t,T} - 1)$  is the excess return from the investment. The sample average of the excess return represents an estimate of the negative of the covariance term in Equation (3.9), which is the risk premium.

Following the methodology of Carr and Wu (2009), in order to understand the dynamic behavior of variance risk premiums, the regression by which the realized variance is regressed on the variance swap rate can be used as follows:

$$RV_{t,T} = a + b \times SW_{t,T} + e \quad (3.10)$$

Under the null hypothesis of zero variance risk premiums in dollar terms:  $Cov_t^P(m_{t,T}, RV_{t,T}) = 0$  as defined in Equation (3.7), we should have  $a = 0$  and  $b = 1$ . In particular, the slope coefficient equal to zero would suggest the absence of the the variance risk premium, whereas the slope coefficient different from zero would imply time-varying risk premium correlated with the variance swap rate.

We also run the expectation hypothesis regression in log terms as follows:

$$\ln RV_{t,T} = a + b \times \ln SW_{t,T} + e \quad (3.11)$$

Under the null hypothesis of zero variance risk premium in return terms:  $Cov_t^P(m_{t,T}, \frac{RV_{t,T}}{SW_{t,T}}) = 0$  as defined in Equation (3.9), the slope estimate  $b$  should be zero and the intercept estimate should be lower than zero due to the convexity term induced by the variance of the log variance risk premiums.

### 3.5 Empirical results

Table 3.1: Descriptive statistics: realized variance (RV), implied variance (IV), and variance risk premium (VRP). This table reports summary statistics for RV, IV, and VRP. Panel A reports the realized variance (RV), implied variance (IV), and the variance risk premium ( $VRP = (RV_{t,T} - IV_{t,T}) \times 100$ ). Panel B reports the logarithm of realized variance ( $\log(RV)$ ), the logarithm of implied variance ( $\log(IV)$ ), and the log variance risk premium ( $LVRP = \log(RV_{t,T}/IV_{t,T})$ ). Statistics include mean, standard deviation, skewness, kurtosis, and serial autocorrelation coefficients with lags of 1 and 20 trading days. The sample period extends from January 1996 through December 2006.

Panel A			
	RV	IV	VRP
Mean	0.328	0.373	-4.549
Std Dev	0.074	0.072	5.714
Skewness	1.548	0.974	-0.235
Kurtosis	6.841	4.287	6.208
$\rho_1$	0.990	0.983	
$\rho_{20}$	0.731	0.827	
Minimum	0.188	0.251	-29.535
Maximum	0.637	0.640	14.810
Panel B			
	$\log(RV)$	$\log(IV)$	LVRP
Mean	-0.494	-0.436	-0.058
Std Dev	0.090	0.080	0.067
Skewness	0.664	0.407	0.040
Kurtosis	4.074	3.100	4.196
Minimum	-0.727	-0.601	-0.269
Maximum	-0.196	-0.194	0.155

Figure 3.1 plots the monthly time series of variance risk premium, implied variance, and realized variance and Table 3.1 reports the summary statistics of the annualized real-

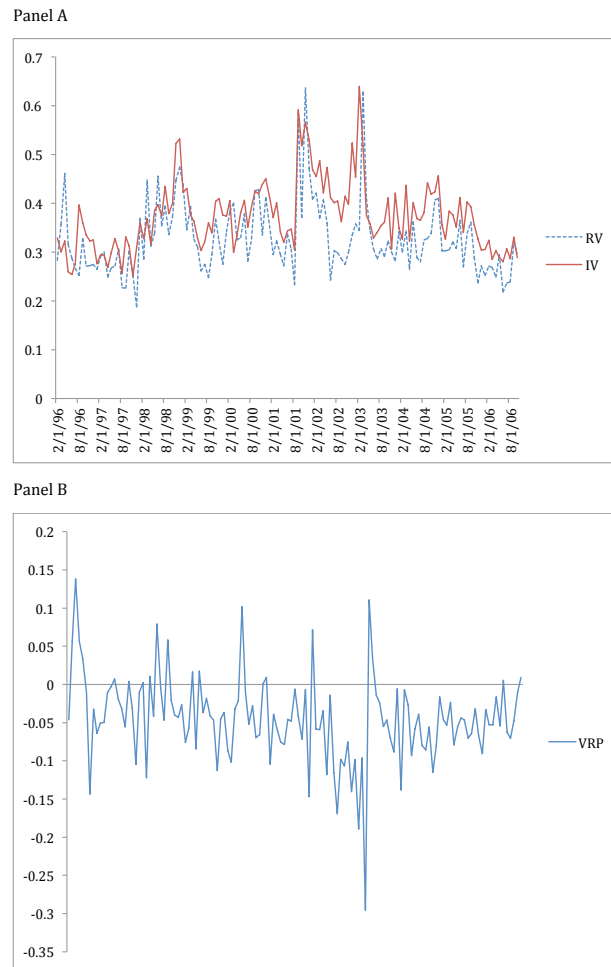


Figure 3.1: Time series plot of the implied, realized variances, and variance risk premium. This figure plots the time series of implied and realized variances (Panel A) and the variance risk premium (Panel B) for the crude oil futures market. The sample period from January 1996 to December 2006.

ized variance (RV), the model-free implied variance (IV), and the variance risk premium (VRP). Comparing the model-free implied variance (MFIV) to the realized volatility, we observe that the model-free implied variance has consistently been higher than the realized volatility throughout the sample period. Since the MFIV is meant to be the risk-neutral expected realized volatility, it suggests that the volatility dynamics under the risk-neutral pricing measure is different from that under the physical probability measure. In other words, the volatility risk has mostly likely been priced by the market.

Both variance swap rates and the realized variance show positive skewness and positive excess kurtosis. The RV is more volatile than the IV, as evidenced by higher standard deviation, skewness and kurtosis statistics. These results are in line with the ones obtained for stock options market. We also observe a very strong degree of persistence in the volatility measures. The autocorrelation patterns decay extremely slowly and are well approximated by a hyperbolic shape. Table 3.1 also reports the summary statistics of the difference between the realized variance and the variance swap rate,  $VRP = (RV_{t,T} - IV_{t,T}) \times 100$  and the log difference  $LVRP = \ln(RV_{t,T}/IV_{t,T})$ . The variance risk premiums VRP show large kurtosis and large skewness. The skewness and kurtosis are smaller for the log variance risk premium LVRP. The sample averages of the variance risk premium and log variance risk premiums are both negative.

The two definitions of variance risk premiums in Table 3.1 represent two ways of computing returns for variance swap investments (Carr and Wu (2009)). The mean estimates in of  $(RV - IV) \times 100$ , represent the average dollar profit and loss for each \$100 notational investment in the variance swap contract. Thus, if we long a 30-day variance swap contract with a notional of \$100 on the crude oil futures and hold the contract to maturity, during our sample period the average return per \$100 notional investment is \$3.46.

Table 3.2 reports the estimates of Equation (3.10) and the  $t$ -statistics under the null hypotheses of  $a = 0$  and  $b = 1$ . The slope estimate is positive, but lower than 1. The  $t$ -statistics confirm these results and do not reject the null that regression slope is significantly lower than the null value of 1.

The estimation results in Table 3.2 show that the slope estimate is closer to unity but still significantly different from one. The difference between the slope estimates of the two regressions indicates that the risk premiums defined in log returns is more constant over time and independent from the log variance swap rate than the risk premiums defined in dollar terms. These results are consistent with those of Carr and Wu (2009) for the

Table 3.2: Time-variation in variance risk premium. The table presents the OLS estimation results of the regressions (3.10) and (3.11). The  $t$ -statistics under the null hypothesis of  $a = 0$  and  $b = 1$  are calculated according to Newey and West (1987) with 30 lags and reported in parentheses. The sample period is from January 1996 through December 2006.

	a	b	$R^2$
(3.10)	0.039 (3.972)	0.610 (-5.692)	0.342
(3.11)	-0.883 (-7.452)	0.698 (-3.582)	0.375

equity markets.

## 3.6 Conclusion

In this chapter, we implemented the model-free method to quantify the variance risk premium in the crude oil futures market from the prices of futures and options on these futures. Since the variance swap rate represents the risk-neutral expected value of the realized variance, the difference between the realized variance and the variance swap rate is used as a measure of the variance risk premium.

We find that the variance risk premium is strongly negative for crude oil futures. The negative sign on the variance risk premiums indicates that investors in this market consider market volatility going up as an unfavorable shock, and thus, are willing to pay a large premium to hedge against market volatility going up. When we regress the realized variance on the variance swap rate, we obtain the slope estimates that are significant and less than one, suggesting the risk premium is time-varying and correlated with the variance swap rate.

There is still a number of directions left for future research. First, given the evidence on negative variance risk premium, it is important to understand which risk factors can account for this premium. Also it would be interesting to study the dynamic interactions of return variance and variance risk premiums, as well as the forecastability of variance risk premium in the crude oil futures market.

The recent studies suggest that the premia to compensate for jump and continuous

volatility risks required by investors in options markets differ (e.g., Todorov (2010)). Thus, by allowing for different risk premia associated with the future risks originating from the continuous sample path price process and the jump component, it may be possible to achieve more accurate volatility risk premium measurement.

Finally, it would also be interesting to more closely compare the estimated volatility risk premium for the crude oil futures market to that of other markets. We leave further work along these lines for future research.



## Bibliography

- 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.
- Aït-Sahalia, Y., Mykland, P. A., and L. Zhang (2005). How often to sample a continuous-time process in the presence of market microstructure noise. *Review of Financial Studies* 18, 351-416.
- Alizadeh, S. (1998). *Essays in Financial Econometrics*. Department of Economics, University of Pennsylvania, PhD Dissertation.
- Alizadeh, S., Brandt, M. W., and F. X. Diebold (2002). Range-based estimation of stochastic volatility models. *Journal of Finance* 57, 1047-1092.
- Amin, K. I., and V. K. Ng (1997). Inferring future volatility from the information in implied volatility in Eurodollar options: a new approach. *Review of Financial Studies* 10, 333-367.
- Andersen, T. G. (1996). Return volatility and trading volume: an information flow interpretation of stochastic volatility. *Journal of Finance* 51, 169-204.
- Andersen, T. G., and T. Bollerslev (1997). Heterogeneous information arrivals and return volatility dynamics: Uncovering the long-run in high frequency returns. *Journal of Finance* 52, 975-1005.
- Andersen, T. G., and T. Bollerslev (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. and H. Ebens (2001). The distribution of stock return volatility. *Journal of Financial Economics* 61, 43-76.
- Andersen, T. G., Bollerslev, T., Diebold, F. X. and P. Labys (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association* 96, 42-55.
- Andersen, T. G., Bollerslev, T., Diebold, F. X. and P. Labys (2003). Modeling and forecasting realized volatility. *Econometrica* 71, 579-625.
- Baillie, R. T., Bollerslev, T., and H. O. Mikkelsen (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 74, 3-30.
- Bakshi, G., and N. Kapadia (2003). Delta-hedged gains and negative market volatility risk premium. *Review of Financial Studies* 16, 527-566.
- Bakshi, G., and D. Madan (2006). A theory of volatility spread. *Management Science* 52, 1945-1956.
- Bali, T. G. (2003). An extreme value approach to estimating volatility and value at risk. *Journal of Business* 76, 83-108.
- Bali, T. G., and D. Weinbaum (2005). Comparative study of alternative extreme value volatility estimators. *Journal of Futures Markets* 25, 873-892.

- Ball, C. A., and W. N. Torous (1984). The maximum likelihood estimation of security price volatility: Theory, evidence, and application to option pricing. *Journal of Business* 57, 97-112.
- Bandi, F., and J. Russell (2006). Separating microstructure noise from volatility. *Journal of Financial Economics* 79, 655-692.
- Barndorff-Nielsen, O. E., and N. Shephard (2002). Estimating quadratic variation using realized variance. *Journal of Applied Econometrics* 2, 1-37.
- Bates, D. S. (1996). Jumps and stochastic volatility: exchange rate process implicit in Deutsche mark options. *Review of Financial Studies* 9, 69-107.
- Bates, D. S. (2000). Post-87 crash fears in S&P 500 futures options. *Journal of Econometrics* 94, 181-238.
- Bates, D. S. (2003). Empirical option pricing: a retrospection. *Journal of Econometrics* 116, 387-404.
- Becker, R., Clements, A. E., and S. I. White (2006). On the informational efficiency of S&P 500 implied volatility. *North American Journal of Economics and Finance* 17, 139-153.
- Beckers, S. (1981). Standard deviations implied in option prices as predictors of future stock price volatility. *Journal of Banking and Finance* 5, 363-381.
- Benzoni, L. (2002). Pricing options under stochastic volatility: an empirical investigation. Working paper, University of Minnesota.
- Black, F., and M. Scholes (1973). The Pricing of Options and Corporate Liabilities. *Journal of Political Economy* 81, 637-659.
- Black, F. (1976). Studies of stock market volatility changes. *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, 177-181.
- Blair, B., Poon, S.-H., and S. J. Taylor (2001). Forecasting S&P 100 volatility: The incremental information content of implied volatilities and high frequency index returns. *Journal of Econometrics* 105, 5-26.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307-327.
- Bollerslev, T., Chou, R. Y., and K. F. Kroner (1992). ARCH modeling in finance: a review of the theory and empirical evidence. *Journal of Econometrics* 52, 5-59.
- Bollerslev, T., Gibson, M., and H. Zhou (2008). Dynamic estimation of volatility risk premia and investor risk aversion from option-implied and realized volatilities. *Journal of Econometrics* 160, 235-245.
- Bollerslev, T., Tauchen, G., and H. Zhou (2009). Expected stock returns and variance risk premia. *Review of Financial Studies* 22, 4463-4492.

- Bollerslev, T., and V. Todorov (2010). Estimating of jump tails. Working paper, Northwestern University.
- Bollerslev, T., and J. M. Wooldridge (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews* 1, 143-173.
- Bollerslev, T., and H. Zhou (2006). Volatility puzzles: a simple framework for gauging return-volatility regressions. *Journal of Econometrics* 131, 123-150.
- Bondarenko, O. (2004). Market price of variance risk and performance of hedge funds. Working paper. University of Illinois.
- Bondarenko, O. (2007). Do hedge funds sell short variance? Working paper, University of Illinois, Chicago.
- Brandt, M. W., and F. X. Diebold (2006). A no-arbitrage approach to range-based estimation of return covariances and correlations. *Journal of Business* 79, 61-74.
- Brandt, M. W., and C. Jones (2006). Volatility forecasting with range-based EGARCH models. *Journal of Business and Economic Statistics* 79, 61-74.
- Breeden, D. T., and R. H. Litzenberger (1978). Prices of state-contingent claims implicit in option prices. *Journal of Business* 51, 621-651.
- Britten-Jones, M., and A. Neuberger (2000). Option prices, implied price processes, and stochastic volatility. *Journal of Finance* 55, 839-866.
- Brown, S. (1990). Estimating volatility. In S. Figlewski, W. Silber and M. Subramanyam (Eds.), *Financial options: from theory to practice*. Business One Irwin, 516-537.
- Brownlees, C. T., and G. M. Gallo (2010). Comparison of volatility measures: a risk management perspective. *Journal of Financial Econometrics* 8, 29-56.
- Campbell, J. Y. (1993). Intertemporal asset pricing without consumption data. *American Economic Review* 83, 487-512.
- Campbell, J. Y. (1996). Understanding risk and return. *Journal of Political Economy* 104, 298-345.
- Campbell, J. Y., and L. Hentschel (1992). No news is good news: an asymmetric model of changing volatility in stock returns. *Journal of Financial Economics* 31, 281-318.
- Canina, L., and S. Figlewski (1993). The informational content of implied volatility. *Review of Financial Studies* 6, 659-681.
- Carr, P., and D. Madan (1998). Towards a theory of volatility trading. In *Volatility: New estimation techniques for pricing derivatives*, R. A. Jarrow (editor), Risk Publications, London.
- Carr, P., and L. Wu (2006). A tale of two indices. *Journal of Derivatives* 13, 13-29.
- Carr, P., and L. Wu (2009). Variance risk premiums. *Review of Financial Studies* 22, 1311-1341.

- Cheong, C. W. (2009). Modeling and forecasting crude oil markets using ARCH models. *Energy Policy* 37, 2346-2355.
- Chernov, M. (2001). Implied volatilities as forecasts of future volatility, the market risk premia, and returns variability. Working paper, Columbia Business School.
- Chernov, M., and E. Ghysels (2000). A study towards a unified approach to the joint estimation of objective and risk-neutral measures for the purpose of options valuation. *Journal of Financial Economics* 56, 407-458.
- Chiras, D., and S. Manaster (1978). The information content of option prices and a test of market efficiency. *Journal of Financial Economics* 6, 213-234.
- Chou, R. (2005). Forecasting financial volatilities with extreme values: the conditional autoregressive range (CARR) model. *Journal of Money, Credit and Banking* 37, 561-582.
- Chou, R. Y., and N. Liu (2010). The economic value of volatility timing using a range-based volatility model. *Journal of Economic Dynamics and Control* 34, 2288-2301.
- Christensen, B. J., and N. R. Prabhala (1998). The relation between implied and realized volatility. *Journal of Financial Economics* 50, 125-150.
- Clark, P. K. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica* 41, 135-155.
- Corrado, C. J., and T. W. Miller (2005). The forecast quality of CBOE implied volatility indexes. *Journal of Futures Markets* 25, 339-373.
- Corrado, C. J., and T. W. Miller (2006). Estimating expected excess returns using historical and option-implied volatility. *Journal of Financial Research* 29, 95-112.
- Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics* 7, 174-196.
- Coval, J., and T. Shumway (2001). Expected option returns. *Journal of Finance* 56, 983-1009.
- Covrig, V., and B. S. Low (2003). The quality of volatility traded on the over-the-counter currency market: A multiple horizons study. *Journal of Futures Markets* 23, 261-285.
- Cox, J. R., and M. Rubinstein (1985). *Options Markets*. Prentice-Hall, NJ.
- Dacorogna, M. M., Müller, U. A., Nagler, R. J., Olsen, R. B., and O. V. Pictet (2001). An introduction to high-frequency finance. Academic Press, San Diego.
- Day, T. E. and C. M. Lewis (1992). Stock market volatility and the information content of stock index options. *Journal of Econometrics* 52, 267-287.
- Day, T. E., and C. M. Lewis (1993). Forecasting futures market volatility. *Journal of Derivatives* 1, 33-50.
- Demeterfi, K., Derman, E., Kamal, M., and J. Zou (1999). More than you ever wanted to know about volatility swaps. *Journal of Derivatives* 6, 9-32.
- Derman, E., and I. Kani (1994). Riding on a smile. *Risk* 7, 32-39.

- Diebold, F. X. (2004). The Nobel memorial prize for Robert F. Engle. *Scandinavian Journal of Economics* 106, 165-185.
- Diebold, F. X., and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics* 13, 253-265.
- Ding, Z., Granger, C. W. J., and R. F. Engle (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance* 1, 83-106.
- Doran, J., and E. Ronn (2008). Computing the market price of volatility risk in the energy commodity markets. *Journal of Banking and Finance*, 2541-2552.
- Dupire, B. (1994). Pricing with a smile. *Risk* 7, 18-20.
- Ederington, H. E., and W. Guan (2002). Measuring implied volatility: is an average better? Which average? *Journal of Futures Markets* 22, 811-837.
- Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of U. K. inflation. *Econometrica* 50, 987-1008.
- Engle, R. F., Lilien, D. M., and R. P. Robbins (1987). Estimating time varying risk premia in the term structure: The ARCH-M model. *Econometrica* 55, 391-407.
- Engle, R. F., and G. M. Gallo (2006). A multiple indicators for volatility using intra-daily data. *Journal of Econometrics* 131, 3-27.
- Engle, R. F., and A. J. Patton (2001). What good is a volatility model? *Quantitative Finance* 1, 237-245.
- Eraker, B. (2004). Do stock prices and volatility jump? Reconciling evidence from spot and option prices. *Journal of Finance* 59, 680-692.
- Figlewski, S. (1997). Forecasting volatility. *Financial Markets, Institutions and Instruments* 6, 1-88.
- Fleming, J. (1998). The quality of market volatility forecasts implied by S&P 100 index option prices. *Journal of Empirical Finance* 5, 317-345.
- Fleming, J., Ostdiek, B., and R. E. Whaley (1996). Trading costs and the relative rates of price discovery in stock futures, and option markets. *Journal of Futures Markets* 16, 353-387.
- Fornari, F. (2010). Assessing the compensation for volatility risk implicit in interest rate derivatives. *Journal of Empirical Finance* 17, 722-743.
- French, K. R., Schwert, G. W., and R. F. Stambaugh (1987). Expected stock returns and volatility. *Journal of Financial Economics* 19, 3-29.
- Gallant, A. R., Hsu, C. T., and G. E. Tauchen (1999). Using daily range data to calibrate volatility diffusions and extract the forward integrated variance. *Review of Economics and Statistics* 81, 617-631.
- Gallant, A. R., Rossi, P. E., and G. E. Tauchen (1992). Stock prices and volume. *Review of*

Financial Studies 5, 199-242.

- Garman, M. B. and M. J. Klass (1980). On the estimation of security price volatilities from historical data. *Journal of Business* 53, 67-78.
- Ghysels, E., Harver, A. C., and E. Renault (1996). Stochastic volatility. In *Statistical Methods in Finance*, C. R. Rao and G. S. Maddala (eds.), North-Holland, 119-191.
- Giot, P., and S. Laurent (2007). The information content of implied volatility in light of the jump/continuous decomposition of realized volatility. *Journal of Futures Markets* 27, 337-359.
- Glosten, L. R., Jagannathan, R., and D. Runkle (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* 48, 1779-1801.
- Granger, C. W. J., and R. Joyeux (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis* 1, 15-29.
- Hansen, P. R., and A. Lunde (2006). Realized variance and market microstructure noise. *Journal of Business and Economic Statistics* 24, 127-161.
- Harris, L. (1987). Transaction data tests of the mixture of distributions hypothesis. *Journal of Financial and Quantitative Analysis* 22, 127-141.
- He, A., Kwok, J., and A. Wan (2010). An empirical model of daily highs and lows of West Texas Intermediate crude oil prices. *Energy Economics* 32, 1499-1506.
- Hosking, J. R. (1981). Fractional differencing. *Biometrika* 68, 165-176.
- Hull, J., and A. White (1987). The pricing of options on assets with stochastic volatilities. *Journal of Finance* 42, 281-300.
- Jackwerth, J. C., and M. Rubinstein (1996). Recovering probability distributions from option prices. *Journal of Finance* 51, 1611-1631.
- Jarque, C. M., and A. K. Bera (1980). Efficient tests for normality, heteroskedasticity, and serial independence of regression residuals. *Economic Letters* 6, 255-259.
- Jiang, G. J., and Y. S. Tian (2005). Model-free implied volatility and its information content. *Review of Financial Studies* 18, 1305-1342.
- Jorion, P. (1995). Predicting volatility in the foreign exchange market. *Journal of Finance* 50, 507-528.
- Kang, S. H., Kang S. M., and S. M. Yoon (2009). Forecasting volatility of crude oil markets. *Energy Economics* 30, 621-641.
- Kroner, K. F., Kneafsey, K. P., and S. Claessens (1995). Forecasting volatility in commodity markets. *Journal of Forecasting* 14, 77-95.
- Kunitomo, N. (1992). Improving the Parkinson method of estimating security price volatilities. *Journal of Business* 65, 292-302.

- Lamoureux, C., and W. Lastrapes (1993). Forecasting stock-return variance: toward an understanding of stochastic implied volatilities. *Review of Financial Studies* 6, 293-326.
- Latane, H., and R. J. Rendleman (1976). Standard deviations of stock price ratios implied in option prices. *Journal of Finance* 31, 369-381.
- Ljung, G. M., and G. E. P. Box (1978). On a measure of lack of fit in time series models. *Biometrika* 65, 297-303.
- Martens, M., and D. van Dijk (2007). Measuring volatility with the realized range. *Journal of Econometrics* 138, 181-207.
- Meddahi, N. (2002). A theoretical comparison between integrated and realized volatility. *Journal of Applied Econometrics* 17, 479-518.
- Mincer, J., and V. Zarnowitz (1969). The evaluation of economic forecasts. In Mincer, J. (ed.), *Economic Forecasts and Expectations*. New York: National Bureau of Economic Research.
- Mohammadi, H., and L. Su (2010). International evidence on crude oil price dynamics: applications of ARIMA-GARCH models. *Energy Economics* 32, 1001-1008.
- Morau, F., Navatte, P. and Villa, C. (1999). The predictive power of the french market volatility index: a multi horizons study. *European Finance Review* 2, 303-320.
- Müller, U. A., Dacorogna, M., Davé, R., Olsen, R. B., Pictet, O. V., and J. von Weizsäcker (1997). Volatilities of different time resolutions - Analysing the dynamics of market components. *Journal of Empirical Finance* 4, 213-239.
- Nelson, D. B. (1990). ARCH models as diffusion approximations. *Journal of Econometrics* 45, 7-38.
- Newey, W. K., and K. D. West (1987). A simple, positive-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703-708.
- Pan, J. (2002). The jump-risk premia implicit in options: evidence from an integrated time-series study. *Journal of Financial Economics* 63, 3-50.
- Parkinson, M. (1980). The extreme value method for estimating the variance of the rate of return. *Journal of Business* 53, 61-65.
- Pong, S. E., Shackleton, M. B., Taylor, S. J., and X. Xu (2004). Forecasting currency volatility: a comparison of implied volatilities and AR(FI)MA models. *Journal of Banking and Finance* 28, 2541-2563.
- Poon, S.-H., and C. W. J. Granger (2003). Forecasting financial market volatility: a review, *Journal of Economic Literature* 41, 478-539.
- Poteshman, A.M. (2000) Forecasting future volatility from option prices, Working paper, University of Illinois at Urbana-Champaign.
- Rogers, L., and S. Satchell (1991). Estimating variance from high, low, and closing prices,

- Annals of Applied Probability 1, 504-512.
- Rosenberg, J. V., and R. F. Engle (2002). Empirical pricing kernels. *Journal of Financial Economics* 64, 341-72.
- Rubinstein, M. (1994). Implied binomial trees. *Journal of Finance* 49, 771-818.
- Rubinstein, M. (1998). Edgeworth binomial trees. *Journal of Derivatives* 5, 20-27.
- Sadorsky, P. (2006). Modeling and forecasting petroleum futures volatility. *Energy Economics* 28, 467-488.
- Schwert, G. W. (1990). Stock volatility and the crash of '87. *Review of Financial Studies* 3, 77-102.
- Schmalensee, R., and R. R. Trippi (1978). Common stock volatility expectations implied by option premia. *Journal of Finance* 33, 129-147.
- Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility models. In Cox, D. R., Hinkley, D. V., Barndorff-Nielsen, O. E. (Eds.), *Time Series Models in Econometrics, Finance and Other Fields*. Chapman and Hall, London, pp. 1-67.
- Shu, J. H., and J. E. Zhang (2006). Testing range estimators of historical volatility. *Journal of Futures Markets* 26, 297-313.
- Simon, D. P. (2003). The Nasdaq Volatility Index During and After the Bubble. *Journal of Derivatives* 11, 9-22.
- Szakmary, A., Ors, E., Kim, J. K., and W. D. Davidson III (2003). The predictive power of implied volatility: Evidence from 35 futures markets. *Journal of Banking and Finance* 27, 2151-2175.
- Taylor, S. J., and X. Xu (1997). The incremental volatility information in one million foreign exchange quotations. *Journal of Empirical Finance* 4, 317-340.
- Thomakos, D. D., and T. Wang (2003). Realized volatility in the futures market. *Journal of Empirical Finance* 10, 321-353.
- Todorov, V. (2010). Variance risk premia dynamics: the role of jumps. *Review of Financial Studies* 23, 345-383.
- Trolle, A., and E. Schwartz (2010). Variance risk premia in energy commodities. *Journal of Derivatives* 18, 1-18.
- Vipul, and J. Jacob (2007). Forecasting performance of extreme-value estimators. *Journal of Futures Markets* 27, 1085-1105.
- Wei, Y., Wang, Y., and D. Huang (2010). Forecasting crude oil market volatility: Further evidence using GARCH class models. *Energy Economics* 32, 1477-1484.
- Yang D., and Q. Zhang (2000). Drift-independent volatility estimation based on high, low, open, and close prices. *Journal of Business* 73, 477-491.