Uppsala University
Department of Statistics

On Statistical Arbitrage:
Cointegration and Vector Error-Correction in the Energy Sector

BY OSCAR NILSSON* AND EMMANUEL LATIM OKUMU†
SUPERVISOR: LARS FORSBERG
Spring 2014

Abstract

This paper provides methods to select pairs potentially profitable within the frame of statistical arbitrage. We employ a cointegration approach on pairwise combinations of five large energy companies listed on the New York Stock Exchange for the period 27th September 2012 to 22nd April 2014. We find one cointegrated pair, for which we further investigate both short and long run dynamics. A vector-error correction model is constructed, supporting a long run relationship between the two stocks, which is also supported by the mean-reverting characteristic of a stationary linear combination of the stocks. Impulse response functions and variance decomposition are also studied to further describe the interrelation of the stocks, supporting a unidirectional causality between the stocks.

Keywords: Mean-reversion, VAR, VECM, impulse response, cointegration

We would like to thank our supervisor, Lars Forsberg, for his support and advice throughout this work.

* Oscar.nilsson.4839@student.uu.se
† Emmanuel.okumu.9117@student.uu.se

Table of contents

Introduction

In the early 80's, Clive Granger (1981) introduced the concept of cointegration. The basic idea is that two variables that each are integrated of some order p may form a linear combination that is integrated of a lower order. In its simplest form, an example is the case where two variables each follow a random walk, but a linear combination of the two is stationary. The stationarity of the combination implies that the variables conform to a long-run equilibrium, and that deviations from the equilibrium are merely temporary. In economics, cointegration is commonly used to describe the relationship between related variables such as short and long interest rates, income and consumption or in the theory of Purchasing Power Parity. This aspect of long-run equilibrium has also gained attention in the financial industry, and the term statistical arbitrage refers to the profitable potential of trading a cointegrated pair of stocks. Consider two securities that are cointegrated and therefore expected to move together in the long run. If any of the securities were to deviate from this co-movement, it would suggest a temporary mispricing. We might not know which security that is actually mispriced, but we know that their relative prices are incorrect. One could then take a short position (sell) in the relatively overvalued stock, and a long position (buy) in the relatively undervalued stock. When the securities return to their long-run equilibrium, the positions are unwinded to a profit. When employing this strategy, it is desirable to enter the positions when the deviation from the equilibrium is at its largest. This is of course unknown beforehand, however, the larger the relative mispricing is when the positions are taken, the larger the potential profit.

Do, Faff and Hamza (2006) examine three pairs of stocks based on industry similarity, seeking to find a pair for which a long-short positioning would be profitable when the pair diverges from its long-run relationship. Because of the proprietary nature of the subject, literature is rather limited, and they therefore present several techniques that may be applied to the study. Through their analysis they find cointegrated stocks, but because of an increasing divergence in the relationship, the pairs would be considered to risky to take action on. Gatev et al. (2006) investigates a large sample of stocks for the period of 1962 – 2002. Employing a minimum distance method when observing normalized prices of stocks that seem to move together, they find that pairs trading is a profitable strategy. The idea of pairs trading can be applied to other assets than stocks, Jacobsen (2008) examines the relationship between two exchange-traded funds (ETF) with large overlap in coverage. Using intraday data (minute by minute) for a period of 3 months and taking transaction costs into account he finds that the strategy is very profitable compared to a buy & hold strategy for the two assets. To assess this relationship he employs the Johansen's test and the Engle and Granger approach, emphasizing the importance of the error-correction model in order to study the short-run and long-run dynamics. The error-correction representation is also adopted by Kniff and Pynnönen (1999) when estimating a VAR for the cross-dependency of eleven different stock market indices. Finding that the Swedish index and the Norwegian index are cointegrated they construct a vector error-correction model to further analyse the relationship.

This paper will aim to describe a methodology of how to find a cointegrated pair of stocks that could be profitable to trade. The search for cointegrated pairs is commonly encouraged to take place between assets that are related by industry and other characteristics that both can support any fundamental relationship and that the cointegration remains significant out-of-sample (Vidyamurthy, 2004). In this paper, five different energy companies listed on the New York Stock Exchange are examined. Data are collected for the period 27 September 2012 to 22 April 2014, and the properties of the time series are inspected. We then test the ten possible pairwise combinations for cointegration by the Johansen's test and the Engle and Granger approach. Of the ten pairs, we find one to be cointegrated over the entire sample period. We investigate the cointegrating relationship by constructing a Vector Error-Correction Model and conduct further analysis of the relationship through impulse response functions and variance decomposition.

The rest of the paper is organized as follows. In section 2 we describe the theoretical framework of cointegration and also explain the error-correction representation. The following section presents the Johansen's test and the Engle & Granger approach, which both will be applied in study. Impulse response functions and variance decomposition are covered and a presentation of the data employed in the study is also given. Section 4 summarizes our results and our conclusions are presented in Section 5.

Theory

Stationarity and Unit root processes

A key concept underlying the idea behind cointegration is stationarity. Stationarity refers to a time series that exhibits certain characteristics that are of great importance to econometric analysis. Distinction is made between strong stationarity and weak stationarity. The latter, also referred to as covariance stationarity, will be of interest throughout the paper, given that strong stationarity is difficult to assess. For a time series to be covariance stationary it requires the following characteristics:

- $E(y_t) = E(y_{t-k}) = \mu$
- $E(y_t - \mu)^2 = E(y_{t-k} - \mu)^2 = \sigma^2$
- $E(y_t - \mu)(y_{t-k} - \mu) = E\big(y_{t-j} - \mu\big)\big(y_{t-j-k} - \mu\big) = \gamma_k.$

The first condition implies a constant mean over time, this ensures that deviations from the mean are temporary and the series is expected to revert back to the long run mean. The second condition concerns the second moment, the variance, which should be finite and constant over time. Lastly, the covariance between different variables is only dependent on the lag length, not on time. A time series generated from a simple AR(1) process, $Y_t = \phi Y_{t-1} + u_t$, is stationary for values of $|\phi| < 1$. The model fulfils the three conditions with an expected mean of zero, a variance of $\sigma_e^2/(1-\phi^2)$ and the covariance $\phi^k\big(\sigma_e^2/(1-\phi^2)\big)$. To extend the above example, the general ARMA(p, q) is given by,

$$Y_t = \sum_{i=1}^{p} \phi_1 Y_{t-i} + u_t + \sum_{i=1}^{q} \theta_1 u_{t-i},$$

using lag operators,

$$\big(1 - \phi_1 L - \cdots - \phi_p L^p\big) Y_t = \big(1 - \theta_1 L - \cdots - \theta_q L^q\big) u_t$$

$$\Phi(L) Y_t = \Theta(L) u_t.$$

Since the ARMA(p, q) is made up of an AR(p) and a MA(q), and the MA(q) is always stationary, the stationarity of the model is dependent on the AR(p)-part. If all the roots of the characteristic equation $\Phi(z) = 0$ are outside the unit circle, then the model is stationary. The ARMA(p, q) is a case of the ARIMA(p, d, q), where I(d), the order of integration, is zero. A time series that is integrated, but stationary when taking the difference *d* times, is integrated of order d. To illustrate the case of an integrated series we can return to the AR(1) model but we let $\phi = 1$. Then the model is simply $Y_t = Y_{t-1} + u_t$, a random walk, and it is no longer stationary as it contains a unit root. The series will have an infinite memory as shocks to the series will have permanent effect, and the variance will grow to infinity as time goes to infinity. (Asteriou & Hall, 2011) Differencing the series once gives $\Delta Y_t = u_t$, simply stationary white noise. In economics it is not unusual to study series that contain unit root, therefore it is common practice to difference the series until stationarity is obtained before conducting regression analysis. If this is ignored, the researcher will likely obtain spurious results, meaning they have no actual meaning at all. It is however not guaranteed that all integrated series can be differenced to become stationary. Stationarity and order of integration play an important role in the concept of cointegration, which will be discussed next.

Cointegration

If conducting regression analysis using integrated series, the researcher runs the risk of obtaining spurious results and integrated residual series that invalidates the estimates. The concept of cointegration was introduced by Granger (1981), the very special case when a linear combination of series that are integrated of the same order, is integrated by a lower order. That is, if two series $\{X, Y\}$ are $I(d)$, but a linear combination of the series is $I(d-b)$, where $d \geq b > 0$, then the series are cointegrated, presented as $\{X, Y\} \sim CI(d, b)$. More generally, let $\boldsymbol{z}_t$ be a $(n * 1)$ vector of series $z_{1t}, z_{2t}, \dots, z_{nt}$ and that each series is $I(d)$. If there exists a $(n * 1)$ vector $\boldsymbol{\beta}$ that gives $\boldsymbol{z}_t' \boldsymbol{\beta} \sim I(d-b)$, we have that $\boldsymbol{z}_t \sim CI(d, b)$. When the series are cointegrated, the analysis of their relationship is not spurious, and no differencing of the variables to reach stationarity is needed before examining their relationship. (Asteriou & Hall, 2011)

Consider the case where $\{X\}$ and $\{Y\}$ both are $I(1)$, but a vector $\{\theta_1, \theta_2\}$ exists that give $\theta_1 Y_t + \theta_2 X_t = e_t$ that is $I(0)$. The series are then cointegrated, resulting in a stationary linear combination. This implies a long-run equilibrium around which the variables are allowed to fluctuate. The shocks that may affect the equilibrium are merely temporary, since the stationarity ensures mean-reversion that restores the long-run equilibrium. Given that our variables are cointegrated it is of interest to study the long-run relation of the variables. The error-correction model (ECM) is a convenient reparametrization of the autoregressive distributed lag model (ADL). The ECM allows us to study both short-run and long-run effects and its popularity is extensive in applied time series econometrics. The ECM is presented in the Granger Representation Theorem by Engle and Granger (1987) for cointegrated series. A general reparametrization of the ADL to the ECM is given below and then a corresponding approach for the multivariate case. The ADL model for Y and X with n and m lags respectively is given in Equation 1 (Asteriou & Hall, 2011).

$$Y_t = \mu + \sum_{i=1}^{n} a_i Y_{t-i} + \sum_{i=0}^{m} \gamma_i X_{t-i} + u_t \qquad 1$$

In the long run, for two cointegrated variables, the relationship can be expressed as,

$$Y^* = \beta_0 + \beta_1 X^*, \qquad 2$$

assuming that $X^* = X_t = X_{t-1} = \cdots = X_{t-m}$ and rearranging the terms,

$$Y^* = \frac{\mu}{1-\sum a_i} + \frac{\sum \gamma_i}{1-\sum a_i} X^* \qquad 3$$

$$Y^* = \frac{\mu}{1-a_1-a_2-\cdots-a_n} + \frac{\gamma_1+\gamma_2+\cdots+\gamma_m}{1-a_1-a_2-\cdots-a_n} X^*. \qquad 4$$

Let $B_0 = \frac{\mu}{1-a_1-a_2-\cdots-a_n}$ and $B_1 = \frac{\gamma_1+\gamma_2+\cdots+\gamma_m}{1-a_1-a_2-\cdots-a_n}$, then Y^* can be expressed as conditional on a constant value of X in time t as in Equation 5. Equation 6 now illustrates the error that occurs when Y_t deviates from the long-run Y^*.

$$Y^* = B_0 + B_1 X_t \qquad 5$$

$$e_t \equiv Y_t - Y^* = Y_t - B_0 - B_1 X_t \qquad 6$$

To move further towards the ECM, Equation 1 is reparametrized to Equation 7 below,

$$\Delta Y_t = \sum_{i=1}^{n-1} a_i \Delta Y_{t-i} + \sum_{i=0}^{m-1} \gamma_i \Delta X_{t-i} + \theta_1 Y_{t-1} + \theta_2 X_{t-1} + u_t\ , \qquad 7$$

where $\theta_2 = \sum \gamma_i$ and $\theta_1 = -(1 - \sum_{i=1}^{n} a_i)$. Let the long-run parameter $B_0 = 1/\theta_1$ and $B_1 = -\theta_2/\theta_1$.

This will lead to Equation 9 by,

$$\Delta Y_t = \mu + \sum_{i=1}^{n-1} a_i \Delta Y_{t-i} + \sum_{i=0}^{m-1} \gamma_i \Delta X_{t-i} + \theta_1 \left(Y_{t-1} - \frac{1}{\theta_1} - \frac{\theta_2}{\theta_1} X_{t-1} \right) + u_t \qquad 8$$

$$\Delta Y_t = \mu + \sum_{i=1}^{n-1} a_i \Delta Y_{t-i} + \sum_{i=0}^{m-1} \gamma_i \Delta X_{t-i} - \theta_1 \left(Y_{t-1} - \hat{\beta}_0 - \hat{\beta}_1 X_{t-1} \right) + u_t \qquad 9$$

$$\Delta Y_t = \mu + \sum_{i=1}^{n-1} a_i \Delta Y_{t-i} + \sum_{i=0}^{m-1} \gamma_i \Delta X_{t-i} - \pi \hat{e}_{t-1} + \varepsilon_t \qquad 10$$

At last, Equation 10 states the ECM in its final representation and we reach the essence of the model, the parameter π. This error-correction coefficient is interpreted as the speed at which the model corrects for the error, i.e. the deviation from the long-run equilibrium, which might have occurred in the previous period. In a cointegrated relationship, it should also be noted that Granger causality must run in at least one direction. Granger causality refers to when a variable can be more accurately predicted if it is not only explained by its own lagged terms, but also by the other variables lagged terms. In the ECM the Granger causality of X on Y can therefore be detected either by a significant coefficient for the lagged terms of Δx or by the coefficient for the lagged error-correction part, which contains the lagged term of x. (Asteriou and Hall, 2011)

Multivariate systems

We will now expand the ideas from the previous example by presenting how we can model multiple time series in a system of equations. Sims (1980) stressed the reasoning of representing multiple economic variables in a dynamic system noting that there is interaction among a set of variables describing the economy, where the dependent variable might also be an important explanatory variable of the initial ‘explanatory variable’. Another view of this is given by Asteriou & Hall (2011), they express that if we do not know if a variable is endogenous or exogenous each variable has to be treated symmetrically. In other words we need to treat the variables as both exogenous and endogenous variables. Since we will try to analyse a set of sequences simultaneously that we hypothesize has some interaction between them, it is reasonable to develop a system of equations in which this interaction is considered.

A widely popular class of these multivariate equations is the vector autoregressive model (VAR). The VAR is a natural extension of the univariate autoregressive model to dynamic multivariate time series. The model has been proven to be superior to its univariate equivalent for modeling multiple economic time series whilst capturing the linear interdependencies. As mentioned earlier, we might have a long run relationship between two sequences by a cointegrated relationship. The multivariate equivalent of the ECM is naturally called a Vector Error-Correction Model (VECM). The VECM allows long-run components of variables to submit to equilibrium constraints while short-run components have a flexible dynamic specification in a multivariate setting. Below we introduce the ideas underlying these techniques. With n variables we have a $(n * 1)$ vector time series $\boldsymbol{x}_t$ where $\boldsymbol{x}_t' = (x_{1t} \ldots x_{nt})$*. To illustrate the construction of VAR models we will simplify the model† by restricting $n = 2$ and later generalizing this result. A bivariate system with one lag could be represented as following:

$$\begin{aligned} x_{1t} &= \phi_{11}x_{1,t-1} + \phi_{12}x_{2,t-1} + \varepsilon_{1t} \\ x_{2t} &= \phi_{21}x_{1,t-1} + \phi_{22}x_{2,t-1} + \varepsilon_{2t}\,, \end{aligned} \qquad 11$$

where ϕ_{ij} are coefficient constants, ε_{1t} and ε_{2t} are white noise but possibly correlated contemporaneously, i.e., ε_{1t} could be correlated with ε_{2t} but not with past values of either ε_{1t} or ε_{2t}. We can rewrite Equation 11 with matrix notation by stacking the equations according to:

$$\boldsymbol{x}_t = \boldsymbol{\Phi}\boldsymbol{x}_{t-1} + \boldsymbol{\varepsilon}_t \qquad 12$$

where $\boldsymbol{\varepsilon}_t' = (\varepsilon_{1t}\ \varepsilon_{2t})$ and $\boldsymbol{\Phi} = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix}$.

Utilizing the lag operator L and the $(2 * 2)$ identity matrix we could represent the equation as following:

$$(\boldsymbol{I}_2 - \boldsymbol{\Phi}L)\boldsymbol{x}_t = \boldsymbol{\varepsilon}_t \qquad 13$$

where $\boldsymbol{\Phi}L = \begin{bmatrix} \phi_{11}L & \phi_{12}L \\ \phi_{21}L & \phi_{22}L \end{bmatrix}$ and $\boldsymbol{I}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

We can see that Equation 12 looks like an AR(1) with vectors instead of scalars, so naturally we can call this a vector autoregressive model of order 1, VAR(1). In general, a VAR of order p is represented as:

$$\boldsymbol{\Phi}(L)\boldsymbol{x}_t = \boldsymbol{\varepsilon}_t\,, \qquad 14$$

where $\boldsymbol{x}_t$ again is a $(n * 1)$ vector of observed variables and $\boldsymbol{\Phi}$ a matrix polynomial of order p in the lag operator L such that:

$$\boldsymbol{\Phi}(L) = \boldsymbol{I}_n - \boldsymbol{\Phi}_1 L^1 - \cdots - \boldsymbol{\Phi}_p L^p = \boldsymbol{I}_n - \sum_{i=1}^{p} \boldsymbol{\Phi}_i L^i, \qquad 15$$

where $\boldsymbol{I}_n$ is the $(n * n)$ identity matrix, $\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2, \ldots, \boldsymbol{\Phi}_p$ $(n * n)$ matrices of parameters and $\boldsymbol{\varepsilon}_t \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{\Omega})$, where $\boldsymbol{\Omega}$ is the contemporaneous covariance matrix of the error-terms. The

* The prime symbol following a vector/matrix indicates the transpose of the vector/matrix.
† Furthermore, we refrain from adding any trend components or intercepts to simplify the illustration of the intuition of the models.

necessary and sufficient condition for the stability of the VAR system is that any given consequences of $\boldsymbol{\varepsilon}_t$ must die out. Consider the proposition (10.1) by Hamilton (1994):

The eigenvalues of the matrix $\boldsymbol{\Phi}$ in Equation 14 satisfy:

$$\left|\boldsymbol{I}_n\boldsymbol{\lambda}^p - \boldsymbol{\Phi}_1\boldsymbol{\lambda}^{p-1} - \boldsymbol{\Phi}_2\boldsymbol{\lambda}^{p-2} - \cdots - \boldsymbol{\Phi}_p\right| = 0. \qquad 16$$

Hence a VAR(p) is covariance-stationary as long as $|\lambda| < 1$ for all values of λ satisfying Equation 16. Equivalently, the VAR is covariance-stationary if all values of z satisfying

$$\left|\boldsymbol{I_n} - \boldsymbol{\Phi}_1 z - \boldsymbol{\Phi}_2 z^2 - \cdots - \boldsymbol{\Phi}_p z^p\right| = 0 \qquad 17$$

lie outside the unit circle.

The preceding assumptions are necessary for the usage of a VAR-model. When dealing with economic theory we usually run into a violation regarding the data, as mentioned previously economic time series usually contain a unit root. Here we present how we incorporate the remedy, differencing, in a VAR model. Writing out $\boldsymbol{\Phi}(L)\boldsymbol{x}_t = \boldsymbol{\varepsilon}_t$ we get:

$$\boldsymbol{x}_t = \sum_{i=1}^{p} \boldsymbol{\Phi}_i \boldsymbol{x}_{t-i} + \boldsymbol{\varepsilon}_t. \qquad 18$$

Subtracting $\boldsymbol{x}_{t-1}$ from both sides and rearranging, the VAR can be represented in differences according to following equation:

$$\boldsymbol{\Delta x}_t = \boldsymbol{\Phi x}_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\Pi}_i \boldsymbol{\Delta x}_{t-i} + \boldsymbol{\varepsilon}_t, \qquad 19$$

where

$$\boldsymbol{\Phi} = -\boldsymbol{I}_n + \sum_{i=1}^{p} \boldsymbol{\Phi}_i = -\boldsymbol{\Phi}(1) \text{ and } \boldsymbol{\Pi}_i = -\sum_{i=1}^{p} \boldsymbol{\Phi}_i,\ i = 1, .., k-1. \qquad 20$$

The rank of the $(n * n)$ matrix $\boldsymbol{\Phi}$ will yield 3 different scenarios of interest in regards to the long-run behaviour of the VAR system:

1. *If the $\boldsymbol{\Phi}$ matrix has rank zero, then there are n unit roots in the system and Equation 19 is simply a VAR represented in differences.*
2. *If the matrix has full rank p, then $\boldsymbol{x}_t$ is an $I(0)$ process, i.e. $\boldsymbol{x}_t$ is stationary in its levels.*
3. *If the rank of $\boldsymbol{\Phi}$ is r with $0 < r < n$ then $\boldsymbol{x}_t$ is said to be cointegrated of order r. This implies that there are $r < n$ linear combinations of $\boldsymbol{x}_t$ that are stationary. (Asteriou & Hall, 2011)*

Engle & Granger (1987) show that if $\boldsymbol{x}_t$ is cointegrated of order r ($\boldsymbol{\Phi}$ has rank r) we can write $\boldsymbol{\Phi} = \boldsymbol{\alpha\beta}'$ where both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $(n * r)$ matrices of full column rank. Substituting this into the VAR in differences (Eq.19) allows us to write the VAR as the following vector error-correction model (VECM):

$$\boldsymbol{\Delta x}_t = \boldsymbol{\alpha\beta' x}_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\Pi}_i \boldsymbol{\Delta x}_{t-i} + \boldsymbol{\varepsilon}_t, \qquad 21$$

where $\boldsymbol{\beta'}$ contains the r cointegrating vectors and $\boldsymbol{\beta' x}_{t-1}$ are the r stationary linear combinations of $\boldsymbol{x}_{t-1}$. Norman Morin of the US Federal Reserve Board offers some intuitive interpretations of the variables. The matrix $\boldsymbol{\beta}$ contains the r equilibrium relationships among the variables. For these relationships, we can interpret the difference between the observed values of $\boldsymbol{\beta' x}_{t-1}$ and their expected values as measures of disequilibrium from the r different long-run relationships. The matrix $\boldsymbol{\alpha}$ measures how quickly $\boldsymbol{\Delta x}_t$ reacts to the deviation from the equilibrium implied by $\boldsymbol{\beta' x}_{t-1}$ (Morin, 2006).

To further explain the error-correction component in Equation 21, $\boldsymbol{\alpha\beta' x}_{t-1}$**,** we will restrict r=2 cointegrating vectors and can then show that:

$$\boldsymbol{\beta' x}_t = \begin{bmatrix} b_{11} & \dots & b_{1n} \\ b_{21} & \dots & b_{2n} \end{bmatrix} \begin{bmatrix} x_{1t} \\ \vdots \\ x_{nt} \end{bmatrix} = \begin{bmatrix} b_{11}x_{1t} & + & \dots & + & b_{1n}x_{nt} \\ b_{21}x_{1t} & + & \dots & + & b_{2n}x_{nt} \end{bmatrix}$$

and thereby we have that:

$$\boldsymbol{\alpha\beta' x}_t = \begin{bmatrix} a_{11} & a_{12} \\ \vdots & \vdots \\ a_{n1} & a_{n2} \end{bmatrix} \begin{bmatrix} \sum_{i=1}^{n} b_{1i}x_{it} \\ \sum_{i=1}^{n} b_{1i}x_{it} \end{bmatrix} = \begin{bmatrix} a_{11} \sum_{i=1}^{n} b_{1i}x_{it} + a_{12} \sum_{i=1}^{n} b_{2i}x_{it} \\ \vdots \\ a_{n1} \sum_{i=1}^{n} b_{1i}x_{it} + a_{n2} \sum_{i=1}^{n} b_{2i}x_{it} \end{bmatrix}.$$

These preceding considerations lead to the foundation of a model useable for the study of pairs of variables. Since we are studying two cointegrated time series, we will now illustrate a bivariate system. Hamilton (1994) lets $\boldsymbol{z}_t \equiv \boldsymbol{\beta' x}_t$ (where $\boldsymbol{z}_t$ is a stationary $(n * 1)$ vector) and shows that the first row of Equation 21 takes the following form:

$$\begin{gathered} \Delta x_{1t} = \Pi_{11}^{(1)} \Delta x_{1,t-1} + \dots + \Pi_{1n}^{(1)} \Delta x_{n,t-1} + \dots \\ + \Pi_{11}^{(2)} \Delta x_{1,t-2} + \cdots + \Pi_{1n}^{(2)} \Delta x_{n,t-2} + \cdots \\ + \Pi_{11}^{(p-1)} \Delta x_{1,t-p+1} + \cdots + \Pi_{1n}^{(p-1)} \Delta x_{n,t-p+1} \\ + a_{11} z_{1,t-1} + a_{12} z_{2,t-1} + \cdots + a_{1n} z_{n,t-1} + \varepsilon_{1t} \end{gathered} \qquad 22$$

where $\Pi_{qj}^{(s)}$ indicates the row q, column j element of the matrix $\mathbf{\Pi}_i$, a_{qj} the row q, column j element of matrix $\boldsymbol{\alpha}$ and z_{qt} represents the *q:th* element of vector $\boldsymbol{z}_t$. Using this result, we can show that with one lag and two rows we get:

$$\begin{aligned}\Delta x_{1t} &= \phi_{11}\Delta x_{1,t-1} + \phi_{12}\Delta x_{2,t-1} + a_1 z_{t-1} + \varepsilon_{1t} \\ \Delta x_{2t} &= \phi_{21}\Delta x_{1,t-1} + \phi_{22}\Delta x_{2,t-1} + a_2 z_{t-1} + \varepsilon_{2t}.\end{aligned} \qquad 23$$

As before, z_t is the stationary linear combination of the two non-stationary dependent variables; $z_t = x_{1t} - \gamma x_{2t}$. Equation 23 represented in matrix notation is:

$$\begin{bmatrix}\Delta x_{1t}\\ \Delta x_{2t}\end{bmatrix} = \begin{bmatrix}\phi_{11} & \phi_{12}\\ \phi_{21} & \phi_{22}\end{bmatrix}\begin{bmatrix}\Delta x_{1,t-1}\\ \Delta x_{2,t-1}\end{bmatrix} + \begin{bmatrix}a_{11}\\ a_{21}\end{bmatrix}(x_{1t} - \gamma x_{2t}) + \begin{bmatrix}\varepsilon_{1t}\\ \varepsilon_{2t}\end{bmatrix}, \qquad 24$$

where a_{11} and a_{21} are the speed of adjustment coefficients describing how quickly deviations from the long run equilibrium are corrected.

Arbitrage Pricing Theory

To see why and how one can hypothesize that two stocks can move synchronously we need to present some economic theory. A particularly famous idea presented by economist Stephen Ross (1976) is the Arbitrage Pricing Theory (APT), which has strong relations to the even more famous Capital Asset Pricing Model* (CAPM). We will here give a brief explanation of this theory without assessing the specific details. The theory holds that the expected return of any financial asset can be modelled as a linear function of a set of macroeconomic factors such as inflation, GNP, investor confidence and so on (Ross et al., 1986). In a multifactor framework, the dynamics of an assets return are usually characterized by its factor exposure/sensitivity profile. Since one can think of APT as an extension of CAPM, we will define the coefficients $(\beta_1, \beta_2, \dots., \beta_k)$ to be the factor exposures and the return contribution to every factor $(X_1, X_2, \dots, X_k)$. Thereby we will have that the total return of the asset is

$$X = \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + X_e, \qquad 25$$

where X_e is the idiosyncratic return not explicable by the factors, which acts as what we have previously called the error-term. From Equation 25 we can see that the contribution to the overall asset return due to each factor is proportional to the exposure/sensitivity of the asset to different factors. Basically we could, according to the theory, know all about the risk and return characteristics of the asset if we knew the risk profile. To illustrate how the risk is

* See Sharpe, William F. (1964), "Capital Asset Prices – A Theory of Market Equilibrium Under Conditions of Risk". *Journal of Finance*, or consult a textbook on Corporate Finance such as 'Corporate Finance' By Stephen Ross and Randolph Westerfield.

measured in the APT framework we will simplify a model to only consist of 2 factors, recall Equation 25 and observe that here $k = 2$. Then the variance is given by,

$$var(X) = \beta_1^2 var(X_1) + \beta_2^2 var(X_2) + 2\beta_1\beta_2 cov(X_1, X_2) + var(X_e). \qquad 26$$

This can be expressed in matrix notation as following,

$$var(X) = [\beta_1\beta_2] \begin{bmatrix} var(X_1) & cov(X_1, X_2) \\ cov(X_2, X_1) & var(X_2) \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + var(X_e) \qquad 27$$

or in compact form,

$$var(X) = \boldsymbol{e\Omega e'} + \ var(X_e). \qquad 28$$

Where $\mathbf{\Omega}$ is the covariance matrix and $\boldsymbol{e}$ the factor exposure vector. Then if $\boldsymbol{e\Omega e'}$ is similar for two variables we should expect the two variables to move together in time. More specifically, two assets will be optimally cointegrated if two assets share common factors and these factors are perfectly aligned. So in the words of the APT, the common factor return of a long-short portfolio of the two stocks is zero. However this is seldom the case in practice except in the case of class A and class B shares. Furthermore, deviation might occur due to X_e being the company-specific risk. It is therefore desirable to study companies that could be expected to move together, this could be done by restricting the selection to companies within the same industry and other similar risk exposures.

Methodology

Augmented Dickey-Fuller

In 1979, Dickey and Fuller developed a test assessing the presence of a unit root in the series. The Dickey-Fuller test starts with Equation 29,

$$\Delta y_t = (\phi - 1)y_{t-1} + u_t, \qquad 29$$

where u_t is white noise, and states the null hypothesis as $H_0: (\phi - 1) = 0$ and $H_1: (\phi - 1) < 0$, which is tested through an ordinary t-test. The test statistic is however not evaluated through the conventional t-values, it requires specific critical values originally developed by Dickey and Fuller. See MacKinnon (1991) for full specification of the critical values. The test is also able to account for an intercept and linear trend. Accepting the null implies that the series contains a unit root and is therefore not stationary. The test was later extended to include p number of lags in the model given the likelihood that the term u_t in Equation 29 is not white noise. This extended version is the Augmented Dicker-Fuller test and the corresponding equations when accounting for several lags in the model are:

$$\Delta y_t = \gamma y_{t-1} + \sum_{i=1}^{p} \beta_i \Delta y_{t-i} + u_t \qquad 30$$

$$\Delta y_t = a_0 + \gamma y_{t-1} + \sum_{i=1}^{p} \beta_i \Delta y_{t-i} + u_t \qquad 31$$

$$\Delta y_t = a_0 + a_1 t + \gamma y_{t-1} + \sum_{i=1}^{p} \beta_i \Delta y_{t-i} + u_t. \qquad 32$$

As for the simple case the test can account for a constant and linear trend. Equation 31 takes into account an intercept through the constant term a_0, whereas Equation 32 accounts for both the intercept as well as for a linear trend by the term $a_1 t$ (Asteriou & Hall, 2011).

The Engle & Granger approach

Before the development of the Johansen's test, a way of assessing a potential cointegration relationship had already been proposed by Engle & Granger (1987). The method is intuitive and simple to perform, but has been criticized on some grounds. The Engle & Granger (EG) approach is able to detect one cointegrating relationship between variables, in contrast to the Johansen's test, which is able to detect more than one, i.e., several cointegrating vectors. If two variables are integrated of the same order, the simple regression $Y_t = \beta_0 + \beta_1 X_t + e_t$ is constructed. If the variables really are cointegrated, the estimated $\hat{\beta}_1$ from the OLS will converge very quickly. The estimates for the error term, $\hat{e}_t$, are then tested for order of integration by the ADF test. If the residuals are integrated of a lower order than the variables themselves, we conclude that the variables are cointegrated and we proceed by estimating an ECM. Given that the procedure involves two estimations before we reach a conclusion, the EG approach is distorted by the fact than any errors from the first estimation of the residuals are carried along to the second estimation where the residuals are tested for the order of integration. Another obvious critique towards the EG approach is that only one cointegrating relationship can be assessed, although there might be more than one cointegrating vector when investigating more than two variables. Furthermore, the question arises for which variable to place as the dependent variable. Asymptotically, the results will be equivalent, but for samples that are not large enough it might give contradicting results. (Asteriou & Hall, 2011) In our analysis we will apply the Johansen's method, crosschecking our finding by the EG-approach.

Johansen's test

In a VAR system where cointegrating relationships exists we found in the previous section that the $(n * n)$ $\mathbf{\Phi}$ matrix can be decomposed into $\boldsymbol{\alpha\beta'}$ where $\boldsymbol{\alpha}$ is a $(n * r)$ matrix and $\boldsymbol{\beta'}$ is a $(r * n)$ matrix. From this decomposition it is possible to obtain important information about the long-run relationship between the cointegrated series, the $\boldsymbol{\beta'}$ matrix contains the coefficients for the cointegrating vectors and the $\boldsymbol{\alpha}$ matrix will give the coefficients for the speed of adjustment. There can exist $(n - 1)$ cointegrating relationships between n variables, and the Johansen's test allows for the detection of more than one cointegrating relationship in contrast to the Engle & Granger approach that can only detect one. The Johansen's test serves to detect the rank of the $\mathbf{\Phi}$ matrix, which is given by the r number of linearly independent columns in the $(n * r) * (r * n)$ decomposition $\boldsymbol{\alpha\beta'}$. If the variables are cointegrated, the rank will be reduced and $r < n$. A full rank would imply that all the variables are $I(0)$, and if the rank $r = 0$, it implies that there are no cointegrated relationships and further research would require differencing to obtain stationarity in the data before estimating a VAR system or other suitable methods. However, if the rank is found to be of reduced order, $0 < r \leq (n - 1)$, there are r number of cointegrating relationships (Asteriou & Hall, 2011).

The Johansen's test investigates whether restriction imposed on the $\mathbf{\Phi}$ matrix that reduces the rank can be rejected. Before performing the Johansen's test, the appropriate lag length and model selection must be set for the model. The lag length determines the number of lags to be included in the VAR system before examining the $\mathbf{\Phi}$ matrix and constructing the VECM.

Commonly, the lag length is determined by constructing an unrestricted VAR and examining the lag length structure suggested by the Akaike information criterion (AIC). Furthermore, the model that best describes the dynamic relationship between the variables must me chosen. This concerns whether the model should include an intercept and/or a trend component in the cointegrating equation and in the VAR respectively. Once these aspects are considered, the Johansen's test can be performed. The two statistics proposed by Johansen (1988) for testing the presence of cointegration are the *trace test* and the *maximum eigenvalues test*,

$$\lambda_{trace}(r) = -T \sum_{i=r+1}^{n} ln(1-\hat{\lambda}_i) \qquad 33$$

$$\lambda_{max}(r, r+1) = -T\ ln(1-\hat{\lambda}_{r+1}). \qquad 34$$

For a sample size of T, each test is a likelihood ratio test based on the eigenvalues ($\hat{\lambda}_i$) from the estimation of the matrix $\mathbf{\Phi}$. The trace test states the null hypothesis that the number of cointegrating vectors is less or equal to r against the alternative hypothesis that there are more than r cointegrating vectors and therefore more than r cointegrating relationships. The maximum eigenvalue test orders the eigenvalues in descending order, and considers if they are different from zero. If none of the $\lambda_1 > \lambda_2 > \cdots > \lambda_n$ are different from zero there is no cointegration among the variables. The null hypothesis is that the number of cointegrating vectors equals to r, which is tested against the alternative hypothesis that the number of cointegrating vectors equals to $r+1$. Both statistics follow non-normal distributions, which are specified in Johansen and Juselius (1990).

Impulse Response

We have established the basis for the model that we will construct and how we will assess it in the form of statistical tests. We will now further introduce concepts of how to study the behaviour of a system through the so-called impulse response function or simply just the impulse response (IR). It can be thought of as the reaction of a dynamic system over time in response to some exogenous shock. Recall that for any stationary VAR(p) there exists an infinite Vector Moving Average (VMA) according to Wold's decomposition, as following:

$$\boldsymbol{x}_t = \boldsymbol{\varepsilon}_t + \boldsymbol{\psi}_1\boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\psi}_2\boldsymbol{\varepsilon}_{t-2} + \cdots = \sum_{i=0}^{\infty} \boldsymbol{\psi}_i \boldsymbol{\varepsilon}_{t-i}. \qquad 35$$

Thus the matrix $\boldsymbol{\psi}_s$ can be interpreted as:

$$\boldsymbol{\psi}_s = \frac{\partial x_{t+s}}{\partial \boldsymbol{\varepsilon}'}, \qquad 36$$

which translates to that the row i and column j element of $\boldsymbol{\psi}_s$ identifies the consequences of a one-unit increase in the j:th variables error-term at date t (ε_{jt}) for the value of the i:th variable at time $t + s$ $(y_{i,t+s})$ at all dates, ceteris paribus (Hamilton, 1994).

Given that we know that the first element of $(\boldsymbol{\varepsilon}_t)$ is changed by η_1, the second by η_2, the third by η_3 and so on we can summarize the combined effect of the vector x_{t+s} as:

$$\Delta \boldsymbol{x}_{t+s} = \frac{\partial x_{t+s}}{\partial \varepsilon_{1t}}\eta_1 + \frac{\partial x_{t+s}}{\partial \varepsilon_{2t}}\eta_2 + \frac{\partial x_{t+s}}{\partial \varepsilon_{3t}}\eta_3 + \cdots + \frac{\partial x_{t+s}}{\partial \varepsilon_{nt}}\eta_n = \boldsymbol{\psi}_s\boldsymbol{\eta}, \qquad 37$$

where $\boldsymbol{\eta} = (\eta_1, \eta_2, \eta_{,3}, \dots, \eta_n)'$, plotting row i and column j element of $\boldsymbol{\psi}_s$ as a function of s yields:

$$\frac{\partial x_{i,t+s}}{\partial \varepsilon_{jt}}, \qquad 38$$

which is known as the impulse response function. From Equation 38 we can see that, as mentioned in the beginning of this section, it describes the response of $x_{i,t+s}$ to a one-time impulse in ε_{jt} holding all other variables dated t or earlier constant. The IR has for a long time been called one of the hallmarks of VAR-analysis, with regards to the illustration of the dynamic characteristics of empirical models in economics (Keating, 1992). The areas of application for the impulse response function come of a wide variety. In economics the impulse would be called a shock and the response a dynamic multiplier. It is usually used to determine whether government spending stimulates other spending in the economy. (Stevans & Sessions, 2010) In physics it could model the decay rate of uranium in a nuclear reactor after adding a kilogram of uranium. Another frequent application area is the technical analysis of stocks, in particular the impact of new information.

Manganelli (2002) utilizes the impulse response function to study how fast cointegrated stocks modelled in a VAR system return to their long-run equilibrium for different levels of liquidity after arrival of new information. Thereby in our dynamical system we could map how the sequence $\{\Delta x_{2t}\}$ responds to a unit-shock to $\{\Delta x_{1t}\}$, which is how we will utilize the method. When computing the IR we are faced with several options. The structural difficulty regarding the simple IR is that we allow for contemporaneous correlation of the error-terms, which means that the error-terms have some overlapping information. Since the IR gives the response of a one-time shock everything else held constant, the contemporaneous correlation will violate the model holding everything else constant. In this paper we will overcome this problem by constraining the contemporaneous effect of the error-term, Hamilton (1994) explains that how we choose to impose our restrictions is based on the type questions we are asking, for example in the sense of forecasting the question would be why and what do we want to forecast? For n variables we have to order the causality of the variables. For example we would have to decide an ordering such that the shock to X_1 feeds contemporaneously to $X_2, X_3, \dots, X_n$and the shock to X_2 feeds contemporaneously to $X_3, X_4, \dots, X_n$, however it would only feed into X_1 with a lag and so on. This way of ordering variables is often called Wold's causality (Demiralp & Hoover, 2003). Since we will analyse the relationship between two variables and would like to know which of two variables leads the other, we utilize our results from Granger causality test, and put restrictions on the contemporaneous effect on the variable which *doesn't* Granger cause the other variable. To separate the individual effects of all the shocks the error-terms must be uncorrelated to yield contemporary independence. Recall that our VAR-model was constructed as

$$x_{1t} = \phi_{11}x_{1,t-1} + \phi_{12}x_{2,t-1} + \varepsilon_{1t} \tag{39}$$

$$x_{2t} = \phi_{21}x_{1,t-1} + \phi_{22}x_{2,t-1} + \varepsilon_{2t}, \tag{40}$$

where $\boldsymbol{\varepsilon}_t'$ =$(\varepsilon_{1t}\ \varepsilon_{2t}) \sim N(\mathbf{0}, \mathbf{\Omega})$ and $\mathbf{\Omega}$ is a contemporaneous variance-covariance matrix of the error-terms such that

$$\boldsymbol{\Omega} = E(\boldsymbol{\varepsilon}_t\ \boldsymbol{\varepsilon}_t') = \begin{bmatrix} var(\varepsilon_{1t}) & cov(\varepsilon_{1t}, \varepsilon_{2t}) \\ cov(\varepsilon_{2t}, \varepsilon_{1t}) & var(\varepsilon_{2t}) \end{bmatrix}. \tag{41}$$

Now if we multiply Equation 40 by $\kappa = \frac{cov(\varepsilon_{1t}, \varepsilon_{2t})}{var(\varepsilon_{2t})}$ and subtract it from Equation 39 this will give us:

$$x_{1t} = \Pi_0 x_{2t} + \Pi_1 x_{2,t-1} + \Pi_2 x_{1,t-1} + \xi_{1t}, \tag{42}$$

where $\Pi_0 = \kappa$, $\Pi_1 = \phi_{11} - \kappa\phi_{22}$, $\Pi_2 = \phi_{12} - \kappa\phi_{22}$ and $\xi_{1t} = \varepsilon_{1t} - \mu\varepsilon_{2t}$.

This effectively gives us a model taking the form of the two linear combinations in Equation 40 and Equation 42. We then have that the error term is now $\boldsymbol{\xi}_t' = (\xi_{1t}, \varepsilon_{2t})$ and no longer has the possibility to correlate contemporaneously, since the covariance matrix is now given by:

$$E(\boldsymbol{\xi}_t \boldsymbol{\xi}_t') = \begin{bmatrix} var(\varepsilon_{1t}) + 2cov(\varepsilon_{1t}, \varepsilon_{2t}) + \mu^2 var(\varepsilon_{2t}) & 0 \\ 0 & var(\varepsilon_{2t}) \end{bmatrix}. \tag{43}$$

This matrix shows how a bivariate Cholesky decomposition allows us to interpret the response of independent exogenous shocks since as we can see there is no contemporaneous relationship between the error-terms that constitute the shocks.

Variance Decomposition

Another important method for describing the behaviour of our system is conducting variance decomposition. The method identifies the importance of each shock as a component of the overall unpredictable variance of each variable over time. It is based on the Moving-Average representation described in the introduction of the impulse response section. Consider the following, if we let $(\hat{\boldsymbol{x}}_{t+s}|\boldsymbol{\Theta})$ denote the s-step ahead forecast of $\boldsymbol{x}_t$ and $\boldsymbol{\Theta}' = (\boldsymbol{x}'_t, \boldsymbol{x}'_{t-1}, \dots)$ we have that:

$$(\hat{\boldsymbol{x}}_{t+s}|\boldsymbol{\Theta}) = \sum_{i=s}^{\infty} \boldsymbol{\psi}_i \, (\boldsymbol{\varepsilon}_{t+s-i}|\boldsymbol{\Theta}) \qquad 44$$

is a s-step ahead forecast, where $\boldsymbol{\Theta} = (\boldsymbol{x}'_t, \boldsymbol{x}'_{t-1}, \dots)$ and $\boldsymbol{\psi}_i$ is the coefficient matrix. Using the notation of Lütkepohl (2005) the optimal s-step forecast error is:

$$\boldsymbol{x}_{t+s} - \boldsymbol{x}_t(s) = \sum_{i=0}^{s-1} \boldsymbol{\psi}_i \boldsymbol{\varepsilon}_{t+s-i}, \qquad 45$$

where in this case, $\boldsymbol{x}_t(s)$ is the optimal forecast $(\hat{\boldsymbol{x}}_{t+s}|\boldsymbol{\Theta})$. To map one row of the matrix $\boldsymbol{\psi}_i$ we use the notation $\theta_{mn,i}$, the s-step forecast error of the j:th column component of $\boldsymbol{x}_t$ is:

$$\begin{aligned} x_{j,t+s} - x_{j,t}(s) &= \sum_{i=0}^{s-1} (\theta_{j1,i}\, \varepsilon_{1,t+s-i} + \dots + \theta_{jN,i} \varepsilon_{N,t+s-i}) \\ &= \sum_{n=1}^{N} (\theta_{jk,0}\, \varepsilon_{n,t+s} + \dots + \theta_{jn,s-1} \varepsilon_{n,t+1}). \end{aligned} \qquad 46$$

From Equation 46 we can see that the forecast error of the j:th column component *could possibly* consist of all the error terms $\varepsilon_{1t}, \varepsilon_{2t}, \dots, \varepsilon_{Nt}$. The reason to why it only *possibly could* consist of all the error terms is that some of the coefficients could be zero. Because all the error-terms are assumed to be uncorrelated, the Mean-Square Error (MSE) of the s-step forecast of $x_{j,t}$ is

$$E(x_{j,t+s} - x_{j,t}(s))^2 = \sum_{n=1}^{N} (\theta_{jn,0}^2 + \dots + \theta_{jn,s-1}^2). \qquad 47$$

Therefore we have that

$$\theta_{jn,0}^2 + \theta_{jn,1}^2 + \dots + \theta_{jn,s-1}^2 = \sum_{i=0}^{s-1} \left(e'_j \boldsymbol{\psi}_i e_n\right)^2, \qquad 48$$

where e_n is the n:th column of the N-dimensional identity matrix I_N, the expression in Equation 48 is the contribution of the n:th shock to the forecast error variance of the j:th variable for a given forecast horizon (Seymen, 2008). This is often interpreted as the MSE of the s-step forecast variable j (Lütkepohl, 2005).

If we proceed to divide Equation 48 by the MSE of $x_{j,t}(s)$ which can be represented as

$$MSE(x_{j,t}(s)) = \sum_{i=0}^{s-1} \sum_{s=1}^{N} \theta_{jk,i}^2, \quad 49$$

this gives us

$$\omega_{jn,s} = \frac{\sum_{i=0}^{s-1}(e'_j \psi_i e_n)^2}{\sum_{i=0}^{s-1} \sum_{s=1}^{N} \theta_{jk,i}^2}, \quad 50$$

which is the proportion of the s-step forecast error variance of variable j, accounted for by ε_{nt} error-terms. If ε_{nt} is connected to variable n, $\omega_{jn,s}$ represents the proportion of the s-step forecast error variance accounted for by error-terms in variable n. Thus, the forecast error *variance* is *decomposed* into components accounted for by error-terms in the different variables of the VAR-system. (Lütkepohl, 2005)

Data

The data that are examined in this paper cover daily closing prices for the stocks of 5 large energy companies listed on the New York Stock Exchange. We restrict the selection of companies to large companies within the same sector, listed on the same exchange. This will further support any cointegrating relationship since the companies are more likely to have similar risk exposures and presumably liquid stocks. For the period of the 27th September 2012 to the 22th April 2014, 409 observations for each stock are collected from Thomson Reuters Datastream. The data is transformed by the natural logarithm, scaling the data but maintaining the characteristics of the data. From now on, each reference to any of the series will be on the logarithm of the series.

Between the five series there can be ten different combinations that could be cointegrated, these results will be presented in the next section. Before performing any tests it is appropriate to conduct some inspection of the data, both visually and formally. In Appendix A, each pair is presented with graphs for the full period, together with descriptive statistics for each series. From the graphs for the pairs of SNP – VLO, PSX – VLO and the last year of SNP – VLO it is possible to see that the series seem to move together, this might however be spurious given that both series are non-stationary. Table 1 presents the companies with their respective symbol and also the results of the Augmented Dickey-Fuller test, showing the probability of the presence of a unit root in the level data and in the differenced data respectively. We cannot, for any series, reject the presence of a unit root in the level data. Once the data is differenced, we can for all series reject the null hypothesis that the series contain a unit root. Thus, all series are integrated of order one. The ten pairs that can be tested for cointegration are: XOM – SNP, XOM – CVX, XOM – PSX, XOM – VLO, SNP – CVX, SNP – PSX, SNP – VLO, CVX – PSX, CVX – VLO and PSX – VLO.

Company	**Symbol**	**Unit root in Y_t**	**Unit root in ΔY_t**	**Order of integration**
Exxon Mobil Corp.	XOM	0.508	0.000	$I(1)$
Sinopec Group	SNP	0.183	0.000	$I(1)$
Chevron	CVX	0.373	0.000	$I(1)$
Phillips 66	PSX	0.683	0.000	$I(1)$
Valero Energy	VLO	0.848	0.000	$I(1)$

Table 1. Reported p-values for the Augmented Dickey-Fuller test on each series, both in level data and in first difference. For all series, it is shown that we can reject the null hypothesis that the series contain a unit root after taking the first difference, therefore all series are integrated of order one. The data comprise 409 observations for each stock between 9/27/2012 – 4/22/2014.

Results

As presented in the previous section, the five time series regarding daily closing prices for different energy companies are all $I(1)$. The aim of our study is to find a pair that could be profitable to trade within the frame of statistical arbitrage, where the concept of mean-reversion is a fundamental aspect of the strategy. In this section we present the results for all pairwise combinations that have been examined by Johansen's test and the EG approach. The tests investigate the existence of cointegrating relationships between the variables. Cointegration implies that the variables are bound to a long-run equilibrium, which can be visualized in the mean-reversion of the error term from the long-run relationship. To perform the Johansen's test, an unrestricted VAR is first estimated for the pair that is examined. This is constructed in order to find the appropriate lag length suggested by AIC. The unrestricted VAR is estimated for a larger number of lags, gradually decreasing the lags to ensure that the lag length suggested by AIC is consistent. We find that for all the pairs, the lag length 1 is consistently suggested by the criterion. Therefore, an unrestricted VAR(1) is estimated for all pairs in order to perform the Johansen's test. Furthermore, it is also necessary to determine the model used to assess the cointegrating rank. All models are simultaneously estimated, and the model suggested by AIC is chosen as the appropriate model for our purpose. The results from the Johansen's test are presented below in Table 2.

Pair	**Johansen's test of cointegrating rank**	**$\hat{e}_t$ stationary applying EG approach**
CVX - PSX	No cointegration, $r = 0$	No
CVX - SNP	No cointegration, $r = 0$	No
CVX - VLO	No cointegration, $r = 0$	No
CVX - XOM	No cointegration, $r = 0$	No
PSX - SNP	No cointegration, $r = 0$	No
PSX - VLO	One cointegrating relationship, $r = 1^{**}$	Yes**
PSX - XOM	No cointegration, $r = 0$	No
SNP - VLO	No cointegration, $r = 0$	No
SNP - XOM	No cointegration, $r = 0$	No
VLO - XOM	No cointegration, $r = 0$	No

Table 2. Each possible pair is tested for cointegration by Johansen's test and by the EG approach. The EG approach is employed by estimating $Y_t = \beta_0 + \beta_1 X_t + e_t$, and testing $\hat{e}_t$ for stationarity using ADF. ** indicates significance on the 1% level.

By the Johansen's test we find that one pair has a cointegrating rank of $r = 1$, significant on the 1% level, whereas we for all other pairs cannot reject the null hypothesis that no cointegrating relationship exists. The results are supported consistently by both the trace test and the maximum eigenvalue test. To further ensure our findings, we employ the EG approach to all pairs. The EG approach is a two-step estimation that involves the estimation of a residual series from a linear combination of the variables in the pair, which is then tested for stationarity by the Augmented Dickey-Fuller test.

Dependent variable: PSX

Independent variable		$Adj. R^2$	Augmented Dickey-Fuller for estimated residuals $(\hat{e}_t)$
C	VLO	0.966	0.000
1.369	0.755		(-4.607)
(53.166)	(107.301)		

Table 3. The pair consisting of PSX and VLO is examined by the EG approach. The linear combination is estimated, obtaining the estimated series of $\hat{e}_t$. The residual series is then tested by the ADF test, the p-value (0.000) strongly rejects the null hypothesis that the series contains a unit root. For regression output, t-statistics are given within parenthesis. The data comprises 409 observations of PSX and VLO respectively.

Both the Johansen's test and the EG approach indicated that the variables PSX and VLO are cointegrated. The results from the EG approach are presented above in Table 3. The linear combination of the two variables shows a very high coefficient of determination, about 97%. The Augmented Dickey-Fuller test strongly rejects the null hypothesis of unit root in the estimated residual series, confirming that the variables are indeed cointegrated and that the relationship is not spurious. As previously mentioned, the cointegrating relationship will imply that there is a long-run equilibrium between the variables and that any deviations from this equilibrium are merely temporary. Figure 1 illustrates this mean-reverting behaviour by plotting the estimated residuals from the linear combination given in Table 3.

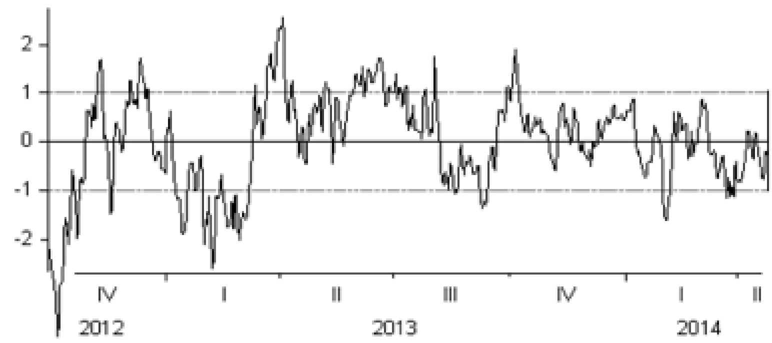

Figure 1. The estimated residuals $(\hat{e}_t)$ from the OLS $PSX = \beta_0 + \beta_1 VLO + e_t$ are standardized and plotted against time for the period 9/27/2012 – 4/22/2014. The graph shows how deviations from the long-run equilibrium are corrected by the mean-reverting behaviour of the series. Figure 14 in Appendix shows the actual series plotted against the fitted series, along with the residuals.

Vector Error-Correction

By now we have obtained significant results from both the Johansen's test and the EG approach that the variables PSX and VLO are cointegrated. Also, the visual inspection of the error term from the linear combination has indicated the mean-reverting characteristic of a cointegrated pair. We are now confident to estimate the VECM from which we can assess both the short-run and the long-run dynamics of the relationship. As indicated in the unrestricted VAR we employ a lag length of one lag and a model that allows for an intercept in the cointegrating equation and in the VAR, but not for a linear trend in the cointegrating equation, nor in the VAR. The estimates of the VECM are given below in Table 4.

Cointegrating equation		
VLO_{t-1}	1.000	
PSX_{t-1}	-1.342 [-26.996]**	
C	1.885	
Error-correction:	ΔVLO	ΔPSX
Cointegrating equation	-0.012[-0.459]	0.066 [2.908]**
ΔVLO_{t-1}	0.023 [0.309]	-0.010 [-0.154]
ΔPSX_{t-1}	-0.008 [-0.090]	0.072 [0.967]
C	0.002 [1.599]	0.001 [1.544]

Table 4. The table reports the estimates for the vector error-correction model for PSX and VLO, t-statistics within brackets. The cointegrating vector that gives a stationary combination of the two non-stationary variables constructs the cointegrating equation. ** indicates significance on the 1% level.

For a cointegrated set of variables, we expect that at least one of the coefficients in the $(n * r)$ vector $\boldsymbol{\alpha}$ is significantly different from zero. The coefficients in the vector give the so-called speed of adjustment when any of the variables in the systems causes a deviation from the long-run equilibrium. If none of the coefficients are different from zero, the error-correction part falls out and the model no longer supports the concept of cointegration. In our VECM, vector $\boldsymbol{\alpha}$ is given by the two coefficients for the Cointegrating equation, -0.012 and 0.066, where only the latter is significantly different from zero. This indicates that we have a long-run relationship between the variables, as we expected given the preceding results. However, only PSX is responsible for correcting the deviations that might occur from the long-run equilibrium. As the coefficient for the Cointegrating equation for ΔVLO is not significantly different from zero, the variable is considered to be weakly exogenous and its whole equation can be dropped, although it can still be a part of the equation for ΔPSX. The significance of the coefficient for the error-correction part in the equation for ΔPSX indicates the unidirectional Granger causality that runs from VLO to PSX.

Thus far we have concluded that the VECM supports a long-run relationship between the variables, where PSX is responsible for correcting any deviations from the long-run equilibrium. The speed of this adjustment is given by its coefficient for the Cointegrating equation. Although this is considered to be the essence of the error-correction representation, the VECM also allows us to assess the short-term dynamics between the variables. By studying the $(2 * 2)$ matrix of β-coefficients for the lagged differences of the variables, the impact of the short-term effects can be measured. In our case none of the coefficients show any significant deviation from zero, thereby suggesting that there is no short-run effect supported by the VECM.

Impulse Response

In the previous section we established that our model is characterized by a mean-reverting behaviour through the error-correcting term in the VECM, which confirmed the long-term relationship of the system. To further investigate the short-term behaviour we assess the Impulse Response Function as mentioned in the theoretical section. Furthermore, we saw that VLO Granger causes PSX, however the direction alone does not tell us the full story regarding the interaction of the variables. If one of our stocks responds to an exogenous shock in the other stock, we may very well draw the conclusion that the latter causes the former, here we expand on this interaction using IR. Recall that the Vector Moving-Average could be obtained through an inversion of the VAR model and that the coefficient matrix of the Moving-Average process contains the responses to a unit-shock. Since the inversion requires stationary variables it is important to point out that the marginal impact of the shock in each subsequent time period will tend towards zero over time. An intuitive way of observing the dynamic interrelationship within the system is to depict a graphical interpretation of a shock in each of the variables and the corresponding responses for all the variables in the system.

As we can see below in Figure 3, a unit standard deviation shock to ΔVLO leads to an expected increase of approximately 130 basis points in ΔPSX after one period and in the subsequent period this significant effect is expected to decrease to approximately 10 basis points. From the third period and on the effect is insignificant as seen in Table 2 in Appendix A. As we can see in both Figure 2 and Figure 3, the autoregressive nature of our system attributes to that there is a convergence to zero rather than the effect seizing immediately. However, the convergence materializes in rapid manner, showcasing the stability of our system.

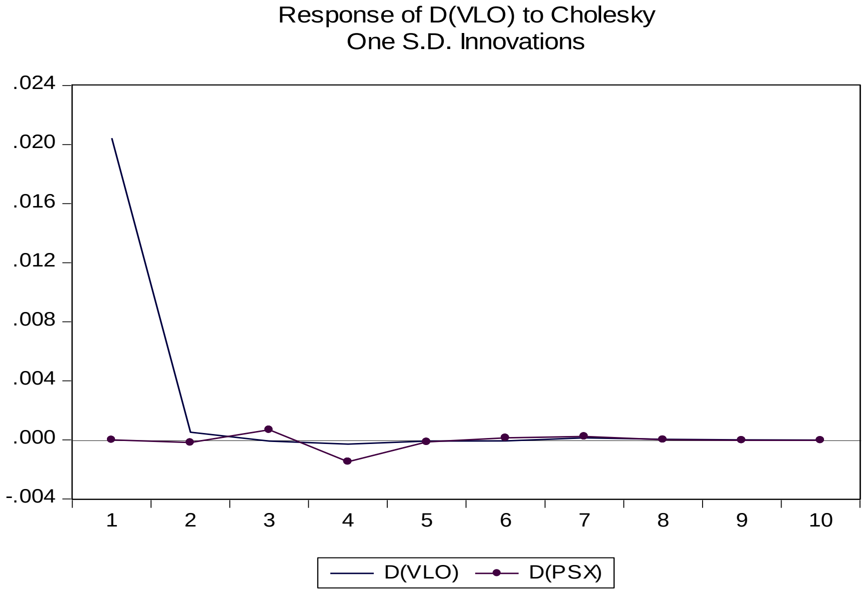

Figure 2. Impulse Response function for ΔVLO responding to a one standard deviation shock in ΔVLO in the solid line and in ΔPSX in the dotted line. We can here see that a unit-shock to ΔPSX has little effect on ΔVLO.

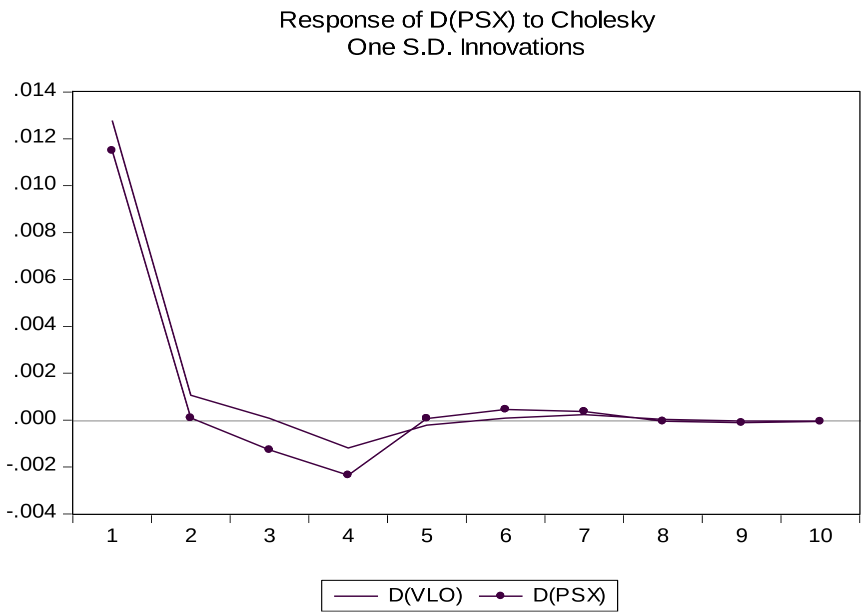

Figure 3. Impulse Response function for ΔPSX responding to a one standard deviation shock in ΔVLO in the solid line and in ΔPSX in the dotted line. We can here see that the to ΔVLO has a significant effect on ΔPSX, confirming that VLO leads PSX in the short-run as well.

Observing Figure 2 we can see that a shock to ΔPSX leads to no significant response in ΔVLO in any of the periods following the shock, as explained this is expected since we established that PSX does not Granger causes VLO. In addition to the interrelationship, Figure 2 and 3 also visualize how the assets respond to exogenous shocks to their own error-terms respectively. The series react in an autoregressive manner with a reaction in the first period, which converges to insignificance over the following periods. If we consider these results in a practical matter, an example could be that a manager of a portfolio consisting of a position in PSX would have to include fundamental news regarding VLO to the information set while calibrating her portfolio for narrow horizons as well.

Variance Decomposition

A natural extension of the impulse response analysis is computing how large portion of each variable's forecast error variance is explained by various shocks at different time horizons. This method is known as the variance decomposition or forecast error variance decomposition. Consistent with the impulse response analysis we can see that ΔVLO evolves independently of ΔPSX. This is observed in Table 5 where we can see that on average ΔVLO accounts for practically all its variance resulting from a shock to itself. In Table 6, we can see that a slight majority of the forecast error variance of ΔPSX is accounted for by shocks to ΔVLO; ranging from 53.95% to 55.10% for the whole horizon. In the same table we can see that $\Delta PSX's$ own shocks to account for 44.90% to 46.05% of the variance. These results are consistent with our findings and provide further robustness to the dynamics of our assessed pair.

Forecast error in ΔVLO		
Forecast Horizon h	Proportion of forecast error variance h periods ahead accounted for by error-terms in	
	ΔVLO	ΔPSX
1	100.000	0.000
2	99.986	0.014
3	99.887	0.113
4	99.330	0.670
5	99.326	0.674
6	99.321	0.679
7	99.307	0.693
8	99.307	0.693
9	99.307	0.693
10	99.306	0.694

Table 5. Variance decomposition of ΔVLO, which shows the expected proportion of its forecast error variance that is due to shocks in ΔPSX in the right column and due to a shock in 'itself' in the left column for 10 periods (days).

Forecast error in ΔPSX		
Forecast Horizon h	Proportion of forecast error variance h periods ahead accounted for by error-terms in	
	ΔVLO	ΔPSX
1	54.966	45.034
2	55.101	44.899
3	54.784	45.216
4	54.004	45.996
5	54.009	45.991
6	53.970	46.030
7	53.954	46.046
8	53.954	46.046
9	53.952	46.048
10	53.951	46.049

Table 6. Variance decomposition of ΔPSX, which shows the expected proportion of its forecast error variance that is due to shocks in ΔVLO in the left column and due to a shock in 'itself' in the right column for 10 periods (days).

Conclusions

The aim of this paper was to describe a set of methods of how to find a pair of stocks that could be profitable to trade within the frame of statistical arbitrage. The framework includes methods of assessing the behaviour of such a relationship for both short and long time horizons. Although previous literature on the subject is rather limited, several methods are proposed. Many seek pairs that move together, tied by a long run relationship. However, not all emphasize the importance of the relationship to be stationary. This paper employs a cointegration approach, testing pairwise combinations of five stocks for cointegration using daily closing prices for the period 9/27/2012 – 4/22/2014. All five companies are within the energy sector and listed on the New York Stock Exchange.

We found the pair consisting of Phillips 66 (PSX) and Valero Energy (VLO) to be cointegrated, supported by both the Johansen's test and the Engle & Granger approach. The cointegrating relationship gives a stationary series that ensures the mean-reverting behaviour, which is essential for the profitability of pairs trading. We then constructed a VECM for the cointegrated variables, the model shows how deviations from the long-run equilibrium are corrected by PSX, and that Granger causality runs from VLO to PSX. The Granger causality confirms that VLO leads PSX, which enables us to study how the stocks react to the arrival of fundamental news through the Impulse Response analysis. The fundamental news in this case is explained by an impulse in the error-term of one stock and the response to the news is the reaction of the other. We find that a shock in VLO yields a significant response from PSX in the two subsequent periods. This implies that an investor's forecast for PSX would need to account for news regarding company specific risk in VLO. As the investor's forecast will be associated with an error, we conducted variance decomposition for the error variance. By this we could see that the forecast error of PSX is highly affected by VLO, but not the other way around. This further confirms the importance of news regarding VLO to the investor of PSX.

Although the framework is referred to as statistical arbitrage, there are several risks involved. Among others, the company specific risks and structural breaks in the relationship that could cause the pair to diverge permanently. Further analysis of the pair would include setting up trading rules and evaluate the profitability of the pair. We could also attempt to model the mean-reversion of the 'spread' to an Ornstein-Uhlenbeck process and examine the parameters. Furthermore, the pair analysis could be extended to a portfolio of assets, allowing for several cointegrating relationships.

Bibliography

Asteriou, D. and Hall, S. (2011), *Applied Econometrics*, Palgrave Macmillan, New York, 2nd Edition

Demiralp, S. and Hoover, K. (2003), "Searching for the Causal Structure of Vector Autoregression", *Oxford bulletin of Economics and Statistics* vol. 65, pp. 745-767

Do, B., Faff, R. and Hamza, K. (2006), "A New Approach to Modelling and Estimation for Pairs Trading", Working Paper, Monash University

Floros, C. and Vougas, D. (2008), "The Efficiency of Greek Stock Index Futures Market", *Managerial Finance* vol. 24, pp. 498-519

Gatev, E., Goetzmann, W. and Rouwenhorst, G. (2006), "Pairs Trading: Performance of a Relative-Value Arbitrage Rule", *The Review of Financial Studies* vol. 19, pp. 797-827

Granger, C. (1981), "Some Properties of Time Series Data and their Use in Econometric Model Specification", *Journal of Econometrics* vol. 16, pp. 131-130

Granger, C. and Engle, R. (1987), "Co-integration and Error Correction: Representation, Estimation, and Testing", *Econometrica* vol. 55, pp. 251-276

Hamilton, J. (1994), *Time Series Analysis*, Princeton University Press, New Jersey, 1st Edition

Jacobsen, B. (2008), "Demonstrating Error-Correction Modelling for Intraday Statistical Arbitrage", *Applied Financial Economics Letters* vol. 4, pp. 287-292

Johansen, S. (1998), "Statistical Analysis of Cointegration Vectors", *Journal of Economics Dynamics and Control* vol. 12, pp. 169-210.

Johansen, S. and Juselius, K. (1990), "The Maximum Likelihood Estimation and Inference on Cointegration – with Application to Demand for Money", *Oxford Bulletin of Economics and Statistics* vol. 52, pp. 169-210

Keating, J. (1992), "Structural Approaches to Vector Autoregressions", *Federal Reserve Bank of St. Louis Review, September/October* vol 72 pp. 37-57

Knif, J. and Pynnönen, S. (1999), "Local and Global Price Memory of International Stock Markets", *Journal of International Financial Markets, Institutions and Money* vol. 9, pp. 129-147

Lütkepohl, H. (2005), *New Introduction to Multiple Time Series Analysis*, Springer-Verlag, Berlin Heidelberg

MacKinnon, J. (1991), "Critical Values for Cointegration Tests", *Long-run Economic Relationships: Readings in Cointegration* Oxford: Oxford University Press

Manganelli, S. (2002), "Duration, Volume and Volatility Impact of Trades", *ECB Working Paper Series* no. 125

Morin, N. (2006), "Likelihood Ratio Tests on Cointegration Vectors, Disequilibrium Adjustment Vectors, and their Orthogonal Complements", *Finance and Economics Discussion Series*

Ross, S. (1976), “The Arbitrage Theory of Capital Asset Pricing”, *Journal of Economic Theory* vol. 13, pp. 341-360

Ross, S., Roll, R. and Chen, N. (1986), “Economic Forces and the Stock Market”, *The Journal of Business* vol. 59, pp. 383-403

Seymen, A. (2008), “A Critical Note on the Forecast Error Variance Decomposition”, *Centre for European Economic Research Discussion Paper* No. 08-065

Sims, C. (1980), “Macroeconomics and Reality”, *Econometrica* vol. 48, pp. 1-48

Stevans, L. and Sessions, D. (2010), “Calculating and Interpreting Multipliers in the Presence of Non-Stationary Time Series: The Case of U.S. Federal Infrastructure Spending”, *American Journal of Social and Management Sciences* vol. 1 No.1, pp. 24-38

Vidyamurthy, G. (2004), *Pairs Trading: Quantitative Methods and Analysis*, John Wiley & Sons, Inc., Hoboken, New Jersey

Appendix A

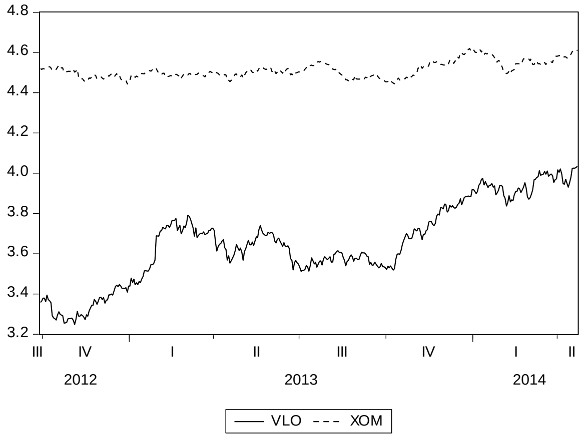

Figure 4. Graph shows VLO in solid line and XOM in dotted line. Prices in logarithm for the 409 observations within the period 9/27/2012 – 4/22/2014.

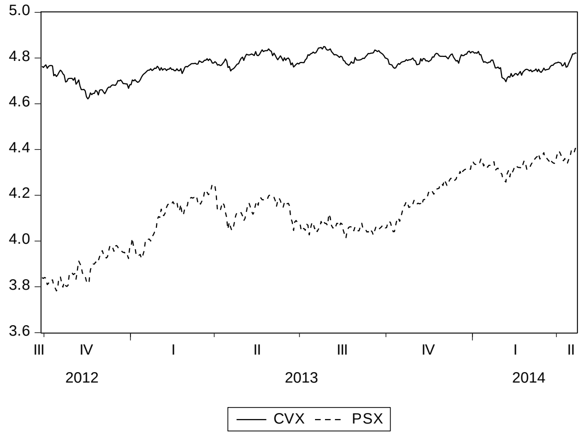

Figure 5. Graph shows CVX in solid line and PSX in dotted line. Prices in logarithm for the 409 observations within the period 9/27/2012 – 4/22/2014.

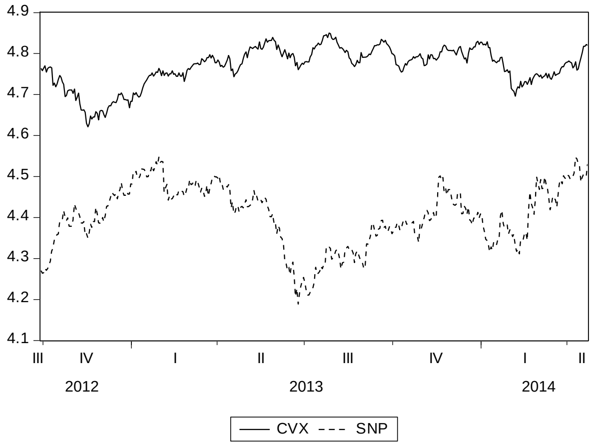

Figure 6. Graph shows CVX in solid line and SNP in dotted line. Prices in logarithm for the 409 observations within the period 9/27/2012 – 4/22/2014.

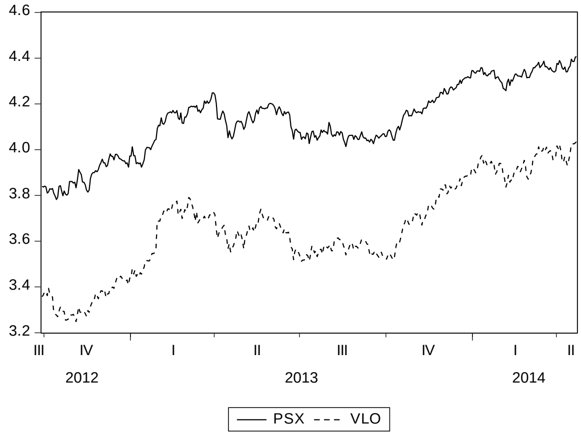

Figure 7. Graph shows PSX in solid line and VLO in dotted line. Prices in logarithm for the 409 observations within the period 9/27/2012 – 4/22/2014.

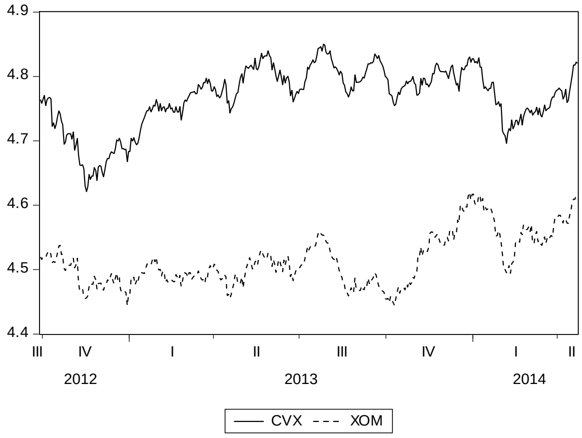

Figure 8. Graph shows CVX in solid line and XOM in dotted line. Prices in logarithm for the 409 observations within the period 9/27/2012 – 4/22/2014.

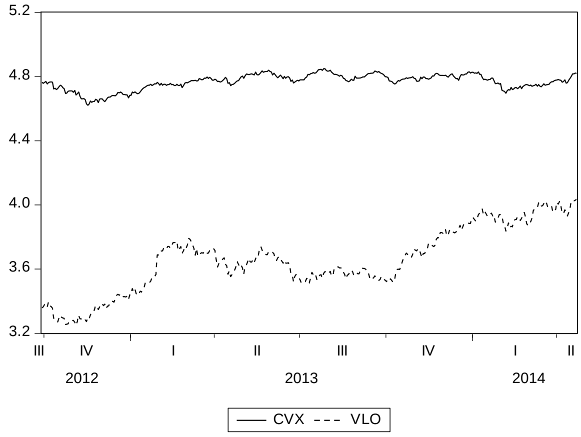

Figure 9. Graph shows CVX in solid line and VLO in dotted line. Prices in logarithm for the 409 observations within the period 9/27/2012 – 4/22/2014.

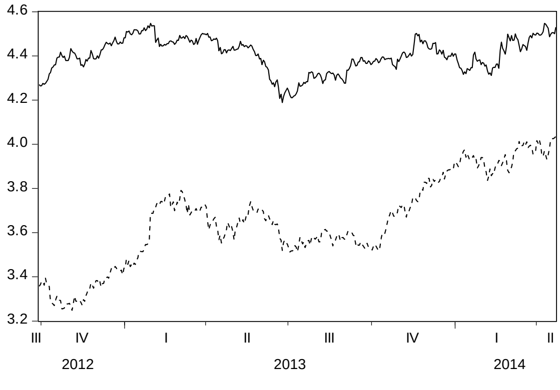

Figure 10. Graph shows SNP in solid line and VLO in dotted line. Prices in logarithm for the 409 observations within the period 9/27/2012 – 4/22/2014.

Figure 11. Graph shows SNP in solid line and XOM in dotted line. Prices in logarithm for the 409 observations within the period 9/27/2012 – 4/22/2014.

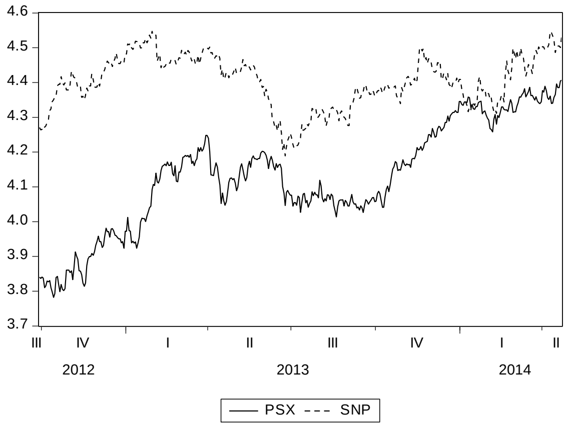

Figure 12. Graph shows PSX in solid line and SNP in dotted line. Prices in logarithm for the 409 observations within the period 9/27/2012 – 4/22/2014.

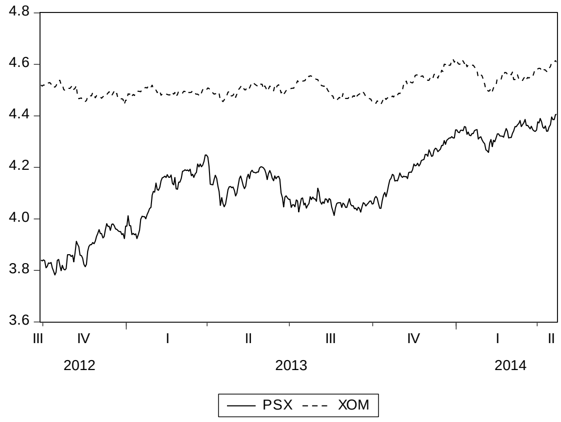

Figure 13. Graph shows PSX in solid line and XOM in dotted line. Prices in logarithm for the 409 observations within the period 9/27/2012 – 4/22/2014.

	CVX	PSX	SNP	VLO	XOM
Mean	4.768	4.128	4.403	3.657	4.514
Std. Dev.	0.048	0.154	0.077	0.202	0.041
Skewness	-0.822	-0.230	-0.441	-0.045	0.702
Kurtosis	3.208	2.341	2.565	2.242	2.718
Jarque-Bera	46.799	10.999	16.457	9.933	34.914
Observations	409	409	409	409	409
Unit root in y_t	0.373	0.683	0.183	0.848	0.508
Unit root in Δy_t	0.0000	0.0000	0.0000	0.0000	0.0000

Table 1. Descriptive statistics for each time series. 409 observations of daily closing prices are collected for each stock within the period 9/27/2012 – 4/22/2014 and the prices are transformed by the natural logarithm. Each series is tested for unit root by the Augmented Dickey-Fuller test for which the p-values are reported in the table. The test shows that all series contain a unit root in the level data, but we reject the null hypothesis that the series contain a unit root after taking the first difference.

Response of ΔPSX

Period	ΔVLO	ΔPSX
1	0.012711	0.011505
	(0.00074)	(0.00041)
2	0.000942	4.69E-05
	(0.00086)	(0.00087)
3	-4.83E-05	-0.001307
	(0.00088)	(0.00086)
4	-0.001303	-0.002393
	(0.00088)	(0.00085)
5	-0.000204	8.74E-05
	(0.00023)	(0.00033)
6	0.000107	0.000480
	(0.00022)	(0.00032)
7	0.000256	0.000381
	(0.00019)	(0.00026)
8	2.75E-05	-5.89E-05
	(7.0E-05)	(0.00012)
9	-4.45E-05	-0.000116
	(6.4E-05)	(0.00011)
10	-4.50E-05	-5.83E-05
	(5.3E-05)	(7.7E-05)

Table 2. Impulse response function for ΔPSX, illustrating the response to a one standard deviation shock in ΔVLO and itself with corresponding standard errors in parenthesis. 10 periods following date t are estimated and the standard errors are calculated using 50000 Monte Carlo repetitions.

Response of ΔVLO

Period	ΔVLO	ΔPSX
1	0.020391	0.000000
	(0.00073)	(0.00000)
2	0.000378	-0.000238
	(0.00103)	(0.00102)
3	-0.000231	0.000643
	(0.00106)	(0.00102)
4	-0.000432	-0.001529
	(0.00105)	(0.00101)
5	-8.50E-05	-0.000124
	(0.00021)	(0.00028)
6	-6.02E-05	0.000151
	(0.00019)	(0.00027)
7	0.000141	0.000239
	(0.00016)	(0.00020)
8	3.08E-05	-1.31E-07
	(5.5E-05)	(7.7E-05)
9	-1.03E-05	-4.50E-05
	(4.8E-05)	(6.8E-05)
10	-2.57E-05	-3.98E-05
	(3.9E-05)	(5.2E-05)

Table 3. Impulse response function for ΔVLO, illustrating the response to a one standard deviation shock in ΔPSX and itself with corresponding standard errors in parenthesis. 10 periods following date t are estimated and the standard errors are calculated using 50000 Monte Carlo repetitions.

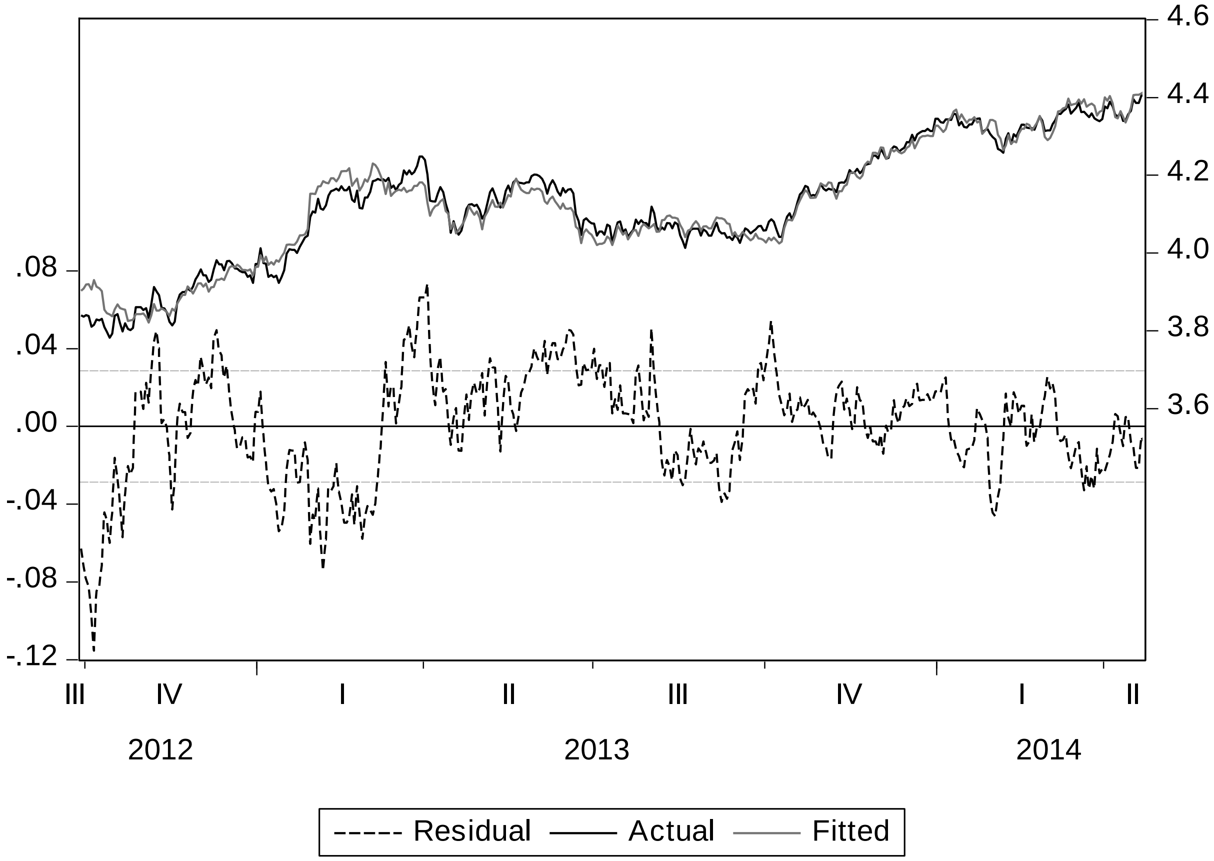

Figure 14. The upper graph shows the estimated values of the equation $PSX = \beta_0 + \beta_1 VLO + e_t$ against the actual values, the residuals are plotted in the lower graph. From the graphs it is seen how deviations from the long-run equilibrium are corrected by the mean-reverting behaviour of the series. The data comprise daily closing prices for the period 9/27/2012 – 4/22/2014 after taking the logarithm. See the Results section for further discussion.