

BANKWEST CURTIN ECONOMICS CENTRE

WORKING PAPER SERIES

13/6: ADVANTAGES OF NON-NORMALITY IN
TESTING COINTEGRATION RANK

Felix Chan

This report was written by researchers affiliated with the Bankwest Curtin Economics Centre ('the Centre'). While every effort has been made to ensure the accuracy of this document, the uncertain nature of economic data, forecasting and analysis means that the Centre, Curtin University and/or Bankwest are unable to make any warranties in relation to the information contained herein. Any person who relies on the information contained in this document does so at their own risk. The Centre, Curtin University, Bankwest, and/or their employees and agents disclaim liability for any loss or damage, which may arise as a consequence of any person relying on the information contained in this document. Except where liability under any statute cannot be excluded, the Centre, Curtin University, Bankwest and/or their advisors, employees and officers do not accept any liability (whether under contract, tort or otherwise) for any resulting loss or damage suffered by the reader or by any other person.

The views in this publication are those of the authors and do not represent the views of Curtin University and/or Bankwest or any of their affiliates. This publication is provided as general information only and does not consider anyone's specific objectives, situation or needs. Neither the authors nor the Centre accept any duty of care or liability to anyone regarding this publication or any loss suffered in connection with the use of this publication or any of its content.

Authorised Use

© Bankwest Curtin Economics Centre, July 2013

Bankwest Curtin Economics Centre Working Paper Series

ISSN: 2202-2791

ISBN: 978-1-925083-05-7

Felix Chan¹

¹ School of Economics and Finance, Curtin Business School, Curtin University, Australia

This publication contains confidential and propriety information of the Bankwest Curtin Economics Centre. All of the material in this publication is for your exclusive use and may not be otherwise used or modified for or by any other person or sold to or otherwise provided in whole or in part to any other person or entity without the prior written consent of the Bankwest Curtin Economics Centre.

For subscribers' use only. NOT TO BE PHOTOCOPIED.

A standard hard copy or electronic subscription to this publication entitles employees of the same organisation and same physical location as the subscriber to the use of its contents for internal reporting purposes only. Multiple user licenses are available for organisations with more than one location.

Advantages of Non-normality in Testing Cointegration Rank

Felix Chan ¹
School of Economics and Finance
Curtin University

First Draft: February 2013

Abstract: Since the seminal work of Engle and Granger (1987) and Johansen (1988), testing for cointegration has become standard practice in analysing economic and financial time series data. Many of the techniques in cointegration analysis require the assumption of normality, which may not always hold. Although there is evidence that these techniques are robust to non-normality, most existing techniques do not seek additional information from non-normality. This is important in at least two cases. Firstly, the number of observations is typically small for macroeconomic time series data, the fact that the underlying distribution may not be normal provides important information that can potentially be useful in testing for cointegrating relationships. Secondly, high frequency financial time series data often shows evidence of non-normal random variables with time-varying second moments and it is unclear how these characteristics affect the standard test of cointegration, such as Johansen's trace and max tests.

This paper proposes a new framework derived from Independent Component Analysis (ICA) to test for cointegration. The framework explicitly exploits processes with non-normal distributions and their independence. Monte Carlo simulation shows that the new test is comparable to the Johansen's trace and max tests when the number of observations is large and has a slight advantage over Johansen's tests if the number of observations is limited. Moreover, the computational requirement for this method is relatively mild, which makes this method practical for empirical research.

Keywords: *Blind Source Separation, Independent Component Analysis, Cointegration Rank*

JEL: C13, C32, C53

¹The author would like to thank Peter Phillips and Weshah Razzak for insightful discussion as well as the participants of the 23th New Zealand Econometrics Study Group (NZESG) meeting. The financial supports from the Australian Research Council and the Bankwest Curtin Economics Centre are gratefully acknowledged.

1 Introduction

Since the seminal work of Engle and Granger (1987) and Johansen (1988), testing for cointegration has become standard practice in analysing economic and financial time series data. While the importance of cointegration analysis is undeniable, the importance of its various assumptions has often been ignored by researchers. Specifically, the assumption of normality and its implication to statistical inference has not been examined sufficiently from the empirical perspective. Although there is evidence that the Johansen's trace and max tests for cointegration ranks are robust to non-normal error, they do not seek additional information from the underlying distribution. However, given the number of observations is typically small for macroeconomic time series data, the fact that the underlying distribution may not be normal provides important information that may be useful in testing for cointegrating relationships. This observation, however, is not restricted to data with limited observations. High frequency financial time series data often shows evidence of random variables with time-varying second moments and excess kurtosis. This has two implications. Firstly, it is unclear how these characteristics affect existing tests of cointegration. Secondly, excess kurtosis implies the assumption of normality will seriously under-estimate financial risk.

In general, it is somewhat unfortunate that the literature tends to view non-normality as an inconvenience, rather than valuable information which can be useful in empirical research. This is reflected by the fact that most studies focused on the robustness of various testing procedures against non-normality instead of trying to seek information from non-normality. This is perhaps not surprising, as analysing non-normal distributions often leads to intractable mathematical exposition. Worse still, it often appears that the marginal benefit in such analyses do not exceed the marginal cost from their mathematical difficulties. This paper aims to demonstrate that taking advantage of non-normality does not have to be a mathematical challenge. By incorporating and extending some fundamental results from Independent Component Analysis (ICA), this paper proposes a new framework for testing cointegration. The framework explicitly exploits stochastic processes with non-normal distributions and their independence. Monte Carlo simulation shows that the test is comparable to

the Johansen's trace and max test when the number of observations is large and has advantage over the Johansen's tests if the number of observations is limited. Moreover, the computational requirement for this method is relatively mild, which makes this method practical for empirical research.

The paper is organised as follows: Section 2 provides the necessary concepts in Independent Component Analysis. This is followed by the introduction of the new co-integration test in Section 3, which also contains an extension on existing results in ICA to accommodate integrated processes. Section 4 investigates the finite sample properties of the proposed test via Monte Carlo experiments. Section 5 provides an empirical example and concluding remarks can be found in Section 6.

2 Independent Component Analysis

The original formulation of Independent Component Analysis can be found in Héroult and Ans (1984) and Héroult et al. (1985) and since then, the field has grown substantially with applications across many different areas such as signal processing, image processing, neural network and machine learning. A comprehensive survey on the development of ICA can be found in Hyvärinen et al. (2001) and some of the latest results in the areas can be found in Common and Jutten (2010). Common (1984) is also a useful and concise introduction to the subject. Given the rapid development of ICA, it is not possible for this section to provide a comprehensive survey of the subject but rather, it aims to provide a concise introduction to some of the crucial elements in ICA relevant to the testing of cointegration.

Consider the following linear equation:

$$\mathbf{y}_t = A\mathbf{x}_t \tag{1}$$

where $\mathbf{y}_t = (y_{1t}, \dots, y_{kt})'$ is a $k \times 1$ vector, A is a $k \times k$ real matrix and $\mathbf{x}_t = (x_{1t}, \dots, x_{kt})'$ is a $k \times 1$ vector of independent random variables with no more than one normal random variate ². Note that $f_{ij}(x_{it}, x_{j\tau}) = f_i(x_{it})f_j(x_{j\tau})$ for all $i \neq j$ and $t, \tau \in \mathbb{Z}^+$, where

²Strictly speaking, it is assumed that no more than one random variate with the same family of

f_{ij} denotes the joint probability density for x_{it} and x_{jt} and f_i denotes the marginal density for x_{it} , for $i, j = 1, \dots, k$. Although not strictly necessary, it is also assumed that x_{it} is strictly stationary for $i = 1, \dots, k$. This assumption will be relaxed in Section 3.

The standard problem is to recover \mathbf{x}_t based on the observed sample of \mathbf{y}_t without any information on A or \mathbf{x}_t . Obviously, this is not possible if more than one component in \mathbf{x}_t is normally distributed. However, it is possible to recover \mathbf{x}_t , subject to scaling and permutation, if only one of its components is normally distributed. The intuition behind this possibility is best explained via central limit theorems. An important implication of most central limit theorems is that a weighted sum of a sequence of random variables converges in distribution to normality. This implies that an additional random variable in an finite sequence will, in a sense, make the sum of that sequence “closer” to normal in distribution. This is an important insight, as it allows the recovery of individual x_{it} if x_{it} is not normally distributed. In order to take advantage of this insight, let consider the following optimisation problem:

$$\begin{aligned} \hat{b}_i &= \arg \max_{b_i} G(b'_i \mathbf{y}_t) \\ \text{s.t } & b'_i b_i = 1 \end{aligned} \tag{2}$$

where $G(b'_i \mathbf{y}_t) : \mathbb{R}^k \rightarrow \mathbb{R}$ is an appropriate measure of non-normality. Let $z_{it} = b'_i \mathbf{y}_t$ and $q'_i = b'_i A$ then $z_{it} = b'_i \mathbf{y}_t = b'_i A \mathbf{x}_t = q'_i \mathbf{x}_t$. Note that for a given element in \mathbf{y}_t , $y_{it} = a_{i.} x_t$, where $a_{i.}$ denotes the i^{th} row of A . Thus, y_{it} is a linear combination of at least $k - 1$ non-normal random variables. Following the intuition as explained before, y_{it} must be “closer” to normal in distribution than any element in \mathbf{x}_t . Hence, in order to maximise the “non-normality” of $z_{it} = b'_i \mathbf{y}_t$, \hat{b}_i must be the vector such that $\hat{q}'_i = \hat{b}'_i A$ has exactly one non-zero element. In other words, $z_{it} = q_j x_{jt}$ for some $j \leq k$. Thus, one of the unobserved components has been recovered subject to a scale factor q_i . The same process can be repeated in order to recover the remaining unobserved variables by utilising the independence nature of \mathbf{x}_t . This can be achieved by solving the same optimisation problem with an additional constraint that the next solution must be

stable distributions

orthogonal to the previous ones.

Under the assumption that A has full rank, there will be exactly k orthogonal vectors, such that $b'_i b_j = 0$ for all $i \neq j$. Hence, the vector \hat{b}_k can be calculated without solving the additional optimisation problem. Instead, it is simply the vector satisfying $\hat{b}'_k \hat{b}_i = 0$ for $i \neq k$. An implication of this result is that one and only one of the unobserved components can be normally distributed.

The following theorem contains a justification for independent component analysis. It provides a sufficient condition on the objective function, G , to ensure that the solution to the optimisation will recover the original independent components, subject to permutation and scaling.

Theorem 1. *Consider the model as defined in equation (1) with each element in \mathbf{y}_t has zero mean and unit variance. Let $G(x)$ be a smooth even function, then the local maxima (minima) of $\mathbb{E}G(b'\mathbf{y}_t)$ under the constraint $\|b\| = 1$ include those rows of A such that the corresponding independent component x_{it} satisfy*

$$\mathbb{E}[g'(x_{it}) - x_{it}g(x_{it})] > 0 \quad \text{for minima} \quad (3)$$

$$\mathbb{E}[g'(x_{it}) - x_{it}g(x_{it})] < 0 \quad \text{for maxima} \quad (4)$$

$$(5)$$

where $g(x) = G'(x)$.

Proof. See Hyvärinen et al. (2001) □

Essentially Theorem 1 shows that any non-quadratic smooth function, G , that divides the space of probability distributions into two half spaces can be used to identify the unobserved variables. Specifically, the variables whose distributions belong in one half space can be estimated by maximising $G(b'_i \mathbf{y}_t)$ and the independent components whose distributions belong to the other half space can be identified by minimising $G(b'_i \mathbf{y}_t)$.

Corollary 1. *Under the assumptions of Theorem (1), let $H(b) = \mathbb{E}G(b'\mathbf{y}_t)$ then the*

solution, \hat{b} , to the following optimisation problem:

$$\begin{aligned} \hat{b} &= \underset{b}{\text{arg opt.}} \mathbb{E}G(b'\mathbf{y}_t) \\ \text{s.t. } & b'b = 1 \end{aligned}$$

satisfy $A'b = q_i\mathbf{e}_i$, for some scalar $q_i \in \mathbb{R}$ and some $i \leq k$ with \mathbf{e}_i denotes the i^{th} column of a $k \times k$ identity matrix.

Proof. See Hyvärinen et al. (2001) □

Remark 1. In practice, the variance of x_{it} is not known and it is not identifiable without further information. Hence, the elements in matrix A can only be identified up to an appropriate scaling. In fact, the square of the unknown scalar, q_i , is proportional to the variance of x_{it} , which cannot be identified without further information. Moreover, the optimisation seeks out the random variable, x_{it} , which is the most “non-normal” random variable as measured by $G(x)$. Hence, it is not possible to sort the order in which each unobserved variable can be identified. Thus, each component in \mathbf{x}_t can only be recovered subject to scaling and permutation. However, these limitations are often unimportant in practice or can be alleviated with additional information provided by the researchers. This issue will be discussed further in relation to identify cointegrating vector in section 3.

The solution to the optimisation as defined in (2) can be calculated using the algorithm below:

Algorithm 1: One Unit Source Separation

Step i. initiate b_i for some $i = 1, \dots, k$.

Step ii. $b_i \leftarrow b_i + B(b_i)$ where

$$B(b_i) = T^{-1} \left[\sum_{t=1}^T \mathbf{y}_t g(b_i \mathbf{y}_t) - \sum_{t=1}^T g'(b_i \mathbf{y}_t) b_i \right]$$

Step iii. $b_i \leftarrow \frac{b_i}{\|b_i\|}$.

Step iv. Repeat the previous Steps (ii)-(iii) until b_i converges, ie $B(b_i) < \varepsilon^*$ where ε^* is some pre-defined tolerance level.

Step v. $z_{it} = b'_i \mathbf{y}_t$.

Define $\mathbf{z}_t = \hat{B} \mathbf{y}_t$ where $\hat{B} = \begin{pmatrix} \hat{b}'_1 \\ \vdots \\ \hat{b}'_k \end{pmatrix}$, the algorithm above computes an element of \mathbf{z}_t . In order to compute the rest of \mathbf{z}_t , one can repeat the algorithm above, with an additional step to ensure that the new vector is orthogonal to the previous solution, that is, $b'_i b_j = 0$ for all $j < i$. This idea can be summarised in the algorithm below:

Algorithm 2: Multiple Units Source Separation

Step i. Initiate $i = 1$.

Step ii. Apply the One Unit Source Separation algorithm as stated in Algorithm 1 for the i^{th} component in \mathbf{y}_t .

Step iii. If $i \neq 1$ then

$$b_i \leftarrow b_i - \sum_{j=1}^{i-1} (b'_i b_j) b_j.$$

Step iv. Normalise b_i by setting $b_i \leftarrow \frac{b_i}{\|b_i\|}$.

Step v. $\hat{B} = (b'_1, b'_2, \dots, b'_k)'$.

Step vi. $\mathbf{z}_t = \hat{B} \mathbf{y}_t$.

The algorithm above incorporates a Gram-Schmidt type method to ensure that each row in \hat{B} is orthogonal to the others. An implication of Proposition 2 is that the unobserved random vector can be recovered, subject to scaling and permutation, by calculating \mathbf{z}_t following Algorithm 2.

The choice of the objective function, G , may seem crucial in the performance of the test but it needs not be the case. While Proposition 1 provided useful sufficient

conditions for G , the performance of different G remains an empirical question. An obvious choice is the excess kurtosis which is defined to be:

$$\kappa(b'\mathbf{y}_t) = \mathbb{E}\mathbf{y}_t^4 - 3[\mathbb{E}\mathbf{y}_t^2]^2. \quad (6)$$

The idea is that any random variable with kurtosis different to three is likely to be non-normal. Excess kurtosis as defined in equation (6) is always non-negative and equals to zero only if the random variable is normally distributed. Therefore, it provides a simple measure of non-normality. A practical problem with kurtosis is that the expectation in equation (6) cannot be calculated directly and must be estimated. Given the estimation of the fourth moment is sensitive to outliers, this means the measure is also extremely sensitive to outliers. An alternative is to utilise the concept of “negentropy”. Recall that the entropy, H , of a random vector, \mathbf{x} , with density $p_{\mathbf{x}}(s)$ is defined to be

$$H(p) = - \int_A p_{\mathbf{x}}(s) \log p_{\mathbf{x}}(s) ds. \quad (7)$$

Define negentropy, $J(p)$, as

$$J(p) = H(\phi_{\mathbf{x}}) - H(p_{\mathbf{x}}) \quad (8)$$

where $\phi_{\mathbf{x}}$ denotes the normal density with the same variance-covariance matrix as density $p_{\mathbf{x}}$, which is the true density of the random vector \mathbf{x} . A fundamental result is that a normal random variable has the largest entropy among all other random variables with equal variance-covariance structure. Thus, $J(p) > 0$ for all density and can be used as a measure of non-normality.

Similar to excess kurtosis, equation (8) cannot be computed directly and the entropy functional must be approximated based on given data. The standard approach is to utilise the higher-order cumulants. Since $b'\mathbf{y}_t$ is an univariate random variable, the following discussion will focus on the univariate case only. The Gram-Charlier expansion of a density, $p(x)$, is given by

$$p_x(s) = \phi(s) \left(1 + \kappa_3(s) \frac{H_3(s)}{3!} + \kappa_4(s) \frac{H_4(s)}{4!} \right) \quad (9)$$

where $\phi(s)$ and $H_i(s)$ denote the normal density and Chebyshev-Hermite polynomial of order i , respectively and $\kappa_i(s)$ denotes the i^{th} cumulant of the random variable, s . Substitute the right hand side of equation (9) to equation (7) and use the approximation $\log(1 + \varepsilon) \approx \varepsilon - \varepsilon^2/2$, $H(p_x)$ can be approximated as

$$H(p_x) \approx - \int_A \phi(s) \log \phi(s) ds - \frac{\kappa_3(s)^2}{12} - \frac{\kappa_4(s)^2}{48}. \quad (10)$$

Subsequently, for a standardised random variable, the negentropy can be approximated by

$$J(p_{\mathbf{x}}) \approx \frac{2}{12} [\mathbb{E}\mathbf{x}^2]^2 + \frac{1}{48}\kappa(\mathbf{x}). \quad (11)$$

The problem of this approximation is that it involves the estimation of kurtosis, which means it is still not robustness to outliers. However, this can be resolved by generalising the approximation method for $J(p_{\mathbf{x}})$. Let $p(x)$ be the maximum entropy density with the moment constraints,

$$\int p(x)F_i(x)dx = c_i \quad i = 1, \dots, m.$$

$p(x)$ can be approximated by

$$\hat{p}(x) = \phi(x) \left(1 + \sum_{i=1}^m c_i F_i(x) \right).$$

Hence, a more flexible approximation for $J(p_{\mathbf{x}})$ can be derived using the approximation above and this gives:

$$J(p_x) \approx \frac{1}{2} \sum_{i=1}^m \mathbb{E}F_i^2(x).$$

Note that if $F_i(x) = x^i$, then the approximation above is algebraically equivalent to equation (11). In other words, the polynomial moment functions in equation (2) can be replaced by more general functions. Specifically, replace x^2 and x^4 by $F_1(x)$ and $F_2(x)$, respectively gives

$$J(p_{\mathbf{x}}) \approx k_1 [\mathbb{E}(F_1(x))]^2 + k_2 [\mathbb{E}F_2(x) - \mathbb{E}F_2(\nu)]^2. \quad (12)$$

Although the accuracy of this approximation depends on the choice of F_i , it is always consistent as a measurement of non-normality in the sense that it is always non-negative and equals to zero only if the random variable is normally distributed. Thus, the choice of F_i can be extremely flexible and can be chosen to alleviate the impact of outliers, which will make the estimation more robust. A popular choice is:

$$F_1(x) = a \log \cosh ax \quad (13)$$

$$F_2(x) = -\sqrt{2\pi}\phi(x), \quad (14)$$

with $1 \leq a \leq 2$ and is often chosen to be 1.

In order to demonstrate the algorithm in practice, let's consider the following simple example. Let $\mathbf{x}_t = (x_{1t}, x_{2t})'$ such that $x_{1t} \sim \chi^2(1)$ and $x_{2t} \sim t(5)$ and

$$A = \begin{pmatrix} 0.2 & 0.98 \\ 0.7 & 0.71 \end{pmatrix}.$$

Set $\mathbf{y}_t = A\mathbf{x}_t$, that is $y_{1t} = 0.2x_{1t} + 0.98x_{2t}$ and $y_{2t} = 0.7x_{1t} + 0.71x_{2t}$, then apply Algorithm 2 to \mathbf{y}_t with the objective function as defined in equations (12)-(14) to obtain \mathbf{z}_t . Figure 1 shows the 3000 random draws of \mathbf{x}_t , the observed (mixed) variables, \mathbf{y}_t , and the recovered variables, \mathbf{z}_t , which is calculated as $\hat{B}\mathbf{y}_t$ following Algorithm 2.

As shown in Figure 1, \mathbf{z}_t resembles the unobserved variables \mathbf{x}_t . Specifically, z_{1t} resembles x_{1t} with a negative scaling factor where z_{2t} resembles x_{2t} with a positive scaling factor. Note that the algorithm does not require any knowledge of \mathbf{x}_t other than the assumption of independence. The matrix \hat{B} is calculated based on \mathbf{y}_t only. Moreover,

$$\hat{B} = \begin{pmatrix} 0.976 & -1.340 \\ 1.022 & -0.293 \end{pmatrix} \quad Q = \hat{B}A = \begin{pmatrix} -0.743 & 0.005 \\ -0.001 & 0.793 \end{pmatrix}$$

provides evidence supporting Proposition 2. Specifically, as $\mathbf{z}_t = \hat{B}\mathbf{y}_t = \hat{B}A\mathbf{x}_t = Q\mathbf{x}_t$, this shows that each component in \mathbf{z}_t corresponds to a unique element in \mathbf{x}_t subject to scaling.

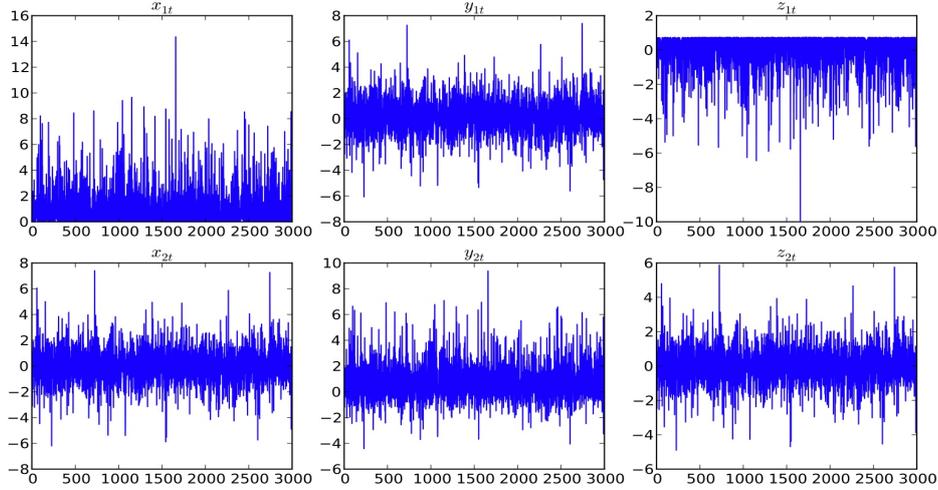


Figure 1: A Simple Demonstration of Algorithm 2

3 Co-integration

This section introduces two new tests of co-integration rank based on ICA as introduced in Section 2. The first test is an application of Theorem 1 and Corollary 1. Let $\mathbf{y}_t = (y_{1t}, \dots, y_{kt})'$ be a $k \times 1$ vector of $I(1)$ variables, such that $\mathbf{y}_t = A\mathbf{x}_t$ where \mathbf{x}_t consists of r $I(0)$ variables and $k - r$ $I(1)$ variables. Note that $\mathbf{y}_t \sim I(1)$ with r cointegrating vectors. The basic idea is to seek a matrix \hat{B} such that $\hat{B}\mathbf{y}_t$ maximizes some measures of non-normality. Note that there are r stationary variables in \mathbf{x}_t , this means there exists r rows in \hat{B} such that $\hat{b}_i\mathbf{y}_t \sim I(0)$ for some i . Thus, the test also estimates the co-integrating vectors and can also potentially identify the underlying $I(1)$ and $I(0)$ variables that govern the co-integrating system.

Formally, consider the following linear system:

$$\mathbf{y}_t = A\mathbf{x}_t \quad (15)$$

where $\mathbf{x}_t = (\mathbf{x}'_{Xt}, \mathbf{x}'_{It})'$ such that $\mathbf{x}_{Xt} = (x_{1t}, \dots, x_{rt})' \sim I(0)$ and $\mathbf{x}_{It} = (x_{r+1t}, \dots, x_{kt})' \sim I(1)$. Obviously, $\mathbf{y}_t \sim I(1)$ and the system as defined in equation (15) can be transformed linearly to the triangular representation as presented in Phillips (1991). To see

this, let V be a $k \times k$ matrix such that

$$\mathbf{u}_t = V\mathbf{y}_t = V\mathbf{A}\mathbf{x}_t$$

with

$$VA = \begin{pmatrix} \mathbf{I}_{r \times r} & \mathbf{C}_{r \times (k-r)} \\ \mathbf{0}_{(k-r) \times r} & \mathbf{I}_{(k-r) \times (k-r)} \end{pmatrix},$$

where \mathbf{I} and $\mathbf{0}$ denote the identity and null matrix, respectively. Note that V always exists since A has full rank. As a result, $\mathbf{u}_t \sim I(1)$ is expressed in the triangular representation as proposed in Phillips (1991) with the matrix \mathbf{C} containing the cointegrating vectors.

Note that if $\mathbf{y}_t = \mathbf{A}\mathbf{x}_t$ then $\Delta\mathbf{y}_t = \mathbf{A}\Delta\mathbf{x}_t$ where Δ denotes the first difference operator. Since $\mathbf{x}_t \sim I(1)$, it implies $\Delta\mathbf{x}_t$ has finite second moment and therefore, \mathbf{A}^{-1} can be recovered subject to scaling and permutation by solving the following optimisation problem using Algorithm 2.

$$\begin{aligned} \hat{B} &= \arg \max_{B \in U} G(B\mathbf{y}_t) \\ \text{s.t. } B'B &= \mathbf{I} \end{aligned} \tag{16}$$

where \mathbf{I} denotes the identity matrix with the appropriate dimension. Let \hat{b}_i denotes the i^{th} row of \hat{B} and define $\mathbf{z}_t = (z_{1t}, \dots, z_{kt})'$, such that

$$z_{it} = \hat{b}_i\mathbf{y}_t = q_{ij}x_{jt}, \tag{17}$$

for some $j = 1, \dots, k$. That is, z_{it} is a scalar multiple of one of the components in \mathbf{x}_t . In addition, due to the constraint $\hat{B}'\hat{B} = \mathbf{I}$, the matrix \hat{B} has full rank, so each element in \mathbf{z}_t is a scalar multiple of a unique element in \mathbf{x}_t .

Since each element in \mathbf{z}_t is a scalar multiple of a unique element in \mathbf{x}_t , there are exactly r stationary variables and $l = k - r$ non-stationary variables in \mathbf{z}_t . Thus, the cointegrating vectors, if exist, are the rows in \hat{B} that correspond to the stationary components in \mathbf{z}_t . Therefore, the test of cointegration can be performed by testing the

stationarity of each row of $\hat{B}\mathbf{y}_t$.

An advantage of this approach is that the test of cointegration is reduced to multiple tests for unit root. More importantly, the order of integration is not required before conducting this test. This can have empirical advantage over the Johansen's approach, as it allows different tests of unit root with each row of \mathbf{z}_t , and can subsequently enhance the test power by accommodating different characteristics of the underlying variables, such as seasonality and structural break.

A potential drawback with this approach is that the test explicitly assumes the components in $\Delta\mathbf{x}_t$ remain non-normally distributed. The reason that the test requires differencing is related to the fact that the proofs of Theorem 1 and Corollary 1 require the unobserved vector, \mathbf{x}_t , to be weakly stationary. However, it is possible to extend Theorem 1 and Corollary 1 to accommodate possible non-stationary components in \mathbf{x}_t . Without loss of generality, let $\mathbf{x}_t = (x'_{It}, x'_{St})'$ such that $x_{It} = (x_{1t}, \dots, x_{lt})' \sim I(1)$ and $x_{St} = (x_{l+1t}, \dots, x_{kt})' \sim I(0)$. Define also, $I = \{1, \dots, l\}$ and $S = \{l+1, \dots, k\}$. Consider the following assumptions:

A1. $\exists T_0 > 0$ such that for any $T > T_0$, $\exists c_T \geq c_0 \geq 0$ and:

$$\left| T^{-1} \sum_{t=1}^T x_{it} \right| \leq c_0.$$

A2. Let $f(x)$ and $g(x)$ be two univariate real-valued and differentiable functions,

$$T^{-1} \left| \sum_{t=1}^T g(x_{it})f(x_{jt}) - \sum_{t=1}^T g(x_{it}) \sum_{t=1}^T f(x_{jt}) \right| \leq c_T \quad \forall i \neq j.$$

A3. $\forall \varepsilon > 0 P(c_T < \varepsilon) = 1$ as $T \rightarrow \infty$.

A4.

$$\begin{aligned}
x_{it} &= x_{it-1} + \varepsilon_{it} && \text{for } i \in I \\
x_{it} &= \Pi(L)\varepsilon_{it} && \text{for } i \in S \\
\varepsilon_{it} &\sim \text{iid}(0, \sigma^2) && \forall i, t \\
\mathbb{E}(\varepsilon_{it}\varepsilon_{j\tau}) &= 0 && \text{for } t \neq \tau, i \neq j.
\end{aligned}$$

where $\Pi(L)$ denotes a lag polynomial with all roots outside the unit circle.

Assumption 1 ensures the sample means are zero for all unobserved components. This can always be achieved and it is therefore assumed for mathematical convenience. Assumption 2 is the sample analogue to independence as required by the standard independent components analysis. Assumption 4 restricts the class of non-stationary process to be unit root, which can be relaxed further to unit root with drift.

Proposition 1. *Consider the model as defined in equation (1) and let $G(x)$ be a twice differentiable functions. Under Assumptions (1) - (3), there exists i such that $1 \leq i \leq k$ and a scalar, q_i such that the solution to the following optimisation problem*

$$\hat{w} = \arg \underset{b}{opt.} \sum_{t=1}^T G(b'y_t) \quad (18)$$

$$s.t. b'b = 1 \quad (19)$$

has the following characteristics:

- i. The absolute value of each element in $A'\hat{b} - q_i\mathbf{e}_i$ is $O_p(1)$
- ii. The absolute value of each element in $\hat{b} - q_i\mathbf{e}_iA^{-1}$ is $O_p(1)$.

Proof. See Appendix. □

This proposition essentially ensures that the unobserved components can still be identified even in the presence of $I(1)$ variables. The following proposition shows the algorithms presented in Section 2 can be used to solve the optimisation problem consider in Proposition 1.

Proposition 2. *Under the assumptions of Proposition 1 and let \hat{B} be the matrix as calculated in Algorithm 2, there exists a permutation matrix \mathbf{K} and a diagonal matrix \mathbf{Q} such that the absolute value of each element in $\mathbf{KQ}\hat{B} - A^{-1}$ is $O_p(1)$.*

Proof. See Appendix. □

Essentially, the results established in Propositions 1 and 2 mean that it is possible to estimate A^{-1} in the presence of integrated unobserved components using existing algorithm such as Algorithm 2. The finite sample performance of these two approaches will be compared with the Johansen max and trace tests in the next section via Monte Carlo experiments.

4 Finite Sample Performance

This section compares the finite sample performance of the proposed tests with the max and trace tests of Johansen (1988) via Monte Carlo simulation. The co-integrating system is generated through the follow Data Generating Process (DGP):

$$\begin{pmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \\ y_{4t} \\ y_{5t} \end{pmatrix} = \begin{pmatrix} 0.2 & 0.4 & 0.7 & 0.1 & 0.2 \\ 0.8 & 0.9 & 1.0 & 0.3 & 0.4 \\ 0.3 & 0.3 & 0.9 & 0.2 & 0.4 \\ 0.72 & 1.2 & 2.3 & 1.4 & 1.0 \\ 1.03 & 1.13 & 2.12 & -1.13 & 3.12 \end{pmatrix} \begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \\ x_{4t} \\ x_{5t} \end{pmatrix} \quad (20)$$

where

$$\begin{pmatrix} x_{1t} \\ \vdots \\ x_{lt} \\ \vdots \\ x_{5t} \end{pmatrix} = \begin{pmatrix} x_{1t-1} \\ \vdots \\ x_{lt-1} \\ 0 \\ \vdots \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \vdots \\ \varepsilon_{lt} \\ \vdots \\ \varepsilon_{5t} \end{pmatrix} \quad (21)$$

with $r < 5$. Each ε_{it} follows a skewed t-distribution with 5 degrees of freedom and skew parameter equals to 2. This section conducts four sets of Monte Carlo experiments with $T = 50$ and $T = 100$ and in each case, $r = 1$ ($l = 4$) and $r = 4$ ($l = 1$). The number of

Table 1: Finite Sample Performance

| T | R | Johansen | | New Test | |
|-----|-----|----------|------|----------------------|----------------------------|
| | | Trace | Max | Using \mathbf{y}_t | Using $\Delta\mathbf{y}_t$ |
| 50 | 1 | 0.85 | 0.67 | 0.91 | 0.96 |
| | 4 | 3.43 | 2.50 | 3.71 | 3.68 |
| 100 | 1 | 1.09 | 1.07 | 1.11 | 1.10 |
| | 4 | 3.88 | 4.05 | 3.86 | 3.91 |

replication is 1500 in each case.

For the first test, \hat{B} is obtained by solving the optimisation problem as defined in equation (16) using Algorithm 2 with $\Delta\mathbf{y}_t$. The objective function G follows the approximation of negentropy as defined in equations (12) - (14). The potential cointegrating vectors are identified via testing the stationarity of each element in $\mathbf{z}_t = \hat{B}\mathbf{y}_t$ using the unit root test proposed in Im et al. (2011). The second test follows the exact same procedure except Algorithm 2 utilises the actual data \mathbf{y}_t rather than its first difference. Table 1 contains the summary of the finite sample performance in each case.

As shown in Table 1, Johansen's Trace and Max tests remain robust to non-normal errors when the sample is moderately large but the new tests performed slightly better when the number of observations is small. This provides some evidence for Propositions (1) and (2) and shows that it is possible to take advantage of non-normal errors when testing for co-integrating ranks.

The next section applies the proposed tests of cointegrating rank to examine the spot and forward relationship in US/AU exchange rates. It also compares the dynamics of the unobserved variables with the dynamics of spot and forward rates. The results provide some interesting insight relating to market efficiency. It also raises some interesting questions on some of the standard assumptions in analysing financial time series.

5 Empirical Example: A Simple Spot and Forward Exchange Rate Model

This section applies the proposed tests of cointegrating rank to examine relationship between spot and forward exchange rates. The forward-spot relationship has been studied extensively in the literature, see for examples, Mark (1990), Naka and Whitney (1995), Liuntel and Paudyal (1998) and Zivot (2000). Let s_t and f_t denote the logarithms of spot and forward exchange rates at time t , respectively, then the forward rate unbiasedness hypothesis (FRUH) suggests that s_{t+1} and f_t should be cointegrated with the cointegrating vector being $(1, -1)'$. Zivot (2000) argued that if $s_t - f_t$ is stationary then $s_{t+1} - f_t$ must also be stationary but the underlying Vector Error Correction Model (VECM) will be much more complicated. Therefore, inference based on a simple VECM model for s_{t+1} and f_t may not be valid. More importantly, testing FRUH using high frequency data with existing co-integration techniques may be problematic for at least two reasons. Firstly, most daily financial time series have high excess kurtosis and hence, the assumption of normality does not hold. Secondly, the conditional variance of most high frequency financial time series are time-varying and its impact on testing co-integrating relationship is still unclear.

In order to address these issues, this section proposes a simple model to analyse spot and forward exchange rates. The model assumes both spot and forward rates are function of two unobserved components. This follows from the theory that the observed price of any asset at any given time can be decomposed into the “efficient price” and market influence. The former is the price of the asset when market is fully efficient and the latter represents external influence and other market behaviours, such as market micro-structure noise. The model implies that both spot and forward rates must be a function of the efficient rate, that is, the exchange rate when market is efficient. It also implies that both spot and forward rates are influenced by the same market noise but with different sensitivity. These can be expressed mathematically as

$$\begin{pmatrix} s_t \\ f_t \end{pmatrix} = A \begin{pmatrix} e_t \\ n_t \end{pmatrix} \quad (22)$$

where s_t and f_t denote the logarithm of spot and forward rates, respectively. A is a

2×2 matrix with full rank. e_t and n_t denote the efficient exchange rate and market noise, respectively. This model has two implications under the assumption that the market noise is stationary with zero mean. Firstly, the non-stationary nature of both spot and forward rates comes from the efficient exchange rate, e_t . Secondly, the market noise, n_t , represents the deviation from the equilibrium. To see this, invert equation (22) yields:

$$\frac{1}{|A|} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix} \begin{pmatrix} s_t \\ f_t \end{pmatrix} = \begin{pmatrix} e_t \\ n_t \end{pmatrix}. \quad (23)$$

Hence, $\left(1, -\frac{a_{22}}{a_{21}}\right)'$ must be the co-integrating vector and n_t represents the error correction term. Note that the efficient rate is often assumed to follow a geometrical brownian motion in continuous time, which is consistent with e_t being non-stationary. More importantly, if it is possible to estimate e_t based on s_t and f_t then the various assumptions on efficient rate can be verified.

Following the model as defined in equation (22), it is possible to apply the proposed test of cointegration rank to determine the cointegrating vector and subsequently estimate e_t and n_t . This is particularly interesting because it is well known that the daily spot and forward data do not follow the normal distribution in either level or first difference. Thus, the proposed test can exploit the non-normality nature of the variables.

This section applies the proposed test of co-integration to examine the forward-spot relationship in the exchange rate between U.S. and Australia. It utilises five-days daily spot and forward exchange rates data from DataStream with the sample starting from 17 September 1999 to 19 September 2012. This gives a total 3394 observations. As mentioned in Section 3, the proposed test can utilise the data at level or first difference. Since both test procedures produce very similar results, this section reports the test based on first differenced data only for brevity.

Note the model as defined in equation (22) implies

$$s_t = \frac{a_{11}}{a_{21}} f_t - \frac{|A|}{a_{21}} n_t.$$

Hence, FRUH holds if $a_{11} = a_{21}$ and $\mathbb{E}n_t = 0$. The implication of the first condition

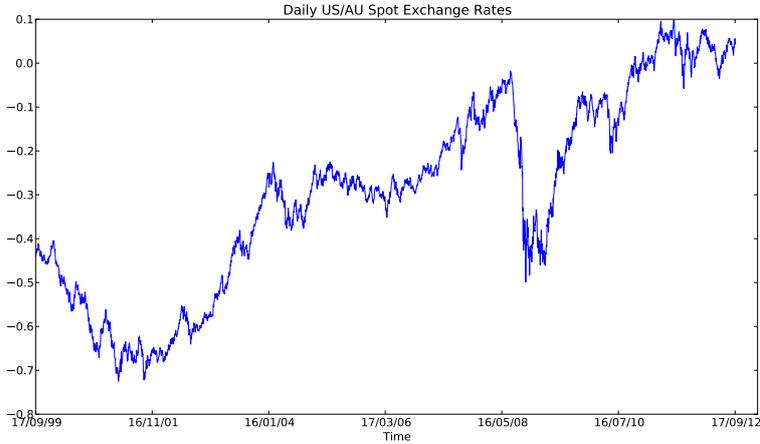


Figure 2: Daily US/AU Spot Exchange Rates

Table 2: Unit Root Tests on US/AU Spot and Forward Rates

| | DF Test | PP Test | KPSS Test |
|--------------|-----------|-----------|-----------|
| s_t | -3.039 | -3.071 | 0.334** |
| f_t | -2.974 | -3.054 | 0.380** |
| Δs_t | -41.782** | -59.112** | 0.064 |
| Δf_t | -43.091** | -63.391** | 0.070 |

is that the cointegrating vector between s_t and f_t must be $(1, -1)'$ and the second condition can be verified by examining n_t .

Figures 2 - 4 contain the plots for daily spot and forward exchange rates as well as the scatter plot between spot and lagged forward exchange rates. As shown in these Figures, both spot and forward exchange rates suffered from steep declines in late 2008 due to the Global Financial Crisis (GFC). However, the scatter plot between the spot and forward shows a strong and rather stable linear relationship. It provides some evidence suggesting that the GFC does not affect the fundamental relationship between US/AU spot and forward exchange rates.

In terms of the order of integration, both spot and forward rates contain unit roots in level but not in first difference, as shown in Table 2. This is confirmed by three

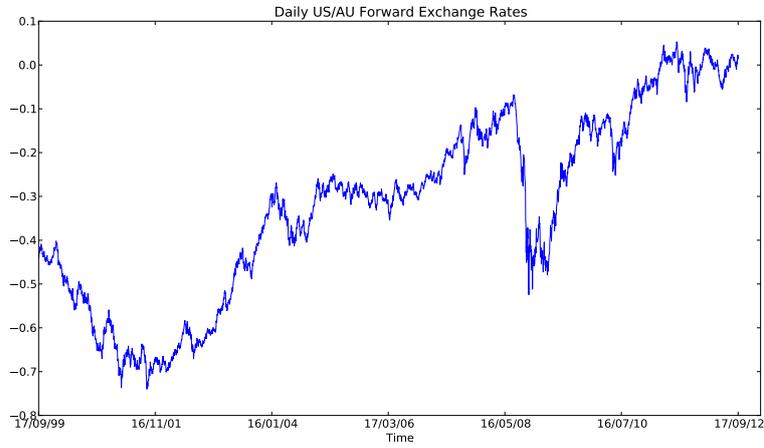


Figure 3: Daily US/AU Future Exchange Rates

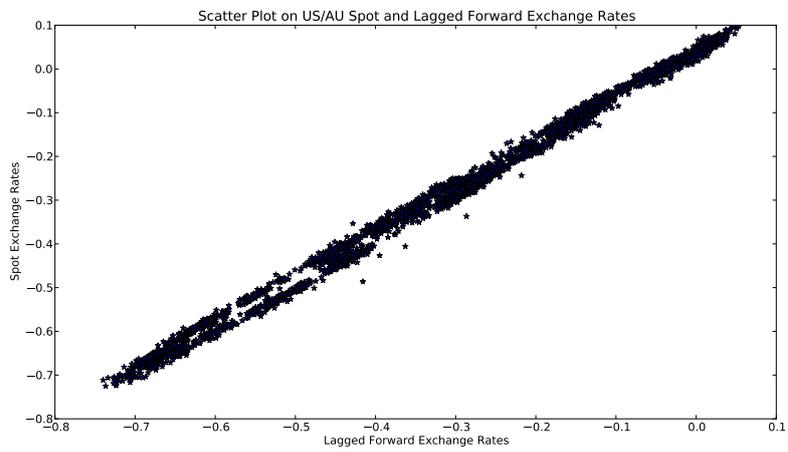


Figure 4: Scatter Plot on Daily US/AU Spot and Future Exchange Rates

Table 3: Unit Root Tests on the Estimated Unobserved Components

| | n_t | e_t |
|-----------|------------|----------|
| DF Test | -7.663*** | -1.743 |
| PP Test | -16.956*** | -2.428 |
| KPSS Test | 0.150 | 0.798*** |

different unit root tests, namely, the augmented Dicky Fuller (DF) test in Dickey and Fuller (1979) and Said and Dickey (1984), the Phillip-Perron (PP) test in Phillips and Perron (1988) and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test in Kwiatkowski et al. (1992).

Apply Algorithm 2 to the data gives ³:

$$\hat{B} = \begin{pmatrix} 1 & -1.0343 \\ 0.595 & 0.256 \end{pmatrix}. \quad (24)$$

Pre-multiplying \hat{B} with $(s_t, f_t)'$ and apply unit root tests to the resulting series gives:

$$\begin{pmatrix} 1 & -1.0343 \\ 0.595 & 0.256 \end{pmatrix} \begin{pmatrix} s_t \\ f_t \end{pmatrix} = \begin{pmatrix} q_1 n_t \\ q_2 e_t \end{pmatrix}.$$

This reflects the fact that the first row of \hat{B} is the co-integrating vector as $s_t - 1.0343f_t \sim I(0)$ based on unit root test result as shown in Table 3. It also reflects the fact that ICA can only recover A up to scaling and permutation.

More importantly, the algorithm produces the two unobserved components as shown in Figure 5. Interestingly, the non-stationary component resembles the dynamics in both spot and forward rates. The stationary component, however, exhibits completely different type of dynamics. Following from the standard assumption in the literature, the non-stationary component can be considered as estimates for e_t , the efficient exchange rate, and the stationary component, n_t , can be considered as the “noise” induced by market behaviour and other external influences. The histogram and de-

³The matrix presented here is being normalised by the (1, 1) element in the original matrix produced by Algorithm 2

scriptive statistics of n_t can be found in Table 5 and Figure 6. As shown in Figure 6, the stationary component seems to have a bi-modal distribution and the descriptive statistics as shown in Table 5 seem to suggest that the mean of n_t is not statistically different from zero and thus provides support of the FRUH. However, a closer examination reveals that n_t is not independent and identically distributed.

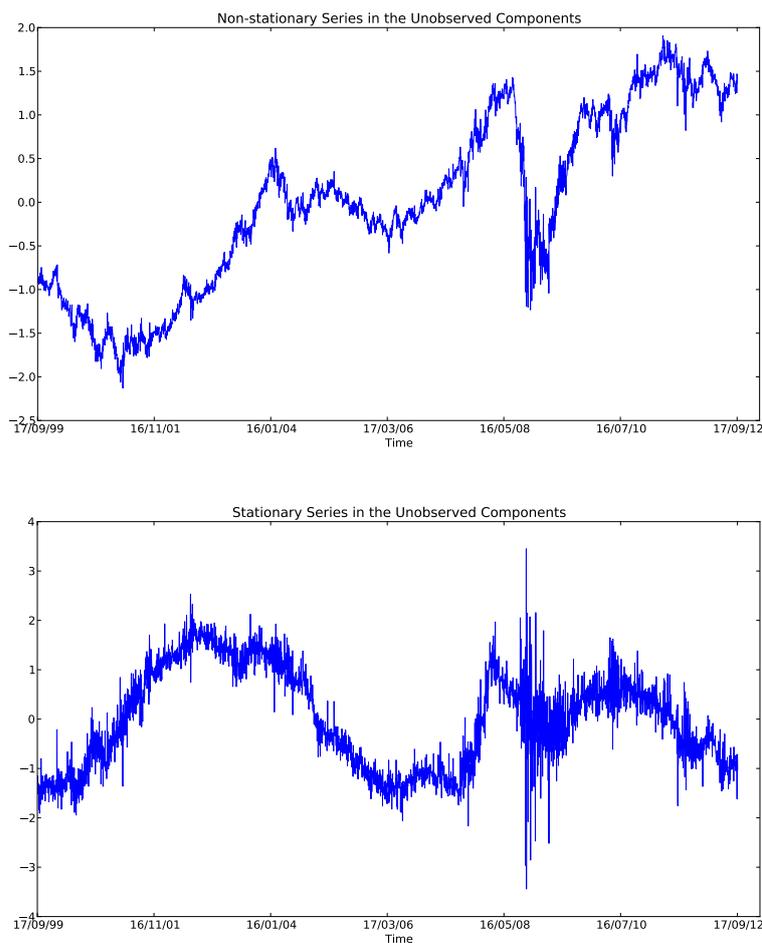


Figure 5: Unobserved Components

In order to compare the dynamic structures between e_t and n_t as well as the original spot and forward exchange rates, consider the following $ARMA(p, q)$ - $GARCH(r, s)$

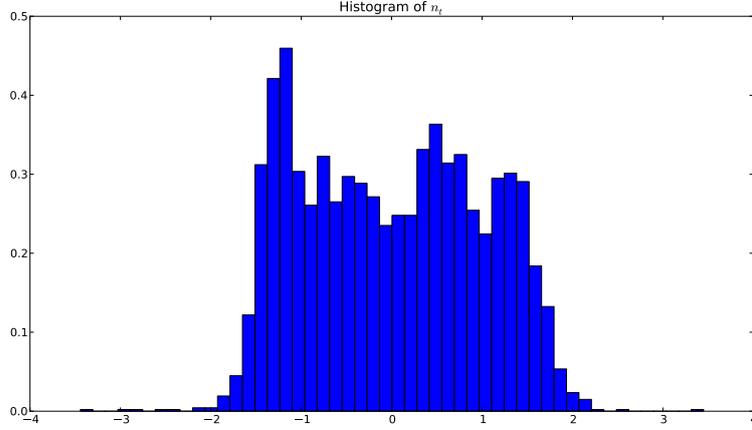


Figure 6: Histogram of estimated n_t

model:

$$\begin{aligned}
 \Phi(L)(y_t - \mu) &= \Theta(L)\varepsilon_t \\
 \varepsilon_t &= \eta_t \sqrt{h_t} \\
 h_t &= \omega + \sum_{i=1}^r \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^s \beta_i h_{t-i}
 \end{aligned} \tag{25}$$

where $\Phi(L) = 1 - \sum_{i=1}^p \phi_i L^i$ and $\Theta(L) = 1 + \sum_{i=1}^q \theta_i L^i$ with L denotes the lag operator such that $Ly_t = y_{t-1}$. It is assumed that $\Phi(L)$ and $\Theta(L)$ has no common or unit root. The structural and statistical properties of this model have been studied extensively in the literature, see for example, Ling and McAleer (2003).

Following the spot-forward rate model as defined in equation (22), it implies that

$$\begin{pmatrix} \Delta s_t \\ \Delta f_t \end{pmatrix} = A \begin{pmatrix} \Delta e_t \\ \Delta n_t \end{pmatrix}.$$

Note that the vectors on both sides of the equation are $I(0)$, this implies the parameters of the ARMA-GARCH model as defined in equation (25) can be consistently estimated

via Quasi Maximum Likelihood Estimator (QMLE). For further details, see Ling and McAleer (2003).

The lag orders for the conditional mean and variance are determined by Schwarz-Bayesian Information Criterion (SBIC). Table 4 contains the estimation results for the four series.

As shown in Table 4, the dynamics between the observed and the unobserved are quite different. Interestingly, the best fit model for forward rate required a higher order lag in the conditional mean while the best fit model for the efficient rate, e_t , required a higher order lag in the conditional variance. More importantly, since both spot and forward rates are linear combination of e_t and n_t , it means that the parameter estimates of spot and forward rates should be interpreted with cautions. This is due to the fact that aggregating a series of GARCH processes does not produce a GARCH process. This implies that the standard GARCH specification is misspecified for spot and forward rates, which casts doubts on the reliability on the parameter estimates for these two series.

Another interesting observation is that the standardised residuals did not seem to be normally distributed for all four series, as shown in Table 5. If the interpretation of efficient rate and market noise was valid in this case, it implied that neither the efficient rates nor the market noise was normally or identically distributed. This raises some interesting questions about the assumption of normality when modelling high frequency financial time series. Moreover, the best fit model for e_t exhibited different dynamics than those implied by the standard Geometrical Brownian Motion (GBH). Thus, the application of GBH might also require further revision if the same empirical evidence found to be common among other financial time series.

The empirical results in this paper is also important for analysing ultra-high frequency (or intra-daily) financial data. The recent development in analysing ultra-high frequency financial data relies on consistent estimation of realized volatility. An important aspect of its estimation is the presence of market micro-structure noise, which is assumed to be normally and identically distributed in most cases. The results established in this paper cast doubts on this assumption and it appears that the ICA method introduced in this paper provides an avenue for future research in this area.

Table 4: Parameters Estimates for the Observed and Unobserved Series

| | Δs_t | Δf_t | Δe_t | Δn_t |
|------------|--------------------|---------------------|---------------------|---------------------|
| μ | 0.024 (1.903) | -0.005 (-46.954) | -0.114 (-2.162) | -0.039 (-0.623) |
| ϕ_1 | -0.206 (-0.871) | 0.342 (23.498) | 1.18e-4 (0.373) | -0.048 (-2.136) |
| ϕ_2 | | 0.656 (28.545) | | |
| θ_1 | 0.225 (0.942) | -0.389 (-22.350) | -0.558 (-41.033) | -0.848 (-67.741) |
| θ_2 | | -0.614 (-18.273) | | |
| ω | 0.007 (3.447) | 0.008 (3.244) | 1.992 (3.969) | 15.600 (3.358) |
| α_1 | 0.057 (5.122) | 0.060 (5.773) | 0.176 (6.407) | 0.076 (4.831) |
| β_1 | 0.933 (89.891) | 0.927 (95.126) | 0.347 (2.194) | 0.906 (53.271) |
| β_2 | | | 0.032 (0.252) | |
| β_3 | | | 0.418 (4.781) | |

Table 5: Descriptive Statistics of the Standardised Residuals

| | Mean | Deviation | Skewness | Kurtosis | Jarque-Bera Statistics |
|--------------|--------|-----------|----------|----------|------------------------|
| Δs_t | -0.014 | 0.998 | -0.441 | 4.740 | 537.820*** |
| Δf_t | -0.011 | 0.996 | -0.378 | 3.994 | 220.547*** |
| Δe_t | 0.008 | 0.999 | 0.416 | 4.786 | 480.431*** |
| Δn_t | 0.011 | 0.998 | 0.245 | 6.461 | 702.206*** |

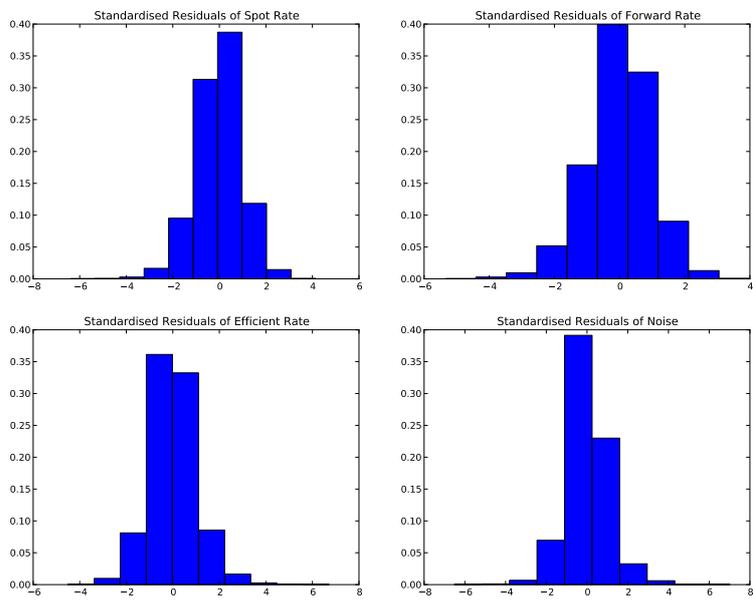


Figure 7: Histogram of the Standardised Residuals

6 Conclusion

This paper presented two new tests of cointegration rank that take advantage of non-normality explicitly. Monte Carlo experiments showed that the new tests performed as well as the Johansen's max and trace tests in large sample, and had slight advantage over the Johansen's tests in small sample. The paper then applied the tests to US/AU spot and forward exchange rates. Empirical results showed that the two exchange rates are cointegrated with the cointegrating vector consistent with the Forward Rate Unbiasedness Hypothesis. Moreover, the testing procedure also produced two independent time series as by products. Interestingly, one of the independent time series could conceivably be interpreted as efficient exchange rate while the other could represent external influence or noise, possibly due to market inefficiency. Closer examination showed that both independent time series contained different dynamic structures. Specifically, both time series contained GARCH errors with different lag orders. This implied that the parameter estimates of GARCH models in the original spot and forward rates might be biased. Moreover, the standardised residuals from both series were non-normal and thus, it provided some evidence against the assumption of normality when modelling financial time series data. More specifically, the assumption that the market noise is a normally and independent distributed random variate might not be appropriate.

Obviously, the empirical finding of this paper had been limited to one particular time series and therefore, the results cannot be generalised without further analysis on broader set of financial time series. However, the intention of this paper was not to challenge the standard practices in the literature, but rather, to demonstrate the usefulness of independent component analysis in analysing economic and financial time series. Specifically, this paper demonstrated that non-normality can be useful in identifying cointegration rank as well as identifying the unobserved factors governing the observed time series.

REFERENCES

- Common, P. (1984). Independent component analysis - a new concept. *Signal Processing* 36, 287 – 314.
- Common, P. and C. Jutten (Eds.) (2010). *Handbook of Blind Source Separation*. Academic Press.
- Dickey, D. and W. Fuller (1979). Distributions of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74, 427–431.
- Engle, R. and C. Granger (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica* 55, 251–276.
- Hérault, J. and B. Ans (1984). Circuits neuronaux à synapses modifiables: décodage de messages composites par apprentissage non supervisé. *C.-R. de l'Académie des Sciences* 299, 525–528.
- Hérault, J., C. Jutten, and B. Ans (1985). Détection de grandeurs primitives dans un message composite par une architecture de calcul neuromimétique en apprentissage non supervisé. In *Actes du Xème colloque GRETSI*, pp. 1017–1022.
- Hyvärinen, A., J. Karhunen, and E. Oja (2001). *Independent Component Analysis*. John Wiley & Sons, Inc.
- Im, K., J. Lee, and M. Tieslau (2011). More powerful unit root tests with non-normal errors. *SSRN Working Papers Series*, 1–36.
- Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* 12, 231–254.
- Kwiatkowski, D., P. Phillips, P. Schmidt, and Y. Shin (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics* 54, 159–178.
- Ling, S. and M. McAleer (2003). Asymptotic theory for a Vector ARMA-GARCH model. *Econometric Theory* 19, 278–308.

- Liuntel, K. and K. Paudyal (1998). Common stochastic trends between forward and spot exchange rate. *Journal of International Money and Finance* 11, 366–381.
- Mark, N. C. (1990). Real and nominal exchange rates in the long run” an empirical investigation. *Journal of International Economics* 28, 115–136.
- Naka, A. and G. Whitney (1995). The unbiased forward rate hypothesis re-examined. *Journal of International Money and Finance* 14, 857–867.
- Phillips, P. (1991). Optimal inference in cointegrated systems. *Econometrica* 59, 283–306.
- Phillips, P. and P. Perron (1988). Testing for a unit root in time series regression. *Biometrika* 75, 335–346.
- Said, S. and D. Dickey (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika* 71, 599–607.
- Zivot, E. (2000). Cointegration and forward and spot exchange rate regressions. *Journal of International Money and Finance* 19, 785–812.

Appendix

Proof of Proposition 1. Note that $\mathbf{y}_t = A\mathbf{x}_t$ and thus $H(w'\mathbf{y}_t) \equiv H(q'\mathbf{x}_t) = H(q)$ where $q = A'w$. Moreover, the constraint, $w'w = 1$ implies there exists a real number, η , such that $q'q = \eta$. The second order Taylor expansion of $H(q)$ is

$$H(q + \varepsilon) = H(q) + \nabla' H(q)\varepsilon + \frac{1}{2}\varepsilon'\nabla^2 H(q)\varepsilon + O(\|\varepsilon\|^2).$$

where ∇ and ∇^2 denotes the gradient and Hessian operators, respectively. These implies:

$$\nabla H(q) = T^{-1} \sum_{t=1}^T g(q'\mathbf{x}_t)\mathbf{x}_t \quad \nabla^2 H(q) = T^{-1} \sum_{t=1}^T g'(q'\mathbf{x}_t)\mathbf{x}_t\mathbf{x}_t' \quad (26)$$

where the functions $g(x)$ and $g'(x)$ are the first and second derivatives of $G(x)$, respectively. Assume that $\varepsilon = (\varepsilon_1, \dots, \varepsilon_k)'$ with $\|\varepsilon\|$ sufficiently small. If $c_T \leq \|\varepsilon\|^2$ then, evaluating the Taylor expansion at $q = \eta \mathbf{e}_i$ gives,

$$\begin{aligned} H(\eta \mathbf{e}_i + \varepsilon) &= H(\eta x_{it}) + T^{-1} \sum_{t=1}^T g(\eta x_{it} x_{it}) \varepsilon_i \\ &\quad + T^{-1} \sum_{j \neq i}^k \left(\sum_{t=1}^T g(\eta x_{it}) x_{jt} \right) \varepsilon_j + \frac{1}{2} T^{-1} \sum_{t=1}^T g'(\eta x_{it}) x_{it}^2 \varepsilon_i^2 \\ &\quad + \frac{1}{2} T^{-1} \sum_{j \neq i}^k \sum_{l \neq j, \neq i}^k \left(\sum_{t=1}^T g'(\eta x_{it}) x_{jt} x_{lt} \right) \varepsilon_j \varepsilon_l + \frac{1}{2} T^{-1} \sum_{t=1}^T \sum_{j \neq i}^k g'(\eta x_{it}) x_{jt}^2 \varepsilon_j^2 \\ &\quad + O(\|\varepsilon\|^2). \end{aligned}$$

Under Assumptions (1) and (2)

$$\begin{aligned} H(\eta \mathbf{e}_i + \varepsilon) &= H(\eta x_{it}) + T^{-1} \sum_{t=1}^T g(\eta x_{it}) x_{it} \varepsilon_t + T^{-1} \sum_{j \neq i}^k \left(\sum_{t=1}^T g(\eta x_{it}) \right) c_T \varepsilon_j \\ &\quad + \frac{1}{2} T^{-1} \sum_{t=1}^T g'(\eta x_{it}) x_{it}^2 \varepsilon_i^2 + \frac{1}{2} T^{-1} \sum_{j \neq i}^k \sum_{l \neq i, l \neq j}^k \left(\sum_{t=1}^T g'(\eta x_{it}) \right) c_T^2 \varepsilon_j \varepsilon_l \\ &\quad + O(\|\varepsilon\|^2). \end{aligned}$$

Since $q'q = \eta$, it implies $\varepsilon_i = \left(\eta - \sum_{j \neq i}^k \varepsilon_j^2 \right)^{\frac{1}{2}} - \eta$. Using the fact that $(\eta - \gamma)^{1/2} = \eta - \gamma/2 + O(\gamma^2)$, it implies $\varepsilon_i = -\frac{1}{2} \sum_{j \neq i}^k \varepsilon_j^2$. Hence ε_i is of higher order and can be neglected. This yields:

$$\begin{aligned} H(\eta \mathbf{e}_i + \varepsilon) &= H(\eta x_{it}) - (2T)^{-1} \sum_{t=1}^T g(\eta x_{it}) x_{it} \sum_{j \neq i}^k \varepsilon_j^2 + (2T)^{-1} \sum_{j \neq i}^k \sum_{t=1}^T g'(\eta x_{it}) x_{jt}^2 \sum_{j \neq i}^k \varepsilon_j^2 \\ &= H(\eta x_{it}) + \frac{1}{2} \left[T^{-1} (\sigma_I^2 + \sigma_S^2) \sum_{t=1}^T g'(\eta x_{it}) - T^{-1} \sum_{t=1}^T g(\eta x_{it}) x_{it} \right] \sum_{j \neq i}^k \varepsilon_j^2 \end{aligned}$$

where σ_I^2 and σ_S^2 denote the variances of the process that belong to group I and S , respectively. The last expression indicates that $\nabla H(\eta \mathbf{e}_i) = 0$, and hence $\eta \mathbf{e}_i$ is clearly an extremum. This completes the proof. \square

Proof of Proposition (2). Following from Proposition (1), the optimisation problem as solved by Algorithm 2 gives a set of orthogonal vectors, such that $\hat{B} = (w'_1, \dots, w'_k)'$ is a $k \times k$ matrix with the property that $A'w_i = q_j \mathbf{e}_j + O_p(1)$ for some $j \in \{1, \dots, k\}$. Since $w_i w_j = 0$, $A'w_i \neq A'w_j$ for $i \neq j$, there exists a matrix $\mathbf{Q} = \text{diag}(q_1, \dots, q_k)$ and a permutation matrix \mathbf{K} such that the absolute value of each element in $\mathbf{KQ}\hat{B} - A^{-1}$ is $O_p(1)$. This completes the proof. \square

The Bankwest Curtin Economics Centre is an independent economic and social research organisation located within the Curtin Business School at Curtin University. The Centre was established in 2012 through the generous support from Bankwest (a division of the Commonwealth Bank of Australia), with a core mission to undertake high quality, objective research on the key economic and social issues of relevance to Western Australia.

The Centre's research and engagement activities are designed to influence economic and social policy debates in state and Federal Parliament, regional and national media, and the wider Australian community. Through high quality, evidence-based research and analysis, our research outcomes inform policy makers and commentators of the economic challenges to achieving sustainable and equitable growth and prosperity both in Western Australia and nationally.

The Centre capitalises on Curtin University's reputation for excellence in economic modelling, forecasting, public policy research, trade and industrial economics and spatial sciences. Centre researchers have specific expertise in economic forecasting, quantitative modelling, microdata analysis and economic and social policy evaluation.

A suite of tailored and national economic models and methods are maintained within the Centre to facilitate advanced economic policy analysis: these include macroeconomic and time series models, micro(simulation) models, computable general equilibrium (CGE) models, spatial modelling methods, economic index analysis, and behavioural modelling methods.

CONTACT

Bankwest Curtin Economics Centre
Curtin University
Kent Street Bentley WA 6102
GPO Box U1987 Perth WA 6845

Tel: +61 8 9266 2873

bankwesteconomicscentre@curtin.edu.au
business.curtin.edu.au/bcec