A STUDY OF COINTEGRATION MODELS WITH APPLICATIONS

by

RAJAB SSEKUMA

submitted in accordance with the requirements for the degree of

MASTER OF COMMERCE

in the subject

STATISTICS

at the

UNIVERSITY OF SOUTH AFRICA

Supervisor: PROFESSOR M D JANKOWITZ

JUNE 2011

Abstract

This study estimates cointegration models by applying the Engle-Granger (1989) two-step estimation procedure, the Phillip-Ouliaris (1990) residual-based test and Johansen's multivariate technique. The cointegration techniques are tested on the $Raotbl3$ data set, the World Economic Indicators data set and the $UKpppuip$ data set using statistical software R. In the $Raotbl3$ data set, we test for cointegration between the consumption expenditure, and income and wealth variables. In the world economic indicators data set, we test for cointegration in three of Australia's key economic indicators, whereas in the $UKpppuip$ data set we test for the existence of long-run economic relationships in the United Kingdom's purchasing power parity. The study finds the three techniques not to be consistent, that is, they do not lead to the same results. However, it recommends the use of Johansen's method because it is able to detect more than one cointegrating relationship if present.

Keywords: cointegration; stationarity; nonstationarity; Augmented Dickey-Fuller test; error-correction model; unit root; Engle-Granger method; Phillip-Ouliaris methods; variance ration test; Johansen's method.

Contents

Acknowledgements

I would like to express my gratitude to all those who assisted me in completing this dissertation.

I am deeply indebted to my supervisor Prof. M Jankowitz whose help, stimulating suggestions and encouragement assited me throughout the research and the writing of this dissertation.

To my colleagues in the Department of Statistics, especially Dr Seeletse, thank you for your help, interest and support with regard to my research.

I would also like to thank my parents, who taught me the value of hard work by their own example. Accordingly, I would like to share this moment of happiness with my mother, my father and my brothers, who gave me enormous support throughout this research.

My acknowledgments would not be complete without the mention of Prof. P Ndlovu for constantly encouraging me.

Finally, I would like to thank all whose direct and indirect advice helped me to complete this dissertation.

DECLARATION BY STUDENT

I declare that the submitted work has been completed by me the undersigned and that I have not used any other than permitted reference sources or materials nor engaged in any plagiarism.

All references and other sources used by me have been appropriately acknowledged in the work.

I further declare that the work has not been submitted for the purpose of academic examination, either in its original or similar form, anywhere else.

Declared on the (date):

Signed:

Name: **R. SSEKUMA**

Chapter 1

Introduction

Empirical research in economics is based on time series. Therefore, it is standard to view time series as the realisation of a stochastic process. Model builders can use statistical inference in constructing and testing the equations that characterise relationships between economic variables. The two central properties of many economic time series are nonstationarity and time-volatility (Wei, 2006). These two properties have led to many applications in both economics and statistics.

Nonstationarity is a property common to many applied time series. This means that a variable has no clear tendency to return to a constant value or linear trend. It is generally correct to assume that economic processes have been generated by a nonstationary process and follow stochastic trends. One major objective of empirical research in economics it to test hypotheses and estimate relationships derived from economic theory, among other such aggregated variables (Pfaff, 2006).

The classical statistical methods used in building and testing large simultaneous equation models, such as Ordinary Least Squares (OLS), were based on the assumption that the variables involved are stationary. The problem is that the statistical inference associated with stationary processes is no longer valid if time series are a realisation of nonstationary processes. If time series are nonstationary it is not possible to use OLS to estimate their long-run linear relationships because it would lead to spurious regression. Spurious regression is a situation in which there appears to be a statistically significant relationship between variables but the variables are unrelated. A few decades ago the difficulty of nonstationarity was not well understood by model builders . However, this is no longer the case because the technique of cointegration has been introduced according to which models containing nonstationary stochastic variables can be constructed in such a way that the results are both statistically and economically meaningful.

Cointegration is an econometric concept which mimics the existence of a long-run equilibrium among economic time series. If two or more series are themselves nonstationary, but a linear combination of them is stationary, then they are said to be cointegrated (Wei, 2006). We should be concerned about cointegration because it is a possible solution to nonstationarity found in many economic time series, and if time series are nonstationary the assumptions upon which OLS estimation rest are violated, rendering its application inappropriate.

Previously, the usual procedure for testing hypotheses concerning the relationship between nonstationary variables was to run OLS regressions on data which had initially been differenced. Data are differenced in order to reduce nonstationary series to stationarity. Although this method is correct in large samples, it may give rise to misleading inferences or spurious regressions in small samples. Moreover, estimation of a single equation framework with integrated or nonstationary variables tends to create the following problems: non-standard distribution of the coefficient estimates generated by the process not being stationary, explanatory variables generated by the process that display autocorrelation, the existence of more than one cointegrated vector and tendency to weak exogeneity (Banerejee et al., 1993).

The remedy for problematic regressions with integrated variables is to test for cointegration and to estimate a vector error-correction model to distinguish between short-run and long-run responses, since cointegration provides more powerful tools when the data sets are of limited length. The technique of cointegration and the error-correction model have both been used before in modelling a number of studies, for example, in modelling Danish gasoline demand (Bentzen et al., 1995), the road transport energy demand for Australia (Samimi, 1995), demand for coal in India (Kulshreshtha and Parikh, 1999), coal demand in China (Chan and Lee, 1997) and the United Kingdom's final user energy demand (Fouguet et al., 1997). In these studies, the multivariate Johansen cointegrating framework was used to ascertain the cointegrating rank.

The main interest in this study is to estimate cointegrating models and explain their applications to different sets of data using the three main methods of testing for cointegration and related relationships. The results of this study can be used to assess the impact of a temporary or permanent shock on economic variables in an economy. These methods are

- the Engle-Granger method (Engle-Granger, 1987)

- the Phillips-Ouliaris residual-based tests, namely a variance ratio and a multivariate trace

statistic (Phillips-Ouliaris, 1988)

- the Johansen's procedure which builds cointegrated variables directly on maximum likelihood estimation instead of relying on OLS procedures (Johansen and Juselius, 1988).

In this study, the above techniques for testing for cointegration are explored using statistical software R. These methods are tested on three data sets:

- the Raotbl3 data set, extracted from the R package urca

- the World Economic Indicators case study on Australia's economic indicator extracted from the United Nations World Economic Indicators

- the United Kingdom purchasing power parity (UKpppuip) data set extracted from the urca R package.

These data sets were selected because of the availability of data from the urca R package and/or on the Internet. In the $Raotbl3$ data set, we test for cointegration between the consumption expenditure, income and wealth series; in the world economic indicators data set, we test for cointegration in three of Australia's key economic indicators; whereas in the $UKpppuip$ data set we test for the existence of long-run economic relationships in the United Kingdom purchasing power parity.

Not all variables in the analysed data sets showed the existence of long-run relationships, and in the World Economic Indicators data set series was cointegrated on second order differencing. These data sets are attached in Appendices A, B and C, respectively.

Chapter 2

A theoretical overview

2.1 Stationarity and nonstationarity

Time series data consist of observations, which are considered as a realisation of random variables that can be described by some stochastic processes. The concept of "stationarity" is related to the properties of this stochastic processes. In this study, the concept of "weak stationarity" is adopted; meaning that the data are assumed to be stationary if the means, variances and covariances of the series are independent of time, rather than the entire distribution.

Nonstationarity in a time series occurs when there is no constant mean μ, no constant variance σ_t^2, or both of these properties. It can originate from various sources but the most important one is the *unit root*.

2.2 Unit root

Any sequence that contains one or more characteristic roots that are equal to one is called a unit root process. The simplest model that may contain a unit root is the AR(1) model.

Consider the autoregressive process of order one, AR(1), below:

$$Y_t = \phi Y_{t-1} + \epsilon_t \tag{2.2.1}$$

where ϵ_t denotes a serially uncorrected white noise error term with a mean of zero and a constant variance.

If $\phi = 1$, equation 2.2.1 becomes a random walk without drift model, that is, a nonstationary process. When this happens, we face what is known as the unit root problem. This means that we are faced with a situation of nonstationarity in the series. If, however, $\phi < 1$, then the series Y_t is stationary. The stationarity of the series is important because correlation could persist in nonstationary time series even if the sample is very large and may result in what is called spurious (or nonsense) regression (Yule, 1989). The unit root problem can be solved, or stationarity can be achieved, by differencing the data set (Wei, 2006).

2.3 The augmented Dickey-Fuller (ADF) test

In section 2.2, it was stated that, if $\phi = 1$, equation 2.2.1 becomes a random walk model without drift, which is known as a nonstationary process. The basic idea behind the ADF unit root test for nonstationarity is to simply regress Y_t on its (one period) lagged value Y_{t-1} and find out if the estimated ϕ is statistically equal to 1 or not. Equation 2.2.1 can be manipulated by subtracting Y_{t-1} from both sides to obtain

$$Y_t - Y_{t-1} = (\phi - 1)Y_{t-1} + \epsilon_t \qquad (2.3.1)$$

which can be written as

$$\triangle Y_t = \delta Y_{t-1} + \epsilon_t \qquad (2.3.2)$$

where $\delta = (\phi - 1)$, and Δ is the first difference operator.

In practice, instead of estimating equation 2.2.1, we shall estimate equation 2.3.2 and test for the null hypothesis of $\delta = 0$ against the alternative of $\delta \neq 0$. If $\delta = 0$, then $\phi = 1$, meaning that we have a unit root problem and the series under consideration is nonstationary. It should be noted that under the null hypothesis $\delta = 0$, the t-value of the estimated coefficient of Y_{t-1} does not follow the t-distribution even in large samples (Erdogdu, 2007). This means that the t-value does not have an asymptotic normal distribution. The decision to reject or not to reject the null hypothesis of $\delta = 0$ is based on the Dickey-Fuller (DF) critical values of the τ(tau) statistic. The DF test is based on an assumption that the errors of term ϵ_t are uncorrelated.

However, in practice, the errors of the term in the DF test usually show evidence of serial correlation. To solve this problem, Dickey and Fuller have developed a test know as the Augmented Dickey-Fuller (ADF) test. In the ADF test, the lags of the first difference are included in the

regression equation in order to make the error term ϵ_t white noise and, therefore, the regression equation is presented in the following form:

$$\Delta Y_t = \delta Y_{t-1} + \alpha_i \sum_{i=1}^{m} \Delta Y_{t-i} + \epsilon_t. \tag{2.3.3}$$

To be more specific, the intercept may be included, as well as a time trend t, after which the model becomes

$$\Delta Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + \alpha_i \sum_{i=1}^{m} \Delta Y_{t-i} + \epsilon_t. \tag{2.3.4}$$

The testing procedure for the ADF unit root test is applied to the following model

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_{j=1}^{\rho} \delta_j \Delta y_{t-j} + \epsilon_{it} \tag{2.3.5}$$

where α is a constant, β the coefficient on a time trend series, γ the coefficient of y_{t-1}, ρ is the lag order of the autoregressive process, $\Delta y_t = y_t - y_{t-1}$ are first differences of y_t, y_{t-1} are lagged values of order one of y_t, Δy_{t-j} are changes in lagged values, and ϵ_{it} is the white noise.

The ADF test can be tested on at least three possible models:

(i) A pure random walk without a drift. This is defined by using the constraint $\alpha = 0$, $\beta = 0$ and $\gamma = 0$ in equation 2.3.5. This leads to the equation

$$\Delta y_t = \Delta y_{t-1} + \epsilon_t. \tag{2.3.6}$$

Equation 2.3.6 is a nonstationary series because its variance grows with time (Pfaff, 2006).

(ii) A random walk with a drift. This is obtained by imposing the constraint $\beta = 0$ and $\gamma = 0$ in equation 2.3.5, which yields to the equation

$$\Delta y_t = \alpha + \Delta y_{t-1} + \epsilon_t. \tag{2.3.7}$$

(iii) A deterministic trend with a drift. For $\beta \neq 0$, equation 2.3.5 becomes the following deterministic trend with a drift model

$$\Delta y_t = \alpha + \beta t + \Delta y_{t-1} + \epsilon_t. \tag{2.3.8}$$

The sign of the drift parameter (α) causes the series to wander upward if positive and downward if negative, whereas the size of the absolute value affects the steepness of the series (Pfaff, 2006).

The parameter of interest in the ADF model is γ. For $\gamma = 0$, the y_t sequence contains the unit root and hence is integrated of order $d = 1$.

The choice of the number of lags (p) in this study is based on the significant lag of the autocorrelation function (ACF) and the partial autocorrelation function $(PACF)$ plots. We take the value of p to be the number of lags at which the ACF cuts off or the number of lags of the $PACF$ that are significant.

The test procedure for unit roots is similar to statistical tests for hypothesis, that is:

(i) Set the null and alternative hypothesis as

$$H_0 : \gamma = 0 \tag{2.3.9}$$

$$H_1 : \gamma < 0 \tag{2.3.10}$$

(ii) Determine the test statistic using

$$F_\tau = \frac{\hat{\gamma}}{SE(\hat{\gamma})} \tag{2.3.11}$$

where $SE(\hat{\gamma})$ is the standard error of γ.

(iii) Compare the calculated test statistic in 2.3.11 with the critical value from Dickey-Fuller table to reject or not to reject the null hypothesis.

(iv) The ADF test is a lower-tailed test, so if F_τ is less than the critical value, then the null hypothesis of unit root is rejected and the conclusion is that the variable of the series does not contain a unit root and is nonstationary.

The DF and ADF tests are similar since they have the same asymptotic distribution. Although there are numerous unit root tests, such as the Phillips-Perron test and the Schmidt-Phillips test, the most notable and commonly used is the ADF test, which will be used in this study.

2.4   Cointegration tests

On the basis of the theory that integrated variables of order one, $I(1)$, may have a cointegration relationship, it is crucial to test for the existence of such a relationship. If a group of variables are individually integrated of the same order and there is at least one linear combination of these

variables that is stationary, then the variables are said to be cointegrated. The cointegrated variables will never move far apart, and will be attracted to their long-run relationship. Testing for cointegration implies testing for the existence of such a long-run relationship between economic variables. This study considers a number of cointegration tests, namely the Engle-Granger method commonly known as the two-step estimation procedure, the Phillips-Ouliaris methods and the Johansen's procedure.

2.4.1 Engle-Granger method

As we have stated, the regression of nonstationary series on other series may produce spurious regression. If each variable of the time series data is subjected to unit root analysis and it is found that all the variables are integrated of order one, $I(1)$, then they contain a unit root. There is a possibility that the regression can still be meaningful (i.e. not spurious) provided that the variables cointegrate. In order to find out whether the variables cointegrate, the least squares regression equation is estimated and the residuals (the error term) of the regression equation are subjected to unit root analysis. If the residuals are stationary, that is $I(0)$, it means that the variables under study cointegrate and have a long-term or equilibrium relationship. The Engle-Granger method is based on the idea described in this paragraph.

In the two-step estimation procedure, Engle-Granger considered the problem of testing the null hypothesis of no cointegration between a set of variables by estimating the coefficient of a statistic relationship between economic variables using the OLS and applying well-known unit root tests to the residuals to test for stationarity. Rejecting the null hypothesis of a unit root is evidence in favour of cointegration.

In the literature, there are a number of studies that apply Engle-Granger's two-step estimation procedure. A summary of some of them follows:

The study by Lee (1993) applied the two-step estimation procedure similar to that used by Engle and Granger to examine cointegration relationships between total consumption and disposable income on Japanese data from January 1961 to April 1987. This study investigated whether seasonality in income cointegrates with that in consumption and identifies reasons in support of the empirical cointegration relationship. The results indicated that income and consumption series are integrated of order one at both the long-run (yearly) and the short-run (seasonal) frequency. Results further indicated that both income and consumption series are nonstationary and that

the seasonal pattern has a significant variation over the period, although the seasonal pattern for consumption was more regular.

There is vast literature that explores whether spot and future prices for oil are linked in a long-run relationship. One particular study was undertaken by Maslyuk and Smyth (2009) to examine whether crude oil spot and future prices of the same and different grades cointegrate. The null hypothesis of no cointegration was tested against the alternative of cointegration in the presence of a regime shift on series monthly data from the United States Western Telematic Inc. and United Kingdom Brent using the two-step estimation procedure. The results revealed that there is a cointegration relationship between spot and future prices of crude oil of the same grade, as well as spot and future prices of different grades. Results further indicated that spot and future prices are governed by the same set of fundamentals, such as the exchange rate of the US dollar, macro-economic variables, and the demand and supply conditions, which are similar and interrelated for crude oil in North American and European markets.

The relationship between female labour force participation (FLFP) and total fertility rate (TFR) has received a lot of attention in the literature on demography and economics. A study of twenty-eight countries from the Organisation for Economic Cooperation and Development (OECD) using panel unit roots, panel cointegration and a panel Granger causality framework revealed a long-run relationship between FLFP and TFR. This study found that there is either uni-directional long-run Granger causality running from FLFP to the TFR or bi-directional Granger causality between the two variables depending on how female labour participation rate is measured and the time period. It was concluded that there is an inverse relationship between FLFP and TFR (Maslyuk and Smyth, 2010).

2.4.2   Phillips-Ouliaris methods

Phillips-Ouliaris introduced two residual-based tests namely:

- the variance ratio test

- the multivariate trace statistics.

These residual-based tests are used in the same way as the unit root tests, but the data are the residuals from the cointegrating regression. These tests seek to test a null hypothesis of no cointegration against the alternative of the presence of cointegration using scalar unit root tests applied

to the residuals. Phillips-Ouliaris methods are based on residuals (differences between the observed and expected values) of the first order autoregression, AR(1), equation. The multivariate trace statistics has the advantage over the variance ratio test in that it is invariant to normalisation, that is, whichever variable is taken to be the dependent variable, the test will yield the same results (Pfaff, 2006).

In the literature, there are no studies directly linked to the application of the Phillips-Ouliaris cointegration test only. However, there are a few studies in which cointegration has been tested using other techniques including the Phillips-Ouliaris methods. These studies will now be discussed.

The study by Cancer (1998) developed the asymptotic theory for residual-based tests and quasi-likelihood ratio tests for cointegration under the assumption of infinite variance errors. He extended the results of the Phillips-Ouliaris methods, which were derived under the assumption of square-integrable errors. He also investigated whether the Phillips-Ouliaris methods are robust to infinite variance errors. The results showed that, regardless of the index of stability α, the residual-based tests are more robust to infinite variance errors than the likelihood ratio-based tests.

Theoretical models of pricing-to-market suggest that the profile of economic exposure may be asymmetric between periods when the real exchange rate appreciates and is symmetric between periods when it depreciates. This hypothesis was tested using time-series data on export prices for eight commodities exported from the United Kingdom to the United States during the period 1981 to 1988 (Kanas, 1997). During this period, there was a long-term real depreciation of the pound against the dollar from January 1985 up to April 1988, followed by a long-term real appreciation from February 1985 to April 1988. Kanas (1997) tested for cointegration between the export price of each commodity and the real exchange rate applying the Phillips-Ouliaris methods. The results were generally in favour of this hypothesis.

The study by Choi (1994) of spurious regression and residual-based tests for cointegration when regressors are cointegrated, analysed the asymptotic null distribution of residual-based cointegration tests such as the Phillips' $\hat{Z}_\alpha$ and the ADF tests when regressors are cointegrated in comparison to the Phillips-Ouliaris methods.

The results showed that the null distributions of residual-based cointegration tests differed from those derived from the use of the Phillips-Ouliaris methods. The practical implication of these

results is that we need to test not only for the presence of a unit root for individual series, but also for the presence of cointegrating vectors for the regressors prior to performing residual-based tests for cointegration.

2.4.3 Johansen's procedure

Johansen's procedure builds cointegrated variables directly on maximum likelihood estimation instead of relying on OLS estimation. This procedure relies heavily on the relationship between the rank of a matrix and its characteristic roots. Johansen derived the maximum likelihood estimation using sequential tests for determining the number of cointegrating vectors. His method can be seen as a secondary generation approach in the sense that it builds directly on maximum likelihood instead of partly relying on least squares. In fact, Johansen's procedure is nothing more than a multivariate generalisation of the Dickey-Fuller test. Consequently, he proposes two different likelihood ratio tests namely

- the trace test

- the maximum eigenvalue test.

This procedure is a vector cointegration test method. It has the advantage over the Engle-Granger and the Phillips-Ouliaris methods in that it can estimate more than one cointegration relationship, if the data set contains two or more time series. In the literature, studies on Johansen's procedure include the following.

The ongoing debate among energy economists about the relationship between energy use and output growth has led to the emergence of many views. One investigation on the causal interaction between energy use and output growth for Canada was undertaken by Khalifa and Sakka (2004). They used time series properties to develop a vector error-correction (VECM) model to test for multivariate cointegration and Granger-causality. The empirical finds from this analysis indicated that output growth, capital, labour and energy use share two common stochastic trends. In particular, output growth and energy use were found to be moving together towards a stable long-run equilibrium relationship, that is, consistent with causality running in both directions.

Masih et al. (1996) used Johansen's cointegration analysis to study the relationship between energy use and gross domestic product (GDP) in a group of six Asian countries, including India, Pakistan, Malaysia, Singapore, Philippines and Indonesia. The results indicated that there were cointegration relationships in energy use and GDP among countries like India, Pakistan and Indonesia.

However, no cointegration was found in the case of Malaysia, Singapore and the Philippines. The flow of causality was found to be running from energy to GDP in India and GDP to energy in Pakistan and Indonesia.

Yang (2000) considered the causal relationship between energy use and GDP in Taiwan. Using different measures of energy consumption, he found a bi-directional causality between energy and the GDP. This result contradicts that of Cheng and Lai (1997), who found that there is a uni-directional causal relationship from GDP.

Belloumi (2009) applied the Johansen's cointegration procedure to examine the causal relationship between per capita energy consumption (PCEC) and per capita gross domestic product (PCGDP) for Tunisia during the 1971-2004 period. In order to test for Granger-causality in the presence of cointegration among variables, a vector error-correction model (VECM) was used instead of a vector autoregressive (VAR) model. His results indicate that the PCGDP and PCEC for Tunisia are related by one cointegrating vector and there is a long-run bi-directional causal relationship between the two series and a short-run uni-directional causality from energy to GDP. The source of causation in the long-run was found to be the error correction terms in both directions. Hence, an important policy implication resulting from this analysis is that energy can be considered as a limiting factor to GDP growth in Tunisia. It was argued that Tunisia's economy is energy dependent and is relatively vulnerable to energy shocks.

Finally, Asufu (2000) tested the cointegration relationship between energy use and income in four Asian countries using cointegration, Johansen's procedure and error-correction analysis. He found that cointegration runs from energy to income in India and Indonesia, and that there is a bi-directional causality in Thailand and the Philippines.

Chapter 3

Testing for cointegration

In this chapter we discuss the various methods of testing for cointegration which include

- the two-step estimation procedure developed by Engle and Granger in 1987

- the Phillips-Ouliaris residual-based tests

- the Johansen procedure which builds cointegrated variables directly on maximum likelihood estimation instead of relying on OLS estimates.

3.1   The Engle-Granger method

Engle and Granger developed this crucial technique in 1987. This technique entails cointegrated variables which are discussed at length including a proof of Granger's representation theorem, which connects the moving average, the autoregressive, and the error correcting representation for cointegrated systems.

According to both statisticians, the steps for determining whether two integrated variables cointegrate of the same order are the following:

- pre-test each variable to determine its order of integration, and

- estimate the error-correction model.

If the integrated variables are found to be integrated to the same order, then it must be tested whether these variables are cointegrated (Johansen, 1988) .

3.1.1   Pre-test each variable

Pre-testing each variable aims at determining the order of integration of each variable. By definition, cointegration necessitates that two variables be integrated of the same order. This is done using the Augmented Dickey-Fuller (ADF) unit root test to infer the number of unit roots (if any) in each of the variables under investigation. The testing procedure for the ADF unit root test is applied to the following model

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_{j=1}^{\rho} \delta_j \Delta y_{t-j} + \epsilon_{it} \tag{3.1.1}$$

where α is a constant, β the coefficient on a time trend series, γ the coefficient of y_{t-1}, ρ is the lag order of the autoregressive process, $\Delta y_t = y_t - y_{t-1}$ are first differences of y_t, y_{t-1} are lagged values of order one of y_t, Δy_{t-j} are changes in lagged values, and ϵ_{it} is the white noise.

Once the hypothesis of the unit root test is rejected, we estimate the long-run equilibrium relationship in the form of an OLS regression line

$$y_t = \beta_0 + \beta_1 x_t + \epsilon_t \tag{3.1.2}$$

where β_0 is the y- intercept, β_1 is the slope, and ϵ_t is the error term.
The parameter estimates in equation 3.1.2 are estimated from

$$\hat{\beta}_1 = \frac{\sum(x_t - \overline{x}_t)(y_t - \overline{y}_t)}{\sum(x_t - \overline{x}_t)^2} \tag{3.1.3}$$

where $\overline{x}_t$ and $\overline{y}_t$ is the mean of x_t and y_t respectively.
The value of β_0 is estimated from

$$\hat{\beta}_0 = \overline{y}_t - \hat{\beta}_1 \overline{x}. \tag{3.1.4}$$

The estimated regression line is then given in the form

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x. \tag{3.1.5}$$

If the variables cointegrate, an OLS regression equation 3.1.5 yields a "super-consistent" estimator (Enders, 2004). This means that there is a strong linear relationship between the variables under study. The strong linear relationship can be tested in either of the following ways.

(a) The value of $\hat{\beta}_1$ falls between 0.5 and 1.

(b) The plot of y_t against x_t shows coordinates appearing in an increasing or decreasing direction.

In order to determine if variables cointegrate, we test for unit roots on the residual sequence in the equation 3.1.2 using the ADF test. The residual sequence, denoted by ϵ_t is a series of estimated values of the deviation from the long-run relationship. They are estimated from

$$\epsilon_t = y_t - \hat{y}_t \tag{3.1.6}$$

where $\hat{y}_t$ are values from the predicted equation 3.1.5. Testing for unit roots on residuals aims at determining whether these deviations are stationary or not. If they are stationary, then the series cointegrate. If the residuals are not stationary, there is no cointegration. The ADF test is performed on the following model

$$\Delta\hat{\epsilon}_t = a_1\hat{\epsilon}_{t-1} + \varepsilon_t \tag{3.1.7}$$

where $\Delta\hat{\epsilon}_t$ are the estimated first differenced residuals, $\hat{\epsilon}_{t-1}$ are the estimated lagged residuals, a_1 is the parameter of interest representing the slope of the line and ε_t are errors obtained in fitting both differenced residuals.

Since the ϵ_t sequences are residuals from a regression equation, there is no need to include the intercept term in equation 3.1.7. To test the hypothesis on a_1 to determine whether the residuals are stationary, we follow the following steps.

(i) Set both the null and alternative hypothesis as

$$H_0 : a_1 = 0 \tag{3.1.8}$$

$$H_1 : a_1 < 0 \tag{3.1.9}$$

(ii) Determine the test statistic using

$$F_{\hat{\epsilon}_t} = \frac{\hat{a}_1}{SE(\hat{a}_1)} \tag{3.1.10}$$

where the value of $SE(\hat{a}_1)$ is the standard error of $\hat{a}_1$, the estimate of a_1.

(iii) Compare the calculated test statistic in 3.1.10 with the critical value from the Dickey-Fuller table to reject or not to reject the null hypothesis.

(iv) If $F_{\hat{\epsilon}_t}$ is greater than the critical value, we do not reject the null hypothesis, H_0.

The rejection of H_0 implies that residuals are stationary. This further implies that the variables under study are cointegrated.

The next step is to estimate the error correction model (ECM) which will be done in the next section.

3.1.2 Estimate the error correction model

An error correction model is defined as a dynamic model in which the movement of a variable in any period is related to the previous period's gap from the long-run equilibrium. Although it may be possible to estimate the long-run or cointegrating relationship, $y_t = \beta x_t + \epsilon_t$ economic systems are rarely in equilibrium, as they are affected by institutional and/or structural changes that might be temporary or permanent. For example, extra income in the form of a birthday bonus may raise someone's expenditure pattern in one or two months and then his/her expenditure gradually goes back to equilibrium. Since equilibrium is rarely observed, the short-run evolution of variables (short-run dynamic adjustment) is important. A simple dynamic model of a short-run adjustment model is given by

$$y_t = \alpha_0 + \gamma_0 x_t + \gamma_1 x_{t-1} + \alpha_1 y_{t-1} + \varepsilon_t \tag{3.1.11}$$

where y_t is the dependent variable, x_t is the independent variable, y_{t-1} and x_{t-1} are lagged values of y_t and x_t respectively, α_0, α_1, γ_0, γ_1 are parameters, and ε_t is the error term assumed to be $\varepsilon_t \sim iN(0, \delta^2)$.

The problems associated with the use of the short-run model are the following:

- Multicollinearity: This is a situation in which two or more independent variables in a multiple regression model are highly correlated.

- Spurious correlation: This is a situation in which two variables have no causal connection, yet it may be inferred that they do as a result of a certain third unseen factor.

These problems are solved by estimating the first differences of equation 3.1.11 to obtain

$$\Delta y_t = \alpha_0 + \gamma_0 \Delta x_{t-1} + \gamma_1 \Delta x_{t-1} + \alpha_1 \Delta y_{t-1} + \varepsilon_t. \tag{3.1.12}$$

This, however, introduces problems of

- loss of information about the long-run equilibrium

- the economic theory is differenced away.

The possible solution is to adopt the error-correction mechanism (ECM) formulation of the dynamic structure. We set up the ECM as follows:

$$y_t = \alpha_0 + \gamma_0 x_t + \gamma_1 x_{t-1} + \alpha_1 y_{t-1} + \varepsilon_t. \tag{3.1.13}$$

Subtracting the term y_t on both sides leads to

$$\Delta y_t = \alpha_0 + \gamma_0 x_t + \gamma_1 x_{t-1} - (1 - \alpha_1) y_{t-1} + \varepsilon_t. \tag{3.1.14}$$

Subtracting the term $\gamma_0 x_{t-1}$ on both sides, equation 3.1.14 becomes

$$\Delta y_t - \gamma_0 x_{t-1} = \alpha_0 + \gamma_0 x_t - \gamma_0 x_{t-1} + \gamma_1 x_{t-1} - (1 - \alpha_1)y_{t-1} + \varepsilon_t, \qquad (3.1.15)$$

$$\Delta y_t = \alpha_0 + \gamma_0 \Delta x_t + (\gamma_0 + \gamma_1)x_{t-1} - (1 - \alpha_1)y_{t-1} + \varepsilon_t. \qquad (3.1.16)$$

Reparameterisation reduces equation 3.1.16 to

$$\Delta y_t = \gamma_0 \Delta x_t - (1 - \alpha_1)\left[y_{t-1} - \frac{\alpha_0}{(1 - \alpha_1)} - \frac{(\gamma_0 + \gamma_1)}{(1 - \alpha_1)}x_{t-1} \right] + \varepsilon_t. \qquad (3.1.17)$$

Taking $\beta_0 = \frac{\alpha_0}{(1-\alpha_1)}$ and $\beta_1 = \frac{(\gamma_0+\gamma_1)}{(1-\alpha_1)}$ the equation 3.1.17 becomes

$$\Delta y_t = \gamma_0 \Delta x_t - (1 - \alpha_1)\left[y_{t-1} - \beta_0 - \beta_1 x_{t-1} \right] + \varepsilon_t \qquad (3.1.18)$$

which is the ECM with $-(1 - \alpha_1)$ as the speed of adjustment, and $\varepsilon_{t-1} = y_{t-1} - \beta_0 - \beta_1 x_{t-1}$ as the error-correction mechanism which measures the distance of the system away from equilibrium. The coefficient of ε_{t-1} should be negative in sign in order for the system to converge to equilibrium. The size of the coefficient $-(1-\alpha_1)$ is an indication of the speed of adjustment towards equilibrium in that

- small values of $-(1 - \alpha_1)$, tending to -1, indicate that economic agents remove a large percentage of disequilibrium in each period

- larger values, tending to 0, indicate that adjustment is slow

- extremely small values, less than -2, indicate an overshooting of economic equilibrium

- positive values would imply that the system diverges from the long-run equilibrium path.

The ECM satisfies the assumptions of classical normal linear regression model (CNLRM). These assumptions include:

- A linear regression model.

- Residuals are normally distributed.

- There is no serial correlation among residuals.

- The number of observations must not exceed the number of parameters to be estimated.

- There is no perfect multicollinearity.

This means that diagnostic tests have to be conducted on the error-correction mechanism in order to determine whether any of these assumptions have not been violated. These tests include

- a normality test

- a heteroscedasticity test

- a serial correlation test

- a misspecification test.

3.1.2.1 Normality test

We use the Jacque-Bera test to determine whether the ECM is normally distributed. This test measures the difference in kurtosis and skewness of a variable compared to those of the normal distribution (Jarque and Bera, 1980).

In the Jacque-Bera test, we set the null and alternative hypothesis as follows:

H_0: The variable is normally distributed.

H_1: The variable is not normally distributed.

The test statistic is

$$JB = \frac{N-k}{6}\left[S^2 + \frac{(K-3)^2}{4}\right] \tag{3.1.19}$$

where N is the number of observations, k is the number of estimated parameters, S is the skewness of a variable, and K is the kurtosis of a variable.

We reject the null hypothesis if the p-value $\leq$ level of significance, or if the $JB > \chi^2(2)$.

3.1.2.2 Heteroscedasticity test

Heteroscedasticity results from a sequence of random variables having different variances. It implies that during regression analysis there is non-consistent variance. Heteroscedasticity is tested using the Langrange multiplier, also known as Engle's Arch LM test (Engle, 1982). The test procedure is as follows:

H_0: There is no heteroscedasticity.

H_1: There is heteroscedasticity.

The test statistic is

$$LM_E = nR^2 \tag{3.1.20}$$

where n is the number of observations, and R^2 is the coefficient of determination of the augmented residual regression.

We reject the null hypothesis if the p-value $\leq$ level of significance and conclude that there is heteroscedasticity.

3.1.2.3 Serial correlation test

Serial correlation is cross-correlation of a signal (white noise) with itself. It may be caused by

- nonstationarity of dependent and explanatory variable

- data manipulation (averaging, interpolation and extrapolation)

- incorrect functional form.

Ljung and Box (1978) suggested the use of Ljung-Box test to test the assumption that the residuals contain no autocorrelation up to any order k. The test procedure is as follows:

H_0: There is no autocorrelation up to order k.

H_1: Autocorrelation exists up to order k.

The test statistic is

$$Q_{LB} = T(T+2) \sum_{j=1}^{k} \frac{r_j^2}{T-j} \qquad (3.1.21)$$

where T is the number of observations, k is the highest order of autocorrelation for which to test, and r_j^2 is the jth autocorrelation.

We reject the null hypothesis if the p-value $\leq$ level of significance and conclude that autocorrelation exists up to order k.

The major drawback of this test is deciding which lag order (k) to use. Ljung and Box (1978) suggested the maximum number of lags to use should be $T^{\frac{1}{3}}$ where T is the number of observations.

3.1.2.4 Misspecification test

Misspecification is as the result of

- incorrect functional form

- inclusion of irrelevant variables

- exclusion of relevant variables.

The consequences of a misspecified regression are

- the residuals are not normally distributed

- there is serial correlation

- regression is inconsistent with actual working of the economy

- the parameter estimates are not robust to samples used.

To test for misspecification, Ramsey (1969) suggested the use of Ramsey's reset test. This is a likelihood test that compares the likelihood function of original regression to an augmented regression.

3.1.3 Limitations of the Engle-Granger method

Although the Engle and Granger procedure is easily implemented, it has several defects:

- The estimation of the long-run equilibrium regression requires that the researcher place one variable on the right-hand side as the dependent variable and use the other variable on the left-hand side as the independent variable. For example, in the case of two variables, it is possible to run the Engle-Granger method for cointegration by using the residuals from either of the following two "equilibrium" regression equation.

$$y_t = \beta_{10} + \beta_{11}x_t + \epsilon_{1t} \tag{3.1.22}$$

or

$$x_t = \beta_{20} + \beta_{21}y_t + \epsilon_{2t} \tag{3.1.23}$$

As the sample size grows infinitely large, the theory indicates that the test for a unit root in the ϵ_{1t} sequence becomes equivalent to the test for a unit root in the ϵ_{2t} sequence. Unfortunately, the properties of large samples on which this result is derived may not be applicable to the sample sizes usually available to economists. In many cases available sample sizes are smaller than the required sample size on which the theory is based.

- The two-step estimation procedure is based on the principle that, irrespective of which variable is chosen for normalisation, the same results will be attained if variables are interchanged. In practice, it is possible to find that one regression indicates that the variables are cointegrated, whereas reversing the order indicates no cointegration. For example, in investigating the relationship between income and expenditure, if income is placed on the left-hand side as the dependent variable, it is possible to conclude that income and expenditure cointegrate, but the reverse is not necessarily true. This is a very undesirable feature of the procedure, because the test for cointegration should be invariant to the choice of variable selected for normalisation.

- Engle-Granger's two-step estimation procedure relies on *a two-step estimator*. Recall that the first step in the two-step estimation method of pre-setting each variable to determine

the order of integration generates residual series $\hat{\epsilon}_t$ used in the second step. This is used to estimate the regression equation of the form $\Delta\hat{\epsilon}_t = a_1\hat{\epsilon}_{t-1} + \varepsilon_t$ as in equation 3.1.7. Thus, the coefficient a_1 is obtained by estimating a regression equation using residuals from another regression. This implies that any error introduced in the first step is carried over into the second step, making the results unreliable.

Fortunately, several methods have been developed that could avoid these defects. These include the Phillip-Ouliaris methods, Johansen's procedure, and the Stock and Watson maximum likelihood estimators. These tests allow the researcher to test for restricted versions of cointegrating vector(s) and the speed of adjustment parameters. They rely heavily on the relationship between the rank of a matrix and its characteristic roots.

3.2   Phillips-Ouliaris methods

In section 3.1, it has been shown that the second step of the Engle-Granger's method is an ADF test applied to the residuals of the long-run equation. Phillips and Ouliaris introduced two residual-based tests, namely the variance ratio test and the multivariate trace statistic (Phillips-Ouliaris, 1988). The latter of these tests has the advantage that it is invariant to normalisation, that is, for whichever variable is taken to be the dependent variable, the test will yield the same results. These tests are used in the same way as unit root tests but the data are residuals from cointegrating regressions. They are implemented on matrices or multivariate series and are both based on residuals of the first order vector autoregression equation

$$\mathbf{z}_t = \hat{\mathbf{\Pi}}\mathbf{z}_{t-1} + \hat{\xi}_t \tag{3.2.1}$$

where $\mathbf{z_t}$ is partitioned as $\mathbf{z_t} = (\mathbf{y_t}, \mathbf{x_t'})$ with a dimension of $\mathbf{x_t}$ equal to $(m = n + 1)$, $\hat{\mathbf{\Pi}}$ is a regression coefficient, and n is equal to the number of variables under study.

3.2.1   The variance ratio test

The variance ratio statistic $\hat{P}_u$ is defined as

$$\hat{P}_u = \frac{T\hat{\omega}_{11.2}}{T^{-1}\sum_{t=1}^{T}\hat{u}_t^2} \tag{3.2.2}$$

where $\hat{u}_t$ are the residuals of the long-run regression equation

$$\mathbf{y_t} = \hat{\beta}\mathbf{x_t} + \hat{\mathbf{u}_t}. \tag{3.2.3}$$

The conditional variance $\hat{\omega}_{11.2}$ is derived from the conditional matrix $\hat{\Omega}$ of $\hat{\xi}$, that is the residuals of equation 3.2.1, and is defined as

$$\hat{\omega}_{11.2} = \hat{\omega}_{11} - \hat{\omega}_{21}' \hat{\Omega}_{22}^{-1} \hat{\omega}_{21}. \tag{3.2.4}$$

The covariance matrix $\hat{\Omega}$ has been partitioned as

$$\hat{\Omega} = \begin{bmatrix} \hat{\omega}_{11} & \hat{\omega}_{21} \\ \hat{\omega}_{21} & \hat{\Omega}_{22} \end{bmatrix} \tag{3.2.5}$$

and is estimated as

$$\hat{\Omega} = T^{-1} \sum_{t=1}^{T} \hat{\xi}_t' \hat{\xi}_t + T^{-1} \sum_{s=1}^{L} \omega_{sl} \sum_{t=1}^{T} (\hat{\xi}_t \hat{\xi}_{t-s} + \hat{\xi}_{t-s} \hat{\xi}_t') \tag{3.2.6}$$

with weighting function $\omega_{sl} = (1 - s)/(l + 1)$.

A variance ratio test is a residual-based test that seeks to test a null hypothesis of **no** cointegration using scalar unit roots applied to equation 3.2.3. The null hypothesis is formulated in terms of the conditional variance parameter $\omega_{11.2}$ as follows:

$$H_0 : \omega_{11.2} \neq 0 \tag{3.2.7}$$

against

$$H_1 : \omega_{11.2} = 0. \tag{3.2.8}$$

Therefore, the variance ratio test measures the size of the residual variance from the cointegrating regression of $\mathbf{y_t}$ on $\mathbf{x_t}$ versus $T^{-1} \sum_{i=1}^{T} \hat{u}_t^2$ against that of a direct estimate of the population conditional variance of $\mathbf{y_t}$, given $\mathbf{x_t}$ versus $\hat{\omega}_{11.2}$. If cointegration exists between variables, the variance ratio should stabilise asymptotically, whereas if a spurious (nonsense) relationship is present, it would be reflected in the cointegrating regression and the variance ratio should diverge.

3.2.2   The multivariate trace statistic

The multivariate trace statistic, denoted as $\hat{P}_z$ is defined as

$$\hat{P}_z = T tr(\hat{\Omega} M_{zz}^{-1}) \tag{3.2.9}$$

with T is the number of observations, $M_{zz}^{-1} = t^{-1} \sum_{t=1}^{T} z_t z_t'$, and $\hat{\Omega}$ is estimated in equation 3.2.6.

The advantage of the Engle-Granger two-step procedure is its ease of implementation. However, its results are dependent on how the long-run equation is specified. In some cases it might be easier

to identify which variable enters on the left as the dependent variable. For example, in accessing the relationship between income and expenditure, it is easy to say that expenditure depends on income. Unfortunately, this is only true in some cases. It is therefore advisable to employ the cointegrating test of Phillips-Ouliaris, which yields the same results irrespective of the variable which enters as the dependent variable, that is, they are invariant to normalisation.

One deficiency of the two methods (two-step procedure and Phillips-Ouliaris) is that one can only estimate a single cointegrating relationship. However, if one deals with more than two time series, it is possible that more than one cointegrating relationship will exist, which calls for the use of vector cointegration techniques like Johansen's procedure.

3.2.3 Limitation of Phillips-Ouliaris methods

This method can only estimate single cointegrating relationships. If one deals with more than two time series, it is possible that more than one cointegrating relationship exists. The remedy to this limitation is to use Johansen's procedures.

3.3 Johansen's procedure

Johansen's method takes as a starting point the vector autoregression (VAR) of order p given by

$$\mathbf{X_t} = \mathbf{\Pi_1 X_{t-1}} + \mathbf{\Pi_2 X_{t-2}} + ... + \mathbf{\Pi_p X_{t-p}} + \mathbf{\underline{u}_t} \tag{3.3.1}$$

where $\mathbf{X_t}$ is an $n \times 1$ vector of variables that are integrated of order one, that is, $I(1)$, $\mathbf{\underline{u}_t}$ is an $n \times 1$ vector of innovations while $\mathbf{\Pi_1}$ through $\mathbf{\Pi_p}$ are $m \times m$ coefficient matrices. Reparameterising the equation 3.3.1, that is, subtracting $\mathbf{X_{t-1}}$ on both sides, leads to

$$\mathbf{\Delta X_t} = \mathbf{\Gamma_1 \Delta X_{t-1}} + \mathbf{\Gamma_2 \Delta X_{t-2}} + ... + \mathbf{\Gamma_{p-1} \Delta X_{t-p+1}} - \mathbf{\Pi X_{t-p}} + \mathbf{\underline{u}_t} \tag{3.3.2}$$

where $\mathbf{\Gamma_1} = \mathbf{\Pi_1} - \mathbf{I}$, $\mathbf{\Gamma_2} = \mathbf{\Pi_2} - \mathbf{\Gamma_1}$, $\mathbf{\Gamma_3} = \mathbf{\Pi_3} - \mathbf{\Gamma_2}$ and $\mathbf{\Pi} = \mathbf{I} - \mathbf{\Pi_1} - \mathbf{\Pi_2} - ... - \mathbf{\Pi_p}$. The matrix $\mathbf{\Pi}$ determines the extent to which the system is cointegrated and is called the *impact matrix*.

Returning to the general reparameterised equation 3.3.2, if we consider the first equation of the system as:

$$\Delta X_{1t} = \gamma'_{11}\Delta X_{t-1} + \gamma'_{12}\Delta X_{t-2} + ... + \gamma'_{1p-1}\Delta X_{t-p+1} - \Pi'_1 X_{t-p} + u_{1t} \tag{3.3.3}$$

where γ'_{ij} is the first row of Γ_j, $j = 1, 2, ...p - 1$, and Π'_1 is the first row of $\mathbf{\Pi}$.

Here ΔX_{1t} is stationary, that is, $I(0)$, $j = 1, 2, ..p - 1$ are all $I(0)$, u_t is assumed to be $I(0)$ and so

for a meaningful equation, $\Pi_1' X_{t-p}$ must be stationary, $I(0)$.

If none of the components of $\mathbf{X_t}$ are cointegrated, they must be *zero*. On the other hand, if they are cointegrated, all the rows of $\mathbf{\Pi}$ must be cointegrated but not necessarily distinct. This is because the number of distinct cointegrating vectors depends on the row rank of $\mathbf{\Pi}$ (Harris, 1995).

The matrix $\mathbf{\Pi}$ is of order $m \times m$. If it has rank m, that is, m number of linearly independent rows or columns, then it forms a basis for m-dimensional vector space. This implies that all $m \times 1$ vectors can be generated as linear combinations of its row. Any of these linear combination of the rows would lead to stationarity, meaning that $\mathbf{X_{t-p}}$ has stationary components if the rank of $\mathbf{\Pi}$ is $r < m$.

We may write $\mathbf{\Pi} = \underline{\beta}\underline{\alpha}'$ for suitable $m \times r$ matrices, $\underline{\beta}$ and $\underline{\alpha}$. Here

$$\underline{\alpha}' = \begin{bmatrix} \alpha_1' \\ \alpha_2' \\ . \\ . \\ . \\ \alpha_r' \end{bmatrix} \tag{3.3.4}$$

$$\underline{\beta} = \begin{bmatrix} \beta_1, \beta_2. & . & . & . & \beta_r \end{bmatrix} \tag{3.3.5}$$

Then $\mathbf{\Pi X_{t-p}} = \underline{\beta}\underline{\alpha}'\mathbf{X_{t-p}}$ and all linear combinations of $\underline{\alpha}'\mathbf{X_{t-p}}$ are stationary. It should be noted that we have to perform the ADF test to access the order of integration of each variable before applying Johansen's procedure. Johansen's procedure estimates the VAR subject to $\mathbf{\Pi} = \underline{\beta}\underline{\alpha}'$ for various values of r number of cointgrating vectors, using the maximum likelihood estimator assuming $u_t \sim iidN(0, \Sigma)$. His estimate can thus be rewritten as

$$\Delta\mathbf{X_t} = \mathbf{\Gamma_1}\Delta\mathbf{X_{t-1}} + \mathbf{\Gamma_2}\Delta\mathbf{X_{t-2}} + ... + \mathbf{\Gamma_{p-1}}\Delta\mathbf{X_{t-p+1}} - \underline{\beta}\underline{\alpha}'\mathbf{X_{t-p}} + \mathbf{u_t}. \tag{3.3.6}$$

The question is, how do we detect the number of cointegrating vectors?

Johansen proposed two likelihood ratio tests namely:

- The trace test

- The maximum eigenvalue.

3.3.1 The trace test

The trace test tests the null hypothesis of r cointegrating vectors against the alternative hypothesis of n cointegrating vectors. The test statistic is given by

$$\jmath_{trace} = -T \sum_{i=r+1}^{n} ln(1 - \hat{\lambda}_i). \tag{3.3.7}$$

3.3.2 The maximum eigenvalue

The maximum eigenvalue test, on the other hand, tests the null hypothesis of r cointegrating vectors against the alternative hypothesis of $(r+1)$ cointegrating vectors. Its test statistic is given by

$$\jmath_{max} = -T(1 - \hat{\lambda}_{r+1}) \tag{3.3.8}$$

where T is the sample size, and $\hat{\lambda}_i$ is the ith largest canonical correlation.

3.3.3 Limitation of Johansen's procedure

This method assumes that the cointegrating vector remains constant during the period of study. In reality, it is possible that the long-run relationships between the underlying variables change. The reason for this might be technological progress, economic crisis, changes in people's preferences and behaviour accordingly, policy or regime alteration and institutional development. This is especially the case if the sample period is long.

To take this into account, Gregory and Hasen (1996) have introduced tests for cointegration with one and two unknown structural break(s). However, such tests do not form part of this study.

Chapter 4

Testing Cointegration Results for the *Raotbl*3 data set

4.1 Introduction

In this chapter we applied the Engle-Granger method, the Phillip-Ouliaris methods and Johansen's procedure to test for the existence of cointegrating relationships in the *Raotbl*3 data set using statistical software R. This data set contains the time series used by Holden and Perman (1994). In this study the *Raotbl*3 data set is included as Appendix A. More details about the data are provided in the data appendix of a book by Bhaskara (1994).

Three time series and three dummy variables are given in this data set. The three dummies are as a result of structural changes such as strikes, wars and natural disasters that did not make it possible to capture data in specific time periods. Data frames are quarterly data (times series objects) from the United Kingdom starting in 1966:4 (1966, April) until 1991:2 (1991, February) for the following six variables all transformed to natural logarithms to reduce variances:

   i. (lc) Real consumption expenditure.

   ii. (li) Real income.

  iii. (lw) Real wealth.

  iv. dd682 Dummy variable for 68:2.

   v. dd792 Dummy variable for 79:2.

vi. dd883 Dummy variable for 88:3.

The three time series (lc, li and lw) above are defined from

(i) CAAB-seasonally adjusted real consumers' expenditure in millions of pounds, 1985 prices

(ii) AIIX-seasonally adjusted nominal consumers' expenditure in millions of pound, current prices

(iii) AIIW-seasonally adjusted nominal personal disposable income in millions of pounds, current prices

The three time series can be found in Table 5 of *Economic Trends Annual Supplement* (1992). The wealth series is defined as seasonally adjusted gross personal financial wealth in millions of pounds, current prices, and is compiled from several editions of *Financial Statistics* in order to obtain a consistent series.

Constant price series for income and wealth for the year 1985 were obtained from the current price series using the implicit consumers' expenditure deflator (AIIX/CAAB). The variables *lc*, *li* and *lw* are the natural logarithm of the constant price consumption, income and wealth series. The dummy variables are defined as follow:

(i) DD682 = 1 in 1968:2, -1 in 1968:3 and 0 otherwise,

(ii) DD792 = 1 in 1979:2, -1 in 1979:3 and 0 otherwise,

(iii) DD883 = 0 up to and including 1988:2 and 1 thereafter.

4.2   Descriptive summary

The following commands import and display the data set described above in a statistical package R. The data set is stored under the file name "*Raotbl3*".

```
> library(urca)
> data(Raotbl3)
> attach(Raotbl3)
>Raotbl3
```

A descriptive summary of the Raotlb3 data set in natural logarithm is provided.

```
> summary(Raotbl3)

       lc               li               lw
 Min.   :10.48    Min.   :10.58    Min.   :12.72
 1st Qu.:10.69    1st Qu.:10.78    1st Qu.:12.92
 Median :10.77    Median :10.90    Median :13.06
 Mean   :10.79    Mean   :10.89    Mean   :13.14
 3rd Qu.:10.89    3rd Qu.:11.00    3rd Qu.:13.30
 Max.   :11.14    Max.   :11.24    Max.   :13.77
```

Figure 4.1: The consumption, income and wealth time series plot.
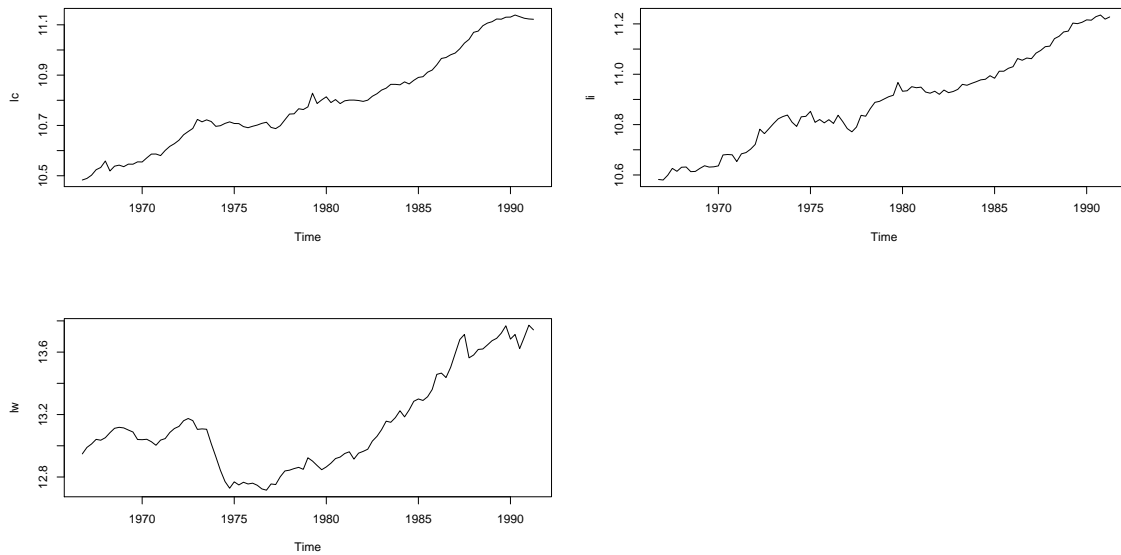
From figure 4.1, the plots of the consumption and income series do not vary about a fixed level, indicating nonstationarity in the mean but not in the variance, while the plot of the wealth series varies about a fixed level with constant variance, indicating nonstationarity in both the mean and the variance. This can be confirmed after testing for unit roots.

Table 4.1 below provides a descriptive summary of the data set.

Table 4.1: Descriptive summary

Variable	Minimum	Median	Mean	Maximum
lc	10.48	10.77	10.79	11.14
li	10.58	10.90	10.89	11.24
lw	12.72	13.06	13.14	13.77

In all three variables, the means and the medians are not far away from each other. This probably indicates that the series in the data set are slightly symmetric. However, this does not indicate normality. A test for normality is performed.

4.3 Test for normality

The null and the alternative hypothesis of this set are the following:

H_0 : The series in the data set are normally distributed.

H_1 : The series in the data set are not normally distributed.

Using the Shapiro test, we test for normality of the variables to see if the series in the data set are well modelled by a normal distribution using the following R command:

```
> shapiro.test(lc)

        Shapiro-Wilk normality test

data:  lc
W = 0.9458, p-value = 0.0004748
> shapiro.test(li)

        Shapiro-Wilk normality test

data:  li
W = 0.9618, p-value = 0.005767
> shapiro.test(lw)

        Shapiro-Wilk normality test

data:  lw
W = 0.9043, p-value = 2.517e-06
```

Table 4.2 provides the summary results of the normality test using Shapiro-Wilk test.

From the above R command, the p-values are very small which means the null hypothesis of normality can be rejected at the 1% level of significance for all series. This implies that the time series

Table 4.2: Normality test on the *Raotbl*3 data set

Variable	Shapiro-Wilk	p-value	Decision
lc	0.9458	4.475×10^{-4}	Reject the null hypothesis
li	0.9618	5.767×10^{-3}	Reject the null hypothesis
lw	0.9043	2.517×10^{-6}	Reject the null hypothesis

of all the variables are not normally distributed.

In the next command, we test for the presence of unit roots in each of the series (consumption, income and wealth) using the ADF unit root test. Testing for unit root implies testing for nonstationarity in the series.

4.4 Augmented Dickey-Fuller (ADF) test for unit roots

We pre-test each variable to determine the order of integration using the ADF unit root test. This is because cointegration necessitates that variables be integrated of the same order. By determining the order of integration, we use the function *ur.df* in R on each variable. For each variable tested for unit root, we set both the null and alternative hypotheses as

$$H_0 : \delta = 0 \tag{4.4.1}$$

$$H_1 : \delta < 0 \tag{4.4.2}$$

After inspection of the behaviour of the ACF plot for the *lc* series, it was found that its residuals become white noise after lag five. This means that the ACF plot for the *lc* series cuts off at lag five. Since this series shows an increasing trend in figure 4.1, we test the ADF test on the trend model and lag five as follows:

```
> lc.ct=ur.df(lc,lags=5,type='trend')
> summary(lc.ct)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.8782632  0.3723647   2.359   0.0206 *
z.lag.1     -0.0837027  0.0355904  -2.352   0.0210 *
tt           0.0005533  0.0002246   2.463   0.0158 *
z.diff.lag1 -0.0505981  0.1040693  -0.486   0.6281
```

```
z.diff.lag2  0.1501023  0.1027518   1.461   0.1478

z.diff.lag3  0.2411832  0.1036023   2.328   0.0223 *

z.diff.lag4 -0.0810911  0.1078641  -0.752   0.4543

z.diff.lag5  0.2129158  0.1067675   1.994   0.0493 *

---

Multiple R-squared: 0.1993,     Adjusted R-squared: 0.1333

Value of test-statistic is: -2.3518 3.6798 3.057
```

We are fitting the ADF model

$$Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + \alpha_i \sum_{i=1}^{m} \Delta Y_{t-i} + \epsilon_t. \tag{4.4.3}$$

In terms of the coefficient estimate, the above model becomes

$$lc_t = 0.8782632 + 0.0005533tt - 0.0837027 lc_{t-1} - 0.0505981 \Delta lc_{t-1} + \ldots + 0.2129158 \Delta lc_{t-5} \tag{4.4.4}$$

where lc_t is the consumption function series, tt is the trend, lc_{t-1} are the lagged values of lc and Δlc_{t-1} are the first difference lagged values of the consumption function. The parameter of interest in equation 4.4.4 is the estimated coefficient of lc_{t-1} which is -0.0837027. The calculated test statistic -2.3518 is derived from

$$F_\tau = \frac{\hat{\delta}}{SE(\hat{\delta})} = \frac{-0.0837027}{0.0355904} = -2.3518. \tag{4.4.5}$$

As extracted from R, the critical values for the unit root test are given in table 4.3.

Table 4.3: Critical values for the unit root test

	1%	5%	10%
τ	-4.04	-3.45	-3.15
α_1	6.50	4.88	4.16
α_2	8.73	6.49	5.47

Since the calculated test statistic -2.3518, falls in the non-rejection region, that is, to the right of the τ(tau) critical values, we cannot reject the null hypothesis for presence of unit roots at the 10% level of significance. This means that the consumption series contains a unit root.

Figure 4.2: A plot of ADF unit root test for consumption series
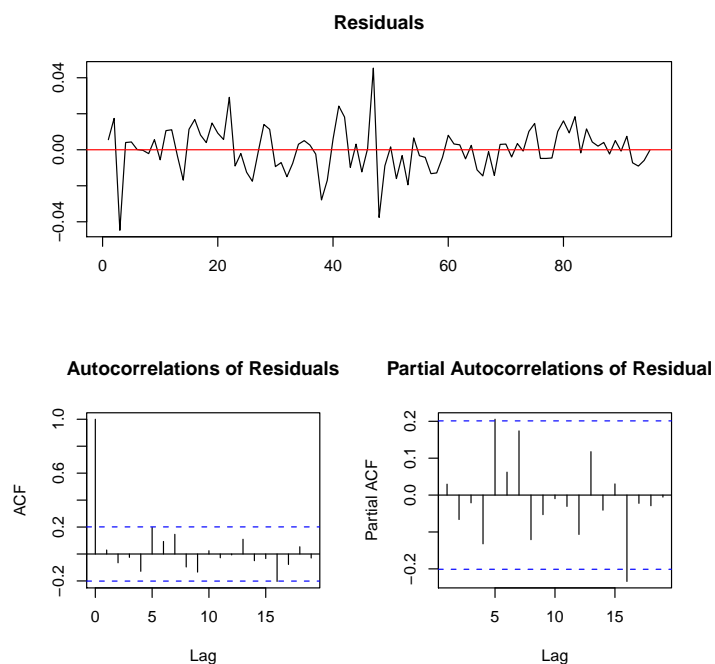
Figure 4.2 shows a plot of the ADF unit root test for the consumption series. From this plot the residuals appear to vary about a fixed level. Time series that exhibit this phenomenon are said to be nonstationary in the mean but not in the variance. This confirms that the consumption series contains a unit root.

The ADF test for unit roots in real income (li) and real wealth (lw) respectively can be obtained from the following commands:

```
> lic=ts(li)
> lcc=ur.df(lic, lags=5, type='trend')
> summary(lcc)
Coefficients:
```

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	1.5208232	0.6263415	2.428	0.0173	*
z.lag.1	-0.1434285	0.0592676	-2.420	0.0177	*
tt	0.0009068	0.0003671	2.470	0.0155	*
z.diff.lag1	-0.0697724	0.1097974	-0.635	0.5268	
z.diff.lag2	0.1145049	0.1118766	1.023	0.3090	
z.diff.lag3	-0.0248040	0.1098202	-0.226	0.8219	

```
z.diff.lag4  0.2221217  0.1090673   2.037   0.0448 *

z.diff.lag5  0.0864713  0.1089450   0.794   0.4296

-----

Multiple R-squared: 0.1359,     Adjusted R-squared: 0.06477

Value of test-statistic is: -2.42 4.4181 3.0514


> kl=ts(lw)

> kl2=ur.df(kl, lags=5, type='trend')

> summary(kl2)

Coefficients:

              Estimate Std. Error t value Pr(>|t|)

(Intercept)  0.3774736  0.2591340   1.457   0.1489

z.lag.1     -0.0302509  0.0203009  -1.490   0.1399

tt           0.0004702  0.0002252   2.088   0.0398 *

z.diff.lag1  0.1644410  0.1067375   1.541   0.1271

z.diff.lag2 -0.0001526  0.1089781  -0.001   0.9989

z.diff.lag3  0.0465937  0.1099422   0.424   0.6728

z.diff.lag4 -0.0509364  0.1163995  -0.438   0.6628

z.diff.lag5  0.2179245  0.1138548   1.914   0.0590

Multiple R-squared: 0.1247,     Adjusted R-squared: 0.0526

Value of test-statistic is: -1.4901 1.7591 2.207
```

Table 4.4: Summary of the ADF test for unit root in the variables (with a trend and intercept)

Variable	ADF test statistic	Decision
Consumption series (lc)	-2.3518	Fail to reject the null hypothesis
Real income (li)	-2.4200	Fail to reject the null hypothesis
Real wealth (lw)	-1.4901	Fail to reject the null hypothesis

The ADF critical value for unit root at the 10% level of significance is -3.15. From table 4.4, it can be deduced that the three variables are nonstationary, each contain at least one unit root, that is, they are integrated of order one, $I(1)$. We now test whether the integration is of the same order.

4.5 Augmented Dickey-Fuller(ADF) test for the order of integration

Here we test whether the series is possible $I(2)$, that is, whether they contain a second order of integration. This test is achieved by supplying the differenced series in the *ur.df* *R* function to test for presence of unit roots.

Figure 4.3: A plot of 1st differences of consumption, income and wealth series
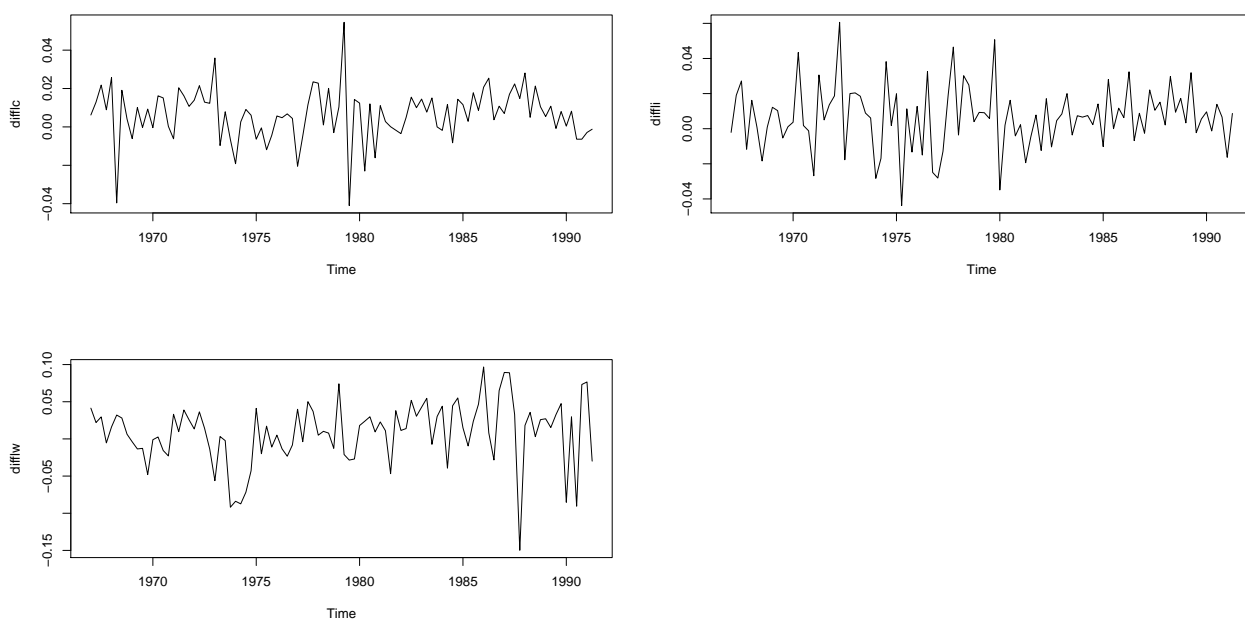
From figure 4.3 it can be observed that the first difference series for the *Raotbl*3 data set are stationary both in the mean and the variance. This is because they fluctuate about the zero mean and exhibit constant variance. Since the ACF and the $PACF$ plots cut off at lag five, we test the ADF test for the possible second order of integration using the no trend model and five lags as follows:

```
> lc2.ct=ur.df(lc2,type="none", lags=5)
> summary(lc2.ct)
Coefficients:
          Estimate Std. Error t value Pr(>|t|)
z.lag.1    -0.28585    0.14698  -1.945  0.05507 .
z.diff.lag1 -0.69758    0.15368  -4.539 1.82e-05 ***
```

```
z.diff.lag2 -0.48591    0.15577  -3.119  0.00247 **

z.diff.lag3 -0.19769    0.15427  -1.281  0.20348

z.diff.lag4 -0.28052    0.13604  -2.062  0.04222 *

z.diff.lag5 -0.05004    0.09784  -0.511  0.61032

Residual standard error: 0.01374 on 90 degrees of freedom

Multiple R-squared: 0.579,      Adjusted R-squared: 0.5602

F-statistic: 30.94 on 4 and 90 DF,   p-value: 3.366e-16

Value of test-statistic is: -1.945
```

Figure 4.4: A plot of 1st difference unit root test for the consumption series
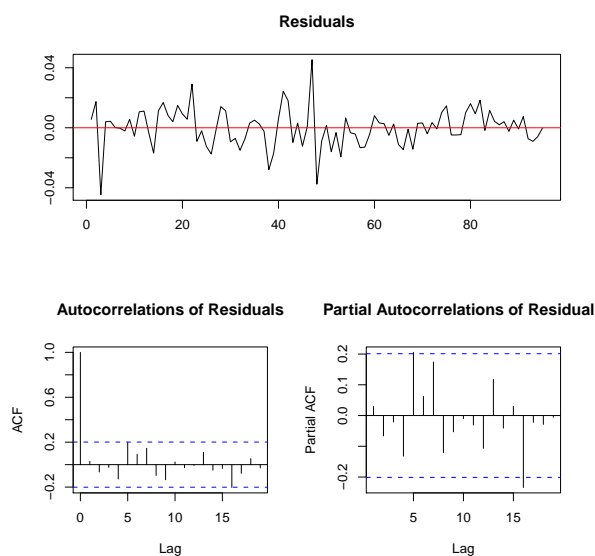
Figure 4.4 exhibits a pattern of series stationary in the mean and in the variance. To verify this assertion we conduct a formal stationarity test (ADF test) on the residuals of the consumption series.

For the differenced real income and differenced real wealth respectively, the following commands are used:

```
> li2=diff(li)

> li2.ct=ur.df(li2,type="none", lags=5)

> summary(li2.ct)

Coefficients:

          Estimate Std. Error t value Pr(>|t|)

z.lag.1      -0.4853     0.2083  -2.329  0.02219 *
```

```
z.diff.lag1  -0.5978      0.2040  -2.931  0.00433 **

z.diff.lag2  -0.5024      0.1993  -2.521  0.01355 *

z.diff.lag3  -0.5068      0.1773  -2.858  0.00535 **

z.diff.lag4  -0.2716      0.1532  -1.774  0.07967 .

z.diff.lag5  -0.1631      0.1052  -1.551  0.12467

Residual standard error: 0.01859 on 90 degrees of freedom

Multiple R-squared: 0.577,      Adjusted R-squared: 0.5582

Value of test-statistic is: -2.3292


> lw2=diff(lw)

> lw2.ct=ur.df(lw2,type="none", lags=5)

> summary(lw2.ct)

Coefficients:

           Estimate Std. Error t value Pr(>|t|)

z.lag.1     -0.51732      0.19326  -2.677   0.0089 **

z.diff.lag1 -0.30022      0.19100  -1.572   0.1197

z.diff.lag2 -0.28261      0.17803  -1.587   0.1161

z.diff.lag3 -0.23015      0.16523  -1.393   0.1672

z.diff.lag4 -0.26658      0.14166  -1.882   0.0632 .

z.diff.lag5 -0.05817      0.11911  -0.488   0.6265

Residual standard error: 0.04313 on 90 degrees of freedom

Multiple R-squared: 0.3959,      Adjusted R-squared: 0.3691

Value of test-statistic is: -2.6769
```

The critical values for the ADF test for the order of integration are shown in the table 4.5 below.

Table 4.5: Critical values for the ADF order of integration test

	1%	5%	10%
τ_1	-2.6	-1.95	-1.61

Table 4.6 summarises the results of the ADF test for the order of integration.

Table 4.6: Summary for ADF test for unit roots in variables (in 1st difference form)

Variable	test statistic	Decision
Differenced Consumption series (Δlc)	-1.9450	Reject the null hypothesis
Differenced Real income (Δli)	-2.3292	Reject the null hypothesis
Differenced Real wealth (Δlw)	-2.6769	Reject the null hypothesis

The ADF critical value at the 10% level of significance is -1.61. Based on the ADF test, the first difference variables are stationary, which implies that variables lc, li and lw are integrated of order one, $I(1)$. Now we can test for the existence of a long-run relationship between the variables, that is, cointegration.

4.6 Tests for cointegration

There are various techniques to test for cointegration. The following techniques, the Engle-Granger method, the Phillips-Ouliaris methods and Johansen's procedure, were applied to the *Raotbl*3 data set.

4.6.1 Engle-Granger method

In R, the Engle-Granger two-step procedure is implemented in R using the following commands.

```
> lc=ts(lc,start=c(1966,4), end=c(1991,2), frequency=4)
> li=ts(li,start=c(1966,4), end=c(1991,2), frequency=4)
> lw=ts(lw,start=c(1966,4), end=c(1991,2), frequency=4)
>ukcons=window(cbind(lc,li,lw),start=c(1962,2),end=c(1991,2))
> lc.eq=summary(lm(lc~li+lw, data =ukcons))
> lc.eq
> li.eq=summary(lm(li~lc+lw, data =ukcons))
> li.eq
> lw.eq=summary(lm(lw~li+lc, data =ukcons))
> lw.eq
```

To implement the Engle-Granger method on *Raotbl*3 data set, we begin by regressing the consumption, income and wealth series on each other and then assess the model fit. If variables cointegrate, the resulting OLS regression yields a "super-consistent" estimator of the cointegrating parameters.

By this we mean that there is a very strong relationship between the estimated parameters. Taking the consumption series as the dependent variable and the other (income and wealth) as the independent variables, we yield the following regression equation

$$lc = -0.178458 + 0.910971li + 0.079761lw. \tag{4.6.1}$$

```
Call:
lm(formula = lc ~ li + lw, data = ukcons)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.178458   0.103491  -1.724    0.088 .
li           0.910971   0.013457  67.697   <2e-16 ***
lw           0.079761   0.007764  10.274   <2e-16 ***
Multiple R-squared: 0.9918,     Adjusted R-squared: 0.9916
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

The p-values of the independent variables are very small (2×10^{-16}); this means that these regression coefficients are statistically significant at the 0.001 level of significance. The R-squared value is 0.9918, meaning that 99.18% of the variations in the consumption series are explained by changes in income and wealth series which reflects the true real-life situation. When income and wealth series are taken to be the dependent variables respectively, the following regression equations are obtained:

$$li = 0.302998 + 1.075667lc - 0.077579lw, \tag{4.6.2}$$

$$lw = 1.0681 - 5.4625li + 6.6315lc. \tag{4.6.3}$$

It is interesting to see that there is a negative relationship between income and wealth, meaning that having income (high or low) does not necessarily mean that you are wealthy. To determine if the variables actually cointegrate, we test whether the residuals from the regression relationship(s) are stationary. To extract and store the residuals of the regression equations, we use the following commands:

```
> error.lc=ts(resid(lc.eq),start=c(1967,2),end=c(1991,2),frequency=4)
> error.li=ts(resid(li.eq),start=c(1967,2),end=c(1991,2),frequency=4)
> error.lw=ts(resid(lw.eq),start=c(1967,2),end=c(1991,2),frequency=4)
```

The residuals of these three long-run relationships are stored as objects *error.lc*, *error.li* and *error.lw* in *R* respectively. When we plot the residuals *error.lc* in figure 4.5 we observe that they are stationary both in mean and variance.

Figure 4.5: A plot of ADF unit root test on residuals for the consumption series
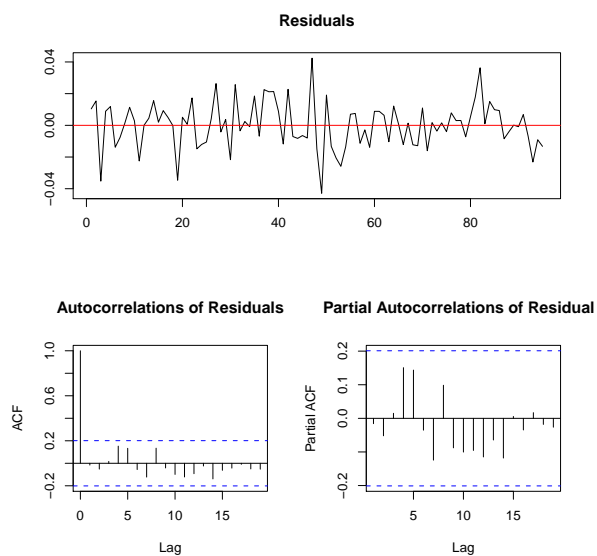
From the plot it is observed that all the lags are insignificant (within the confidence bound), we therefore test the none trend with zero lags ADF test on residuals as follows:

```
> ci.lc=ur.df(error.lc,lags=0,type='none')
Coefficients:
          Estimate Std. Error t value Pr(>|t|)
z.lag.1 -0.62472    0.09605   -6.504 3.63e-09
---
Multiple R-squared: 0.3427,    Adjusted R-squared: 0.3286
Value of test-statistic is: -6.504
> ci.li=ur.df(error.li,lags=0,type='none')
> summary(ci.li)
Coefficients:
          Estimate Std. Error t value Pr(>|t|)
z.lag.1 -0.62593    0.09644    -6.49 3.87e-09 ****
---
Multiple R-squared: 0.3395,    Adjusted R-squared: 0.3253
Value of test-statistic is: -6.49
> ci.lw=ur.df(error.lw,lags=0,type='none')
> summary(ci.lw)
Coefficients:
```

```
        Estimate Std. Error t value Pr(>|t|)

z.lag.1    -0.21469    0.07911  -2.714  0.00793 **

z.diff.lag -0.32469    0.09841  -3.299  0.00137 **

Multiple R-squared: 0.2472,      Adjusted R-squared: 0.231

Value of test-statistic is: -2.7137
```

In performing the ADF test on residuals we are actually fitting the following model

$$\Delta\hat{\epsilon}_t = a_1\hat{\epsilon}_{t-1} + \varepsilon_t \tag{4.6.4}$$

where ϵ_t is the residual from the long-run regression relationship, and a_1 is the estimated regression coefficient of the lagged residuals from the long-run relationship, which is the parameter of interest. The null and alternative hypotheses are set as

$$H_0 : a_1 = 0 \tag{4.6.5}$$

$$H_0 : a_1 < 0 \tag{4.6.6}$$

If we cannot reject the null hypothesis in equation 4.6.5, we conclude that the residuals contain a unit root and, hence the variables do not cointegrate. Instead, rejection of the null hypothesis in equation 4.6.5 above implies that residuals are stationary, hence the variables cointegrate.

As extracted from Hamilton (1994) the critical values for the test are given in table 4.7.

Table 4.7: Critical values of ADF test for residuals

	1%	5%	10%
τ	-4.31	-3.77	-3.45

Table 4.8 summarises the ADF test results on residuals of the regression equation.

Table 4.8: Engle-Granger cointegration test

	ADF test statistic	Results
consumption	-6.50	Reject the null hypothesis
income	-6.49	Reject the null hypothesis
wealth	-2.71	Fail to reject the null hypothesis

The ADF critical value at the 10% level of significance is -3.45. Given that the consumption, income and wealth series are integrated of order one, that is, $I(1)$, and that the residuals are

stationary as shown in the summary in table 4.8, we can conclude that the three variables do not cointegrate.

In the next step, the error correction model (ECM) for the cointegrated series, that is, consumption, wealth and income functions are specified. In other words we fit the ECM model below:

$$\Delta lc_t = \beta_0 + \beta_1 \Delta li_t + \beta_2 \Delta lw + \alpha^* \varepsilon_{t-1} + u_t \qquad (4.6.7)$$

where $\alpha^* = -(1 - \alpha_1)$, $\varepsilon_{t-1} = [lc_{t-1} - \beta_0 - \beta_1 li_{t-1} - \beta_2 lw_{t-1}]$ as in equation 3.1.18. Using the necessary R command the error correction model is set out as follows:

```
> lc.d=diff(lc)
> li.d=diff(li)
> lw.d=diff(lw)
> leq2=lag(error.lc)
> ecm=summary(lm(lc.d~li.d+lw.d+leq2))
Call:
lm(formula = lc.d ~ li.d + lw.d + leq2)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.004774   0.001417    3.369  0.00110 **
li.d        0.235293   0.073202    3.214  0.00179 **
lw.d        0.025621   0.031233    0.820  0.41411
leq2       -0.097765   0.081055    1.206  0.0018**
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
Residual standard error: 0.01299 on 94 degrees of freedom
Multiple R-squared: 0.1242,     Adjusted R-squared: 0.09625
F-statistic: 4.444 on 3 and 94 DF,  p-value: 0.005767
```

The estimated ECM is

$$\Delta \hat{lc}_t = 0.0048 + 0.2353 \Delta \hat{li} + 0.0256 \Delta \hat{lw} - 0.0977 \hat{\varepsilon}_{t-1}. \qquad (4.6.8)$$

From the above equation we see that $\alpha^* = -0.0977$ enters with a correct sign (negative) but is large, that is, tends to 0 indicating that the speed of adjustment to equilibrium is slow. We can conclude that, ceteris paribus (keeping other factors constant), consumption and income series converge to a long-run cointegrating equilibrium.

4.6.1.1 Diagnostic error tests

In R, diagnostic tests are performed using the following commands:

```
>jaque.bera.test(lag(error.lc), lag=3)
>arch(lag(error.lc),lag.single=3)
>box.test(lag(errror.lc), lag=1,type="Ljung-Box")
>reset(lag(error.lc),type="regressors")
```

Table 4.9 summarises the results of the diagnostic test on residuals from the *Raotbl*3 data set.

Table 4.9: Results from the diagnostic error tests

Test	Test statistic	p-value	Conclusion
Jarque-Bera	4.29	0.1169	Normally distributed
ARCH-LM	1.03	0.3101	No heteroscedasticity
Ljung-Box	26.64	0.0420	Serial correlation
Ramsey Reset	2.49	0.288	No misspecification

The p-values in table 4.9 are compared with the 0.10 level of significance.

4.6.2 Phillips-Ouliaris methods

The Phillips-Ouliaris methods are implemented by using two residual-based tests, namely

- the variance ratio test

- the multivariate ratio test.

4.6.2.1 The variance ratio test

In R, the variance ratio test is implemented in the function *ca.po* as follows:

```
> lc=ts(lc,start=c(1966,4),end=c(1991,2),frequency=4)
> li=ts(li,start=c(1966,4),end=c(1991,2),frequency=4)
> lw=ts(lw,start=c(1966,4),end=c(1991,2),frequency=4)
> ukcons=window(cbind(lc,li,lw),start=c(1967,2),end=c(1991,2))
> pu.test=summary(ca.po(ukcons,demean='const',type='Pu'))
Test of type Pu detrending of series with constant only
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept) -0.178458   0.103491  -1.724     0.088 .

z[, -1]li    0.910971   0.013457  67.697    <2e-16 ***

z[, -1]lw    0.079761   0.007764  10.274    <2e-16 ***

---

Multiple R-squared: 0.9918,    Adjusted R-squared: 0.9916

Value of test-statistic is: 58.9108
```

Table 4.10 summarises the results of the variance ratio test from the *Raotbl*3 data set.

Table 4.10: Results from the variance ratio test

Level of significance	10%	5%	1%
Critical value	33.67	40.52	53.87
Decision	Reject H_0	Reject H_0	Reject H_0

The test statistic for the variance ratio test $(\hat{P}_u)$ is 58.91. From table 4.10, we observe that this calculated test statistic is larger than the critical value extracted from R at the 1% level of significance. We therefore reject the null hypothesis of no cointegration against the alternative of the presence of cointegrating variables. This leads to the same conclusion reached while using the Engle-Granger two-step estimation procedure, that is, there is not a long-run relationship (cointegration) between consumption, income and wealth series in the *Raotbl*3 data set.

4.6.2.2   The multivariate trace statistic

In R, the multivariate statistic is implemented in the *ca.po* function as follows:

```
> lc=ts(lc,start=c(1966,4),end=c(1991,2),frequency=4)

> li=ts(li,start=c(1966,4),end=c(1991,2),frequency=4)

> lw=ts(lw,start=c(1966,4),end=c(1991,2),frequency=4)

> ukcons=window(cbind(lc,li,lw),start=c(1967,2),end=c(1991,2)

> Pu.test=summary(ca.po(ukcons,demean='const',type='Pu'))

>Pu.test

Test of type Pu detrending of series with constant only

Coefficients:

            Estimate Std. Error t value Pr(>|t|)

(Intercept) -0.178458   0.103491  -1.724     0.088 .

z[, -1]li    0.910971   0.013457  67.697    <2e-16 ***
```

z[, -1]lw   0.079761   0.007764   10.274   <2e-16 ***

Multiple R-squared: 0.9918,   Adjusted R-squared: 0.9916

Value of test-statistic is: 58.9108

Critical values of Pu are:

                      10pct   5pct   1pct

critical values 33.6955 40.5252 53.8731

The output of the multivariate trace statistic is

> pz.test

Response lc :

Call:

lm(formula = lc ~ zr)

Coefficients:

              Estimate Std. Error t value Pr(>|t|)

(Intercept) -0.022293   0.087754   -0.254   0.80003

zrlc   0.853380   0.085127   10.025   < 2e-16 ***

zrli   0.118152   0.078442   1.506   0.13544

zrlw   0.024679   0.009372   2.633   0.00992 **

Residual standard error: 0.01351 on 92 degrees of freedom

Multiple R-squared: 0.9945,   Adjusted R-squared: 0.9943

F-statistic: 5566 on 3 and 92 DF, p-value: < 2.2e-16

Response li :

Call:

lm(formula = li ~ zr)

Coefficients:

              Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.13002   0.10555   1.232   0.2211

zrlc   0.49383   0.10239   4.823 5.58e-06 ***

zrli   0.52867   0.09435   5.603 2.18e-07 ***

zrlw   -0.02429   0.01127   -2.155   0.0338 *

Residual standard error: 0.01625 on 92 degrees of freedom

Multiple R-squared: 0.9918,   Adjusted R-squared: 0.9915

F-statistic: 3697 on 3 and 92 DF, p-value: < 2.2e-16

Response lw

```
Call:

lm(formula = lw ~ zr)

Coefficients:

            Estimate Std. Error t value Pr(>|t|)

(Intercept) -0.31857    0.27819  -1.145    0.255

zrlc         0.13079    0.26986   0.485    0.629

zrli        -0.07984    0.24867  -0.321    0.749

zrlw         0.98357    0.02971  33.106   <2e-16 ***

---

Value of test-statistic is: 88.0345

Critical values of Pz are:

                 10pct     5pct     1pct

critical values 80.2034 89.7619 109.4525
```

Table 4.11 summarises the results of the long-run relationship of the consumption, income and wealth series using the Phillips-Ouliaris test.

Table 4.11: Phillips-Ouliaris:Cointegration test

Level of significance	10%	5%	1%
Critical value of $\hat{P}_u$	33.69	40.52	53.87
Critical value of $\hat{P}_z$	80.20	89.76	109.45

The calculated test statistics of the variance ratio test $(\hat{P}_u)$ and the multivariate trace $(\hat{P}_z)$ are 58.91 and 88.03 respectively. Since both tests are upper-tailed tests, the null hypothesis is rejected if the test statistic is greater than the critical value. This implies than the null hypothesis is rejected at 1% significance level with the variance ratio test, but can only be rejected at the 10% significance level with the multivariate trace statistic.

4.6.3   Johansen's method

Johansen proposes two likelihood ratio tests namely:

- The trace test

- The maximum eigenvalue.

4.6.3.1 The trace test

The trace test tests the null hypothesis of r cointegrating vectors against the alternative hypothesis of n cointegrating vectors. If $r = 0$, it means that there is no relationship among the variables that is stationary. In R, the trace test is implemented in $ca.jo$ function as follows:

```
> summary(ca.jo(data.frame(lc,li,lw), type="trace",ecdet="const"))
Test type: trace statistic , without linear trend and constant in cointegration
Eigenvalues (lambda):
[1] 3.037824e-01 1.028445e-01 2.280181e-02 4.874698e-16
Values of test statistic and critical values of test:

          test 10pct  5pct  1pct
r <= 2 |  2.24  7.52  9.24 12.97
r <= 1 | 12.76 17.85 19.96 24.60
r = 0  | 47.89 32.00 34.91 41.07
Eigenvectors, normalised to first column:
(These are the cointegration relations)

              lc.l2        li.l2        lw.l2    constant
lc.l2    1.00000000   1.00000000   1.0000000   1.0000000
li.l2   -0.95747202  -0.86796562  -1.8045735  -0.9901210
lw.l2   -0.04852978  -0.10453373   0.2832673  -0.5369134
constant 0.29129705   0.02347248   5.0886246   7.0576721
Weights W:
(This is the loading matrix)

          lc.l2        li.l2         lw.l2        constant
lc.d 0.2422490 -0.18249990 -0.0002770341 -7.625075e-15
li.d 0.4914478  0.05256545  0.0046028620 -2.131011e-14
lw.d 0.3021879  0.10564804 -0.0483327582 -8.724908e-15
```

4.6.3.2 The maximum eigenvalue

The maximum eigenvalue test, on the other hand, tests the null hypothesis of r cointegrating vectors against the alternative hypothesis of $(r + 1)$ cointegrating vectors. In R it is implemented in the $ca.jo$ function as follows:

```
> summary(ca.jo(data.frame(lc,li,lw), type="eigen",ecdet="const"))
Test type: maximal eigenvalue statistic (lambda max) , without linear trend
```

```
and constant in cointegration
Eigenvalues (lambda):
[1] 3.037824e-01 1.028445e-01 2.280181e-02 4.874698e-16
Values of teststatistic and critical values of test:

          test 10pct  5pct  1pct
r <= 2 |  2.24  7.52  9.24 12.97
r <= 1 | 10.53 13.75 15.67 20.20
r = 0  | 35.12 19.77 22.00 26.81
Eigenvectors, normalised to first column:
(These are the cointegration relations)

                lc.l2        li.l2       lw.l2    constant
lc.l2      1.00000000   1.00000000   1.0000000   1.0000000
li.l2     -0.95747202  -0.86796562  -1.8045735  -0.9901210
lw.l2     -0.04852978  -0.10453373   0.2832673  -0.5369134
constant   0.29129705   0.02347248   5.0886246   7.0576721
Weights W:
(This is the loading matrix)

          lc.l2         li.l2          lw.l2         constant
lc.d 0.2422490 -0.18249990 -0.0002770341 -7.625075e-15
li.d 0.4914478  0.05256545  0.0046028620 -2.131011e-14
lw.d 0.3021879  0.10564804 -0.0483327582 -8.724908e-15
```

Table 4.12 summarises results of Johansen's cointegration methodology on the *Raotbl*3 data set.

Table 4.12: Johansen's trace test and maximum eigenvalue results

Null hypothesis	Alternative	test statistic	10%	5%	1%	Results
J_{trace} test						
$r \leq 2$	$r > 2$	2.24	7.52	9.24	12.97	Fail to reject H_0
$r \leq 1$	$r > 1$	12.76	17.85	19.96	24.60	Fail to reject H_0
$r = 0$	$r > 0$	47.89	32.00	34.91	41.07	Reject H_0
J_{max} test						
$r = 2$	$r = 3$	2.24	7.52	9.24	12.97	Fail to reject H_0
$r = 1$	$r = 2$	10.53	13.75	15.67	20.20	Fail to reject H_0
$r = 0$	$r = 1$	35.12	19.77	22.00	26.81	Reject H_0

From table 4.12, it can be observed that

(i) the null hypothesis of no cointegration ($r = 0$) against the alternative of presence of one or more cointegrating vector is rejected at the 10% level of significance in both techniques (trace test and maximum eigenvalue). This implies that cointegration exists between the consumption, income and wealth series of the the *Raotbl3* data set.

(ii) the null hypothesis ($r \leq 1$) and ($r \leq 2$) against the alternative of the existence of two or three cointegrating vectors is not rejected by both tests. This means that there is no more than one cointegration relationship in the *Raotbl3* data set.

4.7 Concluding remarks

Since the three methods used to test for cointegration are not consistent, that is, they do not yield the same results, it can be concluded that there is no cointegration between the consumption, income and wealth time series in the *Raotbl3* data set. The results of this analysis can be used to assess the impact of a temporary shock such as a birthday bonus, or a permanent shock such as annual salary raise on an economy.

Chapter 5

Testing for Cointegration in the World Economic Indicators

5.1 Introduction

World economic indicators are specific indices and measures that not only indicate the overall wealth of the global economy, but also provide some insight into its future. Some economic indicators use statistics to illustrate the ups and downs of particular trends in economic activities. The most commonly used world economic indicators are the rates of inflation, the unemployment rates, the real gross domestic product (GDP) growth rates, GDP per capita, GDP purchasing power parity, amounts of foreign direct investments, populations living below the poverty line, and current account balances. In this data set we assess the existence of long-run equilibrium (cointegration) in the following Australia's economic indicators:

- Total Producer Price Index Manufacturing (PPI): This is an index that shows the cost of resources needed to produce manufactured goods during the previous month. The PPI measures change in effective prices received by domestic producers of the manufacturing sector for that part of their output which is sold on the domestic market.

- Domestic Producer Price Index of Finished goods ($DPPI$): This measures how much money manufacturers and wholesalers pay for finished goods and materials.

- Consumer Price Index (CPI): This is a measure which estimates the average price of the consumer goods and services purchased by households. The CPI measures change in a constant market basket of goods and services from one period to the next within the same

area.

If cointegration exists in these economic indicators, it implies that they are procyclic, that is, they move in the direction of the economic movement (or cycle) of a country. This means that the movement of economic indicators is directly proportional to the trend of Australia's economic performance. If cointegration does not exist, it means that these economic indicators are counter-cyclic, that is, they are inversely related to economic performance (Pauly, 2003). Consequently, the PPI, the $DPPI$ and the CPI series in this data set are transformed to natural logarithms creating new variables, which are defined as:

- ppi equivalent to the natural logarithm of PPI

- $dppi$ equivalent to the natural logarithm of $DPPI$

- cpi equivalent to the natural logarithm of CPI.

This transformation is a remedy to the violation of the assumptions of normality in the dataset such as constant variance and independence of the error term.

More details about the data set can be extracted from United Nations World Economic Indicators (http://quanis1.easydata.co.za/TableViewer/tableView.aspx). In this study the world economic indicators data set is included as Appendix B.

5.2   Descriptive summary

A descriptive summary of the data set in natural logarithm is provided.

```
        ppi              dppi             cpi
Min.   :4.387    Min.    :4.368    Min.    :4.403
1st Qu.:4.501    1st Qu.:4.458    1st Qu.:4.507
Median :4.546    Median :4.560    Median :4.578
Mean   :4.585    Mean    :4.569    Mean    :4.579
3rd Qu.:4.699    3rd Qu.:4.669    3rd Qu.:4.651
Max.   :4.808    Max.    :4.758    Max.    :4.734
```

Figure 5.1 is a plot of the total *ppi* for manufacturing, the *dppi* of finished goods and the *cpi* time series.

Figure 5.1: A plot of *ppi*, *dppi* and *cpi* series.
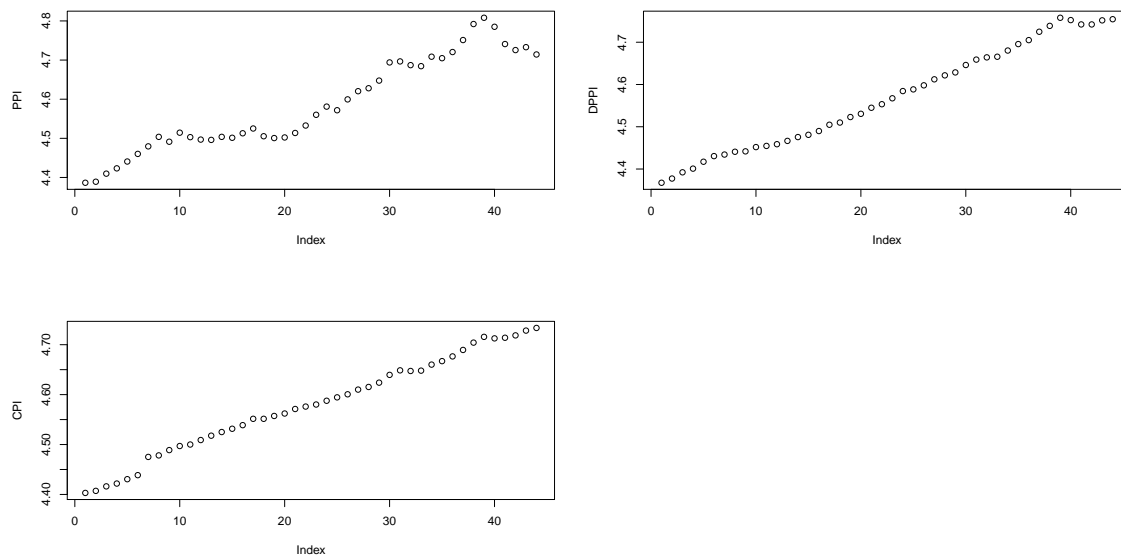
The plot of the *ppi*, *dppi* and *cpi* series indicates nonstationarity in the mean and in the variance with an increasing trend. This implies that transformation and/or differencing is required to reduce the time series to stationarity.

Table 5.1 below provides a descriptive summary of the data set.

Table 5.1: Descriptive summary

	Minimum	Median	Mean	Maximum
ppi	4.387	4.546	4.585	4.808
dppi	4.368	4.560	4.569	4.758
cpi	4.403	4.578	4.579	4.734

From table 5.1, it may be observed that the minimum value, the median, the mean and the maximum value are close to each other. This indicates that this data set is symmetric but does not necessarily show that it is normally distributed.

5.3 Test for normality

We perform a test for normality as follows: Using the Shapiro-Wilk test, we test for normality of variables to see if the series in the data set are well modelled by a normal distribution using the following R command:

```
> shapiro.test(ppi)

        Shapiro-Wilk normality test
data:  ppi
W = 0.8458, p-value = 0.006748
> shapiro.test(dppi)

        Shapiro-Wilk normality test
data:  dppi
W = 0.9618, p-value = 0.004767
> shapiro.test(cpi)

        Shapiro-Wilk normality test
data:  cpi
W = 0.9044, p-value = 2.517e-03
```

Table 5.2 provides the summary results of the normality test using the Shapiro-Wilk test.

Table 5.2: Shapiro-Wilk normality test results

Variable	Shapiro-Wilk	p-value	Decision
ppi	0.8458	6.748×10^{-3}	Reject the null hypothesis
$dppi$	0.9618	4.767×10^{-3}	Reject the null hypothesis
cpi	0.9044	2.517×10^{-3}	Reject the null hypothesis

From the above R command, the p-values are very small (less than 0.01) which means the null hypothesis of normality can be rejected at the 1% level of significance for all series. This implies that the time series of the variables are not normally distributed. However, this was expected, since all variables in this data set were transformed by taking a natural logarithm.

5.4 Augmented Dickey-Fuller (ADF) test for unit roots

In this section we perform the ADF test for unit roots to determine if the ppi, the $dppi$ and the cpi series are nonstationary. After inspection of the behaviour of the ACF plot for the ppi series,

it was found that its residuals become white noise after lag five. This means that the *ACF* plot
for the *ppi* series cuts off at lag five. Since this series shows an increasing trend in figure 5.1, we
test the ADF test on the trend model and lag five as follows:

```
>la=ts(ppi)
> df=ur.df(ppi,lags=5,type='trend')
> summary(df)
##################################################
# Augmented Dickey-Fuller Test Unit Root Test #
##################################################
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.929926   0.465990   1.996   0.0540
la.lag.1    -0.210870   0.106813  -1.974   0.0565
tt           0.001775   0.001031   1.723   0.0940
la.diff.lag1 0.440848   0.173157   2.546   0.0156 *
la.diff.lag2 0.004497   0.186146   0.024   0.9809
la.diff.lag3 0.138793   0.182940   0.759   0.4533
la.diff.lag4 0.024497   0.176146   0.139   0.6809
la.diff.lag5 0.338793   0.192940   1.756   0.5433
Multiple R-squared: 0.216,     Adjusted R-squared: 0.1007
F-statistic: 1.874 on 5 and 34 DF,  p-value: 0.1249
Value of test-statistic is: -1.9742
Value of the test statistic is: -2.1568
Value of the test statistic is: -3.1018
Critical values for test statistics:
      1pct  5pct 10pct
tau3 -4.15 -3.50 -3.18
phi2  7.02  5.13  4.31
phi3  9.31  6.73  5.61
```

Table 5.3 provides summary results of the ADF unit root test.

Table 5.3: Summary of the ADF unit root test

Variable	Test statistic	Decision
ppi	-1.9742	Fail to reject the null hypothesis
$dppi$	-2.1568	Fail to reject the null hypothesis
cpi	-3.1018	Fail to reject the null hypothesis

The critical value at the 10% level of significance of tau3 is -3.18. The results of the ADF unit root test indicate that all variables in this data set are nonstationary. This is because we fail to reject the hypothesis of the presence of unit roots.

Figure 5.2: The plot of the ADF unit root test for the ppi series
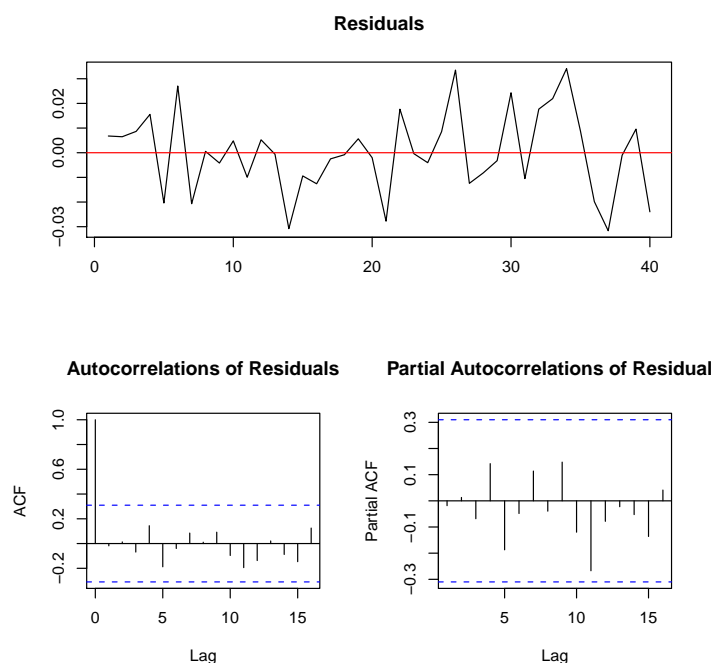
From figure 5.2, it can be observed that the residuals of the ppi series are nonstationary both in the mean and the variance. This suggests differencing and/or transformation to reduce nonstationarity to stationarity. In this way we are determining the order of integration.

Recall that cointegration requires that the series in the data set should be integrated (nonstationary) of the same order and their linear combination must be stationary. Now that we have shown in table 5.3 that all the variables are integrated, in the next section we determine the order of integration.

5.5 Augmented Dickey-Fuller (ADF) test for order of integration

Here we test whether the series is possible $I(2)$, that is, whether it contains a second order of integration. This test is achieved by supplying the differenced series in the *ur.df* R function to test for the presence of unit roots. In R, the order of integration is determined by using the ADF unit root test on a differenced series. We difference the series to transform a nonstationary series to stationarity. Recall that the original series portrayed an increasing trend in figure 5.1 and the ACF is significant at lag five. Therefore, the ADF test for the possible second order of integration is tested on trend model with lag five as follows:

```
> la.ct=diff(ppi)
> la.ct1=ur.df(la.ct,lags=5,type='trend')
> summary(la.ct1)
#################################################
# Augmented Dickey-Fuller Test Unit Root Test #
#################################################
Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.0088469  0.0076142    1.162   0.2536
la.ct.lag.1 -0.7113754  0.3132107   -2.271   0.0298 *
tt          -0.0001833  0.0002734   -0.670   0.5072
la.ct.diff.lag1  0.0707510  0.2680152    0.264   0.7934
la.ct.diff.lag2 -0.0457867  0.2363041   -0.194   0.8476
la.ct.diff.lag3 -0.0499768  0.2039751   -0.245   0.8080
la.ct.diff.lag4 -0.03972     0.13448   -0.295   0.7690
la.ct.diff.lag5 -0.06029     0.13181   -0.457   0.6495
Value of test-statistic is: -2.3812
Value of test-statistic is: -2.2717
Value of test-statistic is: -3.8251
Critical values for test statistics:
      1pct  5pct 10pct
tau3 -4.15 -3.50 -3.18
phi2  7.02  5.13  4.31
phi3  9.31  6.73  5.61
```

56

Figure 5.3 is a plot of the first difference residuals of the *ppi* series. From this plot it can be observed that the mean of the residuals is constant and the $PACF$ is significant at lag five, that is, the $PACF$ cuts off at lag five. This confirms that the trend model with lag five is appropriate to this series in testing for the order of integration.

Figure 5.3: A plot of the $1st$ difference residuals of the *ppi* series
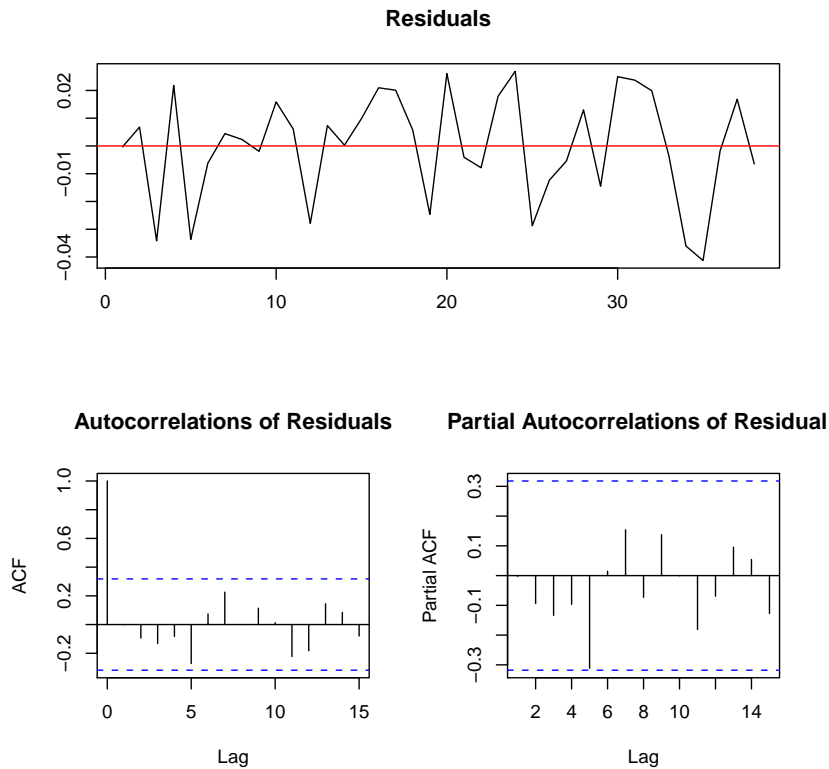
Table 5.4 provides summary results of the first order differenced ADF unit root test.

Table 5.4: Summary of the 1st order difference ADF test

Variable	Test statistic	Decision
Differenced *ppi*	-2.3812	Fail to reject H_0
Differenced *dppi*	-2.2717	Fail to reject H_0
Differenced *cpi*	-3.8251	Reject H_0

Using the critical value at the 5% level of significance, table 5.4 indicates that the null hypothesis of the presence of the first order integration for the *cpi* is rejected. This implies that only the *cpi* is integrated of order one, that is, $I(1;1)$. To obtain the same order of integration for all the variables in the data set, we re-difference the series and perform the unit root test again as follows:

57

```
> la.ct2=diff(la.ct)

> la.ct3=ur.df(la.ct2,lags=5,type='trend')

> summary(la.ct3)

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.0010109  0.0077187   0.131 0.896621
la.ct2.lag.1-2.2249989  0.5516184  -4.034 0.000319 ***
tt          -0.0001356  0.0003125  -0.434 0.667217
la.ct2.diff.lag1  0.7918639  0.4474334   1.770 0.086291 .
la.ct2.diff.lag2  0.3490917  0.3247018   1.075 0.290363
la.ct2.diff.lag3  0.0099522  0.1869345   0.053 0.957873
la.ct2.diff.lag4 -0.0810911  0.1078641  -0.752   0.4543
la.ct2.diff.lag5  0.2129158  0.1067675   1.994   0.0493 *

Value of test-statistic is: -4.0336

Value of test-statistic is: -4.2299

Value of test-statistic is: -4.7550

Critical values for test statistics:
      1pct   5pct  10pct
tau3 -4.15 -3.50 -3.18
phi2  7.02  5.13  4.31
phi3  9.31  6.73  5.61
```

Table 5.5 provides the summary results of the ADF unit root test for the second order of integration.

Table 5.5: Summary of the 2st order difference ADF test

Variable	Test statistic	Decision
Second difference ppi	-4.0336	Reject the null hypothesis
Second difference $dppi$	-4.2299	Reject the null hypothesis
Second difference cpi	-4.7550	Reject the null hypothesis

Using the critical value at the 5% level of significance, table 5.5 indicates that the null hypothesis of the second order difference unit root test is rejected in all series of this data set. This means that all the variables are integrated of order two, that is, $I(2; 2)$. In the following subsections, tests for the presence of long-run relationships (cointegration) are provided by testing for the stationarity

of the residuals of the series.

5.6 Tests for cointegration

5.6.1 Engle-Granger method

This two-step estimation procedure starts with fitting the linear regression models of the series, taking each variable, one at a time, as the dependent variable and the rest as independent variables. Residuals from the regression models are extracted and stored in statistical software R. The ADF test for unit root is then tested on the estimated residuals. If the residuals are stationary, the variables under investigation have long-run relationships. Hence, cointegration exists among variables.

Taking *ppi* as the dependent variable, the regression model is fitted as follows:

```
> data2=cbind(ppi,dppi,cpi)
> ppi.eq=summary(lm(ppi~dppi+cpi),data=data2)
> ppi.eq
Call:
lm(formula = ppi ~ dppi + cpi)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.5497     0.3105   1.770   0.0841 .
dppi          1.1963     0.2169   5.515   2.12e-06 ***
cpi          -0.3124     0.2743  -1.139   0.2614
Multiple R-squared: 0.9572,     Adjusted R-squared: 0.9552
F-statistic:   459 on 2 and 41 DF,  p-value: < 2.2e-16
```

The resulting model is

$$ppi = 0.5497 + 1.1963dppi - 0.3124cpi. \tag{5.6.1}$$

Equation 5.6.1 shows that there is an inverse relationship between the *ppi* and the *cpi*. This implies that as one variable increases the other one decreases. It can also be deduced that the model fits the data set well, $R^2 = 0.9572$. This means that 95.72% of the variations in the *ppi* are explained by changes in the *dppi* and the *cpi*.

When the *dppi* and the *cpi* are taken to be the dependent variables, the following regression models are fitted:

$$dppi = -0.86152 + 0.35597ppi + 0.82953cpi, \tag{5.6.2}$$

$$cpi = 1.02681 - 0.09816ppi + 0.87593dppi. \tag{5.6.3}$$

The estimated residuals of equations 5.6.1, 5.6.2 and 5.6.3 are extracted and stored as:

```
> error.ppi=(resid(ppi.eq))
> error.dppi=(resid(dppi.eq))
> error.cpi=(resid(cpi.eq))
```

The ADF test for unit root on the estimated residuals is performed as:

```
> summary(ur.df(error.ppi,lags=1,type='none'))
######################################################################
# Augmented Dickey-Fuller Test Unit Root / Cointegration Test #
######################################################################
Coefficients:
Estimate Std. Error t value Pr(>|t|)
error.PPI.lag.1    -0.21045    0.08867  -2.373   0.0225 *
error.PPI.diff.lag  0.38141    0.15300   2.493   0.0169 *
The value of the test statistic is: -2.3734
The value of the test statistic is: -2.4588
The value of the test statistic is: -2.0856
```

As extracted from Hamilton (1994) the critical values for the test are given in table 5.6.

Table 5.6: Critical values of ADF test for residuals

	1%	5%	10%
τ	-4.31	-3.77	-3.45

Table 5.7 provides summary results for the ADF unit root test for the nonstationarity of residuals in the estimated regression model.

Table 5.7: Engle-Granger cointegration test

Variable	Test statistic	Decision
error.ppi	-2.3734	Fail to reject the null hypothesis
error.dppi	-2.4588	Fail to reject the null hypothesis
error.cpi	-2.0856	Fail to reject the null hypothesis

The ADF critical value at the 10% level of significance is -3.45. From table 5.7 the null hypothesis of the presence of unit roots in the residuals of the regression model is not rejected at the 10% level of significance. This implies that the residuals of the estimated regression models are nonstationary.

Figure 5.4: A plot of the ADF unit root test on the residuals of the *ppi* series
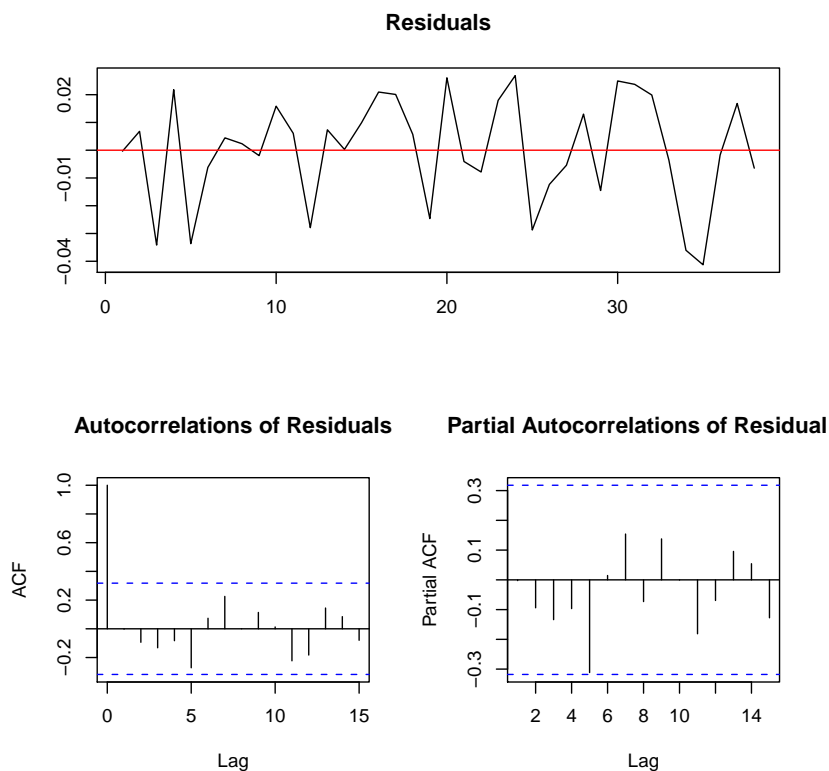
Figure 5.4 shows that the residuals of the *ppi* series are nonstationary. This confirms that the variables in the data set do not cointegrate.

Based on the two-step estimation procedure, it can be concluded that Australia's PPI for manu-

facturing, the $DPPI$ for finished goods and its CPI do not cointegrate.

In the next step, the error correction model (ECM) is fitted as follows:

$$\Delta ppi_t = \beta_0 + \beta_1 \Delta dppi_t + \beta_2 \Delta cpi + \alpha^* \varepsilon_{t-1} + u_t \qquad (5.6.4)$$

where $\alpha^* = -(1 - \alpha_1)$, $\varepsilon_{t-1} = [ppi_{t-1} - \beta_0 - \beta_1 dppi_{t-1} - \beta_2 cpi_{t-1}]$ as in equation 3.1.18. Using the necessary R command the error correction model is set out as follows:

```
> ppi.d=diff(ppi)
> dppi.d=diff(dppi)
> cpi.d=diff(cpi)
> leq1=lag(error.ppi)
> ecm=summary(lm(ppi.d~dppi.d+cpi.d+leq1))
Call:
lm(formula = ppi.d ~ dppi.d + cpi.d + leq1)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.005678   0.001417   4.369  0.00510
dppi.d      0.978583   0.073202   9.214  0.00379 **
cpi.d       0.679321   0.031233   0.120  0.41411
leq1       -0.790765   0.081055   1.606  0.0018**
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
Residual standard error: 0.7899 on 44 degrees of freedom
Multiple R-squared: 0.7642,     Adjusted R-squared: 0.09625
F-statistic: 5.444 on 3 and 44 DF,  p-value: 0.005767
```

The estimated ECM is

$$\Delta \hat{ppi}_t = 0.0057 + 0.9785 \Delta \hat{dppi} + 0.6793 \Delta \hat{cpi} - 0.7908 \hat{\varepsilon}_{t-1}. \qquad (5.6.5)$$

From the above equation, $\alpha^* = -0.7908$ enters with a small correct sign (negative), that is, tends to -1 indicating that the speed of adjustment to equilibrium is high. We can conclude that, ceteris paribus (keeping other factors constant), Australia's economic indicators do not cointegrate and the ECM pushes the economy back to equilibrium at a high rate.

5.6.1.1 Diagnostic error tests

In R, diagnostic tests are performed using the following commands:

```
>jaque.bera.test(lag(error.ppi), lag=3)

>arch(lag(error.ppi),lag.single=3)

>box.test(lag(errror.ppi), lag=1,type="Ljung-Box")

>reset(lag(error.ppi),type="regressors")
```

Table 5.8 summarises the results of the diagnostic test on residuals from the *Raotbl*3 data set.

Table 5.8: Results from the diagnostic error tests

Test	Test statistic	*p*-value	Conclusion
Jarque-Bera	3.26	0.2116	Normally distributed
ARCH-LM	2.07	0.2101	No heteroscedasticity
Ljung-Box	13.64	0.0020	Serial correlation
Ramsey Reset	4.45	0.0328	Misspecification

The *p*-values in table 5.8 are compared with the 0.10 significance level.

5.6.2 Phillips-Ouliaris methods

The Phillips-Ouliaris methods are implemented by using two residual-based tests.

5.6.2.1 The variance ratio test

In *R*, the variance ratio test is implemented in the function *ca.po* as follows:

```
> Pu.test=summary(ca.po(data2,demean='const',type='Pu'))

> Pu.test

###########################################

# Phillips and Ouliaris Unit Root Test #

###########################################

Test of type Pu detrending of series with constant only

Coefficients:

            Estimate Std. Error t value Pr(>|t|)

(Intercept)   0.5497     0.3105   1.770   0.0841 .

z[, -1]dppi   1.1963     0.2169   5.515 2.12e-06 ***

z[, -1]cpi   -0.3124     0.2743  -1.139   0.2614

Value of test-statistic is: 5.5393

Critical values of Pu are:
```

```
                10pct     5pct     1pct
critical values 33.6955 40.5252 53.8731
```

Table 5.9 summarises the results of the variance ratio test from the economic indicators data set.

Table 5.9: Results from the variance ratio test

Level of significance	10%	5%	1%
Critical value	33.67	40.52	53.87
Decision	Fail to reject H_0	Fail to reject H_0	Fail to reject H_0

The test statistic for the variance ratio test $(\hat{P}_u)$ is 5.54. From the above analysis, we observe that this calculated test statistic is smaller than the critical values as extracted from R at the 10% level of significance. Therefore, at all significance levels the null hypothesis of no cointegration is not rejected.

5.6.2.2   The multivariate trace statistic

In R, the multivariate trace statistic is implemented in the function *ca.po* as follows:

```
> Pz.test=summary(ca.po(data2,demean='const',type='Pz'))
> Pz.test
Response ppi:
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.4780     0.2235   2.139   0.0388 *
zrppi         0.7959     0.1122   7.095 1.59e-08 ***
zrdppi        0.4239     0.1989   2.131   0.0395 *
zcpi         -0.3213     0.1921  -1.673   0.1024
Response dppi :
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.09041    0.08092   1.117    0.271
zrppi       -0.06707    0.04062  -1.651    0.107
zrdppi       1.08963    0.07201  15.131   <2e-16 ***
zrcpi       -0.04005    0.06954  -0.576    0.568
Response cpi :
Coefficients:
```

```
           Estimate Std. Error t value Pr(>|t|)

(Intercept)  0.174910   0.075116   2.329   0.0252 *

zrppi       -0.006555   0.037702  -0.174   0.8629

zrdppi       0.104021   0.066848   1.556   0.1278

zrcpi        0.866235   0.064553  13.419 3.38e-16 ***

Value of test-statistic is: 15.1071

Critical values of Pz are:

                 10pct    5pct     1pct

critical values 80.2034 89.7619 109.4525
```

Table 5.10 summarises the results of the multivariate trace statistic test from the economic indicators data set.

Table 5.10: Results from the multivariate trace statistic test

Level of significance	10%	5%	1%
Critical value	80.2034	89.7619	109.4525
Decision	Fail to reject H_0	Fail to reject H_0	Fail to reject H_0

The test statistic for the multivariate test statistic ($\hat{P}_z$) is 15.11. From the above analysis, we observe that this calculated test statistic is smaller than the critical values extracted from R at the 10% level of significance. Therefore, at all levels of significance the null hypothesis of presence of no cointegration is not rejected.

5.6.3 Johansen's method

Johansen's method is implemented on two likelihood ratio tests, namely:

- The trace test

- The maximum eigenvalue.

5.6.3.1 The trace test

In R, the trace statistic is implemented as follows:

```
> summary(ca.jo(data2,type='trace',ecdet='const'))
######################
# Johansen-Procedure #
######################
```

```
Trace statistic , without linear trend and constant in cointegration

Eigenvalues (lambda):

[1] 5.875376e-01 2.607680e-01 1.616397e-01 1.135093e-15

Values of teststatistic and critical values of test:

           test 10pct  5pct  1pct

r <= 2 |   7.40   7.52   9.24 12.97

r <= 1 |  20.09  17.85 19.96 24.60

r = 0  |  57.29  32.00 34.91 41.07

Eigenvectors, normalised to first column:

(These are the cointegration relations)

              ppi.l2      dppi.l2   cpi.l2   constant

ppi.l2     1.0000000   1.000000   1.0000000  1.000000

dppi.l2   -0.4848253  -1.893377   0.6822461 -4.221935

cpi.l2    -0.3169917   1.282587  -2.2073541  2.920107

constant  -1.8389176  -1.808761   2.4083450  1.317851

Weights W:

(This is the loading matrix)

         ppi.l2         dppi.l2        cpi.l2        constant

ppi.d -0.007918325 -0.31229813 -0.02543182 -2.390012e-13

dppi.d -0.010170771 -0.06330496 -0.03353216 -2.505351e-15

cpi.d -0.011029486 -0.05050685  0.03556958 -9.709437e-14
```

5.6.3.2   The maximum eigenvalue

The multivariate trace statistic is implemented as follows:

```
> summary(ca.jo(data2,type='eigen',ecdet='const'))

######################

# Johansen-Procedure #

######################

Maximal eigenvalue statistic (lambda max) , without linear trend and constant in cointegration

Eigenvalues (lambda):

[1] 5.875376e-01 2.607680e-01 1.616397e-01 1.135093e-15

Values of teststatistic and critical values of test:

           test 10pct  5pct  1pct

r <= 2 |   7.40   7.52   9.24 12.97
```

```
r <= 1 | 12.69 13.75 15.67 20.20

r = 0  | 37.20 19.77 22.00 26.81
```

Eigenvectors, normalised to first column:

(These are the cointegration relations)

```
               ppi.l2     dppi.l2    cpi.l2   constant
ppi.l2     1.0000000   1.000000   1.0000000   1.000000
dppi.l2   -0.4848253  -1.893377   0.6822461  -4.221935
cpi.l2    -0.3169917   1.282587  -2.2073541   2.920107
constant  -1.8389176  -1.808761   2.4083450   1.317851
```

Weights W:

(This is the loading matrix)

```
          ppi.l2       dppi.l2       cpi.l2       constant
ppi.d -0.007918325 -0.31229813 -0.02543182 -2.390012e-13
dppi.d -0.010170771 -0.06330496 -0.03353216 -2.505351e-15
cpi.d -0.011029486 -0.05050685  0.03556958 -9.709437e-14
```

Table 5.11 provides summary results of Johansen's cointegration method.

Table 5.11: Johansen's trace test and maximum eigenvalue results

Null hypothesis	Alternative	Test statistic	10%	5%	1%	Results
$\jmath_{trace}$ test						
$r \leq 2$	$r > 2$	7.40	7.52	9.24	12.97	Do not reject H_0
$r \leq 1$	$r > 1$	20.09	17.85	19.96	24.60	Reject H_0
$r = 0$	$r > 0$	57.29	32.00	34.91	41.07	Reject H_0
$\jmath_{max}$ test						
$r = 2$	$r = 3$	7.40	7.52	9.24	12.97	Do not reject H_0
$r = 1$	$r = 2$	12.69	13.75	15.67	20.20	Reject H_0
$r = 0$	$r = 1$	37.20	19.77	22.00	26.81	Reject H_0

From table 5.11 it can be observed that

- cointegration exists among Australia's economic indicators, since the null hypothesis of no cointegration ($r = 0$) is rejected

- there is at most one cointegration vector relationship between these variables, since the null hypothesis of $r \leq 2$ is not rejected.

5.7 Concluding remarks

From the results of the analysis, it is observed that the Engle-Granger method and the Phillips-Ouliaris methods indicate no cointegration yet the Johansen's method indicates presence of cointegration in Australia's economic indicators. Since the three methods are not consistent, that is, they do not yield the same results, it can be concluded that the PPI (Manufacturing), the PPI (Finished goods) and the CPI do not cointegrate. This means that these economic indicators are countercyclic, that is, they move in opposite directions of Australia's economy.

Chapter 6

Testing Cointegration Results for the $UKpppuip$ Data Set

6.1   Introduction

In this data set we test for cointegration in the purchasing power parity (PPP) and the uncovered interest rate parity (UIP) for the United Kingdom. The PPP is the economic concept that continuously adjusts exchange rates between countries in order to denote the purchasing power of each country; that is, PPP refers to the use of the long-term equilibrium exchange rate of two countries to equalise purchasing power.

Purchasing power is the number of goods/services that can be purchased with a unit of currency. A parity condition occurs when the difference in the interest rate between two countries is equal to the expected change in exchange rate between the countries' currencies. It can be expressed as

$$(i_1 - i_2) = E(e) \tag{6.1.1}$$

where, i_1 represents the interest rate in country 1, i_2 represents the interest rate in country 2, and $E(e)$ represents the expected rate of change in the exchange rate.

In the $UKpppuip$ data set, we use quarterly data spanning a range from 1972:Q1 to 1987:Q2 where Q1 and Q2 are the first and second quarter respectively. In this study the $UKpppuip$ data set is included as Appendix C.

The variables are

p_1 : UK wholesale price index.

p_2 : Trade weighted foreign wholesale price.

i_1 : Three month Treasury Bill rate.

i_2 : Three month Eurodollar rate.

e_{12} : UK effective exchange rate.

$dpoilp_0$: World oil price at period t.

$dpoilp_1$: World oil price at period t_{-1}.

Figure 6.1 is a plot of the $UKpppuip$ data set.

Figure 6.1: A plot of p_1, p_2, e_{12}, i_1, i_2, $doilp_0$ and $doilp_1$ series
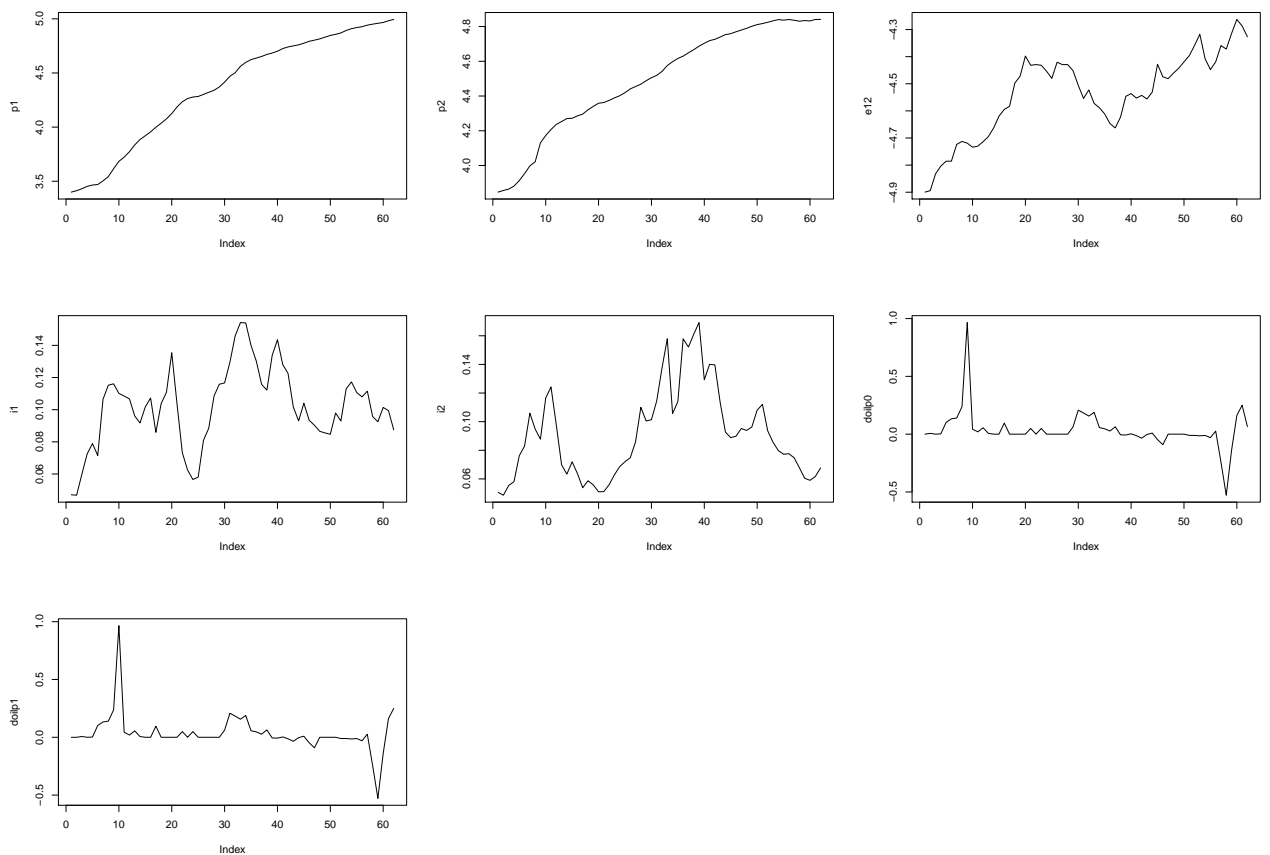
Figure 6.1 indicates nonstationarity in the mean with an increasing trend for the p_1, p_2 and the e_{12} series. For the i_1 and the i_2 series the plot indicates nonstationarity in both the mean and the variance. Moreover, the series of the $doilp_0$ and the $doilp_1$ appear to be stationary. These results can only be confirmed after performing a formal unit root test for nonstationarity.

6.2   Descriptive summary

A descriptive summary of the $UKpppuip$ data set in natural logarithm is provided.

```
> library(urca)
>attach(UKpppuip)
> summary(UKpppuip)
      p1               p2              e12               i1
 Min.   :3.400   Min.   :3.847   Min.   :-4.900   Min.   :0.04679
 1st Qu.:3.967   1st Qu.:4.288   1st Qu.:-4.641   1st Qu.:0.08771
 Median :4.486   Median :4.530   Median :-4.514   Median :0.10368
 Mean   :4.361   Mean   :4.494   Mean   :-4.538   Mean   :0.10182
 3rd Qu.:4.801   3rd Qu.:4.777   3rd Qu.:-4.429   3rd Qu.:0.11560
 Max.   :4.995   Max.   :4.841   Max.   :-4.263   Max.   :0.15418
      i2              doilp0            doilp1
 Min.   :0.04860   Min.   :-0.52843   Min.   :-0.52843
 1st Qu.:0.06471   1st Qu.: 0.00000   1st Qu.: 0.00000
 Median :0.08687   Median : 0.00000   Median : 0.00000
 Mean   :0.09105   Mean   : 0.03581   Mean   : 0.03476
 3rd Qu.:0.10953   3rd Qu.: 0.05625   3rd Qu.: 0.05384
 Max.   :0.16924   Max.   : 0.96554   Max.   : 0.96554
```

Table 6.1 provides a descriptive summary of the $UKpppuip$ data set.

Table 6.1: Descriptive summary of the $UKpppuip$ data set

Variable	Minimum	Median	Mean	Maximum
p_1	3.400	4.486	4.361	4.995
p_2	3.847	4.530	4.494	4.841
e_{12}	-4.900	-4.514	-4.538	-4.263
l_1	0.048	0.104	0.102	0.154
i_2	0.049	0.087	0.091	0.169
$doilp_0$	-0.528	0.000	0.036	0.966
$doilp_1$	-0.528	0.000	0.035	0.966

6.3 Test for normality

The null and the alternative hypothesis of this set are set as follows:

H_0 : The series in the data set are normally distributed

H_1 : The series in the data set are not normally distributed

Using the Shapiro-Wilk test, we test for the normality of the variables to see if the series in the data set are well modelled by a normal distribution using the following R command:

```
> shapiro.test(p1)

        Shapiro-Wilk normality test

data:  normality
W = 0.8997, p-value = 9.973e-05
> shapiro.test(p2)

        Shapiro-Wilk normality test

data:  p2
W = 0.9025, p-value = 0.0001275
> shapiro.test(e12)

        Shapiro-Wilk normality test

data:  e12
W = 0.9637, p-value = 0.06355
> shapiro.test(i1)

        Shapiro-Wilk normality test

data:  i1
W = 0.9823, p-value = 0.5102
> shapiro.test(i2)

        Shapiro-Wilk normality test

data:  i2
W = 0.9309, p-value = 0.001783
> shapiro.test(doilp0)

        Shapiro-Wilk normality test

data:  doilp0
W = 0.6285, p-value = 2.931e-11
> shapiro.test(doilp1)

        Shapiro-Wilk normality test
```

```
data:  doilp1

W = 0.6237, p-value = 2.439e-11
```

Table 6.2 provides the summary results of the normality test using the Shapiro-Wilk test.

Table 6.2: Normality test on the *Raotbl*3 data set

Variable	Shapiro-Wilk	p-value	Decision
p_1	0.8997	9.973×10^{-5}	Reject the null hypothesis
p_2	0.9025	1.275×10^{-4}	Reject the null hypothesis
e_{12}	0.9637	0.6355×10^{-1}	Reject the null hypothesis
i_1	0.9823	5.102×10^{-1}	Do not reject the null hypothesis
i_2	0.9309	1.783×10^{-3}	Reject the null hypothesis
$doilp_0$	0.6285	2.931×10^{-11}	Reject the null hypothesis
$doilp_1$	0.6237	2.439×10^{-11}	Reject the null hypothesis

Since the p-values are very small (less than 0.01), the null hypothesis of normality can be rejected at the 1% level of significance for all variables except the i_1. This implies that most of these time series variables are not all normally distributed.

6.4 Augmented Dickey-Fuller (ADF) test for unit roots

The aim of the unit root test is to test for nonstationarity of the variables in the data set. A nonstationarity test is performed by testing for the existence of unit roots in each variable of the data set. We set the null and the alternative hypotheses as

$$H_0 : \delta = 0 \tag{6.4.1}$$

$$H_1 : \delta < 0. \tag{6.4.2}$$

If the null hypothesis of $\delta = 0$ is not rejected, it means that the variable contains a unit root, that is, it is nonstationary.

After inspection of the behaviour of the ACF plot for the p_1 series, it was found that its residuals become white noise after lag six. This means that the ACF plot for the p_1 series cuts off at lag six. Since this series shows an increasing trend in figure 6.1, we test the ADF test on the trend model and lag six as follows:

```
> p1.ct=ur.df(p1,lags=6,type="trend")

> summary(p1.ct)

Test regression trend

Call:

lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)

Coefficients:

              Estimate Std. Error t value Pr(>|t|)

(Intercept)  0.0968845  0.0503226   1.925  0.06039 .

z.lag.1     -0.0146446  0.0145885  -1.004  0.32070

tt          -0.0002650  0.0004393  -0.603  0.54932

z.diff.lag1  0.4869411  0.1425790   3.415  0.00134 **

z.diff.lag2 -0.0922546  0.1518551  -0.608  0.54650

z.diff.lag3 -0.0151649  0.1525544  -0.099  0.92125

z.diff.lag4  0.0150175  0.1524567   0.099  0.92196

z.diff.lag5 -0.0708829  0.1510372  -0.469  0.64107

z.diff.lag6 -0.1614905  0.1248796  -1.293  0.20241

Multiple R-squared: 0.7543,      Adjusted R-squared: 0.7115

Value of test-statistic is: -1.0038 8.5796 10.7345
```

The estimated model is

$$p1_t = 0.09869 - 0.000265tt - 0.014645p1_{t-1} - 0.48694\Delta p1_{t-2} - \ldots - 0.16149\Delta p1_{t-6}. \quad (6.4.3)$$

The calculated test statistic is derived from

$$F_\tau = \frac{\hat{\delta}}{SE(\hat{\delta})} = \frac{-0.0146446}{0.0145885} = -1.0038. \quad (6.4.4)$$

As extracted from R, the critical values for the ADF unit root test are given in table 6.3.

Table 6.3: Critical values for the test

	1%	5%t	10%
τ	-4.04	-3.45	-3.15
α_1	6.50	4.88	4.16
α_2	8.73	6.49	5.47

Since the calculated test statistic, -1.0038, falls within the non-rejection region, that is, to the right of the τ(tau) critical values, we cannot reject the null hypothesis for the presence of unit

roots at the 10% level of significance. This means that the variable, UK wholesale price index, p_1, in the data set contains a unit root.

Figure 6.2 is a plot of the residuals of the foreign wholesale price ($p1$). From the figure, we observe that the series is nonstationary in the mean but not in the variance.

Figure 6.2: A plot of the ADF unit root test for the $p1$ series
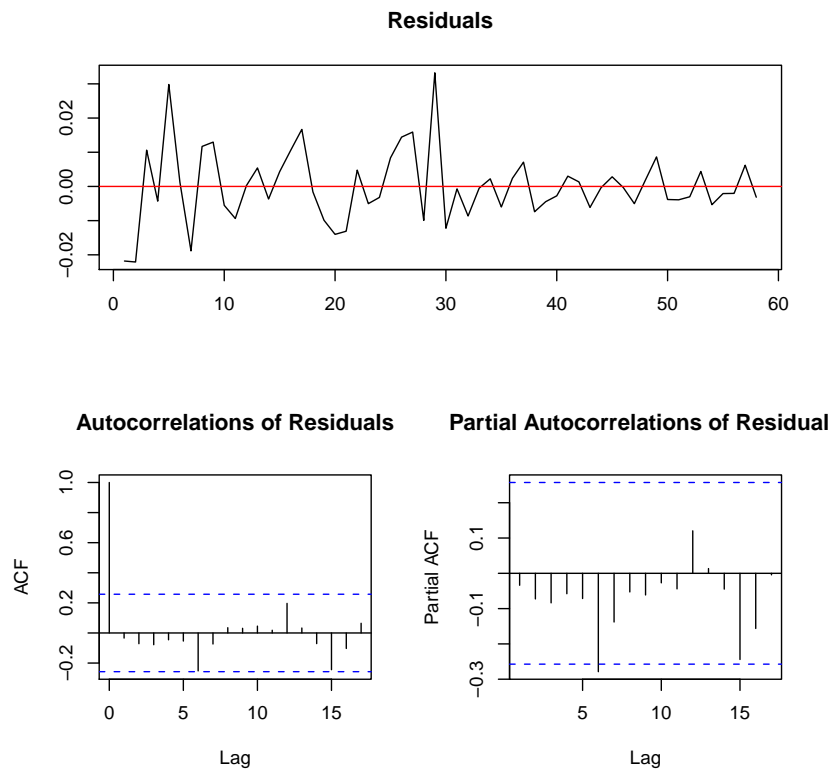
The ADF test output for unit roots for all other variables in this data set is obtained using the following R commands:

```
> p2.ct=ur.df(p2,lags=6,type='trend')
> summary(p2.ct)
>plot(p1.ct)
> e12.ct=ur.df(e12,lags=6,type='trend')
> summary(e12.ct)
> i1.ct=ur.df(i1,lags=6,type='none')
> summary(i1.ct)
> i2.ct=ur.df(i2,lags=6,type='none')
```

```
> summary(i2.ct)

> doilp0.ct=ur.df(doilp0,lags=6,type=none')

> summary(doilp0.ct)

> doilp1.ct=ur.df(doilp1,lags=6,type='none')

> summary(doilp1.ct)
```

Table 6.4 summarises the results of the ADF unit root test for the $UKpppuip$ data set.

Table 6.4: Summary of the ADF test for unit roots in the variables

Variable	ADF test statistic	Decision
Wholesale price index (p_1)	-1.0038	Fail to reject the null hypothesis
Weighted foreign wholesale price (p_2)	-1.4112	Fail to reject the null hypothesis
Effective exchange rate (e_{12})	-2.0664	Fail to reject the null hypothesis
Treasury bill rate (i_1)	-0.4923	Fail to reject the null hypothesis
Eurodollar rate (i_2)	-0.9492	Fail to reject the null hypothesis
World oil price at period t $(doilp_0)$	-4.1297	Reject the null hypothesis
Word oil price at period t_1 $(doilp_1)$	-4.1042	Reject the null hypothesis

The ADF critical value at the 10% level of significance is -3.15. Table 6.4 shows that the null hypothesis of presence of unit roots for variables $doilp0$ and $doilp1$ is rejected. This means that all other variables are nonstationary except for the world oil price.

It is possible for the world oil price to be stationary since its price is fixed by international organisations outside a country, for example the Organisation of Petroleum Exporting Countries (OPEC). Such prices are relatively stable because they are not affected by structural fluctuations in a country's economic activities. For example, if employees strike in a country like South Africa, the price of a barrel of oil on the international market is not affected, although the exchange rates in South Africa may be affected.

Since the first five variables contain a unit root, it is possible that they are integrated. In the next section, the ADF test for the order of integration is performed.

6.5 Augmented Dickey-Fuller (ADF) test for order of integration

In this test, each variable in the data set is differenced to reduce the nonstationary series to stationarity, and the ADF test is performed on the differenced series for a possible second order of integration, $I(2)$, using the following R commands:

```
> p1t=diff(p1)
> p1t.ct=ur.df(p1t)
> plot(p1t.ct)
```

Figure 6.3: A plot of the first difference p_1, p_2, e_{12}, i_1, i_2, $doilp_0$ and $doilp_1$ series
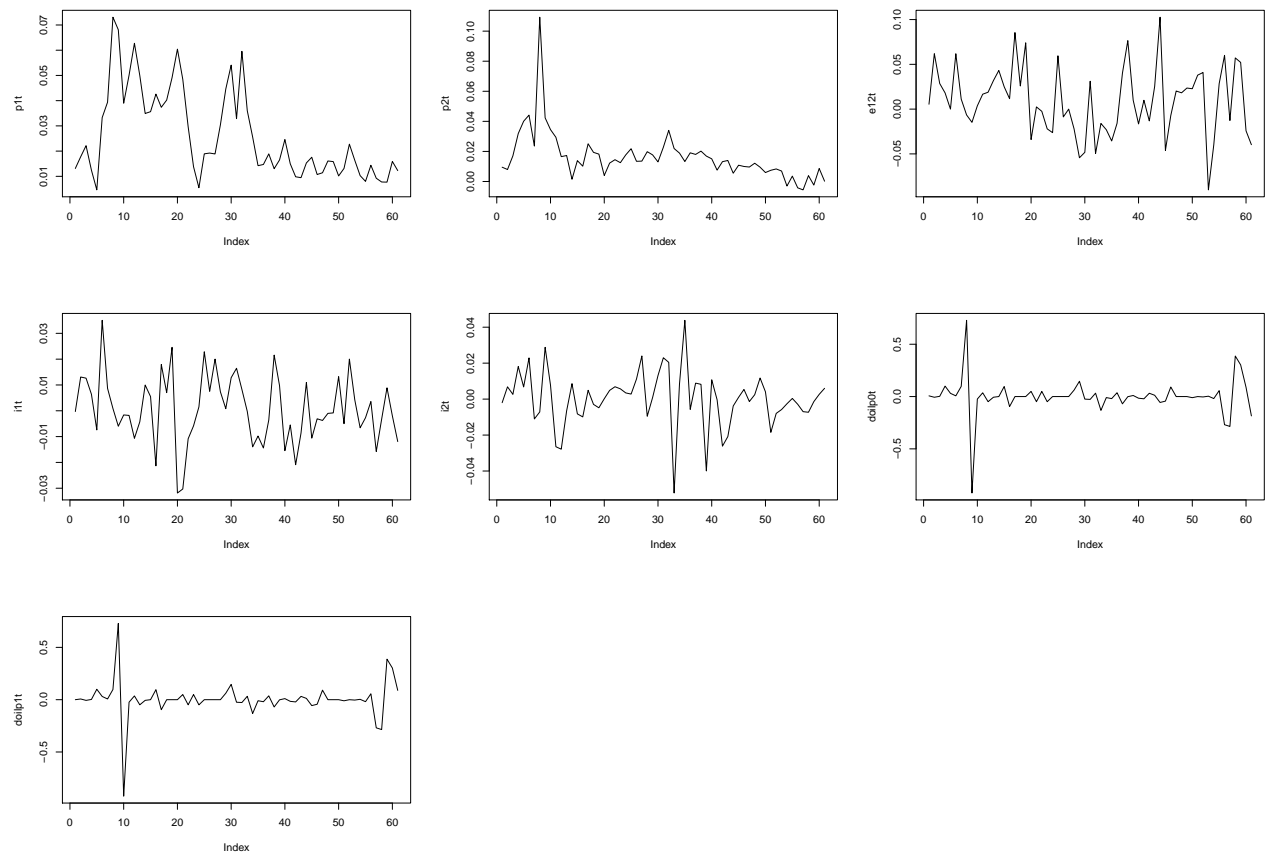
From the plot, it can be observed that the first difference series are nonstationary in the mean but not in the variance. This means that the no trend model is appropriate in testing for the unit root. Inspection of the ACF plots showed that the ACF's of the series in this data set are significant at

lag six. We therefore test for the possible second order of integration using the ADF test for unit root on the no trend model with six lags are follows:

```
> p1t=diff(p1)
> p1t.ct=ur.df(p1t, lags=6,type='none')
> summary(p1t.ct)
> p2t=diff(p2)
> p2t.ct=ur.df(p2t, lags=6,type='none')
> summary(p2t.ct)
> e12t=diff(e12)
> e12t.ct=ur.df(e12t, lags=6,type='none')
> summary(e12t.ct)
> i1t=diff(i1)
> i1t.ct=ur.df(i1t, lags=6,type='none')
> summary(i1t.ct)
> i2t=diff(i2)
> i2t.ct=ur.df(i2t, lags=6,type='none')
> summary(i2t.ct)
```

For the UK wholesale price index, the output is given by:

```
> summary(p1t.ct)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
z.lag.1     -0.04719    0.05086  -0.928   0.3582
z.diff.lag1 -0.05744    0.14086  -0.408   0.6853
z.diff.lag2 -0.11241    0.13492  -0.833   0.4089
z.diff.lag3 -0.10667    0.13526  -0.789   0.4343
z.diff.lag4 -0.03972    0.13448  -0.295   0.7690
z.diff.lag5 -0.06029    0.13181  -0.457   0.6495
z.diff.lag6 -0.26352    0.13173  -2.000   0.0513 .
Multiple R-squared: 0.1351,     Adjusted R-squared: 0.006268
Value of test-statistic is: -0.928


Critical values for test statistics:
     1pct  5pct 10pct
```

```
tau1 -2.6 -1.95 -1.61
```

Figure 6.4 is a plot of the ADF test on the first difference p_1 series.

Figure 6.4: A plot of the ADF unit root test for the first difference p_1 series
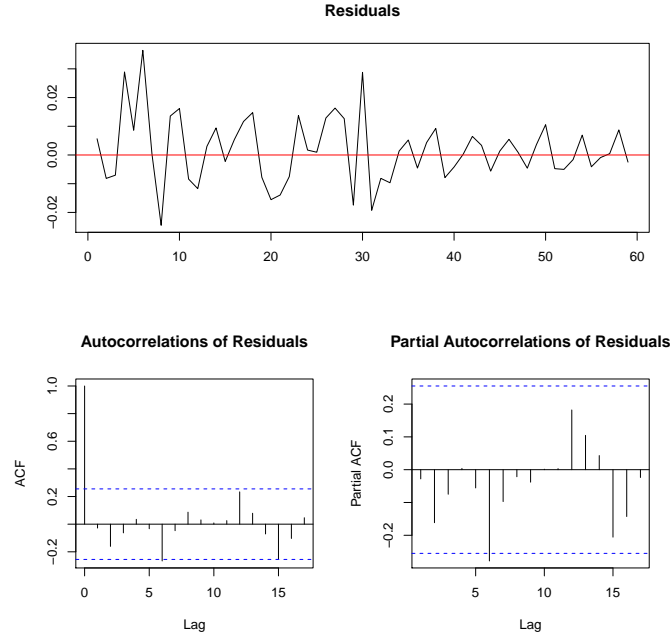
Table 6.5 summarises the results of the ADF test order of integration.

Table 6.5: Summary for ADF test for the order of integration

Variable	Test statistic	Decision
Differenced Wholesale price index (Δp_1)	-0.928	Fail to reject the null hypothesis
Differenced Weighted foreign wholesale price (Δp_2)	-1.5796	Fail to reject the null hypothesis
Differenced Effective exchange rate (Δe_{12})	-2.1458	Reject the null hypothesis
Differenced Treasury bill rate Δi_1	-3.6266	Reject the null hypothesis
Differenced Eurodollar rate Δi_2	-2.7766	Reject the null hypothesis

The ADF critical value at the 10% level of significance is -1.61. Only three variables in table 6.5 are integrated of order one, $I(1,1)$, the presence of a long-run equilibrium relationship (cointegration) is tested for in the following sections.

6.6 Tests for cointegration

6.6.1 Engle-Granger method

In the two-step estimation procedure, we fit a regression line of nonstationary variables taking each variable, one at a time, to be the dependent variable and the rest as independent variables. Thereafter, we test for the stationarity of the residuals from each estimated regression line. If the residuals are found to be stationary, then the variables cointegrate. In R this is done as follows:

```
> data1=cbind(p1,p2,e12,i1,i2)
> p1.eq=summary(lm(p1~p2+e12+i1+i2),data=data1)
> p1.eq
Call:
lm(formula = p1 ~ p2 + e12 + i1 + i2)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.30081    0.80040  -2.875  0.00568 **
p2           1.61307    0.06309  25.566  < 2e-16 ***
e12          0.11967    0.12278   0.975  0.33383
i1          -0.70801    0.46090  -1.536  0.13004
i2           0.31163    0.44207   0.705  0.48372
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
Residual standard error: 0.06547 on 57 degrees of freedom
Multiple R-squared: 0.9847,     Adjusted R-squared: 0.9836
F-statistic: 918.2 on 4 and 57 DF,  p-value: < 2.2e-16
Correlation of Coefficients:
    (Intercept) p2     e12    i1
p2  -0.94
e12  0.99        -0.88
i1  -0.17         0.05 -0.18
i2   0.63        -0.60  0.63 -0.55
```

Considering p_1 as the dependent variable, the estimated regression line is

$$\hat{p}_1 = -2.30081 + 1.61307p_2 + 0.11967e_{12} - 0.70801i_1 + 0.31163i_2. \tag{6.6.1}$$

From the output we observe that only p_2 is statistically significant ($p\text{-}value = 2 \times 10^{-16}$) at the 1% significance level. Since the $R^2 = 0.98$ is higher and the correlation coefficient between p_2 and

e_{12} is also high (-0.88), we suspect serial correlation or misspecification or both in the residuals of p_1, but we have to perform diagnostic tests to confirm this.

Taking the other remaining variables as dependent variables, we fit the following regression lines:

$$\hat{p}_2 = 2.27064 + 0.57021p_1 + 0.06968e_{12} + 0.375181i_1 + 0.15738i_2. \tag{6.6.2}$$

$$\hat{e}_{12} = -6.0199 + 0.1370p_1 + 0.2256p_2 + 0.76101i_1 - 2.2714i_2. \tag{6.6.3}$$

$$\hat{i}_1 = 0.15960 - 0.05615p_1 + 0.08417p_2 + 0.05272e_{12} + 0.52827i_2. \tag{6.6.4}$$

$$\hat{i}_2 = -1.06968 + 0.02773p_1 + 0.03962p_2 - 0.17659e_{12} + 0.59285i_1. \tag{6.6.5}$$

The residuals of the estimated regression lines above are extracted from the following commands:

```
> error.p1=(resid(p1.eq))
> error.p2=(resid(p2.eq))
> error.e12=(resid(e12.eq))
> error.i1=(resid(i1.eq))
> error.i2=(resid(i2.eq))
```

Nonstationarity in the estimated residuals of the regression lines is tested for using the the ADF test as follows:

```
> p1.lc=ur.df(error.p1,type='none')
> summary(p1.lc)
>plot(pl.lc)
```

Figure 6.5: A plot of ADF test on residuals of wholesale price.
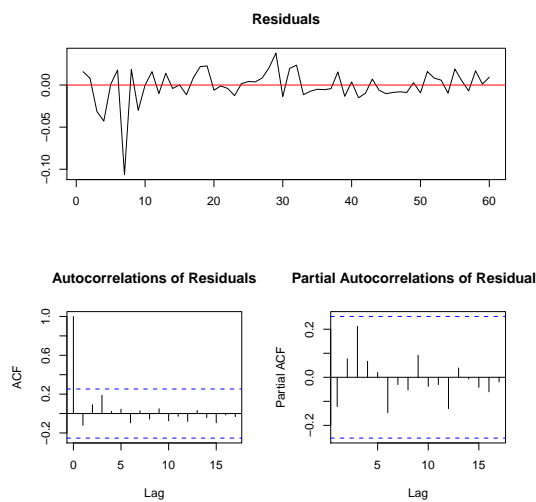
Figure 6.5 shows that the residuals of long-run relationships are stationary in mean but not in variance. This suggests that taking p_1 as the normalisation variables there is no evidence of cointegration in the $UKpppuip$ data set. The nonstationarity of the estimated residuals of the regression lines for other variables is tested using the the ADF test as follows:

```
> p2.lc=ur.df(error.p2,type='none')
> p2.lc
> e12.lc=ur.df(error.e12,type='none')
> e12.lc
> i1.lc=ur.df(error.i1,type='none')
> i1.lc
> i2.lc=ur.df(error.i2,type='none')
> i2.lc
####################################################################
# Augmented Dickey-Fuller Test Unit Root / Cointegration Test #
####################################################################
The value of the test statistic is: -2.4662
The value of the test statistic is: -2.1394
The value of the test statistic is: -3.5934
The value of the test statistic is: -4.1602
The value of the test statistic is: -4.0841
```

As extracted from Hamilton (1994) the critical values for the test are given in table 6.6.

Table 6.6: Critical values of ADF test for residuals

	1%	5%	10%
τ	-4.31	-3.77	-3.45

Table 6.7 summarises the results of the ADF test for the stationarity of residuals.

Table 6.7: Engle-Granger cointegration test

	ADF test statistic	Results
Wholesale price index (p_1)	-2.4662	Fail to reject the null hypothesis
Weighted foreign wholesale price (p_2)	-2.1394	Fail to reject the null hypothesis
Effective exchange rate (e_{12})	-3.5934	Fail to reject the null hypothesis
Treasury bill rate (i_1)	-4.1602	Fail to reject the null hypothesis
Eurodollar rate (i_2)	-4.0841	Fail to reject the null hypothesis

The ADF critical value at the 1% level of significance is -4.31. Based on Engle-Granger's two-step procedure, we cannot reject the null hypothesis of no cointegration in all five variables.

In the next step, we estimate the following the error-correction model:

$$\Delta p_1 = \beta_0 + \beta_1 \Delta p_2 + \beta_2 \Delta e_{12} + \beta_3 \Delta i_1 + \beta_4 \Delta i_2 + \alpha^* \varepsilon_{t-1} + \epsilon_t \tag{6.6.6}$$

where $\alpha^* = -(1 - \alpha_1)$ as the speed of adjustment to equilibrium, and ε_{t-1} is the error-correction mechanism. In R this is done as follows:

```
> p1.l=lag(error.p1)
> p1.d=diff(p1)
> p2.d=diff(p2)
> e12.d=diff(e12)
> i1.d=diff(i1)
> i2.d=diff(i2)
> ecm1=summary(lm(p1.d~p2.d+e12.d+i1.d+i2.d+p1.l))
Call:
lm(formula = p1.d ~ p2.d + e12.d + i1.d + i2.d + p1.l)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.004774    0.001417    3.369   0.00110 **
p2.d        0.235293    0.073202    3.214   0.00179 **
e12.d       0.027621    0.031233    0.820   0.41411
i1.d 0.02623    0.01324    0.987   0.43211
i2.d 0.07624    0.013726    0.107   0.43211
p1.1        0.07865    0.081055    1.206   0.08812
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1    1
```

```
Residual standard error: 0.01299 on 94 degrees of freedom

Multiple R-squared: 0.1242,    Adjusted R-squared: 0.09625

F-statistic: 4.444 on 3 and 94 DF,  p-value: 0.005767
```

The estimated ECM is

$$\Delta \hat{p}_{1t} = 0.0048 + 0.2353\Delta \hat{p}_2 + 0.0276\Delta \hat{e}_{12} + 0.0262\Delta \hat{i}_1 + 0.0762\Delta \hat{i}_2 + 0.0787\hat{\varepsilon}_{t-1}. \qquad (6.6.7)$$

From the above equation we observe that $\alpha^* = 0.0787$ enters with the wrong sign (positive); this is an indication that the system diverges from equilibrium with time.

6.6.1.1 Diagnostic error tests

In R, diagnostic tests are performed using the following commands:

```
>jaque.bera.test(lag(error.p1), lag=3)
>arch(lag(error.p1),lag.single=3)
>box.test(lag(errror.p1), lag=1,type="Ljung-Box")
>reset(lag(error.p1),type="regressors")
```

Table 6.8 summarises the results of the diagnostic test on residuals.

Table 6.8: Results from the diagnostic error tests

Test	Test statistic	p-value	Conclusion
Jarque-Bera	18.267	0.3169	Normally distributed
ARCH-LM	52.724	0.000	Heteroscedasticity
Ljung-Box	140.19	0.9640	No serial correlation
Ramsey Reset	34.7334	0.8754	No misspecification

The p-values in table 6.8 are compared with the 5% level of significance.

6.6.2 Phillips-Ouliaris methods

The Phillips-Ouliaris methods are implemented by using two residual-based tests.

6.6.2.1 The variance ratio test

In R, the variance ratio test is implemented in the function $ca.po$ as follows:

```
> pu.test=summary(ca.po(data1,demean='const',type='Pu'))
```

```
> pu.test

##########################################

# Phillips and Ouliaris Unit Root Test #

##########################################

Test of type Pu

detrending of series with constant only

Call:

lm(formula = z[, 1] ~ z[, -1])

Coefficients:

             Estimate Std. Error t value Pr(>|t|)

(Intercept) -2.30081     0.80040  -2.875  0.00568 **

z[, -1]p2    1.61307     0.06309  25.566  < 2e-16 ***

z[, -1]e12   0.11967     0.12278   0.975  0.33383

z[, -1]i1   -0.70801     0.46090  -1.536  0.13004

z[, -1]i2    0.31163     0.44207   0.705  0.48372

Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.06547 on 57 degrees of freedom

Multiple R-squared: 0.9847,    Adjusted R-squared: 0.9836

F-statistic: 918.2 on 4 and 57 DF,  p-value: < 2.2e-16

Value of test-statistic is: 1.2458

Critical values of Pu are:

                 10pct    5pct    1pct

critical values 45.3308 53.2502 71.5214
```

Table 6.9 summarises the results of the variance ratio test from the $UKpppuip$ data set.

Table 6.9: Results from the variance ratio test

Level of significance	10%	5%	1%
Critical value	45.3308	53.2502	71.5214
Decision	Fail to reject H_0	Fail to reject H_0	Fail to reject H_0

The test statistic for the variance ratio test $(\hat{P}_u)$ is 1.2458. From table 6.9, we observe that this calculated test statistic is smaller than the critical values extracted from R at the 10% level of significance. Therefore, the null hypothesis of no cointegration is not rejected at all levels of significance.

6.6.2.2   The multivariate trace statistic

In R, the multivariate trace statistic is implemented in the function *ca.po* as follows:

```
> pz.test=summary(ca.po(data1,demean='const',type='Pz'))
> pz.test
Value of test-statistic is: 55.1917
Critical values of Pz are:
                  10pct     5pct     1pct
critical values 168.8572 182.0749 209.8054
```

Table 6.10 summarises the results of the multivariate trace statistic from the $UKpppuip$ data set.

Table 6.10: Results of the multivariate trace statistic test

Level of significance	10%	5%	1%
Critical value	168.8572	182.0749	209.8054
Decision	Fail to reject H_0	Fail to reject H_0	Fail to reject H_0

The test statistic for the multivariate test statistic $(\hat{P}_z)$ is 55.1917. From table 6.10, we observe that this calculated test statistic is smaller than the critical values extracted from R at the 10% level of significance. Therefore, the null hypothesis of no cointegration is not rejected at all levels of significance.

6.6.3   Johansen's method

Johansen's method is implemented on two likelihood ratio tests namely:

- The trace test

- The maximum eigenvalue.

6.6.3.1   The trace test

```
> summary(ca.jo(data1,type="trace",ecdet="const"))
######################
# Johansen-Procedure #
######################
Test type: trace statistic , without linear trend and constant in cointegration
Eigenvalues (lambda):
```

```
[1]  5.214764e-01  3.304515e-01  2.932623e-01  1.667568e-01  8.128293e-02 -1.392413e-16
```

Values of teststatistic and critical values of test:

```
           test 10pct  5pct  1pct
r <= 4 |    5.09  7.52  9.24 12.97
r <= 3 |   16.03 17.85 19.96 24.60
r <= 2 |   36.86 32.00 34.91 41.07
r <= 1 |   60.93 49.65 53.12 60.16
r = 0  |  105.15 71.86 76.07 84.45
```

Eigenvectors, normalised to first column:

(These are the cointegration relations)

```
              p1.l2       p2.l2       e12.l2      i1.l2       i2.l2      constant
p1.l2     1.0000000   1.0000000   1.000000   1.0000000   1.0000000   1.0000000
p2.l2    -0.7346918  -0.9207675  -1.361829  -1.9915177  -1.4252911  -1.0358045
e12.l2   -0.9704284  -1.7829624  -2.109588   0.6873137  -0.7242021  -0.2570326
i1.l2    -2.8848474  -7.3432095  23.583651  -0.7875323   2.5025109  -0.4401157
i2.l2    -2.8286769  -0.4397012 -21.948283   1.7337995   0.8246376  -0.4631123
constant -5.1797144  -7.5542800  -7.790391   7.5913467  -1.6406824  -0.6387833
```

Weights W:

(This is the loading matrix)

```
              p1.l2         p2.l2         e12.l2         i1.l2         i2.l2      constant
p1.d  -0.063116180 -0.008089024  0.0028462619  0.0005926796 -0.011717664  1.330204e-15
p2.d  -0.087829681  0.016994213  0.0013870250  0.0329533062  0.007204453 -5.855598e-15
e12.d -0.047349896  0.089044841 -0.0003203234 -0.1010515588 -0.029084086  2.014059e-14
i1.d  -0.007442938  0.025918893 -0.0056791100  0.0131966312 -0.020372446 -2.054232e-15
i2.d   0.039061585  0.009907056  0.0132306063  0.0454082868 -0.010762483 -9.438451e-15
```

6.6.3.2 The maximum eigenvalue

```
> summary(ca.jo(data1,type="eigen",ecdet="const"))
######################
# Johansen-Procedure #
######################
 maximal eigenvalue statistic (lambda max) , without linear trend and constant in cointegration
Eigenvalues (lambda):
[1]  5.214764e-01  3.304515e-01  2.932623e-01  1.667568e-01  8.128293e-02 -1.392413e-16
```

```
Values of teststatistic and critical values of test:

         test 10pct  5pct  1pct

r <= 4 |  5.09  7.52  9.24 12.97

r <= 3 | 10.95 13.75 15.67 20.20

r <= 2 | 20.83 19.77 20.00 20.81

r <= 1 | 26.07 25.56 24.14 23.24

r = 0  | 44.22 31.66 34.40 39.79

Eigenvectors, normalised to first column:

(These are the cointegration relations)

               p1.l2       p2.l2       e12.l2       i1.l2       i2.l2      constant

p1.l2     1.0000000   1.0000000   1.000000   1.0000000   1.0000000   1.0000000

p2.l2    -0.7346918  -0.9207675  -1.361829  -1.9915177  -1.4252911  -1.0358045

e12.l2   -0.9704284  -1.7829624  -2.109588   0.6873137  -0.7242021  -0.2570326

i1.l2    -2.8848474  -7.3432095  23.583651  -0.7875323   2.5025109  -0.4401157

i2.l2    -2.8286769  -0.4397012 -21.948283   1.7337995   0.8246376  -0.4631123

constant -5.1797144  -7.5542800  -7.790391   7.5913467  -1.6406824  -0.6387833

Weights W:

(This is the loading matrix)

              p1.l2          p2.l2          e12.l2            i1.l2           i2.l2        constant

p1.d  -0.063116180 -0.008089024  0.0028462619  0.0005926796 -0.011717664  1.330204e-15

p2.d  -0.087829681  0.016994213  0.0013870250  0.0329533062  0.007204453 -5.855598e-15

e12.d -0.047349896  0.089044841 -0.0003203234 -0.1010515588 -0.029084086  2.014059e-14

i1.d  -0.007442938  0.025918893 -0.0056791100  0.0131966312 -0.020372446 -2.054232e-15

i2.d   0.039061585  0.009907056  0.0132306063  0.0454082868 -0.010762483 -9.438451e-15
```

Tables 6.11 summarises results of Johansen's cointegration method on the $UKpppuip$ data set.

Table 6.11: Johansen's trace tests and maximum eigenvalues results

Null hypothesis	Alternative	Test statistic	10%	5%	1%	Results
J_{trace} test						
$r \leq 4$	$r > 4$	5.09	7.52	9.24	12.97	Fail to reject H_0
$r \leq 3$	$r > 3$	16.03	17.85	19.96	24.60	Fail to reject H_0
$r \leq 2$	$r > 2$	36.86	32.00	34.91	41.07	Reject H_0
$r \leq 1$	$r > 1$	60.93	49.65	53.12	60.16	Reject H_0
$r = 0$	$r > 0$	105.15	71.86	76.07	84.45	Reject H_0
J_{max} test						
$r = 4$	$r = 5$	5.09	7.52	9.24	12.97	Fail to reject H_0
$r = 3$	$r = 4$	10.95	13.75	15.67	20.20	Fail to reject H_0
$r = 2$	$r = 3$	20.83	19.77	20.00	20.81	Reject H_0
$r = 1$	$r = 2$	26.07	25.56	24.14	23.24	Reject H_0
$r = 0$	$r = 1$	44.22	31.66	34.40	39.79	Reject H_0

From table 6.11, it can be observed that

(i) the null hypothesis of **no** cointegration ($r = 0$) against the alternative of the presence of one or more cointegrating vector is rejected at the 10% level of significance in both techniques (trace test and maximum eigenvalue). This implies that cointegration exists among all variables of the data set

(ii) the null hypotheses of ($r \leq 1$) and ($r \leq 2$) are rejected at the 10% level of significance. This implies that there are at most two cointegration vectors in the $UKpppuip$ data set

(iii) the results of Johansen's method further imply that multicointegrating relationships can be tested for in the $UKpppuip$ data set. However, testing for multicointegration is not part of this study.

6.7 Concluding remarks

In the $UKpppuip$ data set, seven variables are tested for cointegration, of which five showed the existence of a long-run relationship. Although the Engle-Granger and the Phillips-Ouliaris methods indicated that there is no cointegration amongst the five variables, Johansen's method indicated presence of cointegration and further showed the existence of at most two cointegration vectors.

Chapter 7

Conclusions and recommendations

In conclusion, the Engle-Granger method requires testing whether variables are integrated of the same order. This is done using the ADF unit root test. If they are integrated and of the same order, we examine the long-run equilibrium relationship by fitting the OLS estimator in order to assess whether there is a linear relationship. Using the error correction model, we test whether the residuals of the long-run relationship are stationary. If all conditions are satisfied, then the two or more variables under investigation cointegrate.

The advantage of the Engle-Granger method over the other techniques is its ease of implementation. However, its results are dependent on how the long-run equilibrium equation is specified. In some cases it might not be easy to identify which variable enters on the left as the dependent variable. It is therefore advisable to employ the cointegration tests of the Phillips-Ouliaris methods, which yield the same results irrespective of the variable which enters as the dependent variable, that is, they are invariant to normalisation.

According to the Phillips-Ouliaris methods, two residual-based tests, namely the variance ratio test and the multivariate trace statistic, are employed to test for cointegration. These tests measure the size of the residual variance from the cointegrating regression of the variables under study. If the residual variance is greater than the conditional variance, then the variables cointegrate. However, one deficiency of the Phillips-Ouliaris methods is that one can only estimate a single cointegrating relationship. Nevertheless, if one deals with more than two time series, it is possible that more than one cointegrating relationship exists which calls for the use of vector cointegration techniques like Johansen's procedure.

Johansen's method employs the trace test and maximum eigenvalues to test for the existence of one or more cointegrating vectors in the data set. This procedure is used mostly on multivariate data sets where we suspect the existence of more than one cointegrating relationship. However, it can also be used to verify the results of other cointegration techniques. This method assumes that the cointegrating vector is constant during the period of study. In reality, it is possible that the long-run relationships between the underlying variables change because of changes in technological progress and/or economic crises. In order to remedy this limitation, Gregory and Hasen (1996) have introduced tests for cointegration with one and two unknown structural break(s).

This study finds all the cointegration techniques tested not to be consistent. That is, all three methods do not lead to the same results. In the data analysis performed in chapter 4, 5, and 6, the Engle-Granger method and the Phillips-Ouliaris methods indicated no cointegration whileas the Johansen's method indicated presence of cointegration. We recommend the use of Johansen's method because it is able to detect more than one cointegrating relationship if present.

Chapter 8

Appendix

8.1 Appendix A

The Raotbl3 data

```
> Raotbl3
           lc       li       lw dd682 dd792 dd883
1966.4 10.4831 10.5821 12.9481    NA    NA    NA
1967.1 10.4893 10.5800 12.9895     0     0     0
1967.2 10.5022 10.5990 13.0115     0     0     0
1967.3 10.5240 10.6262 13.0411     0     0     0
1967.4 10.5329 10.6145 13.0357     0     0     0
1968.1 10.5586 10.6307 13.0518     0     0     0
1968.2 10.5190 10.6316 13.0839     1     0     0
1968.3 10.5381 10.6132 13.1120    -1     0     0
1968.4 10.5422 10.6141 13.1183     0     0     0
1969.1 10.5361 10.6263 13.1144     0     0     0
1969.2 10.5462 10.6366 13.1009     0     0     0
1969.3 10.5459 10.6313 13.0882     0     0     0
1969.4 10.5552 10.6324 13.0402     0     0     0
1970.1 10.5548 10.6361 13.0391     0     0     0
1970.2 10.5710 10.6795 13.0417     0     0     0
1970.3 10.5861 10.6813 13.0261     0     0     0
1970.4 10.5864 10.6801 13.0032     0     0     0
```

1971.1	10.5802	10.6533	13.0364	0	0	0
1971.2	10.6006	10.6839	13.0461	0	0	0
1971.3	10.6168	10.6889	13.0850	0	0	0
1971.4	10.6275	10.7025	13.1107	0	0	0
1972.1	10.6414	10.7212	13.1241	0	0	0
1972.2	10.6629	10.7818	13.1605	0	0	0
1972.3	10.6758	10.7641	13.1748	0	0	0
1972.4	10.6881	10.7841	13.1612	0	0	0
1973.1	10.7240	10.8045	13.1050	0	0	0
1973.2	10.7143	10.8230	13.1082	0	0	0
1973.3	10.7222	10.8319	13.1059	0	0	0
1973.4	10.7156	10.8380	13.0140	0	0	0
1974.1	10.6964	10.8097	12.9301	0	0	0
1974.2	10.6990	10.7928	12.8427	0	0	0
1974.3	10.7081	10.8310	12.7710	0	0	0
1974.4	10.7142	10.8328	12.7281	0	0	0
1975.1	10.7078	10.8527	12.7692	0	0	0
1975.2	10.7073	10.8089	12.7492	0	0	0
1975.3	10.6954	10.8202	12.7664	0	0	0
1975.4	10.6910	10.8069	12.7554	0	0	0
1976.1	10.6967	10.8196	12.7605	0	0	0
1976.2	10.7015	10.8046	12.7471	0	0	0
1976.3	10.7083	10.8372	12.7238	0	0	0
1976.4	10.7127	10.8123	12.7156	0	0	0
1977.1	10.6922	10.7842	12.7555	0	0	0
1977.2	10.6874	10.7713	12.7517	0	0	0
1977.3	10.6989	10.7904	12.8018	0	0	0
1977.4	10.7224	10.8369	12.8388	0	0	0
1978.1	10.7452	10.8333	12.8438	0	0	0
1978.2	10.7462	10.8635	12.8540	0	0	0
1978.3	10.7663	10.8884	12.8618	0	0	0
1978.4	10.7633	10.8924	12.8491	0	0	0
1979.1	10.7737	10.9017	12.9232	0	0	0
1979.2	10.8282	10.9108	12.9022	0	1	0

1979.3	10.7872	10.9166	12.8737	0	-1	0
1979.4	10.8015	10.9673	12.8467	0	0	0
1980.1	10.8139	10.9324	12.8647	0	0	0
1980.2	10.7909	10.9344	12.8885	0	0	0
1980.3	10.8029	10.9506	12.9183	0	0	0
1980.4	10.7868	10.9465	12.9277	0	0	0
1981.1	10.7979	10.9488	12.9505	0	0	0
1981.2	10.8007	10.9294	12.9615	0	0	0
1981.3	10.8008	10.9248	12.9147	0	0	0
1981.4	10.7991	10.9326	12.9527	0	0	0
1982.1	10.7956	10.9202	12.9641	0	0	0
1982.2	10.8005	10.9373	12.9780	0	0	0
1982.3	10.8160	10.9269	13.0299	0	0	0
1982.4	10.8260	10.9315	13.0604	0	0	0
1983.1	10.8405	10.9399	13.1031	0	0	0
1983.2	10.8482	10.9599	13.1577	0	0	0
1983.3	10.8633	10.9563	13.1504	0	0	0
1983.4	10.8633	10.9637	13.1805	0	0	0
1984.1	10.8615	10.9703	13.2245	0	0	0
1984.2	10.8732	10.9778	13.1852	0	0	0
1984.3	10.8649	10.9801	13.2298	0	0	0
1984.4	10.8793	10.9942	13.2849	0	0	0
1985.1	10.8909	10.9840	13.2999	0	0	0
1985.2	10.8938	11.0120	13.2904	0	0	0
1985.3	10.9116	11.0120	13.3140	0	0	0
1985.4	10.9202	11.0237	13.3606	0	0	0
1986.1	10.9409	11.0300	13.4574	0	0	0
1986.2	10.9663	11.0624	13.4655	0	0	0
1986.3	10.9700	11.0556	13.4371	0	0	0
1986.4	10.9808	11.0644	13.5020	0	0	0
1987.1	10.9878	11.0618	13.5914	0	0	0
1987.2	11.0048	11.0839	13.6804	0	0	0
1987.3	11.0272	11.0944	13.7131	0	0	0
1987.4	11.0420	11.1095	13.5633	0	0	0

```
1988.1 11.0701 11.1116 13.5814     0     0     0
1988.2 11.0751 11.1413 13.6171     0     0     0
1988.3 11.0964 11.1507 13.6201     0     0     1
1988.4 11.1069 11.1680 13.6460     0     0     1
1989.1 11.1123 11.1713 13.6731     0     0     1
1989.2 11.1231 11.2032 13.6884     0     0     1
1989.3 11.1223 11.2009 13.7211     0     0     1
1989.4 11.1303 11.2064 13.7686     0     0     1
1990.1 11.1307 11.2160 13.6833     0     0     1
1990.2 11.1389 11.2147 13.7130     0     0     1
1990.3 11.1325 11.2286 13.6225     0     0     1
1990.4 11.1261 11.2352 13.6957     0     0     1
1991.1 11.1232 11.2189 13.7723     0     0     1
1991.2 11.1220 11.2276 13.7424     0     0     1
```

8.2   Appendix B

Australia's Economic Indicators

```
 > data1
   Quarter    ppi      dppi      cpi
1   1999Q1 4.386517 4.367547 4.403054
2   1999Q2 4.389126 4.377516 4.407207
3   1999Q3 4.409763 4.392224 4.416186
4   1999Q4 4.423289 4.400971 4.421848
5   2000Q1 4.440767 4.417273 4.430579
6   2000Q2 4.460260 4.430579 4.438643
7   2000Q3 4.479494 4.434263 4.475175
8   2000Q4 4.503802 4.440885 4.478245
9   2001Q1 4.491329 4.441827 4.488861
10  2001Q2 4.514479 4.451902 4.497028
11  2001Q3 4.503027 4.454696 4.500032
12  2001Q4 4.496805 4.458756 4.508990
13  2002Q1 4.496025 4.466598 4.517759
14  2002Q2 4.503802 4.475631 4.525044
```

```
15  2002Q3 4.501475 4.480967 4.531631

16  2002Q4 4.512945 4.489872 4.538817

17  2003Q1 4.525152 4.504797 4.551558

18  2003Q2 4.505350 4.509980 4.551558

19  2003Q3 4.500698 4.522984 4.557240

20  2003Q4 4.502251 4.530662 4.562158

21  2004Q1 4.513713 4.544995 4.571200

22  2004Q2 4.532707 4.553350 4.576050

23  2004Q3 4.560173 4.567364 4.580160

24  2004Q4 4.581082 4.584457 4.587719

25  2005Q1 4.571820 4.588431 4.594514

26  2005Q2 4.599554 4.598045 4.600660

27  2005Q3 4.620453 4.612245 4.610058

28  2005Q4 4.628007 4.621634 4.615319

29  2006Q1 4.647559 4.628594 4.623992

30  2006Q2 4.694005 4.646120 4.639572

31  2006Q3 4.696564 4.658901 4.648613

32  2006Q4 4.687027 4.664194 4.647367

33  2007Q1 4.684443 4.665701 4.648038

34  2007Q2 4.708629 4.680370 4.660132

35  2007Q3 4.704835 4.695651 4.667112

36  2007Q4 4.720550 4.705016 4.676560

37  2008Q1 4.751173 4.724818 4.689511

38  2008Q2 4.792065 4.738739 4.704201

39  2008Q3 4.808111 4.757891 4.715727

40  2008Q4 4.785072 4.752383 4.712678

41  2009Q1 4.740924 4.742146 4.713935

42  2009Q2 4.725528 4.742146 4.718677

43  2009Q3 4.732948 4.751778 4.728272

44  2009Q4 4.714294 4.754452 4.733563
```

8.3   Appendix C

United Kingdom purchasing power parity

```
> UKpppuip
```

	p1	p2	e12	i1	i2	doilp0	doilp1
1	3.399837	3.846749	-4.899593	0.04707441	0.05059805	0.000000000	0.000000000
2	3.412952	3.856261	-4.894152	0.04678816	0.04859967	0.006686741	0.000000000
3	3.430709	3.864228	-4.832260	0.05987140	0.05543471	0.000000000	0.006686741
4	3.452861	3.881285	-4.803663	0.07250669	0.05808021	0.002213712	0.000000000
5	3.465303	3.913175	-4.785582	0.07890360	0.07622003	0.102026935	0.002213712
6	3.469929	3.953242	-4.785582	0.07148310	0.08296152	0.132782319	0.102026935
7	3.503215	3.997386	-4.723762	0.10651984	0.10589038	0.139537134	0.132782319
8	3.542608	4.020950	-4.712609	0.11520193	0.09485553	0.236910345	0.139537134
9	3.615693	4.130217	-4.719131	0.11609272	0.08773611	0.965538339	0.236910345
10	3.683799	4.172550	-4.733877	0.11010922	0.11653782	0.043433914	0.965538339
11	3.722781	4.207048	-4.730211	0.10849560	0.12425096	0.019527335	0.043433914
12	3.772879	4.236469	-4.713544	0.10669962	0.09776172	0.055380463	0.019527335
13	3.835640	4.253006	-4.694693	0.09603719	0.06989913	0.006386841	0.055380463
14	3.885599	4.270279	-4.662791	0.09175843	0.06335032	0.000000000	0.006386841
15	3.920515	4.271766	-4.619633	0.10174398	0.07194850	0.000000000	0.000000000
16	3.956147	4.285636	-4.594710	0.10723875	0.06372569	0.095711863	0.000000000
17	3.998746	4.295829	-4.583076	0.08590243	0.05382509	0.000000000	0.095711863
18	4.036117	4.320907	-4.497610	0.10381933	0.05874050	0.000000000	0.000000000
19	4.076345	4.340277	-4.471787	0.11082556	0.05581306	0.000000000	0.000000000
20	4.125455	4.358491	-4.397771	0.13540464	0.05097824	0.000000000	0.000000000
21	4.185891	4.362472	-4.431758	0.10354888	0.05116829	0.049157883	0.000000000
22	4.234323	4.374696	-4.429279	0.07325046	0.05600219	0.000000000	0.049157883
23	4.263965	4.389198	-4.431758	0.06241126	0.06288090	0.049220550	0.000000000
24	4.277778	4.401695	-4.453790	0.05656935	0.06859279	0.000000000	0.049220550
25	4.283250	4.419375	-4.480076	0.05808021	0.07204155	0.000000000	0.000000000
26	4.302169	4.441176	-4.420557	0.08093462	0.07482914	0.000000000	0.000000000
27	4.321394	4.454597	-4.429279	0.08846865	0.08599419	0.000000000	0.000000000
28	4.340257	4.468155	-4.429279	0.10849560	0.11001965	0.000000000	0.000000000
29	4.370716	4.488001	-4.451366	0.11582557	0.10047862	0.061622239	0.000000000

```
30 4.415330 4.505689 -4.505689 0.11662681 0.10138261  0.207787798  0.061622239

31 4.469429 4.518725 -4.553932 0.12944807 0.11448873  0.183289180  0.207787798

32 4.502292 4.541496 -4.522783 0.14591687 0.13749851  0.156218845  0.183289180

33 4.561852 4.575517 -4.572380 0.15417925 0.15794348  0.188576302  0.156218845

34 4.597851 4.597449 -4.588381 0.15392208 0.10571045  0.056546043  0.188576302

35 4.623419 4.616333 -4.611395 0.13993584 0.11404272  0.046682730  0.056546043

36 4.637637 4.629624 -4.646949 0.13015068 0.15785808  0.027065257  0.046682730

37 4.652340 4.648700 -4.662791 0.11573650 0.15212005  0.063683392  0.027065257

38 4.671239 4.666700 -4.622706 0.11216730 0.16101280 -0.005968468  0.063683392

39 4.684259 4.686902 -4.546235 0.13374387 0.16923632 -0.007273117 -0.005968468

40 4.700753 4.703789 -4.536252 0.14349410 0.12927234  0.002852345 -0.007273117

41 4.725350 4.718869 -4.552836 0.12804135 0.14002278 -0.013059408  0.002852345

42 4.740313 4.726432 -4.542918 0.12257155 0.13967498 -0.034575983 -0.013059408

43 4.750136 4.739622 -4.556120 0.10165365 0.11350724 -0.002991714 -0.034575983

44 4.759607 4.753597 -4.530662 0.09303487 0.09267033  0.009277172 -0.002991714

45 4.774913 4.759102 -4.428038 0.10399959 0.08883471 -0.047274182  0.009277172

46 4.792479 4.769929 -4.474163 0.09339926 0.08984070 -0.091113230 -0.047274182

47 4.803201 4.779921 -4.481255 0.09029764 0.09521927  0.000000000 -0.091113230

48 4.814620 4.789540 -4.461028 0.08654460 0.09385458  0.000000000  0.000000000

49 4.830711 4.801698 -4.442835 0.08553529 0.09630968  0.000000000  0.000000000

50 4.846547 4.811212 -4.419304 0.08470873 0.10804690  0.000000000  0.000000000

51 4.856707 4.817162 -4.396546 0.09794307 0.11207790 -0.010630026  0.000000000

52 4.869839 4.824625 -4.358630 0.09294375 0.09358141 -0.010766150 -0.010630026

53 4.892602 4.832971 -4.317755 0.11297148 0.08562709 -0.014522756 -0.010766150

54 4.908972 4.839915 -4.407938 0.11724956 0.07982730 -0.011065893 -0.014522756

55 4.919251 4.836898 -4.447697 0.11064652 0.07723878 -0.030087322 -0.011065893

56 4.927254 4.840464 -4.419322 0.10795714 0.07760898  0.026371311 -0.030087322

57 4.941642 4.836249 -4.359270 0.11154137 0.07473635 -0.242824909  0.026371311

58 4.950885 4.830666 -4.371976 0.09576462 0.06775210 -0.528430180 -0.242824909

59 4.958640 4.834564 -4.314818 0.09248802 0.06043637 -0.140982497 -0.528430180

60 4.966335 4.832189 -4.262680 0.10138261 0.05902334  0.161746159 -0.140982497

61 4.982236 4.840834 -4.286891 0.09939274 0.06165938  0.250437899  0.161746159

62 4.994506 4.841025 -4.326646 0.08746128 0.06765865  0.065340376  0.250437899
```

Bibliography

Asufu-Adjaye D., 2000 The relationship between energy consumption, energy prices and economic growth: time series evidence from Asian developing countries. *Energy economics,* 22:615−625.

Banerejee A, Galbabraith W, Dolado J.J. and Hendry D.F., 1993. *Cointegration, error-correction, and ecometric analysis of nonstationary data.* London: Oxford University Press.

Belloumi M., 2009. Energy consumption and GDP in Tunisia: Cointegration and causality analysis. *Energy policy,* 37:2745−2753.

Bentzen J., 1995. An empirical of gasoline in Denmark using cointegration approach. *Energy Economics,* 17:329−339.

Bhaskara R., 1994. *Cointegration for applied economist.* London:Springer + Business Media.

Cancer M., 1998 Tests for cointegration with infinite variance errors. *Econometrics,* 86:155−175.

Chan H.L and Lee S.K., 1997. Modeling and forecasting and demand for China. *Energy Economics* 19:149−168.

Cheng B.S. and Lai T.W., 1997.An investigation of cointegration and causality between energy consumption and economic activity in Taiwan. *Energy Economics,* 19:435−444.

Choi I., 1994 Spurious regression and residual based tests for cointegration when regressors are cointegrated. *Econometrics,* 60:331−320.

Enders W., 2004. *Applied econometrics time series.* Wiley series in Probability and Statistics.

Engle R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica,* 50:987−1007.

Engle R.F. and Granger C.W.J., 1987. Cointegration and error correction: Representation, estimation and testing. *Econometrica,* 55:251−276.

Erdogdu E., 2007. Electricity demand analysis using cointegration ARIMA modeling: A case study of Turkey. *Energy Policy,* 35:1129−1146.

Fouquest R, Pearson A, Hawdon P.D. Robinson C. and Stevens P., 1997. The future of UK final user energy demand. *Energy Policy,* 25:231−240.

Gregory A.W. and Hasen J.M., 1996. Testing for structural breaks in cointegrated relationships. *Econometrics,* 71:321−341.

Hamilton J.D., 1994. *Time Series Analysis.* New Jersey:Princeton University Press.

Harris R., 1995. *Using cointegration analysis in econometric modelling.* London:Oxford university press.

Holden D. and Perman R., 1994. *Unit roots and cointegration for the economist.* London:Oxford university press.

Jarque C.M. and Bera A.K., 1980. Efficient tests for normality, homoskedasticity and serial independence of regression residuals. *Economics letters,* 6:255−259.

Johansen R. and Juselius K., 1998. Testing structural hypothesis in multivariate cointegration of the PPP and the UIP for UK. *Econometrics,* 53:211−244.

Johansen S., 1988. Statistical analysis of cointegration vectors. *Economic Dynamic control,* 12:231−254.

Kanas A., 1997. Is economic exposure asymmetric between long-run depreciations and appreciation? Testing using cointegration analysis. *Multinational Functional Management,* 7:27−42.

Khalifa H. and Sakka M., 2004. Energy use and output in Canada: a multivariate cointegration analysis. *Energy Economics,* 25:225−238.

Kulshreshtha M. and Parikh J.K., 1999. Modeling demand coal India: vector autoregressive models with cointegration variables. *Energy Economics* 26:149−168.

Lee H., 1993. Seasonal cointegration:− The Japanese consumption function. *Econometrics,* 55:275−298.

Ljung G.M. and Box G.E.P., 1978. On a measure of lack of fit in time series models. *Biometrika,* 65:297−303.

Masih A. and Mashi R., 1996. Energy consumption and real income temporal causality, results for a multi-country study based on cointegration and error correction techniques. *Energy Economics,* 1:165−183.

Maslyuk S. and Smyth R., 2009. Cointegration between oil spot and future prices of the same and different grades in the presence of structural change. *Economics and Business,* 65:1−63.

Maslyuk S. and Smyth R., 2010. Female labor force participation and total fertility rates in the OECD: New evidence from panel coitegration and Granger causality testing *Economics and Business,* 65:48−64.

Pauly P., 2003. Hypothesis test. Lecture notes for econometrics 815, taught at the University of Pretoria in March 2003.

Pfaff B., 2006. *Analysis of integrated and cointegrated time series with R.* London:Springer + Business Media.

Phillips P.C.B. and Oualiaris S., 1998. Testing for cointegration using principal component methods. *Economics Dynamic and Control,* 12:205−230.

Ramsey J.B., 1978. Tests for specification errors in classical linear least squares regression analysis. *Royal statistical society,* 31:350−371.

Samimi R., 1995. Road transport energy demand in Austaralia: a cointegration approach. *Energy Economics* 17:349−339.

United Nations World Economic Indicators *http://quanis1.easydata.co.za/TableViewer/tableView.aspx.*

Wei W.S., 2006. *Time series analysis: univariate and multivariate.* Boston: Pearson.

Yang H., 2000. A note on casual relationship between energy and GNP in Taiwan. *Energy Economics,* 22:309−317.

Yule G.U., 1989. Why do we sometimes get nonsense-corrections between time series?−a study in sampling and the nature of time series. *Royal Statistical Society,* 1:1−63.