Testing for Purchasing Power Parity and Uncovered Interest parity in the Presence of Monetary and Exchange Rate Regime Shifts

Miguel Lacerda[1], Johannes Fedderke[2] and Linda Haines[3]

Occasional Paper Number 01

[1] School of Economics, University of Cape Town
[2] School of Economics, University of Cape Town
[3] School of Economics, University of Cape Town

Testing for Purchasing Power Parity and Uncovered Interest Parity in the Presence of Monetary and Exchange Rate Regime Shifts

Miguel Lacerda, Johannes Fedderke and Linda Haines

Abstract

Testing for purchasing power parity (PPP) and uncovered interest parity (UIP) has been the focus of many empirically oriented studies. Whilst these simple economic theories of exchange rate and interest rate determination are theoretically attractive, the empirical support for these equilibrium conditions is at best mixed. Many potential reasons have been cited in the literature for the failure of such studies, ranging from market imperfections to inappropriate modelling strategies. The current state-of-the-art procedure involves testing for two cointegrating vectors in a multivariate error correction model which may be economically identified as the PPP and UIP relations. However, such a procedure does not account for the policy regime shifts which often characterise economic time series. It is proposed that such regime shifts distort the underlying PPP and UIP relations, making them difficult to detect when modelled within this conventional framework. In this paper, a Markov-switching vector error correction model (VECM) is considered for time series data in which monetary and exchange rate regime shifts are known *a priori* to be present. Weak evidence in favour of PPP and UIP is established in a standard linear VECM, although the residuals of this model clearly indicate that it is inappropriate in terms of functional form. The Markov-switching VECM, however, provides convincing evidence in favour of both the PPP and UIP relations and a marked improvement in the residual distributions. An enlightening by-product of this method is the data-based estimation of regimes which appear to conform to those suggested by economic history.

KEY WORDS: Purchasing power parity; Uncovered interest parity; Markov-switching vector error correction model; Multivariate cointegration

Contents

List of Figures

List of Tables

1 Introduction

1.1 Problem Description

Purchasing power parity (PPP) was first pondered by scholars in the sixteenth century and is the surprisingly simple proposition that, once converted to a common currency, national price levels should be equal [47]. The basic idea stems from the law of one price which states that in competitive markets free of transactions costs and trade barriers, arbitrage in the goods market will ensure that identical goods are sold for the same price in different countries when their prices are expressed in terms of the same currency [24]. Of course, if goods market arbitrage enforces broad parity in prices across a sufficiently large range of individual goods, then there should also be a high correlation in aggregate price levels. This result is exactly what is posited by the theory of PPP [47].

Meanwhile, in the capital market, a similar arbitrage condition is also thought to prevail. The uncovered interest parity (UIP) requires that the expected returns on deposits in any two currencies should be equal in the absence of transactions costs, risk premia and speculative effects when measured in a common currency [16]. If expected returns do differ between countries, rational, profit-seeking investors will choose to invest in the country with the highest expected return on deposits *ceteris paribus*. The resulting capital flows will put pressure on the bilateral exchange rate and on the prices of the two financial instruments to adjust in such a way as to eliminate the arbitrage opportunity in the asset market.

The two equilibrium conditions mentioned above do not, however, prevail independently. Indeed, one might expect interactions in the determination of exchange rates, prices and interest rates in the goods and asset markets [30]. Interest rates, adjusting according to UIP, affect capital flows and the real demand for liquidity. The strength and direction of capital flows influence the demand for currencies and hence induce movements in the exchange rate. However, a change in the exchange rate would require a counterbalancing movement in national price levels in order for PPP to hold. On the other hand, changes in the real demand for liquidity will affect price levels directly and should therefore translate into a movement in the exchange rate in accordance with PPP [43]. The PPP and UIP relations are therefore intrinsically interwoven.

Whilst UIP may be regarded as a short-term market clearing mechanism under the efficient market hypothesis, few empirically literate economists take PPP seriously as a short-term proposition. Instead, most instinctively believe in some variant of purchasing power parity as a long-run determinant of the exchange rate [47]. This paper will attempt to establish whether the economic theories of PPP and UIP are supported by firm empirical evidence when South Africa and the United States are considered as trading partners.

1.2 Background to the Investigation

Whilst the PPP and UIP conditions are theoretically attractive, the empirical support for these relations is at best mixed. Slight variations in data definitions, time periods and estimation methods are known to have substantive effects on results. The search for methods which are robust to such variations has led to an enormous and ever-growing empirical literature on the topic. The fact that so much research has been reported on this subject indicates, to some extent, the reluctance of researchers to part with the commonsense notions of PPP and UIP [43].

The difficulties faced in ascertaining firm empirical evidence in favour of these theories should come as no surprise. The very definitions of PPP and UIP given above are riddled with caveats such as "assuming competitive markets free of transactions costs and trade barriers" and "in the absence of transactions costs, risk premia and speculative effects." Clearly, these assumptions are not consistent with the real world and it is the violations of these assumptions which are often cited as the cause of failure of many empirically oriented studies.

Johansen and Juselius attribute the failure of many of the earlier studies of PPP and UIP in which these relations are treated as disjoint to the neglect of the important interactions between the goods and asset markets alluded to earlier [30]. Instead, the authors propose modelling both relations simultaneously. Although such a multivariate approach seems far more convincing, the results unfortunately have not been. Some authors find evidence for PPP and UIP, whilst others do not.

What might be concluded from the literature as a whole is that if PPP does in fact hold as a long-run relation, the speed of adjustment back to equilibrium is very slow, making it difficult to establish this condition empirically [47]. Additionally, most empirical work finds evidence in support of UIP as a *long-run* relation, rather than an immediate market clearing mechanism as might be expected in efficient markets [30]. As such, the empirical evidence often does not support the relevant economic theory in its strictest form.

1.3 Purpose of the Research

This paper will consider an alternative method for assessing the evidence in favour of PPP and UIP. The current state-of-the-art approach which tests for these relations within a multivariate error correction framework will be adapted to explicitly allow for regime shifts over the time period considered. Indeed, the time period considered in this study is characterised by substantial changes in monetary and exchange rate policy. The manner in which the South African Reserve Bank provided accommodation to commercial banks and achieved monetary targets has certainly changed substantially throughout the twentieth century. Exchange rate policy too has evolved considerably from the Bretton Woods fixed exchange rate system introduced in the wake of the second World War to the present-day floating exchange rate. Moreover, intervention by the Reserve Bank to stabilise the real exchange rate has diminished substantially over the years. It is proposed that these shifts in monetary and exchange rate policies cloud the underlying PPP and UIP relations, making them difficult to detect when modelled within the conventional multivariate framework. Instead, it is suggested that PPP and UIP hold *conditional on* the underlying regimes, but not necessarily across regimes. In this sense, the PPP and UIP relations considered here are of a weaker form than those implied by the formal definitions given above.

The underlying regimes will be estimated from the data together with the parameters of the multivariate error correction model. These estimated regimes will then be assessed against the monetary and exchange rate regimes in South Africa as suggested by historical evidence. The economic and political developments that typify the regimes should provide further insight into why it is that PPP and UIP are difficult to demonstrate over the entire period. The parameters obtained in this model will also be compared to those obtained by following the standard approach in order to determine whether there are any important differences in the adjustment mechanisms implied by these models.

1.4 Layout of the Paper

This paper is laid out as follows. Section 2 presents the theoretical concepts necessary to understand the modelling strategy to follow. An overview of the relevant properties of stochastic processes is provided, followed by a description of the multivariate models employed. Parameter estimation in these models, although interesting, is relegated to the appendices. The section is concluded with a more detailed review of the economic theories of purchasing power parity and uncovered interest parity. The dataset employed in this study is discussed in Section 3, paying particular attention to the economic and political developments that characterise the studied time period. The methodological approach adopted in this paper is also outlined in this section. The empirical findings are presented in Section 4. Finally, conclusions and recommendations for further research are given in Section 5.

2 Theoretical Background

This section covers the necessary statistical and economic theory required for testing the purchasing power and uncovered interest parities empirically. It commences by considering some of the basic properties of stochastic processes that are relevant in the time series context. The vector autoregressive model is introduced together with its vector error correction representation appropriate for modelling non-stationary processes. These models are then adapted to allow for hidden regimes which are assumed to evolve according to a first-order Markov process. Finally, the various forms of PPP and UIP are discussed in the context of the aforementioned statistical methods for testing their data admissibility.

2.1 Basic Notions

2.1.1 Stochastic Processes

Consider the probability space $(\Omega, \mathcal{F}, P)$ where Ω is the sample space, $\mathcal{F}$ is the event space or σ-algebra consisting of subsets of Ω and P is a probability measure which maps $\mathcal{F} \mapsto [0, 1]$. Then the collection of random variables $\{Y_t : t \in \mathbb{Z}\}$ defined relative to the probability space $(\Omega, \mathcal{F}, P)$ is referred to as a discrete-time *stochastic process*, where $\mathbb{Z}$ is the set of all real integers [23]. Note that whilst it is customary to denote random variables with capital letters and their realisations with lower case letters, the notation in this paper will in general not distinguish between the two. The distinction should, however, be clear from the context.

A special stochastic process of relevance in time series analysis is *white noise*. Let $\{u_t\}$ be a zero mean, finite variance stochastic process, that is $\mathrm{E}[u_t] = 0$ and $\mathrm{Var}[u_t] < \infty$. Then $\{u_t\}$ is referred to as *weak* white noise if $\mathrm{Cov}[u_t, u_s] = 0$ for $t \neq s$. If, in addition, the stochastic process $\{u_t\}$ is also homoscedastic, that is $\mathrm{Var}[u_t] = \sigma_u^2$ for all t and some finite constant σ_u^2, then the process is referred to as *strong* white noise. Finally, if the random variables in a stochastic process are identically and independently distributed, denoted as i.i.d, then this constitutes a *strict* white noise process since independence is a stricter condition than the zero correlation implied by $\mathrm{Cov}[u_t, u_s] = 0$ [23].

2.1.2 Autocorrelation

If $\mathrm{Cov}[u_t, u_s] \neq 0$, then the series $\{u_t\}$ is said to be *autocorrelated* or *serially correlated*. Autocorrelation introduces time dependence into the stochastic process in that the realisation at a given point in time is dependent upon the previously realised values [23]. A classic example of a stochastic process which demonstrates this property is the autoregressive process

$$y_t = \phi y_{t-1} + \varepsilon_t, \tag{2.1}$$

with autoregressive coefficient ϕ and strict white noise innovations ε_t; that is, ε_t is an i.i.d random variable with zero mean and variance σ^2, written more compactly as $\varepsilon_t \overset{iid}{\sim} \mathrm{rv}\,(0, \sigma^2)$. More specifically, this process is known as a first-order autoregressive process, denoted as AR(1), and may be easily generalised to the pth order. An AR(p) process has the form

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t \tag{2.2}$$

with $\{\varepsilon_t\}$ representing a strict white noise process as before and $\phi_1, \ldots, \phi_p$ representing the partial autocorrelation coefficients. These autocorrelation coefficients are *partial* in the sense that they measure the correlation between the current and lagged values after removing the predictive power of all the values of the series with smaller lags. Clearly, an AR(p) process would be expected to have zero partial autocorrelations for lags greater than p [49].

It is also useful at this point to introduce the concept of a lag operator $\mathbb{L}$, which shifts variables back in time such that $\mathbb{L}^i y_t = y_{t-i}$. Using the lag operator, the AR(p) process in Equation (2.2) has the equivalent representation

$$(1 - \phi_1 \mathbb{L} - \phi_2 \mathbb{L}^2 - \ldots - \phi_p \mathbb{L}^p) y_t = \varepsilon_t.$$

The pth order polynomial $p(\mathbb{L}) = 1 - \sum_{i=1}^{p} \phi_i \mathbb{L}^i$ is referred to as the (reverse) *characteristic polynomial* of the AR(p) process [23].

2.1.3 Stationarity

The notion of stationarity is fundamental to the analysis of all time series processes. In a univariate context, *strict stationarity* implies that the joint distribution of $\{y_{t_1}, y_{t_2}, \ldots, y_{t_k}\}$ is identical to that of $\{y_{t_1-h}, y_{t_2-h}, \ldots, y_{t_k-h}\}$ for any collection of time points $t_1, t_2, \ldots, t_k$, for any number $k = 1, 2, \ldots,$

and for any shift $h = 0, \pm 1, \pm 2, \ldots$ [49]. The extension to a multivariate time series is analogous, replacing the single observation y_t with a vector of observations $\boldsymbol{y}_t$ drawn randomly from a multivariate distribution at time t. Since strict stationarity is difficult to show in practice, one must typically settle for a weaker form thereof. More specifically, a stochastic process $\{y_t : t = 1, 2, \ldots\}$ is said to be *weakly stationary* if the joint distribution of $\{y_{t_1}, y_{t_2}, \ldots, y_{t_k}\}$ is identical to that of $\{y_{t_1-h}, y_{t_2-h}, \ldots, y_{t_k-h}\}$ with respect to their first two product moments and these two moments are finite; that is, the mean and variance are time invariant and correlations depend only on the time lag between observations, but not on time itself [23].

As an illustration, consider the AR(1) process in Equation (2.1). By recursive substitution, this process may be written as

$$y_t = \phi^n y_{t-n} + \sum_{i=0}^{n-1} \phi^i \varepsilon_{t-i}.$$

Recalling that $\{\varepsilon_t : t = 1, 2, \ldots\}$ is a sequence of identically and independently distributed random variables with $\mathrm{E}[\varepsilon_t] = 0$ and $\mathrm{Var}[\varepsilon_t] = \sigma^2$ for all t, it therefore follows that

$$\mathrm{E}[y_t] = \phi^n \mathrm{E}[y_{t-n}] \tag{2.3}$$

and

$$\mathrm{Var}[y_t] = \phi^{2n} \mathrm{Var}[y_{t-n}] + \sigma^2 \sum_{i=0}^{n-1} \phi^{2i}. \tag{2.4}$$

Now consider what happens to these two moments under each of the following three conditions as $n \to \infty$.

(a) Explosive Autoregressive Coefficient $|\phi| > 1$

If the autoregressive coefficient ϕ is greater than 1, both the mean and variance will grow exponentially as n increases. Consequently, these two moments are nonconstant over time and therefore the process cannot be regarded as stationary even in the weak sense. On the other hand, an autoregressive coefficient ϕ less than -1 will produce an oscillating time series. This condition implies that the absolute value of the mean will grow exponentially as n increases, as will the variance. Clearly, such a process will also be non-stationary. In light of the implications for the mean and variance of an autoregressive process with $|\phi| > 1$, such processes are said to be *explosive* and are generally not sustainable in the long-run [23].

(b) Unitary Autoregressive Coefficient $|\phi| = 1$

A special case of an AR(1) process, known as the pure *random walk*, is obtained when $\phi = 1$ and may be expressed as

$$y_t = y_{t-1} + \varepsilon_t, \tag{2.5}$$

or equivalently,

$$(1 - \mathbb{L})y_t = \varepsilon_t.$$

Clearly, the characteristic polynomial $p(\mathbb{L}) = (1 - \mathbb{L})$ has a *unit root* such that $p(1) = 0$.

From Equations (2.3) and (2.4), the first two moments of a pure random walk process which started in the infinite past may be established as

$$\lim_{n\to\infty} \mathrm{E}[y_t] \to \lim_{n\to\infty} \mathrm{E}[y_{t-n}]$$

and

$$\lim_{n\to\infty} \mathrm{Var}[y_t] = \lim_{n\to\infty} \mathrm{Var}[y_{t-n}] + n\sigma^2 \to \infty.$$

Hence, whilst the pure random walk is indeed stationary in mean, its variance accumulates over time and it is therefore non-stationary. An extension of this model is the *random walk with drift*

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

which allows for a stochastic trend through the drift term ν. By successive substitution, the random walk with drift may be rewritten as

$$y_t = n\nu + y_{t-n} + \sum_{i=0}^{n-1} \varepsilon_{t-i}$$

and is therefore neither stationary in mean nor variance in the limit as $n \to \infty$. Consequently, it is concluded that unit root processes are non-stationary [38].

Although extremely rare in economic applications, an autoregressive process with $\phi = -1$ is also non-stationary. From Equation (2.3), it follows that such a process will exhibit an oscillating mean, although the mean will be constant in absolute value. In addition, the variance of such a process will accumulate over time as is the case with a pure random walk process. Such a process would oscillate indefinitely without ever converging to a stable equilibrium and it would therefore seem unsurprising that an autoregressive process with $\phi = -1$ is seldom encountered in practice.

(c) Stable Autoregressive Coefficient $|\phi| < 1$

Suppose the AR(1) process given by Equation (2.1) has an autoregressive coefficient $|\phi| < 1$. Then, assuming this process began in the infinite past, it is easy to verify that

$$\lim_{n\to\infty} \mathrm{E}[y_t] = \lim_{n\to\infty} \phi^n\, \mathrm{E}[y_{t-n}] \to 0$$
$$\lim_{n\to\infty} \mathrm{Var}[y_t] = \lim_{n\to\infty} \phi^{2n}\mathrm{Var}[y_{t-n}] + \frac{\sigma^2(1-\phi^{2n})}{1-\phi^2} \to \frac{\sigma^2}{1-\phi^2}$$

provided $\lim_{n\to\infty} \mathrm{E}[y_{t-n}] < \infty$ and $\lim_{n\to\infty} \mathrm{Var}[y_{t-n}] < \infty$. In this case, both the mean and variance are time invariant and finite. Furthermore, the covariance between observations is also constant over time, which can be shown by multiplying Equation (2.1) through by y_t and taking expectations to yield

$$\mathrm{E}[y_t y_{t-1}] = \phi \mathrm{E}[y_{t-1}^2] + \mathrm{E}[\varepsilon_t y_{t-1}]$$
$$\mathrm{Cov}[y_t, y_{t-1}] = \frac{\phi\sigma^2}{1-\phi^2}$$

since $\mathrm{E}[y_{t-1}] = 0$ and $\mathrm{Cov}[\varepsilon_t, y_{t-1}] = 0$. The general case follows by induction with

$$\mathrm{Cov}[y_t, y_{t-n}] = \frac{\phi^n\sigma^2}{1-\phi^2},$$

which is also time invariant. The correlation coefficient ρ_i between y_t and y_{t-i} for $i = 1, 2, \ldots$ is found by simply dividing $\mathrm{Cov}[y_t, y_{t-i}]$ by the variance of y_t to obtain the Yule-Walker system of recursive equations. The first-order correlation coefficient is $\rho_1 = \mathrm{Cov}[y_t, y_{t-1}]/\, \mathrm{Var}[y_t] = \phi$ and it can be easily shown that the correlation ρ_n between y_t and y_{t-n} equals ϕ^n in general. Correlations therefore depend upon the time lag between observations, but not on time itself. Hence one may conclude that an AR(1) process with $|\phi| < 1$ and finite first two moments is at least weakly stationary [23]. It is worth noting that this is a limiting result obtained by assuming an infinite time line. It is therefore perhaps more accurate to refer to such a process as being *asymptotically stationary* [38].

One of the more common tests for stationarity is the Dickey-Fuller test. This test proceeds by subtracting y_{t-1} from both sides of Equation (2.1) to produce

$$(1 - \mathbb{L})y_t = \gamma y_{t-1} + \varepsilon_t, \tag{2.6}$$

where $\gamma = \phi - 1$. The null and alternative hypotheses for this test are then given as

$$H_0 : \gamma = 0$$
$$H_1 : \gamma < 0$$

where stationarity is implied by the alternative hypothesis and non-rejection of the null hypothesis suggests that the characteristic polynomial has a unit root. Note that this test was developed specifically for economic time series which are typically modelled as autoregressive processes with $0 \leq \phi \leq 1$. Explosive processes with $|\phi| > 1$ and oscillating processes with $\phi < 0$ are not common in economics and are therefore disregarded in this test.

While it may appear that this test can be carried out by performing a standard t-test on the ordinary least squares estimate of the γ coefficient, it can be shown that the test statistic does not have a conventional Student t distribution under the null hypothesis of a unit root [11]. Indeed, an implicit assumption of the t-test is that the data are drawn from a single population distribution, which is clearly contrary to the null hypothesis since non-stationarity implies that the parameters of the distribution of y_t depend on the time t. Consequently, Dickey and Fuller simulated critical values for this test for time series of various lengths [9]. MacKinnon later produced a much larger set of simulated critical values than those tabulated by Dickey and Fuller [41].

The above test of γ in Equation (2.6) is only applicable to AR(1) processes. This simple test has, however, been augmented to allow for higher-order autoregressive processes as well as the inclusion of intercept and trend terms. As an example of this augmentation, consider an AR(2) process with an intercept ν and deterministic trend δt written in terms of the lag operator as

$$(1 - \phi_1\mathbb{L} - \phi_2\mathbb{L}^2)y_t = \nu + \delta t + \varepsilon_t.$$

Now consider factorising the reverse characteristic polynomial of this AR(2) process as follows

$$(1 - \upsilon\mathbb{L})(1 - \varphi\mathbb{L})y_t = \nu + \delta t + \varepsilon_t, \tag{2.7}$$

where $\upsilon \geq \varphi$ with $0 \leq \upsilon \leq 1$ such that explosive and oscillating processes are again excluded from further consideration. Multiplying out Equation (2.7) and rearranging yields

$$\begin{aligned} y_t &= \upsilon\mathbb{L}y_t + \varphi\mathbb{L}y_t - \upsilon\varphi\mathbb{L}^2 y_t + \nu + \delta t + \varepsilon_t \\ (1-\mathbb{L})y_t &= -\mathbb{L}y_t + \upsilon\mathbb{L}y_t + \varphi\mathbb{L}y_t - \upsilon\varphi\mathbb{L}y_t + \upsilon\varphi\mathbb{L}y_t - \upsilon\varphi\mathbb{L}^2 y_t + \nu + \delta t + \varepsilon_t \\ &= (\upsilon - 1)(1-\varphi)y_{t-1} + \upsilon\varphi(1-\mathbb{L})y_{t-1} + \nu + \delta t + \varepsilon_t \\ &= \gamma y_{t-1} + \beta(1-\mathbb{L})y_{t-1} + \nu + \delta t + \varepsilon_t, \end{aligned}$$

where $\gamma = (\upsilon - 1)(1-\varphi)$ and $\beta = \upsilon\varphi$. The existence of at least one unit root in this AR(2) process would imply $\upsilon = 1$ and hence $\gamma = 0$. On the other hand, a stationary process is characterised by $|\upsilon| < 1$, which translates into $\gamma < 0$, ignoring the unlikely possibility of an explosive or oscillating process.

More generally, consider the AR(p) process

$$(1 - \phi_1\mathbb{L} - \phi_2\mathbb{L}^2 - \ldots - \phi_p\mathbb{L}^p)y_t = \nu + \delta t + \varepsilon_t$$

which, through a similar factorisation of the reverse characteristic polynomial, may be rewritten as

$$(1 - \upsilon\mathbb{L})(1 - \varphi_1\mathbb{L} - \varphi_2\mathbb{L}^2 - \ldots - \varphi_{p-1}\mathbb{L}^{p-1})y_t = \nu + \delta t + \varepsilon_t, \qquad (2.8)$$

where υ is the largest inverse root of the reverse characteristic polynomial of the AR(p) process with $0 \leq \upsilon \leq 1$. Equation (2.8) may be manipulated in a similar manner to the AR(2) process above in order to obtain the following formulation

$$(1-\mathbb{L})y_t = \gamma y_{t-1} + \sum_{i=1}^{p-1} \beta_i (1-\mathbb{L})y_{t-i} + \nu + \delta t + \varepsilon_t, \qquad (2.9)$$

where $\gamma = (\upsilon - 1)(1 - \sum_{i=1}^{p-1} \varphi_i)$ and $\beta_i = f_i(\upsilon, \varphi_1, \ldots, \varphi_{p-1})$ for $i = 1, \ldots, p-1$. Clearly, this process will also have a unit root under the null hypothesis $H_0 : \gamma = 0$, whilst the alternative hypothesis $H_1 : \gamma < 0$ implies stationarity if one disregards explosive and oscillating processes. This test of γ in Equation (2.9) is known as the *Augmented* Dickey-Fuller (ADF) test [5].

Since the characteristic polynomial of higher-order processes may have more than one unit root, the interpretation of the null hypothesis must be modified such that the non-rejection of $H_0 : \gamma = 0$ implies *at least* one unit root in the characteristic polynomial of y_t. If this is found to be the case, the ADF test can be repeated on the first differences of y_t, where the null hypothesis of *at* *least* two unit roots is tested against the alternative of exactly one unit root. One could continue in this manner taking differences of y_t and repeating the ADF test until the number of unit roots is established through the eventual rejection of the null hypothesis [5].

As was the case with the simple Dickey-Fuller test, a t-test of the ordinary least squares estimate of γ in Equation (2.9) is inappropriate and hence simulated critical values are utilised to test the statistical significance of the alternative hypothesis. Importantly, these critical values depend on whether or not intercept and trend terms are included in Equation (2.9), as well as the number of lags in this model. It is generally recommended to include a sufficient number of lags in the model to remove any serial correlation in the residuals. The inclusion of intercept and trend terms in the model is usually informed by their plausibility as a description of the data generating process [19]. Including irrelevant regressors in the regression will, however, reduce the power of the test, making it more likely to reject stationarity [11].

2.1.4 Integrated Processes

Integrated processes are a class of trending processes which may be rendered stationary by differencing. A stochastic process $\{y_t : t = 1, \ldots, T\}$ is said to be *integrated of order zero*, denoted $I(0)$, if it is stationary in the strict sense with a finite and non-zero long-run variance defined as the limit as $T \to \infty$ of $\text{Var}[\sqrt{T}\bar{y}]$ where $\bar{y} = T^{-1} \sum_{t=1}^{T} y_t$. Now, let $\Delta \equiv (1 - \mathbb{L})$ denote the difference operator such that

$$\Delta^d y_t = (1 - \mathbb{L})^d y_t.$$

Then the stochastic process $\{y_t\}$ is said to be integrated of order d, denoted $I(d)$ for $d = 1, 2, \ldots$, if its dth difference $\Delta^d y_t$ is $I(0)$ [21]. Alternatively, a univariate $I(d)$ process may be defined as having d unit roots in its characteristic polynomial [38]. The order of integration of a stochastic process may therefore be established by performing the sequence of ADF tests described earlier.

As an example, consider the pure random walk process $\{y_t\}$ in Equation (2.5). This process was shown to be non-stationary in levels. Taking first differences yields $\Delta y_t = (1 - \mathbb{L})y_t = \varepsilon_t \sim I(0)$ since $\{\varepsilon_t\}$ is white noise. Consequently, a non-stationary random walk process is rendered stationary by taking first differences and hence $y_t \sim I(1)$.

More generally, consider an AR(p) process written in the form given by Equation (2.8). Now suppose the characteristic polynomial of this process has a unit root implying $\upsilon = 1$. Then, omitting the intercept and trend terms, Equation (2.8) may be expressed as

$$\Delta y_t = \varphi_1 \Delta y_{t-1} + \varphi_2 \Delta y_{t-2} + \ldots + \varphi_{p-1} \Delta y_{t-p+1} + \varepsilon_t,$$

which is an AR($p-1$) process in first differences. If the remaining roots of $p(\mathbb{L}) = (1 - \varphi_1 \mathbb{L} - \varphi_2 \mathbb{L}^2 - \ldots - \varphi_{p-1} \mathbb{L}^{p-1})$ are all greater than one in absolute value, y_t will be stationary in first differences; that is $y_t \sim I(1)$. If $p(1) = 0$, then Δy_t is not stationary (due to a unit root in its characteristic polynomial) and it will be possible to factorise $p(\mathbb{L})$ further and take differences of Δy_t. If $\Delta^2 y_t$ is stationary, then $y_t \sim I(2)$. Otherwise, assuming $\{y_t\}$ is in fact an integrated process, one could conceivably continue differencing until $\Delta^d y_t$ is stationary and conclude that $y_t \sim I(d)$.

2.1.5 Cointegrated Processes

Equilibrium relationships are assumed to exist between many economic variables. Suppose k such time-dependent variables are collected in the vector $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{kt})'$ and their long-run equilibrium relation is given by the linear combination $\mathrm{E}[\boldsymbol{\beta}'\boldsymbol{y}_t] = \mathrm{E}[\beta_1 y_{1t} + \ldots + \beta_k y_{kt}] = 0$ where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)'$. In any particular period, this relation may not be satisfied exactly so that $\boldsymbol{\beta}'\boldsymbol{y}_t = z_t$, where $\{z_t\}$ is a stochastic process representing the short-run departures from equilibrium. Now suppose the variables in $\boldsymbol{y}_t$ are all $I(1)$. Then, assuming no equilibrium relationship exists between the variables in $\boldsymbol{y}_t$, a linear combination of these $I(1)$ processes will itself be $I(1)$ implying $z_t \sim I(1)$ [23]. However, if there is in fact an equilibrium relationship between the variables in $\boldsymbol{y}_t$, it seems quite plausible that these variables may wander extensively as a group and that $\{z_t\}$ is stationary at the same time. Hence, although each process is integrated, the process generated by a linear combination of the variables in $\boldsymbol{y}_t$ may be stationary. Integrated processes exhibiting this property are said to be *cointegrated* [38].

More generally, the variables in the k dimensional vector process $\boldsymbol{y}_t$ are *cointegrated of order* (d, b), denoted $CI(d, b)$, if all the components of $\boldsymbol{y}_t$ are $I(d)$ and there exists a linear combination $\boldsymbol{\beta}'\boldsymbol{y}_t$ with $\boldsymbol{\beta} \neq \mathbf{0}$ which is $I(d-b)$ for $b > 0$. For example, if $\boldsymbol{y}_t$ is a vector process of $I(1)$ variables and $\boldsymbol{\beta}'\boldsymbol{y}_t \sim I(0)$, then $\boldsymbol{y}_t \sim CI(1, 1)$. The vector $\boldsymbol{\beta}$ is referred to as a *cointegrating vector* and the stochastic process consisting of the cointegrated variables is said to be a *cointegrated process* [38].

2.2 The Vector Autoregressive Model

2.2.1 Definition

The vector autoregressive, or VAR, model arises frequently in the modelling of multivariate relationships and is a natural extension of its univariate autoregressive counterpart. The simplest case, a bivariate first-order VAR model, may be expressed mathematically as follows

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \nu_1 \\ \nu_2 \end{pmatrix} + \begin{bmatrix} \pi_{11.1} & \pi_{12.1} \\ \pi_{21.1} & \pi_{22.1} \end{bmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

$$\boldsymbol{y}_t = \boldsymbol{\nu} + \boldsymbol{\Pi}_1 \boldsymbol{y}_{t-1} + \boldsymbol{\varepsilon}_t$$

where $\boldsymbol{\nu} = (\nu_1, \nu_2)'$ is a vector of constants known as drifts and $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t})'$ is an identically and independently distributed random vector of innovations at time t relative to $\boldsymbol{y}_{t-1} = (y_{1,t-1}, y_{2,t-1})'$ at time $t-1$. More specifically, $\boldsymbol{\varepsilon}_t$ is usually regarded as an independent draw from a multivariate Gaussian distribution, denoted $\mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$, with a zero mean vector and variance-covariance matrix $\boldsymbol{\Omega}$. Although y_{1t} and y_{2t} could be modelled as two separate univariate autoregressions on $y_{1,t-1}$ and $y_{2,t-1}$ respectively, this approach would not capture any interactions that may be present between the two variables. The multivariate model, however, allows for such interactions by modelling each variable as a function of both its own lags as well as that of all other variables in the response vector $\boldsymbol{y}_t$ [43].

In general, a pth order VAR model in k variables may be expressed as

$$\boldsymbol{y}_t = \boldsymbol{\nu} + \boldsymbol{\Pi}_1 \boldsymbol{y}_{t-1} + \boldsymbol{\Pi}_2 \boldsymbol{y}_{t-2} + \ldots + \boldsymbol{\Pi}_p \boldsymbol{y}_{t-p} + \boldsymbol{\varepsilon}_t,$$

where $\boldsymbol{y}_t = (y_{1t}, y_{2t}, \ldots, y_{kt})'$, $\boldsymbol{\nu} = (\nu_1, \nu_2, \ldots, \nu_k)'$, $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, \ldots, \varepsilon_{kt})' \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$ and

$$\boldsymbol{\Pi}_j = \begin{bmatrix} \pi_{11.j} & \pi_{12.j} & \cdots & \pi_{1k.j} \\ \pi_{21.j} & \pi_{22.j} & \cdots & \pi_{2k.j} \\ \vdots & \vdots & \ddots & \vdots \\ \pi_{k1.j} & \pi_{k2.j} & \cdots & \pi_{kk.j} \end{bmatrix}.$$

This VAR model is said to be in *reduced form* in that no current dated values of the k variables appear as regressors in any of the equations. When current values are included, the model is said to be in *structural form* in the sense that the equations directly represent behavioural relationships, rather than merely the interactions of such relationships [43].

2.2.2 Stationarity of a VAR Process

Consider the VAR(1) model without drift written in terms of the lag operator as

$$(\boldsymbol{I} - \boldsymbol{\Pi}_1 \mathbb{L})\, \boldsymbol{y}_t = \boldsymbol{\epsilon}_t.$$

Multiplying through by $(\boldsymbol{I} - \boldsymbol{\Pi}_1 \mathbb{L})^{-1}$ yields

$$\begin{aligned} \boldsymbol{y}_t &= (\boldsymbol{I} - \boldsymbol{\Pi}_1 \mathbb{L})^{-1}\, \boldsymbol{\epsilon}_t \\ |\boldsymbol{I} - \boldsymbol{\Pi}_1 \mathbb{L}| \boldsymbol{y}_t &= \text{adj}\,[\boldsymbol{I} - \boldsymbol{\Pi}_1 \mathbb{L}]\, \boldsymbol{\epsilon}_t \end{aligned} \tag{2.10}$$

where adj $[\boldsymbol{I} - \boldsymbol{\Pi}_1 \mathbb{L}]$ is the adjoint matrix.

Now consider the eigenvalue problem $\boldsymbol{\Pi}_1 \boldsymbol{x} = \lambda \boldsymbol{x}$ with characteristic equation

$$|\boldsymbol{\Pi}_1 - \lambda \boldsymbol{I}| = 0,$$

where the polynomial $p(\lambda) = |\boldsymbol{\Pi}_1 - \lambda \boldsymbol{I}|$ is known as the *characteristic polynomial* of $\boldsymbol{\Pi}_1$. The real and complex roots $\lambda_1, \lambda_2, \ldots, \lambda_k$ of this polynomial are referred to as the eigenvalues of $\boldsymbol{\Pi}_1$, each of which has an associated eigenvector $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k$ such that $\boldsymbol{\Pi}_1 \boldsymbol{x}_i = \lambda_i \boldsymbol{x}_i$ for $i = 1, \ldots, k$. Alternatively, the eigenvalues of $\boldsymbol{\Pi}_1$ may be found by considering the *reverse* characteristic equation

$$|\boldsymbol{I} - \boldsymbol{\Pi}_1 \mathbb{L}| = 0,$$

where $\mathbb{L} = \lambda^{-1}$. The reverse characteristic polynomial is then defined as $q(\mathbb{L}) = |\boldsymbol{I} - \boldsymbol{\Pi}_1 \mathbb{L}|$ which has roots equal to the inverse eigenvalues of $\boldsymbol{\Pi}_1$ [20].

Recall from Section 2.1.3 that the univariate AR(1) process

$$y_t = \phi y_{t-1} + \epsilon_t \Leftrightarrow (1 - \phi \mathbb{L}) y_t = \epsilon_t$$

is stationary if its autoregressive coefficient ϕ is less than one in absolute value. This condition is clearly equivalent to stating that the polynomial $(1 - \phi \mathbb{L})$ must have a root greater than one in modulus. Now consider the VAR(1) model in Equation (2.10) written in terms of the reverse characteristic polynomial as

$$q^*(\mathbb{L})(1 - \upsilon \mathbb{L}) \boldsymbol{y}_t = \text{adj}\,[\boldsymbol{I} - \boldsymbol{\Pi}_1 \mathbb{L}]\, \boldsymbol{\epsilon}_t,$$

where $q(\mathbb{L}) = q^*(\mathbb{L})(1 - \upsilon\mathbb{L})$ and υ is a real or complex root of the polynomial $q(\mathbb{L})$. Rearranging terms and defining $\boldsymbol{\zeta}_t = q^*(\mathbb{L})^{-1} \operatorname{adj}[\boldsymbol{I} - \boldsymbol{\Pi}_1\mathbb{L}] \boldsymbol{\epsilon}_t$ yields

$$\boldsymbol{y}_t = \upsilon \boldsymbol{y}_{t-1} + \boldsymbol{\zeta}_t. \tag{2.11}$$

It therefore follows that each of the stacked univariate processes in $\boldsymbol{y}_t$ will only be stationary if $|\upsilon| < 1$ or, equivalently, the absolute root of the polynomial $(1 - \upsilon\mathbb{L})$ is greater than one. Moreover, this must be the case for *all* roots of the reverse characteristic polynomial $q(\mathbb{L})$, since for each real or complex root, it is possible to rewrite the VAR(1) process in the form given by Equation (2.11). Since a vector stochastic process is only stationary if each of its univariate constituents is stationary, it follows that a sufficient condition for stationarity in the VAR(1) model is that all the roots of the reverse characteristic polynomial must be greater than one in absolute value or, equivalently, lie outside the unit circle [38].

Although the above discussion has focused exclusively on first-order vector autoregressions, the results may be easily generalised to higher-order VAR(p) processes by noting that any VAR(p) process can be expressed as a first-order system by rewriting it in what is known as the *companion form*. To illustrate, consider the VAR(p) process

$$\boldsymbol{y}_t = \boldsymbol{\nu} + \boldsymbol{\Pi}_1 \boldsymbol{y}_{t-1} + \boldsymbol{\Pi}_2 \boldsymbol{y}_{t-2} + \ldots + \boldsymbol{\Pi}_p \boldsymbol{y}_{t-p} + \boldsymbol{\epsilon}_t.$$

This process may be expressed in companion form as

$$\boldsymbol{Y}_t = \boldsymbol{A}_0 + \boldsymbol{A}_1 \boldsymbol{Y}_{t-1} + \boldsymbol{E}_t,$$

where $\boldsymbol{Y}_t = \left(\boldsymbol{y}_t', \boldsymbol{y}_{t-1}', \ldots, \boldsymbol{y}_{t-p+1}'\right)'$, $\boldsymbol{A}_0 = (\boldsymbol{\nu}', \mathbf{0}', \ldots, \mathbf{0}')'$, $\boldsymbol{E}_t = (\boldsymbol{\epsilon}_t', \mathbf{0}', \ldots, \mathbf{0}')'$ and

$$\boldsymbol{A}_1 = \begin{bmatrix} \boldsymbol{\Pi}_1 & \boldsymbol{\Pi}_2 & \cdots & \boldsymbol{\Pi}_{p-1} & \boldsymbol{\Pi}_p \\ \boldsymbol{I} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{I} & \mathbf{0} \end{bmatrix}.$$

Stationarity in this context is therefore ensured when all the eigenvalues of the companion matrix $\boldsymbol{A}_1$ have modulus less than one [43].

2.3 The Vector Error Correction Model

2.3.1 Definition

The vector error correction model (VECM) is a convenient reparameterisation of the VAR model discussed above. To illustrate its derivation, consider the simple VAR(3) process without an intercept or other deterministic terms

$$\boldsymbol{y}_t = \boldsymbol{\Pi}_1 \boldsymbol{y}_{t-1} + \boldsymbol{\Pi}_2 \boldsymbol{y}_{t-2} + \boldsymbol{\Pi}_3 \boldsymbol{y}_{t-3} + \boldsymbol{\varepsilon}_t.$$

Subtracting $\boldsymbol{y}_{t-1}$ from both sides and simplifying, the VAR(3) process may be rewritten as follows

$$\begin{aligned}
\Delta \boldsymbol{y}_t &= (\boldsymbol{\Pi}_1 - \boldsymbol{I})\, \boldsymbol{y}_{t-1} + \boldsymbol{\Pi}_2 \boldsymbol{y}_{t-2} + \boldsymbol{\Pi}_3 \boldsymbol{y}_{t-3} + \boldsymbol{\varepsilon}_t \\
&= (\boldsymbol{\Pi}_1 - \boldsymbol{I})\, \boldsymbol{y}_{t-1} + \boldsymbol{\Pi}_2 \boldsymbol{y}_{t-2} + \boldsymbol{\Pi}_3 \boldsymbol{y}_{t-2} - \boldsymbol{\Pi}_3 \boldsymbol{y}_{t-2} + \boldsymbol{\Pi}_3 \boldsymbol{y}_{t-3} + \boldsymbol{\varepsilon}_t \\
&= (\boldsymbol{\Pi}_1 - \boldsymbol{I})\, \boldsymbol{y}_{t-1} + (\boldsymbol{\Pi}_2 + \boldsymbol{\Pi}_3)\, \boldsymbol{y}_{t-2} - \boldsymbol{\Pi}_3 \Delta \boldsymbol{y}_{t-2} + \boldsymbol{\varepsilon}_t \\
&= (\boldsymbol{\Pi}_1 - \boldsymbol{I})\, \boldsymbol{y}_{t-1} + (\boldsymbol{\Pi}_2 + \boldsymbol{\Pi}_3)\, \boldsymbol{y}_{t-1} - (\boldsymbol{\Pi}_2 + \boldsymbol{\Pi}_3)\, \boldsymbol{y}_{t-1} + (\boldsymbol{\Pi}_2 + \boldsymbol{\Pi}_3)\, \boldsymbol{y}_{t-2} \\
&\quad - \boldsymbol{\Pi}_3 \Delta \boldsymbol{y}_{t-2} + \boldsymbol{\varepsilon}_t \\
&= (\boldsymbol{\Pi}_1 + \boldsymbol{\Pi}_2 + \boldsymbol{\Pi}_3 - \boldsymbol{I})\, \boldsymbol{y}_{t-1} - (\boldsymbol{\Pi}_2 - \boldsymbol{\Pi}_3)\, \Delta \boldsymbol{y}_{t-1} - \boldsymbol{\Pi}_3 \Delta \boldsymbol{y}_{t-2} + \boldsymbol{\varepsilon}_t.
\end{aligned} \tag{2.12}$$

Equation (2.12) is the error correction representation of a VAR(3) process. In general, the error correction representation of a k dimensional VAR(p) process is

$$\Delta \boldsymbol{y}_t = \boldsymbol{\Pi} \boldsymbol{y}_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i \Delta \boldsymbol{y}_{t-i} + \boldsymbol{\Phi} \boldsymbol{D}_t + \boldsymbol{\varepsilon}_t \tag{2.13}$$

where $\boldsymbol{\Pi} = \sum_{i=1}^{p} \boldsymbol{\Pi}_i - \boldsymbol{I}$, $\boldsymbol{\Gamma}_i = -\sum_{j=i+1}^{p} \boldsymbol{\Pi}_j$ and as before it is assumed that $\boldsymbol{\varepsilon}_t \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$. Note that Equation (2.13) allows for deterministic terms and $I(0)$ exogenous variables in the model through the $d \times 1$ vector $\boldsymbol{D}_t$ which may include a constant (drift), trend, seasonal dummies, intervention dummies or other regressors which are considered exogenous to the system. The $k \times d$ matrix $\boldsymbol{\Phi}$ includes the coefficients associated with these deterministic and exogenous variables [27]. Observe that a cointegrated VAR(p) process translates into a VECM of order $p-1$, since $p-1$ lags of the response vector $\Delta \boldsymbol{y}_t$ appear as regressors in the VECM given as Equation (2.13).

To illustrate the usefulness of this representation, suppose that the characteristic polynomial of a VAR(p) process has exactly one unit root such that

$\boldsymbol{y}_t$ is $I(1)$ and therefore non-stationary. One possible response to such a case may be to model the first differences as a VAR(p) process of the form

$$\Delta \boldsymbol{y}_t = \boldsymbol{\Pi}_1 \Delta \boldsymbol{y}_{t-1} + \boldsymbol{\Pi}_2 \Delta \boldsymbol{y}_{t-2} + \ldots + \boldsymbol{\Pi}_p \Delta \boldsymbol{y}_{t-p} + \boldsymbol{\Phi} \boldsymbol{D}_t + \boldsymbol{\varepsilon}_t^*.$$

Statistically, this model may seem acceptable since the first differences of $I(1)$ variables are by definition stationary. However, such a formulation provides no information on the relationship between the levels of the variables in the VAR model, yet it is these precise relationships which are usually of economic interest [43].

An alternative arises when stationary relations exist between the non-stationary levels of the variables in $\boldsymbol{y}_t$; that is, the k variables in $\boldsymbol{y}_t$ are *cointegrated*. Suppose $\boldsymbol{y}_t$ is integrated of order one as described above. Recall that $\boldsymbol{y}_t$ is said to be cointegrated, denoted $CI(1,1)$, if there exists a linear combination of these variables which is $I(0)$. Now suppose $0 < r < k$ such linearly independent combinations exist so that $\boldsymbol{\beta}_1 \boldsymbol{y}_t, \ldots, \boldsymbol{\beta}_r \boldsymbol{y}_t$ are all stationary processes. Then the $k \times r$ matrix $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_r)$ is referred to as the cointegration matrix of $\boldsymbol{y}_t$ where each column represents a unique cointegrating relation between the variables in $\boldsymbol{y}_t$. The r dimensional vector process $\boldsymbol{\beta}' \boldsymbol{y}_t$ is therefore integrated of order zero [43].

Returning to the VECM($p-1$) given as Equation (2.13), if $\boldsymbol{y}_t \sim I(1)$ and hence $\Delta \boldsymbol{y}_t \sim I(0)$, then at first glance this model would appear to be unbalanced in its time series properties; the left hand side is clearly $I(0)$, whilst the right hand side includes the $I(1)$ vector $\boldsymbol{y}_{t-1}$ and, since a linear combination of $I(1)$ variables is in general also $I(1)$, the right hand side would therefore appear to be $I(1)$. However, suppose that $\boldsymbol{\Pi} = \boldsymbol{\alpha} \boldsymbol{\beta}'$ with $\boldsymbol{\beta}$ representing the matrix of r cointegrating vectors alluded to earlier. Then the VECM($p-1$) process may be rewritten as

$$\Delta \boldsymbol{y}_t = \boldsymbol{\alpha} \boldsymbol{\beta}' \boldsymbol{y}_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i \Delta \boldsymbol{y}_{t-i} + \boldsymbol{\Phi} \boldsymbol{D}_t + \boldsymbol{\varepsilon}_t. \tag{2.14}$$

If the variables in $\boldsymbol{y}_t$ are indeed cointegrated, then the linear combination of $I(1)$ variables $\boldsymbol{\beta}' \boldsymbol{y}_{t-1}$ has a reduced order of integration and is therefore $I(0)$ in this case. Hence, the right hand side of Equation (2.14) becomes a linear combination of $I(0)$ variables and the model would therefore be balanced in its time series properties provided that such cointegrating relations exist [43].

The interpretation of the error correction model in Equation (2.14) is indeed most appealing. The cointegrating combinations $\boldsymbol{\beta}' \boldsymbol{y}_{t-1}$ represent the long-run or equilibrium relationships among the levels of the variables. Recall that these relationships are often of distinct interest in empirical studies. Nonzero values of $\boldsymbol{\beta}'\boldsymbol{y}_{t-1}$ indicate lagged disequilibria which are eradicated through the adjustment coefficients in the $k \times r$ matrix $\boldsymbol{\alpha}$, with each column corresponding to one of the r cointegrating relations. For this reason, the model is sometimes, and perhaps more appropriately, referred to as the vector *equilibrium* correction model. Short-run adjustments are also captured in the model through the elements in the $\boldsymbol{\Gamma}_i$ matrices [43]. A detailed account of parameter estimation in the VECM is provided in Appendix A.

2.3.2 Hypothesis Tests of the Cointegrating Rank

The rank of $\boldsymbol{\beta}$ is referred to as the *cointegrating rank* and represents the number of linearly independent cointegrating relations. Since $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$, where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are both $k \times r$ matrices with $r < k$, this is equivalent to finding the rank of $\boldsymbol{\Pi}$. By excluding the possibility that $r \geq k$, it should be immediately obvious that $\boldsymbol{\Pi}$ is singular or *rank deficient*. In fact, this must be the case if the characteristic polynomial of the VAR process has a unit root such that $\boldsymbol{y}_t \sim I(1)$. To understand why this is so, consider the now familiar VAR(p) process without drift or deterministic terms

$$\boldsymbol{y}_t = \boldsymbol{\Pi}_1\boldsymbol{y}_{t-1} + \boldsymbol{\Pi}_2\boldsymbol{y}_{t-2} + \ldots + \boldsymbol{\Pi}_p\boldsymbol{y}_{t-p} + \boldsymbol{\varepsilon}_t$$

which, using a previous trick, may be rewritten as

$$\begin{aligned}\left|\boldsymbol{I} - \boldsymbol{\Pi}_1\mathbb{L} - \boldsymbol{\Pi}_2\mathbb{L}^2 - \ldots - \boldsymbol{\Pi}_p\mathbb{L}^p\right| \boldsymbol{y}_t & \\ = \operatorname{adj}\left[\boldsymbol{I} - \boldsymbol{\Pi}_1\mathbb{L} - \boldsymbol{\Pi}_2\mathbb{L}^2 - \ldots - \boldsymbol{\Pi}_p\mathbb{L}^p\right] \boldsymbol{\varepsilon}_t &\end{aligned}$$

with reverse characteristic polynomial $|\boldsymbol{I} - \boldsymbol{\Pi}_1\mathbb{L} - \boldsymbol{\Pi}_2\mathbb{L}^2 - \ldots - \boldsymbol{\Pi}_p\mathbb{L}^p|$. Note that this approach is equivalent to rewriting the VAR(p) process in companion form, in which case the reverse characteristic polynomial is found to be $|\boldsymbol{I} - \boldsymbol{A}_1\mathbb{L}| = |\boldsymbol{I} - \boldsymbol{\Pi}_1\mathbb{L} - \boldsymbol{\Pi}_2\mathbb{L}^2 - \ldots - \boldsymbol{\Pi}_p\mathbb{L}^p|$. Now suppose this polynomial has a unit root. Then

$$|\boldsymbol{I} - \boldsymbol{A}_1| = |\boldsymbol{I} - \boldsymbol{\Pi}_1 - \boldsymbol{\Pi}_2 - \ldots - \boldsymbol{\Pi}_p| = 0$$

and, noting that $\boldsymbol{\Pi} = \boldsymbol{\Pi}_1 + \boldsymbol{\Pi}_2 + \ldots + \boldsymbol{\Pi}_p - \boldsymbol{I}$, it therefore follows that $|\boldsymbol{\Pi}| = 0$ so that $\boldsymbol{\Pi}$ will be rank deficient [43]. The case where $r \geq k$ would therefore imply that $\boldsymbol{y}_t$ is stationary, in which case cointegration is irrelevant.

In the event that there are no cointegrating relations amongst the $I(1)$ variables in $\boldsymbol{y}_t$, the rank of $\boldsymbol{\Pi}$ will be zero. However, $\boldsymbol{\Pi}$ can only be completely rank deficient if it is the null matrix and the VECM($p-1$) would therefore reduce to a VAR($p-1$) model in first differences [27].

In order to establish a formal testing procedure for the cointegrating rank, define $H(r)$ as the submodel

$$\Delta \boldsymbol{y}_t = \boldsymbol{\Pi} \boldsymbol{y}_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i \Delta \boldsymbol{y}_{t-i} + \boldsymbol{\Phi} \boldsymbol{D}_t + \boldsymbol{\varepsilon}_t \tag{2.15}$$

under the condition that $\text{rank}(\boldsymbol{\Pi}) \leq r$. This formulation leads to the nested sequence of models

$$H(0) \subset \cdots \subset H(r) \subset \cdots \subset H(k).$$

The models $H(1), \ldots, H(k-1)$ imply cointegrating relations between the $I(1)$ variables in the response vector. The model $H(k)$ is the unrestricted VAR or $I(0)$ model, whilst $H(0)$ imposes the restriction $\boldsymbol{\Pi} = \boldsymbol{0}$ corresponding to a VAR model in first differences. The nested sequence ranges from a VAR model in differences through to a VAR model in stationary levels and therefore provides a means for investigating the coefficient matrix $\boldsymbol{\Pi}$ in terms of the information it may include with respect to the long-run equilibria hidden in the data [27].

The above formulation allows for the construction of likelihood ratios to test the null hypothesis H_{0r} : model $H(r)$ against the alternative hypotheses H_{1k} : model $H(k)$ and $H_{1,r+1}$: model $H(r+1)$. The full derivation of these hypothesis tests from the likelihood function of the general VECM is given in Appendix A. Briefly, these tests are derived by first concentrating the likelihood function for the VECM process

$$\Delta \boldsymbol{y}_t = \boldsymbol{\alpha} \boldsymbol{\beta}' \boldsymbol{y}_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i \Delta \boldsymbol{y}_{t-i} + \boldsymbol{\Phi} \boldsymbol{D}_t + \boldsymbol{\varepsilon}_t$$

with respect to the parameters $\{\boldsymbol{\Gamma}_1, \ldots, \boldsymbol{\Gamma}_{p-1}, \boldsymbol{\Phi}\}$ by regressing $\Delta \boldsymbol{y}_t$ and $\boldsymbol{y}_{t-p}$ on $\Delta \boldsymbol{y}_{t-1}, \ldots, \Delta \boldsymbol{y}_{t-p+1}$ and $\boldsymbol{D}_t$ [30]. Define the residual vectors from these two regressions at time t as $\boldsymbol{R}_{0t}$ and $\boldsymbol{R}_{1t}$ respectively with residual product moment matrices

$$\boldsymbol{S}_{ij} = T^{-1} \sum_{t=1}^{T} \boldsymbol{R}_{it} \boldsymbol{R}'_{jt} \qquad \text{for } i, j = 0, 1,$$

where T is the length of the observed data sequence. Further define $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots \geq \hat{\lambda}_k$ as the ordered eigenvalues associated with the eigenvalue problem

$$\left|\lambda \boldsymbol{S}_{11} - \boldsymbol{S}_{10}\boldsymbol{S}_{00}^{-1}\boldsymbol{S}_{01}\right| = 0.$$

The likelihood ratio statistic for the test of H_{0r} against H_{1k} may then be constructed by dividing the maximised likelihood function for the model with the restriction rank($\boldsymbol{\Pi}$) $= r$ by the maximised likelihood of the unrestricted VAR model with $\boldsymbol{\Pi}$ of full rank; that is,

$$Q\left(H(r)|H(k)\right) = \frac{\left(|\boldsymbol{S}_{00}|\prod_{i=1}^{r}(1-\hat{\lambda}_i)\right)^{-T/2}}{\left(|\boldsymbol{S}_{00}|\prod_{i=1}^{k}(1-\hat{\lambda}_i)\right)^{-T/2}}$$

where $Q(\cdot)$ is the likelihood ratio. This expression may be rearranged as follows

$$\begin{aligned} Q\left(H(r)|H(k)\right)^{-2/T} &= \frac{|\boldsymbol{S}_{00}|\prod_{i=1}^{r}(1-\hat{\lambda}_i)}{|\boldsymbol{S}_{00}|\prod_{i=1}^{k}(1-\hat{\lambda}_i)} \\ &= \prod_{i=r+1}^{k}(1-\hat{\lambda}_i)^{-1} \end{aligned}$$

which, after taking logarithms, leads to the so-called trace statistic [27]

$$-2\log Q\left(H(r)|H(k)\right) = -T\sum_{i=r+1}^{k}\log(1-\hat{\lambda}_i).$$

A "large" value of the trace statistic is evidence against the hypothesis H_{0r}; that is, the cointegrating rank is greater than r. A "small" value of the trace statistic indicates a lack of evidence against H_{0r} and hence would lead to the conclusion that the cointegrating rank is less than or equal to r. In order to determine the exact cointegrating rank, the following sequence of hypothesis tests would need to be performed:

1. Begin by testing the model $H(0)$ against the model $H(k)$. If $H(0)$ is not rejected, then the cointegrating rank $r = 0$ and the sequence stops. If $H(0)$ is rejected, then conclude that the cointegrating rank is greater than zero and move on to the next test.

2. Test the model $H(1)$ against the model $H(k)$. If $H(1)$ is not rejected, then $r \leq 1$ and, since $r = 0$ was rejected in the previous test, conclude that the cointegrating rank $r = 1$ and the sequence stops. If $H(1)$ is rejected, then conclude that the cointegrating rank is greater than one and move on to the next test.

3. Continue this process with the last possible test considering the model $H(k-1)$ against the model $H(k)$. If $H(k-1)$ is not rejected, then $r \leq k-1$ and, since $r \leq k-2$ was rejected in the previous test, conclude that the cointegrating rank $r = k-1$. Otherwise, if $H(k-1)$ is rejected, conclude that $r = k$ [43].

Another likelihood ratio test that may be employed to ascertain the cointegrating rank is the test of H_{0r} : model $H(r)$ against $H_{1,r+1}$: model $H(r+1)$. The likelihood ratio test statistic is constructed as before, except the denominator now becomes the maximised likelihood function for the model $H(r+1)$

$$\begin{aligned} Q\left(H(r)|H(r+1)\right)^{-2/T} &= \frac{|\boldsymbol{S}_{00}| \prod_{i=1}^{r}(1-\hat{\lambda}_i)}{|\boldsymbol{S}_{00}| \prod_{i=1}^{r+1}(1-\hat{\lambda}_i)} \\ &= (1-\hat{\lambda}_{r+1})^{-1}. \end{aligned}$$

Taking logarithms and rearranging produces what is known as the maximal eigenvalue or $\lambda_{\max}$ test statistic [27]

$$-2\log Q\left(H(r)|H(r+1)\right) = -T\log(1-\hat{\lambda}_{r+1}).$$

The testing procedure using the $\lambda_{\max}$ statistic is almost identical to that employed when working with the trace statistic, with a minor modification. Begin by testing the model $H(0)$ against the model $H(1)$, rather than $H(k)$ as previously. If $H(0)$ is not rejected, stop the sequence and conclude that the cointegrating rank $r = 0$. If $H(0)$ is rejected, proceed to test the model $H(1)$ against the model $H(2)$. Continue in this manner if necessary until the last test of the model $H(k-1)$ against the model $H(k)$, in which case the $\lambda_{\max}$ and trace statistics will be identical [43]. The aforementioned procedures based on the $\lambda_{\max}$ and trace statistics were largely developed by Søren Johansen and consequently have become known as the *Johansen approach* to rank selection.

The asymptotic distributions of both the trace and $\lambda_{\max}$ statistics are non-standard and depend upon the deterministic terms included in the model.

The asymptotic distributions and critical values have, however, been tabulated from simulations for various values of $k-r$ and under different assumptions regarding the deterministic terms [43].

An alternative approach to choosing the cointegrating rank is based on familiar model selection criteria, such as Akaike's information criterion (AIC), the Hannan-Quinn information criterion (HQC) and the Schwarz Bayesian information criterion (SBC). At the outset, this approach would seem advantageous in that the order p and rank r of the VAR model written in its error correction form as in Equation (2.15) may be chosen simultaneously [38]. Indeed, the specified lag length in the VAR model forms the basis of inference on the cointegrating rank [43].

A Monte Carlo study conducted by Reimers using each of the above three information criteria to select model order and cointegrating rank revealed mixed results [46]. Changing the order, rank and length of the observed time series seems to have a large effect on the performance of each criterion in correctly identifying model order, rank and the correct combination of the two [43]. Furthermore, these changes appear to affect the performance of each criterion differently. Consequently, it is possible to arrive at different model specifications depending on the choice of information criterion [38]. Overall, it is suggested that, if an information criterion is to be used, the SBC is probably the best choice to simultaneously estimate order and cointegrating rank. However, the dominant practice at present is to first choose the lag length based on one or more of the information criteria and then to decide upon the cointegrating rank following the Johansen approach [43].

2.3.3 Intercepts and Trends in the VECM

Consider the VECM process

$$\Delta \boldsymbol{y}_t = \boldsymbol{\alpha} \begin{pmatrix} \boldsymbol{\beta}' & \boldsymbol{\nu}_1 & \boldsymbol{\delta}_1 \end{pmatrix} \begin{pmatrix} \boldsymbol{y}_{t-1} \\ 1 \\ t \end{pmatrix} + \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i \Delta \boldsymbol{y}_{t-i} + \boldsymbol{\varepsilon}_t \quad \text{for } t = 1, 2, \ldots, T, \tag{2.16}$$

which allows for an intercept term and a deterministic trend in the equilibrium relations through the inclusion of the r dimensional vectors $\boldsymbol{\nu}_1$ and $\boldsymbol{\delta}_1$ in the cointegration space. The inclusion of these intercepts and trends in the cointegration space will affect the likelihood ratio statistics employed in the hypothesis tests of the cointegrating rank and hence also the critical values used to determine statistical significance in such tests.

The VECM in Equation (2.16) with an intercept and trend in the cointegration space may be rewritten in the more familiar form

$$\Delta \boldsymbol{y}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{y}_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i \Delta \boldsymbol{y}_{t-i} + \boldsymbol{\Phi}\boldsymbol{D}_t + \boldsymbol{\varepsilon}_t \qquad \text{for } t = 1, 2, \ldots, T, \tag{2.17}$$

with $\boldsymbol{\Phi} = (\boldsymbol{\nu}, \boldsymbol{\delta})$ and $\boldsymbol{D}_t = (1, t)'$, where $\boldsymbol{\nu} = \boldsymbol{\alpha}\boldsymbol{\nu}_1$ and $\boldsymbol{\delta} = \boldsymbol{\alpha}\boldsymbol{\delta}_1$ in this case. Note, however, that in this example it is assumed that there is no intercept or trend in the model itself; that is, the intercept and trend are restricted to the cointegration space only. Of course, it is possible to have constants and trends arising from both of these sources simultaneously. In this more general case, it will initially be unclear whether the intercept and trend terms, $\boldsymbol{\nu}$ and $\boldsymbol{\delta}$, pertain to the cointegration space, the actual data generating process or both when the model is estimated in the form given by Equation (2.17). In order to draw inferences concerning the cointegrating rank of the model, however, it is necessary to partition the deterministic component according to the two sources from which it may originate [43].

More generally then, suppose $\boldsymbol{\Phi} = (\boldsymbol{\nu}, \boldsymbol{\delta})$, $\boldsymbol{D}_t = (1, t)'$ and $\boldsymbol{\alpha}\boldsymbol{\nu}_1$ and $\boldsymbol{\alpha}\boldsymbol{\delta}_1$ are constant and trend terms generated in the cointegration space as before. Now define $\boldsymbol{\alpha}_\perp$ as a $k \times (k - r)$ matrix of full rank which is orthogonal to $\boldsymbol{\alpha}$. Further define $\boldsymbol{\nu}_2$ and $\boldsymbol{\delta}_2$ as $k-r$ dimensional vectors. Then the deterministic component of the model may be decomposed into two unrelated or orthogonal parts

$$\boldsymbol{\Phi}\boldsymbol{D}_t = \boldsymbol{\nu} + \boldsymbol{\delta}t = \boldsymbol{\alpha}(\boldsymbol{\nu}_1 + \boldsymbol{\delta}_1 t) + \boldsymbol{\alpha}_\perp(\boldsymbol{\nu}_2 + \boldsymbol{\delta}_2 t), \tag{2.18}$$

where $\boldsymbol{\alpha}(\boldsymbol{\nu}_1 + \boldsymbol{\delta}_1 t)$ is generated in the cointegration space and $\boldsymbol{\alpha}_\perp(\boldsymbol{\nu}_2 + \boldsymbol{\delta}_2 t)$ is generated in the model. Clearly, $\boldsymbol{\nu} = \boldsymbol{\alpha}\boldsymbol{\nu}_1 + \boldsymbol{\alpha}_\perp\boldsymbol{\nu}_2$ and $\boldsymbol{\delta} = \boldsymbol{\alpha}\boldsymbol{\delta}_1 + \boldsymbol{\alpha}_\perp\boldsymbol{\delta}_2$ such that Equation (2.16) holds when there is an intercept term and a trend term in the cointegration space, but not in the data generating process; that is, $\boldsymbol{\nu}_2 = \boldsymbol{\delta}_2 = \boldsymbol{0}$.

The deterministic components arising from the cointegration space may be extracted by premultiplying Equation (2.18) through by $(\boldsymbol{\alpha}'\boldsymbol{\alpha})^{-1}\boldsymbol{\alpha}'$ and noting that $\boldsymbol{\alpha}'\boldsymbol{\alpha}_\perp = \boldsymbol{0}$ due to orthogonality

$$\begin{aligned}(\boldsymbol{\alpha}'\boldsymbol{\alpha})^{-1}\boldsymbol{\alpha}'\boldsymbol{\Phi}\boldsymbol{D}_t &= (\boldsymbol{\alpha}'\boldsymbol{\alpha})^{-1}\boldsymbol{\alpha}'\boldsymbol{\alpha}(\boldsymbol{\nu}_1 + \boldsymbol{\delta}_1 t) + (\boldsymbol{\alpha}'\boldsymbol{\alpha})^{-1}\boldsymbol{\alpha}'\boldsymbol{\alpha}_\perp(\boldsymbol{\nu}_2 + \boldsymbol{\delta}_2 t)\\ &= \boldsymbol{\nu}_1 + \boldsymbol{\delta}_1 t\end{aligned}$$

so that $\boldsymbol{\nu}_1 = (\boldsymbol{\alpha}'\boldsymbol{\alpha})^{-1}\boldsymbol{\alpha}'\boldsymbol{\nu}$ and $\boldsymbol{\delta}_1 = (\boldsymbol{\alpha}'\boldsymbol{\alpha})^{-1}\boldsymbol{\alpha}'\boldsymbol{\delta}$. Similarly, premultiplying through by $(\boldsymbol{\alpha}_\perp'\boldsymbol{\alpha}_\perp)^{-1}\boldsymbol{\alpha}_\perp'$ reveals that $\boldsymbol{\nu}_2 = (\boldsymbol{\alpha}_\perp'\boldsymbol{\alpha}_\perp)^{-1}\boldsymbol{\alpha}_\perp'\boldsymbol{\nu}$ and $\boldsymbol{\delta}_2 = (\boldsymbol{\alpha}_\perp'\boldsymbol{\alpha}_\perp)^{-1}\boldsymbol{\alpha}_\perp'\boldsymbol{\delta}$. Hence it is a straightforward exercise to decompose the deterministic component of the model into its two constituents, namely that which is attributable to the data generating process and that which is inherent in the long-run relationships between the variables [27].

In deciding whether or not to allow for intercept and trend terms in either the model, the cointegration space or both, one needs to consider the roles that such terms fulfill. As it turns out, the practical implications of a constant and deterministic trend are quite different depending on whether these terms are assumed to characterise the data generating process or the cointegrating relations. Suppose for simplicity that the variables $\boldsymbol{y}_t$ in the VECM process defined by Equation (2.17) are in logarithms so that $\Delta\boldsymbol{y}_t$ may be interpreted as a rate of growth. Including a constant term in the model, but not in the cointegration space, therefore implies “autonomous drift” or constant growth over time which leads to a deterministic trend in the levels of $\boldsymbol{y}_t$. By contrast, including a constant term in the cointegration space simply allows for a nonzero intercept term in the equilibrium relationships between the variables in $\boldsymbol{y}_t$. Indeed, one would ordinarily allow for a nonzero intercept term in a linear model such that a constant in the cointegration space would seem to be a reasonable *a priori* specification. Note, however, that including a constant in the cointegration space only will *not* account for a linear trend in the data. Of course, a constant in both the cointegration space and the data generating process will allow for both a nonzero intercept term in the cointegrating relations and a linear trend in the data. Indeed, such a model would seem practically relevant [43].

Inclusion of a deterministic trend in the VECM data generating process implies that the growth rates $\Delta\boldsymbol{y}_t$ themselves grow or shrink over time. Clearly, a linear trend in growth rates will produce a quadratic trend in levels. Since time series which explode exponentially are rare, the inclusion of a time trend in the model itself (not in the cointegration space) is unlikely to be justified in practice. By contrast, the inclusion of a time trend in the cointegration space implies a linear (rather than quadratic) time trend in the levels of $\boldsymbol{y}_t$. Recall from the preceding paragraph that a linear trend in levels may also be accounted for by including a constant in the data generating process. Hence, if a time series includes a linear trend in levels so as to justify a constant term in the data generating process, then it may be necessary to consider a time trend in the cointegrating combination. Note, however, that it is not always necessary to include a linear trend in the cointegration space if there is a linear trend in the levels of $\boldsymbol{y}_t$. Variables exhibiting a linear trend which are (stochastically) cointegrated may in addition also be deterministically cointegrated in that the linear trend cancels out in the cointegrating combination [43].

2.3.4 Identification of the Cointegrating Vectors

Consider the VECM process

$$\Delta \boldsymbol{y}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{y}_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i \Delta \boldsymbol{y}_{t-i} + \boldsymbol{\Phi}\boldsymbol{D}_t + \boldsymbol{\varepsilon}_t$$

and suppose that $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$, where $\boldsymbol{\beta}$ is a $k \times r$ cointegration matrix and $\boldsymbol{\alpha}$ is a $k \times r$ matrix of equilibrium adjustment coefficients, also known as the *loading matrix*. Clearly, this decomposition of $\boldsymbol{\Pi}$ into $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}'$ is not unique. One could of course set $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\kappa}^{-1}\boldsymbol{\kappa}\boldsymbol{\beta}'$ for some non-singular $r \times r$ matrix $\boldsymbol{\kappa}$. In this case, the loading matrix is $\boldsymbol{\alpha}\boldsymbol{\kappa}^{-1}$, whilst the cointegration matrix becomes $\boldsymbol{\beta}\boldsymbol{\kappa}'$. Clearly, if $\boldsymbol{\beta}'\boldsymbol{y}_t \sim I(0)$, then $\boldsymbol{\kappa}\boldsymbol{\beta}'\boldsymbol{y}_t$ must also be stationary, confirming that $\boldsymbol{\beta}\boldsymbol{\kappa}'$ is also a cointegration matrix for $\boldsymbol{y}_t$. Hence, whilst the data may be used to isolate the cointegration space $sp(\boldsymbol{\beta})$, it does not ensure the uniqueness or economic relevance of the estimated cointegrating vectors. It therefore becomes necessary to impose identifying restrictions on the cointegration space [43].

In order to identify the cointegrating relations, define $\boldsymbol{R}_i$ as a $k \times g_i$ matrix of g_i linearly independent restrictions of the form

$$\boldsymbol{R}_i'\boldsymbol{\beta}_i = \boldsymbol{0} \qquad \text{for } i = 1, \ldots, r \tag{2.19}$$

where $\boldsymbol{\beta}_i$ denotes the ith column or cointegrating vector in $\boldsymbol{\beta}$. The number of linearly independent restrictions g_i is assumed to be strictly less than k, since for $g_i = k$ all the elements of $\boldsymbol{\beta}_i$ are constrained to zero implying that $\text{rank}(\boldsymbol{\beta}) < r$, whilst $g_i > k$ implies redundant restrictions in $\boldsymbol{R}_i$. The matrix $\boldsymbol{R}_i$ is therefore assumed to be of full column rank $g_i < k$ [43].

Now let $\boldsymbol{H}_i \equiv \boldsymbol{R}_i^{\perp}$ be a $k \times (k - g_i)$ matrix of full column rank $k - g_i$ such that $\boldsymbol{R}_i'\boldsymbol{H}_i = \boldsymbol{0}$ and

$$\boldsymbol{\beta}_i = \boldsymbol{H}_i\boldsymbol{\varphi}_i$$

for some $k - g_i$ dimensional vector $\boldsymbol{\varphi}_i$. The cointegrating vector $\boldsymbol{\beta}_i$ therefore belongs to the space orthogonal to $\boldsymbol{R}_i$. Since $\boldsymbol{\beta}_i$ is a k dimensional vector of freely varying parameters, imposing g_i restrictions reduces the number of free parameters to the $k - g_i$ included in $\boldsymbol{\varphi}_i$. The indirect parameterisation in Equation (2.19) may easily be retrieved from this last equation through premultiplication by $\boldsymbol{R}_i'$ [43].

A direct parameterisation of the cointegration matrix is given by

$$\boldsymbol{\beta} = (\boldsymbol{H}_1\boldsymbol{\varphi}_1, \ldots, \boldsymbol{H}_r\boldsymbol{\varphi}_r). \tag{2.20}$$

Letting $\boldsymbol{H} = (\boldsymbol{H}_1, \ldots, \boldsymbol{H}_r)$, it should be clear that the above conditions restrict the cointegration space to

$$sp(\boldsymbol{\beta}) \subset sp(\boldsymbol{H}).$$

Hence, whilst the restriction $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$ leads to the estimation of an r dimensional subspace chosen in $\mathcal{R}^k$, the restriction imposed by Equation (2.20) forces this subspace to lie in the given subspace $sp(\boldsymbol{H})$ of $\mathcal{R}^k$ [27].

To illustrate, consider the $I(1)$ vector process $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{5t})'$ and suppose economic theory posits two cointegrating relations between these variables $\varphi_1 y_{1t} - \varphi_1 y_{2t} - \varphi_1 y_{3t} = z_{1t}$ and $\varphi_2 y_{4t} - \varphi_2 y_{5t} = z_{2t}$, where z_{1t} and z_{2t} are $I(0)$ processes. Imposing these cointegrating relations would involve restricting the elements in the cointegration matrix as follows

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \\ \beta_{31} & \beta_{32} \\ \beta_{41} & \beta_{42} \\ \beta_{51} & \beta_{52} \end{bmatrix} = \begin{bmatrix} \varphi_1 & 0 \\ -\varphi_1 & 0 \\ -\varphi_1 & 0 \\ 0 & \varphi_2 \\ 0 & -\varphi_2 \end{bmatrix}, \tag{2.21}$$

such that

$$\boldsymbol{\beta}'\boldsymbol{y}_t = \begin{bmatrix} \varphi_1 & -\varphi_1 & -\varphi_1 & 0 & 0 \\ 0 & 0 & 0 & \varphi_2 & -\varphi_2 \end{bmatrix} \begin{pmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \\ y_{4t} \\ y_{5t} \end{pmatrix} = \begin{pmatrix} \varphi_1 y_{1t} - \varphi_1 y_{2t} - \varphi_1 y_{3t} \\ \varphi_2 y_{4t} - \varphi_2 y_{5t} \end{pmatrix}$$

is an $I(0)$ vector process. Now consider the cointegrating vector $\boldsymbol{\beta}_1$ corresponding to the first column of $\boldsymbol{\beta}$. In order to identify the first cointegrating relation, this vector is subject to the following set of $g_1 = 4$ linear restrictions

$$\begin{matrix} \boldsymbol{R}_1' & \boldsymbol{\beta}_1 & = & \boldsymbol{0} \\ \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} & \begin{pmatrix} \beta_{11} \\ \beta_{21} \\ \beta_{31} \\ \beta_{41} \\ \beta_{51} \end{pmatrix} & = & \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix} \end{matrix}.$$

In this case, $\boldsymbol{H}_1$ will be a five dimensional vector (since $k = 5$ and $g_1 = 4$) orthogonal to $\boldsymbol{R}_1$ such that

$$\begin{array}{ccccc} \boldsymbol{\beta}_1 & = & \boldsymbol{H}_1 & \boldsymbol{\varphi}_1 \\ \begin{pmatrix} \beta_{11} \\ \beta_{21} \\ \beta_{31} \\ \beta_{41} \\ \beta_{51} \end{pmatrix} & = & \begin{pmatrix} 1 \\ -1 \\ -1 \\ 0 \\ 0 \end{pmatrix} & \varphi_1. \end{array}$$

Clearly there is only one free scalar parameter φ_1 which requires estimation in this cointegrating relation, which may seem obvious from Equation (2.21).

Similarly, the second cointegrating vector $\boldsymbol{\beta}_2$ is subject to the following $g_2 = 4$ linear restrictions

$$\begin{array}{ccccc} \boldsymbol{R}_2' & & \boldsymbol{\beta}_2 & = & \boldsymbol{0} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} & & \begin{pmatrix} \beta_{12} \\ \beta_{22} \\ \beta_{32} \\ \beta_{42} \\ \beta_{52} \end{pmatrix} & = & \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix} \end{array}$$

and it therefore follows that

$$\begin{array}{ccccc} \boldsymbol{\beta}_2 & = & \boldsymbol{H}_2 & \boldsymbol{\varphi}_2 \\ \begin{pmatrix} \beta_{12} \\ \beta_{22} \\ \beta_{32} \\ \beta_{42} \\ \beta_{52} \end{pmatrix} & = & \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ -1 \end{pmatrix} & \varphi_2. \end{array}$$

As with the previous cointegrating vector, there is only a single free scalar parameter φ_2 which requires estimation in this cointegrating relation. Defining

$$\boldsymbol{H} = (\boldsymbol{H}_1, \boldsymbol{H}_2) = \begin{bmatrix} 1 & 0 \\ -1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 0 & -1 \end{bmatrix},$$

it should be clear that the restriction implied by Equation (2.21) ensures $sp(\boldsymbol{\beta}) \subset sp(\boldsymbol{H})$.

An obvious question at this point is how many linearly independent restrictions are necessary in order to uniquely identify the cointegrating relations. The answer is given by the so-called *rank condition* for the identification of cointegrating vectors [27]. More specifically, the rank condition states that provided there are no further identifying conditions on the parameters in the model (for example, the cointegrating vectors have been normalised appropriately), then $\boldsymbol{\beta}_i$ is identified if and only if the parameters satisfy the conditions

$$\text{rank}\,(\boldsymbol{R}_i'\boldsymbol{\beta}) = r - 1 \quad \text{for } i = 1, \ldots, r, \tag{2.22}$$

implying $\boldsymbol{R}_i'\boldsymbol{\beta}_i = \mathbf{0}$ for all i and $\boldsymbol{R}_i'\boldsymbol{\beta}_j \neq \mathbf{0}$ for all $i \neq j$. Since $\boldsymbol{R}_i'\boldsymbol{\beta}$ is a $g_i \times r$ dimensional matrix, the rank condition requires $g_i \geq r - 1$ for $i = 1, \ldots, r$, which implies that the cointegrating vectors are identified if there are at least $r - 1$ linear restrictions on each cointegrating vector [43]. Furthermore, since $\text{rank}(\boldsymbol{R}_i'\boldsymbol{\beta}_i) = \text{rank}(\mathbf{0}) = 0$, it follows that when the restrictions associated with the ith cointegrating vector are applied to the remaining $r - 1$ cointegrating vectors, the resulting matrix will have rank $r - 1$. Hence, it is impossible to create a vector which is restricted in the same way as $\boldsymbol{\beta}_i$ from a linear combination of the remaining $r-1$ cointegrating vectors. The ith cointegrating vector $\boldsymbol{\beta}_i$ may therefore be identified as the only vector among all linear combinations of $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_r$ that satisfies the restriction $\boldsymbol{R}_i$ or, equivalently, is in the space spanned by the columns of $\boldsymbol{H}_i$; that is, $\boldsymbol{\beta}_i \in sp(\boldsymbol{H}_i)$ [28].

However, the true parameter value $\boldsymbol{\beta}$ is not known in general and hence the rank condition as presented in Equation (2.22) is not operational in practice. Consequently, it would be far more useful to establish a set of conditions on the linear restrictions $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_r$ such that they uniquely identify the cointegrating vectors $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_r$. From the rank condition, it follows that the cointegrating relations are identified if

$$\text{rank}\,(\boldsymbol{R}_i'\boldsymbol{H}_1\boldsymbol{\varphi}_1, \ldots, \boldsymbol{R}_i'\boldsymbol{H}_r\boldsymbol{\varphi}_r) = r - 1 \quad \text{for } i = 1, \ldots, r$$

where the $r - 1$ column vectors $\{\boldsymbol{R}_i\boldsymbol{H}_j\boldsymbol{\varphi}_j,\ j = 1, \ldots, r,\ j \neq i\}$ are linearly independent. This result implies the following necessary and sufficient condition for a set of restrictions to be identifying as given by Johansen [28]:

$$\text{rank}(\boldsymbol{R}_i'\boldsymbol{H}_{i_1}, \ldots, \boldsymbol{R}_i'\boldsymbol{H}_{i_s}) \geq s$$

for $s = 1, \ldots, r - 1$ and any set of indices $1 \leq i_1 \leq \ldots \leq i_s \leq r$ excluding i. Note that this condition does not involve the parameters and is therefore immediately operational given a set of restrictions. For this reason, the cointegrating relations are said to be *generically identified*; that is, the statistical model satisfies the aforementioned set of algebraic conditions which ensure that the cointegrating vectors are distinguishable regardless of the parameter values [28].

In the example above,

$$\boldsymbol{R}_1'\boldsymbol{H}_2 = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \end{pmatrix}$$

and

$$\boldsymbol{R}_2'\boldsymbol{H}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{pmatrix} 1 \\ -1 \\ -1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ -1 \\ 0 \end{pmatrix}.$$

Hence $\text{rank}(\boldsymbol{R}_1'\boldsymbol{H}_2) = 1$ and $\text{rank}(\boldsymbol{R}_2'\boldsymbol{H}_1) = 1$ and therefore both cointegrating relations are generically identified by the chosen set of restrictions in this example.

Since the choice of $\boldsymbol{H}_1, \ldots, \boldsymbol{H}_r$ is not unique, the parameters $\boldsymbol{\varphi}_1, \ldots, \boldsymbol{\varphi}_r$ are, however, only determined up to a constant factor. Consequently, each cointegrating vector is typically normalised on a different variable by setting the coefficient on the normalising variable to one, where the normalising variable is usually chosen to aid economic interpretation. Such a normalisation is of the form

$$\boldsymbol{\beta}_i = \boldsymbol{h}_i + \boldsymbol{H}_i^*\boldsymbol{\psi}_i,$$

where $\boldsymbol{\psi}_i$ is a $k - g_i - 1$ dimensional vector of free parameters, with one less freely varying parameter relative to $\boldsymbol{\varphi}_i$ due to the normalisation on one of the variables. Clearly, $\boldsymbol{h}_i \in sp(\boldsymbol{H}_i)$ and $sp(\boldsymbol{h}_i, \boldsymbol{H}_i^*) = sp(\boldsymbol{H}_i)$. The normalisation may therefore be achieved by choosing $\boldsymbol{h}_i$ as a unit vector with $\boldsymbol{H}_i^*$ having zeros in the row corresponding to the unitary element in $\boldsymbol{h}_i$ [27].

In the example above, it should be noted that the choice of $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$ from all possible vectors in the set of orthogonal complements to $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ was essentially arbitrary. Indeed, any $\boldsymbol{H}_1^\bullet \in sp(\boldsymbol{H}_1)$ and $\boldsymbol{H}_2^\bullet \in sp(\boldsymbol{H}_2)$ (for example, $\boldsymbol{H}_1^\bullet = (2, -2, -2, 0, 0)'$ and $\boldsymbol{H}_2^\bullet = (0, 0, 0, -0.5, 0.5)'$) would be orthogonal to $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ respectively and could very well be chosen instead. A normalisation is therefore necessary in order to uniquely determine the parameters φ_1 and φ_2. For example, normalising on y_{1t} in the first cointegrating relation and y_{4t} in the second cointegrating relation involves setting $\beta_{11} = 1$ and $\beta_{42} = 1$. Given the choice of $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$, doing so is equivalent to setting $\varphi_1 = 1$ and $\varphi_2 = 1$. Clearly, if $\boldsymbol{H}_1^\bullet$ and $\boldsymbol{H}_2^\bullet$ were chosen instead of $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$, the same normalisation would imply $\varphi_1 = 0.5$ and $\varphi_2 = -2$. In both cases, there are no free parameters in the cointegrating relations after normalisation and the imposition of the specified restrictions.

Whilst generic identification is concerned with the correct model specification such that the parameters may be uniquely identified, Johansen and Juselius suggest replacing the requirement of correct model specification with the assumption of a reasonably well-structured model which is consistent with several theories [31]. Consequently, they propose that the concept of identification be broadened to include the notions of *empirical* and *economic* identification, in addition to generic identification. Whilst empirical identification focuses on the estimated parameters and whether the data supports the generic restrictions, economic identification relates to the interpretability of the estimated parameter values in light of relevant economic theory. Hence, whilst the mathematical condition above is necessary and sufficient for the generic identification of the cointegrating relations, it does not guarantee that the model is either empirically or economically identified [31].

2.3.5 Testing Linear Restrictions on the Cointegrating Vectors

The rank condition given by Equation (2.22) implies that at least $r-1$ linear restrictions must be imposed on each cointegrating vector for the long-run relations to be generically identified. The case where exactly $r-1$ restrictions are placed on each column in $\boldsymbol{\beta}$, that is, $\boldsymbol{R}_i$ is a $k \times (r-1)$ matrix for $i = 1, \ldots, r$, is referred to as *exact* or *just identification*, since it involves the minimum number of restrictions required for identification. By contrast, when more than $r-1$ restrictions are imposed on each cointegrating vector, the restrictions are said to be *overidentifying*, since they are not required in order to uniquely determine the cointegrating vectors in the generic sense. In the example above, four restrictions were imposed on each of the two cointegrating relations when in fact only one linear restriction on each is necessary for just identification. Consequently, both sets of linear restrictions imposed on the cointegrating vectors in that example were overidentifying. Such additional restrictions are usually based on theoretical considerations and it would therefore seem appropriate to test whether these restrictions are in fact consistent with the data [43].

Testing the overidentifying restrictions may be achieved by means of a standard hypothesis test of

$$\begin{aligned} H_0 &: \text{Overidentified model} \\ H_1 &: \text{Exactly identified model} \end{aligned}$$

by comparing the likelihood of the overidentified model to that of the just identified model. The asymptotic distribution of the appropriate likelihood ratio statistic for this test can be shown to be chi-squared with v degrees of freedom equal to the number of overidentifying restrictions, provided $\boldsymbol{\beta}$ is identified. Noting that the maximum number of freely varying coefficients in each cointegrating vector equals $k - (r - 1)$ and letting s_i denote the actual number of freely estimated coefficients in the ith cointegrating vector, the total number of overidentifying restrictions relative to the number required for exact identification is

$$v = \sum_{i=1}^{r} \left[k - (r - 1) - s_i \right].$$

A statistically significant χ^2_v value would suggest that the additional restrictions are not supported by the data, whilst non-rejection of the null hypothesis would imply that the overidentifying restrictions are consistent with the data. Note that an nonsignificant test statistic does not imply *acceptance* of the null hypothesis since it is possible to formulate several different sets of overidentifying restrictions. The acceptability of a particular set will depend upon whether the restrictions lead to an economically identified structure [43].

Additionally, it should be noted that overidentifying restrictions may result in a model which is no longer generically identified. In this case, if the χ^2_v test statistic is found to be nonsignificant, one could conclude that the economic structure imposed is not supported by the data. Hence, even though the economic structure can be described by a generically identified statistical model, the data may suggest that this structure cannot be empirically identified [31].

2.3.6 Partial Models and Exogeneity

Economic systems often comprise of many interrelated variables and modelling such multivariate time series may therefore require the estimation of complex systems of equations of relatively high dimensionality. Consequently, it is often tempting to consider partially specified systems, where only some of the variables are treated as endogenous conditional on the remaining variables [25].

To this effect, consider the full VECM

$$\Delta \boldsymbol{y}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{y}_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i \Delta \boldsymbol{y}_{t-i} + \boldsymbol{\Phi}\boldsymbol{D}_t + \boldsymbol{\varepsilon}_t \tag{2.23}$$

where $\boldsymbol{D}_t$ includes only deterministic terms such as an intercept, trend and seasonal dummy variables and $\boldsymbol{\varepsilon}_t \overset{iid}{\sim} \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Omega})$. Now suppose the k dimensional vector process $\Delta \boldsymbol{y}_t$ is decomposed into the vectors $\Delta \boldsymbol{y}_{1t}$ and $\Delta \boldsymbol{y}_{2t}$ of dimensions k_1 and k_2 respectively, where $k_1 + k_2 = k$. Such a decomposition may be achieved, for example, by defining the $k \times k_1$ matrix $\boldsymbol{a} = (\boldsymbol{I}_{k_1}, \boldsymbol{0})'$ such that $\boldsymbol{a}'\Delta \boldsymbol{y}_t = \Delta \boldsymbol{y}_{1t}$ and $\boldsymbol{a}_\perp' \Delta \boldsymbol{y}_t = \Delta \boldsymbol{y}_{2t}$, where $\boldsymbol{I}_{k_1}$ is a k_1 dimensional identity matrix and $\boldsymbol{a}_\perp$ is a $k \times k_2$ matrix of full column rank orthogonal to $\boldsymbol{a}$ [25]. This choice of $\boldsymbol{a}$ implies that $\Delta \boldsymbol{y}_{1t}$ includes the first k_1 variables in $\Delta \boldsymbol{y}_t$ and $\Delta \boldsymbol{y}_{2t}$ therefore comprises of the remaining $k - k_1 = k_2$ variables, although clearly $\boldsymbol{a}$ could be constructed such that $\Delta \boldsymbol{y}_{1t} = \boldsymbol{a}'\Delta \boldsymbol{y}_t$ may include any set of k_1 variables from $\Delta \boldsymbol{y}_t$. Let the model parameters $(\boldsymbol{\alpha}, \boldsymbol{\Gamma}_1, \ldots, \boldsymbol{\Gamma}_{p-1}, \boldsymbol{\Phi})$ be decomposed correspondingly with $(\boldsymbol{\alpha}_1, \boldsymbol{\Gamma}_{1,1}, \ldots, \boldsymbol{\Gamma}_{1,p-1}, \boldsymbol{\Phi}_1) = \boldsymbol{a}'(\boldsymbol{\alpha}, \boldsymbol{\Gamma}_1, \ldots, \boldsymbol{\Gamma}_{p-1}, \boldsymbol{\Phi})$ and $(\boldsymbol{\alpha}_2, \boldsymbol{\Gamma}_{2,1}, \ldots, \boldsymbol{\Gamma}_{2,p-1}, \boldsymbol{\Phi}_2) = \boldsymbol{a}_\perp'(\boldsymbol{\alpha}, \boldsymbol{\Gamma}_1, \ldots, \boldsymbol{\Gamma}_{p-1}, \boldsymbol{\Phi})$. Similarly, define $\boldsymbol{\varepsilon}_{1t} = \boldsymbol{a}'\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\varepsilon}_{2t} = \boldsymbol{a}_\perp'\boldsymbol{\varepsilon}_t$ with variance-covariance matrices $\boldsymbol{\Omega}_{11} = \boldsymbol{a}'\boldsymbol{\Omega}\boldsymbol{a}$ and $\boldsymbol{\Omega}_{22} = \boldsymbol{a}_\perp'\boldsymbol{\Omega}\boldsymbol{a}_\perp$ respectively and $\mathrm{Var}[\boldsymbol{\varepsilon}_{1t}, \boldsymbol{\varepsilon}_{2t}] = \boldsymbol{\Omega}_{12} = \boldsymbol{a}'\boldsymbol{\Omega}\boldsymbol{a}_\perp$. Then the VECM given as Equation (2.23) may be decomposed into a partial or *conditional* model for $\Delta \boldsymbol{y}_{1t}$ given $\Delta \boldsymbol{y}_{2t}$ and past information as follows

$$\begin{aligned}
\boldsymbol{a}'\Delta \boldsymbol{y}_t &= \boldsymbol{\omega}\boldsymbol{a}_\perp'\Delta \boldsymbol{y}_t + (\boldsymbol{a}' - \boldsymbol{\omega}\boldsymbol{a}_\perp')\Delta \boldsymbol{y}_t \\
&= \boldsymbol{\omega}\boldsymbol{a}_\perp'\Delta \boldsymbol{y}_t + (\boldsymbol{a}' - \boldsymbol{\omega}\boldsymbol{a}_\perp')\Big(\boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{y}_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i \Delta \boldsymbol{y}_{t-i} + \boldsymbol{\Phi}\boldsymbol{D}_t + \boldsymbol{\varepsilon}_t\Big) \\
\Delta \boldsymbol{y}_{1t} &= \boldsymbol{\omega}\Delta \boldsymbol{y}_{2t} + (\boldsymbol{\alpha}_1 - \boldsymbol{\omega}\boldsymbol{\alpha}_2)\boldsymbol{\beta}'\boldsymbol{y}_{t-1} + \sum_{i=1}^{p-1} (\boldsymbol{\Gamma}_{1,i} - \boldsymbol{\omega}\boldsymbol{\Gamma}_{2,i})\Delta \boldsymbol{y}_{t-i} \\
&\quad + (\boldsymbol{\Phi}_1 - \boldsymbol{\omega}\boldsymbol{\Phi}_2)\boldsymbol{D}_t + \boldsymbol{\varepsilon}_{1t} - \boldsymbol{\omega}\boldsymbol{\varepsilon}_{2t}
\end{aligned} \tag{2.24}$$

where $\boldsymbol{\omega} = \boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22}^{-1}$, together with a *marginal* model for $\boldsymbol{y}_{2t}$ given as

$$\boldsymbol{a}'_{\perp}\Delta\boldsymbol{y}_t = \boldsymbol{a}'_{\perp}\Big(\boldsymbol{\alpha\beta}'\boldsymbol{y}_{t-1} + \sum_{i=1}^{p-1}\boldsymbol{\Gamma}_i\Delta\boldsymbol{y}_{t-i} + \boldsymbol{\Phi D}_t + \boldsymbol{\varepsilon}_t\Big)$$
$$\Delta\boldsymbol{y}_{2t} = \boldsymbol{\alpha}_2\boldsymbol{\beta}'\boldsymbol{y}_{t-1} + \sum_{i=1}^{p-1}\boldsymbol{\Gamma}_{2,i}\Delta\boldsymbol{y}_{t-i} + \boldsymbol{\Phi}_2\boldsymbol{D}_t + \boldsymbol{\varepsilon}_{2t}. \tag{2.25}$$

Notice that the cointegrating relations $\boldsymbol{\beta}'\boldsymbol{y}_{t-1}$ feature in both the conditional and marginal models such that both $\boldsymbol{y}_{1t}$ and $\boldsymbol{y}_{2t}$ will in general include information relevant to the estimation of the cointegrating coefficients. Furthermore, it is observed that whilst the adjustment coefficients corresponding to $\Delta\boldsymbol{y}_{1t}$ were defined as $\boldsymbol{\alpha}_1$ in the full model given by Equation (2.23), the loading matrix in Equation (2.24) for the conditional model of $\Delta\boldsymbol{y}_{1t}$ given $\Delta\boldsymbol{y}_{2t}$ is $\boldsymbol{\alpha}_1 - \boldsymbol{\omega\alpha}_2$. Inferences regarding the manner in which $\boldsymbol{y}_{1t}$ adjusts to departures from the equilibrium relations $\boldsymbol{\beta}'\boldsymbol{y}_{t-1}$ may therefore differ depending on whether the full or conditional model is considered. In general, these interrelations between the parameters of the marginal and conditional models imply that a full system analysis is necessary in order to draw inferences which are efficient in the sense that all the available information included in the data is utilised [26].

There is, however, one situation in which the conditional model includes as much information as the full system about the cointegrating relations and adjustment coefficients, such that an analysis of the partial model given in Equation (2.24) will be efficient with respect to inferences on $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. An efficient analysis of the partial model is possible if the rows of $\boldsymbol{\alpha}$ corresponding to the $\Delta\boldsymbol{y}_{2t}$ equations in the full model are zero, in which case $\boldsymbol{\alpha}_2 = \boldsymbol{0}$ and the conditional and marginal models given as Equations (2.24) and (2.25) reduce to

$$\Delta\boldsymbol{y}_{1t} = \boldsymbol{\omega}\Delta\boldsymbol{y}_{2t} + \boldsymbol{\alpha}_1\boldsymbol{\beta}'\boldsymbol{y}_{t-1} + \sum_{i=1}^{p-1}(\boldsymbol{\Gamma}_{1,i} - \boldsymbol{\omega\Gamma}_{2,i})\Delta\boldsymbol{y}_{t-i} + (\boldsymbol{\Phi}_1 - \boldsymbol{\omega\Phi}_2)\boldsymbol{D}_t + \boldsymbol{\varepsilon}_{1t} - \boldsymbol{\omega\varepsilon}_{2t} \tag{2.26}$$

and

$$\Delta\boldsymbol{y}_{2t} = \sum_{i=1}^{p-1}\boldsymbol{\Gamma}_{2,i}\Delta\boldsymbol{y}_{t-i} + \boldsymbol{\Phi}_2\boldsymbol{D}_t + \boldsymbol{\varepsilon}_{2t} \tag{2.27}$$

respectively. It should be clear from this formulation that the adjustment coefficients for $\Delta\boldsymbol{y}_{1t}$ are the same in both the full model and the conditional model given by Equation (2.26) where $\boldsymbol{\alpha}_2 = \mathbf{0}$. In addition, the cointegrating relations fall away in Equation (2.27) and the marginal model is therefore not relevant to the estimation of $\boldsymbol{\beta}$ when $\boldsymbol{\alpha}_2 = \mathbf{0}$. Hence, the maximum likelihood estimator for $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ in the full model given by Equation (2.23) will be the same as the maximum partial likelihood estimator in the conditional model presented as Equation (2.26). If this condition holds, the variables in $\Delta \boldsymbol{y}_{2t}$ are said to be *weakly exogenous* for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ [25]. It is worth noting at this point that the notion of weak exogeneity depends on the explicit choice of parameters of interest [26]. Hence, although $\Delta \boldsymbol{y}_{2t}$ is weakly exogenous for $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ in the VECM given by Equation (2.23) with $\boldsymbol{\alpha}_2 = \mathbf{0}$, it is obvious upon examination of Equations (2.26) and (2.27) that $\Delta \boldsymbol{y}_{2t}$ is *not* weakly exogenous for the parameters $\boldsymbol{\Gamma}_1, \ldots, \boldsymbol{\Gamma}_{p-1}$ and $\boldsymbol{\Phi}$ in this same model.

The weak exogeneity of $\Delta \boldsymbol{y}_{2t}$ for $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ implies that $\Delta \boldsymbol{y}_{2t}$ does not react to disequilibria in the cointegrating relations, but may still react to lagged changes of $\boldsymbol{y}_{1t}$ as is evident from the marginal model in Equation (2.27). If the lagged changes of $\boldsymbol{y}_{1t}$ include information relevant for forecasting $\Delta \boldsymbol{y}_{2t}$ that is not present in any other covariates, then $\Delta \boldsymbol{y}_{1t}$ is said to *Granger cause* $\Delta \boldsymbol{y}_{2t}$ [17]. If the columns of $\boldsymbol{\Gamma}_{2,i}$ corresponding to $\Delta \boldsymbol{y}_{1,t-i}$ in the marginal model are equal to zero for $i = 1, \ldots, p-1$ so that $\Delta \boldsymbol{y}_{1t}$ does not Granger cause $\Delta \boldsymbol{y}_{2t}$, then $\Delta \boldsymbol{y}_{2t}$ is said to be *strongly exogenous* for $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ [26].

Testing for the weak exogeneity of $\Delta \boldsymbol{y}_{2t}$ with respect to $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ in the VECM given by Equation (2.23) is equivalent to testing the following set of linear restrictions on the loading matrix $\boldsymbol{\alpha}$

$$
\begin{aligned}
H_0 &: \boldsymbol{a}_\perp' \boldsymbol{\alpha} = \mathbf{0} \;\Leftrightarrow\; \boldsymbol{\alpha}_2 = \mathbf{0} \\
H_1 &: \boldsymbol{a}_\perp' \boldsymbol{\alpha} \neq \mathbf{0} \;\Leftrightarrow\; \boldsymbol{\alpha}_2 \neq \mathbf{0},
\end{aligned}
$$

where weak exogeneity is implied under the null hypothesis. The test for strong exogeneity is defined analogously by imposing additional zero restrictions on the columns of $\boldsymbol{\Gamma}_{2,i}$ corresponding to $\Delta \boldsymbol{y}_{1,t-i}$ for $i = 1, \ldots, p-1$ and testing the jointly constrained $\boldsymbol{\alpha}_2$ and $\boldsymbol{\Gamma}_{2,i}$ matrices for data admissibility. In both cases, the restrictions on the coefficient matrices can be tested by means of a standard likelihood ratio or Wald test comparing the restricted or partial model with the unrestricted, full model [29].

2.4 The Markov-Switching VAR Model

The Markov-switching vector autoregressive model, denoted as MS-VAR for short, may be regarded as an extension of a VAR(p) process to time series which are subject to shifts in regime. In such instances, the standard VAR model with its time invariant parameters is likely to be inappropriate. The basic idea behind this class of regime-switching models is that the parameters of the underlying data generating process of the *observed* multivariate time series $\boldsymbol{y}_t$ depend upon an *unobserved* regime variable s_t. Since the regimes are unobserved, a model for the regime generating process must first be formulated such that the observed data may be used to draw inferences about the state of the world at each time point [35].

2.4.1 The Regime Generating Process

Markov-switching models assume that the evolution of the regime variable $s_t \in \{1, \ldots, m\}$ can be described by a discrete time, discrete state Markov stochastic process with transition probabilities

$$p_{ij} = \Pr[s_{t+1} = j | s_t = i], \quad \sum_{j=1}^{m} p_{ij} = 1 \quad \text{for all } i, j \in \{1, \ldots, m\}.$$

In particular, it is assumed that s_t follows an irreducible, ergodic and finite Markov process with transition probability matrix

$$\mathbb{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{bmatrix}.$$

The defining properties of any transition probability matrix are that all its elements are non-negative and, since the system must move to some state from any state i, all rows sum to unity such that $\mathbb{P}\mathbf{1} = \mathbf{1}$, where $\mathbf{1} = (1, \ldots, 1)'$ is an m dimensional vector [7]. The assumptions of ergodicity and irreducibility are fundamental to the theoretical properties of MS-VAR models and are therefore considered formally in Appendix B.

2.4.2 The Data Generating Process

The most general specification of a Markov-switching vector autoregressive process of order p with m regimes is given by

$$\boldsymbol{y}_t = \boldsymbol{\nu}(s_t) + \boldsymbol{\Pi}_1(s_t)\boldsymbol{y}_{t-1} + \boldsymbol{\Pi}_2(s_t)\boldsymbol{y}_{t-2} + \ldots + \boldsymbol{\Pi}_p(s_t)\boldsymbol{y}_{t-p} + \boldsymbol{\varepsilon}_t, \qquad (2.28)$$

where $\boldsymbol{\varepsilon}_t|s_t \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Omega}(s_t)\right)$ and $\boldsymbol{\nu}(s_t), \boldsymbol{\Pi}_1(s_t), \ldots, \boldsymbol{\Pi}_p(s_t), \boldsymbol{\Omega}(s_t)$ reflect the dependence of the parameters on the realised regime $s_t \in \{1, \ldots, m\}$.

The model given by Equation (2.28) specifies that all parameters of the stochastic process are conditional on the state s_t of the Markov chain. Of course, it may be the case that only some of the parameters are conditioned on the state, whilst others may be regime invariant. In order to establish a unique notation for each model, it is conventional to specify the regime dependent parameters in the acronym describing the model. For example, the aforementioned model is an MSIAH(m)-VAR(p) process, revealing the dependence of the intercept (I), autoregressive coefficients (A) and error variance structure (H for heteroscedasticity) on the underlying regime [34].

As with the linear VAR model, an MS-VAR process may be reparameterised as a Markov-switching vector error correction model of the form

$$\Delta\boldsymbol{y}_t = \boldsymbol{\nu}(s_t) + \boldsymbol{\alpha}(s_t)\boldsymbol{\beta}'\boldsymbol{y}_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i(s_t)\Delta\boldsymbol{y}_{t-i} + \boldsymbol{\varepsilon}_t, \qquad (2.29)$$

where all the parameters are as previously defined and $\boldsymbol{\varepsilon}_t|s_t \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Omega}(s_t)\right)$. Note that the regime dependent intercept in the above model may be decomposed as $\boldsymbol{\nu}(s_t) = \boldsymbol{\alpha}(s_t)\boldsymbol{\nu}_1(s_t) + \boldsymbol{\alpha}_\perp(s_t)\boldsymbol{\nu}_2(s_t)$, provided the intercept term is not restricted to the cointegration space. Hence, this model allows for regime shifts in the intercept term $\boldsymbol{\nu}_1$ in the cointegrating relations, whilst the cointegrating coefficients in $\boldsymbol{\beta}$ are assumed constant across regimes. The error correction coefficients $\boldsymbol{\alpha}$ may also be considered regime dependent permitting different adjustment dynamics between regimes.

Krolzig justifies a two stage procedure for the estimation of an MS-VECM where the regime generating process is described by an irreducible, ergodic Markov chain [33]. The first stage involves the estimation and identification of the cointegrating relations over the entire time period under analysis following the methods discussed above. The regime generating process is not considered at this point. Then, conditional on the cointegration matrix $\boldsymbol{\beta}$, the remaining parameters are estimated treating $\boldsymbol{z}_{t-1} = \boldsymbol{\beta}'\boldsymbol{y}_{t-1}$ as a vector of exogenous $I(0)$ variables [36]. In addition to the estimation of the model parameters in Equation (2.29), the probability $\Pr[s_t = i|\boldsymbol{Y}_T]$ of being in regime $i = 1, \ldots, m$ at each time $t = 1, \ldots, T$ is also estimated based on the entire dataset $\boldsymbol{Y}_T$. These estimated probabilities are referred to as *smoothed probabilities* since they incorporate all the available information in the data and the estimated regime s_t at time t is that regime with the highest smoothed probability at time t for $t = 1, \ldots, T$ [34]. The estimation of the model parameters and the smoothed probabilities in this second stage is accomplished via the EM algorithm outlined in Appendix C.

2.5 Economic Theory

2.5.1 Purchasing Power Parity

Let P_t^{SA} and P_t^{US} denote the prices of a reference basket of tradable commodities in South Africa and the United States at time t respectively. Then, in the absence of transactions costs and barriers to trade, the absolute version of purchasing power parity (PPP) states that the nominal rand/dollar exchange rate $E_t^{R/\$}$ must equal the ratio of the countries' price levels; that is,

$$E_t^{R/\$} = \frac{P_t^{SA}}{P_t^{US}}. \tag{2.30}$$

Alternatively, PPP asserts that the price of the basket of goods will be equal across the two countries when measured in a common currency, such that

$$P_t^{SA} = E_t^{R/\$} \times P_t^{US}.$$

A departure from this equilibrium condition would lead to an arbitrage opportunity in the commodity market and the fulfillment thereof would put pressure on the exchange rate to enforce parity in the national price levels. Hence, deviations from PPP should be self-correcting [32].

However, the absolute version of PPP may only be expected to hold when the two baskets whose prices are compared in Equation (2.30) are identical and include only tradable goods which are subject to international arbitrage. Since the coverage and composition of a country's basket is chosen to reflect the consumption specific to that country, the absolute version of the PPP theory therefore cannot be immediately assessed by comparing the national price levels across two countries [43]. Furthermore, even if an internationally standardised basket of tradable goods is available for the countries under consideration, information disparities, transactions costs and the effects of tariff and non-tariff trade barriers are likely to result in considerable deviations from absolute PPP in the short run. Such market imperfections, with the likely exception of information disparities, may persist indefinitely making absolute PPP a somewhat elusive concept [16].

Suppose, however, that the short-run deviations from absolute PPP due to transactions costs, information disparities and the effects of tariff and non-tariff trade barriers form a stationary stochastic process, say Q_t, such that

$$E_t^{R/\$} = Q_t \times \frac{P_t^{SA}}{P_t^{US}}. \tag{2.31}$$

In this case, percentage changes in relative price levels may still approximate percentage changes in the exchange rate provided that the factors causing deviations from absolute PPP are relatively stable over time. Equation (2.31) is known as the relative version of purchasing power parity, which converts the absolute version from a statement about price and exchange rate *levels* into one about price and exchange rate *changes*. This weaker form of PPP requires only that Q_t is stationary with some constant mean reflecting the average departure from absolute PPP due to market imperfections and differences in the coverage and composition of the commodity baskets under comparison. In contrast, Q_t must have a unit mean in order for absolute PPP to hold. Consequently, relative PPP may be valid even when absolute PPP is not [51].

Absolute PPP as presented in Equation (2.30) is related to the monetary approach to exchange rate determination, which posits that national price levels are fully determined by the supply and (real) demand for money in the long run. Changes in the interest rate and output levels affect the exchange rate only through their influences on money demand. As a long-run theory of exchange rate determination, relative PPP also allows for the impact of changes in money supply and demand on the exchange rate, but corrects the monetary approach by allowing for nonmonetary factors through the additional term Q_t. By rearranging Equation (2.31), it is easy to verify that

$$Q_t = \frac{E_t^{R/\$} \times P_t^{US}}{P_t^{SA}}, \tag{2.32}$$

which is the price of the American commodity basket expressed in rands relative to the price of the South African commodity basket. This quantity is referred to as the *real* exchange rate to differentiate it from the *nominal* exchange rate $E_t^{R/\$}$ at which currency is traded [37].

Taking logarithms on both sides of Equation (2.32) and using lower case letters to denote the variables in logarithmic form yields the following expression for the relative PPP condition

$$p_t^{SA} - p_t^{US} - e_t^{R/\$} = -q_t. \tag{2.33}$$

Now, in order for relative PPP to hold, q_t must be a stationary process. However, the variables on the left hand side of Equation (2.33) cannot be assumed to be stationary in practice. At first glance, this may seem to refute the relative PPP hypothesis. Suppose, however, that p_t^{SA}, p_t^{US} and $e_t^{R/\$}$ are $I(1)$ processes. Then relative PPP may still hold if the linear combination of these processes on the left hand side of Equation (2.33) results in a cointegrated process. This condition would imply that the real exchange rate is stationary as required.

The above argument suggests the following testing strategy for this strict form of relative PPP. First, determine whether p_t^{SA}, p_t^{US} and $e_t^{R/\$}$ are all $I(1)$ processes. If this condition holds, then proceed to test whether the vector $\boldsymbol{\beta} = (1, -1, -1)'$ is a cointegrating vector for the $I(1)$ vector process $\boldsymbol{y}_t = (p_t^{SA}, p_t^{US}, e_t^{R/\$})'$; that is, determine whether $q_t \sim I(0)$. If the cointegrating vector is supported by the data such that the real exchange rate is stationary, the evidence is supportive of the view that there is a long-run tendency for relative PPP to hold. On the other hand, if the real exchange rate is $I(1)$, it will exhibit random walk behaviour which suggests that deviations from PPP are persistent, rather than self-correcting [43].

Letting $-q_t = -\mu^{(1)} + \xi_t^{(1)}$, where $\mu^{(1)}$ is the mean natural logarithm of the real exchange rate and $\xi_t^{(1)}$ is a white noise process, yields an alternative, but synonymous, expression for the relative PPP hypothesis

$$p_t^{SA} - p_t^{US} - e_t^{R/\$} + \mu^{(1)} = \xi_t^{(1)}. \tag{2.34}$$

Clearly, this formulation is equivalent to that presented above, except in that the intercept in the cointegration space is now made explicit.

Assessing the validity of PPP as a determinant of exchange rates in the long run has been the focus of many empirically oriented studies. Surprisingly, the empirical evidence in support of the cointegrating vectors necessary to establish both the absolute and relative versions of PPP is somewhat scarce. Many explanations for this lack of evidence have been proposed in the literature. One possible explanation that was mentioned earlier is the inclusion of non-tradable goods in commodity baskets and the presence of trade barriers that are high enough to inhibit international arbitrage [50]. Another potential reason is that the price elasticity of demand for goods and services may vary substantially between countries, leading firms to set market-specific prices, practicing so-called "pricing to market" [1]. Additionally, monopolistic and oligopolistic practices may interact with transactions costs and trade barriers to further weaken the link between prices of similar goods sold in different countries [42]. As a result, empirical studies often fail to establish $\boldsymbol{\beta} = (1, -1, -1)'$ as a cointegrating vector for $\boldsymbol{y}_t = (p_t^{SA}, p_t^{US}, e_t^{R/\$})'$. However, a non-stationary real exchange rate does not preclude a long-run relationship between relative prices and the nominal exchange rate. Such a long-run relationship may still exist, although not the one-for-one relationship implicit in the calculation of the real exchange rate [50]. It may therefore be more viable to test for a weaker form of relative PPP such as

$$p_t^{SA} - \beta_1^{(1)} p_t^{US} - \beta_2^{(1)} e_t^{R/\$} + \mu^{(1)} = \xi_t^{(1)},$$

where $\beta_1^{(1)}$ and $\beta_2^{(1)}$ are close to one. Departures in these coefficients from unity may be attributable to the aforementioned market frictions [48]. Note that

$$\mu^{(1)} = \mathrm{E}\left[\ln \frac{\left(P_t^{US}\right)^{\beta_1^{(1)}} \times \left(E_t^{R/\$}\right)^{\beta_2^{(1)}}}{P_t^{SA}}\right]$$

in this case. Consequently, $\mu^{(1)}$ will only approximate the mean real exchange rate in logarithmic form when $\beta_1^{(1)}$ and $\beta_2^{(1)}$ are not exactly equal to one, provided these coefficients are at least positive.

Johansen and Juselius suggest that the failure of many of the earlier studies of PPP may be ascribed to the omission of important short-run determinants of the exchange rate in the modelling procedure [30]. In particular, arbitrage opportunities in the asset market are also likely to affect the exchange rate and, under the assumption of market efficiency, the asset market is likely to clear much faster than the goods market in which participants are bound by contracts. Furthermore, arbitrage is more costly in the goods market relative to the asset market. Consequently, it is reasonable to assume that exchange rates are affected by short-run fluctuations arising from highly volatile asset markets and by long-run effects from interrelated goods markets. In addition, the influence of commodity prices on the exchange rate cannot be ignored, particularly in a resource-based economy such as South Africa.

2.5.2 Uncovered Interest Parity

Just as the PPP condition in Equation (2.30) implies equilibrium in the goods market in the sense that no arbitrage opportunities prevail, the uncovered interest parity (UIP) ensures equilibrium in the asset market. This latter theory states that the expected returns on deposits of any two currencies should be equal when measured in a common currency. Let r_t^{SA} and r_t^{US} be the rand and dollar interest rates at time t respectively. Then the UIP condition may be stated mathematically as

$$r_t^{SA} = r_t^{US} + \frac{E_{t+1}^{R/\$} - E_t^{R/\$}}{E_t^{R/\$}} + r_t^{US} \times \frac{E_{t+1}^{R/\$} - E_t^{R/\$}}{E_t^{R/\$}}. \tag{2.35}$$

The last term on the right hand side is usually negligibly small and is therefore typically ignored, such that the UIP condition reduces to

$$r_t^{SA} = r_t^{US} + \frac{E_{t+1}^{R/\$} - E_t^{R/\$}}{E_t^{R/\$}}. \tag{2.36}$$

Since the future nominal exchange rate $E_{t+1}^{R/\$}$ cannot be known in advance, it is usually replaced by its expected value. The right hand side of Equation (2.36) implies that the expected rand rate of return on dollar deposits is approximately equal to the dollar interest rate plus the rate of depreciation in the rand/dollar exchange rate. For UIP to hold, this rand rate of return on dollar deposits must equal the rate of return on rand deposits, such that no type of deposit is in excess demand or excess supply. In the event that this equality does not hold, an arbitrage opportunity will exist in the asset market and the exchange rate will be forced to adjust accordingly as profit-seeking investors take advantage of it. As with PPP, departures from UIP should therefore also be self-correcting, albeit at a faster pace than PPP [37].

Most empirical works on UIP have, however, not been able to verify this relation as an immediate market clearing mechanism. Instead, empirical evidence seems to suggest that UIP holds as a long-run relation [30]. The presence of transactions costs, risk premia and speculative effects are all possible explanations for the observed short-run deviations from UIP [16]. Now suppose these effects are captured in the stationary process $u_t = \mu^{(2)} + \xi_t^{(2)}$, where $\mu^{(2)}$ is the mean of these effects and $\xi_t^{(2)}$ is white noise. Then, the long-run, steady state solution implied by UIP is given as

$$r_t^{SA} - r_t^{US} - \mu^{(2)} = \xi_t^{(2)}. \tag{2.37}$$

Note that the expected depreciation in the exchange rate from Equation (2.36) falls away in this expression since the exchange rate is assumed constant in the steady state [30].

In order for UIP to hold, it must be the case that $\xi_t^{(2)}$ is stationary. However, interest rates are typically $I(1)$ processes when examined over a specified time period, despite being implicitly bound between zero and some, usually small, positive number. It therefore follows that $r_t^{SA} - r_t^{US}$ must be a cointegrated process. Hence, it is necessary to establish that $\boldsymbol{\beta} = (1, -1)'$ is a cointegrating vector for the $I(1)$ vector process $\boldsymbol{y}_t = (r_t^{SA}, r_t^{US})'$ such that the linear combination $\boldsymbol{\beta}'\boldsymbol{y}_t$ is stationary. If the data support this proposition, then this would constitute evidence in favour of UIP as a long-run relation.

Note that the UIP condition given as Equation (2.37) was derived from Equation (2.35) by assuming that the last term on the right hand side of the latter equation is a zero mean process. This assumption is, however, only valid if the exchange rate does not exhibit trending behaviour over the time period considered. Rewriting Equation (2.35) as

$$r_t^{SA} = \Big(1 + \frac{E_{t+1}^{R/\$} - E_t^{R/\$}}{E_t^{R/\$}}\Big) r_t^{US} + \frac{E_{t+1}^{R/\$} - E_t^{R/\$}}{E_t^{R/\$}},$$

it should be clear that the coefficient on r_t^{US} will not be unitary if the exchange rate is non-stationary in mean, since the expected value of $E_{t+1}^{R/\$} - E_t^{R/\$}$ will be non-zero. In such a case, it may be more reasonable to consider a weaker form of UIP given by

$$r_t^{SA} - \beta_1^{(2)} r_t^{US} - \mu^{(2)} = \xi_t^{(2)},$$

where

$$\beta_1^{(2)} = \Big(1 + \frac{E_{t+1}^{R/\$} - E_t^{R/\$}}{E_t^{R/\$}}\Big)$$

may deviate from one due to a trending exchange rate and $\xi_t^{(2)}$ is a white noise process as before.

Note that

$$\mu^{(2)} = \mathrm{E}\left[u_t + \frac{E_{t+1}^{R/\$} - E_t^{R/\$}}{E_t^{R/\$}}\right]$$

in this formulation, where u_t is a stochastic process which captures the short-run deviations from UIP due to market imperfections.

Testing for PPP and UIP individually as presented above implicity ignores the substantive impact of capital markets on the exchange rate. Consequently, it may be difficult or impossible to find evidence in favour of one relation without simultaneously accounting for the other [30]. Most recent empirical studies therefore cast this problem within the multivariate framework of the VECM and test for two cointegrating relations amongst $p_t^{SA}, p_t^{US}, e_t^{R/\$}, r_t^{SA}$ and r_t^{US} which may be economically identified as the PPP and UIP conditions. This testing strategy will be employed in this paper conditional on the monetary and exchange rate regimes which characterise the time period under analysis.

3 Data Issues and Methodological Approach

3.1 Time Period

The empirical analyses to follow in Section 4 were conducted using quarterly data from the first quarter of 1972 to the first quarter of 2007 obtained from the International Financial Statistics online database published by the International Monetary Fund. This time period is characterised by a number of significant monetary and exchange rate regime shifts together with a substantial degree of financial and external liberalisation.

The choice of time period for analysis was motivated by the change in exchange rate policy that took place toward the end of 1971. As is evident from the top panel of Figure 1, the rand/dollar exchange rate was fixed prior to 1972 under the Bretton Woods fixed exchange rate system. In order to ease pressures on the exchange rate, relative price levels had to be controlled accordingly during this period. From 1972, however, the exchange rate was allowed to vary such that arbitrage opportunities in the goods market could be eliminated by exchange rate movements, thereby restoring parity in national price levels. The studied time period was therefore restricted to the post Bretton Woods era in which exchange rate adjustment is expected to bring about purchasing power parity, albeit subject to intervention by the South African Reserve Bank at times.

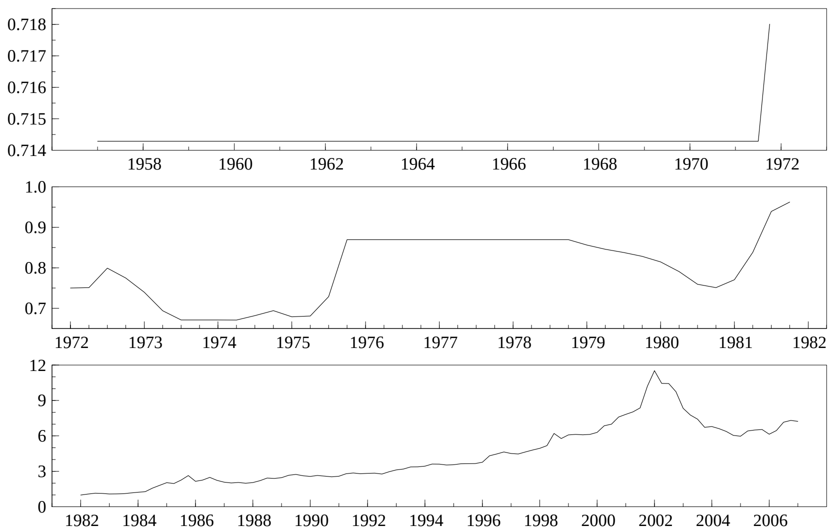

Figure 1: Nominal rand/dollar exchange rate

During the 1970s, the rand was pegged to either the US dollar or the pound sterling. As a consequence, South Africa's exchange rate policy mirrored the volatile developments on the international front during this period. The level of the peg was adjusted relatively frequently by policy-makers and took the form of discrete step-changes, as is evident from the middle panel of Figure 1. Annual inflation over this decade averaged 9%, with major price shocks attributable to the 1973 and 1979 oil price shocks and the sharp rise in the gold price from 1979 to 1983 [4].

Import surcharges were introduced in 1977 following the cessation of capital inflows due to the Soweto riots in 1976 and remained in effect until 1980 when the high gold price took pressure off the current account [4]. In addition to the strict exchange controls on the capital flows of residents, nonresidents were required to place the proceeds from sales of South African assets in blocked rand accounts for a period of five years [32].

From the 1960s to early 1980s, monetary policy in South Africa may be characterised as a liquid asset ratio system [4]. During this period, liquidity conditions in the country were controlled mainly by means of changes to the amount of cash and liquid assets that commercial banks were required to hold as a percentage of total deposits [32]. The interest rate, on the other hand, was not regarded as an important corrective tool. The intention was that the limited supply and low yield of such liquid assets would curtail bank lending and money supply, thereby easing inflationary pressures [4].

Greater flexibility was introduced into the foreign exchange market in 1979 with the adoption of a managed float, dual exchange rate system. Under this system, the financial transactions of nonresidents were valued at a discounted exchange rate, known as the financial rand, whilst current account transactions were subject to a commercial exchange rate which was announced on a daily basis in line with market forces. The intended impact of this dual exchange rate system was to break the direct link between foreign and domestic interest rates as well as to insulate the capital account from particular categories of capital flows [4]. The financial rand was later abolished in 1983 in line with the liberalisation objectives of the Reserve Bank. At this time, controls on nonresident capital flows were also lifted and a more lenient attitude was taken to applications from residents for direct investment abroad. The unified exchange rate remained stable for a few months, after which it depreciated sharply on the back of the gold price decline in 1983 [3].

The early 1980s also saw a monetary regime shift to a system based on cash reserves, with the South African Reserve Bank maintaining indirect control of the money supply by influencing the money market shortage through the interest rate at which banking institutions were provided with discount-window accommodation against suitable collateral [32]. The climbing gold price in 1979 put downward pressure on interest rates in an attempt to prevent an excessive appreciation of the newly floating exchange rate. By the first quarter of 1980, the interest rate had dropped to 4.7%, while annual inflation rose to almost 15%, averaging around 13% from 1980 to 1985. This inflationary trend was, however, later reversed after a very sharp rise in interest rates in the aftermath of the gold shocks in 1981 and 1982 [4].

The mid-1980s were characterised by prolonged social and political unrest in South Africa and a deteriorating sovereign risk rating which led to large capital outflows and disinvestment from the country. The tense political climate coupled with the gold price decline in 1983 caused the unified rand/dollar exchange rate to fall sharply. In 1985, the rand fell even further following the debt crisis induced by the refusal of a number of international banks to roll over South Africa's short-term loans. The resulting financial sanctions led to a foreign debt standstill and a subsequent series of rescheduling agreements over the period from 1985 to 1995. The financial rand was reintroduced in 1985 and capital controls were once again tightened [32]. The feed-through to inflation became evident toward the end of 1984 which, together with strong consumer demand, led to a sharp rise in interest rates peaking at almost 22% in early 1985 [4]. Within the next year, however, interest rates were more than halved in an attempt to stimulate domestic demand on the back of declining investor confidence. With inflation rising to 16.4% in 1986, the Reserve Bank retaliated by imposed target ranges for the growth of broad money, which were announced annually from 1986 to 1998 [3].

Aron, Elbadawi and Kahn suggest that exchange rate intervention during the 1980s was directed at maintaining profitability and stability in the gold mining industry [3]. The authors indicate that the real rand gold price was fairly stable over this period, despite large fluctuations in the dollar gold price. Furthermore, any instability in the real rand price of gold over this period can be attributed to real shocks in the form of the excessive rise in the dollar gold price and the debt crisis shock of 1985. Whether the stable real rand gold price was deliberate or not, the authors argue that the outcome was a highly volatile real exchange rate that protected the gold mining industry, but had a negative impact on the manufacturing export sector.

From 1988, the real rand gold price has been allowed to fall as the importance of gold as a proportion of foreign exchange earnings declines. The result was a far more stable real exchange rate, appreciating between 1988 and 1992 and then falling from 1992 to mid 1994 due to the capital outflows that resulted from political uncertainty ahead of South Africa's first democratic elections. Following the general elections in 1994, capital inflows resumed strongly as liberalisation efforts were once again intensified. Capital controls on residents were gradually relaxed and almost all controls on nonresidents were removed. In addition, the lowering of trade tariffs led to a sharp growth in exports and imports [32]. The intervention efforts that followed in order to prevent an excessive appreciation of the rand were largely successful, albeit at the expense of monetary targets [4]. In March 1995, the financial and commercial rands were once again unified resulting in highly volatile real and nominal exchange rates from 1996 to 1999 [3].

Interest rates were raised substantially in 1988, partly influenced by the increases in world interest rates and partly to curb inflation and maintain a current account surplus. Interest rates were kept high in subsequent years by the new Reserve Bank governor, Dr Chris Stals, appointed in 1989. Nonetheless, inflation remained persistent until 1992, exceeding 15% for the most part. From late 1992, however, inflation finally began to decline reaching single figures from early 1993. Money growth too declined sharply, falling within the official targets from mid-1990. Between 1992 and 1993, there was a gradual decline in the interest rate to 12% which was sustained until the 1994 elections [4].

The currency crisis that began in February 1995 resulted in a sharp depreciation in the rand/dollar exchange rate which was met with massive intervention from the Reserve Bank. Further currency crises occurred in October 1996, November 1997 and April 1998, triggered largely by contagion effects from the Asian crisis and a fall in the price of gold and other metals [15]. Following the April 1998 crisis, the exchange rate was around 40% below its average value between the elections in 1994 and the first crisis in 1995. The interest rate rose to 17% after the first crisis, falling to 15% prior to the April 1998 crisis and then rising to 20% thereafter. Inflation, however, averaged just over 6% in 1998 [4].

In early 1998, a new regime of monetary accommodation was introduced, whereby the Reserve Bank offered daily tenders of liquidity to banks through repurchase transactions. By controlling the amount of liquidity available for tender each day, the Reserve Bank would signal its intentions with respect to the repurchase rate, where an excess demand for liquidity would cause the repurchase rate to rise and an excess supply would cause the repurchase rate to fall. The repurchase rate could therefore fluctuate on a daily basis according to market forces. However, the rate was fixed at 12% toward the end of 1999 in order to provide some certainty during the transition into the new millennium amidst Y2K fears. The repurchase rate was again allowed to fluctuate in January 2000, but was subsequently fixed in October 2001. The fixed repurchase system in which the Reserve Bank announces the repurchase rate periodically is still in effect to date [4].

In February 2000, the South African Reserve Bank introduced a formal inflation-targeting framework in which it employs the repurchase rate as its sole mechanism for influencing general price levels in the country. Hence, the repurchase rate is determined by the Reserve Bank in order to restrict inflation to within the current target range of 3% to 6%. The turn of the century also brought with it a sharp real depreciation in the rand from 1999 to 2002 due to a decline in mineral exports, followed by a substantial real appreciation precipitated by the natural resource boom between 2002 and 2006 [15]. These movements in the exchange rate are reflected in the bottom panel of Figure 1.

3.2 Data Description

The variables of interest in the present study are as follows:

- The natural logarithm of the Producer Price Index in South Africa, denoted as p_t^{SA} at time t, with 2000 as the base year
- The natural logarithm of the Producer Price Index in the United States, denoted as p_t^{US} at time t, with 2000 as the base year
- The natural logarithm of the nominal rand/dollar exchange rate, denoted as $e_t^{R/\$}$ at time t
- The average discount rate of the three month Treasury Bill in South Africa, denoted as r_t^{SA} at time t
- The average discount rate of the three month Treasury Bill in the United States, denoted as r_t^{US} at time t
- The natural logarithm of the average dollar gold price, denoted as g_t at time t
- The natural logarithm of the average crude oil price, denoted as c_t at time t

Note that the Consumer Price Index (CPI) and Gross Domestic Product (GDP) deflator were also considered as indicators of the general price levels in South Africa and the United States. These series together with the Producer Price Index (PPI) are illustrated in Figure 2 for the two countries. It should be noted that the three price indices do not appear to differ substantially over the time period considered, particularly in the case of South Africa. The results of this paper therefore appear to be fairly robust to the choice of price index. The PPI was ultimately chosen over the CPI as the latter index is likely to bias the results in favour of PPP, since South Africa's consumer commodity bundle includes imported goods from the United States which are also included in the consumer commodity bundle of the United States and vice versa.

The relevance of the gold price in the present study is motivated by the fact that South Africa is a predominately resource-based economy and one of the largest exporters of gold in the world. The ratio of gold exports to non-gold merchandise exports excluding services averaged 0.93 between 1980 and 1985, although this figure has declined considerably in recent years. Nonetheless, movements in the gold price continue to have marked effects on the exchange rate in South Africa. The inclusion of the crude oil price as a variable of interest is prompted by the influence of this variable on the price of goods and services which require oil and oil products as inputs to production. Since most of South Africa's oil stocks are imported, fluctuations in the oil price can have profound effects on inflation levels in the country, whilst also influencing the exchange rate [4]. All series and their first differences are plotted in Figures 3 and 4 respectively.

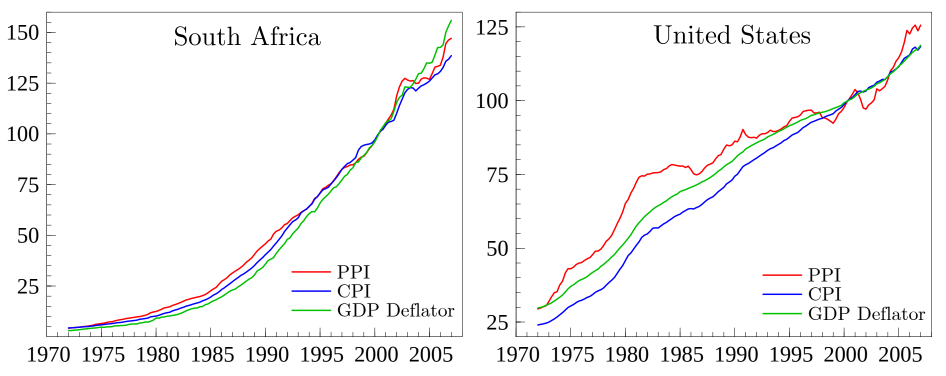

Figure 2: Producer Price Index (PPI), Consumer Price Index (CPI) and Gross Domestic Product (GDP) deflator for South Africa and the United States

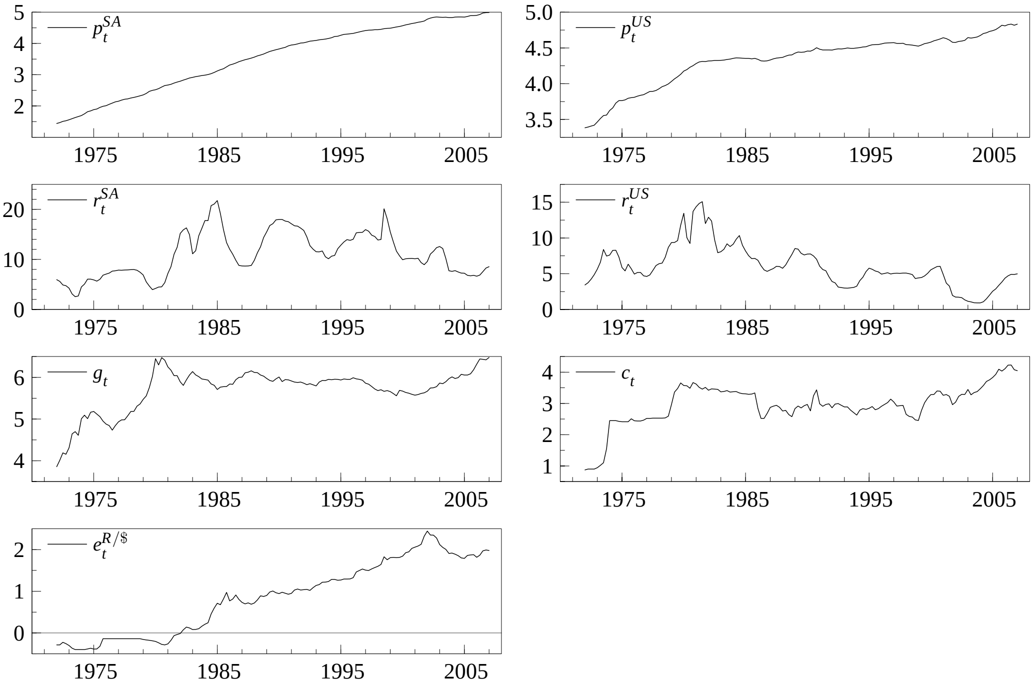

Figure 3: Time series plots of variables under analysis

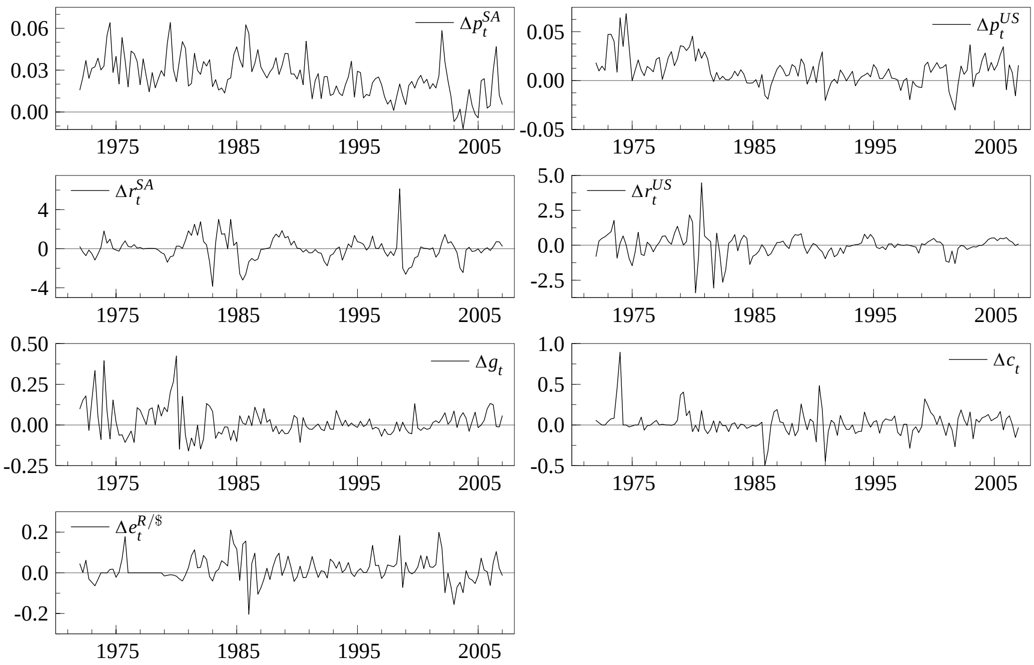

Figure 4: Time series plots of differenced variables under analysis

3.3 Methodology

The Augmented Dickey-Fuller tests presented in Table 1 indicate that all series are non-stationary in levels, but stationary in first differences. Although there is evidence at the 5% significance level to suggest that the natural logarithm of the average dollar gold price g_t is stationary in levels, this hypothesis is rejected at the 1% significance level. Given that the ADF test lacks statistical power, g_t will be treated as non-stationary in levels. Consequently, it is concluded that all series have exactly one unit root and are therefore $I(1)$ processes. The non-stationarity of the data suggests that an unrestricted vector autoregressive model, coupled with Johansen's multivariate cointegration tests, may be an appropriate point of departure for analysing these data.

Levels	Test Statistic	Model*
p_t^{SA}	-0.283	AR(4) with intercept and trend
p_t^{US}	-3.363	AR(2) with intercept and trend
$e_t^{R/\$}$	-2.985	AR(4) with intercept and trend
r_t^{SA}	-2.733	AR(2) with intercept
r_t^{US}	-2.696	AR(8) with intercept and trend
g_t	-3.598†	AR(4) with intercept and trend
c_t	-3.044	AR(2) with intercept and trend
Differences	**Test Statistic**	**Model***
Δp_t^{SA}	-3.599‡	AR(3) with intercept
Δp_t^{US}	-4.747‡	AR(2) with intercept
$\Delta e_t^{R/\$}$	-4.991‡	AR(3) with intercept
Δr_t^{SA}	-7.366‡	AR(1)
Δr_t^{US}	-5.294‡	AR(7)
Δg_t	-3.974‡	AR(3)
Δc_t	-8.065‡	AR(2)

† and ‡ denote significance at the 5% and 1% levels respectively

* The lag length was chosen to remove serial correlation and is supported by the Hannan-Quinn information criterion

Intercepts and trends were included where these appeared commensurate with the data generating process

Table 1: Augmented Dickey-Fuller Tests of Integrating Order

Since South Africa represents a small economy relative to the United States, it may be the case that price levels and interest rates in the United States do not adjust to disequilibria in the long-run PPP and UIP relations between South Africa and the United States. Consequently, the series p_t^{US} and r_t^{US} will be tested for weak exogeneity with respect to the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. The full vector error correction model comprising a system of five equations for each of Δp_t^{SA}, Δp_t^{US}, $\Delta e_t^{R/\$}$, Δr_t^{SA} and Δr_t^{US} could therefore potentially be reduced to a partial model of only three or four equations which includes as much information as the full system about the cointegrating relations and adjustment coefficients [26]. The weakly exogenous variables, if any, will then feature only in the cointegrating relations and as lagged covariates (after taking first differences) in the model.

The influence of the gold and oil prices on the rand/dollar exchange rate and general price levels as described in Sections 3.1 and 3.2 is likely to warrant their inclusion in the model. Following the workings of Johansen and Juselius, these variables and their lags will be treated as exogenous and tested for statistical significance as differenced $I(0)$ covariates in the proposed VECM [30]. It will therefore be assumed that the gold and oil prices do not feature in the long-run cointegrating relations.

The economic theories of PPP and UIP imply two cointegrating relations between the variables p_t^{SA}, p_t^{US}, $e_t^{R/\$}$, r_t^{SA} and r_t^{US}. The strict versions of these two equilibrium conditions were defined earlier as

$$\begin{aligned} p_t^{SA} - p_t^{US} - e_t^{R/\$} + \mu^{(1)} &= \xi_t^{(1)} \\ r_t^{SA} - r_t^{US} - \mu^{(2)} &= \xi_t^{(2)}, \end{aligned}$$

where $\xi_t^{(1)}$ and $\xi_t^{(2)}$ are white noise processes. The term $\mu^{(1)}$ in the PPP equation represents the mean natural logarithm of the real exchange rate and the term $\mu^{(2)}$ in the UIP equation is the mean interest rate differential, which may be non-zero due to a country-specific risk premium, speculative effects and other imperfections in the asset market. Recall, however, that relative PPP may not be expected to hold in the strict sense above due to the presence of transactions costs, trade barriers and non-tradable goods *inter alia*. In addition, the strict form of UIP assumes that the exchange rate is stationary in mean, which does not appear to be the case upon examination of the exchange rate series plotted in Figure 3. It may therefore be more feasible to consider the weak forms of relative PPP and UIP defined as

$$\begin{aligned} p_t^{SA} - \beta_1^{(1)} p_t^{US} - \beta_2^{(1)} e_t^{R/\$} + \mu^{(1)} &= \xi_t^{(1)} \\ r_t^{SA} - \beta_1^{(2)} r_t^{US} - \mu^{(2)} &= \xi_t^{(2)}, \end{aligned}$$

where the β coefficients may deviate from unity due to the aforementioned market frictions. As with the strict versions of PPP and UIP, the series $\xi_t^{(1)}$ and $\xi_t^{(2)}$ must be stationary. Hence, the unrestricted VAR model will be tested for a cointegrating rank of two. The linear restrictions necessary to economically identify the two cointegrating vectors as the weak and strict forms of PPP and UIP will then be tested for data admissibility. In the event that the data do support these long-run relations, the adjustment coefficients in the VECM will be examined to determine the manner in which equilibrium is restored.

Sustained departures from PPP and UIP may, however, be expected *a priori* in light of the notable monetary and exchange rate regime shifts discussed earlier. Indeed, many empirical studies have found the real value of the rand to be non-stationary when large periods of time are considered, such as MacDonald and Ricci who examine the real effective exchange rate from 1970 to 2002 [40]. Furthermore, Fedderke and Pillay find evidence in favour of a time-varying risk premium for South Africa [12]. It may therefore be anticipated that the mean logarithm of the real exchange rate $\mu^{(1)}$ and the mean interest rate differential $\mu^{(2)}$ are not constant over the entire time period as is required for PPP and UIP to hold. Moreover, large changes in these means will lead to the rejection of the PPP and UIP hypotheses when tested in the standard VECM framework. Suppose, however, that $\mu^{(1)}$ and $\mu^{(2)}$ are constant within distinct regimes, but differ between regimes. In this case, a Markov-switching VECM with regime-dependent intercepts in the cointegration space will be more likely to establish PPP and UIP as cointegrating relations conditional on the underlying regimes. Furthermore, the regime-dependent means associated with these relations may shed light on the economic developments that led to the break-down of PPP and UIP in the conventional sense. Since the underlying regimes are estimated simultaneously with the parameters in the MS-VECM, it will also be interesting to note if the regimes estimated from the data coincide with the historical regimes outlined in Section 3.1.

In light of the economic and political developments that took place in South Africa, there would appear to be four distinct regimes in the 35-year time period. The first regime spans the 1970s and is characterised by a fixed exchange rate system and monetary policy based on the liquid asset ratio system. The early to mid 1980s may be classified as another distinct period in South Africa's economic history. A managed float exchange rate system was adopted and the liquid asset ratio system of accommodation was substituted with a system based on cash reserves. Politically, this regime is characterised by uncertainty and social unrest. Consequently, investors demanded a higher risk premium for investing in South Africa during this regime relative to regime 1. The mean interest rate differential $\mu^{(2)}$ should therefore be larger in this regime. Similarly, the large capital outflows following the debt crisis in 1985 produced a real depreciation in the rand/dollar exchange rate such that $\mu^{(1)}$ should also be larger in regime 2 relative to regime 1.

The late 1980s to late 1990s may be regarded as another distinct economic regime. During this period, the new governor of the South African Reserve Bank, Dr Chris Stals, was concerned largely with stabilising the real effective exchange rate in the wake of several currency crises from 1995 to 1998. After the huge capital outflows during the tumultuous second regime, a lower real exchange rate was observed throughout regime 3, despite substantial real appreciations and depreciations at times. Consequently, $\mu^{(1)}$ might be expected to decline from regime 2 to regime 3. Target ranges for broad money were announced annually throughout regime 3 to be achieved by manipulating the interest rate. Nonetheless, the interest rate remained high throughout this regime. The higher interest rate, together with the rise in investor risk prior to the elections in 1994, is likely to result in a higher mean interest rate differential $\mu^{(2)}$ in this regime relative to regime 2.

Finally, the period from 1998 to 2007 may be classified as a fourth regime, coinciding with the introduction of the repurchase system of accommodation in early 1998. This final regime also saw the introduction of a formal inflation-targeting policy, which has successfully brought inflation under control in recent years. Lower interest rates and an improved country-specific risk profile are likely to have led to a decline in the mean interest rate differential $\mu^{(2)}$ in this regime. Meanwhile, the real exchange rate depreciated sharply in the first half of this regime following a decline in mineral exports, but recovered dramatically in subsequent years. The net effect on the mean real exchange rate over this regime is therefore indeterminate.

Due to software restrictions, the cointegrating rank and economically identifying constraints on the cointegrating vectors cannot be tested in the MS-VECM. Instead, these will be tested in a linear VECM with dummy variables for each regime. The regime-dependent intercepts in the cointegration space will be computed and discussed and the loading matrix in the MS-VECM will be compared to that obtained in the linear VECM to determine whether the two models imply different adjustment dynamics to disequilibria in the long-run relations. Residual diagnostics will also be performed for both models and any discrepancies noted.

3.4 Software

Microfit 4.1 was utilised for the cointegration analyses and estimation of the standard vector error correction models to follow. The Markov-switching model was estimated using the MSVAR 1.30 package in Ox 3.00. All graphics were generated in GiveWin 2.02.

4 Empirical Findings

In this section, an attempt will be made to establish data-based evidence in support of the PPP and UIP relations proposed by economic theory. The section will commence by determining the appropriate lag order of an unrestricted VAR model with five endogenous variables: the natural logarithm of the Producer Price Index in South Africa p_t^{SA}, the natural logarithm of the Producer Price Index in the United States p_t^{US}, the natural logarithm of the nominal rand/dollar exchange rate $e_t^{R/\$}$, the average discount rate of the three month Treasury Bill in South Africa r_t^{SA} and the average discount rate of the three month Treasury Bill in the United States r_t^{US}. Having established in Section 3.3 that these five series are in fact $I(1)$, the VAR model will then be reparameterised as a VECM under the presumption that the matrix $\mathbf{\Pi}$ is rank deficient. The number of cointegrating relations will be assessed by considering both the data and economic theory. Linear restrictions that economically identify the PPP and UIP relations will then be imposed on the cointegrating vectors and tested for data admissibility. The variables p_t^{US} and r_t^{US} will then be tested for weak exogeneity with respect to $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ in an attempt to reduce the full VECM to a more parsimonious partial model without adversely affecting inferences about the cointegrating relations. Finally, a Markov-switching VECM, which explicitly accounts for the regime shifts noted in Section 3.3, will be proposed for the data. The linear and Markov-switching models will be compared in terms of the evidence furnished by each with respect to PPP and UIP as well as the equilibrium adjustment mechanisms implied by the two models. The section will conclude with an examination of the residuals of the Markov-switching model relative to those obtained in the linear model.

4.1 The Unrestricted VAR Model

In order to determine an appropriate lag order p for subsequent models, the following unrestricted VAR model was estimated for $p = 1, \ldots, 5$,

$$\boldsymbol{y}_t = \sum_{i=1}^{p} \mathbf{\Pi}_i \boldsymbol{y}_{t-i} + \mathbf{\Phi} \boldsymbol{D}_t + \boldsymbol{\varepsilon}_t, \tag{4.1}$$

where $\boldsymbol{y}_t = (p_t^{SA}, p_t^{US}, e_t^{R/\$}, r_t^{SA}, r_t^{US})'$. The deterministic and exogenous variables in the vector $\boldsymbol{D}_t$ were selected based on their statistically significant contribution to likelihood. These variables are presented in Table 2 together with their respective likelihood ratio statistics. Note that the inclusion of a deterministic trend in the model is warranted in terms of likelihood, although its economic interpretation is questionable. It may, however, be argued that this term acts as a proxy for other observed and unobserved variables which have not been included in the analysis, but which exert an influence on the endogenous variables in the response vector $\boldsymbol{y}_t$.

The Akaike information criterion (AIC), Schwarz Bayesian information criterion (SBC) and likelihood ratio tests for lag order selection in the unrestricted VAR model given by Equation (4.1) are presented in Table 3. Note that

$$\text{AIC} = \log L(\hat{\boldsymbol{\theta}}) - q$$

and

$$\text{SBC} = \log L(\hat{\boldsymbol{\theta}}) - \frac{q}{2} \log T,$$

where $\hat{\boldsymbol{\theta}}$ is the maximum likelihood estimate of the model parameters, q is the number of freely estimated parameters in the system and $T = 141$ is number of observations used for estimation in each equation, such that larger values of these criteria imply better models. Note too that the likelihood ratio test statistic is only asymptotically chi-squared distributed and tends to over-reject the null hypothesis when the length of the observed data sequence is small. Consequently, a modified version of the likelihood ratio statistic is also provided in Table 3, which attempts to account for the finite number of observations by making an adjustment for the degrees of freedom in the estimated model [44].

Variables in $\boldsymbol{D}_t$		χ^2_d	d	P-value*
Intercept	1	12.95	5	0.024
Differenced gold price at time t	Δg_t	23.15	5	0.000
Differenced oil price at time t	Δc_t	35.16	5	0.000
Differenced oil price at time $t-1$	Δc_{t-1}	11.48	5	0.043
Centred seasonal dummies	d_t^1, d_t^2, d_t^3	27.38	15	0.026
Linear trend	t	25.74	5	0.000

* Test of the null hypothesis that the variable does not make a significant contribution to likelihood against the alternative hypothesis that the variable does make a significant contribution to likelihood

Table 2: Deterministic and exogenous variables in the unrestricted VAR

p	$\log L$	AIC	SBC	H_0	H_1	χ^2_{25}	χ^{2*}_{25}
1	767.053	702.053	606.218	$p=0$	$p=1$	2204.27‡	2001.04‡
2	812.121	722.121	589.427	$p=1$	$p=2$	90.14‡	78.63‡
3	831.171	716.171	546.617	$p=2$	$p=3$	38.10†	31.89
4	857.222	717.222	510.809	$p=3$	$p=4$	52.10‡	41.76†
5	874.449	709.449	466.176	$p=4$	$p=5$	34.45	26.39

† and ‡ denote significance at the 5% and 1% levels respectively

* Adjusted for small data sequences

Table 3: Lag order selection criteria in the unrestricted VAR model

Both of the information criteria suggest a fairly low lag order for the unrestricted VAR model with the AIC favouring a VAR(2) model and the more conservative SBC suggesting a lag order of one. The likelihood ratio test clearly supports the second-order VAR model relative to its first-order counterpart. On the other hand, the adjusted likelihood ratio test comparing the model with three lags to one with only two lags suggests that these two models are equivalent in terms of likelihood. Furthermore, the adjusted likelihood ratio test comparing the VAR(2) model with a VAR(4) model which may be appropriate for quarterly data yields a test statistic of 72.29 with 50 degrees of freedom and a corresponding p-value of 0.021. There is therefore no overwhelming evidence to suggest that more than two lags are necessary for these data and hence a lag order of two will be adopted in the unrestricted VAR model.

The results of the residual diagnostic tests associated with each of the five equations in the unrestricted VAR(2) model specified by Equation (4.1) with $p = 2$ are given in Table 4. The residuals of all equations fail the Jarque-Bera test for normality, although this violation does not appear particularly concerning when examining the distribution of the residuals corresponding to the p_t^{SA} and $e_t^{R/\$}$ equations. The residuals of the p_t^{US} equation and more notably the interest rate equations are, however, severely leptokurtic. Heteroscedasticity, based on the regression of squared residuals on squared fitted values, does not appear to be problematic, except in the case of the residuals of the r_t^{US} equation which also exhibit significant autoregressive conditional heteroscedasticity (ARCH). There is also evidence of significant ARCH effects in the residuals of the p_t^{US} equation and, to a lesser extent, the $e_t^{R/\$}$ and r_t^{SA} equations. Note, however, that non-normality and ARCH effects in the residuals may be expected when regime shifts are present, since the

		Equation				
		p_t^{SA}	p_t^{US}	$e_t^{R/\$}$	r_t^{SA}	r_t^{US}
Autocorrelation	χ_4^2	6.985	$12.511^{\dagger}$	$12.213^{\dagger}$	6.346	$20.154^{\ddagger}$
Functional form	χ_1^2	1.541	2.912	0.075	1.960	$5.843^{\dagger}$
Normality	χ_2^2	$7.241^{\dagger}$	$166.473^{\ddagger}$	$19.200^{\ddagger}$	$907.845^{\ddagger}$	$443.904^{\ddagger}$
Heteroscedasticity	χ_1^2	0.722	2.691	1.261	$4.833^{\dagger}$	$24.080^{\ddagger}$
ARCH(2)	χ_2^2	0.142	1.016	$6.947^{\dagger}$	$8.431^{\dagger}$	$31.632^{\ddagger}$
ARCH(4)	χ_4^2	1.364	$14.778^{\ddagger}$	8.021	8.801	$31.782^{\ddagger}$

† and ‡ denote significance at the 5% and 1% levels respectively

Table 4: Model diagnostics for unrestricted VAR(2) model

distribution of the residuals may then be a mixture of zero mean Gaussian distributions with different variances for each regime.

The Lagrange multiplier test indicates statistically significant serial correlation amongst the residuals pertaining to the p_t^{US}, $e_t^{R/\$}$ and r_t^{US} equations, although this is only especially concerning in the latter case. Ramsey's regression specification error test (RESET) of functional form does not reject the null hypothesis of zero mean Gaussian residuals with constant variance in all equations except that of r_t^{US} [43]. This result might be expected given that the residuals of this equation violate all the requirements of the linear model considered in Table 4. However, if interest rates in the United States are weakly exogenous with respect to the PPP and UIP relations between South Africa and the United States, r_t^{US} may be excluded from the vector of endogenous variables without biasing inferences on these long-run relations. The problematic residuals associated with this equation would therefore only be of concern if r_t^{US} is not weakly exogenous for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ in the VECM.

The 10×10 companion matrix of the VAR(2) process with five endogenous variables is defined as

$$\begin{bmatrix} \boldsymbol{\Pi}_1 & \boldsymbol{\Pi}_2 \\ \boldsymbol{I} & \boldsymbol{0} \end{bmatrix}.$$

The eigenvalues of this estimated matrix are given in Table 5. These eigenvalues all lie within the unit circle, although the largest absolute eigenvalue is very close to unity implying an extremely slow reversion to the mean. From Section 2.2.2, this implies that $\boldsymbol{y}_t$ may be treated as $I(1)$ and modelled as a non-stationary time series.

Eigenvalues	
$0.98 \pm 0.01i$	0.47
$0.85 \pm 0.11i$	$0.29 \pm 0.17i$
0.78	$0.12 \pm 0.07i$

Table 5: Eigenvalues of the companion matrix of the VAR(2) model where $i = \sqrt{-1}$

4.2 The Vector Error Correction Model

The unrestricted VAR model in Equation (4.1) with $p = 2$ may be rewritten in its error correction form as a VECM(1) process

$$\Delta \boldsymbol{y}_t = \boldsymbol{\Pi} \boldsymbol{y}_{t-1} + \boldsymbol{\Gamma} \Delta \boldsymbol{y}_{t-1} + \boldsymbol{\Phi} \boldsymbol{D}_t + \boldsymbol{\varepsilon}_t, \tag{4.2}$$

where $\boldsymbol{\Gamma} = -\boldsymbol{\Pi}_2$ and all other parameters are as previously defined.

When parameterised in this manner, it is necessary to specify explicitly whether the constant term and linear trend which were present in the unrestricted VAR model pertain to the cointegration space, the proposed data generating process or both. Of course, it is also possible to discard these terms altogether. Whilst all these cases were explored, unrestricted intercepts would seem to be justified by the statistical significance of the intercept term and linear trend in the unrestricted VAR model. Hence the vector of deterministic and exogenous variables $\boldsymbol{D}_t$ in Equation (4.2) will include all the variables presented in Table 2 except the trend term. This specification allows for a non-zero intercept in the PPP and UIP relations which conforms to Equations (2.34) and (2.37), whilst simultaneously permitting a linear trend in the levels of the endogenous variables through the intercept term in the data generating process of $\Delta \boldsymbol{y}_t$. Note, however, that this specification does not allow for a deterministic trend in the cointegration space since this would violate the PPP and UIP conditions. The trend is therefore assumed to be deterministically cointegrated in the sense that it cancels out in the cointegrating relations [43].

4.2.1 Determining the Cointegrating Rank of the VECM

The number of cointegrating relations is given by the rank of the matrix $\boldsymbol{\Pi}$, which has an upper bound of five in this study. The eigenvalues corresponding to each of the five potential cointegrating vectors are 0.29, 0.16, 0.08, 0.03 and 0.02, which at first glance would seem to suggest a cointegrating rank of two. However, Johansen's $\lambda_{\max}$ and trace test statistics presented in Table 6 imply only a single cointegrating relation at the 5% significance level, although the hypothesis of two cointegrating relations is only marginally nonsignificant in both tests at this significance level. Indeed, the $\lambda_{\max}$ test statistic provides evidence in favour of two cointegrating relations at the 10% significance level, whilst the trace statistic for this test is only negligibly smaller than its 90% critical value. Furthermore, the Monte Carlo simulations of Persaran, Shin and Smith indicate that these test statistics tend to under-reject when observed data sequences are small [45]. Hence Johansen's test statistics would appear to support the notion of two cointegrating relations between the variables in $\boldsymbol{y}_t$.

The information criteria presented in Table 7 provide somewhat conflicting evidence with respect to the cointegrating rank. Whilst the liberal AIC favours a rank of three, the conservative SBC suggests that there is only a single cointegrating relation. The Hannan-Quinn information criterion (HQC),

H_0	H_1	$\lambda_{\max}$	95% CV	90% CV
$r = 0$	$r = 1$	47.45	33.64	31.02
$r \leq 1$	$r = 2$	25.25	27.42	24.99
$r \leq 2$	$r = 3$	12.19	21.12	19.02
$r \leq 3$	$r = 4$	4.57	14.88	12.98
$r \leq 4$	$r = 5$	3.17	8.07	6.50

H_0	H_1	Trace	95% CV	90% CV
$r = 0$	$r \geq 1$	92.63	70.49	66.23
$r \leq 1$	$r \geq 2$	45.18	48.88	45.70
$r \leq 2$	$r \geq 3$	19.93	31.54	28.78
$r \leq 3$	$r \geq 4$	7.74	17.86	15.75
$r \leq 4$	$r = 5$	3.17	8.07	6.50

Table 6: Johansen's test statistics for selection of cointegrating rank in the VECM(1) with unrestricted intercepts

Rank	AIC	SBC	HQC
$r = 0$	692.94	604.47	656.99
$r = 1$	707.66	605.93	666.32
$r = 2$	713.29	601.23	667.75
$r = 3$	714.38	594.96	665.85
$r = 4$	713.67	589.82	663.34
$r = 5$	714.25	588.93	663.33

Table 7: Information criteria for selection of cointegrating rank in the VECM(1) with unrestricted intercepts

which is defined as

$$\text{HQC} = \log L(\hat{\boldsymbol{\theta}}) - q \log(\log T),$$

provides a compromise between these two extremes, suggesting a cointegrating rank of two in the VECM(1) with unrestricted intercepts [44]. The information criteria therefore complement Johansen's test statistics in support of two cointegrating relations. Hence this specification will be adopted in the hope that these two long-run relations correspond to the economic theories of PPP and UIP.

4.2.2 Testing for PPP and UIP

Assuming two cointegrating relations and an intercept in the cointegration space and the data generating process, the VECM in Equation (4.2) may be re-expressed as

$$\Delta \boldsymbol{y}_t = \boldsymbol{\alpha}\boldsymbol{\beta}' \boldsymbol{y}_{t-1} + \boldsymbol{\Gamma}\Delta \boldsymbol{y}_{t-i} + \boldsymbol{\Phi}\boldsymbol{D}_t + \boldsymbol{\varepsilon}_t, \tag{4.3}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are 5×2 matrices of adjustment and cointegration coefficients respectively, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Phi}$ are coefficient matrices to be estimated and $\boldsymbol{D}_t = (1, \Delta g_t, \Delta c_t, \Delta c_{t-1}, d_t^1, d_t^2, d_t^3)'$ is a vector of deterministic and exogenous variables as defined in Table 2. In order for the cointegrating relations to be identified, the rank condition requires the imposition of at least one linear restriction on each of the two normalised cointegrating vectors in $\boldsymbol{\beta}$. To this effect, the zero constraints $\beta_{41} = 0$ and $\beta_{12} = 0$ were imposed on the first and second columns of the cointegration matrix $\boldsymbol{\beta}$ respectively and each cointegrating vector was normalised by setting $\beta_{11} = \beta_{42} = 1$. From Section 2.3.4, this is equivalent to imposing exactly identifying linear restrictions $\boldsymbol{R}_1 = (0,0,0,1,0)'$ and $\boldsymbol{R}_2 = (1,0,0,0,0)'$ on the normalised cointegrating vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ respectively such that

$$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ \beta_{21} \\ \beta_{31} \\ \beta_{41} \\ \beta_{51} \end{pmatrix} = 0 \quad \text{and} \quad \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \beta_{12} \\ \beta_{22} \\ \beta_{32} \\ 1 \\ \beta_{52} \end{pmatrix} = 0.$$

The resulting normalised and constrained cointegration matrix is therefore

$$\boldsymbol{\beta}' = \begin{bmatrix} 1 & \beta_{21} & \beta_{31} & 0 & \beta_{51} \\ 0 & \beta_{22} & \beta_{32} & 1 & \beta_{52} \end{bmatrix},$$

which ensures that the cointegrating relations are exactly identified.

Now recall that $\boldsymbol{y}_t = (p_t^{SA}, p_t^{US}, e_t^{R/\$}, r_t^{SA}, r_t^{US})'$. Hence the first cointegrating vector has been normalised on p_t^{SA} with a zero constraint on the coefficient corresponding to r_t^{SA}. In order to examine whether purchasing power parity holds in the weakest sense, the coefficient β_{51} corresponding to r_t^{US} in the first cointegrating vector $\boldsymbol{\beta}_1$ must be constrained to zero. To achieve this, $\boldsymbol{R}_1$ is augmented by an additional linearly independent column such that

$$\begin{matrix} \boldsymbol{R}_1' & & \boldsymbol{\beta}_1 & = & \boldsymbol{0} \\ \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} & & \begin{pmatrix} 1 \\ \beta_{21} \\ \beta_{31} \\ \beta_{41} \\ \beta_{51} \end{pmatrix} & = & \begin{pmatrix} 0 \\ 0 \end{pmatrix} \end{matrix}. \tag{4.4}$$

Note that this additional restriction implies that the first cointegrating vector is overidentified since only a single linear restriction on each cointegrating vector is required for just identification in this case. If the imposition of this overidentifying restriction is supported by the data and β_{21} and β_{31} are close to -1, this would constitute evidence in favour of the weak form of PPP. Under the hypothesis that weak PPP holds, the cointegration matrix would therefore take the form

$$H_{\text{WPPP}} : \boldsymbol{\beta}' = \begin{bmatrix} 1 & \beta_{21} & \beta_{31} & 0 & 0 \\ 0 & \beta_{22} & \beta_{32} & 1 & \beta_{52} \end{bmatrix},$$

where the subscript WPPP stands for weak purchasing power parity. In order to test for PPP in the strict sense, a further two overidentifying linear restrictions on the first normalised cointegrating vector $\boldsymbol{\beta}_1$ are necessary. Consequently, $\boldsymbol{R}_1$ will now comprise of four linearly independent columns with

$$\begin{array}{ccc} \boldsymbol{R}_1' & \boldsymbol{\beta}_1 & = \quad \mathbf{0} \\ \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} & \begin{pmatrix} 1 \\ \beta_{21} \\ \beta_{31} \\ \beta_{41} \\ \beta_{51} \end{pmatrix} & = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \end{array} \tag{4.5}$$

such that β_{21} and β_{31} are now constrained to -1 and $\boldsymbol{\beta}_1'\boldsymbol{y}_t = p_t^{SA} - p_t^{US} - e_t^{R/\$}$ corresponds to the strict PPP relation. The strict purchasing power parity hypothesis may therefore be tested by constraining the normalised cointegration matrix as

$$H_{\text{PPP}} : \boldsymbol{\beta}' = \begin{bmatrix} 1 & -1 & -1 & 0 & 0 \\ 0 & \beta_{22} & \beta_{32} & 1 & \beta_{52} \end{bmatrix}$$

and testing the three overidentifying restrictions for data admissibility. If the data supports the overidentifying restrictions in Equation (4.5), then $\boldsymbol{\beta}_1'\boldsymbol{y}_t = p_t^{SA} - p_t^{US} - e_t^{R/\$}$ is $I(0)$ and hence PPP holds in the strictest sense.

Similarly, the second cointegrating vector has been normalised on r_t^{SA} with the coefficient on p_t^{SA} constrained to zero. The weak form of the uncovered interest parity may therefore be represented by the second normalised cointegrating vector subject to the following three linear restrictions

$$\begin{array}{ccc} \boldsymbol{R}_2' & \boldsymbol{\beta}_2 & = \quad \mathbf{0} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} & \begin{pmatrix} \beta_{12} \\ \beta_{22} \\ \beta_{32} \\ 1 \\ \beta_{52} \end{pmatrix} & = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \end{array} \tag{4.6}$$

such that $\boldsymbol{\beta}_2'\boldsymbol{y}_t = r_t^{SA} + \beta_{52} r_t^{US}$ corresponding to the weak UIP relation. If the overidentifying restrictions on $\boldsymbol{\beta}_2$ given by Equation (4.6) are supported by the data and β_{52} is close to -1, then this would constitute evidence in favour of the weak form of the UIP relation. The resulting cointegration matrix under the hypothesis that weak UIP holds would have the form

$$H_{\text{WUIP}} : \boldsymbol{\beta}' = \begin{bmatrix} 1 & \beta_{21} & \beta_{31} & 0 & \beta_{51} \\ 0 & 0 & 0 & 1 & \beta_{52} \end{bmatrix},$$

where the subscript WUIP is an acronym for weak uncovered interest parity. In order to test for UIP in the strict sense, it is necessary to impose a further restriction on the second cointegrating vector $\boldsymbol{\beta}_2$ constraining the coefficient β_{52} to -1. This constraint is achieved by augmenting the matrix of linearly independent restrictions $\boldsymbol{R}_2$ by an additional column such that

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}}_{\boldsymbol{R}_2'} \underbrace{\begin{pmatrix} \beta_{12} \\ \beta_{22} \\ \beta_{32} \\ 1 \\ \beta_{52} \end{pmatrix}}_{\boldsymbol{\beta}_2} = \underbrace{\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}}_{\mathbf{0}}. \tag{4.7}$$

The cointegration matrix under the hypothesis of strict UIP would therefore have the form

$$H_{\text{UIP}} : \boldsymbol{\beta}' = \begin{bmatrix} 1 & \beta_{21} & \beta_{31} & 0 & \beta_{51} \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}.$$

If the three overidentifying restrictions in Equation (4.7) are supported by the data, then this would imply that $\boldsymbol{\beta}_2'\boldsymbol{y}_t = r_t^{SA} - r_t^{US}$ is an $I(0)$ process and hence UIP holds in the strictest sense.

Finally, the hypothesis that the weak forms of PPP and UIP hold jointly may be tested by considering the empirical evidence in support of Equations (4.4) and (4.6) as simultaneous overidentifying constraints, leading to a normalised cointegration matrix of the form

$$H_{\text{WPPP}} \cap H_{\text{WUIP}} : \boldsymbol{\beta}' = \begin{bmatrix} 1 & \beta_{21} & \beta_{31} & 0 & 0 \\ 0 & 0 & 0 & 1 & \beta_{52} \end{bmatrix},$$

with β_{21}, β_{31} and β_{52} close to -1. Similarly, the hypothesis that the strict forms of PPP and UIP hold jointly may be tested by constraining the cointegration matrix as

$$H_{\text{PPP}} \cap H_{\text{UIP}} : \boldsymbol{\beta}' = \begin{bmatrix} 1 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}$$

and testing the simultaneous imposition of the overidentifying restrictions implied by Equations (4.5) and (4.7) for data admissibility.

The results obtained from testing the above hypotheses in the VECM(1) given by Equation (4.3) with unrestricted intercepts are presented in Table 8. The overidentifying restrictions implied under H_{UIP} are clearly not rejected at any reasonable significance level, confirming the presence of the UIP relation in its strictest form. On the other hand, H_{PPP} is decisively rejected and even the weak form of PPP implied under H_{WPPP} lacks support from the data as an individual hypothesis.

Jointly, the strict forms of PPP and UIP are rejected due to the lack of evidence in favour of H_{PPP}. However, if the coefficient corresponding to the rand/dollar exchange rate in the first cointegrating vector is left unconstrained under $H_{\text{PPP}} \cap H_{\text{UIP}}$, the resulting set of overidentifying restrictions, denoted by the hypothesis $H^{*}_{\text{PPP}} \cap H_{\text{UIP}}$ in Table 8, are only marginally nonsignificant at the 5% level. Moreover, the Monte Carlo experiments of Gredenhoff and Jacobson indicate that the size of the asymptotic likelihood ratio test can be substantially upwardly biased when the length of the observed data sequence is small, such that the true p-values are likely to be larger than those reported here [18]. Consequently, these results do provide some evidence in favour of the joint hypothesis of strict UIP together with a weaker form of PPP such as that implied under $H^{*}_{\text{PPP}} \cap H_{\text{UIP}}$. In addition, the p-value of 0.034 corresponding to the test of the joint hypothesis $H_{\text{WPPP}} \cap H_{\text{WUIP}}$ may be regarded as supportive of this hypothesis when the results of Gredenhoff and Jacobson are considered [18]. Hence, although there is clearly no evidence in favour of PPP as an individual long-run relation, the data seems to provide some support for PPP and UIP as *joint* long-run relations. This finding reinforces the claims of Johansen and Juselius who stress the importance of modelling both relations simultaneously in a full system of equations, thereby allowing for interactions in the determination of prices, interest rates and the exchange rate [30].

4.2.3 Testing for Weak Exogeneity for $(\boldsymbol{\alpha}, \boldsymbol{\beta})$

Since South Africa represents a small economy relative to the United States, it may be possible to exclude Δp_t^{US} and Δr_t^{US} as endogenous variables in the VECM(1) above without adversely affecting inferences concerning the adjustment coefficients $\boldsymbol{\alpha}$ and cointegration matrix $\boldsymbol{\beta}$. If either or both of these variables are weakly exogenous for $(\boldsymbol{\alpha}, \boldsymbol{\beta})$, a more parsimonious partial system of fewer equations may suffice for testing the PPP and UIP relations.

Exactly Identified

$$\begin{bmatrix} 1.00 & 0.00 \\ \underset{(0.98)}{-0.36} & \underset{(16.80)}{13.82} \\ \underset{(0.36)}{-0.97} & \underset{(6.08)}{-2.92} \\ 0.00 & 1.00 \\ \underset{(0.11)}{-0.12} & \underset{(1.95)}{-2.60} \end{bmatrix}$$

$\log L = 789.29$

H_{WPPP}

$$\begin{bmatrix} 1.00 & 0.00 \\ \underset{(0.18)}{-1.29} & \underset{(6.87)}{3.95} \\ \underset{(0.10)}{-0.57} & \underset{(3.36)}{1.15} \\ 0.00 & 1.00 \\ 0.00 & \underset{(0.66)}{-1.35} \end{bmatrix}$$

$\log L = 786.57$
$\chi^2_1 = 5.44$
P-value $= 0.020$

H_{WUIP}

$$\begin{bmatrix} 1.00 & 0.00 \\ \underset{(0.65)}{-0.65} & 0.00 \\ \underset{(0.26)}{-0.91} & 0.00 \\ 0.00 & 1.00 \\ \underset{(0.07)}{-0.09} & \underset{(0.45)}{-1.25} \end{bmatrix}$$

$\log L = 787.54$
$\chi^2_2 = 3.50$
P-value $= 0.174$

$H_{\text{WPPP}} \cap H_{\text{WUIP}}$

$$\begin{bmatrix} 1.00 & 0.00 \\ \underset{(0.17)}{-1.33} & 0.00 \\ \underset{(0.08)}{-0.58} & 0.00 \\ 0.00 & 1.00 \\ 0.00 & \underset{(0.37)}{-1.24} \end{bmatrix}$$

$\log L = 784.97$
$\chi^2_3 = 8.64$
P-value $= 0.034$

H_{PPP}

$$\begin{bmatrix} 1.00 & 0.00 \\ -1.00 & \underset{(12.12)}{13.89} \\ -1.00 & \underset{(6.35)}{-3.27} \\ 0.00 & 1.00 \\ 0.00 & \underset{(1.30)}{-2.62} \end{bmatrix}$$

$\log L = 779.85$
$\chi^2_3 = 18.87$
P-value $= 0.000$

H_{UIP}

$$\begin{bmatrix} 1.00 & 0.00 \\ \underset{(0.76)}{-0.53} & 0.00 \\ \underset{(0.30)}{-0.94} & 0.00 \\ 0.00 & 1.00 \\ \underset{(0.09)}{-0.10} & -1.00 \end{bmatrix}$$

$\log L = 787.37$
$\chi^2_3 = 3.83$
P-value $= 0.281$

$H_{\text{PPP}} \cap H_{\text{UIP}}$

$$\begin{bmatrix} 1.00 & 0.00 \\ -1.00 & 0.00 \\ -1.00 & 0.00 \\ 0.00 & 1.00 \\ 0.00 & -1.00 \end{bmatrix}$$

$\log L = 777.16$
$\chi^2_6 = 24.24$
P-value $= 0.000$

$H^*_{\text{PPP}} \cap H_{\text{UIP}}$

$$\begin{bmatrix} 1.00 & 0.00 \\ -1.00 & 0.00 \\ \underset{(0.12)}{-0.61} & 0.00 \\ 0.00 & 1.00 \\ 0.00 & -1.00 \end{bmatrix}$$

$\log L = 783.69$
$\chi^2_5 = 11.20$
P-value $= 0.048$

Asymptotic standard errors are given in parentheses

Table 8: Estimated cointegration matrix $\boldsymbol{\beta}$ with linear restrictions in the VECM(1) with unrestricted intercepts

The variables Δp_t^{US} and Δr_t^{US} were therefore tested for weak exogeneity with respect to $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ in the VECM(1) given by Equation (4.3) with the cointegration matrix $\boldsymbol{\beta}$ restricted according to the hypothesis $H^*_{\text{PPP}} \cap H_{\text{UIP}}$. In order to test the weak exogeneity of Δp_t^{US} for $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ in this model, a Wald test of the following hypotheses was performed

$$H_0 : \boldsymbol{\alpha}_2 = \mathbf{0}$$
$$H_1 : \boldsymbol{\alpha}_2 \neq \mathbf{0},$$

where $\boldsymbol{\alpha}_2$ denotes the second row of $\boldsymbol{\alpha}$ and weak exogeneity is implied under the null hypothesis. A χ^2_2 test statistic of 15.90 was obtained for this hypothesis test, which leads to a firm rejection of the null hypothesis. Consequently, Δp_t^{US} cannot be treated as weakly exogenous for $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ and must therefore remain as an endogenous variable in the model. A possible explanation for this finding is that the performance of the South African economy is to some extent indicative of general world economic conditions which do exert an influence on aggregate price levels in the United States.

The variable Δr_t^{US} was tested for weak exogeneity with respect to $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ by conducting a Wald test of the following hypotheses

$$H_0 : \boldsymbol{\alpha}_5 = \mathbf{0}$$
$$H_1 : \boldsymbol{\alpha}_5 \neq \mathbf{0},$$

where $\boldsymbol{\alpha}_5$ denotes the fifth row of $\boldsymbol{\alpha}$. In this case, a χ^2_2 test statistic of 1.57 was obtained with a p-value of 0.457, which clearly does not reject the hypothesis that Δr_t^{US} is weakly exogenous for $(\boldsymbol{\alpha}, \boldsymbol{\beta})$. Accordingly, the interest rate in the United States may be treated as an $I(1)$ exogenous variable, appearing only in the cointegrating relations and as a lagged differenced covariate in the VECM(1). The full model given by Equation (4.3) may therefore be reduced to a partial system of four equations by excluding Δr_t^{US} as an endogenous variable without compromising the efficiency of the estimates of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

4.2.4 The Partial Model

Following the findings of the weak exogeneity tests above, the partial model

$$\Delta \boldsymbol{y}^*_t = \boldsymbol{\alpha}\boldsymbol{\beta}' \boldsymbol{y}_{t-1} + \boldsymbol{\Gamma} \Delta \boldsymbol{y}_{t-1} + \boldsymbol{\Phi} \boldsymbol{D}_t + \boldsymbol{\varepsilon}_t, \tag{4.8}$$

was estimated with unrestricted intercepts, where $\boldsymbol{y}^*_t = (p_t^{SA}, p_t^{US}, e_t^{R/\$}, r_t^{SA})'$ and the variables in $\boldsymbol{y}_t$ and $\boldsymbol{D}_t$ are as defined previously for the full VECM(1) given by Equation (4.3).

Johansen's test statistics and the information criteria for the selection of the cointegrating rank in this partial model are presented in Tables 9 and 10 respectively. As expected, the results do not differ substantively from those of the full model. Johansen's $\lambda_{\max}$ and trace test statistics both indicate a single cointegrating relation at the 5% significance level, although the evidence in favour of two cointegrating relations is only marginally nonsignificant at the 10% level. Since Johansen's test statistics have been shown to be downwardly biased in small datasets, these results may be construed as evidence in favour of two long-run relations between the variables in $\boldsymbol{y}_t$ [45]. In addition, both the AIC and HQC favour a cointegrating rank of two, although the more conservative SBC suggests only one cointegrating relation. Consequently, a cointegrating rank of two will be specified in the partial model.

The overidentifying restrictions on the 5×2 cointegration matrix $\boldsymbol{\beta}$ representing the weak and strict forms of PPP and UIP alluded to earlier were tested for data admissibility in the partial model given by Equation (4.8). The constrained cointegration matrices and their corresponding likelihood ratio statistics are presented in Table 11. Once again, UIP is supported by the data as a long-run relation, even in its strictest form. Furthermore, the first cointegrating vector in the partial model with $\boldsymbol{\beta}$ constrained under the hypothesis H_{UIP} has elements very close to what would be expected of the

H_0	H_1	$\lambda_{\max}$	95% CV	90% CV
$r = 0$	$r = 1$	45.31	30.71	28.27
$r \leq 1$	$r = 2$	22.14	24.59	22.15
$r \leq 2$	$r = 3$	5.62	18.06	15.98
$r \leq 3$	$r = 4$	4.53	11.47	9.53

H_0	H_1	Trace	95% CV	90% CV
$r = 0$	$r \geq 1$	77.60	58.63	54.84
$r \leq 1$	$r \geq 2$	32.29	38.93	35.88
$r \leq 2$	$r \geq 3$	10.15	23.32	20.75
$r \leq 3$	$r = 4$	4.53	11.47	9.53

Table 9: Johansen's test statistics for selection of cointegrating rank in the partial VECM(1) with unrestricted intercepts

Rank	AIC	SBC	HQC
$r = 0$	864.06	793.29	835.30
$r = 1$	878.71	796.15	845.16
$r = 2$	883.78	792.37	846.64
$r = 3$	882.59	785.28	843.05
$r = 4$	882.86	782.60	842.12

Table 10: Information criteria for selection of cointegrating rank in the partial VECM(1) with unrestricted intercepts

PPP relation. In fact, this model does not reject H_{WPPP}, although it was firmly rejected in the full model with Δr_t^{US} as an endogenous variable. The strict form of PPP, however, is still not supported as an individual hypothesis in the partial model.

The partial model also provides sound evidence in favour of weak PPP and weak UIP as joint long-run relations. The additional restrictions implied under $H^*_{\text{PPP}} \cap H_{\text{UIP}}$ are also well supported in this model. Recall that both of these hypotheses were previously rejected in the full model at the conventional 5% significance level, albeit only marginally. The joint hypothesis of strict PPP and strict UIP is, however, rejected in both the full and partial models.

The aforementioned results pertain to models in which an intercept term is assumed to be present in both the cointegration space and the data generating process. The findings of cointegration analyses are, however, known to be sensitive to intercept specifications. In order to assess the robustness of the above results to other assumptions regarding the intercept term, the partial VECM(1) was re-estimated without an intercept term and with an intercept term restricted to the cointegration space only. The likelihood ratio statistics associated with each of the hypotheses discussed earlier in each of these models are presented in Table 12.

The results of the partial VECM(1) with intercepts restricted to the cointegration space are very similar to those obtained earlier with unrestricted intercepts. Indeed, the same conclusions are reached in both models when testing all hypotheses at the standard 5% significance level. In contrast, the model without intercepts rejects all the individual and joint PPP and UIP hypotheses, with the exception of H_{WPPP} at the 5% significance level. Note that this implies that the absolute version of strict PPP corresponding to

Exactly Identified		H_{WPPP}		H_{WUIP}		$H_{\text{WPPP}} \cap H_{\text{WUIP}}$	
1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
−1.03	5.31	−1.31	1.54	−1.12	0.00	−1.32	0.00
(0.41)	(8.67)	(0.19)	(5.85)	(0.31)		(0.17)	
−0.64	1.32	−0.54	2.62	−0.66	0.00	−0.57	0.00
(0.15)	(3.29)	(0.11)	(2.99)	(0.12)		(0.09)	
0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00
−0.04	−1.58	0.00	−1.08	−0.03	−1.18	0.00	−1.13
(0.04)	(0.91)		(0.49)	(0.03)	(0.39)		(0.33)
$\log L = 945.78$		$\log L = 945.08$		$\log L = 943.95$		$\log L = 943.23$	
		$\chi^2_1 = 1.42$		$\chi^2_2 = 3.67$		$\chi^2_3 = 5.11$	
		P-value = 0.234		P-value = 0.159		P-value = 0.164	

H_{PPP}		H_{UIP}		$H_{\text{PPP}} \cap H_{\text{UIP}}$		$H^*_{\text{PPP}} \cap H_{\text{UIP}}$	
1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
−1.00	3.85	−1.09	0.00	−1.00	0.00	−1.00	0.00
	(7.89)	(0.33)					
−1.00	1.16	−0.65	0.00	−1.00	0.00	−0.59	0.00
	(5.29)	(0.13)				(0.13)	
0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00
0.00	−1.36	−0.03	−1.00	0.00	−1.00	0.00	−1.00
	(0.66)	(0.03)					
$\log L = 936.38$		$\log L = 943.83$		$\log L = 935.35$		$\log L = 942.16$	
$\chi^2_3 = 18.81$		$\chi^2_3 = 3.90$		$\chi^2_6 = 20.87$		$\chi^2_5 = 7.24$	
P-value = 0.000		P-value = 0.272		P-value = 0.002		P-value = 0.203	

Asymptotic standard errors are given in parentheses

Table 11: Estimated cointegration matrix $\boldsymbol{\beta}$ with linear restrictions in the partial VECM(1) with unrestricted intercepts

Hypothesis	d	No Intercepts χ^2_d	P-value	Restricted χ^2_d	P-value
H_{WPPP}	1	3.47	0.062	1.67	0.197
H_{WUIP}	2	11.69	0.003	3.80	0.150
$H_{\text{WPPP}} \cap H_{\text{WUIP}}$	3	26.83	0.000	5.49	0.139
H_{PPP}	3	19.76	0.000	26.61	0.000
H_{UIP}	3	13.58	0.004	3.96	0.266
$H_{\text{PPP}} \cap H_{\text{UIP}}$	6	36.05	0.000	33.88	0.000
$H^*_{\text{PPP}} \cap H_{\text{UIP}}$	5	16.24	0.006	7.53	0.184

d is the number of overidentifying linear restrictions on $\boldsymbol{\beta}$

Table 12: Likelihood ratio statistics for tests of linear restrictions on $\boldsymbol{\beta}$ in the partial VECM(1) with no intercepts and intercepts restricted to the cointegration space

H_{PPP} in the model without intercepts does not hold over the given time period. Furthermore, the pure version of UIP which implies that the interest rate differential between South Africa and the United States should be a zero mean $I(0)$ process is firmly rejected in the model without intercepts. Instead, it would appear necessary to include an intercept term in the cointegrating relations in order to allow the interest rate differential to have a non-zero mean capturing the collective effects of transactions costs, riskiness and speculation. Such a non-zero mean is permitted in the VECM(1) with restricted intercepts as well as in the previous VECM(1) which allowed for intercepts in both the cointegration space and the data generating process. In both cases, H_{UIP} is clearly not rejected. The empirical evidence would therefore seem to justify the inclusion of an intercept term at least in the cointegration space. Moreover, the findings based on the partial models with restricted and unrestricted intercepts are coherent.

The estimates of the parameters in the partial VECM(1) with unrestricted intercepts and $\boldsymbol{\beta}$ constrained under $H^*_{\text{PPP}} \cap H_{\text{UIP}}$ are presented in Table 13. The two rows corresponding to $\xi^{(1)}_{t-1}$ and $\xi^{(2)}_{t-1}$ include the elements of the loading matrix $\boldsymbol{\alpha}$, which adjust the endogenous variables in response to lagged disequilibria in the cointegrating relations. The first cointegrating vector in this model may be regarded as an imperfect version of relative PPP, where the cointegrating coefficient on the exchange rate variable is not exactly equal to -1 due to transactions costs, trade barriers, non-tradable goods and other market imperfections which distort this relation. The significantly negative

	Δp_t^{SA}		Δp_t^{US}		$\Delta e_t^{R/\$}$		Δr_t^{SA}	
ν	-0.003	(0.006)	0.005	(0.002)	0.032	(0.036)	2.048‡	(0.629)
$\xi_{t-1}^{(1)}$	-0.017‡	(0.005)	-0.003	(0.005)	-0.016	(0.026)	1.018†	(0.455)
$\xi_{t-1}^{(2)}$	-0.0005	(0.0003)	-0.0008‡	(0.0003)	-0.002	(0.001)	-0.118‡	(0.025)
Δp_{t-1}^{SA}	0.344‡	(0.074)	-0.024	(0.073)	-0.435	(0.409)	2.035	(7.220)
Δp_{t-1}^{US}	-0.109	(0.086)	0.188†	(0.086)	-1.203†	(0.480)	-13.464	(8.461)
$\Delta e_{t-1}^{R/\$}$	0.056‡	(0.016)	-0.020	(0.016)	0.176	(0.090)	0.533	(1.589)
Δr_{t-1}^{SA}	-0.001	(0.0008)	0.0007	(0.0008)	-0.001	(0.004)	0.403‡	(0.076)
Δr_{t-1}^{US}	-0.002	(0.001)	0.002	(0.001)	0.005	(0.006)	0.046	(0.114)
Δg_t	-0.017	(0.010)	0.030‡	(0.010)	-0.180‡	(0.058)	-3.297‡	(1.023)
Δc_t	0.008	(0.006)	0.039‡	(0.006)	-0.022	(0.035)	0.347	(0.610)
Δc_{t-1}	0.020‡	(0.007)	0.010	(0.007)	0.007	(0.040)	0.489	(0.705)
d_t^1	-0.0005	(0.002)	0.003	(0.003)	-0.011	(0.014)	0.293	(0.242)
d_t^2	0.007‡	(0.003)	0.002	(0.002)	0.010	(0.014)	0.164	(0.248)
d_t^3	0.005†	(0.002)	-0.416‡	(0.111)	0.026	(0.014)	0.358	(0.240)
$\hat{\sigma}$	0.010		0.010		0.056		0.993	
R^2	0.558		0.623		0.242		0.383	
χ_4^2 (A)	10.06†		10.86†		10.90†		5.55	
χ_1^2 (F)	1.08		4.99†		7.31‡		9.23‡	
χ_2^2 (N)	2.02		220.37‡		28.77‡		797.74‡	
χ_1^2 (H)	10.02‡		1.38		2.87		0.15	
ARCH(2)	0.02		0.89		5.52		9.87‡	
ARCH(4)	1.44		10.25†		9.45		10.20†	

† and ‡ denote significance at the 5% and 1% levels respectively
Standard errors are given in parentheses

The error correction terms are given as
$\xi_{t-1}^{(1)} = p_{t-1}^{SA} - p_{t-1}^{US} - 0.59 e_{t-1}^{R/\$}$
$\xi_{t-1}^{(2)} = r_{t-1}^{SA} - r_{t-1}^{US}$

Model diagnostics
χ_4^2 (A): Lagrange multiplier test for residual autocorrelation
χ_1^2 (F): Ramsey's RESET test of functional form
χ_2^2 (N): Jarque-Bera test for residual normality
χ_1^2 (H): Heteroscedasticity test based on regression of squared residuals on squared fitted values
ARCH(2): χ_2^2 statistic for test of second-order autoregressive conditional heteroscedasticity in the residuals
ARCH(4): χ_4^2 statistic for test of fourth-order autoregressive conditional heteroscedasticity in the residuals

Table 13: Parameter estimates in the VECM(1) with unrestricted intercepts

adjustment coefficient for this PPP relation in the Δp_t^{SA} equation implies that when relative prices exceed the nominal exchange rate, the general price level in South Africa adjusts downwards to restore equilibrium as expected. Note, however, that the negative coefficients corresponding to this relation in the Δp_t^{US} and $\Delta e_t^{R/\$}$ equations are not of the expected sign, although they are statistically nonsignificant. The nonsignificance of the former adjustment coefficient is not surprising given that prices in a large economy such as the United States are unlikely to adjust to the PPP relation between itself and a small economy such as South Africa.

Notice that there is also a statistically significant and positive adjustment coefficient corresponding to the PPP relation in the South African interest rate equation. If relative prices exceed the nominal exchange rate, investors may anticipate a future depreciation (increase) in the nominal rand/dollar exchange rate. However, this would make South African investments less attractive relative to foreign investments such that domestic interest rates may rise in future in order to restore international competitiveness in the asset market. An alternative explanation for this positive coefficient is that during the later part of the time period considered, monetary policy in South Africa dictated an interest rate hike following a period of inflationary pressure where relative prices are likely to have exceeded the nominal exchange rate. Consequently, a positive PPP adjustment coefficient might be expected in the interest rate equation.

The significantly negative UIP adjustment coefficient in the interest rate equation implies that when the interest rate differential exceeds the risk premium for investing in South Africa, interest rates in South Africa adjust downwards to restore equilibrium. Although statistically nonsignificant, the negative UIP adjustment coefficient in the exchange rate equation suggests that a higher interest rate in South Africa relative to the United States will lead to capital inflow into South Africa, which will in turn cause the rand/dollar exchange rate to appreciate (decrease). Hence, arbitrage in the asset market restores UIP as expected.

Note also that price levels in the United States adjust significantly to departures from UIP. In fact, this is the very reason why Δp_t^{US} cannot be treated as weakly exogenous for $(\boldsymbol{\alpha}, \boldsymbol{\beta})$. The significantly negative UIP adjustment coefficient in the Δp_t^{US} equation is not entirely unanticipated. High interest rates in South Africa in excess of the country's risk premium relative to the United States would lead investors to favour the more attractive South African economy, inducing capital flows from the United States to South Africa. However, the small South African economy alone is unlikely to invite capital inflows from the United States that are of such a magnitude as to significantly reduce the US money supply and hence US price levels. It is instead more plausible that the interest rate in South Africa is acting as a proxy for those of emerging markets as a whole. Higher returns on investments in emerging markets could indeed lead to a significant shift in capital from the United States to the emerging economies, thereby leading to a large reduction in the money supply and consequently also prices in the United States.

The residuals of the individual equations comprising the VECM(1) do not appear to exhibit serious autocorrelation. Although there is some evidence of serial correlation in the residuals pertaining to the two price equations and the exchange rate equation at the 5% significance level, it is unlikely that these violations are severe enough to adversely affect inferences. However, highly significant non-normality is evident in all equations except the price equation for South Africa. Whilst the residual distributions of each equation are fairly symmetric, they exhibit rather severe excess kurtosis in the equations for Δp_t^{US} and Δr_t^{SA} and to a much lesser extent $\Delta e_t^{R/\$}$. The residuals of the Δp_t^{SA} equation are significantly heteroscedastic and there is evidence of ARCH effects in the residuals corresponding to the Δp_t^{US} equation and in particular the Δr_t^{SA} equation. As mentioned earlier, however, non-normality and ARCH effects may be anticipated from data collected over several distinct regimes. It is therefore not surprising that Ramsey's RESET test of functional form rejects the null hypothesis of zero mean Gaussian residuals with constant variance in all equations, except that corresponding to domestic price levels.

4.3 The Markov-Switching VECM

The linear VECM presented above provides empirical evidence in favour of two cointegrating relations which may be economically identified as the weak form of PPP and the strict form of UIP. However, the residuals corresponding to the equations in this model clearly indicate that the functional form of the linear VECM is inappropriate. Indeed, this finding might be anticipated in light of the significant monetary and exchange rate regime shifts that took place over the studied time period. Ignoring such regime changes is likely to dilute the evidence in favour of PPP and UIP as equilibrium conditions. Consequently, it would seem desirable to incorporate these regime shifts explicitly in the modelling procedure, which may be achieved by considering a Markov-switching VECM for these data. In doing so, it is hoped that the evidence in favour of the various forms of PPP and UIP will be enhanced and that the residuals of the resulting model will better satisfy the assumptions necessary for sound statistical inference.

The partial VECM(1) presented in Equation (4.8) is considered as the point of departure for the analysis to follow. This previous linear VECM(1) may be adapted to allow for non-linearities due to regime shifts by rewriting it as an MS(m)-VECM(1) process of the following form

$$\Delta \boldsymbol{y}_t^* = \boldsymbol{\nu}(s_t) + \boldsymbol{\alpha}(s_t)\boldsymbol{\beta}'\boldsymbol{y}_{t-1} + \boldsymbol{\Gamma}(s_t)\Delta \boldsymbol{y}_{t-1} + \boldsymbol{\Phi}(s_t)\boldsymbol{D}_t + \boldsymbol{\varepsilon}_t, \tag{4.9}$$

where $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}(s_t))$. Note that since the unrestricted intercept term $\boldsymbol{\nu}(s_t)$ is explicit in this formulation, $\boldsymbol{D}_t = (\Delta g_t, \Delta c_t, \Delta c_{t-1}, d_t^1, d_t^2, d_t^3)'$ with coefficient matrix $\boldsymbol{\Phi}(s_t)$ appropriately redefined. All other parameters are as previously defined in the VECM(1) given by Equation (4.8), but are now allowed to depend upon the state $s_t \in \{1, \ldots, m\}$ at time t. In light of the economic developments that took place in South Africa over the studied time period as described in Section 3.1, it was decided to consider a three-regime and four-regime model for the data. An attempt was also made at fitting a five-regime model, although the estimation procedure failed to converge to a solution. Of course, it need not be the case that all the parameters in this model are regime dependent. In order to determine which parameters may be regarded as regime invariant, the information criteria presented in Table 14 for $m = 3$ and $m = 4$ may be utilised. Both the SBC and HQC prefer the MSIH model formulation, whilst the AIC favours the most complex MSIAH specification for both the three-regime and four-regime models. Note, however, that allowing the coefficient matrices $\boldsymbol{\alpha}$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Phi}$ to vary by regime leads to an enormous increase in the number of parameters to be estimated. Moreover, the estimated regimes obtained by allowing these parameters to

VECM(1)	Parameters	AIC	SBC	HQC
Linear	66	869.23	771.92	829.69
MSI(3)	80	885.83	767.88	837.90
MSIH(3)	100	945.74	798.30	885.83
MSIAH(3)	204	975.73	674.95	853.50
MSI(4)	90	885.16	752.46	831.23
MSIH(4)	120	957.95	781.03	886.06
MSIAH(4)	276	997.66	590.73	832.29

Table 14: Information criteria for Markov-switching model specifications

vary were not economically meaningful. Consequently, only the intercept term $\boldsymbol{\nu}(s_t)$ and residual variance-covariance matrix $\boldsymbol{\Omega}\,(s_t)$ will be regarded as regime dependent, in line with the recommendations of the SBC and HQC.

A decision must also be made with respect to the number of regimes in the time period under analysis. Using the information criteria in Table 14, it is noted that the AIC and HQC favour the MSIH(4)-VECM(1) over the MSIH(3)-VECM(1), whereas the latter is preferable in terms of the SBC. After fitting both of these models, it was decided that the regimes obtained in the estimation of the MSIH(4)-VECM(1) are more consistent with the economic and political developments that took place in South Africa over the studied time period than was the case with the MSIH(3)-VECM(1). Indeed, prior knowledge of this time period suggests the presence of four distinct monetary and exchange rate regimes as outlined in Section 3.3. Accordingly, the four-regime model will be adopted, motivated by the data in terms of the AIC and HQC as well as the known economic events that took place during the studied time period.

The computer code employed to estimate the MSIH(4)-VECM(1) using the MS-VAR 1.30 package in Ox 3.00 is given in Appendix D. Note that this software package was actually designed to estimate MS-VAR models of which the MS-VECM is simply a reparameterisation. This fact has two important implications for the estimation of an MS-VECM. Firstly, the software only allows for unrestricted intercepts in the model given by Equation (4.9). This restriction is, however, not of particular concern since the regime-dependent intercepts in the cointegration space can easily be recovered from the unrestricted intercepts $\boldsymbol{\nu}(s_t)$. Secondly, $\xi^{(1)}_{t-1} = p^{SA}_{t-1} - p^{US}_{t-1} - e^{R/\$}_{t-1}$ and $\xi^{(2)}_{t-1} = r^{SA}_{t-1} - r^{US}_{t-1}$ must be specified as exogenous $I(0)$ variables in the model, which presumes that the processes defining the PPP and UIP relations are in fact cointegrated conditional on the underlying regimes. Of course, there is no way to establish this *a priori* since the regimes themselves are unknown. However, Krolzig suggests that if $\xi^{(1)}_{t-1}$ and $\xi^{(2)}_{t-1}$ can be shown to be stationary in a linear VECM, their inclusion as $I(0)$ variables in the MS-VECM is likely to be valid [33]. Since the previous modelling procedure established firm evidence in favour of strict UIP as a long-run relation, treating $\xi^{(2)}_{t-1}$ as an exogenous $I(0)$ variable would seem justified. On the other hand, the linear restrictions on the cointegration matrix $\boldsymbol{\beta}$ necessary to establish strict PPP were rejected in the linear VECM. However, this does not preclude strict PPP from holding in the MS-VECM conditional on the underlying regimes. Indeed, the linear model did provide empirical support for the long-run relation $p^{SA}_{t-1} - p^{US}_{t-1} - 0.59e^{R/\$}_{t-1}$ and it is hoped that restricting the coefficient on the exchange rate to -1 will be permissable in the MS-VECM such that strict PPP will hold after conditioning on the regimes. Consequently, it will be assumed for the moment that $\xi_{t-1}^{(1)}$ is an $I(0)$ exogenous variable conditional on the underlying regimes. An attempt will be made to confirm this assumption post estimation.

4.3.1 Estimated Regimes

The Markov chain governing the regime generating process may be described by the following estimated transition probability matrix

$$\mathbb{P} = \begin{bmatrix} 0.928 & 0.000 & 0.000 & 0.072 \\ 0.001 & 0.764 & 0.120 & 0.115 \\ 0.000 & 0.042 & 0.958 & 0.000 \\ 0.139 & 0.075 & 0.000 & 0.786 \end{bmatrix}.$$

The large probabilities on the diagonal indicate that once the chain enters a particular state, it is likely to stay in that state for some time before moving to another state. The estimated regimes are therefore persistent, but not absorbing. This result might be expected of economic regimes which are defined in terms of distinct policy decisions and frameworks such that exogenous shocks are unlikely to induce regime shifts. Note that the elements in $\mathbb{P}$ which appear as 0.000 are only zero due to rounding. In fact, none of the elements in $\mathbb{P}$ are exactly equal to zero, such that it is possible to move from any regime to any other regime. The regime generating process is therefore said to be irreducible. Furthermore, $\mathbb{P}$ has eigenvalues of 1, 0.97, 0.81 and 0.65, which implies that the Markov chain is also ergodic as is required of an MS-VECM.

The regime classifications for the MSIH(4)-VECM(1) based on the smoothed regime probabilities defined in Section 2.4.2 and estimated by the EM algorithm are given in Table 15. The smoothed probabilities of being in each regime at each time point are illustrated in Figure 5. The four estimated regimes correspond roughly to the 1970s and 2005 to 2007 (regime 1), the mid 1980s (regime 2), the late 1980s to late 1990s (regime 3) and the late 1990s to 2005 (regime 4). Indeed, these periods would appear to represent the distinct monetary and exchange rate regimes in South Africa outlined in Section 3.3.

Regime 1 spans roughly 1972 to the end of 1981, but also includes the last two years of the studied time period from mid-2005 to 2007. This regime is

Regime 1	Regime 2	Regime 3	Regime 4
1972:1–1975:3	1982:3–1983:3	1983:4–1984:2	1975:4–1975:4
1976:1–1981:4	1984:3–1986:3	1986:4–1998:2	1982:1–1982:2
2003:3–2003:4	1998:3–1998:4		1999:1–2001:4
2005:3–2007:1	2002:1–2002:1		2002:2–2003:2
			2004:1–2005:2

Table 15: Regime classifications in the MSIH(4)-VECM(1)

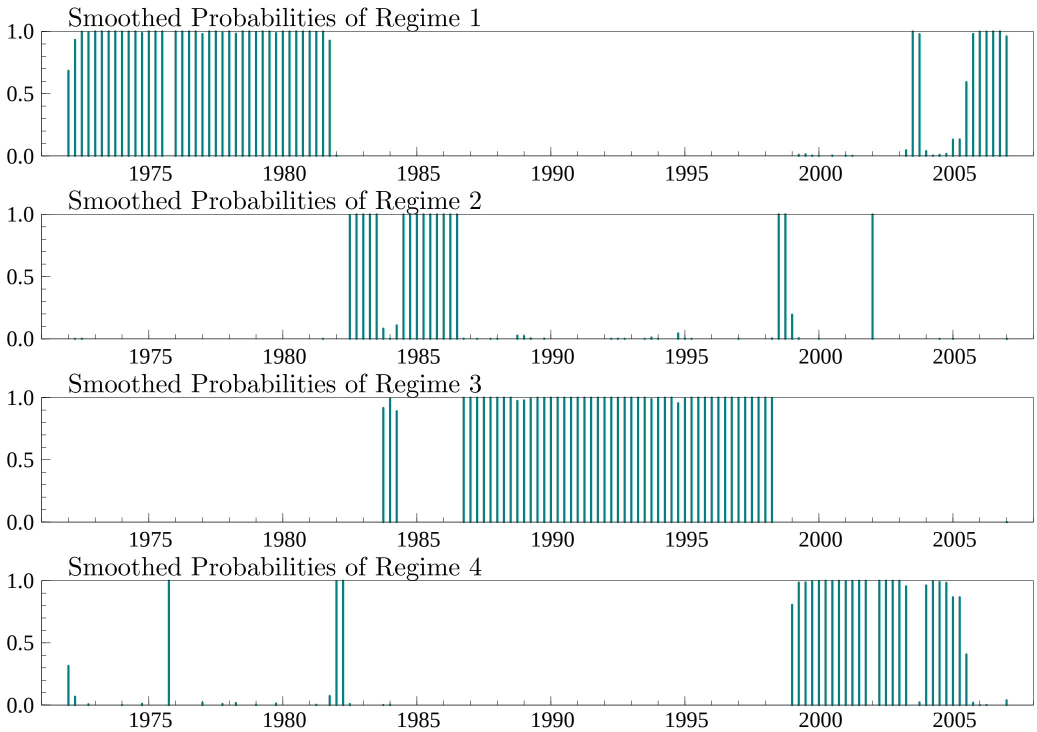

Figure 5: Smoothed probabilities of each regime in the MSIH(4)-VECM(1)

characterised by a fairly stable nominal rand/dollar exchange rate. Following the abolishment of the Bretton Woods fixed exchange rate system at the end of 1971, the rand was pegged to either the US dollar or the pound sterling until 1979. As a direct consequence, movements in the exchange rate took the form of discrete step-changes, as is evident from the middle panel of Figure 1 in Section 3.1. From 2005 to 2007, the currently floating rand/dollar exchange rate stabilised considerably. This later period may therefore be regarded as similar to the 1970s in terms of exchange rate stability, which would appear to motivate its classification as part of regime 1.

Regime 2 covers roughly the period from 1982 to the end of 1986, which might be regarded as a period of substantial political unrest and uncertainty in South Africa. The fixed exchange rate system of the 1970s was finally replaced with a managed float exchange rate system in the early 1980s. A dual exchange rate system was in effect until 1983, at which point the exchange rates were unified for two years before reverting back to the dual exchange rate system in 1985. The rand depreciated sharply during this regime due to the gold price decline in 1983 and the debt crisis of 1985. The real exchange rate was highly volatile during the 1980s, possibly as a result of attempts to maintain a stable real rand gold price over this period.

Monetary policy also shifted from the liquid asset system of the 1970s to a cash reserves system of accommodation in regime 2. Interest rates were extremely volatile during this second regime, as the Reserve Bank juggled between stimulating domestic demand in an atmosphere of investor uncertainty and counteracting rising prices, with inflation exceeding 10% throughout the 1980s. Interestingly, the last two quarters of 1998 and the first quarter of 2002 are also classified as regime 2. Indeed, these too were periods characterised by immense uncertainty amongst investors brought about by the international financial crises which affected emerging markets in the immediately preceding quarters.

The third estimated regime covers the period from 1987 to 1998, which coincides roughly with Dr Chris Stals' term as governor of the South African Reserve Bank. In contrast to the second regime, this regime is characterised by a fairly stable real exchange rate. The stability of the real exchange rate might be attributed to less stringent controls over the real rand gold price and direct intervention by the Reserve Bank in times of currency crises. The dual exchange rate system, which was reintroduced in 1985, was in effect for most of this period before the financial rand was finally terminated in 1995. A series of foreign debt repayments were also made over this period following the financial sanctions imposed against South Africa in 1985. After temporarily relaxing capital controls in the early 1980s, these controls were once again tightened for most of this period. Pre-announced monetary targets were used for the first time from 1986 to be achieved indirectly by adjusting interest rates. Nonetheless, high interest rates are a defining feature of regime 3 for the most part.

A repurchase system of accommodation was employed throughout regime 4, which covers most of the period from 1999 to 2005. This regime also saw the introduction of a formal inflation-targeting policy in South Africa, amidst increased liberalisation efforts by the South African Reserve Bank. Consequently, inflation declined over this period. Interest rates too began to fall as inflation came under control and as South Africa's risk profile improved under the new democratic government. A sharp real depreciation in the currency was observed in the first half of this regime, although the real exchange rate recovered substantially between 2002 and 2006.

4.3.2 Testing for Cointegration in the MS-VECM

As mentioned earlier, there is currently no formal procedure for testing for cointegration in a Markov-switching model. Consequently, the estimated regimes presented in Table 15 were instead included as exogenous dummy variables in the linear VECM(1) as follows

$$\Delta \boldsymbol{y}_t^* = \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{y}_{t-1} + \boldsymbol{\Gamma}\Delta\boldsymbol{y}_{t-1} + \boldsymbol{\Phi}\boldsymbol{D}_t + \boldsymbol{\nu} + \sum_{i=1}^{3} \boldsymbol{\eta}_i \vartheta_{it} + \boldsymbol{\varepsilon}_t, \tag{4.10}$$

where

$$\vartheta_{it} = \begin{cases} 1 \text{ if } s_t = i \\ 0 \text{ otherwise} \end{cases}$$

with coefficient vector $\boldsymbol{\eta}_i$ for $i = 1, 2, 3$. Note that the intercept term $\boldsymbol{\nu}$ again appears explicitly in this model formulation such that $\boldsymbol{D}_t = (\Delta g_t, \Delta c_t, \Delta c_{t-1}, d_t^1, d_t^2, d_t^3)'$ in Equation (4.10). The inclusion of the dummy variables ϑ_{1t}, ϑ_{2t} and ϑ_{3t} in the linear model with an unrestricted intercept $\boldsymbol{\nu}$ effectively allows for a regime-dependent intercept in the cointegration space and the data generating process. In this way, the cointegrating rank of $\boldsymbol{\beta}$ and the validity of the linear restrictions defining the weak and strict forms of the PPP and UIP relations might be tested within the standard VECM framework *conditional on* the estimated regimes. Of course, the VECM(1) in Equation (4.10) does not allow the residual variance-covariance matrix to depend upon the underlying regimes as is the case in the MSIH(4)-VECM(1). Nonetheless, this procedure should shed some light on whether PPP and UIP hold as long-run relations in the Markov-switching model.

The cointegrating rank of the VECM(1) with regime dummies given by Equation (4.10) was tested using Johansen's $\lambda_{\max}$ and trace test statistics together with the information criteria. The former are presented in Table 16, whilst the latter are given in Table 17.

Both the $\lambda_{\max}$ and trace statistics clearly indicate two cointegrating relations in this model at the 5% significance level. Recall that these statistics only supported a single cointegrating relation in the partial model without regime dummies at this significance level. In that case, the hypothesis of two cointegrating relations was found to be only marginally nonsignificant at the 10% level. The VECM(1) with regime dummies would therefore appear to be more supportive of two cointegrating relations relative to the model without regime dummies based on Johansen's statistics for the test of cointegrating rank.

The AIC favours three cointegrating relations, whilst the SBC now suggests a cointegrating rank of two. The HQC also provides evidence in favour of two or three cointegrating relations, with little difference between these two

H_0	H_1	$\lambda_{\max}$	95% CV	90% CV
$r = 0$	$r = 1$	51.99	30.71	28.27
$r \leq 1$	$r = 2$	33.49	24.59	22.15
$r \leq 2$	$r = 3$	13.85	18.06	15.98
$r \leq 3$	$r = 4$	2.08	11.47	9.53

H_0	H_1	Trace	95% CV	90% CV
$r = 0$	$r \geq 1$	101.41	58.63	54.84
$r \leq 1$	$r \geq 2$	49.42	38.93	35.88
$r \leq 2$	$r \geq 3$	15.93	23.32	20.75
$r \leq 3$	$r = 4$	2.08	11.47	9.53

Table 16: Johansen's test statistics for selection of cointegrating rank in the partial VECM(1) with regime dummies

Rank	AIC	SBC	HQC
$r = 0$	865.30	776.84	829.35
$r = 1$	883.29	783.03	842.55
$r = 2$	894.04	784.93	849.70
$r = 3$	896.96	781.96	850.23
$r = 4$	896.00	778.05	848.07

Table 17: Information criteria for selection of cointegrating rank in the partial VECM(1) with regime dummies

specifications in terms of this criterion. Consequently, a cointegrating rank of two will once again be specified, which would appear to garner support from the data as well as being consistent with the economic theories of PPP and UIP.

Having established the presence of two cointegrating vectors in the VECM(1) with regime dummies, the weak and strict forms of PPP and UIP were tested individually and jointly by constraining the cointegration matrix $\boldsymbol{\beta}$ in the same manner as was considered for the VECM(1) without regime dummies and testing these linear restrictions for data admissibility. The results are presented in Table 18. As a first observation, it should be noted that the p-values associated with all the tests of the linear restrictions on $\boldsymbol{\beta}$ considered in this table are larger than those presented in Table 11 corresponding to the equivalent model without regime dummies. Hence, the inclusion of the regime dummies in the VECM(1) leads to an overall improvement in the empirical evidence in support of all the forms of PPP and UIP considered here.

More specifically, the VECM(1) with regime dummies does not reject UIP as an individual long-run relation even in its strictest form, as was the case in the model without regime dummies. This result is also true of the weaker form of PPP implied under H_{WPPP}, which is also supported by both models. Unsurprisingly, weak PPP and weak UIP can therefore not be rejected as joint equilibrium conditions.

Some support for the strict form of PPP is also evident in the model with regime dummies. Notice that the overwhelmingly significant χ^2_3 statistic of 18.81 obtained when testing H_{PPP} in the VECM(1) without regime dummies drops considerably when regime dummies are added to a value of 9.34, which is statistically significant at the 5% level, but not at the 1% level. Moreover,

Exactly Identified

$$\begin{bmatrix} 1.00 & 0.00 \\ & \\ -1.28 & 4.00 \\ (0.10) & (3.27) \\ -0.77 & -1.40 \\ (0.05) & (1.69) \\ 0.00 & 1.00 \\ & \\ 0.01 & -1.00 \\ (0.01) & (0.28) \end{bmatrix}$$

$\log L = 968.04$

H_{WPPP}

$$\begin{bmatrix} 1.00 & 0.00 \\ & \\ -1.23 & 3.63 \\ (0.07) & (3.24) \\ -0.79 & -1.22 \\ (0.04) & (1.69) \\ 0.00 & 1.00 \\ & \\ 0.00 & -0.95 \\ & (0.27) \end{bmatrix}$$

$\log L = 967.89$
$\chi^2_1 = 0.30$
P-value $= 0.584$

H_{WUIP}

$$\begin{bmatrix} 1.00 & 0.00 \\ & \\ -1.23 & 0.00 \\ (0.10) & \\ -0.78 & 0.00 \\ (0.05) & \\ 0.00 & 1.00 \\ & \\ 0.00 & -0.80 \\ (0.01) & (0.20) \end{bmatrix}$$

$\log L = 967.18$
$\chi^2_2 = 1.72$
P-value $= 0.423$

$H_{\text{WPPP}} \cap H_{\text{WUIP}}$

$$\begin{bmatrix} 1.00 & 0.00 \\ & \\ -1.21 & 0.00 \\ (0.08) & \\ -0.79 & 0.00 \\ (0.04) & \\ 0.00 & 1.00 \\ & \\ 0.00 & -0.78 \\ & (0.19) \end{bmatrix}$$

$\log L = 967.14$
$\chi^2_3 = 1.80$
P-value $= 0.614$

H_{PPP}

$$\begin{bmatrix} 1.00 & 0.00 \\ & \\ -1.00 & -0.47 \\ & (4.02) \\ -1.00 & 2.27 \\ & (2.73) \\ 0.00 & 1.00 \\ & \\ 0.00 & -0.91 \\ & (0.27) \end{bmatrix}$$

$\log L = 963.37$
$\chi^2_3 = 9.34$
P-value $= 0.025$

H_{UIP}

$$\begin{bmatrix} 1.00 & 0.00 \\ & \\ -1.23 & 0.00 \\ (0.10) & \\ -0.78 & 0.00 \\ (0.05) & \\ 0.00 & 1.00 \\ & \\ 0.01 & -1.00 \\ (0.01) & \end{bmatrix}$$

$\log L = 966.72$
$\chi^2_3 = 2.63$
P-value $= 0.453$

$H_{\text{PPP}} \cap H_{\text{UIP}}$

$$\begin{bmatrix} 1.00 & 0.00 \\ & \\ -1.00 & 0.00 \\ & \\ -1.00 & 0.00 \\ & \\ 0.00 & 1.00 \\ & \\ 0.00 & -1.00 \\ & \end{bmatrix}$$

$\log L = 961.65$
$\chi^2_6 = 12.77$
P-value $= 0.047$

$H^*_{\text{PPP}} \cap H_{\text{UIP}}$

$$\begin{bmatrix} 1.00 & 0.00 \\ & \\ -1.00 & 0.00 \\ & \\ -0.88 & 0.00 \\ (0.04) & \\ 0.00 & 1.00 \\ & \\ 0.00 & -1.00 \\ & \end{bmatrix}$$

$\log L = 964.64$
$\chi^2_5 = 6.80$
P-value $= 0.236$

Asymptotic standard errors are given in parentheses

Table 18: Estimated cointegration matrix $\boldsymbol{\beta}$ with linear restrictions in the partial VECM(1) with regime dummies

Gredenhoff and Jacobson have shown through Monte Carlo experimentation that the χ^2 statistic associated with this likelihood ratio test can be substantially upwardly biased when the length of the observed data sequence is small, such that the observed p-value of 0.025 for this hypothesis test is likely to be an underestimate of the true p-value [18]. Consequently, strict PPP cannot be firmly rejected as an individual long-run relation in the VECM(1) with regime dummies.

As was the case in the model without regime dummies, the linear restrictions implied under $H^*_{\text{PPP}} \cap H_{\text{UIP}}$ are not rejected in the model with regime dummies. Furthermore, the only unrestricted cointegrating coefficient β_{13} corresponding to the exchange rate in the PPP relation is -0.88 in the model with regime dummies, which is closer to the desired -1 under strict PPP than the -0.59 obtained in the VECM(1) without regime dummies. Constraining this coefficient to -1 to obtain $H_{\text{PPP}} \cap H_{\text{UIP}}$ would therefore seem more plausible in the latter model. Recall that this hypothesis of strict PPP and strict UIP as joint long-run relations was decisively rejected in the previous VECM(1) without regime dummies. Indeed, this is not the case in the model with regime dummies. The likelihood ratio test of $H_{\text{PPP}} \cap H_{\text{UIP}}$ in the model given by Equation (4.10) yielded a p-value of 0.047. At the 5% significance level, the joint hypothesis of strict PPP and strict UIP is therefore only marginally rejected, although it should be borne in mind that the use of a 5% significance level is merely a “convenient convention” [14]. Furthermore, the simulations of Gredenhoff and Jacobson suggest that the true p-value of this test is likely to be larger than that reported here such that it may very well exceed the conventional 0.05 cut-off [18]. It may therefore be concluded that the strict forms of PPP and UIP do exist jointly as long-run relations conditional on the underlying regimes. Failure to condition on the regimes, however, leads to an outright rejection of $H_{\text{PPP}} \cap H_{\text{UIP}}$.

4.3.3 Model Parameter Estimates

The parameter estimates of the MSIH(4)-VECM(1) are given in Table 19. Comparing the long-run adjustment coefficients in the MS-VECM with those obtained previously in the linear VECM given in Table 13 is enlightening.
Firstly, note that when relative prices exceed the nominal exchange rate, the adjustment coefficients in the MS-VECM imply that domestic prices fall, US prices rise and the nominal rand/dollar exchange rate depreciates to restore PPP. The signs of the PPP adjustment coefficients in the price and exchange rate equations are therefore as expected. Both the PPP adjustment

	Δp_t^{SA}		Δp_t^{US}		$\Delta e_t^{R/\$}$		Δr_t^{SA}	
$\hat{\nu}(s_t = 1)$	0.011	(0.009)	0.034‡	(0.008)	0.203‡	(0.052)	-0.433	(0.737)
$\hat{\nu}(s_t = 2)$	0.021†	(0.009)	0.029‡	(0.007)	0.251‡	(0.059)	-0.068	(0.940)
$\hat{\nu}(s_t = 3)$	0.013	(0.008)	0.028‡	(0.007)	0.202‡	(0.047)	0.294	(0.671)
$\hat{\nu}(s_t = 4)$	0.004	(0.010)	0.031‡	(0.008)	0.240‡	(0.058)	-0.024	(0.826)
$\xi_{t-1}^{(1)}$	-0.007	(0.006)	0.012‡	(0.005)	0.115‡	(0.032)	-0.404	(0.463)
$\xi_{t-1}^{(2)}$	-0.001‡	(0.0003)	-0.001‡	(0.0002)	-0.0002	(0.001)	-0.113‡	(0.020)
Δp_{t-1}^{SA}	0.333‡	(0.049)	0.004	(0.046)	-0.124	(0.273)	3.042	(4.055)
Δp_{t-1}^{US}	-0.110	(0.073)	0.040	(0.077)	-0.663†	(0.309)	-11.650†	(4.658)
$\Delta e_{t-1}^{R/\$}$	0.037‡	(0.013)	-0.003	(0.009)	0.229‡	(0.071)	0.641	(1.088)
Δr_{t-1}^{SA}	0.0003	(0.001)	0.0003	(0.0004)	0.005	(0.004)	0.496‡	(0.057)
Δr_{t-1}^{US}	-0.001	(0.001)	0.002‡	(0.001)	0.004	(0.004)	-0.019	(0.062)
Δg_t	-0.030‡	(0.011)	0.011	(0.009)	-0.069	(0.039)	-1.614‡	(0.604)
Δc_t	0.018‡	(0.005)	0.048‡	(0.004)	-0.036	(0.023)	0.111	(0.355)
Δc_{t-1}	0.029‡	(0.006)	0.018‡	(0.005)	0.030	(0.024)	0.475	(0.369)
d_t^1	0.001	(0.002)	0.003†	(0.001)	-0.001	(0.009)	0.108	(0.134)
d_t^2	0.009‡	(0.002)	0.003†	(0.001)	0.016	(0.009)	-0.079	(0.138)
d_t^3	0.001	(0.002)	-0.001	(0.001)	0.012	(0.009)	-0.141	(0.151)
$\hat{\sigma}(s_t = 1)$	0.013		0.014		0.031		0.479	
$\hat{\sigma}(s_t = 2)$	0.012		0.003		0.105		2.427	
$\hat{\sigma}(s_t = 3)$	0.007		0.005		0.034		0.508	
$\hat{\sigma}(s_t = 4)$	0.006		0.010		0.063		0.601	
R^2	0.510		0.603		0.249		0.329	
χ_4^2 (A)	11.51†		3.57		6.23		2.91	
χ_2^2 (N)	2.96		12.92‡		5.98		0.48	
χ_1^2 (H)	1.29		0.14		0.01		2.07	
ARCH(2)	0.27		1.46		0.19		0.17	
ARCH(4)	0.41		4.68		0.98		2.17	

† and ‡ denote significance at the 5% and 1% levels respectively

Standard errors are given in parentheses

The error correction terms are given as

$\xi_{t-1}^{(1)} = p_{t-1}^{SA} - p_{t-1}^{US} - e_{t-1}^{R/\$}$

$\xi_{t-1}^{(2)} = r_{t-1}^{SA} - r_{t-1}^{US}$

Model diagnostics based on standardised residuals

χ_4^2 (A): Lagrange multiplier test for residual autocorrelation

χ_2^2 (N): Jarque-Bera test for residual normality

χ_1^2 (H): Heteroscedasticity test based on regression of squared residuals on squared fitted values

ARCH(2): χ_2^2 statistic for test of second-order autoregressive conditional heteroscedasticity in the residuals

ARCH(4): χ_4^2 statistic for test of fourth-order autoregressive conditional heteroscedasticity in the residuals

Table 19: Parameter estimates in the MSIH(4)-VECM(1)

coefficients in the Δp_t^{US} and $\Delta e_t^{R/\$}$ equations are statistically different from zero in this model. Recall that in the linear VECM, the PPP adjustment coefficient was only statistically significant in the Δp_t^{SA} equation, whilst the nonsignificant adjustment coefficients in the Δp_t^{US} and $\Delta e_t^{R/\$}$ equations were of the incorrect sign.

Both the linear and Markov-switching vector error correction models yield a highly significant, negative UIP adjustment coefficient in the interest rate equation. This finding confirms economic theory, which suggests that domestic interest rates should adjust downwards when the interest rate differential exceeds the relative risk premium. Significantly negative UIP adjustment coefficients are also observed in both the price equations. Indeed, higher interest rates in South Africa will restrict the purchasing power of consumers and the resulting drop in demand should ease inflationary pressure in the country. Furthermore, if South African interest rates are higher than those in the United States after taking into account expected depreciation and country-specific risk, South Africa will attract foreign investment from the United States thereby leading to a reduction in the US money supply. Prices in the United States may therefore be expected to fall, particularly if the South African interest rate is considered as a proxy for those of emerging economies as a whole. Note that the increased demand for South African investments would also put pressure on the rand/dollar exchange rate to appreciate (decrease), which would appear to motivate the negative, although statistically nonsignificant, UIP adjustment coefficient in the exchange rate equation.

The regime-dependent intercepts in the cointegration space are also of substantive interest. Since the MSIH(4)-VECM(1) was estimated with unrestricted intercepts, the intercepts in the cointegration space must be derived analytically as

$$\boldsymbol{\Xi}^* = (\boldsymbol{\alpha}'\boldsymbol{\alpha})^{-1}\boldsymbol{\alpha}'\boldsymbol{\Xi},$$

with $\boldsymbol{\Xi} = (\boldsymbol{\nu}(s_t = 1), \boldsymbol{\nu}(s_t = 2), \boldsymbol{\nu}(s_t = 3), \boldsymbol{\nu}(s_t = 4))$ and $\boldsymbol{\Xi}^* = (\boldsymbol{\nu}^*(s_t = 1), \boldsymbol{\nu}^*(s_t = 2), \boldsymbol{\nu}^*(s_t = 3), \boldsymbol{\nu}^*(s_t = 4))$, where $\boldsymbol{\nu}^*(s_t = i)$ is the vector of intercepts in the cointegration space for regime $i = 1, 2, 3, 4$. Solving for $\boldsymbol{\Xi}^*$ in the MSIH(4)-VECM(1) yields

$$\boldsymbol{\Xi}^* = \begin{bmatrix} 1.77 & 2.17 & 1.74 & 2.07 \\ -2.49 & -7.14 & -8.84 & -7.20 \end{bmatrix},$$

where the first row includes the intercepts in the PPP relations and the second row includes the intercepts in the UIP relations for each of the four regimes. Hence, the estimated relative PPP and UIP relations conditional on the underlying regimes are given as follows.

	PPP	UIP
Regime 1	$E_t^{R/\$} = 5.85 \times P_t^{SA}/P_t^{US}$	$r_t^{SA} - r_t^{US} = 2.49$
Regime 2	$E_t^{R/\$} = 8.72 \times P_t^{SA}/P_t^{US}$	$r_t^{SA} - r_t^{US} = 7.14$
Regime 3	$E_t^{R/\$} = 5.72 \times P_t^{SA}/P_t^{US}$	$r_t^{SA} - r_t^{US} = 8.84$
Regime 4	$E_t^{R/\$} = 7.94 \times P_t^{SA}/P_t^{US}$	$r_t^{SA} - r_t^{US} = 7.20$

The relative PPP equations indicate a real depreciation in the rand/dollar exchange rate from regime 1 to regime 2. Indeed, this would appear to have been the case based on the historical evidence outlined in Section 3.1. The gold price decline of 1983 and the capital outflows due to political uncertainty during regime 2 led to the sharp depreciation of the exchange rate observed here.

The mean real exchange rate appears to have recovered somewhat during regime 3 from 1987 to 1998, as less emphasis was placed on maintaining a stable real rand gold price. This result is hardly surprising given that the real exchange rate reached record highs during the previous regime on the back of major capital withdraws from the politically turbulent South African economy. Hence, despite major depreciations and appreciations in the real exchange rate during regime 3, the level of the real exchange rate was on average lower than that of the previous regime.

Finally, the intercept in the first cointegrating vector implies a depreciation in the real exchange rate from regime 3 to regime 4, where the latter regime covers roughly the period from 1999 to 2005. Indeed, the real exchange rate did depreciate heavily from 1999 to 2002 due to the decline in mineral exports, although the currency has appreciated in more recent times. It is therefore unsurprising that the last two years of the time period considered, that is 2006 and 2007, are classified as regime 1, rather than regime 4. Moreover, the real exchange rate in regime 1 is lower than the real exchange rate in regime 4, implying a real appreciation in the currency in the later years of the studied time period as expected.

Although the intercept term in the UIP equation does to some extent account for imperfect asset substitutability and transactions costs, its interpretation as the risk premium sought by investors in the South African asset market seems natural in this context. Indeed, the political uncertainty associated with regime 2 would have led investors to demand a high risk premium in this regime relative to regime 1. This presumption is confirmed by the mean interest rate differential which rises dramatically from regime 1 to regime 2. The uncertainty experienced during this tumultuous regime is further reflected in the large standard error of the residuals associated with this regime in the interest rate equation in Table 19. Recall that interest rates were raised substantially in 1985 to almost 22% in response to inflationary pressures and then more than halved in the subsequent year to stimulate domestic demand as investors withdrew from South Africa. Hence, this high standard error is not unexpected.

The risk premium demanded by foreign investors continued to rise from regime 2 to regime 3 in light of the additional political uncertainty prior to the elections in 1994. Subsequently, South Africa's risk profile has improved somewhat in the wake of the country's first democratic government and the adoption of more stable monetary and fiscal policies. The country's improved risk profile is reflected in the lower risk premium demanded in regime 4 relative to regime 3. Once again, it should be noted that the years 2006 and 2007 are actually classified as regime 1, which has a very low mean interest rate differential relative to the other three regimes. Consequently, the results of the Markov-switching model indicate that South Africa is perceived as a less risky investment environment post 1994 and that such perceptions have improved considerably in recent years. The real exchange rate and interest rate differential are plotted in Figure 6 with the estimated regimes superimposed.

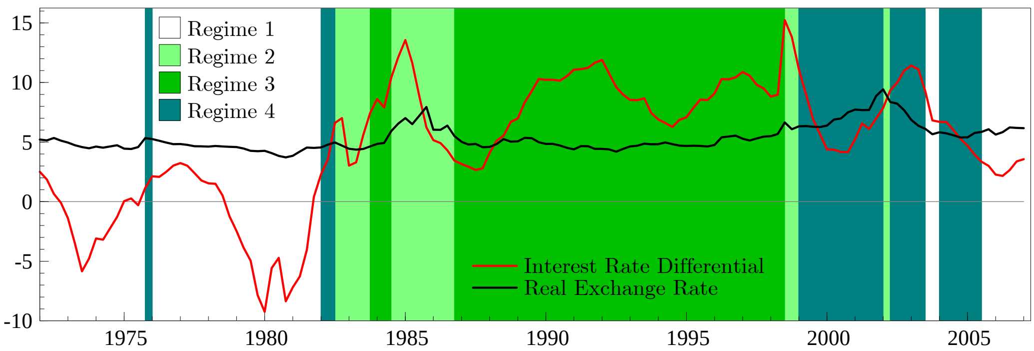

Figure 6: Real exchange rate and interest rate differential with estimated regimes superimposed

4.3.4 Model Diagnostics

The actual and fitted values obtained for each of the four variables in the MSIH(4)-VECM(1) are given in Figure 7. The corresponding residuals are illustrated in Figure 8. The residual plots on the left in Figure 8 display the unstandardised prediction errors associated with the model. Note that the variance of these residuals is assumed to differ between regimes such that heteroscedasticity and ARCH effects might be expected in examining these graphs. Non-constant variance is particularly notable in the case of the unstandardised residuals associated with the interest rate equation, where greater variability is observed in the mid 1980s and again in 1998 relative to the rest of the series. Indeed, the estimated standard deviation of these residuals in regime 2 is 2.427, which is much larger than the standard deviations of 0.479, 0.508 and 0.601 obtained for regimes 1, 3 and 4 respectively.

The standardised residuals on the right hand panel of Figure 8 were obtained by normalising the residuals with respect to their regime-specific standard deviations. As such, these residuals should be homoscedastic over the entire time period under analysis. Examination of the time series plots does not reveal any immediate concern for the violation of this condition. In order to test this assumption more formally, the squared fitted values were regressed on the squared standardised residuals to obtain the χ^2_1 statistics presented in Table 19. The null hypothesis of homoscedasticity cannot be rejected for all four equations. Furthermore, the test statistics associated with the ARCH tests in Table 19 indicate no evidence of second-order or fourth-order autoregressive conditional heteroscedasticity in the residuals. The significant ARCH effects detected particularly in the residuals of the interest rate equation in the linear VECM(1) have therefore been appropriately captured by the Markov-switching residual variance-covariance matrix.

Table 19 also provides the test statistic for the Jarque-Bera test of normality for the standardised residuals in each equation. Note that by specifying regime-dependent residual variances in the model, the *unstandardised* residuals are implicitly assumed to be non-normal, arising instead from a mixture of four Gaussian distributions with zero means, but different variances. The significant non-normality of the residuals in all but the domestic price equation in the linear model may therefore have been expected *a priori* in the presence of such regime-dependent residual variability. In fact, non-normality was found to be most severe in the residuals associated with the interest rate equation in the linear VECM and, unsurprisingly, the residuals corresponding to this same equation in the Markov-switching model exhibit the most notable shifts in residual variance between regimes. After accounting

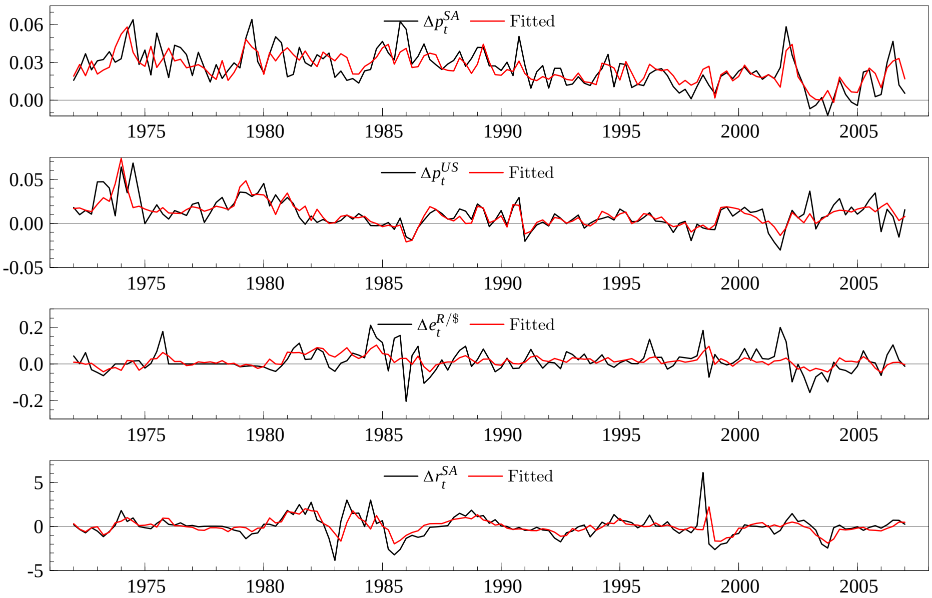

Figure 7: Fitted values according to MSIH(4)-VECM(1)

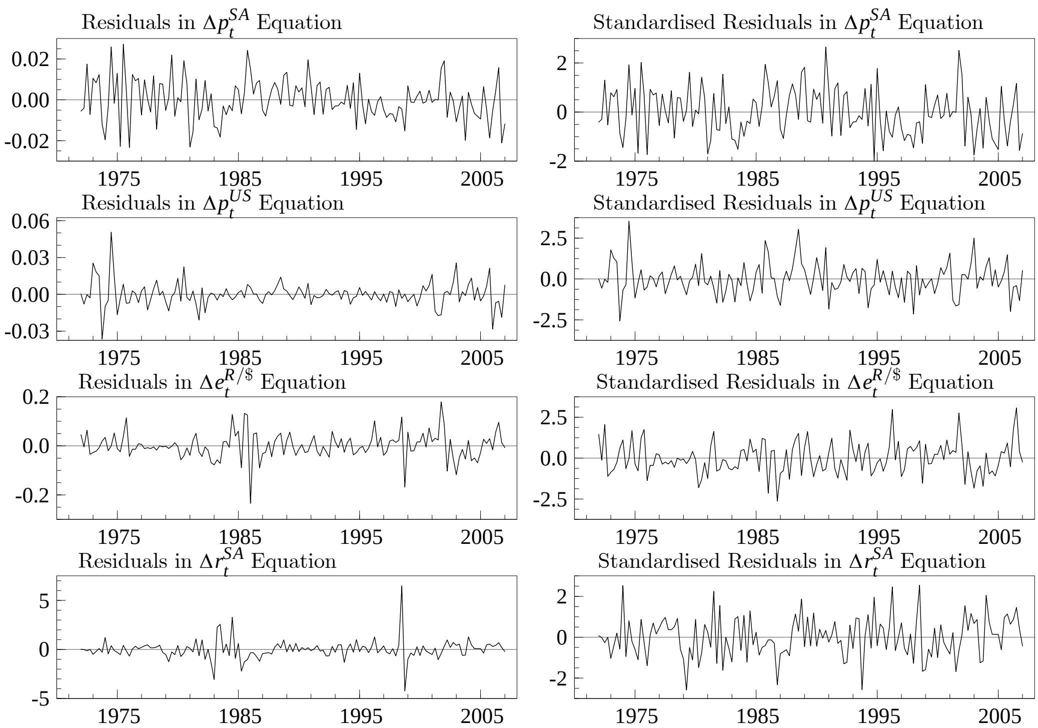

Figure 8: Residuals obtained in the MSIH(4)-VECM(1)

for regime-dependent residual variability, the standardised residuals pass the Jarque-Bera test for normality in all equations except that pertaining to US price levels. Further examination of the empirical density of the standardised residuals associated with the Δp_t^{US} equation given in Figure 9 reveals that the distribution of residuals is only sightly leptokurtic relative to the normal density. These residuals were found to have an excess kurtosis of 1.14, which is not serious enough to adversely affect statistical inferences based on this model. The QQ plots presented in Figure 9 further confirm that the residuals are satisfactorily normal.

Figure 9 also includes plots of the autocorrelation functions (ACF) and partial autocorrelation functions (PACF) of the standardised residuals in each equation. Examination of these plots does not suggest any drastic autocorrelation amongst the residuals. The Lagrange multiplier tests for serial correlation presented in Table 19 confirm this observational conclusion for all equations except that corresponding to domestic prices, where autocorrelation is found to be statistically significant at the 5% level. The autocorrelations (AC), partial autocorrelations (PAC) and Ljung-Box Q statistics for

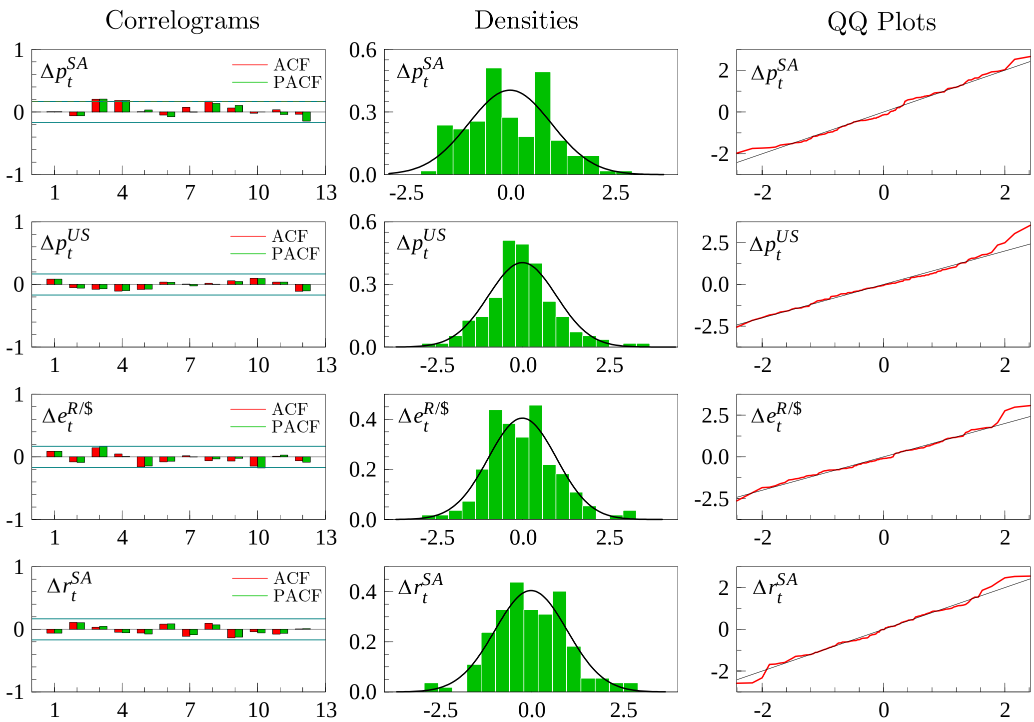

Figure 9: Diagnostic plots for the standardised residuals in the MSIH(4)-VECM(1)

the residuals of this equation are given up to lag 12 in Table 20 [11]. It is noted that the moderately large partial autocorrelations at lags three and four produce significant Q statistics in subsequent lags. However, these partial autocorrelations are not large enough to warrant concern. Indeed, none of the Q statistics are significant at the 1% level. The partial autocorrelations for the residuals of each of the other three equations were also examined and found to be nonsignificant in all cases.

Overall, the properties of the residuals from the Markov-switching model are far more satisfactory than those of the linear VECM, suggesting that the appropriate treatment of regime shifts is necessary for valid statistical inferences. In addition, the MSIH(4)-VECM(1) is preferred to its linear counterpart in terms of all the three information criteria presented in Table 14. The test statistic for the likelihood ratio test comparing the linear and Markov-switching models is $\chi^2_{54} = 285.45$, which firmly rejects the null hypothesis of linearity, even when the upper bound of Davies is invoked [8]. Hence, it would appear that significant non-linearities induced by regime shifts are a pertinent feature of these data and should therefore be incorporated appropriately in the modelling procedure.

Lag	AC	PAC	Q Statistic	P-value*
1	0.006	0.006	0.01	0.939
2	-0.061	-0.061	0.54	0.763
3	0.205	0.206	6.67	0.083
4	0.185	0.184	11.70	0.020
5	0.006	0.034	11.71	0.039
6	-0.049	-0.075	12.07	0.060
7	0.076	0.000	12.94	0.074
8	0.172	0.137	17.43	0.026
9	0.065	0.105	18.07	0.034
10	-0.021	0.002	18.14	0.053
11	0.036	-0.041	18.35	0.074
12	-0.035	-0.145	18.54	0.100

* Test of null hypothesis that there is no autocorrelation up to lag i against the alternative hypothesis that at least one autocorrelation coefficient is statistically different from zero

Table 20: Ljung-Box Q statistics for residuals of Δp_t^{SA} equation in MSIH(4)-VECM(1)

5 Conclusions

In this paper, an attempt was made to establish firm empirical evidence in favour of PPP and UIP as long-run relations when South Africa and the United States are considered as trading partners. Two forms of these equilibrium conditions were defined. Strict purchasing power parity is given by

$$p_t^{SA} - p_t^{US} - e_t^{R/\$} + \mu^{(1)} = \xi_t^{(1)}, \tag{5.1}$$

where p_t^{SA}, p_t^{US} and $e_t^{R/\$}$ are aggregate price levels in South Africa, aggregate price levels in the United States and the nominal rand/dollar exchange rate respectively after taking natural logarithms. The series $\xi_t^{(1)}$ is stationary with a zero mean if PPP holds. If $\mu^{(1)} = 0$, then Equation (5.1) corresponds to the absolute version of purchasing power parity. On the other hand, the relative version of strict PPP is defined by Equation (5.1) when $\mu^{(1)} \neq 0$, which may be expected *a priori* when the coverage and composition of reference baskets of commodities differ between countries.

Transactions costs, tariff and non-tariff trade barriers, the presence of non-tradable goods in commodity bundles and market-specific pricing are some of the reasons why even the relative version of strict PPP may not hold in practice. Consequently, a weaker form of PPP was defined as

$$p_t^{SA} - \beta_1^{(1)} p_t^{US} - \beta_2^{(1)} e_t^{R/\$} + \mu^{(1)} = \xi_t^{(1)},$$

where the coefficients $\beta_1^{(1)}$ and $\beta_2^{(1)}$ may deviate from unity due to the aforementioned market frictions.

The strict form of the uncovered interest parity is given by

$$r_t^{SA} - r_t^{US} - \mu^{(2)} = \xi_t^{(2)},$$

where r_t^{SA} and r_t^{US} are the returns on deposits bearing the same risk and with the same time to maturity in South Africa and the United States respectively. The series $\xi_t^{(2)}$ is a white noise process if UIP holds. The uncovered interest parity, in its purest sense, requires that $\mu^{(2)} = 0$, but the presence of transactions costs, country-specific risk premia and speculative effects imply that the interest rate differential is likely to be non-zero in practice. Furthermore, this strict form of UIP requires that the exchange rate does not exhibit a trend over the time period under analysis. If this condition is violated, the weaker form of UIP given as

$$r_t^{SA} - \beta_1^{(2)} r_t^{US} - \mu^{(2)} = \xi_t^{(2)}$$

is more likely to hold, where the coefficient $\beta_1^{(2)}$ may deviate from unity.

The paper commenced by adopting the standard VECM approach in order to ascertain whether the weak and strict forms of PPP and UIP are supported by the data. Quarterly data were employed from the first quarter of 1972 to the first quarter of 2007. A partial VECM(1) with four equations corresponding to Δp_t^{SA}, Δp_t^{US}, $\Delta e_t^{R/\$}$ and Δr_t^{SA} was constructed after establishing the weak exogeneity of Δr_t^{US} with respect to the cointegrating relations and adjustment coefficients in the model. This model did indeed yield some evidence in favour of PPP and UIP as long-run relations. The overidentifying restrictions on the cointegration matrix necessary to establish the weak forms of PPP and UIP could not be rejected individually or jointly. Furthermore, the strict form of UIP could also not be rejected in this model. On the other hand, strict PPP was firmly rejected, both as an individual relation and as a joint relation together with strict UIP.

These findings were established in a VECM(1) with unrestricted intercepts. It should be noted, however, that the results are dependent upon whether or not an intercept term is included in the model. Excluding the intercept altogether led to a firm rejection of both the weak and strict forms of PPP and UIP, both as individual and joint long-run relations. The only exception in this regard was the weak form of PPP, although the evidence in support of this condition as an individual long-run relation was hardly convincing. This finding is akin to rejecting the absolute version of PPP and the pure form of UIP which assumes perfect asset substitutability. On the other hand, permitting an intercept term confined to the cointegration space leads to the same set of conclusions with respect to the PPP and UIP relations as those outlined above for the model with unrestricted intercepts. An intercept term in the cointegration space or in the cointegration space and the data generating process would therefore appear to be warranted, suggesting that only relative PPP and the UIP relation with a non-zero interest rate differential are empirically relevant.

The conventional VECM approach to testing for PPP and UIP was subsequently adapted to allow for regime shifts in the data. Indeed, the time period under review is characterised by a number of monetary and exchange rate regime changes. It was proposed that these regime shifts lead to persistent changes in the mean natural logarithm of the real exchange rate $\mu^{(1)}$ and the mean interest rate differential $\mu^{(2)}$, thereby undermining the ability of the standard VECM approach to establish PPP and UIP as long-run relations over the entire time period under analysis. A Markov-switching VECM was therefore considered, wherein the parameters of the usual VECM are allowed to depend upon underlying regimes which are modelled by a hidden Markov chain. A four-regime model with regime-dependent intercepts and residual variance-covariance matrix was deemed most appropriate based on both the data through the information criteria and the historical evolution of monetary and exchange rate policy in South Africa.

The estimation of the hidden state at each time point is an enlightening by-product of this model specification. The four estimated regimes obtained cover roughly the 1970s as well as 2005 to 2007, the mid 1980s, the late 1980s to late 1990s and the late 1990s to 2005. These regimes were found to be economically meaningful with respect to the developments in monetary and exchange rate policy that took place in South Africa over the studied time period. The first regime defines a period of exchange rate stability with the rand pegged to either the US dollar or pound sterling throughout the 1970s. The second regime, which covers the mid 1980s, was a period of much political uncertainty in the country. As a result, the rand depreciated sharply on the back of large capital outflows. There was also a shift in monetary policy from a system of commercial bank accommodation based on liquid asset requirements during the first regime to a cash reserves system of accommodation in the second regime. The third regime coincides approximately with the period in which Dr Chris Stals served as governor of the South African Reserve Bank. During this regime, emphasis was placed on achieving a stable real exchange rate, often at the expense of monetary targets. Finally, the last regime from 1999 to 2005 is typified by greater political stability and a firm inflation-targeting policy.

Since there is as yet no formal procedure for testing for cointegration in a Markov-switching model, the estimated regimes were added as dummy variables to the previous linear VECM(1) and cointegration tests performed within this familiar framework. The results were extremely positive, despite the fact that this model does not account for the regime-dependent residual variance structure assumed in the Markov-switching model. The evidence in support of the weak and strict forms of PPP and UIP, both as individual and joint long-run relations, was in all cases more statistically assertive. The weak forms of PPP and UIP were once again established as individual and joint equilibrium conditions, whilst strict UIP could not be rejected as an individual long-run relation. Although strict PPP is rejected as an individual hypothesis at the 5% significance level, this decision is borderline at best with non-rejection of strict PPP at the 1% significance level. Moreover, strict PPP and strict UIP cannot be rejected as joint long-run relations in the VECM(1) with regime dummies. This result confirms the promptings of Johansen and Juselius who first proposed that if PPP is a valid description of international price dependence, information on this relation can be found in both the goods and asset markets [30]. Explicitly accounting for regimes in the modelling procedure would therefore appear to dramatically improve the evidence in support of the PPP and UIP relations.

The inclusion of a Markov-switching intercept in the cointegration space also has economically informative implications. First and foremost, large changes in this intercept between regimes suggest reasons why it may be difficult or impossible to establish the necessary cointegrating relations over the entire time period. Secondly, in the context of PPP and UIP, the regime-dependent intercepts in these cointegrating relations have distinct economic interpretations. In the case of the log-linear representation of the PPP relation given by Equation (5.1), a simple monotonic transformation of the intercept term yields the real exchange rate. Based on the regime-dependent intercepts in the cointegration space, the real exchange rate was shown to have depreciated from regime 1 to regime 2, recovering in regime 3 and then depreciating again in regime 4. These movements in the real exchange rate were found to be consistent with that expected based on the economic and political developments that took place over the studied time period. Clearly, such changes in the mean real exchange rate will prohibit cointegration over the entire time period under analysis. However, if the real exchange rate is stationary within distinct regimes, as was found to be the case in this paper, then relative PPP will still hold conditional on the underlying regimes.

A regime-dependent interest rate differential in the UIP relation will also compromise evidence in favour of this relation over the entire time period. If this intercept is interpreted as a risk premium, then UIP can only hold when the relative riskiness of the two countries under consideration is constant. Since this is not the case when comparing South Africa with the United States from 1972 to 2007, it cannot be anticipated that UIP will hold over this time period. However, introducing a regime-dependent intercept into this relation allows for changes in the risk premium across regimes. In this study, the risk premium was found to be lowest in the first regime, rising in regimes 2 and 3 and then declining in regime 4. Indeed, this result is to be expected given the social unrest associated with regime 2 and the political uncertainty which characterises regime 3 ahead of the elections in 1994. An improvement in South Africa's risk profile is observed in regime 4 in the wake of the country's transition to a democratic society. Note too that since regime 1 includes the period from 2005 to 2007, it would appear that the perceived riskiness of South Africa as an investment hub has improved somewhat in these recent years.

The Markov-switching model was uniformly preferred to its linear counterpart in terms of the likelihood ratio test and all the information criteria considered. Furthermore, the properties of the residuals of the MSIH(4)-VECM(1) were far more satisfactory than those obtained in the linear model. Indeed, the residuals of the latter model exhibited significant non-normality and ARCH effects which appear to be remedied in the Markov-switching model.

This paper therefore demonstrates the importance of incorporating regime shifts appropriately into the modelling procedure adopted in testing for PPP and UIP. The regime-dependent forms of relative PPP and UIP implied by the "conditional cointegration" alluded to above appear to garner more support from the data than the standard definitions of these theories. In addition, the estimated regimes and regime-dependent parameters shed further light on the empirical problem at hand. The effects of changes in monetary and exchange rate policies should therefore not be discounted when modelling the PPP and UIP relations.

Further research into the theoretical properties of the Markov-switching vector error correction model is still necessary. In particular, formal tests of the cointegrating rank and the imposition of overidentifying linear restrictions on the cointegrating vectors have yet to be adapted for the Markov-switching model. Additionally, an option to restrict the intercept term to the cointegration space only in the Markov-switching model could be advantageous. It would also be interesting to consider how well this model performs on other datasets and time frames where regime changes are known to exist. Indeed, the success of cointegration tests in establishing firm evidence in favour of PPP and UIP is known to vary greatly between studies [32]. At the very least, however, this paper has demonstrated that accounting for the monetary and exchange rate regime shifts observed in South Africa between 1972 and 2007 leads to a marked improvement in the empirical evidence in support of PPP and UIP over this period.

References

[1] Alessandria, G. and Kaboski, J. (2005). *Violating Purchasing Power Parity.* Meeting Paper No. 807, Society for Economic Dynamics.

[2] Anderson, T.W. (1951). Estimating Linear Restrictions on Regression Coefficients for Multivariate Normal Distributions. *The Annals of Mathematical Statistics*, 22, 327–351.

[3] Aron, J., Elbadawi, I.A. and Kahn, B. (2000). Real and Monetary Determinants of the Real Exchange Rate in South Africa, in *Development Issues in South Africa*, edited by Elbadawi, I.A and Hartzenberg, T. London: MacMillan.

[4] Aron, J. and Muellbauer, M. (2001). *Estimating Monetary Policy Rules for South Africa.* Working Paper No. 145, The Centre for the Study of African Economies, University of Oxford. http://www.bepress.com/csae/paper145

[5] Barr, G.D.I. and Kahn, B. (1994). *Testing for Purchasing Power Parity in South Africa in the Presence of Real Shocks.* Discussion Paper, Centre for Research into Economics and Finance in Southern Africa, London School of Economics and Political Science.

[6] Cappé, O., Moulines, E. and Rydén, T. (2005). *Inference in Hidden Markov Models.* New York: Springer.

[7] Cox, D.R. and Miller, H.D. (1965). *The Theory of Stochastic Processes.* London: Methuen.

[8] Davies, R.B. (1987). Hypothesis Testing When a Nuisance Parameter is Present Only Under the Alternative. *Biometrika*, 74(1), 33–43.

[9] Dickey, D.A. and Fuller, W.A. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, 74, 427-431.

[10] Debreu, G. and Herstein, I.N. (1953). Non-Negative Square Matrices. *Econometrica*, 21, 597–607.

[11] E-Views 4 User's Guide. (2000). United States of America: Quantitative Micro Software.

[12] Fedderke, J.W. and Pillay, N. (2007). *A Theoretically Defensible Measure of Risk: Using Financial Market Data from a Middle Income Context.* Working Paper No. 64, Economic Research South Africa, School of Economics, University of Cape Town.

[13] Feller, W. (1968). *An Introduction to Probability Theory and Its Applications, 3rd edition.* New York: John Wiley and Sons.

[14] Fisher, R.A. (1951). *The Design of Experiments, 6th edition.* Great Britain: Oliver and Boyd.

[15] Frankel, J. (2007). *On the Rand: Determinants of the South African Exchange Rate.* CID Working Paper No. 139, Centre for International Development, Harvard University.

[16] Garratt, A., Lee, K., Pesaran, M.H. and Shin, Y. (2003). A Long Run Structural Macroeconometric Model of the UK. *The Economic Journal*, 113, 412–455.

[17] Granger, C.W.J. and Lin, J-L. (1995). Causality in the Long-Run. *Econometric Theory*, 11(3), 530–536.

[18] Gredenhoff, M. and Jacobson, T. (2001). Bootstrap Testing Linear Restrictions on Cointegrating Vectors. *Journal of Business and Economic Statistics*, 19(1), 63–72.

[19] Hamilton, J.D. (1994). *Time Series Analysis.* New Jersey: Princeton University Press.

[20] Harville, D.A. (1997). *Matrix Algebra from a Statistician's Perspective.* New York: Springer-Verlag.

[21] Hayashi, F. (2000). *Econometrics.* Princeton: Princeton University Press.

[22] Henderson, H.V. and Searle, S.R. (1979). Vec and Vech Operators for Matrices, with Some Uses in Jacobians and Multivariate Statistics. *The Canadian Journal of Statistics*, 7, 65–81.

[23] Hendry, D.F. (1995). *Dynamic Econometrics.* Oxford: Oxford University Press.

[24] Isard, P. (1977). How Far Can We Push the "Law of One Price"? *The American Economic Review*, 67(5), 942–948.

[25] Johansen, S. (1992a). Cointegration in Partial Systems and the Efficiency of Single-Equation Analysis. *Journal of Econometrics*, 52, 389–402.

[26] Johansen, S. (1992b). Testing Weak Exogeneity and the Order of Cointegration in UK Money Demand Data. *Journal of Policy Modelling*, 14, 313–334.

[27] Johansen, S. (1995a). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models.* New York: Oxford University Press.

[28] Johansen, S. (1995b). Identifying Restrictions of Linear Equations with Applications to Simultaneous Equations and Cointegration. *Journal of Econometrics*, 69, 111–132.

[29] Johansen, S. and Juselius, K. (1990). Maximum Likelihood Estimation and Inference on Cointegration — With Applications to the Demand for Money. *Oxford Bulletin of Economics and Statistics*, 52, 169–210.

[30] Johansen, S. and Juselius, K. (1992). Testing Structural Hypotheses in a Multivariate Cointegration Analysis of the PPP and the UIP for UK. *Journal of Econometrics*, 53, 211–244.

[31] Johansen, S. and Juselius, K. (1994). Identification of the Long-Run and the Short-Run Structure: An Application to the ISLM Model. *Journal of Econometrics*, 63, 7–36.

[32] Jonsson, G. (2001). *Inflation, Money Demand, and Purchasing Power Parity in South Africa.* International Monetary Fund Staff Paper, Vol. 48, No. 2.

[33] Krolzig, H-M. (1996). *Statistical Analysis of Cointegrated VAR Processes with Markovian Regime Shifts.* SFB 373 Discussion Paper 25/1996, Humboldt Universität zu Berlin.

[34] Krolzig, H-M. (1997). *Markov-Switching Vector Autoregressions: Modelling, Statistical Inference and Application to Business Cycle Analysis.* Berlin: Springer-Verlag.

[35] Krolzig, H-M. (1998). *Econometric Modelling of Markov-Switching Vector Autoregressions using MSVAR for Ox.* Discussion Paper, Institute of Economics and Statistics, University of Oxford. http://www.econ.ox.ac.uk/research/hendry/krolzig.

[36] Krolzig, H-M. and Toro, J. (2001). *A New Approach to the Analysis of Business Cycle Transitions in a Model of Output and Employment.* Discussion Paper 59, Department of Economics, University of Oxford.

[37] Krugman, P.R. and Obstfeld, M. (2006). *International Economics: Theory and Policy, 7th edition.* United States: Pearson.

[38] Lütkepohl, H. (1993). *Introduction to Multiple Time Series Analysis, 2nd edition.* Berlin: Springer-Verlag.

[39] Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis.* Berlin: Springer-Verlag.

[40] MacDonald, R. and Ricci, L.A. (2004). Estimation of the Equilibrium Real Exchange Rate for South Africa. *The South African Journal of Economics*, 72(2), 282–304.

[41] MacKinnon, J.G. (1991). Critical Values for Cointegration Tests, in *Long-Run Economic Relationships: Readings in Cointegration*, edited by Engle, R.F. and Granger, C.W.J. New York: Oxford University Press.

[42] O'Connell, P.G.J. (1998). Market Frictions and Real Exchange Rates. *Journal of International Money and Finance*, 17, 71–95.

[43] Patterson, K. (2000). *An Introduction to Applied Econometrics: A Time Series Approach.* New York: Palgrave.

[44] Pesaran, M.H. and Pesaran, B. (1997). *Working with Microfit 4.0: Interactive Econometric Analysis.* Great Britain: Oxford University Press.

[45] Pesaran, M.H., Shin, Y. and Smith, R.J. (2000). Structural Analysis of Vector Error Correction Models with Exogenous $I(1)$ Variables. *Journal of Econometrics*, 97, 293–343.

[46] Reimers, H.E. (1992). Comparisons of Tests for Multivariate Cointegration. *Statistical Papers*, 33, 335–359.

[47] Rogoff, K. (1996). The Purchasing Power Parity Puzzle. *Journal of Economic Literature*, 34, 647–668.

[48] Sercu, P., Uppal, R. and Van Hulle, C. (1995). The Exchange Rate in the Presence of Transaction Costs: Implications for Tests of Purchasing Power Parity. *The Journal of Finance*, 50(4), 1309–1319.

[49] Shumway, R.H. and Stoffer, D.S. (2000). *Time Series Analysis and Its Applications.* New York: Springer-Verlag.

[50] Strauss, J. (1995). Real Exchange Rates, PPP and the Relative Price of Nontraded Goods. *Southern Economic Journal*, 61(4), 991–1005.

[51] Taylor, A.M. and Taylor, M.P. (2004). The Purchasing Power Parity Debate. *The Journal of Economic Perspectives*, 18(4), 135–158.

A Parameter Estimation in the VECM

The vector error correction model is given as

$$\Delta \boldsymbol{y}_t = \boldsymbol{\alpha}\boldsymbol{\beta}' \boldsymbol{y}_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i \Delta \boldsymbol{y}_{t-i} + \boldsymbol{\Phi} \boldsymbol{D}_t + \boldsymbol{\varepsilon}_t \qquad \text{for } t = 1, 2, \ldots, T \tag{A.1}$$

where $\boldsymbol{\varepsilon}_t$ are identically and independently distributed $\mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$ random vectors and $\{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Gamma}_1, \ldots, \boldsymbol{\Gamma}_{p-1}, \boldsymbol{\Phi}, \boldsymbol{\Omega}\}$ are freely varying parameters as defined in the main text [27].

For the discussion to follow, it is useful to introduce the notation $\boldsymbol{Z}_{0t} = \Delta \boldsymbol{y}_t$, $\boldsymbol{Z}_{1t} = \boldsymbol{y}_{t-1}$ and let $\boldsymbol{Z}_{2t} = \left(\Delta \boldsymbol{y}_{t-1}, \ldots, \Delta \boldsymbol{y}_{t-p+1}, \boldsymbol{D}_t\right)'$ with associated coefficient matrix $\boldsymbol{\Psi} = (\boldsymbol{\Gamma}_1, \ldots, \boldsymbol{\Gamma}_{p-1}, \boldsymbol{\Phi})$. $\boldsymbol{Z}_{0t}$ and $\boldsymbol{Z}_{1t}$ are therefore k dimensional vectors, $\boldsymbol{Z}_{2t}$ is a $k(p-1) + d$ dimensional vector and accordingly $\boldsymbol{\Psi}$ is a $k \times [k\,(p-1) + d]$ dimensional matrix. In this notation, the VECM in Equation (A.1) becomes

$$\boldsymbol{Z}_{0t} = \boldsymbol{\alpha}\boldsymbol{\beta}' \boldsymbol{Z}_{1t} + \boldsymbol{\Psi} \boldsymbol{Z}_{2t} + \boldsymbol{\varepsilon}_t \qquad \text{for } t = 1, 2, \ldots, T. \tag{A.2}$$

Equation (A.2) represents a non-linear regression model since it involves the product of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}'$. The parameters $\boldsymbol{\Psi}$ may vary freely and no prior imposition or assumption is made about the rank of $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$ or equivalently the number of cointegrating relations. In this sense, the model is said to be *unrestricted* [43].

The likelihood function is given as

$$L\left(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Psi}, \boldsymbol{\Omega}\right) = \prod_{t=1}^{T} (2\pi)^{-k/2} \left|\boldsymbol{\Omega}\right|^{-1/2} \times \\ \exp\left\{-\frac{1}{2}\left(\boldsymbol{Z}_{0t} - \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{Z}_{1t} - \boldsymbol{\Psi}\boldsymbol{Z}_{2t}\right)' \boldsymbol{\Omega}^{-1} \left(\boldsymbol{Z}_{0t} - \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{Z}_{1t} - \boldsymbol{\Psi}\boldsymbol{Z}_{2t}\right)\right\}$$

which leads to the log likelihood

$$\ln L\left(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Psi}, \boldsymbol{\Omega}\right) = \frac{-kT}{2} \ln 2\pi - \frac{T}{2} \ln \left|\boldsymbol{\Omega}\right| \\ - \frac{1}{2} \sum_{t=1}^{T} \left(\boldsymbol{Z}_{0t} - \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{Z}_{1t} - \boldsymbol{\Psi}\boldsymbol{Z}_{2t}\right)' \boldsymbol{\Omega}^{-1} \left(\boldsymbol{Z}_{0t} - \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{Z}_{1t} - \boldsymbol{\Psi}\boldsymbol{Z}_{2t}\right).$$

The first order conditions for estimating $\boldsymbol{\Psi}$ may be found by differentiating this log likelihood with respect to $\boldsymbol{\Psi}$ and setting this derivative equal to zero to produce

$$\sum_{t=1}^{T} \left(\boldsymbol{Z}_{0t} - \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{Z}_{1t} - \hat{\boldsymbol{\Psi}}\boldsymbol{Z}_{2t} \right) \boldsymbol{Z}_{2t}' = 0. \tag{A.3}$$

where $\hat{\boldsymbol{\Psi}}$ is the maximum likelihood estimate of $\boldsymbol{\Psi}$. Defining the product moment matrices as

$$\boldsymbol{M}_{ij} = T^{-1} \sum_{t=1}^{T} \boldsymbol{Z}_{it}\boldsymbol{Z}_{jt}' \qquad \text{for } i, j = 0, 1, 2,$$

Equation (A.3) may be rewritten as

$$\boldsymbol{M}_{02} - \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{M}_{12} - \hat{\boldsymbol{\Psi}}\boldsymbol{M}_{22} = 0$$

such that

$$\hat{\boldsymbol{\Psi}}\left(\boldsymbol{\alpha}, \boldsymbol{\beta}\right) = \boldsymbol{M}_{02}\boldsymbol{M}_{22}^{-1} - \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{M}_{12}\boldsymbol{M}_{22}^{-1}. \tag{A.4}$$

Now consider the regression of $\Delta\boldsymbol{y}_t$ on the lagged differences $\Delta\boldsymbol{y}_{t-1}, \ldots,$ $\Delta\boldsymbol{y}_{t-p+1}$ and $\boldsymbol{D}_t$ or equivalently $\boldsymbol{Z}_{0t}$ on $\boldsymbol{Z}_{2t}$ given as

$$\boldsymbol{Z}_{0t} = \boldsymbol{\Theta}_{02}\boldsymbol{Z}_{2t} + \boldsymbol{\eta}_t$$

where $\boldsymbol{\eta}_t \overset{iid}{\sim} \mathcal{N}(\boldsymbol{0}, \boldsymbol{V}_{02})$ with variance-covariance matrix $\boldsymbol{V}_{02}$ [43]. The maximum likelihood estimates for the regression parameters $\boldsymbol{\Theta}_{02}$ may be derived from the log likelihood

$$L\left(\boldsymbol{\Theta}_{02}, \boldsymbol{V}_{02}\right) = \prod_{t=1}^{T} (2\pi)^{-k/2} \left|\boldsymbol{V}_{02}\right|^{-1/2} \times \\ \exp\left\{ -\frac{1}{2} \left(\boldsymbol{Z}_{0t} - \boldsymbol{\Theta}_{02}\boldsymbol{Z}_{2t}\right)' \boldsymbol{V}_{02}^{-1} \left(\boldsymbol{Z}_{0t} - \boldsymbol{\Theta}_{02}\boldsymbol{Z}_{2t}\right) \right\}$$

$$\ln L\left(\boldsymbol{\Theta}_{02}, \boldsymbol{V}_{02}\right) = \frac{-kT}{2} \ln 2\pi - \frac{T}{2} \ln\left|\boldsymbol{V}_{02}\right| \\ - \frac{1}{2} \sum_{t=1}^{T} \left(\boldsymbol{Z}_{0t} - \boldsymbol{\Theta}_{02}\boldsymbol{Z}_{2t}\right)' \boldsymbol{V}_{02}^{-1} \left(\boldsymbol{Z}_{0t} - \boldsymbol{\Theta}_{02}\boldsymbol{Z}_{2t}\right)$$

Taking the derivative with respect to $\mathbf{\Theta}_{02}$ and equating this derivative to zero produces the normal equations

$$\sum_{t=1}^{T} \left(\boldsymbol{Z}_{0t} - \hat{\mathbf{\Theta}}_{02} \boldsymbol{Z}_{2t} \right) \boldsymbol{Z}'_{2t} = 0$$

and hence, using the previous notation, the maximum likelihood estimate

$$\hat{\mathbf{\Theta}}_{02} = \boldsymbol{M}_{02} \boldsymbol{M}_{22}^{-1}.$$

Finally, define the residuals obtained in this regression as

$$\begin{aligned} \boldsymbol{R}_{0t} &= \boldsymbol{Z}_{0t} - \hat{\mathbf{\Theta}}_{02} \boldsymbol{Z}_{2t} \\ &= \boldsymbol{Z}_{0t} - \boldsymbol{M}_{02} \boldsymbol{M}_{22}^{-1} \boldsymbol{Z}_{2t}. \end{aligned}$$

Next, consider the regression of $\boldsymbol{y}_{t-1}$ on the lagged differences $\Delta \boldsymbol{y}_{t-1}, \ldots,$ $\Delta \boldsymbol{y}_{t-p+1}$ and $\boldsymbol{D}_t$ or simply $\boldsymbol{Z}_{1t}$ on $\boldsymbol{Z}_{2t}$

$$\boldsymbol{Z}_{1t} = \mathbf{\Theta}_{12} \boldsymbol{Z}_{2t} + \boldsymbol{\varsigma}_t$$

with $\boldsymbol{\varsigma}_t \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{V}_{12})$ [43]. In a similar vein, it may be shown that the maximum likelihood estimate for the regression parameters $\mathbf{\Theta}_{12}$ is given by

$$\hat{\mathbf{\Theta}}_{12} = \boldsymbol{M}_{12} \boldsymbol{M}_{22}^{-1}$$

so that the residuals in this regression are

$$\boldsymbol{R}_{1t} = \boldsymbol{Z}_{1t} - \boldsymbol{M}_{12} \boldsymbol{M}_{22}^{-1} \boldsymbol{Z}_{2t}.$$

Rearranging terms in the VECM given as Equation (A.2), it is noted that

$$\boldsymbol{\varepsilon}_t = \boldsymbol{Z}_{0t} - \boldsymbol{\alpha} \boldsymbol{\beta}' \boldsymbol{Z}_{1t} - \mathbf{\Psi} \boldsymbol{Z}_{2t} \qquad \text{for } t = 1, 2, \ldots, T.$$

Replacing $\mathbf{\Psi}$ with its maximum likelihood estimate in the above expression yields the residual estimate

$$\begin{aligned} \hat{\boldsymbol{\varepsilon}}_t &= \boldsymbol{Z}_{0t} - \boldsymbol{\alpha} \boldsymbol{\beta}' \boldsymbol{Z}_{1t} - \left(\boldsymbol{M}_{02} \boldsymbol{M}_{22}^{-1} - \boldsymbol{\alpha} \boldsymbol{\beta}' \boldsymbol{M}_{12} \boldsymbol{M}_{22}^{-1} \right) \boldsymbol{Z}_{2t} \\ &= \boldsymbol{Z}_{0t} - \boldsymbol{M}_{02} \boldsymbol{M}_{22}^{-1} \boldsymbol{Z}_{2t} - \boldsymbol{\alpha} \boldsymbol{\beta}' \left(\boldsymbol{Z}_{1t} - \boldsymbol{M}_{12} \boldsymbol{M}_{22}^{-1} \boldsymbol{Z}_{2t} \right) \\ &= \boldsymbol{R}_{0t} - \boldsymbol{\alpha} \boldsymbol{\beta}' \boldsymbol{R}_{1t}. \end{aligned}$$

Thus, the parameters $\boldsymbol{\Psi}$ are eliminated in the regression of the residuals

$$\boldsymbol{R}_{0t} = \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{R}_{1t} + \hat{\boldsymbol{\varepsilon}}_t. \tag{A.5}$$

since the log likelihood function

$$\begin{aligned}\ln L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Omega}) = -\frac{kT}{2}\ln 2\pi - \frac{T}{2}\ln|\boldsymbol{\Omega}| \\ - \sum_{t=1} T\left(\boldsymbol{R}_{0t} - \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{R}_{1t}\right)' \boldsymbol{\Omega}^{-1}\left(\boldsymbol{R}_{0t} - \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{R}_{1t}\right)\end{aligned}$$

does not feature $\boldsymbol{\Psi}$ [27]. This method is known as *reduced rank regression*, as detailed in Anderson [2].

Given $\boldsymbol{\beta}$, it is straightforward to estimate $\boldsymbol{\alpha}$ and $\boldsymbol{\Omega}$ from this likelihood. First, define

$$\boldsymbol{S}_{ij} = T^{-1}\sum_{t=1}^{T}\boldsymbol{R}_{it}\boldsymbol{R}'_{jt} \qquad \text{for } i, j = 0, 1.$$

Then, it can be easily shown that

$$\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}) = \boldsymbol{S}_{01}\boldsymbol{\beta}\left(\boldsymbol{\beta}'\boldsymbol{S}_{11}\boldsymbol{\beta}\right)^{-1} \tag{A.6}$$

$$\begin{aligned}\hat{\boldsymbol{\Omega}}(\boldsymbol{\beta}) &= \boldsymbol{S}_{00} - \boldsymbol{S}_{01}\boldsymbol{\beta}\left(\boldsymbol{\beta}'\boldsymbol{S}_{11}\boldsymbol{\beta}\right)^{-1}\boldsymbol{\beta}'\boldsymbol{S}_{10} \\ &= \boldsymbol{S}_{00} - \hat{\boldsymbol{\alpha}}(\boldsymbol{\beta})\left(\boldsymbol{\beta}'\boldsymbol{S}_{11}\boldsymbol{\beta}\right)^{-1}\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta})'.\end{aligned} \tag{A.7}$$

Substituting the maximum likelihood estimators for $\boldsymbol{\alpha}$ and $\boldsymbol{\Omega}$ into the likelihood function, one can express the likelihood in terms of $\boldsymbol{\beta}$ only

$$\begin{aligned}L_{\max}(\boldsymbol{\beta}) &= \prod_{t=1}^{T}(2\pi)^{-k/2}|\hat{\boldsymbol{\Omega}}|^{-1/2}\exp -\frac{1}{2}\mathrm{tr}\big[\hat{\boldsymbol{\Omega}}^{-1}\left(\boldsymbol{R}_{0t} - \hat{\boldsymbol{\alpha}}\boldsymbol{\beta}'\boldsymbol{R}_{1t}\right)\times \\ &\quad \left(\boldsymbol{R}_{0t} - \hat{\boldsymbol{\alpha}}\boldsymbol{\beta}'\boldsymbol{R}_{1t}\right)'\big] \\ &= (2\pi)^{-kT/2}|\hat{\boldsymbol{\Omega}}|^{-T/2}\exp -\frac{1}{2}\sum_{t=1}^{T}\mathrm{tr}\big[\hat{\boldsymbol{\Omega}}^{-1}\big(\boldsymbol{R}_{0t}\boldsymbol{R}'_{0t} - \hat{\boldsymbol{\alpha}}\boldsymbol{\beta}'\boldsymbol{R}_{1t}\boldsymbol{R}'_{0t} \\ &\quad - \boldsymbol{R}_{0t}\boldsymbol{R}'_{1t}\boldsymbol{\beta}\hat{\boldsymbol{\alpha}}' + \hat{\boldsymbol{\alpha}}\boldsymbol{\beta}'\boldsymbol{R}_{1t}\boldsymbol{R}'_{1t}\boldsymbol{\beta}\hat{\boldsymbol{\alpha}}'\big)\big]\end{aligned}$$

$$
\begin{aligned}
&= (2\pi)^{-kT/2}\,|\hat{\boldsymbol{\Omega}}|^{-T/2} \exp -\frac{T}{2}\mathrm{tr}\big[\hat{\boldsymbol{\Omega}}^{-1}\big(\boldsymbol{S}_{00} - \boldsymbol{S}_{01}\boldsymbol{\beta}\,(\boldsymbol{\beta}'\boldsymbol{S}_{11}\boldsymbol{\beta})^{-1}\,\boldsymbol{\beta}'\boldsymbol{S}_{10} \\
&\quad - \boldsymbol{S}_{01}\boldsymbol{\beta}\,(\boldsymbol{\beta}'\boldsymbol{S}_{11}\boldsymbol{\beta})^{-1}\,\boldsymbol{\beta}'\boldsymbol{S}_{01}' + \boldsymbol{S}_{01}\boldsymbol{\beta}\,(\boldsymbol{\beta}'\boldsymbol{S}_{11}\boldsymbol{\beta})^{-1}\,\boldsymbol{\beta}'\boldsymbol{S}_{01}'\big)\big] \\
&= (2\pi)^{-kT/2}\,|\hat{\boldsymbol{\Omega}}|^{-T/2} \exp -\frac{T}{2}\mathrm{tr}\big[\hat{\boldsymbol{\Omega}}^{-1}\big(\boldsymbol{S}_{00} - \boldsymbol{S}_{01}\boldsymbol{\beta}\,(\boldsymbol{\beta}'\boldsymbol{S}_{11}\boldsymbol{\beta})^{-1}\,\boldsymbol{\beta}'\boldsymbol{S}_{10}\big)\big] \\
&= (2\pi)^{-kT/2}\,|\hat{\boldsymbol{\Omega}}|^{-T/2} \exp -\frac{T}{2}\mathrm{tr}\big[\hat{\boldsymbol{\Omega}}^{-1}\hat{\boldsymbol{\Omega}}\big] \\
&= (2\pi e)^{-kT/2}\,|\hat{\boldsymbol{\Omega}}|^{-T/2}.
\end{aligned}
$$

Hence,

$$
L_{\max}^{-2/T}(\boldsymbol{\beta}) \propto |\hat{\boldsymbol{\Omega}}(\boldsymbol{\beta})\,| = |\boldsymbol{S}_{00} - \boldsymbol{S}_{01}\boldsymbol{\beta}\,(\boldsymbol{\beta}'\boldsymbol{S}_{11}\boldsymbol{\beta})^{-1}\,\boldsymbol{\beta}'\boldsymbol{S}_{10}|.
$$

Now consider the following determinant

$$
\begin{vmatrix} \boldsymbol{S}_{00} & \boldsymbol{S}_{01}\boldsymbol{\beta} \\ \boldsymbol{\beta}'\boldsymbol{S}_{10} & \boldsymbol{\beta}'\boldsymbol{S}_{11}\boldsymbol{\beta} \end{vmatrix} = |\boldsymbol{S}_{00}||\boldsymbol{\beta}'\left(\boldsymbol{S}_{11} - \boldsymbol{S}_{10}\boldsymbol{S}_{00}^{-1}\boldsymbol{S}_{01}\right)\boldsymbol{\beta}| \tag{A.8}
$$

$$
= |\boldsymbol{\beta}'\boldsymbol{S}_{11}\boldsymbol{\beta}||\boldsymbol{S}_{00} - \boldsymbol{S}_{01}\boldsymbol{\beta}\,(\boldsymbol{\beta}'\boldsymbol{S}_{11}\boldsymbol{\beta})^{-1}\,\boldsymbol{\beta}'\boldsymbol{S}_{10}| \tag{A.9}
$$

applying the identity

$$
\begin{vmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{vmatrix} = |\boldsymbol{\Sigma}_{11}||\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}| = |\boldsymbol{\Sigma}_{22}||\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}|.
$$

From Equations (A.8) and (A.9), it therefore follows that

$$
\begin{aligned}
L_{\max}^{-2/T}(\boldsymbol{\beta}) &\propto |\boldsymbol{S}_{00} - \boldsymbol{S}_{01}\boldsymbol{\beta}\,(\boldsymbol{\beta}'\boldsymbol{S}_{11}\boldsymbol{\beta})^{-1}\,\boldsymbol{\beta}'\boldsymbol{S}_{10}| \\
&\propto |\boldsymbol{S}_{00}| \times \frac{|\boldsymbol{\beta}'\left(\boldsymbol{S}_{11} - \boldsymbol{S}_{10}\boldsymbol{S}_{00}^{-1}\boldsymbol{S}_{01}\right)\boldsymbol{\beta}|}{|\boldsymbol{\beta}'\boldsymbol{S}_{11}\boldsymbol{\beta}|}.
\end{aligned} \tag{A.10}
$$

The likelihood function is maximised when the last factor in Equation (A.10) is minimised with respect to $\boldsymbol{\beta}$. This maximisation of the likelihood function may be achieved by noting the following result in linear algebra [27].

Result 1 *Let $\boldsymbol{M}$ be symmetric and positive semi-definite and $\boldsymbol{N}$ symmetric and positive definite. Then the function*

$$
f(\boldsymbol{X}) = \frac{|\boldsymbol{X}'\boldsymbol{M}\boldsymbol{X}|}{|\boldsymbol{X}'\boldsymbol{N}\boldsymbol{X}|}
$$

is maximised amongst all $k \times r$ matrices by $\hat{\boldsymbol{X}} = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r)$ with maximum $\prod_{i=1}^{r} \lambda_i$, where λ_i and $\boldsymbol{v}_i$ are solutions to the eigenvalue problem

$$|\lambda \boldsymbol{N} - \boldsymbol{M}| = 0,$$

which has k eigenvalues ordered such that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k > 0$ with corresponding eigenvectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$.

The following results hold:

$$\begin{aligned} \boldsymbol{N}\boldsymbol{V}\boldsymbol{\Lambda} &= \boldsymbol{M}\boldsymbol{V} \\ \boldsymbol{V}'\boldsymbol{N}\boldsymbol{V} &= \boldsymbol{I} \\ \boldsymbol{V}'\boldsymbol{M}\boldsymbol{V} &= \boldsymbol{\Lambda} \end{aligned}$$

where $\boldsymbol{V} = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$ and $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_k)$.

Note that $\hat{\boldsymbol{X}}\boldsymbol{A}$ is also a maximising argument for $f(\boldsymbol{X})$ for any non-singular $r \times r$ matrix $\boldsymbol{A}$.

Minimising $|\boldsymbol{\beta}' \left(\boldsymbol{S}_{11} - \boldsymbol{S}_{10}\boldsymbol{S}_{00}^{-1}\boldsymbol{S}_{01}\right) \boldsymbol{\beta}| / |\boldsymbol{\beta}'\boldsymbol{S}_{11}\boldsymbol{\beta}|$ is equivalent to maximising its reciprocal, that is

$$f(\boldsymbol{\beta}) = \frac{|\boldsymbol{\beta}'\boldsymbol{S}_{11}\boldsymbol{\beta}|}{|\boldsymbol{\beta}' \left(\boldsymbol{S}_{11} - \boldsymbol{S}_{10}\boldsymbol{S}_{00}^{-1}\boldsymbol{S}_{01}\right) \boldsymbol{\beta}|}.$$

Maximising this function may be achieved by means of the aforementioned algebraic result and solving the eigenvalue problem

$$|\psi(\boldsymbol{S}_{11} - \boldsymbol{S}_{10}\boldsymbol{S}_{00}^{-1}\boldsymbol{S}_{01}) - \boldsymbol{S}_{11}| = 0 \tag{A.11}$$

for the eigenpairs $(\psi_i, \boldsymbol{v}_i)$. From the mathematical result earlier, $f(\boldsymbol{\beta})$ will clearly attain a maximum value of $\prod_{i=1}^{r} \psi_i$ when $\boldsymbol{\beta} = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r)$.

Multiplying Equation (A.11) through by $-\psi^{-1}$ yields

$$\left|\psi^{-1}\boldsymbol{S}_{11} - \left(\boldsymbol{S}_{11} - \boldsymbol{S}_{10}\boldsymbol{S}_{00}^{-1}\boldsymbol{S}_{01}\right)\right| = 0$$

and, putting $\lambda = \psi^{-1}(\psi - 1)$, the eigenvalue problem may be restated as

$$\left|\lambda \boldsymbol{S}_{11} - \boldsymbol{S}_{10}\boldsymbol{S}_{00}^{-1}\boldsymbol{S}_{01}\right| = 0$$

with eigenpairs $(\lambda_i, \boldsymbol{v}_i)$. Therefore, expressed in terms of λ, $f(\boldsymbol{\beta}) = |\boldsymbol{\beta}'\boldsymbol{S}_{11}\boldsymbol{\beta}|/|\boldsymbol{\beta}'\left(\boldsymbol{S}_{11} - \boldsymbol{S}_{10}\boldsymbol{S}_{00}^{-1}\boldsymbol{S}_{01}\right)\boldsymbol{\beta}|$ will attain a maximum value of $\prod_{i=1}^{r}(1-\lambda_i)^{-1}$ when $\boldsymbol{\beta} = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r)$. Hence it follows that

$$\underset{\hat{\boldsymbol{\beta}} \,\in\, \boldsymbol{\beta}}{\arg\min} \frac{|\hat{\boldsymbol{\beta}}'\left(\boldsymbol{S}_{11} - \boldsymbol{S}_{10}\boldsymbol{S}_{00}^{-1}\boldsymbol{S}_{01}\right)\hat{\boldsymbol{\beta}}|}{|\hat{\boldsymbol{\beta}}'\boldsymbol{S}_{11}\hat{\boldsymbol{\beta}}|} = (\hat{\boldsymbol{v}}_1, \ldots, \hat{\boldsymbol{v}}_r)$$

and

$$\begin{aligned}\min L_{\max}^{-2/T} &\propto |\boldsymbol{S}_{00}| \frac{|\hat{\boldsymbol{\beta}}'\left(\boldsymbol{S}_{11} - \boldsymbol{S}_{10}\boldsymbol{S}_{00}^{-1}\boldsymbol{S}_{01}\right)\hat{\boldsymbol{\beta}}|}{|\hat{\boldsymbol{\beta}}'\boldsymbol{S}_{11}\hat{\boldsymbol{\beta}}|} \\ &\propto |\boldsymbol{S}_{00}| \prod_{i=1}^{r}(1-\hat{\lambda}_i).\end{aligned}$$

The space spanned by the eigenvectors corresponding to the r largest eigenvalues, denoted $sp(\{\hat{\boldsymbol{v}}_1, \ldots, \hat{\boldsymbol{v}}_r\})$, is referred to as the cointegration space $sp(\hat{\boldsymbol{\beta}})$. For $0 < r < k$, this space will clearly include any linear combination or scalar multiple of these r eigenvectors and will therefore require further restrictions in order to uniquely identify the cointegrating relations. In the case where $r = 0$, $sp(\hat{\boldsymbol{\beta}}) = \emptyset$ and consequently $\boldsymbol{\Pi} = \boldsymbol{0}$ as before. At the other extreme, when $r = k$, the cointegration space is the set of all k dimensional vectors $\mathcal{R}^k$ so that any linear combination of the variables in $\boldsymbol{y}_t$ is stationary; that is, $\boldsymbol{y}_t \sim I(0)$. Thus all possible cases from $r = 0$ through to $r = k$ are dealt with by solving a single eigenvalue problem [27].

Now suppose that the cointegrating relations are uniquely identified as

$$\boldsymbol{\beta} = (\boldsymbol{H}_1\boldsymbol{\varphi}_1, \ldots, \boldsymbol{H}_r\boldsymbol{\varphi}_r), \tag{A.12}$$

where $\boldsymbol{H}_i$ is a $k \times (k - g_i)$ matrix indicating which elements in the ith cointegrating vector $\boldsymbol{\beta}_i$ vary without restriction and $\boldsymbol{\varphi}_i$ is a $k - g_i$ dimensional vector of freely varying parameters associated with $\boldsymbol{\beta}_i$. Note that this implies that g_i restrictions have been imposed on the ith cointegrating vector, where $g_i \geq r - 1$ for generic identification.

The free parameters $\boldsymbol{\varphi}_1, \ldots, \boldsymbol{\varphi}_r$ associated with each of the r cointegrating vectors in the VECM may be estimated by a switching algorithm comprising an iterative sequence of reduced rank regressions

$$\boldsymbol{R}_{0t} = \boldsymbol{\alpha}_1\boldsymbol{\varphi}_1'\boldsymbol{H}_1'\boldsymbol{R}_{1t} + \ldots + \boldsymbol{\alpha}_r\boldsymbol{\varphi}_r'\boldsymbol{H}_r'\boldsymbol{R}_{1t} + \hat{\varepsilon}_t,$$

obtained by substituting Equation (A.12) into Equation (A.5) and decomposing $\boldsymbol{\alpha}$ into its k dimensional column vectors $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r$. For fixed values of $\boldsymbol{\varphi}_2, \ldots, \boldsymbol{\varphi}_r$ or equivalently $\boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_r$, the above equation is simply a reduced rank regression of $\boldsymbol{R}_{0t}$ on $\boldsymbol{\varphi}_1'\boldsymbol{H}_1'\boldsymbol{R}_{1t}$ corrected for $\boldsymbol{\alpha}_2\boldsymbol{\varphi}_2'\boldsymbol{H}_2'\boldsymbol{R}_{1t}, \ldots, \boldsymbol{\alpha}_r\boldsymbol{\varphi}_r'\boldsymbol{H}_r'\boldsymbol{R}_{1t}$. Performing this regression yields an estimate of $\boldsymbol{\varphi}_1$ and $\boldsymbol{\beta}_1 = \boldsymbol{H}_1\boldsymbol{\varphi}_1$ This procedure is then repeated, fixing $\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_3, \ldots, \boldsymbol{\varphi}_r$ or equivalently $\boldsymbol{\beta}_1, \boldsymbol{\beta}_3, \ldots, \boldsymbol{\beta}_r$ and estimating $\boldsymbol{\varphi}_2$ and hence also $\boldsymbol{\beta}_2$ through a reduced rank regression of $\boldsymbol{R}_{0t}$ on $\boldsymbol{\varphi}_2'\boldsymbol{H}_2'\boldsymbol{R}_{1t}$ corrected for $\boldsymbol{\alpha}_1\boldsymbol{\varphi}_1'\boldsymbol{H}_1'\boldsymbol{R}_{1t}, \boldsymbol{\alpha}_3\boldsymbol{\varphi}_3'\boldsymbol{H}_3'\boldsymbol{R}_{1t}, \ldots, \boldsymbol{\alpha}_r\boldsymbol{\varphi}_r'\boldsymbol{H}_r'\boldsymbol{R}_{1t}$. Running this algorithm until convergence will produce the maximum likelihood estimates of the cointegrating relations under the specified restrictions [27].

Whilst the estimates of the cointegrating vectors $\hat{\boldsymbol{\beta}}$ obtained in the unrestricted model may be used as starting values, this is generally not recommended since the ordering of the unrestricted eigenvectors may not correspond to the ordering given by $\boldsymbol{H}_1, \ldots, \boldsymbol{H}_r$. Instead, it is preferable to find an initial value for $\hat{\boldsymbol{\beta}}_i$ which is closest to $sp(\boldsymbol{H}_i)$ by taking linear combinations of the unrestricted estimates. This initial value is found by solving the eigenvalue problem

$$|\lambda\hat{\boldsymbol{\beta}}'\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}'\boldsymbol{H}_i\left(\boldsymbol{H}_i'\boldsymbol{H}_i\right)^{-1}\boldsymbol{H}_i'\hat{\boldsymbol{\beta}}| = 0$$

for the r eigenvalues $\lambda_1, \ldots, \lambda_r$ and eigenvectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r$ and choosing $\hat{\boldsymbol{\beta}}\boldsymbol{v}_1$ as the starting value for $\boldsymbol{\beta}_i = \boldsymbol{\varphi}_i\boldsymbol{H}_i$ [31].

After estimating $\boldsymbol{\beta}$ in this manner, it is possible to return to Equation (A.6) and estimate $\boldsymbol{\alpha}$. Finally, with $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$ in hand, one can estimate $\boldsymbol{\Omega}$ and $\boldsymbol{\Psi}$ by Equations (A.7) and (A.4) respectively [43].

B Properties of Markov Chains

Suppose $s_t \in \{1, \ldots, m\}$ is a finite first-order Markov process with transition probability matrix $\mathbb{P}$. Let

$$\vartheta\left(s_t = i\right) = \begin{cases} 1 \text{ if } s_t = i \\ 0 \text{ otherwise} \end{cases}$$

be an indicator variable for $i = 1, \ldots, m$ and define

$$\boldsymbol{\xi}_t = \begin{pmatrix} \vartheta(s_t = 1) \\ \vdots \\ \vartheta(s_t = m) \end{pmatrix}.$$

Hence $\boldsymbol{\xi}_t$ represents the unobserved state of the system. Note that all relevant information about the future of a Markovian process is included in the present state such that

$$\Pr[\boldsymbol{\xi}_{t+n} | \boldsymbol{\xi}_t, \boldsymbol{\xi}_{t-1}, \ldots; \boldsymbol{y}_t, \boldsymbol{y}_{t-1}, \ldots] = \Pr[\boldsymbol{\xi}_{t+n} | \boldsymbol{\xi}_t], \tag{B.1}$$

where the past and additional variables such as $\boldsymbol{y}_t$ include no relevant information beyond the current state. The assumption of a first-order Markov process is not especially restrictive, since a higher-order Markov chain can be reparameterised as a higher dimensional first-order Markov process [34]. Furthermore, it is assumed that the Markov chain is *homogenous* or possesses a stationary transition mechanism so that the probability in Equation (B.1) depends only on the time interval n, but not on the time t. It therefore follows that the n-step transition probability given above may be re-expressed as

$$\Pr[\boldsymbol{\xi}_{t+n} | \boldsymbol{\xi}_t] = \Pr[\boldsymbol{\xi}_n | \boldsymbol{\xi}_0]$$

in the case of a homogenous Markov chain [7].

Since $\boldsymbol{\xi}_t$ consists of binary variables, it follows that

$$\mathrm{E}[\boldsymbol{\xi}_{t+1} | \boldsymbol{\xi}_t = \boldsymbol{\iota}_i] = \begin{pmatrix} \Pr[\boldsymbol{\xi}_{t+1} = \boldsymbol{\iota}_1 | \boldsymbol{\xi}_t = \boldsymbol{\iota}_i] \\ \vdots \\ \Pr[\boldsymbol{\xi}_{t+1} = \boldsymbol{\iota}_m | \boldsymbol{\xi}_t = \boldsymbol{\iota}_i] \end{pmatrix} = \begin{pmatrix} \Pr[s_{t+1} = 1 | s_t = i] \\ \vdots \\ \Pr[s_{t+1} = m | s_t = i] \end{pmatrix},$$

where $\boldsymbol{\iota}_i$ denotes the ith column of the identity matrix. Note that this conditional expectation is simply the ith row of the transition probability matrix $\mathbb{P}$ which, together with the Markovian property of Equation (B.1), implies

$$\mathrm{E}[\boldsymbol{\xi}_{t+1} | \boldsymbol{\xi}_t, \boldsymbol{\xi}_{t-1}, \ldots] = \mathbb{P}' \boldsymbol{\xi}_t.$$

From this last equation, it is possible to express the regime generating process as a first-order vector autoregression of the form

$$\boldsymbol{\xi}_{t+1} = \mathbb{P}' \boldsymbol{\xi}_t + \boldsymbol{\eta}_{t+1}. \tag{B.2}$$

with error term $\boldsymbol{\eta}_{t+1}$ at time $t+1$. By recursive substitution, it can be shown that

$$\boldsymbol{\xi}_{t+n} = \mathbb{P}'^n \boldsymbol{\xi}_t + \mathbb{P}'^{n-1} \boldsymbol{\eta}_{t+1} + \ldots + \mathbb{P}' \boldsymbol{\eta}_{t+n-1} + \boldsymbol{\eta}_{t+n}$$

and, taking expectations, the n-period ahead forecast for the Markov chain may be established as

$$\mathrm{E}[\boldsymbol{\xi}_{t+n} | \boldsymbol{\xi}_t, \boldsymbol{\xi}_{t-1}, \ldots] = \mathbb{P}'^n \boldsymbol{\xi}_t. \tag{B.3}$$

Hence, the probabilities associated with the realisation of each regime in n periods time, given that regime i is currently observed, are collected in the ith column of $\mathbb{P}'^n$ [19].

B.1 Recurrence and Transience

The states of a Markov chain may be classified into distinct types according to their limiting behaviour. Suppose a chain is initially in state i and let $f_{ii}^{(n)}$ denote the probability that the next occurrence of state i is at time n; that is, $f_{ii}^{(1)} = p_{ii}$ and

$$f_{ii}^{(n)} = \Pr[s_n = i \cap s_t \neq i, t = 1, \ldots, n-1 | s_0 = i] \qquad \text{for } n = 2, 3, \ldots$$

Given that the chain starts in state i, the infinite sum

$$f_i = \sum_{n=1}^{\infty} f_{ii}^{(n)}$$

represents the probability that state i will eventually be re-entered. If the ultimate return to state i is a certain event, then $f_i = 1$ and the state is said to be *recurrent*. On the other hand, if there is a positive probability that the state will never be re-entered, then $f_i < 1$ and the state is called *transient*. In the case of a recurrent state, the probabilities $f_{ii}^{(n)}$ sum to unity over n and hence $f_{ii}^{(n)}$ may be regarded as the probability mass function of the recurrence time n, with the mean recurrence time of state i given as

$$\mathrm{E}_i[n] = \sum_{n=1}^{\infty} n f_{ii}^{(n)}.$$

If the mean recurrence time is infinite, the state is said to be *null recurrent*. In contrast, if the state is expected to reoccur in finite time, the state is called *positive recurrent* [7].

At this point, it is worthwhile reviewing the Decomposition Theorem of Markov chains, as presented in Cox and Miller [7].

Theorem 1 *The states of an arbitrary Markov chain may be divided into two sets (one of which may be empty), one set being comprised of all the recurrent states and the other set comprising all the transient states. The recurrent states may be further decomposed into unique closed sets. Within each closed set, all states can be reached from every other state and they are all of the same type and period. A state in one closed set cannot, however, be reached from a state in another closed set.*

As an example, consider the following transition probability matrix

$$\mathbb{P} = \begin{array}{c} \\ \text{C1} \\ \text{C2} \\ \text{C3} \end{array} \left[\begin{array}{c|c|c|c} \begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \mathbb{P}_1 & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & \mathbb{P}_2 & \mathbf{0} \\ \hline \boldsymbol{A} & \boldsymbol{B} & \boldsymbol{C} & \boldsymbol{D} \end{array} \right].$$

The first two rows represent *absorbing states*, which persist indefinitely once entered. C1 and C2 represent two closed sets of recurrent states with transition matrices $\mathbb{P}_1$ and $\mathbb{P}_2$ respectively. C3 represents the transient states. The matrices $\boldsymbol{B}$ and $\boldsymbol{C}$ include the transition probabilities from the transient states to the recurrent sets C1 and C2 respectively. The matrix $\boldsymbol{D}$ includes the transition probabilities within the transient states, whilst $\boldsymbol{A}$ includes the transition probabilities from the transient states to the absorbing states [7].

B.2 Periodicity

Suppose that when a chain starts in state i, subsequent occupations of state i can only occur at times $t, 2t, 3t, \ldots$ where t is an integer greater than one. Such a state is said to be *periodic* with period t and implies that $p_{ii}^{(n)} = 0$ except when n is a multiple of t, where $p_{ii}^{(n)}$ denotes the ith diagonal element of $\mathbb{P}^n$ [7]. The *period* of state i is defined as the largest common devisor of the set $\{n > 0 : p_{ii}^{(n)} > 0\}$. On the other hand, a state i is defined as *aperiodic* if $p_{ii}^{(n)} > 0$ for all n sufficiently large or, equivalently, if the state has a period of one [6].

B.3 Irreducibility

Consider the Markov chain $\{s_t : t = 1, 2, \ldots\}$ and define the first hitting time of state i as

$$\tau_i = \inf\{t \geq 0 : s_t = i\},$$

where $\inf \emptyset = +\infty$ by convention. For two states i and j, state i is said to *lead to* state j, written $i \to j$, if $\Pr[\tau_j < \infty | s_0 = i] > 0$. Hence, state i leads to state j if state j can be reached from state i in finite time. Additionally, if both i leads to j and j leads to i, then states i and j are said to *communicate*, denoted $i \leftrightarrow j$. If state i communicates with state j for all states $i, j \in \{1, 2, \ldots, m\}$, then the Markov chain is called *irreducible*, otherwise it is a *reducible* chain [6].

An irreducible chain forms a single closed set and all its states are of the same type. If a chain is irreducible, it is therefore also possible to refer to the *entire chain* as being recurrent, periodic and so on as the case might be. Furthermore, since a finite Markov chain cannot consist only of transient states and cannot include any null recurrent states, an irreducible finite chain is necessarily positive recurrent [7]. This result is particularly relevant in the context of Markov-switching models where a finite number of regimes is assumed.

An m state Markov chain is reducible if there exists a way to label the states such that the transition probability matrix is of the form

$$\mathbb{P} = \begin{bmatrix} \boldsymbol{B} & \boldsymbol{0} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix}, \tag{B.4}$$

where $\boldsymbol{B}$ is a square matrix of dimension $1 \leq k < m$. In this case, the states corresponding to $\boldsymbol{B}$ clearly form a closed set from which the remaining states cannot be reached. Observe that if $\mathbb{P}$ is lower block-triangular, then so is $\mathbb{P}^n$ for all n. Hence, once such a process enters a state $i \leq k$, there is no possibility of ever returning to one of the states $k+1, k+2, \ldots, m$ [19].

B.4 Ergodicity

A Markov chain comprised of only positive recurrent and aperiodic states is said to be *ergodic*. It therefore follows that in the case of a finite Markov chain, a necessary requirement for ergodicity is irreducibility. If an irreducible finite Markov chain is in addition aperiodic, then it is ergodic.

Interestingly, the ergodicity of a finite Markov chain may be established by examining the eigenvalues of the transition probability matrix. Recall that a transition probability matrix is necessarily square and can therefore be reduced to a lower block-triangular matrix of the form

$$\mathbb{P} = \begin{bmatrix} \mathbb{P}_{11} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbb{P}_{21} & \mathbb{P}_{22} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & & \vdots \\ \mathbb{P}_{m1} & \mathbb{P}_{m2} & \mathbb{P}_{m3} & \cdots & \mathbb{P}_{mm} \end{bmatrix} \tag{B.5}$$

by a suitable relabeling of the states, where the matrices $\mathbb{P}_{ii}$ on the diagonal are square and irreducible [7]. Now since

$$|\,\mathbb{P} - \lambda \boldsymbol{I}| = \prod_{i=1}^{m} |\,\mathbb{P}_{ii} - \lambda \boldsymbol{I}|,$$

it follows that the spectrum of the eigenvalues of $\mathbb{P}$ is equal to the union of the spectra of the irreducible matrices $\mathbb{P}_{11}, \ldots, \mathbb{P}_{mm}$ [20]. Consequently, one need only consider the properties of the eigenvalues of irreducible, non-negative matrices in this context. In particular, the Perron-Frobenius Theorem for irreducible, non-negative matrices will prove useful. Selected results stemming from this theorem are presented below. The reader is referred to Cox and Miller [7] for a more complete account and Debreu and Herstein [10] for the proof.

Theorem 2 *Suppose* $\boldsymbol{A} \geq \mathbf{0}$ *and irreducible. Then*

1. $\boldsymbol{A}$ *has a real positive eigenvalue* λ_1 *with the following properties:*

   (a) λ_1 *has a corresponding eigenvector whose elements are strictly positive*

   (b) If λ *is any other eigenvalue of* $\boldsymbol{A}$*, then* $|\lambda| \leq \lambda_1$

(c) $\lambda_1 \leq \max_i \sum_j a_{ij}, \;\; \lambda_1 \leq \max_j \sum_i a_{ij}$

2. *if* $\boldsymbol{A}$ *has* t *eigenvalues equal in modulus to* λ_1*, these eigenvalues are all different. If* $t > 1$*, then* $\boldsymbol{A}$ *can be reduced to the following cyclic form*

$$\boldsymbol{A} = \begin{bmatrix} \mathbf{0} & \boldsymbol{A}_{12} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{A}_{13} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & & \boldsymbol{A}_{t-1,t} \\ \boldsymbol{A}_{t1} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}$$

by a permutation applied to the rows and columns of $\boldsymbol{A}$*. The submatrices on the main diagonal are square.*

Now consider the transition probability matrix $\mathbb{P}$, which may be rewritten in the lower block-triangular form of Equation (B.5) such that the eigenvalues of each square matrix on the diagonal satisfy the results of the Perron-Frobenius Theorem above. Part 1c of this theorem places an upper bound of unity on the maximum eigenvalue of a transition probability matrix $\mathbb{P}$, since the column sums of the irreducible matrices $\mathbb{P}_{ii}$ along the diagonal of $\mathbb{P}$ in Equation (B.5) must be less than or equal to unity (strictly equal to one in the case of $\mathbb{P}_{11}$). Additionally, since $\mathbb{P}\mathbf{1} = \mathbf{1}$, it follows that any transition probability matrix $\mathbb{P}$ must have an eigenvalue of one with a corresponding eigenvector proportional to $\mathbf{1}$. These results, together with Part 1b of the Perron-Frobenius Theorem, imply that every transition probability matrix must have at least one eigenvalue equal to unity in modulus with all other eigenvalues lying within the unit circle [7].

Next, note that since

$$|\mathbb{P}' - \lambda \boldsymbol{I}| = |(\mathbb{P} - \lambda \boldsymbol{I})'| = |\mathbb{P} - \lambda \boldsymbol{I}|\,,$$

$\mathbb{P}'$ shares the same spectrum as $\mathbb{P}$ and every eigenvalue in the spectrum of $\mathbb{P}$ has the same algebraic multiplicity when regarded as an eigenvalue of $\mathbb{P}'$ [20]. Hence, $\mathbb{P}'$ will have a unit eigenvalue with an associated eigenvector, say $\boldsymbol{\pi}$, such that $\mathbb{P}'\boldsymbol{\pi} = \boldsymbol{\pi}$ with $\boldsymbol{\pi}$ appropriately scaled so that $\mathbf{1}'\boldsymbol{\pi} = 1$. It follows that $\mathbb{P}'^2\,\boldsymbol{\pi} = \mathbb{P}'\,\boldsymbol{\pi} = \boldsymbol{\pi}$ and by induction $\mathbb{P}'^n\,\boldsymbol{\pi} = \boldsymbol{\pi}$. Hence, if the initial distribution of the states is $\boldsymbol{\pi}$, this distribution will persist for all subsequent time periods. The eigenvector $\boldsymbol{\pi}$ corresponding to the unit eigenvalue of $\mathbb{P}'$ therefore represents the *stationary distribution* of the states [7]. However, it should be noted that $\boldsymbol{\pi}$ is not necessarily unique, since $\mathbb{P}$ may have more than one unit eigenvalue. In order to ensure that the stationary distribution of the states is in fact unique, one must therefore restrict the algebraic multiplicity of the unit eigenvalue of $\mathbb{P}$ to one. Together with the restriction that the elements of $\boldsymbol{\pi}$ sum to unity, $\boldsymbol{\pi}$ will then represent a unique stationary distribution.

Moreover, since $\boldsymbol{\pi}$ is the eigenvector corresponding to the largest eigenvalue of $\mathbb{P}'$, it follows from Part 1a of the Perron-Frobenius Theorem that the elements of $\boldsymbol{\pi}$ are strictly positive. Now suppose that state j of a finite Markov chain is transient. Then it must be the case that $p_{ij}^{(n)} \to 0$ as $n \to \infty$ for all i so that state j is eventually exited. However, this would imply that the jth element of $\boldsymbol{\pi}$ equals zero, since $\mathbb{P}'^n \boldsymbol{\pi} = \boldsymbol{\pi}$ [13]. Hence, the existence of a positive stationary distribution ensures that no states are transient and therefore $\mathbb{P}$ cannot be reduced to the form given by Equation (B.4), where the sub-matrix $\boldsymbol{C}$ is assumed to include at least one element greater than zero. The case where $\boldsymbol{C} = \boldsymbol{0}$ would result in two closed sets with transition matrices $\boldsymbol{B}$ and $\boldsymbol{D}$ respectively. However, $\mathbb{P}$ would then have at least two unit eigenvalues, since $\boldsymbol{B}\boldsymbol{1} = \boldsymbol{1}$ and $\boldsymbol{D}\boldsymbol{1} = \boldsymbol{1}$. This case is therefore excluded by the restriction imposed on the algebraic multiplicity of the unit eigenvalue of $\mathbb{P}$. Consequently, a transition probability matrix with a single unit eigenvalue cannot be reduced to the form given by Equation (B.4) for any non-negative sub-matrix $\boldsymbol{C}$. Such a transition probability matrix is therefore irreducible.

Now consider an irreducible Markov chain; that is, a chain with exactly one eigenvalue of $\mathbb{P}$ equal to unity and all other eigenvalues on or within the unit circle. Part 2 of the Perron-Frobenius Theorem implies that if $\mathbb{P}$ has exactly t eigenvalues on the unit circle, then the states can be partitioned into t mutually exclusive and exhaustive subsets $S_1, \ldots, S_t$ such that a one-step transition from a state in S_i can only lead to a state in S_{i+1} (if $i = t$, then let $t + 1 = 1$). Hence the period of an irreducible Markov chain is given by the number of eigenvalues of $\mathbb{P}$ which have unit modulus [7]. If $t > 1$, the Markov chain is periodic with period t. On the other hand, if $t = 1$, then the Markov chain is of period one and is therefore aperiodic. It follows that a sufficient condition for ergodicity in a Markov chain is that the transition probability matrix $\mathbb{P}$ has exactly one unit eigenvalue with *all* other eigenvalues less than one in modulus [19].

In order to examine the consequences of ergodicity, consider first the following result from linear algebra.

Result 2 *Let* $\boldsymbol{A}$ *be a* $k \times k$ *matrix with* $n \leq k$ *distinct eigenvalues* $\lambda_1, \lambda_2, \ldots, \lambda_n$*. Then there exists a non-singular matrix* $\boldsymbol{H}$ *such that*

$$\boldsymbol{H}\boldsymbol{A}\boldsymbol{H}^{-1} = \boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{\Lambda}_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \boldsymbol{\Lambda}_n \end{bmatrix} \qquad \text{or} \qquad \boldsymbol{A} = \boldsymbol{H}^{-1}\boldsymbol{\Lambda}\boldsymbol{H}$$

where

$$\boldsymbol{\Lambda}_i = \begin{bmatrix} \lambda_i & 1 & 0 & \cdots & 0 \\ 0 & \lambda_i & 1 & & \vdots \\ 0 & 0 & \lambda_i & \ddots & 0 \\ \vdots & & \ddots & \ddots & 1 \\ 0 & \cdots & \cdots & 0 & \lambda_i \end{bmatrix} \tag{B.6}$$

is a square matrix of dimension r_i *equal to the algebraic multiplicity of the* i*th distinct eigenvalue. If the algebraic multiplicity of* λ_i *is unity, then* $\boldsymbol{\Lambda}_i$ *reduces to the scalar* λ_i*.*

This decomposition of $\boldsymbol{A}$ *is known as the* Jordan canonical form.

It can be further shown that

$$\boldsymbol{\Lambda}_i^t = \begin{bmatrix} \lambda_i^t & \binom{t}{1}\lambda_i^{t-1} & \cdots & \binom{t}{r_i-1}\lambda_i^{t-r_i+1} \\ 0 & \lambda_i^t & \cdots & \binom{t}{r_i-2}\lambda_i^{t-r_i+2} \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_i^t \end{bmatrix}. \tag{B.7}$$

Now suppose a finite Markov chain is ergodic with transition probability matrix $\mathbb{P}$. Then, the Jordan canonical decomposition of $\mathbb{P}$ is given by

$$\mathbb{P} = \boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}^{-1}$$

where

$$\boldsymbol{\Lambda} = \begin{bmatrix} 1 & & & \mathbf{0} \\ & \boldsymbol{\Lambda}_2 & & \\ & & \ddots & \\ \mathbf{0} & & & \boldsymbol{\Lambda}_n \end{bmatrix}.$$

The columns of $\boldsymbol{H}$ are the eigenvectors of $\mathbb{P}$ and the $\boldsymbol{\Lambda}_i$ matrices are as defined by Equation (B.6) with diagonal elements strictly less than one in modulus [7]. This result will prove useful in examining the limiting behaviour of $\mathbb{P}^n$ since clearly

$$\mathbb{P}^n = \boldsymbol{H}\boldsymbol{\Lambda}^n\boldsymbol{H}^{-1}.$$

Since the first diagonal element of $\boldsymbol{\Lambda}$ is unity and the eigenvalues corresponding to each of the $\boldsymbol{\Lambda}_i$ matrices are less than one in modulus, it follows from Equation (B.7) that $\boldsymbol{\Lambda}^n$ converges to a matrix with a first diagonal element of one and zeros elsewhere as $n \to \infty$. Hence,

$$\lim_{n\to\infty} \mathbb{P}^n = \boldsymbol{x}\boldsymbol{y}'$$

where $\boldsymbol{x}$ is the first column of $\boldsymbol{H}$ and $\boldsymbol{y}'$ is the first row of $\boldsymbol{H}^{-1}$. Now the first column of $\boldsymbol{H}$ is the eigenvector of $\mathbb{P}$ corresponding to the unit eigenvalue, which was shown to be proportional to $\mathbf{1}$ so that $\boldsymbol{x} \propto \mathbf{1}$. Furthermore, the transpose of the first row of $\boldsymbol{H}^{-1}$ is simply the eigenvector of $\mathbb{P}'$ associated with its unit eigenvalue, since

$$\begin{aligned}
\mathbb{P}\,\boldsymbol{H} &= \boldsymbol{H}\boldsymbol{\Lambda} \\
\boldsymbol{H}'\,\mathbb{P}' &= \boldsymbol{\Lambda}\boldsymbol{H}' \\
\mathbb{P}'\,\boldsymbol{H}'^{-1} &= \boldsymbol{H}'^{-1}\boldsymbol{\Lambda}.
\end{aligned}$$

Hence, $\boldsymbol{y} = \boldsymbol{\pi}$ with $\mathbf{1}'\boldsymbol{\pi} = 1$ and therefore

$$\lim_{n\to\infty} \mathbb{P}^n \propto \mathbf{1}\boldsymbol{\pi}'. \tag{B.8}$$

From the earlier discussion on the properties of a Markov chain, it was noted that $\mathbb{P}^n$ represents the transition probabilities after n periods, the rows of which must therefore sum to unity [19]. Consequently, the proportionality sign in Equation (B.8) can be replaced with an equality such that

$$\lim_{n\to\infty} \mathbb{P}^n = \mathbf{1}\boldsymbol{\pi}'.$$

From the previous discussion on the n-period ahead forecast of a Markov chain, it therefore follows that

$$\lim_{n\to\infty} \boldsymbol{p}^{(n)} = \lim_{n\to\infty} \mathbb{P}'^n\,\boldsymbol{p}^{(0)} = \boldsymbol{\pi}\mathbf{1}'\boldsymbol{p}^{(0)} = \boldsymbol{\pi},$$

for any initial probability distribution vector $\boldsymbol{p}^{(0)}$. It has thus been demonstrated that $\boldsymbol{\pi}$ is not only a stationary distribution, but also the *limiting distribution* of the states of an ergodic Markov chain. In this case, the elements of $\boldsymbol{\pi}$ are referred to as *ergodic probabilities.*

Note that in Equation (B.3), $\boldsymbol{p}^{(0)} = \boldsymbol{\xi}_t$ so that the system was assumed to be in a specific state at time t with probability one. Letting $n \to \infty$ in this equation, $\boldsymbol{\pi}$ may therefore be regarded as the expected state of the world in the infinite future which, under the assumption of ergodicity, is invariant to the initial state. Hence, a finite ergodic chain settles down in the long run to a condition of statistical equilibrium independent of the initial conditions [7]. This result is a remarkable property of ergodic chains.

Note that although an irreducible finite chain will possess a stationary distribution, irreducibility in itself is not sufficient to ensure that this distribution is also limiting. In particular, the n-step transition probability matrix $\mathbb{P}^n$ of an irreducible periodic chain cannot converge to any fixed limit as the chain jumps from one subset of states to another indefinitely [19]. Ergodicity is therefore a stronger assumption than irreducibility for finite Markov chains.

C Parameter Estimation in the MS-VAR Model

The Expectation-Maximisation (EM) algorithm is an iterative maximum likelihood estimation technique which has gained momentum in a number of application areas in recent years. In the time series context, the EM algorithm has proved useful in estimating the parameters of observed processes which depend upon unobserved stochastic variables or states. Its adoption in the estimation of regime-switching regressions should therefore seem unsurprising.

The Markov-switching vector autoregressive model has three sets of parameters which require estimation. For given states $\boldsymbol{\xi}_t$ and lagged endogenous variables $\boldsymbol{Y}_{t-1} = (\boldsymbol{y}'_{t-1}, \boldsymbol{y}'_{t-2}, \ldots, \boldsymbol{y}'_1, \boldsymbol{y}'_0, \ldots, \boldsymbol{y}'_{1-p})'$, let $p(\boldsymbol{y}_t|\boldsymbol{\xi}_t, \boldsymbol{Y}_{t-1})$ be the conditional probability density function of the observed data $\boldsymbol{y}_t$ at time $t = 1, \ldots, T$. Now suppose $\boldsymbol{y}_t$ can be described by the MS-VAR model

$$\boldsymbol{y}_t = \boldsymbol{\nu}(s_t) + \boldsymbol{\Pi}_1(s_t)\boldsymbol{y}_{t-1} + \boldsymbol{\Pi}_2(s_t)\boldsymbol{y}_{t-2} + \ldots + \boldsymbol{\Pi}_p(s_t)\boldsymbol{y}_{t-p} + \boldsymbol{\varepsilon}_t,$$

with innovations $\boldsymbol{\varepsilon}_t|s_t = i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Omega}_i)$. Then the conditional distribution of $\boldsymbol{y}_t$ will itself be multivariate Gaussian, given as

$$p(\boldsymbol{y}_t|\boldsymbol{\xi}_t = \boldsymbol{\iota}_i, \boldsymbol{Y}_{t-1}) = (2\pi)^{-k/2}|\boldsymbol{\Omega}_i|^{-1/2} \exp -\frac{1}{2} (\boldsymbol{y}_t - \boldsymbol{\mu}_{it})' \boldsymbol{\Omega}_i^{-1} (\boldsymbol{y}_t - \boldsymbol{\mu}_{it}) \tag{C.1}$$

with

$$\boldsymbol{\mu}_{it} = \mathrm{E}[\boldsymbol{y}_t|\boldsymbol{\xi}_t = \boldsymbol{\iota}_i, \boldsymbol{Y}_{t-1}] = \boldsymbol{\nu}(s_t = i) + \sum_{j=1}^{p} \boldsymbol{\Pi}_j(s_t = i)\boldsymbol{y}_{t-j}.$$

Define $\boldsymbol{\theta}_i$ as a vector which stacks all the parameters $\boldsymbol{\nu}, \boldsymbol{\Pi}_1, \ldots, \boldsymbol{\Pi}_p, \mathrm{vech}(\boldsymbol{\Omega})$ describing the observed time series in regime i, where the vech operator stacks the columns of the square matrix $\boldsymbol{\Omega}$, starting each column at its diagonal element [22]. Collect the parameters for each regime in the vector $\boldsymbol{\theta} = (\boldsymbol{\theta}'_1, \ldots, \boldsymbol{\theta}'_m)'$. The parameter vector $\boldsymbol{\theta}$ therefore fully defines the generating process of the observed data. However, since the true regime at any point in time is unknown, it is insufficient to only estimate $\boldsymbol{\theta}$. It is of course also necessary to estimate the parameters in the model of the regime generating process. Recall from Equation (B.2) that the Markov-switching regime generating process may be expressed as a first-order vector autoregression with parameters $\boldsymbol{\rho} = \text{vec}(\mathbb{P})$ and $\boldsymbol{\xi}_0$ representing the transition probabilities and initial state vector respectively, where the vec operator stacks the columns of $\mathbb{P}$ [22]. Consequently, the MS-VAR model is fully specified by both the regime generating parameters and those parameters governing the observed data, which may be assembled in a single vector $\boldsymbol{\lambda} = (\boldsymbol{\theta}', \boldsymbol{\rho}', \boldsymbol{\xi}_0')'$ [34].

C.1 The E-Step: Estimation of the Regimes

The expectation step in the EM algorithm is concerned with the determination of the sequence of true state vectors $\{\boldsymbol{\xi}_t : t = 1, \ldots, T\}$ given an observed time series of T observations $\boldsymbol{Y}_T$ and known parameters $\boldsymbol{\lambda} = (\boldsymbol{\theta}', \boldsymbol{\rho}', \boldsymbol{\xi}_0')'$ of the MS-VAR model. The algorithm must therefore be initialised by specifying starting values $\boldsymbol{\theta}^{(0)}$ and $\boldsymbol{\rho}^{(0)}$ for $\boldsymbol{\theta}$ and $\boldsymbol{\rho}$ respectively, whilst the parameter vector $\boldsymbol{\xi}_0$ may be avoided initially by specifying starting values for $\Pr[\boldsymbol{\xi}_1 = \boldsymbol{\iota}_i | \boldsymbol{Y}_0]$ for $i = 1, \ldots, m$, where $\boldsymbol{Y}_0 = (\boldsymbol{y}_0', \ldots, \boldsymbol{y}_{1-p}')'$.

At the outset, each state vector $\boldsymbol{\xi}_t$ is estimated by

$$\hat{\boldsymbol{\xi}}_{t|t} = \text{E}[\boldsymbol{\xi}_t | \boldsymbol{Y}_t] = \begin{pmatrix} \Pr[\boldsymbol{\xi}_t = \boldsymbol{\iota}_1 | \boldsymbol{Y}_t] \\ \vdots \\ \Pr[\boldsymbol{\xi}_t = \boldsymbol{\iota}_m | \boldsymbol{Y}_t] \end{pmatrix},$$

conditioning on the observed data up to and including time point t. Since each element of $\boldsymbol{\xi}_t$ is binary, $\hat{\boldsymbol{\xi}}_{t|t}$ has a dual interpretation. On the one hand, $\hat{\boldsymbol{\xi}}_{t|t}$ may interpreted as the conditional mean of $\boldsymbol{\xi}_t$ and is therefore an unbiased estimator of the true state vector $\boldsymbol{\xi}_t$ based on the observed data up to and including time point t. On the other hand, since the elements of $\hat{\boldsymbol{\xi}}_{t|t}$ sum to unity, $\hat{\boldsymbol{\xi}}_{t|t}$ may be regarded as the conditional probability mass function from which the regime at time t arises. The probabilities included within $\hat{\boldsymbol{\xi}}_{t|t}$ are referred to as *filtered probabilities* and may be derived analytically by applying Bayes' rule to obtain

$$\Pr[\boldsymbol{\xi}_t | \boldsymbol{Y}_t] \equiv \Pr[\boldsymbol{\xi}_t | \boldsymbol{y}_t, \boldsymbol{Y}_{t-1}] = \frac{p(\boldsymbol{y}_t | \boldsymbol{\xi}_t, \boldsymbol{Y}_{t-1}) \Pr[\boldsymbol{\xi}_t | \boldsymbol{Y}_{t-1}]}{p(\boldsymbol{y}_t | \boldsymbol{Y}_{t-1})}. \tag{C.2}$$

Noting that

$$p(\boldsymbol{y}_t | \boldsymbol{Y}_{t-1}) = \sum_{\boldsymbol{\xi}_t} p(\boldsymbol{y}_t | \boldsymbol{\xi}_t, \boldsymbol{Y}_{t-1}) \Pr[\boldsymbol{\xi}_t | \boldsymbol{Y}_{t-1}]$$

and

$$\Pr[\boldsymbol{\xi}_t|\boldsymbol{Y}_{t-1}] = \sum_{\boldsymbol{\xi}_{t-1}} \Pr[\boldsymbol{\xi}_t|\boldsymbol{\xi}_{t-1}] \Pr[\boldsymbol{\xi}_{t-1}|\boldsymbol{Y}_{t-1}],$$

where $\Pr[\boldsymbol{\xi}_t|\boldsymbol{\xi}_{t-1}, \boldsymbol{Y}_{t-1}] = \Pr[\boldsymbol{\xi}_t|\boldsymbol{\xi}_{t-1}]$ follows from the assumed Markovian nature of the regime generating process, Equation (C.2) may be rewritten as

$$\Pr[\boldsymbol{\xi}_t|\boldsymbol{Y}_t] = \frac{p(\boldsymbol{y}_t|\boldsymbol{\xi}_t, \boldsymbol{Y}_{t-1}) \Pr[\boldsymbol{\xi}_t|\boldsymbol{Y}_{t-1}]}{\sum_{\boldsymbol{\xi}_t} p(\boldsymbol{y}_t|\boldsymbol{\xi}_t, \boldsymbol{Y}_{t-1}) \Pr[\boldsymbol{\xi}_t|\boldsymbol{Y}_{t-1}]} \tag{C.3}$$

$$= \frac{\sum_{\boldsymbol{\xi}_{t-1}} p(\boldsymbol{y}_t|\boldsymbol{\xi}_t, \boldsymbol{Y}_{t-1}) \Pr[\boldsymbol{\xi}_t|\boldsymbol{\xi}_{t-1}] \Pr[\boldsymbol{\xi}_{t-1}|\boldsymbol{Y}_{t-1}]}{\sum_{\boldsymbol{\xi}_t} \sum_{\boldsymbol{\xi}_{t-1}} p(\boldsymbol{y}_t|\boldsymbol{\xi}_t, \boldsymbol{Y}_{t-1}) \Pr[\boldsymbol{\xi}_t|\boldsymbol{\xi}_{t-1}] \Pr[\boldsymbol{\xi}_{t-1}|\boldsymbol{Y}_{t-1}]}. \tag{C.4}$$

Starting at time $t = 1$, Equation (C.3) is immediately operational, since $p(\boldsymbol{y}_1|\boldsymbol{\xi}_1, \boldsymbol{Y}_0)$ is fully specified by Equation (C.1) with parameters $\boldsymbol{\theta}^{(0)}$ and initial values are also assumed for $\Pr[\boldsymbol{\xi}_1|\boldsymbol{Y}_0]$. Thereafter, the filtered probabilities can be determined by solving Equation (C.4) sequentially for $t = 2, 3, \ldots, T$. On the first run of the algorithm, the transition probabilities $\Pr[\boldsymbol{\xi}_t|\boldsymbol{\xi}_{t-1}]$ in this equation are included in the starting vector $\boldsymbol{\rho}^{(0)}$. The required filtered probabilities $\Pr[\boldsymbol{\xi}_{t-1}|\boldsymbol{Y}_{t-1}]$ are the solutions to Equation (C.4) from the previous iteration. Proceeding in this manner from $t = 1$ through to $t = T$ will thus produce the desired sequence of filtered probabilities $\{\hat{\boldsymbol{\xi}}_{t|t} : t = 1, \ldots, T\}$ [34].

The aforementioned filter recursions produce estimates of $\{\boldsymbol{\xi}_t : t = 1, \ldots, T\}$ based on information up to and including time point t. This procedure is clearly a limited information technique since it neglects the information $\boldsymbol{Y}_{t+1.T} = (\boldsymbol{y}'_{t+1}, \ldots, \boldsymbol{y}'_T)'$ included in the observed time series in inferences about $\boldsymbol{\xi}_t$ at each time point $t < T$. A more desirable estimator of the unobserved regimes would therefore be

$$\hat{\boldsymbol{\xi}}_{t|T} = \mathrm{E}[\boldsymbol{\xi}_t|\boldsymbol{Y}_T] = \begin{pmatrix} \Pr[\boldsymbol{\xi}_t = \boldsymbol{\iota}_1|\boldsymbol{Y}_T] \\ \vdots \\ \Pr[\boldsymbol{\xi}_t = \boldsymbol{\iota}_m|\boldsymbol{Y}_t] \end{pmatrix},$$

which incorporates all the information included in the observed time series in inferences about $\boldsymbol{\xi}_t$. The probabilities in $\hat{\boldsymbol{\xi}}_{t|T}$ are referred to as *smoothed probabilities* and may be derived in terms of the filtered probabilities as follows

$$
\begin{aligned}
\Pr[\boldsymbol{\xi}_t|\boldsymbol{Y}_T] &= \sum_{\boldsymbol{\xi}_{t+1}} \Pr[\boldsymbol{\xi}_t|\boldsymbol{\xi}_{t+1},\boldsymbol{Y}_T]\Pr[\boldsymbol{\xi}_{t+1}|\boldsymbol{Y}_T] \\
&= \sum_{\boldsymbol{\xi}_{t+1}} \Pr[\boldsymbol{\xi}_t|\boldsymbol{\xi}_{t+1},\boldsymbol{Y}_t]\Pr[\boldsymbol{\xi}_{t+1}|\boldsymbol{Y}_T] \\
&= \sum_{\boldsymbol{\xi}_{t+1}} \frac{\Pr[\boldsymbol{\xi}_t|\boldsymbol{Y}_t]\Pr[\boldsymbol{\xi}_{t+1}|\boldsymbol{\xi}_t]}{\Pr[\boldsymbol{\xi}_{t+1}|\boldsymbol{Y}_t]}\Pr[\boldsymbol{\xi}_{t+1}|\boldsymbol{Y}_T] \\
&= \sum_{\boldsymbol{\xi}_{t+1}} \frac{\Pr[\boldsymbol{\xi}_t|\boldsymbol{Y}_t]\Pr[\boldsymbol{\xi}_{t+1}|\boldsymbol{\xi}_t]}{\sum_{\boldsymbol{\xi}_t}\Pr[\boldsymbol{\xi}_{t+1}|\boldsymbol{\xi}_t]\Pr[\boldsymbol{\xi}_t|\boldsymbol{Y}_t]}\Pr[\boldsymbol{\xi}_{t+1}|\boldsymbol{Y}_T]. \qquad \text{(C.5)}
\end{aligned}
$$

The second step above follows since

$$
\begin{aligned}
\Pr[\boldsymbol{\xi}_t|\boldsymbol{\xi}_{t+1},\boldsymbol{Y}_T] &\equiv \Pr[\boldsymbol{\xi}_t|\boldsymbol{\xi}_{t+1},\boldsymbol{Y}_t,\boldsymbol{Y}_{t+1.T}] \\
&= \frac{p(\boldsymbol{Y}_{t+1.T}|\boldsymbol{\xi}_t,\boldsymbol{\xi}_{t+1},\boldsymbol{Y}_t)\Pr[\boldsymbol{\xi}_t|\boldsymbol{\xi}_{t+1},\boldsymbol{Y}_t]}{p(\boldsymbol{Y}_{t+1.T}|\boldsymbol{\xi}_{t+1},\boldsymbol{Y}_t)} \\
&= \Pr[\boldsymbol{\xi}_t|\boldsymbol{\xi}_{t+1},\boldsymbol{Y}_t],
\end{aligned}
$$

where $p(\boldsymbol{Y}_{t+1.T}|\boldsymbol{\xi}_t,\boldsymbol{\xi}_{t+1},\boldsymbol{Y}_t) = p(\boldsymbol{Y}_{t+1.T}|\boldsymbol{\xi}_{t+1},\boldsymbol{Y}_t)$ is a Markov property of the regime generating process.

The smoothed probabilities $\{\hat{\boldsymbol{\xi}}_{t|T} : t = 1,\ldots,T\}$ may therefore be found by iteratively solving Equation (C.5) backward from $t = T-1$ through to $t = 1$ and using the already calculated filtered probabilities together with the smoothed probability vector from the previous iteration. The smoothing algorithm is initialised with the final filtered probability vector $\hat{\boldsymbol{\xi}}_{T|T}$, which is of course equivalent the smoothed probability vector since $t = T$ [34].

The above algorithm, known as the BLHK filter and smoother after its developers Baum, Lindgren, Hamilton and Kim, delivers unbiased estimates $\{\hat{\boldsymbol{\xi}}_{t|T} : t = 1,\ldots,T\}$ of the unobserved regimes $\{\boldsymbol{\xi}_t : t = 1,\ldots,T\}$ based on all the information included in the observed time series, assuming the parameters $\boldsymbol{\lambda} = (\boldsymbol{\theta}',\boldsymbol{\rho}',\boldsymbol{\xi}_0')'$ are known. Of course, the parameters in $\boldsymbol{\lambda}$ are not known and are therefore assigned (arbitrary) starting values in order to initialise the algorithm. The scientific estimation of these parameters was not possible at the outset since at least some of the parameters in $\boldsymbol{\lambda}$ are assumed to be dependent on unobserved regimes. The expectation step of the EM algorithm, however, produces estimates of these unobserved regimes. Treating these estimates as the true state of the world, at least momentarily, enables the maximum likelihood estimation of $\boldsymbol{\lambda}$. This estimation is the focus of the M-step in the EM algorithm, which is considered next.

C.2 The M-Step: Estimation of the Model Parameters

The maximisation step in the EM algorithm is, as the name would suggest, concerned with the maximisation of the (log-) likelihood function with respect to the model parameters $\boldsymbol{\lambda}$. This likelihood function conditional on the observations $\boldsymbol{Y}_T$ may be expressed as

$$\begin{aligned} L(\boldsymbol{\lambda}|\boldsymbol{Y}_T) &= p(\boldsymbol{Y}_T|\boldsymbol{\lambda}) \\ &= \sum_{\boldsymbol{\xi}} p(\boldsymbol{Y}_T, \boldsymbol{\xi}|\boldsymbol{\lambda}) \\ &= \sum_{\boldsymbol{\xi}} p(\boldsymbol{Y}_T|\boldsymbol{\xi}, \boldsymbol{\theta}) \Pr[\boldsymbol{\xi}|\boldsymbol{\rho}, \boldsymbol{\xi}_0], \end{aligned}$$

with state vector $\boldsymbol{\xi} = \boldsymbol{\xi}_T \otimes \boldsymbol{\xi}_{T-1} \otimes \ldots \otimes \boldsymbol{\xi}_1$. Since the elements of each of the true state vectors in $\{\boldsymbol{\xi}_t : t = 1, \ldots, T\}$ are binary, this Kronecker product will produce a vector with zero elements in all but one of its rows. The non-zero element will be unitary and the position of this element within the vector $\boldsymbol{\xi}$ uniquely determines the states at all time points. The summation is therefore taken over all possible combinations of states in the observed time frame.

The above formulation of the likelihood function is relevant in that it illustrates that the likelihood can be factorised into two terms, where one term depends solely on the parameters $\boldsymbol{\theta}$ of observed time series and the other depends exclusively on the Markov chain parameters $\boldsymbol{\rho}$ and $\boldsymbol{\xi}_0$ associated with the regime generating process. From the properties of Markov processes, these two terms may be expressed as

$$p(\boldsymbol{Y}_T|\boldsymbol{\xi}, \boldsymbol{\theta}) = \prod_{t=1}^{T} p(\boldsymbol{y}_t|\boldsymbol{\xi}_t, \boldsymbol{Y}_{t-1}, \boldsymbol{\theta})$$

and

$$\Pr[\boldsymbol{\xi}|\boldsymbol{\rho}, \boldsymbol{\xi}_0] = \prod_{t=1}^{T} \Pr[\boldsymbol{\xi}_t|\boldsymbol{\xi}_{t-1}, \boldsymbol{\rho}].$$

The maximum likelihood estimates of the parameters are found by maximising the likelihood function $L(\boldsymbol{\lambda}|\boldsymbol{Y}_T)$ with respect to $\boldsymbol{\lambda}$ subject to the adding-up restrictions

$$\begin{aligned} \mathbb{P}\,\mathbf{1} &= \mathbf{1} \\ \mathbf{1}'\boldsymbol{\xi}_0 &= 1 \end{aligned}$$

and non-negativity constraints

$$\boldsymbol{\rho} \geq \mathbf{0}, \boldsymbol{\xi}_0 \geq \mathbf{0}, \boldsymbol{\Omega}_{ii.j} \geq \mathbf{0},$$

for $i = 1, \ldots, k$ and states $j \in \{1, \ldots, m\}$.

If the non-negativity restrictions can be ensured, then the maximum likelihood estimator of $\boldsymbol{\lambda}$ is given by the first-order condition of the constrained log-likelihood function

$$\ln L^*(\boldsymbol{\lambda}|\boldsymbol{Y}_T) = \ln L(\boldsymbol{\lambda}|\boldsymbol{Y}_T) - \boldsymbol{\kappa}_1'(\mathbb{P}\,\mathbf{1} - \mathbf{1}) - \kappa_2(\mathbf{1}' - 1),$$

where $\boldsymbol{\kappa}_1$ and κ_2 are the Lagrange multipliers associated with the adding-up restrictions on $\boldsymbol{\rho}$ and $\boldsymbol{\xi}_0$ respectively. The normal equations for each of the three distinct parameter sets $\boldsymbol{\theta}, \boldsymbol{\rho}$ and $\boldsymbol{\xi}_0$ included in $\boldsymbol{\lambda}$ are then given by the set of simultaneous equations

$$\begin{aligned} \frac{\partial \ln L(\boldsymbol{\lambda}|\boldsymbol{Y}_T)}{\partial \boldsymbol{\theta}'} &= \mathbf{0} \\ \frac{\partial \ln L(\boldsymbol{\lambda}|\boldsymbol{Y}_T)}{\partial \boldsymbol{\rho}'} - \boldsymbol{\kappa}_1'(\mathbf{1}' \otimes \boldsymbol{I}) &= \mathbf{0} \\ \frac{\partial \ln L(\boldsymbol{\lambda}|\boldsymbol{Y}_T)}{\partial \boldsymbol{\xi}_0'} - \kappa_2 \mathbf{1}' &= \mathbf{0}. \end{aligned}$$

Closed form expressions for the maximum likelihood estimators of $\boldsymbol{\theta}, \boldsymbol{\rho}$ and $\boldsymbol{\xi}_0$ may be found by solving the above set of equations, the details of which are given in Krolzig [34]. The dependence of these solutions on the unobserved regimes $\boldsymbol{\xi}_t$ is resolved by substituting in the regime estimates obtained through the smoothing algorithm of the previous expectation step. The M-step thereby produces a set of maximum likelihood estimates $\hat{\boldsymbol{\lambda}}$ for the model parameters.

The above discussion has described only a single iteration of the EM algorithm. Subsequent iterations proceed by updating the filtered and smoothed probabilities in the E-step using the estimated parameter vector $\hat{\boldsymbol{\lambda}}$ from the previous M-step in place of the true, but unknown parameter vector. The updated regime estimates are then substituted into the normal equations in the M-step to obtain a new set of maximum likelihood estimates for $\boldsymbol{\lambda}$ [35]. The algorithm continues to iterate through the expectation and maximisation steps in this manner until the gain in likelihood is negligible and convergence is achieved.

D Computer Code for MS-VECM

The following computer code estimates the MSIH(4)-VECM(1) referred to in Section 4.3 using the MS-VAR 1.30 package in Ox 3.00. The code has largely been adapted from Krolzig [35].

```
#include <oxstd.h>
#import  <C:\ProgramFiles\OxMetrics4\Ox\packages\msvar130>

main() {
    decl time=timer();
    decl msvar = new MSVAR();
    msvar->IsOxPack(FALSE);
    msvar->Load("C:\Data.xls");
    msvar->SetOptions(TRUE,TRUE,TRUE);
    // settings (StdErrors, DrawResults, Save)
    msvar->SetPrint(TRUE,TRUE);
    // all results are printed
    msvar->SetEmOptions(1e-6, 5000, 10);
    // EmAlg specification (tolerance, max iters, MSteps)

    decl M=4;                           //  number of regimes
    decl p=1;                           //  number of lags
    decl fModel=MSIH;                   //  model type

    // ENDOGENOUS VARIABLES:
    msvar->Select(Y_VAR,{"DLNPPISA", 0, p, "DLNPPIUS", 0, p,
                         "DLNEXCH", 0, p, "DTBILLSA", 0, p});

    // EXOGENOUS VARIABLES:
    msvar->Select(X_VAR,{"DTBILLUS", 1, p, "DLNGOLD", 0, 0,
                         "DLNOIL", 0, 1, "SC1", 0, 0, "SC2",
                         0, 0, "SC3", 0, 0,  "PPP", 1, 1,
                         "UIP", 1, 1});

    // INITIAL VALUES FOR REGIMES:
    msvar->Select(S_VAR,{"REGIME", 0 ,0});

    msvar->SetSample(1972,1,2007,1);    // time period
    msvar->SetModel(fModel, M);
```

```
print(msvar->Estimate());
println("\nIsConverged=",msvar->IsConverged());

// STANDARD ERRORS:
println("\nStandard errors:");
msvar->StdErr();
println("\nStandard errors:");
msvar->PrintStdErr();
println("\nVariance-covariance matrix:");
print(msvar->GetCovar());
msvar->PrintCovar();

// GRAPHICS:
msvar->DrawResults();
SaveDrawWindow("results.gwg");
msvar->DrawErrors(TRUE);
SaveDrawWindow("errors.gwg");
msvar->DrawFit();
SaveDrawWindow("fitted.gwg");
msvar->DrawModelAnalysis();
SaveDrawWindow("analysis.gwg");
msvar->CycleDating();

// REGIME PROBABILITIES:
println("\nSmoothed regime probabilities:");
println("%10.4f", msvar->GetProbSt()');



// RESIDUALS:
decl resid = msvar->GetU();
println("Residuals:");
print(resid');

delete msvar;
print("\n\n****\ttime passed: ", timespan(time),
      "\t****\n");

}
```