Forecasting with EC-VARMA Models

George Athanasopoulos
Monash University

Don Poskitt
Monash University

Farshid Vahid
Monash University

Wenying YAO
University of Tasmania

Forecasting with EC-VARMA Models[*]

George Athanasopoulos[♯], Don Poskitt[♯], Farshid Vahid[♯] and Wenying Yao[‡]

♯ Monash University
‡ University of Tasmania

February 22, 2014

Abstract

This article studies error correction vector autoregressive moving average (EC-VARMA) models. A complete procedure for identifying and estimating EC-VARMA models is proposed. The cointegrating rank is estimated in the first stage using an extension of the non-parametric method of Poskitt (2000). Then, the structure of the VARMA model for variables in levels is identified using the scalar component model (SCM) methodology developed in Athanasopoulos and Vahid (2008), which leads to a uniquely identifiable VARMA model. In the last stage, the VARMA model is estimated in its error correction form. Monte Carlo simulation is conducted using a 3-dimensional VARMA(1,1) DGP with cointegrating rank 1, in order to evaluate the forecasting performances of the EC-VARMA models. This algorithm is illustrated further using an empirical example of the term structure of U.S. interest rates. The results reveal that the out-of-sample forecasts of the EC-VARMA model are superior to those produced by error correction vector autoregressions (VARs) of finite order, especially in short horizons.

Keywords: cointegration, VARMA model, iterative OLS, scalar component model.

JEL: C1, C32, C53

---

[*]Corresponding author: Wenying Yao, Tasmanian School of Business and Economics, Private Bag 84, University of Tasmania, Hobart TAS 7001, Australia, Email: Wenying.Yao@utas.edu.au, Tel: +61 3 62267141, Fax: +61 3 62267587.

1 Introduction

Cointegration refers to situations where several $I(1)$ variables share at least one common stochastic trend. The Granger Representation Theorem (Engle and Granger, 1987) states that all cointegrated time series have a vector error correction representation. Since most studies on cointegration are set within the context of finite vector autoregressive (VAR) models, the error correction VARs are commonly referred to as vector error correction models (VECMs). However, the Granger Representation Theorem allows for the time series of interest to have vector autoregressive moving average (VARMA) dynamics. In this paper we provide a methodology for the identification and estimation of the error correction VARMA models. While we could legitimately call such models VECMs as well, we refer to them as EC-VARMA models, and use the term "VECM" exclusively for the error correction VARs of finite order throughout the paper.

The literature on EC-VARMA models is quite limited. Kascha and Trenkler (2011) generalize the final moving average (FMA) representation proposed by Dufour and Pelletier (2008) to cointegrated VARMA models, and use an information criterion to choose the AR and MA orders for the cointegrated VARMA model in levels. They find promising results relative to multivariate random walk and standard VECM for predicting U.S. interest rates. Their specification strategy is simpler by using the FMA representation, but similar to Dufour and Pelletier (2008), they focus on a special subset of VARMA models which the MA operator is scalar. Furthermore, the rank of the cointegration space is taken as given. These limitations restrict the applicability of their methodology to empirical analyses.

Lütkepohl and Claessen (1997) consider a four variable EC-VARMA model for U.S. money demand. They find that in general the EC-VARMA model substantially outperforms the VECM in terms of mean squared errors and mean absolute errors. They also examine a restricted version of the EC-VARMA model by dropping the insignificant parameters, which leads to an even better forecasting performance. Poskitt (2003) uses a six variable model with U.S. macroeconomic data to illustrate the Echelon form EC-VARMA model. He observes an improvement in the forecasting performance of an EC-VARMA model over a VECM and a VAR in levels. He also points out that:

> *"The acquisition of additional hands on experience with EC-ARMA$_E$ forecasting systems would be desirable in order to gain further insight into their practical merits and possible pitfalls."*[1]

This is precisely what we pursue in this paper. By proposing a complete algorithm for identifying and estimating EC-VARMA models, we expect these models to be utilized more broadly in macroeconomic modelling and forecasting.

The first stage of the proposed algorithm determines the number of cointegration relationships. The usual practice is to use the Johansen procedure (Johansen, 1988, 1991,

---

[1]EC-ARMA$_E$ stands for the error correction, Echelon canonical form VARMA model.

1995) for this stage. In particular, Lütkepohl and Claessen (1997) use a likelihood ratio (LR) type test which is based on the ideas of Johansen (1988). The Johansen procedure was originally developed as a likelihood-based method, assuming that variables have a finite VAR representation. Although this can be justified as a valid method for infinite VARs under certain specific assumptions (see e.g. Lütkepohl and Saikkonen, 1999),[2] we show in this paper that it suffers from size and power distortions in such situations.[3] Therefore, we extend the non-parametric approach of Poskitt (2000) to choose the cointegrating rank for any VARMA process. This selection procedure is strongly consistent, and does not require any assumptions about the functional form of the underlying data generating mechanism.

The second stage identifies a canonical VARMA model for variables in levels. We use the scalar component model (SCM) methodology which was originally proposed by Tiao and Tsay (1989) and further developed by Athanasopoulos and Vahid (2008). One of the main contribution of this paper is that we establish the validity of the *SCM* methodology for non-stationary *VARMA* models. Specifically, we show that the *SCM* methodology can be applied to partially non-stationary time series in exactly the same way as to stationary time series.[4]

The third stage of the proposed algorithm puts the uniquely specified VARMA model into an error correction form that imposes the cointegrating rank restriction, and then estimates this model using full information maximum likelihood (FIML).

We use Monte Carlo simulation to evaluate the finite sample performance of the extended Poskitt's procedure of selecting the cointegrating rank. We also examine the predictive ability of EC-VARMA models and VECMs when forecasting data generated from an EC-VARMA DGP. The computational demands of maximum likelihood estimation are impractical for Monte Carlo simulations, so we replace FIML estimation with iterative OLS (IOLS) suggested by Kapetanios (2003), which we extend to EC-VARMA models. The proposed algorithm is applied to the modelling of the term structure of U.S. interest rates. We find that the EC-VARMA models produce forecasts that are superior to those produced by VECMs, especially in short horizons.

The remainder of this paper is organized as follows. Section 2 defines the notation used in this paper. We propose the estimation algorithm for the EC-VARMA model in Section 3. The Monte Carlo simulation is conducted in Section 4 in order to examine the loss in forecasting accuracy from using VECMs, when the true DGP is an EC-VARMA process. Section 5 presents an empirical application to forecasting the term structure of interest rates using the EC-VARMA models. Section 6 concludes.

---

[2]Lütkepohl and Saikkonen (1999) show that as long as a consistent model selection criterion is used to choose the AR lag order in the VAR model, the asymptotic distribution of the LR test statistic for the cointegration rank remains valid even if the true DGP is of infinite order.

[3]Lütkepohl and Saikkonen (1999) also find that the small sample properties of the cointegration tests are strongly dependent on the choice of the AR lag length.

[4]Alternatively, the canonical Echelon form VARMA model (Poskitt, 1992; Lütkepohl and Poskitt, 1996) or reverse Echelon form Lütkepohl and Claessen (1997) can also be applied in order to obtain a uniquely identifiable VARMA structure.

2 Notation

The general form of a VARMA(p,q) process is

$$\boldsymbol{\Phi}_0\boldsymbol{y}_t = \boldsymbol{\Phi}_1\boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Phi}_p\boldsymbol{y}_{t-p} + \boldsymbol{\Theta}_0\boldsymbol{u}_t + \boldsymbol{\Theta}_1\boldsymbol{u}_{t-1} + \cdots + \boldsymbol{\Theta}_q\boldsymbol{u}_{t-q}, \tag{1}$$

where $\boldsymbol{y}_t$ is a K-dimensional time series, $\boldsymbol{\Phi}_i$ and $\boldsymbol{\Theta}_j$ are $K \times K$ matrices, $i = 0, 1, \ldots, p$, $j = 0, 1, \ldots, q$, and $\boldsymbol{u}_t$ is a K-dimensional vector of *i.i.d.* Gaussian white noise process with mean zero and nonsingular covariance matrix $\Sigma = \mathbb{E}(\boldsymbol{u}_t\boldsymbol{u}_t')$. $\boldsymbol{\Phi}_0$ and $\boldsymbol{\Theta}_0$ are nonsingular, and hence, can be normalized to identity matrices without the loss of generality. Alternatively, the process in equation (1) can be written as

$$\boldsymbol{\Phi}(L)\,\boldsymbol{y}_t = \boldsymbol{\Theta}(L)\,\boldsymbol{u}_t, \tag{1'}$$

where $\boldsymbol{\Phi}(L) = \boldsymbol{\Phi}_0 - \boldsymbol{\Phi}_1 L - \cdots - \boldsymbol{\Phi}_p L^p$, $\boldsymbol{\Theta}(L) = \boldsymbol{\Theta}_0 + \boldsymbol{\Theta}_1 L + \cdots + \boldsymbol{\Theta}_q L^q$, and L is the lag operator, such that $L\,\boldsymbol{y}_t = \boldsymbol{y}_{t-1}$. The matrix polynomials satisfy

$$\det \boldsymbol{\Phi}(z) \neq 0 \quad |z| \leq 1, z \neq 1, \qquad \text{and} \qquad \det \boldsymbol{\Theta}(z) \neq 0 \quad |z| \leq 1. \tag{2}$$

We allow for the AR operator $\boldsymbol{\Phi}(z)$ to have roots at $z = 1$, to account for the integrated and cointegrated components of $\boldsymbol{y}_t$. Each individual time series in $\boldsymbol{y}_t$ is at most $I(1)$. The possibility that some elements in $\boldsymbol{y}_t$ may be stationary without first differencing is not excluded.

Notice that there are no deterministic terms in the underlying true DGP. This assumption is retained to simplify the exposition of our proposed algorithm, although empirical applications usually require the deterministic terms. In general, our algorithm will not be affected if the deterministic terms are pre-filtered.

We obtain the EC-VARMA form representation from equation (1) by subtracting $\boldsymbol{\Phi}_0\boldsymbol{y}_{t-1}$ from both sides:

$$\boldsymbol{\Phi}_0\Delta\boldsymbol{y}_t = \boldsymbol{\Pi}\boldsymbol{y}_{t-1} + \boldsymbol{\Psi}_1\Delta\boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Psi}_{p-1}\Delta\boldsymbol{y}_{t-p+1} + \boldsymbol{\Theta}_0\boldsymbol{u}_t + \boldsymbol{\Theta}_1\boldsymbol{u}_{t-1} + \cdots + \boldsymbol{\Theta}_q\boldsymbol{u}_{t-q}, \tag{3}$$

where $\boldsymbol{\Pi} = -(\boldsymbol{\Phi}_0 - \boldsymbol{\Phi}_1 - \cdots - \boldsymbol{\Phi}_p)$, and $\boldsymbol{\Psi}_i = -(\boldsymbol{\Phi}_{i+1} + \cdots + \boldsymbol{\Phi}_p)$ for $i = 1, \ldots, p-1$. Denote the true cointegrating rank by ρ_0, *i.e.* there exist ρ_0 linear combinations of the components in $\boldsymbol{y}_t$ that are stationary, which imposes the restriction that $\text{rank}(\boldsymbol{\Pi}) = \rho_0$. Hence, $\boldsymbol{\Pi}$ can be decomposed into $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$, where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are both matrices of dimension $K \times \rho_0$ of full column rank. $\boldsymbol{\beta}$ represents the cointegrating relationships in $\boldsymbol{y}_t$ (Granger, 1981; Engle and Granger, 1987).

We use the Johansen procedure (Johansen, 1988, 1991, 1995) to test the cointegrating rank for the VECMs. The hypothesis used in this paper is

$$H_0 : \rho \leq r \quad \text{or} \quad \text{rank}(\boldsymbol{\Pi}) \leq r. \tag{4}$$

4

The LR test statistic for the above hypothesis is

$$-T \sum_{i=r+1}^{K} \ln(1 - \hat{\lambda}_i), \qquad (5)$$

where $\hat{\lambda}_{r+1}$, ..., $\hat{\lambda}_K$ are the $K - r$ smallest sample squared partial canonical correlations between Δy_t and y_{t-1} after the effects of lagged differences and a constant term have been removed.

The three stage algorithm proposed in this paper is based on a realization of time series with sample size T, $\{ y_1, y_2, \ldots, y_T \}$, generated from equation (1). For ease of notation, we use y_t to denote both the underlying data generating mechanism, and the realization generated from the true DGP. $\{ y_1, y_2, \ldots, y_T \}$ and $\{ y_t \}_{t=1}^{T}$ are used interchangeably throughout the paper. Further, symbols of the parameters in the true DGP are also used to denote the unknown coefficients in the model to be estimated.

3 A Proposed Algorithm for Estimating an EC-VARMA Model

Given $\{ y_1, y_2, \ldots, y_T \}$, the algorithm for identifying and estimating an EC-VARMA model consists of the following three stages. The details of each stage are presented in Sections 3.1, 3.2 and 3.3 respectively.

Stage 1 Use an extended version of the non-parametric approach of Poskitt (2000) to obtain a super-consistent estimate of the cointegrating rank ρ. We extend the selection procedure of Poskitt (2000) to include the possibility that $\rho = K$, *i.e.*, $\{ y_1, y_2, \ldots, y_T \}$ is stationary in levels, in which case a VARMA model for variables in levels should be estimated.

Stage 2 Identify a canonical SCM VARMA(p, q) representation for y_t in levels, which takes the form of equation (1)

$$\mathbf{\Phi}_0 y_t = \mathbf{\Phi}_1 y_{t-1} + \cdots + \mathbf{\Phi}_p y_{t-p} + u_t + \mathbf{\Theta}_1 u_{t-1} + \cdots + \mathbf{\Theta}_q u_{t-q}. \qquad (1)$$

$\mathbf{\Phi}_0$ is a nonsingular matrix with unit diagonal elements. We use the SCM methodology (Athanasopoulos and Vahid, 2008) in this stage. This method imposes normalization restrictions and zero restrictions on the coefficients in $\mathbf{\Phi}_0, \ldots, \mathbf{\Phi}_p$ and $\mathbf{\Theta}_1, \ldots, \mathbf{\Theta}_q$, in order to achieve unique identification of a VARMA(p,q) model. The coefficient of u_t, $\mathbf{\Theta}_0$ is normalized to identity.

Stage 3 Estimate an EC-VARMA model in the form of equation (3)

$$\mathbf{\Phi}_0 \Delta y_t = \mathbf{\Pi} y_{t-1} + \mathbf{\Psi}_1 \Delta y_{t-1} + \ldots + \mathbf{\Psi}_{p-1} \Delta y_{t-p+1}$$
$$+ u_t + \mathbf{\Theta}_1 u_{t-1} + \cdots + \mathbf{\Theta}_q u_{t-q} \qquad (3)$$

5

using FIML. Both the estimated cointegrating rank from stage 1 and the coefficient constraints in the canonical VARMA representation from stage 2 will carry over to the error correction form in equation (3).

3.1 Stage 1: Determining the Cointegrating Rank

The usual Johansen procedure for testing the cointegrating rank in the context of VECMs will be disadvantageous for EC-VARMA models due to the presence of the moving average component. Lütkepohl and Saikkonen (1999) show that the asymptotic distribution of the test statistic in Johansen's sequential LR test for cointegrating rank remains unchanged even if the true DGP is an infinite order VAR, or equivalently, a VARMA process. However, the power and size of the test in finite samples depend crucially on the choice of the lag length. The lag length of the truncated VAR should be chosen using a consistent information criterion, which ensures that the chosen lag length goes to infinity with the sample size. Unfortunately this condition cannot be satisfied in practice. In particular, the available sample size is rather limited for applied macroeconomic research, which is typically less than 400 observations. In such circumstances, the empirical distribution of the LR test statistic is likely to be distant from its asymptotic distribution. Hence, we resort to alternative methods that have better finite sample properties. In this paper we use the non-parametric method of Poskitt (2000) to determine the cointegration rank.

Lütkepohl and Poskitt (1998) and Gonzalo and Pitarakis (1995) point out that the statistics for testing the cointegration rank of a multivariate system can be used to construct model selection criteria for estimating the rank consistently. The Poskitt (2000) method that we adopt here is such a model selection procedure, built on a canonical correlation based testing procedure which was first proposed by Yang and Bewley (1996). This method does not require that the true DGP is a finite order VAR process. In fact it does not make any assumptions about the short run dynamics due to its non-parametric nature.

One of the advantages of using a model selection procedures is that the probability that this method chooses the correct cointegrating rank, $\hat{\rho} = \rho_0$, converges to 1 as the sample size goes to infinity. In contrast, the performances of hypothesis testing type methods are bounded by $1 - \alpha$, where α is the test size.

Given a sample of T observations $\{y_t\}_{t=1}^T$, denote the sample squared canonical correlations between y_t and y_{t-1} (both in levels), in ascending order, as

$$\lambda_{(1),T} \leq \lambda_{(2),T} \leq \cdots \leq \lambda_{(K),T}. \tag{6}$$

For $\rho = 0, \ldots, K - 1$, let $\Lambda_T(\rho)$ be the ratio of the arithmetic to the geometric mean of

the $K - \rho$ largest squared canonical correlations, $\Lambda_T(\rho) = \bar{\lambda}_{\rho,T} / \bar{\lambda}^g_{\rho,T}$, where

$$\bar{\lambda}_{\rho,T} = (K - \rho)^{-1} \sum_{i=\rho+1}^{K} \lambda_{(i),T}, \qquad \text{and} \qquad \bar{\lambda}^g_{\rho,T} = \left(\prod_{i=\rho+1}^{K} \lambda_{(i),T} \right)^{1/(K-\rho)}. \qquad (7)$$

We choose the cointegrating rank $\hat{\rho}$ to be the one that minimizes the following criterion function:

$$\zeta_T(\rho) = T(K - \rho) \ln(\Lambda_T(\rho)) + \rho(2K - \rho + 1)P_T/2, \qquad \text{for} \quad \rho = 0, \ldots, K - 1. \qquad (8)$$

The choice of the penalty term P_T in equation (8) should satisfy the following conditions (see Poskitt, 2000, Theorem 1.2):

$$\lim_{T \to \infty} P_T/T = 0, \qquad \text{and} \qquad \lim_{T \to \infty} \ln(\ln T)/P_T = 0. \qquad (9)$$

Under condition (9), the value of $\hat{\rho}$ that minimizes equation (8) will converge to the true cointegrating rank ρ_0 with probability 1 under certain regularity conditions. We set $P_T = \ln T$ throughout the paper.

Notice that in the construction of the criteria in equation (7), ρ cannot take the value K, to ensure that there are some $\lambda_{(i),T}$ to calculate $\Lambda_T(\rho)$. Hence the primary drawback of this selection criterion is that it rules out the possibility that $\rho = K$, *i.e.*, all of the individual components in the process y_t are $I(0)$ series. Although in practice, one always starts with univariate unit root tests, and no one would consider cointegration when all series are $I(0)$, we extend the selection procedure to deal with this scenario for completeness, using the same rationale as the Poskitt (2000) method. The following lemma is utilized in constructing this selection criterion (see Poskitt, 2000, Lemma 1.1).

Lemma 1 *Let $\lambda_{(i),T}$, $i = 1, \ldots, K$ be the ordered sample squared canonical correlations in equation (6), and denote their population counterparts by $\lambda_{(1)} \leq \lambda_{(2)} \leq \cdots \leq \lambda_{(K)}$. Then, with probability 1,*

$$\lambda_{(i),T} = \lambda_{(i)} + O\left(\left(\frac{\ln T}{T}\right)^{1/2}\right), \qquad \text{for} \quad i = 1, \ldots, \rho_0,$$

$$\lambda_{(i),T} = 1 + O\left(\left(\frac{\ln \ln T}{T}\right)^{1/2}\right), \qquad \text{for} \quad i = \rho_0 + 1, \ldots, K,$$

where $0 \leq \lambda_{(i)} < 1$, for $i = 1, \ldots, \rho_0$ when the cointegrating rank in the true DGP is ρ_0.

A significant consequence of Lemma 1 is that for large values of T, $\lambda_{(\rho_0+1),T}$, $\ldots$, $\lambda_{(K),T}$ can be arbitrarily close to unity, while $\lambda_{(1),T}, \ldots, \lambda_{(\rho_0),T}$ are strictly less than unity. Hence, to decide whether $\rho = K$, we can simply take the largest squared canonical

7

correlation $\lambda_{(K),T}$, and compare it to $1 - C(\ln T/T)^{1/2}$. The decision rule is

$$\hat{\rho} = K \qquad \text{if} \qquad \lambda_{(K),T} \leq 1 - C\left(\frac{\ln T}{T}\right)^{1/2}, \qquad (10)$$

where C is some positive constant. We choose $C = 1$ for ease of exposition. According to Lemma 1, in the situation when $\rho_0 = K$, the criterion in equation (10) will choose $\hat{\rho} = \rho_0$ with probability 1. This criterion is designed to be an extra step of the original selection criterion of Poskitt (2000). They can be used in combination, as specified in the following steps:

Step 1 For a given sample of K-dimensional time series $\{y_t\}_{t=1}^T$, we first determine the sample squared canonical correlations between y_t and y_{t-1}, in ascending order, as $\lambda_{(1),T} \leq \lambda_{(2),T} \leq \cdots \leq \lambda_{(K),T}$.

Step 2 Compare $\lambda_{(K),T}$ to $1 - (\ln T/T)^{1/2}$. If $\lambda_{(K),T} \leq 1 - (\ln T/T)^{1/2}$, let $\hat{\rho} = K$. Otherwise, go to step 3.

Step 3 Construct the criterion in equation (8), and choose the cointegrating rank $\hat{\rho}$ such that

$$\hat{\rho} = arg \min_{\rho \in \{0,1,\dots,K-1\}} \zeta_T(\rho). \qquad (11)$$

This procedure for selecting the cointegration rank is not confined to the class of VARMA models due to its non-parametric nature. This is preferable from both the theoretical and practical perspectives, because it allows us to determine the cointegrating rank consistently without specifying the form of the short run dynamics.[5]

3.2 Stage 2: Specifying the VARMA Model in Levels

The scalar component model (SCM) is initiated by Tiao and Tsay (1989) and further developed by Athanasopoulos and Vahid (2008). We adopt this methodology in this paper, because the canonical SCM VARMA representation is generally more parsimonious than the canonical Echelon form (Hannan and Kavalieris, 1984; Hannan and Deistler, 1988; Poskitt, 1992; Lütkepohl and Poskitt, 1996; Lütkepohl, 2005) by allowing for different AR and MA orders in each row. More importantly, it will become clear later that the SCM methodology shares the same theoretical foundation with Poskitt's method of selecting the cointegrating rank. This section demonstrates that the testing procedure of the SCM methodology is still valid for nonstationary VARMA models.

---

[5]There are also other system cointegration tests that use a VAR as an adjustment for short run dynamics, in order to eliminate the effect of the unknown nuisance parameters (e.g. the principal components test of Stock and Watson, 1988). Generalizing from a VAR to a VARMA adjustment may potentially improve the performances of such tests as well, but we do not explore this possibility here.

3.2.1 Canonical Correlations Framework

We first present the SCM methodology for a stationary and ergodic process y_t. Extension to nonstationary systems is discussed in Section 3.2.2 and onwards.

Definition 1 (SCM) *For a K-dimensional process y_t, a non-zero linear combination $z_{i,t} = a'_{(i)}y_t$ is said to follow an SCM(p_i,q_i) structure if there exist p_i K-dimensional vectors $v_1,\ldots,v_{p_i}$ where $v_{p_i} \neq 0$, such that the linear combination*

$$\xi_{i,t} = a'_{(i)}y_t + \sum_{s=1}^{p_i} v'_s y_{t-s} \tag{12}$$

satisfies the condition

$$\mathbb{E}(y_{t-j}\xi_{i,t}) \begin{cases} \neq 0 & \text{if } j = q_i \\ = 0 & \text{if } j > q_i \end{cases}. \tag{13}$$

It follows from equation (13) that $z_{i,t} = a'_{(i)}y_t$ has an SCM(p_i,q_i) structure, with $v_1,\ldots,v_{p_i}$ being the associated vectors, if and only if

$$\mathbb{E}\left(y_{t-j}[y'_t a_{(i)} + \sum_{s=1}^{p_i} y'_{t-s}v_s]\right) = 0 \qquad \text{for } j > q_i. \tag{14}$$

Let $\Gamma_j = \mathbb{E}(y_{t-j}y'_t)$ be the j-th lag autocovariance matrix of y_t. Then equation (14) becomes

$$\Gamma_j a_{(i)} + \sum_{s=1}^{p_i} \Gamma_{j-s}v_s = 0 \qquad \text{for } j > q_i. \tag{15}$$

We construct the following vectors and matrices to utilize the relationship in equation (15) in the specification procedure. First, let $v = \left(a'_{(i)}, \ v'_1, \ \ldots, v'_{p_i}\right)'$. For integers $m \geq 0$ and $l \geq 0$, denote the two $K(m+1)$-dimensional vectors $Y_{m,t}$ and $Y_{m,t-l-1}$ by

$$Y_{m,t} = \left(y'_t, \ \cdots, \ y'_{t-m}\right)', \tag{16}$$

$$\text{and} \quad Y_{m,t-l-1} = \left(y'_{t-l-1}, \ \cdots, \ y'_{t-l-1-m}\right)', \tag{17}$$

where t and $t-l-1$ denote the starting points and m denotes the number of lags. It will become clear later that m controls the order of the autoregressive component of the underlying SCM, and l controls the order of the moving average component.

Then, let the $K(m+1)$-dimensional square matrix

$$\Gamma(m,l) = \mathbb{E}(Y_{m,t-l-1}Y'_{m,t}) = \begin{pmatrix} \Gamma_{l+1} & \Gamma_l & \cdots & \Gamma_{l+1-m} \\ \Gamma_{l+2} & \Gamma_{l+1} & \cdots & \Gamma_{l+2-m} \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_{l+1+m} & \Gamma_{l+m} & \cdots & \Gamma_{l+1} \end{pmatrix} \tag{18}$$

be the covariance matrix of $Y_{m,t}$ and $Y_{m,t-l-1}$. Combining equations (15) and (18), there exists an SCM(p_i,q_i), if and only if

$$\Gamma(p_i, q_i)\, v = \begin{pmatrix} \Gamma_{q_i+1} & \Gamma_{q_i} & \cdots & \Gamma_{q_i+1-p_i} \\ \Gamma_{q_i+2} & \Gamma_{q_i+1} & \cdots & \Gamma_{q_i+2-p_i} \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_{q_i+1+p_i} & \Gamma_{q_i+p_i} & \cdots & \Gamma_{q_i+1} \end{pmatrix} \begin{pmatrix} a_{(i)} \\ v_1 \\ \vdots \\ v_{p_i} \end{pmatrix} = \mathbf{0}. \tag{19}$$

Such a vector v is called a right vector of $\Gamma(p_i, q_i)$ corresponding to zero.

Definition 2 (Right Vector) *For a real matrix A of dimension $m \times n$, $m \geq n$, the n-dimensional non-zero vector x is a right vector of A corresponding to zero if $Ax = \mathbf{0}$. Further, $\text{rank}(A) = n - r$, where r is the number of linearly independent right vectors of A.*

According to Definition 2, if $\text{rank}[\Gamma(p_i, q_i)] = K(p_i + 1) - r$, there exist r linearly independent vectors $a_{(1)}, \ldots, a_{(r)}$, such that the linear transformations $z_{j,t} = a'_{(j)} y_t \sim$ SCM(p_i,q_i), $j = 1, \ldots, r$. Hence, we can design a procedure to test the rank of the series of $\Gamma(m, l)$, or equivalently, the number of non-zero eigenvalues of $\Gamma(m, l)$, $m = 0, 1, \ldots, p$ and $l = 0, 1, \ldots, q$, in order to detect the number of SCMs.

In practice, we estimate the rank of $\Gamma(m, l)$ within the canonical correlation framework. Consider the matrix

$$A(m, l) = \left[\mathbb{E}\left(Y_{m,t} Y'_{m,t}\right) \right]^{-1} \Gamma'(m, l) \left[\mathbb{E}\left(Y_{m,t-l-1} Y'_{m,t-l-1}\right) \right]^{-1} \Gamma(m, l). \tag{20}$$

The eigenvalues of $A(m, l)$ are the squared population canonical correlations of $Y_{m,t}$ and $Y_{m,t-l-1}$ (Anderson, 2003). Furthermore, assuming that u_t, the innovation process has a nonsingular covariance matrix Σ, the multiplicity of zero eigenvalues of $A(m, l)$ is the same as the number of linearly independent right vectors of $\Gamma(m, l)$ corresponding to zero, and the linearly independent right eigenvectors of $A(m, l)$ corresponding to the zero eigenvalues are the right vectors of $\Gamma(m, l)$. Hence testing for SCM(p_i,q_i) is equivalent to testing for zero eigenvalues in $A(p_i, q_i)$.

We can decompose $A(m, l) = \beta^*(m, l) \beta(m, l)$, where

$$\beta^*(m, l) = \left[\mathbb{E}\left(Y_{m,t} Y'_{m,t}\right) \right]^{-1} \Gamma'(m, l), \tag{21}$$

$$\beta(m, l) = \left[\mathbb{E}\left(Y_{m,t-l-1} Y'_{m,t-l-1}\right) \right]^{-1} \Gamma(m, l). \tag{22}$$

Note that $\beta^*(m, l)$ and $\beta(m, l)$ are the probability limits of the OLS estimators $\hat{\beta}^*(m, l)$ and $\hat{\beta}(m, l)$, which are the regression coefficients of the following forward and backward autoregressions:

$$Y'_{m,t-l-1} = Y'_{m,t}\, \beta^*(m, l) + e^*_t, \tag{23}$$

$$Y'_{m,t} = Y'_{m,t-l-1}\, \beta(m, l) + e_t. \tag{24}$$

10

For a given finite sample of size T, the sample counterparts are $\hat{A}(m,l) = \hat{\beta}^*(m,l)\hat{\beta}(m,l)$, where $\hat{\beta}^*(m,l)$ and $\hat{\beta}(m,l)$ are the OLS estimates of equations (23) and (24). The sequence of hypothesis tests are constructed using the series of $\hat{A}(m,l)$ for different values of m and l.

Denote the ordered eigenvalues of $\hat{A}(m,l)$, *i.e.* the ordered sample squared canonical correlations between $Y_{m,t}$ and $Y_{m,t-l-1}$, as

$$\hat{\lambda}_1(m,l) \le \hat{\lambda}_2(m,l) \le \cdots \le \hat{\lambda}_{K(m+1)}(m,l). \tag{25}$$

Generally the null and alternative hypotheses that test for r zero eigenvalues in $\hat{A}(m,l)$ are as follows:

$$\boldsymbol{H}_0 \; : \; \hat{\lambda}_r(m,l) = 0; \qquad \text{against} \qquad \boldsymbol{H}_1 \; : \; \hat{\lambda}_r(m,l) \ne 0. \tag{26}$$

The test statistic is

$$C(m,l) = -(T - m - l) \sum_{j=1}^{r} \ln\left\{1 - \frac{\hat{\lambda}_j(m,l)}{d_j(m,l)}\right\} \stackrel{a}{\sim} \chi_{r^2}^2 \quad \text{under} \quad \boldsymbol{H}_0, \tag{27}$$

where

$$d_j(m,l) = 1 + 2 \sum_{i=1}^{l} \hat{\rho}_i(\hat{f}_i' Y_{m,t}) \hat{\rho}_i(\hat{g}_i' Y_{m,t-1-l}); \tag{28}$$

$\hat{\rho}_i(x_t)$ is the i-th lag sample autocorrelation of the process x_t; and $\hat{f}_i$ and $\hat{g}_i$ are the left and right canonical covariates corresponding to the eigenvalue $\hat{\lambda}_j(m,l)$.

3.2.2 Nonstationary Environment

Given a sample of T observations, we denote the sample counterpart of $A(m,l)$ by

$$\hat{A}(m,l) = \left(\frac{1}{T}\sum_{t=1}^{T} Y_{m,t} Y_{m,t}'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} Y_{m,t} Y_{m,t-l-1}'\right)$$
$$\left(\frac{1}{T}\sum_{t=1}^{T} Y_{m,t-l-1} Y_{m,t-l-1}'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} Y_{m,t-l-1} Y_{m,t}'\right). \tag{29}$$

We can now turn to investigate the nonstationary situation. If y_t has some nonstationary components, $\mathbb{E}(y_{t-j}y_t')$ depends on both t and j, and hence a time invariant $\Gamma(m,l)$ does not exist. Rather than solving the eigenvalue-eigenvector problem of $\hat{A}(m,l)$ directly, we consider a different normalization of the process y_t.

For a time series y_t that comes from a VARMA(p,q) process, there exists a $K \times K$ nonsingular transformation matrix H such that $Hy_t = (\, s_t' \;\; n_t' \,)'$, where s_t is a purely stationary process. Given the cointegrating relationship in y_t, s_t has dimension $\rho_0 \times 1$. Thus, the difference stationary component n_t has dimension $(K - \rho_0) \times 1$ (see Poskitt, 2000, for an example of such a transformation). For any scalar of m, let I_m denote the

11

$m \times m$ identity matrix; then let

$$x_t = G_T H y_t = \begin{pmatrix} s_t \\ n_t/T^{1/2} \end{pmatrix}, \qquad \text{where} \quad G_T = \begin{pmatrix} I_{\rho_0} & 0 \\ 0 & I_{K-\rho_0}/T^{1/2} \end{pmatrix}. \tag{30}$$

Consider the sample moment matrices for y_t and the normalized process x_t as defined above:

$$\hat{\Gamma}_{i,j}^y = \frac{1}{T} \sum_{t=1}^{T} y_{t-i} y_{t-j}', \tag{31a}$$

$$\hat{\Gamma}_{i,j}^x = \frac{1}{T} \sum_{t=1}^{T} x_{t-i} x_{t-j}' = G_T H \hat{\Gamma}_{i,j}^y H' G_T', \tag{31b}$$

where $\hat{\Gamma}_{i,j}^x$ has a well defined probability limit for any finite numbers i, j, denoted by $\Gamma_{i,j}^x := \text{plim}_{T \to \infty} \hat{\Gamma}_{i,j}^x$; see Hamilton (1994), Proposition 18.1, or Phillips and Durlauf (1986). We can construct $\hat{A}_x(m, l)$ in the same way as in equation (29), replacing $Y_{m,t}$ and $Y_{m,t-l-1}$ with $X_{m,t}$ and $X_{m,t-l-1}$.

3.2.3 Testing for SCM(0,0)

First consider the case of testing for SCM(0,0), *i.e.* $m = l = 0$. By making use of the relationship in equation (31), $\hat{A}_x(0,0)$ can be expressed as

$$\hat{A}_x(0,0) = (\hat{\Gamma}_{0,0}^x)^{-1} \hat{\Gamma}_{0,1}^x (\hat{\Gamma}_{1,1}^x)^{-1} \hat{\Gamma}_{1,0}^x = (H' G_T')^{-1} \hat{A}(0,0) (H' G_T'). \tag{32}$$

If the eigendecomposition of $\hat{A}(0,0)$ is $\hat{A}(0,0) = F \Lambda F^{-1}$, where Λ is a diagonal matrix with the eigenvalues of $\hat{A}(0,0)$ being the diagonal elements, and the i-th column of F is the eigenvector corresponding to $[\Lambda]_{ii}$, then the eigendecomposition of $\hat{A}_x(0,0)$ is

$$\hat{A}_x(0,0) = F_x \Lambda F_x^{-1}, \qquad \text{where} \quad F_x = (H' G_T')^{-1} F. \tag{33}$$

Hence, the eigenvalues of $\hat{A}(0,0)$ are the same as those of $\hat{A}_x(0,0)$. Notice that $\hat{A}_x(0,0)$ also has a well defined probability limit. Due to the fact that eigenvalues are continuous functions of the matrix, we only examine the eigenvalues of the limiting matrix $A_x(0,0) := \text{plim}_{T \to \infty} \hat{A}_x(0,0)$.

We analyze the rank property of $A_x(0,0)$ in order to detect SCM(0,0) for the original process y_t. Using the notation for stationary VARMA process in Section 3.2.1, there exist r linearly independent vectors $a_{(1)}, \ldots, a_{(r)}$, such that $z_{j,t} = a_{(j)}' y_t \sim SCM(0,0)$, $j = 1, \ldots, r$, if and only if $\text{rank}[A_x(0,0)] = K - r$, or equivalently, the multiplicity of the zero eigenvalues of $A_x(0,0)$ is r. The eigenvalues of $A_x(0,0)$ are the squared population canonical correlations between x_t and x_{t-1}, which are the same as the squared population canonical correlations between y_t and y_{t-1}.

Based on the result of Lemma 1, the ordered eigenvalues of $A_x(0,0)$, $\lambda_{(1)} \leq \lambda_{(2)} \leq$

$\cdots \leq \lambda_{(K)}$, can be classified into two groups in the following fashion. When the true cointegrating rank of y_t is ρ_0,

$$\lambda_{(\rho_0+1)} = \cdots = \lambda_{(K)} = 1, \tag{34}$$

are the squared canonical correlations between $n_t/T^{1/2}$ and $n_{t-1}/T^{1/2}$; and

$$\lambda_{(1)} \leq \cdots \leq \lambda_{(\rho_0)} < 1, \tag{35}$$

are the squared canonical correlations between s_t and s_{t-1}.[6] Hence, the multiplicity of zero eigenvalues and the rank of $A_x(0,0)$ are determined by s_t, the stationary components of y_t, simply because the canonical correlations between the normalized nonstationary components are unity in the limit. The true cointegrating rank ρ_0 implies that the dimension of the stationary component, and hence the number of SCM(0,0) in y_t should not exceed ρ_0. These parameter restrictions will be discussed further in Section 3.3.1.

Above analysis shows that if there is some nonstationary component in y_t, the behavior of the canonical correlations are the same as in the stationary case, when the sample size T goes infinity. However, the SCM methodology is valid in this case only if the test statistic retains the same distribution and degree of freedom. In what follows we establish the normality of the canonical correlations that are used in the construction of the test statistic in equation (27), and then examine the degree of freedom of the test statistic.

We first examine the asymptotic distributions of $\lambda_{(1)} \leq \cdots \leq \lambda_{(\rho_0)}$ when the non-stationary component n_t exists. Since s_t is purely stationary, it has well defined population autocovariance matrices $\Gamma_{i,j}^s := \mathbb{E}(s_{t-i}s_{t-j}')$, for any i and j. Further, s_t has a stationary VMA(∞) representation,

$$s_t = \sum_{c=0}^{\infty} Y_c \varepsilon_{t-c}, \tag{36}$$

where ε_t is a ρ_0-dimensional vector of *i.i.d.* $\mathcal{N}(0, \Sigma_\varepsilon)$ random variables. Then, we can examine the finite sample property of s_t. First, it is straightforward that the following result holds:

$$\plim_{T\to\infty} \hat{\Sigma}_\varepsilon = \Sigma_\varepsilon, \qquad \hat{\Sigma}_\varepsilon = \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t \varepsilon_t', \tag{37}$$

where $T^{1/2}(\hat{\Sigma}_\varepsilon - \Sigma_\varepsilon)$ has a limiting normal distribution. Based on the analysis above, in finite samples, the ρ_0 smallest eigenvalues of $\hat{A}_x(0,0)$ converge in probability to the squared canonical correlations between s_t and s_{t-1}, *i.e.* the eigenvalues of

$$\hat{A}_s(0,0) = (\hat{\Gamma}_{0,0}^s)^{-1} \hat{\Gamma}_{0,1}^s (\hat{\Gamma}_{1,1}^s)^{-1} \hat{\Gamma}_{1,0}^s, \qquad \text{where} \quad \hat{\Gamma}_{i,j}^s = \frac{1}{T} \sum_{t=1}^{T} s_{t-i} s_{t-j}'. \tag{38}$$

---

[6]The proof of this result is provided in the appendix of Poskitt (2000).

Given the relationship in equation (36), it follows immediately that

$$\underset{T \to \infty}{\text{plim}} \, \hat{\mathbf{\Gamma}}^s_{i,j} = \mathbf{\Gamma}^s_{i,j}, \qquad (39)$$

and further, $T^{1/2}(\hat{\mathbf{\Gamma}}^s_{i,j} - \mathbf{\Gamma}^s_{i,j})$ has a limiting normal distribution. Hence, the asymptotic distributions of $\lambda_{(1),T} \leq \cdots \leq \lambda_{(\rho_0),T}$ are not affected by the existence of the nonstationary component $\boldsymbol{n}_t$.

We then examine the degrees of freedom of the test statistic given in equation (27). For a stationary K-dimensional time series, the degrees of freedom for testing the existence of r SCM(0,0) are calculated on the basis of the $K \times K$ square matrix $\hat{A}(0,0)$ having rank $K - r$. However, for a nonstationary K-dimensional process, there is a separation between $I(0)$ and $I(1)$ variables based on the above results. Furthermore, the multiplicity of zero eigenvalues of $A(0,0)$ only depends on $\boldsymbol{s}_t$, which is of dimension $\rho_0 \times 1$. Thus, in this situation, the degrees of freedom for testing r SCM(0,0) is based on a $\rho_0 \times \rho_0$ matrix $\hat{A}_s(0,0)$ having rank $\rho_0 - r$. Fortunately, they both have r^2 degrees of freedom according to the test statistic given in Section 3.2.1. Therefore, we have verified that the testing procedure of Tiao and Tsay (1989) for SCM(0,0) is still valid in the nonstationary case.

3.2.4 Testing for SCM(p_i,q_i)

The same reasoning will carry through to the test for SCM(m,l) in general, $m = 0, 1, \ldots, p$ and $l = 0, 1, \ldots, q$. The eigenvalues of $\hat{A}(m,l)$, are the same as those of $\hat{A}_x(m,l)$, and the probability limit of the latter, $A_x(m,l) := \underset{T \to \infty}{\text{plim}} \, \hat{A}_x(m,l)$, exists. The matrix $A_x(m,l)$ is a $K(m+1)$-dimensional square matrix. There exists a nonsingular $K(m+1) \times K(m+1)$ transformation matrix $\mathcal{H}$ to partition $Y_{m,t}$ into a purely stationary part and a difference stationary part,

$$\mathcal{H} Y_{m,t} = \begin{pmatrix} \tilde{\boldsymbol{s}}_{m,t} \\ \tilde{\boldsymbol{n}}_{m,t} \end{pmatrix}; \qquad \text{and hence,} \quad \mathcal{H} Y_{m,t-l-1} = \begin{pmatrix} \tilde{\boldsymbol{s}}_{m,t-l-1} \\ \tilde{\boldsymbol{n}}_{m,t-l-1} \end{pmatrix}. \qquad (40)$$

For any finite number l, $\tilde{\boldsymbol{s}}_{m,t-l-1}$ and $\tilde{\boldsymbol{n}}_{m,t-l-1}$ will have the same stationarity properties as $\tilde{\boldsymbol{s}}_t$ and $\tilde{\boldsymbol{n}}_t$, respectively. As with the rank property of $A_x(0,0)$, the nonstationary components $\tilde{\boldsymbol{n}}_{m,t}$ will not affect the number of zero canonical correlations between $Y_{m,t}$ and $Y_{m,t-l-1}$, and thus the multiplicity of zero eigenvalues in $A_x(m,l)$.

Therefore, the nonstationary components have no effect on the rank property of $\hat{A}_x(m,l)$. The statistical procedures needed to test for zero canonical correlations for a given sample are the same for both stationary and nonstationary cases. We do not discuss the testing procedure further, because it is not the focus of the present paper. Interested readers can refer to the work of Athanasopoulos and Vahid (2008); Athanasopoulos (2007) and Tiao and Tsay (1989) for details of the steps for the SCM testing procedure.

3.3 Stage 3: Estimating the EC-VARMA Model

Suppose that the identified canonical SCM VARMA model for y_t in levels is

$$\mathbf{\Phi}_0 y_t = \mathbf{\Phi}_1 y_{t-1} + \cdots + \mathbf{\Phi}_p y_{t-p} + u_t + \mathbf{\Theta}_1 u_{t-1} + \cdots + \mathbf{\Theta}_q u_{t-q}, \tag{41}$$

where $\mathbf{\Phi}_0$ is a nonsingular matrix with unit diagonal elements. The error correction model can be obtained from equation (41) by subtracting $\mathbf{\Phi}_0 y_{t-1}$ from both sides of the equation and rearranging terms:

$$\mathbf{\Phi}_0 \Delta y_t = \mathbf{\Pi} y_{t-1} + \mathbf{\Psi}_1 \Delta y_{t-1} + \cdots + \mathbf{\Psi}_{p-1} \Delta y_{t-p+1} + u_t + \mathbf{\Theta}_1 u_{t-1} + \cdots + \mathbf{\Theta}_q u_{t-q}, \tag{42}$$

where

$$\mathbf{\Pi} = -(\mathbf{\Phi}_0 - \sum_{j=1}^{p} \mathbf{\Phi}_i), \quad \text{and} \quad \mathbf{\Psi}_i = -\sum_{j=i+1}^{p} \mathbf{\Phi}_j, \qquad i = 1, \ldots, p-1. \tag{43}$$

3.3.1 Parameter Restrictions in the EC-VARMA Model

In general, there are certain zero restrictions placed on $\mathbf{\Phi}_i$, $i = 1, \ldots, p$. According to the relationships in equation (43), $\mathbf{\Psi}_i$ will satisfy the same identification restrictions as $\sum_{j=i+1}^{p} \mathbf{\Phi}_i$, $i = 1, \ldots, p-1$, and the zero rows of $\mathbf{\Pi}$ will be the same as those of $\mathbf{\Phi}_0 - \mathbf{\Phi}_1$. Hence, in the case where all of the elements in $\mathbf{\Phi}_1$ are free-varying parameters, there will be no zero elements in $\mathbf{\Pi}$, keeping in mind that $\mathbf{\Phi}_0$ has full rank.

The cointegrating rank ρ and the number of zero rows in $\mathbf{\Phi}_1$ are related, because they both imply parameter constraints on $\mathbf{\Pi}$. Denote the number of zero rows in $\mathbf{\Phi}_1$ by τ. It implies that $\text{rank}(\mathbf{\Pi}) = \rho_0 \geq \tau$, because the τ rows that come from $\mathbf{\Phi}_0$ are linearly independent. In the context of SCM procedure, τ is the number of SCM(0,j) process in y_t, $j = 0, 1, \ldots, q$. Given the derivations in Section 3.2, the number of SCM(0,j) cannot exceed ρ_0, *i.e.* the dimension of the stationary component s_t. Hence, $\tau \leq \rho_0$ should always hold for a canonical SCM VARMA representation.

3.3.2 An Iterative Procedure

The VARMA(p,q) model to be estimated is in the error correction form:

$$\mathbf{\Phi}_0 \Delta y_t = \boldsymbol{\alpha} \boldsymbol{\beta}' y_{t-1} + \mathbf{\Psi}_1 \Delta y_{t-1} + \cdots + \mathbf{\Psi}_{p-1} \Delta y_{t-p+1} + u_t + \mathbf{\Theta}_1 u_{t-1} + \cdots + \mathbf{\Theta}_q u_{t-q}, \tag{44}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are both $K \times \hat{\rho}$ dimensional matrices of full column rank, and there are proper restrictions imposed on the coefficient matrices in order to ensure unique identification. The traditional approach in the literature is to estimate all unknown parameters simultaneously using FIML, which is the exact method that we use for the empirical application in Section 5. However, it is both infeasible and computationally inefficient to use FIML with large scale simulations in Section 4, because it may occasionally fail to converge. Thus, we use the following iterative procedure to estimate

model (44) in the Monte Carlo simulation. This procedure is built upon the iterative OLS (IOLS) estimation suggested by Kapetanios (2003) for stationary VARMA models.

The initial estimate of the error sequence $\hat{u}_t^0$ is obtained from the residual of a VAR, where the lag length of the VAR is an increasing function of the sample size T, and is larger than the AR order of the identified VARMA DGP. We let the lag length be $\lceil \ln T \rceil$, *i.e.* the smallest integer that is greater than $\ln T$, as was suggested by Lütkepohl and Poskitt (1996). The residual obtained from this VAR($\lceil \ln T \rceil$), namely $\hat{u}_t^0$, is a consistent estimate of the true error u_t.

The cointegrating vectors in the error correction model are estimated in the first step of the IOLS procedure. We calculate the partial canonical correlations between Δy_t and y_{t-1} after controlling for Δy_{t-1}, ..., Δy_{t-p+1} and $\hat{u}_{t-1}^0$, ..., $\hat{u}_{t-q}^0$. The canonical covariates corresponding to the largest $\hat{r}$ squared partial canonical correlations are taken as the estimated cointegrating vectors, $\hat{\beta}^0$. $(\hat{\beta}^0)' y_{t-1}$ is commonly referred to as the error correction term. The rest of the parameters are then estimated by the OLS regression of equation (44) with $(\hat{\beta}^0)' y_{t-1}$ and $\hat{u}_{t-1}^0$, ..., $\hat{u}_{t-q}^0$ taken as known, subject to its zero restrictions in equation (44).

There are a few important issues which should be noted in the OLS estimation of equation (44). First, the zero restrictions imposed by the SCMs on the coefficient matrices $\mathbf{\Psi}_i$ and $\mathbf{\Theta}_j$ should be taken into account in the estimation, $i = 1, \ldots, p$, $j = 1, \ldots, q$. To put it differently, if some elements of $\mathbf{\Psi}_i$ or $\mathbf{\Theta}_j$ are restricted to be zero, then the corresponding variables need to be excluded from the OLS estimation.

More importantly, the restrictions on $\mathbf{\Phi}_0$ should be reflected in the estimation as well. Recall that $\mathbf{\Phi}_0$ is a non-singular matrix with unit diagonal elements for the SCM representation. Consider the OLS estimation of the i-th row of the system equation (44), $i = 1, \ldots, K$. If the ij-element of $\mathbf{\Phi}_0$ is non-zero, $j = 1, \ldots, K$ and $j \neq i$, the j-th contemporaneous variable, $\Delta y_{j,t}$ should be put on the right hand side as an explanatory variable. Specifically, the OLS estimation should be conducted using the following equation

$$\Delta y_t = \alpha (\hat{\beta}^0)' y_{t-1} + \mathbf{\Psi}_1 \Delta y_{t-1} + \cdots + \mathbf{\Psi}_{p-1} \Delta y_{t-p+1}$$
$$+ (I - \mathbf{\Phi}_0) \Delta y_t + \mathbf{\Theta}_1 \hat{u}_{t-1}^0 + \cdots + \mathbf{\Theta}_q \hat{u}_{t-q}^0 + u_t, \qquad (45)$$

where u_t is the residual.

The contemporaneous variables $(I - \mathbf{\Phi}_0) \Delta y_t$ are not included in the estimation of the cointegrating vectors, because it will not affect the estimated values of $\hat{\beta}^0$. This can be seen by pre-multiplying both sides of equation (44) by $\mathbf{\Phi}_0^{-1}$:

$$\Delta y_t = \mathbf{\Phi}_0^{-1} \alpha \beta' y_{t-1} + \mathbf{\Phi}_0^{-1} \mathbf{\Psi}_1 \Delta y_{t-1} + \cdots + \mathbf{\Phi}_0^{-1} \mathbf{\Psi}_{p-1} \Delta y_{t-p+1}$$
$$+ \mathbf{\Phi}_0^{-1} u_t + \mathbf{\Phi}_0^{-1} \mathbf{\Theta}_1 u_{t-1} + \cdots + \mathbf{\Phi}_0^{-1} \mathbf{\Theta}_q u_{t-q}$$
$$= \tilde{\alpha} \beta' y_{t-1} + \widetilde{\mathbf{\Psi}}_1 \Delta y_{t-1} + \cdots + \widetilde{\mathbf{\Psi}}_{p-1} \Delta y_{t-p+1} + \tilde{u}_t + \widetilde{\mathbf{\Theta}}_1 u_{t-1} + \cdots + \widetilde{\mathbf{\Theta}}_q u_{t-q}, \qquad (46)$$

where $\widetilde{\alpha} = \Phi_0^{-1}\alpha$, $\widetilde{u}_t = \Phi_0^{-1}u_t$, $\widetilde{\Psi}_i = \Phi_0^{-1}\Psi_i$, and $\widetilde{\Theta}_j = \Phi_0^{-1}\Theta_j$, $i = 1, \ldots, p-1$, $j = 1, \ldots, q$. Hence equation (46) will give rise to different estimates of the coefficient matrices, but the estimates of interest— $\hat{\beta}^0$ will not change.

The same set of rules applies to each iteration of the OLS estimation hereafter, although we do not state this explicitly in each case. The estimated residual is denoted by $\hat{u}_t^1$. In the subsequent iteration of estimating $\hat{\beta}^1$ and the OLS regression in the form of equation (44), $\hat{u}_t^1$ is used in place of $\hat{u}_t^0$. Formally, suppose that the j-th iteration is evaluated and $\hat{u}_t^j$ is obtained. Let $\hat{\Omega}^j$ be the sample covariance matrix of $\hat{u}_t^j$. The IOLS procedure takes the following steps for an error correction VARMA model.

In the $(j+1)$-th iteration, we first calculate the partial canonical correlation between Δy_t and y_{t-1} after controlling for Δy_{t-1}, …, Δy_{t-p+1} and $\hat{u}_{t-1}^j$, …, $\hat{u}_{t-q}^j$. The estimated cointegrating vectors $\hat{\beta}^{j+1}$ are formed by the canonical covariates correspond to the largest $\hat{\rho}$ sample squared partial canonical correlations. We then use OLS to estimate the regression model of the following form:

$$\Delta y_t = \alpha(\hat{\beta}^{j+1})'y_{t-1} + \Psi_1\Delta y_{t-1} + \cdots + \Psi_{p-1}\Delta y_{t-p+1}$$
$$+ (I - \Phi_0)\Delta y_t + \Theta_1\hat{u}_{t-1}^j + \cdots + \Theta_q\hat{u}_{t-q}^j + u_t. \qquad (47)$$

Denote the residual estimates obtained from equation (47) by $\hat{u}_t^{j+1}$, and its covariance matrix estimate by $\hat{\Omega}^{j+1}$. If the iterative procedure converges such that $\| \ln|\hat{\Omega}^{j+1}| - \ln|\hat{\Omega}^j| \| < \epsilon$ for some pre-specified constant $\epsilon > 0$, then the OLS estimates of the coefficients in equation (47) are adopted. Otherwise, we should proceed to evaluate the $(j+2)$-th iteration.

The sequence of the residual $\hat{u}_t^j$ is redefined with each iteration j, and therefore there is no guarantee that this iterative process will converge. Kapetanios (2003) points out that iterations of $\hat{u}_t^j$ will converge if this procedure produces a contraction mapping. Hence, he suggests to check the eigenvalues of the Jacobian at each iteration. If any of these eigenvalues are greater than unity, then this signals that the iterative process is unlikely to converge. However, it is difficult to implement this procedure in practice when the dimension of the parameter space is high. Hence, it is necessary to set a pre-specified maximum number of iterations, M_{max}. If the convergence condition $\| \ln|\hat{\Omega}^{j+1}| - \ln|\hat{\Omega}^j| \| < \epsilon$ cannot be achieved within M_{max} iterations, there are a few possible solutions to resort to.

Similar to the numerical maximum likelihood methods, good starting values of the parameters are important for convergence of the iterative algorithm. We can perturb the initial estimates of the coefficients using $\hat{u}_t^0$, and repeat the iterative procedure a few times. Alternatively, we can use other estimators as the starting values. To name a few, the Hannan-Rissanen method (Hannan and Rissanen, 1982), the Hannan-Kavalieris procedure (Hannan and Kavalieris, 1984) and the generalized least squares procedure proposed by Koreisha and Pukkila (1990) can all serve this purpose. Kascha (2012) conducts an extensive comparison of these estimators for stationary VARMA DGPs via

Monte Carlo simulations. His results suggest that the algorithm of Hannan and Kavalieris (1984) is generally preferable to other algorithms. Hence, if all attempts to initialize the IOLS procedure with good starting values fail and convergence still cannot be achieved, we suggest to use the estimator given by Hannan and Kavalieris (1984) to produce the final estimates.

4 Monte Carlo Simulation

We use Monte Carlo simulation to evaluate the predictive ability of EC-VARMA model and VECMs, when the data are in truth generated from an EC-VARMA DGP. This simulation also serves the purpose of examining the finite sample performance of our extension of Poskitt's selection criterion for determining the cointegrating rank, compared to the usual Johansen procedure.

DGP used in the simulation is a 3-dimensional cointegrated VARMA(1,1) process:

$$
\boldsymbol{y}_t = \begin{pmatrix} 0.75 & 0.25 & 0 \\ 0.11 & 0.89 & 0 \\ -0.1 & 0.1 & 1 \end{pmatrix} \boldsymbol{y}_{t-1} + \boldsymbol{u}_t + \begin{pmatrix} -0.35 & 0.2 & -0.54 \\ 0.7 & 0.5 & 0.1 \\ -0.4 & 0.75 & 0.6 \end{pmatrix} \boldsymbol{u}_{t-1}, \tag{48}
$$

where $\boldsymbol{u}_t$ is *i.i.d.* $\mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_3)$. The AR and MA orders are both one to simplify the illustration. The EC-VARMA(0,1) representation is $\Delta \boldsymbol{y}_t = \boldsymbol{\Pi} \boldsymbol{y}_{t-1} + \boldsymbol{u}_t + \boldsymbol{\Theta}_1 \boldsymbol{u}_{t-1}$, where $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$, $\boldsymbol{\alpha} = (-0.25, 0.11, -0.1)'$ and $\boldsymbol{\beta} = (1, -1, 0)'$. Hence the true cointegrating rank $\rho_0 = 1$. All three eigenvalues of $\boldsymbol{\Theta}_1$ are close to 0.8, indicating the presence of a relatively strong propagation mechanism in the MA dynamics.

We simulate four different sample sizes: $T = 100$, 200, 400 and 1000, and perform 100 replications for each sample size. The forecasting horizons are $h = 1, \ldots, 24$. We consider two measures of forecasting accuracy: the trace of mean squared forecast errors (tr(MSFE)) for $\boldsymbol{y}_t$ in levels and the generalized forecast error second moment (GFESM). The latter metric of forecasting accuracy is proposed by Clements and Hendry (1993). The GFESM is the determinant of the forecast error second moment matrix pooled across all horizons,

$$
\text{GFESM}_h = \left(\det \left(\mathbb{E}[\text{vec}(\mathbf{e}_1, \ldots, \mathbf{e}_h) \, \text{vec}(\mathbf{e}_1, \ldots, \mathbf{e}_h)'] \right) \right)^{1/h}, \tag{49}
$$

where $\mathbf{e}_i$ is the $K \times 1$ dimensional vector of i-th step ahead forecasting error, $i = 1, \ldots, h$. The main advantage of GFESM is that it is invariant to non-singular, scale preserving linear transformations for all forecast horizons (see Clements and Hendry, 1993, for details).

The forecast errors generated from the estimated VECMs and EC-VARMA model are compared to the theoretical forecast errors, which are generated from the EC-VARMA(0,1) models with the true parameters. We refer to these models as the "oracle".

4.1 Selection of Cointegrating Rank, Lag Length and the SCM Structure

The first step of the propose algorithm is to determine the cointegrating rank using the extended Poskitt (2000) procedure for the VARMA models. We compare its finite sample performance with that of the usual Johansen procedure combined with three different information criteria — AIC, HQ and BIC — for the VARs in the simulation. Due to the non-parametric nature of the modified Poskitt (2000) procedure, it can be used in conjunction with either VARs or VARMA models. Given the DGP in equation (48) and the simulation setting outlined above, the Poskitt's procedure select the true cointegrating rank $\hat{\rho} = \rho_0 = 1$ for all simulated sample paths, even using only 100 observations.[7]

The Johansen procedure is dependent on the lag length of the VECMs, hence we first specify the lag lengths for VARs using the three information criteria. Figure 1 plots the distribution of the estimated lag length for VARs with different sample sizes. The maximum lag length is set to 20 considering the available sample size. All three of the information criteria choose longer lags as the sample size T increases. AIC has the tendency to choose very long lags when the sample size $T = 100$. The empirical distributions of the estimated cointegrating rank $\hat{\rho}$ selected by the Johansen procedure are plotted in Figure 2.

Figure 1: Distribution of the estimated lag length for VAR with different sample sizes
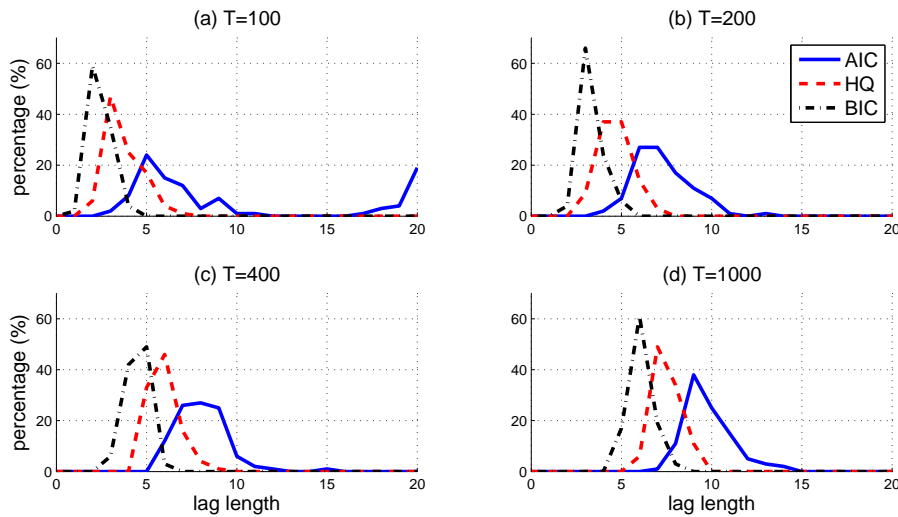
It is evident from Figure 2 that the actual size of the Johansen procedure is far from its nominal size of 5%. This phenomenon in finite samples is in accord with the observation by Lütkepohl and Saikkonen (1999) when the true DGPs are cointegrated VARMA processes.

Another interesting observation from these two plots is the case when $T = 100$. AIC

---

[7]A more comprehensive simulation using 100 different DGPs shows that the extended Poskitt (2000) procedure can choose the true cointegrating rank for at least 95% of the time when $T = 100$. These simulation results are available upon request.

Figure 2: Distribution of the estimated cointegrating rank $\hat{\rho}$ with different sample sizes
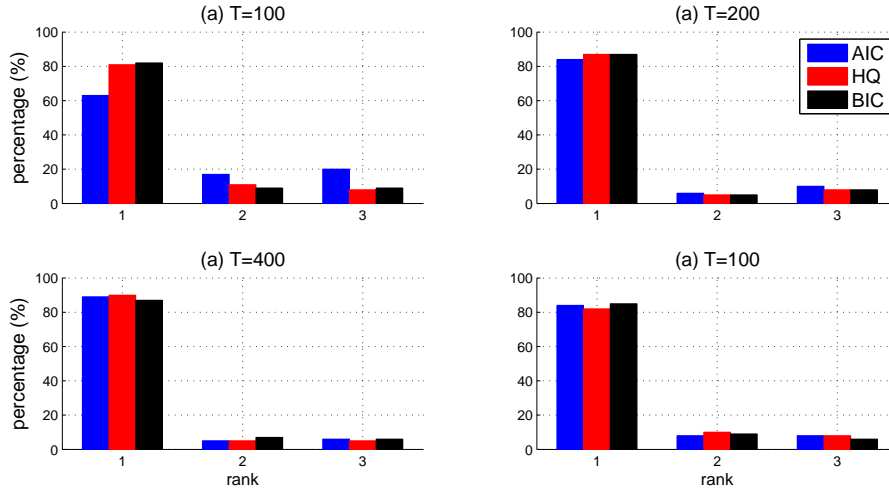
chooses very long lags (*i.e.* higher than 12) more than 20% of the time. In line with the findings of Vahid and Issler (2002), we would expect such an over-parameterization by AIC to cause a large estimation error, especially in small samples. Correspondingly, Figure 2 shows that when $T = 100$, the Johansen procedure with lag length selected using AIC chooses the correct specification of the cointegrating rank for roughly 60% of the time. This phenomenon also reveals that the cointegrating rank estimated using the Johansen procedure depends crucially on the lag length of the VECM.

Overall, Figures 2 provides supporting evidence in favor of the modified Poskitt (2000) method for choosing the cointegrating rank when the true DGP is a VARMA process. We then proceed to the second stage of the proposed algorithm to identify the underlying SCM structure of each simulated path.

Various studies (see e.g. Athanasopoulos and Vahid, 2008; Athanasopoulos et al., 2012) have found that the identification procedure for specifying the canonical SCM VARMA models is quite successful. We find similar results using the cointegrated VARMA DGP in equation (48). The true SCM structure implied by equation (48) is three SCM(1,1). The testing procedure outlined in Section 3.2 is able to pick up the correct structure for more than 95% of the time. Other identified SCM specifications are listed in Appendix A.

4.2 Forecasting with EC-VARMA models and VECMs

Tables 1 and 2 present the percentage differences in tr(MSFE) and GFESM between the estimated models and the "oracle" at each forecasting horizon. For instance, the first number in Table 1 denotes that when the sample size is $T = 100$, the trace of the one-step MSFE obtained from the EC-VARMA models estimated using IOLS is 16.3% larger than the trace of one-step MSFE from the "oracle". Similar interpretations can be drawn from Table 2 for GFESM.

20

Table 1: Percentage difference in tr(MSFE) for y_t in levels between the estimated models and the "oracle"

h	ECVARMA $\hat{\rho}_P$	AIC $\hat{\rho}_J$	AIC $\hat{\rho}_P$	HQ $\hat{\rho}_J$	HQ $\hat{\rho}_P$	BIC $\hat{\rho}_J$	BIC $\hat{\rho}_P$
T = 100							
1	16.3♮	63.1	61.5	28.8	26.2	29.4	28.5
4	1.0♮	68.0	51.7	21.1	15.0	7.5	6.1
8	6.7♮	76.7	50.4	16.3	12.3	10.5	8.9
12	10.0	97.3	59.9	11.8	11.4	5.5♮	9.8
16	11.2	138.2	77.5	14.6	14.3	8.1♮	12.6
20	17.0	175.4	93.1	19.5	17.9	12.9♮	15.6
24	18.6	211.3	98.9	21.8	20.8	16.3♮	18.0
T = 400							
1	5.6♮	9.0	8.5	11.6	11.1	12.6	12.6
4	1.8♮	9.7	11.2	7.5	8.6	4.8	6.3
8	3.2♮	9.8	10.1	7.2	8.0	5.5	5.9
12	2.4♮	7.9	6.6	5.0	4.5	3.1	3.2
16	1.1♮	8.5	5.3	6.9	4.4	4.1	1.8
20	1.6♮	8.9	5.2	7.3	4.9	5.2	2.7
24	2.6♮	8.3	5.6	5.9	5.3	5.5	3.9

h	ECVARMA $\hat{\rho}_P$	AIC $\hat{\rho}_J$	AIC $\hat{\rho}_P$	HQ $\hat{\rho}_J$	HQ $\hat{\rho}_P$	BIC $\hat{\rho}_J$	BIC $\hat{\rho}_P$
T = 200							
1	10.7♮	15.5	16.4	19.6	19.2	21.8	22.2
4	5.4♮	21.9	21.1	16.2	12.6	13.1	11.7
8	6.0♮	17.0	12.5	15.5	8.3	11.0	7.6
12	6.7	13.9	7.6	13.0	5.8♮	9.8	7.7
16	5.9	15.6	7.8	15.4	5.6♮	12.2	7.2
20	8.4	14.6	8.9	14.2	7.4♮	12.8	8.3
24	9.7♮	15.2	10.7	14.3	10.0	14.2	10.0
T = 1000							
1	0.8♮	4.3	4.4	1.7	1.9	5.0	5.1
4	-2.4♮	4.5	4.8	3.0	3.2	4.0	3.9
8	-1.7♮	2.8	3.6	1.5	2.3	1.4	1.9
12	-1.7	-0.4	0.2	-2.0	-0.6	-2.6♮	-1.4
16	-2.9♮	0.1	0.1	-2.1	-0.4	-2.4	-0.9
20	-2.2	-0.6	-1.1	-2.3♮	-1.6	-1.3	-1.1
24	-2.9♮	-0.2	-0.8	-2.2	-1.9	-0.8	-0.6

[1] $\hat{\rho}_P$: using the cointegrating rank selected by the extended Poskitt's method;
[2] $\hat{\rho}_J$: using the cointegrating rank selected by the Johansen procedure;
[3] ♮: indicating the smallest measure of forecasting accuracy among the estimated models.

Table 2: Percentage difference in GFESM between the estimated models and the "oracle"

h	ECVARMA $\hat{\rho}_P$	AIC $\hat{\rho}_J$	AIC $\hat{\rho}_P$	HQ $\hat{\rho}_J$	HQ $\hat{\rho}_P$	BIC $\hat{\rho}_J$	BIC $\hat{\rho}_P$
	$T=100$						
1	48.1♮	185.7	179.3	82.9	75.7	85.3	82.9
4	46.1♮	154.6	145.7	54.0	51.4	66.6	66.3
8	37.2	124.8	116.9	36.2	34.4♮	45.9	44.9
12	29.5	106.8	100.7	26.5	25.6♮	34.5	34.6
16	25.7	97.2	88.2	20.3	20.2♮	28.9	28.8
20	24.1	91.5	83.0	16.1	15.9♮	24.0	23.6
24	23.1	87.6	79.6	12.9♮	13.4	20.0	19.9
	$T=400$						
1	12.9♮	24.8	23.5	32.6	31.6	35.5	35.4
4	12.6♮	23.8	24.9	26.8	27.6	29.2	30.2
8	9.0♮	22.0	22.1	21.5	21.8	20.1	20.0
12	8.0♮	16.8	16.7	15.1	15.3	14.6	14.5
16	5.5♮	13.1	13.1	11.3	11.3	11.6	11.2
20	6.3♮	12.2	12.3	10.3	10.5	9.5	9.2
24	7.1♮	10.5	10.6	9.7	9.6	8.5	8.2
	$T=200$						
1	29.4♮	43.8	46.6	57.1	56.0	63.7	65.2
4	21.4♮	44.4	44.5	39.9	39.7	50.7	50.7
8	23.5♮	30.4	29.1	25.8	24.8	32.3	31.8
12	18.6♮	23.4	22.6	20.2	19.8	25.2	25.2
16	18.6	16.8	16.7	15.1	14.9♮	19.9	20.1
20	15.3	15.4	15.8	13.7	13.7♮	15.3	15.9
24	14.7	14.0	14.3	11.5	11.1♮	13.0	13.0
	$T=1000$						
1	-0.8♮	9.9	10.3	2.0	2.8	12.5	12.9
4	1.7♮	16.1	16.5	12.6	12.9	19.8	19.9
8	1.2♮	13.7	14.0	12.2	12.4	14.9	15.0
12	2.0♮	10.0	10.1	9.3	9.5	10.5	10.7
16	2.0♮	8.7	8.7	7.2	7.4	8.0	8.2
20	0.2♮	7.2	7.0	6.6	6.4	7.1	6.8
24	0.6♮	6.6	6.4	5.6	5.3	5.7	5.4

[1] $\hat{\rho}_P$: using the cointegrating rank selected by the extended Poskitt's method;
[2] $\hat{\rho}_J$: using the cointegrating rank selected by the Johansen procedure.
[3] ♮: indicating the smallest measure of forecasting accuracy among the estimated models.

In addition to the estimated EC-VARMA models, we also fit VECMs with lag lengths selected by AIC, HQ or BIC to each simulated data path separately. Conditional on the selected AR lag length, the cointegrating rank is then chosen by the Johansen procedure. VECMs with the cointegrating rank determined by the modified Poskitt (2000) approach are also estimated here. The tr(MSFE) and GFESM are reported in Tables 1 and 2 for both specifications of the cointegrating rank, in columns labeled $\hat{\rho}_J$ and $\hat{\rho}_P$ respectively. The symbol "♮" indicates the model specification that produces the most accurate forecast. The evidences provided in Tables 1 and 2 are conclusive, that in general, given the typical sample sizes available for macroeconomic data, using EC-VARMA models reduces the forecasting error, especially with large sample sizes. For small sample sizes, the advantages of EC-VARMA models are quite substantial in short horizons.

The columns labeled $\hat{\rho}_J$ and $\hat{\rho}_P$ in Tables 1 and 2 allow us to examine the effects on the forecasting accuracy of using different cointegrating ranks. The lag lengths of VECMs are selected by the same information criteria, but the cointegrating ranks are chosen by the Johansen procedure and the extended Poskitt (2000) method, respectively. The tables show that using the extended Poskitt (2000) method produces smaller forecast error mostly when the sample size is small.

5   Term Structure of Interest Rates

It is commonly accepted that interest rates with different maturities are cointegrated (see Hall et al., 1992). The cointegrating vector between any two interest rate series should be close to $(1, -1)$, *i.e.* the interest rate spreads should be stationary, despite the fact that most interest rates are regarded as $I(1)$ series. Many studies of interest rates have been conducted within the VECM framework, with the aim of forecasting interest rates. To name one among others, Hall et al. (1992) find that yields to maturity of U.S. treasury bills specify an error correction model with post-1982 data, which proves to be useful in forecasting changes in yields.

The use of VARMA models to capture the dynamics in the term structure of interest rates is not new. Kascha and Trenkler (2011) show that a cointegrated VARMA model generates superior forecasts of U.S. interest rates. Nevertheless, our study differs from theirs in several aspects. First, Kascha and Trenkler (2011) extend the FMA representation of Dufour and Pelletier (2008) to specify their VARMA model. This is simpler but less general and parsimonious than the SCM representation (Athanasopoulos and Vahid, 2008) employed here. Moreover, Kascha and Trenkler (2011) take the cointegrating rank as given ($\rho = K - 1$) for their forecasting evaluation, whereas we test for the cointegrating rank for each sample.

5.1   Data

We use monthly data of the U.S. federal funds rate, and 3-month and 6-month treasury bill rates to form a three variable system. Let $\boldsymbol{y}_t = (ff_t, i3_t, i6_t)'$. The available sample

period is from 1958:12 to 2011:09, which leads to a total of 634 observations. Figure 3 plots the three interest rate series over the entire sample period. The movements in the three series clearly share a similar pattern, especially for the 3-month and 6-month treasury bill rates.[8]

Figure 3: The three interest rate series (%)
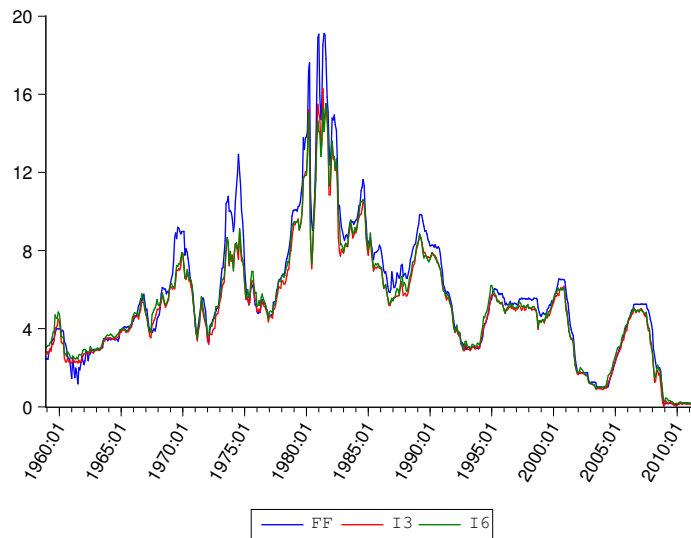
We use the first 400 observations as the initial estimation sample to forecast future interest rates for up to 12-steps ahead. We use an expanding window,[9] and the same forecasting exercise is repeated 222 times. In addition to tr(MSFE) for y_t in levels and GFESM, the determinant of MSFE (det(MSFE)) for y_t in levels is also examined.

5.2 Selection of Cointegrating Rank and Lag Length

We first estimate the VECMs and EC-VARMA models using the theoretical cointegrating relationships — for a K-dimensional model, the cointegrating rank is fixed to be $\rho = K - 1$. Furthermore, the cointegrating vectors are specified as follows:

$$\beta = \left[\left(1, \quad -1, \quad 0\right)', \quad \left(1, \quad 0, \quad -1\right)' \right]. \tag{50}$$

The AR order of the finite VAR is determined using AIC, HQ and BIC. The maximum AR lag is set to be 24 which is two years for monthly data. The distributions of the

---

[8]One may want to drop the observations during the last global financial crisis (2008:01-2011:09, the last 45 observations), due to the possibility of a structural break. We experiment with this shorter sample as well, and it produces qualitatively similar results. The forecast errors are smaller in almost all cases, but the ranking of the competing models does not change.

[9]We also use rolling window to generate forecasts of future interest rates, which can account for the possible structural breaks over the time span that we considered. The forecasting results are very similar to what are reported below using expanding window, and hence are omitted here. Those results are available upon request.

selected AR lag length for 222 sets of estimation samples are tabulated in Table 3. All of the models specified by AIC are heavily parameterized, choosing a VAR(21) for y_t in levels for all 222 sets of estimation samples. BIC seems to choose a much shorter lag length for this dataset. Surprisingly, HQ only chooses three different lag lengths — 3, 9 and 16, which do not increase gradually, but this is what is observed from the data using an expanding window.

Table 3: Distribution of the estimated lag length for VAR
with theoretical cointegration

Information Criteria	AR Lag Length				
	2	3	9	16	21
AIC	0	0	0	0	100%
HQ	0	64.9%	12.1%	23.0%	0
BIC	70.3%	29.7%	0	0	0

Table 4: Distribution of the estimated lag length and
cointegrating rank (data-specified cointegration)

Information Criteria	Lag	Cointegration Rank		
		0	1	2
AIC	21	48.2%	51.8%	
HQ	3			64.9%
	9		0.9%	11.3%
	16			23.0%
BIC	2			70.3%
	3			29.7%
Poskitt's Method				100%

Empirical researchers usually estimate the cointegrating relationships from the data rather than fix them beforehand. Thus, we also investigate the forecasting performances of models with data-specified cointegration. We use the algorithm proposed in Section 3 to identify the structure of the EC-VARMA models. The first step of the procedure determines the cointegrating rank. The extended non-parametric approach (Poskitt, 2000) applied to this three variable system chooses $\hat{\rho} = 2$ consistently for all 222 sets of estimation samples. On the other hand, the Johansen procedure chooses different cointegrating ranks based on the lag lengths selected by different information criteria. The lag length and cointegrating rank specified by the two methods are tabulated in Table 4. AIC once again shows a tendency toward over-parameterization. The joint use of the Johansen procedure and AIC chooses $\hat{\rho} = 0$ or $\hat{\rho} = 1$. Neither of these comply

with economic theory. VECM with lag length selected using BIC chooses a reasonable lag length and an expected cointegrating rank $\hat{\rho} = 2$.

5.3 Canonical SCM VARMA Representation

There is no evidence in the literature showing that a system of interest rates follows a VARMA process. Hence, the second stage of identifying the VARMA model is necessary in order to justify the choice of this model. We use the SCM methodology to determine the lag orders and the corresponding canonical structure of the VARMA model, keeping in mind that even if the true DGP is a finite order VAR process, the SCM searching procedure is able to identify a VARMA(p,0) model asymptotically.

Conditional on the identified tentative overall order VARMA(1,1) for $\boldsymbol{y}_t$ in levels, we search for each individual SCM. Starting from the most parsimonious SCM(0,0), the underlying SCMs are identified as SCM(1,1) $\sim$ SCM(1,1) $\sim$ SCM(1,0). After testing the SCM structure of the sub-systems and imposing identification restrictions on $\boldsymbol{\Phi}_0$ (see Athanasopoulos, 2007; Athanasopoulos and Vahid, 2008), the canonical SCM VARMA model has the following error correction representation:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ a_0 & 0 & 1 \end{pmatrix} \Delta \boldsymbol{y}_t = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \end{pmatrix} \begin{pmatrix} 1 & \beta_{12} \\ \beta_{21} & 1 \\ \beta_{31} & \beta_{32} \end{pmatrix}' \boldsymbol{y}_{t-1} + \boldsymbol{u}_t + \begin{pmatrix} \theta^1_{11} & \theta^1_{12} & \theta^1_{13} \\ \theta^1_{21} & \theta^1_{22} & \theta^1_{23} \\ 0 & 0 & 0 \end{pmatrix} \boldsymbol{u}_{t-1}. \quad (51)$$

We find exactly the same SCM structure for all 222 sets of estimation samples. There are no zero restrictions imposed on $\boldsymbol{\Phi}_1$ by the canonical SCM structure. Hence, the rank restrictions on $\boldsymbol{\Pi}$ only come from the cointegration relationships. We estimate equation (51) using FIML. In order to provide good initial estimates for maximum likelihood estimation, we put the algorithm of Hannan and Rissanen (1982) into the context of EC-VARMA models, as the starting values of the unknown parameters for the maximum likelihood iteration. This initial estimate works well for this empirical example. The model structure in equation (51) is also used for the estimation with the theoretical cointegrating vectors.

5.4 Forecast Evaluation of Interest Rates

The measures of forecasting accuracy calculated from all types of models are presented in Table 5. We take the EC-VARMA model with theoretical cointegration relationships as the benchmark. The first three columns are models with theoretical cointegration relationships, and the last four columns are models with data-specified cointegration, denoted by the subscript "d". It is worth reminding ourselves that the modified Poskitt's method always chooses a cointegrating rank of $\hat{\rho} = 2 = K - 1$. Hence, the only difference between SCM$_d$ and the benchmark model in this application is due solely to the estimation of the two cointegrating vectors.

Table 5: Percentage difference in measures of forecast accuracy between other models and the EC-VARMA models with theoretical cointegration

	Theoretical Cointegration			Data-specified Cointegration			
	VECMs				VECMs		SCM_d
h	AIC	HQ	BIC	AIC_d	HQ_d	BIC_d	
tr(MSFE)							
1	56.6	31.5	14.7	59.0	33.1	16.3	0.9
4	40.1	12.1	6.1	43.1	13.8	5.1	3.8
8	23.0	1.8	3.8	21.3	6.8	4.7	6.9
12	17.8	$-1.1^\natural$	4.4	11.9	5.6	6.0	7.3
det(MSFE)							
1	112.4	52.0	17.7	118.9	54.7	28.8	$-4.8^\natural$
4	72.4	35.3	16.9	85.4	25.5	14.6	$-22.6^\natural$
8	20.0	16.7	17.0	37.2	1.9	1.5	$-35.8^\natural$
12	-0.0	0.8	16.5	14.5	-26.4	-19.9	$-48.9^\natural$
GFESM							
1	112.4	52.0	17.7	118.9	54.7	28.8	$-4.8^\natural$
4	101.5	39.0	6.2	105.2	36.4	6.2	$-6.9^\natural$
8	97.9	29.9	2.0	99.7	29.2	1.0	$-3.9^\natural$
12	90.3	23.9	0.8	23.8	23.8	-0.0	$-2.2^\natural$

$\natural$: indicates the smallest measure of forecasting accuracy in each row.

The numbers in Table 5 indicate the percentage by which the measures of forecasting accuracy calculated from each type of model are larger than the measures calculated from the benchmark model. Hence, a negative number indicates an improvement in forecasting accuracy over EC-VARMA model with theoretical cointegration on average. The symbol "$\natural$" denotes the type of model that produces the smallest measure of forecasting accuracy in each row. No "$\natural$" in a row indicates that the benchmark model is most accurate.

In Table 5, VECMs with lag lengths chosen using AIC usually lead to the largest forecast error. Such results are not surprising, taking into account the presumably large estimation error caused by the over-parameterization of AIC. It is evident that if the theoretical cointegrating relationships are imposed, EC-VARMA models have smaller tr(MSFE) than VECMs in most of the scenarios considered here. Their advantages over VECMs are more pronounced in the short run. EC-VARMA models with data-specified cointegrating vectors generate the smallest det(MSFE) and GFESM of all of the different model specifications up to 12-step ahead forecasts. In terms of the determinant of the MSFE, there are considerable gains from estimating the cointegrating vectors from the

data rather than using the theoretical ones. The values of det(MSFE) can be reduced by nearly 50% by the use of the EC-VARMA models with estimated cointegrating vectors. The other two measures calculated from the EC-VARMA models with either estimated or theoretical cointegrating relationships are of roughly the same sizes.

5.5 Diebold-Mariano Tests

One may be concerned about the importance of the results reported in Tables 5, since most of the differences between these measures of forecasting accuracy generated from VECMs and EC-VARMA models are very small in magnitude. We use the Diebold-Mariano test (Diebold and Mariano, 1995; West, 1996; Giacomini and White, 2006) to compare the predictive accuracies of these two classes of models. The hypotheses are

$$H_0 : \mathbb{E}\left[e^2_{1,i,h}\right] - \mathbb{E}\left[e^2_{2,i,h}\right] = 0, \quad \text{against} \quad H_1 : \mathbb{E}\left[e^2_{1,i,h}\right] - \mathbb{E}\left[e^2_{2,i,h}\right] < 0, \quad (52)$$

where $e_{1,i,h}$ and $e_{2,i,h}$ denote the h-step ahead forecast errors of the i-th component of y_t, generated from the estimated EC-VARMA model and VECMs, respectively, with $i = 1, 2, 3$. The forecast errors of ff_t, $i6_t$ and $i3_t$ are tested individually. Newey-West heteroskedasticity and auto-correlation robust standard errors are used in these tests, because the forecast errors are generally correlated with each other for $h \geq 2$. The test statistics are tabulated in Table 6. A negative test statistic indicates a smaller MSFE from the EC-VARMA model than from the VECM.

Examining the MSFEs of the three interest rate series individually, the gain from using EC-VARMA models is most substantial when forecasting the federal funds rate, especially in the short to medium horizons. At 10% significance level, the MSFE of the federal funds rate generated from the estimated EC-VARMA models is significantly lower than the MSFEs of VECMs with BIC up to 4-steps ahead, lower than VECMs with HQ up to 5-steps ahead, and lower than VECMs with AIC up to 8-steps ahead. Statistically significant differences are also seen for 3-month and 6-month treasury bill rates. VECMs with AIC produce largest forecast errors in most cases, possibly due to their over-parameterization and mis-specification in the long run dynamics among interest rate series. We also observe that in the cases where the EC-VARMA models produce larger MSFEs than the VECMs, the differences are all insignificant.

Table 6 also reports the results of a generalization of the Diebold-Mariano test to multivariate models to test the differences in the values of tr(MSFE). The null and alternative hypotheses for the h-steps ahead tr(MSFE) are

$$H_0 : \mathbb{E}\left[\sum_{i=1}^{3} e^2_{1,i,h}\right] - \mathbb{E}\left[\sum_{i=1}^{3} e^2_{2,i,h}\right] = 0, \quad (53a)$$

$$H_1 : \mathbb{E}\left[\sum_{i=1}^{3} e^2_{1,i,h}\right] - \mathbb{E}\left[\sum_{i=1}^{3} e^2_{2,i,h}\right] < 0. \quad (53b)$$

Table 6: Test statistics for testing the differences between $e^2_{1,i,h}$ and $e^2_{2,i,h}$

h	EC-VARMA vs. VECM(AIC)				EC-VARMA vs. VECM(HQ)				EC-VARMA vs. VECM(BIC)			
	ff_t	$i3_t$	$i6_t$	tr(MSFE)	ff_t	$i3_t$	$i6_t$	tr(MSFE)	ff_t	$i3_t$	$i6_t$	tr(MSFE)
1	−4.1**	−4.7**	−5.8**	(−6.0)**	−2.2*	−3.3**	−15.2**	(−3.6)**	−1.8*	−1.2	−2.1*	(−1.8)*
2	−3.9**	−4.1**	−3.3**	(−4.8)**	−1.9*	−3.3**	−2.7**	(−2.3)**	−1.8*	0.8	−0.3	(−2.8)**
3	−4.7**	−3.5**	−2.9**	(−4.1)**	−2.7**	−1.9*	−1.0	(−2.5)**	−2.3*	−0.5	−0.4	(−2.4)**
4	−4.1**	−3.0**	−2.3*	(−3.6)**	−2.3*	−0.8	−0.3	(−1.3)†	−1.8*	0.2	0.4	(−0.3)
5	−2.8**	−2.4**	−1.8*	(−2.5)**	−1.5†	−0.3	−0.0	(−0.6)	−0.7	0.4	0.6	(0.2)
6	−2.2*	−1.9*	−1.5†	(−2.0)*	−1.1	−0.0	0.2	(−0.3)	−0.4	0.4	0.6	(0.2)
7	−1.8*	−1.5†	−1.2	(−1.6)†	−0.8	0.1	0.2	(−0.2)	−0.2	0.5	0.6	(0.3)
8	−1.5†	−1.2	−0.8	(−1.3)	−0.6	0.2	0.4	(0.0)	−0.2	0.5	0.7	(0.3)
9	−1.1	−0.8	−0.4	(−0.8)	−0.4	0.3	0.4	(0.1)	−0.3	0.4	0.6	(0.3)
10	−0.8	−0.6	−0.2	(−0.5)	−0.2	0.4	0.5	(0.2)	−0.3	0.4	0.6	(0.3)
11	−0.7	−0.4	−0.1	(−0.4)	−0.2	0.4	0.5	(0.2)	−0.4	0.4	0.7	(0.3)
12	−0.7	−0.5	−0.2	(−0.5)	−0.2	0.3	0.5	(0.2)	−0.3	0.4	0.6	(0.2)

[1] $e^2_{1,i,h}$ denotes the h-step ahead forecast errors of the i-th component of y_t, generated from the estimated EC-VARMA models.

[2] $e^2_{2,i,h}$ denotes the h-step ahead forecast errors of the i-th component of y_t, generated from the estimated VECMs.

[3] Significance levels: † : 10%, * : 5%, ** : 1%.

The testing outcomes of tr(MSFE) are consonant with those for individual series. All in all, we can conclude that EC-VARMA models can forecast at least as accurately as VECMs for all of the forecast horizons considered here. Furthermore, EC-VARMA models produce significantly smaller MSFEs than VECMs over short horizons.

6   Conclusion

This paper combines cointegration relationship among nonstationary time series with VARMA models, which have been shown to improve the forecasting accuracy of time series models. We then investigate whether the use of EC-VARMA models can improve the predictive ability of nonstationary time series.

Applied researchers have tended to favor finite order VARs and VECMs rather than the more general and flexible VARMA models in macroeconomic modelling. This is due mainly to the specification and estimation difficulties encountered with VARMA models. We propose a complete three stage algorithm for identifying and estimating an EC-VARMA model, in order to overcome this problem. The proposed procedure extends the non-parametric approach for determining the cointegrating rank of Poskitt (2000) to include the possibility of $I(0)$ time series. Simulations show that the modified version of the Poskitt (2000) method has good finite sample performances in correctly specifying the true cointegrating rank. Its advantage over the traditional Johansen procedure is manifest for cointegrated VARMA DGP. In order to avoid the computational inefficiency of maximum likelihood estimation, the iterative OLS procedure proposed by Kapetanios (2003) is generalized to estimate EC-VARMA models in the Monte Carlo simulation. When evaluating the forecasting accuracies of the estimated EC-VARMA models and VECMs, the loss from using finite order VECMs is evident when the true DGP is a VARMA process.

We apply the proposed algorithm for identifying and estimating the EC-VARMA model to an empirical dataset for illustration. In this model of the term structure of interest rates, EC-VARMA models and VECMs are estimated with the cointegration relationships being either dictated by economic theory or specified from the data. The EC-VARMA models with either theoretical or data-specified cointegration always produce the most accurate forecasts. Diebold-Mariano tests show that the EC-VARMA model is superior to finite order VECMs in forecasting future interest rates, especially in the short run.

This paper contributes to the growing body of literature on the identification and estimation of VARMA models, and suggests that VARMA models can be both beneficial and relatively straightforward to estimate using our proposed algorithm. In particular, we focus on the error correction VARMA models in this paper, which incorporate cointegrating relationships among multiple time series. Given the empirical evidence in favor of EC-VARMA models, these models are expected to be utilized more comprehensively in macroeconomic modelling and forecasting.

A Identified SCM Structure in the Monte Carlo Simulation

%	SCM Structure	%	SCM Structure
$T = 100$		**$T = 200$**	
95	(1,1)∼(1,1)∼(1,1)	95	(1,1)∼(1,1)∼(1,1)
4	(1,2)∼(1,1)∼(1,1)	3	(1,2)∼(1,1)∼(1,1)
1	(1,2)∼(1,2)∼(1,1)	1	(1,2)∼(1,2)∼(1,1)
		1	(2,3)∼(2,3)∼(1,2)
$T = 400$		**$T = 1000$**	
96	(1,1)∼(1,1)∼(1,1)	96	(1,1)∼(1,1)∼(1,1)
2	(2,2)∼(1,1)∼(1,1)	3	(1,2)∼(1,1)∼(1,1)
1	(2,2)∼(2,2)∼(1,1)	1	(2,2)∼(1,1)∼(1,1)
1	(1,3)∼(1,3)∼(1,1)		

References

Anderson, T. W. (2003), *An Introduction to Multivariate Statistical Analysis*, Wiley, New York.

Athanasopoulos, G. (2007), Essays on Alternative Methods of Identification and Estimation of Vector Autoregressive Moving Average Models, PhD thesis, Monash University.

Athanasopoulos, G., Poskitt, D. S. and Vahid, F. (2012), 'Two Canonical VARMA Forms: Scalar Component Models Vis-à-Vis the Echelon Form', *Econometric Reviews* **31**(1), 60–83.

Athanasopoulos, G. and Vahid, F. (2008), 'A Complete VARMA Modelling Methodology Based on Scalar Components', *Journal of Time Series Analysis* **29**(3), 533–554.

Clements, M. P. and Hendry, D. F. (1993), 'On the Limitations of Comparing Mean Squared Forecast Errors (with discussion)', *Journal of Forecasting* **12**, 617–637.

Diebold, F. X. and Mariano, R. S. (1995), 'Comparing Predictive Accuracy', *Journal of Business and Economic Statistics* **13**, 253–263.

Dufour, J.-M. and Pelletier, D. (2008), 'Practical Methods for Modelling Weak VARMA Processes: Identification, Estimation and Specification with a Macroeconomic Application', *Discussion Paper,* Department of Economics, McGill University, CIREQ and CIRANO .

Engle, R. F. and Granger, C. W. J. (1987), 'Co-integration and Error Correction: Representation, Estimation, and Testing', *Econometrica* **55**, 251–276.

Giacomini, R. and White, H. (2006), 'Tests of Conditional Predictive Ability', *Econometrica* **74**, 1545–1578.

Gonzalo, J. and Pitarakis, J. (1995), 'Specification via Model Selection in Vector Error Correction Models', *Economic Letters* **60**, 321–328.

Granger, C. (1981), 'Some Properties of Time Series Data and Their Use in Econometric Model Specification', *Journal of Econometrics* **16**(1), 121 – 130.

Hall, A. D., Anderson, H. M. and Granger, C. W. J. (1992), 'A Cointegration Analysis of Treasury Bill Yields', *The Review of Economics and Statistics* **74**(1), pp. 116–126.

Hamilton, J. D. (1994), *Time Series Analysis*, Princeton University Press, Princeton.

Hannan, E. J. and Deistler, M. (1988), *The Statistical Theory of Linear Systems*, John Wiley.

Hannan, E. J. and Kavalieris, L. (1984), 'Multivariate Linear Time Series Models', *Advances in Applied Probability* **16**, 492–561.

Hannan, E. J. and Rissanen, J. (1982), 'Recursive Estimation of Mixed Autoregressive-Moving Average Order', *Biometrika* **69**, 81–94.

Johansen, S. (1988), 'Statistical Analysis of Cointegrating Vectors', *Journal of Economic Dynamics and Control* **12**, 231–254.

Johansen, S. (1991), 'Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models', *Econometrica* **59**, 1551–1580.

Johansen, S. (1995), *Likelihood-based Inference on Cointegrated Vector Auto-regressive Models*, Oxford University Press, Oxford.

Kapetanios, G. (2003), 'A Note on an Iterative Least-squares Estimation Method for ARMA and VARMA Models', *Economics Letters* **79**, 305–312.

Kascha, C. (2012), 'A Comparison of Estimation Methods for Vector Autoregressive Moving Average Models', *Econometric Reviews* **31**(3), 297–324.

Kascha, C. and Trenkler, C. (2011), Cointegrated VARMA models and Forecasting US Interest Rates, ECON - Working Papers econwp033, Department of Economics - University of Zurich.

Koreisha, S. and Pukkila, T. (1990), 'A Generalized Least-squares Approach for Estimation of Autoregressive Moving Average Models', *Journal of Time Series Analysis* **11**(2), 139–151.

Lütkepohl, H. (2005), *The New Introduction to Multiple Time Series Analysis*, Springer-Verlag, Berlin.

Lütkepohl, H. and Claessen, H. (1997), 'Analysis of cointegrated VARMA processes', *Journal of Econometrics* **80**, 223–239.

Lütkepohl, H. and Poskitt, D. S. (1996), 'Specification of Echelon Form VARMA Models', *Journal of Business and Economic Statistics* **14**, 69–79.

Lütkepohl, H. and Poskitt, D. S. (1998), Consistent Estimation of the Number of Cointegration Relations in a Vector Autoregressive Model, Working papers, Humboldt University, Sonderforschungsbereich 373.

Lütkepohl, H. and Saikkonen, P. (1999), 'Order Selection in Testing for the Cointegration Rank of a VAR Process', *Cointegration, Causality and Forecasting. A Festscherift in Honour of Clive W.J. Granger* pp. 168–199.

Phillips, P. C. B. and Durlauf, S. N. (1986), 'Multiple Time Series Regression with Integrated Processes', *Review of Economic Studies* **53**(4), 473–95.

Poskitt, D. S. (1992), 'Identification of Echelon Canonical Forms for Vector Linear Processes Using Least Squares', *The Annals of Statistics* **20**, 195–215.

Poskitt, D. S. (2000), 'Strongly Consistent Determination of Cointegrating Rank via Canonical Correlations', *Journal of Business and Economic Statistics* **18**, 77–90.

Poskitt, D. S. (2003), 'On the Specification of Cointegrated Autoregressive Moving Average Forecasting Systems', *International Journal of Forecasting* **19**, 503–519.

Stock, J. H. and Watson, M. W. (1988), 'Testing for Common Trends', *Journal of the American Statistical Association* **83**(404), pp. 1097–1107.

Tiao, G. C. and Tsay, R. S. (1989), 'Model Specification in Multivariate Time Series (with discussion)', *Journal of the Royal Statistical Society B* **51**, 157–213.

Vahid, F. and Issler, J. V. (2002), 'The Importance of Common Cyclical Features in VAR Analysis: A Monte-Carlo Study', *Journal of Econometrics* **109**, 341–363.

West, K. D. (1996), 'Asymptotic Inference About Predictive Ability', *Econometrica* **64**, 1067–1084.

Yang, M. and Bewley, R. (1996), 'On Cointegration Tests for VAR Models with Drift', *Economics Letters* **51**(1), 45 – 50.

Tasmanian School of Business and Economics
Working Paper Series

2014-08   How Many Stocks are Enough for Diversifying Canadian Institutional Portfolios? **Vitali Alexeev** and Francis Tapon

2014-07   Forecasting with EC-VARMA Models, George Athanasopoulos, Don Poskitt, Farshid Vahid, **Wenying Yao**

2014-06   Canadian Monetary Policy Analysis using a Structural VARMA Model, **Mala Raghavan**, George Athanasopoulos, Param Silvapulle

2014-05   The sectorial impact of commodity price shocks in Australia, S. Knop and **Joaquin Vespignani**

2014-04   Should ASEAN-5 monetary policymakers act pre-emptively against stock market bubbles? **Mala Raghavan** and **Mardi Dungey**

2014-03   Mortgage Choice Determinants: The Role of Risk and Bank Regulation**, Mardi Dungey, Firmin Doko Tchatoka, Graeme Wells, Maria B. Yanotti**

2014-02   Concurrent momentum and contrarian strategies in the Australian stock market, Minh Phuong Doan, **Vitali Alexeev**, Robert Brooks

2014-01   A Review of the Australian Mortgage Market**, Maria B. Yanotti**

2013-20   Towards a Diagnostic Approach to Climate Adaptation for Fisheries, P. Leith, E. Ogier, G. Pecl, **E. Hoshino**, J. Davidson, M. Haward

2013-19   Equity Portfolio Diversification with High Frequency Data, **Vitali Alexeev** and **Mardi Dungey**

2013-18   Measuring the Performance of Hedge Funds Using Two-Stage Peer Group Benchmarks, Marco Wilkens, Juan Yao, **Nagaratnam Jeyasreedharan** and Patrick Oehler

2013-17   What Australian Investors Need to Know to Diversify their Portfolios, **Vitali Alexeev** and Francis Tapon

2013-16   Equity Portfolio Diversification: How Many Stocks are Enough? Evidence from Five Developed Markets**, Vitali Alexeev** and **Francis Tapon**

2013-15   Equity market Contagion during the Global Financial Crisis: Evidence from the World's Eight Largest Economies, **Mardi Dungey** and **Dinesh Gajurel**

2013-14   A Survey of Research into Broker Identity and Limit Order Book, **Thu Phuong Pham** and P Joakim Westerholm

2013-13   Broker ID Transparency and Price Impact of Trades: Evidence from the Korean Exchange, **Thu Phuong Pham**

2013-12   An International Trend in Market Design: Endogenous Effects of Limit Order Book Transparency on Volatility, Spreads, depth and Volume, **Thu Phuong Pham** and P Joakim Westerholm

UNIVERSITY *of* TASMANIA
AUSTRALIA