Price dynamics under the presence of threshold effects in WTI crude oil markets

Eunyoung Kim*

< Abstract >

We analyzed long-run adjustment process to equilibrium and short-run dynamics with WTI spot and futures prices using bivariate 3-regime TVECM. After dividing the entire sample period into 5 sub-samples, we applied this model to each sub-sample. This allows us to figure out the differentiable effects of market conditions which can vary by sub-samples and regimes, on investors' behaviors. The estimation results showed quite interesting points. First, the middle regimes of all 5 sub-samples were not targeted by investors, so we could find their activities only in lower and/or upper regimes. Second, we can make 3 different groups with 5 sub-samples. Period 1 showed brisk movements in futures markets under the price information leadership of spot markets and relatively longer adjustment time compared to the other 4 periods. Period 2, 3, and 4 showed the opposite phenomena to period 1 probably caused by sporadic big shocks on world economy. Period 5 showed mixed results in lower and upper regimes. We could analyze quite differentiable and opposite adjustments according to regimes in this period even under the dominated contango condition, which could be regarded as the precious harvest of employing a bivariate 3-regime TVECM in this paper.

Keywords : 3-Regime Threshold Vector Error Correction Model, Structural Breaks, Nonlinearity, WTI Crude Oil Markets

* Part-time lecturer, Department of Statistics, Pusan National University
E-mail : keyivy@empal.com; Tel: +82-51-510-1784

Ⅰ. Introduction

WTI(West Texas Intermediate) crude oil spot and futures prices and their volatilities have been one of the most favorite research subjects in the industries and academia along with Brent and Dubai crude oils. Most of the previous studies on this subject used conventional methodologies including cointegration analysis, VAR(Vector AutoRegressive) models, VECM(Vector Error Correction Models), and so on to find out lead-lag relationships, the adjustment processes to long-run equilibrium and short-run dynamics between spot and futures prices. These methodologies have the linear analysis in common, but unfortunately the data of interest in this field tend to show nonlinearity. Therefore many scholars has been trying to construct more advanced models which can reflect this nonlinearity of the data. As a result, a multiple-regime TVECM(Threshold VECM) has been developed and several cointegration tests for linear cointegration null against nonlinear cointegration alternative have been devised based on the established multiple-regime TVECM. We can take Hansen and Seo(2002) test and Seo(2006) test as an example. Hansen and Seo(2002) proposed a SupLM type cointegration test statistic based on a 2-regime TVECM and Seo(2006) went several more steps forward by developing a Sup-Wald type cointegration test statistic based on a 3-regime TVECM.

The previous literature includes Huang, Yang, and Hwang(2009), Mamatzakis and Remoundos(2010), and Hammoudeh, Chen, and Fattouh(2010) if we confine its coverage to recently issued studies whose subject and mehodologies are relatively similar to ours. Huang, Yang, and Hwang(2009) focus their study on the long-run adjustment processes to equilibrium and causal relationships between WTI spot and futures prices with a MVTAR(MultiVariate Threshold AutoRegressive) model proposed by Tsay(1998). They also compare the predictive power between the linear and nonlinear models and have a conclusion that the in-sample prediction of the nonlinear model is clearly superior to that of the linear model. Mamatzakis and Remoundos(2010) analyse the long-run adjustment process to equilibrium of Brent

crude oil spot and futures markets along with short-run dynamics using Hansen and Seo(2002) methodology. They make the following three conclusions. First, Brent crude oil markets follow a gradual integration path. Second, Brent crude oil spot and futures prices are cointegrated, though two regimes are clearly identified. Third, adjustment costs in the error correction are present and valid at the dominant regime. Hammoudeh, Chen, and Fattouh(2010) find that spot and futures prices in each of the four widely traded commodities, copper, gold, WTI crude oil and silver are asymmetrically cointegrated using the threshold cointegration methods, Enders-Siklos(2001) and Hansen and Seo(2002). They conclude that the different adjustments among 4 commodities imply different trading strategies, depending on whether the faster adjustment happened from above or below the threshold.

The above previous studies show that (WTI) crude oil spot and futures prices are nonlinearly cointegrated and the estimation and/or prediction performance on this data can be improved when the nonlinear models such as a multiple-regime TVECM are used.

We made our model specifications after considering the results of previous studies. First, we divided the entire sample period into 5 sub-samples via structural break point detection method proposed by Bai and Perron(2003) to provide the general market condition(backwardation or contango) to each sub-sample. This allows us to get more enriched interpretations along with the presence of thresholds. Second, we took basis as a threshold variable based on "cost-of-carry relationship". Third, we set the number of thresholds to 2 so there are 3 regimes in our model. This allows us to identify the regimes with/without investment activities. Forth, We performed BDS nonlinearity test and cointegration test based on 3-regime TVECM proposed by Seo(2006) before the estimation of bivariate 3-regime TVECM. This allows us to believe that fitting nonlinear models can produce more reliable estimation results.

We had nonlinearly cointegrated data like the results of previous studies. The division of the entire sample period into 5 sub-samples and setting 3 regimes seems to help draw interesting and useful conclusions along with underlying causes even compared with the results of previous studies.

This paper is organized as follows: we explained some important methodologies used in this paper in section 2. We described the structure and nature of the data in section 3. We reported all the test results including a bivariate 3-regime TVECM in section 4, and section 5 concludes this study.

Ⅱ. Research Methodologies

1. Multiple structural change point estimation

Bai and Perron(2003) dealt with the problem of structural break point estimation and presented an efficient algorithm called 'the dynamic programming algorithm' to obtain global minimizers of the sum of squared residuals. This algorithm uses at most least-squares operations of order $O(T^2)$ for any number of structural changes m, while a standard grid procedure does least-squares operations of order $O(T^m)$. With this algorithm they successfully performed the sequential detection of structural change points by increasing the number of breaks one at a time.

They considered the following multiple linear regression with m structural breaks ($m+1$ sub-samples):

$$y_t = x_t{}'\beta + z_t{}'\delta_j + u_t \quad (t = T_{j-1}+1, \cdots, T_j) \tag{1}$$

for $j = 1, \cdots, m+1$. y_t is the observed response variable at time t, $x_t(p \times 1)$ and $z_t(q \times 1)$ are vectors of covariates, β and $\delta_j(j = 1, \cdots, m+1)$ are corresponding vectors of coefficients and u_t is the disturbance at time t. The break points $(T_1, \cdots, T_m)$ are treated as unknown($T_0 = 0$ and $T_{m+1} = T$). They tried to estimate the unknown regression coefficients together with the break points when T observations on (y_t, x_t, z_t) are in their hands.

Before using dynamic programming algorithm, the construction of the triangular matrix of sums of squared residuals for all possible segments should be made. The vertical axis of this matrix indicates the initial date of a segment and the horizontal axis the ending date, so each element means an estimated sum of squared residuals

corresponding to the associated segment. The global sum of squared residuals for any m-partition$(T_1, \cdots, T_m)$ and for any value of m should be a particular combination of these $T(T+1)/2$ sums of squared residuals. The estimates of break dates $(\hat{T}_1, \cdots, \hat{T}_m)$ correspond to this linear combination with a minimal value. They wrote that in practice, less than $T(T+1)/2$ segments were possible in this paper. For this procedure, some minimum distance h between each breaks may be imposed which can be quite helpful for reducing the number of segments along with other factors.

If the sum of squared residuals of the relevant segments have been calculated and stored, a dynamic programming algorithm can be applied. This method examines optimal one-break partitions sequentially. Let $SSR(\{T_{r,n}\})$ be the sum of squared residuals associated with the optimal partition having r breaks using the first n observations. The optimal partition solves the following recursive problem:

$$SSR(\{T_{r,n}\}) = \min_{mh \le j \le T-h} [SSR(\{T_{m-1,j}\}) + SSR(j+1, T)] \qquad (2)$$

They also constructed confidence intervals for the parameters β and δ, and the break dates. They proposed their own test statistic for multiple breaks as follows:

$$F_T(\lambda_1, \cdots, \lambda_k; q) = \frac{1}{T}\left(\frac{T-(k+1)q-p}{kq}\right)\hat{\delta}' R'(R\hat{V}(\hat{\delta})R')^{-1} R\hat{\delta} \qquad (3)$$

This is a sup$F$ type test of no structural break$(m=0)$ versus $m=k$ breaks. $(T_1, \cdots, T_k)$ is a partition such that $T_i = [T\lambda_i] (i=1, \cdots, k)$, where $\lambda_i = \frac{T_i}{T}$ can be called 'break fractions'. R is the conventional matrix such that $(R\delta)' = (\delta_1' - \delta_2', \cdots, \delta_k' - \delta_{k+1}')$. $\hat{V}(\hat{\delta})$ is the estimate of the variance covariance matrix of $\hat{\delta}$ which is robust to serial correlation and heteroscedasticity. According to Andrews(1993) and others, sup$F_T(k;q) = F_T(\hat{\lambda}_1, \cdots, \hat{\lambda}_k; q)$ holds. Here $\hat{\lambda}_1, \cdots, \hat{\lambda}_k$ minimize the global sum of squared residuals which has the same meaning of maximizing the F-test assuming spherical errors. In this paper, we'll use 'dynamic programming algorithm' presented by Bai and Perron(2003) to detect the multiple structural break points.

2. Nonlinearity test(Brock, Dechert, and Scheinkman, 1987)

Brock, Dechert, and Scheinkman(1987) suggested a new test statistic for checking whether the I.I.D. assumption of any time series could be satisfied or not. This test statistic was shown to be used for detecting serial dependence of any time series in the residual analysis(Kim et al., 2003). Two different hypotheses used in hypothesis testing are as follows:

H_0: This time series is I.I.D..

H_1: This time series is not I.I.D..

BDS test statistic can be calculated like this. There is a time series with N observations, which should be the first difference of the natural logarithms of raw data in a time series.

$$\{x_i\} = [x_1, x_2, \cdots, x_N]$$

You can select m, embedding dimension and embed the time series into m-dimensional vectors by taking each m successive points in this time series. This procedure turns the series of scalars into a series of vectors with overlapping elements.

$$\begin{aligned} x_1^m &= (x_1, x_2, \cdots, x_m) \\ x_2^m &= (x_2, x_3, \cdots, x_{m+1}) \\ &\vdots \\ x_{N-m}^m &= (x_{N-m}, x_{N-m+1}, \cdots, x_N) \end{aligned}$$

You can compute the correlation integral which measures the spatial correlation among the points by adding the number of pairs of points (i, j), where $1 \le i \le N$ and $1 \le j \le N$, in the m-dimensional space which are "close" in the sense that the points are within a radius or tolerance ϵ of each other.

$$C_{\epsilon,m} = \frac{1}{N_m(N_m - 1)} \sum_{i \ne j} I_{i,j;\epsilon} \tag{4}$$

where $I_{i,j;\epsilon} = 1$ if $\| x_i^m - x_j^m \| \le \epsilon$

$I_{i,j;\epsilon} = 0$ if $\| x_i^m - x_j^m \| > \epsilon$

Brock, Dechert, and Scheinkman(1987) proved that if this time series is I.I.D.

$$C_{\epsilon,m} \approx \left[C_{\epsilon,1} \right]^m \tag{5}$$

If $\frac{N}{m}$ is greater than 200, the values of $\frac{\epsilon}{\sigma}$ are from 0.5 to 2(Lin, 1997) and m can have its values between 2 and 5(Brock, Dechert, and Scheinkman, 1988) and the quantity $\left[C_{\epsilon,m} - (C_{\epsilon,1})^m \right]$ has an asymptotic normal distribution with mean zero and variance $V_{\epsilon,m}$ defined as follows:

$$V_{\epsilon,m} = 4\left[K^m + 2\sum_{j=1}^{m-1} K^{m-j} C_\epsilon^{2j} + (m-1)^2 - m^2 K C_\epsilon^{2m-2} \right] \tag{6}$$

where $K = K_\epsilon = \frac{6}{N_m(N_m-1)(N_m-2)} \sum_{i<j<N} h_{i,j,N;\epsilon}$

$$h_{i,j,N;\epsilon} = \frac{\left[I_{i,j;\epsilon} I_{j,N;\epsilon} + I_{i,N;\epsilon} I_{N,j;\epsilon} + I_{j,i;\epsilon} I_{i,N;\epsilon} \right]}{3}$$

Consequently, the BDS test statistic can be formulated like this:

$$BDS_{\epsilon,m} = \frac{\sqrt{N}\left[C_{\epsilon,m} - (C_{\epsilon,1})^m \right]}{\sqrt{V_{\epsilon,m}}} \tag{7}$$

It has a limiting standard normal distribution under the null hypothesis of I.I.D. as $N \to \infty$ and obtains its critical values using the standard normal distribution. Due to the type of alternative hypothesis BDS test becomes a two-tailed test and the rejection regions are placed at the two tails of probability density curve. Therefore if you can reject the null hypothesis at a given significance level α, the time series is not I.I.D., so this data should be fitted with nonlinear models.

3. Threshold Vector Error Correction Models(TVECM)

Threshold Vector Error Correction Models(TVECM) can be defined as the Vector Error Correction Models(VECM) using a threshold variable as an error correction term. The general bivariate threshold vector error correction models can be specified as follows:

$$\triangle f_t = \alpha_i^f ecm_{t-d} + \sum_{j=1}^{p} \beta_{i,t-j}^f \triangle f_{t-j} + \sum_{j=1}^{q} \gamma_{i,t-j}^f \triangle s_{t-j} + \epsilon_t^f \tag{11}$$

$$\triangle s_t = \alpha_i^s ecm_{t-d} + \sum_{j=1}^{p} \beta_{i,t-j}^s \triangle f_{t-j} + \sum_{j=1}^{q} \gamma_{i,t-j}^s \triangle s_{t-j} + \epsilon_t^s \quad (i = 1, \cdots, k) \quad (12)$$

where $ecm_{t-d} = b_{t-d} (= f_{t-d} - s_{t-d},\ \tau_{i-1} \le b_{t-d} < \tau_i)$[1] is a threshold variable. In this study, f_t and s_t are natural logarithmic values of F_t(the closing price of the nearby WTI futures contract at time t) and S_t(the spot price at time t), respectively. τ_i's are integers such as $-\infty = \tau_0 < \tau_1 < \cdots < \tau_{k-1} < \tau_k = \infty$. k and d are all positive integers and d is called 'a delay parameter'. (i) means regimes, Δ is a difference operator, and p and q are lags of AR(AutoRegressive) models. $\{\epsilon_t^{(i)}\}$ is assumed to be I.I.D. with mean 0 and variance σ_i^2 and independent for different (i) regimes.

4. The theoretical background for selecting basis as a threshold variable

We can find out the theoretical background for selecting basis as a threshold variable from the theory of storage. This theory means that the marginal convenience yield on inventory falls at a decreasing rate as aggregate inventory increases.[2] It has been the dominant model of commodity forward and futures prices[Working(1949), Brennan(1958), and Telser(1958)]. We can write down the following equation about commodity futures pricing according to this theory.

$$F(t, T) = S(t)[1 + R(t, T)] + W(t, T) - C(t, T) \quad (13)$$

where $F(t, T)$ is the futures price at time t for delivery of a commodity at T and $S(t)$ is the spot price at t. $R(t, T)$ is the interest rate at which market participants can borrow or lend over the period between t and T. $W(t, T)$ is the marginal warehousing cost and $C(t, T)$ is the marginal convenience yield.

---

1) We used basis(b_{t-d}) as an error correction term(ecm) in this study. Basis can be defined as the difference between futures and spot prices at time $t-d$ in finance.

2) The intuition of convenience yield is that an uncompensated carrying cost-a futures price does not exceed the spot price by enough to cover interest and warehousing costs-implies that storers get some other return from inventory. For example, a convenience yield can arise when holding inventory of an input lowers unit output costs and replenishing inventory involves lumpy costs. Alternatively, time delays, lumpy replenishment costs, or high costs of short-term changes in output can lead to a convenience yield on inventory held to meet customer demand for spot delivery(Fama and French, 1988).

As we are interested in basis as an threshold variable, equation (13) can be written again like this.

$$F(t,T) - S(t) = S(t)R(t,T) + W(t,T) - C(t,T) \tag{14}$$

Equation (14) is also known as "the cost-of-carry pricing relationship" and equates basis with the cost of carry. As a result, arbitrage is not profitable. Dividing both sides of equation (14) by $S(t)$, we can have the following equation.

$$\frac{F(t,T) - S(t)}{S(t)} - R(t,T) = \frac{W(t,T)}{S(t)} - \frac{C(t,T)}{S(t)} \tag{15}$$

Equation (15) shows that the observed quantity on the left-hand side which can be called "interest-adjusted basis" is the same as the difference between the relative warehousing cost, $\frac{W(t,T)}{S(t)}$ and the relative convenience yield, $\frac{C(t,T)}{S(t)}$. Assuming that the marginal warehousing cost is roughly constant, the marginal convenience yield declines at a decreasing rate with increases in inventory, and variation in the marginal convenience yield dominates variation in the marginal warehousing cost.

If we assume that $R(t,T)$ is fixed at a certain level during a certain period, we can have not exactly the same but similar results in the case of using "basis" instead of "interest-adjusted basis." On the right-hand side of equation (15), $\frac{W(t,T)}{S(t)} - \frac{C(t,T)}{S(t)}$ can be regarded as total cost(tc), so we can use this as the value of threshold parameter in TVECM.

Under the Cost-of-Carry framework and our 3-regime TVECM, if the basis is larger than the threshold band[3], an arbitrage can be profitable. Therefore, a long arbitrage position(positive basis) is the case of $b_t > tc$[4] and a short arbitrage position(negative basis) is the case of $b_t < -tc$(Huang, Yang, and Hwang, 2009).

3) We're considering bivariate TVECM with 3 regimes, so there are 2 limits which are upper limit and lower limit of threshold band.

4) In this expression, we used b_t as the basis at time t and interest-adjusted basis was replaced by basis. tc represents total cost.

5. Cointegration test based on the TVECM representation

We used the sup-Wald type test statistic proposed by Seo(2006) for finding out the property of adjustment process toward long-run equilibrium because the significantly greater power of this test against the threshold cointegration alternative has been proven compared to conventional cointegration tests. The other reason for us to choose this test is that Seo(2006) considers a Band-TVECM whose model specification is exactly the same as ours. The model is as follows:

$$\Phi(L)\Delta x_t = \alpha_1 z_{t-1} 1\{z_{t-1} \le \gamma_1\} + \alpha_2 z_{t-1} 1\{z_{t-1} > \gamma_2\} + \mu + \epsilon_t \tag{16}$$

where $t = 1, \cdots, n$ and $\Phi(L)$ is a qth-order polynomial in the lag operator defined as $\Phi(L) = I - \Phi_1 L^1 - \cdots - \Phi_q L^q$. The error correction term is defined as $z_t = x_t{}'\beta$ for a known cointegrating vector β. The threshold parameter $\gamma = (\gamma_1, \gamma_2)$ satisfying $\gamma_1 \le \gamma_2$ takes values on a compact set Γ(Seo, 2006). $1\{z_{t-1} \le \gamma_1\}$ and $1\{z_{t-1} > \gamma_2\}$ are indicator functions.

Seo(2006) develops a test for the linear cointegration null hypothesis:

$$H_0 : \alpha_1 = \alpha_2 = 0 \tag{17}$$

The test statistic which is calculated under this null hypothesis in Seo(2006) is the supremum of the Wald statistic. When γ is given, the least-squares estimators for the coefficients are the OLS estimators.

$$\begin{aligned} \Delta x_t = {} & \hat{\alpha}_1(\gamma) z_{t-1} 1\{z_{t-1} \le \gamma_1\} + \hat{\alpha}_2(\gamma) z_{t-1} 1\{z_{t-1} > \gamma_2\} \\ & + \hat{\mu}(\gamma) + \hat{\Phi}_1(\gamma)\Delta x_{t-1} + \cdots + \hat{\Phi}_q(\gamma)\Delta x_{t-q} + \hat{\epsilon}_t(\gamma) \end{aligned} \tag{18}$$

and let

$$\hat{\Sigma}(\gamma) = \frac{1}{n}\sum_{t=1}^{n} \hat{\epsilon}_t(\gamma)\hat{\epsilon}_t(\gamma)'. \tag{19}$$

The Wald statistic testing null hypothesis (17) with a fixed γ is

$$W_n(\gamma) = vec(\hat{A}(\gamma))' var(vec(\hat{A}(\gamma)))^{-1} vec(\hat{A}(\gamma))$$

$$= vec\left((Z_\gamma' M_{-1} Z_\gamma)^{-1}(Z_\gamma' M_{-1}\epsilon)\right)' \left[(Z_\gamma' M_{-1} Z_\gamma)^{-1} \otimes \hat{\Sigma}(\gamma)\right]^{-1}$$
$$\times vec\left((Z_\gamma' M_{-1} Z_\gamma)^{-1}(Z_\gamma' M_{-1}\epsilon)\right)$$
$$= tr\left\{\left(Z_\gamma' M_{-1}\epsilon \hat{\Sigma}(\gamma)^{-1/2}\right)'(Z_\gamma' M_{-1} Z_\gamma)^{-1}\left(Z_\gamma' M_{-1}\epsilon \hat{\Sigma}(\gamma)^{-1/2}\right)\right\} \quad (20)$$

where $A = (\alpha_1, \alpha_2)'$, and Z_γ and ϵ are matrices stacking $(z_{t-1}1\{z_{t-1} \le \gamma_1\}, z_{t-1}1\{z_{t-1} > \gamma_2\})$ and ϵ_t', respectively. M_{-1} is projection onto the orthogonal space of the constant and the lagged terms $\Delta x_{t-1}, \cdots, \Delta x_{t-q}$. Then the supremum statistic can be defined as

$$sup\ W = \sup_{\gamma \in \Gamma} W_n(\gamma). \quad (21)$$

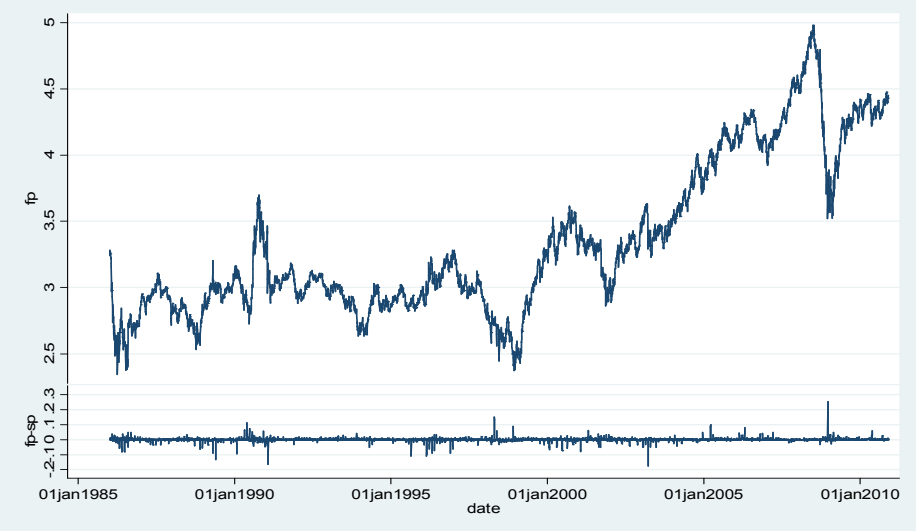
Ⅲ. Data

The raw data for this research are WTI crude oil closing prices of the nearby futures contracts traded in NYMEX and corresponding spot prices[5]. The sample period is from January 2, 1986 to December 30, 2010 and data type is daily time series. All the prices are expressed in dollar terms per barrel. We considered the trading days when both futures and spot markets opened and the final data set has 6,244 observations in total. We downloaded these data from the homepage of the Energy Information Administration and edited them for the purpose of this analysis.

<Figure 1> is the plot of log-transformed WTI crude oil closing prices of the nearby futures contracts(f_t) and basis$(f_t - s_t)$ used in this study.

5) The nearby futures series are constructed from the daily closing prices on futures contracts one month prior to the expiration month in this study.

<Figure 1> WTI crude oil closing prices of the nearby futures contracts and basis(log-transformed data)



Ⅳ. Empirical Analysis

1. Stationarity test on the entire sample period

We planned to divide the entire sample into several sub-samples with Bai and Perron(2003)'s structural break point detection method for more detailed analysis. Hence the stationarity test on the entire sample should be a preliminary one because Bai and Perron(2003)'s test can be performed on the stationary time series. The conventional stationarity tests such as Augmented Dickey-Fuller(ADF) test and Phillips-Perron(PP) test were made on log-transformed price data $(f_t = \ln(F_t),\ s_t = \ln(S_t))$ and their first differenced data $(\triangle f_t = f_t - f_{t-1},\ \triangle s_t = s_t - s_{t-1})$. <Table 1> shows the stationarity test results on the entire sample period.

Stationarity test results indicate that log-transformed price series are not stationary but their first differenced series become stationary. This condition allows

us to perform cointegration tests on these two log-transformed data.

<Table 1> Stationarity test results on the entire sample period

test	f_t	$\triangle f_t$	s_t	$\triangle s_t$
ADF	-2.9634	-17.2893^{***}	-3.0008	-17.1793^{***}
PP	-17.5387	-5678.4850^{***}	$-18.1682\ *$	-5618.7520^{***}

Notes.
1. The null hypothesis of these two tests is 'the time series has a unit root' which means that it is not stationary.
2. *** and * indicate that the null hypothesis of this test can be rejected at the 1% and 10% significance levels, respectively.

2. The detection of structural break points

We applied 'dynamic programming algorithm' presented by Bai and Perron(2003) to the stationary first-differenced data of log-transformed WTI crude oil spot and futures price time series, the same structural break point $m=4$ in both spot and futures log-return series was chosen by the criterion of the minimum RSS(Residual Sum of Squares) like <Table 2>.

<Table 2> The choice of the most appropriate structural change points m by RSS

m	0	1	2	3	4	5
RSS	4.081173	4.079751	4.078949	4.078245	4.078125	4.078174

Notes.
1. The selection criterion of Residual Sum of Squares(RSS) has been used for the detection of structural break point.
2. The break point $m=4$ corresponding to the minimum RSS value has been chosen so there are 5 sub-samples in our analysis.

<Table 3> The sub-samples divided by test result

	sub-sample 1	sub-sample 2	sub-sample 3	sub-sample 4	sub-sample 5
period by dates	1986/01/03~ 1990/10/11	1990/10/12~ 1994/08/22	1994/08/23~ 1998/12/10	1998/12/11~ 2005/08/30	2005/08/31~ 2010/11/30
period by obs. #	1~1199	1200~2167	2168~3248	3249~4924	4925~6243
# of obs. (ratio, %)	1199 (19.21%)	968 (15.51%)	1081 (17.32%)	1676 (26.85%)	1319 (21.13%)
basis (ratio, %) (+)	465(38.78)	398(41.12)	382(35.34)	738(44.03)	750(56.86)
0	86(7.17)	90(9.30)	104(9.62)	73(4.36)	82(6.22)
(−)	648(54.05)	480(49.59)	595(55.04)	865(51.61)	487(36.92)

Notes. # of obs. means the number of observations in each sub-sample.

The entire sample period was divided into 5 sub-samples according to the most appropriate structural change points $m = 4$. <Table 3> shows what's all about 5 sub-samples. Sub-sample period 2 is the smallest with 968 observations(15.51%) while sub-sample period 4 the largest with 1,676 observations(26.85%).

The five fractioned periods represent relatively different market conditions by the sign of basis which can be calculated as ($b_t = f_t - s_t$). If spot prices are above futures ones during a specific period(negative basis), the market can be called 'backwardation.' On the other hand, if futures prices are above spot ones during a specific period(positive basis), the market can be called 'contango.' The first four sub-samples have relatively more ratios of negative basis and the last sub-sample is the opposite. Hence we can recognize the first four sub-samples as a relative backwardation and the last one as a relative contango, which is consistent with many analyses that the crude oil markets have been turned into contango since 2004.

We can find several possible reasons for this phenomena in the case of WTI crude oil markets. First, the increasing inventory in US has pushed down the marginal convenience yield, which has led to the fall of the relative convenience yield and finally the increase of basis. Second, the raise of federal funds rate increased the relative warehousing cost through the increment of cost in terms of

interest rates. The up-turn of the relative warehousing cost led to positive basis, namely contango(refer to Eq.(15) in this paper).

2. Descriptive statistics for 5 sub-samples

<Table 4> reports descriptive statistics for 5 sub-samples.

<Table 4> The descriptive statistics for 5 sub-samples

	$\triangle f_t$					$\triangle s_t$				
	SS1	SS2	SS3	SS4	SS5	SS1	SS2	SS3	SS4	SS5
mean	0.00037	−0.00088	−0.00043	0.00111	0.00015	0.00038	−0.00088	−0.00042	0.00111	0.00015
maximum	0.1403	0.1268	0.1423	0.0811	0.1641	0.1915	0.1412	0.1538	0.1244	0.1641
minimum	−0.1676	−0.4005	−0.0912	−0.1654	−0.1307	−0.1834	−0.4064	−0.1409	−0.1709	−0.1283
s.d.	0.0287	0.0257	0.0213	0.0257	0.0270	0.0292	0.0257	0.0228	0.0259	0.0271
skewness	−0.3141	−4.2384	0.4427	−0.6534	0.1703	−0.1212	−4.3664	0.1799	0.1378	−4.3664
kurtosis	8.5066	67.8943	7.8081	5.8828	7.6410	9.7393	70.9493	9.5100	6.8457	7.5427
Jarque −Bera	1533.30 (0.0000)	172574.1 (0.0000)	1075.57 (0.0000)	699.179 (0.0000)	1189.21 (0.0000)	2270.07 (0.0000)	189103.5 (0.0000)	1912.94 (0.0000)	1150.64 (0.0000)	1137.44 (0.0000)

Notes.
1. The study data comprise two log−returns series based on the daily closing prices of WTI crude oil nearly futures contracts and its corresponding daily spot prices. The entire data cover January 3, 1986 through November 30, 2010. All the prices are stated dollar terms per barrel.
2. The Jarque−Bera test is a goodness−of−fit measure of departure of normality based on the sample kurtosis and skewness. The null hypothesis of this test is that 'the data are from a normal distribution.' The test results showed that p−values are really close to zero so we can reject the null hypothesis, which means that all the series for this analysis don't follow a normal distribution.

The two interesting results were found from the descriptive statistics. First, five sub-samples can be grouped into the following two: the first group(sub-sample 2 and sub-sample 3) has negative means and relatively lower level of volatility but the second group(sub-sample 1, sub-sample 4, and sub-sample 5) has positive means and relatively higher level of volatility. Sub-sample 4 and sub-sample 2 have the highest and lowest means, respectively. Sub-sample 1 and sub-sample 3 have the highest and lowest level of volatility, respectively. Second, all the series don't follow a normal distribution as Jarque-Bera test results represent. Most of the series have a kurtosis

significantly higher than 3, implying that extreme market movements in either direction(gains or losses) occur in WTI spot and futures markets with greater frequency in practice than would be predicted by the normal distribution(Hammoudeh, Chen, and Fattouh, 2010). Especially, sub-sample 2 has remarkably smaller and negative skewness and much higher kurtosis than the others, which reflect it's mean and standard deviation rank nearly the lowest.

3. Stationarity test

Augmented Dickey-Fuller(ADF) test and Phillips-Perron(PP) test were performed on the log-transformed data$\left(f_t = \ln\left(F_t\right),\ s_t = \ln\left(S_t\right)\right)$ and their first-differenced data for checking stationarity of these time series. The results of these two unit root tests in 5 sub-samples are reported in <Table 5>.

The null hypothesis of these unit root tests is 'this time series has a unit root.' While the unit root test results show that log-transformed time series is not stationary except for PP test results of sub-sample 2, their first-differenced time series become stationary at the 5% significance level. Fortunately ADF test results on log-transformed time series for sub-sample 2 show the disability to reject the null hypothesis at the 5% significance level, so we can say that all the series are turned into stationary ones after first difference. This condition can guarantee that the two log-transformed time series are ready for cointegration test.

<Table 5> Unit root test results

period	test	f_t	$\triangle f_t$	s_t	$\triangle s_t$
1	ADF	−2.2330	−10.3837***	−2.2347	−10.4685***
	PP	−14.8043	−1151.049***	−14.9887	1150.5130***
2	ADF	−3.2271 *	−11.8356***	−3.2057 *	−12.0641***
	PP	−27.0754 **	−805.7402***	−26.0839 **	−786.6005***
3	ADF	−0.9247	−11.2769***	−0.9708	−11.4296***
	PP	−5.4429	−921.8112***	−6.2630	−900.4872***
4	ADF	−2.4246	−12.7612***	−2.4375	−13.1990***
	PP	−13.0693	−1570.485***	−13.4312	−1628.5740 ***
5	ADF	−1.6268	−11.6684***	−1.6367	−11.7932***
	PP	−6.5561	−1335.6430 ***	−7.2301	−1307.1040 ***

Notes.
1. The null hypothesis of these two unit root tests is 'the time series has a unit root' which means that it is not stationary.
2. ***, **, and * indicate that the null hypothesis of this test can be rejected at the 1%, 5%, and 10% significance levels, respectively.

4. Nonlinearity test

BDS test was performed on $\triangle f_t[= f_t - f_{t-1} = \ln(F_t) - \ln(F_{t-1})]$ and $\triangle s_t[= s_t - s_{t-1} = \ln(S_t) - \ln(S_{t-1})]$ series of 5 sub-samples to check the appropriate model selection(linear models vs. nonlinear models). We chose the linear model(VECM) with the highest goodness of fit, extracted its residual series and applied BDS nonlinearity test on it. The null and alternative hypotheses of this nonlinearity test are as follows:

H_0: This time series is I.I.D..

H_1: This time series is not I.I.D..

<Table 6> shows the nonlinearity test results. Although we can witness a rather doubtful result on $\triangle f_t$ series of sub-sample 4, we can reject the null hypothesis of BDS test with full confidence. It can be therefore valid to fit nonlinear models to these data.

<Table 6> Nonlinearity test results

m	Δf_t (ss1)		Δs_t		Δf_t (ss2)		Δs_t		Δf_t (ss3)		Δs_t	
	ϵ/σ	t.s.	ϵ/σ	t.s.	ϵ/σ	t.s.	ϵ/σ	t.s.	ϵ/σ	t.s.	ϵ/σ	t.s.
2	0.0142	8.9132***	0.0144	11.6625***	0.0119	5.0993***	0.0118	5.0111***	0.0106	2.9155***	0.0110	3.6094**
	0.0285	9.6624***	0.0288	11.9504***	0.0238	5.0466***	0.0237	5.7293***	0.0212	3.3947***	0.0220	4.2558***
	0.0427	9.8262***	0.0432	10.9781***	0.0357	5.6566***	0.0355	6.8341***	0.0317	4.1004***	0.0330	4.8459***
	0.0570	9.5431***	0.0576	9.3032***	0.0476	5.1857***	0.0474	7.7658***	0.0423	4.5944***	0.0440	5.3583***
3	0.0142	13.0175***	0.0144	16.0968***	0.0119	7.4760***	0.0118	6.1677***	0.0106	3.7272***	0.0110	4.8095***
	0.0285	12.5436***	0.0288	15.1655***	0.0238	7.5595***	0.0237	7.3426***	0.0212	4.4044***	0.0220	5.7447***
	0.0427	11.5291***	0.0432	13.2289***	0.0357	8.4524***	0.0355	8.6602***	0.0317	5.4335***	0.0330	6.7331***
	0.0570	10.5564***	0.0576	11.0665***	0.0476	8.0589***	0.0474	9.8152***	0.0423	6.5013***	0.0440	7.9082***

m	Δf_t (ss4)			Δs_t	Δf_t (ss5)			Δs_t
	ϵ/σ	t.s.	ϵ/σ	t.s.	ϵ/σ	t.s.	ϵ/σ	t.s.
2	0.0122	0.9992	0.0127	2.5640**	0.0135	6.2499***	0.0133	6.3237***
	0.0244	0.9294	0.0254	2.9327***	0.0270	7.6362***	0.0265	8.2998***
	0.0366	1.5547	0.0381	4.2644***	0.0404	9.7136***	0.0398	10.4274***
	0.0489	2.8553***	0.0508	6.0795***	0.0539	10.7567***	0.0531	11.6111***
3	0.0122	1.0342	0.0127	3.2332***	0.0135	8.3087***	0.0133	8.0427***
	0.0244	1.2354	0.0254	3.6768***	0.0270	9.9273***	0.0265	10.2354***
	0.0366	1.8413*	0.0381	5.0227***	0.0404	12.6898***	0.0398	13.0201***
	0.0489	3.3090***	0.0508	7.0165***	0.0539	14.1030***	0.0531	14.1962***

Notes.
1. The null hypothesis of BDS nonlinearity test is 'this time series is I.I.D.' which means that it is fit for linear modelling, while the alternative hypothesis is 'this time series is not I.I.D.' which means that it is fit for nonlinear modelling.
2. m represents embedding dimension.
3. t.s. represents a test statistic.
4. ***, **, and * indicate that the null hypothesis of this test can be rejected at the 1%, 5%, and 10% significance levels, respectively.

5. The cointegration test based on TVECM: Seo(2006)

We chose the methodology developed by Seo(2006) as a cointegration test for the following reasons. First, Seo(2006) designed his cointegration test based on a 3-regime or band TVECM[6] which is exactly the same as ours. Second, his test method has significantly greater power than conventional cointegration tests.

Seo(2006) proposed a sup-Wald type test statistic(refer to Eq.(20) and Eq.(21) in this paper) and tested the linear cointegration null hypothesis against the threshold cointegration in the following form:

$$H_0 : \alpha_1 = \alpha_2 = 0$$

<Table 7> shows cointegration test results. We are sure about the presence of threshold cointegration relationships between the log-transformed spot and futures prices by the rejection of null hypothesis even at the 1% significance level. This test results can make our model specification(a bivariate 3-regime TVECM) really fit for our data along with the results of BDS nonlinearity test. We added 4 different critical values calculated by residual-based bootstrapping to test statistics for your reference.

<Table 7> Seo(2006) Threshold Cointegration test results

period	test statistic	critical value			
		90%	95%	97.5%	99%
1	389.3841***	45.2542	47.2749	47.5551	47.7233
2	325.4701***	26.7578	29.3784	29.7153	29.9174
3	399.9170***	49.5577	50.0360	50.2967	50.4531
4	560.6767***	65.3884	65.5913	66.9630	67.7861
5	535.8078***	24.0345	24.0818	24.3638	24.5329

Notes.
*** indicates that the null hypothesis of this test can be rejected at the 1% significance level.

6) Hansen and Seo(2002) have also proposed a similar cointegration test statistic but their test method is based on a vector error correction model with 2 regimes.

6. The application of a bivariate 3-regime TVECM

The entire sample has been divided into 5 sub-samples via the detection of structural break points($m=4$). This allows us not to miss any information about sub-divisions in this analysis. The 5 sub-samples can be classified into two parts, that is, the first 4 sub-samples with backwardation market character and the last sub-sample with contango market character(see <Table 3> and its related explanations).

Moreover, we applied a bivariate 3-regime TVECM to each sub-sample to infer appropriate investment strategies of arbitrageurs. Some explanations are needed about our model specification. First, we didn't include constant term and time trend term in this model. Second, we assumed the error correction terms without constant terms, and a cointegrating vector of $(1,-1)'$ to equate an error correction term with a basis. Third, the lags of each equation has been chosen by the model selection criteria such as AIC and BIC, and the goodness of fit measure like SSR.

We can expect the sign of some parameters, especially α_i^f and α_i^s, in the following models.

$$\triangle f_t = \alpha_i^f ecm_{t-1} + \beta_i^f(L)\triangle f_{t-1} + \gamma_i^f(L)\triangle s_{t-1} + \epsilon_t^f \tag{22}$$

$$\triangle s_t = \alpha_i^s ecm_{t-1} + \beta_i^s(L)\triangle f_{t-1} + \gamma_i^s(L)\triangle s_{t-1} + \epsilon_t^s \tag{23}$$

When $ecm_{t-1}(=b_{t-1}=f_{t-1}-s_{t-1})>0$, f_t should decrease and s_t should increase to return to the price relationship to the long-run equilibrium, and the opposite being the case when $ecm_{t-1}(=b_{t-1}=f_{t-1}-s_{t-1})<0$. Therefore, $ecm_{t-1}(=b_{t-1}=f_{t-1}-s_{t-1})<0$ should be negative and positive, respectively(Li, 2009).

If we consider the investment strategies of an arbitrageur upon the above setting, the statistical significance of estimates should be another important factor. If an arbitrageur is in contango[$ecm_{t-1}(=b_{t-1}=f_{t-1}-s_{t-1})>0$, positive basis], he'll sell his futures contracts and buy spots. This moves spot prices upward and futures prices downward. That is, there can be a positive

relationship between basis and spot prices, so we can expect α^s to be statistically significant and positive. On the contrary, if he is in backwardation [$ecm_{t-1}(= b_{t-1} = f_{t-1} - s_{t-1}) < 0$, negative basis], he'll sell his spots and buy futures contracts. This moves futures prices upward and spot prices downward. That is, there can be a negative relationship between basis and futures prices, so we can expect α^f to be statistically significant and negative. The absolute magnitude of α^f and α^s indicates the adjustment speed to the long-run equilibrium in the WTI crude oil futures and spot markets, respectively, while β and γ do the dynamics in the short run.

<Table 8>~<Table 12> summarize the estimation results of a bivariate 3-regime TVECM for each sub-sample.

<Table 8> Bivariate 3-regime TVECM estimation results for sub-sample 1

(Threshold values: 0.001033, 0.003305)

sub-sample 1		lower regime (69.6%)	middle regime (14.4%)	upper regime (16.0%)
$\triangle f_t$	α^f	-0.2917^{***}	-2.8235	-0.0780
	β_{t-1}	0.0494	-0.1441	-0.0903
	γ_{t-1}	-0.0046	-0.1929	0.1139
$\triangle s_t$	α^s	0.3775^{***}	-1.6427	0.4393^{***}
	β_{t-1}	-0.0209	0.0138	-0.0453
	γ_{t-1}	0.0469	-0.2049	0.0965

Notes.
1. The figures in parenthesis represent the ratio of the number of observations in this regime to the total number of observations.
2. *** indicates that this estimate is statistically significant with not being zero at the 1% significance level.

The ratio of the number of observations in lower regime($b_{t-1} < 0.001033$, roughly negative basis, backwardation) to the total number of observations is comparatively greater than the cases of sub-samples 2, 3, and 4 whose dominant regime is a middle regime. As expected, α^f in lower regime(negative basis) and α^s in upper regime(positive basis) are significantly negative and positive, respectively. The absolute magnitude of α^s is larger than that of α^f so the adjustment speed to

the long-run equilibrium in spot markets is faster than that in futures markets, but the direction of the long-run adjustment in these two markets are the contrary. When futures prices are higher than spot prices in upper regime(contango markets) of spot markets, we can find faster adjustment speed compared to the opposite situation of the same markets. If we look at short-run dynamics, we cannot find out any significant influence flow.

<Table 9> Bivariate 3-regime TVECM estimation results for sub-sample 2

(Threshold values: -0.002582, 0.009939)

sub-sample 2		lower regime (17.8%)	middle regime (77.1%)	upper regime (5.1%)
$\triangle f_t$	α^f	0.0207	−0.1927	−0.0304
	β_{t-1}	−0.2462 *	0.1026	−0.9425***
	γ_{t-1}	−0.0586	0.0855	0.8213 **
$\triangle s_t$	α^s	0.4686***	0.4416	0.5864 *
	β_{t-1}	−0.5183***	0.2022	−0.7422 **
	γ_{t-1}	0.2169 *	−0.0504	0.7416 **

Notes.
1. The figures in parenthesis represent the ratio of the number of observations in this regime to the total number of observations.
2. ***, ** and * indicate that this estimate is statistically significant with not being zero at the 1%, 5% and 10% significance levels, respectively.

α^f in lower regime(negative basis) isn't significantly negative while α^s in upper regime(positive basis) is significantly positive in sub-sample 2. The upper regime(contango) of spot markets shows brisker adjustment to the long-run equilibrium than the lower regime. The upper regime of futures and spot markets contains much energetic short-run dynamics. The only difference is the direction of influence flow. That is to say, the futures markets give the adverse influence on both markets while the spot markets the favorable influence on them.

<Table 10> Bivariate 3-regime TVECM estimation results for sub-sample 3

(Threshold values: -0.008863, 0.003872)

sub-sample 3		lower regime (9.1%)	middle regime (78.7%)	upper regime (12.2%)
$\triangle f_t$	α^f	-0.0320	-0.1379	0.0085
	β_{t-1}	-0.0785	0.0318	-0.4231^{***}
	γ_{t-1}	0.0626	0.0508	0.5044^{***}
$\triangle s_t$	α^s	0.5238^{***}	0.4741	0.9225^{***}
	β_{t-1}	-0.3807^{***}	0.0362	-1.0723^{***}
	γ_{t-1}	0.2210^{*}	0.0698	1.0919^{***}

Notes.
1. The figures in parenthesis represent the ratio of the number of observations in this regime to the total number of observations.
2. *** and * indicate that this estimate is statistically significant with not being zero at the 1% and 10% significance levels, respectively.

α^f in lower regime(negative basis) isn't significantly negative while α^s in upper regime(positive basis) is significantly positive in sub-sample 3. The upper regime(contango) of spot markets shows much brisker adjustment to the long-run equilibrium than the lower regime. Specifically, α^s of upper regime in spot markets(0.9225) is slightly less than twice the magnitude of α^s of lower regime(0.5238). The upper regime of futures and spot markets contains very active short-run dynamics. The only difference is the direction of influence. In other words, the futures markets give the adverse influence on both markets while the spot markets the favorable influence on them.

<Table 11> Bivariate 3-regime TVECM estimation results for sub-sample 4

(Threshold values: -0.009841, 0.007600)

sub-sample 4		lower regime (6%)	middle regime (87.6%)	upper regime (6.3%)
$\triangle f_t$	α^f	0.1889	−1.4601	0.4601 *
	β_{t-1}	−0.3945 **	0.1300	−0.7609***
	β_{t-2}	−0.3196 *	0.0021	−0.4190
	γ_{t-1}	0.2432	−0.1173	0.8239***
	γ_{t-2}	0.2720	−0.0103	0.4151
$\triangle s_t$	α^s	1.1994	−0.6180	1.3285***
	β_{t-1}	−1.0054***	0.0831	−0.9749***
	β_{t-2}	−1.0772***	0.0071	−0.6575 **
	γ_{t-1}	0.7949***	−0.0833	1.0404***
	γ_{t-2}	0.7950***	−0.0105	0.6077 **

Notes.
1. The figures in parenthesis represent the ratio of the number of observations in this regime to the total number of observations.
2. ***, **, and * indicate that this estimate is statistically significant with not being zero at the 1%, 5% and 10% significance levels, respectively.

α^f in lower regime(negative basis) isn't significantly negative while α^s in upper regime(positive basis) is significantly positive in sub-sample 4. The upper regime of futures and spot markets contains very active short-run dynamics. The only difference is the direction of influence flow. Namely, the futures markets give the adverse influence on both markets while the spot markets the favorable influence on them. The short-run adjustment coefficients have one thing in common, that is, the magnitude of influence at $t-1$ is larger than that at $t-2$.

<Table 12> Bivariate 3-regime TVECM estimation results for sub-sample 5

(Threshold values: -0.003066, -0.001044)

sub-sample 5		lower regime (5.4%)	middle regime (7.8%)	upper regime (86.8%)
$\triangle f_t$	α^f	-1.1444^{**}	3.3505	0.1301
	β_{t-1}	1.6663^{***}	-2.8486	-0.1448^{*}
	β_{t-2}	1.3636^{**}	-1.5144	-0.2564^{***}
	γ_{t-1}	-2.0532^{***}	2.9356	-0.1097
	γ_{t-2}	-1.4670^{**}	1.1349	0.2648^{***}
$\triangle s_t$	α^s	-0.3303	4.7121	0.5499^{***}
	β_{t-1}	1.7092^{***}	-2.6415	-0.4411^{***}
	β_{t-2}	0.8776	-1.4497	-0.1281
	γ_{t-1}	-2.0456^{***}	2.7168	0.4264^{***}
	γ_{t-2}	-1.0404	1.1050	0.1601^{*}

Notes.
1. The figures in parenthesis represent the ratio of the number of observations in this regime to the total number of observations.
2. ***, **, and * indicate that this estimate is statistically significant with not being zero at the 1%, 5% and 10% significance levels, respectively.

The ratio of the number of observations in upper regime($b_{t-1} > -0.001044$, roughly positive basis, contango) to the total number of observations is comparatively greater than the cases of sub-samples 2, 3, and 4 whose dominant regime is a middle regime. As expected, α^f in lower regime(negative basis) and α^s in upper regime(positive basis) are significantly negative and positive, respectively.

The upper regime(roughly contango) shows quite different short-run dynamics according to markets. Futures markets give negative influence on both markets while spot markets do the opposite, but the influence-receiving timing is 2 lags and 1 lag after the occurrence of impulse in futures and spot markets, respectively. Compared with short-run dynamics in upper regime, lower regime shows many eye-catching movements. The magnitude of short-run adjustment coefficients is far greater and the direction of influence flow is exactly the opposite compared to the cases of upper regime.

Ⅴ. Conclusions

VECM is a conventional and effective statistical model for analyzing both the adjustment process toward long-run equilibrium and short-run dynamics with cointegrated data. We employed a bivariate 3-regime Threshold VECM to check whether there is any differentiable analysis results or arbitrageurs' investment behaviors after considering the different 3 regimes which can be classified by basis. Moreover we divided the entire sample into 5 sub-samples using Bai and Perron(2003)'s structural break point detection methods not to miss any market information on each sub-sample. which to each sub-period for enriching the interpretation of estimation results. The two test results including BDS nonlinearity test and Seo(2006)'s cointegration test guaranteed the preference of nonlinear Threshold VECM over linear VECM.

The estimation results of our bivariate 3-regime TVECM showed quite interesting points. First, the middle regimes of all 5 sub-samples were not targeted by arbitragers so we could find their activities only in lower and/or upper regimes. Second, we found investors' arbitrage behaviors in the upper regimes(positive basis, contango) of all the 5 sub-samples but in the lower regimes(negative basis, backwardation) of only sub-samples 1 and 5. Third, most of the adjustments tends to happen in upper regime(contango) and spot markets, which means the price leadership of futures markets especially in sub-samples 1, 2, and 3. We found much brisker feedbacks of price information in sub-samples 4 and 5 where positive means and relatively higher level of volatility in terms of returns were recorded.

This paper is on the way to completion. We would like to add the estimation of a linear VECM as a benchmark even though we fully guaranteed the preference of a nonlinear 3-regime TVECM over a linear VECM in this paper this time. This job is also needed for the comparative study of forecasting performance of these two different models afterwards.

References

Andrews D.W.K. (1993), "Tests for parameter instability and structural change with unknown change point," *Econometrica* 61, pp.821-856.

Andrews D.W.K. and Ploberger W. (1994), "Optimal tests when a nuisance parameter is present only under the alternative," *Econometrica* 62, pp.1383-1414.

Bai J. and Perron P. (1998), "Estimating and testing linear models with multiple structural changes," *Econometrica* 66, pp.47-78.

Brennan M.J. (1958), "The Supply of Storage," *American Economic Review* 48, pp.50-72.

Brock W.A., Dechert W.D., and Scheinkman J.A. (1987), "A test for independence based on the correlation dimension," Working paper, Department of Economics, University of Wisconsin, Madison.

Enders W. and Siklos P.L. (2001), "Cointegration and threshold adjustment," *Journal of Business and Economic Statistics* 19, pp.166-176.

Fama E.F. and French K.R. (1988), "Business Cycles and the Behavior of Metals Prices," *Journal of Finance* 43(5), pp.1075-1093.

Hammoudeh S., Chen L.H., and Fattouh B. (2010), "Asymmetric Adjustments in Oil and Metals Markets," *Energy Journey* 31(4), pp.183-203.

Hansen B.E. (1992), "Tests for parameter instability in regressions with I(1) processes," *Journal of Business and Economic Statistics* 10, pp.321-335.

Hansen B.E. and Seo B. (2002), "Testing for two-regime threshold cointegration in vector error-correction models," *Journal of Econometrics* 110(2), pp.293-318.

Huang B.N., Yang C.W., and Hwang M.J. (2009), "The dynamics of a nonlinear relationship between crude oil spot and futures prices: A multivariate threshold regression approach," *Energy Economics* 31, pp.91-98.

Kim H.S., Kang D.S., and Kim J.H. (2003), "The BDS statistic and residual test," *Stochastic Environmental Research and Risk Assessment* No. 17, pp.104-115.

Kuan C.M. and Hornik K. (1995), "The generalized fluctuation test: A unifying view," *Econometric Reviews* 14, pp.135-161.

Li M. (2009), "The dynamics of the relationship between spot and futures markets under high and low variance regimes," *Applied Stochastic Models in Business and*

Industry 25, pp.696-718.

Lin, K. (1997), "The ABC's of BDS," *Journal of Computational Intelligence in Finance,* 97(July/August), pp.23-26.

Mamatzakis E., Remoundos P. (2010), "Threshold Cointegration in Brent crude futures market," *MPRA Paper* 19978(12), pp.1-30.

Ploberger W., Kramer W. (1992), "The CUSUM test with OLS residuals," *Econometrica* 60(2), pp.271-285.

Seo M. (2006), "Bootstrap testing for the null of no cointegration in a threshold vector error correction model," *Journal of Econometrics* 134, pp.129-150.

Telser L.G. (1958), "Futures Trading and the Storage of Cotton and Wheat," *Journal of Political Economy* 66, pp.233-255.

Tong, H. (1978), "On a threshold model," In *Pattern recognition and signal processing*, (ed. C. H. Chen). Sijthoff and Noordhoff, Amsterdam.

Tsay R.S. (2009), "Testing and Modeling multivariate threshold models," *Journal of the American Statistical Association* 93(443), pp.1188-1202.

Working H. (1949), "The Theory of the Price of Storage," *American Economic Review* 39, pp.1254-1262.