

# A Co-Integration Analysis of the Interdependencies between Crude Oil and Distillate Fuel Prices

Jane Aduda<sup>1</sup>, Patrick Weke<sup>2</sup>, Philip Ngare<sup>2</sup>

<sup>1</sup>Department of Statistics and Actuarial Sciences, Jomo Kenyatta university of Agriculture and Technology, Nairobi, Kenya

<sup>2</sup>School of Mathematics, University of Nairobi, Nairobi, Kenya

Email: jaduda@jkuat.ac.ke, pweke@uonbi.ac.ke, pngare@uonbi.ac.ke

**How to cite this paper:** Aduda, J., Weke, P. and Ngare, P. (2018) A Co-Integration Analysis of the Interdependencies between Crude Oil and Distillate Fuel Prices. *Journal of Mathematical Finance*, 8, 478-496. <https://doi.org/10.4236/jmf.2018.82030>

**Received:** October 16, 2017

**Accepted:** May 28, 2018

**Published:** May 31, 2018

Copyright © 2018 by authors and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

The co-evolution and co-movement of financial time series are of utmost importance in contemporary finance, especially when considering the joint behaviour of asset price realizations. The ability to model interdependencies and volatility spill-over effects introduces interesting dimensions in finance. This paper explores co-integrating relationships between crude oil and distillate fuel prices. Existence of multivariate co-integrating relations and bidirectional Granger-Causality is established among the series. It is also established that even after fitting a full VECM, the residuals are not necessarily multivariate normal suggesting the noise could as well be multivariate GARCH.

## Keywords

Co-Integration, Futures and Spot Prices, Granger-Causality, ECM, VECM, Volatility Spill-Over, Johansen's Test, Engle-Granger Test

---

## 1. Introduction

In empirical finance, most realistic applications are actually multivariate in nature. The co-evolution and covariation between the prices of financial assets is essentially important in Finance. Often, the interest is not merely in the behaviour of a single stock but rather in the joint behaviour of several stocks, and such joint behaviour is described using multivariate distributions. Multivariate analysis provides a framework for describing the properties of individual series as well as any possible correlations among series that are interrelated both contemporaneously and across time lags [1]. Co-integration, which provides a sound methodology for modelling both short-run and long-run dynamics in a

system, has emerged as a powerful technique for investigating common trends in multivariate time series [2]. It was developed as a means of modelling dynamic co-dependencies in multivariate time series [3] and it deals with the common behaviour of a multivariate time series where each component series may be non-stationary, but certain linear combinations of these components are stationary [3].

Co-integrated processes are characterized by short-term dynamics and long-run equilibria. Economic theory supposes a long term pricing relationship between prices of crude oil and revenues from distillate fuels. This implies that profit spreads tend to converge to a long term average. Co-integration theory, is useful in estimating and testing long-term equilibrium relationships among non-stationary asset prices and making meaningful statistical inference. Co-integrated series have stationary co-integrating residuals, and various spreads will be stationary [4]. This study uses co-integration to analyse the long-term equilibrium relationship between the prices of crude oil and the distillate fuels. The findings from this study will aid in gaining insights into the interdependencies between energy input and output products, information which is useful in portfolio diversification from the refiner's perspective, both for purposes of profit maximization and risk management.

The positive correlation of price variations or volatility clustering, evidenced in speculative markets motivated the introduction of the autoregressive conditional heteroscedastic (ARCH) process by Engle [5], and it's generalization to the generalised autoregressive conditional heteroscedastic (GARCH) by Bollerslev [6] later. Given their univariate nature, these models neglect the possibility of any further information embedded in multiple measurements with temporal and cross-sectional dependence in empirical stock price variation and the contemporaneous cross correlation implied by economic theory, such as a set of asset prices, exchange and interest rates, stock market indices among other macroeconomic variables.

For dynamic volatilities, multivariate models provide the natural framework to account for cross sectional information. Haigh and Holt [7] employ a Multivariate GARCH (MGARCH) model that allows direct incorporation of the time to maturity effect in accounting for the time-varying volatility spill overs between related markets when considering the possibility of hedging crude oil, unleaded gasoline and heating oil price risk simultaneously in a time varying setting. Their results show the importance of cross market linkages between crude oil, unleaded gasoline and heating oil markets. Veiga and McAleer [8] use the vector autoregressive moving average asymmetric generalised autoregressive conditional heteroscedasticity (VARMA-AGARCH) model of Hoti *et al.* [9] to show the existence of volatility spillovers between the Standard & Poor's 500 Index (S & P 500), the Financial Times Stock Exchange 100 Index (FTSE 100) and Nikkei 225.

This study builds on the work of Aduda *et al.* [10] in detailing empirical analysis of financial time series data. Official daily closing prices from the trading

floor of the New York Mercantile Exchange (NYMEX) for a specific delivery month for Cushing Oklahoma West Texas Intermediate (OK WTI), Reformulated Blendstock for Oxygenate Blending (RBOB), and the number 1 heating oil futures contracts are considered. The existence of co-integration is established and causal relationships are investigated. The analysis is done using MATLAB [11] and R. The rest of the document is organized as follows: Section 2 discusses the methods used in empirical analysis, Section 3 gives a brief discussion of the data used, and the empirical analysis and the results obtained and finally Section 4 gives a summary of the work and the findings.

## 2. Methodology

In this section, we test for cointegration using the trace test, fit a full VECM and test for Grangercausality. The Durbin-Watson (D-W) is also used to check for spurious regression.

### 2.1. Co-Integration

The error processes from two non-stationary series can be represented as a combination of two cumulated error processes. These cumulated error processes form stochastic trends which produce another non-stationary process if combined. If two series, say  $x_t$  and  $y_t$ , are related, they would be expected to move together and their two stochastic trends would be similar. If when combined, it would be possible to find a combination of them which eliminates the non-stationarity, then the two series would be co-integrated.

Co-integrated series evolve together, staying close to each other, even if individually, their systems drift about. If the series are not co-integrated, then, their spreads can deviate without bounds and spread trading for risk management would not be optimal [4]. If spreads are mean-reverting, and asset prices are tied together by a common stochastic trend in the long-term, then, the prices are said to be co-integrated. This means there are system feedbacks that keep variables mutually aligned [12]. Co-integration simply augments correlation analysis to include a first stage where the price data are analysed and a second stage that includes dynamic analysis of correlations which inform about any lead-lag behaviour between returns [13].

According to Engle and Granger [14], two series  $\{x_t\}$  and  $\{y_t\}$  are co-integrated if  $x_t, y_t \sim I(1)$  but there exists  $\alpha$  such that  $u_t = x_t - \alpha y_t \sim I(0)$ . Generally, the variables in a  $k$ -dimensional process  $\{p_t\}$  are called co-integrated of order  $(d, b)$ , written briefly as,  $p_t \sim CI(d, b)$ , if all components of  $\{p_t\}$  are  $I(d)$  and there exists a linear combination  $a_t := \beta' p_t$  with  $\beta = (\beta_1, \dots, \beta_k)' \neq 0$  such that  $a_t \sim I(d-b)$  [15]. It is possible to have up to  $n - 1$  linearly independent co-integration vectors.

Co-integration analysis seeks to detect any common stochastic trends in price data and use these trends for dynamic analysis of correlations in returns, through, the error correction model (ECM). If two variables are  $I(1)$  and

co-integrated, they can be modelled as having been generated by an ECM which corrects the deviations from the long-run equilibrium [2]. Multivariate time series are best considered as components of some vector-valued process  $\{p_t\}$ , which has both serial dependence within each component of the series  $\{p_{it}\}$  and interdependence between different components  $\{p_{ij}\}$  and  $\{p_{ji}\}$ , for  $i \neq j$  [3] [16].

For a  $k$ -dimensional time series  $\{p_t = (p_{t1}, \dots, p_{tk})'\}$ , the mean vector is defined by  $\mu = E[p_t] = (E(p_{t1}), \dots, E(p_{tk}))'$  and the covariance matrix

$$\Gamma(t+h, t) = \text{Cov}(p_{t+h}, p_t) = \begin{pmatrix} \gamma_{11}(t+h, t) & \dots & \gamma_{1k}(t+h, t) \\ \vdots & \ddots & \vdots \\ \gamma_{k1}(t+h, t) & \dots & \gamma_{kk}(t+h, t) \end{pmatrix} \quad \text{where}$$

$$\gamma_{ij}(t+h, t) = \text{Cov}(p_{t+h,i}, p_{t,j}).$$

This which can be written in compact form as

$$\Gamma(t+h, t) = E(p_{t+h} - \mu_{t+h})(p_t - \mu_t)'. \tag{1}$$

The diagonal elements of the matrix in equation (1) are the auto-covariance functions of the univariate series  $\{p_{it}\}$ , whereas the off-diagonal elements are the cross-covariances between  $p_{t+h,i}$  and  $p_{t,j}$   $i \neq j$ . Note also that  $\gamma_{ij}(h) = \gamma_{ji}(-h)$

and the autocorrelation matrix is then defined as  $R(h) = \begin{pmatrix} \rho_{11}(h) & \dots & \rho_{1k}(h) \\ \vdots & \ddots & \vdots \\ \rho_{k1}(h) & \dots & \rho_{kk}(h) \end{pmatrix}$ ,

where the cross-correlation function  $\rho_{ij}(h) = \gamma_{ij}(h)(\gamma_{ii}(0)\gamma_{jj}(0))^{-1/2}$ . Unlike the case of univariate stationary time series for which the auto-covariances of lag  $h$  and lag  $-h$  are identical, one must take the transpose of a positive-lag cross-covariance matrix to obtain the negative-lag cross-covariance matrix.

The cross-covariance matrices  $\Gamma(h)$  and cross-correlation matrices  $\rho(h)$  are positive definite since  $\sum_{i=1}^n \sum_{j=1}^n b_i' \Gamma(i-j) b_j \geq 0$  for all positive integers  $n$  and all

$k$ -dimensional vectors  $b_1, \dots, b_n$ , which follows since  $\text{Var}\left(\sum_{i=1}^n b_i' p_{t-i}\right) \geq 0$ .

### 2.2. Error Correction Model

The time paths of co-integrated variables are influenced by the extent of any deviation from long-run equilibrium. In any case, for the system to return to the long-run equilibrium, the movements of at least some of the variables must respond to the magnitude of the disequilibrium. These movements are captured using an ECM. When two random walk  $I(1)$  variables are co-integrated, an ECM can be formulated to study their short-run dynamics as influenced by the deviations from equilibrium [17]. An ECM is a dynamic model in which the movement of a variable in any period is related to the previous period's departure from the long-run equilibrium. In such a model, the changes in a variable depend on the deviations from some equilibrium relation [15].

Consider a simple long-term equilibrium model

$$y_t = \beta_0 + \beta_1 x_t. \tag{2}$$

A simple dynamic model of short-run adjustment is given by

$$y_t = \alpha_0 + \gamma_0 x_t + \gamma_1 x_{t-1} + \alpha_1 y_{t-1} + z_t \tag{3}$$

where  $z_t \sim N(0, \sigma^2)$ ,  $\gamma_0$  denotes the short-term reaction of  $y_t$  to changes in  $x_t$ . This implies that in the long run elasticity between  $y$  and  $x$ , assuming  $\alpha < 1$ , is  $\beta_1 = \gamma_0 + \frac{\gamma_1}{1-\alpha}$ . However, with this form of the dynamic model, there are several potential problems including the likelihood of a high level of correlation between current and lagged values of a variable, which will result in problems such as multicollinearity, non-standard distributed parameter estimates and spurious correlation [18]. These problems could be solved by estimating the first differences of Equation (3) to obtain  $\Delta y_t = \alpha_0 + \gamma_0 \Delta x_t + \gamma_1 \Delta x_{t-1} + \alpha_1 \Delta y_{t-1} + z_t$ . This, however, introduces problems of loss of information about the long-run equilibrium. A more suitable approach is to adopt the ECM, which is set up by subtracting  $y_{t-1}$  from both sides of the short-run model in Equation (3) and further subtracting  $\gamma_0 x_{t-1} - 1$  from both sides of the resulting equation and then re-parametrizing to give

$$\begin{aligned} \Delta y_t &= \gamma_0 x_t - (1-\alpha_1) \left[ y_{t-1} - \frac{\alpha_0}{1-\alpha_1} - \frac{\gamma_0 + \gamma_1}{1-\alpha_1} x_{t-1} \right] + z_t \\ &= \gamma_0 x_t - (1-\alpha_1) [y_{t-1} - \beta_0 - \beta_1 x_{t-1}] + z_t \end{aligned} \tag{4}$$

if we take  $\beta_0 = \frac{\alpha_0}{1-\alpha_1}$  and  $\beta_1 = \frac{\gamma_0 + \gamma_1}{1-\alpha_1}$ . This representation now incorporates

both short-run and long-run effects so that should the equilibrium hold, then the term  $z_{t-1} = [y_{t-1} - \beta_0 - \beta_1 x_{t-1}]$  indicating a period of disequilibrium will equal 0. The introduction of this equilibrium error of the previous period as an explanatory variable in this representation allows movement into a new equilibrium. The term  $(1-\alpha_1)$  measures the speed of adjustment back to the long-term equilibrium depicted by Equation (2). For the system to converge to equilibrium, the coefficient of  $z_t - 1$  must be negative. If the values of  $(1-\alpha_1) \rightarrow 1$ , it indicates that economic agents eliminate large percentages of disequilibrium in each period, if  $(1-\alpha_1) \rightarrow 0$ , it indicates that adjustment is slow and values close to 2, indicate an overshooting of economic equilibrium. Positive values would imply that the system diverges from the long-run equilibrium path.  $z_t$  here is the disequilibrium error or the co-integrating residual. The expected value of  $z_t$  defines a long term equilibrium relationship between  $x_t$  and  $y_t$  and the periods of disequilibrium occur as the observed value of  $z_t$  varies around its expected value.

All terms in the ECM are stationary, so standard regression techniques are valid, assuming co-integration and that we have estimates for  $\beta_0$  and  $\beta_1$  [18]. According to Engle and Granger [14], if  $y_t, x_t \sim CI(1,1)$ , then an ECM must exist,

and conversely, the ECM generates a co-integrated series.

The ECM can also be specified in multivariate form. In order to do this, we consider vector autoregressive and moving average (VARMA) or multivariate autoregressive and moving average (MARMA) models. A general VARMA (pq) is written as

$$p_t = \phi_0 + \sum_{i=1}^p \phi_i p_{t-i} + \sum_{j=1}^q \theta_j a_{t-j} + a_t \tag{5}$$

with  $p$  and  $q \geq 0$ ,  $\phi_0 = k$ -dimensional constant vector,  $\phi_i$  and  $\theta_j = k \times k$  constant matrices and  $\{a_t\} \sim \text{i.i.d}(0, \Sigma_a)$ . Using the back shift operator, we can write  $\phi(B)p_t = \phi_0 + \theta(B)a_t$  where  $\phi(B) = I_k - \phi_1(B) - \dots - \phi_p B^p$  and  $\theta(B) = I_k - \theta_1(B) - \dots - \theta_q B^q$  are matrix polynomials in  $B$ . In a vector autoregressive (VAR) model, each variable is expressed by its own lagged values and the lagged values of all the other variables in the system. If the variables are co-integrated vector autoregressive (CVAR), we also include the co-integrating vectors that pull the entire system towards a long run equilibrium. When the variables of a VAR model are co-integrated, we use a vector error correction model (VECM).

Co-integrated variables are generally unstable in their levels, but exhibit mean-reverting “spreads” (generalized by the co-integrating relation) that force the variables to move around common stochastic trends. Modification of the VAR model to include co-integrated variables balances the short-term dynamics of the system with long-term tendencies. For the general VAR(p) model,

$$p_t = \sum_{i=1}^p \phi_i p_{t-i} + a_t, \text{ then, } \Delta p_t = p_t - p_{t-1} = \Pi p_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta p_{t-i} + a_t \text{ where}$$

$\Pi = (1 - \phi_1 - \phi_2 - \phi_3 - \dots - \phi_{p-1} - \phi_p)$  and  $\Gamma_i = -(\phi_i + \dots + \phi_p)$ . If the rank  $r$  of the matrix  $\Pi$  is zero, there is no co-integration, no stable long-run relationship between variables and VECM is not possible, only VAR in first differences, so that the equation  $\Delta p_t = \Pi p_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta p_{t-i} + a_t$  reduces to a VAR model for the

differences  $\Delta p_t = \sum_{i=1}^{p-1} \Gamma_i \Delta p_{t-i} + a_t$ . If the matrix is full rank *i.e.*  $r = k$ , then all the variables in  $p_t \sim I(0)$ . If  $0 < r < k$ , there are  $r$  co-integrating vectors describing the long-run relationships between variables. In this case VECM suffices, and the  $k \times k$  matrix  $\Pi$  can be written as

$$\Pi = \alpha\beta', \tag{6}$$

where  $\alpha$  and  $\beta$  are  $k \times r$  matrices, though this representation is of course not unique. For interpretations, it is often convenient to normalize or identify the co-integrating vectors by choosing a specific coordinate system in which to express the variables. An arbitrary normalization, suggested by Johansen [19], is to solve for the triangular representation of the co-integrated system. The elements of  $\beta p_t - 1$  may be interpreted as equilibrium relations and the elements of  $\alpha$ , as adjustment coefficients which multiply the co-integrating relationship  $\beta p_t - 1$  to help counterbalance the deviations from the equilibrium.  $\alpha$  can also be considered

a loading matrix since it determines into which equation the co-integrating vectors enter and with what magnitudes. Normally, these coefficients would be expected to be negative. Suppose  $p_t$  is  $I(1)$  with  $E[p_t] = 0$ , then Equation (6) represents a VECM and the rank  $r$  of the matrix  $\Pi$  determines the number of co-integrating relationships.

The ECM also satisfies the assumptions of classical normal linear regression model (CNLRM) which include a linear regression model, normally distributed residuals, no serial autocorrelation of residuals, and no multicollinearity. As such, diagnostic tests must be carried out to ensure these assumptions are not violated. Among the tests employed are Jarque-Bera (JB) [20] to determine the normality of the ECM, Lagrange multiplier (LM) [21] test for ARCH effects and Ljung-Box (LB) [22] for serial autocorrelation or cross correlation in the residuals.

### 2.3. Estimation and Testing for Co-Integration

Testing for co-integration is necessary in checking if the models being built are empirically meaningful. With no evidence of co-integration, time series data is considered in differences. Spurious regression can also occur when completely unrelated time series appear to be related [23]. According to [23], an  $R^2 > d$  where  $d$  is the Durbin-Watson (D-W) [24] [25] statistic can be a good signal of a spurious regression. The D-W statistic tests for autocorrelation in the residuals from a regression analysis. It lies between 0 and 4. A value of 2 implies no autocorrelation in the sample. Small values of  $d$  indicate that successive error terms are positively correlated.

Co-integration allows for the regression of one integrated series over other integrated series [26]. Testing for co-integration implies testing the existence of a long run relationship between two or more non stationary series. Because of this, it is important that a sufficiently long period of data is used, in order to detect the common long term trends among the series under consideration [13]. The two most common methods used for testing co-integration among series are the Engle-Granger (E-G) [14] test, which is based on an ordinary least squares (OLS) regression model, and [27] test, also found in [28] [29], which is based on eigenvalue analysis.

#### 2.3.1. Engle-Granger (E-G) Test for Co-Integration

The residual based E-G test [14] follows in two steps. The first step involves fitting the static OLS regression (after confirming that the series are  $I(1)$ ) which captures any potential long-run relationship between the series and then carrying out a stationarity test on the residuals of this OLS regression. The second step describes the dynamic adjustment of the series towards an equilibrium. Since the E-G method produces only one co-integrating vector, it involves pairwise comparison of two co-integrating regressions and it is affected by the choice of the dependent variable. Testing for co-integration using the E-G test is essentially equivalent to testing for unit roots in the estimated residual series us-

ing the augmented Dickey-Fuller (ADF) test. If the unit root hypothesis is rejected then  $x_t$  and  $y_t$  are co-integrated.

### 2.3.2. Johansen's Test for Co-Integration

The Johansen [27] test provides a means for testing for co-integration in a multivariate context. It allows for more than one co-integrating vector since the variables in the model might form several equilibrium relationships. The procedure builds co-integrated variables directly on maximum likelihood estimation (MLE) instead of just relying on the OLS estimates. If the variables are  $I(1)$ , the Johansen maximum likelihood procedure can be used to determine the existence of a stable long-run relationship between variables and the number of co-integrating vectors present. Generally, there will be a possible  $k - 1$  vectors, where  $k$  is the number of variables included in the model.

This test is based on the examination of the long-run coefficient matrix  $\Pi = \alpha\beta$ , described in Equation (6), so that testing for co-integration between variables is achieved by examining the rank of  $\Pi$  through the eigenvalues. These Johansen tests are the likelihood ratio test based on maximal eigenvalue of the stochastic matrix and the test based on the trace on the stochastic matrix. Before estimating the parameters of a VECM, you must choose the number of lags in the underlying VAR, the trend specification, and the number of co-integrating equations.

The basic steps in Johansen's methodology include 1) testing the order of integration of all variables, 2) setting the appropriate lag length of the model, 3) choose an appropriate model with regard the deterministic components in the multivariate system, 4) Construct likelihood ratio tests for the rank of  $\Pi$  to determine the number of co-integrating vectors, 5) impose normalization and identifying restrictions on the co-integrating vectors, and 6) given the normalized co-integrating vectors estimate the resulting co-integrated VECM by maximum likelihood (ML). The test procedure produces two statistics useful in determining the number of co-integrating vectors.

#### The trace statistic

The trace test is a test whether the rank of the matrix  $\Pi = r$ . The null hypothesis of the trace statistic is that there are no more than  $r$  co-integrating relations. Restricting the number of co-integrating equations to be  $r$  or less implies that the remaining  $k - r$  eigenvalues are zero, where  $k$  is the maximum number of possible co-integrating vectors. Johansen [27] derives the distribution of the trace statistic

$$LR_{\text{trace}} = -T \sum_{i=r+1}^k \ln(1 - \hat{\lambda}_i) \quad (7)$$

where  $T$  is the number of observations and the  $\hat{\lambda}_i$  are the  $k - r$  estimated eigenvalues. For any given value of  $r$ , large values of the trace statistic are evidence against the null hypothesis that there are  $r$  or fewer co-integrating relations in the VECM. Note: The test is not based on the trace of  $\Pi$ .



### The maximum eigenvalue statistic

To test the null hypothesis of  $r$  co-integrating vectors versus the alternative of  $(r + 1)$  co-integrating vectors the test statistic is

$$LR_{\max} = -T \ln(1 - \hat{\lambda}_{r+1}) \quad (8)$$

where  $T$  is the number of observations and  $\hat{\lambda}_r$  is the  $r^{\text{th}}$  largest canonical correlation.

### 2.4. Granger-Causality

If two or more time-series are co-integrated, then there must be Granger causality between them—either one-way or in both directions. However, the converse is not true. Co-integration usually indicates the existence of a long-run relationship between variables. Even when the variables are not co-integrated in the long-run, they could still be related in the short-run. If the prediction of one time series is improved by incorporating the knowledge of a second time series, then the latter is said to have a causal influence on the first. Granger [30] proposed a time-series data based approach that can be used to determine causality. Specifically, two autoregressive (AR) models are fitted to the first time series—with and without including the second time series—and the improvement of the prediction is measured by the ratio of the variance of the error terms. A ratio larger than one signifies an improvement, hence a causal connection. At worst, the ratio is 1 which signifies causal independence from the second time series to the first. The Granger causality test is based on a standard F-test which seeks to determine if changes in one variable cause changes in another variable.

### 3. Data and Results

This study explores the co-integration relationships and interdependencies between the Cushing OK WTI and RBOB and the number 1 heating oil traded in the NYMEX, for the period running from 2<sup>nd</sup> January 2006 to 22<sup>nd</sup> May 2015. The data was obtained from the U.S. Energy Information Administration (EIA), the principal agency of the U.S. Federal Statistical System responsible for collecting, analysing, and disseminating energy information [31]. EIA can be accessed at <http://www.eia.gov/petroleum/data.cfm>.

Spot and futures prices of crude oil and distillate fuels are non-stationary and integrated of order one [10], and therefore co-integration analysis is suitable. We use the E-G method to test for the existence of a co-integrating relationship between the six series. In this test, we take the crude futures  $CF$  series as the dependent variable and obtain the results summarised in **Table 1**.

Though the  $R^2 = 0.9994$ , the D-W statistic is 0.7545 signalling a spurious regression. The null hypothesis of no co-integration is rejected with a  $p$ -value  $< 0.001$  implying that indeed these series are co-integrated with the single co-integrating vector given by

$$\beta' = [1, 0.986809, -0.02571, 0.001709, -0.027940, 0.035578] \quad (9)$$

and the long-run relationship given by

$$CF = 0.5118 + 0.986809CS - 0.02571GF + 0.001709GS - 0.027940HF + 0.035578HS \tag{10}$$

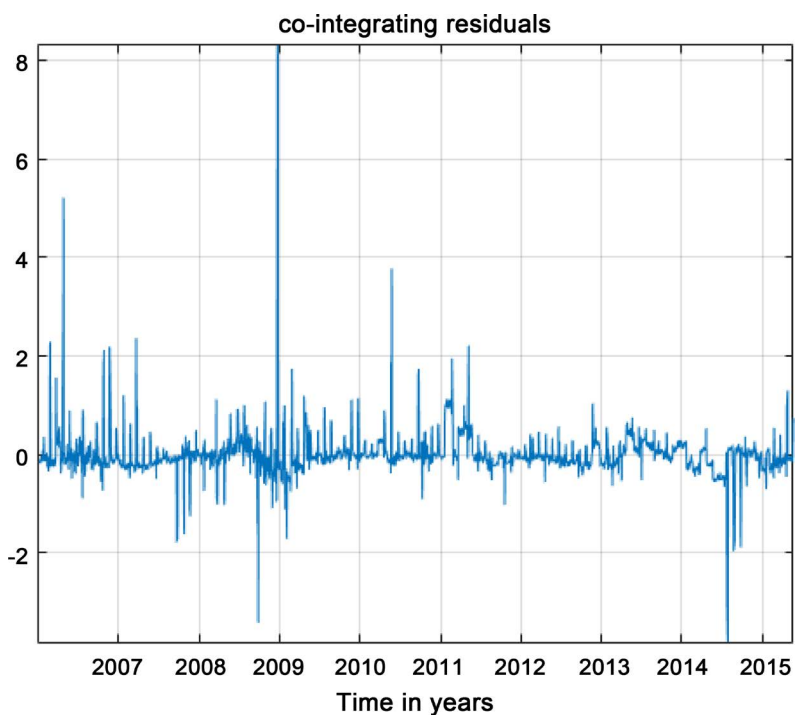
The co-integrating residuals which give the co-integrating relation are given by the representation

$$\hat{a}_t = CF - 0.5118 - 0.986809CS + 0.02571GF - 0.001709GS + 0.027940HF - 0.035578HS, \tag{11}$$

and the plot of the co-integrating relations (the error correction term), which depicts stationarity, is shown in **Figure 1**. Testing for stationarity using the ADF test gives a  $p$ -value  $< 0.001$  indicating we reject the null hypothesis of existence of unit roots. However, as has been discussed previously, one of the main drawbacks of the E-G technique of testing for co-integration is that it identifies only a

**Table 1.** Results obtained from the E-G test.

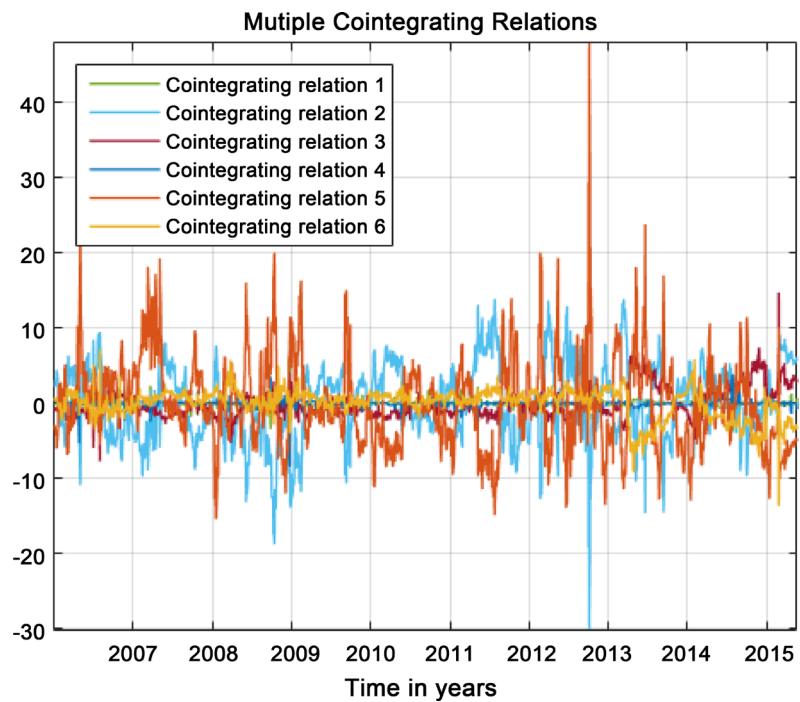
Parameter	Value	$t$ -statistic	$p$ -value	SE
$\alpha_0$	0.511765	10.69652	0.0000	0.047844
$\alpha_1$	0.986809	625.7656	0.0000	0.001577
$\alpha_2$	-0.002571	-1.251454	0.2109	0.002055
$\alpha_3$	0.001709	1.004433	0.3153	0.001701
$\alpha_4$	-0.027940	-5.154723	0.0000	0.005420
$\alpha_5$	0.035578	6.805134	0.0000	0.005228



**Figure 1.** Co-integrating residuals for the co-integrating relationship in the six series taking crude futures prices as the dependent variable.

single co-integrating relation, among what might be many such relations. When we consider more than two series, we will have a different co-integrating relationship for every dependent variable specified as shown in **Figure 2** where each of the six price series under consideration is taken as the dependent variable. These relations are not unique and change according to how the dependent variables are chosen. In all the six relationships, we reject the null hypothesis of no co-integration with  $p$ -values  $< 0.001$  in all the six instances indicating six co-integrating relations.

For the Johansen co-integration trace test, we examine whether the rank of the matrix  $\Pi = r$ . As such, testing proceeds sequentially for  $r = 1, 2, \dots, k$  and the first non-rejection of the null is taken as an estimate of smallest lag for which we obtain a rank of  $r$ . In our case this happens for  $r = 4$  at 44, making this the optimal lag length. The trace test indicates 4 lags with a  $p$ -value of 0.0746 as shown in **Table 2**. Lag 4 is the co-integrating equations at the  $\alpha = 0.05$  level and \* denotes



**Figure 2.** Six co-integrating relations obtained when all six series are considered as dependent variables.

**Table 2.** Results obtained from Johansen co-integration test.

Rank	Trace statistic	Critical value	$p$ -value	Eigenvalue
0*	444.1103	95.7541	0.0010	0.1150
1*	155.9466	69.8187	0.0010	0.0376
2*	65.5717	47.8564	0.0010	0.0132
3*	34.3072	29.7976	0.0143	0.0084
4	14.3214	15.4948	0.0746	0.0044
5*	3.9785	3.8415C	0.0461	0.0017

rejection of the null hypothesis at the 0.05 level.

The co-integrating relations are as shown in **Figure 3** and the resultant parameters from MLE are the co-integrating speed parameters

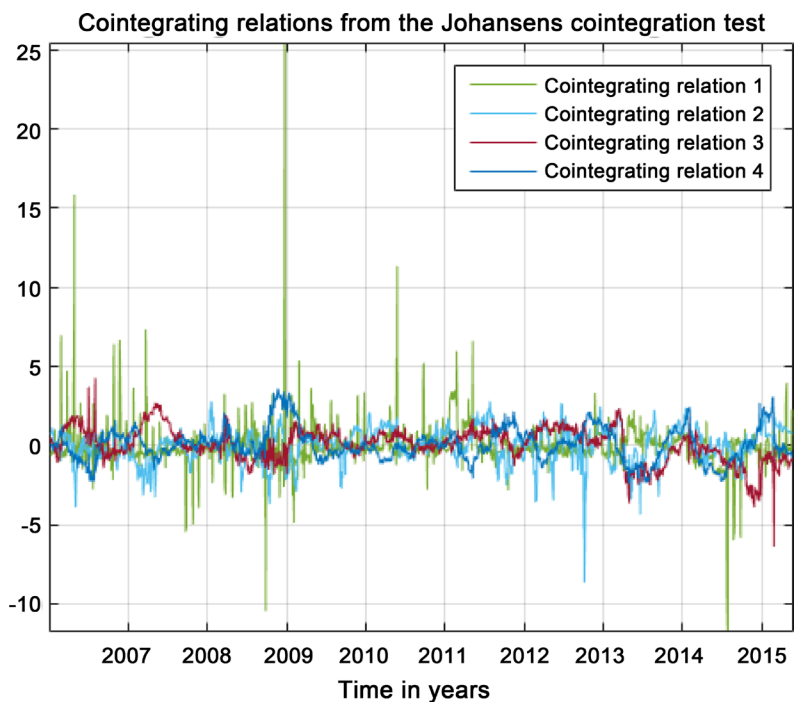
$$\alpha = \begin{pmatrix} 0.0991 & 0.0008 & 0.1401 & 0.0100 \\ 0.1803 & -0.0336 & 0.1195 & 0.1037 \\ 0.0847 & 0.0396 & 0.1639 & -0.0170 \\ 0.2289 & 0.0022 & 0.1364 & 0.0050 \\ 0.1508 & 0.4087 & 0.0775 & 0.0913 \\ 0.1001 & 0.0305 & 0.1070 & -0.0324 \end{pmatrix}$$

and the vector of co-integrating relations

$$\beta = \begin{pmatrix} 3.0080 & 0.1463 & -0.1334 & -0.1482 \\ -0.0017 & 0.1818 & 0.0643 & -0.0817 \\ 0.0870 & -0.1898 & -0.4645 & -0.1300 \\ -2.9847 & -0.1302 & 0.0974 & 0.0821 \\ 0.0066 & -0.1813 & 0.0136 & -0.0022 \\ -0.1008 & 0.1706 & 0.4150 & 0.2435 \end{pmatrix}$$

so that for most cases, the  $6 \times 6$  matrix  $\Pi$  represented in Equation (6) would be

$$\Pi = \alpha\beta' = \begin{pmatrix} 0.2782 & 0.0082 & -0.0579 & -0.2816 & 0.0024 & 0.0507 \\ 0.5062 & -0.0072 & -0.0469 & -0.5137 & 0.0087 & 0.0509 \\ 0.2412 & 0.0190 & -0.0741 & -0.2434 & -0.0043 & 0.0621 \\ 0.6699 & 0.0084 & -0.0445 & -0.6698 & 0.0029 & 0.0351 \\ 0.4897 & 0.0715 & -0.1123 & -0.4884 & -0.0722 & 0.1089 \\ 0.2962 & 0.0149 & -0.0426 & -0.2951 & -0.0033 & 0.0317 \end{pmatrix} \quad (12)$$



**Figure 3.** Four co-integrating relations obtained using the Johansen’s trace test.

$$\Delta p_t = \Gamma_0 + \sum_{i=1}^4 \Gamma_i \Delta p_{t-i} + \alpha (\beta' p_{t-1} + \delta) + a_t \tag{13}$$

where  $\alpha$  and  $\beta$  are given as above,  $p_t' = [CF_t, CS_t, GF_t, GS_t, HF_t, HS_t]$ ,  $\delta' = [-1.4215, 1.6461, 0.9019, 2.5075]$  are intercepts in the co-integrating relations,  $\Gamma_0' = [-0.0045, 0.0017, 0.0021, -0.0009, -0.0002, 0.0021]$ , and  $a_t$  are the shocks or innovations associated with each of the relations,

$$\Gamma_1 = \begin{pmatrix} -0.3080 & 0.0253 & -0.1005 & 0.2390 & -0.0317 & 0.1120 \\ -0.5324 & 0.0092 & -0.1000 & 0.4972 & -0.0102 & 0.0732 \\ -0.2669 & 0.0049 & -0.3733 & 0.2746 & -0.0364 & 0.3705 \\ -0.3995 & 0.0306 & -0.1024 & 0.3167 & -0.0291 & 0.1200 \\ -0.3504 & 0.0815 & 0.1126 & 0.3070 & -0.0287 & -0.1938 \\ -0.2601 & 0.0460 & 0.1269 & 0.2649 & -0.0490 & -0.1698 \end{pmatrix}$$

$$\Gamma_2 = \begin{pmatrix} -0.1898 & 0.0159 & -0.1731 & 0.1933 & 0.0085 & 0.1007 \\ -0.4415 & -0.0219 & -0.2134 & 0.4629 & 0.0007 & 0.1372 \\ -0.1942 & 0.0021 & -0.3479 & 0.2948 & 0.0129 & 0.2319 \\ -0.2013 & 0.0159 & -0.1646 & 0.1999 & 0.0075 & 0.0996 \\ -0.1517 & 0.0158 & -0.1349 & 0.2206 & -0.0216 & 0.0236 \\ -0.1870 & 0.0092 & -0.1063 & 0.2752 & 0.0168 & -0.0145 \end{pmatrix}$$

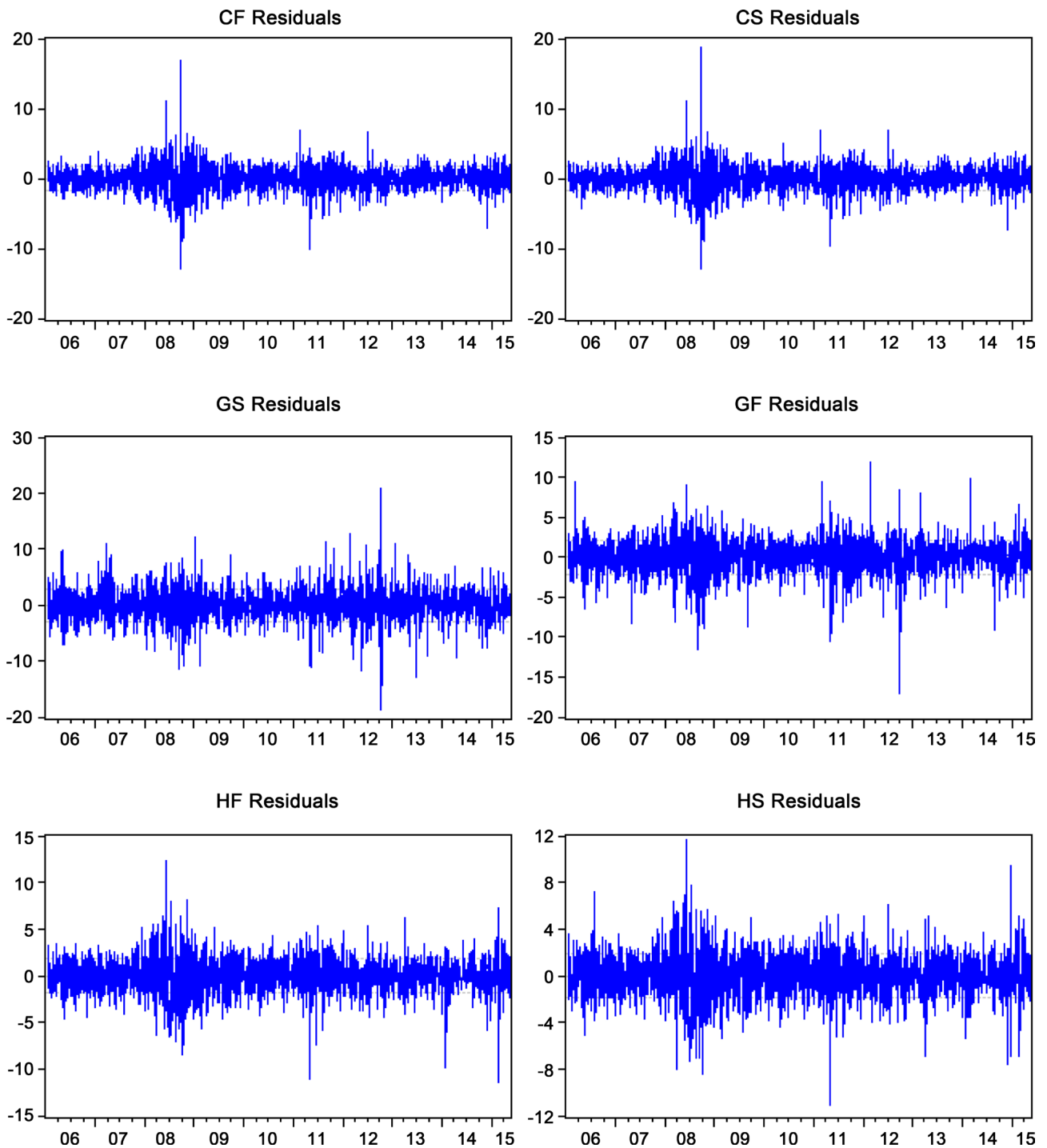
$$\Gamma_3 = \begin{pmatrix} -0.2912 & -0.0254 & -0.0248 & 0.3225 & 0.0006 & 0.0213 \\ -0.4268 & -0.0866 & -0.0535 & 0.5186 & 0.0095 & 0.0366 \\ -0.2768 & -0.0503 & -0.2320 & 0.3808 & 0.0160 & 0.1827 \\ -0.1805 & -0.0330 & -0.0340 & 0.2683 & -0.0029 & 0.0077 \\ -0.2061 & -0.1020 & -0.0417 & 0.3759 & 0.0643 & -0.0527 \\ -0.3183 & -0.0349 & -0.0491 & 0.4017 & 0.0128 & 0.0058 \end{pmatrix}$$

and

$$\Gamma_4 = \begin{pmatrix} -0.0854 & 0.0188 & -0.0183 & 0.1668 & -0.0166 & -0.0293 \\ -0.0418 & -0.0058 & -0.0020 & 0.1972 & -0.0123 & -0.0532 \\ -0.2077 & -0.0020 & -0.0483 & 0.2911 & -0.0197 & 0.0685 \\ -0.0743 & 0.0190 & -0.0155 & 0.1458 & -0.0163 & -0.0289 \\ 0.0470 & -0.0729 & 0.0273 & 0.1029 & 0.0394 & -0.0047 \\ -0.0756 & 0.0060 & 0.0202 & 0.1828 & -0.0084 & -0.0508 \end{pmatrix}$$

$\alpha (\beta' \Delta p_{t-1} + \delta)$  represents the error correction term.

After fitting this model, an analysis of the residuals show that they are not white noise. **Figure 4** shows plots for the residuals from the VECM. They do not seem to be multivariate normal, and this is confirmed by the results shown on **Table 3** which reports the multivariate extensions of the JB residual normality test, which is a test of the hypothesis that the skewness of the underlying distribution is zero, and the kurtosis is three. This test also compares the third and fourth moments of the residuals to those from the normal distribution. These two measures allow one to test the hypotheses that are compatible with the assumption of multivariate normality.



**Figure 4.** Residuals from the full VECM.

This test is applicable since individual variables from a set of variables that are jointly multinomial distributed are also normally distributed, although if a number of variables are normally distributed individually, they are not necessarily also multivariate normal [32]. The test statistic is given by equation (14) below where  $S^*$  and  $K^*$  represent the sample skewness and sample kurtosis respectively.

$$JB = n \left\{ \frac{\hat{S}^2}{6} + \frac{(\hat{K} - 3)^2}{24} \right\} \tag{14}$$

The null hypothesis for this test is that the residuals are multivariate normal and this is rejected as can be seen from the *p*-values reported in **Table 3**.

A visual inspection of **Figure 4** suggests the presence of volatility clustering. Further investigation of the residual series reveals the presence of arch effects as shown in **Table 4**. The null hypothesis of no ARCH effects is rejected. The noise is therefore MGARCH.

For this study, the results for Granger-causality in levels are reported in **Table 5** and results for causality in differences are recorded in **Table 6**.

From **Table 5**, we see bidirectional causality between almost all the series except for gasoline futures which doesn't Granger cause gasoline spot though gasoline spot Granger causes gasoline futures. This result is confirmed by the joint

**Table 3.** Multivariate normality test on residual series.

Series	Skewness	$\chi^2$ statistic	df	<i>p</i> -value
CF	0.056937	1.274020	1	0.2590
CS	-4.240484	7066.810	1	0.0000
GF	-0.589093	136.3830	1	0.0000
GS	0.592101	137.7792	1	0.0000
HF	-1.138971	509.8211	1	0.0000
HS	-0.125048	6.145373	1	0.0132
		7858.213	6	0.0000
Series	Kurtosis	$\chi^2$ statistic	df	<i>p</i> -value
CF	10.21375	5112.752	1	0.0000
CS	101.6174	955,519.8	1	0.0000
GF	17.20323	19,820.15	1	0.0000
GS	15.34322	14,968.89	1	0.0000
HF	24.22178	44,248.25	1	0.0000
HS	31.65732	80,687.03	1	0.0000
		1,120,357	6	0.0000
Series	JB statistic	df	<i>p</i> -value	
CF	5114.026	2	0.0000	
CS	962,586.6	2	0.0000	
GF	19,956.53	2	0.0000	
GS	15,106.67	2	0.0000	
HF	44,758.07	2	0.0000	
HS	80,693.18	2	0.0000	
Joint	1,128,215	12	0.0000	

**Table 4.** Test for ARCH effects on residual series.

Joint test:					
$\chi^2$ statistic	df	<i>p</i> -value			
6278.398	1176	0.0000			
Individual components:					
Dependent	$R^2$	$F(56, 2301)$	<i>p</i> -value	$\chi^2(56)$	<i>p</i> -value
res1 * res1	0.233479	12.51559	0.0000	550.5426	0.0000
res2 * res2	0.211971	11.05254	0.0000	499.8271	0.0000
res3 * res3	0.085742	3.853501	0.0000	202.1805	0.0000
res4 * res4	0.260093	14.44374	0.0000	613.2987	0.0000
res5 * res5	0.142756	6.842568	0.0000	336.6190	0.0000
res6 * res6	0.100405	4.586018	0.0000	236.7544	0.0000
res2 * res1	0.222538	11.76124	0.0000	524.7442	0.0000
res3 * res1	0.145531	6.998214	0.0000	343.1617	0.0000
res3 * res2	0.146352	7.044489	0.0000	345.0987	0.0000
res4 * res1	0.134115	6.364236	0.0000	316.2435	0.0000
res4 * res2	0.131406	6.216239	0.0000	309.8558	0.0000
res4 * res3	0.097981	4.463300	0.0000	231.0398	0.0000
res5 * res1	0.127599	6.009785	0.0000	300.8780	0.0000
res5 * res2	0.126302	5.939869	0.0000	297.8197	0.0000
res5 * res3	0.133311	6.320207	0.0000	314.3473	0.0000
res5 * res4	0.119110	5.555916	0.0000	280.8617	0.0000
res6 * res1	0.100783	4.605250	0.0000	237.6472	0.0000
res6 * res2	0.100042	4.567614	0.0000	235.8994	0.0000
res6 * res3	0.101842	4.659118	0.0000	240.1439	0.0000
res6 * res4	0.107252	4.936366	0.0000	252.9014	0.0000
res6 * res5	0.100744	4.603270	0.0000	237.5553	0.0000

test also. This actually shows the presence of spillover effects across the series. From **Table 6**, we see, returns from gasoline futures do not Granger cause gasoline spot returns and we also note that gasoline spot returns only Granger causes heating oil futures and maybe heating oil spot returns.

#### 4. Discussion

In this study, both the E-G test [14] and Johansen's test [27] are used to explore the presence of co-integrating relations between the daily prices of crude oil and distillate fuels. Given the multivariate nature of the data under consideration, Johansen's test becomes more efficient given its ability to test for co-integration



**Table 5.** Results obtained from Granger-causality ( $F$ ) test in levels.

	CF	CS	GF	GS	HF	HS	Joint
CF	2356 (<0.001)	254 (<0.001)	239 (<0.001)	195 (<0.001)	1044 (<0.001)	951 (<0.001)	1819.2 (<0.001)
CS	321 (<0.001)	382 (<0.001)	293 (<0.001)	159 (<0.001)	833 (<0.001)	732 (<0.001)	1564 (<0.001)
GF	40 (<0.001)	46 (<0.001)	33369 (<0.001)	6 <b>(0.2132)</b>	54 (<0.001)	40 (<0.001)	146.6 (<0.001)
GS	25 (<0.001)	19 (<0.001)	49 (<0.001)	6077 (<0.001)	36 (<0.001)	34 (<0.001)	103 (<0.001)
HF	61 (<0.001)	125 (<0.001)	110 (<0.001)	109 (<0.001)	13655 (<0.001)	148 (<0.001)	458.9 (<0.001)
HS	102 (<0.001)	122 (<0.001)	78 (<0.001)	104 (<0.001)	48 (<0.001)	14178 (<0.001)	455.3 (<0.001)

$p$ -values are indicated in the parenthesis.

**Table 6.** Results obtained from Granger-causality ( $F$ ) test in differences.

	$\Delta$ CF	$\Delta$ CS	$\Delta$ GF	$\Delta$ GS	$\Delta$ HF	$\Delta$ HS	Joint
$\Delta$ CF	51.5493 (<0.001)	81.5114 (<0.001)	158.8971 (<0.001)	231.7504 (<0.001)	871.1525 (<0.001)	598.9422 (<0.001)	1463.8 (<0.001)
$\Delta$ CS	37.6260 (<0.001)	28.3827 (<0.001)	210.8375 (<0.001)	176.4722 (<0.001)	857.9510 (<0.001)	579.0799 (<0.001)	1212.4 (<0.001)
$\Delta$ GF	23.9313 (<0.001)	32.1628 (<0.001)	22.5417 (<0.001)	3.8876 <b>(0.4214)</b>	51.2044 (<0.001)	30.3673 (<0.001)	175 (<0.001)
$\Delta$ GS	6.3281 <b>(0.176)</b>	2.8095 <b>(0.5902)</b>	13.2817 <b>(0.1)</b>	6.7035 <b>(0.1524)</b>	21.7478 (<0.001)	14.5890 <b>(0.0056)</b>	72.2 (<0.001)
$\Delta$ HF	61.3676 (<0.001)	192.8939 (<0.001)	45.3097 (<0.001)	122.6379 (<0.001)	173.5127 (<0.001)	112.1761 (<0.001)	498.4 (<0.001)
$\Delta$ HS	34.5446 (<0.001)	90.3696 (<0.001)	61.5954 (<0.001)	114.1517 (<0.001)	30.1818 (<0.001)	27.3170 (<0.001)	451.6 (<0.001)

$p$ -values are indicated in the parenthesis.

in a multivariate context.

Under the E-G test, the null hypothesis of no co-integration is rejected and this result is backed by the DW statistic. Similarly, Johansen's test results reveal up to 4 co-integrating vectors with an optimal lag length of 4 if all six series are considered simultaneously. After fitting the full VECM, a residual analysis is carried out and it suggests some heteroscedasticity in the noise process, contravening the Gaussian assumption on the residuals. Tests reveal ARCH effects in

the residuals, and the JB hypothesis of multivariate normality is rejected. This suggests a VECM with GARCH errors could have been better. For noise that is MGARCH, fitting a VEC-GARCH. This involves writing the VEC part as a seemingly unrelated regression (SUR) model and combining to obtain SUR-GARCH model. That is however not covered in this study.

Causal relationships are also explored and bidirectional Granger-causality is exhibited for all cases except for the gasoline futures which do not seem to Granger cause gasoline spot prices. This underscores the importance of the spill-over effects in the volatilities of these price series. It confirms the presence of cointegrating relations and hence spill-over effects.

## References

- [1] Reinsel, G.C. (1993) Elements of Multivariate Time Series Analysis. Springer US. <https://doi.org/10.1007/978-1-4684-0198-1>
- [2] Alexander, C., Giblin, I. and Weddington, W. (2002) Co-Integration and Asset allocation: A New Active Hedge Fund Strategy. ISMA Centre Discussion Papers in Finance Series.
- [3] Chan, N.H. (2011) Time Series: Applications to Finance with R and S-Plus. John Wiley & Sons, Hoboken.
- [4] Girma, P.B. and Paulson, A.S. (1999) Risk Arbitrage Opportunities in Petroleum Futures Spreads. *Journal of Futures Markets*, **19**, 931-955. [https://doi.org/10.1002/\(SICI\)1096-9934\(199912\)19:8<931::AID-FUT5>3.0.CO;2-L](https://doi.org/10.1002/(SICI)1096-9934(199912)19:8<931::AID-FUT5>3.0.CO;2-L)
- [5] Engle, R.F. (1982) Autoregressive Conditional Heteroscedasticity with Estimates of the variance of United Kingdom Inflation. *Econometrica*, **50**, 987-1007. <https://doi.org/10.2307/1912773>
- [6] Bollerslev, T. (1986) Generalized Autoregressive Conditional Heteroscedasticity. *Journal of Economics*, **31**, 307-327. [https://doi.org/10.1016/0304-4076\(86\)90063-1](https://doi.org/10.1016/0304-4076(86)90063-1)
- [7] Haigh, M.S. and Holt, M.T. (2002) Crack Spread Hedging: Accounting for Time-Varying Volatility Spill Overs in the Energy Futures Markets. *Journal of Applied Econometrics*, **17**, 269-289. <https://doi.org/10.1002/jae.628>
- [8] Veiga, B. and McAleer, M. (2004) Multivariate Volatility and Spillover Effects in Financial Markets. <https://scholarsarchive.byu.edu/iemssconference/2004/all/26/>
- [9] Hoti, S., Chan, F., McAleer, M., et al. (2002) Structure and Asymptotic Theory for Multivariate Asymmetric Volatility: Empirical Evidence for Country Risk Ratings. *Australasian Meeting of the Econometric Society*, Brisbane. <https://core.ac.uk/download/pdf/6390132.pdf>
- [10] Aduda, J., Weke, P., Ngare, P. and Mwaniki, J. (2016) Financial Time Series Modelling of Trends and Patterns in the Energy Markets. *Journal of Mathematical Finance*, **6**, 324-337. <https://doi.org/10.4236/jmf.2016.62027>
- [11] The MathWorks (2013) MATLAB Version 7.14.0.739 R2013a. The MathWorks Inc., Torrance.
- [12] Rachev, S.T., Mittnik, S., Fabozzi, F.J., Focardi, S.M. and Jašić, T. (2007) Financial Econometrics: From Basics to Advanced Modeling Techniques. John Wiley & Sons, Hoboken.
- [13] Alexander, C. (2008) Practical Financial Econometrics. Wiley, Hoboken.
- [14] Engle, R.F. and Granger, C.W.J. (1987) Co-Integration and Error Correction: Re-

- presentation, Estimation, and Testing. *Econometrica: Journal of the Econometric Society*, **55**, 251-276. <https://doi.org/10.2307/1913236>
- [15] Lütkepohl, H. (2007) New Introduction to Multiple Time Series Analysis. Springer Science & Business Media, Berlin.
- [16] Tsay, R.S. (2013) Multivariate Time Series Analysis: With R and Financial Applications. Wiley, Hoboken.
- [17] Enders, W. (2008) Applied Econometric Time Series. John Wiley & Sons, Hoboken.
- [18] Harris, R.I.D. and Sollis, R. (2003) Applied Time Series Modelling and Forecasting. John Wiley & Sons, Hoboken.
- [19] Johansen, S. (1995) Identifying Restrictions of Linear Equations with Applications to Simultaneous Equations and Co-Integration. *Journal of Econometrics*, **69**, 111-132. [https://doi.org/10.1016/0304-4076\(94\)01664-L](https://doi.org/10.1016/0304-4076(94)01664-L)
- [20] Jarque, C.M. and Bera, A.K. (1987) A Test for Normality of Observations and Regression Residuals. *International Statistical Review/Revue Internationale de Statistique*, **55**, 163-172.
- [21] Engle, R.F. (1984) Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics. In: *Handbook of Econometrics*, Vol. 2, North Holland, New York, 775-826.
- [22] Ljung, G.M. and Box, G.E.P. (1978) On a Measure of Lack of Fit in Time Series Models. *Biometrika*, **65**, 297-303. <https://doi.org/10.1093/biomet/65.2.297>
- [23] Granger, C.W.J. and Newbold, P. (1974) Spurious Regressions in Econometrics. *Journal of Econometrics*, **2**, 111-120. [https://doi.org/10.1016/0304-4076\(74\)90034-7](https://doi.org/10.1016/0304-4076(74)90034-7)
- [24] Durbin, J. and Watson, G.S. (1950) Testing for Serial Correlation in Least Squares Regression: I. *Biometrika*, **37**, 409-428.
- [25] Durbin, J. and Watson, G.S. (1951) Testing for Serial Correlation in Least Squares Regression. II. *Biometrika*, **38**, 159-178. <https://doi.org/10.1093/biomet/38.1-2.159>
- [26] Fabozzi, F.J., Focardi, S.M. and Kolm, P.N. (2006) Financial Modelling of the Equity Market: From CAPM to Co-Integration. Wiley, Hoboken.
- [27] Johansen, S. (1988) Statistical Analysis of Co-Integration Vectors. *Journal of Economic Dynamics and Control*, **12**, 231-254. [https://doi.org/10.1016/0165-1889\(88\)90041-3](https://doi.org/10.1016/0165-1889(88)90041-3)
- [28] Johansen, S. and Juselius, K. (1990) Maximum Likelihood Estimation and Inference on Cointegration with Applications to the Demand for Money. *Oxford Bulletin of Economics and Statistics*, **52**, 169-210. <https://doi.org/10.1111/j.1468-0084.1990.mp52002003.x>
- [29] Johansen, S. (1991) Estimation and Hypothesis Testing of Co-Integration Vectors in Gaussian Vector Autoregressive Models. *Econometrica*, **59**, 1551-1580. <https://doi.org/10.2307/2938278>
- [30] Granger, C.W.J. (1969) Investigating Causal Relations by Econometric Models and Cross-Spectral Methods. *Econometrica*, **37**, 424-438. <https://doi.org/10.2307/1912791>
- [31] EIA (2015) U.S. Energy Information Administration. <http://www.eia.gov/about/>
- [32] Von Eye, A. and Bogat, A.G. (2004) Testing the Assumption of Multivariate Normality. *Psychology Science*, **46**, 243-258.