JOURNAL OF APPLIED MATHEMATICS AND DECISION SCIENCES, 7(1), 29–48


# A Case Study of the Residual-Based Cointegration Procedure

RICCARDO BIONDINI riccardo@uow.edu.au
*Department of Accounting & Finance, the University of Wollongong, Sydney, Australia.*

YAN-XIA LIN† yanxia@uow.edu.au
*School of Mathematics & Applied Statistics, the University of Wollongong, Sydney, Australia.*

MICHAEL MCCRAE mccrae@uow.edu.au
*Department of Accounting & Finance, the University of Wollongong,Sydney, Australia.*

**Abstract.** The study of long-run equilibrium processes is a significant component of economic and finance theory. The Johansen technique for identifying the existence of such long-run stationary equilibrium conditions among financial time series allows the identification of all potential linearly independent cointegrating vectors within a given system of eligible financial time series. The practical application of the technique may be restricted, however, by the pre-condition that the underlying data generating process fits a finite-order vector autoregression (VAR) model with white noise. This paper studies an alternative method for determining cointegrating relationships without such a pre-condition. The method is simple to implement through commonly available statistical packages. This 'residual-based cointegration' (RBC) technique uses the relationship between cointegration and univariate Box-Jenkins ARIMA models to identify cointegrating vectors through the rank of the covariance matrix of the residual processes which result from the fitting of univariate ARIMA models. The RBC approach for identifying multivariate cointegrating vectors is explained and then demonstrated through simulated examples. The RBC and Johansen techniques are then both implemented using several real-life financial time series.

**Keywords:** Equilibrium processes, Cointegration, Johansen, ARIMA models.

## 1. Introduction

The long-run equilibrating potential of financial time series such as foreign exchange rates, stock prices, interest rates and their relation to market efficiency in both spot and forward markets are probably the most intensively researched topics in cointegration analysis (Copland, 1991; Crowder, 1996). Cointegration occurs if attractor forces within a vector of financial

† Requests for reprints should be sent to Yan-Xia Lin, School of Mathematics & Applied Statistics, the University of Wollongong, Sydney, Australia.

time series keep the series in close proximity or 'long-run equilibrium' so that a linear combination of component series forms a stationary series in itself (Granger, 1981; Engle and Granger, 1987; Johansen, 1988; Johansen and Juselius, 1990; Layton and Tan, 1992). Should the variables drift away from equilibrium for a certain period of time, equilibrating economic forces are expected to act, thus restoring equilibrium.

The concept of cointegration, introduced by Granger (1981) and further developed by Engle and Granger (1987), defined long-run equilibrium between elements of a time series vector as the existence of a linear combination of vector elements that is shown to be stationary. Cointegration overcame several limitations of classical univariate inference analysis in studying these processes by incorporating non-stationarity, long-term relations and short-run properties into the one modelling process (Engle and Yoo, 1987). The notion of equilibrium has formally specified the definition of a long-run equilibrium process. The importance of such a precise definition of long-run equilibrium is two-fold. Firstly, it may be exploited to test whether economic and finance theory actually holds. Secondly, cointegration has allowed the modelling of long-run behaviour especially in the area of derivative markets.

Before cointegration was formally specified, the usual practice in univariate model fitting procedures was to difference the data to achieve stationarity, resulting in a loss of information about long-run trends. The technique of cointegration thus allowed the simultaneous investigation into both the long-run relationships and short-run dynamic adjustments to such long-run relationships. Furthermore, Engle and Yoo (1987) and Diebold, Gardeazabal and Yilmaz (1994), amongst others, have shown that the imposition of cointegration (via the error correction model) leads to superior forecasting accuracy over the longer term.

The original Engle and Granger (1987) technique only allowed identification of a single cointegrating vector within a system. Johansen (1988) derived a procedure which overcame this limitation by being able to identify multiple linearly independent cointegrating vectors, if they exist, within a system. In the Johansen method the evolution of eligible time series is assumed to be well specified by a finite-order VAR model with white noise. Letting $\mathbf{X}_t$ be a $p \times 1$ vector of $I(1)$ variables (i.e. variables which become stationary after first differences);

$$\mathbf{X}_t = \Pi_1 \mathbf{X}_{t-1} + \Pi_2 \mathbf{X}_{t-2} + ... + \Pi_k \mathbf{X}_{t-k} + \epsilon_t, \quad t = 1, 2, ..., T, \tag{1}$$

where each of the $\Pi_k$ is a $(p \times p)$ matrix of parameters and $\epsilon_t \sim IN(0, \Sigma)$.

Equation (1) may be transformed into the following form;

$$\Delta \mathbf{X}_t = \Gamma_1 \Delta \mathbf{X}_{t-1} + ... + \Gamma_{k-1} \Delta \mathbf{X}_{t-k+1} + \Pi \mathbf{X}_{t-k} + \epsilon_t,$$

where $\Gamma_i = -(1-\Pi_1-...-\Pi_i)$, $i = 1, 2, ..., k-1$ and $\Pi = -(1-\Pi_1-...-\Pi_k)$.

The matrix $\Pi$ contains information about the long-run properties of the model. If $\Pi$ has rank 0, then the system is not cointegrated (all the variables in $\mathbf{X}_t$ are integrated of order one or higher). If $\Pi$ has rank $p$ (i.e. full rank), the variables in $\mathbf{X}_t$ are stationary. If $\Pi$ has rank $r$ (where $0 < r < p$), $\Pi$ may be decomposed into two distinct $(p \times r)$ matrices $\alpha$ and $\beta$ such that $\Pi = \alpha\beta^T$ (i.e. there are $r$ cointegrating vectors, which are given by $\beta$).

The restriction that the time series must be well specified by a finite-order VAR model may be particularly severe in the case of 'chaotic' financial time series which appear random but actually have some deterministic elements or when the order of the moving-average component is greater than zero and may not be adequately proxied by a finite-order pure autoregressive model (see examples in section three).

An alternative approach to identifying cointegrating vectors, under certain conditions, has been developed by Lin and McCrae (1999, 2001a). The procedure may be easily applied using widely-available standard statistical packages. This 'residual-based cointegration' (RBC) approach relies on the relationship between cointegrating vectors and residual processes obtained from the fitting of univariate Box-Jenkins ARIMA models. Lin and McCrae (1999, 2001a) show that, in theory, the number and identity of the linearly independent cointegrating vectors may be determined via the rank and content of the residual covariance matrix respectively. The technique is relatively simple to implement and is able to be executed using commonly available statistical packages.

Strictly speaking, the cointegration approaches of Lin and McCrae (1999, 2001a) and Johansen (1988) are not directly comparable since they are based on two different methodologies and each method is valid under different conditions. As mentioned previously, the Johansen (1988) technique models the series of interest by a VAR process and tests for cointegration by examining whether $\Pi$ is of full rank. The RBC procedure on the other hand may be applied when the variance/covariance matrix of residual processes (obtained when an appropriate univariate ARIMA model is applied to each time series) may be accepted as having reduced rank or asymptotically reduced rank.

The purpose of this paper is to test the practical implementation of the RBC procedure to both simulated and real-life data and is organised as follows; section two introduces the RBC approach to cointegration. The fundamental link between cointegrating vectors and residual processes obtained from the fitting of univariate ARIMA models, which motivates the RBC technique, is explained. This section also describes the steps in the practical application of the RBC approach. In section three, examples of autoregressive and moving-average series are analysed to show the applicability of the RBC procedure. In section four, the accuracy of the cointegration estimates is investigated for simulated series drawn from a model example in section three. Section five contains real-life applications of the RBC approach to foreign exchange rate data. The last section summarises the potential contribution of the RBC approach in financial time series analysis and suggests further avenues of research.

## 2. An Alternative Approach to Cointegration

A new procedure for determining cointegrating vectors, when the components of the system of interest satisfy certain conditions, was introduced by Lin and McCrae (1999, 2001a). This method allows the underlying time series to be fitted by ARIMA models other than finite-order VAR's only. Using this innovative procedure, the significance of the cointegration behaviours may be identified and estimated easily. Although the idea is similar to that of Engle and Yoo (1987) and Bierens (1997), the difference between this proposed procedure and that of both Engle and Yoo (1987) and Bierens (1997) is that the latter focus their attention on standard VAR models in which the covariance matrix has full rank. The novelty of the RBC approach is that the method involves an analysis of the relationship between multivariate cointegration and univariate ARIMA modelling where each individual time series is modelled independently. In such univariate modelling, investigation is limited to the analysis of the covariances between individual series. The covariance matrix of the residual processes given by the ARIMA models may not be of full rank, resulting in the corresponding time series being cointegrated.

The RBC approach utilises the theoretical relationship between the Engle-Granger and Johansen cointegration procedures and univariate ARIMA modelling fitting techniques for individual time series. The approach provides a method for examining;

- whether a relationship between cointegration and univariate ARIMA modelling exists, and
- whether the cointegrating vectors for the system may be determined by univariate ARIMA model fitting procedures, or more specifically, the covariance matrix of the residual processes.

Firstly, the RBC model will be introduced by reviewing the definitions of $I(d)$ time series and cointegrating vectors.

DEFINITION 1: A time series $X_t$, which has a stationary ARMA representation after differencing $d$ times but with the term $(1-B)^{d-1}X_t$ being non-stationary, is integrated of order $d$ and is denoted by $X_t \sim I(d)$, where $B$ is the back-shift operator.

DEFINITION 2: Let $\mathbf{X}_t = (X_{1,t}, \cdots, X_{p,t})^T$ be a (transposed) time series vector, $t = 1, 2, ..., n$. If each component of $\mathbf{X}_t$ is $I(1)$ and there exists a vector $\xi$ such that $\xi^T\mathbf{X}_t \sim I(0)$, $X_{1,t}, \cdots, X_{p,t}$ are said to be cointegrated and $\xi$ is called a cointegrating vector for the system $\mathbf{X}_t$.

Under certain weak conditions, Lin and McCrae (1999, 2001a) show that the number of cointegrating vectors may be determined via the rank of the covariance matrix of the residual processes. Given a system $\mathbf{X}_t = (X_{1t}, ..., X_{pt})^T$, assume that all of the elements of $\mathbf{X}_t$ are $I(1)$. If $\mathbf{X}_t$ may be accepted as a cointegrated system, the following procedure may be applied to determine all linear independent cointegrating vectors. The procedure consists of the following five steps.

(1) Fit $\mathbf{X}_t$ by an appropriate ARIMA model, say

$$\Phi(B)(1-B)\mathbf{X}_t = \mu + \Theta(B)\epsilon_t, \qquad t = 1, 2, ..., n$$

where $\epsilon_t = (\epsilon_{1t}, ..., \epsilon_{pt})^T$ are white noise, $\mu = (\mu_1, ..., \mu_p)^T$ and

$$\mathbf{\Phi}(B) = \begin{pmatrix} \Phi_1(B) & & 0 \\ & \ddots & \\ 0 & & \Phi_p(B) \end{pmatrix}, \mathbf{\Theta}(B) = \begin{pmatrix} \Theta_1(B) & & 0 \\ & \ddots & \\ 0 & & \Theta_p(B) \end{pmatrix}$$

where all $\Phi_i(B)$ and $\Theta_i(B)$ are finite-order polynomial functions of $B$ with roots outside the unit circle.

(2) If the residual vectors $\epsilon_t = (\epsilon_{1t}, ..., \epsilon_{pt})^T$ and residual cross-products $\epsilon_{it}\epsilon_{jt}$, $i < j \le p$, $t = 1, 2, ..., n$ are both stationary (or have ergodic properties) then the residual vectors may be analysed to obtain the sample covariance matrix for $\epsilon_t$. The sample covariance matrix is denoted by $\hat{\Sigma}_n$, and is used to estimate var$(\epsilon_t)$.

(3) Determine the eigenvalues of $\hat{\Sigma}_n$, say $\lambda_1 \ge \lambda_2 \ge ... \ge \lambda_{p-r} > \lambda_{p-r+1} \ge ... \ge \lambda_p$, and corresponding eigenvectors. The eigenvectors form a matrix denoted by $\mathbf{A}$. The matrix $\mathbf{A}$ may be re-written as $\mathbf{A} = (\mathbf{A_1}, \mathbf{A_2})$, where $\mathbf{A_2}$ is formed by those eigenvectors corresponding to the smaller eigenvalues $\lambda_{p-r+1} \ge \lambda_{p-r+2} \ge ... \ge \lambda_p$.

(4) Let $v_t^{(1)} = (v_{1,t}, ..., v_{p-r,t})^T$ and $v_t^{(2)} = (v_{p-r+1,t}, ..., v_{p,t})^T$ satisfy the following equation

$$\Phi(B)(1-B)\mathbf{X}_t = \mu + \Theta(B)\epsilon_t, \qquad t = 1, 2, ..., n$$

where

$$\epsilon_t = (\mathbf{A_1} \;\; \mathbf{A_2}) \begin{pmatrix} v_t^{(1)} \\ v_t^{(2)} \end{pmatrix}.$$

Now $\mathbf{X}_t$ may be expressed in the following form;

$$(1-B)\mathbf{X}_t = \tilde{\mu} + C(B)v_t^{(1)} + C_1(B)v_t^{(2)}$$

with $C(B) = \mathbf{\Phi}(B)^{-1}\mathbf{\Theta}(B)\mathbf{A_1}$, $C_1(B) = \mathbf{\Phi}(B)^{-1}\mathbf{\Theta}(B)\mathbf{A_2}$ and $\tilde{\mu} = \mathbf{\Phi}(B)^{-1}\mu$.

Therefore,

$$\begin{aligned}\mathbf{X}_t &= \tilde{\mu}t + \mathbf{X_0} + (1-B)\frac{C(B)-C(1)}{1-B}\Sigma_{i=1}^{t} v_t^{(1)} \\ &+ C(1)\Sigma_{i=1}^{t} v_t^{(1)} + C_1(B)\Sigma_{i=1}^{t} v_t^{(2)},\end{aligned}$$

where $C(1) = C(B)$ with $B = 1$. Upon solving for $\xi^T C(1) = 0$, $\xi$ is obtained such that $\xi^T\mathbf{X}_t = \xi^T\mathbf{X_0} - \xi^T\mathbf{W_0} + \xi^T\mathbf{W_t} + \xi^T C_1(B)\Sigma_{i=1}^{t} v_i^{(2)}$, where $\mathbf{W}_t = \frac{C(B)-C(1)}{1-B}v_t^{(1)}$ and the term $\xi^T\mathbf{X_0} - \xi^T\mathbf{W_0} + \xi^T\mathbf{W_t}$ is $I(0)$ (Lin and McCrae, 1999).

Therefore, $\xi$ may be accepted as a cointegrating vector for $X_t - \tilde{\mu}t$ if the impact of the non-stationary component $\xi^T C_1(B)\Sigma_{i=1}^{t} v_i^{(1)}$ is not significant. An important issue is how to determine whether or not the non-stationary component is in fact significant. The size of the impact of this non-stationary component may be measured by the ratio of variances given by the non-stationary and stationary components in the manner of Lin and McCrae (2001b).

(5) Verify that the linear combination of the cointegrating vector $\xi^T(\mathbf{X}_t - \tilde{\mu}t)$ is indeed stationary via the Augmented Dickey-Fuller (ADF) test as well as by graphical procedures.

## 3. Application of the RBC Procedure to Simulated Data

This section will outline how the method of RBC may be applied in practice via the use of simulated data. In theory this alternative approach is theoretically valid but a sub-issue is whether this translates into practical significance. There are two stages in the application of the procedure. By using such simulated data one may test whether the RBC theory shows improvement in the experimental stage. If this procedure fails then there is a strong indication that the theoretical results will not translate into practical significance. If the results do show promise then the next logical step would be to implement the method to real-life data. Simulated data is considered first due to the fact that real-life data may be "dirty" and thus may be susceptible to conditioning factors (i.e. other influences on data that may complicate the data modelling). Such external factors may be controlled by the use of simulated data, as well as providing knowledge of the true cointegrating relationship present. Furthermore, if the RBC theory translates well to simulated data, then providing external factors are minimised in real-life data the method will be equally valid for this type of data.

Three examples using simulated data will be considered. The first example is considered because it is generated by two autoregressive processes and shows the possible inaccuracy of the resulting estimates if each and every series is not fitted properly. The second example is included because it results in the failure of the Johansen (1988) technique whilst the third example is considered as three variables are included in the system.

The true theoretical value of the cointegrating vector is firstly derived and then the estimates of each of the cointegrating vectors are compared with the true values. This is done for one simulation from each of the three examples.

Example 1: Consider two time series, $X_{1,t}$ and $X_{2,t}$, which are generated from the following models:

$$\begin{aligned} X_{1,t} &= 1.4X_{1,t-1} - 0.2X_{1,t-2} - 0.2X_{1,t-3} + \epsilon_t, \\ X_{2,t} &= 1.2X_{2,t-1} - 0.2X_{2,t-2} + \epsilon_t, \end{aligned}$$

where $\epsilon_t$ is white noise with mean 0 and variance 1. It may be seen that after differencing both series $X_{1,t}$ and $X_{2,t}$ are generated by an autoregressive model.

The time series may be written as follows;

$$(1-B)\begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix} = \begin{pmatrix} \frac{1}{1-0.4B-0.2B^2} & 0 \\ 0 & \frac{1}{1-0.2B} \end{pmatrix} \begin{pmatrix} \epsilon_t \\ \epsilon_t \end{pmatrix}.$$

The true variance/covariance matrix of $(\epsilon_{1t}, \epsilon_{2t})^T$ is given by $\Sigma_{2\times 2}$ with all entries equal to 1. The matrix $\Sigma$ has resulting eigenvalues of 2 and 0. Since one eigenvalue is equal to zero, following the discussion in Lin and McCrae (1999, 2001a), $X_{1,t}$ and $X_{2,t}$ are cointegrated. The true cointegrating vector of the system $X_{1,t}$ and $X_{2,t}$ is given by $(1, -2)^T$ (for details on how to determine the true cointegrating vector see Lin and McCrae, 1999).

In the following, it is of interest to examine whether, given a sample $X_{1,t}$ and $X_{2,t}$, the RBC procedure may be implemented to estimate the cointegrating vector. A sample $(X_{1,t}, X_{2,t})$ of size 1000 are simulated from the models in Example 1. The RBC and Johansen procedures are then applied to the samples. The first step in the RBC procedure is to fit univariate ARIMA models to each of the time series $X_{1,t}$ and $X_{2,t}$, both of which are $I(1)$. The appropriate models to be fitted, using Box-Jenkins methods, are ARIMA(2,1,0) and ARIMA(1,1,0) respectively.[1] The true value of the coefficients of the autoregressive parameters of $X_{1,t}$ in Example 1 are 0.4 and 0.2 respectively, while the true value of the coefficient of the autoregressive parameter of $X_{2,t}$ is 0.2. Furthermore, there are no moving-average parameters. The estimated coefficients of the autoregressive parameters of $X_{1,t}$ are 0.3442 and 0.2354. Likewise, the estimated coefficient for the autoregressive parameter of $X_{2,t}$ is 0.1658.

The second step is to construct the sample covariance matrix $\hat{\Sigma}_n$, given that the residual vectors $\epsilon_t = (\epsilon_{1t}, ..., \epsilon_{pt})^T$ and residual cross-products $\epsilon_{it}\epsilon_{jt}$, $i < j \leq p$, $t = 1, 2, ..., n$ are both stationary. After fitting the ARIMA models to $X_{1,t}$ and $X_{2,t}$, the residual vectors $\epsilon_t = (\epsilon_{1t}, ..., \epsilon_{pt})^T$ and residual cross-products $\epsilon_{it}\epsilon_{jt}$, $i < j \leq p$, $t = 1, 2, ..., n$ are both stationary. Therefore, $\hat{\Sigma}_n$ may be used to estimate the true covariance matrix $\Sigma$ and is given by

$$\hat{\Sigma}_n = \begin{pmatrix} 1.073226 & 1.073925 \\ 1.073925 & 1.075640 \end{pmatrix}.$$

The third step is to determine the eigenvalues of $\hat{\Sigma}_n$. These estimated eigenvalues are 2.148360 and 0.000508 respectively. Since one estimated eigenvalue is approximately equal to zero and the ratio of it to the sum of all eigenvalues is negligible (0.000236), the RBC procedure may be implemented to estimate the cointegrating vector. The resulting eigenvectors form the matrix $\mathbf{A}$ which is given by

$$\mathbf{A} = \begin{pmatrix} 0.706710 & 0.707504 \\ 0.707504 & -0.706710 \end{pmatrix}.$$

Furthermore, $\mathbf{A}$ may be decomposed into $(\mathbf{A_1}, \mathbf{A_2})$, where $\mathbf{A_1}$ is formed by those eigenvectors corresponding to the largest eigenvalue $\lambda_1$ (=2.148360), i.e.

$$\mathbf{A_1} = \begin{pmatrix} 0.706710 \\ 0.707504 \end{pmatrix}.$$

The fourth step involves the estimation of the cointegrating vector via the solving of the equation $\xi^T C(1) = 0$. As mentioned previously, $C(1)$ is given by $C(B)$ with $B = 1$. In this example $C(1)$ is equal to

$$C(1) = \begin{pmatrix} \frac{1}{0.4204} & 0 \\ 0 & \frac{1}{0.8342} \end{pmatrix} \begin{pmatrix} 0.706710 \\ 0.707504 \end{pmatrix}.$$

The true cointegrating vector may be seen to be theoretically equal to (1, -2)$^T$. The estimated cointegrating vector in this simulated example is $\hat{\xi}$=(1, -1.981889)$^T$, a very accurate estimate notwithstanding the fact that $X_{1,t}$ has estimated coefficients which are not very close to the true coefficients.

The fifth and last step in the procedure involves the testing for stationarity of the linear combination of the cointegrating vector. The ADF test shows that the linear combination is stationary as does the time series plot of the linear combination. The Johansen procedure yields an estimate of $\hat{\xi}$=(1, -2.0003)$^T$ in this example.

It is worth noting that if inappropriate ARIMA models are used to fit the time series an inaccurate estimate of the cointegrating vector will, in all likelihood, be obtained. This may be seen in Table 1 where the most accurate estimate of the cointegrating vector is obtained when both time series are fitted correctly.

*Table 1.* Estimates of the Cointegrating Vector for Example 1

| $X_{1,t}$ | $X_{2,t}$ | Estimate of Cointegrating Vector |
|---|---|---|
| ARIMA(1,1,0) | ARIMA(1,1,0) | (1, -1.558540) |
| ARIMA(2,1,0) | ARIMA(1,1,0) | (1, -1.981889) |
| ARIMA(1,1,0) | ARIMA(2,1,0) | (1, -1.507315) |
| ARIMA(2,1,0) | ARIMA(2,1,0) | (1, -1.915948) |

Example 2: Consider two time series, $X_{1,t}$ and $X_{2,t}$, which are generated from the following models:

$$\begin{aligned} X_{1,t} &= X_{1,t-1} + \epsilon_t - 0.2\epsilon_{t-1}, \\ X_{2,t} &= X_{2,t-1} + \sqrt{2}\epsilon_t + 0.2\epsilon_{t-1}, \end{aligned}$$

where $\epsilon_t$ is white noise with mean 0 and variance 1. It may be seen that after differencing both series $X_{1,t}$ and $X_{2,t}$ are generated by a moving-average model.

The time series may be written as follows;

$$(1-B)\begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix} = \begin{pmatrix} 1-0.2B & 0 \\ 0 & \sqrt{2}+0.2B \end{pmatrix} \begin{pmatrix} \epsilon_t \\ \epsilon_t \end{pmatrix}.$$

The true variance/covariance matrix of $(\epsilon_{1t}, \epsilon_{2t})^T$ is given by $\Sigma_{2\times 2}$ with all entries equal to 1. The matrix $\Sigma$ has resulting eigenvalues of 2 and 0.

Since one eigenvalue is equal to zero, $X_{1,t}$ and $X_{2,t}$ are cointegrated (Lin and McCrae, 1999, 2001a). The true cointegrating vector is given by (1, -0.4956)$^T$.

Data is simulated from the models in Example 2 and the RBC and Johansen procedures are applied to the simulated data, the sample size being 1000. In this instance, the appropriate ARIMA model to be fitted to both the time series $X_{1,t}$ and $X_{2,t}$ is ARIMA(0,1,1). From the fitting of this model to $X_{1,t}$ and $X_{2,t}$ the residual vectors $\epsilon_t = (\epsilon_{1t}, ..., \epsilon_{pt})^T$ and residual cross-products $\epsilon_{it}\epsilon_{jt}$, $i < j \leq p$, $t = 1, 2, ..., n$ are both stationary. The sample covariance matrix is given by

$$\hat{\Sigma}_n = \begin{pmatrix} 1.074328 & 1.518333 \\ 1.518333 & 2.146540 \end{pmatrix}.$$

The eigenvalues of $\hat{\Sigma}_n$ are 3.220634 and 0.000234 respectively. The second estimated eigenvalue is negligible (and the ratio of it to the sum of all of the eigenvalues is also negligible) and so an estimate of the cointegrating vector may be obtained by the RBC procedure. The resulting eigenvectors form the matrix $\mathbf{A}$, where

$$\mathbf{A} = \begin{pmatrix} 0.577519 & 0.816377 \\ 0.816377 & -0.577519 \end{pmatrix},$$

the eigenvector $(0.577519, 0.816377)^T$ corresponding to the largest eigenvalue, 3.220634.

The estimated cointegrating vector in this simulated example is $\hat{\xi}$=(1, -0.494087)$^T$, a very accurate estimate noting that the true cointegrating vector has been shown to be theoretically equal to (1, -0.4956)$^T$. The graph of the linear combination of the cointegrating vector is stationary. The ADF test also reveals stationarity of the linear combination. The traditional cointegration approach of Johansen implemented via use of the PcFIML package yields invalid estimates in this instance.[2] The invalid estimates arise from the singularity of the $\Sigma$ matrix (Johansen, 1988).[3]

Example 3: Consider three time series, $X_{1,t}$, $X_{2,t}$ and $X_{3,t}$, which are generated from the following models:

$$\begin{aligned} X_{1,t} &= 1.4X_{1,t-1} - 0.2X_{1,t-2} - 0.2X_{1,t-3} + \epsilon_{1,t}, \\ X_{2,t} &= 1.3X_{2,t-1} - 0.3X_{2,t-2} + \epsilon_{1,t} + \epsilon_{2,t}, \\ X_{3,t} &= 1.6X_{3,t-1} - 0.6X_{3,t-2} + \epsilon_{2,t} - 0.8\epsilon_{2,t-1}, \end{aligned}$$

where $\epsilon_{1,t}$ is white noise with mean 0 and variance 0.64 and $\epsilon_{2,t}$ is white noise with mean 0 and variance 1. It may be seen that after differencing the series $X_{1,t}$ and $X_{2,t}$ are generated by an autoregressive model and the series $X_{3,t}$ contains both autoregressive and moving-average components.

The time series may be written as follows;

$$(1-B)\begin{pmatrix} X_{1,t} \\ X_{2,t} \\ X_{3,t} \end{pmatrix} = \begin{pmatrix} \frac{1}{1-0.4B-0.2B^2} & 0 & 0 \\ 0 & \frac{1}{1-0.3B} & 0 \\ 0 & 0 & \frac{1-0.8B}{1-0.6B} \end{pmatrix} \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{1,t}+\epsilon_{2,t} \\ \epsilon_{2,t}\epsilon_{2,t-1} \end{pmatrix}.$$

The true variance/covariance matrix is given by

$$\Sigma = \begin{pmatrix} 0.64 & 0.64 & 0 \\ 0.64 & 1.64 & 1 \\ 0 & 1 & 1 \end{pmatrix},$$

which has resulting eigenvalues of 2.517268, 0.762732 and 0. Since one eigenvalue is equal to zero, following the discussion in Lin and McCrae (1999, 2001a), $X_{1,t}$, $X_{2,t}$ and $X_{3,t}$ are cointegrated. The true cointegrating vector of the system $X_{1,t}$, $X_{2,t}$ and $X_{3,t}$ is given by (1, -1.75, 5)$^T$.

Data is simulated from the models in Example 3 and the RBC and Johansen procedures are applied to the simulated data, the sample size being 1000. In this instance, the appropriate ARIMA model to be fitted to $X_1$, $X_2$ and $X_3$ are ARIMA(2,1,0), ARIMA(1,1,0) and ARIMA(1,1,1) respectively. After fitting the ARIMA models to $X_{1,t}$, $X_{2,t}$ and $X_{3,t}$, the residual vectors $\epsilon_t = (\epsilon_{1t}, ..., \epsilon_{pt})^T$ and residual cross-products $\epsilon_{it}\epsilon_{jt}$, $i < j \leq p$, $t = 1, 2, ..., n$ are stationary. The sample covariance matrix may then be estimated and is given by

$$\hat{\Sigma}_n = \begin{pmatrix} 0.686866 & 0.691890 & 0.006158 \\ 0.691890 & 1.647333 & 0.951617 \\ 0.006158 & 0.951617 & 0.945766 \end{pmatrix}.$$

The eigenvalues of $\hat{\Sigma}_n$ are 2.497964, 0.780245 and 0.001755 respectively. The last estimated eigenvalue is negligible (and the ratio of it to the sum of all of the eigenvalues is also negligible) and so an estimate of the cointegrating vector may be obtained by the RBC procedure. The resulting eigenvectors form the matrix **A**, where

$$\mathbf{A} = \begin{pmatrix} 0.311021 & -0.754901 & 0.577399 \\ 0.809702 & -0.107642 & -0.576885 \\ 0.497644 & 0.646944 & 0.577766 \end{pmatrix},$$

the eigenvector (0.311021, 0.809702, 0.497644)$^T$ corresponds to the largest eigenvalue (2.497964) and the eigenvector (-0.754901, -0.107642, 0.646944)$^T$ corresponds to the second largest eigenvalue, 0.780245. The first two columns of **A** are used to compose $C(1)$.

The estimated cointegrating vector in this simulated example is $\hat{\xi}$=(1, -1.647948, 4.863616)$^T$. The graph of the linear combination of the cointegrating vector is stationary. The ADF test also reveals stationarity of the linear combination. The traditional cointegration approach of Johansen implemented via use of the PcFIML package yields an estimate of the cointegrating vector of $\hat{\xi}$= (1, -1.7664, 5.0294)$^T$. The resultant linear combination of cointegrating vector is also seen to be stationary in this instance (both by graphical procedures and the ADF test).

## 4. Simulated Examples

In this section ten thousand independent simulations of the model in Example 1 are carried out to investigate whether the cointegrating vector estimates from the RBC approach vary significantly between simulations for a given model. The RBC procedure described previously is then implemented to each simulated data set to obtain an estimate of the cointegrating vector. It is to be noted that the first component of the cointegrating vector $\xi$ (i.e. $\xi_1$) in all cases is fixed (equal to one) and only the second component of $\xi$ (i.e. $\xi_2$) is estimated.

According to the procedure outlined in section two, the negligibility of the smallest estimated eigenvalue is of interest. Therefore, for each simulation, the proportion contribution of the smallest eigenvalue (to the sum of eigenvalues) is calculated (see Examples 1-3 above). The estimated eigenvalues are not identical between replications due to the residual covariances between $X_{1,t}$ and $X_{2,t}$ not being equal.

Consider ten thousand independent samples, $(X_{1,t}, X_{2,t})$, which are generated from the following ARIMA models described in Example 1. The correlation between the magnitude of the contribution of the smallest eigenvalue and the most accurate estimates may be evidenced by the cross-plot in Figure 1. It may be seen that the smaller the proportion contribution of the smallest eigenvalue, the more accurate the resulting estimate of the cointegrating vector. It appears as though the resulting estimates of the cointegrating vector are symmetrically dispersed around the central true value (-2). Figure 1 appears to be bimodal in the sense that, when the contribution of the smallest eigenvalue (relative to the sum of all eigenvalues) is relatively large, two possible events might occur; either the resulting estimate of the cointegrating vector might over-estimate or the estimate may under-estimate the true cointegrating vector.

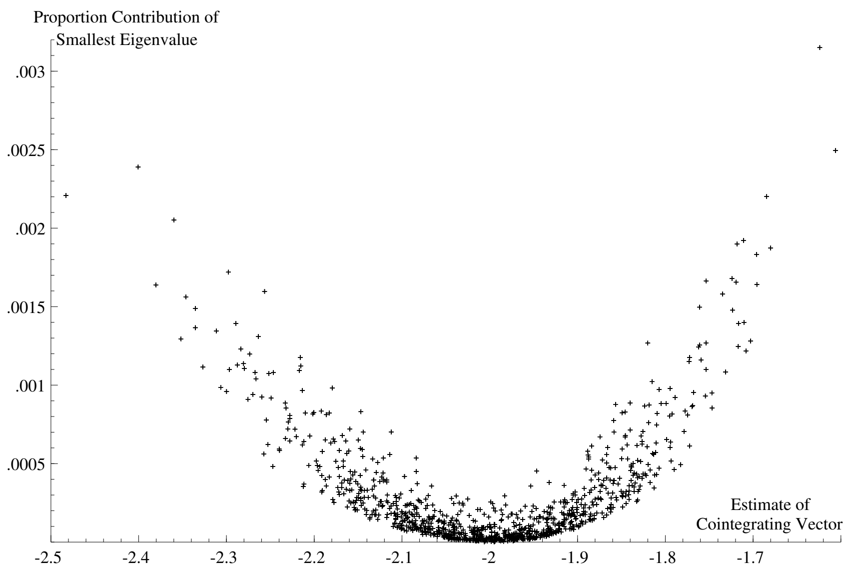


*Figure 1.* Cross-plot of the estimate of the cointegrating vector via the RBC procedure and the proportion contribution of the smallest eigenvalue for 10000 simulations of Example 1.

## 5. Application of the Proposed Method to Financial Data

This section will apply the RBC procedure to real-life data. The purpose is to show whether, in practical situations, the method of RBC may be invoked as an alternative to the Johansen procedure. The aim is to see whether the improvement in the alternative technique that was gained with "clear" series (i.e. generated series) may be translated to real-life data. The examples both apply to foreign exchange rate data, the first example tests for cointegration between spot and forward prices. Derivatives such as forward contracts are commonly traded by hedgers who transfer the associated price risk to others, thus lessening their own exposure to such risk. Hedging exchanges spot risk, which is defined as the chance of movement in prices in the underlying instrument, for basis risk (the chance that forward prices may move out of line with spot prices over time). However, basis risk is usually much less than spot risk because, in efficient markets, there is a theoretical relationship between spot and forward prices which means they should not move too far apart from one another. Therefore, one would expect forward and spot prices to be cointegrated in the long-run. The second example tests for cointegration between three (spot) exchange rate series in the Asian region.

The first example relates to an analysis of potential long-run equilibrium between daily spot and one week forward rates for the US dollar (relative to the UK pound) exchange rate. The period examined is between October 27, 1997 and December 22, 1999 inclusive, which provides 563 observations. Denoting $X_{1,t}$ to be the natural logarithm of the spot rates of the US Dollar (with respect to the British Pound) and $X_{2,t}$ as the natural logarithm of the one week forward rates of the US Dollar (with respect to the British Pound), the ARIMA(3,1,0) or ARIMA(0,1,3) time series models seem applicable models to fit $X_{1,t}$ and $X_{2,t}$. The fitting of these ARIMA models to $X_{1,t}$ and $X_{2,t}$ result in the residual vectors $\epsilon_t = (\epsilon_{1t}, ..., \epsilon_{pt})^T$ and residual cross-products $\epsilon_{it}\epsilon_{jt}$, $i < j \leq p$, $t = 1, 2, ..., n$ both being stationary. Therefore, the residual vectors may be exploited to obtain the sample covariance matrix for $\epsilon_t$. All four combinations of models are fitted to the series $X_{1,t}$ and $X_{2,t}$ to determine which combination provides the most accurate estimate of the cointegrating vector. The determination of the most accurate estimate of the cointegrating vector is provided via both an ADF test and a time series plot on the resulting linear combination of cointegrating vector. Both series are non-stationary in levels and stationary in first differences. The time series graphs of $X_{1,t}$ and $X_{2,t}$ are shown in Figure 2. The estimates of the cointegrating vector for each combination are shown in Table 2.[4]

*Table 2.* Estimates of the Cointegrating Vector for the spot and forward data

| $X_{1,t}$ | $X_{2,t}$ | Estimate of Cointegrating Vector |
|---|---|---|
| ARIMA(3,1,0) | ARIMA(3,1,0) | (1, -0.999145) |
| ARIMA(0,1,3) | ARIMA(0,1,3) | (1, -0.999345) |
| ARIMA(3,1,0) | ARIMA(0,1,3) | (1, -1.016763) |
| ARIMA(0,1,3) | ARIMA(3,1,0) | (1, -0.981885) |

The third estimate is the most accurate and the graph of the linear combination of cointegrating vector is shown in Figure 3. Via the ADF test, the linear combination of cointegrating vector for the third potential cointegrating vector is the only one which reveals stationarity, the other linear combinations are not stationary. Therefore, the estimate of the cointegrat-

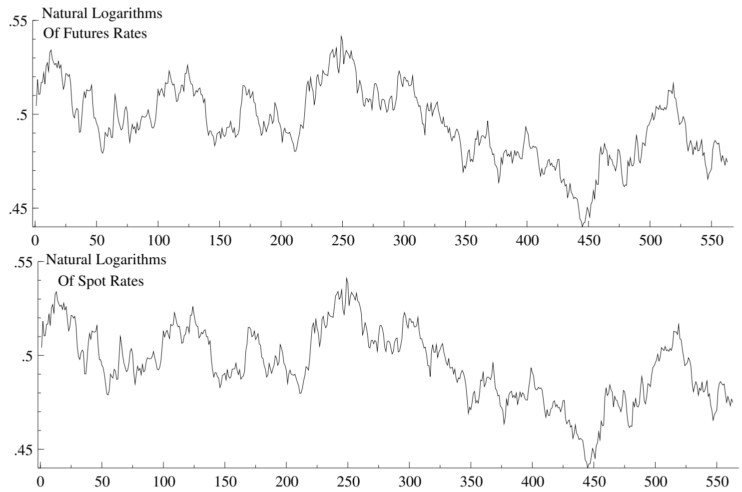


*Figure 2.* Time series plots of the natural logarithms of the forward and spot exchange rates respectively (of the US Dollar with respect to the British Pound).

ing vector is $(1, \text{-}1.016763)^T$. The cointegrating vector when the method of Johansen is applied is $(1, \text{-}1.0118)^T$. The PcFIML output of the cointegrating analysis is reported in Table 3 whilst the graph of the linear combination of the cointegrating vector via this procedure is shown in Figure 3. The ADF test for stationarity of the linear combination of the cointegrating vector obtained via the RBC procedure reveals significance at the 1% level of significance, thus inferring stationarity. Analysis on the linear combination of cointegrating vector via the Johansen procedure also reveals stationarity at the 1% level of significance.

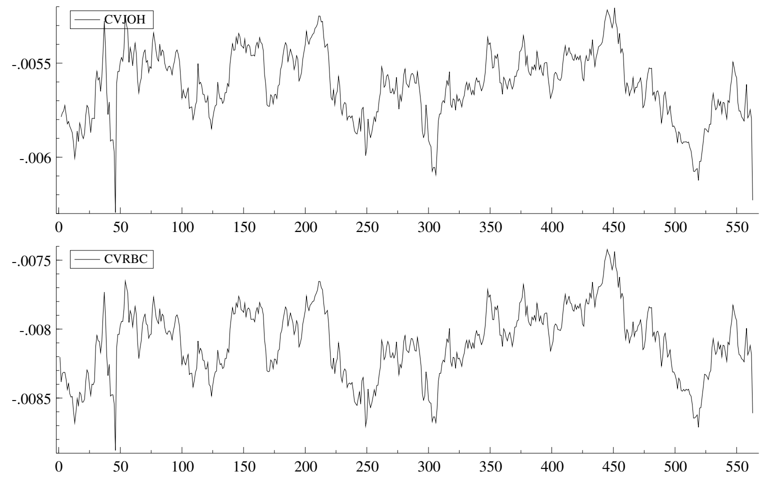


*Figure 3.* Time series plots of the linear combinations of the cointegrating vector via the Johansen Method and RBC procedure respectively for the spot and forward data.

*Table 3.* PcFIML output of cointegration analysis via the Johansen procedure for the spot and forward exchange rate example

```
Cointegration analysis 2 to 563

Ho:rank=p  -Tlog(1-\mu) using T-nm  95% -T\Sum log(.) using T-nm 95%
 p ==  0      15.92*     15.87*    14.1    16.74*     16.68*     15.4
 p <=  1      0.8205     0.8175     3.8    0.8205     0.8175      3.8

standardized \beta' eigenvectors
      LSSpot    LSForward
       1.0000     -1.0118
      -1.0025      1.0000
```

The method of RBC is now applied to foreign exchange rates in the Asian region. In this example there are three series of (spot) exchange rates. The first is the Malaysian Ringgit, the second is the Philippines Peso and the third series is the Thai Baht. All series are expressed in terms of the US Dollar. The time period examined commences on January 1, 1985 and terminates on December 30, 1994 which provides ten years data and 2610 observations for each series. The natural logarithms of the exchange rates are once again analysed and the time series graphs of each of the three variables are shown in Figure 4. The most appropriate univariate Box-Jenkins ARIMA model for the Malaysian Ringgit is the ARIMA(4,1,0) model. The most appropriate Box-Jenkins ARIMA time series models for the Philippine Peso and the Thai Baht are ARIMA(1,1,2) and ARIMA(0,1,4) respectively. The resulting residual vectors $\epsilon_t = (\epsilon_{1t}, ..., \epsilon_{pt})^T$ and residual cross-products $\epsilon_{it}\epsilon_{jt}$, $i < j \leq p$, $t = 1, 2, ..., n$ are both stationary and the residual vectors may therefore be exploited to obtain the sample covariance matrix for $\epsilon_t$.

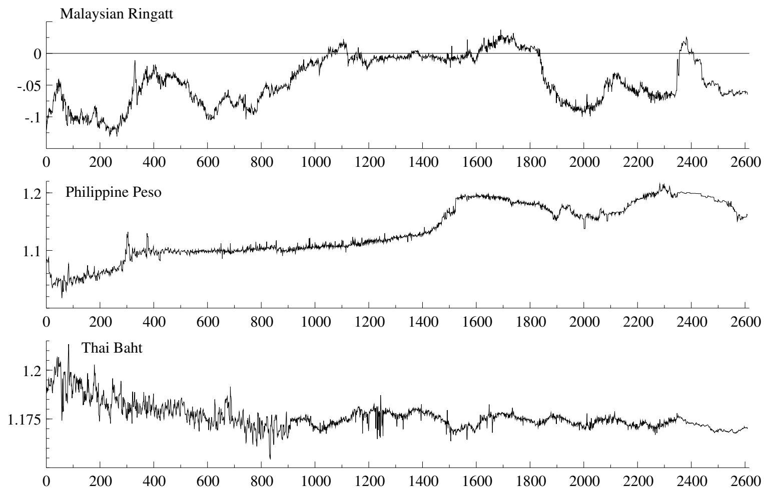


*Figure 4.* Time series plots of the natural logarithms of the Malaysian Ringgit, Philippine Peso and Thai Baht exchange rates respectively.

The resultant eigenvalues obtained from the fitting of the three ARIMA models specified above are 0.00002498, 0.00001171, 0.00000364. From these eigenvalues it may be expected that there are either zero or one cointegrating vectors since the second smallest eigenvalue appears to be much larger than the smallest eigenvalue whilst not being much smaller than the largest eigenvalue. When the technique of RBC is applied the resulting estimate of the cointegrating vector is $(1, 3.964026, -23.342658)^T$. This linear cointegrating vector is found to be stationary using the ADF test at the 1% level of significance. The same conclusion may be reached with the Johansen technique, where the estimate of the cointegrating vector is given by $(1, 28.229, 298.19)^T$. The linear combination of cointegrating vector obtained via both the RBC and Johansen techniques are shown in Figure 5. By noting that the cointegrating vector given by the Johansen method assigns heavy weight on the Thai Baht, this currency plays a major role in the system as defined by the Johansen method and the impact given by the Malaysian Ringgit and Philippine Peso on the system becomes insignificant. From this point of view, the cointegration system determined by RBC seems more appropriate.

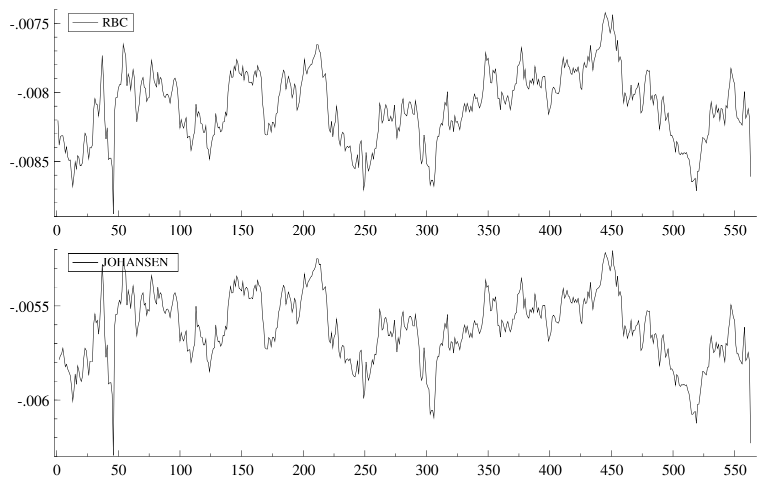


*Figure 5.* Time series plots of the linear combinations of the RBC and Johansen cointegrating vectors respectively for the Asian exchange rate data.

## 6. Discussion

This paper has shown the applicability of the procedure of RBC through both simulated and real-life examples. There is still an outstanding issue which may be examined in future study. This issue involves the criteria for determining whether the smallest eigenvalue provides negligible contribution with respect to the larger eigenvalue(s). Lin and McCrae (2001b) have shown that it is possible to provide a relative comparison of the variances of the non-stationary and stationary components respectively and whether or not the non-stationary component is negligible determines whether the system is cointegrated or not. It is seen in the simulations performed in this paper that the more negligible the contribution of the smallest eigenvalue, the more likely the estimate of the cointegrating vector is closer to the true theoretical value. Given a multivariate time series and estimated eigenvalues, how could one be able to tell if in fact the smaller estimated eigenvalue is small enough with respect to the larger estimated eigenvalue(s), so that cointegration is a distinct possibility? A simple plot of the time series of the linear combination of the cointegrating vector(s) may well not be enough.

By taking into account the relationship between cointegration and univariate ARIMA models, the number of cointegrating vectors may be determined via the rank of the covariance matrix of the residual processes of univariate ARIMA models. Furthermore, the more negligible the contribution of the smallest estimated eigenvalue, the more accurate the resulting estimates of the cointegrating vector. As seen in this paper, the RBC approach may, in certain instances, be an alternative to the Johansen procedure. This appears to be true when the underlying time series may not be modelled appropriately by a finite order autoregression model but rather a finite order moving-average model.

## Notes

1. Obviously these are the two correct ARIMA models for $X_{1,t}$ and $X_{2,t}$ from the generation of the two particular time series.
2. PcFIML is an econometric package for determining cointegrating vectors via the Johansen procedure.
3. The resulting output is made redundant (in actual fact, the output does not reveal cointegration when clearly there is).
4. The estimates in Table 2 arise from both time series being fitted using ARIMA models without the constant term.

## References

1. Bierens, H.J., 1997, Nonparametric cointegration analysis, Journal of Econometrics 77, 379-404.
2. Copland, L., 1991, Cointegration with daily exchange rate data, Oxford Bulletin of Economics 53, 185-199.
3. Crowder, W., 1996, A note on cointegration and international capital market efficiency, Journal of International Money and Finance 15, 661-664.
4. Diebold, F.X., J. Gardeazabal and K. Yilmaz, 1994, On cointegration and exchange rate dynamics, Journal of Finance 49, 253-263.
5. Engle, R.F. and C.W. Granger, 1987, Cointegration and error correction: representation, estimation and testing, Econometrica 35, 251-276.
6. Engle, R.F. and B.S. Yoo, 1987, Forecasting and testing in cointegrated systems, Journal of Econometrics 35, 143-159.
7. Granger, C.W.J., 1981, Some properties of time series data and their use on econometric model specification, Journal of Econometrics 16, 121-130.

8. Johansen, S., 1988, Statistical analysis of cointegration vectors, Journal of Economic Dynamics and Control 12, 231-254.
9. Johansen, S. and K. Juselius, 1990, Maximum likelihood estimation and inference on cointegration - with applications to the demand for money, Oxford Bulletin of Economics and Statistics 52, 169-210.
10. Layton, A.P. and A. Tan, 1992, Multivariate cointegration testing of the efficiency of Australia's spot forex market, Accounting and Finance 32, 63-70.
11. Lin, Y-X. and M. McCrae, 1999, A new approach to determining cointegrating vectors, Proceedings of the International Congress on Modelling and Simulation 2, 265-270.
12. Lin, Y-X. and M. McCrae, 2001a, Estimation of cointegrating vectors based on residuals of ARIMA models, System Analysis Modelling Simulation (to appear).
13. Lin, Y-X. and M. McCrae, 2001b, The effects of $I(1)$ series on cointegration inference (working paper).