



Munich Personal RePEc Archive

# **Cointegration Analysis of Oil Prices and Consumer Price Index in South Africa using STATA Software**

Sukati, Mphumuzi

Ministry of Agriculture

13 September 2013

Online at <https://mpra.ub.uni-muenchen.de/49797/>

MPRA Paper No. 49797, posted 14 Sep 2013 05:41 UTC

# **Cointegration Analysis of Oil Prices and Consumer Price Index in South Africa using STATA Software**

***By***

***Mphumuzi Angelbert Sukati\****

## **Abstract**

This paper investigates the concept of vector autoregression (VAR) and cointegration using a bivariate model of global oil prices and headline Consumer Price Index (CPI) in South Africa. The study aims to determine how much of inflation is driven by oil prices. Particular attention is paid to the theoretical underpinnings of cointegration analysis and the application of STATA software to undertake such analysis and perform test statistics. Contrary to the popular myth that a rise in global oil prices fuels inflation, this study has observed that global oil prices are not the drivers of inflation in South Africa. In this way, other macroeconomic indicators and policy developments need to be integrated in analyzing the determinants of South African inflation.

*Key words: Consumer Price Index, Oil Prices, Vector Autoregression, Cointegration, STATA Software, South Africa*

*\*Mphumuzi Sukati is an independent researcher working for the Ministry of Agriculture in the Kingdom of Swaziland. His research interests are in global food markets and drivers of food prices, mainly using CGE models and time series analysis. He holds a PhD in economics from the University of Nottingham.*

## 1. Introduction

Oil prices are a key driver of economic activities, with high prices perceived as being unfavorable for global economic growth. Popular myth is that high oil prices are generally associated with high consumer prices. The linkage between oil prices and CPI is especially important for the South African economy for two reasons. Firstly, in terms of income, South Africa is one of the most unequal countries in the world with a Gini coefficient of 63.1 in 2009<sup>1</sup>. This means that inflation disproportionately affect larger sectors of the population that do not have enough income to keep up with rising prices. Further, South Africa is an oil importing country and as such it is exposed to external shocks of rising oil prices. For these reasons, it is important to determine the role of imported inflation (via rising global oil prices) in the economy.

Many studies have used the concept of VAR and cointegration to investigate the link between oil prices and inflation. For example, Cologni and Manera (2005) used a structural cointegrated VAR model to study the effects of oil price shocks on output and prices in G-7 countries. Their key finding was that for most of the countries considered, there seems to be an impact of unexpected oil price shocks on interest rates, suggesting a contractionary monetary policy response directed to fight inflation.

Çelik and Akgül (2011) studied the relationship between CPI and oil prices in Turkey using the Vector Error Correction Model (VECM). Their study revealed that a 1% increase in fuel prices caused the CPI to rise by 1.26% with an approximate one year lag.

Ansar and Asaghar (2013) analyzed the impact of oil prices on stock exchange and CPI in Pakistan and concluded that there was no strong relationship between oil prices, CPI and KSE-100 Index.

LeBlanc and Chinn (2004) estimated the effects of oil price changes on inflation for the United States, United Kingdom, France, Germany and Japan using an augmented Phillips curve framework. Their study found that oil price increases of as much as 10 % will lead to direct inflationary increases of about 0.1-0.8 % in the U.S. and the E.U, which showed a modest response.

Cunado and Perezde (2003) analyzed the effect of oil prices on inflation and industrial manufacturing for several European countries for the period of 1960 to 1999. Their findings were that there is an asymmetric effect of oil price on production and inflation. Their findings suggest that there are expected differences in countries' responses to changes in global oil

---

<sup>1</sup> The Gini index measures the area between the Lorenz curve and a hypothetical line of absolute equality, expressed as a percentage of the maximum area under the line. Thus a Gini index of 0 represents perfect equality, while an index of 100 implies perfect inequality (<http://data.worldbank.org/indicator/SI.POV.GINI>).

prices depending on their macroeconomic status, whether the country is an oil importer or exporter, and the monetary policies adopted by a given country in response to global oil prices and other trends like exchange rate variations.

Niyimbanira (2013) has analyzed the relationship between oil prices and inflation in South Africa. The difference between his work and ours is that in his paper he modeled inflation as the dependant variable which is driven by oil prices. However, our approach firstly uses headline CPI and not inflation. Secondly, our approach tests the myth that high oil prices drive up prices in the economy, such that oil prices are the dependant variables in our analysis. In this way, there is no need to conduct an Engle Granger causality test.

Our approach is also supported by the work of Lescaoux and Mignon (2008) who noted that concerning the short term analysis, results indicate that when causality exists between oil prices and other macroeconomic variables, it generally runs from oil prices to the other considered variables.

Using the uncorrected or headline CPI and oil prices carries a risk of endogeneity. However, the direct link between oil price inflation and headline CPI is mainly through the price of petrol and this accounts for only 4.07% of the total CPI according to the CPI country weights of 2008 (Statistics South Africa, 2008)<sup>2</sup>. Further, cointegration analysis removes endogeneity and autocorrelation as we will discuss later.

Our analysis investigate the theoretical foundations of VAR processes and cointegration and their economic interpretation using the South African CPI monthly data from May 1987- 2013 and global oil prices for the same period.<sup>3</sup> Our study approach specifically highlight the STATA commands used in such analysis and supported by the theoretical foundations of the analytical framework, STATA language and test statistics used<sup>4</sup>.

---

<sup>2</sup> The complete contribution of goods and services to the CPI are as follows: Food and non alcoholic beverages 20.6%, alcoholic beverages and tobacco 6.26%, clothing and footwear 4.98%, housing and utilities 11.03%, household contents, equipment and maintenance 6.92%, health 1.67%, transport 20.04%, communication 3.52%, recreation and culture 4.43%, education 2.43%, restaurants and hotels 3.14% and miscellaneous goods and services 14.98%.

<sup>3</sup> In January 2013, Statistics SA revised the basket of goods and services used to measure CPI, in order to measure consumer inflation more precisely. Among these changes are: food prices were gathered from rural areas, the fixed fruit basket was altered to a seasonal one, reduced weightings of automobiles, furniture and appliances whose prices have been falling in previous years, and increased weight was given to petrol, transport costs, electricity, education and medical insurance (Dhliwayo, 2013)

<sup>4</sup> STATA statistical software is a complete, integrated statistical software package that is user friendly and readily available for purchase. It is versatile and has many techniques for data analysis for a wide range of fields. In economics it can be used to analyze for example survival models, panel data, generalized estimating equations, multilevel mixed models, models with sample selection, ARCH and GARCH, OLS, logit/probit regressions ANOVA/MANOVA, ARIMA and others. The software also facilitates the presentation of summary results in clear tabulated forms with strong graphical capabilities.

The rest of the paper is organized as follows; section 2 presents the modeling approach and tests for unit root. In section 3 we test for cointegration in the bivariate model and discuss the results. Section 4 presents the VECM estimates and discusses their implications while section 5 concludes.

## 2. Modeling approach

Before working with our bivariate model we have to test the variables for unit root. Following Hendry and Juselius (2000), data can be unit root i.e. integrated of degree 1 (denoted as I(1)). Such data cannot be used to investigate relationships between the variables because of spurious regression and OLS estimates become invalid.

However, data showing such properties can be made stationary by first differencing. If a series is such that its first difference is stationary (and has positive spectrum at zero frequency) then the series has an exact (or pure) unit root (Granger and Swanson, 1996).

The test for unit root starts with Equation 1 below, which is an autoregressive process of degree one, denoted as AR(1) process.

$$y_t = y_{t-1} + \varepsilon_t \tag{1}$$

With;

$$\varepsilon_t \sim IN [0, \sigma_\varepsilon^2]$$

From this equation it can be shown that subtracting  $y_t$  (as data) on both sides will result in a stationary process even though  $y_t$  is non stationary, i.e.

$$y_t - y_{t-1} = \Delta y_t = \varepsilon_t \tag{2}$$

Therefore;

$$\Delta y_t \sim IN [0, \sigma_\varepsilon^2]$$

Such differencing can be extended to twice-integrated series i.e. I(2), in which case it must be differenced twice to deliver a stationary process *etc.*

It is visually difficult to predict the nature of variables in an economic process i.e. whether they are stationary or not. Figure 1 below is a plot of monthly data of oil prices and CPI for the South

African economy from 1987 to 2013 (changed to natural logarithm) with 309 observations and their first difference.

The oil prices data has been obtained from Europe Brent Spot Price FOB (Dollars per Barrel) (<http://tonto.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=RB RTE&f=M>).

The headline CPI has been obtained from statistics South Africa, available at [www.statssa.gov.za](http://www.statssa.gov.za).

**Figure 1: Monthly data of oil prices and consumer price index for the South African economy from 1987 to 2013 and their first difference**



It is not obvious from the graphs if the processes are unit root or follow a random walk. To determine their true nature requires the application of the relevant statistical analytical tools. As mentioned above, taking first difference should result in stationary processes but still this

stationary cannot be determined from the graphs of the first differences above. We therefore need to undertake a formal test for unit root of the data, which we do next.

## 2.1 Testing for unit root

A test whether variable has a unit root (random walk) was developed by Dickey and Fuller (1979). The null hypothesis for this test is that the variable under analysis has a unit root.

To develop this test, we repeat the simple AR(1) model shown in Equation 1 above but including a constant term  $\alpha$ , time trend  $\partial t$  and a coefficient  $\rho$ , all that is important in the test statistics to be developed. This extended model is shown in Equation 3 below:

$$y_t = \alpha + \rho y_{t-1} + \partial t + \varepsilon_t \quad (3)$$

With  $\varepsilon_t$  as described previously.

The regression in Equation 3 can also be extended to remove possibilities of serial correlation in the lagged variables by taking  $p$  lagged differences and fitting a model as shown in Equation 4 below<sup>5</sup>:

$$\Delta y_t = \alpha + \beta y_{t-1} + \partial t + \varphi_1 \Delta y_{t-1} + \varphi_2 \Delta y_{t-2} + \dots + \varphi_k \Delta y_{t-p} + u_t \quad (4)$$

In STATA, these lags are specified in the *lags(p)* command. Equation 4 above is the augmented Dickey-Fuller regression.

STATA command<sup>6</sup> facilitates putting constraints on the augmented Dickey-Fuller regression. The *noconstant* option eliminates  $\alpha$  while the *trend* option includes the time trend  $\partial t$ . Equations 3 and 4 means that testing if  $\beta=0$  is the same as testing if  $\rho = 1$ , or that  $y_t$  is a unit root process.

Four possibilities can arise depending on constraints placed on the constant and time trend and these possibilities are summarized in Table 1<sup>7</sup>:

---

<sup>5</sup> This is the Augmented Dickey Fuller Regression that is used to test for unit root

<sup>6</sup> STATA commands will be shown in italics and underlined to differentiate them from the main text.

<sup>7</sup> The critical values of the Dickey Fuller test are adapted from tables in Fuller (1996) reported as one-sided critical values, with the p-values for the test of  $H_0$  against the one-sided  $H_a: \beta < 0$ , which is equivalent to  $\rho < 1$ , while MacKinnon (1994) reports the p-values on the basis of a regression surface.

**Table 1: Constrains on constant and time trend in augmented Dickey-Fuller unit root test**

Possibilities	Process under $H_0$	Regression restrictions	dfuller option
1	Random walk without drift	$\alpha=0, \partial=0$	<i>noconstant</i> (default)
2	Random walk without drift	$\partial=0$	<i>drift</i>
3	Random walk with drift	$\partial=0$	<i>trend</i>
4	Random walk with or without drift	none	

The choice of which constraint to choose depends on economic theory and trending behavior of the data. For example, if the data shows an increasing time trend, then case four may be preferred.

Visual inspection of the data in Figure 1 shows a clear upward trend for both the oil prices and the CPI. Because of this we will therefore use the *trend* option with *dfuller* to include a constant and a time trend in the augmented Dickey-Fuller regression.

## 2.2 Selecting the number of lags

The need for the lags arises because values in the past affect today's values for a given variable. This is to say the variable in question is persistent. There are various methods to determine how many lags to use. The two most commonly encountered in time series analysis are the Akaike Information Criterion (AIC) and the Schwarz' Bayesian Information Criterion (SBIC). These rules choose lag length  $p$  to minimize:  $\log(SSR(p)/n) + (p + 1)C(n)/n$ , where  $SSR(p)$  is the sum or squared residuals for the VAR with  $p$  lags and  $n$  is the number of observations, with  $C(n) = 2$  for AIC and  $C(n) = \log(n)$  for SBIC.

STATA *varsoc* command facilitates the calculation of these lags for the various selection criterions. *varsoc* reports the final prediction error (FPE), Akaike's information criterion (AIC), Schwarz's Bayesian information criterion (SBIC), the Hannan Quinn Information Criterion (HQIC), the log likelihood (LL) and likelihood-ratio (LR)<sup>8</sup>.

<sup>8</sup>  $LL = -\left(\frac{T}{2}\right) \{Ln(|\hat{\Sigma}|) + KLn(2\pi) + K\}$ , where T is the number of observations and K is the number of equations and  $\hat{\Sigma}$  is the maximum likelihood estimate of  $[\varepsilon_t \varepsilon_t']$ .

$LR(p) = 2\{LL(p) - LL(p - 1)\}$ , where  $p$  is the number of lags.

$FPE = |\sum \varepsilon| \left(\frac{T+\bar{m}}{T-\bar{m}}\right)^K$ , where  $\bar{m}$  is the average number of parameters over the K equations.

$AIC = -2\left(\frac{LL}{T}\right) + \frac{2tp}{T}$ .



Therefore to determine the number of lags to use in our bivariate model we run the *varsoc* command and the results are shown in Table 2 below.

**Table 2: Number of lags for a VAR of oil prices and CPI**

. varsoc LogOilPrice

Selection-order criteria  
Sample: 1987m9 - 2013m5

Number of obs = 309

lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	-73.3034				.094709	.480928	.485758	.49301
1	568.09	1282.8	1	0.000	.001501	-3.66401	-3.65435	-3.63985
2	579.81	23.44*	1	0.000	.0014*	-3.7334*	-3.7189*	-3.69715*
3	579.994	.36899	1	0.544	.001407	-3.72812	-3.7088	-3.67979
4	580.074	.15884	1	0.690	.001416	-3.72216	-3.69801	-3.66175

Endogenous: LogOilPrice  
Exogenous: \_cons

. varsoc LogCPI

Selection-order criteria  
Sample: 1987m9 - 2013m5

Number of obs = 309

lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	17.7336				.05254	-.108308	-.103478	-.096226
1	1450.76	2866.1	1	0.000	5.0e-06	-9.37708	-9.36742	-9.35292
2	1456.91	12.307	1	0.000	4.8e-06	-9.41044	-9.39594	-9.37419
3	1461.25	8.6732*	1	0.003	4.7e-06	-9.43203	-9.41271*	-9.3837*
4	1462.88	3.253	1	0.071	4.7e-06*	-9.43609*	-9.41193	-9.37568

Endogenous: LogCPI  
Exogenous: \_cons

The results reported in Table 2 above shows that the maximum number of lags to include for a VAR in oil prices is 2, as all the selection criterion show significant values at this lag. For the CPI, the maximum number of VAR lags is 3 as determined by significant levels of LR, HQIC and SBIC.

Having determined the number of lags to use in the VAR we then run the Augmented Dickey-Fuller test to determine if the two processes are unit root. For the Dickey-Fuller test, if the test statistics is smaller (larger) than the critical values we do not reject (reject) the null hypothesis of unit root in the data.

Computation of the unit root test statistics starts from the Augmented Dickey-Fuller expression as shown in Equation 4 above, i.e.

$$SBIC = -2 \left( \frac{LL}{T} \right) + \frac{\ln(T)}{T} t_p.$$

$$HQIC = -2 \left( \frac{LL}{T} \right) + \frac{2\ln\{\ln(T)\}}{T} t_p, \text{ where } t_p \text{ is the total number of parameters in the model.}$$

$$\Delta y_t = \alpha + \beta y_{t-1} + \partial t + \sum_{j=1}^p \varphi_j \Delta y_{t-j} + \varepsilon_t$$

The test statistic for  $H_0: \beta = 0$  is given by  $z_t = \hat{\beta} / \hat{\sigma}_{\beta}$ , where  $\hat{\sigma}_{\beta}$  is the standard error of  $\hat{\beta}$ .

STATA selects the augmented Dickey-Fuller test from a drop down menu. The results for unit root test for oil prices and CPI are shown in Table 3 below:

**Table 3: Augmented Dickey-Fuller test for unit root test of oil prices and CPI**

```
. dfuller LogOilPrice, lags(2)
```

Augmented Dickey-Fuller test for unit root                      Number of obs =                      310

	Test Statistic	1% Critical Value	Interpolated Dickey-Fuller 5% Critical Value	10% Critical Value
z(t)	-0.967	-3.455	-2.878	-2.570

Mackinnon approximate p-value for z(t) = 0.7652

```
. dfuller LogCPI, lags(3)
```

Augmented Dickey-Fuller test for unit root                      Number of obs =                      309

	Test Statistic	1% Critical Value	Interpolated Dickey-Fuller 5% Critical Value	10% Critical Value
z(t)	-4.151	-3.455	-2.878	-2.570

Mackinnon approximate p-value for z(t) = 0.0008

The results reported in Table 3 above show that oil prices follow a unit root process while CPI is a stationary process. This was not apparent from the graphical plot of the two processes as shown in Figure 1 above, and statistical analysis was necessary to determine stationary conditions of the two time series variables under analysis.

### 3. Testing for Cointegration

In a bivariate model with  $y_t$  and  $x_t$  variables, there exist a  $\beta$  such that  $y_t - \beta x_t$  is  $I(0)$  even though  $x_t$  and  $y_t$  are non stationary processes. This means the two variables are cointegrated or have a stationary long run relationship even though individually they are stochastic. Investigation of such processes can starts with the concept of VAR.

Generally, a VAR model with  $p$  lags can be represented as shown in Equation 5 below, which is an extension of Equation 3:

$$y_t = \rho_1 y_{t-1} + \rho_2 y_{t-2} + \dots + \rho_p y_{t-p} + \varphi \partial_t + \varepsilon_t \quad (5)$$

In the above equation,

$y_t$  is an  $k \times 1$  vector of  $I(1)$  variables

$\partial_t$  is an  $k \times 1$  vector of deterministic variable

$\rho_i (i=1 \dots p)$  is an  $k \times k$  and  $\varphi$  is an  $k \times n$  matrix of coefficients to be determined for a given data set

$\varepsilon_t$  is an  $k \times 1$  vector of identically and normally distributed errors with mean of zero and non-diagonal covariance matrix,  $\Sigma$ .

Given that the variables are cointegrated, equation 5 can be represented by an equilibrium correction model shown in Equation 6 below, which is an extension of Equation 4 discussed previously:

$$\Delta y_t = \alpha \beta' y_{t-p} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \delta t + v + \varepsilon_t \quad (6)$$

Economic importance is placed on the  $\alpha$  and  $\beta$  coefficients.  $\beta$  is an  $k \times r$  matrix of cointegrating vectors that explain the long-run relationship of the variables.  $\alpha$  is also an  $k \times r$  matrix that explains long run disequilibrium of the variables.  $\Gamma_i$  are coefficients that estimate short-run shock effects on  $\Delta y_t$  and these explain the differences between the short-run and long-run responses. It is important to note that for cointegration to exist, matrices  $\alpha$  and  $\beta$  should have reduced rank  $r$ , where  $r < k$ . The identification of the cointegrating vectors in STATA uses maximum likelihood (ML) method developed by Johansen (1988, 1991, and 1995).

$v$  and  $\delta t$  are the deterministic trend components which can be written as;

$$v = \alpha \mu + \gamma \quad (7)$$

$$\delta t = \alpha \rho t + \tau t \quad (8)$$

Where  $\mu$  and  $\rho$  are  $r \times 1$  vectors of parameters.  $\gamma$  and  $\tau$  are also  $k \times 1$  vectors of parameters.  $\gamma$  is orthogonal to  $\alpha \mu$  and  $\tau$  is orthogonal to  $\alpha \rho$ , such that  $\gamma' \alpha \mu = 0$  and  $\tau' \alpha \rho = 0$ .

Following this motivation, equation (6) can be written as VECM shown below;

$$\Delta y_t = \alpha(\beta y_{t-p} + \mu + \rho t) + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \gamma + \tau t + \varepsilon_t \quad (9)$$

There are 5 possibilities that the trend terms in equation (9) can take and they are the following;

**Possibility 1: unconstrained trend**

If there are no constraints on the trend parameters, this means that there exist quadratic trends in the levels of the variables but cointegrating equations are still stationary.

**Possibility 2: constrained trend,  $\tau = 0$**

Setting  $\tau = 0$ , means that there are only linear trends in the levels of the data, with no quadratic expressions. Cointegrating equations are still trend stationary.

**Possibility 3: unconstrained constant  $\tau = 0$  and  $\rho=0$**

Setting  $\tau = 0$  and  $\rho=0$  eliminates quadratic trends in the level variables and cointegrating variables are stationary around the constant means. However, since  $\gamma$  is not zero, this model still places a linear time trend in the levels of the data.

**Possibility 4: constrained constant,  $\tau = 0$ ,  $\rho=0$  and  $\gamma = 0$**

Inclusion of  $\gamma = 0$  eliminates all linear time trends in the levels of the data. Cointegrating equations are still stationary around a constant mean with no other trends.

**Possibility 5: constrained trend  $\tau = 0$ ,  $\rho=0$ ,  $\gamma = 0$  and  $\mu=0$**

This model eliminates all means or trends i.e. reduce them to zero. Cointegrating equations are stationary around a mean of zero.

In the determination of cointegration or long run relationship in our bivariate model, we still need to determine the number of lags to be included in the VECM, as it was the case for the VAR discussed previously.

Again in working with STATA we apply the varsoc commands to statistically select the number of lags of a VECM model, as is built on a paper by Tsay (1984) and Paulsen (1984).

The STATA output from running a varsoc command is shown in Table 4 below;

**Table 4: Lag determination of VECM of Oil Prices and CPI in South Africa**

`varsoc` LogOilPrice LogCPI

Selection-order criteria  
 Sample: 1987m9 - 2013m5  
 Number of obs = 309

lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	101.902				.001796	-.646617	-.636956	-.622453
1	2034.32	3864.8	4	0.000	6.8e-09	-13.1283	-13.0993	-13.0558
2	2050.52	32.394*	4	0.000	6.3e-09	-13.2072	-13.1589*	-13.0864*
3	2054.8	8.5606	4	0.073	6.3e-09	-13.2091	-13.1414	-13.0399
4	2059.37	9.1461	4	0.058	6.3e-09*	-13.2128*	-13.1258	-12.9953

Endogenous: LogOilPrice LogCPI  
 Exogenous: \_cons

The results above show that the HQIC, SBIC and the LR test all chose two lags. This means our oil prices and CPI bivariate model will be explained by two lags.

Once we have determined the number of lags, our next task is to test for cointegration amongst the variables.

STATA has inbuilt test for cointegration. This test is performed via the `vecrank` command. The command `vecrank` produces statistics used to determine the number of cointegrating equations in a VECM i.e., is used to determine the value of  $r$ .

The `vecrank` command solves for the rank of the model using three methods which are the Johansen's static method, the maximum eigenvalue statistic method and lastly the choice of  $r$  to minimize an information criterion. All these methods are based on Johansen's maximum likelihood (ML) estimator of the parameters of a cointegrating VECM.

The ML estimator is based on papers by Anderson (1984) and Johansen (1995) who derived the ML estimator for the parameters and LR test for inference on  $r$ . These LR tests are known as the trace statistics and the maximum-eigenvalue statistic. For the trace statistics as derived in Johansen (1995)<sup>9</sup>, large values are evidence against the null hypothesis that there are  $r$  or fewer cointegrating relations in the VECM.

For the eigenvalue statistics, letting  $\lambda_1, \dots, \lambda_k$  be  $k$  eigenvalues used in computing the log likelihood at the optimum and assuming that these eigenvalues are sorted from largest  $\lambda_1$  to smallest it follows that if there are  $r < k$  cointegrating equations,  $\alpha$  and  $\beta$  have rank  $r$  and the rest of the eigenvalues beyond  $r$ , i.e.  $\lambda_{r+1}, \dots, \lambda_k$  are zero.

<sup>9</sup> Johansen (1995) derives the distribution of the trace statistics as  $= -T \sum_{i=r+1}^K \ln(1 - \hat{\lambda}_i)$ , where  $T$  is the number of observations and the  $\hat{\lambda}_i$  are the estimated eigenvalues.

The test for cointegration therefore is based on the log likelihood findings in the model. The null hypothesis is that the log likelihood of the unconstrained model including the cointegrating equations is not significantly different from the log likelihood of the constrained model that does not include the cointegrating equations. If the two models are significantly different then we reject the null hypothesis and conclude that there is statistical evidence of cointegration amongst the variables. In other words, the test begins from  $r=0$  where there is no cointegration amongst the variables and accepts the first null hypothesis that is not rejected.

The results of the `vecrank` command are shown in Table 5 below:

**Table 5: Johansen test for Cointegration in oil prices and CPI in South Africa**

`vecrank LogOilPrice LogCPI`

Johansen tests for cointegration					
Trend: constant			Number of obs =		311
Sample: 1987m7 - 2013m5			Lags =		2
maximum rank	parms	LL	eigenvalue	trace statistic	5% critical value
0	6	2037.3234	.	54.4031	15.41
1	9	2062.7725	0.15097	3.5049*	3.76
2	10	2064.5249	0.01121		

The header produces information about the sample, the trend specification, and the number of lags included in the model. The main table contains a separate row for each possible value of  $r$ , the number of cointegrating equations. In our model, when  $r=2$ , all variables in the model are stationary.

In the above table, the trace statistics at  $r=0$  of 54.4031 exceeds its critical value of 15.41, we reject the null hypothesis of no cointegrating equations. The trace statistics at  $r=1$  of 3.5049 is less than the critical value of 3.76; we cannot reject the null hypothesis that there is one cointegration relationship between oil prices and CPI in South Africa.

As discussed above, another alternative to the determination of the rank of the model is the use of LR test that there are  $r+1$  cointegrating equations, which is the maximum eigenvalue test statistics. The results of this test are shown in Table 6 below:

**Table 6: Alternative test for cointegration in oil prices and CPI in South Africa**

```
. vecrank LogOilPrice LogCPI, lags(5) max levela notrace
```

Johansen tests for cointegration

Trend: constant Number of obs = 308  
Sample: 1987m10 - 2013m5 Lags = 5

---

maximum rank	parms	LL	eigenvalue	max statistic	5% critical value	1% critical value
0	18	2045.6655		21.5810	14.07	18.63
1	21	2056.456	0.06767	2.6794	3.76	6.65
2	22	2057.7957	0.00866			

---

The above output also confirms the number of rank to be 1 in the model at both 5% and 1% level. After determining that there is indeed a long run cointegration relationship between the prices of oil and CPI, the next step is to collect the VECM estimates.

To find out if we have correctly specified the number of cointegrating equations, we use the vecstable command. The companion matrix of a VECM with  $m$  endogenous variables and  $r$  cointegrating equations has  $m-r$  unit eigenvalues. The results of the stability conditions are shown in Table 7 and Figure 2 below:

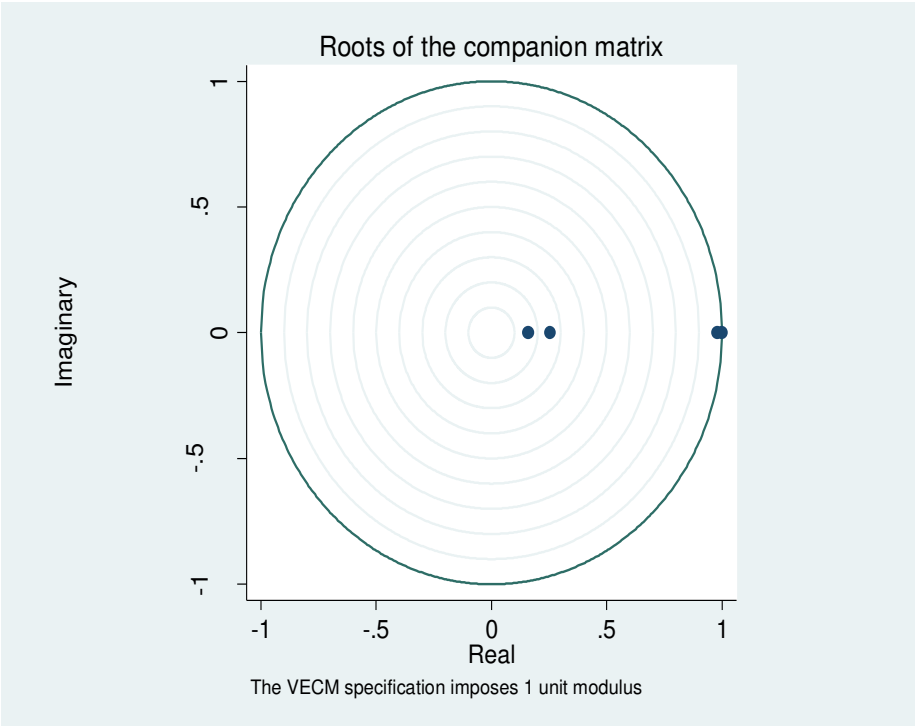
**Table 7: Stability test for the cointegration relationship between oil prices and CPI**

**Eigenvalue stability condition**

Eigenvalue	Modulus
1	1
.9797972	.979797
.2550938	.255094
.1582474	.158247

The VECM specification imposes a unit modulus.

**Figure 2: Stability test for the cointegration relationship between oil prices and CPI**



Our results show that 2 eigenvalues are strictly less than one, thus confirming stability of our bivariate model.

**4. Estimation of the VECM parameters**

For population of the VECM cointegration estimates, we use STATA vec command. vec simply runs a VAR of the cointegrated variables using Johansen’s (1995) maximum likelihood method as discussed above. From Equation 6 above, our estimates of interest are the matrix  $\beta$  which contain the cointegrating parameters,  $\alpha$  which is the adjustment coefficient and the short run coefficients,  $\Gamma$ .

The STATA vec command output is shown in Table 7 below:



**Table 7: VECM estimates for oil prices and CPI in South Africa**

vec LogOilPrice LogCPI

Vector error-correction model

Sample: 1987m7 - 2013m5  
 No. of obs = 311  
 AIC = -13.20754  
 Log likelihood = 2062.772  
 HQIC = -13.16428  
 Det(Sigma\_ml) = 5.94e-09  
 SBIC = -13.09931

Equation	Parms	RMSE	R-sq	chi2	P>chi2
D_LogOilPrice	4	.037114	0.0824	27.58133	0.0000
D_LogCPI	4	.002129	0.6687	619.5162	0.0000

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
D_LogOilPrice _ce1 L1.	-.0081944	.0064279	-1.27	0.202	-.0207928	.004404
LogOilPrice LD.	.2692116	.0551547	4.88	0.000	.1611104	.3773127
LogCPI LD.	-.4574795	.9869151	-0.46	0.643	-2.391798	1.476839
_cons	.0009504	.0042948	0.22	0.825	-.0074672	.0093681
D_LogCPI _ce1 L1.	.0025802	.0003687	7.00	0.000	.0018575	.0033029
LogOilPrice LD.	.0050637	.0031641	1.60	0.110	-.0011377	.0112651
LogCPI LD.	.1383145	.0566164	2.44	0.015	.0273484	.2492805
_cons	.0030185	.0002464	12.25	0.000	.0025356	.0035014

Cointegrating equations

Equation	Parms	chi2	P>chi2
_ce1	1	133.5188	0.0000

Identification: beta is exactly identified

Johansen normalization restriction imposed

beta	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
_ce1 LogOilPrice	1	.	.	.	.	.
LogCPI	-2.400303	.2077279	-11.56	0.000	-2.807442	-1.993164
_cons	2.236263	.	.	.	.	.

The short run estimates are read from the first part of Table 7 above. The two coefficients on L\_ce1 make up the long run disequilibrium adjustment matrix  $\alpha$  for our model. The second part of the Table presents the  $\beta$  parameters of the cointegrating vector. The short run coefficients

contained in  $\Gamma$  are collected from the row coefficients of the lagged differences (LD) and the constant matrix is read from the row of constants (`_cons`) in the first part of the table.

The matrix estimates are summaries below:

$$\hat{\alpha} = (-0.00819, 0.00258)$$

$$\hat{\beta} = (1, -2.4)$$

$$\hat{\nu} = (0.00095, 0.003019)$$

and

$$\hat{\Gamma} = \begin{pmatrix} 0.2692 & -0.4575 \\ 0.005064 & 0.13831 \end{pmatrix}$$

The assumption that the errors are independently, identically and normally distributed with zero mean and finite variance allows for the derivation of the likelihood function. If the errors do not come from a normal distribution but are just independently and identically distributed with zero mean and finite variance, the parameter estimates are still consistent, but they are not efficient. We use `vecnorm` command to test the null hypothesis that the errors are normally distributed and the results are shown in Table 9 below:

**Table 9: Test for distribution of the error terms of the bivariate oil and CPI model**

**Jarque-Bera test**

Equation	chi2	df	Prob > chi2
D_LogOilPrice	55.420	2	0.00000
D_LogCPI	25.360	2	0.00000
ALL	80.780	4	0.00000

**Skewness test**

Equation	Skewness	chi2	df	Prob > chi2
D_LogOilPrice	.05454	0.154	1	0.69455
D_LogCPI	.55487	15.958	1	0.00006
ALL		16.113	2	0.00032

**Kurtosis test**

Equation	Kurtosis	chi2	df	Prob > chi2
D_LogOilPrice	5.0652	55.266	1	0.00000
D_LogCPI	3.8518	9.402	1	0.00217
ALL		64.667	2	0.00000

The results above show that the errors are not normally distributed but show some evidence of skewness and kurtosis.

As alluded to earlier, one of the problems in statistics is autocorrelation amongst the variables. However, autocorrelation is not a problem in cointegration analysis in that, beginning with a simple OLS estimation of an AR(1) process,

$$y_t = \rho y_{t-1} + \varepsilon_t$$

Where  $\varepsilon_t$  are independently and identically distributed as  $N(0, \sigma^2)$  and  $y_0 = 0$ , the OLS estimate of  $n$  time series observations, the autocorrelation parameter  $\rho$  is given by:

$$\hat{\rho}_n = \frac{\sum_{t=1}^n y_{t-1} y_t}{\sum_{t=1}^n y_t^2}$$

If  $|\rho| < 1$ , then

$$\sqrt{n}(\hat{\rho}_n - \rho) \rightarrow N(0, 1 - \rho^2)$$

For the data to have unit root it must be that  $\rho=1$  so that the variance of the distribution is zero. However, in cointegration analysis, the data used has been corrected for unit root processes and therefore autocorrelation. This means that even though the CPI used in our data also contain the price of transport fuel, which is expected to have strong correlation to the price of oil, the removal of unit root and also autocorrelation in long run cointegration analysis removes this problem. This means our estimates of the cointegration relationship between oil prices and CPI based on our data are valid.

The adjustment parameters in general are small, implying a slow correction to equilibrium. The adjustment parameter on the CPI is small but significant, meaning that the CPI does not adjust contemporaneously to changes in the prices of oil as expected. The estimate of the coefficient for the CPI is 0.00258, meaning when the price of oil is high, the consumer price index slowly adjust upwards to match the oil prices, while the oil price attempts to adjust down, probably due to high commodity prices and reduced consumer demand thus leading to reduced demand for oil. Oil prices are not the only drivers of CPI but also other factors like exchange rate fluctuations and production cost like labor, electricity and land rent.

The long run relationship between oil prices (OP) and CPI is summarized in the equation below:

$$OP = -2.4CPI + 2.2$$

The long run relationship between oil prices and CPI is surprising in that it predicts that a 1% increase in the price of oil is associated with a 2.4% decrease in the CPI. However, this supports the observation the CPI is not only driven by oil prices but by other internal developments in the country. To better understand the relationship, other macroeconomic variables like the exchange rate and interest rate should be factored in.

It is known that when there is inflationary pressure monetary policy tend to increase interest rates to decrease demand, which may deflate the increase in prices. As observed by Tresor Economics (2012), oil price increases can also feed expectations of monetary tightening, which could reinforce the negative impact of the higher price of oil on aggregate supply and demand.

Other studies also support the observation that oil prices are not the sole drivers of inflation in the economy. Responses to oil price shocks are also country specific depending on internal variables that drive consumer demand.

Due to anti-inflationary tendencies, monetary authorities generally adopt contractionary policy after the impact of oil shocks, and this is a possible reason behind deepening of economic recession (Kuo-Wei and Yi-heng, 2011) and thus the need to also factor in GDP and employment rates in such analysis.

Blomberg and Harris (1995) observed that commodity prices should remain a secondary indicator of future inflation. Inflation hawks might more profitably focus on the unemployment rate and other indicators for signs of future inflation.

The implication for empirical work is that commodity prices' influence on consumer prices may not be captured adequately by mechanical pass-through effects from the commodity market to the final goods market and a richer, monetary-based characterization and modeling of their relationship is required (Brown and Cronin, 2007).

Further, oil prices could be associated with a weaker dollar, on which the price of oil is denominated. This could lead to a relatively strong Rand thus resulting in reduced inflation. This means that other macroeconomic indicators like the exchange rate should be included in the analysis of determinants of inflation. This observation is also supported by Thrung and Vinh (2011) who noted that Vietnamese economic activity is influenced more by changes of value of the Vietnamese currency than the fluctuations of oil prices.

As noted by Jordan (2011), rising oil prices in isolation are not recessionary (technically, they are not even inflationary, but rather represent a relative price increase). This therefore means that headline CPI is as a result of a cocktail of internal and external policies and shock rather than oil prices alone.

## **5. Conclusion**

This paper has investigated the long run relationship between global oil prices and the headline consumer price index in South Africa using STATA software. The paper has highlighted the flexibility and ease of using this software for cointegration analysis supported by the theoretical foundations of such analysis. It has also shown that there is a long run relationship between global oil prices and headline consumer price index in the country. Contrary to belief that oil prices drive up inflation, the paper has shown that in fact in the long run, increase in prices decrease inflation in South Africa. The analysis of the relationship between oil prices and inflation should therefore factor in other macroeconomic indicators like GDP, employment, exchange rate variations and interest rates for it to be conclusive.

## References

1. Anderson, T.W. (1984). *An introduction to multivariate statistical analysis*. (John Wiley & Sons, New York).
2. Ansar, I. and Asaghar, M. N. (2013). *The impact of oil prices on stock exchange and CPI in Pakistan*. IOSR Journal of Business and Management (IOSR-JBM). E-ISSN: 2278-487X. Volume 7, Issue 6 (Jan. - Feb. 2013), PP 32-36.
3. Blomberg, S. B. and Harris, E. S. (1995). *The commodity consumer price connection: Fact or Fable*. FRBNY Economic Policy Review / October 1995.
4. Brown, F. and Cronin, D. (2007). *Commodity Prices, Money and Inflation*. Working paper series No 738/March 2007. European Central Bank.
5. Çelik, T. and Akgül, B. (2011). *Changes in Fuel Prices in Turkey: An Estimation of the inflation effect using VAR Analysis*. Journal of Economic and Business, 14(2), 11-21.
6. Cologni, A. and Manera, M. (2005). *Oil Prices, Inflation and Interest Rates in a Structural Cointegrated VAR Model for the G-7 Countries*. NOTA DI LAVORO 101.2005. IEM – International Energy Markets.
7. Cunado, J. and Gracia, F. P. (2005). *Oil prices, economic activity and inflation: evidence for some Asian countries*. The Quarterly Review of Economics and Finance, 45, 65–83.
8. Dhliwayo, 2013: *Consumer Inflation Report – April 2013*
9. Dickey, D. A. and Fuller, W. A. (1979). *Distribution of the estimators for autoregressive time series with a unit root*. Journal of the American Statistical Association 74: 427-431.
10. Fuller, W.A. (1976). *Introduction to statistical time series*. John Wiley, New York.
11. Granger, C. J. W. and Swanson, N. R. (1996). *An Introduction to Stochastic Unit Root Processes*.
12. Hamilton, J. D. (1984). *Time Series Analysis*. Princeton: Princeton University Press
13. Hendry, D. F. and Juselius, K. (2000). *Explaining cointegration analysis: Part I*. Energy Journal, 21, 1-42.
14. World Bank Report: Available at <http://data.worldbank.org/indicator/SI.POV.GINI>
15. Energy Information Administration. Available at <http://tonto.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=RB RTE&f=M>

16. Johansen, S. (1991). *Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models*. *Econometrica* 59: 1551-80.
17. Johansen, S. (1988). *Statistical analysis of cointegrated vectors*. *Journal of Economic Dynamics and Controls* 12: 231-254.
18. Johansen, S. (1995). *Likelihood-based Inference in Cointegration Vector Autoregressive Models*. Oxford: Oxford University Press.
19. Johansen, S. and Juselius, K. (1992). *Testing structural hypothesis in a multivariate cointegration analysis of the PPP and the UIP for UK*. *Journal of Econometrics* 53 (1-3): 211-244.
20. Jordan, B. (2011). *Oil Price: Myth and Reality*. Framing the Economy: Nationwide Investment: First Quarter 2011. V4 Issue 2.
21. Kuo-Wei, C. and Yi-heng, T. (2011). *Pass-Through of Oil Prices to CPI Inflation in Taiwan*. *International Research Journal of Finance and Economics*. ISSN 1450-2887 Issue 69 (2011).
22. Le Blanc, M. and Chinn, M. (2004). *Do high oil price presage inflation? The evidence from G-5 countries*. Santa Cruz Center for International Economics Working Paper, WP1021.
23. Lescaroux, F. and Mignon, V. (2008). *On the Influence of Oil Prices on Economic Activity and Other Macroeconomic and Financial Variables*. CEPII, Working Paper N° 2008 - 05
24. MacKinnon, J. G. (1994). *Approximate asymptotic distribution functions for unit-root and cointegration test*. *Journal of Business and Economic Statistics* 12: 167-176.
25. Niyimbanira, F. (2013). *An overview of methods for testing short and long run equilibrium with time series data: Cointegration and error correction mechanisms*. *Mediterranean Journal of Social Sciences*, 4(4), 151-156.
26. Paulsen, J. (1984). *Order determination of multivariate autoregressive time series with unit root*. *Journal of time series analysis* 5: 115-127.
27. Statistics South Africa: available at <http://www.statssa.gov.za>.
28. Thrun, Le Viet and Vinh, N. T.T. (2011). *The impact of oil prices, real effective exchange rate and inflation on economic activity: novel evidence from Vietnam*. Discussion paper series: DP2011-09, RIEB. Kobe University.
29. Tresor Economics (2012). TRÉSOR-ECONOMICS No. 106 – October 2012: available at <http://www.tresor.economie.gouv.fr/tresor-economics>.
30. Tsay, R. S. (1984). *Order selection in nonstationary autoregressive models*. *Annals of Statistics* 12: 1425-1433.