ESSAYS ON FORECASTING AND HEDGING MODELS IN THE OIL MARKET AND CAUSALITY ANAYSIS IN THE KOREAN STOCK MARKET

A Dissertation

by

HANKYEUNG CHOI

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2012

Major Subject: Agricultural Economics

Essays on Forecasting and Hedging Models in the Oil Market and Causality Analysis in the Korean Stock Market

ESSAYS ON FORECASTING AND HEDGING MODELS IN THE OIL MARKET AND CAUSALITY ANAYSIS IN THE KOREAN STOCK MARKET

A Dissertation

by

HANKYEUNG CHOI

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	David J. Leatham
Committee Members,	Henry Bryant
	Ximing Wu
	Samiran Sinha
Head of Department,	John P. Nichols

August 2012

Major Subject: Agricultural Economics

ABSTRACT

Essays on Forecasting and Hedging Models in the Oil Market and Causality Analysis in the Korean Stock Market. (August 2012)

Hankyeung Choi, B.A., Yonsei University;

M.E., Korean Advanced Institute of Science and Technology

Chair of Advisory Committee: Dr. David J. Leatham

In this dissertation, three related issues concerning empirical time series models for energy financial markets and the stock market were investigated. The purpose of this dissertation was to analyze the interdependence of price movements, focusing on the forecasting models for crude oil prices and the hedging models for gasoline prices, and to study the change in the contemporaneous causal relationship between investors' activities and stock price movements in the Korean stock market.

In the first essay, the nature of forecasting crude oil prices based on financial data for the oil and oil product market is examined. As crack spread and oil-related Exchange-Traded Funds (ETFs) have enabled more consumers and investors to gain access to the crude oil and petroleum products markets, I investigated whether crack spread and oil ETFs were good predictors of oil prices and attempted to determine whether crack spread or oil ETFs were better at explaining oil price movements.

In the second essay, the effectiveness of diverse hedging models for the unleaded gasoline price is examined using futures and ETFs. I calculated the optimal hedge ratios

for gasoline futures and gasoline ETF utilizing several advanced econometric models and then compared their hedging performances.

In the third essay, the contemporaneous causal relationship between multiple players' activities and stock price movements in the Korean stock market was investigated using the framework of a DAG model. The causal impacts of three players' activities in regard to stock return and stock price volatility are examined, concentrating on foreign investor activities. Within this framework, two Korean stock markets, the KSE and KOSDAQ markets, are analyzed and compared. Recognizing the global financial crisis of 2008, the change in casual relationships was examined in terms of pre- and post-break periods.

In conclusion, when a multivariate econometric model is developed for multi-markets and multi-players, it is necessary to consider a number of attributes on data relations, including cointegration, causal relationship, time-varying correlation and variance, and multivariate non-normality. This dissertation employs several econometric models to specify these characteristics. This approach will be useful in further studies of the information transmission mechanism among multi-markets or multi-players.

DEDICATION

To my beloved family

ACKNOWLEDGEMENTS

I would like to express special thanks to the inspirational instruction and insightful guidance of Dr. Leatham. Without his continued support and firm belief on me, I could not have finished my academic process at Texas A&M University. From beginning to end in my life in the Department of Agricultural Economics, he has been a great teacher and special advisor to me. Thanks also go to my committee members, Dr. Bryant, Dr. Wu, and Dr. Sinha for their guidance and help throughout the course of this research. I am also obligated to Dr. Siebert and Dr. Power for their encouragement to my study. I owe much to my colleagues Sungwook Hong, Chanhee Rhew, Beomsu Park, Sungju Cho, Hyoungil Lee, Hojung Yoon, Wujin Choi, and Hyunsoo Kim.

The encouragement of my parents Jangjong Choi and Jongrae Lee towards my academic achievements was a motivating force whenever I faced difficulties. My parents-in-law Chokwang Cho and Bangja Park also cared much about me and my family. I would like to reserve special thanks for my wife, Eunhee Cho. She is all my love in my life. Lastly thanks to my two lovely sons, Minsik and Minjae, for revitalizing my life every day.

TABLE OF CONTENTS

	Page
ABSTRACT	iii
DEDICATION	v
ACKNOWLEDGEMENTS	vi
TABLE OF CONTENTS	vii
LIST OF FIGURES	ix
LIST OF TABLES	x
CHAPTER	
I INTRODUCTION	1
II OIL PRICE FORECASTING USING CRACK SPREAD FUTURES AND OIL EXCHANGE-TRADED FUNDS	5
Introduction	5
Theoretical Background	12
Methodology	16
Data	24
Empirical Results and Discussion	26
Summary and Concluding Remarks	38
III OPTIMAL DYNAMIC HEDGING OF UNLEADED GASOLINE USING FUTURES AND EXCHANGE-TRADED FUNDS	40
Introduction	40
Theoretical Background on Copula Methodology	45
Estimation Methodology for Optimal Hedge Ratios	53
Data and Empirical Results	62
Summary and Concluding Remarks	73

CHAPTER Page

IV CAUSALITY ANALYSIS ON STOCK PRICES AND STOCK MARKET PARTICIPANTS IN CASE OF THE KOREAN STOCK MARKET 75

Introduction and Background 75
Literature Review 78
Data and Methodology 83
Empirical Results and Discussion 89
Summary and Concluding Remarks 92

V CONCLUSION 95

REFERENCES 98

APPENDIX A 104

APPENDIX B 120

VITA 144

LIST OF FIGURES

FIGURE		Page
A-1	Trade Volumes of Three Oil ETFs from 2009 to 2011	104
A-2	Price Movements on Crude Oil and Crack Spread	105
A-3	Squared Returns of OPEC Crude Oil and a 2-1-1 Crack Spread	106
A-4	Unconditional Correlations and Estimated Conditional Correlations	107
A-5	Log Prices and First Differences of Log Prices	108
A-6	Impulse Response to One Standard Error in ECM Models	109
A-7	Residuals for ECM and Standardized Residual for ECM-MGARCH	110
A-8	Conditional Variances of Oil, Crack Spread, ETF Spread	111
A-9	Price Level Changes of Gasoline Spot, Futures and ETF Prices	112
A-10	Estimated Conditional Correlations in DCC MGARCH	113
A-11	Estimated Conditional Variance in DCC MGARCH	114
A-12	Hedge Ratios in case of Futures Hedging	115
A-13	Hedge Ratios in case of ETF Hedging	116
A-14	Estimated Time-varying Dependences of Lower and Upper Tails by Dynamic SJC Copula Function in Case of Futures Hedging	117
A-15	KOSPI and KOSDAQ Indices from 2005 to 2010	118
A-16	Contemporaneous Causal Relationships in the Korean Stock Market	119

LIST OF TABLES

TABLE		Page
B-1	Descriptive Statistics	120
B-2	Unit Root Test	121
B-3	Johansen Cointegration Maximum Likelihood Test	122
B-4	Wald Test-Granger Causality	123
B-5	Error Correction Models with Structural Break	124
B-6	Estimation Results of Multivariate GARCH Models	125
B-7	Forecast Error Statistics for Crude Oil Forecasting	126
B-8	Model Framework	127
B-9	Summary Statistic and Correlation	128
B-10	Unit Root Test	129
B-11	EG-ADF Test and Johansen ML Test for Cointegration	130
B-12	Wald Test-Granger Causality	131
B-13	Estimation of VEC and DCC MGARCH Models in Futures Hedging	132
B-14	Estimation of VEC and DCC MGARCH Models in ETF Hedging	133
B-15	Comparisons of Copula Functions in Case of Futures Hedging	134
B-16	Comparisons of Copula Functions in Case of ETF Hedging	135
B-17	Estimation of Copula-GARCH Model in the Futures Hedging	136

TABLE		Page
B-18	Estimation of Copula-GARCH Model in the ETF Hedging	137
B-19	Comparison of Hedge Performance of Futures and ETF Hedging	138
B-20	Ceiling of Foreign Ownership in the Korean Stock Exchange	139
B-21	Proportion of Trading Volume by Investor Type in the KSE Market	140
B-22	Results of Augmented Dickey-Fuller Test	141
B-23	DAG Result on Foreign Investor's Activities to Stock Returns and Volatility	142
B-24	Information Flow among Three Types of Investors	143

CHAPTER I

INTRODUCTION

Modeling and analyzing the price relationships in the oil markets reflecting the increase in information transmission between the spot and derivatives markets is currently one of the most interesting subjects in empirical economics research. An understanding of the individual price movements of vertical chains in the oil market is important in its own right. Especially, when these markets experience an increase of interaction and interdependence between the markets and a high volatility in price movement, this will drive the demand for more accurate forecasting and more elaborate hedging models. For this reason, diverse financial instruments are utilized, including futures contracts, options, and additional new financial tools like Exchange-Traded Funds (ETFs). Especially, the recent advent of oil ETF market enables consumers and investors to access crude oil and petroleum products in diverse ways, since the hedging and investment effects of these funds are very similar to those of futures contracts. Therefore, for traders seeking the cheapest means to reduce the uncertainty of their market exposures, in this study, crack spread futures and ETF spread were evaluated as predictors of crude oil prices, and gasoline futures and its corresponding ETFs were assessed as hedging tools.

This dissertation follows the style of *American Journal of Agricultural Economics*.

In addition, using the Korean stock market data, the empirical causal relationships between stock price movement and the activities of three types of investors are investigated using the Directed Acyclical Graph (DAG). This study divides investors or players in the Korean stock market into three types: foreign investors, domestic institutional investors, and domestic individual investors. It is of interest to determine which of the trading activities associated with these three groups are highly related with stock price movements and the manner in which traders' activities affect each other, taking into consideration of the recent fluctuation in the stock market during the 2008 global financial crisis.

The purpose of this dissertation is to analyze the interdependence of price movements by developing a forecasting model for crude oil prices using diverse oil financial derivatives and a hedging model for gasoline spot prices, and to study the dynamic changes in the causal relationship between investors' activities and stock price movements, in the case of Korean stock market from 2005 to 2010. In the first essay, the nature of forecasting crude oil prices based on financial data for oil and the oil products market is examined. Traditionally, petroleum refiners have used crack spread futures as an effective risk management tool and a good indicator of oil market prices, because the crack spread intrinsically represents one of a refiner's goals, that of protecting the margin between the crude oil and oil product. Recently, the advent of diverse oil-related ETFs has enabled more consumers and investors to gain access to the crude oil and petroleum product market. In this research, I investigated whether crack spread and oil ETFs were good predictors of oil prices and attempted to determine whether crack spread or oil ETFs were better at explaining oil price movements. Based on the Error Correction Model (ECM) and ECM Multivariate Generalized Autoregressive Conditional Heteroskedasticity Model (GARCH), I examined the causal relationships between crude oil and both crack spread and oil ETFs and the forecasting abilities of these two tools.

The purpose of the second essay was to examine the effectiveness of a dynamic hedging model for unleaded gasoline spot price, using gasoline futures and gasoline exchange-traded funds (ETFs). I calculated the most efficient hedge ratio for gasoline futures and gasoline ETF utilizing several advanced econometric models and then compared their hedging performance. As the relationship between spot and futures data for gasoline is cointegrated, the basic static hedge model was based on the Vector Error Correction (VEC) model. The Dynamic Conditional Correlation Multivariate GARCH (DCC MGARCH) model, in which time-varying dependence allows the use of the conditional covariance to capture the updated information, derives the dynamic hedge ratio. Compared to these symmetric and multivariate normal distribution-based VEC and DCC MGARCH models, the copula function allows the development of an asymmetric and non-normal multivariate distribution-based model. Both static and dynamic copula-based GARCH models were exploited to analyze time-varying optimal hedge ratios in the case of gasoline.

While the previous two essays focused on multi-markets analyses, such as an examination of crude oil and its products markets or spot, futures and ETF markets of unleaded gasoline, in the third essay, the causal relationship between multiple players in a single market is investigated within the framework of the DAG model. In this research, the contemporaneous causal impacts of three market players' activities on stock returns and stock price volatilities were investigated. The question I sought to answer in this essay was how did the dynamics of causal relationships between stock price movement and three investors' activities change during the period from 2005 to 2010. The fifth, and final, chapter of this dissertation provides the summary of this research as well as discussing its implications.

CHAPTER II

OIL PRICE FORCASTING USING CRACK SPREAD FUTURES AND OIL EXCHANGE-TRADED FUNDS

Introduction

Investors, who are interested in entering the oil market, resort to many diverse traditional types of trading in the market, such as purchasing the stock of oil firms, investing in oil-related mutual funds, and trading on the commodity futures market for crude oil or petroleum products. However, the stock price for each oil company reflects not only prospective oil prices, but also the individual company's diverse issues. For example, investing in mutual funds is very restrictive in terms of liquidity. The futures market usually requires all participants to open a futures account with margin requirements, which entails relatively high transaction costs, in order to maintain the timely implementation of transactions in the market. All of these conditions present a strong entry barrier to private investors. Thus, most of the transactions in the futures market are completed by refinery companies or institutional investors.

Especially, from the perspective of the refiners, there is more interest in crack spread futures rather than individual commodity futures for crude oil, heating oil, and unleaded gasoline. This is because crack spread entails the simultaneous purchase or sale of crude futures against the sale or purchase of refined petroleum products futures. Simply put, the difference in price between crude oil and its derived products is called “crack spread”. Therefore, refineries are naturally more concerned about the difference between input and output prices, rather than solely the price of crude oil, since the profits of refineries are tied directly to the crack spread. Crack spread derivatives aid market participants to better manage the inherent price risks of the energy market. In addition, crack spread has been reported to be a good predictor of spot oil prices, because the refineries are major participants in oil markets, and they are primarily concerned with crack spread (Murat and Tokat 2009).

However, the recent advent of energy Exchange-Traded Funds (ETFs) has enabled diverse investors to directly access the energy market with strong liquidity and without high entry costs, as with futures accounts. Originally, an ETF was an investment fund traded on the stock exchange, much like trading in shares, however, commodity ETFs invest in commodities, such as precious metals, oil, and agricultural products. The number of ETFs is still increasing, since they are attractive to investors because of their low costs, tax efficiency, and stock-like features. After the first oil ETF, USO, was introduced in April 2006, various other oil-related ETFs have been introduced in the market[1], reflecting the high volatility and price level hikes in the oil market. In addition, the use of combinations of individual oil-related EFTs[2] also enables the cracking margin for refineries with traditional crack spreads to be locked in; therefore, the cracking margin is no longer the exclusive property of refiners. After the last ETF (SCO) utilizing

[1] United States Gasoline Fund (NYSE, UGA)**,** United States Heating Oil Fund (NYSE, UHN)**,** and ProShares UltraShort DJ-UBS Crude Oil ETF (NYSE, SCO) were launched Feb. 26, 2008, Apr. 29, 2008, and Nov. 24, 2008, respectively.

[2] Buying equal lots of the SCO, the UHN and the UGA effectively puts an investor short two units of crude oil and long one unit each of gasoline and heating oil. In other words, this three-legged purchase simulates buying a 2:1:1 crack spread. (Hereafter, an ETF spread will indicate an ETF version of crack spread with combined trades of SCO, UGA and UHN).

this combination was introduced on November 2008, a substantial trade volume for SCO was recorded at the beginning of Jan. 2009[3]. Figure A-1 shows that the trade volumes of all three ETFs have fluctuated, with SCO having a trend of increasing volume during the period from January 2009 to December 2011. Considering the convenient trading system for ETFs and increased interest by investors in ETFs, oil ETFs are becoming one of the most important factors in understanding the spot and futures oil markets.

Recently, this increased interest in the energy market has resulted in a high volume of ETF and crack spread future trading and a high volatility in the price levels and returns for these markets. Considering the dynamic changes in the oil market, I further examined the interaction between the oil market and related finance markets, such as in crack spreads and oil ETFs. The questions to be posed are whether these ETFs and crack spread futures will be good predictors of spot price movement for crude oil and whether crack spread and oil ETF spreads are better at explaining oil price movement focusing on the recent fluctuations in the oil market. Therefore, the data for this research examining the effect on the market of holding ETFs was collected for the period of 2009-2011. To further study the interaction and interdependence between the oil spot market and oil-related financial markets, the spot price of crude oil, the crack spread, and the ETF spread are explained sequentially in more detail in the following sections.

---

[3] SCO was introduced on Nov. 14, 2008, and the average trade volume of SCO in 2008 was 18,339. However, trade volume has increased to more than 200,000 since Jan. 2009. Therefore, this research deals with ETF trade data from Jan. 2009 to Dec. 2011.

Crude Oil

In 2008, the US economy experienced a serious financial crisis as a result of the mortgage market collapse and market unrest, and this shock spread globally as an economic depression. Due to the strong dollar and reduced consumption, most of the resource markets also experienced a plunge in prices. The spot price of crude oil also dropped from $140 per barrel (June 2008) to less than $40 per barrel (February 2009). With the recovery of the world economy since 2009, the spot prices of crude oil have shown an increasing trend. In Figure 2-2, the spot prices of West Texas Intermediate (WTI) crude oil and Organization of the Petroleum Exporting Countries (OPEC) Reference Basket Price (ORB)[4] are compared. As lighter and sweet crude oil, like WTI's, usually yields more gasoline than heavier crude oil like that of ORB, WTI’s spot price is higher than OPEC’s. Especially, in 2011, a series of unsettled political situations in the Middle East and North African area, such as the revolution in Tunisia in 2010 and domestic turmoil in Egypt and Libya in 2011, increased the uncertainty of the supply-side of crude oil in some OPEC-member countries, while the US crude oil inventory of WTI was stably maintained at a sufficient level. This difference in certainty on the supply side between the two main crude oil producers is the main reason for the price reverse phenomenon in 2011. In addition, structural breaks in oil prices were observed around the time of the 2008 financial crisis. These structural breaks will be tested in

[4] The new OPEC Reference Basket (ORB) introduced on June 16, 2005, is currently made up of the crude oil from OPEC members, such as Saharan Blend (Algeria), Girassol (Angola), Oriente (Ecuador), Iran Heavy (Islamic Republic of Iran), Basra Light (Iraq), Kuwait Export (Kuwait), Es Sider (Libya), Bonny Light (Nigeria), Qatar Marine (Qatar), Arab Light (Saudi Arabia), Murban (UAE) and Merey (Venezuela). OPEC collects price data on this “basket” of crude oils, and uses average prices of these oils to develop an OPEC reference price. The ORB price is considered as representative of heavier oil as compared to light oil like WTI and Brent (http://www.opec.org/opec_web/en/data_graphs/40.htm).

more detail to determine the change in causal relationship among oil price, crack spread, and ETF spread.

Crack Spread

The most popular crack spread contract is the 3-2-1 crack spread, which is computed from the daily futures prices of crude oil, heating oil and unleaded gasoline of the same term structures, and involves three contracts of crude oil, two contracts of unleaded gasoline, and one contract of heating oil. Essentially, traders buying or selling 3-2-1 crack spreads take advantage of 75% margin credits, which is very attractive to traders in the futures market. The refining of lighter, sweeter crudes, such as those produced by WTI, are best represented by the 3-2-1 spread. Alternative ratios, such as 2-1-1 and 5-3-2, may also be utilized for crack spread margins. Especially, the 2-1-1 crack spread, signifying that two barrels of crude yield a barrel each of gasoline and heating oil, is a better description of the case of heavy crude oils like OPEC basket grades[5], because heavy crudes do not yield as much gasoline as light crude.

In Figure A-2, the 3-2-1 and 2-1-1 crack spreads are compared during January 2005 to November 2011. Although crude yields varied depending upon the refining model employed, the two crack spreads models show similar movement patterns and had a very strong correlation of 0.98. One of their interesting features is the fact that 3-2-1 crack spreads are more volatile than 2-1-1 ones before the summer driving season,

[5] Crude oil is traded on a barrel basis, while heating oil and gasoline are traded on a gallon basis. The 2-1-1 crack spread is calculated by the formula; 2-1-1 Crack spread (\$/barrel) = Gasoline price (\$/gal) × 21 + Heating oil price (\$/gal) × 21 – Crude oil price (\$/bbl). The 3-1-1 crack spread (\$/bbl) is derived by the following formula; 3-1-1 Crack spread (\$/bbl) = Gasoline price (\$/gal) × 28 + Heating oil price (\$/gal) × 14 – Crude oil price (\$/bbl).

because the 3-2-1 spread, double-weighted in gasoline, tends to outperform the 2-1-1 spread when gasoline prices rise in relation to heating oil. On the contrary, the 2-1-1 crack spread commands a premium over the 3-2-1 spread typically in the fall and winter, as the demand for heating oil increases.

Oil ETF Spread

Instead of trading crack spreads in the futures market, a trader can capture spread change without resorting to the futures market with the advent of diverse ETFs and combinations of three specific ETFs (SCO, UHN, and UGA). ETF spread[6] allows investors to trade the spread margin free in at least the 2-1-1 version of crack spread, in which two contracts of crude oil are converted into one contract of heating oil and one contract of unleaded gasoline. Investors without a futures account can use ETFs in their portfolio to trade margin variances. While futures are traded with significant margin requirements for entering this market, there are few restrictions in this ETFs market. Contracts don't have to be rolled over, there's no contango or backwardation to deal with and investors don't have to worry about changes in margin requirements. Even the small private investor can be a hydrocarbon cracker using ETF spreads. Traditional mutual

[6] The concept of ETF crack spread was suggested in 2009 by Brad Zigler, who is the managing editor of Hard Assets Investor and the alternative investments editor of Registered Rep. magazine. Conceptually, he combined three ETFs; UGA for gasoline prices, UHN for heating oil prices, and SCO for crude oil. While UGA and UHN track the near month future price of gasoline and heating oil, respectively, SCO corresponds to twice the inverse of crude oil prices. The simultaneous purchase of these three ETFs is conceptually similar to buying a 2-1-1 crack spread, which indicates the trade of selling a crude oil contract and purchasing heating oil and gasoline contracts simultaneously. He proposed this ETF spread as a substitute investment vehicle for crack spreads. For a consistent value of ETF spread, the SCO index is modified to reflect the ratio change from the reverse stock split of SCO in Feb. 2011. (http://www.hardassetsinvestor.com/interviews/1450-accounting-for-crack-spread-differences.html?showall=&start=1)

funds for energy commodities, which were heavily invested in by private investors before the advent of ETFs, are usually contract based with restrictions for not selling within a certain period of time. An ETF investor, however, can easily buy and sell this security without any time limits. There are also a lot of hedging and investing tools available to refiners in the oil market.

The price of crude oil is basically established by supply and demand conditions in the global market overall, and more particularly, in the main refining centers of the US Gulf Coast, Northwest Europe, and Singapore. Demand for petroleum products by consumers, as well as for agricultural, manufacturing, household heating, and transportation uses, determines the demand for crude oil by refiners. Product demand is also linked to economic conditions and may also be influenced by other factors, like weather conditions. Therefore, there are many ways to forecast demand for crude oil. However, this research focuses on the financial instruments crack spread and ETF spread, which have recently received intense scrutiny from refiners and oil market investors as hedging and investment tools. In addition, both crack spread and ETF spread may reflect the daily change in spot price of crude oil effectively and comprehensively, compared to other less flexible factors such as oil stock changes, capacity change, and capacity utilization.

While the spot price of OPEC crude oils and 2-1-1 crack spread data can be compared for the periods covering 2005 to 2011, ETF spread data only covers the period from 2009 to 2011, considering that the launch of SCO was in November 2008. In previous research, Murat and Tokat (2009) showed that a 3-2-1 crack spread can be a good predictor of WTI based on weekly data from January 2001 to February 2008. However, the current research investigates whether a 2-1-1 crack spread and ETF spread, utilized as daily financial tools of the oil market, are good predictors of crude oil prices, in the case of OPEC Reference Basket Price (ORB).

Theoretical Background

In order to investigate the relationships among spot price (crude oil), commodity futures (crack spread), and commodity ETFs (ETF spread), we selected two cases, in particular, to examine in further detail; crude oil and crack spread, and crack spread and ETF spread.

Relationship between Crude Oil and Crack Spread

The relationship between the spot price and commodity futures is generally defined as convenience yield (Heinkel, Howe, and Hughes 2006), which is defined as the benefit of owing a particular good physically rather than owning a futures contract for that good (Working 1949; Brennan and Schwartz 1985). These researchers proposed that there is a positive relationship between marginal production costs and convenience yield, because if the marginal cost of production is relatively low, unexpected demands arising in the market can be met by immediate production of the commodity. Conversely, if marginal production costs are relatively expensive, increased demand can only be compensated by taking from inventory. Edwards and Ma (1992) found that variation in profit margins can be attributed mainly to a change in cost factors. The fluctuations in crack spread can be explained as changes in the cost of production of heating oil and gasoline and are used to obtain information about marginal production costs (Murat and Tokat 2009). Kocagil (2004) demonstrated a positive relationship between convenience yield and production cost by examining the convenience yield behavior for crack spread futures. In addition, (Anon.)Zigler (2009) suggested that crude oil tends to re-price more quickly than its products, so crack spreads tend to widen or narrow when crude oil prices move precipitously. He noted that there is a negative correlation between crude oil price and crack spread, especially when oil prices increase and decrease suddenly, because there is a difference in adjustment rates for prices of crude oil and its products. Therefore, obtaining data from the energy financial market, which is more responsive than the commodity market, is a good source of getting updated information on crude oil for analysis.

Relationship between Crack Spread and Oil ETF Spread

Crack spread is calculated from the individual futures prices of crude oil, heating oil, and gasoline. Both heating oil and gasoline must use the same contract month as crude oil, because these two petroleum products contracts follow the same contract. Therefore, front month data from the futures market are usually used for calculation of crack spread. As is the case in many commodity ETFs, the so-called front month futures contracts are simply rolled from month to month. In the case of UGA, the trust invests in futures contracts on unleaded gasoline delivered to the New York harbor and traded on the NYMEX that is near the month contract to expire. The UHN fund basically seeks to track the movement of heating oil prices; this fund consists of listed heating oil futures contracts, and other heating oil-related futures, forwards, and swaps contracts. Unlike UGA and UHN, SCO is designed to track the daily performance that corresponds to twice (200%) the opposite of the performance of the Dow Jones-AIG Crude Oil Sub-index, which is intended to reflect the performance of crude oil as measured by the price of future contracts of crude oil traded on the NYMEX. This fund invests in any one of or combination of futures contracts, forward contracts, swap contracts, and option contracts.

As crack spread is based on the date of the front month for crude oil futures, the base prices of crude oil utilized in calculation of crack spread and ETF spread have very similar data sources and a strong correlation. The correlation level of data for gasoline futures and UGA is 0.9746 and the correlation of daily returns is 0.9098. In the case of heating oil futures and UHN, the correlation level and daily return are 0.9707 and 0.9043, respectively. Due to the inverse relationship between crude oil and SCO, the correlation level is -0.9873, but the correlation of daily returns is 0.9043. However, as there are some differences between gasoline futures and UGA, in terms of trading markets, management cost, the main investor types, and so forth, daily returns and the volatility of these two financial tools are not the same. To obtain crack spread futures prices, most researchers use the daily settlement prices for all NYMEX-traded futures contracts with front months on crude oil, heating oil, and unleaded gasoline. Therefore, the positive relationship between crack spread and ETF spread can easily be expected and the ETF spread is observed to track the crack spread fairly well, as shown in Figure A-2, during the period from January 2009 to November 2011, although it is not exactly the same. In addition, both crack spread and ETF spread serve the similar functions of diversification and as hedging instruments in the market. However, there are some factors that differ between crack spread and ETF spread. For one thing, as the entry barrier in the ETF market is not as strict as in the futures market, more diverse types of investors may enter the ETF market. In the futures market, the data is based on settlement values, while the individual ETF index typically reflects the last sale data available to retail investors. In addition, no contango is reflected in the ETF market, as the product ETFs are designed to be continuously invested in front-month futures, rather than the back-month contracts dictated by refiners.

In addition to a review of the relationships among oil, crack spread, and ETF spread, previous research on crack spread and ETFs are presented here in more detail. First of all, there has been little research on the effect of oil ETFs on the energy market considering their recent advent in the market. After USO, the first oil ETF, was introduced to the market in April 2006, many oil ETFs have been added, especially since the oil price hike in 2008. Only a little empirical research on financial EFTs has been conducted, with the focus on volatility increases by leveraged ETFs and the effect of inverse ETFs. Most of the literature on crack spread has mainly focused on modeling of hedging strategies with the use of crack spread derivatives (Haigh and Holt 2002; Carmona and Durrleman 2003). Murat and Tokat (2009) showed that crack spread futures are almost as good as crude oil futures in predicting oil prices, using data from January 2000 to February 2008. In addition, they used 3-2-1 crack spreads for forecasting West Texas Intermediate (WTI) crude oil prices.

Based on this literature review, the objective of the current research is to further explore the interaction between the oil and oil product markets, specifically focusing on two relationships; the first, between crude oil and crack spread prices as traded on the New York Mercantile Exchange (NYMEX), and the second, between oil prices and ETFs traded on the New York Stock Exchange (NYSE). Based on the structural breaks observed during the sample periods, we evaluated the forecasting performances of ETFs and crack spread futures. While Murat and Tokat (2009) showed that 3-2-1crack spreads were almost as good as crude oil futures in predicting the movement in spot prices of a light and sweet crude oil, West Texas Intermediate (WTI) crude oil, in this study I investigated whether a 2-1-1 crack spread and its corresponding oil ETFs are good predictors of the price of a relatively heavy crude oil, as in the case of the Organization of the Petroleum Exporting Countries (OPEC) Reference Basket Price.

Methodology

The current research studies the possibility that a 2-1-1 crack spread and oil ETF spread are significant predictors of crude oil prices based on the Error Correction Model (ECM) and additionally the Multivariate GARCH (Generalized Autoregressive Conditional Heteroskedasticity) model. Murat and Tokat's study (2009) on forecasting WTI crude oil prices with a 3-2-1 crack spread utilized the Error Correction Model (ECM), which defines the mean equation using the homoskedasticity of variance. If hetroskedastic features of volatility are observed in the data, adding a volatility equation might be a more appropriate model specification. In Figure A-3, a volatility-clustering characteristic in the financial market was observed in the data for OPEC crude oil and a 2-1-1 crack spread. Daily squared returns of crude oil and crack spread are one evaluator of time varying volatility features. Specifically, the Dynamic Conditional Correlation (DCC) MGARCH model, among the various MGARCH models, will be applied to explain the volatility characteristic. Therefore, in this research, I analyzed the data, first, utilizing the Error Correction Model and, second, by applying the ECM-MGARCH model. Additional reviews of methodological tools for the structural break test and forecasting methods are presented sequentially.

Error Correction Model

In the previous research of Murat and Tokat (2009), as the estimation of an ECM model requires the data series to be cointegrated, they investigated the unit root behavior of a series; a 3-1-1 crack spread and WTI oil prices. Based on weekly oil price and crack spread data from January 2000 to February 2008, they showed that the WTI oil series is integrated on the order of one, I(1), while crack spread futures are on the order of zero, I(0). In the current research, the first bivariate case of the ECM model for OPEC crude oil and a 2-1-1 crack spread is explored in equation (2.1). Following Engle and Granger (1987), if both the log of OPEC crude oil prices, oil_t, and the log of 2-1-1 crack spread futures prices, cs_t, are integrated on the order of one and the stochastic error term is stationary, then oil_t and cs_t are said to be cointegrated, and an error correction representation must be made that may take the following form in the mean equation:

$$(2.1)\quad \begin{aligned} \Delta oil_t &= \alpha_{oil} + \sum_{l=1}^{n} \alpha_{11}(l)\Delta oil_{t-l} + \sum_{l=1}^{n} \alpha_{12}(l)\Delta cs_{t-l} + \tau_{oil} ECT_{t-1} + \varepsilon_{oil,t} = \mu_{oil,t} + \varepsilon_{oil,t} \\ \Delta cs_t &= \alpha_{cs} + \sum_{l=1}^{n} \alpha_{21}(l)\Delta oil_{t-l} + \sum_{l=1}^{n} \alpha_{22}(l)\Delta cs_{t-l} + \tau_{cs} ECT_{t-1} + \varepsilon_{cs,t} = \mu_{cs,t} + \varepsilon_{cs,t} \end{aligned}$$

where μ_t is the conditional mean based on the past information set I_{t-1}, ε_t is the stationary disturbance term with conditional heteroskedasticity, and ECT_t is the error correction term. In this frame, τ_{oil} and τ_{cs} is the estimated coefficient of the long run error term, which reflects the adjustment of the short run to long run equilibrium.

In equation (2.2), the second ECM model, for OPEC crude oil and oil ETF spread, has a bivariate equation form with the replacement of *etf* for *oil* in equation (2.1). As both the log of OPEC crude oil prices, oil_t, and the log of ETF spread prices, etf_t, are integrated on the order of one and the stochastic error term is stationary, then oil_t and etf_t are said to be cointegrated, and an error correction representation must be made that may take the following form in the mean equation:

$$(2.2)\quad \begin{aligned} \Delta oil_t &= \alpha_{oil} + \sum_{l=1}^{n} \alpha_{11}(l)\Delta oil_{t-l} + \sum_{l=1}^{n} \alpha_{12}(l)\Delta etf_{t-l} + \tau_{oil} ECT_{t-1} + \varepsilon_{oil,t} = \mu_{oil,t} + \varepsilon_{oil,t} \\ \Delta etf_t &= \alpha_{etf} + \sum_{l=1}^{n} \alpha_{21}(l)\Delta oil_{t-l} + \sum_{l=1}^{n} \alpha_{22}(l)\Delta etf_{t-l} + \tau_{etf} ECT_{t-1} + \varepsilon_{etf,t} = \mu_{etf,t} + \varepsilon_{etf,t} \end{aligned} .$$

where μ_t is the conditional mean based on the past information set I_{t-1}, ε_t is the stationary disturbance term with conditional heteroskedasticity, and ECT_t is the error correction term. τ_{oil} and τ_{etf} is the estimated coefficient of the long run error term, which explain the adjustment speed. In this system, $\tau_i = 0$ is similar to a Vector

Autoregressive (VAR) model for the first difference.

DCC MGARCH Model with Error Correction Term

The modeling of volatility as a time-varying function is required to explain this characteristic of heterogeneous volatility. Engle (1982) proposed an autoregressive conditional heteroskedastic (ARCH) process to describe the time-varying essence of conditional variances which depend on past information. Bollerslev (1986) proposed a more parsimonious and flexible model than Engle’s ARCH model, that is, the Generalized ARCH (GARCH) model. This model takes past error terms and conditional variances into its variance equation simultaneously, to avoid the problem of the number of parameters to be estimated becoming too large, as the number of lagging periods to be considered increases in the ARCH model. The multivariate GARCH class of models was first introduced and formulated empirically by Bollerslev, Engle, and Wooldridge (1988). In a multivariate sense, Bollerslev (1990) extended a seemingly unrelated regression model which parameterized each conditional variance as a single univariate GARCH process. Multivariate GARCH (MGARCH) model allows the conditional covariance matrix of the dependent variables to follow a flexible dynamic structure and allows the conditional mean to follow a vector-autoregressive (VAR) form. As the general MGARCH model is too flexible to perform parameterization, many restricted MGARCH models have been introduced. Among these models, the dynamic conditional correlation (DCC) MGARCH model and the constant conditional correlation (CCC) MGARCH model are investigated to explain the variance equation in this study. Specifically, the

DCC MGARCH model was introduced to explain the conditional quasicorrelations that follow a GARCH (1, 1) process. The DCC model is more flexible than the CCC model and does not introduce an inestimable number of parameters for a reasonable number of series (Engle 2002). In this study, the first bivariate case for OPEC crude oil and a 2-1-1 crack spread uses equation (2.1) as a mean equation and equation (2.3) as a variance equation, which follows the DCC MGARCH formulation proposed by Engle (2002). The variance equation of this first bivariate case can be written in the following form:

$$
\begin{aligned}
&\varepsilon_t = [\varepsilon_{oil,t}, \varepsilon_{cs,t}]^T \mid I_{t-1} \sim N(0, H_t) \\
&H_t = \begin{bmatrix} h_{oil,t}^2 & h_{oil,cs,t} \\ h_{oil,cs,t} & h_{cs,t}^2 \end{bmatrix} = \begin{bmatrix} h_{oil,t} & 0 \\ 0 & h_{cs,t} \end{bmatrix} \begin{bmatrix} 1 & \rho_{oil,cs,t} \\ \rho_{oil,cs,t} & 1 \end{bmatrix} \begin{bmatrix} h_{oil,t} & 0 \\ 0 & h_{cs,t} \end{bmatrix} = D_t R_t D_t \\
&D_t = \text{diag}[h_{oil,t}, h_{cs,t}] \\
&\tilde{\varepsilon}_t = D_t^{-1} \varepsilon_t
\end{aligned} \quad . \tag{2.3}
$$

where H_t is a time-varying conditional covariance matrix, D_t is a time-varying diagonal matrix of conditional standard deviation, R_t is a time varying conditional correlation matrix, $h_{i,t}^2$ is the estimated conditional variance from the individual univariate GARCH model, and $\tilde{\varepsilon}_t$ is a standardized residual vector with a mean of zero and variance of one, which, in this research, is a 2×1 vector of normal , independent, and identically distributed (*iid*) innovation.

In order to explore the vector representation of equation (2.3) in more detail, the variance equation, dependence equation, and conditional correlation are introduced sequentially. First, with regard to the conditional variance, each $\sigma_{i,t}^2$ $(i = oil, cs)$ evolves according to a univariate GARCH model of the following form:

$$(2.4)\quad \begin{aligned} h_{oil,t}^2 &= \beta_{oil} + \sum_{k=1}^{p_{oil}} \beta_{11}(k)\varepsilon_{oil,t-k}^2 + \sum_{k=1}^{q_{oil}} \beta_{12}(k)h_{oil,t-k}^2 \\ h_{cs,t}^2 &= \beta_{cs} + \sum_{k=1}^{p_{cs}} \beta_{21}(k)\varepsilon_{cs,t-k}^2 + \sum_{k=1}^{q_{cs}} \beta_{22}(k)h_{cs,t-k}^2 \end{aligned}.$$

where β_i is a constant term, β_{11} and β_{21} are ARCH parameters, and β_{12} and β_{22} are GARCH parameters. In this model, I assume each conditional variance follows the GARCH (1, 1) model. Therefore, *k* equals 1 in the variance equation (2.4). From the above basic construction, secondly, I note that the dynamic conditional correlation coefficient matrix (R_t) of the DCC model has a time varying form, as a $n \times n$ symmetric positive definite matrix ($\mathrm{Q}_t = q_{ij,t}$) that is specified in the following formula:

$$(2.5)\quad R_t = J_t\mathrm{Q}_tJ_t = \mathrm{diag}(\mathrm{Q}_t)^{-1/2}\mathrm{Q}_t\mathrm{diag}(\mathrm{Q}_t)^{-1/2}$$

where $\mathrm{Q}_t = (q_{ij,t})_{2\times 2}$ is a positive definite matrix, $J_t = \mathrm{diag}[q_{s,t}^{-1/2}, q_{f,t}^{-1/2}]$, and Q_t satisfies:

$$(2.6)\quad \mathrm{Q}_t = (1-\lambda_1-\lambda_2)\bar{\mathrm{Q}} + \lambda_1\tilde{\varepsilon}_{i,t-1}\tilde{\varepsilon}_{j,t-1} + \lambda_2\mathrm{Q}_{t-1}$$

where $\tilde{\varepsilon}_t (= D_t^{-1}\varepsilon_t)$ is the standardized disturbance vector, $\bar{\mathrm{Q}}$ is the unconditional correlation matrix of the standardized residual ($\tilde{\varepsilon}_t$), and λ_1 and λ_2 are parameters that govern the dynamics of conditional quasicorrelation. λ_1 and λ_2 are nonnegative and satisfy the formula, $0 \le \lambda_1 + \lambda_2 < 1$.

In sum, MGARCH models are dynamic multivariate regression models in which the conditional variances and covariances of the errors follow an autoregressive-moving-average structure. The DCC MGARCH model uses a nonlinear combination of univariate GARCH models with time-varying cross equation weights to model the conditional covariance matrix (H_t) of the errors. The diagonal elements of H_t are modeled as univariate GARCH models, whereas the off-diagonal elements are modeled as nonlinear functions of diagonal terms. Bollerslev (1990) proposed a constant conditional correlation MGARCH model in which the correlation matrix is time invariant. For this reason the model is known as a constant conditional correlation (CCC) MGARCH model. Restricting R_t to a constant matrix (R) reduces the number of parameters and simplifies estimations but may be too strict for many empirical applications. I also compared the CCC MGARCH and DCC MGARCH models; however, the DCC MGARCH model is more flexible than the CCC MGARCH model.

Evaluating Forecasting Accuracy

Forecasting model may be selected utilizing diverse criteria. However, Smeral, Witt, and Witt (1992) listed the error magnitude accuracy, directional change accuracy, and turning point accuracy as forecasting accuracy criteria. Most researchers have used error magnitude accuracy for evaluating forecasting performance. In this research, the evaluation of forecasting ability by error magnitude accuracy is based on the root mean squared error (RMSE) and the mean absolute error (MAE). Both are somewhat similar measures and generally give comparable results. While the MAE is simply the actual error without regard to sign, the RMSE takes into account the greater penalty associated with very large forecasting errors. A forecasting error is simply the difference between the forecasted value ($\hat{y}_t$) and actual value (y_t).

(2.7) $$MAE = \sum_{t=1}^{n} \left|\hat{y}_t - y_t\right| / n \qquad RMSE = \sqrt{\sum_{t=1}^{n} (\hat{y}_t - y_t)^2 / n}$$

Within the time frame T_0 to T_k, basic model estimations and within sample forecasting evaluations are done. Model fitness within the sample period is usually evaluated by measuring a residual, the difference between the fitted value and actual value ($\varepsilon_t = \hat{y}_t - y_t$). Within the time frame T_{k+1} to T_n, the forecasting error is evaluated by measuring the difference between the forecasted and actual values. One step ahead of the forecasted value is derived by applying the recursive forecasting model at time *k+1* to *n*. Therefore, this residual concept may be applied both within and outside of the sample periods when I use the ECM.

Normally, residual can be thought of as an element of variation that is not explained by the fitted model. Therefore, the hetroskedastic variance in the MGARCH model could reduce the unexplained variance, or residual, in the ECM. While the ECM model assumes homoskedasticity, the MGARCH model basically enables the conditional covariance matrix of the dependent variables to follow the flexible dynamic structure of heteroskedasticity. In order to measure statistical loss by the MGARCH forecasting model, I used the standardized residual ($\tilde{\varepsilon}_t$) in place of the residual (ε_t) of ECM. In equation (2.3), the standardized residual is derived by dividing the residual by the square root of conditional variance ($\tilde{\varepsilon}_t = D_t^{-1}\varepsilon_t$).This is one of the modified methods for scaling residuals, which usually divide the forecasting error by the square rooted value of variance. This change could reduce the effect of strong outliers in forecasting performance evaluations and penalize these outliers by dividing hetroskedastic variance.

Data

The data set includes daily spot prices for the OPEC Reference Basket (ORB), daily time series for prices of NYMEX futures contracts written on a 2-1-1 crack spread, and daily close prices of oil ETFs trading on the NYSE. Spot prices for the ORB were comprised of averages of the OPEC member’s crude oil prices, as the OPEC collects price data on a ‘basket’ of crude oils and calculates an average price from these oils to use as the OPEC reference price. For the crack spread futures price, I utilized the daily settlement price for all NYMEX-traded futures contracts on crude oil, heating oil and unleaded gasoline. Specifically, a 2-1-1 crack spread indicates that two barrels of crude oil yield one barrel, each, of gasoline and heating oil, which is characteristic of heavy crude oils, similar to OPEC basket grades, because heavy crude oils do not yield as much gasoline as light crude oils.

As the use of a combination of individual oil-related ETFs enables us to lock in the cracking margins of refiners by traditional crack spread trading, I derived the daily ETF crack spread value using the daily close price of three oil ETFs**,** Pro Shares Ultra Short DJ-UBS Crude Oil ETF (NYSE Arca: SCO), United States Gasoline Fund (NYSE Arca: UGA)**,** and United States Heating Oil Fund (NYSE Arca: UHN)**.** Buying equal lots of SCO**,** UHN and UGA effectively puts an investor short two units of crude oil and long one unit, each, of gasoline and heating oil, which simulates the buying of a 2-1-1 crack spread in the futures market. Each data set was obtained from OPEC, NYMEX, and NYSE data sources, respectively.

The oil ETFs data spans the period from January 2009 to December 2011, encompassing the launch of the SCO ETF on November 2008. Basic statistics on level data and first difference data is described in Table B-1. Dependant variables in the ECM comprised the first difference data, and three series in the first difference of log prices exhibited negative skewness and higher kurtosis than in the normal distribution. After performing the Jarque-Berra test for normality of distribution, all three series were significantly different from a normal return distribution.

In order to examine the relationships among oil prices, crack spread, and ETF spread, I plotted two correlations: one was between oil prices and crack spread, and the other was between oil prices and ETFs, on the left side of Figure A-4. Using the 50 daily data points, I calculated the correlation by the moving window method. The dashed line represents the unconditional correlation, which is derived from all the data. The overall correlation between oil prices and ETFs is higher than the one between oil prices and crack spread. In addition, correlations using the moving window method change very quickly in both of the two relationships. Therefore, the time-varying correlation model was more appropriate for use than the time-invariant correlation model. Among the diverse multivariate correlation model of MGARCH, time-varying conditional correlation model, which includes dynamic conditional correlation or varying conditional correlation model, is more appropriate than constant correlation model. This topic will be revisited later by an examination of MGARCH model specifications for conditional correlation.

Empirical Results and Discussion

Estimation of an ECM model requires that the data series be cointegrated. In the current research, the unit root behavior of a series was investigated; the Dickey-Fuller and Kiwatkowski-Phillips-Schmidt-Shin (KPSS) tests were applied to measure the unit root and first difference of log data. These two tests were utilized to investigate the opposite null hypotheses: the ADF test used the unit test as the null hypothesis and the KPSS test used the stationarity test as the null hypothesis. Table B-2 shows the unit root test results. Based on the KPSS test, a basically level data series indicates that the stationarity hypothesis is rejected at the 1% significance level, but the stationarity hypothesis cannot be rejected by the first difference data. The results of the ADF test are similar to the KPSS results, except for the crack spread level data. The ADF test suggests that the unit root hypothesis is rejected at the 5% significance level by the crude oil level data series, which indicates that the crack spread futures series are integrated on the order of zero, I(0). This finding corresponds with the unit root test results for the crack spread 3-2-1 by Murat and Tokat (2009).

Therefore, the ADF test suggests that the crude oil and ETF series are integrated on the order of one I(1), while crack spreads are integrated on the order of zero, I(0). However, the KPSS test results suggest that all three level series are integrated on the order of one I(1). In sum, although there is a dispute regarding the stationarity of crack spread, the crude oil and ETF level data are integrated on the order of I(0). Johansen

(1995) defined a stochastic vector process to be I(1), if the highest order of integration of any of its elements was I(1).

Cointegration of the series was assessed by the Johansen Maximum Likelihood (ML) Test. The null hypothesis that no cointegration occurred was tested in two systems of equations, first oil prices and crack spread, and, second, oil prices and ETFs. In the first set of equations, for oil prices and crack spread, the trace statistic of 19.7725 at rank = 0 exceeds the critical value of 15.41. Thus, the null hypothesis that no cointegration equations exist is rejected. However, since the trace statistic of 2.1320 at rank = 1 is less than its critical value of 3.76, the null hypothesis that there are one or fewer cointegrating equations cannot be rejected from the test result. Since Johansen's method for estimating rank entails accepting the first rank at which the null hypothesis is not rejected, I accepted rank 1 as an estimate of the number of cointegrating equations for these two variables. In the case of the second set of equations, for oil prices and ETFs, the no cointegration hypothesis was also rejected at the 5% significance level. In order to perform the Johansen ML procedure in Table B-3 for the system of two equations, the lag selection was based on the Final Prediction Error, using Akaike, Schwarz and Hannan-Quinn information criteria. A lag structure is selected as a result of majority rule among four criteria. Based on this criterion, the optimum lag length for the system is 2.

Granger-Causality Test

In this research, the data sample periods were from October 2005 to December 2011. However, if a structural break occurred within a sample period, an empirical estimation using the entire sample might fail to provide reliable results (Clements and Hendry 2006). As the sample periods included the global financial crisis in 2008, taking into account any changes in the financial environment would be desirable for parameter stability. To explain the major changes occurring during the financial crisis and identify other unknown changes within the sample periods, the model coefficient should be flexible for one, or several, dates. And, since the actual dates were unknown, I had to estimate them as well as the model parameters. In this study, I adopted Zivot and Andrews' model (1992) to determining the break point endogenously from the data. Zivot and Andrews (1992) endogenous structural break test is a sequential test which utilizes the full sample and uses a different dummy variable for each possible break date. The break date is selected where t-statistic from the Augmented Dickey-Fuller test of unit root is at a minimum. Consequently a break date will be chosen where the evidence is least favorable for the unit root null.

When the first structural break point is detected and identified, we still do not know whether more than one break exists. After identifying the first break point, a second break point among unstable parameters is tested to investigate whether more than one break exists. Also, when structural breaks occur within the sample period, an Error Correction Model analysis for each subsample period will provide further insights into the structural relationship between crude oil and other oil-related financial investment tools such as crack spread and oil ETF spread.

Based on the results of the unknown structural break test, which is described in the methodology section, two structural break points were found; one on September 2, 2008 and the other on April 29, 2009. Those times are consistent with the period of the 2008 financial crisis, which was related to subprime mortgages. On September 2008, Lehman Brothers submitted a bankruptcy petition and Merrill Lynch was sold to Bank of America. During this month, the global stock markets, including the Dow Jones Industrial Average, FTSE of England, CAC40 of France, Dax30 of Germany, and Hang Seng of HongKong, dropped precipitously. Likewise, the price of oil decreased significantly, after a record peak of US$145 in July 2008. On December 23, 2008, the WTI crude oil spot price fell to US$30.28 a barrel, the lowest since the financial crisis of 2008 began, and traded at between US$35 and US$82 a barrel in 2009. After April 2009, the global oil market recovered from the price collapse resulting from the financial crisis. The possibility of multiple structural breaks was investigated by applying the structural break test for each sub-sample period. As no further break points were detected, the entire sample period was divided into three-sub groups; the 1st period (October 2005 to September 2008), 2nd period (October 2008 to April 2009), and 3rd period (May 2009 to December 2011). These divisions can be denoted as the pre-crisis, crisis, and post-crisis periods.

Based on the detected structural breaks, I conducted a Granger-causality test for oil prices and crack spread on the entire period and the three sub-sample periods. As ETF spread data could only be obtained beginning in January 2009, the causality test between oil prices and ETFs was only done in the 3rd period (May 2009 to December 2011). A variable, *x*, can be said to Granger-cause a variable, *y*, if, given the past values of y, the past values of x are useful in predicting y. A common method for testing

Granger causality is to regress *y* on the lagged values of *x* that are jointly zero. Failure to reject the null hypothesis is equivalent to failing to reject the hypothesis that *x* does not Granger-cause *y*. Table B-4 gives the Wald test-Granger Causality statistics. The null hypothesis is that the coefficients of all the lags of an endogenous variable are jointly zero. In the whole sample, where no structural breaks were accounted for, I found that crack spread had no causal impact on crude oil prices, but the causal impact of oil prices on crack spread was observed. Based on the structural breaks, the 1st and 2nd periods exhibited the same results as the whole sample, while in the 3rd period the dynamic between oil prices and crack spread started to change and crack spread futures became a significant leading factor in the crude oil market. The causality test for oil prices and ETFs resulted in the 3rd period also showing a strong unidirectional causal relationship from ETFs to the crude oil market.

This result is consistent with our expectations. After the financial crisis, the global economy experienced a dramatic drop in stock prices and commodity prices both before and after the crisis. This trend of synchronization between financial and commodity markets is accelerated by increased trading of oil-related futures and the advent of new financial investment tools, like oil ETFs, that link the two markets. Especially, while the futures market has a relatively limited number of investors due to margin requirements, ETF investors were able to easily enter the financial market by selling and buying on the NYSE market. Dynamic changes in the global oil market has caused oil market investors to depend on more sophisticated financial tools, such as crack spread and oil ETFs, for both investing and hedging. For the analysis of the 3rd

period, the level and first difference data for the three series are indicated in Figure A-5. Compared to oil prices and ETF spread, the crack spread shows more pronounced fluctuations.

Estimation of the ECM model

Based on the structural breaks and Granger-causality results, ECM estimations for oil prices and crack spread during the entire sample period, 1^{st}-2^{nd} period, and 3^{rd} period were given in Table B-5. Contrary to the 1^{st} and 2^{nd} periods, the crack spread at t-1 in equation Δoil_t had a significant positive coefficient (0.0219) during the 3^{rd} period. Considering that data for ETFs was only available after 2009, results of the estimation of ECM for oil prices and ETFs series were also obtained. In the 3^{rd} period, ETFs at t-1 in equation Δoil_t had a significant positive coefficient (0.7450), which indicates the robust effect of the independent variables on the dependent variable. This estimation result is consistent with the Granger-causality results. In order to discuss and compare the forecasting performance of crack spread and ETFs, I focused on the 3^{rd} period data to extend this analysis. Hence, the MGARCH analysis and predictions are based on the 3^{rd} period hereafter. The estimate coefficient results were consistent with the Granger causality results. In the 3^{rd} period, the coefficient Δcs_{t-1} in the equation Δoil_t of ECM 3 and the coefficient Δetf_{t-1} in the equation Δoil_t of ECM 4 had positive significant values, which indicates positive causal relationships between crack spread and crude oil, and between ETFs and oil, respectively. In addition, the coefficient of the error correction term for both ECM 3 and ECM 4 indicated the adjustment speed.

Generalized impulse responses to one standard error shock are estimated in order to analyze the relative responses of oil prices by crack spread and ETF spread. The estimations for ECM 3 and ECM 4 are compared in Figure A-6. As crack spread and ETF spread both have Granger causality to oil prices in this period, an increase in the orthogonal shock in crack spread and ETF spread causes an increase in oil price movement. However, a shock in ETF spread has a relatively stronger effect on oil prices than a shock in crack spread. The speed of adjustment toward the long-run equilibrium of crack spread and ETF spread is the same for the two periods. In addition, in terms of a shock from oil prices, there is little effect on crack spread and ETF, but ETF response disappeared after two periods while crack spread response disappeared after one.

After applying the two ECM models (ECM 3 and ECM 4) to the 3rd period, the residuals of oil and CS in the ECM 3 model and the residuals of oil and ETFs in the ECM 4 model were plotted on the left side of Figure A-7. The heteroskedasticity of residuals were tested using the Breusch-Pagan (BP) and White tests. While the BP test measures whether the estimated variance of the residuals from a regression is dependent on the values of the independent variable, the White test is a statistical test that establishes whether the residual variance of a variable is constant. The results of both tests suggest that the null hypothesis of homoskedasticity is rejected at the 1% significance level for the residual of CS in ECM 3, residual of oil in ECM 4, and residual of ETF in ECM 4, and the null hypothesis of homoskedasticity is rejected at the 5% significance level for the residual of oil in ECM 3. Therefore, I utilized an estimation model to estimate the time varying variance.

Estimation of ECM-MGARCH model

As heteroskedasticity, shown on the left side of Figure A-7, and volatility clustering, shown in Figure A-3, were observed, the application of a time varying variance model was needed. In this study, I applied the MGARCH model, which enabled the conditional variance to be dynamic. Next, it was important to define the correlation in a multivariate model. However, the general MGARCH model is too flexible. Among the following diverse models, the diagonal vech model (DVECH), the constant conditional correlation model (CCC), the dynamic conditional correlation model (DCC), and the time-varying conditional correlation model (VCC), I compared the correlations of the CCC MGARCH and DCC MGARCH models in order to decide which specification of correlation was appropriate for this model. The DCC MGARCH model was as flexible as the closely related VCC model, more flexible than the CCC model, and more parsimonious than the DVECH model. I finally utilized the DCC MGARCH model, in which the conditional variances are modeled as univariate generalized autoregressive conditionally hetroskedastic models and conditional covariances are modeled as nonlinear functions of the conditional variances. We used the ECM-DCC MGACH model to investigate the dynamic interaction between oil prices and crack spread, and between oil prices and ETFs. Thus, the ECM framework was applied to investigate the causality relationship among variables and a DCC MGARCH model took into account the variables' hetroskedastic properties of variances and covariance. I first used the bivariate ECM-MGARCH (1, 1) model to probe the transmission effects among oil prices and crack spread in their first and second moments. Secondly, the other bivariate ECM-MGARCH (1, 1) model was estimated for crude oil spot prices and ETF spread.

The mean equation estimation and test statistics for the ECM-DCC MGARCH model are presented in Table B-6. As this ECM-MGARCH model focuses on the volatility model, I omitted the condition mean by the ECM model. For the MGARCH model comparison, the ECM-MGARCH constant conditional correlation (CCC) model is additionally estimated, which basically assumes a constant conditional correlation. The estimated correlation results for CCC were different from those of ECM-MGARCH 1 and ECM-MGARCH 2. If the sum of the coefficients of arch (1) and garch (1) are estimated as being close to 1, which implies that shocks cause a high persistence in volatility. In the ECM-MGARCH model 1 for oil prices and crack spread, this sum in oil equation is lower than that in crack spread equation. The shock effect of crack spread is more persistent than the shock effect of oil. This relationship is also observed in the ECM-MGARCH model 2. In addition, the $\lambda_1 + \lambda_2$ estimates of the both DCC ECM-MGARCH model are close to (but less than) 1, which implies that the correlations between oil prices and ETFs in case of ECM MGARCH 2 are highly persistent. Such high persistence means that a shock can move the correlation away from its long-run average for a considerable time, although the correlation is eventually mean-reverting. Therefore, the DCC MGARCH model may capture the variation in correlation between crude oil spot prices and ETF spread more effectively than the CCC MGARCH model does.

On the right side of Figure A-4, the straight line represents the conditional correlation of the ECM-CCC MGARCH model and the time-varying correlation line represents the conditional correlation of the model. While the correlation between oil prices and crack spread in the CCC model is very low (0.1051), the correlation between oil prices and ETFs in the CCC model is relatively high (0.5990). In terms of the DCC, the conditional correlation between oil prices and crack spread is low and sometimes negative, while the conditional correlation between oil prices and ETFs, although a little low after the financial crisis increased to more than 0.5 as the economy recovered. As the dynamic correlations change strongly in both ECM-MGARCH 1 and ECM-MGARCH 2 models, the DCC model is more appropriate than the CCC model, which assumes the correlation is constant. This result for conditional correlation is consistent with comparisons of unconditional correlation among oil prices, crack spread and ETFs, as shown on the left side of Figure A-4.

Additionally, the time-varying variance of the ECM-MGARCH model and the time invariant variance of the ECM model are compared in Figure A-8. The graph shows relatively large fluctuations in the volatility of crack spread. The conditional variances of the same variable, oil prices, were different for the ECM-MGARCH 1 and ECM-MGARCH 2 models. Oil return which is estimated by crack spread is more volatile than oil return as estimated by ETFs. From Figures A-4 and A-8, ETFs can be used to obtain a relatively higher dynamic correlation with less volatility to estimate an oil series than for crack spread.

Forecasting Performance

The predictive ability of a model to fit and explain oil price movements is assessed by using the random walk model (RWM) as a benchmark. The RWM is a univariate model of oil return data, which moves randomly around the mean and is widely used in the area of finance. Therefore, a better prediction of performance than the univariate approach, such as the random walk model, supports the validity of the multivariate approach for forecasting oil price returns. In this study, two multivariate approaches are suggested for oil prices and crack spread, and for oil prices and ETFs. Dynamic forecasts of oil returns were computed based on crack spread, first, and ETFs, second, and then these forecasts were compared with those provided by the RWM.

For outside of sample forecasting, the data was divided into two periods, the first was May 2009 to September 2011, and the second was October 2011 to December 2011. Outside of sample forecasting may be made one step ahead by the recursive method. Forecasting performance was evaluated on the basis of MAE and RMSE in equation (2.7). Table B-7 reports the forecasting error statistics.

In terms of the ECM, the evaluation of the forecasting performance of the three models shows that, first, two ECMs for crack spread and ETFs outperform the random walk model and, second, ETFs exhibit a better ability to predict oil prices than crack spread does. Therefore, the ECM model with a 2-1-1crack spread shows superior predictive ability for heavier crude oils than the RWM, which is consistent with the results of Murat and Tokat (2009), who showed that an ECM model for a 3-2-1crack spread outperformed the RWM model. In addition, ETFs are better predictors of oil price movement than crack spread. The Diebold and Mariano (1995) test is used to check if the difference between the MSE produced by the two alternative model forecasts are statistically significant. The result of the DM test indicates the better forecasting performance of ETF as compare to CS because the difference between the models is statistically significant at 0.1% level in favor of ETF forecasting model. The null hypothesis of DM test can also be rejected at the 5% significance level in comparison with RWM.

In terms of the ECM-MGARCH model, it is difficult to derive consistent results among the three models. Overall, the evaluation of forecasting performance for ETFs does not differ from the crack spread model, both in the ECM and ECM-MGARCH models. However, by taking into account heteroskedasticity and volatility clustering by implementation of the MGARCH model, less volatile ETFs shows a better predictive ability than relatively strong volatile crack spread. In addition, this result is consistent with the fact that the relationship between ETFs and oil prices shows a higher dynamic correlation than the one between crack spread and oil prices.

Several limitations of this study are noteworthy. First, the daily data of this research covers only a period of 32 months, from May 2009 to December 2011, which also corresponds to the post-2008 financial crisis period. Most of the previous studies on crack spread used a relatively long period of data reported on a weekly basis (Haigh and Holt 2002, 1984 to 1997; Murat and Tokat 2009, 2000 to 2008). In fact, the data used in this study reflects the launch of oil ETFs around 2008 and the causal relationship change in crude oil prices in April 2009. Second, the forecasting performance was evaluated by both the ECM model and ECM MGARCH model framework. The results of the ECM MGARCH model confirm the results of the ECM model; however, recent forecasting models based on the GARCH model focus on volatility forecasting, which require the realized volatility level as actual data. There are diverse disputed methods used to define the realized volatility, such as computing the difference between weekly maximum and minimum data. Taking into account the 32 months of data utilized, this study focus only on forecasting price level. Therefore, further research on forecasting crude oil prices using longer periods of ETF data would be helpful in identifying the unique effect of ETFs in the oil market. More research is recommended to include volatility forecasting by crack spread and ETF spread to further test the forecasting performance of volatility-based models.

Summary and Concluding Remarks

A number of studies have investigated the relationship between spot oil prices in the oil market and crack spread in the futures market. In this study, I call attention to a new financial instrument; oil-related Exchange-Traded Funds (ETFs), in predicting the movement of spot oil prices. To evaluate the performance of crack spread and ETFs in managing oil price risk, the random walk model (RWM) was applied as a benchmark. Furthermore, I compared the predictive abilities of crack spread and ETFs, both in the error correction model (ECM) and ECM-multivariate GARCH (ECM-MGARCH) model.

In this study, based on the unknown structural break test, a change in causal relationship between crack spread and oil was observed after a break point. The relationship between ETFs and oil prices, however, remained the same after a break point. The results of this study reveal that crack spread futures and oil ETF spread are good predictors of oil price movement and, in a comparison of crack spread and ETFs, ETFs are better predictors than crack spread. In case of crude oil forecasting by crack spread, the Granger causal relationship was compared depending on the sub-sample periods. The change in this causal relationship can be explained by the fact of the increasing need of the oil-related financial market for oil price hedging and investments. The break points corresponded to the beginning of the global financial crisis in 2008 and the start of the recovery in 2009. As a result of the economic crisis in 2008, the financial and commodity markets experienced a similar pattern of price decreases and increases. This was also observed in the oil market, which indicates the synchronization of price movement and volatility. In addition, the entry of new investors in the oil market with new financial products like oil ETFs, and the increasing need of refiners for hedging and speculation in the more volatile oil market with crack spread, has increased the importance of these financial tools in the entire oil market. In summary, the ECM and MGARCH models were used to provide valuable information on the relationship between crack spread and oil-related ETFs, and the forecasting performance of these two models for oil-related financial markets was compared.

CHAPTER III

OPTIMAL DYNAMIC HEDGING OF UNLEADED GASOLINE USING FUTURES AND EXCHANGE-TRADED FUNDS

Introduction

The purpose of this study was to examine the effectiveness of dynamic hedging of the unleaded gasoline spot price, using gasoline futures and gasoline exchange-traded funds (ETFs). Basically, futures contracts are standardized forward contracts with an inherent obligation to take delivery of or to deliver a set quantity of a specific financial instrument at an agreed price on a specific date. Therefore, gasoline futures can be utilized as a direct hedging or investment tool for the spot price for unleaded gasoline. Since 2006, diverse oil related-ETFs have been introduced in the market and they can be used in an active or passive way to construct a portfolio. An ETF market can function as an alternative to hedging tools for the traders who seek the cheapest means to reduce the uncertainty of their market exposures.

The participants in an oil ETF market may buy or sell on their own account to counteract temporary imbalances in supply and demand and hence stabilize prices. While a long position for an ETF produces a return similar to that of an index or the underlying portfolio, a short position for an ETF offers inverse returns. The ability to short sell, coupled with low transaction costs, has fuelled a significant increase in ETF usage. An underlying feature of most ETFs is an index, and the price of an ETF basically follows the movement of its underlying index. For example, if an index loses 5% within a one day period, a short ETF with this underlying index would rise by 5%. In addition, the ability to trade an ETF with put or call options at a given time in the future also indicates that ETFs are robust alternative tool for futures trading.

In the futures market, high trading costs and costly trading information has caused temporary divergences in the equilibrium relationship between spot and futures prices. Alexander and Barbosa (2007) addressed the effect of an ETF launch on spot and futures relationships. As transaction costs drop and spot-futures arbitrage is facilitated by ETFs, the correlation between spot and futures returns increases and basis risk declines.

In the unleaded gasoline market, gasoline futures generally provide an opportunity to hedge risk, but they are volatile and are not suitable for all investors. However, the introduction of diverse oil ETFs has enabled a wide array of investors to access the energy market with strong liquidity and without high entry costs. Especially, the United States Gasoline Fund (UGA)[7] has been traded actively on the New York Stock Exchange market (NYSE) since February 26, 2008. The UGA is designed to reflect changes in percentage of the price of gasoline futures contracts traded on the New York Mercantile Exchange (NYMEX). Therefore, portfolio selection of futures or ETFs involves not only the maximization of return, but also the issue of risk management.

In this study, first, I calculated the most efficient hedge ratio for gasoline futures utilizing several diverse advanced econometric models and then compared their hedging

---

[7] This fund tracks the change in percentage of the price of gasoline and invests in the nearest futures contract on unleaded gasoline delivered to the New York harbor traded on the New York Mercantile Exchange (NYMEX).

performance. Second, using this framework of futures hedging, I extended the current research by comparing the hedging performance of ETFs to that of futures hedging. In Figure A-9, it is apparent that these three data series exhibit very similar price movements. During the period of data collection, the financial crisis in 2008 resulted in the sudden price drop, around September 2008, and a slow price recovery, beginning in 2009.

An evaluation of futures hedging is principally of interest in the oil market due to problems with the limited storage capacity of oil and its products, high volatility in prices, and the recent price hike. The optimal proportion of futures contracts that should be held to offset a spot position is called the optimal hedge ratio. The optimal hedge ratio is traditionally estimated by calculating the ratio of the unconditional covariance between spot and futures prices and the unconditional variance in the price of futures. This method is called static hedging, and the optimal hedge ratio can be estimated by utilizing ordinary least squares (OLS) regression. Although the extant literature places emphasis on estimating a static hedge ratios using the ordinary least square technique, more recent studies employ various bivariate conditional volatility models to estimate a time-varying hedge ratio and have demonstrated that a dynamic hedging strategy can result in greater risk reduction that a static one (Hsu, Tseng, and Wang 2008; Liu, Jian, and Wang 2010). The static regression method is often criticized because all financial assets and commodities have time-varying second moments, and, thus, hedge ratios will be time-varying and arguably best modeled in a dynamic framework (Chen, Lee, and Shrestha 2003).

There are many methods available for estimating optimal hedge ratios. Among these methods, multivariate generalized autoregressive conditional heteroskedasticity (MGARCH) models has demonstrated the superiority of time varying hedge ratios by taking into account the changing joint distribution of spot and futures returns. Most of the earlier dynamic hedge ratio models were based on the dynamic conditional correlation multivariate GARCH (DCC MGARCH) model (Engle and Sheppard 2001; Engle 2002). Hedging performance depends on the marginal distributions and dependence structure of the variables. However, most of these dynamic hedging models assume that the spot and futures returns follow a multivariate normal distribution with linear dependence. As the typical assumption of joint multivariate normality is made, this is an important limitation of the multivariate GARCH model, because this assumption is at odds with numerous empirical studies, in which it has been shown that many asset returns are skewed, leptokurtic, and asymmetrically dependent. Recently, an important development in modeling dependence structures, known as a copula, was proposed by Sklar (1959). The copula function completely describes the dependence between N variables. If we do not make the assumption of multivariate normality, a joint distribution can be decomposed into its marginal distributions and a copula, which can be considered both separately and simultaneously. The various copula functions allow great flexibility in modeling joint distributions. Moreover, especially in this study, static copula models are extended to dynamic copula functions in order to apply a dynamic optimizing hedge model. This dynamic copula theory was recently developed in an analysis of time-varying conditional dependence (Patton 2006). This research investigates the diverse range of copula functions and compares the hedging strategies of each model using futures and ETF prices.

Table B-8 shows the basic framework of the models examined in this research. As the relationship between spot and futures data for gasoline is cointegrated, the basic static hedge model is based on the Vector Error Correction (VEC) model. The time invariant dependence allows the use of the conditional covariance to capture the updated information. In this frame work, the DCC MGARCH model generates the time varying covariance and variance structure in order to derive the dynamic hedge ratio. Compared to these symmetric and multivariate normal distribution-based VEC and DCC MGARCH models, the copula function allows the consideration of an asymmetric and non-normal multivariate distribution-based model. Both static and dynamic copula-based GARCH models are exploited to analyze the time varying optimal hedge ratio in the case of gasoline.

This study is organized as follows. After the introduction, the second section discusses the theoretical background of copula methodology. The third section describes the estimation methods for the models considered, including the Vector Error Correction Model (VEC), the dynamic conditional correlation multivariate GARCH model (DCC MGARCH), and the copula-based GARCH model (CGARCH), divided into static and dynamic copula models. The fourth section details the data used in this study and provides a comparison of the empirical results from the different hedging models, presenting the optimal hedge ratio and hedge effectiveness. The conclusions drawn from this study are presented in the last section.

Theoretical Background on Copula Methodology

The copula function describes the dependence structure of a set of variables and this has become a standard tool in modeling dependence among time series, especially when the researcher does not want to make the highly debatable assumption of multivariate normality. In addition, dynamic copula functions can describe the time-varying dependence without making the multivariate normality assumption. Examples of the application of dynamic copula functions to the field of finance can be found in (Patton 2006, 2009; Rodriguez 2007; Hsu, Tseng, and Wang 2008; Liu, Jian, and Wang 2010). In this section, I introduce the basic concept of the copula function, the bivariate static copula functions, and the bivariate dynamic copula functions.

The Copula Function

A copula function represents a flexible dependence structure for a set of variables, and Sklar's theorem (1959) provides a link between a joint distribution and the corresponding copula. As a hedge ratio is derived from the relationship between two variables, I focus on the bivariate case of the copula function. According to Sklar's theorem, for every *n* dimensional distribution function *F*, with marginal distribution $F_i (i = 1, \ldots, n)$, there exists a copula *C*, such that:

(3.1) $$F(x_1, \ldots, x_n) = C(F_1(x_1), \ldots F_n(x_n))$$

where $F_i(x_i)$ is the cumulative distribution function (CDF) respectively, $F(x_1,\ldots,x_n)$ is a joint CDF of x_1, …, x_n with margin $F_1(x_1)$, …, $F_n(x_n)$. The copula defined in (3.1) will be unique, if all marginal distributions are continuous.

In order to apply the copula function in bivariate cases of spot and futures returns In hedging application, I define sp_t and fu_t as random variables denoting spot returns, and futures returns of gasoline at period *t*, and denote $F_{s,t}(sp_t \mid I_{t-1})$ and $F_{f,t}(fu_t \mid I_{t-1})$ as their conditional cumulative distribution functions (CDF), respectively, where I_{t-1} denotes all past return information. The conditional copula function $C_t(u_t, v_t \mid I_{t-1})$ is defined by the time varying CDF of spot returns and returns of its hedging tools, futures or ETFs, in equation (3.2), where $u_t = F_{s,t}(sp_t \mid I_{t-1})$ and $v_t = F_{f,t}(fu_t \mid I_{t-1})$ are distributed as continuous uniform variables on (0, 1). From Sklar's theorem, I know the bivariate conditional CDFs of sp_t and fu_t can be written as

$$F(sp_t, fu_t \mid I_{t-1}) = C_t(u_t, v_t \mid I_{t-1}) \tag{3.2}$$

Assuming all *F* (CDFs) in equation (3.2) are differentiable, the joint density can be obtained by;

$$\begin{aligned} f(sp_t, fu_t) &= \frac{\partial^2 F(sp_t, fu_t \mid I_{t-1})}{\partial s_t \partial f_t} \\ &= c_t(u_t, v_t \mid I_{t-1}) \times f_{s,t}(sp_t \mid I_{t-1}) \times f_{f,t}(fu_t \mid I_{t-1}) \end{aligned} \tag{3.3}$$

where $c_t(u_t, v_t \mid I_{t-1}) = \partial^2 C_t(u_t, v_t \mid I_{t-1}) / \partial u_t \partial v_t$ is the conditional copula density function.

From (3), the log-likelihood functions for u_t and v_t is;

$$(3.4)\quad \begin{aligned} L(\theta) &= \sum_{t=1}^{T} l_t = \sum_{t=1}^{T} \ln f(sp_t, fu_t; \theta) \\ &= \sum_{t=1}^{T} \ln f_{s,t}(sp_t; \theta_1) + \sum_{t=1}^{T} \ln f_{f,t}(fu_t; \theta_2) + \sum_{t=1}^{T} \ln c_t(F_{s,t}(sp_t; \theta_1), F_{f,t}(fu_t; \theta_2); \theta_3) \end{aligned}$$

where T is the number of observations and $\theta = (\theta_1, \theta_2, \theta_3)$ is the parameters in the marginal densities $f_{s,t}(\cdot)$ and $f_{f,t}(\cdot)$, and the copula shape parameter.

The log-likelihood is decomposed into two parts, the first two terms related to the marginal distributions and the last term related to the copula. In order to apply the bivariate case of spot and ETF returns, the etf_t, ETF return at period *t*, is used in place of fu_t in equations (3.1), (3.2), (3.3) and (3.4).

Bivariate Static Copula Functions

There are many types of copula functions; in generally, they can be divided into elliptical and Archimedean copulas, and each category has a number of specific connection functions. Since the copula function determines the dependence structure, its selection should depend on the type of dependence observed in the data set. The usual choice is an elliptical copula such as a Normal (Gaussian) copula or Student's t copula. In the case of the Gaussian copula, following equation (3.2), the bivariate Gaussian copula is defined as:

$$(3.5)\quad C_t^{Gaussian}(u_t, v_t; R) = \Phi_R(\Phi^{-1}(u_t), \Phi^{-1}(v_t))$$

where R is a correlation matrix and $\Phi(\cdot)$ represents the CDF of the standard normal distribution.

The density function of the Gaussian copula in equation (3.5) is:

(3.6) $$c_t^{Gaussian}(u_t, v_t; R) = \frac{1}{|R|^{1/2}} \exp\{-\frac{1}{2}\eta'(R^{-1} - I)\eta\}$$

where R is a correlation matrix, $\eta = (\Phi^{-1}(u_t), \Phi^{-1}(v_t))$, and $\Phi^{-1}(\cdot)$ is the inverse of the univariate normal CDF.

The density function of the bivariate Gaussian copula using correlation (ρ) rather than the correlation matrix (R) is:

(3.7) $$c_t^{Gaussian}(u_t, v_t; \rho) = \frac{1}{\sqrt{1-\rho^2}} \exp\{-\frac{1}{2(1-\rho^2)}[a_t^2 + b_t^2 - 2\rho a_t b_t] + \frac{1}{2}[a_t^2 + b_t^2]\}$$

where ρ is the linear correlation coefficient and constrained within the interval (-1, 1), $a_t = \Phi^{-1}(u_t)$ and $b_t = \Phi^{-1}(v_t)$.

The family of Archimedean copulas, in contrast to the elliptical ones, could be used to describe the asymmetric tail dependency of variables. An Archimedean copula can be expressed as:

(3.8) $$C(u_t, v_t) = \varphi^{-1}(\varphi(u_t) + \varphi(v_t))$$

where φ is a convex decreasing function, called a generator. Different generators will induce different copulas in the family of Archimedean copulas.

To capture potential asymmetric tail dependency, the Clayton and Gumbel copulas, in the Archimedean family, have been used broadly. While the Clayton copula considers the lower tail, the Gumbel copula focuses on the upper tail. In these two Archimedean copulas, there is a one-to-one mapping relationship between Kendall's τ and θ (parameter), which clearly shows that the copula shape parameter determines the dependence structure. The generator for the Clayton copula is $\varphi_{\theta_C}(x) = (x^{-\theta_C} - 1)/\theta_C$. For $\theta_C > 0$, the CDF and PDF for the Clayton copula are

(3.9) $$C_t^{Clayton}(u_t, v_t; \theta_C) = (u_t^{-\theta_C} + v_t^{-\theta_C} - 1)^{-1/\theta_C}$$

(3.10) $$c_t^{Clayton}(u_t, v_t; \theta_C) = \frac{(1+\theta_C)(u_t^{-\theta_C} + v_t^{-\theta_C} - 1)^{-\frac{1}{\theta_C}-2}}{(u_t v_t)^{1+\theta_C}}$$

Kendall's τ for the Clayton copula is $\theta_C / (\theta_C + 2)$. The upper tail dependence is $\lambda_U^{Clayton} = 0$ and the lower tail dependence is $\lambda_L^{Clayton} = 2^{-1/\theta_C}$.

The generator for the Gumbel copula is $\varphi_{\theta_G}(x) = (-\ln x)^{\theta_G}$. For $\theta_G \geq 1$ (θ_G =1 for independence and $\theta_G \to \infty$ for increased dependence), the CDF and PDF for the Gumbel copula are

(3.11) $$C_t^{Gumbel}(u_t, v_t; \theta_G) = \exp\{-[(-\ln u_t)^{\theta_G} + (-\ln v_t)^{\theta_G}]^{1/\theta_G}\}$$

(3.12) $$\begin{aligned} c_t^{Gumbel}(u_t, v_t; \theta_G) &= \frac{\exp\{-[[-\ln(u_t)]^{\theta_G} + [-\ln(v_t)]^{\theta_G}]^{1/\theta_G}\}}{u_t v_t [[-\ln(u_t)]^{\theta_G} + [-\ln(v_t)]^{\theta_G}]^{2-(1/\theta_G)}} \\ &\times [\ln(u_t)\ln(v_t)]^{\theta_G - 1}\{[[-\ln(u_t)]^{\theta_G} + [-\ln(v_t)]^{\theta_G}]^{1/\theta_G} + \theta_G - 1\} \end{aligned}$$

Kendall's τ for the Gumbel copula is $1 - 1/\theta_G$. The upper tail dependence is $\lambda_U^{Gumbel} = 2 - 2^{1/\theta^G}$ and the lower tail dependence is $\lambda_L^{Gumbel} = 0$.

While the Clayton and Gumbel copulas only consider asymmetric cases, the symmetrized Joe-Clayton (SJC) copula is regarded as a more comprehensive Archimedean copula that takes into consideration both symmetric and asymmetric cases, as defined in Patton (2006). The SJC copula is based on the Joe-Clayton (JC) copula, as it is derived by taking a particular Laplace transformation of Clayton's copula. Thus, the

SJC copula is a more comprehensive specification model than the JC copula. SJC copulas have two parameters, τ^U and τ^L, which are measures of dependence known as upper tail dependence and lower tail dependence, respectively. While an SJC copula nests symmetry as a special case, by constructing it symmetrically when $\tau^U = \tau^L$, it does not impose symmetric dependence on the variables as do the Normal copula and the Student's t copula.

The CDF and PDF for an SJC copula are:

(3.13) $C_t^{SJC}(u_t, v_t; \tau^U, \tau^L)$

$$= \frac{1}{2}(C^{JC}(u_t, v_t \mid \tau^U, \tau^L) + C^{JC}(1-u_t, 1-v_t \mid \tau^U, \tau^L) + u_t + v_t - 1)$$

where $C_t^{JC}(u_t, v_t; \tau^U, \tau^L) = 1 - \left(1 - \frac{1}{\left(\frac{1}{(1-(1-u_t)^k)^\gamma} + \frac{1}{(1-(1-v_t)^k)^\gamma} - 1\right)^{1/\gamma}}\right)^{1/k}$

with $k = 1/\log_2(2-\tau^U),\ \ \gamma = -1/\log_2(\tau^L)\ \ and\ \ \tau^U\ \&\ \ \tau^U \in (0,1)$

(3.14) $c_t^{SJC}(u_t, v_t; \tau^U, \tau^L) = \frac{\partial^2 C_t^{SJS}}{\partial u_t \partial v_t}$

$$= \frac{1}{2}\left(\frac{\partial^2 C^{JC}(u_t, v_t \mid \tau^U, \tau^L)}{\partial u_t \partial v_t} - \frac{\partial^2 C^{JC}(1-u_t, 1-v_t \mid \tau^U, \tau^L)}{\partial u_t \partial v_t}\right)$$

where $\frac{\partial^2 C^{JC}(u_t, v_t \mid \tau^U, \tau^L)}{\partial u_t \partial v_t} = A - B$

$$A = \frac{\gamma \cdot k(1-1/Z^{1/\gamma})^{-2+1/k}(1/\gamma - 1)(1-u_t)^{k-1} \cdot (1-v_t)^{k-1}}{(1-(1-u_t)^k)^{1+\gamma} \cdot (1-(1-v_t)^k)^{\gamma+1} \cdot Z^{2+1/\gamma}}$$

$$B = \frac{k(1-1/Z^{1/\gamma})^{-2+1/k}(1/k - 1)(1-u_t)^{k-1} \cdot (1-v_t)^{k-1}}{(1-(1-u_t)^k)^{1+\gamma} \cdot (1-(1-v_t)^k)^{\gamma+1} \cdot Z^{2+2/\gamma}}$$

with $Z = \frac{1}{(1-(1-u_t)^k)^\gamma} + \frac{1}{(1-(1-v_t)^k)^\gamma} - 1$

Bivariate Dynamic Copula Functions

While the static copula function models estimate the time invariant parameters of each model based on the unconditional copula theory, the dynamic copula model investigates the time-varying conditional dependence. This research attempts to capture time-varying dependence by allowing the parameters of copulas to evolve over time. Time variation in the conditional first and second moments of financial data are reported in many research studies and may be explained by diverse GARCH models. In this study, I used four time-varying copula functions. While time-varying Gaussian copula can be used to capture the symmetric behaviors of linear correlation, time-varying Clayton and Gumbel copulas may be used to capture the asymmetric behavior of lower or upper tail dependence. And, time-varying SJC copula models can reflect the symmetric and asymmetric tail dependence for both the upper and lower tails.

Patton (2006) proposed the several time-varying copula models utilizing a parametric function of transformations of the lagged data and an autoregressive term. I followed the functional form of the evolution equation suggested by Patton (2006) and consider four types of dynamic copula models, which correspond to four forms of static copula models. The evolution model for the dynamics of the correlation for a bivariate Normal (Gaussian) copula model is:

(3.15) $$\rho_t = \Lambda_1\left(\omega_{\rho,1} + \omega_{\rho,2}\rho_{t-1} + \omega_{\rho,3} \cdot \frac{1}{10}\sum_{j=1}^{10}\Phi^{-1}(u_{t-j}) \cdot \Phi^{-1}(v_{t-j})\right)$$

where ρ_t is the time-varying correlation coefficient, constrained within the interval (-1, 1), $\Phi^{-1}(\cdot)$ is the inverse of the standard univariate normal CDF, and $\Lambda_1(x)=(1-e^{-x})(1-e^{-x})^{-1}$ is the modified logistic transformation, designed to keep ρ_t constrained within (-1, 1) at all times.

The evolution equations for a time-varying Clayton copula model and a time-varying Gumbel copula model are:

$$\tau_t = \Lambda_3\left(\omega_{\tau,1}+\omega_{\tau,2}\tau_{t-1}+\omega_{\tau,3}\cdot\frac{1}{10}\sum_{j=1}^{10}|u_{t-j}-v_{t-j}|\right) \tag{3.16}$$

where τ_t is the time-varying Kendall tau coefficient , and $\Lambda_3(x)=(1+e^{-x})^{-1}$ is the modified logistic transformation, designed to keep the parameters within the interval (0, 1).

$$\tau_t = \Lambda_3\left(\omega_{\tau,1}+\omega_{\tau,2}\tau_{t-1}+\omega_{\tau,3}\cdot\frac{1}{10}\sum_{j=1}^{10}|u_{t-j}-v_{t-j}|\right) \tag{3.17}$$

where τ_t is the time varying Kendall tau coefficient , and $\Lambda_3(x)=1+\exp(x)$ is the modified logistic transformation, designed to keep the parameters in (0, 1).

The evolution equations of the upper and lower tail dependences for a time-varying SJC copula model are:

$$\begin{aligned}\tau_t^U &= \Lambda_4\left(\omega_{U,1}+\omega_{U,2}\tau_{t-1}^U+\omega_{U,3}\cdot\frac{1}{10}\sum_{j=1}^{10}|u_{t-j}-v_{t-j}|\right)\\ \tau_t^L &= \Lambda_4\left(\omega_{L,1}+\omega_{L,2}\tau_{t-1}^L+\omega_{L,3}\cdot\frac{1}{10}\sum_{j=1}^{10}|u_{t-j}-v_{t-j}|\right)\end{aligned} \tag{3.18}$$

where τ_t is the time varying tail dependence , and $\Lambda_4(x) = (1+e^{-x})^{-1}$ is the modified logistic transformation, used to keep the parameters within the interval (0, 1) at all times.

All of the evolution equations contain an autoregressive term and a forcing variable for a time varying limit probability over the previous 10 observations. The unconditional correlation from the static Gaussian copula, and the upper and lower tail dependences from the static SJC copula are used as initial values for equations (3.18). Especially, we expect that this distance measure in the SJC evolution equation would be inversely related to the concordance ordering of the copulas; under perfect positive dependence it would equal zero, under independence it would equal 1/3, and under perfect negative dependence it would equal 1/2.

Estimation Methodology for Optimal Hedge Ratios

In this study, three categorical models, the Vector Error Correction (VEC), Multivariate GARCH (MGARCH), and static and dynamic copula-based GARCH (CGARCH) models, were employed to estimate optimal hedge ratios. The VEC model estimates a constant hedge ratio whereas time-varying optimal hedge ratios are calculated using MGARCH and CGARCH models. Before detailing the model specifications for each of three categorical models, I discuss hedge ratios and hedge-effectiveness.

In portfolio theory, hedging with futures can be regarded as a portfolio selection issue and futures may be chosen as one of the assets in the portfolio to minimize the overall risk or to maximize utility function. From this perspective, hedging with an ETF may be considered as an alternative asset in portfolio selection. The optimal hedge ratio may be defined as the ratio of futures holdings to a spot position that minimizes the total risk of the hedged portfolio. In this research, traders use futures or ETFs as hedging instruments of the spot market. Let $pf_{Fu,t}$ and $pf_{ETF,t}$ represent the return of a portfolio with futures and the return of a portfolio with ETFs, respectively. Therefore, each portfolio return may be given as follows:

(3.19) $pf_{Fu,t} = sp_t - \delta_{Fu} fu_t$

(3.20) $pf_{ETF,t} = sp_t - \delta_{ETF} etf_t$

where sp_t is the return of the spot price, which is derived from the first difference of the natural logarithm of spot price, fu_t is the return of futures prices, δ_{Fu} is the optimal hedge ratio of futures, and δ_{ETF} is the optimal hedge ratio of ETFs.

As a tool for evaluating hedge performance, the variances of the hedged portfolios, in both cases are:

(3.21) $Var(pf_{Fu,t}) = Var(sp_t - \delta_{Fu} fu_t)$

(3.22) $Var(pf_{Fu,t}) = Var(sp_t - \delta_{ETF} etf_t)$

In this study, using the calculated optimal hedge ratio, I compared the variance of hedged portfolios to evaluate the hedging effectiveness. The hedging effectiveness of a dynamic hedge model can be evaluated by the proportional reduction in variance of the hedged portfolio in comparison to that of the hedged position of a static VEC model.

Hedge ratio calculation depends on how we define the multivariate relationship between spot return and futures return. Therefore, the specific defining hedge ratio is presented for each estimation methodology. Time-varying minimum variance hedge ratios are also obtained using two different frameworks of dynamic econometric models; a multivariate GARCH model (MGARCH) and a dynamic copula-based GARCH model (CGARCH) model.

V*ector Error Correction Model*

As time invariant optimal hedge ratio models, the conventional Ordinary Least Squares Model (OLS) and Vector Autoregressive Model (VAR) do not take into consideration the possibility of long term integration between spot and futures returns, which are, in fact, widely observed in the relationship between spot and futures returns. Therefore, if two prices are co-integrated in the long run, then the Vector Error Correction Model (VEC) is more appropriate (Lien and Luo 2006). If the futures and spot series are co-integrated on the order of one and each series is not dependent on its autoregressive value, the VEC model of the spot and futures series is given as:

(3.23) $$\begin{aligned} sp_t &= \alpha_{0s} + \alpha_{1s} ECT_{t-1} + \varepsilon_{s,t} \\ fu_t &= \alpha_{0f} + \alpha_{1f} ECT_{t-1} + \varepsilon_{f,t} \end{aligned}$$

where ECT_{t-1} is the error correction term, which is a function of $Sp_{t-1} - \lambda_0 Fu_{t-1} +$ *Constant*. Sp_{t-1} and Fu_{t-1} are the spot and futures prices, respectively.

I utilized a bivariate error correction model to obtain the static hedge ratio, δ_{Fu}, following Kroner and Sultan (1993). They proposed the following bivariate error-correction model for sp_t and fu_t. The error terms in the equations, $\varepsilon_{s,t}$ and $\varepsilon_{f,t}$, are independently identically distributed (*iid*) random vectors. The minimum variance hedge ratio for futures hedging is calculated as:

(3.24) $$\delta_{Fu}^{ECM} = \frac{Cov(\varepsilon_{s,t}, \varepsilon_{f,t})}{Var(\varepsilon_{f,t})}$$

Multivariate GARCH Model

A multivariate GARCH model enables the conditional covariance matrix of the dependent variables to follow a flexible dynamic structure and allow the conditional mean to follow a vector autoregressive or vector error corrections structure. In the mean equation, the non-stationarity and cointegration relationship has been observed in previous research on spot and futures prices of oil commodities.

I used a bivariate error correction model as a mean equation, which is the same as equation (3.23) in the VEC model. Kroner and Sultan (1993) proposed the following bivariate error correction model for sp_t and fu_t with a constant conditional correlation (CCC) GARCH (1, 1) structure for the estimation of hedge ratios:

(3.25) $$\begin{aligned} sp_t &= \alpha_{0s} + \alpha_{1s} ECT_{t-1} + \varepsilon_{s,t} \\ fu_t &= \alpha_{0f} + \alpha_{1f} ECT_{t-1} + \varepsilon_{f,t} \end{aligned}$$

where ECT_{t-1} is the error correction term, which is a function of $Sp_{t-1} - \lambda_0 Fu_{t-1} +$ *Constant*. Sp_{t-1} and Fu_{t-1} are the spot and futures prices, respectively.

(3.26) $$\begin{bmatrix} \varepsilon_{s,t} \\ \varepsilon_{f,t} \end{bmatrix} | I_{t-1} \sim N(0, H_t)$$

where $\varepsilon_{s,t}$ and $\varepsilon_{f,t}$ are error terms following the GARCH (1,1) model with a zero mean and a conditional covariance matrix H_t with a constant correlation ρ.

$$(3.27)\quad H_t = \begin{bmatrix} h_{s,t}^2 & h_{sf,t} \\ h_{sf,t} & h_{s,t}^2 \end{bmatrix} = \begin{bmatrix} h_{s,t} & 0 \\ 0 & h_{f,t} \end{bmatrix} \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \begin{bmatrix} h_{s,t} & 0 \\ 0 & h_{f,t} \end{bmatrix}$$

$$(3.28)\quad \begin{aligned} h_{s,t}^2 &= \beta_{0s} + \beta_{1s}\varepsilon_{s,t-1}^2 + \beta_{2s}h_{s,t-1}^2 \\ h_{f,t}^2 &= \beta_{0f} + \beta_{1f}\varepsilon_{f,t-1}^2 + \beta_{2f}h_{f,t-1}^2 \end{aligned}$$

where $h_{s,t}^2$ and $h_{f,t}^2$ are the conditional variance for the spot and futures returns, which follows the GARCH (1, 1) model.

Since the assumption of constant correlation may be too restrictive to fit reality, I adopted the DCC MGARCH model based on the multivariate normal distribution to remove this restriction and improve the flexibility of the hedging models. In contrast to the CCC MGARCH model, the DCC MGARCH model allows a time-varying correlation ρ_t:

$$(3.29)\quad H_t = \begin{bmatrix} h_{s,t} & 0 \\ 0 & h_{f,t} \end{bmatrix} \begin{bmatrix} 1 & \rho_t \\ \rho_t & 1 \end{bmatrix} \begin{bmatrix} h_{s,t} & 0 \\ 0 & h_{f,t} \end{bmatrix}$$

where $h_{i,t}$ is the diagonal matrix of the conditional standard deviation matrix from the univariate GARCH model, ρ_t satisfies:

$$(3.30)\quad \rho_t = (1-\lambda_1-\lambda_2)\rho + \lambda_1\rho_{t-1} + \lambda_2\tilde{\varepsilon}_{i,t-1}\tilde{\varepsilon}_{j,t-1}$$

where $\tilde{\varepsilon}_t (= \hat{\varepsilon}_t / h_t^2)$ is the standardized disturbance vector, ρ is the unconditional correlation of the standardized residual ($\tilde{\varepsilon}_t$), and λ_1 and λ_2 are parameters that govern the dynamics of a conditional quasi-correlation. λ_1 and λ_2 are nonnegative and satisfy

$0 \le \lambda_1 + \lambda_2 < 1$.

Engle (2002) proposed the use of a two-step maximum likelihood method for the estimation of the parameters for the DCC MGARCH model. In the first step, the parameters in the univariate GARCH models are estimated for each residual series. As a second step, the parameters of the dynamic correlation are estimated using the results of the first step and the transformed residuals $\hat{\hat{\varepsilon}}_{i,t} = \hat{\varepsilon}_{i,t} / h_{i,t}^2$.

Given the estimates $\hat{H}_t$ obtained in the DCC MGARCH models, the optimal dynamic hedge ratio in the case of futures hedging is estimated by:

(3.31) $$\delta_{Fu,t}^{DCC} = \frac{h_{sf,t}}{h_{f,t}^2} = \rho_t \frac{h_{s,t}}{h_{f,t}}$$

where $h_{sf,t}$ is the conditional covariance between the errors $\varepsilon_{s,t}$ and $\varepsilon_{f,t}$, $h_{s,t}^2$ and $h_{f,t}^2$ are the conditional variances of the errors $\varepsilon_{s,t}$ and $\varepsilon_{f,t}$, respectively, and, ρ_t is the conditional correlation.

Regarding the bivariate case of spot and ETF returns, the etf_t, ETF return at period *t*, is used in place of fu_t in equations (3.25) - (3.31). Generally, returns for financial data in time series possesses a time-varying hetroskedastic volatility structure. Due to this ARCH effect (Bollerslev, Engle, and Wooldridge 1988) on the returns of spot and futures prices and their time-varying joint distribution, the simple estimation of hedge ratios and hedge effectiveness using the conventional OLS method may be inappropriate. Therefore, I modeled time-varying hedge ratios based on a conditional bivariate GARCH of unexpected returns, exploring a variety of GARCH (1,1) parameterizations such as DCC MGARCH(1,1) (Engle 2002) and copula MGARCH(1,1) (Hsu, Tseng, and Wang 2008). Alexander and Barbosa (2007) define the dynamic hedge ratio as the minimum variance hedge ratio for each day, which determines the position to be taken at the end of the day through the following day.

Copula-based GARCH Model

Though MGARCH models consider the time-varying characteristics of hedge ratios in their dynamic frameworks, an important limitation of MGARCH models is the typical assumption of joint multivariate normality. Methodological developments on defining diverse joint distributions by the copula function enable it to reflect the asymmetry of the dependence structure. The use of diverse copula functions including asymmetric equity correlation and financial contagion, and the calculation of the Value at Risk (VaR) for a portfolio of assets has been actively studied. In addition, a recent study on time-varying copula models has presented time variation as a conditional dependence without the restriction of the multivariate normal distribution assumption.

Marginal distribution and copula function may be defined in two ways, parametrically or non-parametrically. I choose the parametric approach for both marginal distribution and copula function, based on the statistics for each data series; the proposed hedging model uses the GARCH-Student’s t specification for marginal distributions in the first step, and four static copula functions (Gaussian, Clayton, Gumbel, and SJC copulas) and four dynamic copula functions (time-varying Gaussian, time-varying Clayton, time-varying Gumbel, and time-varying SJC copulas) for joint distributions, to allow for a wide range of possible dependence structures in the second step.

First, I specify the conditional marginal density for spot and futures returns of gasoline prices using a Student's t distribution and GARCH (1,1) framework, defined by

$$\begin{aligned} & sp_t = \alpha_{0s} + \alpha_{1s} ECT_{t-1} + \varepsilon_{s,t} \\ (3.32) \quad & \varepsilon_{s,t} \mid I_{t-1} \sim t(0, h_{s,t}^2; d_s) \\ & h_{s,t}^2 = \beta_{0s} + \beta_{1s} \varepsilon_{s,t-1}^2 + \beta_{2s} h_{s,t-1}^2 \end{aligned}$$

$$\begin{aligned} & fu_t = \alpha_{0f} + \alpha_{1f} ECT_{t-1} + \varepsilon_{f,t} \\ (3.33) \quad & \varepsilon_{f,t} \mid I_{t-1} \sim t(0, h_{f,t}^2; d_f) \\ & h_{f,t}^2 = \beta_{0f} + \beta_{1f} \varepsilon_{f,t-1}^2 + \beta_{2f} h_{f,t-1}^2 \end{aligned}$$

where ECT_{t-1} is the error correction term with unconditional correlation, which is a function of $Sp_{t-1} - \lambda_0 Fu_{t-1} + \text{Constant}$. The two individual marginal equations are the same as the mean equations of the VEC and MGARCH models. $\varepsilon_{s,t}$ and $\varepsilon_{f,t}$ are the error terms following Student's t distribution with degrees of freedom, d_s and d_f, respectively. $h_{s,t}^2$ and $h_{f,t}^2$ are the conditional variances for the spot and futures returns of gasoline.

Second, I selected some appropriate copula functions for the dependence structure of standardized spot and futures innovations. The standard for judging which copula function is a good fitted consists of the log likelihood values followed by the Akaike Information Criterion (AIC) and Bayes Information Criterion (BIC). Copula parameters were estimated by optimizing the log-likelihood functions; however, there are difficult to optimize when the number of model parameters is large. As the dimensions of the estimated equation may be quite large, it is difficult in practice to achieve a simultaneous maximization of log-likelihood functions for all of the parameters. To effectively solve this problem, the two-stage estimation procedure of Joe and Xu (1996) that was adopted by Patton (2006) and Bartram, Taylor, and Wang (2007) is followed. This method, called inference for the margins (IFM), in which the marginal densities and copula density can be estimated separately partially resolves the problem. Joe and Xu (1996) also demonstrated the high efficiency of the easily-implemented IFM method, compared with the customary maximum likelihood method.

In the first stage, the parameters of the marginal distribution were estimated from the univariate time series by:

$$\begin{aligned} \hat{\theta}_s &\equiv \arg\max \sum_{t=1}^{T} \ln f_{s,t}(sp_t \mid I_{t-1}; \theta_s) \\ \hat{\theta}_f &\equiv \arg\max \sum_{t=1}^{T} \ln f_{f,t}(fu_t \mid I_{t-1}; \theta_f) \end{aligned} \tag{3.34}$$

In the second stage, the marginal CDFs are applied to the standardized residuals, using the estimates from (34), to provide estimates of the probabilities u_t and v_t, which are then used to estimate the copula parameters by:

$$\hat{\theta}_c \equiv \arg\max \sum_{t=1}^{T} \ln c_t(u_t, v_t \mid I_{t-1}; \theta_c) \tag{3.35}$$

After estimating the parameters in the different copula-based GARCH (CGARCH) models, the conditional variances, $h_{s,t}^2$ and $h_{f,t}^2$, were obtained from the equations (3.32) and (3.33), and the unconditional dependence or conditional linear dependence was generated by the evolving equations of the dynamic copula, (3.15) –

(3.18). The optimal dynamic hedge ratio in the case of futures hedging was calculated from equation (3.31). The proposed hedge ratios of dynamic CGARCH models consider time-varying dependence and asymmetric specifications in the joint and marginal distribution of assets.

Data and Empirical Results

This section examines the basic hedge model for gasoline spot prices and futures prices, and the alternative hedge model for gasoline spot prices and a gasoline ETF (UGA). The data set includes daily spot prices for unleaded gasoline, daily prices for gasoline in a New York Mercantile Exchange (NYMEX) futures contract, and daily close prices of the UGA ETF trading on the New York Stock Exchange (NYSE). Considering the timing of the launch of the United States Gasoline Fund (NYSE Arca: UGA) in the oil ETFs market, on February 26, 2008, and the gasoline price breakpoints from the 2008 financial crisis, the three data spans the period from March 2008 to December 2011. Spot prices for gasoline were obtained from Los Angeles Reformulated RBOB Regular Gasoline Spot Price in the U.S. Energy Information Administration. For the gasoline futures price, I utilized the daily settlement price of futures contracts on unleaded gasoline delivered to the New York harbor and traded on the NYMEX that were near month contracts set to expire. The daily close price of the United States Gasoline Fund (NYSE Arca: UGA) on the NYSE was used as a representation of ETF prices. Each data set was obtained from the U.S. Energy Information Administration (EIA), NYMEX, and NYSE data sources, respectively.

Basic statistics on level data and first difference data are presented in Table B-9. In three series the return data in the first difference of log prices exhibited skewness and kurtosis that was greater than in the normal distribution. After performing the Jarque-Berra test for normality of distribution, all three series were significantly different from a normal return distribution. In order to examine the relationships among spot price returns, futures returns, and ETF returns, correlations and Kendall's Tau were computed in Table B-9.

In the current research, the unit root behavior of a series was investigated; the Augmented Dickey-Fuller (ADF) and Kiwatkowski-Phillips-Schmidt-Shin (KPSS) tests were conducted to measure the unit root and first difference of log data. These two tests were utilized to investigate the opposite null hypotheses: the ADF test used the unit test as the null hypothesis and the KPSS test used the stationarity test as the null hypothesis. Table B-10 shows the unit root test results. Based on the KPSS test, a basically level data series indicates that the stationarity hypothesis is rejected at the 1% significance level, but the stationarity hypothesis cannot be rejected by the first difference data. The results of the ADF test are consistent with the KPSS test results, which indicates that all log data are unit root.

Based on the unit root test result, the cointegration relationships were examined. Cointegration refers to the fact that two or more series share a stochastic trend. Engle and Granger (1987) suggested that a two-step process be used to test for cointegration, the EG-ADF test. In Table B-11, the EG-ADF test results cause us to reject the hypothesis of no cointegration at the 1% significance level in both cases. This cointegration test is confirmed by the criteria of the Johansen Maximum Likelihood (ML) Test. The null hypothesis that no less cointegration equations existed than at the maximum rank level was tested in two systems of equations, first spot and futures prices, and, second, spot and ETF prices. In the first set of equations, for spot and futures prices, the trace statistic of 2.0846 at the maximum rank level 1 does not exceed the critical value of 3.76 of the 5% significance level. Thus, I could reject the null hypothesis that there is one or fewer cointegration equations. Since Johansen's method for estimating rank entails accepting the first rank at which the null hypothesis is not rejected, I accepted rank 1 as an estimate of the number of cointegrating equations for these two variables. In the case of the second set of equations, for spot and ETF prices, the null hypothesis of one or fewer cointegration equations was not rejected at the 5% significance level. Based on the cointegration test results, I added the error correction term in the mean equation in the MGARCH and CGARCH models.

Estimation Results of VEC and MGARCH Models

In this research, the data sample periods ranged from March 2008 to December 2011. However, if a structural break occurred within a sample period, an empirical estimation using the entire sample might fail to provide reliable results (Clements and Hendry 2006). As the global financial crisis in 2008 occurred during the sample periods, taking into account any changes in the financial environment would be desirable for parameter stability. To explain the major changes occurring during the financial crisis and identify other unknown changes within the sample periods, the model coefficient should be flexible for one, or several, dates. And, since the actual dates were unknown, I had to estimate them as well as the model parameters.

In this study, I adopted Zivot and Andrews's model (1992) to determining these unknown dates for structural breaks. Based on the results of the unknown structural break test, two structural break points were found; one on September 20, 2008 and the other on April 29, 2009. Based on the unknown structural break test results and causal relationship changes given in Table B-12, the entire sample period was divided into two-sub groups; the 1st period from March 2008 to April 2009, and the 2nd period from May 2009 to December 2011. Based on these sub groups, the results of the VEC (Vector Error Correction) model estimation are presented in Table B-13. These two periods have strictly different data characteristics; the 1st period includes the majority of the duration of the 2008 financial crisis, in which highly volatile price movements were experienced, and the 2nd period covers the post-financial crisis period having relatively stable price volatility.

For my first dynamic model, I applied the MGARCH model, which allowed for the conditional variance to be dynamic. Next, it was important to define the correlation in a multivariate model; however, the general MGARCH model was too flexible. Among the following diverse models, the diagonal vech model (DVECH), the constant conditional correlation model (CCC), the dynamic conditional correlation model (DCC), and the time-varying conditional correlation model (VCC), I utilized the DCC

MGARCH model, in which the conditional variances are modeled as univariate generalized autoregressive conditionally hetroskedastic models and conditional covariances are modeled as nonlinear functions of the conditional variances. I used the DCC MGARCH model to investigate the dynamic interaction between spot gasoline prices and futures prices, and between gasoline spot prices and ETF prices. Thus, the mean equation with an error correction term (ECT) was applied to investigate the causality relationship among variables and a DCC MGARCH model was used to take into account the hetroskedastic properties of the variances and covariances of variables. I first used the bivariate DCC MGARCH (1, 1) model with an error correction term (ECT) to probe the hedging effects by futures prices in their first and second moments. This model may be identified as a VEC DCC MGARCH model. Second, another bivariate VEC DCC MGARCH (1, 1) model was estimated for gasoline spot prices and ETF prices. The parameter estimation and test statistics for the VEC and VEC DCC MGARCH model are presented in Table B-13 and Table B-14. The results of the parameter estimation from the VEC are used again in the VEC DCC MGARCH model.

In addition, in the top graphs of Figure A-10, the conditional correlations of spot returns and futures returns for the two sub periods are compared. In the bottom graphs of Figure A-10 the conditional correlations of spot returns and ETF returns are exhibited. The straight line represents the conditional correlation of the CCC MGARCH model and the time-varying correlation line represents the conditional correlation of the DCC MGARCH model. The conditional correlations of the CCC model in the 2nd period are lower than those in the 1st period in the cases of both the futures and ETF hedging.

Comparing the conditional correlations between futures and ETF hedging in the CCC MGARCH model, the estimated results for ETF hedging (1st: 0.8191 and 2nd: 0.7660) are higher than those in the case of futures hedging (1st: 0.7606 and 2nd: 0.7228).

Concerning the estimated results of the DCC MGARCH model, the conditional correlation between spot returns and futures returns exhibits very similar movement to that between spot returns and ETF returns in both periods. As the dynamic correlations change strongly in both futures hedging and ETF hedging models, the DCC model is more appropriate than the CCC model, which assumes the correlation is constant. In Figure A-11, the conditional variances of the DCC MGARCH model are compared for the two sub periods. The conditional variances of the three series in the 1st period are significantly high as a result of the financial crisis in 2008. Based on the estimated conditional variance and correlation, the time-varying hedge ratios are exhibited in Figures A-12 and A-13.

Estimation Results of Copula GARCH Model

Following the inference for the margins (IFM), I first estimated the marginal GARCH (1,1) Student's t model. The independently and identically distributed (iid) standardized residuals obtained from the marginal assumption were transformed to a uniform (0, 1) by a probability integral transformation (PIT), prior to the estimation of the copula parameters. In order to find the dependence structure for spot and futures innovations, I estimated four static and four dynamic copula specifications. The results of these estimations are presented in Table B-15. The three time-varying copula models (time-varying Normal, time-varying Gumbel and time-varying SJC copulas) commonly exhibit a better fit than static copula functions. Comparing the 1st and 2nd periods for the case of futures hedging, Table B-15 shows that all time-varying copula functions for futures hedging exhibit better goodness of fit than the static copula functions. In the 1st period, the time-varying Gumbel copula exhibited the best goodness of fit, followed by the time-varying SJC and time-varying Gaussian copulas, according to ranking. In the 2nd period, the time-varying SJC copula exhibited the best fit, followed by the time-varying Gumbel and time-varying Gaussian copula functions.

As the SJC copula function can explain both the lower and upper tail dependences for symmetric and asymmetric cases, its features differ from the Clayton and Gumbel copulas; it can analyze the dynamic patterns of lower and upper tail dependences at the same time as it provides the dependence relationship for the given time periods. The SJC copula was developed from the Joe-Clayton (JC) copula. While the JC copula explains asymmetric tail behavior for upper and lower tail dependences, the SJC copula nests symmetry as a special case. In the SJC copula, a static copula, the parameters of the upper and lower tail dependences in the 1st period are estimated as 0.6581 and 0.6736, respectively, and the upper and lower tail dependences in the 2nd period are estimated as 0.6061 and 0.6487, respectively. From a static point of view, the upper tail dependence is more correlated than the lower tail dependence in both of the two sub-periods. However, when these dependences are revisited in terms of a time-varying copula, the results can be interpreted differently. In Figure A-14, while the time-varying movements between the lower and upper tail dependences are very similar in the

2nd period, the time-varying dependence of the upper tail fluctuates a lot, but the lower tail dependence shows similar dynamic movements to the static copula estimation results for the 1st period.

I also compare the copula functions in the case of ETF hedging, in Table A-16. The time-varying copula functions all provide a better fit than their corresponding static functions. Based on the results of Tables B-15 and B-16, the three static and three dynamic copula functions were selected to derive the static and dynamic optimal hedge ratios. For the copula-based GARCH models, Panel A of Table B-17 presents the estimates of parameters for the conditional means, variance, and marginal distribution. Panels B through D of Table B-17 present the estimates of parameters for the three dynamic copula functions. Estimation result in case of ETF hedging is summarized in Table B-18.

Comparisons of Hedging Performances

The hedge ratio was derived based on the estimation results. First, regarding the hedge ratios of the 1st sub-period for the case of futures hedging, the static hedge ratios of the VEC model is 0.8092, and those of three static copula are 0.8797 (Normal copula), 0.8094 (Gumbel copula), and 0.7627 (SJC copula) respectively. In addition, the dynamic hedge ratios, which were derived from the DCC MGARCH and three dynamic copula models, have time-varying values using a conditional information-based approach. The averages of the time-varying hedge ratios for the DCC MGARCH, time-varying Normal, time-varying Gumbel, and time-varying SJC copulas were 0.7870, 0.7650, 0.7106, and

0.8093, respectively. In Figures A-12 and A-13, the dynamic changes of these four time-varying optimal hedge ratios are presented together with the basic constant hedge ratio of the VEC model.

Based on the derived values for optimal hedge ratios, I evaluated the hedging performance of the different models. A hedge portfolio is composed of a spot asset and δ units of futures; for purposes of comparison, the variances of the returns of these portfolios over the two sub-sample periods were calculated and presented in Table B-19. The hedging effectiveness of the DCC model, three static copulas, and three dynamic copulas are reported in the form of variance reduction over the VEC measure, which is used as a benchmark. Some of the other findings are summarized as follows.

First, regarding the percentage of variance reduction occurring during the 1st sub-sample period, a static hedging strategy (VEC and three static copula strategies) was better than a dynamic hedging strategy (DCC and three time-varying copula strategies) in both cases of hedging. This can be explained by the fact that the dynamic models used to capture time-varying trends did not effectively reflect the strong volatility and unexpected price movement actually observed in the 1st sub-sample period data, which included highly volatile data from the financial crisis of 2008. Second, concerning the relatively stable 2nd sub-sample period, most of the dynamic hedging strategies with specific time-varying copulas exhibited better hedging performance. In the case of futures hedging, the time-varying Gumbel copula outperformed the VEC and other static copula models, while the time-varying Normal copula performed much better than the VEC and other static copula models, in the case of ETF hedging. Third, regarding the comparison of hedging effectiveness for futures and ETFs, the average of all portfolio variance from ETFs is lower than the average variance in a futures portfolio. This supports the possibility that ETF's hedging capability is as good as that for futures. Fourth, while asymmetric copula-based models, such as the Gumbel and SJC copulas, perform more effectively in futures hedging, the symmetric copula-based models perform better in ETF hedging.

Recent studies on dynamic hedging models with DCC or time-varying copulas have shown better fitness and more effective hedging performances (Hsu, Tseng, and Wang 2008; Liu, Jian, and Wang 2010). However, dividing the sample period into two sub-samples, the crisis and post-crisis periods, provides an additional explanation for the effectiveness of dynamic hedging models, signifying that dynamic hedging models might also show better hedging performances in normal data periods, even though they have a limited ability to explain highly volatile data periods. This can be explained by the fact that as all dynamic hedging models assume an evolutionary pattern of dependence based on autoregressive information, the conditional information used in the dependence evolution does not effectively explain the frequent unexpected price movements during a crisis period.

In normal data periods, like the post-crisis period, the copula-based GARCH models provide the best performance in terms of both futures and ETF hedging. By specifying a joint distribution as spot and futures or spot and ETF returns with full flexibility, the copula-based GARCH models can be used to effectively reduce risk in hedged portfolios. In addition, ETF-hedged portfolios can be an effective alternative to futures-hedged portfolios in regard to gasoline spot returns.

Several limitations of this study are noteworthy. First, as the data used in this study reflects the launch of the UGA ETF in February 2008, the data spans a period of only 45 months, from March 2008 to December 2011. This relatively short period of data collection offers limited possibilities in generalizing the results of this research. Therefore, further research, covering a greater data collection period, would be helpful in analyzing the hedging performance for an optimal dynamic hedging ratio approach. Second, in this study, hedging performance was evaluated by the minimum variance hedging model (Ederington 1979). However, there are other ways to evaluate hedge performance, such as the certainty equivalent (CE), derived from an exponential utility for hedging. This approach considers not only risk but also risk averseness in measuring hedge effectiveness and may be included in additional further research using a dynamic hedging approach. Third, unlike futures hedging with specific delivery objectives, ETF hedging is a conceptual approach that models ETF price movements for the purpose of hedging gasoline spot prices. Therefore, as various put or call options for ETFs have been widely used, these option values should be added to calculate the variance or return of an ETF-hedged portfolio. Therefore, further research on hedging models using data collected over longer periods and the use of diverse dynamic hedging models would all be helpful in examining the hedging performance of various investment instruments in the diverse commodity futures and ETF market.

Summary and Concluding Remarks

This study employs four categories of models to evaluate the optimal hedge ratio for gasoline returns on spot and futures, and on spot and ETFs. As a basic benchmark system, the Vector Error Correction (VEC) model estimates the static hedge ratio based on the multivariate normal distribution, which does not take into account dynamic changes in the hedge ratio. In order to explain dynamic changes in the hedge ratio, first, a dynamic conditional correlation multivariate GARCH (DCC MGARCH) model is used to estimate the conditional hedge ratio in the market. However, the DCC MGARCH model also assumes multivariate normality. Therefore, static and dynamic CGARCH models are also used to estimate constant and dynamic hedge ratios based on multivariate non-normal distributions. As the dynamic copula approach gives more flexibility to modeling time-varying dependence based on multivariate non-normal distributions, the symmetric and asymmetric joint time-varying relationships of spot and futures returns are analyzed. The analysis framework is also applied to both spot and ETF hedging cases.

Based on structural breaks and causal relationship changes, the data was divided into crisis and post-crisis sub-sample periods. In the post-crisis period, the CGARCH dynamic hedging model exhibited the best hedging performance in comparison to the other models. However, during the crisis period, the VEC or static copula models provided the best hedge ratio for risk reduction in comparison to the other alternative models.

From this research, we can conclude that in analyzing any given data period it is very important to select the correct model to estimate the hedge ratio. Although the dynamic copula approach may have limited power to examine extreme data periods, like that experienced during the 2008 financial crisis, in more stable and normal data periods, the risk exposure of a portfolio may be effectively managed by a dynamic copula model with precise specifications for the joint distribution of assets. In addition, considering the increased interactions among the spot, futures and matching commodity ETF markets, the use of an appropriate hedge model to create a diverse hedged portfolio may have crucial implications for risk management.

CHAPTER IV

CAUSALITY ANALYSIS ON STOCK PRICES AND STOCK MARKET PARTICIPANTS IN CASE OF THE KOREAN STOCK MARKET

Introduction and Background

Causal relationships can be used to investigate the information flows and directions of control in the market. In this study, I used daily Korean stock market data to investigate the contemporaneous causal relationships between the stock price movement and the activities of stock market participants based on Direct Acyclic Graphs (DAGs). After the financial turmoil experienced by East Asian countries circa 1997, the impact of foreign investors in the stock markets of emerging market economies, including the Korean stock market, has been disputed. Choe, Kho, and Stulz (1999) examined the relationship between foreign investor’s trading and stock returns for the Korean stock market around 1997. They found a causal relationship between stock return and the foreign investor trading during the period before the crisis, but did not find any evidence of a linkage between foreign investor trading and stock price movements after the crisis period. In fact, most of the literature concerning the Korean stock market has focused on the causal relationship between the trading volume of foreign investor and returns, or foreign investor trading volume and the volatility of the stock price with data of the 1997 Asian financial crisis (Silvapulle and Choi 1999; Pyun, Lee, and Nam 2000; J. Kim, Kartsaklas, and Karanasos 2005).

However, a limitation of this causal analysis occurred when a legal restriction was placed on foreign investor trading in the Korean stock market, through which the Korean stock exchange placed a ceiling[8] on foreign ownership of individual companies. Through the crisis recovery program of the International Monetary Fund (IMF), this ceiling on foreign ownership in the Korean stock market was lifted completely in May 1998[9] to bring about the full effect of financial liberalizations. In this research, I conducted a causal analysis of the relationship between stock market between stock price movement and investors' trading activities using the Korean stock market from 2005 to 2010, which were expected to reflect the full effect of financial liberalization and also include an important historical event, the global financial crisis in 2008. In addition, I used the trading data of not only foreign investors, but also domestic institutional investors and domestic individual investors. For this study, daily trading data was utilized and sorted according to these three types of investors. Therefore, this study enables a more comprehensive analysis of the dynamics among different market players' activities in the Korean stock market. Furthermore, the data from two Korean stock exchange markets were utilized in this analysis and compared with the causality results. The Korean Stock Exchange (KSE), a market that corresponds to the New York Stock Exchange (NYSE), and the Korea Securities Dealers Automated Quotation

[8] The ownership limit for each individual foreign investor was 5% of a affirm shares until May 2, 1997, when it was increased to 6%. It was then increased to 7% on November 3, 1997, to 50% on December 11, 1997, and to 100% on May 25, 1998. In addition, the aggregate ownership, representing foreign investor as a group, limit on foreign investors increased to 23% on May 2, 1997, to 26% on November 3, 1997, to 50% on December 11, 1997, to 55% on December 30, 1997, and finally to 100% on May 25, 1998.

[9] After the abolishment of the ceiling restriction for foreign investment, the ratio of foreign investment to total market value increased from 18.43% in January 1999 to 40.10% in January 2004.

(KOSDAQ) Exchange, a counterpart of the NASDAQ exchange, were employed in this framework.

The KSE, which was established in 1956, is a well-known stock exchange equipped with most of the hardware and software features common to advanced stock markets. Most of the conventional and large Korean firms are traded on the KSE. The KOSDAQ exchange, a Korean version of the NASDAQ stock market, was formally established in 1996 by gathering less structured and relatively dormant over-the-counter (OTC) stock together to form a more lively market. Although the KOSDAQ exchange only began drawing investor attention gradually after its formal establishment, it grew rapidly with the recovery from the economic crisis of 2000. KOSDAQ firms are generally much smaller and less well-known to the public than those traded on the KSE. In addition, many regulations are looser for KOSDAQ firms than for KSE firms. Thus, more lenient oversight or regulation of the KOSDAQ exchange may lead investors in the KOSDAQ market to pursue greater earning that carry higher risk in comparison to KSE market. Therefore, a causality comparison by investor types for the KSE and KOSDAQ markets may explain the trading pattern for each type of investor in these two markets and also provide implications concerning investment strategies in each market.

In addition, most previous literatures on causal analysis mainly focused on investigating the lagging causality between foreign investor trading and stock price using Granger causality test. However, with the development of information carriers and transaction technology, stock price is more and more sensitive to the information release on market player, including the amount of trading volume, the relative ratio of purchase and sales, trading pattern change, and so on, which can cause stock price fluctuation correspondingly on the date of issue. Additionally, stock price of one country fluctuated severely with stock market of other countries in the same time for the promotion of global market integration. The contemporaneous relationship between stock price and investor trading pattern has become more important in case of the Korean stock market, but fewer researches have worked on it before. Therefore, after the 2008 global financial crisis, researches on change in causal relationship between investor trading and stock price would further an understanding of the Korean stock market and would help to predict stock price movement and one of its driving factors in the market.

This study is organized as follows. After the introduction, the second section reviews previous studies on causal analysis of stock markets, and the third section details the Korean stock market data used in this study and contemporaneous causality analysis methods for the model utilized, the Direct Acyclic Graph (DAG) with Vector Autoregressive Model (VAR) innovations series. The fourth section describes the empirical results of the DAG and discusses the manner in which the contemporaneous causalities of the KSE and KOSDAQ markets change around the time of 2008 financial crisis. In the last section, I present my conclusion.

Literature Review

There are several explanations for the existence of a causal relationship between stock prices, volatility and trading activities. Karpoff (1987) suggested four reasons why the relationship between stock price and trade volume data is important: it provides insight into the structure of a financial market, it is important for event studies that use a combination of price and volume data from which to draw inferences, it is critical to the debate over empirical distribution of speculative prices and it has significant implications for future research studies. There are several measures of stock price, volatility and trading activities. The trading activities of investors are mainly measured in two ways; the daily net purchase (Dornbusch and Park 1995) or the total transaction volume (Brooks 1998; Hiemstra and Jones 1994).

Especially, first, concerning the relationship between stock prices and net purchases by foreign investors, Dornbusch and Park (1995) referred to positive feedback trading in which investor buy when prices increase and sell they fall. This model has also shown that investors who buy when stock prices increase and sell when they decrease can have a destabilizing influence on the stock market. In some models, positive feedback trading leads to both bubbles, where prices depart from fundamentals, and crashes when the bubbles burst. Dornbusch and Park (1995) and Choe, Kho, and Stulz (1999) investigated whether foreign investors engaged in positive feedback trading in emerging country stock markets. In this perspective, researchers have focused on the causal relationship between the lagged information of stock prices and investor trading activities. Second, a number of studies have also examined the relationship between stock returns and total transaction volume. Bohn and Tesar (1996) and Clark and Berko (1997) demonstrated a positive relationship between equity flow from trading and stock returns using monthly data. Froot, O’Connell, and Seasholes (2001) examined the relationship between equity flows and stock index returns using trade data from the institutions and showed a positive feedback trading effect. In the case of the Korean stock market, Silvapulle and Choi (1999) examined the dynamic relationship between daily aggregate Korean stock returns and trading volume. After controlling for volatility persistence in both series and filtering for linear dependence, they found evidence of non-linear bidirectional causality between stock returns and volume series. Choe, Kho, and Stulz (1999) investigated the causal relationship between net purchase of foreign investors and stock returns using with 1996-1997 Korean stock market data and dividing the data into two periods: before the Korean financial crisis and during the crisis.

In examining the relationship between volatility and trading volume, Karpoff (1987) proposed a model which links trading volume and volatility and predicted a positive but asymmetric relationship between trading volume and the absolute value of returns. Concerning volatility, four different measures have been used in previous studies, including the difference between the daily or weekly high and low prices (Alizadeh, Brandt, and Diebold 2002; Gallant, Hsu, and Tauchen 1999), the absolute value of the return series (Saatcioglu and Starks 1998), the squared return series (Brooks 1998) and the conditional variance from a given type of ARCH model (Tse 1998). Some researchers have studied this relationship based on information economics approach, analyzing the impact of information arrival for trading on price changes and price volatility. Some models suggest that trading volume and variance of price changes move together (Karpoff 1987), while another one suggests that there is no relationship between stock price changes and trading volume (Brailsford 2009). Two recent studies have examined the volatility-volume relationship in the Korean stock market. Pyun, Lee, and Nam (2000) examined the relationship between information flows and return volatility for individual companies actively traded in the Korean stock exchange, whereas Kim, Kartsaklas, and Karanasos (2005) investigated the causal relationship between volatility and trade volumes for two market players, such as domestic investors and foreign investors.

Most of the literature on causality analysis of the stock market have centered on the causal relationship between stock returns and trading volume, or between stock volatility and trading volume based on Granger causality, which is the most widely used approach in economics for identifying dynamic causality (Engle and Granger 1987). In addition, most studies of the Korean stock market have focused on the trading data of foreign investors in the KSE market. Since Granger causality is based on the lag relationship inherent in time-series data, it has little to say about contemporaneous causation.

In cases of high frequency financial data, contemporaneous causality explains the pattern of information flow between market participants more effectively. Given the recent developments in information technology (IT), market information now spreads rapidly to all market participants and information asymmetries between domestic and foreign investors has greatly been reduced compared to a decade ago. Also, the amount of intra-day trading has increased with IT development of financial market. Therefore, an analysis of the causal relationship in contemporaneous information transmission mechanism has meaningful implications for the analysis of recent stock market data.

However, to the best of my knowledge, the contemporaneous causality of stock price data and all stock market participants' trading activities has not yet been investigated.

Taking into account the recent market synchronization and the decline of information asymmetries between diverse market participants, this study investigates the contemporaneous causal relationship between stock returns, volatility and the trade activity data for three types of investors: institutional investors, individual investors, and foreign investors, simultaneously. First, the contemporaneous causal relationship between the KSE and KOSDAQ markets are examined and compared for the period covering 2005 to 2010. As there were no restriction on the trading activities of foreign investors during this period and the sample data includes all market participants' activity data, this causality analysis fully reflects the impact of foreign investors on the Korean stock market and the multiple dynamics occurring among the three types of investors. Based on the structural break during the 2008 financial crisis, the changes in these causal relationships are also studied. Second, considering the interactions and effects of multiple players' activities, this study identifies the types of investor whose activities can be considered a root cause or a sink in contemporaneous information flow in the two markets and examines the manner in which this finding changes from pre- to post-break periods. This analysis could give implication on diverse contemporaneous interaction among the market participants, which broaden the previous research on focusing on foreign investor's causal impact to stock price movement in case of the Korean stock market.

Data and Methodology

The data set used in this study comprises 1494 daily stock data items from the KSE and KOSDAQ markets, spanning the period from January 3, 2005 to December 29, 2010. In the case of the KSE market, trading activities for three types of investors and the Korean Composite Stock Price Index (KOSPI[10]) were used. In the case of the KOSDAQ market, trading activities for three types of investors and the composite index of the KOSDAQ[11] were analyzed. The three types of investors were domestic institutional investors (INS), domestic individual investors (IND), and foreign investors (FOR)[12]. Stock price returns (RET) data was calculated by taking the log difference of the closing price of indices. Stock price volatility (VOL) was derived by taking the difference between the daily high and low prices for each market. Trade activity data included daily net purchase (NP) and daily total trade volume (TV) for each type of investor. All data were supported from the Korean Exchange (KRX)[13]. Figure A-15 shows the daily stock price movements of KOSPI and KOSDAQ indices from 2005 to

---

[10] The KOSPI is a comprehensive measure of the general market trend in Korean, and is measured as a price-weighted index based on the aggregate market value using the base date January 4, 1980 with the base index of 100. A total of 902 stocks of 704 companies were listed on the Korean Stock Exchange (KSE) with the market capitalization reaching 1888 trillion Korean won by year end of 2000 (Jeon and Jang 2004).

[11] The KOSDAQ index is a capitalization-weighted index that measures the performance of the KOSDAQ market. The index was developed with a base value of 100 as of July 1, 1996. The base value changed to 1000 as of January 26, 2004.

[12] Foreign investors in Korea must register with the Financial Supervisory Board (FSB) and obtain an ID number before they can start trading stock.

[13] Korea Exchange (KRX) is the sole securities exchange operator in South Korea. The Korea Exchange was created through the integration of Korea Stock Exchange (KSE), Korea Futures Exchange and KOSDAQ Stock Market in 2007. As of 31 December 2007, Korea Exchange had 1,757 listed companies with a combined market capitalization of $1.1 trillion. The exchange has normal trading sessions from 09:00 am to 03:00 pm on all days of the week except Saturdays, Sundays and holidays declared by the Exchange in advance

2010. Both stock indices experienced a price drop around the time of the 2008 financial crisis.

The Korean market is classified as one of the emerging markets as it has experienced significant economic growth and development in recent decades. Emerging market countries often present various barriers that hinder international portfolio investment. And, even though the foreign ownership of domestic firms may not be a complete measure of stock market openness, the lifting of a foreign investment ceiling enables the liberalization of market conditions to enhance participation by foreign investors. Historically, the Korea stock market strictly limited foreign investment at the 10% level, and this ceiling was increased very carefully in a step-by-step manner, as shown in Table B-20. Financial reforms implemented by the International Monetary Funds (IMF) played a large role in Korean financial liberalization after the Korean financial crisis in 1997, and the Korean stock market was completely opened up to foreign investment, without any ownership ceiling, in May 1998, eight months after the financial crisis. The program of reforms implemented by the Korean government, under IMF supervision, has succeeded in restoring market confidence. In addition, the IMF aided the Korean government in revising existing laws and regulations to further induce capital inflow. Table B-21 details the proportion of the daily trading volume attributable to each of the three types of players in the of the KSE market from 2001 to 2010. With respect to the trading activities of foreign investors, the average proportion increased gradually from 4.86% in 1995, to 7.47% in 1998, to 10.89% in 2001, to 15.68% in 2003, to 21.16% in 2005, to 25.52% in 2008, and to 19.47% in 2010 (J. Kim, Kartsaklas, and

Karanasos 2005). The sample data collection period for this study was from 2005 to 2010, a time period that covered the full effect of stock market liberalization on foreign investors.

In the current research, the stationarity of each series was investigated first. In Table B-22, the results of the Augmented Dickey-Fuller test are exhibited. There is no ground for suspicion of a unit root with respect to the eight data series of the two markets. In general, I reject the null hypothesis that each series contains a unit root. Following previous research using the time series data in the contemporaneous causal study (Bryant, Bessler, and Haigh 2006; Kim, Leatham, and Bessler 2007), I filtered the time series for the two markets through a vector autoregressive (VAR) model. Swanson and Granger (1997) introduce the use of graphical methods to contemporaneous causal ordering of VAR models. In addition, Pearl (2000) and Spirtes, Glymour, and Scheines (2000) developed Direct Acyclic Graphs (DAGs) utilizing conditional probabilities and graph theory to identify contemporaneous causality. Contemporaneous causality analysis is conducted over the innovation arising from VAR. In the KSE and KOSDAQ markets, a VAR was estimated that includes the eight time series data items; RET (stock price returns), VOL (stock price volatility), INS-NP (net purchase of institutional investors), INS-TV (total trade volume of institutional investors), IND-NP (net purchase of individual investor), IND-TV (total trade volume of individual investors), FOR-NP (net purchase of foreign investors), and FOR-TV (total trade volume of foreign investors). An optimal lag length for VAR was selected using the Schwarz (1978) information criterion.

DAGs are diagrams that use arrows and variables to represent the contemporaneous causal flow among or between a set of variables based on observed and partial correlation (Pearl 2000). Two popular algorithms were utilized to search for patterns in the DAGs. The first candidate was the PC algorithm (Pearl 2000) which assesses particular independence and conditional interdependence using the null hypothesis test. The second candidate was the Greedy Equivalence Search (GES), which is a score-based search algorithm. Dash and Druzdzel (1999) provided a constraint-based search that is relatively rapid but has two well-known weaknesses, one of which arises from the treatment of latent variables. The constraint-based search tends to portray causal relationship using a bi-directional arrow when a latent variable exists. The other weakness is an instability problem concerning the sample size.

The GES algorithm detects the causal pattern using the following systematic search algorithm. Starting with an undirected DAG, a two-step algorithm is used. In the first step, the Forward Search step calculates the goodness of fit among all equivalence classes with a single additional edge (acyclic) and selects the class having the highest score. This procedure is then repeated until no further improvements can be made to the score. In the second step, the Backward Equivalence Search utilizes with the results of the first step, and, then, it repeatedly searches among equivalent classes with a single edge less and selects the graph with the highest Schwarz Bayesian Score until no further improvements can be made. Thus, the best fit model is chosen from among the structural equation models using the innovations of VAR. Tetrad IV software was used to generate the GES algorithm in this study. However, in this GES algorithm, it could not accommodate latent variables, which certainly exist given the limited observed variables employed in this study.

Since causalities may respond to exogenous shocks or change through time and the data collection for this study included the 2008 financial crisis, the possible structural change resulting from this economic crisis may change the contemporaneous causality. A traditional approach would be to pick an arbitrary sample breakpoint, often the midpoint of the sample, and use a Chow test for structural change. This could be further refined by associating breakpoints with major events relevant to the data series. However, either of these approaches suffers from the arbitrary nature of the selected breakpoints. In this research, the Quandt-Likelihood Ratio (QLR) test, which is based on Andrew's approach to the unknown structural break test (1993), is used for detecting structural change of unknown timing. The QLR test consists of calculating Chow breakpoint tests at every observation, while ensuring that subsample points are not too close the end-points of the sample. The QLR test was applied to the pooled data in this study with 20-percent trimming. The probabilities for these statistics were calculated using Hansen's (1997) method. The critical value of the QLR statistic at the 90 percent significance level was 3.26 (Stock 2007), which indicates that the null hypothesis of no structural change is rejected. The maximum statistic of 3.32 was observed on November 11, 2007, which indicates the breakpoint location.

This structural break was likely caused by a combination of domestic and international economic factors. As the Korean economy experienced a continuous price hike in domestic real estate market from 2000 to 2006, concerns regarding economic bubble gradually increased and affected the financial market beginning in the fourth quarter of 2007. In addition, more bad news related to the sub-prime mortgage problem in the US, also affected the Korean stock market. In sequence, New Century Financial, a big mortgage lender, filed for the bankruptcy on April 2007, two hedge funds within Bear Stearns also filed for bankruptcy on July 2007, and several big financial companies including Morgan Stanley, Merrill Lynch, and Bear Stearns began layoffs after announcing third quarter declines in business performance. The KSE market reached a maximum on October 31, 2007, following a period that coincided with some of the biggest shocks form the US economic crisis.

Based on these result, the data is divided into pre-break and post-break periods. I conducted the Box-M test to validate the structural change. If there is a structural change, the two covariance matrices based on the estimated VAR in the pre-break and post-break periods will differ significantly from each other. I employed the Box-M test (Box 1949) to measure equality of the two covariance matrices. I found that the statistic for the Box-M test (690.25), in case of KSE market, exceeds the critical value ($\chi^2(36) = 58.62$), and does the statistic for the KOSDAQ market (489.74) at 1% significance level. Hence, the two covariance matrices of the KSE for the pre- and post-break periods differ from each other, as do the two covariance matrices of the KOSDAQ market indicating the structural change.

Empirical Results and Discussion

Figure A-16 shows the four DAGs, generated using the TETRAD IV GES algorithm, representing the direction of contemporaneous causal flows among variables in the pre- and post-break periods in both the KSE and KOSDAQ markets. Comparison of the two DAGs in Figure A-16 (a) suggests that causalities change after a financial crisis. A striking finding is that more contemporaneous causal relationships appear to be present in the post-break periods, which implies that information flow is faster and/or more effective within the KSE market during post-break in comparison to the pre-break period. In the case of the KOSDAQ market, the most evident finding, shown in Figure A-16 (b), is that the simple and obvious unidirectional contemporaneous causal relationship appears to be present in the post-break, which is the opposite result with the KSE market case. This can be interpreted as the dominance of contemporaneous information flow is clearly constructed by one player in this period. This is highly related with another finding that contemporaneous causal relations between domestic investor's activities and foreign investor's activities are reversed from pre- to post-break periods. This implies an important change in information flow in the KOSDAQ market, whereas the KSE market maintained a relatively constant information flow among player's activities.

Focusing on the impact of foreign investors on stock price movement, the causal relationships in each of the four cases is summarized in Table B-23. Especially, the net purchase of foreign investor can be explained as the capital inflow from outside of market. In the KSE market, it does not appear that the activities of foreign investors impacted stock returns or volatility movements, directly or indirectly, in pre-break period. However, in the post-break period, the activities of foreign investors directly affected stock returns and indirectly impacted volatility in the KSE market. In the KOSDAQ market, findings similar to those of the KSE market were observed in the causal relationship from foreign investor's activities to stock price movement. This implies that the impact of foreign investors on stock price returns and volatility became more evident in the post- than in the pre-break periods both for the KSE and KOSDAQ markets. However, in order to explain whether the foreign investors play a more dominant role in contemporaneous information transmission especially in the post-break periods, the dynamic interrelation among the three types of investors has to be examined.

For this, the contemporaneous causalities analysis was the identification of the type of investor whose activities represented the information sink or information root cause for each period. As the interaction and interdependency among the three types of players' activities increased and information asymmetry among the player was reduced due to IT development in the stock market, the role of information flow also changed depending on the market type and sample period. Table B-24 exhibits the root cause and sink of information for each market and each sample period.

First, in the KSE market, the trading volume for individual investors was a root cause in terms of information discovery and the net purchases of foreign investors was a sink in the pre-break period. In the post-break period, the information root cause was the trading volume of institutional investors and the information sink was the activities of foreign investors. From these results of Tables B-23 and B-24, it appears that, even though the impact of foreign investor on the stock price movements became more evident in the post- than the pre-break period, the role of foreign investors is still relatively less important both before and after the crisis periods in the KSE market. Contrarily, it appears that the dominant role of the information flow mechanism in the KSE market changed the individual investors in the before the crisis period to the institutional investors in the after the crisis period. This can be explained by the typical behavioral pattern of individual investors, in which they became more careful in investment decision and were easy to follow other trustful market player, the institutional investors in the KSE market, as they investors experienced the sudden drops and high volatility of stock price after the crisis.

In the KOSDAQ market, the most interesting results were observed. In the pre-break period, the trading volume of institutional investors was a root cause and the trading volume of foreign investors was an information sink. However, in the post-break period, the activities of foreign investors were root causes in new influence mechanism and the net purchases of individual investors was an information sink. Compared to the KSE market, the impact of foreign investors in the post-break period was much stronger than other investors in the post-break period. As in the NASDAQ market in the US, KOSDAQ firms are generally similar to venture companies.

In addition, government regulations are less strict in the KOSDAQ market than the KSE market, and this looser oversight or regulation of the KOSDAQ exchange may induce KOSDAQ investors to pursue highly returns with higher risk than KSE investor.

Therefore, the individual investors in the KOSAQ market centered on the short term investment, while the individual investors in the KSE market tried to invest in the long term perspective. In the post-crisis period, it was easy for the individual investor in the KOSDAQ market to invest careful and to follow the other more trustful investor's activities. After crisis, individual investors were more dependent on the foreign investors in the KOSDAQ market, while they followed the institutional investors in the KSE market. The results of this DAG result provide a good explanation of the market characteristics of the KOSDAQ.

Summary and Concluding Remarks

In this study, contemporaneous causal relationships among stock returns, volatility and three types of investors' activities in the Korean stock markets are investigated using Directed Acyclical Graph (DAG). Because of the dispute regarding the impact of foreign investors on the Asian stock markets, this study focused the causal relationships among foreign investor's activities, stock returns and volatility movement. In contrast to a previous analysis of the 1997 Asian financial crisis, I used 2005-2010 stock market data of foreign investor's activities, which reflect the financial market liberalization with the lifting of restrictions to foreign investment in Korean firms listed on the stock markets. In addition, this analysis was conducted over the pre- and post-break periods of two Korean stock markets, KSE (a Korean version of the NYSE) and

KOSDAQ (a Korean version of the NASDAQ), and reflected the structural change resulting from the 2008 global financial crisis.

This empirical approach to an analysis of the Korean stock markets investigates the causal relationships between market participants' activities, and stock price movement. Results of the current research suggest the following: first, based on the unknown structural break test of Andrews (1993) and knowledge of historical events occurring in the Korean stock market, a significant structural change in the Korean stock markets was detected in November 2007 that corresponds to the beginning of the decline in prices on the Korean stock markets and the sequential news regarding the financial crisis resulting from subprime mortgage problems in the US. This structural change is supported by the results of the Box-M test on covariance matrices between the pre- and post-break periods.

Second, I found that, in both the KSE and KOSDAQ markets, the contemporaneous causal influence of foreign investor's activities to stock return and volatility appears to more evident in the post-break period rather than the pre-break period. The strong contemporaneous causality of foreign investor's activity after the structural break implies that new information or shock emanating from foreign investor is more quickly and effectively transmitted thereby affecting stock price movement.

Third, while individual investor's activities in the pre-break period and institutional investor's activities in the post-break periods are a root cause of information flow in the KSE market, the information root cause in the KOSDAG market changes from institutional investor's activities in the pre-breaks period to foreign investor's activities in the post-break period. After the financial crisis, these causal effects of foreign investors were exhibits more intensively in the KOSDAQ market than in the KSE market. One might speculate that, when economic shocks or crises occur, domestic investors maintain their usual investment pattern in more stable market like the KSE market, but can be easily induced to investigate and follow foreign investor's patterns in risky market like the KOSDAQ.

Clearly, even though the financial shocks occurring around 2008 mainly originated from outside of the Korean market, they influenced these stock market participant's activities. The patterns of influence also differed depending on the market's characteristics. Through contemporaneous causal analyses, an improved understanding of the causal linkages among the different market players in the Korean stock market provides rich implication for market participants. While it remains a challenge to discover causal relationships among variables based on observational data, the methods available today, such as DAG, offer us greater opportunities to increase our knowledge on this issue.

CHAPTER V

CONCLUSION

The goal of this dissertation was to explore and increase our understanding of the recent oil market changes as a result of the utilization of diverse financial tools as elaborate predictors of oil prices and also as enhanced risk management tools. First, two essays examine an oil price forecasting model and several gasoline price hedging models from this perspective. The third essay studies the contemporaneous causal relationships among stock price movements and market participants' activities in the Korean stock markets.

In Chapter II, I provide the motivation and develop a model for using the crack spread and ETF spread for accurate forecasting purposes, especially in view of recent trends in the highly volatile prices of oil and its products. Based on the Error Correction Model (ECM) and Multivariate Generalized Autoregressive Conditional Heteroskedasticity Model (MGARCH), I examined the causal relationships between crude oil and both crack spread and oil ETFs and the forecasting abilities of these two tools. The results of this study reveal that crack spread futures and oil ETF spread are good predictors of oil price movement and, in a comparison of crack spread and ETFs, that ETFs are better predictors than crack spread. The change in causal relationship can be explained by the fact of the increasing need of the oil-related financial market for oil price hedging tools and investments.

In Chapter III, I incorporated diverse optimal hedging models for unleaded gasoline spot prices using gasoline futures with static-symmetric, time varying-symmetric, static-asymmetric, and time varying-asymmetric dependencies for the purpose of risk minimization in portfolio. In addition, the alternative hedging performance of ETFs was compared using futures' performance in the framework of four dependency cases. I conclude that in analyzing any given data period it is very important to select the correct model to estimate the hedge ratio. Although the dynamic copula approach may have limited power to examine atypical data period, the risk exposure of a portfolio may be effectively managed by the use of a dynamic copula model with precise specifications for the joint distribution of assets in more stable and typical data periods. In addition, considering the increasing interactions among the spot, futures and matching commodity ETF markets, the use of an appropriate hedge model to create a diverse hedged portfolio may have crucial implications for risk management.

In Chapter IV, the final essay, I examined contemporaneous causal relationships in the Korean stock markets, focusing on the contemporaneous causal changes occurring at the time around the financial crisis in 2008. The causal influence of foreign investor's activities on stock returns and volatility appears to more evident in the post-break period than in the pre-break periods for both the KSE and KOSDAQ markets. After the crisis, these causal effects of foreign investor exhibits more intensively in the KOSDAQ market than in the KSE. The results of this causal analysis will provide implications regarding the three types of investors' investment strategies depending on the stage of the economic business cycle.

These essays build on our understanding of the diverse financial tools utilized in the oil and oil products market, such as crack spread, futures, and ETFs, in terms of investment and risk management, and offer suggestions regarding the multiple players' contemporaneous causal dynamics in the stock market.

REFERENCES

Alexander, Carol, and Andreza Barbosa. 2007. “Effectiveness of Minimum-Variance Hedging.” *The Journal of Portfolio Management* 33 (2) (January): 46–59. doi:10.3905/jpm.2007.674793.

Alizadeh, Sassan, Michael W Brandt, and Francis X Diebold. 2002. “Range-Based Estimation of Stochastic Volatility Models.” *The Journal of Finance* 57 (3) (December 17): 1047–1091. doi:10.1111/1540-6261.00454.

Andrews, Donald W. K. 1993. “Tests for Parameter Instability and Structural Change With Unknown Change Point.” *Econometrica* 61 (4) (July 1): 821–856. doi:10.2307/2951764.

Anon. “321energy :: Time (Again) For Crack Trades? :: Brad Zigler.” http://www.321energy.com/editorials/zigler/zigler103009.html.

Bartram, Söhnke M., Stephen J. Taylor, and Yaw-Huei Wang. 2007. “The Euro and European Financial Market Dependence.” *Journal of Banking & Finance* 31 (5) (May): 1461–1481. doi:10.1016/j.jbankfin.2006.07.014.

Bohn, Henning, and Linda L. Tesar. 1996. “U.S. Equity Investment in Foreign Markets: Portfolio Rebalancing or Return Chasing?” *The American Economic Review* 86 (2) (May 1): 77–81.

Bollerslev, Tim. 1986. “Generalized Autoregressive Conditional Heteroskedasticity.” *Journal of Econometrics* 31 (3) (April): 307–327. doi:10.1016/0304-4076(86)90063-1.

———. 1990. “Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized Arch Model.” *The Review of Economics and Statistics* 72 (3): 498–505. doi:10.2307/2109358.

Bollerslev, Tim, Robert F. Engle, and Jeffrey M. Wooldridge. 1988. “A Capital Asset Pricing Model with Time-Varying Covariances.” *Journal of Political Economy* 96 (1) (February 1): 116–131.

Box, G. E. P. 1949. “A General Distribution Theory for a Class of Likelihood Criteria.” *Biometrika* 36 (3/4) (December 1): 317–346. doi:10.2307/2332671.

Brailsford, Timothy J. 2009. “The Empirical Relationship Between Trading Volume, Returns and Volatility.” *Accounting & Finance* 36 (1) (February 25): 89–111. doi:10.1111/j.1467-629X.1996.tb00300.x.

Brennan, Michael J., and Eduardo S. Schwartz. 1985. “Evaluating Natural Resource Investments.” *The Journal of Business* 58 (2) (April 1): 135–157.

Brooks, Chris. 1998. “Predicting Stock Index Volatility: Can Market Volume Help?” *Journal of Forecasting* 17 (1) (December 4): 59–80. doi:10.1002/(SICI)1099-131X(199801)17:1<59::AID-FOR676>3.0.CO;2-H.

Bryant, Henry L, David A Bessler, and Michael S Haigh. 2006. “Causality in Futures Markets.” *Journal of Futures Markets* 26 (11) (September 22): 1039–1057. doi:10.1002/fut.20231.

Carmona, René, and Valdo Durrleman. 2003. “Pricing and Hedging Spread Options.” *SIAM Review* 45 (4) (December 1): 627–685.

Chen, S. S, C. Lee, and K. Shrestha. 2003. “Futures Hedge Ratios: a Review.” *Quarterly Review of Economics and Finance* 43 (3): 433–465.

Choe, Hyuk, Bong-Chan Kho, and René M Stulz. 1999. “Do Foreign Investors Destabilize Stock Markets? The Korean Experience in 1997.” *Journal of Financial Economics* 54 (2) (October): 227–264. doi:10.1016/S0304-405X(99)00037-9.

Clark, John, and Elizabeth Berko. 1997. “Foreign Investment Fluctuations and Emerging Market Stock Returns: The Case of Mexico.” *SSRN eLibrary* (May). http://papers.ssrn.com/sol3/papers.cfm?abstract_id=993813.

Clements, Michael P., and David F. Hendry. 2006. “Chapter 12 Forecasting with Breaks.” In *Handbook of Economic Forecasting*, Volume 1:605–657. Elsevier. http://www.sciencedirect.com/science/article/pii/S1574070605010128.

Dash, Denver, and Marek J. Druzdzel. 1999. “A Hybrid Anytime Algorithm for the Construction of Causal Models From Sparse Data.” In *Proceedings of the Fifteenth International Conference on Uncertainty in Artificial Intelligence*, 142–149. Morgan Kaufmann.

Diebold, Francis X., and Roberto S. Mariano. 1995. “Comparing Predictive Accuracy.” *Journal of Business & Economic Statistics* 13 (3) (July 1): 253–263. doi:10.2307/1392185.

Dornbusch, R., and Y.C Park. 1995. “Financial Integration in a Second-best World: Are We Still Sure About Our Classical Prejudices.” In *Inancial Opening: Policy Lessons for Korea*. Seoul, South Korea: Korea Institute of Finance.

Ederington, Louis H. 1979. “The Hedging Performance of the New Futures Markets.” *The Journal of Finance* 34 (1) (March 1): 157–170. doi:10.2307/2327150.

Edwards, F.R., and C. Ma. 1992. *Futures and Options*. New York: McGraw-Hill.

Engle, Robert. 2002. "Dynamic Conditional Correlation." *Journal of Business & Economic Statistics* 20 (3) (July): 339–350. doi:10.1198/073500102288618487.

Engle, Robert F. 1982. "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation." *Econometrica* 50 (4) (July 1): 987–1007. doi:10.2307/1912773.

Engle, Robert F., and C. W. J. Granger. 1987. "Co-Integration and Error Correction: Representation, Estimation, and Testing." *Econometrica* 55 (2) (March 1): 251–276. doi:10.2307/1913236.

Engle, Robert F., and Kevin Sheppard. 2001. "Theoretical and Empirical Properties of Dynamic Conditional Correlation Multivariate GARCH." *National Bureau of Economic Research Working Paper Series* No. 8554. http://www.nber.org/papers/w8554.

Froot, Kenneth A., Paul G.J. O'Connell, and Mark S. Seasholes. 2001. "The Portfolio Flows of International Investors." *Journal of Financial Economics* 59 (2) (February): 151–193. doi:10.1016/S0304-405X(00)00084-2.

Gallant, A. Ronald, Chien-Te Hsu, and George Tauchen. 1999. "Using Daily Range Data to Calibrate Volatility Diffusions and Extract the Forward Integrated Variance." *Review of Economics and Statistics* 81 (4): 617–631. doi:10.1162/003465399558481.

Haigh, Michael S, and Matthew T Holt. 2002. "Crack Spread Hedging: Accounting for Time-varying Volatility Spillovers in the Energy Futures Markets." *Journal of Applied Econometrics* 17 (3) (March 4): 269–289. doi:10.1002/jae.628.

Hansen, Bruce E. 1997. "Approximate Asymptotic P Values for Structural-Change Tests." *Journal of Business & Economic Statistics* 15 (1) (January 1): 60–67. doi:10.2307/1392074.

Heinkel, Robert, Maureen E Howe, and John S Hughes. 2006. "Commodity Convenience Yields as an Option Profit." *Journal of Futures Markets* 10 (5) (August 25): 519–533. doi:10.1002/fut.3990100508.

Hiemstra, Craig, and Jonathan D. Jones. 1994. "Testing for Linear and Nonlinear Granger Causality in the Stock Price- Volume Relation." *The Journal of Finance* 49 (5) (December 1): 1639–1664. doi:10.2307/2329266.

Hsu, Chih-Chiang, Chih-Ping Tseng, and Yaw-Huei Wang. 2008. "Dynamic Hedging with Futures: A Copula-based GARCH Model." *Journal of Futures Markets* 28 (11) (September 9): 1095–1116. doi:10.1002/fut.20345.

Jeon, Bang Nam, and Beom-Sik Jang. 2004. "The Linkage Between the US and Korean Stock Markets: The Case of NASDAQ, KOSDAQ, and the Semiconductor Stocks." *Research in International Business and Finance* 18 (3) (September): 319–340. doi:10.1016/j.ribaf.2004.04.006.

Joe, H., and J.J. Xu. 1996. "The Estimation Method of Inference Functions for Margins for Multivariate Models". Department of Statistics, University of British Columbia.

Johansen, Søren. 1995. "Identifying Restrictions of Linear Equations with Applications to Simultaneous Equations and Cointegration." *Journal of Econometrics* 69 (1) (September): 111–132. doi:10.1016/0304-4076(94)01664-L.

Karpoff, Jonathan M. 1987. "The Relation Between Price Changes and Trading Volume: A Survey." *Journal of Financial and Quantitative Analysis* 22 (01): 109–126. doi:10.2307/2330874.

Kim, J., A. Kartsaklas, and M. Karanasos. 2005. "The Volume–volatility Relationship and the Opening of the Korean Stock Market to Foreign Investors After the Financial Turmoil in 1997." *Asia-Pacific Financial Markets* 12 (3): 245–271. doi:10.1007/s10690-006-9024-7.

Kim, J.W., D.J. Leatham, and D.A. Bessler. 2007. "REITs' Dynamics Under Structural Change with Unknown Break Points." *Journal of Housing Economics* 16 (1) (March): 37–58. doi:10.1016/j.jhe.2007.04.001.

Kocagil, Ahmet E. 2004. "Optionality and Daily Dynamics of Convenience Yield Behavior: An Empirical Analysis." *Journal of Financial Research* 27 (1) (February 9): 143–158. doi:10.1111/j.1475-6803.2004.00082.x.

Kroner, Kenneth F., and Jahangir Sultan. 1993. "Time-Varying Distributions and Dynamic Hedging with Foreign Currency Futures." *Journal of Financial and Quantitative Analysis* 28 (04): 535–551. doi:10.2307/2331164.

Lien, Donald, and Xiangdong Luo. 2006. "Multiperiod Hedging in the Presence of Conditional Heteroskedasticity." *Journal of Futures Markets* 14 (8) (August 28): 927–955. doi:10.1002/fut.3990140806.

Liu, S. D., J. B. Jian, and Y. Y. Wang. 2010. "Optimal Dynamic Hedging of Electricity Futures Based on Copula-GARCH Models." In *Industrial Engineering and*

Engineering Management (IEEM), 2010 IEEE International Conference On, 2498–2502.

Murat, Atilim, and Ekin Tokat. 2009. “Forecasting Oil Price Movements with Crack Spread Futures.” *Energy Economics* 31 (1) (January): 85–90. doi:10.1016/j.eneco.2008.07.008.

Patton, Andrew J. 2006. “Modelling Asymmetric Exchange Rate Dependence*.” *International Economic Review* 47 (2) (April 11): 527–556. doi:10.1111/j.1468-2354.2006.00387.x.

Patton, Andrew J. 2009. “Copula–Based Models for Financial Time Series.” In *Handbook of Financial Time Series*, ed. Thomas Mikosch, Jens-Peter Kreiß, Richard A. Davis, and Torben Gustav Andersen, 767–785. Springer Berlin Heidelberg. http://www.springerlink.com/content/lr77032062096102/abstract/.

Pearl, Judea. 2000. *Causality: Models, Reasoning, and Inference*. New York: Cambridge University Press.

Pyun, Chong Soo, Sa Young Lee, and Kiseok Nam. 2000. “Volatility and Information Flows in Emerging Equity Market: A Case of the Korean Stock Exchange.” *International Review of Financial Analysis* 9 (4): 405–420. doi:10.1016/S1057-5219(00)00037-5.

Rodriguez, Juan Carlos. 2007. “Measuring Financial Contagion: A Copula Approach.” *Journal of Empirical Finance* 14 (3) (June): 401–423. doi:10.1016/j.jempfin.2006.07.002.

Saatcioglu, Kemal, and Laura T Starks. 1998. “The Stock Price–volume Relationship in Emerging Stock Markets: The Case of Latin America.” *International Journal of Forecasting* 14 (2) (June 1): 215–225. doi:10.1016/S0169-2070(98)00028-4.

Silvapulle, Param, and Jong-Seo Choi. 1999. “Testing for Linear and Nonlinear Granger Causality in the Stock Price-volume Relation: Korean Evidence.” *The Quarterly Review of Economics and Finance* 39 (1): 59–76. doi:10.1016/S1062-9769(99)80004-0.

Sklar, A. 1959. “Fonctions De Répartition Í¿ n Dimensions Et Leurs Marges.” In *Publications De l’Institut Statistique De l’Université De Paris 8*, 229–231.

Smeral, Egon, Stephen F. Witt, and Christine A. Witt. 1992. “Econometric Forecasts: Tourism Trends to 2000.” *Annals of Tourism Research* 19 (3): 450–466. doi:10.1016/0160-7383(92)90130-H.

Spirtes, Peter, Clark N. Glymour, and Richard Scheines. 2000. *Causation, Prediction, and Search*. Cambridge: MIT Press.

Stock, James H. 2007. *Introduction to Econometrics*. 2nd ed. Addison-Wesley Series in Economics. Boston: Pearson/Addison Wesley.

Swanson, Norman R., and Clive W. J. Granger. 1997. "Impulse Response Functions Based on a Causal Approach to Residual Orthogonalization in Vector Autoregressions." *Journal of the American Statistical Association* 92 (437) (March 1): 357–367. doi:10.2307/2291481.

Tse, Y. K. 1998. "The Conditional Heteroscedasticity of the Yen-dollar Exchange Rate." *Journal of Applied Econometrics* 13 (1): 49–55.

Working, Holbrook. 1949. "The Theory of Price of Storage." *The American Economic Review* 39 (6) (December 1): 1254–1262.

Zigler, Brad. "Time (Again) For Crack Trades?" *321energy*. Accessed October 30, 2009. http://www.321energy.com/editorials/zigler/zigler103009.html.

Zivot, Eric, and Donald W. K. Andrews. 1992. "Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis." *Journal of Business & Economic Statistics* 10 (3) (July 1): 251–270. doi:10.2307/1391541.

APPENDIX A

FIGURES

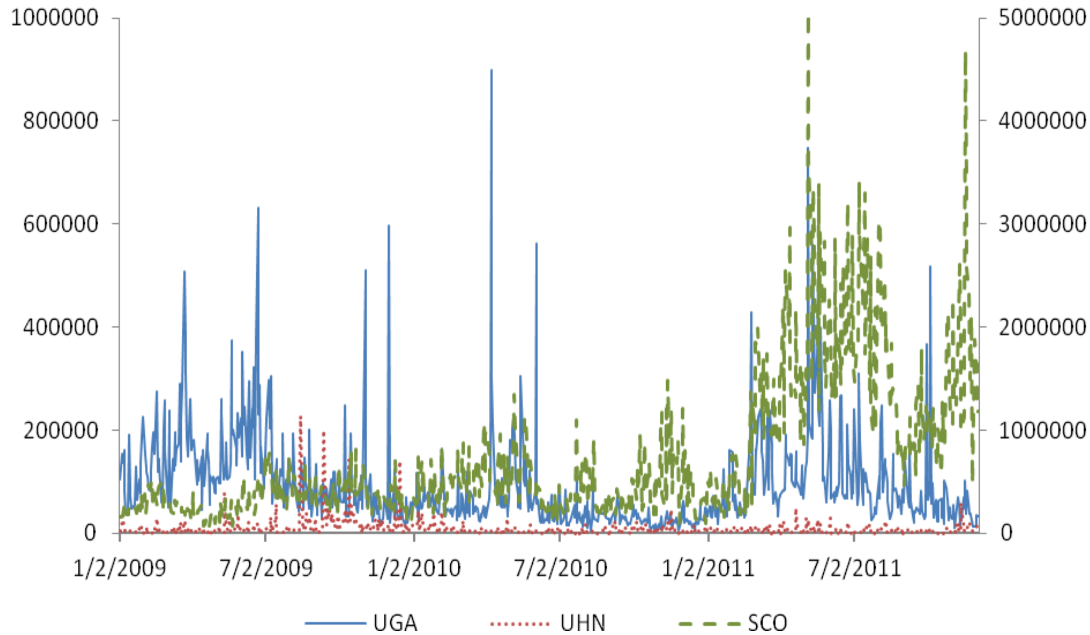
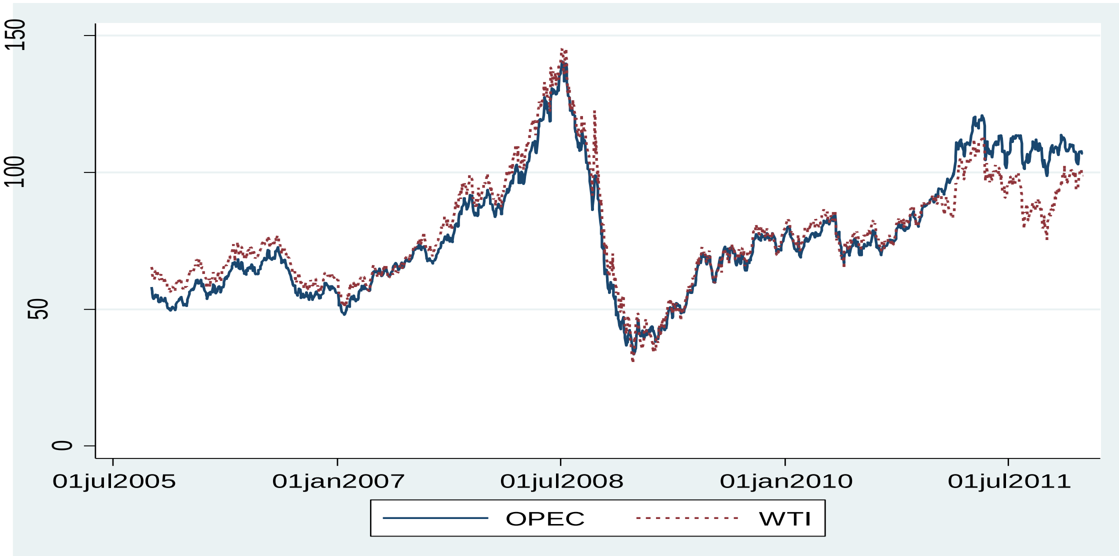


Figure A-1. Trade Volumes of Three Oil ETFs from 2009 to 2011

Note: The abbreviations are UGA (United States Gasoline Fund), UHN (United States Heating Oil Fund), and SCO (ProShares UltraShort DJ-UBS Crude Oil Exchange-Traded Funds). Left y-axis represents the trade volumes of UGA and UHN, while right y-axis represents the trade volume of SCO.

OPEC vs WTI



CS 211 vs CS321

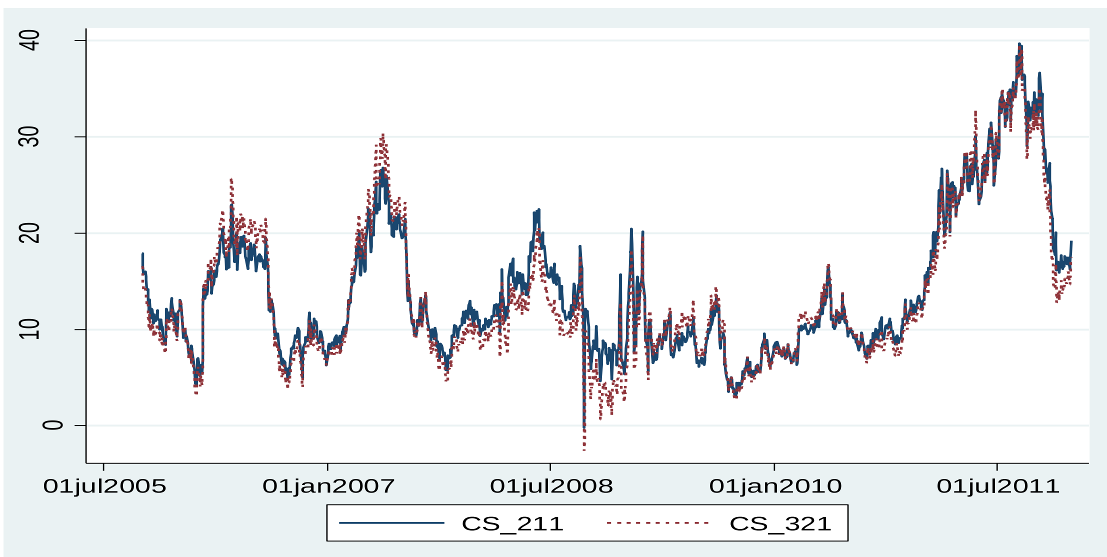
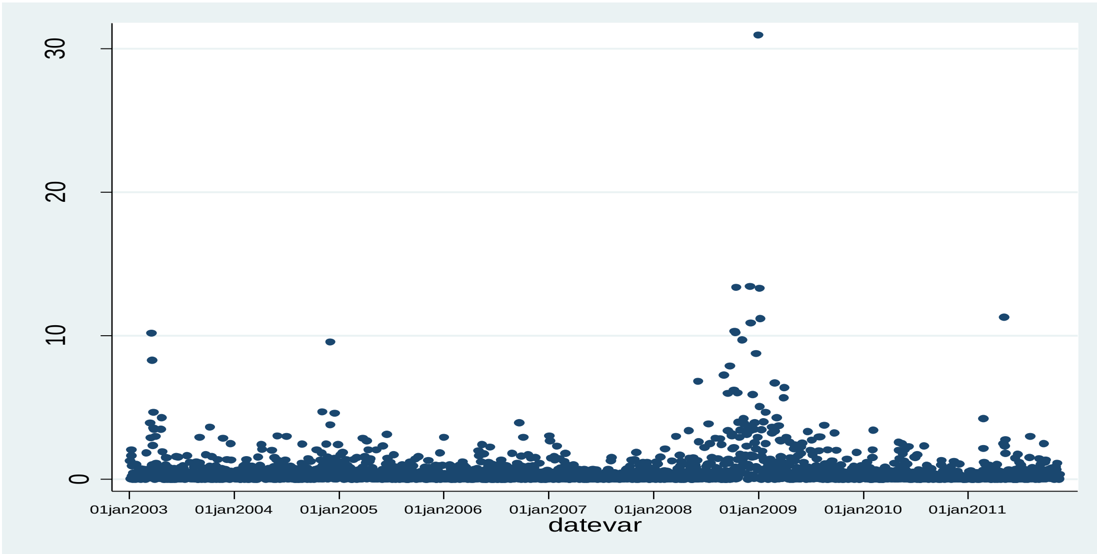


Figure A-2. Price Movements on Crude Oil and Crack Spread

Note: The abbreviations are OPEC (Organization of the Petroleum Exporting Countries' Reference Basket Price), WTI (West Texas Intermediate Price), CS_211 (2-1-1 version of Crack Spread) and CS_321 (3-2-1 version of Crack Spread).

Squared Returns of Crude oil



Squared Returns of 2-1-1 Crack Spread

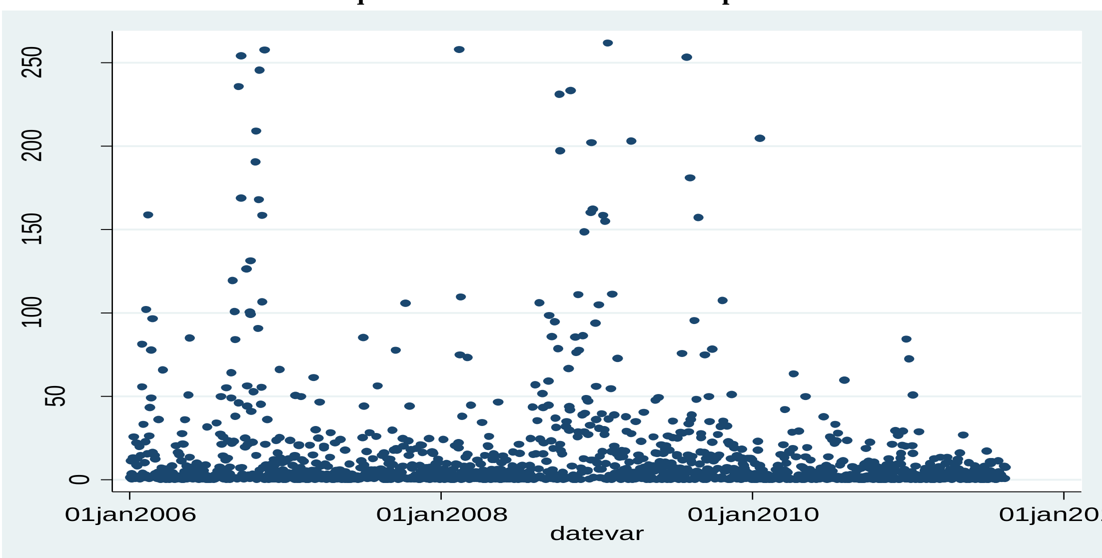


Figure A-3. Squared Returns of OPEC Crude Oil and a 2-1-1 Crack Spread

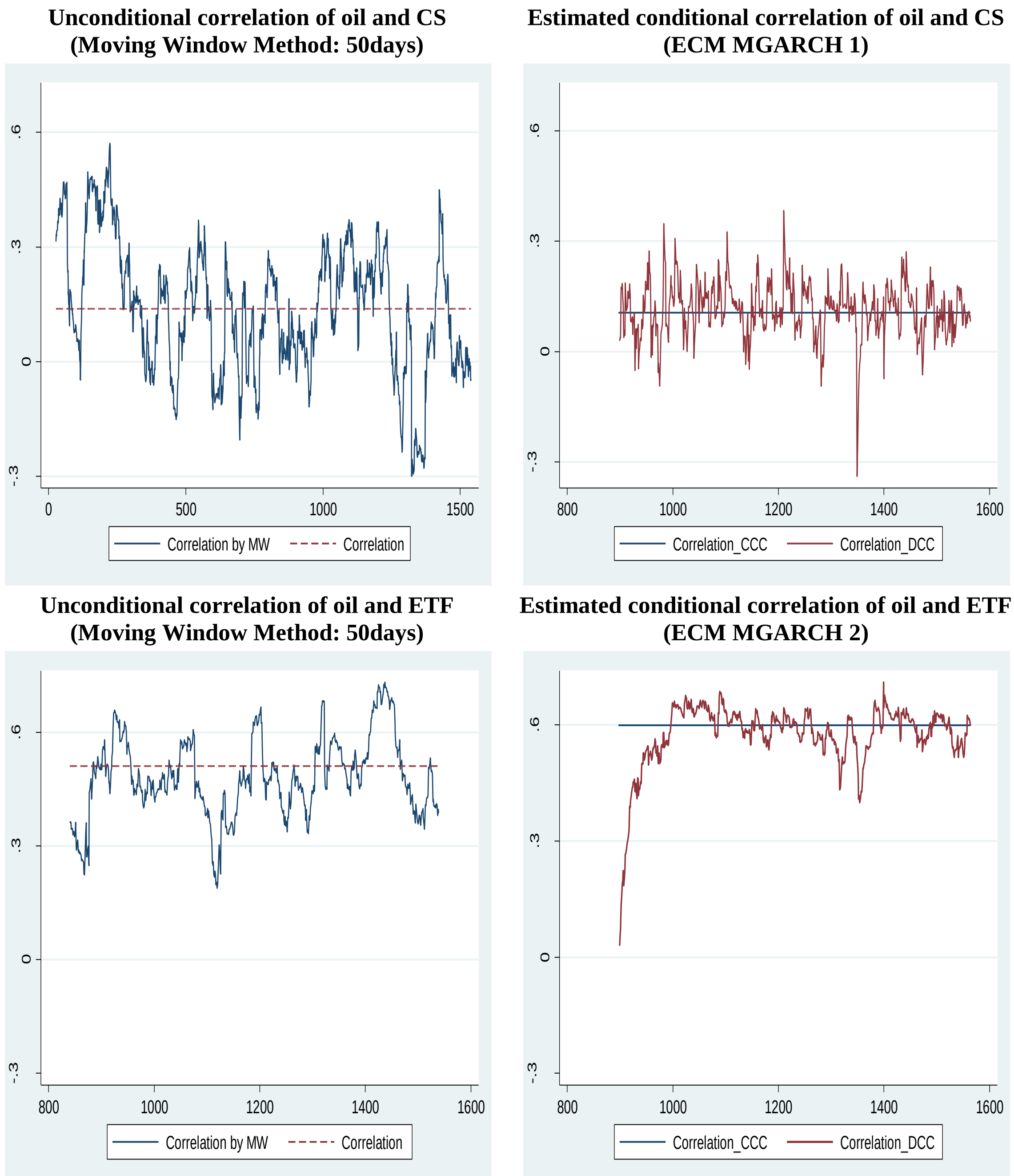


Figure A-4. Unconditional Correlations and Estimated Conditional Correlations

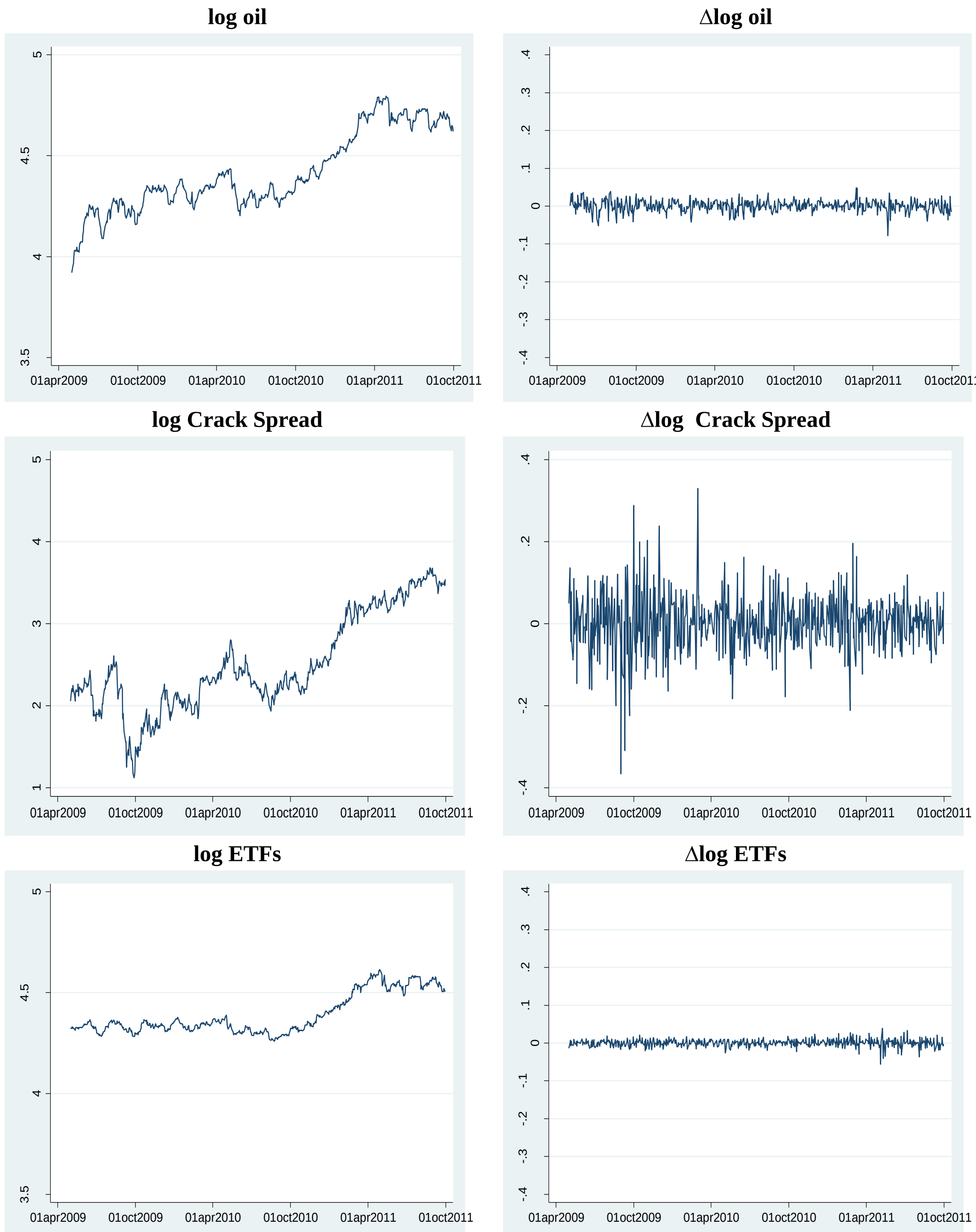
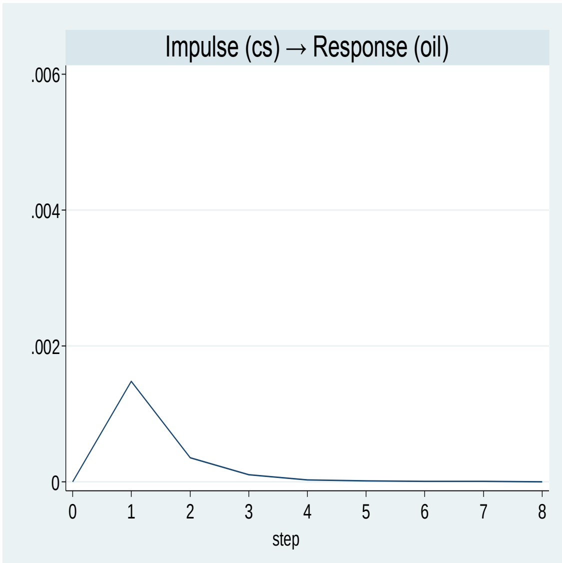
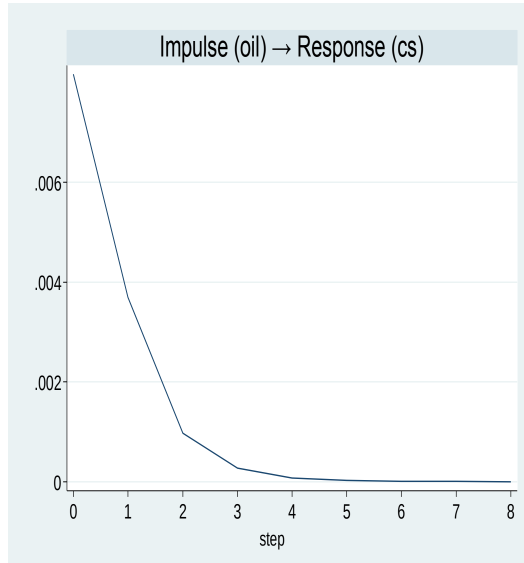


Figure A-5. Log Prices and First Differences of Log Prices

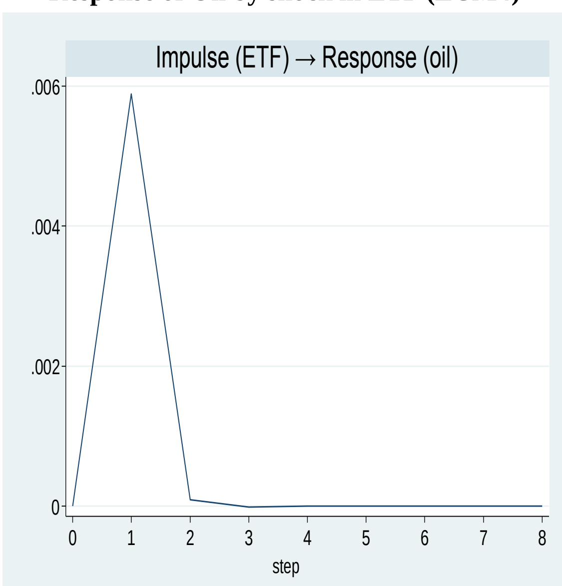
Response of Oil by shock in CS (ECM3)



Response of CS by shock in Oil (ECM3)



Response of Oil by shock in ETF (ECM4)



Response of ETF by shock in Oil (ECM4)

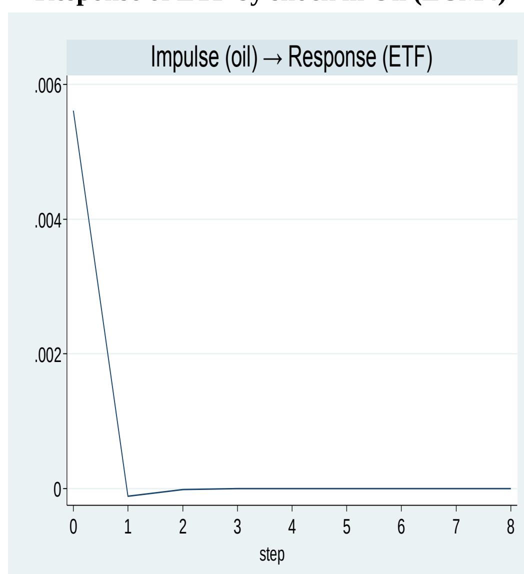


Figure A-6. Impulse Response to One Standard Error in ECM Models

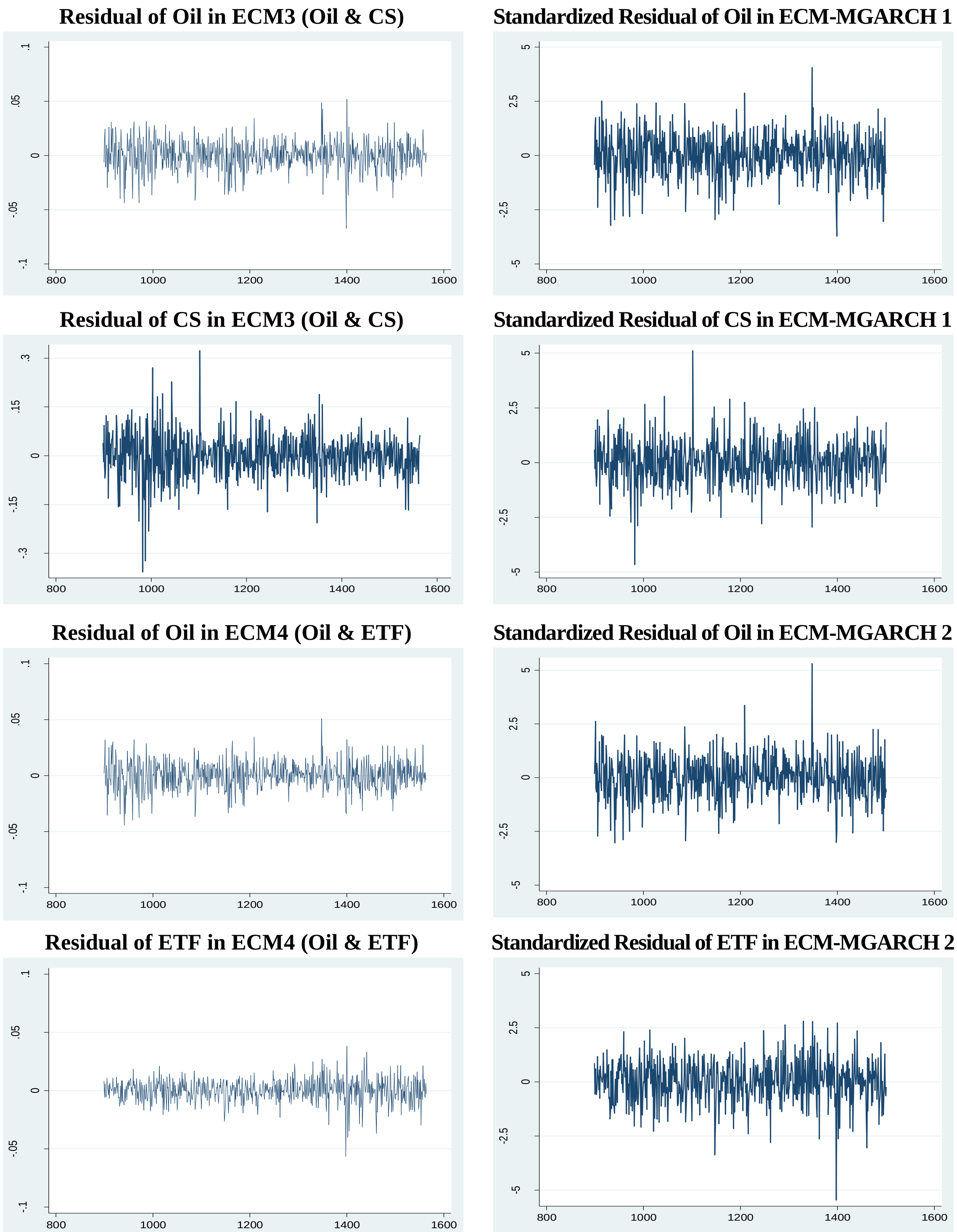


Figure A-7. Residuals for ECM and Standardized Residual for ECM-MGARCH

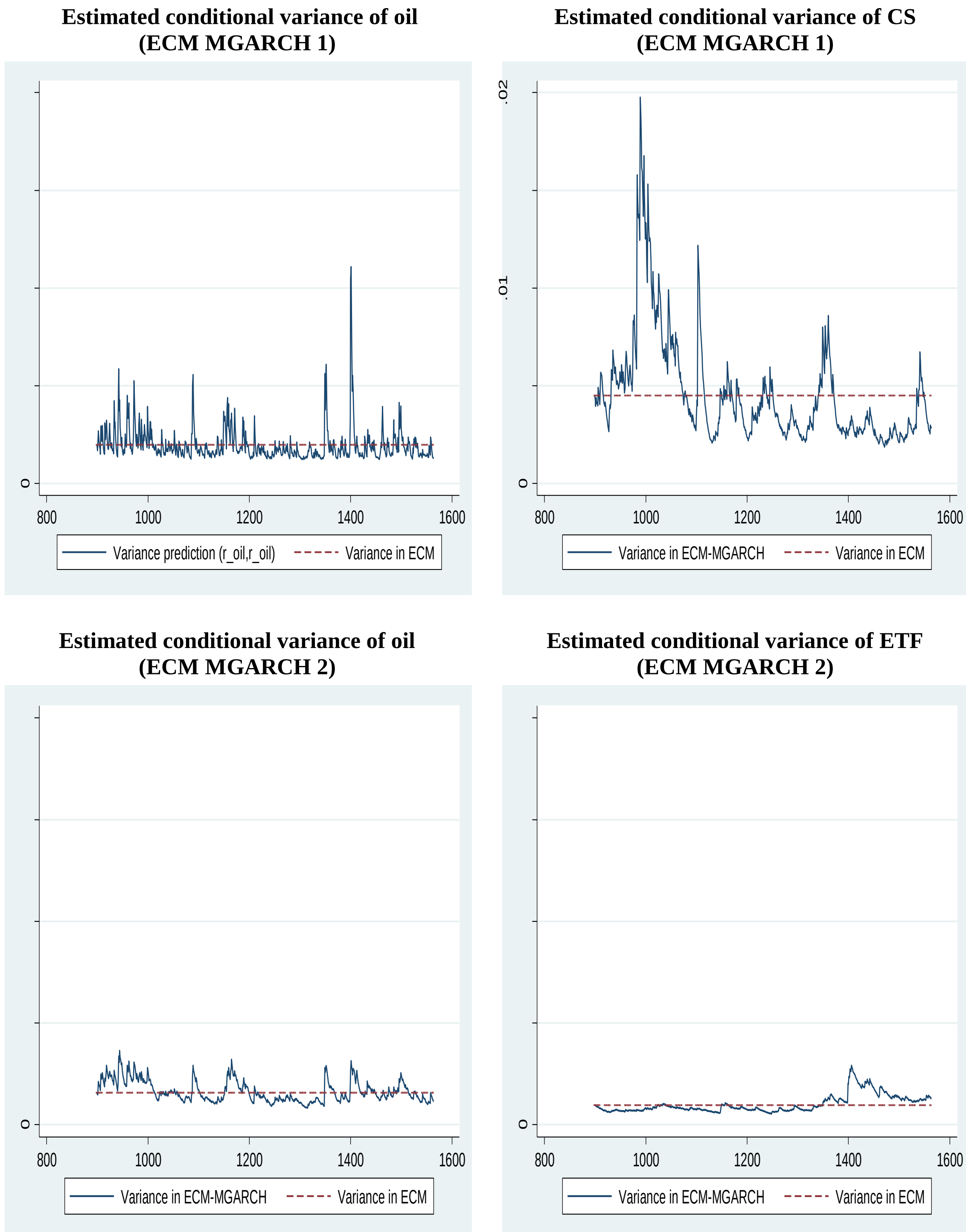
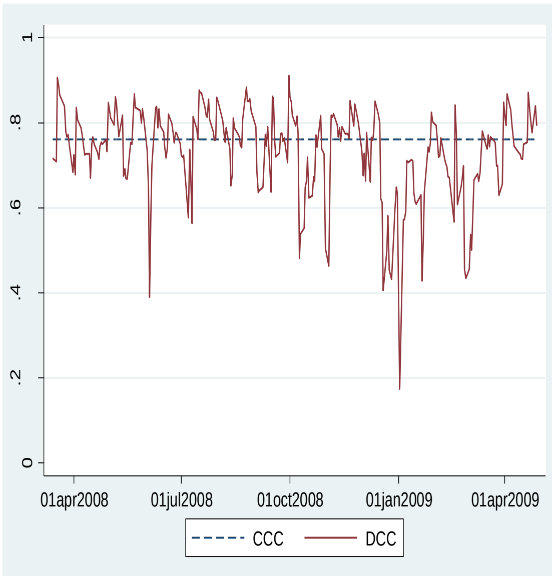


Figure A-8. Conditional Variances of Oil, Crack Spread, and ETF Spread

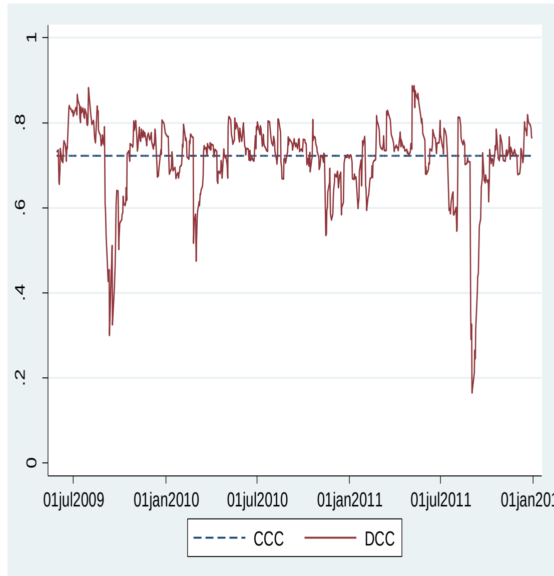
Figure A-9. Price Level Changes of Gasoline Spot, Futures and ETF Prices

Note: Left y-axis represents the log value of Spot and Futures of unleaded gasoline prices, while right y-axis represents the log value of UGA (Gasoline ETF) price.
.

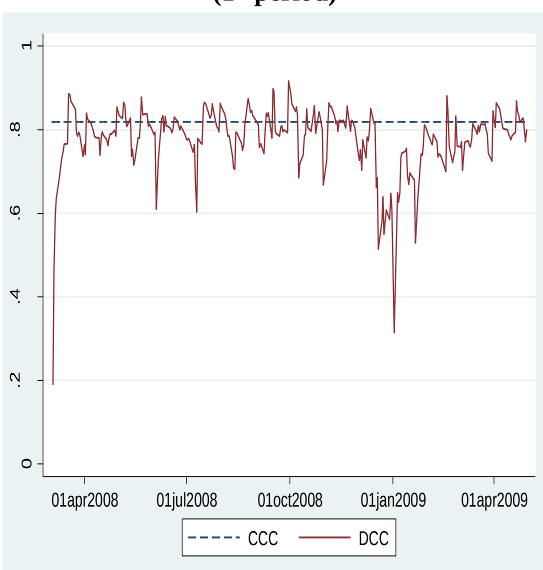
Conditional correlation of spot and futures (1st period)



Conditional correlation of spot and futures (2nd period)



Conditional correlation of spot and EFT (1st period)



Conditional correlation of spot and EFT (2nd period)

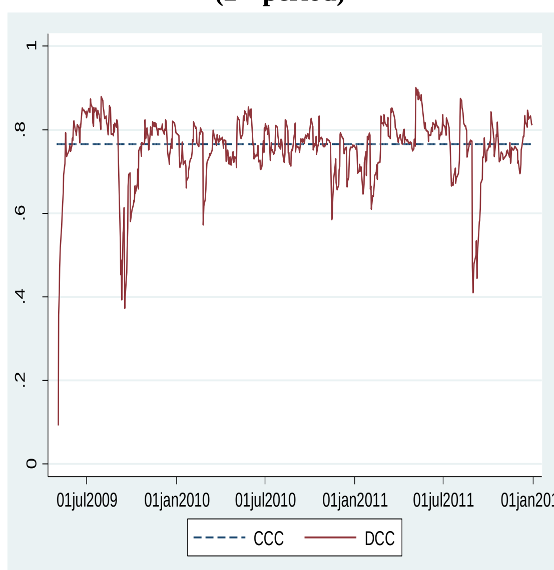


Figure A-10. Estimated Conditional Correlations in DCC MGARCH

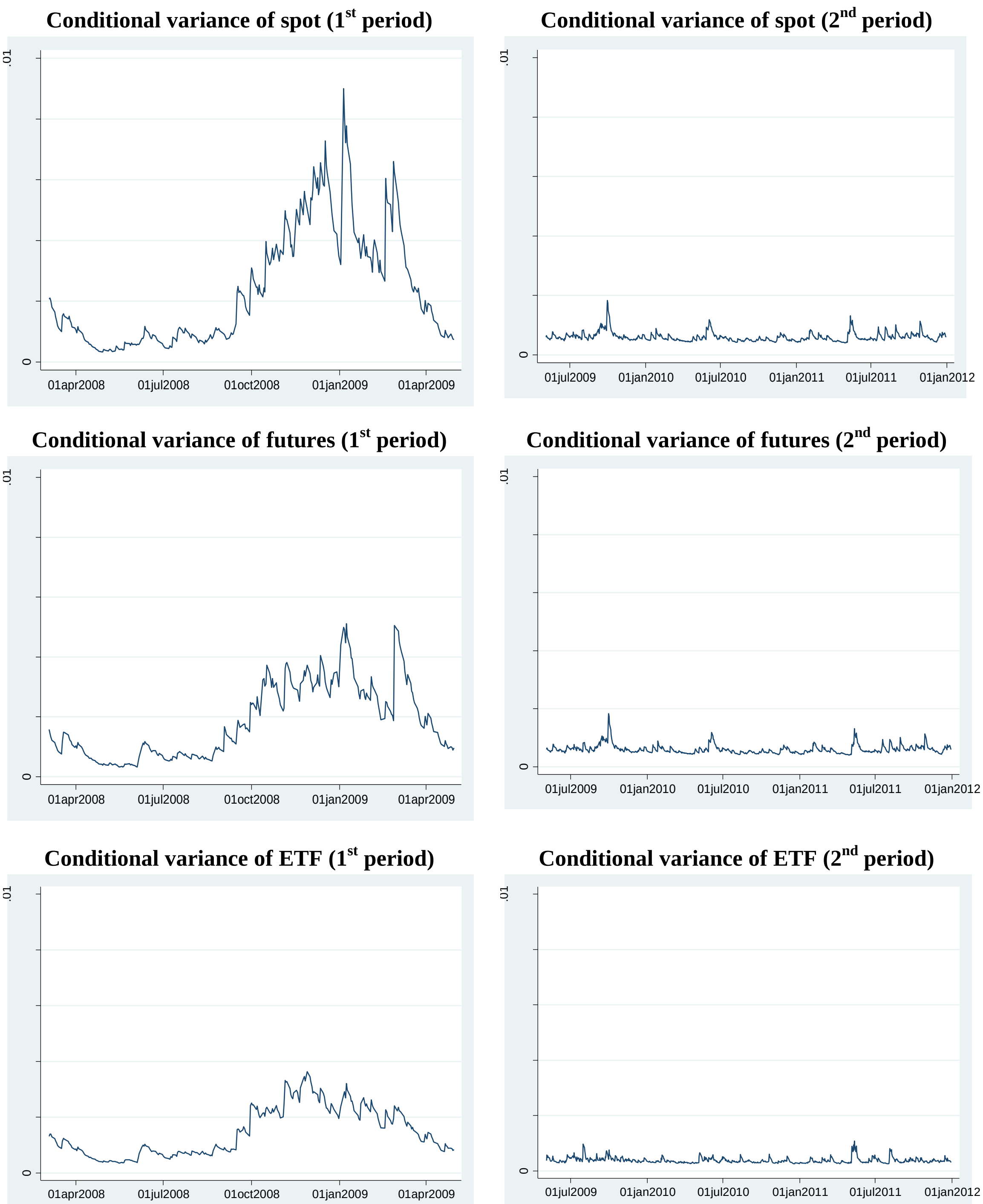
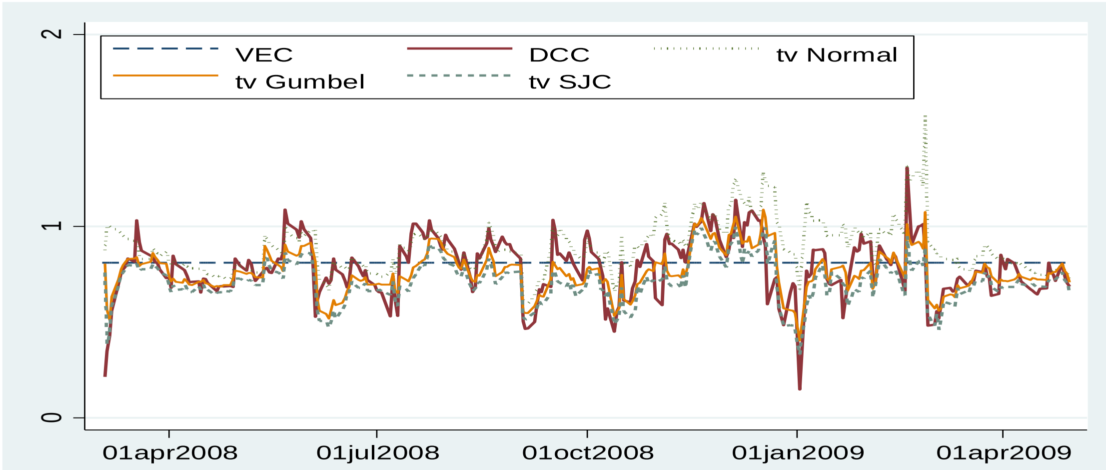


Figure A-11. Estimated Conditional Variance in DCC MGARCH

Hedge ratio in 1st period



Hedge ratio in 2nd period

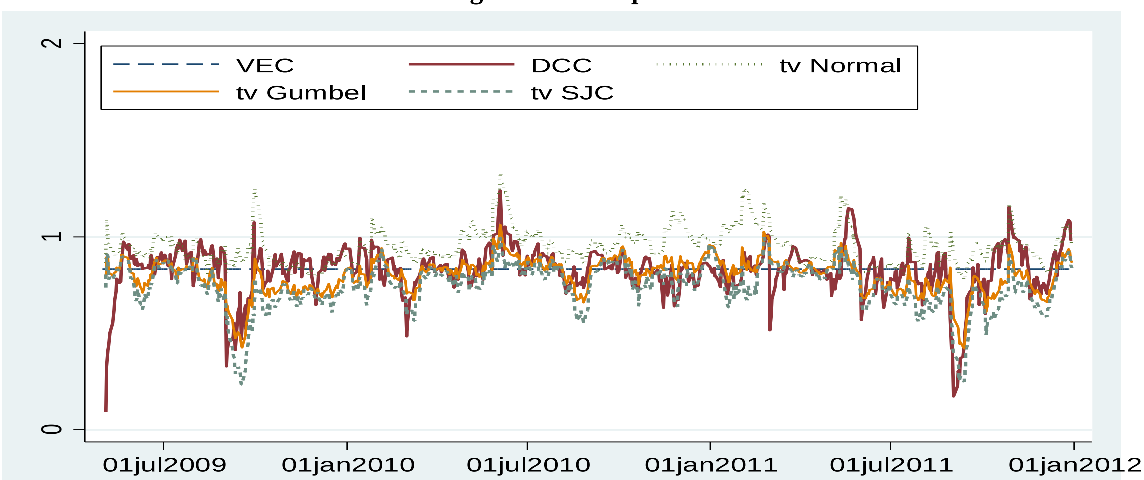
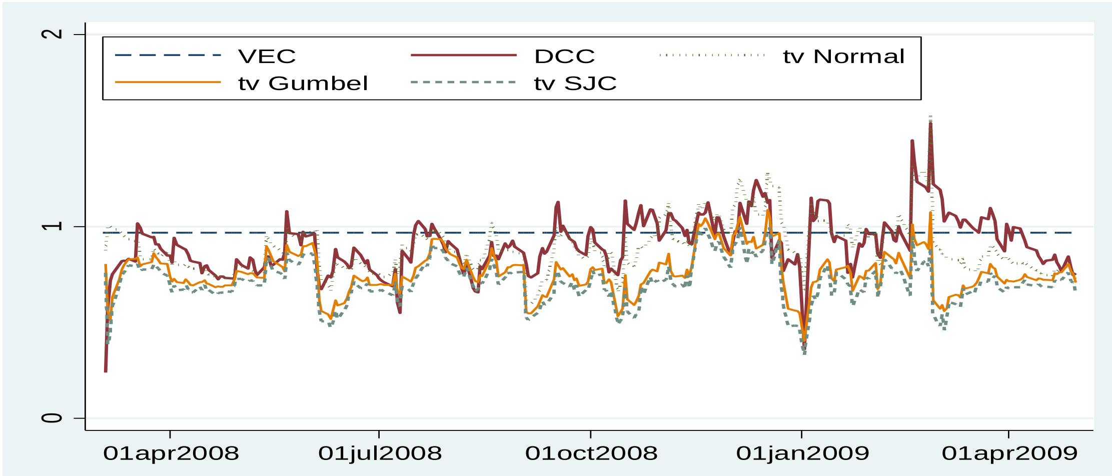


Figure A-12. Hedge Ratios in Case of Futures Hedging

Note: Static hedge ratio is derived only from the Vector Error Correction (VEC) model. Other models except the VEC model, Dynamic Conditional Correlation (DCC) MGARCH, time-varying Normal Copula (tv Normal), time-varying Gumbel Copula (tv Gumbel), and time-varying SJC Copula (tv SJC), generate the time-varying hedge ratios.

Hedge ratio in 1st period



Hedge ratio in 2nd period

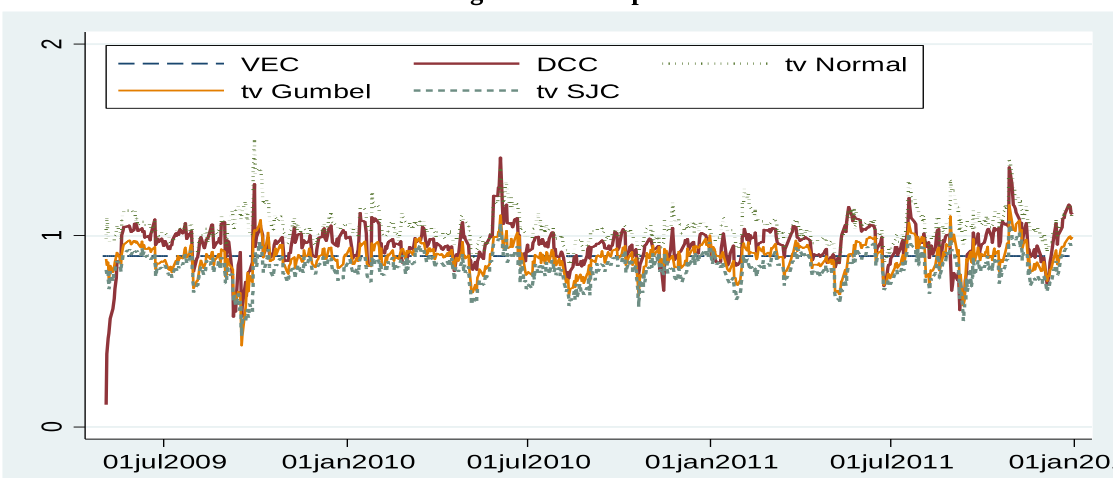


Figure A-13. Hedge Ratios in Case of ETF Hedging

Note: Static hedge ratio is derived only from the Vector Error Correction (VEC) model. Other models except the VEC model, Dynamic Conditional Correlation (DCC) MGARCH, time-varying Normal Copula (tv Normal), time-varying Gumbel Copula (tv Gumbel), and time-varying SJC Copula (tv SJC), generate the time-varying hedge ratios.

Time-varying dependence of SJC copula (1st period) | **Time-varying dependence of SJC copula (2nd period)**

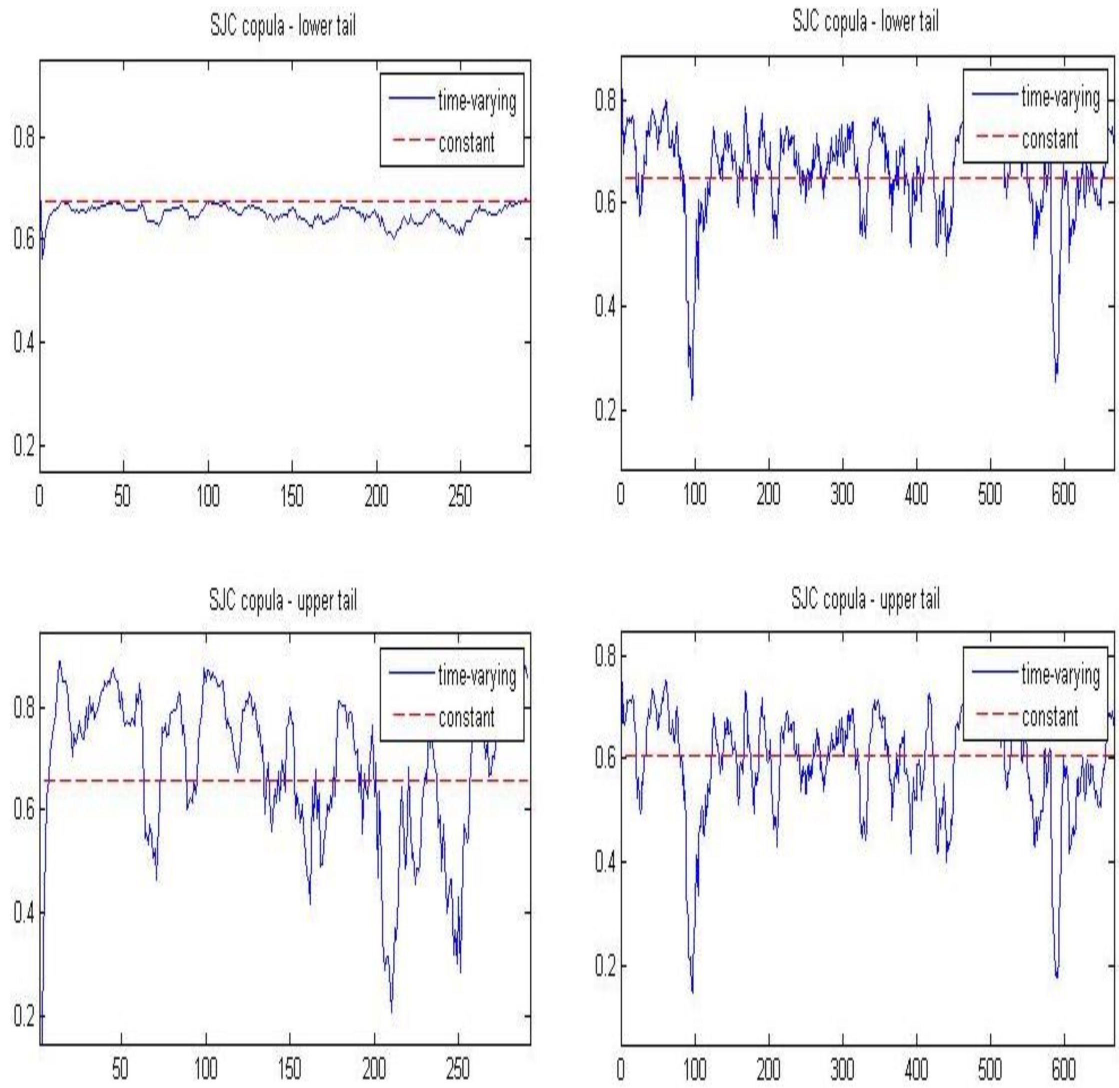


Figure A-14. Estimated Time-Varying Dependences of Lower and Upper Tails by Dynamic SJC Copula Function in Case of Futures Hedging

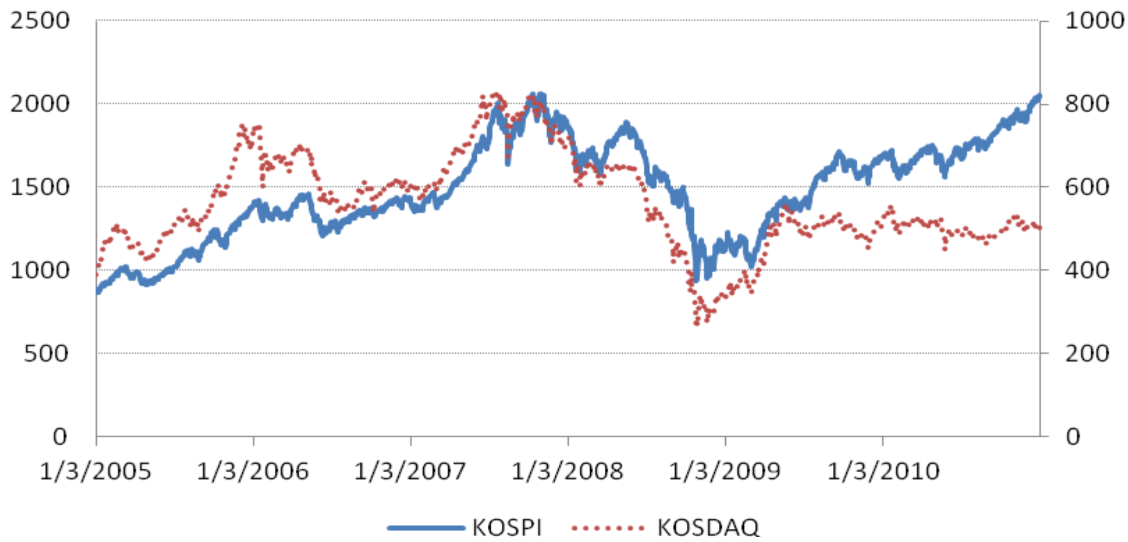


Figure A-15. KOSPI and KOSDAQ Indices from 2005 to 2010

Note: The abbreviations are KOSPI (Korean Composite Stock Price Index) and KOSDAQ (Korea Securities Dealers Automated Quotation). Left y-axis represents the KOSPI, while right y-axis represents the composite index of the KOSDAQ market.

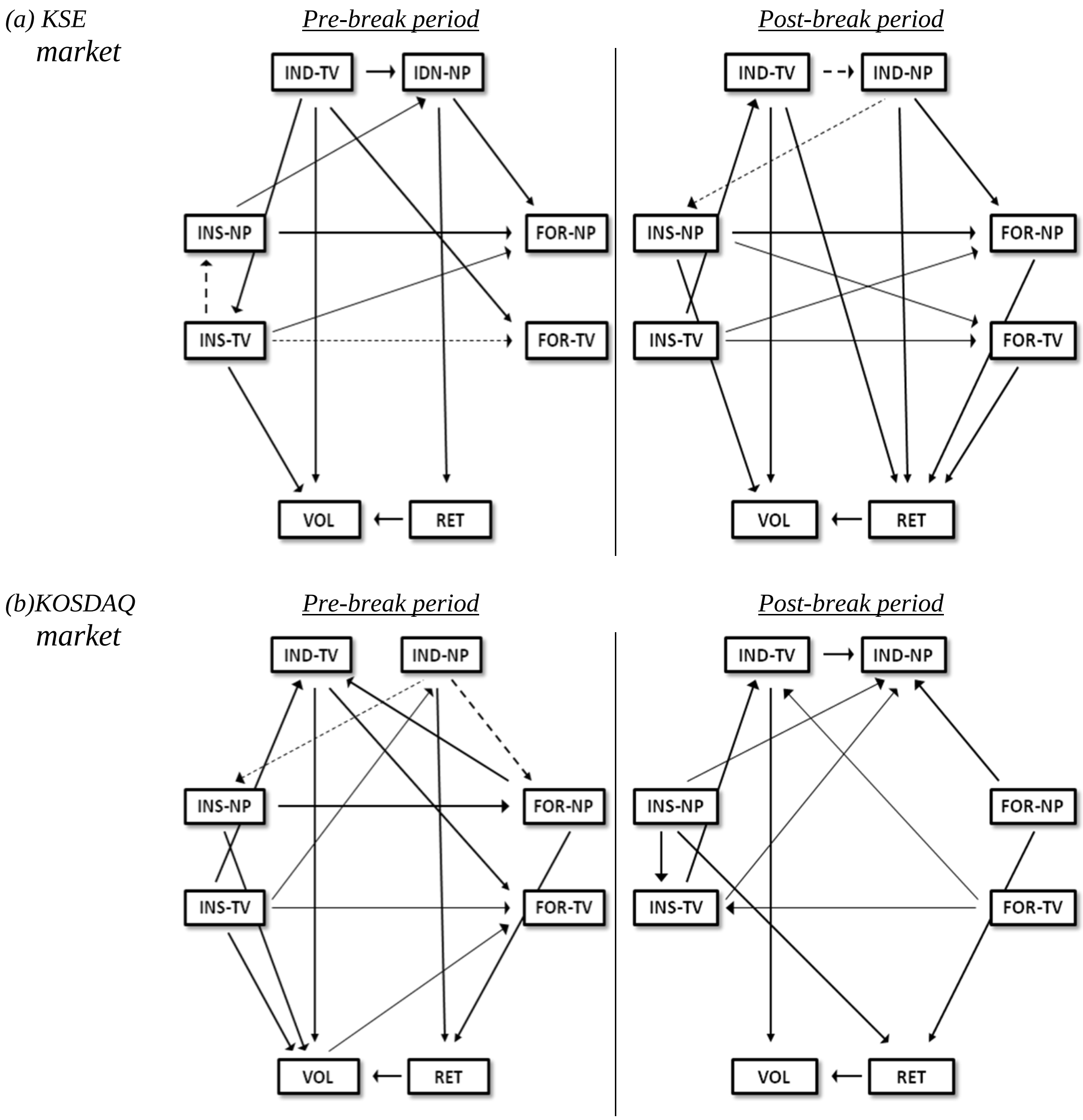


Figure A-16. Contemporaneous Causal Relationships in the Korean Stock Market

Note: The abbreviations are RET (return rate), VOL (volatility), INS-NP (institutional investor's net purchase), INS-TV (institutional investor's trade volume), IND-NP (individual investor's net purchase), IND-TV (individual investor's trade volume), FOR-NP (foreign investor's net purchase), and FOR-TV (foreign investor's trade volume). The dotted line is not assigned a direction by the TETRAD IV.

APPENDIX B

TABLES

Table B-1. Descriptive Statistics

Statistic	log price			first difference of log price (return)		
	oil	cs	etf	oil	cs	etf
Mean	4.3077	2.5145	4.3974	.0004	.0001	.0003
Median	4.2856	2.4348	4.3502	.0013	.0024	.0005
St Dev	.2892	.4820	.0976	.0178	.0862	.0119
Skewness	-.0163	.2501	.6531	-.3442	-.2550	-.7277
Kurtosis	2.390	2.6487	1.9142	4.1110	9.8810	5.1606
JB test (p value)	10.3553 (1.0e-003)	10.3677 (1.0e-003)	80.0621 (1.0e-003)	47.40 (1.0e-003)	1321.128 (1.0e-003)	187.528 (1.0e-003)

Note: In the JB test, the *p*-value is the probability that the data conform to the Normal distribution.

Table B-2. Unit Root Test

	Variable (X) and their first difference (ΔX)	ADF	KPSS
log(crude oil)	X	-1.757	2.64**
	ΔX	-25.233**	.103
log(crack spread)	X	-3.663*	4.80**
	ΔX	-28.374 **	.037
log(ETF)	X	-2.214	3.51**
	ΔX	-20.664 **	.049

Note: The ADF test is based on lag (2) with trend on level data and lag (1) without trend on first difference data. The null hypothesis of the ADF tests is the non-stationarity of the series, while the null hypothesis of the KPSS test is the stationarity of the series. *, ** denotes the rejection of the null hypothesis at the 5% and 1% levels, respectively.

Table B-3. Johansen Cointegration Maximum Likelihood Test

	maximum rank	Eigen value	Trace statistic	5% critical value
	0		19.7725	15.41
crude oil and crack spread	1	.0112	2.1320*	3.76
	2	.0014		
	0		25.8011	15.41
crude oil and ETF	1	.0066	1.9651*	3.76
	2	.0041		

Note: The null hypothesis of the Johansen test is there is no less cointegration equation than maximum rank level. The cointegrating vector of oil and crack spread is log(oil) – 1.073log(cs) -1.615 = 0, and the cointegrating vector of oil and ETF is log(oil)- 1.756(ETF) +3.225 = 0.

Table B-4. Wald Test-Granger Causality

	Entire period (Whole sample)	1st period (2005:10-2008:8)	2nd period (2008:8-2009:4)	3rd period (2009:5-2011:12)
$\Delta oil \rightarrow \Delta cs$	4.4691*	6.8845**	5.5237*	2.5905
$\Delta cs \rightarrow \Delta oil$	.8674	1.7873	0.7612	8.6193**
$\Delta oil \rightarrow \Delta ETF$				0.0303
$\Delta ETF \rightarrow \Delta oil$				178.91**

Note: Wald tests report the marginal probabilities associated with the Granger-causality test. *, ** denotes the rejection of the null hypothesis at the 5% and 1% levels, respectively. The null hypothesis is that all coefficients on the lag of the endogenous variable are jointly zero.

Table B-5. Error Correction Models with Structural Break

	ECM (oil&CS)						ECM (oil & ETF)	
	ECM1 model (Whole sample)		ECM2 (2005:10-2009:4)		ECM3 (2009:5-2011:12)		ECM4 (2009:5-2011:12)	
	Δoil_t	Δcs_t	Δoil_t	Δcs_t	Δoil_t	Δcs_t	Δoil_t	Δetf_t
ECT_{t-1}	-.0006	.0197**	-.0007	.0166**	-.0141*	.0759**	-.0233**	.0044*
Δoil_{t-1}	.2455**	.2869*	.2401**	.3180*	.2569**	.2742	.0241	-.0038
Δcs_{t-1}	.0043	-.0022	-.0031	.0138	.0219**	-.0220	-	-
Δetf_{t-1}	-	-	-	-	-	-	.7450**	- .0113
constant	.0003	9.9e-06	.0001	-3.4e-06	.0089	.0002	.0001	.0004

Note: *, ** denotes the *p*-value of the 5% and 1% levels, respectively. The cointegrating vector of ECM 3 is *log(oil)* -0.310*log(cs)* – 3.636 = 0, and ECM 4 is *log(oil)* – 1.745*log(ETF)* + 3.208 = 0.

Table B-6. Estimation Results of Multivariate GARCH Models

		ECM-MGARCH 1 (oil & CS) (2009:5-2011:12)		ECM-MGARCH 2 (oil & ETF) (2009:5-2011:12)	
		Δoil_t	Δcs_t	Δoil_t	Δetf_t
DCC	*arch (1)*	.1800**	.0815**	.0734**	.0312**
	garch(1)	.5923**	.8952**	.8743**	.9624**
	constant	.0000*	.0001*	9.1e-0.6*	8.6e-0.7
	λ_1	.0490		.0286*	
	λ_2	.7083**		.9218**	
CCC	*arch (1)*	.1759**	.0805**	.0624**	.026**
	garch(1)	.5397**	.8968**	.8915**	.9618**
	constant	.0000*	.0001*	7.1e-0.6*	9.1e-0.7
	conditional correlation	.1051**		.5990**	

Note: *, ** denotes the *p*-value of the 5% and 1% levels, respectively.

Table B-7. Forecast Error Statistics for Crude Oil Forecasting

		by Crack spread	by ETF	by RWM
ECM	*MAE*	.008965	.007734	.009264
	RMSE	.010914	.009943	.011206
	DM	-	.0000*	
ECM-MGARCH	*MAE*	.684738	.649659	.820236
	RMSE	.836930	.846540	.992240

Note: MAE and RMSE of the RWM model are derived by dividing the residual by the time invariant variance. DM reports the p-values of the Diebold and Mariano (1995) test, where the null hypothesis is that the related models provide equal forecast accuracy with oil:CS model. This H_0 is rejected at 0.1% significance level. Forecasting error has to be evaluated only in the same row in the table.

Table B-8. Model Framework

	Time invariant dependence	Time-varying dependence
Multivariate normal distribution	VEC	DCC MGARCH
Multivariate non-normal distribution	Static Copula GARCH	Dynamic Copula GARCH

Note: Static Copula GARCH include Gaussian, Clayton, Gumbel and SJC Copulas, and dynamic copula GARCH include their corresponding time-varying Copula models, such as time-varying Gaussian, time-varying Clayton, time-vary Gumbel, and time-varying SJC Copulas. However, as static Copula GARCH model also derive the time-varying hedge ratio, only VEC model is the static hedging model in this framework.

Table B-9. Summary Statistic and Correlation

	log price			first difference of log price (return)		
	spot	futures	ETF	spot	futures	ETF
Statistic						
Mean	.8382	.7818	3.6507	.0096	.0133	.0000
St Dev	.2739	.3012	.2898	.5351	.5703	.0257
Skewness	-.7094	-.7642	-.4182	7.2094	9.3204	-.4300
Kurtosis	3.5394	3.2996	2.6720	301.3519	272.6466	5.2243
JB test (*p value*)	92.1646 (1.0e-003)	97.0514 (1.0e-003)	32.2963 (1.0e-003)	3.56e+006 (1.0e-003)	2.92e+006 (1.0e-003)	227.49 (1.0e-003)
Correlation						
spot		.8651	.8286		.5669 (.5765)	.5830 (.5994)
futures	.9724		.8719	.1496 (.6758)		.7314 (.7789)
ETF	.9645	.9791		.0587 (.5950)	.2089 (.2754)	

Note: In the JB test, the *p*-value is the probability that the data conform to the Normal distribution. In correlation, lower triangle is the linear correlation and upper triangle is the Kendall's Tau. The figures in parenthesis are the value in the periods from 2009 to 2011.

Table B-10. Unit Root Test

	Variable (X) and first difference (ΔX)	ADF	KPSS
log(spot price)	X	-1.072	5.08**
	ΔX	-36.868**	.0105
log(futures price)	X	-1.271	4.9**
	ΔX	-21.977 **	.0184
log(ETF price)	X	-1.339	5.79**
	ΔX	-29.782 **	.0921

Note: The ADF test is based on lag (1) with trend on level data and lag (0) without trend on first difference data. The null hypothesis of the ADF tests is the non-stationarity of the series, while the null hypothesis of the KPSS test is the stationarity of the series. *, ** denotes the rejection of the null hypothesis at the 5% and 1% levels, respectively. In order to perform the ADF and the KPSS procedures for the system of two equations, the lag selection was based on the Final Prediction Error, using Akaike, Schwarz and Hannan-Quinn information criteria. A lag structure is selected as a result of majority rule among four criteria.

Table B-11. EG ADF Test and Johansen ML Test for Cointegration

	EG-ADF Test			Johansen Test	
	Test statistic	1% critical value	5% critical value	Trace Statistic (if max rank=1)	5% critical value
spot and futures	-4.913	-3.430	-2.860	2.0846*	3.76
spot and ETF	-4.913	-3.430	-2.860	2.3943*	3.76

Note: The null hypothesis of the Engle and Granger (1987) Cointegration Test (EG-ADF test) is there is no cointegration equation. The null hypothesis of the Johansen test is there is no less cointegration equation than maximum rank level. The cointegrating vector of spot and futures is log(spot) –0.8907log(futures) -0.1424 = 0, and the cointegrating vector of spot and ETF is log(spot)-0.9122(ETF) +2.4907 = 0.

Table B-12. Wald Test-Granger Causality

	Entire period (Whole sample)	1st period (2008:3-2009:4)	2nd period (2009:5-2011:12)
$\Delta spot \rightarrow \Delta futures$	4.192	1.296	1.253
$\Delta future \rightarrow \Delta spot$	100.230**	30.338**	1.871
$\Delta spot \rightarrow \Delta ETF$	12.415**	6.012*	1.531
$\Delta ETF \rightarrow \Delta spot$	100.230**	4.083	1.340

Note: Wald tests report the marginal probabilities associated with the Granger-causality test. * and ** denote the rejection of the null hypothesis at the 5% and 1% levels, respectively. The null hypothesis is that all coefficients on the lag of the endogenous variable are jointly zero, which means the no Granger casual relationship.

Table B-13. Estimation of VEC and DCC MGARCH Model in Futures Hedging

	Parameters	1st period (2008:3-2009:4) Spot (i=s)	1st period (2008:3-2009:4) Futures (i=f)	2nd period (2009:5-2011:12) Spot (i=s)	2nd period (2009:5-2011:12) Futures (i=f)
VEM	α_{0i}	-.0011	-.0024	.0009	.0008
		(.0027)	(.0024)	(.0009)	(.0008)
	α_{1i}	-.0471	.0228	-.0368	.0406*
		(.0299)	(.0261)	(.0208)	(.0177)
DCC	β_{0i}	.0000	.0000	.0001*	.0000**
		(.0000)	(.0000)	(.0000)	(.0000)
	β_{1i}	.1209**	.1025**	.0700**	.0594**
		(.0248)	(.0237)	(.0215)	(.0125)
	β_{2i}	.8879**	.9092**	.8109**	.9233**
		(.0196)	(.0181)	(.0587)	(.0167)
	λ_1	.2031**		.0763**	
		(.0457)		(.0292)	
	λ_2	.5874**		.8203**	
		(.0832)		(.0310)	

Note: *, ** denotes the *p*-value of the 5% and 1% levels, respectively. Figures in parenthesis are the standard error.

Table B-14. Estimation of VEC and DCC MGARCH Model in ETF Hedging

		1^{st} period (2008:3-2009:4)		2^{nd} period (2009:5-2011:12)	
	Parameters	*Spot (i=s)*	*ETF (i=etf)*	*Spot (i=s)*	*ETF (i=etf)*
VEM	α_{0i}	-.0022	-.0021	.0009	.0008
		(.0027)	(.0021)	(.0009)	(.0007)
	α_{1i}	-.0233	.0248	-.0227	.0235*
		(.0027)	(.0022)	(.0174)	(.0137)
DCC	β_{0i}	.0000	.0000	.0000**	.0000**
		(.0000)	(.0000)	(.0000)	(.0000)
	β_{1i}	.1201**	.0829**	.0737**	.1100**
		(.0257)	(.0194)	(.0202)	(.0309)
	β_{2i}	.8868**	.9197**	.8250**	.6906**
		(.0200)	(.0176)	(.0483)	(.0835)
	λ_1	.1403*		.0699**	
		(.0571)		(.0177)	
	λ_2	.6490**		.8168**	
		(.2093)		(.0368)	

Note: *, ** denotes the *p*-value of the 5% and 1% levels, respectively. Figures in parenthesis are the standard error.

Table B-15. Comparison of Copula Functions in Case of the Futures Hedging

Copula	1st period (2008:3-2009:4) LL	AIC	BIC	2nd period (2009:5-2011:12) LL	AIC	BIC
Gaussian	129.674	-259.351	-259.358	259.807	-519.617	-519.624
Clayton	114.121	-228.245	-228.252	235.196	-470.395	-470.402
Gumbel	134.646	-269.295	-269.302	269.649	-539.301	-539.308
SJC	136.068	-272.137	-272.146	268.319	-536.639	-536.648
Time-varying Gaussian	136.069	-272.141	-272.148	272.186	-544.375	-544.382
Time-varying Clayton	118.009	-236.021	-236.028	254.183	-508.369	-508.376
Time-varying Gumbel	<u>148.152</u>	-296.307	-296.314	284.102	-568.207	-568.214
Time-varying SJC	144.672	-289.345	-289.354	<u>285.107</u>	-570.215	-570.224

Note: LL means the log-likelihood values, AIC and BIC represent the Akaike information criterion and Bayes information criterion respectively. The underlined copula function shows the best fitness level in the each sub period.

Table B-16. Comparison of Copula Functions in Case of ETF Hedging

Copula	1st period (2008:3-2009:4)			2nd period (2009:5-2011:12)		
	LL	AIC	BIC	LL	AIC	BIC
Gaussian	154.679	-309.361	-309.368	303.815	-607.633	-607.640
Clayton	126.151	-252.305	-252.312	262.739	-525.481	-525.488
Gumbel	154.647	-309.297	-309.304	312.649	-625.301	-625.308
SJC	149.669	-299.339	-299.348	304.001	-608.003	-608.012
Time-varying Gaussian	160.266	-320.535	-320.542	314.753	-629.509	-629.516
Time-varying Clayton	132.074	-264.151	-264.158	271.792	-543.587	-543.594
Time-varying Gumbel	<u>167.199</u>	-334.401	-334.408	<u>319.025</u>	-638.053	-638.060
Time-varying SJC	163.052	-326.105	-326.114	311.653	-623.307	-623.316

Note: LL means the log-likelihood values, AIC and BIC represent the Akaike information criterion and Bayes information criterion respectively. The underlined copula function shows the best fitness level in the each sub period.

Table B-17. Estimation of C-GARCH Model in the Futures Hedging

Parameters	1st period (2008:3-2009:4)		2nd period (2009:5-2011:12)	
	Spot (*i=s*)	*Futures* (*i=f*)	*Spot* (*i=s*)	*Futures* (*i=f*)
Panel A:Estimates of marginal process				
α_{0i}	.0026 (.0008)	.0032 (.0009)	-.0004 (.0008)	.0006 (.0009)
α_{1i}	-.0471 (.0299)	.0228 (.0261)	-.0482** (.0162)	.0196 (.0141)
β_{0i}	.0000 (.0000)	.0000 (.0000)	.0000* (.0000)	.0000* (.0000)
β_{1i}	.1240* (.0505)	.1056* (.0475)	.0402** (.0205)	.0144** (.0464)
β_{2i}	.8746** (.0407)	.8942** (.0407)	.8745** (.0235)	.9760** (.1669)
df	6.8119 (3.030)	5.5617 (2.187)	6.0062 (1.485)	7.3435 (1.485)
Panel B: Estimates of time varying Gaussian dependence process				
ω_1	3.225** (.0180)		2.2224** (.5240)	
ω_2	-1.098 (.0368)		-.4688 (.8643)	
ω_3	.1512** (.0310)		.5963** (.1908)	
Panel C: Estimates of time varying Gumbel dependence process				
ω_1	2.048** (.4100)		2.7430** (.0276)	
ω_2	-5.002** (2.570)		-5.002** (1.2403)	
ω_3	-0.027 (0.105)		-.3407 (0.1146)	
Panel D: Estimates of time varying SJC dependence process				
ω_{U1}	3.676** (.4840)		.9146** (.0102)	
ω_{U2}	-1.157** (1.1130)		-.7132** (1.8234)	
ω_{U3}	-16.567* (1.8210)		-7.106* (1.0102)	
ω_{L1}	1.5160* (2.144)		2.7062* (2.004)	
ω_{L2}	-1.0751* (0.7602)		-1.138* (0.2562)	
ω_{L3}	-1.6141 (0.8935)		-9.2959 (0.3835)	

Note: *, ** denotes the *p*-value of the 5% and 1% levels, respectively. Figures in parenthesis are the standard error.

Table B-18. Estimation of Copula GARCH Model in the ETF Hedging

Parameters	1st period (2008:3-2009:4) *Spot* (*i=s*)	*ETF* (*i=etf*)	2nd period (2009:5-2011:12) *Spot* (*i=s*)	*ETF* (*i=etf*)
Panel A: Estimates of marginal process				
α_{0i}	.0033	.0029	.0003	.0005
	(.0018)	(.0018)	(.0010)	(.0008)
α_{1i}	-.0233	.0248	-.0227	.0235*
	(.0027)	(.0022)	(.0174)	(.0137)
β_{0i}	.0000	.0000	.0001	.0001
	(.0000)	(.0000)	(.0000)	(.0310)
β_{1i}	.1245*	.0882*	.0407	.0714*
	(.0512)	(.0464)	(.0244)	(.0216)
β_{2i}	.8741**	.9054**	.8267**	.7586**
	(.0414)	(.0378)	(.0916)	(.1384)
df	6.789	10.147	6.0326	11.345
	(2.870)	(6.604)	(1.654)	(4.881)
Panel B: Estimates of time varying Gaussian dependence process				
c	.7039**		4.661**	
	(.0312)		(.0251)	
ω_2	1.8047		-3.482	
	(.0254)		(.0235)	
ω_3	.4589**		.8243**	
	(2.621)		(2.264)	
Panel C: Estimates of time varying Gumbel dependence process				
ω_1	1.8605**		2.2146**	
	(.4100)		(.5905)	
ω_2	-4.356**		-3.4039**	
	(2.570)		(3.1746)	
ω_3	.0156		-.1992	
	(0.105)		(0.2703)	
Panel D: Estimates of time varying SJC dependence process				
ω_{U1}	-.6353**		.3.9200**	
	(.5528)		(.2319)	
ω_{U2}	-4.241**		-3.4212**	
	(.0918)		(1.0162)	
ω_{U3}	2.9673*		-9.5185*	
	(2.924)		(1.0861)	
ω_{L1}	3.2529*		-1.2832*	
	(1.285)		(1.687)	
ω_{L2}	-1.9405*		-3.2959*	
	(0.027)		(0.0069)	
ω_{L3}	-11.949		1.2566	
	(8.245)		(6.8672)	

Note: *, ** denotes the *p*-value of the 5% and 1% levels, respectively. Figures in parenthesis are the standard error.

Table B-19. Comparison of Hedge Performance of Futures and ETF Hedging

	Portfolio Variance		Variance Reduction over the VEC model (%)	
Model	*1st period*	*2nd period*	*1st period*	*2nd period*
Panel A: Futures Hedging				
VEC	0.939137	0.000470		
DCC GARCH	0.943552	0.000476	-0.004414 (-0.4700)	-0.000006 (-1.2783)
Normal copula	0.939050	0.000474	<u>0.000086</u> (0.0092)	-0.000004 (0.7988)
Gumbel copula	0.939137	0.000467	0.000000 (0.0000)	<u>0.000003</u> (0.5446)
SJC copula	0.939202	0.000468	-0.000064 (-0.0068)	<u>0.000002</u> (0.3328)
Time-varying Normal copula	0.943190	0.000505	-0.004053 (-0.4316)	-0.000035 (-7.4613)
Time-varying Gumbel copula	0.942906	0.000466	-0.003768 (-0.4012)	<u>0.000003</u> (0.7258)
Time-varying SJC copula	0.942918	0.000468	-0.003780 (-0.4025)	<u>0.000002</u> (0.3669)
average	0.941137	0.000474		
Panel B: ETF Hedging				
VEC	0.938962	0.000451		
DCC GARCH	0.943804	0.000449	-0.004842 (-0.5157)	<u>0.000002</u> (0.4657)
Normal copula	0.938920	0.000436	<u>0.000041</u> (.0044)	<u>0.000015</u> (3.3175)
Gumbel copula	0.939027	0.000455	<u>0.000064</u> (-0.0069)	-0.000004 (-0.7956)
SJC copula	0.939084	0.000467	-0.000122 (-0.0130)	-0.000016 (-3.6197)
Time-varying Normal copula	0.943548	0.000432	-0.004586 (-0.4884)	<u>0.000019</u> (4.1538)
Time-varying Gumbel copula	0.943052	0.000454	-0.004090 (-0.4356)	-0.000003 (-0.6046)
Time-varying SJC copula	0.942891	0.000467	-0.003929 (-0.4185)	-0.000017 (-3.6695)
average	0.941161	0.000451		

Note: Figures in parenthesis are the percentage of variance reduction compared to VEC. The underlined models show better hedge performance than basic VEC model.

Table B-20. Ceiling of Foreign Ownership in the Korean Stock Exchange

	Dec. 3. 1994	Jun.1. 1995	Apr.1. 1996	Oct.1. 1996	May 2. 1997	Nov.3. 1997	Nov.11. 1997	Dec.30. 1997	May 25. 1998
Collective ceiling	12%	15%	18%	20%	23%	26%	50%	55%	100%
Individual ceiling	3%	3%	4%	5%	6%	7%	50%	50%	100%

Table B-21. Proportion of Trading Volume by Investor Type in the KSE Market

	Institutional Investor (INS)		Individual Investor (IND)		Foreign Investor (FOR)	
Year	Purchase	Sales	Purchase	Sales	Purchase	Sales
2001	13.78%	14.01%	72.69%	73.23%	11.34%	10.36%
2002	13.94%	14.00%	71.53%	71.23%	11.64%	12.09%
2003	15.09%	16.64%	64.91%	65.60%	16.63%	14.74%
2004	15.47%	16.05%	57.42%	58.75%	23.32%	21.68%
2005	15.85%	14.82%	59.54%	60.81%	21.00%	21.34%
2006	20.46%	19.02%	49.63%	49.87%	25.85%	27.51%
2007	19.21%	18.91%	51.67%	51.28%	24.86%	26.32%
2008	22.77%	21.00%	49.44%	49.03%	24.11%	26.94%
2009	21.52%	23.08%	57.84%	57.93%	18.16%	16.29%
2010	21.44%	22.04%	54.31%	54.75%	20.98%	19.47%

Table B-22. Results of Augmented Dickey-Fuller Test

Market	RET	VOL	INS-NP	INS-TV	IND-NP	IND-TV	FOR-NP	FOR-TV
KSE	-5.64(7)	-27.20(1)	-11.04(4)	-4.11(5)	-22.75(1)	-4.24(4)	-10.06(4)	-4.43(8)
KOSDAQ	-8.89(5)	-10.07(9)	-18.68(1)	-5.77(4)	-21.09(1)	-4.68(4)	-12.76(4)	-3.82(8)

Note: The null hypothesis is that the series has a unit root. This hypothesis is rejected if the ADF test statistics is less than the critical value -3.43 (1%) given in Fuller (1976). Both an intercept and a time trend were included in the tests. The optimal lag length given in parenthesis was chosen using the Schwartz (1978) information criterion.

Table B-23. DAG Result of Foreign Investor's Activity to Stock Returns and Volatility

	KSE		KOSDAQ	
	Pre-break	Post-break	Pre-break	Post-break
Trade volume of foreign investor → Stock return	-	directly	-	-
Trade volume of foreign investor → Volatility	-	indirectly	-	indirectly
Net purchase of foreign investor → Stock return	-	directly	directly	directly
Net purchase of foreign investor → Volatility	-	indirectly	indirectly	indirectly

Table B-24. Information Flow among Three Types of Investors

	KSE		KOSDAQ	
	Pre-break	Post-break	Pre-break	Post-break
Root Cause	IND-TV	INS-TV	INS-TV	FOR-NP FOR-TV
Sink	FOR-NP	-	FOR-TV	IND-NP

VITA

Name:	Hankyeung Choi
Address:	Ministry of Strategy and Finance. Kyunggi-Do Kwacheon-Si Kwanmoon-ro 88, Republic of Korea 427-719
Email Address:	felixchoi20@gmail.com

Education:	B.A., Economics, The Yonsei University, 1995 M.E., Management Engineering, The Korean Advanced Institute of Science and Technology, 1998 Ph.D., Agricultural Economics, Texas A&M University, 2012