An Analysis of Oil ETFs and Crude Oil Price

by

Wenpei Yu

A thesis submitted in partial fulfillment
of the requirements for the
Degree of Doctor of Philosophy in Business Administration
in the Graduate College of the
Univeristy of Illinois at Chicago, 2015
Chicago, Illinois

Defense Committee:

Houston Stokes, Chair and Advisor
Stanley Sclove, IDS
Fangfang Wang, IDS
Matthew Liotine, IDS
Jim Man Lee, Economics

Summary

The purpose of this paper is to study the statistical relation between the price of oil ETFs and the underlying benchmark crude oil price.

Results from this study show that the oil ETFs price and underlying crude oil price are co-integrated in a nonlinear form. This finding is important and is consistent with the widely-held view that the dynamic relationship between time series in economics and financel is usually nonlinear. In applying tests for investigating the cointegration relationship, the classic Engle-Granger two-step method, dynamic approach and the nonlinear extension of cointegration have been used.

Granger Causality analysis suggests a two-way direction causal and feedback relationship between oil ETFs and the underlying crude oil price. This study also pays attention to the controversy whether its necessary to remove the unit root in series when testing Granger-causality. Spectral analysis and impulse response function are used to conduct the analysis. There is strong evidence that low frequency information has been discarded if using transformed data in research.

Finally, the linear and nonlinear Error Correction Model (ECM) has been established to estimate the interaction of oil ETFs and WTI, further to determine the lead and lag position. In the long-term relationship, oil ETFs (especially DBO) takes the lead to dominate the adjustment of price change to the long run equilibrium, while in the short term, crude oil price takes the lead in response of the price change of previous day.

Oil ETFs usually consist of oil future contracts, and the futures contracts prices are considered to provide information about future spot oil prices. Meanwhile, oil ETFs are tracking the price and movement of the current oil price. This makes the relationship between oil ETF and the underlying crude oil complex and interesting.

This paper studies the most popular oil ETF, United States Oil Fund (USO) and PowerShares DB Oil Fund (DBO) in the US market and their benchmark, The West Texas Intermediate (WTI) crude oil to investigate the equilibrium relationship between oil ETFs and the underlying benchmark. The data series is on a daily basis, and covers from the

launch date of oil ETF to March 31, 2012. Recursive residuals test and Stock-Watson test were conducted to test for stability of data series, and a structural break point has been detected over the whole period data. The breakpoint is around December 31, 2008, which is at the financial crisis in 2008. Therefore, the following analysis is performed over three data periods: the whole data period, sub-period(I) and sub-period(II), which are separated by the breakpoint.

Contents

Chapter 1. Introduction **1**

1.1 Motivation 1

1.2 Crude Oil Benchmark 2

1.3 ETFs and Oil ETFs Overview 2

1.4 Oil Futures 4

Chapter 2. Literature Review **6**

2.1 Literature on ETF 6

2.2 Literature on Oil Spot and Futures Market 7

Chapter 3. Data **9**

Chapter 4. Structural Breakpoint **12**

4.1 Structural Break 12

4.1.1 Testing for Structural Stability 12

4.1.2 Break Point for Data 14

4.2 Stock-Watson Test 18

4.3 Conclusion 23

Chapter 5. Granger Causality Analysis **24**

5.1 Order Selection for VAR Mode 25

5.2 Results of Granger Causality Test 27

5.3 Controversy regarding Stationary of Series 29

5.3.1 Spectral Analysis 29

5.3.2 Impulse Response Results 32

5.4 Conclusion 33

Chapter 6. Cointegration **36**

6.1 Cointegration Theory 36

6.2 Linear Cointegration 37

6.2.1 Engle and Granger Two-Step Methodology 37

6.2.2 Cointegration Testing Result 39

6.3 Dynamic Approach to Cointegration 43

6.3.1 Auto-regression Distributed Lag Model 43

6.3.2 Testing Results for Dynamic Model 44

6.4 Non-linear Cointegration Relationship 48

6.4.1 Non-linear Cointegration Generalization 48

6.4.2 ACE Algorithm 49

6.4.3 Nonlinear Cointegration testing results 50

Chapter 7. Error Correction Model (ECM) **53**

7.1 Error Correction Model (ECM) 53

7.2 Linear Error Correction Model 53

7.3 Nonlinear Error Correction - MARS 54

7.4 Results of Error Correction Model 55

7.4.1 Linear Model 55

7.4.2 ECM with MARS model 58

7.5 Alternative Nonlinear Approach - GAM 65

Chapter 8. Conclusion and Future Work **67**

8.1 Conclusion 67

8.2 Future Work 69

References **71**

Appendices **75**

Appendix A - VAR Lag Selection 76

Appendix B - VMA Coefficient for Impulse Response Function 79

(1) Level Data 80

(2) Differencel Data 86

Appendix C - Coefficients for Dynamic Model (6.17) & (6.18) 92

Appendix D - ACE Transformation Graph: Figure _ 6.6 & Figure _ 6.7 94

Appendix E - Knots and Slops for MARS (mi=2) model 96

Appendix F - Table _ 7.7 Coefficients for GAM ECM 103

List of Figures

1. ETFs Overview 3

2. Price on USO, DBO & WTI 10

3. Monthly Return For Oil ETFs and WTI 11

4. CUSUM plot for USO & WTI 14

5. CUSUMSQ plot for USO & WTI 15

6. QLR plot for USO & WTI 15

7. CUSUM, CUSUMSQ and QLR plot for DBO & WTI 16

8. Data separated into two sub-periods 17

9. Stock-Watson Value 21

10. Spectrums of Series (level data vs. difference data 30

11. Graph of Monte Carlo Integration 34

12. Scatter plot of residuals using Engle-Granger Cointegration model 41

13. Residual Correlogram for Dynamic Model 45

List of Tables

1. Oil ETFs Overview 4

2. Data Summary 10

3. Factual & Counterfactual Data for Stock-Watson Test 20

4. VAR Lag Selection 26

5. Granger Causality F Test 28

6. Dickey-Fuller Test for Engle-Granger Cointegration 40

7. Engle-Granger Test 41

8. ADF Test for Dynamic Model 48

9. R^2 value for linear model and ACE transformation 51

10. DF Test for ACE Residual 52

11. Coefficients for linear ECM model 56

12. MARS ECM (mi=1) for ΔUSO & ΔWTI in sample (1) 59

14. MARS ECM (mi=1) for ΔUSO & ΔWTI in sample (2) 61

15. MARS ECM (mi=1) for ΔUSO & ΔWTI in sample (3) 61

16. MARS ECM (mi=1) for ΔDBO & ΔWTI in sample (1) 62

17. MARS ECM (mi=1) for ΔDBO & ΔWTI in sample (2) 63

18. MARS ECM (mi=1) for ΔDBO & ΔWTI in sample (3) 64

19. R^2 value for linear and nonlinear ECM 66

Chapter 1. Introduction

1.1 Motivation

Crude oil is one of the most important and actively traded commodities in the world. The demand for oil is closely related to global economic growth. Crude oil also can be an important investment instrument for institutional and retail investors to diversify their portfolios or hedge against market changes. However, direct investment in physical oil is quite costly and not that practical. Oil Exchange Traded Funds offer investors an efficient way to gain exposure to various oil products without actually owning the oil itself. Most of these Oil ETFs have been designed to track crude oil price by investing in future contracts on crude oil.

The first oil exchange-traded fund, the United States Oil Fund, was introduced on April 10th, 2006. Since then, the oil exchange-traded fund (ETF) became a new investment instrument in financial market. Usually, the oil ETF invests in energy futures and other oil related futures, designed to track the movement of a specified benchmark oil price.

The oil ETF is a quite new instrument in the market, and in the family of Commodity ETFs. It is different from the traditional exchange-traded funds in asset holdings and tracking strategy. There are few quantitative studies to investigate their equilibrium relationship with the benchmark crude oil.

Most academic and empirical research on ETFs has been focused on the Index ETF, efficiency and performance, in comparison to other investment products, such as hedge funds and mutual funds. Furthermore, plenty of research has been done on the spot and future prices of crude oil, placing emphasis on the determinants of the spot price of crude oil, such as supply and demand, storage cost, and other economic factors. Additional work has been done whether or if the futures prices could be used as a predictor for spot oil prices, and the relationship between oil futures prices and oil spot prices. Research connecting the oil ETF and crude oil price, and concerning the dynamic relationship between the price of ETF and the underlying index is limited.

The main objective of this paper is to perform quantitative research on the relation between oil ETF and the crude oil price by investigating their equilibrium relationship. The ability of Oil ETFs to track the underlying Crude Oil Benchmark is also studied.

1.2 Crude Oil Benchmark

"Crude Oil, which is also known by the name Petroleum, is the most actively traded Commodity in energy Markets. The Largest Market for Oil is located in London and New York. There are a couple of popular benchmarks around the world that investors use for tracking the Market Price of Oil. The most recognized and widely quoted oil index in North America is known as The West Texas Intermediate or WTI. This benchmark reflects the market price for a single barrel of light sweet crude oil, most of which is pumped and refined in Texas and other locations along the Gulf Coast." - *from multiple sources*

1.3 ETFs and Oil ETFs Overview

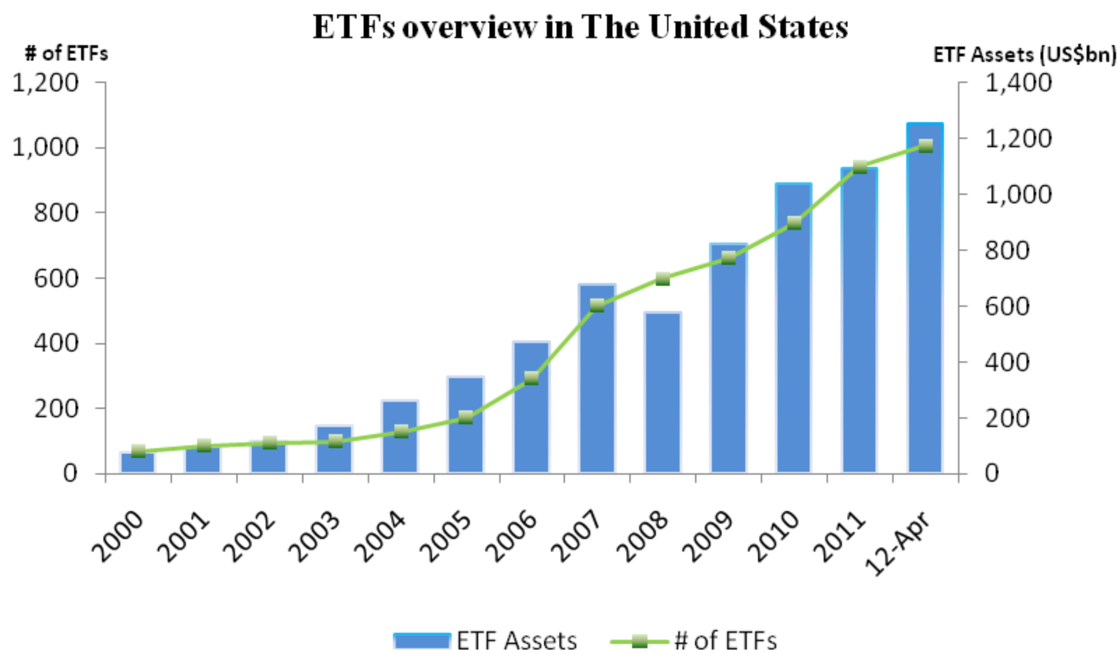
An Exchange-traded Fund (ETF) is a basket of securities, holding assets like stocks, commodities (such as precious metals and futures) and bonds. An ETF is traded at stock market like individual stocks. The most popular strategy of an ETF is to track a particular index, including broad market indices, such as FTSE All-World index and the MSCI US Broad Market index, major-indices, such as the S&P 500, Dow Jones Industrial Average and some are tracking the country index and cap-size index.

ETFs became immediately popular after they were made available in the US in 1993, and have become one of the fastest growing sectors in the financial market due to their advantages and features. ETFs have several advantages compared to the traditional mutual funds, they have lower operating expenses, can be traded more flexibly, as ETFs can trade throughout the trading day and the mutual funds only can trade at the end of trading day at their net asset value (NAV), and enjoy the tax-efficiency.

By the end of April 2012, the number of ETFs reached 1,175 in the United States with an estimated $1,075 billion in assets under management.

Figure_1.1 shows the explosive growth of the ETF sector in the United States regarding the number of ETFs and value of ETF assets since 2000.

Figure_1.1



Assets (US$bn)	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	Difference
ETF Assets	65.6	84.6	102.3	150.3	227.7	299.4	406.8	580.7	497.1	705.5	891	940.4	1074.5	1008.9
# of ETFs	81	101	113	117	152	201	343	601	698	772	896	1098	1175	1094

**Source: ETF Landscape April 2012, BlackRock*

ETFs can be divided into several types based on their structure, such as Index ETFs, Commodity ETFs, Bond ETFs, Currency ETFs and Leveraged ETFs. Oil ETFs are in the category of Commodity ETFs. Commodity ETFs invest in commodities, including energy, precious metals, softs and agriculture. Oil ETFs track the underlying index in the same way as other exchanged funds do. But the main difference between Oil ETFs and traditional ETFs is that traditional ETFs usually hold a basket of securities which comprise the underlying index, while Oil ETFs invest in the near term futures contracts of corresponding oil prices. The very first ETF for Oil is The United States Oil Fund ETF, which was launched on April 10, 2006.

The following are the four most popular ETFs which track the performance of crude oil, using the West Texas Intermediate (WTI) Benchmark:

Table _ 1.1 Oil ETFs Overview

ETF name	Trading Symbol	Launch Date	Market Cap	AverageDaily Trading Volume
United States Oil Fund	USO	April 10th, 2006	$1,210.00M	6,176,730
PowerShares DB Oil Fund	DBO	January 5th, 2007	$720.06M	392,592
United States 12 Month Oil Fund	USL	December 5th, 2007	$101.14M	16,241
Teucrium WTI Crude Oil Fund	CRUD	February 23rd, 2011	$2.04M	536

**Source: seekingalpha.com on 1/22/2013*

In this paper, we will focus our study on the USO and DBO oil ETFs, both of which track the movement of the WTI, crude oil benchmark, and are actively traded with high market caps.

1.4 Oil Futures

To understand the Oil ETF, one must understand the nature of oil futures contracts, since Oil ETFs are mainly composed of oil futures in different maturity periods. Oil futures contracts are agreements between buyers and sellers on the price of the oil upon delivery at a future designated date. The prices of futures reflect the market expectation for the spot oil price in the future. The unit of the oil futures contract is 1,000 barrels of oil. The New York Mercantile Exchange (NYMEX) and the Intercontinental Exchange (ICE) are the major markets for the trading of oil futures contracts.

The relationship between oil futures prices and spot oil price has drawn a lot of interest in academics and empirical research. Oil futures prices are widely used as predictors of spot oil prices and thought to be better than forecasts using economic models. However, Alquist and Kilian (2010) looked at it in another light, they concluded that the price of crude oil futures is not the most accurate predictor of the spot price of crude oil in practice.

The remainder of this study is organized as follows: Chapter 2 provides a literature review on ETF and energy futures research. Chapter 3 describes the data used for this study. The statistical analysis begins in Chapter 4 with the testing of structural stability of the data. Chapter 5 performs Granger Causality analysis. In Chapters 6 and 7, we develop models using three samples for the whole periods and sub-periods of data. A non-linear co-integration relationship is developed in Chapter 6 between ETFs and underlying crude oil price, and error correction models are applied for estimation in Chapter 7. Chapter 8 presents the conclusion.

Chapter 2. Literature Review

Prior to the year 2000, there was little research on ETFs. Laurent Deville (2006) noticed that despite the increasing importance of ETF markets, literature on these topics was still scarce, although research perspectives were promising. Research on ETFs considered mainly of empirical studies and were mostly focused on ETF characteristics and performance (i.e. return and tracking error) relative to other investment instruments, such as mutual funds, Index funds and etc. To our knowledge, there is no cointegration study on oil ETFs and crude oil. Following are some of the studies on ETFs and quantitative research on oil spot and futures price.

2.1 Literature on ETF

Elton, Gruber, Comer and Li (2002) examined the performance of SPDR or Spiders[1], which is the most actively traded exchange-traded Index fund to replicate the S&P Index. They found out that Spiders underperformed the S&P Index by 28 basis points and low-cost index funds by 18 points over the 1993 1998 period. Gallagher and Segara (2006) investigated the ability of classical ETFs to track underlying equity benchmarks on the Australian Stock Exchange. They examined the tracking errors of ETFs on the Australian stock exchange and compared the tracking error volatility between ETFs and equity index funds operated off-market. They argued that ETFs are better at tracking their benchmarks than off-market index funds, and concluded that classical ETFs in Australia provided investors with returns commensurate with the underlying benchmark before costs. Patrick Chu (2011) studied the magnitude of tracking error[2] and the determinants of tracking errors using the daily figures of the ETFs traded in the Hong Kong (HK) stock market. He found out that the tracking errors for ETFs traded in HK stock exchange are comparatively higher than those documented in US and Australia. Shin and Soydemir (2011) estimate tracking errors and relative performance of 26 ETFs over their benchmark indexes. They found that tracking errors are significantly different from zero and display

[1] SPDR is an exchange-traded fund to track the S&P 500 index, also known as Spiders. The symbol on NYSE is 'SPY'. It is one of the largest ETFs in the world.

[2] tracking error measures the divergence between a ETF and benchmark

persistence. They also examined the factors affecting tracking errors, such as expense ratio, dividends, exchange rate and spreads of trading prices and found out that the main factor driving tracking errors is the change in the exchange rate.

Gastineau (2004) compared the benchmark index ETFs to indexed mutual funds by investigating the difference in returns between the iShares Russell 2000 ETFs and Vanguard Small Cap Index Fund over 1994 to 2002. The results show that ETFs underperform their corresponding mutual fund. Kostovetsky (2003) developed a simple one-period model to examine the major differences between ETFs and index funds. The key areas of difference were management fees, share-holder transaction fees, taxation efficiency, as well as other qualitative differences. All these features are attractive to more active larger investors.

Murdock and Richie (2008) checked the correlation between the oil ETF USO and crude oil futures contracts to determine if USO could be an effective hedging instrument against oil price. They found that the spread of USO and crude oil futures deviate more during periods of contango.

2.2 Literature on Oil Spot and Futures Market

P. Silvapulle and I. A. Moosa (1999) used the Baek-Brock nonparametric test[3] to detect the presence of nonlinear causal relations between spot and futures crude oil prices, and mentioned the nonlinear relation is mainly due to transaction cost functions, marerket noises, etc. They concluded that the feedback between spot and futures market was bidirectional, and that both spot and futures markets react to market information simultaneously.

Westgaard, Estenstad, Seim and Frydenberg (2011) investigated the co-integration relationship between gas oil futures and Brent crude oil futures prices. Daily data was used for five different futures price with maturity of one, two, three, six and twelve months, covering the period from 1994 to 2009. Pair-wise data for different maturity periods were tested for co-integration using the Engle-Granger methodology and Johansen approach. A

[3] Proposed by Baek-Brock in 1992. It is designed to test for nonlinear causal relations by using concept of correlation integral.

co-integration relationship was found for 1 and 2 month contracts over the data period. But no co-integration existed in sub-periods from 2002 to 2009 for all pair-wise data series with five different contracts length. The author explained that the result for the sub-period was mainly due to the market volatility during the period, caused by certain events such as hurricane Katrina, the economic boom and the financial crises.

Bekiros and Diks (2008) examined the linear and nonlinear causality relationship between daily spot and futures prices for contract lengths of one, two, three and four months of WTI crude oil. They split the data into two sample periods of 10/21/1991 to 10/29/1999 and 11/1/1999 to 10/30/2007. The results showed that Granger causality between spot and futures prices in both periods is in bi-direction. Moreover, if account for the nonlinear effects, the leads or lags of spot and futures in crude oil market changes over time.

Alquist and Kilian (2010) concluded that the price of crude oil futures tends to be less accurate in predicting the spot price of crude oil. They used the data from two countries, the United States and Saudi Arabia to build a general equilibrium model. The futures-based forecast (based on oil futures and oil futures spread) and the no-change forecast, which was estimated under a loss function, were made over 5 periods from 1 month to 12 months on oil prices in spot and future markets. The most robust finding was that the no-change forecast performs better than futures-based forecast in predicting the spot price of crude oil.

Chapter 3. Data

The data for this study is comprised of the daily closing prices of two ETFs - the United States Oil Fund (USO), the Deutsche Bank's PowerShares DB Oil Fund(DBO)which track the price of oil the light, sweet crude oil benchmark WTI. The data series for USO and DBO are the daily closing prices downloaded from finance.yahoo.com, and the daily crude oil price / per barrel for WTI are obtained from the U.S. Energy Administration website.

The USO Fund is the first crude oil based fund launched on April 10, 2006 at a price of $67.84 / share. The fund is designed to track the movement of WTI. The USO portfolio consists of crude oil futures, other oil related futures, and some short-term US Treasury Securities. The principle mechanism of USO tracking WTI is the percentage change of the net asset value (NAV) on a daily basis to reflect the daily price / barrel change in the spot price of crude oil WTI.

Another popular oil ETF is the DBO Fund. The fund is introduced by Deutsche Bank as:" The fund is designed to track the market performance of crude oil, which it achieves by following the performance of a benchmark known as the Deutsche Bank Liquid Commodity Oil Index. This index is comprised of light sweet crude oil futures contracts as well as investments made in highly liquid short-term financial instruments such as 3 month United States Treasury Bills. These ETF Shares first began trading on the New York Stock Exchange on January 5th, 2007".

WTI light, sweet crude oil is the benchmark for crude oil price in the US. Another primary US benchmark for oil price is the prices of Brent crude from the North Sea. There are some other important oil benchmarks, including the Dubai Crude and the OPEC Reference Basket.

The sample period of USO and DBO is different with respect to the launch date of the ETF. The sample period for the USO data set is from April 12, 2006 to March 31, 2012. It includes 1,501 observations. The sample period for the DBO data set is from January 5, 2007 to March 31, 2012 with 1,319 observations. In addition the corresponding ranges of WTI are used to match the sample periods for USO and DBO.

Table 1 gives the mean prices for the USO and DBO ETFs and the crude oil WTI benchmark.

Table _ 3.1 Data Summary

Variable	Obs.	Start	End	Mean	Std Dev.	Max.	Min.	DF test
USO	1501	April 10, 2006	March 31, 2012	50.28	19.92	117.48	22.86	-1.034
DBO	1319	January 5, 2007	March 31, 2012	28.82	6.85	55.01	15.83	-0.589
WTI	1501	April 10, 2006	March 31, 2012	80.80	20.59	145.31	30.28	-0.004

A plot of the raw data in Figure_3.1 shows that USO, DBO and WTI tend to move together at the beginning, and then deviate more and more gradually. The effects of the financial crisis and recession in year 2008 are quite visible. The financial crisis in 2008 had a big impact on the prices for USO, DBO and WTI, as all of them suffered a huge drop.

Figure_3.1 produced by using raw data series for price of USO, DBO & WTI Figure_3.1

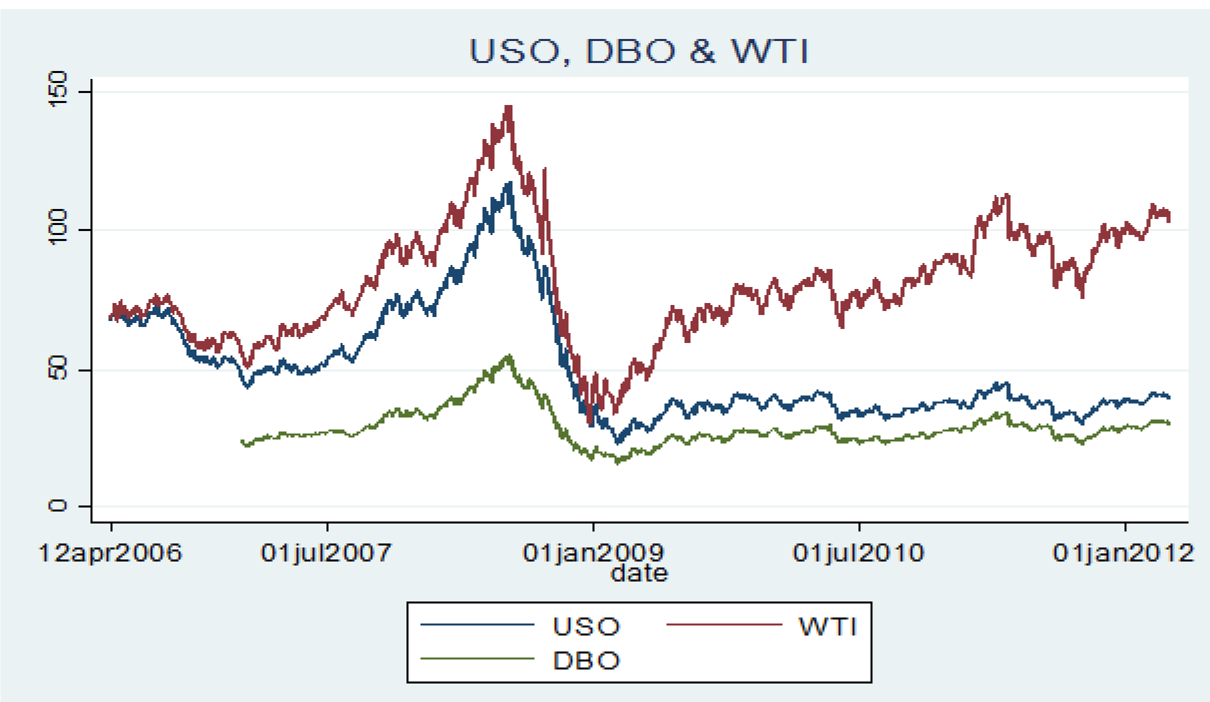
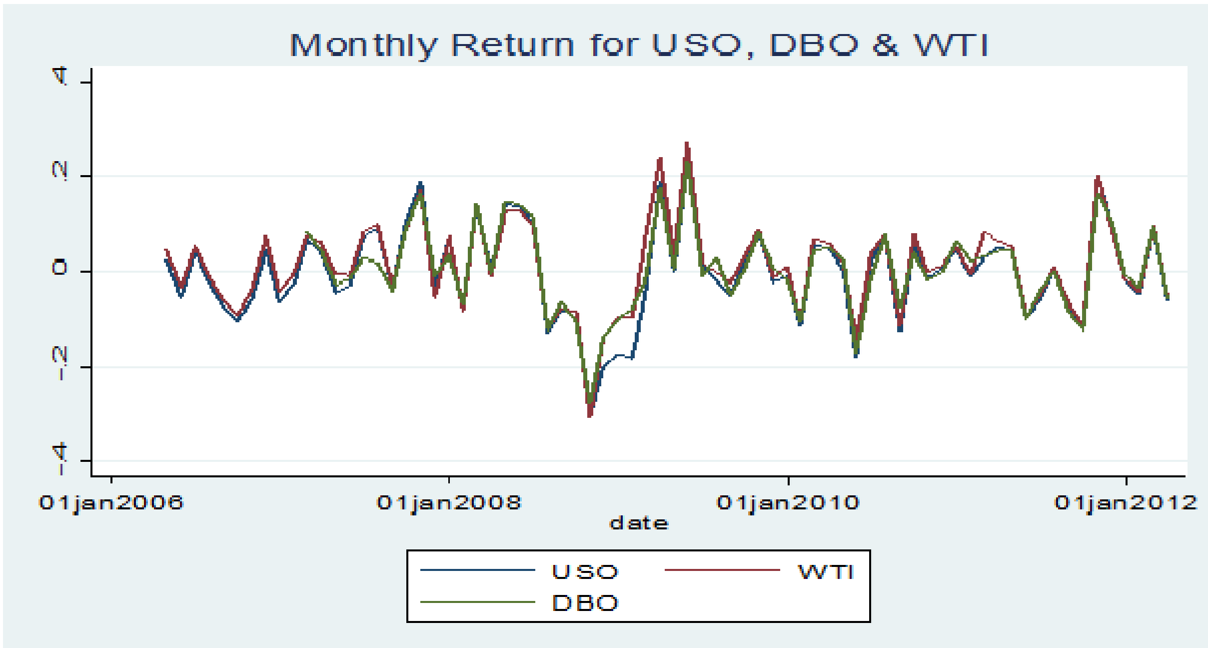


Figure 3.2 Gives the Monthly Return for Oil ETFs and Benchmark Figure 3.2



Chapter 4. Structural Breakpoint

4.1 Structural Break

4.1.1 Testing for Structural Stability

We begin our analysis with the test for structural stability of data series for USO and DBO ETF, and crude oil WTI benchmark. This step is the most important because the estimates can be severely biased if the time series are indeed realizations of not a stable process.

Tests for structural stability are much discussed in the literature. Nyblom (1989) proposed the sup-F test to detect possible changes in parameters. Brown, Durbin, and Evans (1975) made an important contribution by assessing the constancy of regression coefficients, calculating updated coefficient vectors as additional observations are added to the regression. Stokes (1997) discussed thoroughly the Recursive Residuals (RR) procedure to detect the locations of potential structural breaks in a series. For an Ordinary Least Squares (OLS) procedure, the OLS residuals can be heteroscedastic and auto-correlated even when the true errors are white noise. The Recursive Residuals procedure transforms the OLS residuals, since they are not BLUE so that they do satisfy the OLS assumptions. The technique begins with estimating OLS and then calculating updated coefficient vectors as additional observations are added to the regression, while the recursive residuals satisfy OLS properties, and are identically and independently distributed as normal with mean and standard deviation σ, which we will denote as i.i.d.$\sim$ N(0, σ).

Stokes(1997) mentioned that the cumulated sum of recursive residuals test (CUSUM), cumulated sum of squared standardized recursive residuals test (CUSUMSQ) and the Harvey-Collier(1977) test are the three most important summary tests for parameter stability. CUSUM test and CUSUMSQ test were proposed by Brown, Durbin and Evans (1975). If the break is not known, CUSUM and CUSUMSQ will be more appropriate. Also, Quandt log-likelihood ratio test (QLR) is another important technique suggested by Quandt(1960) to detect the unknown breakpoint.

Here, we will use CUSUM, CUSUMSQ and QLR test to perform the breakpoint analysis.

(i) CUSUM test computes: $\Gamma_i = \dfrac{\sum_{j=K+1}^{i} \omega_j}{\hat{\sigma}}$, $i = K+1 \ldots T$

ω_j is the standardized recursive residual, $\hat{\sigma}^2$ is estimated variance of ω_j.
If the series is stationary, $E(\Gamma_i) = 0$

(ii) CUSUMSQ based on $\Gamma_i^* = \dfrac{\sum_{j=K+1}^{i} \omega_j^2}{\sum_{j=K+1}^{T} \omega_j^2}$, $i = K+1 \ldots T$

ω_j is the standardized recursive residual.
If the series is stationary, $E(\Gamma_i^*) = \dfrac{i-K}{T-K}$ varies from 0 (i=K) to 1 (i=T)

(iii) Quandt log-likelihood ratio test involves the calculation of λ_i

$\lambda_i = 0.5 * i * ln(\sigma_1^2) + 0.5 * (T-i) * ln(\sigma_2^2) - 0.5 * T * ln(\sigma^2)$

Where σ_1^2, σ_2^2, and σ^2 are the variances of regressions fitted to the first i observations up to the breakpoint, the last T-i observations after breakpoint and the whole T observations, respectively.

If the series is stationary, λ_i will be close to 0. The minimum value for λ_i can be used to determine the breakpoint.

The CUSUM test detects breaks in the model itself, while CUSUMSQ examines breaks in the variance. The CUSUM test is particularly good at detecting systematic departure of the coefficients that results in a systematic sign on the first step ahead forecast error. The CUSUMSQ test is useful when the departure of the coefficients from constancy is haphazard rather than systematic but that there involves a systematic change in the accuracy of the estimated equation as observations are added. The plotting of Quandts log-likelihood ratio (QLR) statistic is used to detect the single time-point at which there is a discontinuous change from one constant set of regression parameters to another.

The advantage of these test statistics is that they can be graphed, and we can identify not only their significance but also at what time point a possible break occurred.

4.1.2 Break Point for Data

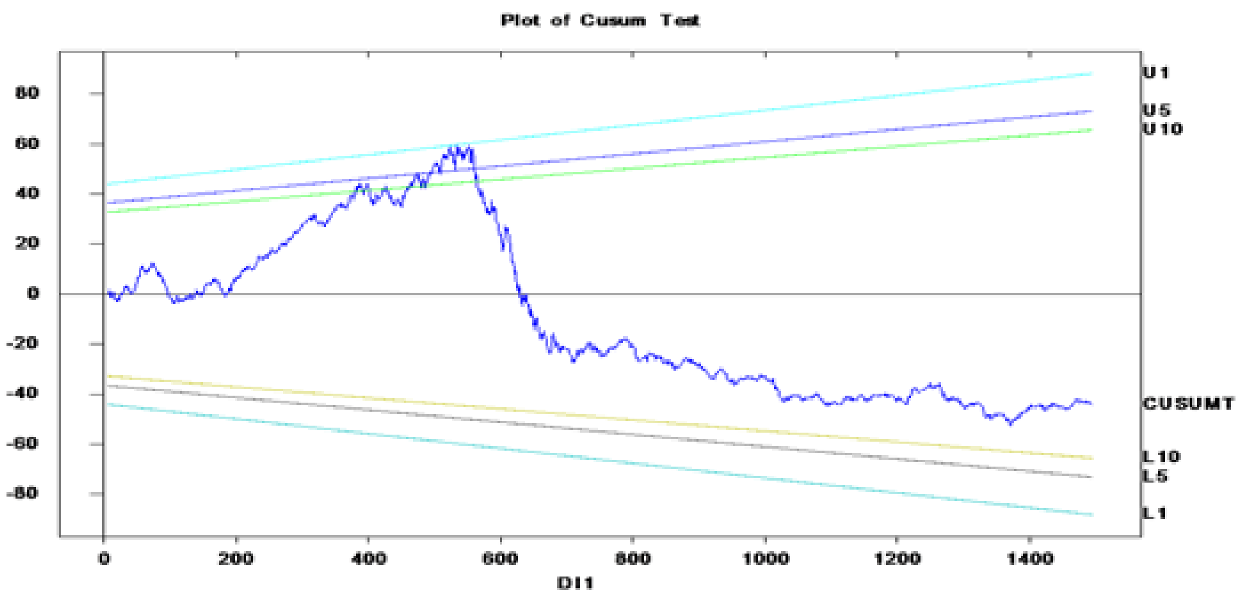
The recursive residuals plots in Figure 4.1, 4.2, 4.3 and Figure 4.4 below show the results of CUSUM, CUSUMSQ and QLR test for price of USO & WTI and price of DBO & WTI to examine the breakpoint in the data for the complete sample.

Plots in Figure 4.1 - 4.3 are based on the OLS model (4.1):

$$USO_t = 1.007 * USO_{t-1} + 0.1853 * USO_{t-5} - 0.1949 * USO_{t-6} - 0.1625 * WTI_{t-5} + 0.1625 * WTI_{t-6} \quad (4.1)$$

(t=74.53) (3.33) (-3.55) (-4.76) (4.77)

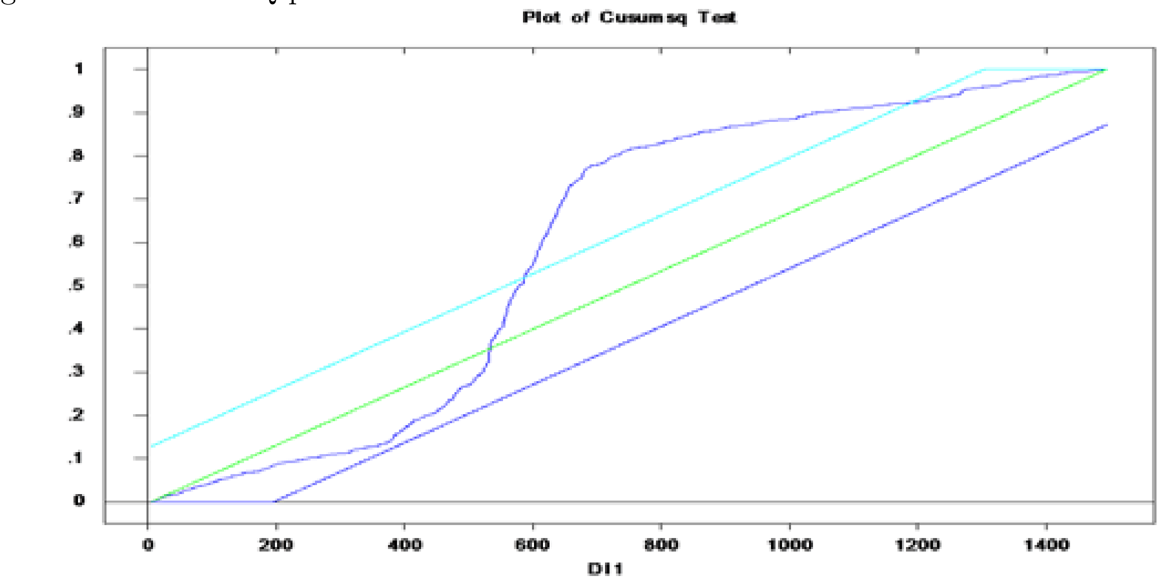
Figure_4.1 CUSUM plot for USO & WTI



CUSUM plot for USO price and crude oil price during 4/12/2006 to 3/31/2012

The CUSUM test statistics plotted in Figure 4.1 are almost inside the CUSUM confidence bounds. But we noticed that during Apr. 1st 2008 to Aug. 31st 2008 (around n=450 to n=600), it shows instability, although still within 99% confidence level, but out of the 95% bound. The plot experienced a sudden change around n=600 to n=700, the corresponding tie period is August 30, 2008 to January 23, 2009.

Figure_4.2 CUSUMSQ plot for USO & WTI

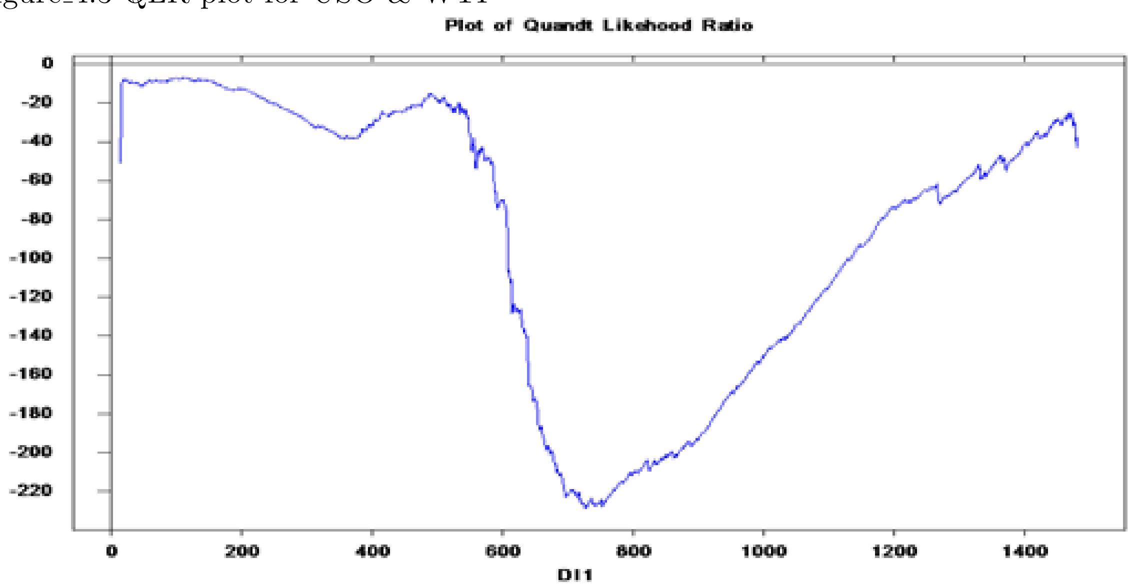


CUSUMSQ plot for USO price and crude oil price during 4/12/2006 to 3/31/2012

The CUSUMSQ plot clearly shows the instability as it breaks the upper confidence bound, which suggests that there should be model instability at some time-points during time period we study.

The maximum statistics moves away from the upper bound around n=680 to n=700, which corresponds to the time period of December 23, 2008 to January 23, 2009.

Figure_4.3 QLR plot for USO & WTI



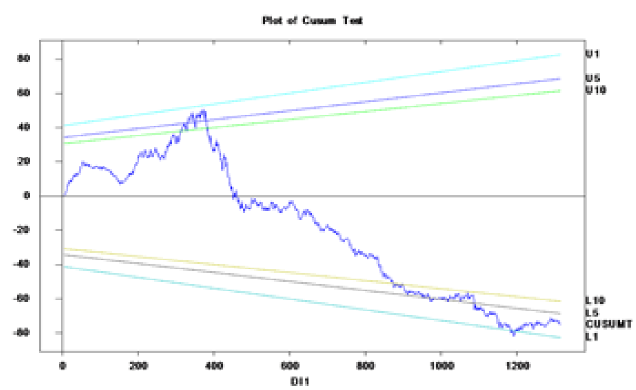
QLR plot for USO price and crude oil price during 4/12/2006 to 3/31/2012

The QLR shows that there is a dramatic drop from around n=575 to n=730, corresponding to the time period July 8th, 2008 to March 9th, 2009, indicating that the coefficient shifts during this period.

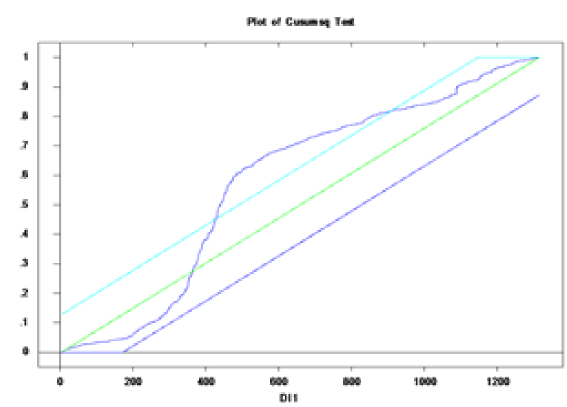
A further check of the USO and WTI prices in this period show that prices of USO experienced a tremendous decrease from $117.48/share on July 14, 2008 to $22.86/share on February 18, 2009, which represents a cumulative loss of 80.54% over 7 month and price of WTI dropped 79.16% from $145.31/barrel on July 3, 2008 to $30.28/barrel on December 23, 2008. This suggests breaking the data series into two periods: April 12th, 2006 to December 31, 2008 and January 2nd, 2009 to March 31, 2012.

Examining the data series for DBO and WTI:

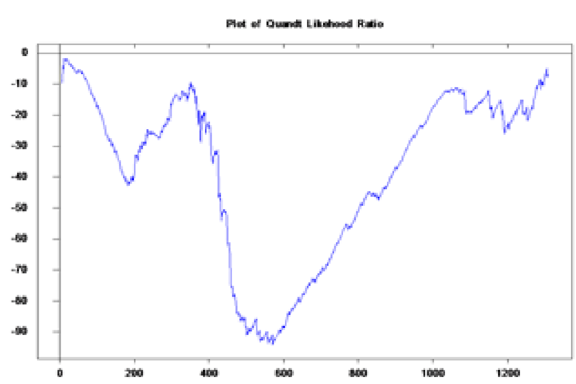
Figure_ 4.4 CUSUM, CUSUMSQ and QLR plot for DBO & WTI



CUSUM plot for DBO price and crude oil price during 1/5/2007 to 3/31/2012



CUSUMSQ plot for DBO price and crude oil price during 1/5/2007 to 3/31/2012



QLR plot for DBO price and crude oil price during 1/5/2007 to 3/31/2012

Plots in Figure 4.4 are based on the OLS model (4.2):

$$DBO_t = 0.8628 * DBO_{t-1} + 0.1304 * DBO_{t-4} + 0.0360 * WTI_{t-1} - 0.0350 * WTI_{t-5} \quad (4.2)$$

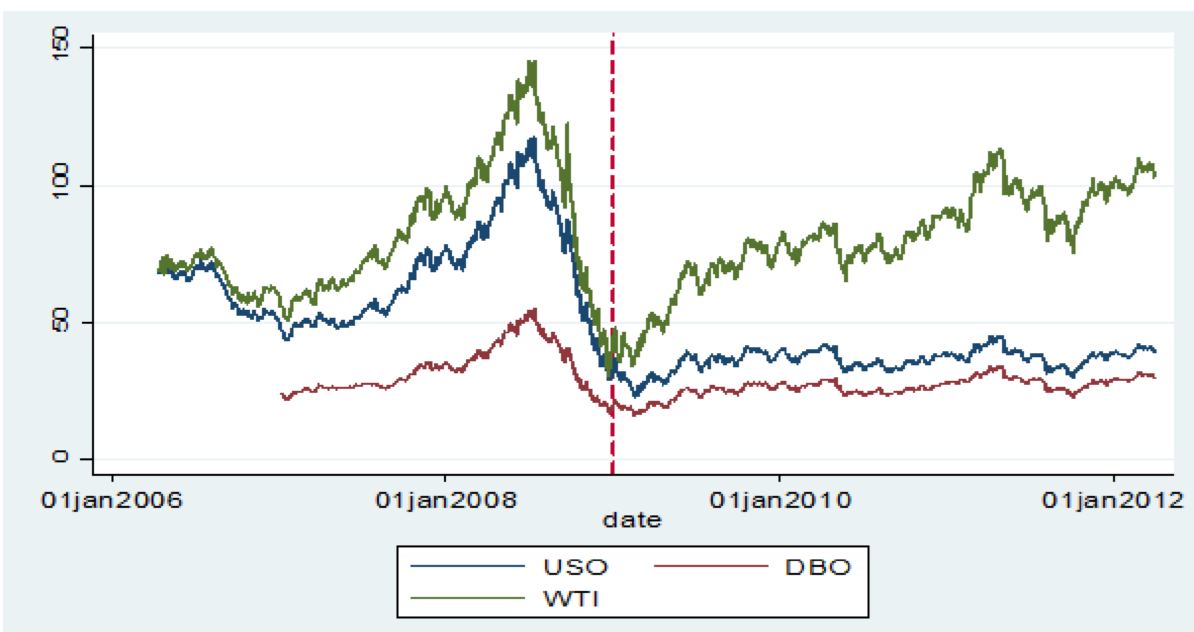
(t=32.40) (4.95) (5.13) (-5.03)

The recursive residuals plots for DBO & WTI are very similar to those of USO & WTI. CUSUM is almost within the confidence bound, CUSUMSQ goes outside of the upper bound to indicate the coefficient shift. QLR experienced tremendous drop during July 2008 to March 2009. The price of DBO dropped from $55.01/share on July 14th, 2008 to $15.83/share on February 18th, 2009. Again, the breakpoint would be set on December 31, 2008. The data series would then be truncated into two subsets for January 5th, 2007 to December 31st, 2008 and January 5th, 2009 to March 31st, 2012.

The breakpoint at December 31, 2008 just dropped in the outbreak period of 2008 financial crisis, and separated the full sample into two sub samples of pre-crisis and post-crisis.

Figure_4.5 shows the entire data series separated into two sub-periods by Dec. 31, 2008

Figure_4.5



4.2 Stock-Watson Test

In previous section, a distinct break in volatility for data series USO, DBO and WTI has been detected, which is located around December 31, 2008. The main focus of this section is on whether the change of volatility pre and post the break point is associated with the shocks or the structure.

Stock and Watson (2002) proposed a test using counter-factual Vector Auto-regression (VAR) analysis to distinguish between changes in the coefficients or the changes in the variance of a VAR model.

The two variables VAR model will be estimated in this study.

Suppose y_t is a vector time series, $y_t = \begin{pmatrix} USO_t \\ WTI_t \end{pmatrix}$, or $y_t = \begin{pmatrix} DBO_t \\ WTI_t \end{pmatrix}$, the VAR model for y_t:

$$y_t = \Phi(B)y_{t-1} + a_t \tag{4.3}$$

where $Var(a_t) = \Sigma$, $t = 1 \dots n$

Denote τ the break point, the full sample can be separated into two periods: sub-sample periods (1) when $t \leq \tau$, and sub-sample period (2) when $t > \tau$.

The VAR model for each period has the form:

$$y_{1t} = \Phi_1(B)y_{1,t-1} + a_{1t}, \quad Var(a_{1t}) = \Sigma_1, \quad t \leq \tau \tag{4.4}$$

$$y_{2t} = \Phi_2(B)y_{2,t-1} + a_{2t}, \quad Var(a_{2t}) = \Sigma_2, \quad t > \tau \tag{4.5}$$

When $\Phi_1 = \Phi_2$, there is no change in coefficient over the sub-periods. When $\Sigma_1 = \Sigma_2$, that means the variance has not changed from sub-sample (1) to sub-sample (2). The purpose of the Stock-Watson test is to check if $\Phi_1 = \Phi_2$ or $\Sigma_1 = \Sigma_2$ for equation (4.4) and (4.5).

The counter-factual procedure is to evaluate the equations (4.4) and (4.5), and obtain the factual standard deviation for y_{1t} and y_{2t}, then allow the estimated $\Phi_1(B)$ to replace $\Phi_2(B)$

in (4.5) to get the counter-factual standard deviation for y_{2t}, and use $\Phi_2(B)$ in (4.5) to replace $\Phi_1(B)$ in (4.4) to get the counter-factual standard deviation for y_{1t}.

Define $Var(y_{it}) = \sigma(\Phi(i), \Sigma_j)^2 \equiv \sigma_{ij}, \quad i = 1, 2 \quad and \quad j = 1, 2$

σ_{11} and σ_{22} are the variance for y_{1t} and y_{2t} that actually occurred in sub-sample period (I) and (II), respectively. σ_{12} and σ_{21} represent the counter-factual variance which happened that the coefficient and innovation variance come from different time periods.
$\sigma_{12} = \sigma(\Phi_1, \Sigma_2)^2$ represents the counter-factual variance estimated by using $\Phi_1(B)$ in equation (4.4), and $\sigma_{21} = \sigma(\Phi_2, \Sigma_1)^2$ represents a counterfactual variance produced by using the combination of second period dynamics and first-period shocks.

If $\sigma_{11} = \sigma_{12}, \quad \sigma_{21} = \sigma_{22}$, then there is no change in coefficient. If $\sigma_{11} = \sigma_{21}, \quad \sigma_{12} = \sigma_{22}$, that variance is stable over the two time periods. So the appropriate values to test are $\mid \sigma_{11} = \sigma_{21} \mid, \quad \mid \sigma_{12} = \sigma_{22} \mid$ testing for coefficient shifts, $\mid \sigma_{11} = \sigma_{12} \mid, \quad \mid \sigma_{21} = \sigma_{22} \mid$ testing for variance shifts.

The null hypothesis for test is $\mid \sigma_{ij} - \sigma_{kl} \mid = 0, \quad i, \quad j, \quad k, \quad l$, equal to 1, 2 for no coefficient shifts or variance shifts. The Bootstrapping technique will be used to produce the critical value of $\mid \sigma_{ij} - \sigma_{kl} \mid = 0$.

In our case, a VAR(16) model was estimated for USO & WTI, DBO & WTI in sub-period (I) and sub-period (II). Different length of lags has been tried, and the lag length of 16 was selected because the error term of the VAR model a_{ti} turns to be clean with such lags. The results are summarized in Table 4.1.

Table _ 4.1 Factual & Counterfactual data for Stock-Watson Test

Data Set	**Factual**	**Counter factural**	**Factual**	**Counter factural**	**Differences by shock**		**Differences by structure**		**Differences by factual**
	σ_{11}	σ_{12}	σ_{21}	σ_{22}	$\mid \sigma_{11} - \sigma_{12} \mid$	$\mid \sigma_{21} - \sigma_{22} \mid$	$\mid \sigma_{11} - \sigma_{21} \mid$	$\mid \sigma_{12} - \sigma_{22} \mid$	$\mid \sigma_{11} - \sigma_{22} \mid$
USO (w/ 1501 obs)	95.11	21.63	46.19	9.43	73.48***	36.75***	48.91**	12.19	85.67***
WTI (w/ 1501 obs)	148.9	48.83	97.61	42.37	100***	55.24***	51.25	6.47	106.5**
DBO (w/ 1319 obs)	19.14	7.92	10.70	4.17	11.23***	6.53***	8.44	3.75	14.98***
WTI (w/ 1319 obs)	166.70	75.74	108.8	51.1	90.95***	57.69***	57.89	24.64	115.6**

$^{*}p < 0.1,\ ^{**}p < 0.05,\ ^{***}p < 0.01$

Let's consider the results for DBO as an example to discuss the Stock-Watson test results. In this case, $y_t = \begin{pmatrix} DBO_t \\ WTI_t \end{pmatrix}$, the factual variance for the first period is 19.14, for second period 4.17, and the difference is 14.98, which is significant. The counterfactual variance produced by combination of first period dynamics and second period shocks σ_{12} is 7.92, and by second period dynamics and first period shocks σ_{21} 10.70. The shock change for the first period $\mid \sigma_{11} - \sigma_{12} \mid$ is 11.23, for the second period $\mid \sigma_{21} - \sigma_{22} \mid$ is 6.53. The changes in two periods are both significant at 1% level. Then look at the structural change in two periods: $\mid \sigma_{11} - \sigma_{21} \mid = 8.44$ for the first period, $\mid \sigma_{12} - \sigma_{22} \mid = 3.75$ for second period, both of which are not significant. Therefore, we conclude that the volatility change in series DBO by the break point is mainly attributed to the shocks or impulses.

In a similar way, we check the shock changes and structure changes during two periods for USO and WTI, and the result indicates that it was an innovation change rather than the coefficient change for USO, WTI as well to cause the volatility change for pre-crisis period and post-crisis period by the breakpoint.

Figures 4.6 and 4.7 show the plots of factual and counterfactual data for Stock-Watson Test. We compare the graphs in pairs vertically for structural change and horizontally for shock change.

The following graphs confirm our prior statements that it was mainly change in the variance, as shapes changed more if compared horizontally than vertically.

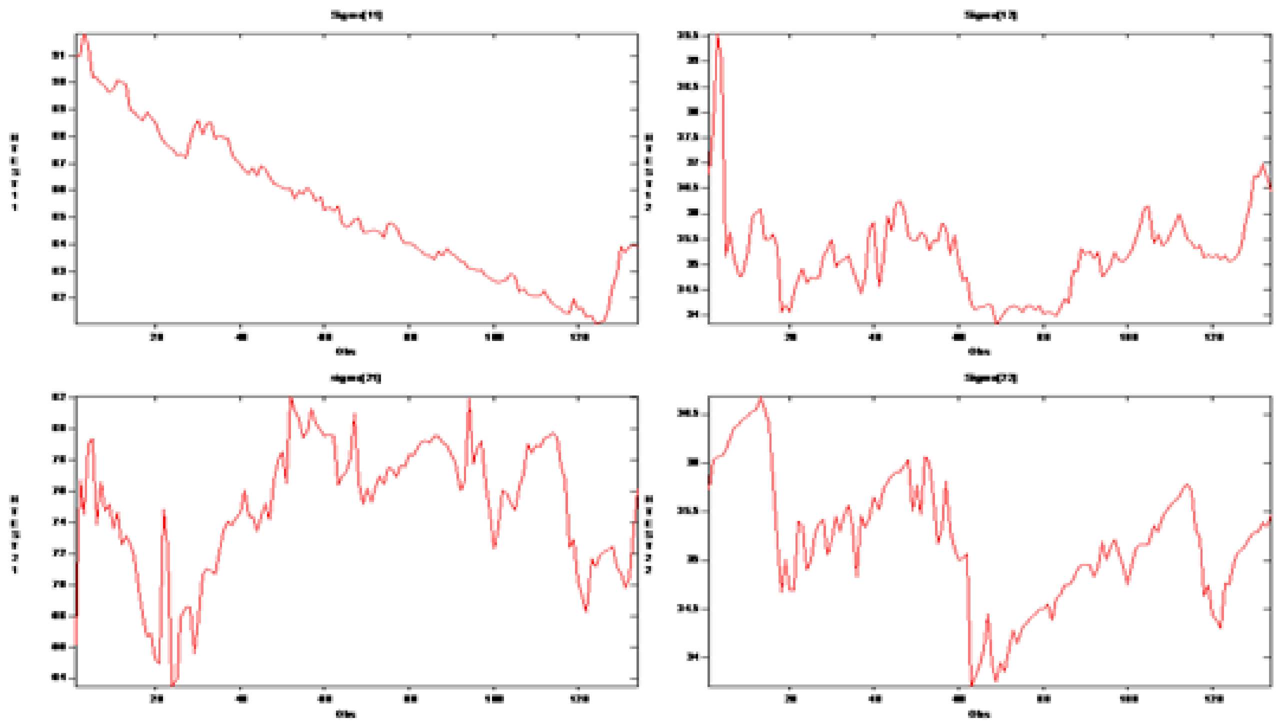


Figure 4.6 Stock-Watson values for USO & WTI (raw data)

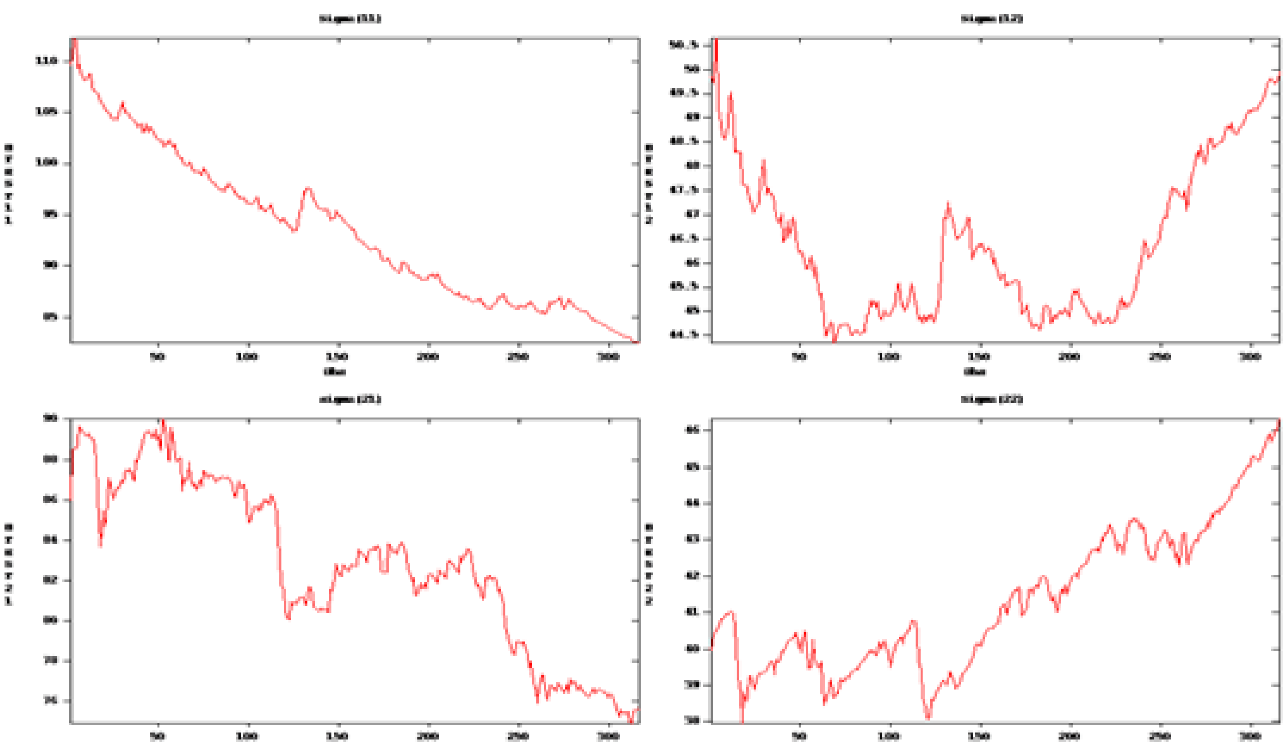


Figure 4.7 Stock-Watson values for DBO & WTI (raw data)

4.3 Conclusion

The Recusive Residuals (RR) test has been used to test the structural stability for data series USO, DBO and WTI. After finding evidence of structural break, we set the breakpoint on the date of December 31, 2008. The Stock-Watson test indicates that the volatility change in series for pre and post-crisis period by the breakpoint is mainly attributed to innovation change (or shocks).

In the following sections, we will have our discussion based on the full sample and sub-samples as follows:

Sample (I): Full Sample
for USO: 4/12/2006 3/31/2012
for DBO: 1/5/2007 3/31/2012

Sample (II): Sub-period (I) or pre-crisis period
for USO: 4/12/2006 - 12/31/2008
for DBO: 1/5/2007 12/31/2008

Sample (III): Sub-period (II) or post-crisis period
for both USO & DBO: 1/2/2009 3/31/2012

Chapter 5. Granger Causality Analysis

In Chapter 4, we found out that there is a breakpoint in data series of USO, DBO and WTI due to the financial crisis in 2008. From this chapter, we will investigate the causal and equilibrium long-term relationship between two pairs of variables USO & WTI and DBO & WTI (Chapter 5 & 6), and model the relationship by using Error Correction Model (ECM) (Chapter 7).

The Granger-Causality test was proposed by Granger (1969) to assess the causality and feedbacks between two related series. The basic idea of Granger Causality is that for series x_t and y_t, if y_t could be better predicted using the information of y_{t-i} and x_{t-i}, $(i = 1, 2, \ldots)$ than just using y_{t-i} alone, then we say variable X Granger-causes variable Y.

Granger (1969) suggested the Causality testing based on a bivariate VAR representation:

$$y_t = \alpha_0 + \sum_{i=1}^{n} \alpha_i y_{t-i} + \sum_{j=1}^{n} \beta_j x_{t-j} + e_{ty} \tag{5.1}$$

$$x_t = \beta_0 + \sum_{i=1}^{n} \alpha_i x_{t-i} + \sum_{j=1}^{n} \beta_j y_{t-j} + e_{tx} \tag{5.2}$$

Here y_t represents the price of an ETF, x_t is the price in crude oil price. x_{t-j} and y_{t-j} contains information which is statistically significant to predict the value of y_t and x_t, respectively. If $\beta_j \neq 0$, that means x_t and y_t will be helpful in estimating y_t in (5.1) and x_t in (5.2), respectively. In other words, if $\beta_j \neq 0$, the variance of e_t produced by (5.1) and (5.2) will be significantly lower than the $var(e_t)$ produced when restricting $\beta_j = 0$.

Testing for X Granger-causes Y is based on equation (5.1) and Y Granger-causes X on (5.2).

The null hypothesis for Granger Causality F- test is

$H_0 : \beta_1 = \beta_2 = \ldots = \beta_n = 0$ (for non-causality)

To test for Granger causality, we need to specify the lags of the VAR model first.

5.1 Order Selection for VAR Model

VAR model for USO & WTI and DBO & WTI:

$$\begin{pmatrix} G_{11}(B) & G_{12}(B) \\ G_{21}(B) & G_{22}(B) \end{pmatrix} \begin{pmatrix} USO_t \\ WTI_t \end{pmatrix} = \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix} \tag{5.3}$$

$$\begin{pmatrix} D_{11}(B) & D_{12}(B) \\ D_{21}(B) & D_{22}(B) \end{pmatrix} \begin{pmatrix} DBO_t \\ WTI_t \end{pmatrix} = \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \tag{5.4}$$

B is the backshift operator. Lag selection for the VAR (5.3) and (5.4) is tested by AIC, BIC and HQ, and the testing results are presented in Appendix A. Table_5.1 summarizes the lag selection results by using different criteria.

Since the orders selected by different criteria are not even close, we would prefer long lags to short lags. The reason is that the data used in this study is on a daily basis, and there is a unit root existed in the raw data series. Therefore, we will select the 16 lags for VAR model (5.3) and (5.4) over three sample periods according to the lag selection result of AIC.

Also, inspection of the residual cross correlation function(CCF) matrix for VAR models of USO & WTI and DBO & WTI over three samples with 2 lags, 3 lags, 6 lags and 16 lags shows VAR order of 16th is appropriate. When use short lags, at the 12th to 16th order VAR, there are still spikes in the cross correlations for the residuals. If use order of 16, the significant autocorrelation and cross correlation have been removed in residual.

Table _ 5.1 VAR Lag Selection

	(I)			**(II)**			**(III)**		
	4/12/2006 - 3/31/2012			**4/12/2006 - 12/31/2008**			**1/2/2009 - 3/31/2012**		
USO & WTI	AIC	BIC	HQ	AIC	BIC	HQ	AIC	BIC	HQ
Lags	16	2	5	16	3	6	2	2	2
	1/5/2007 - 3/31/2012			**1/5/2007 - 12/31/2008**			**1/2/2009 - 3/31/2012**		
DBO & WTI	AIC	BIC	HQ	AIC	BIC	HQ	AIC	BIC	HQ
Lags	16	2	6	6	3	3	2	2	2

AIC: Akaikes information criteria

BIC: Schwarzs information criteria

HQ: Hannan-Quinn information criteria

5.2 Results of Granger Causality Test

Table 5.2 reports the F-statistics and the significance of the Granger Causality test for ETFs USO & DBO and the underlying crude oil price WTI. The results suggest that the price of ETFs and the price of WTI are inter-related. In sample (1), all the F-test statistics are significant at 1% level for ETFs USO_t and DBO_t Granger cause WTI_t and WTI_t causes USO_t and DBO_t. In sample (2), the situations are similar, only the effect of DBO to Granger cause in the WTI_t is significant at 5% level. But in sample (3), there is an interesting result for both USO_t and DBO_t. The significance of the F test for WTI_t Granger-Causality of USO_t is at 10% level and DBO_t 5%. Seems the causality relationship for WTI_t mapping to oil ETFs is not as strong as in full sample and pre-crisis sample. As we know, the price of USO and DBO are tracking the performance of WTI, so the price of WTI should have impact on the price of ETFs. The sample period in sample (3) is just during the time post the 2008 financial crisis, when the market was highly volatile and unpredictable, which weakens the impact of crude oil on the price of ETFs.

The results indicate that the feedbacks between ETFs and WTI are not only from WTI to ETFs, but also from ETFs to WTI, the mappings are bi-directional. This result is in keeping with our original thought because oil ETFs are tracking the performance of WTI, meanwhile the oil futures consisted in ETFs could be a predictor to WTI. In Chapter 6 & 7, we will investigate more in equilibrium relationship and figure out the lead and lag position in oil ETFs and WTI.

Before doing that, we will discuss an argument regarding whether the non-stationary series could be involved in testing the Granger-causality.

Table_5.2 Granger Causality F test for ETFs and WTI

	(I)		**(II)**		**(III)**	
USO & WTI	**4/12/2006 - 3/31/2012**		**4/12/2006 - 12/31/2008**		**1/2/2009 - 3/31/2012**	
Causality Relation	F(16, 1452)	P-value	F(16, 635)	P-value	F(2, 810)	P-value
$WTI_t \rightarrow USO_t$	3.25	0.0000	3.08	0.0000	2.94	*0.0532*
$USO_t \rightarrow WTI_t$	4.19	0.0000	3.52	0.0000	13.97	0.0000
DBO & WTI	**1/5/2007 - 3/31/2012**		**1/5/2007 - 12/31/2008**		**1/2/2009 - 3/31/2012**	
Causality Relation	F(16, 1270)	P-value	F(6, 483)	P-value	F(2, 810)	P-value
$WTI_t \rightarrow DBO_t$	2.59	0.0006	3.92	0.0007	4.19	0.0155
$DBO_t \rightarrow WTI_t$	3.48	0.0000	2.62	0.0164	7.23	0.0008

$*p < 0.1,\ **p < 0.05,\ ***p < 0.01$

5.3 Controversy regarding Stationary of Series

There is a controversy regarding if it is necessary to use transformed data when the unit root existed, as non-stationarity may cause bias in result.

Granger (1969) mentioned that the assumption in the Granger-causality test is that the series involved are stationary. The reason of the requirement of stationarity is that in the case of non-stationary, the variance of e_t in (5.1) and (5.2) is not defined, therefore the existence of causality might change over time as well.

Sims, Stock and Watson (1990) showed by using the example of a vector auto-regression model that it is not necessary to transform the series to be stationary, since residuals can be made white, the Granger tests will still "be asymptotically valid, and which will have nonstandard limiting distributions". Sims worked on transformed and untransformed models to compare the results, and concluded "the common practice of attempting to transform models to stationary form by difference or cointegration operators whenever it appears likely that the data are integrated is in many cases unnecessary."

Stokes and McDonald (2013) investigated a VAR based Granger causal relationship among monetary policy, mortgage rates and the housing bubble, and tested the effect of the various transformations on the series. Their findings suggested that low frequency information is removed by the differencing transformation.

We will use the spectral analysis and impulse response function with the data untransformed and transformed (by difference) in this study to compare the results.

5.3.1 Spectral Analysis

The spectrum is the distribution of variance of the series in a frequency domain. In the frequency domain, we study the variance of the series as a function of frequency. The spectrum graphs the variance contributed at frequency f_j, and the figure of variance plotted against f_j indicates which frequencies are the most important to the variability of the time series.

For spectra with level data for USO, DBO and WTI in three samples, we noticed that the variance tends to be higher at the low frequencies than at the high frequencies. That indicates the series contain low frequency information, which is non-stationary or long-range dependence. Then we checked the spectrum with differenced data, and the low frequency information disappeared. This finding suggests that using filtered data will also remove or attenuate the low frequency information in series.

Figure_5.1 to Figure_5.3 graphs the spectra of series in level vs. in difference.

Figure_5.1 Spectra of series (level data vs. difference data) in Sample (I)

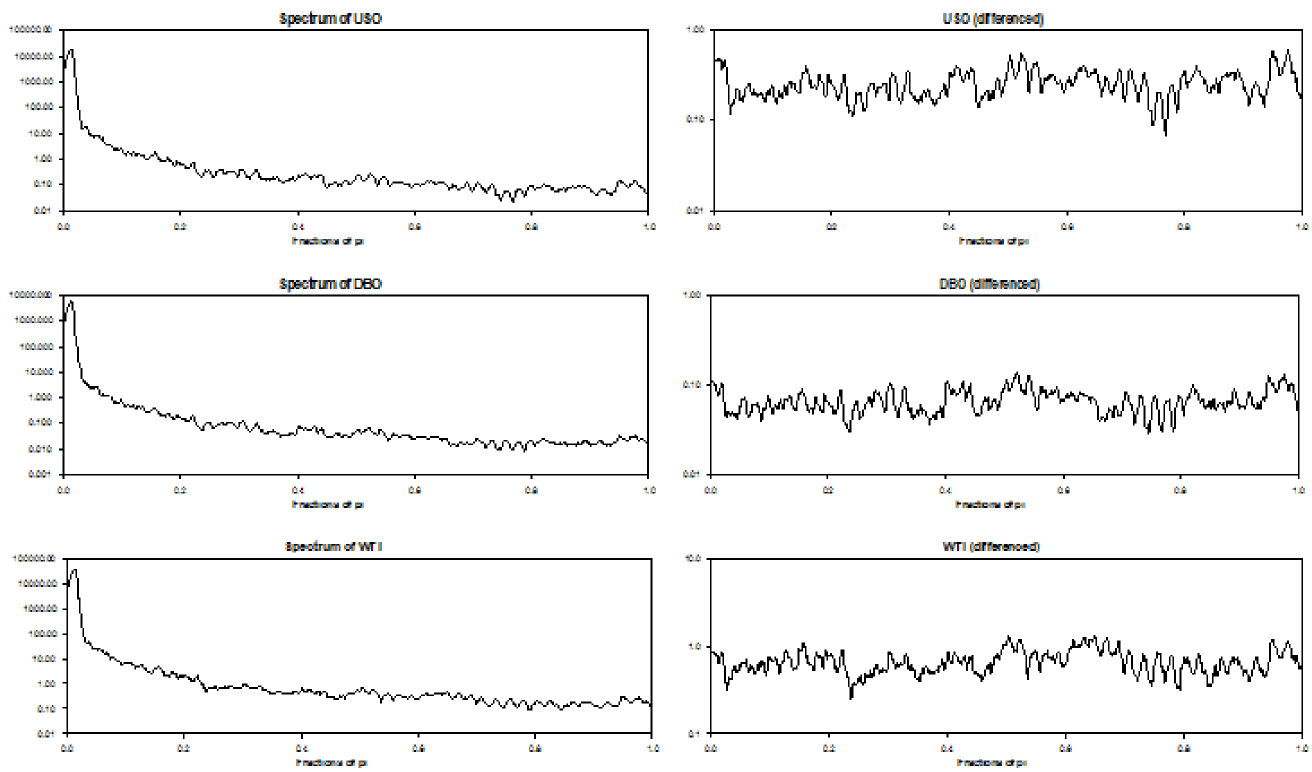


Figure_5.2 Spectrums of series (level data vs. difference data) in Sample (II)

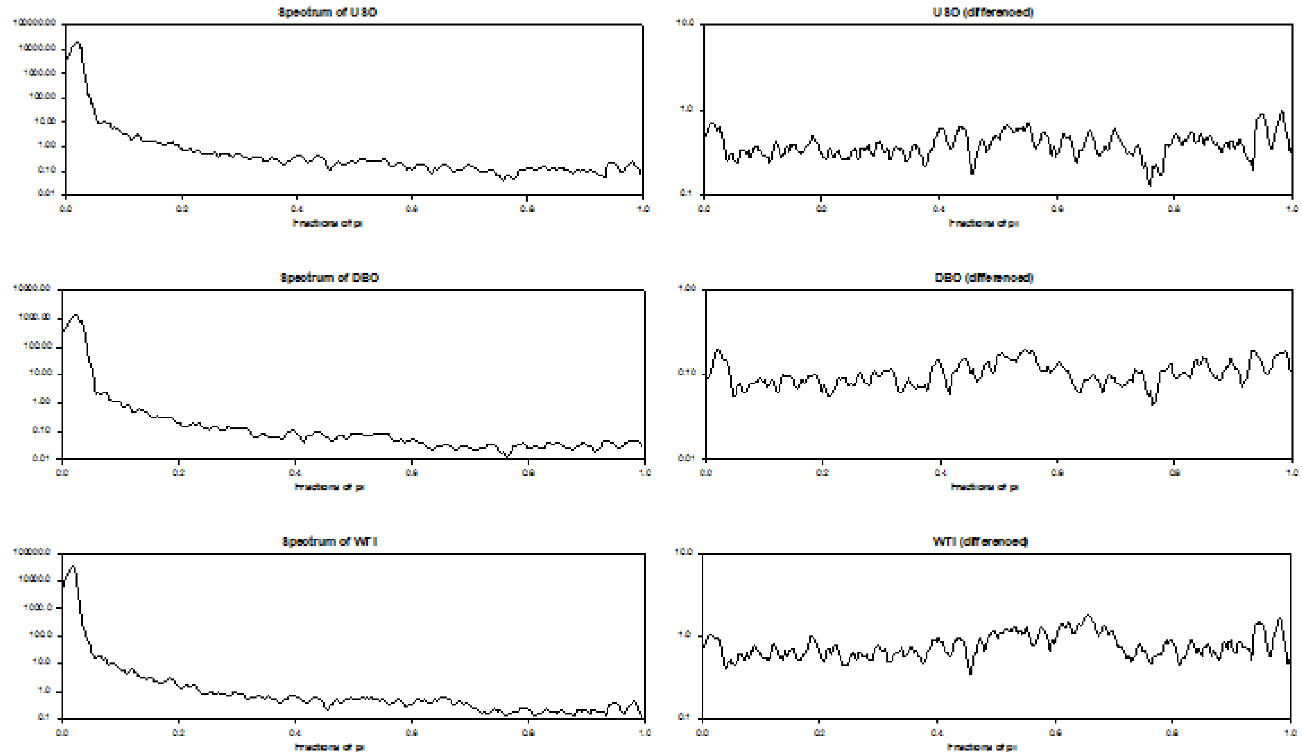
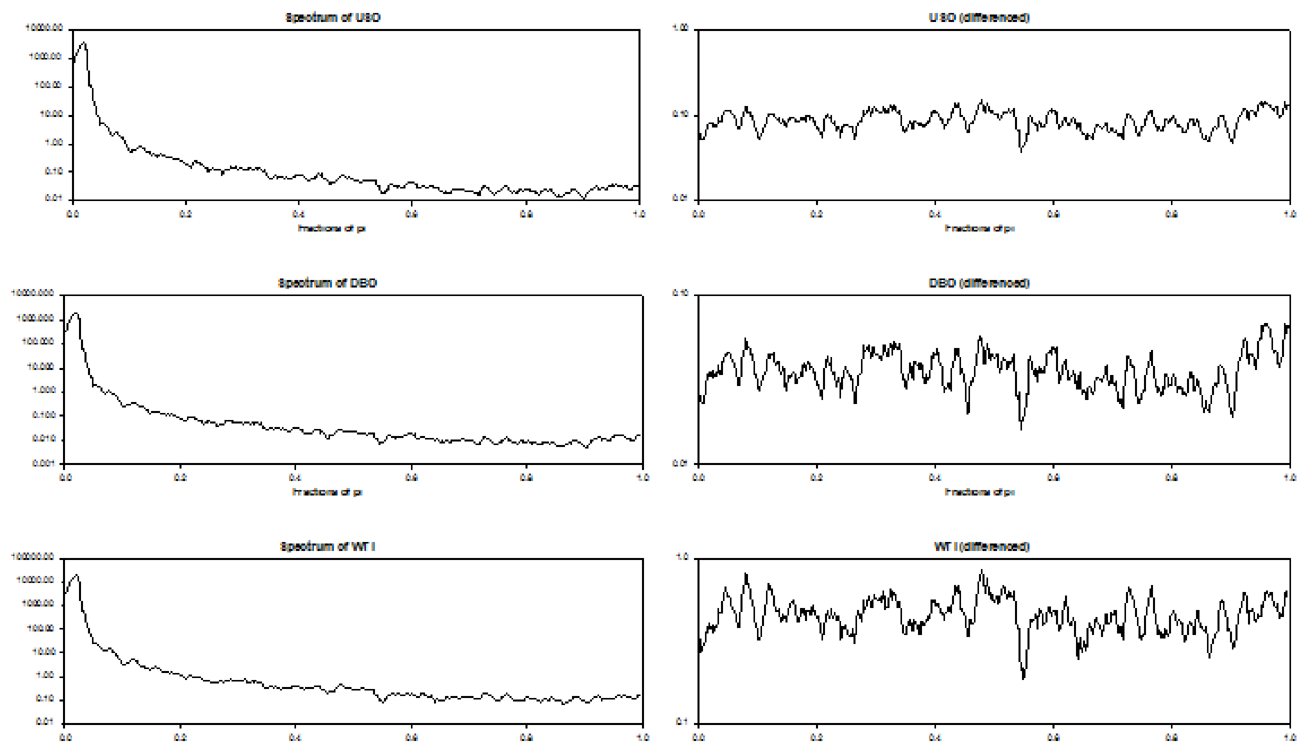


Figure_5.2 Spectrums of series (level data vs. difference data) in Sample (III)



5.3.2 Impulse Response Results

The VAR form of (5.3) and (5.4) can be transformed to a vector moving average form (VMA) to estimate the impulse response in the shocks, and VMA allows the measurement of shock going in two ways.

Provided $G(B) = \begin{pmatrix} G_{11}(B) & G_{12}(B) \\ G_{21}(B) & G_{22}(B) \end{pmatrix}$ and $D(B) = \begin{pmatrix} D_{11}(B) & D_{12}(B) \\ D_{21}(B) & D_{22}(B) \end{pmatrix}$ in (5.3) and (5.4), respectively, are convertible, (5.3) (5.4) can be transformed in form of VMA model:

$$\begin{pmatrix} USO_t \\ WTI_t \end{pmatrix} = \Theta(B) \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix} \tag{5.5}$$

$$\begin{pmatrix} DBO_t \\ WTI_t \end{pmatrix} = \Omega(B) \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \tag{5.6}$$

where $\Theta(B) \equiv [G(B)]^{-1}$, $\Omega(B) \equiv [D(B)]^{-1}$.

$\Theta(B)$ and $\Omega(B)$ measures the dynamic responses of the price of USO and DBO or WTI to a shock in the model. (5.5) and (5.6) can be expanded to:

$$\begin{pmatrix} USO_t \\ WTI_t \end{pmatrix} = \begin{pmatrix} \theta_{11}(B) & \theta_{12}(B) \\ \theta_{2}1(B) & \theta_{22}(B) \end{pmatrix} \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix} \tag{5.7}$$

$$\begin{pmatrix} DBO_t \\ WTI_t \end{pmatrix} = \begin{pmatrix} \omega_{11}(B) & \omega_{12}(B) \\ \omega_{2}1(B) & \omega_{22}(B) \end{pmatrix} \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \tag{5.8}$$

In detail, $\theta_{12}(B)$ and $\omega_{12}(B)$ measures the effect of shocks in WTI on price of USO and DBO, respectively. $\theta_{21}(B)$ and $\omega_{21}(B)$ measures the effect of shocks in USO and DBO on price of WTI.

The VMA coefficients for (5.7) and (5.8) are provided in Appendix B(1). The lags with highlight indicate the maximum magnitude of shock effect.

The Monte Carlo Integration in RATS was applied to estimate a 95% confidence interval of the point values that are generated by impulse. Figture_5.4 is based on the model (5.7) and (5.8) for lag of 16 to show the effects of shocks for level data of USO & WTI and DBO & WTI over three sample period. The impulses on diagonal are the effects of shocks of oil ETFs and WTI on itself. Those on off-diagonal are shocks of USO or DBO on WTI or WTI on USO or DBO. For all the samples, the effects of shocks of oil ETFs on WTI ($\theta_{21}(B)$ and $\omega_{21}(B)$) are positive and significant. Now look at the change of price of WTI on USO and DBO ($\theta_{12}(B)$ and $\omega_{12}(B)$), it varies around positive and negative. Anyway, all of the effects of shocks are significant.

In term of the transformed data, the VMA form of the model can be written as:

$$\begin{pmatrix} DUSO_t \\ DWTI_t \end{pmatrix} = \begin{pmatrix} \tilde{\theta}_{11}(B) & \tilde{\theta}_{12}(B) \\ \tilde{\theta}_2 1(B) & \tilde{\theta}_{22}(B) \end{pmatrix} \begin{pmatrix} \tilde{e}_{1t} \\ \tilde{e}_{2t} \end{pmatrix} \tag{5.9}$$

$$\begin{pmatrix} DDBO_t \\ DWTI_t \end{pmatrix} = \begin{pmatrix} \tilde{\omega}_{11}(B) & \tilde{\omega}_{12}(B) \\ \tilde{\omega}_2 1(B) & \tilde{\omega}_{22}(B) \end{pmatrix} \begin{pmatrix} \tilde{u}_{1t} \\ \tilde{u}_{2t} \end{pmatrix} \tag{5.10}$$

The VMA coefficients for (5.9) and (5.10) are provided in Appendix B(2). Figture_5.5 is based on the first difference data of USO & WTI and DBO & WTI for three samples. We noticed that all the effects of shocks converged to zero. Compared to the results of level data, Figture_5.5 indicates that vital information has been discarded by differencing. This finding is consistent with the results from section 5.3.1. In section 5.3.1 of spectral analysis, we concluded that the low frequency information has been removed when performing the filter of differencing.

5.4 Conclusion

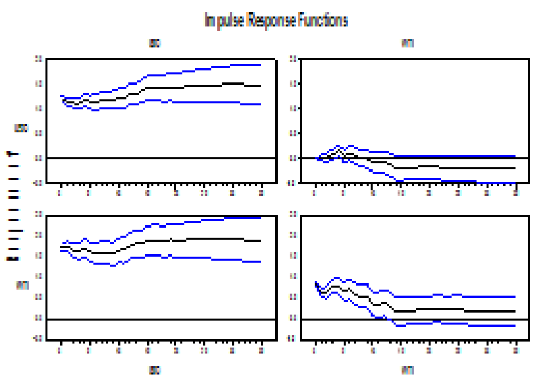
Based on result of spectral analysis and impulse response function by using transformed and untransformed data, we found that the vital information (mainly low frequency) is discarded if we filtered data in first difference. Therefore, we still use the level data to perform Granger-causality test in this study.

Figure_5.4 Graph of Monte Carlo integration with level data (lags=16)

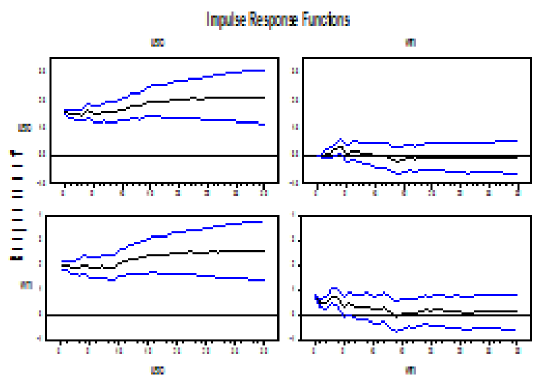
(# of response steps = 36, # of keeper draws = 800)

Impulse Response Function for USO & WTI

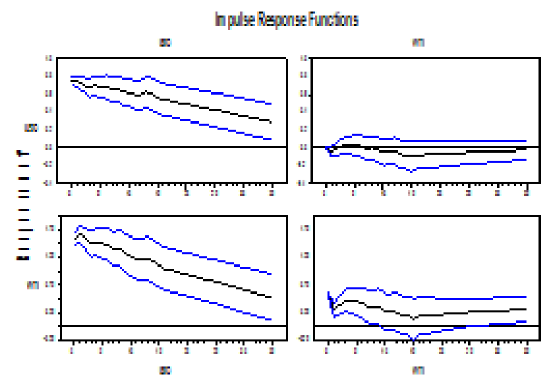
Sample (I)



Sample (II)

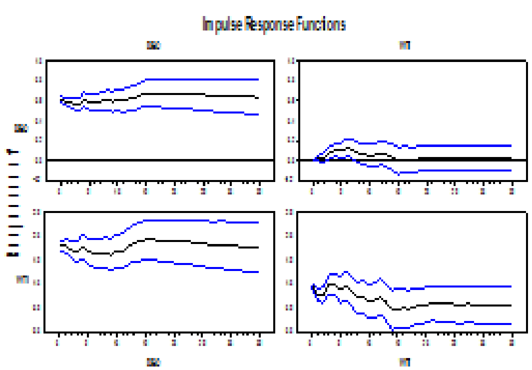


Sample (III)

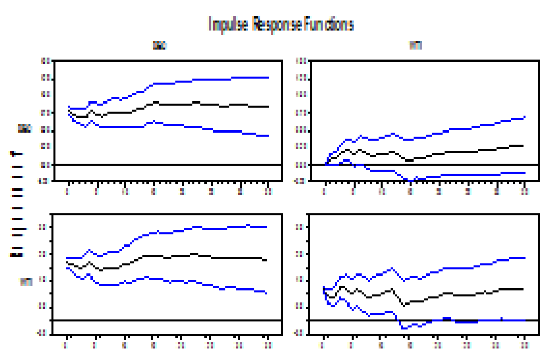


Impulse Response Function for DBO & WTI

Sample (I)



Sample (II)



Sample (III)

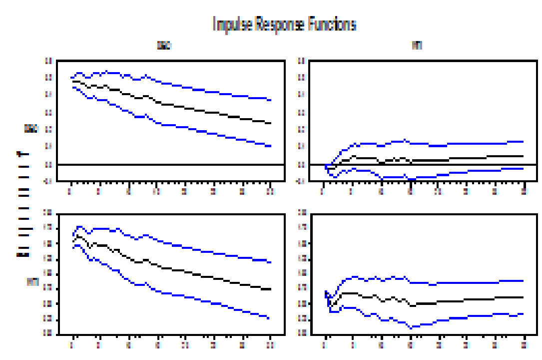
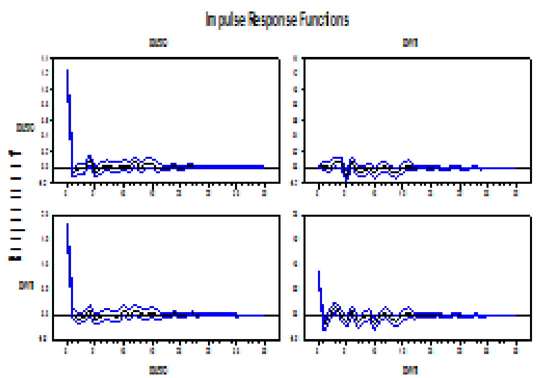


Figure_5.5 Monte Carlo integration with differencing data (lags=16)

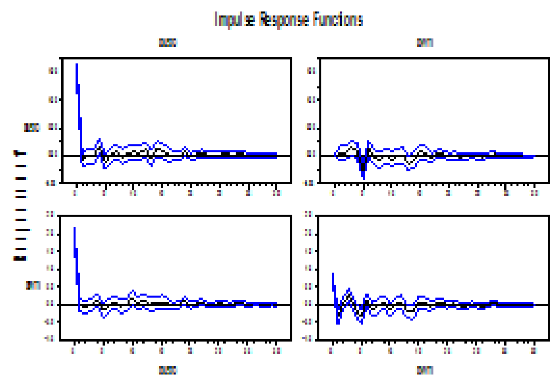
(# of response steps = 36, # of keeper draws = 800)

Impulse Response Function for USO & WTI

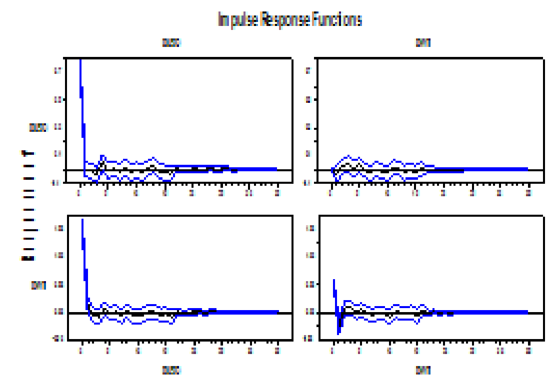
Sample (I)



Sample (II)

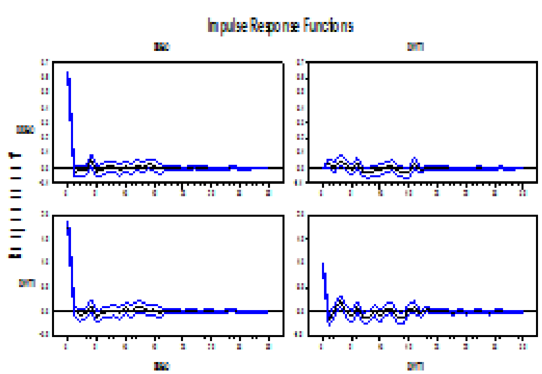


Sample (III)

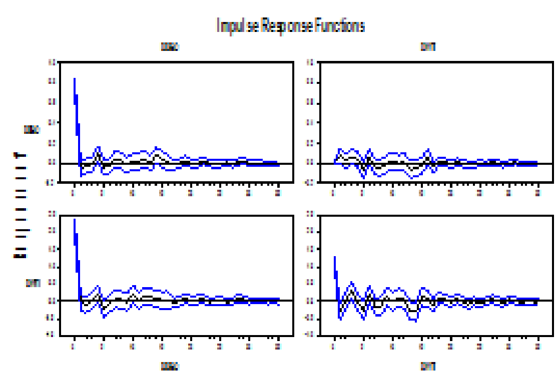


Impulse Response Function for DBO & WTI

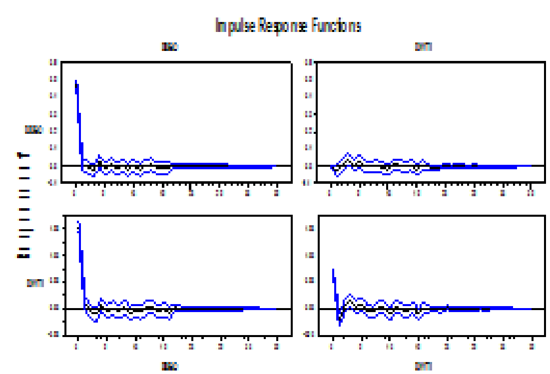
Sample (I)



Sample (II)



Sample (III)



Chapter 6. Cointegration

6.1 Cointegration Theory

The concept of cointegration was introduced by Granger (1981). Originally, it is to solve the problem of so-called "spurious regression". In macroeconomics, it has been a common practice to estimate a model involving non-stationary variables by linear regression process for a long time. The problem is that even if the result suggests there be a statistally significant relationship between variables, there could be none existed due to the non-stationary of series. So the relationship indicated by the result might be well spurious. Granger's solution to this problem is to develop the concept of cointegration.

Cointegration therefore describes whether or not two (or more) non-stationary series follow the same long-run trends by an equilibrium relationship. Ender (2004) stated "Equilibrium theories involving non-stationary variables require the existence of a process of combination of the variables that is stationary". Cointegration often means that a linear combination of individually unit-root non-stationary time series becomes a stationary and invertible series. One of the most commonly employed procedures to test the existence of a co-integration relationship is Engle and Granger two-step methodology.

Granger (1981) defined the concept of integration:

A time series x_t is said to be integrated of order d, $x_t \sim I(d)$, if $(1-B)^d x_t$ is stationary and invertible, where $d > 0$.

In a multivariate case, a vector x_t is said to be cointegrated of order d, b, $x_t \sim$ CI(d,b), if (i) all components of x_t are I(d); (ii) there exists a vector $\alpha(\neq 0)$ so that $z_t = \alpha' x_t$ is integrated of order I(d b), $b > 0$. The vector is called the cointegrating vector.

If interpreting $\alpha' x_t = 0$ as a long run equilibrium, cointegration implies that deviations from equilibrium are stationary.

Since $z_t = \alpha' x_t$ is stationary, the l-step ahead forecast of z_{T+l} at the forecast origin T satisfies

$z_{T(l)} \longrightarrow E(z_t) = u_z,\, l \to \infty$

This also implies that $\alpha' z_{T(l)} \longrightarrow u_z$ as l increases. Then the point forecasts of x_t satisfy a long-term stable constraint.

In the case of Oil ETF and Crude Oil, we noticed that USO, DBO and WTI move dependently with each other, so next, we will determine whether there exists an equilibrium relationship between USO & WTI and DBO & WTI.

6.2 Linear Cointegration

6.2.1 Engle and Granger Two-Step Methodology

Enders (2004) illustrated two methodologies to test for cointegration. One is the Engle-Granger testing procedure, which was initially proposed by Engle and Granger (1987).This methodology seeks to determine whether the residuals of the equilibrium relationship are stationary. The other is the Johansen (1988) and Stock-Watson (1988) methodologies, which determine the rank of cointegration by using the maximum likelihood estimator.

Engle and Granger (1987) suggested the following two-step estimator.

The first step is to determine the order of integration for each variable and generate the error series $\{\hat{e}_t\}$. The Dickey-Fuller (DF) or augmented Dickey-Fuller (ADF) test can be used to detect the number of unit roots in each variable. It is important because if the variables are integrated of different orders, its possible to conclude that they are not cointegrated. If the results indicate that two series $\{y_t\}$ and $\{z_t\}$ are I(1), the long-run equilibrium relationship can be estimated using OLS:

$$y_t = \beta_0 + \beta_1 z_t + e_t \tag{6.1}$$

Then define

$$\hat{e}_t = y_t - \hat{\beta}_0 - \hat{\beta}_1 z_t \tag{6.2}$$

$\{\hat{e}_t\}$ is the series of the estimated residuals of the long-run relationship. If the deviations from long-run equilibrium are stationary, then $\{y_t\}$ and $\{z_t\}$ are co-integrated of order

(1,1). We could perform the Dickey-Fuller test and Augmented Dickey-Fuller test on the residuals by using the following two equations to test if $\alpha_1 = 0$

$$\Delta \hat{e}_t = \alpha_1 \hat{e}_{t-1} + \varepsilon_t \tag{6.3}$$

$$\Delta \hat{e}_t = \alpha_1 \hat{e}_{t-1} + \sum \alpha_{i+1} \Delta \hat{e}_{t-i} + \varepsilon_t \tag{6.4}$$

According to Ender (2004): "The only difference from the traditional ADF to (this version of) the Engle-Granger test are the critical values. The critical values to be used here are no longer the same provided by Dickey-Fuller, but instead provided by Engle and Yoo (1987). This happens because the residuals above are not the actual error terms, but estimated values from the long run equilibrium equation."

The second step is to use the residuals $\{\hat{e}_t\}$ to estimate the error-correction model, then estimate the long-run equilibrium relationship. If $\{y_t\}$ and $\{z_t\} \sim$ CI(1,1), the variables have the error-correction form:

$$\Delta y_t = \alpha_{01} + \alpha_y \hat{e}_{t-1} + \sum_{i=1} \alpha_{11}(i) \Delta y_{t-i} + \sum_{i=1} \alpha_{12}(i) \Delta z_{t-i} + \varepsilon_{yt} \tag{6.5}$$

$$\Delta z_t = \alpha_{02} + \alpha_z \hat{e}_{t-1} + \sum_{i=1} \alpha_{21}(i) \Delta y_{t-i} + \sum_{i=1} \alpha_{22}(i) \Delta z_{t-i} + \varepsilon_{zt} \tag{6.6}$$

The residual ε_{yt}, ε_{zt} will be checked whether they are serially correlated. The model should be re-estimated by using longer lag lengths if the residuals are serially correlated until they yield serially uncorrelated errors. Then we may test the speed of adjustment parameters α_y and α_z, and if Δy_t and Δz_t converge to the long-run equilibrium relation.

Enders (2004) mentioned that although the Engle-Granger procedure is convenient, there are two important defects. First, the procedure requires placing one variable on the left-hand side and using the others as regressors on the right-hand side. If three or more variables are used since any of the variables can be selected as the left-hand side variable, the result of the test will be different. Second, the coefficient is obtained by estimating a regression using the residuals from another regression, so any error introduced in step 1 is carried into step 2.

6.2.2 Cointegration Testing Result

The variables USO, DBO and WTI for the whole period and sub-period were all tested using the Augmented Dickey-Fuller test (ADF). The results are reported in Table 6.1. The 95% critical value for for ADF using 4 lags is -2.865. All the absolute values of t-statistics for USO, DBO and WTI are below the critical value, so we cannot reject the null hypothesis that there is a unit root in any of the series. To prove the series is I(1), we still need to check if (1-B)x_t (or Δx_t) is stationary. The results are also reported in Table 6.1. There is no evidence indicating a unit root in the differenced series, as the t-statistics of ADF test are all much greater than 95% critical value of 2.865. Therefore, the whole period and sub-periods series of USO, DBO and WTI are I(1). Then the long-run equilibrium regression can be estimated. The estimates of the long-run relationship for USO & WTI, DBO & WTI with whole data and sub-period data:

$$\text{(Sample I)} \quad USO_t = \underset{(3.67)}{6.3163} + \underset{(26.34)}{0.5441} * WTI_t + e_{1t} \tag{6.7}$$

$$\text{(Sample I)} \quad DBO_t = \underset{(14.38)}{5.0661} + \underset{(69.60)}{0.2872} * WTI_t + e_{2t} \tag{6.8}$$

$$\text{(Sample II)} \quad USO_t = \underset{(9.82)}{5.8391} + \underset{(107.09)}{0.7527} * WTI_t + e_{1t} \tag{6.9}$$

$$\text{(Sample II)} \quad DBO_t = \underset{(56.27)}{22.2154} + \underset{(36.74)}{0.1761} * WTI_t + e_{2t} \tag{6.10}$$

$$\text{(Sample III)} \quad USO_t = \underset{(19.68)}{4.145} + \underset{(142.72)}{0.3340} * WTI_t + e_{1t} \tag{6.11}$$

$$\text{(Sample III)} \quad DBO_t = \underset{(55.71)}{11.4465} + \underset{(73.93)}{0.1844} * WTI_t + e_{2t} \tag{6.12}$$

where e_{1t} and e_{2t} are the residuals from the equilibrium regressions.

The question of the greatest interest is whether or not the residuals $\{\hat{e}_{1t}\}$ and $\{\hat{e}_{2t}\}$ are stationary. If the residuals are I(0), then the variables are said to be co-integrated of order (1,1).

Table_6.1 Unit root test for series USO, DBO and WTI

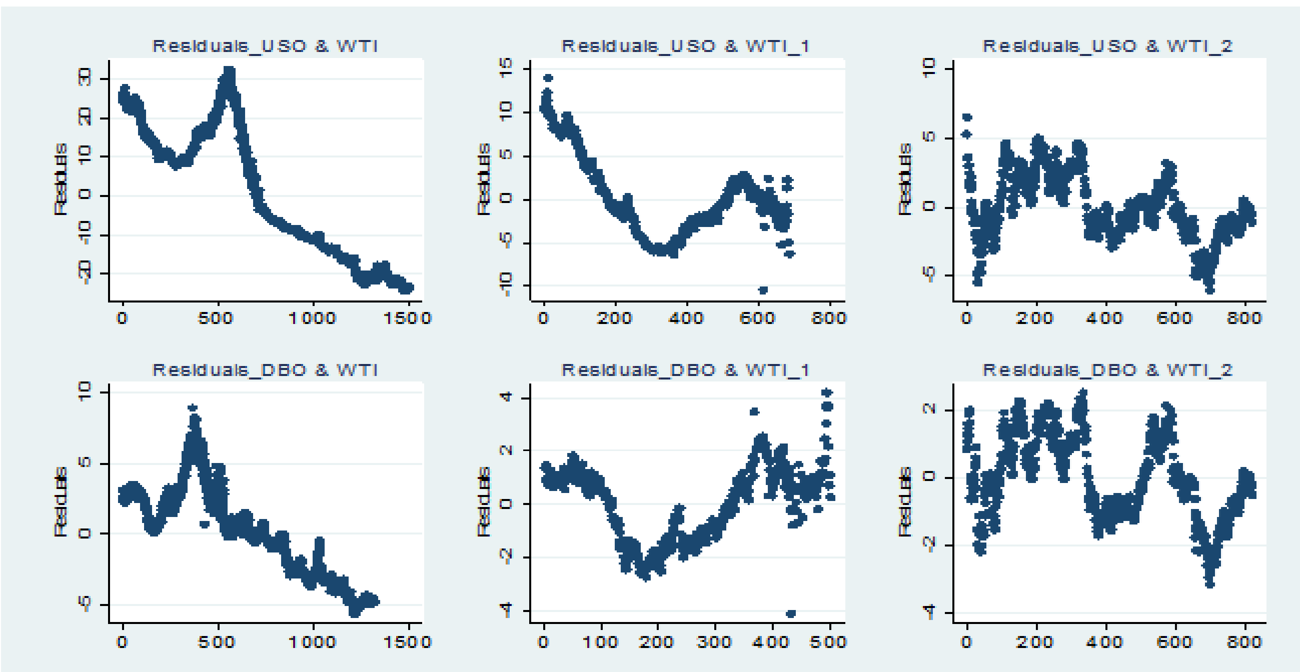
	Sample (I)	**Sample (II)**	**Sample (III)**
ADF with 4 lags	**4/12/2006 - 3/31/2012**	**4/12/2006 - 12/31/2008**	**1/2/2009 - 3/31/2012**
USO_t	-1.395	-0.473	-2.849
ΔUSO_t	-21.610	-15.860	-9.498
WTI_t	-1.760	-0.797	-2.081
ΔWTI_t	-16.797	-17.424	-7.593
	1/5/2007 - 3/31/2012	**1/5/2007 - 12/31/2008**	**1/2/2009 - 3/31/2012**
DBO_t	-1.804	-0.816	-2.243
ΔDBO_t	-26.468	-12.170	-7.907
WTI_t	-1.988	-0.951	-2.081
ΔWTI_t	-27.348	-14.750	-7.593

The Associated t-statistic for $\Delta\hat{e}_{1t}$ and $\Delta\hat{e}_{2t}$ are reported in Table 6.2. The plots in Figure 6.2 are of scatter plot of residuals $\{\hat{e}_{1t}\}$ and $\{\hat{e}_{2t}\}$ for full samples and subsamples.

Table_6.2 The Associated t-statistic for $\Delta\hat{e}$

		Sample (I)	**Sample (II)**	**Sample (III)**
		4/12/2006 - 3/31/2012	**4/12/2006 - 12/31/2008**	**1/2/2009 - 3/31/2012**
$\Delta\hat{e}_{1t}$	No Lags	-0.754	-2.676	-3.548
for USO_t	4 Lags	-0.771	-2.334	-3.066
		1/5/2007 - 3/31/2012	**1/5/2007 - 12/31/2008**	**1/2/2009 - 3/31/2012**
$\Delta\hat{e}_{2t}$	No Lags	-1.668	-3.907	-2.912
for DBO_t	4 Lags	-0.963	-2.149	-2.793

Figure_6.1 Scatter Plot of Residuals



The null hypothesis is that there is a unit root in residuals. Here, the critical value from an ordinary Dickey-Fuller table might not be appropriate to test whether the residual series is stationary. The reason is that the $\{\hat{e}_t\}$ sequence is generated from the regression $y_t = \beta_0 + \beta_1 z_t + e_t$, and we do not know the actual error e_t, only the estimate of the error $\hat{e}_t$. Only if β_0 and β_1 were known in advance and used to construct the true $\{\hat{e}_t\}$, the

Dickey-Fuller test can be used. Engle and Yoo (1987) provided a table of critical values for Engle-Granger cointegration test. To test for cointegration between two variables, the critical value of DF test at the 5 percent significance level is 3.35, at 1 percent level is 4.00, and critical value for ADF at 5% is 3.25, 1% 3.78.

Obviously, based on the Engle and Granger methodology of cointegration testing, variable USO and WTI for the whole period from 4/12/2006 to 3/31/2012 are not cointegrated at any significance level. DF and ADF statistics -0.754 and -0.771 are less than the critical values 3.35 and 3.25 at 5% level, respectively. It is the same for the variable DBO and WTI from 1/5/2007 to 3/31/2012. DF statistics is -1.668, ADF statistics -0.963, which is far below the DF critical value at 5% level 3.35, and ADF 3.25. So we cannot reject the null of I(1) that residuals from the equilibrium regression are non-stationary.

For the sub-periods, the ADF test shows all variables are not co-integrated for sub-period I and sub-period II. If using DF statistics, USO and WTI are cointegrated at the 5% level for period of 1/2/2009 to 3/31/2012. Also, we noticed that the DF statistics for DBO and WTI at period 1/7/2007 to 12/31/2008 is -3.907, which is greater than 5% level of 3.35 and close to 1% level of 4.00, implying that the linear relationship of DBO & WTI for this period is quite strong. The difference in results for full sample and the subsamples is mainly due to the structural break cross the series, and is consistent with previous conclusions in section 3.

In general, by using Engle and Granger two-step methodology, under the null hypothesis, $\{\hat{e}_t\}$ is I(1), so that cointegration relationship is not found for the variable USO & WTI and DBO & WTI in either whole period or sub-periods.

This result may give rise to an argument that the first step in Engle and Granger methodology is just a simple regression model, so the estimates can be substantially biased partly due to the serial correlation in residuals. Even if the results show that the residuals from equation (6.1) have a unit root, it still may incorrectly assume the non-stationarity. In the next section, we will discuss a dynamic model, which removes the serial correlation from the residuals, and test the co-integration based on a linear dynamic model.

6.3 Dynamic Approach to Cointegration
6.3.1 Auto-regression Distributed Lag Model

David F. Hendry (1999) suggested that we could try an autoregressive-distributed lag model (6.13) to reduce the potential bias caused by auto-correlation in residuals if use a simple regression model.

$$y_t = \beta_0 + \beta_1 z_t + \beta_2 y_{t-1} + \beta_3 z_{t-1} + u_t \tag{6.13}$$

The linear dynamic model (6.13) can be rewritten in an equilibrium-correction form:

(i) subtract y_{t-1} in both sides:

$$\Delta y_t = \beta_0 - (1-\beta_2)y_{t-1} + \beta_1 z_t + \beta_3 z_{t-1} + u_t \tag{6.14}$$

(ii) add and subtract $\beta_1 z_{t-1}$ in right side:

$$\begin{aligned} \Delta y_t &= \beta_0 - (1-\beta_2)y_{t-1} + \beta_1 z_t - \beta_1 z_{t-1} + \beta_1 z_{t-1} + \beta_3 z_{t-1} + u_t \\ &= \beta_0 - (1-\beta_2)y_{t-1} + \beta_1 \Delta z_t + (\beta_1 + \beta_3) z_{t-1} + u_t \\ &= \beta_1 \Delta z_t - (1-\beta_2)(y_{t-1} - \frac{\beta_0}{1-\beta_2} - \frac{\beta_1+\beta_3}{1-\beta_2} z_{t-1}) + u_t \end{aligned} \tag{6.15}$$

Let $\alpha_0 = \dfrac{\beta_0}{1-\beta_2}$, $\alpha_1 = \dfrac{\beta_1+\beta_3}{1-\beta_2}$, where $\beta_2 \neq 1$

(6.15) is the error-correction form for dynamic model (6.13). $(y_{t-1} - \alpha_0 - \alpha_1 z_{t-1})$ is the lagged equilibrium error, which captures the deviations from long-run equilibrium. The speed of adjustment toward the steady process is $-(1-\beta_2)$.

Therefore, the test for co-integration of (6.13) is to test whether the error correction term $(y_t - \alpha_0 - \alpha_1 z_t)$ is a stationary process. If define $v_t = (y_t - \alpha_0 - \alpha_1 z_t)$, the error correction term from (6.15) can be represented as:

$$y_t = \alpha_0 + \alpha_1 z_t + v_t \tag{6.16}$$

(6.16) transformed the co-integration test for a dynamic model in the form of a basic regression model in (6.1). The cointegrating vector for (6.16) will be $(1, -\alpha_1)$, where $\alpha_1 = \dfrac{\beta_1+\beta_3}{1-\beta_2}$

6.3.2 Testing Results for Dynamic Model

(I) serial correlation in residuals The purpose of using a dynamic model(including lag terms in a simple regression model) to test the co-integration relationship is to remove the potential auto-correlation in residuals from simple regression model.

We applied the data to equation (6.13) for USO & WTI and DBO & WTI over three time periods, and checked residuals u_t to see if dynamic model will make any improvement to reduce the serial correlation in residuals.

$$USO_t = \beta_{10} + \beta_{11}WTI_t + \beta_{12}USO_{t-1} + \beta_{13}WTI_{t-1} + u_{1t} \tag{6.17}$$

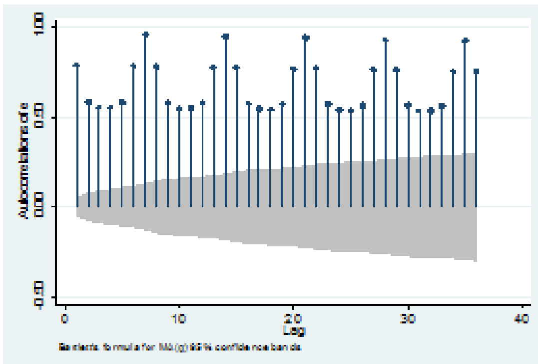
$$DBO_t = \beta_{20} + \beta_{21}WTI_t + \beta_{22}DBO_{t-1} + \beta_{23}WTI_{t-1} + u_{2t} \tag{6.18}$$

The residual autocorrelation function (acf) for regression model (6.1) and dynamic model (6.13) for two pairs of data USO & WTI and DBO & WTI in three samples are plotted in Figure_6.2 to Figure_6.5.

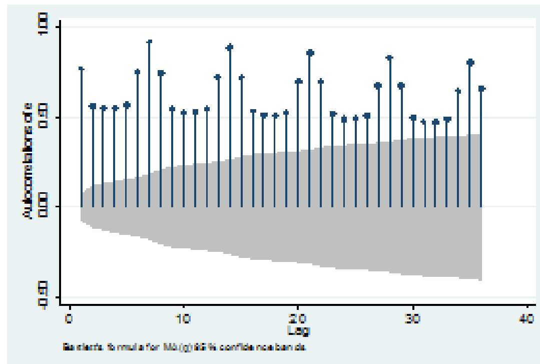
The grey areas in the graphs indicate the 95% confidence interval with mean of zero. The acf of residuals from simple regression model show persistent correlation out of the 95% confidence bound up to lag 36 for two pairs of data from sample (I) to sample (III), the residuals are strongly auto-correlated. On the other hand, most of the plots of residual acf from dynamic model are within the 95% confidence. That clearly shows that dynamic model do help to remove the serial correlation in residuals. Therefore, if using the dynamic model to test the co-integration relationship, the residual autocorrelation is no longer a problem.

Figure_6.2 Residual Correlogram for USO & WTI using simple regression model of (6.1)

Sample (I)



Sample (II)



Sample (III)

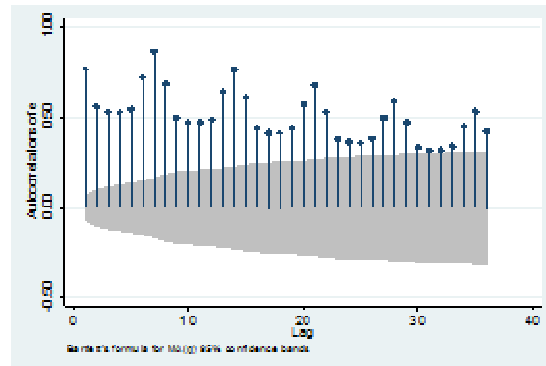
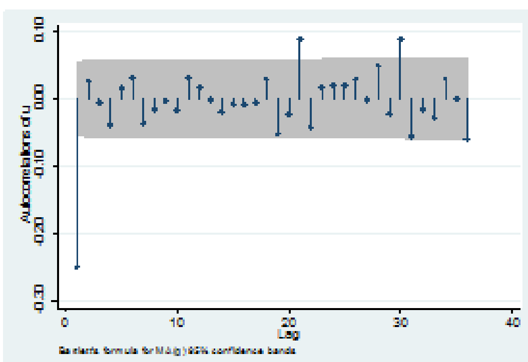
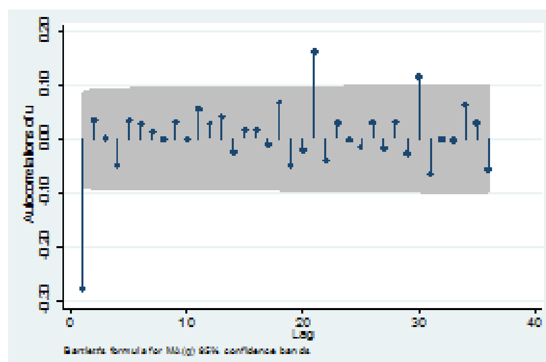


Figure_6.3 Residual Correlogram for USO & WTI using dynamic model of (6.17)

Sample (I)



Sample (II)



Sample (III)

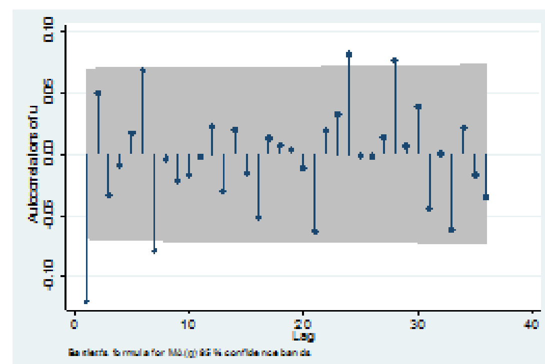


Figure_6.4 Residual Correlogram for DBO & WTI using simple regression model of (6.1)

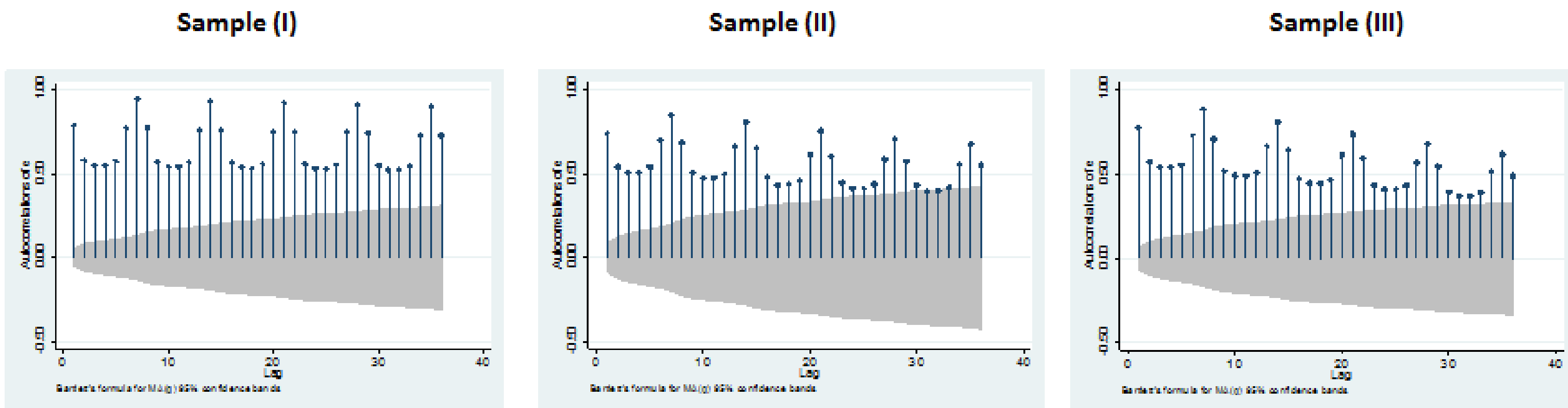
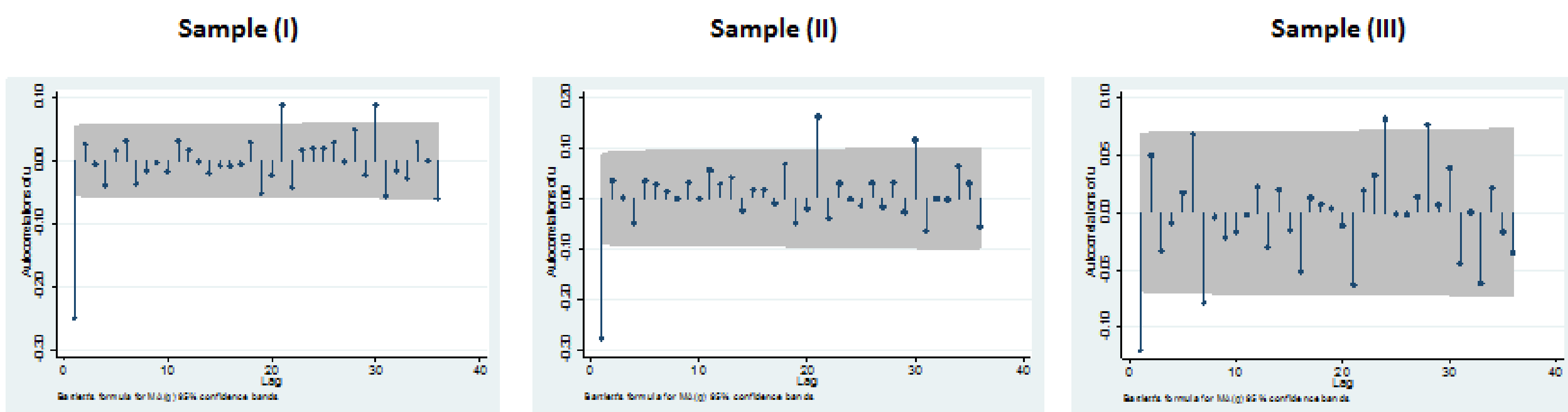


Figure_6.5 Residual Correlogram for DBO & WTI using dynamic model of (6.18)



(II) empirical test for cointegration

The estimates of coefficients for dynamic model (6.17) and (6.18) over three sample periods are reported in Appendix D.

The error-correction form of (6.17) and (6.18) will be:

$$\Delta USO_t = \beta_{11}\Delta WTI_t - (1-\beta_{12})[USO_{t-1} - \frac{\beta_{10}}{1-\beta_{12}} - \frac{\beta_{11}+\beta_{13}}{1-\beta_{12}}\Delta WTI_{t-1}] + u_{1t}$$

$$\Delta DBO_t = \beta_{21}\Delta WTI_t - (1-\beta_{22})[DBO_{t-1} - \frac{\beta_{20}}{1-\beta_{22}} - \frac{\beta_{21}+\beta_{23}}{1-\beta_{22}}\Delta WTI_{t-1}] + u_{2t}$$

It is to test whether the error-correction term in (6.19) and (6.20) are stationary.

$$V_{1t} = USO_{t-1} - \frac{\beta_{10}}{1-\beta_{12}} - \frac{\beta_{11}+\beta_{13}}{1-\beta_{12}}\Delta WTI_{t-1} \tag{6.19}$$

$$V_{2t} = DBO_{t-1} - \frac{\beta_{20}}{1-\beta_{22}} - \frac{\beta_{21}+\beta_{23}}{1-\beta_{22}}\Delta WTI_{t-1} \tag{6.20}$$

If USO & WTI and DBO & WTI are cointegrated, the long-run equilibrium-correction term V_{1t} and V_{2t} must be stationary. We still will use ADF testing for unit root in V_{1t} and V_{2t}.

Table 6.3 presents the Augmented Dickey-Fuller test results for V_{1t} and V_{2t}. We noticed that all the t-stat for ADF test are not significant except for V_{2t} in sample (II). So for V_{1t} in three samples and V_{2t} in sample (I) and sample (III), no rejection of null hypothesis of I(1), there is unit root existed in these samples, and no cointegration relationship in series. The interesting point here is for V2t in sample (II). If we go back to compare the unit root test results by using simple regression model in Table 6.2, the t-stat of $\Delta\hat{e}_{2t}$ in sample (II) for DF test is significant at 5% level. The results are quite consistent by using simple regression model and dynamic model.

Table_6.3 ADF test for error correction term

	Sample (I)	Sample (II)	Sample (III)
	4/12/2006 - 3/31/2012	**4/12/2006 - 12/31/2008**	**1/2/2009 - 3/31/2012**
V_{1t} DF	-1.623	-2.532	-2.487
	1/5/2007 - 3/31/2012	**1/5/2007 - 12/31/2008**	**1/2/2009 - 3/31/2012**
V_{2t} DF	-1.296	-3.937	-2.604
$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$			

The test results for co-integration based on a linear dynamic model show no evidence of linear cointegration in most of the samples.

After removing autocorrelation to avoid bias caused by spurious regression, there is still no presence of linear cointegration relationship in series USO & WTI and DBO & WTI.

But the daily price of USO and DBO tracks the performance of WTI, and oil ETFs consist of oil futures contract, we believe there must be some potential relation between the price of ETFs and the underlying crude oil price of WTI.

In next section, we will investigate the non-linear relationship in prices of USO and DBO with WTI.

6.4 Non-linear Cointegration Relationship

6.4.1 Non-linear Cointegration Generalization

In Engle-Granger methodology, cointegration refers to a linear combination of non-stationary variables $z_t = x_t A y_t$ that is stationary. Actually, in many macroeconomic and financial situations, a linear relationship is not found in non-stationary contexts, but it is possible that a nonlinear long-run equilibrium exists among the integrated variables, even if the variables are not linearly cointegrated.

Escanciano and Escribano (2011) defined nonlinear co-integration as if two or more series are of extended memory, but a nonlinear transformation of them is short memory, then the series are said to be nonlinearly co-integrated. The transformation also could be case of taking difference, but the thing is the vital information can be removed if differencing. Granger (1991) proposed generalizations extended to nonlinear co-integration. The first generalization is that nonlinear transformation of the time series that will be co-integrated in $g(x)$ and $h(y)$, and the linear combination of nonlinearly transformed variables $z_t = g(x) - Ah(y)$ is short memory in mean.

Here Granger (1991) also defined the variable that is short memory / long memory in mean. Given information I_t at time t, if the conditional mean of a variable x at time $t+h$, $E(x_{t+h}|I_t)$ converges to a constant, when $h \to \infty$, then we say the variable x is short memory in mean (SMM). If $E(x_{t+h}|I_t)$ depends on I_t for all h, variable x is long memory in mean (LMM). In long memory series, the shocks have persistent effects.

A second generalization is using time-varying parameters, and the error-correction model equations are in form of:

$$\Delta x_t = \rho_t(t) z_{t-1} + lags \Delta x_t, \Delta y_t + residual$$

$\rho_1(t)$ is the speed of adjustment parameter that is allowed to change over time.

Michael, Nobay and Peel(1997) used nonlinear error-correction in the residuals from linear cointegration to capture the deviations from purchasing power parity (PPP). The nonlinear adjustment process was characterized in terms of an exponential smooth transition autoregressive (ESTAR) model, and concluded that The failure to find co-integration on the basis of a linear model does not necessarily invalidate long-run PPP.

6.4.2 ACE Algorithm

Granger and Hallman (1991) suggested that two series are not cointegrated linearly, but if there exists a nonlinear attractor, it can be viewed as a nonlinear co-integration. In the linear case, if x_t, y_t are I(1) and there exists a linear combination $z_t = x_t - Ay_t$ which is I(0), the line $x = Ay$ can be thought of an attractor. In the nonlinear case, if x_t, y_t are not linearly cointegrated, but we have $q_t = g(x_t) - h(y_t) \sim I(0)$, we define $A = (x, y : g(x) = h(y) \quad \text{or} \quad f(x, y) = 0)$, then A is a nonlinear attractor for x_t and y_t.

Granger and Hallman (1991) showed that the ACE algorithm provides a practical estimation to obtain the nonlinear attractor if there is no prior information about the shape of a possible attractor.

The Alternating Conditional Expectations (ACE) algorithm was originally proposed by Breiman and Friedman (1985). The ACE model can be written as:

$\Theta(y) = \alpha_0 + \sum_{j=1}^{k} \alpha_j(x_j)$

Where $\alpha_j(.)$ is the unknown smooth function. The ACE algorithm maximize the correlation between $\Theta(y)$ and $\sum_{j=1}^{k} \alpha_j(x_j)$, which is equivalent to minimizing the squared error $E\{\Theta(y) - \alpha_0 - \sum_{j=1}^{k} \alpha_j(x_j)\}^2$ subject to $var\{\Theta(y)\} = 1$.

The procedure to estimate ACE algorithm includes four steps:

(i) Initialize to set $\Theta(y) = \dfrac{y - \bar{y}}{\sqrt{var(y)}}$, and set $\alpha_j(x_j)$ as the regression of y on x_j

(ii) Fit an additive model to $\Theta(y)$ to obtain new function $\alpha_1(x_1) \ldots \alpha_k(x_k)$

(iii) Compute $\hat{\Theta}(y) = E\{\sum_{j=1}^{k} \alpha_j(x_j)|y\}$ and standardize the new $\Theta(y)$,

$$\Theta(y) = \frac{\hat{\Theta}(y)}{\sqrt{var(\hat{\Theta}(y))}}$$

(iv) Alternate by repeating (ii) and (iii) until $E\{\Theta(y) - \alpha_0 - \sum_{j=1}^{k} \alpha_j(x_j)\}^2$ converges

6.4.3 Nonlinear Cointegration testing results

In section 6.2, we analyzed data by using Engle and Granger methodology to estimate the linear cointegration relationship between series USO & WTI and DBO & WTI. Although whole period and sub-period datasets of USO, DBO and WTI are I(1), the results suggested there is no linear cointegration in these series for either the full period or sub-periods. In this section, we are interested in finding out if there is any nonlinear cointegration relationship in series.

According to Granger's generalization of nonlinear co-integration, if the residual of transformed series x, y, $q_t = g(x_t) - h(y_t) \sim I(0)$, then we say x and y are non-linearly cointegrated. Figure 6.6 in Appendix D presents the graph of ACE transformations of series $g_1(USO_t)$ and $h_1(WTI_t)$ in full data period from 4/12/2006 to 3/31/2012 and sub-data periods from 4/12/2006 to 12/31/2008 and 1/2/2009 to 3/31/2012, respectively. Similarly, Figure 6.7 in Appendix D shows the ACE transformations of the series $g_2(DBO_t)$ and $h_2(WTI_t)$ from whole period 1/7/2007 to 3/31/2012 and sub-periods 1/7/2007 to 12/31/2008 and 1/2/2009 to 3/31/2012.

Almost all of the transformed series clearly show evidence of nonlinearity, except for series DBO and WTI in sample period of 1/7/2007 to 12/31/2008. Here lines are almost straight and suggest the linearity of the series. In section 6.2, the test for linear cointegration between the series indicated that the linear relationship of DBO & WTI for period of 1/7/2007 to 12/31/2008 is quite strong, which is consistent with the result of ACE transformation.

Table 6.1 compares the R^2 for the linear model and ACE transformation. The ACE transformation produces a better fit than linear model, as all the R^2 from ACE transformation is greater than that of from the estimated linear model of (6.7) (6.12), which suggests that the nonlinear transformation is necessary.

Table _ 6.4 Comparison of R^2 value for linear model and ACE Transformation

		Sample (I)	**Sample (II)**	**Sample (III)**
		4/12/2006 - 3/31/2012	**4/12/2006 - 12/31/2008**	**1/2/2009 - 3/31/2012**
R2	Linear Model	0.3160	0.9438	0.6231
	ACE	0.9841	0.9968	0.8580
		1/5/2007 - 3/31/2012	**1/5/2007 - 12/31/2008**	**1/2/2009 - 3/31/2012**
R2	Linear Model	0.7861	0.976	0.8701
	ACE	0.9762	0.9936	0.969

Table 6.5 shows the DF statistics in the three samples for

$q_{1t} = g(USO_t) - h(WTI_t)$

$q_{2t} = g(DBO_t) - h(WTI_t)$

Table _ 6.5 DF test statistics for q_{1t} and q_{2t}

	Sample (I)	Sample (II)	Sample (III)
	4/12/2006 - 3/31/2012	**4/12/2006 - 12/31/2008**	**1/2/2009 - 3/31/2012**
q_{1t} DF	-3.550***	-4.861***	-5.448***
	1/5/2007 - 3/31/2012	**1/5/2007 - 12/31/2008**	**1/2/2009 - 3/31/2012**
q_{2t} DF	-4.797***	-9.061***	-4.638***

$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

Critical Values for DF test at 1% is -2.569. All the t- statistics are significant at 1% level, suggesting that the null hypothesis of long memory (ie. I(1)) for q_{1t} and q_{2t} should be rejected. Therefore, we conclude that both

$$\left.\begin{array}{l} q_{1t} = g(USO_t) - h(WTI_t) \\ q_{2t} = g(DBO_t) - h(WTI_t) \end{array}\right\} \sim I(0)$$

at full period and sub periods. If we define $A_1 = (USO, WTI : g(USO) = h(WTI))$ and $A_2 = (DBO, WTI : g(DBO) = h(WTI))$, A_1 and A_2 are the nonlinear attractor for series USO_t and WTI_t, and DBO_t and WTI_t.

Previously, we concluded that the data series USO and WTI are not linearly cointegrated, as the residuals of the linear combination of these two series comes out to be I(1). In this section, we extend the concept of linear cointegration to generalization of nonlinear cointegration, and transform the two series by use of the ACE algorithm. Results show that $q_{1t} = g(USO_t) - h(WTI_t)$ is stationary, and a nonlinear attractor A_1 was found as well, which indicates the presence of a nonlinear cointegration relation in series USO and WTI. Since similar situation exists for series DBO and WTI, we also obtain the nonlinear attractor A_2. Therefore, the data series USO & WTI and DBO & WTI are said to be cointegrated nonlinearly, in both the whole date period and sub data periods.

In next chapter, we will use the error correction model to model the cointegration process.

Chapter 7. Error Correction Model (ECM)

7.1 Error Correction Model (ECM)

The Error Correction Model (ECM) is used to model co-integrated processes by estimating the short-term and long-term effects of X on Y between two cointegrated series $\{x_i\}$ and $\{y_j\}$, and the speed that Y returns to the equilibrium after a deviation occurred. Here, we will estimate the ECM with ΔUSO and ΔDBO by introducing both linear and nonlinear models, and to see if the nonlinear model will be better off in the nonlinear context. To estimate the nonlinear ECM model, we are going to use the MARS approach. The reason for choosing MARS approach is because the MARS procedure is powerful to detect and fit models in situations where there are distinct breaks in the model, such as a change of the coefficients. As evidence showed in section 4, there is a distinct breakpoint in the data series of USO, DBO and WTI. Both ECM models will be estimated for USO & WTI and DBO & WTI over three time periods, which had been identified in section 4.1.2. Also, we will use an alternative nonlinear approach GAM, to compare with MARS and linear model OLS.

7.2 Linear Error Correction Model

With the ACE transformation of the series, we obtain the long-run equilibrium for USO & WTI and DBO & WTI. Therefore, in the linear ECM model, we will include the nonlinear cointegration residuals instead of the residuals of the linear combination of $(y_t - \hat{\beta} z_t)$.

The linear form of ECM for USO & WTI:

$$\Delta USO_t = \alpha_{01} + \alpha_{uso}\hat{q}_{1,t-1} + \sum_{i=1} \alpha_{11}(i)\Delta USO_{t-i} + \sum_{i=1} \alpha_{12}(i)\Delta WTI_{t-i} + \varepsilon_{yt} \tag{7.1}$$

$$\Delta WTI_t = \alpha_{02} + \alpha_{wti}\hat{q}_{1,t-1} + \sum_{i=1} \alpha_{21}(i)\Delta USO_{t-i} + \sum_{i=1} \alpha_{22}(i)\Delta WTI_{t-i} + \varepsilon_{zt} \tag{7.2}$$

where $\hat{q}_{1,t-1}$ are the residuals of the ACE algorithm $q_{1t} = g(USO_t) - h(WTI_t)$ at the time t-1, and α_{uso} and α_{wti} are the speed of adjustment.

In equation (7.1) and (7.2), α_{uso} and α_{wti} are the parameters to adjust the change of USO and WTI in response to the previous periods deviation from long-term equilibrium $g(USO_t) - h(WTI_t)$. Similarly, for DBO & WTI, the ECM in linear form:

$$\Delta DBO_t = \alpha_{01} + \alpha_{dbo}\hat{q}_{1,t-1} + \sum_{i=1} \alpha_{11}(i)\Delta DBO_{t-i} + \sum_{i=1} \alpha_{12}(i)\Delta WTI_{t-i} + \varepsilon_{yt}$$

$$\Delta WTI_t = \alpha_{02} + \alpha_{wti}\hat{q}_{1,t-1} + \sum_{i=1} \alpha_{21}(i)\Delta DBO_{t-i} + \sum_{i=1} \alpha_{22}(i)\Delta WTI_{t-i} + \varepsilon_{zt}$$

$\hat{q}_{2,t-1}$ are the residuals of $q_{2t} = g(DBO_t) - h(WTI_t)$

7.3 Nonlinear Error Correction - MARS

The multivariate adaptive regression splines (MARS) approach is a procedure to describe nonlinear relationship between the response variable and set of explanatory variables by defining the spline knots, which are breakpoints or changes in a model coefficient. MARS can be written in the form of

$y = \alpha + c_1(x - \tau*)_+ - c_2(\tau * -x)_+ + e$

Where $\tau*$ is the knot point, $(.)_+$ is the right truncated spline function which takes the maximum value on max(0, (.)).

The MARS model also can identify the complex nonlinear interactions between variables. An interaction model for $y = f(x, z)$ can be written:

$y = \alpha + c_1(x - \tau_1*)_+ - c_2(\tau_1 * -x)_+ + c_3(x - \tau_1*)_+(z - \tau_2*)_+ + e$

when $x > \tau_1*$ and $z > \tau_2*, y = \alpha + c_1(x - \tau_1*)_+ + c_3(x - \tau_1*)_+(z - \tau_2*)_+ + e$

when either $x < \tau_1*$ or $z < \tau_2*$, or both $x < \tau_1*, z < \tau_2*$, the interaction term $c_3(x - \tau_1*)_+(z - \tau_2*)_+$ equals 0.

Friedman introduced the MARS approach in 1991. Stokes (2005) explained in detail how to perform the MARS method by using B34S ProSeries Econometric System and SCA WorkBench software. The MARS approach is such a powerful data mining methodology that it has extensive and increasing applications in different fields, such as macro economy, finance and social science.

In our case, to study the dynamic relationship between the prices of ETFs and the oil price WTI, we will involve the same variables and same order of lags in the MARS ECM model as those of the linear ECM models. For example, when estimating the ECM for USO & WTI in equation (7.1), the left hand-side variable is ΔUSO_t, and right-hand side variables include $\hat{q}_{1,t-1}$, lags in ΔUSO_t and lags in ΔWTI_t, equation(7.2) the left hand-side variable is ΔWTI_t, and right-hand side variables $\hat{q}_{1,t-1}$, lags in ΔUSO_t and lags in ΔWTI_t. Then we can compare the results of ECM by using linear and nonlinear models.

7.4 Results of Error Correction Model

7.4.1 Linear Model

To estimate the ECM, we use one lag in the model for both linear and nonlinear ECM over three sample periods, as the residuals appear to be clean with such number of lags. The models are as follows:

$$\Delta USO_t = \alpha_{01} + \alpha_{uso}(1)\hat{q}_{1,t-1} + \alpha_{11}(1)\Delta USOt - 1 + \alpha_{12}(1)\Delta WTI_{t-1} + \varepsilon_{yt} \tag{7.3}$$

$$\Delta WTI_t = \alpha_{02} + \alpha_{wti}(1)\hat{q}_{1,t-1} + \alpha_{21}(1)\Delta USOt - 1 + \alpha_{22}(1)\Delta WTI_{t-1} + \varepsilon_{zt} \tag{7.4}$$

$$\Delta DBO_t = \alpha_{01} + \alpha_{dbo}(2)\hat{q}_{2,t-1} + \alpha_{11}(2)\Delta DBOt - 1 + \alpha_{12}(2)\Delta WTI_{t-1} + \varepsilon_{yt} \tag{7.5}$$

$$\Delta WTI_t = \alpha_{02} + \alpha_{wti}(2)\hat{q}_{2,t-1} + \alpha_{21}(2)\Delta DBOt - 1 + \alpha_{22}(2)\Delta WTI_{t-1} + \varepsilon_{zt} \tag{7.6}$$

Table 7.3 shows the estimates of coefficients for linear ECM models with associated t-value

Table _ 7.3 Coefficients for Linear ECM models for sample (I) (II) (III)

Sample	**(I)**		**(II)**		**(III)**	
	4/12/2006 - 3/31/2012		**4/12/2006 - 12/31/2008**		**1/2/2009 - 3/31/2012**	
	ΔUSO_t	ΔWTI_t	ΔUSO_t	ΔWTI_t	ΔUSO_t	ΔWTI_t
α	-0.0220	0.0345	-0.0574	-0.0305	0.0028	0.0841
	(-0.69)	(0.67)	(-0.92)	(-0.35)	(0.10)	(1.41)
q_{t-1}	-0.0010	-0.0022	-0.0162	0.0034	-0.0519***	-0.0954***
	(-0.44)	(-0.62)	(-0.97)	(0.15)	(-4.03)	(-3.37)
ΔUSO_{t-1}	-0.1042*	0.3116***	-0.1325	0.3414***	0.0258	0.8092***
	(-1.89)	(3.52)	(-1.54)	(2.85)	(0.29)	(4.12)
ΔWTI_{t-1}	0.0381	-0.2082***	0.0512	-0.2936***	-0.0149	-0.3161***
	(1.12)	(-3.79)	(0.83)	(-3.42)	(-0.37)	(-3.57)
	1/5/2007 - 3/31/2012		**1/5/2007 - 12/31/2008**		**1/2/2009 - 3/31/2012**	
	ΔDBO_t	ΔWTI_t	ΔDBO_t	ΔWTI_t	ΔDBO_t	ΔWTI_t
α	0.0077	0.0477	-0.0123	-0.0259	0.0113	0.0717
	(0.44)	(0.83)	(-0.35)	(-0.23)	(0.67)	(1.19)
q_{t-1}	-0.0128*	-0.0262	-0.2338***	-0.1723	-0.0477***	-0.1675***
	(-1.74)	(-1.07)	(-4.29)	(-0.98)	(-3.05)	(-3.01)
ΔDBO_{t-1}	-0.1366**	0.4718***	-0.1663**	0.4030	0.1377*	0.8834***
	(-2.50)	(2.61)	(-1.97)	(1.47)	(1.76)	(3.17)
ΔWTI_{t-1}	0.0335**	-0.1632***	0.0228	-0.2039**	-0.0380*	-0.2063***
	(2.03)	(-2.98)	(0.82)	(-2.26)	(-1.74)	(-2.65)

*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$

The q_{t-1} is the deviation from previous periods equilibrium. The prices of USO and WTI or DBO and WTI changed in response to the previous periods deviation by the speed of adjustment parameters α_{uso} or α_{dbo} and α_{wti}.

In long-term relationship for oil ETFs and WTI, both α_{uso} and $\alpha_{dbo} \in [-1, 0]$ for three samples, which imply that price of USO and DBO converge or balance back to an equilibrium in the long run. The speed of adjustment parameters α_{uso} and α_{wti} have low t-values in sample (1) and sample (2), but statistically significant at 1% level in sample (3). For price of DBO and WTI, α_{dbo} and α_{wti} are significant at 1% level in sample (3) as well. The results indicate that the adjustment to long-term equilibrium is more notably for period after 2008 financial crisis. One reasonable interpretation is that the higher market volatility of post crisis caused the greater deviation from long-term equilibrium, so the system reacted more actively to move back to the equilibrium. Also, for DBO and WTI, α_{dbo} is significant in sample (1) and (2), while α_{wti} is insignificant, suggest DBO plays a dominant role in interaction relationship to adjust the price in response to the equilibrium error.

In short-term relationship, for USO and WTI, the coefficient of ΔUSO_{t-1} in (7.4) is greater than that of in (7.3) ($\alpha_{21} > \alpha_{11}$), and similarly for coefficient of ΔWTI_{t-1}, $\alpha_{22} > \alpha_{12}$ in all three samples. We also noticed that all α_{21} and α_{22} have a significant t-statistics over three samples. This result suggests that the price change of WTI respond more significantly to past changes in USO price and lags of its own. The t-statistics for DBO and WTI also suggest that ΔWTI responds significantly to past changes in DBO and its own price, especially in sample (1) and sample (3). In sample (2), seems ΔWTI_{t-1} is not a notable effect for the price change of WTI. Compared to ΔUSO, ΔDBO responds more significantly to the past price change of WTI and its own. But the coefficients in (7.6) are still greater than that of in (7.5), $\alpha_{21} > \alpha_{11}$ and $\alpha_{22} > \alpha_{12}$ imply that ΔWTI responses more strongly to the price change of previous period.

So from the linear error correction model, we can conclude:

1) For both USO & WTI and DBO & WTI, the adjustment to long-term equilibrium is more notably for period after 2008 financial crisis. Learning from previous discussion, there is a structure break in the entire data from 4/12/2006 to 3/31/2012, and the break point is

exactly at the time of 2008 financial crisis. During that time, the crude oil price experienced a dramatic drop, and it began to recover gradually since 2009. But for both the crude oil market and stock market, it turned at to be more volatile and unpredictable after the financial crisis due to economic uncertainties. Since the oil ETFs consists of listed crude oil futures and other oil related futures, and tries to follow the performance of crude oil, its reasonable to find that the price of oil ETFs and WTI in sample (3) (which is from 1/2/2009 3/31/2012, after the financial crisis period) has the different movement toward equilibrium point when the market is greatly volatile.

2) In short-term, the price change of WTI responds more significantly than oil ETFs to its own lags and the lags of oil ETFs. But in the long-term relationship, oil ETFs (especially DBO) dominates the adjustment to the long run equilibrium.

3) The responses of ΔUSO and ΔDBO to price change in lag of WTI and its own lags are not the same. The difference in strategy of USO and DBO may explain why it happens. For ETF USO, it consists of near-month futures contracts of crude oil. Usually the maturity of the near-month futures contracts is less than 22 days. USO rolls over the near-month futures contracts about two weeks in advance before its expiration into the next month contracts. On the other hand, the investment of DBO on futures contracts is not limited to the near-month futures, but comprises of a portfolio of futures contracts with maturity from near-month to 13 month, to follow the Optimum Yield Crude oil Index, in order to produce the best possible roll yield. The specific futures it holds are not fully revealed for each month.

7.4.2 ECM with MARS model

As we already know, the MARS function uses the combination of truncated spline functions to estimate the model, and the knots represent the potential thresholds in the independent variables. When identifying a MARS model, we need to determine the maximum number of basis functions or knots (set as nk) and the maximum number of variables in each basis function (set as mi). In this study, we set nk=20, mi=1 with no interactions, and then try mi=2.

For the oil ETFs and crude oil price response to previous period changes, we get more detail information by using MARS model. Table 7.4A listed the knots, coefficient and associated t-value estimated for MARS model with left hand side variable ΔUSO_t and ΔWTI_t in sample (1).

Table _7.4A MARS ECM for ΔUSO & ΔWTI in sample (1)

(1)			
4/12/2006 - 3/31/2012			
ΔUSO_t **(left-hand variable)**	**Slope**	ΔWTI_t **(left-hand variable)**	**Slope**
$\Delta USO\{1\} < -3.58$	0.672	$\Delta USO\{1\} > -3.58$	0.374
	(3.36)		(3.92)
$\Delta WTI\{1\} > -2.10$	-7.931	$\Delta USO\{1\} > 1.34$	-0.440
	(-3.12)		(-2.14)
$\Delta WTI\{1\} > -5.12$	0.211	$\Delta WTI\{1\} > 4.53$	-5.904
	(2.88)		(-3.88)
		$\Delta WTI\{1\} > 4.13$	5.284
			(3.54)
		$\Delta WTI\{1\} > -2.10$	-0.247
			(-3.64)
MARS model estimated with 20 knots, 1 lag and no interaction term (mi=1)			

The corresponding ECM in MARS models are:

$$\Delta USO_t = 0.672 * (-3.58 - \Delta USO_{t-1})_+ - 0.258 * (\Delta WTI_{t-1} + 2.10)_+ + 0.211 * (\Delta WTI_{t-1} + 5.12)_+ \tag{7.7}$$

$$\Delta WTI_t = 0.374 * (\Delta USO_{t-1} + 3.58)_+ - 0.440 * (\Delta USO_{t-1} - 1.34)_+ - 5.904 * (\Delta WTI_{t-1} - 4.53)_+ + 5.284 * (\Delta WTI_{t-1} - 4.13)_+ - 0.247 * (\Delta WTI_{t-1} + 2.10)_+ \tag{7.8}$$

Model (7.7) tells us when price change of WTI in previous day is less than -5.12, ΔWTI_{t-1} has no impact on ΔUSO_t, and if ΔWTI_{t-1} is between -2.10 and -5.12, the impact is positive. If the price change of USO in previous day exceeds the threshold of -3.58, the impact to ΔWTI in current period is positive. If ΔUSO_{t-1} is less than -3.58, there is limit influence to ΔWTI_t. The findings from (7.7) and (7.8) says that the interaction between price change in USO and WTI become weak when yesterdays price change in USO is less than -3.58, and in WTI less than -5.12.

In linear ECM model, ΔWTI_{t-1} has no significant impact on USOt. But here estimated by MARS model, we get more accurate information, even at what range of price changes in WTI yesterday will affect the change in USO today. MARS model explain more in detail the short-term dynamics between two variables.

Table 7.4B & 7.4C listed the knots, coefficient and associated t-value estimated for MARS model with left hand side variable ΔUSO_t and ΔWTI_t in sample (2) and (3).

Another finding by MARS model which is consistent with the linear ECM is that in sample (1) and sample (2), the long-term cointegration relationship between USO and WTI is insignificant, but statistically significant in sample (3).

Table _ 7.4B MARS ECM for ΔUSO & ΔWTI in sample (2)

(2)			
4/12/2006 - 12/31/2008			
ΔUSO_t **(left-hand variable)**	**Slope**	ΔWTI_t **(left-hand variable)**	**Slope**
$\Delta USO\{1\} < -3.58$	0.805	$\Delta WTI\{1\} > 4.42$	-0.775
	(3.00)		(-5.53)
$\Delta WTI\{1\} > -2.62$	-0.661		
	(-3.35)		
$\Delta WTI\{1\} > -4.57$	0.580		
	(3.18)		
MARS model estimated with 20 knots, 1 lag and no interaction term (mi=1)			

Table _ 7.4C MARS ECM for ΔUSO & ΔWTI in sample (3)

(3)			
1/2/2009 - 3/31/2012			
ΔUSO_t **(left-hand variable)**	**Slope**	ΔWTI_t **(left-hand variable)**	**Slope**
$\hat{q}_t\{1\} > 3.91$	-0.424	$\Delta USO\{1\} > -1.52$	1.128
	(-3.47)		(5.31)
$\hat{q}_t\{1\} < 3.91$	0.039	$\Delta WTI\{1\} > 3.34$	1.043
	(2.86)		(3.45)
		$\Delta WTI\{1\} > -3.77$	-0.490
			(-5.05)
		$\hat{q}_t\{1\} > 3.91$	-0.788
			(-2.93)
		$\hat{q}_t\{1\} < 3.91$	0.065
			(2.17)
MARS model estimated with 20 knots, 1 lag and no interaction term (mi=1)			

In MARS ECM for DBO, the t-statistics for the deviation from the previous equilibrium $\hat{q}_{t-1}$ is significant for both ΔDBO and ΔWTI over three samples. In linear model, the long-term equilibrium is not significant for ΔWTI in sample (1) and (2). MARS finds the significant knots for ΔWTI at 4.87, 6.25 and 3.90 in sample (1) and 1.56, -0.12, and -0.88 in sample (2). This is because MARS estimate the model flexibly at different threshold levels, which is more advanced than the regular linear model. MARS also tells us in sample (1) when the price change of WTI in previous day is less than -3.12, the impact of ΔWTI_{t-1} on ΔDBO_t is negative.

Table 7.5A - C shows the result of MARS ECM for DBO & WTI in three samples.

Table _ 7.5A MARS ECM for ΔDBO & ΔWTI in sample (1)

(1)			
1/5/2007 - 3/31/2012			
ΔDBO_t **(left-hand variable)**	**Slope**	ΔWTI_t **(left-hand variable)**	**Slope**
$\Delta DBO\{1\} > 1.51$	-1.197	$\Delta WTI\{1\} > 4.65$	-4.05
	(-3.40)		(-3.64)
$\Delta DBO\{1\} < 1.51$	0.777	$\Delta WTI\{1\} > 4.15$	3.125
	(2.89)		(3.02)
$\Delta DBO\{1\} > -1.65$	0.742	$\hat{q}_t\{1\} > 4.87$	3.515
	(2.73)		(3.83)
$\Delta WTI\{1\} < -3.12$	-0.128	$\hat{q}_t\{1\} > 6.25$	-2.326
	(-3.27)		(-3.28)
$\hat{q}_t\{1\} > 4.166$	-0.465	$\hat{q}_t\{1\} > 3.90$	-1.702
	(-4.47)		(-3.60)
$\hat{q}_t\{1\} > 5.756$	-6.627		
	(-6.66)		
$\hat{q}_t\{1\} > 5.555$	6.623		
	(6.42)		
MARS model estimated with 20 knots, 1 lag and no interaction term (mi=1)			

Table _ 7.5B MARS ECM for ΔDBO & ΔWTI in sample (2)

(2)			
4/12/2006 - 12/31/2008			
ΔDBO_t **(left-hand variable)**	**Slope**	ΔWTI_t **(left-hand variable)**	**Slope**
$\Delta DBO\{1\} < -1.65$	1.393	$\Delta WTI\{1\} > 4.18$	0.912
	(3.63)		(2.59)
$\Delta DBO\{1\} > -1.25$	-0.165	$\hat{q}_t\{1\} > 1.555$	2.885
	(-3.10)		(2.40)
$\Delta WTI\{1\} > -4.69$	1.617	$\hat{q}_t\{1\} > -0.118$	-1.019
	-2.89		(-3.84)
$\Delta WTI\{1\} > -4.29$	-1.558	$\hat{q}_t\{1\} < -0.879$	-7.421
	(-2.74)		(-5.13)
$\hat{q}_t\{1\} > -0.388$	-0.364		
	(-5.77)		
$\hat{q}_t\{1\} < -0.388$	-0.386		
	(-2.32)		
MARS model estimated with 20 knots, 1 lag and no interaction term (mi=1)			

Table _ 7.5C MARS ECM for ΔDBO & ΔWTI in sample (3)

(3)			
1/2/2009 - 3/31/2012			
ΔDBO_t **(left-hand variable)**	**Slope**	ΔWTI_t **(left-hand variable)**	**Slope**
$\Delta DBO\{1\} < 0.550$	-0.295	$\Delta DBO\{1\} > -1.13$	0.964
	(-3.48)		(3.53)
$\Delta WTI\{1\} > 3.34$	0.273	$\Delta DBO\{1\} < -1.13$	-2.518
	(3.42)		(-3.13)
$\Delta WTI\{1\} < 3.34$	0.087	$\Delta WTI\{1\} < 1.520$	0.307
	(3.80)		(3.41)
$\hat{q}_t\{1\} > -1.246$	-0.385	$\hat{q}_t\{1\} > 1.899$	-4.359
	(-2.94)		(-3.24)
$\hat{q}_t\{1\} > 0.158$	-13.34	$\hat{q}_t\{1\} > -0.691$	5.456
	(-3.70)		-2.95
$\hat{q}_t\{1\} > 0.360$	-3.12	$\hat{q}_t\{1\} > -0.427$	-2.591
	(-3.42)		(-2.28)
$\hat{q}_t\{1\} > 0.201$	16.189	$\hat{q}_t\{1\} > -1.011$	-2.968
	(3.73)		(-3.55)
MARS model estimated with 20 knots, 1 lag and no interaction term (mi=1)			

More information will be obtained in response of price change in oil ETFs and WTI to the change in previous period if MARS model including interaction terms (mi-2) are used. Knots and Slopes for MARS (mi=2) are presented in Appendix E.

In next section, we will discuss an alternative nonlinear approach, the GAM (General Additive Model), to further determine if MARS could outperform other nonlinear alternative method.

7.5 Alternative Nonlinear Approach - GAM

The basic idea of GAM (Generalized Additive Model) approach, proposed by Hastie and Tibshirani (1986), is to use the additive predictor $\Sigma\alpha_j(x_j)$ replacing linear predictor $\Sigma\beta_j x_j$ in linear regression model, x_j is the explanatory variable. The relax of the linear predictor to be smooth function could improve the prediction accuracy in nonlinear case. The GAM model can be written in form of conditional expectation:

$$E(y|x_1, x_2, \ldots, x_k) = \alpha_0 + \sum_{j=1}^{k} \alpha_j(x_j)$$

where $E\alpha_j(x_j) = 0$, $\alpha_j(.)$ are smooth function estimated by a scatterplot smoother.

The estimated coefficients for GAM model are presented in Appendix F. The results show that the significance of α_{uso} and α_{dbo} has not changed compared to linear ECM model, but $e'e$ has been reduced in all equations over three samples.

Table 7.6 shows the RSS for linear and nonlinear (GAM and MARS) models.

All $e'e$ from nonlinear model are less than that of linear model. Also note that MARS(mi=1) produces a slightly lower $e'e$ in 10 equations out of 12 than GAM, and MARS(mi=2) gives a even lower $e'e$ to all the other method. This indicates that MARS model out performed the other methods.

(All computations in this study have been done with B34S version 8.13 and RATS 8.0)

Table _ 7.6 Residual Sum of Squared for OLS, GAM and MARS

		RSS		
		Sample (I) **4/12/2006 - 3/31/2012**	**Sample (II)** **4/12/2006 - 12/31/2008**	**Sample (III)** **1/2/2009 - 3/31/2012**
ΔUSO_t	OLS	2286.9	1777.6	484.3
	GAM	2269.7	1754.3	476.7
	MARS (mi=1)	2268.6	1748.4	478.9
	MARS (mi=2)	2196.4	1724.0	456.3
ΔWTI_t	OLS	5909.7	3467.2	2347.8
	GAM	5765.8	3317.8	2311.8
	MARS (mi=1)	5752.6	3377.6	2290.3
	MARS (mi=2)	5428.5	2730.3	2192.9
		1/5/2007 - 3/31/2012	**1/5/2007 - 12/31/2008**	**1/2/2009 - 3/31/2012**
ΔDBO_t	OLS	514.4	307.4	187.7
	GAM	509.3	294.0	184.5
	MARS (mi=1)	492.0	287.6	178.3
	MARS (mi=2)	481.30	282.5	180.0
ΔWTI_t	OLS	5657.7	3223.8	2376.3
	GAM	5511.7	2950.9	2332.1
	MARS (mi=1)	5471.6	2910.2	2298.8
	MARS (mi=2)	4958.6	2760.3	2225.9

Chapter 8. Conclusion and Future Work

8.1 Conclusion

Crude oil is one of the most heavily traded commodities in the world market. The topic of crude oil, crude oil futures and oil ETFs draws great interest from economists, financial investors and researchers. Studies show the price of crude oil is impacted by the world economy, and even political influence. It's not easy to forecast the price of crude oil. Some people think oil futures, which is an energy derivative product, could provide information on prospective oil prices. The expansion of oil ETFs brings the investor more choices, to a large extent, in energy markets. So new questions emerge: what is the relationship between oil ETFs and the underlying crude oil price? Could oil ETFs be a source of information on the crude oil price?

This paper investigated the performance of the most popular oil ETFs in the US market, USO and DBO, and the underlying crude oil WTI, in conjunction with statistical analysis to shed light upon the dynamic relationship between oil ETF USO and DBO with the benchmark crude oil WTI. We used daily time series data on the price of USO in the period of April 12, 2006 to March 31, 2012, DBO January 5, 2007 to March 31, 2012 and the corresponding periods of the price of WTI to identify the co-integration relationship between oil ETFs and WTI, estimating the linear and nonlinear ECM and further examined the reaction behavior between these variables.

(I) nonlinear cointegration relationship between oil ETF and WTI
There are several important findings of this study. In light of the results, no matter what the strategy employed for the oil ETFs and which asset holdings are in security basket, a nonlinear co-integration relationship between oil ETFs and WTI has been detected. Our findings show using an ACE model, the nonlinear long-run equilibrium was found. This finding was based on the theory developed by Granger, who extends generalizations of co-integration relationship to nonlinear contexts in 1991.

(II) The feedback between Oil ETF and WTI is bi-directional
Granger causality analysis was employed to test if the series of ETFs or WTI could be

useful in forecasting the other. The result suggests the feedbacks in both directions are strong and statistically significant, except for the series of USO & WTI in the period after the financial crisis.

The Error Correction Model was estimated on the premise that a co-integration relationship exists. Both linear and nonlinear ECM models were estimated in order to compare two types of models and to test whether the nonlinear model will be better off under the nonlinear contexts. Here, the MARS model was used for nonlinear estimation. MARS ECM model provides more accurate information of the interaction of variables. Another important finding is the interaction between ETFs and WTI in the period after the financial crisis in late 2008 is different from that of the whole data period and period before the financial crisis. The explanation of this is the impact of the global economic turmoil during and after year 2008. The financial crisis in 2008 had a great impact on the price of crude oil. During that time, the demand collapsed, which caused increase in storage level, creating the price fluctuations in crude oil and oil futures.

The results of ECM estimation also suggest that the feedback between oil ETFs and WTI is in two directions. As we know, the impact of oil ETFs on WTI stem from the nature of the oil ETFs, which consist of oil futures contracts. So the price of oil ETFs exerts the influence on the price of WTI through the oil futures. On the other hand, oil ETFs is tracking the performance of WTI, thus, the impact of WTI on oil ETFs must exist.

(III) lead and lag position

This paper also proposes a comparatively accurate description of the dynamic relationship between oil ETFs and WTI by using linear and nonlinear (MARS model) Error Correction Model. Generally, in long-term relationship, oil ETFs (especially DBO) takes the lead to dominate the adjustment of price change to the long run equilibrium. While in short term, the price change of crude oil price responds more significantly to the changes in price of oil ETFs for the previous day. In other words, WTI takes the lead in short term in response of the price change of previous day.

With more accurate estimation for the full data sample, when price of WTI decreases in the range of \$2.10 to \$5.12 the day before, price of USO will increase this day. But for DBO, the share of DBO will decrease if WTI decreases more than \$3.12 the previous day. If the price of USO goes down less than $3.58 in previous day, share of WTI will go up this day.

8.2 Future Work

In discussion of this paper, there are some limitations which are left for future work.

1) *Impact of roll cost.* The roll penalty occurs when funds sell the expiring futures and replace the new contract. For example, USO is using near-month futures contract. Shortly before the futures contracts expire, the fund will sell the futures and pay for the next-month contract. Usually each months contract is a bit more expensive than the previous month due to the cost of storing a commodity. For example, in year 2010, US Oil Fund holds up to 24 million barrels of oil in futures contracts. The roll penalties generated are from $0.30 to $4 per barrel at different points. Actually, US Oil fund realized such potential problem, and tried to reduce the impact of roll penalty in 2009: they changed the strategy from rolling the entire futures position in a single day to over four days period each month. For a single day strategy, because the Fund roll date is publicly available, oil traders would sell the front month futures contract ahead of that date, that would push the price down. Since oil for delivery in the future is higher than spot prices, then USO would pay more to buy the next month contract. That will raise the roll cost.

As for Deutsche Banks PowerShares DB Oil Fund (DBO), it has used an optimal rolling strategy instead of rolling the expiring contract to the next month. The strategy is more flexible and searches for the best possible roll yield.

From the scatter plot on Figure_1 in Chapter 1, we noticed that the price of oil ETFs match the price of crude oil quite well at the beginning, but deviate and underperform more and more gradually. Table 8.1 below shows the average price change in percentage.

The average price for WTI goes up 28.29% and 19.37% in 2010 and 2011, while USO and DBO shares just up 6.44% and 10.90%, 3.26% and 9.71% in 2010 and 2011, respectively. The oil ETFs underperformed the underlying oil price in year 2010 and 2011, but nevertheless the performance of DBO is a bit better than that of USO.

Table _ 8.1 Percentage Change in Average Price

Yearly Average Price			
Year	**USO**	**DBO**	**WTI**
2009	34.28	23.46	61.95
2010	36.49	26.01	79.48
2011	37.68	28.54	94.87
Average Price Change in Percentage			
2010	6.44%	10.90%	28.29%
2011	3.26%	9.71%	19.37%

One reason for the underperformance is that of roll penalty or roll cost. Is it that the roll penalty to cause the spread deviations or are there any other factors? Also, is that the flexible roll strategy to make the performance of DBO better off USO? In future study, the fund strategy and roll cost will be taken into account to conduct the research.

2) *Separate the data into three periods pre-crisis, crisis and post-crisis instead of two to study the 2008 financial crisis.* In this paper, we mentioned there is an obvious structure breakpoint during the financial crisis in 2008. In the period after the financial crisis in 2008, the interaction of oil ETF and crude oil present different results in the previous period and the whole period. Salisu and Fasanya (2012) used the asymmetric GARCH models to compare the volatility of oil price over three samples: pre-crisis, during crisis and after crisis, and found out that the volatility of the oil price during the financial crisis is the greatest. So how on earth does the financial crisis and the volatility of the oil price will impact the relationship among the price of crude oil, oil futures, and then the oil ETF, should be an extensive topic for this research.

3) *Role of Volume in lead-lag position.* Third, in this paper, we have never mentioned the trading volume of oil ETF, crude oil and oil futures. What the role of volume is in the relationship of oil ETFs and oil futures would be an interesting question. For example, the oil ETFs consist of the oil futures and track the crude oil price, but if a significant volume player enters the market and there is enough volume in an oil ETF, it could be the price leader of the oil futures, and then generate the impulse in the interaction relationship.

References

Alquist, R and Lutz Kilian (2010), What do we learn from the price of crude oil futures, *Journal of Applied Econometrics,* J. Appl. Econ. 25: 539-573 (2010)

Bekiros, S. D. and Diks, C. G.H. (2008), The relationship between crude oil spot and futures prices: Cointegration, linear and nonlinear causality, *Energy Economics*, Volume 30, Issue 5, September 2008, Pages 2673-2685

Blume, Marshall E. , and Roger M. Edelen (2004), S&P 500 Indexers, Tracking Error, and Liquidity, *The Journal of Portfolio Management,* spring 2004

Breiman, L. and Friedman, J.H. (1985), Estimating optimal transformations for multiple regression and correlation, *Journal of the American Statistical Association*, September 1985, Vol. 80, No. 391

Brown, R.L., Durbin, J. and Evans, J.M., (1975), Techniques for Testing the Constancy of Regression Relationships over Time, *Journal of Royal Statistical Society,* B, Vol. 37 pp149-192

Chatrath, A., Miao, H. and Ramchander, S. (2012), Does the price of crude oil respond to macroeconomic news?, *The Journal of Futures Markets,* Vol, 32, No. 6, 536-559 (2012)

Chu, Patrick Kuok-Kun (2011), Study on the tracking errors and their determinants: evidence from Hong Kong exchange traded funds, *Applied Financial Economics,* Vol 21, Issue 5, 2011

Deville, L (2008), Exchange Traded Funds: History, Trading and Research, *Handbook of Financial Engineering,* Springer (Ed.),(2008) 1-37

Elton, E., Gruber, M., Comer, G., Li, K. (2002), Spiders: Where are the Bugs?,*Journal of Business*, Vol. 75(3): pp453-472

Elton, E., Gruber, M., Busse, J. (2004), Are Investors Rational? Choices Among Index Funds, *Journal of Finance,* Vol. 59(1): pp261-288

Enders, W. (2004), Book: *Applied Econometric Time Series*

Engle R. F., and Granger C.W.J (1987), Co-Integration and Error Correction: Representation, Estimation, and Testing, *Econometric: Journal of the Econometric Society,* Vol. 55, No.2 (Mar., 1987), pp. 251-276

Engle R. F., and Yoo B.S (1987), Forecasting and Testing in Cointegrated Systems, *Journal of Econometrics,* 35, (1987) 143-159

Escanciano, J.C. and Escribano, A., (2011), Econometrics: Non-linear Cointegration, Complex Systems in *Finance and Econometrics,* 2011, pp 203-215

Frino, A., Gallagher, D. (2001), Tracking S&P 500 Index Funds, *Journal of Portfolio Management,* Vol. 28(1): pp44-55

Gastineau, G. L. (2004), The Benchmark Index ETF Performance Problem, *Journal of Portfolio Management,* v30(2), 96-103.

Granger, C. W. J. (1969), Investigating Causal Relations by Econometric Models and Cross-spectral methods, *Econometrica,* vol. 37, No.3 (Aug., 1969), pp. 424-438

Granger, C. W. J. (1981), Some Properties of Time Series Data and Their Use in Econometric Model Specification, *Journal of Econometrics* 16 (1): 121130.

Granger, C. W. J. and Hallman, J. (1991), Long Memory Series with Attractors, *Oxford Bulletin of Economics and Statistics,* 53, 1 (1991)

Hastie, T., and Tibshirani, R. (1986), Generalized Additive Model, *Statistical Science,* 1986, Vol. 1, No. 3, 297-318

Hendry, D.F. (1995), *Dynamic Econometrics.* Oxford: Oxford University Press

Hendry, D. F., and Juselius, K (1999), Explaining Cointegration Analysis: Part I

Kostovetsky, L. (2003), Index Mutual Funds and Exchange-Traded Funds, *Journal of Portfolio Management*, v29(4), 80-92

Murdock, M. and Richie, N. (2008), The United States Oil Fund as a hedging instrument, *Journal of Asset Management*, Vol. 9, 5, 222-346, 2008

Nyblom, J. (1989), Testing for the Constancy of Parameters Over time, *Journal of the American Statistical Association,* Vol. 84, No. 405, pp 223-230

Pinnuck, M., Gallagher, D. (2006), Seasonality in Fund Performance: An Examination of the Portfolio Holdings and Trades of Investment Managers, *Journal of Business, Finance and Accounting*, Vol. 33 (7-8): 1240-1266,

Pope, P., Yadav, P. (1994), Discovering Errors in Tracking Error, *Journal of Portfolio Management,* Vol. 20(2): pp27-32

Roll, R. (1992), A Mean/Variance Analysis of Tracking Error, *Journal of Portfolio Management,* Vol. 18(4): pp13-22

Sadorsky, P. (2012), Correlations and volatility spillovers between oil prices and the stock prices of clean energy and technology companies, *Energy Economics,* Volume 34, Issue 1, January 2012, Pages 248-255

Salisu, A. A. and Ismail O. Fasanya, (2012), Comparative Performance of Volatility Models for Oil Price, *International Journal of Energy Economics and Policy*, Vol.2, No. 3, 20

Segara, R., Gallagher,. D. (2006), The Performance and Trading Characteristics of Exchange-Traded Funds, *Journal of Investment Strategy,*

Shin, S., Gke, Soydemir (2010), Exchange-traded funds, persistence in tracking errors and information dissemination, *Journal of Multinational Financial Management*, Volume 20, Issues 45, December 2010, Pages 214-234

Silvapulle, P. and Moosa, I. A. (1999), The relationship between spot and futures prices: evidence from the crude oil market, *The Journal Futrues Markets*, Vol. 19, No.2, 175-193 (1999)

Sims, C., J. Stock and M. Watson (1990), Inference in Linear Time Series Models with some Unit Roots, Econometrica, Vol. 58, No. 1 (January, 1990), p.113-144

Stock, J.H., and Watson, M. W. (2002), Has the Business Cycle Changed and Why? *NBER Macroeconomics Annual 2002,* Volume 17

Stokes, H.H. (1997), Specifying and Diagnostically Testing Econometric Models, New York: Quorum Books, 1997

Stokes, H.H., and McDonald J.F. (2013), The housing price bubble, the monetary policy and the foreclosure crisis in the US, *Applied Economics Letters*

Westgaard, S, Estenstad, M, Seim, M and S. Frydenberg, (2011), Co-integration of ICE Gas oil and Crude oil futures, *Energy Economics,* Vol 33, Issue#2, March 2011, Pages 311-320

Appendices

Appendix A - VAR Lag Selection

VAR Lag Selection for USO & WTI in Sample (I)			
Lags	**AICC**	**SBC/BIC**	**HQ**
0	25654.5157	25665.1126	25658.4634
1	8812.9207	8844.6951	8824.7475
2	8680.4105	**8733.3406***	8700.0947
3	8671.8883	8745.9524	8699.4081
4	8664.729	8759.9051	8700.0624
5	8635.4535	8751.7195	**8678.5784***
6	8628.9682	8766.3022	8679.8627
7	8634.7294	8793.1090	8693.3712
8	8631.6187	8811.0217	8697.9855
9	8634.0493	8834.4532	8708.1187
10	8632.7918	8854.1743	8714.5413
11	8629.8895	8872.2279	8719.2966
12	8629.6090	8892.8806	8726.6509
13	8630.0227	8914.2049	8734.6768
14	8634.1707	8939.2405	8746.4141
15	8625.1389	8951.0735	8744.9486
16	**8620.2798***	8967.0561	8747.6328
17	8627.8155	8995.4103	8762.6887
18	8626.7240	9015.1142	8769.0942
19	8630.9023	9040.0644	8780.7460
20	8628.6890	9058.5997	8785.9830

VAR Lag Selection for DBO & WTI in Sample (I)			
Lags	**AICC**	**SBC/BIC**	**HQ**
0	18284.3729	18294.7069	18288.2475
1	6309.6945	6340.6782	6321.2999
2	6210.402	**6262.0104***	6229.7132
3	6199.2441	6271.4524	6226.2364
4	6189.6248	6282.4079	6224.273
5	6176.4824	6289.815	6218.7612
6	6165.7243	6299.5813	**6215.6085***
7	6166.8301	6321.186	6224.2944
8	6161.7571	6336.5864	6226.7759
9	6158.9984	6354.2755	6231.5461
10	6161.3418	6377.0408	6241.3926
11	6164.8400	6400.935	6252.3679
12	6172.9089	6429.3741	6267.8881
13	6177.7452	6454.5543	6280.1494
14	6176.4321	6473.5588	6286.2351
15	6154.0724	6471.4904	6271.2478
16	**6147.4574***	6485.1402	6271.9787
17	6151.1777	6509.0987	6283.0183
18	6154.3691	6532.5015	6293.5022
19	6162.4550	6560.7719	6308.8537
20	6149.4258	6567.9002	6303.0631

Appendix A - VAR Lag Selection (continued)

VAR Lag Selection for USO & WTI in Sample (II)

Lags	AICC	SBC/BIC	HQ
0	9880.2232	9889.2107	9883.7003
1	4416.4662	4443.3923	4426.8610
2	4312.7377	4357.5535	4330.0014
3	4274.2230	**4336.8790***	4298.3062
4	4272.5343	4352.9808	4303.3871
5	4240.8060	4338.9928	4278.3782
6	4226.1270	4342.0032	**4270.3677***
7	4230.5847	4364.0991	4281.4429
8	4223.2577	4374.3586	4280.6816
9	4221.4548	4390.0901	4285.3922
10	4223.7420	4409.8589	4294.1402
11	4225.3326	4428.8781	4302.1386
12	4229.6424	4450.5628	4312.8024
13	4232.3922	4470.6333	4321.8521
14	4225.4363	4480.9434	4321.1414
15	4221.6437	4494.3617	4323.5389
16	**4217.7389***	4507.6120	4325.7684
17	4219.3804	4526.3524	4333.4880
18	4220.6212	4544.6353	4340.7500
19	4226.5406	4567.5395	4352.6334
20	4226.5221	4584.4479	4358.5209

VAR Lag Selection for DBO & WTI in Sample (II)

Lags	AICC	SBC/BIC	HQ
0	6174.6334	6182.9768	6177.9048
1	2839.9758	2864.9557	2849.7399
2	2780.6324	2822.1810	2796.8213
3	2763.2943	**2821.3429***	**2785.8392***
4	2764.6983	2839.1775	2793.5300
5	2760.8932	2851.7325	2795.9411
6	**2753.2998***	2860.4280	2794.4927
7	2759.2848	2882.6295	2806.5503
8	2761.0918	2900.5800	2814.3570
9	2762.7841	2918.3416	2821.9746
10	2768.8236	2940.3754	2833.8645
11	2773.8882	2961.3582	2844.7034
12	2781.2232	2984.5344	2857.7357
13	2785.4283	3004.5026	2867.5601
14	2789.4252	3024.1836	2877.0971
15	2773.0263	3023.3885	2866.1582
16	2770.8086	3036.6934	2869.3192
17	2779.2568	3060.5819	2883.0638
18	2787.5123	3084.1942	2896.5322
19	2792.2567	3104.2109	2906.4050
20	2784.9272	3112.0680	2904.1182

Appendix A - VAR Lag Selection (continued)

VAR Lag Selection for USO & WTI in Sample (III)			
Lags	**AICC**	**SBC/BIC**	**HQ**
0	10302.7743	10312.1284	10306.3637
1	3314.2430	3342.2752	3324.9810
2	**3208.4910***	**3255.1605***	**3226.3369***
3	3213.1006	3278.3666	3238.0135
4	3214.8650	3298.6861	3246.8036
5	3218.6609	3320.9955	3257.5835
6	3222.8315	3343.6377	3268.6963
7	3228.6163	3367.8519	3281.3810
8	3234.3097	3391.9322	3293.9318
9	3239.6782	3415.6446	3306.1148
10	3246.1854	3440.4524	3319.3932
11	3246.7488	3459.2730	3326.6843
12	3254.9648	3485.7023	3341.5842
13	3258.1107	3507.0172	3351.3696
14	3257.6903	3524.7212	3357.5442
15	3258.9544	3544.0648	3365.3583
16	3265.1653	3568.3099	3378.0740
17	3271.1704	3592.3037	3390.5383
18	3277.1683	3616.2441	3402.9492
19	3282.1840	3639.1560	3414.3317
20	3280.8235	3655.6450	3419.2912

VAR Lag Selection for DBO & WTI in Sample (III)			
Lags	**AICC**	**SBC/BIC**	**HQ**
0	9278.6377	9287.9919	9282.2272
1	2825.3257	2853.3579	2836.0637
2	**2793.2041***	**2839.8737***	**2811.0501***
3	2800.4712	2865.7371	2825.3841
4	2798.6549	2882.4760	2830.5935
5	2803.7560	2906.0906	2842.6786
6	2805.3503	2926.1565	2851.2151
7	2811.1531	2950.3888	2863.9179
8	2814.8796	2972.5021	2874.5017
9	2822.0994	2998.0657	2888.5359
10	2821.8039	3016.0710	2895.0117
11	2822.4590	3034.9832	2902.3945
12	2830.3664	3061.1039	2916.9858
13	2837.5941	3086.5006	2930.8530
14	2834.7922	3101.8232	2934.6461
15	2842.4534	3127.5638	2948.8573
16	2844.4162	3147.5609	2957.3249
17	2851.4374	3172.5706	2970.8053
18	2856.2318	3195.3076	2982.0128
19	2857.2909	3214.2629	2989.4386
20	2849.2285	3224.0500	2987.6963

Appendix B - VMA Coefficient for Impulse Response Function

(1) Level Data

Sample (I) for USO & WTI

Lags=16

Responses to Shock in USO		
Entry	USO	WTI
1	1.17852	1.67814
2	1.09858	1.68065
3	1.06940	1.56684
4	1.01507	1.50065
5	1.05824	1.53377
6	0.99414	1.41346
7	0.96779	1.36998
8	1.01079	1.39741
9	1.03314	1.44369
10	1.03174	1.36358
11	1.08148	1.49773
12	1.11880	1.60403
13	1.18077	1.62702
14	1.13134	1.61661
15	1.17910	1.58748
16	1.22263	1.63530
17	1.21804	1.63903
18	1.22983	1.64547
19	1.21369	1.62763
20	1.24665	1.67564
21	1.23906	1.66938
22	1.23645	1.65180
23	1.26302	1.69205
24	1.27318	1.70132
25	1.27713	1.69532
26	1.27128	1.69268
27	1.29070	1.70673
28	1.29070	1.70957
29	1.29874	1.71582
30	1.31137	**1.72853**
31	1.30872	1.72433
32	1.31151	1.71815
33	1.30776	1.71290
34	1.30796	1.70724
35	1.31149	1.70833
36	1.31026	1.70359

Responses to Shock in WTI		
Entry	USO	WTI
1	0.00000	0.83542
2	0.03477	0.65180
3	0.01742	0.60512
4	0.07895	0.74671
5	0.15720	0.73837
6	0.04034	0.59286
7	0.08856	0.64069
8	0.04318	0.55369
9	-0.05465	0.42836
10	-0.06755	0.42525
11	-0.20058	0.18088
12	-0.21568	0.13119
13	-0.23218	0.13117
14	-0.27618	0.05155
15	**-0.41025**	-0.13818
16	-0.40010	-0.14822
17	-0.36377	-0.07644
18	-0.37146	-0.06586
19	-0.34692	-0.04827
20	-0.34312	-0.01759
21	-0.34261	-0.02155
22	-0.34320	-0.02454
23	-0.36088	-0.03341
24	-0.35857	-0.03502
25	-0.36387	-0.03980
26	-0.36559	-0.04210
27	-0.37194	-0.04563
28	-0.37241	-0.05143
29	-0.37103	-0.03868
30	-0.37912	-0.04581
31	-0.38009	-0.05450
32	-0.38377	-0.05585
33	-0.38361	-0.05960
34	-0.38789	-0.06723
35	-0.39057	-0.06941
36	-0.38901	-0.06770

Sample (II) for USO & WTI

Lags=16

Responses to Shock in USO		
Entry	USO	WTI
1	1.59111	2.02185
2	1.53758	2.03696
3	1.55756	2.07005
4	1.50901	2.00396
5	1.60017	2.03743
6	1.51150	1.88050
7	1.44185	1.73430
8	1.47784	1.73227
9	1.49245	1.80688
10	1.49400	1.73515
11	1.41560	1.79740
12	1.29990	1.65793
13	1.44555	1.79420
14	1.33480	1.76413
15	1.52084	1.83532
16	1.71094	2.09199
17	1.74453	2.15513
18	1.76453	2.16598
19	1.68272	2.06483
20	1.76062	**2.18389**
21	1.71398	2.12120
22	1.68996	2.05921
23	1.72215	2.12803
24	1.71269	2.07164
25	1.71627	2.08349
26	1.64467	2.01724
27	1.65334	2.00397
28	1.65468	2.02125
29	1.66268	2.02697
30	1.71235	2.09880
31	1.71953	2.09685
32	1.75896	2.14124
33	1.73568	2.12707
34	1.72698	2.09381
35	1.73227	2.10364
36	1.72267	2.08872

Responses to Shock in WTI		
Entry	USO	WTI
1	0.00000	0.84872
2	0.05982	0.49129
3	0.13914	0.52150
4	0.27630	0.82504
5	**0.44242**	0.81370
6	0.13595	0.45872
7	0.28950	0.57112
8	0.26512	0.46857
9	0.02969	0.22912
10	0.05163	0.35692
11	-0.02769	0.09632
12	-0.02285	0.15186
13	-0.02241	0.23496
14	-0.11190	0.09447
15	-0.16884	-0.03460
16	-0.08437	0.00504
17	-0.01619	0.22000
18	-0.06808	0.14026
19	0.03279	0.22456
20	0.08525	0.35283
21	0.08268	0.32561
22	0.07295	0.33351
23	0.05840	0.32086
24	0.07789	0.26016
25	0.04905	0.24930
26	0.04829	0.29015
27	0.02435	0.20185
28	0.01227	0.18213
29	0.03828	0.25932
30	-0.00120	0.21323
31	-0.00286	0.19269
32	0.03156	0.23082
33	0.04022	0.23719
34	0.04214	0.24633
35	0.04932	0.25235
36	0.06430	0.25687

Sample (III) for USO & WTI

Lags=16

Responses to Shock in USO		
Entry	USO	WTI
1	0.75379	1.57508
2	0.73973	**1.68823**
3	0.72475	1.59757
4	0.66886	1.49444
5	0.65384	1.43973
6	0.62625	1.33140
7	0.65617	1.38231
8	0.61717	1.30741
9	0.62679	1.31162
10	0.59621	1.23468
11	0.55599	1.11217
12	0.49883	0.98750
13	0.51794	1.03372
14	0.56654	1.11087
15	0.53711	0.99461
16	0.50509	0.96906
17	0.48251	0.91430
18	0.47029	0.90255
19	0.46859	0.90050
20	0.47045	0.90957
21	0.46170	0.89933
22	0.46315	0.91123
23	0.43896	0.85639
24	0.43281	0.83839
25	0.42587	0.83565
26	0.42670	0.84729
27	0.41867	0.82780
28	0.40723	0.80024
29	0.39086	0.77372
30	0.38028	0.75414
31	0.36680	0.72312
32	0.35876	0.70711
33	0.34992	0.68778
34	0.34218	0.67277
35	0.33092	0.65163
36	0.31795	0.62400

Responses to Shock in WTI		
Entry	USO	WTI
1	0.00000	0.56815
2	-0.04173	0.31995
3	-0.02232	0.33586
4	0.00071	0.39629
5	-0.00345	0.34042
6	-0.04139	0.24216
7	0.00450	0.33680
8	-0.04756	0.23249
9	0.01500	0.35113
10	-0.01868	0.34203
11	-0.07640	0.20533
12	**-0.10291**	0.11912
13	-0.08051	0.14231
14	-0.10165	0.10349
15	-0.10244	0.09762
16	-0.10129	0.07055
17	-0.08457	0.12885
18	-0.08123	0.13837
19	-0.07931	0.14617
20	-0.07904	0.15738
21	-0.07435	0.16310
22	-0.06239	0.18659
23	-0.05586	0.19349
24	-0.06349	0.18867
25	-0.05585	0.19982
26	-0.04403	0.22600
27	-0.04173	0.23117
28	-0.03971	0.23831
29	-0.03788	0.23881
30	-0.03388	0.24480
31	-0.03311	0.24427
32	-0.03074	0.24403
33	-0.02934	0.24735
34	-0.02620	0.25311
35	-0.02394	0.25491
36	-0.02112	0.25851

Sample (I) for DBO & WTI

Lags=16

Responses to Shock in DBO				Responses to Shock in WTI		
Entry	DBO	WTI		Entry	DBO	WTI
1	0.61348	1.79453		1	0.00000	0.97356
2	0.56512	1.72217		2	0.02279	0.74320
3	0.54258	1.62126		3	-0.00377	0.61730
4	0.52175	1.56094		4	0.03331	0.78275
5	0.59043	1.69203		5	0.03406	0.71237
6	0.57820	1.60309		6	0.01565	0.58121
7	0.58637	1.61887		7	0.00927	0.56253
8	0.58218	1.57738		8	-0.01551	0.44174
9	0.56945	1.53780		9	-0.06003	0.29004
10	0.57841	1.56839		10	-0.08288	0.34153
11	0.60393	1.65987		11	-0.07851	0.29809
12	0.57899	1.62570		12	-0.10415	0.27068
13	0.61549	1.69261		13	-0.07905	0.36641
14	0.63456	1.84743		14	-0.07103	0.31363
15	0.65368	1.83281		15	-0.10250	0.16044
16	0.67167	1.87461		**16**	**-0.14074**	0.20325
17	0.67483	**1.89126**		17	-0.11008	0.21180
18	0.67170	1.87751		18	-0.12307	0.16523
19	0.66891	1.85902		19	-0.12208	0.21574
20	0.67844	1.88278		20	-0.12135	0.20115
21	0.67510	1.86948		21	-0.12558	0.18820
22	0.67044	1.85063		22	-0.12198	0.21343
23	0.67279	1.86660		23	-0.11706	0.20278
24	0.67326	1.86185		24	-0.12736	0.16283
25	0.67060	1.84568		25	-0.13367	0.17494
26	0.67268	1.86036		26	-0.12803	0.15873
27	0.67091	1.85412		27	-0.13787	0.14550
28	0.67102	1.84043		28	-0.13546	0.17076
29	0.66798	1.85372		29	-0.12926	0.16177
30	0.67361	1.85506		30	-0.13474	0.14434
31	0.66993	1.84088		31	-0.14055	0.15769
32	0.66898	1.84129		32	-0.13201	0.15879
33	0.66743	1.83522		33	-0.13516	0.14779
34	0.66527	1.82375		34	-0.13573	0.15987
35	0.66392	1.82015		35	-0.13450	0.15156
36	0.66244	1.81378		36	-0.13648	0.14962

Sample (II) for DBO & WTI

Lags=16

Responses to Shock in DBO		
Entry	DBO	WTI
1	0.82352	2.15962
2	0.68372	1.99993
3	0.69161	1.86852
4	0.69015	1.99065
5	0.82667	2.29667
6	0.76086	1.98139
7	0.73611	1.93797
8	0.82879	2.16519
9	0.88510	2.34663
10	0.85182	2.22878
11	0.93125	**2.71041**
12	0.96147	2.69618
13	1.00345	2.68726
14	0.86059	2.46829
15	0.96932	2.65369
16	0.92980	2.39751
17	0.92728	2.43727
18	0.88952	2.34784
19	0.91219	2.42912
20	0.90935	2.42578
21	0.91517	2.47867
22	0.89560	2.46263
23	0.90991	2.50528
24	0.88667	2.36296
25	0.87642	2.32178
26	0.83680	2.27451
27	0.87487	2.29358
28	0.82167	2.13841
29	0.82952	2.23306
30	0.82842	2.24374
31	0.83267	2.23487
32	0.81247	2.19888
33	0.81489	2.21813
34	0.79784	2.15733
35	0.78958	2.11482
36	0.77251	2.04555

Responses to Shock in WTI		
Entry	DBO	WTI
1	0.00000	1.20807
2	0.14950	0.97509
3	0.15054	1.00853
4	0.19800	1.27229
5	0.23658	1.27809
6	0.22097	1.08794
7	0.27157	1.19352
8	0.27147	1.15565
9	0.22550	1.00961
10	0.18885	1.01240
11	0.29270	1.20685
12	0.29064	1.39812
13	0.30138	1.43785
14	0.31866	1.34433
15	0.25331	1.03642
16	0.25429	1.23044
17	0.28193	1.14062
18	0.26027	1.03071
19	0.26576	1.09392
20	0.27936	1.14890
21	0.28095	1.19407
22	0.30720	1.26466
23	0.33692	1.33195
24	0.31937	1.30557
25	0.31326	1.26768
26	0.32625	1.18943
27	0.31600	1.22176
28	0.33511	1.27262
29	0.34539	1.27265
30	0.34195	1.25668
31	0.35971	1.36401
32	0.37559	1.39742
33	0.37901	1.39889
34	0.38636	1.40326
35	**0.39146**	1.39608
36	0.38319	1.38258

Sample (III) for DBO & WTI

Lags=16

Responses to Shock in DBO		
Entry	DBO	WTI
1	0.46194	1.45400
2	0.48690	**1.62994**
3	0.45972	1.51223
4	0.42055	1.35545
5	0.45501	1.42264
6	0.41715	1.33078
7	0.41324	1.25917
8	0.39338	1.17748
9	0.38518	1.19056
10	0.36949	1.10333
11	0.37454	1.04665
12	0.35013	0.98332
13	0.32322	0.92167
14	0.33206	0.94433
15	0.31281	0.86576
16	0.28347	0.82626
17	0.27127	0.77681
18	0.27069	0.79798
19	0.25589	0.73635
20	0.25121	0.72266
21	0.24773	0.70281
22	0.23775	0.68156
23	0.23257	0.64939
24	0.23078	0.64104
25	0.22069	0.61692
26	0.20870	0.58247
27	0.20656	0.56785
28	0.19915	0.54210
29	0.18857	0.50636
30	0.18408	0.48851
31	0.17900	0.46338
32	0.17274	0.44300
33	0.16679	0.41995
34	0.16205	0.40535
35	0.15486	0.38244
36	0.14966	0.36697

Responses to Shock in WTI		
Entry	DBO	WTI
1	0.00000	0.71826
2	-0.02857	0.48761
3	-0.02302	0.62554
4	0.00166	0.66974
5	0.00844	0.70985
6	0.04504	0.72627
7	0.04506	0.69850
8	0.05357	0.62143
9	0.07243	0.69424
10	**0.08232**	0.78150
11	0.05495	0.64971
12	0.06860	0.71679
13	0.04860	0.69045
14	0.02291	0.54609
15	0.04660	0.61927
16	0.04239	0.55136
17	0.04920	0.57739
18	0.05485	0.58079
19	0.06218	0.61989
20	0.05937	0.60444
21	0.06577	0.66481
22	0.06151	0.65302
23	0.06133	0.65920
24	0.06542	0.67753
25	0.06664	0.66996
26	0.07202	0.67792
27	0.07661	0.69320
28	0.08121	0.69535
29	0.08611	0.71348
30	0.08649	0.71892
31	0.08698	0.72067
32	0.08827	0.72470
33	0.08861	0.72963
34	0.08781	0.71940
35	0.09075	0.72668
36	0.09166	0.72522

(2) Difference Data

Sample (I) for DUSO & DWTI

Lags=16

Responses to Shock in DUSO			Responses to Shock in DWTI		
Entry	DUSO	DWTI	Entry	DUSO	DWTI
1	1.20258	1.73209	1	0.00000	0.85905
2	-0.07925	-0.04062	2	0.02690	-0.23212
3	-0.03611	-0.07459	3	-0.02956	-0.00959
4	-0.00254	-0.00547	4	0.13363	0.26713
5	0.10315	0.12141	**5**	0.09377	0.06384
6	-0.03769	-0.06497	6	-0.12703	-0.13829
7	0.00269	-0.02965	7	0.07286	0.09205
8	-0.01288	-0.01123	8	-0.06320	-0.14537
9	-0.01136	-0.03919	9	-0.05306	-0.03971
10	-0.04041	-0.10885	10	0.03100	0.06274
11	-0.02047	0.04405	11	-0.05128	-0.16763
12	-0.01078	-0.01692	12	0.00969	0.01370
13	0.05810	0.05090	13	0.02234	0.03671
14	0.05237	0.11908	14	-0.05111	-0.07694
15	0.05737	-0.01258	15	-0.02310	-0.02149
16	0.11386	0.14764	16	-0.00171	-0.02705
17	0.02325	0.06562	17	0.04465	0.08910
18	-0.00673	-0.02591	18	-0.00473	-0.00713
19	-0.02267	-0.03966	19	0.01947	0.00598
20	0.03872	0.06140	20	0.00848	0.02905
21	-0.00890	-0.02463	21	-0.00346	-0.00885
22	-0.00845	-0.01694	22	-0.00457	-0.00207
23	0.00436	0.01116	23	-0.00705	0.00312
24	-0.00954	-0.01939	24	-0.00274	-0.01543
25	-0.01109	-0.01001	25	0.00088	0.00305
26	-0.00035	0.01178	26	0.00272	0.00962
27	0.00472	0.00325	27	-0.00797	-0.01789
28	0.01989	0.02269	28	-0.00562	-0.00557
29	0.01953	0.02620	29	0.00067	0.00622
30	0.01598	0.01743	30	-0.00527	-0.00714
31	0.00755	0.00679	31	0.00130	-0.00112
32	0.00481	0.00670	32	0.00826	0.01137
33	-0.00011	0.00243	33	0.00146	0.00041
34	-0.00297	-0.00820	34	0.00193	0.00207
35	0.00430	0.00448	35	-0.00075	-0.00050
36	0.00008	-0.00018	36	-0.00105	-0.00153

Sample (II) for DUSO & DWTI

Lags=16

Responses to Shock in DUSO		
Entry	DUSO	DWTI
1	1.54621	2.03087
2	-0.05766	-0.02369
3	0.08073	0.00197
4	-0.05034	0.02423
5	0.12052	0.02646
6	-0.15278	-0.20133
7	0.04266	0.06104
8	0.20762	0.24708
9	-0.09668	-0.07927
10	0.09299	0.09285
11	0.01718	0.09192
12	0.03669	0.01423
13	0.21629	0.25971
14	-0.00004	0.10167
15	0.23546	0.15426
16	0.11827	0.13606
17	0.04554	0.08786
18	-0.06395	-0.12092
19	-0.00476	-0.00077
20	0.08752	0.13260
21	-0.04489	-0.03496
22	0.09089	0.10306
23	0.03333	0.07532
24	0.01859	0.00537
25	0.04431	0.04359
26	0.02938	0.06210
27	0.06230	0.04604
28	0.01090	-0.00807
29	0.06646	0.09638
30	0.01962	0.02032
31	0.00637	0.00657
32	0.02258	0.04597
33	-0.00162	0.00258
34	0.03172	0.03430
35	0.01502	0.02772
36	0.02019	0.01633

Responses to Shock in DWTI		
Entry	DUSO	DWTI
1	0.00000	0.82174
2	0.01265	-0.39184
3	0.11124	0.10111
4	0.13052	0.28685
5	0.04595	-0.15262
6	-0.30498	-0.40939
7	0.16147	0.19124
8	-0.06021	-0.15623
9	0.04628	0.11645
10	-0.06783	-0.02799
11	-0.00199	-0.08792
12	-0.00412	0.06129
13	0.12755	0.23522
14	-0.04275	0.01838
15	-0.01931	-0.20408
16	0.08947	0.12758
17	0.00729	-0.02766
18	-0.07761	-0.09785
19	-0.01208	-0.01730
20	0.02241	0.04620
21	-0.02616	-0.00164
22	0.03725	0.08128
23	-0.02686	-0.03604
24	0.02050	0.00040
25	0.02205	0.05308
26	0.01881	0.02939
27	-0.03315	-0.06566
28	0.02722	-0.02478
29	0.00376	0.03810
30	-0.02054	-0.03426
31	-0.00917	0.00354
32	0.00065	0.00961
33	0.01415	0.02757
34	0.00202	0.01035
35	0.00721	0.01194
36	-0.01018	-0.03376

Sample (III) for DUSO & DWTI

Lags=16

Responses to Shock in DUSO		
Entry	DUSO	DWTI
1	0.73541	1.52989
2	-0.00538	0.10332
3	-0.02100	-0.02658
4	-0.03435	-0.08399
5	0.08585	0.14798
6	-0.00120	0.00137
7	0.00625	0.01760
8	-0.02621	-0.08526
9	0.03471	0.06611
10	-0.01518	-0.02931
11	0.00270	-0.02553
12	-0.00510	0.03001
13	-0.02895	-0.07085
14	0.06700	0.02388
15	0.01759	0.03584
16	0.02538	0.06881
17	-0.02919	-0.06717
18	0.01062	0.01890
19	-0.00713	-0.02480
20	0.00377	0.01497
21	-0.00310	-0.00438
22	0.00859	0.01578
23	-0.00619	-0.01735
24	0.00952	0.00833
25	-0.00084	0.00784
26	-0.01394	-0.01822
27	-0.00598	-0.03037
28	0.00938	0.01325
29	0.01707	0.03902
30	-0.00350	-0.00672
31	0.00077	0.00347
32	-0.00179	-0.00632
33	0.00220	0.00467
34	-0.00012	0.00086
35	0.00033	0.00100
36	-0.00323	-0.00659

Responses to Shock in DWTI		
Entry	DUSO	DWTI
1	0.00000	0.57691
2	-0.02839	-0.20417
3	0.03126	0.09444
4	0.07255	0.12186
5	-0.01695	-0.05273
6	0.02999	0.09723
7	0.01599	-0.00153
8	-0.03115	-0.05841
9	0.00170	-0.01981
10	0.00642	0.06948
11	-0.01545	-0.04177
12	0.01939	0.03768
13	0.04004	0.04761
14	0.02573	0.09059
15	-0.04766	-0.08604
16	-0.03976	-0.12138
17	0.00681	0.03090
18	0.00445	-0.01284
19	0.00347	0.02277
20	0.00089	-0.00652
21	-0.00564	-0.00842
22	0.00613	0.00882
23	0.00573	0.01373
24	-0.00815	-0.00679
25	-0.00560	-0.01900
26	0.00364	0.00166
27	0.01147	0.01557
28	-0.00174	0.01123
29	-0.00825	-0.01755
30	-0.00179	-0.00216
31	0.00054	0.00027
32	0.00104	-0.00039
33	0.00117	0.00610
34	-0.00051	-0.00263
35	0.00027	0.00045
36	0.00178	0.00000

Sample (I) for DDBO & DWTI

Lags=16

Responses to Shock in DDBO		
Entry	DDBO	DWTI
1	0.61472	1.73847
2	-0.00370	0.15495
3	0.01959	0.03404
4	0.00055	0.01304
5	0.04495	0.08257
6	-0.02166	-0.13137
7	0.00113	-0.02131
8	0.02423	0.00175
9	0.02694	0.06878
10	0.00641	0.02361
11	0.03723	0.15584
12	0.01206	0.05348
13	0.04291	0.17136
14	0.01210	0.13088
15	0.02208	-0.02957
16	0.04574	0.11117
17	-0.00425	-0.05948
18	0.00622	0.02895
19	0.00311	-0.00257
20	0.00972	0.03039
21	0.00336	0.00869
22	0.01022	0.04741
23	0.01335	0.03541
24	0.00712	0.01738
25	0.00817	0.01756
26	0.00497	0.01493
27	-0.00160	-0.01640
28	0.00447	0.01340
29	0.00041	0.02261
30	0.01115	0.02255
31	0.00383	0.01221
32	0.00471	0.00796
33	0.00280	0.01189
34	0.00359	0.00832
35	0.00136	0.00289
36	0.00092	-0.00026

Responses to Shock in DWTI		
Entry	DDBO	DWTI
1	0.00000	0.94750
2	0.04769	-0.17900
3	-0.01890	-0.04089
4	0.04924	0.19217
5	0.04216	0.05747
6	-0.00892	-0.07057
7	0.01229	0.00797
8	-0.00966	-0.09204
9	0.00345	0.00446
10	-0.00571	0.01908
11	-0.00905	-0.07078
12	0.02578	0.10127
13	0.01117	0.00176
14	-0.03182	-0.13567
15	-0.00413	-0.07850
16	-0.02971	0.01808
17	0.02256	0.06908
18	-0.01348	-0.05943
19	0.00478	0.04376
20	0.01143	0.04013
21	0.00509	0.01605
22	0.00208	0.01494
23	0.00139	-0.00726
24	0.00045	-0.01028
25	-0.00242	0.01024
26	-0.00405	-0.02656
27	0.00336	0.01054
28	0.00507	0.02024
29	-0.00238	-0.01627
30	-0.00032	-0.00571
31	-0.00359	-0.00668
32	0.00280	0.01616
33	-0.00122	-0.01126
34	-0.00115	-0.00005
35	0.00055	0.00131
36	0.00124	0.00334

Sample (II) for DDBO & DWTI

Lags=16

Responses to Shock in DDBO		
Entry	DDBO	DWTI
1	0.80650	2.22941
2	-0.03556	-0.02958
3	-0.07170	-0.21426
4	0.01824	0.02957
5	0.16669	0.49206
6	-0.03252	-0.25126
7	-0.00020	-0.04277
8	0.04266	0.18097
9	-0.01867	-0.15984
10	-0.04177	-0.16525
11	0.08233	0.45819
12	0.00040	0.05427
13	-0.02371	-0.11580
14	-0.01071	0.01445
15	0.06019	0.16010
16	0.00439	-0.14690
17	0.00544	0.03722
18	-0.01367	0.00473
19	-0.00446	-0.07442
20	0.01071	0.04978
21	0.03096	0.13084
22	-0.01434	-0.04848
23	0.01680	0.06834
24	0.01303	0.05281
25	-0.02286	-0.13097
26	-0.01752	-0.06498
27	0.02331	0.08984
28	-0.01286	-0.05757
29	-0.00963	-0.03332
30	0.02342	0.11112
31	0.00901	0.02747
32	-0.02103	-0.07625
33	0.01177	0.06687
34	0.01163	0.03112
35	-0.01331	-0.07026
36	-0.00422	-0.01735

Responses to Shock in DWTI		
Entry	DDBO	DWTI
1	0.00000	1.24034
2	0.06958	-0.38736
3	0.03683	0.07584
4	0.07256	0.46021
5	-0.00352	-0.25573
6	-0.09410	-0.25904
7	0.11478	0.29001
8	-0.05346	-0.23161
9	-0.01350	-0.05555
10	0.00974	0.12240
11	0.02668	0.00065
12	0.02725	0.17220
13	0.04011	0.07520
14	-0.05593	-0.13124
15	-0.07800	-0.50768
16	-0.00119	0.11242
17	0.10959	0.37554
18	-0.07982	-0.40073
19	-0.02082	0.05959
20	0.06155	0.20563
21	-0.02662	-0.06827
22	0.00498	0.02758
23	0.02982	0.13810
24	-0.01457	-0.09339
25	-0.02223	-0.07685
26	0.00345	-0.03106
27	0.00382	0.03621
28	0.00553	0.00126
29	0.00785	0.07748
30	-0.00417	0.00717
31	-0.01446	-0.09006
32	0.02350	0.12352
33	0.01337	0.02228
34	-0.03227	-0.11498
35	0.00825	0.01155
36	0.00468	0.01447

Sample (III) for DDBO & DWTI

Lags=16

Responses to Shock in DDBO		
Entry	DDBO	DWTI
1	0.47450	1.51576
2	-0.00007	0.08985
3	-0.01325	-0.11028
4	-0.03411	-0.15072
5	0.02711	0.08677
6	-0.03416	-0.09788
7	0.01076	-0.02024
8	-0.00356	-0.04126
9	0.00674	0.05819
10	0.02234	0.06466
11	0.00664	-0.01603
12	-0.01925	-0.07995
13	0.01453	0.09295
14	0.01314	0.04084
15	-0.01280	-0.06479
16	-0.01497	-0.03755
17	-0.00276	-0.04330
18	0.00679	0.04390
19	0.00286	0.00949
20	-0.00297	-0.01054
21	0.00531	0.01105
22	0.00124	0.02438
23	0.00355	0.00967
24	-0.00454	-0.01684
25	-0.00025	0.00274
26	0.00197	0.00169
27	0.00392	0.01561
28	-0.00120	-0.00336
29	-0.00441	-0.01742
30	0.00309	0.01192
31	0.00061	0.00444
32	0.00017	-0.00001
33	-0.00086	-0.00401
34	-0.00010	0.00029
35	0.00022	-0.00151
36	0.00031	0.00157

Responses to Shock in DWTI		
Entry	DDBO	DWTI
1	0.00000	0.69560
2	-0.02073	-0.20451
3	0.04175	0.16457
4	0.02607	0.10190
5	-0.01066	-0.03343
6	-0.00245	-0.03561
7	0.01282	-0.01919
8	-0.02172	-0.13476
9	0.00594	-0.00190
10	-0.03070	-0.03056
11	-0.01676	-0.13338
12	-0.00202	0.00017
13	0.00233	-0.04369
14	-0.00847	-0.08885
15	0.01703	0.06690
16	-0.01585	-0.10104
17	0.03265	0.08000
18	-0.00808	0.00520
19	0.00290	0.02469
20	-0.00153	0.00727
21	0.00997	0.05792
22	-0.00499	-0.00532
23	0.00782	0.02761
24	-0.00311	0.01916
25	0.00355	-0.00094
26	0.00133	0.01918
27	0.00021	0.01205
28	-0.00447	-0.01172
29	0.00137	0.01339
30	-0.00529	-0.01413
31	0.00198	-0.00579
32	-0.00468	-0.01134
33	0.00071	-0.00900
34	-0.00201	-0.01651
35	0.00202	0.00163
36	-0.00304	-0.01384

Appendix C - Coefficients for Dynamic Model (6.17) & (6.18)

$$USO_t = \beta_{10} + \beta_{11}WTI_t + \beta_{12}USO_{t-1} + \beta_{13}WTI_{t-1} + u_{1t} \tag{6.11}$$

	Sample (I)	**Sample (II)**	**Sample (III)**
	4/12/2006 - 3/31/2012	**4/12/2006 - 12/31/2008**	**1/2/2009 - 3/31/2012**
β_{10}	**-0.1485**	**-0.0536**	**0.0815**
t-stat	*(-2.43)*	*(-0.52)*	*(0.72)*
β_{11}	**0.5482**	**0.6378**	**0.4159**
t-stat	*(72.84)*	*(52.58)*	*(67.14)*
β_{12}	**0.9993**	**0.9823**	**0.9909**
t-stat	*(1096.74)*	*(157.43)*	*(217.52)*
β_{13}	**-0.5463**	**-0.6229**	**-0.4131**
t-stat	*(-72.3)*	*(-47.83)*	*(-65.39)*

$$DBO_t = \beta_{20} + \beta_{21}WTI_t + \beta_{22}DBO_{t-1} + \beta_{23}WTI_{t-1} + u_{2t} \tag{6.12}$$

	Sample (I)	**Sample (II)**	**Sample (III)**
	1/5/2007 - 3/31/2012	**1/5/2007 - -12/31/2008**	**1/2/2009 - -3/31/2012**
β_{20}	**-0.0174**	**0.1983**	**0.1176**
t-stat	*(-0.46)*	*(2.28)*	*(1.46)*
β_{21}	**0.2603**	**0.2663**	**0.2501**
t-stat	*(61.90)*	*(35.96)*	*(56.58)*
β_{22}	**0.9949**	**0.9381**	**0.9866**
t-stat	*(360.40)*	*(67.70)*	*(157.94)*
β_{23}	**-0.2584**	**-0.2450**	**-0.2473**
t-stat	*(-60.12)*	*(-27.70)*	*(-53.24)*

Appendix D - ACE Transformation Graph: Figure_6.6 & 6.7

Figure_6.6 ACE transformation of series USO and WTI over three samples

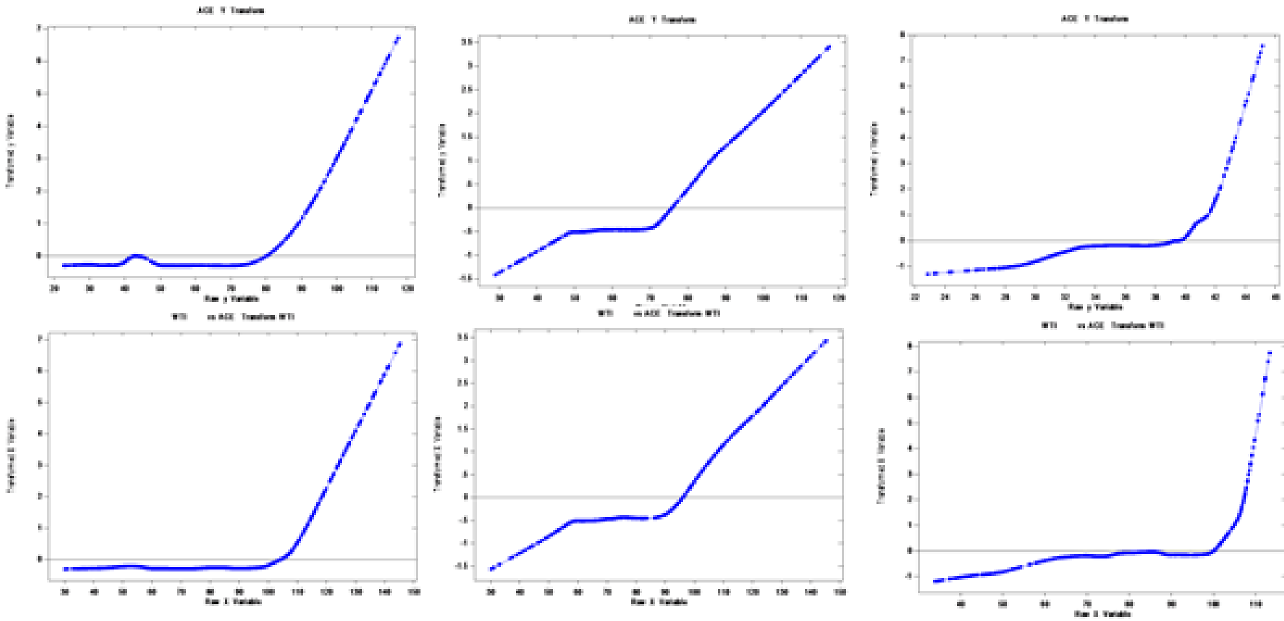
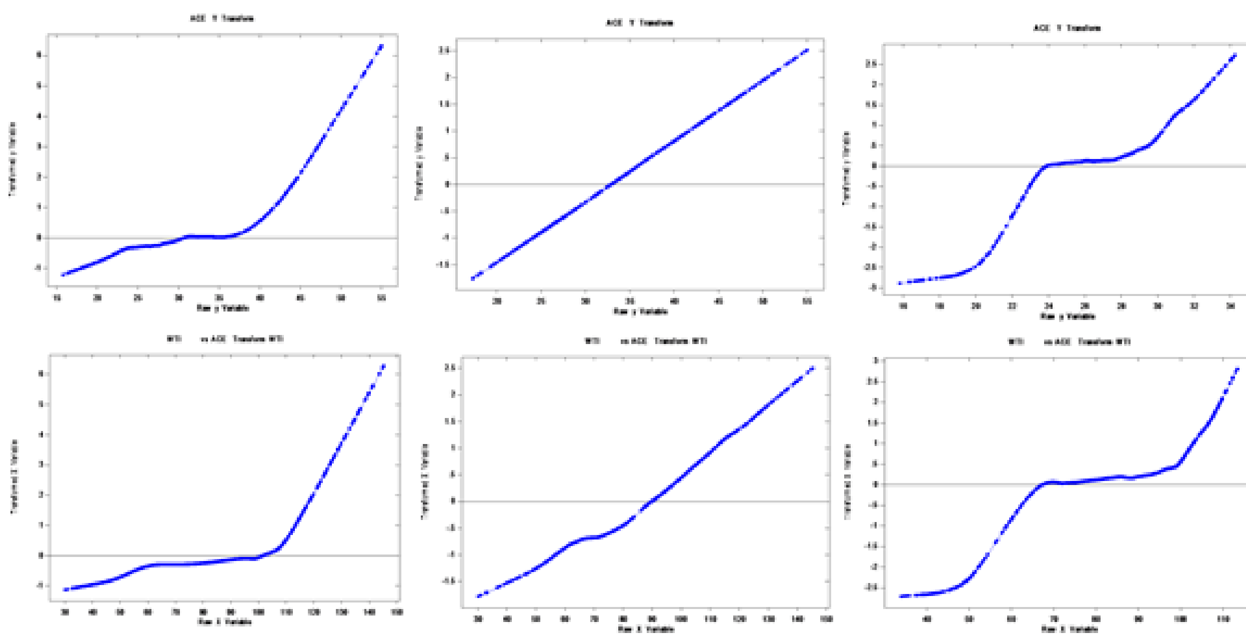


Figure_6.7 ACE transformation of series DBO and WTI over three samples



Appendix E - Knots and Slops for MARS (mi=2) model

Sample (1)

Left-hand side variable: ΔUSO_t, ΔWTI_t

(1)			
4/12/2006 - 3/31/2012			
ΔUSO_t **(left-hand variable)**	**Slope**	ΔWTI_t **(left-hand variable)**	**Slope**
$\Delta USO\{1\} < -3.58$	2.137	$\Delta USO\{1\} > 3.13$	4.949
	(3.98)		(5.99)
$\Delta USO\{1\} < -3.58$	-0.297	$\Delta USO\{1\} < 3.13$	-2.429
$\Delta WTI\{1\} > -14.76$	(-3.26)		(-7.04)
$\Delta USO\{1\} > -0.98$	-2.315	$\Delta WTI\{1\} < 4.53$	0.275
$\Delta WTI\{1\} < -2.10$	(-4.92)		(3.14)
$\Delta WTI\{1\} > -2.10$	0.756	$\Delta USO\{1\} > -3.58$	-0.234
$\hat{q}_t\{1\} > 7.25$	(4.93)	$\Delta WTI\{1\} < 4.53$	(-5.46)
$\Delta WTI\{1\} > -2.10$	-0.608	$\Delta USO\{1\} < -3.58$	0.191
$\hat{q}_t\{1\} > 6.61$	(-4.96)	$\Delta WTI\{1\} < 4.53$	(6.29)
$\Delta WTI\{1\} > -2.10$	-0.150	$\Delta USO\{1\} > -3.58$	-0.283
$\hat{q}_t\{1\} > 10.17$	(-4.04)	$\Delta WTI\{1\} > -2.95$	(-7.11)
		$\Delta USO\{1\} < 3.13$	0.117
		$\Delta WTI\{1\} > -0.20$	(2.62)
		$\Delta WTI\{1\} > 4.53$	0.237
		$\hat{q}_t\{1\} > 0.026$	(9.12)
		$\Delta WTI\{1\} > 4.53$	0.207
		$\hat{q}_t\{1\} < 0.026$	(4.24)
MARS model estimated with 20 knots, 1 lag with interaction term (mi=2)			

Sample (2)

Left-hand side variable: ΔUSO_t, ΔWTI_t

(2) 4/12/2006 - 12/31/2008			
ΔUSO_t (left-hand variable)	**Slope**	**ΔWTI_t (left-hand variable)**	**Slope**
$\Delta USO\{1\} > -3.58$	0.988	$\Delta USO\{1\} > 3.34$	-9.894
	(2.89)		(-3.55)
$\Delta USO\{1\} > -2.45$	-1.127	$\Delta WTI\{1\} > 4.42$	9.225
	(3.10)		(5.43)
$\Delta WTI\{1\} < -4.57$	0.691	$\Delta WTI\{1\} > 3.41$	-3.023
	(3.75)		(-4.14)
$\Delta USO\{1\} > -3.58$	-0.349	$\Delta USO\{1\} > 3.25$	-5.467
$\Delta WTI\{1\} < -2.06$	(-4.57)	$\Delta WTI\{1\} > 3.41$	(-4.82)
		$\Delta USO\{1\} > 2.46$	4.119
		$\Delta WTI\{1\} > 3.41$	(3.36)
		$\Delta USO\{1\} > 3.34$	0.694
		$\hat{q}_t\{1\} > -10.93$	(2.78)
		$\Delta WTI\{1\} > 4.42$	-124.885
		$\hat{q}_t\{1\} > 0.59$	(-5.99)
		$\Delta WTI\{1\} > 4.42$	-0.594
		$\hat{q}_t\{1\} < 0.59$	(-9.07)
		$\Delta WTI\{1\} > 4.42$	24.309
		$\hat{q}_t\{1\} > 0.36$	(2.89)
MARS model estimated with 20 knots, 1 lag with interaction term (mi=2)			

Sample (3)

Left-hand side variable: ΔUSO_t, ΔWTI_t

(3) 1/2/2009 - 3/31/2012			
ΔUSO_t **(left-hand variable)**	**Slope**	ΔWTI_t **(left-hand variable)**	**Slope**
$\Delta USO\{1\} < 0.20$	1.601	$\Delta USO\{1\} > 0.93$	-4.369
	(3.73)		(-3.20)
$\Delta USO\{1\} > -1.52$	1.156	$\Delta USO\{1\} < 0.93$	3.492
	(4.24)		(3.06)
$\Delta WTI\{1\} > 3.34$	0.352	$\Delta USO\{1\} > 0.20$	2.322
	(2.62)		(3.74)
$\hat{q}_t\{1\} > 3.91$	-1.029	$\Delta USO\{1\} > -1.52$	3.395
	(-5.32)		(3.79)
$\hat{q}_t\{1\} < 3.91$	0.033	$\Delta WTI\{1\} > -8.90$	-1.058
	(2.46)		(-7.00)
$\Delta USO\{1\} < 0.23$	-0.132	$\Delta WTI\{1\} > 3.34$	1.716
$\Delta WTI\{1\} < 3.34$	(-3.54)		(4.85)
$\Delta USO\{1\} < 0.20$	0.215	$\hat{q}_t\{1\} > 3.91$	-1.854
$\Delta WTI\{1\} > -1.22$	(2.07)		(-4.41)
$\Delta WTI\{1\} > -0.59$	0.360	$\Delta USO\{1\} < 0.93$	1.289
$\hat{q}_t\{1\} > 3.91$	(3.78)	$\Delta WTI\{1\} > 1.81$	(2.94)
		$\Delta USO\{1\} < 0.93$	-0.404
		$\Delta WTI\{1\} < 1.81$	(-4.54)
		$\Delta USO\{1\} > -1.52$	0.456
		$\hat{q}_t\{1\} < -4.27$	(2.97)
MARS model estimated with 20 knots, 1 lag with interaction term (mi=2)			

Sample (1)

Left-hand side variable: ΔDBO_t, ΔWTI_t

(1)			
1/5/2007 - 3/31/2012			
ΔDBO_t **(left-hand variable)**	**Slope**	ΔWTI_t **(left-hand variable)**	**Slope**
$\hat{q}_t\{1\} > 6.68$	-1.903	$\Delta WTI\{1\} > 4.65$	-41.703
	(-6.88)		(-7.68)
$\hat{q}_t\{1\} > 5.29$	0.435	$\Delta WTI\{1\} > 4.15$	-11.621
	(2.83)		(-6.73)
$\Delta DBO\{1\} > 0.28$	1.697	$\hat{q}_t\{1\} > 6.68$	-2.160
$\hat{q}_t\{1\} > 5.29$	(6.04)		(-4.44)
$\Delta DBO\{1\} < 0.28$	0.850	$\Delta DBO\{1\} > 0.28$	2.0157
$\hat{q}_t\{1\} > 5.29$	(5.37)	$\hat{q}_t\{1\} > 5.29$	(5.31)
$\Delta DBO\{1\} > 0.54$	-0.202	$\Delta DBO\{1\} < 0.28$	0.494
$\hat{q}_t\{1\} < 0.27$	(-4.86)	$\hat{q}_t\{1\} > 5.29$	(2.37)
$\Delta DBO\{1\} < 0.54$	-0.189	$\Delta WTI\{1\} > 4.65$	5.185
$\hat{q}_t\{1\} > 3.50$	(-3.36)	$\hat{q}_t\{1\} > -3.78$	(8.45)
$\Delta WTI\{1\} > 0.03$	-0.439		
$\hat{q}_t\{1\} > 5.29$	(-3.75)		
$\Delta WTI\{1\} < 0.03$	-0.087		
$\hat{q}_t\{1\} > 5.29$	(-3.41)		
$\Delta WTI\{1\} > 2.91$	0.480		
$\hat{q}_t\{1\} > 5.29$	(2.57)		
MARS model estimated with 20 knots, 1 lag with interaction term (mi=2)			

Sample (2)

Left-hand side variable: ΔDBO_t, ΔWTI_t

(2) 4/12/2006 - 12/31/2008			
ΔDBO_t (left-hand variable)	**Slope**	**ΔWTI_t (left-hand variable)**	**Slope**
$\Delta DBO\{1\} < -1.65$	1.393	$\hat{q}_t\{1\} > -0.879$	-0.401
	(3.63)		(-2.40)
$\Delta DBO\{1\} > -1.25$	-0.165	$\Delta DBO\{1\} > -0.60$	-15.154
	(-3.10)	$\Delta WTI\{1\} < -3.06$	(-4.24)
$\Delta WTI\{1\} > -4.69$	1.617	$\Delta DBO\{1\} > 1.04$	0.499
	-2.89	$\Delta WTI\{1\} > 2.46$	(3.08)
$\Delta WTI\{1\} > -4.29$	-1.558	$\Delta DBO\{1\} > 0.17$	-5.160
	(-2.74)	$\hat{q}_t\{1\} < -0.879$	(-7.04)
$\hat{q}_t\{1\} > -0.388$	-0.364	$\Delta WTI\{1\} > 4.18$	6.751
	(-5.77)	$\hat{q}_t\{1\} > -0.879$	(3.44)
		$\Delta WTI\{1\} > 3.73$	-5.173
		$\hat{q}_t\{1\} > -0.879$	(-3.23)

MARS model estimated with 20 knots, 1 lag with interaction term (mi=2)

Sample (3)

Left-hand side variable: ΔDBO_t, ΔWTI_t

(3)			
1/2/2009 - 3/31/2012			
ΔDBO_t **(left-hand variable)**	**Slope**	ΔWTI_t **(left-hand variable)**	**Slope**
$\Delta DBO\{1\} < 0.550$	-0.416	$\Delta DBO\{1\} > -1.13$	-2.171
	(-4.06)		(-2.25)
$\Delta WTI\{1\} < 3.34$	0.140	$\Delta DBO\{1\} > -0.62$	5.219
	(4.71)		(3.88)
$\hat{q}_t\{1\} > 1.899$	-1.126	$\Delta WTI\{1\} > -2.64$	0.941
	(-3.17)		(3.95)
$\hat{q}_t\{1\} < 1.899$	0.066	$\hat{q}_t\{1\} > 1.899$	-4.431
	(3.62)		(-3.54)
$\Delta DBO\{1\} < 0.550$	0.057	$\hat{q}_t\{1\} < 1.899$	0.138
$\Delta WTI\{1\} > -2.64$	(2.02)		(2.38)
$\Delta WTI\{1\} > 3.34$	0.101	$\Delta DBO\{1\} > -1.13$	-1.825
$\hat{q}_t\{1\} < 1.86$	(2.91)	$\Delta WTI\{1\} > -1.46$	(-4.61)
$\Delta WTI\{1\} > 0.02$	0.069	$\Delta DBO\{1\} > -1.13$	2.595
$\hat{q}_t\{1\} > -0.725$	(4.11)	$\Delta WTI\{1\} < -1.46$	(5.29)
		$\Delta DBO\{1\} > -0.62$	1.640
		$\Delta WTI\{1\} > -0.32$	(3.98)
		$\Delta WTI\{1\} > 1.52$	0.596
		$\hat{q}_t\{1\} < 0.43$	(3.10)
MARS model estimated with 20 knots, 1 lag with interaction term (mi=2)			

Appendix F - Coefficients for GAM ECM: Table _ 7.7

Table _ 7.7 Coefficients for GAM ECM models for sample (I) (II) (III)

Sample	**(I)**		**(II)**		**(III)**	
	4/12/2006 - 3/31/2012		**4/12/2006 - 12/31/2008**		**1/2/2009 - 3/31/2012**	
	ΔUSO_t	ΔWTI_t	ΔUSO_t	ΔWTI_t	ΔUSO_t	ΔWTI_t
α	-0.0220	0.0330	-0.0581	-0.0334	0.0033	0.0859
	(-0.69)	(0.64)	(-0.94)	(-0.39)	(0.12)	(1.44)
q_{t-1}	-0.0010	-0.0017	-0.0173	-0.0017	-0.0513***	-0.0918***
	(-0.46)	(-0.48)	(-1.03)	(-0.15)	(-4.00)	(-3.26)
ΔUSO_{t-1}	-0.1066*	0.2745***	-0.1590*	0.2334**	0.0486	0.8820***
	(-1.94)	(3.13)	(-1.86)	(1.98)	(0.55)	(4.51)
ΔWTI_{t-1}	0.0391	-0.1885***	0.0686	-0.2241***	-0.0243	-0.3461***
	(1.15)	(-3.47)	(1.12)	(-2.66)	(-0.61)	(-3.92)
	1/5/2007 - 3/31/2012		**1/5/2007 - 12/31/2008**		**1/2/2009 - 3/31/2012**	
	ΔDBO_t	ΔWTI_t	ΔDBO_t	ΔWTI_t	ΔDBO_t	ΔWTI_t
α	0.0075	0.0461	-0.0122	-0.0255	0.0115	0.0723
	(0.46)	(0.81)	(-0.35)	(-0.23)	(0.69)	(1.21)
q_{t-1}	-0.0124*	-0.0246	-0.2199***	-0.1191	-0.0471***	-0.1657***
	(-1.67)	(-1.01)	(-4.10)	(-0.70)	(-3.02)	(-2.99)
ΔDBO_{t-1}	-0.1518***	0.2705	-0.2807***	-0.1651	0.1841**	1.0179***
	(-2.79)	(1.51)	(-3.37)	(0.80)	(2.36)	(3.67)
ΔWTI_{t-1}	0.0372**	-0.1087**	0.0645**	-0.0010	-0.0487**	-0.2370***
	(2.26)	(-2.01)	(2.35)	(-0.01)	(-2.24)	(-3.06)

$^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

Wendy Yu

Email: ywenpei@hotmail.com

SUMMARY

- Specialized in time series, statistical modeling, financial econometrics, regression analysis
- Strength in sophisticated exploratory data analysis and predictive modeling
- Experience with data mining solutions for business analytic problems (logistic regression, decision / regression trees, multivariate regression, generalized linear model(GLM))
- Advanced skills in statistical software: R, SAS, SQL, RATS, Stata, SPSS

WORK EXPERIENCE

Research Assistant 08/2009 - present

Departments of Information & Decision Sciences, Economics. UIC

- Studied dynamic relation between oil ETFs and underlying benchmark crude oil price
- Conducted linear and nonlinear cointegration test on non-stationary data series
- Developed nonlinear error correction model(ECM), and generalized additive models(GAM)
- Analyzed spectrum and impulse response function to study non-stationarity

Teaching Assistant 08/2009 - present

Department of Information & Decision Sciences. UIC

- Instructed lab sessions for the course of Business Statistics, and Operations Management
- Graded / evaluated assignments, quizzes and exams

Data Analytic Projects 01/2012 - 05/2012

Department of Information & Decision Sciences. UIC

- Analyzed major demographic segments and zip code data using k-means cluster analysis
- Created model to cut mailing quantities by 25% and retain 96% responses for insurance firm
- Designed competing models to select members for screening within budget of federal grant

EDUCATION

Ph.D., Business Statistics University of Illinois at Chicago 2015

Dissertation: An Analysis of Oil ETFs and Crude Oil Price

MBA, International Business University of Illinois at Chicago