Exact Maximum Likelihood Estimation of Partially Nonstationary Vector ARMA Models

José Alberto Mauricio

Universidad Complutense de Madrid

Abstract: A useful class of partially nonstationary vector autoregressive moving average (VARMA) models is considered with regard to parameter estimation. An exact maximum likelihood (EML) approach is developed on the basis of a simple transformation applied to the error-correction representation of the models considered. The employed transformation is shown to provide a standard VARMA model with the important property that it is stationary. Parameter estimation can thus be carried out by applying standard EML methods to the stationary VARMA model obtained from the error-correction representation. This approach resolves at least two problems related to the current limited availability of EML estimation methods for partially nonstationary VARMA models. Firstly, it resolves the apparent impossibility of computing the exact log-likelihood for such models using currently available methods. And secondly, it resolves the inadequacy of considering lagged endogenous variables as exogenous variables in the error-correction representation. Theoretical discussion is followed by an example using a popular data set. The example illustrates the feasibility of the EML estimation approach as well as some of its potential benefits in cases of practical interest which are easy to come across. As in the case of stationary models, the proposed EML method provides estimated model structures that are more reliable and accurate than results produced by conditional methods.

Keywords: Cointegration; Exact maximum likelihood estimation; Partially nonstationary model; Unit roots; Vector autoregressive moving average model; Vector error-correction model.

First published in *Computational Statistics and Data Analysis*, Vol. 50, No. 12, pp. 3644-3662

August 2006

1. INTRODUCTION

Considerable interest in describing joint nonstationarity among components of multiple time series processes has arisen, especially for economic time series, since the work on cointegration of the 2003 Nobel Laureates in Economics, Robert F. Engle and Clive W.J. Granger (Engle and Granger, 1987). Both earlier (especially Box and Tiao, 1977) and subsequent research on the subject (see, from a long list of contributions, the surveys by Watson, 1994; Reinsel, 1997, Chapter 6; Johansen, 2001; and the references cited therein) have shown the importance of representing adequately the nonstationary structure among components of a vector time series process in order to (i) ensure proper model estimation and testing, (ii) increase understanding of the process under study, (iii) forecast such process efficiently, and, in general, (iv) avoid difficulties in parameter identification that may result from multivariate overdifferencing.

Among the many techniques developed to cope with the above aims, inference on nonstationary systems in the context of partially nonstationary vector autoregressive (VAR) models has played a prominent role in applied research since the works of Johansen (1988, 1991), Ahn and Reinsel (1990), and Reinsel and Ahn (1992). More recently, Yap and Reinsel (1995), Saikkonen and Lütkepohl (1996), Lütkepohl and Claessen (1997), Saikkonen and Luukkonen (1997), and Bauer and Wagner (2002), have extended those earlier works to the context of partially nonstationary vector autoregressive moving average (VARMA) models, which, in practice, may provide more parsimonious representations for many vector time series processes than pure VAR models. Asymptotic optimal inference has been established in every case in line with results given in Phillips (1991), some of which have been summarized recently by Tanaka (1996) and White (2001).

Although the works mentioned above differ in scope and complexity, all of them share one common feature, namely the estimation of parameters through some suitable adaptation of either conditional least squares or conditional maximum likelihood. It is well known from the existing literature on stationary VARMA models (see, for example, Mauricio, 1995, 2002, and the references cited therein) that conditional estimation methods may, and often do, lead to inefficient use of available information, which in turn may, and often does, result in unreliable and inaccurate estimated model structures. However, unconditional, exact maximum likelihood (EML) estimation methods have received very limited attention in the existing literature on partially nonstationary VARMA models. Mélard et al. (2006) appears to be the only reference in this respect. Although asymptotically equivalent to conditional methods, EML estimation of such models may serve in many practical situations as a helpful instrument for providing as much information as possible from a single data set about its generating time series process, which is especially important for small data sets.

There seem to exist at least three reasons for the current limited availability of EML estimation methods for partially nonstationary models. Firstly, it is impossible to compute the exact log-likelihood associated with such models through available methods for stationary models. Secondly, it is quite inadequate to consider the vector error-correction (VEC) representation of partially nonstationary VARMA models as a standard, stationary VARMA model with exogenous variables. And thirdly, it is a firmly established belief that conditional methods are as reliable in practice as unconditional methods because of their asymptotic equivalence. This last reason alone seems to render both the development and the usage of unconditional estimation methods unnecessarily complicated.

This article shows that none of the preceding reasons justifies, as a general rule in applied work, routine and uncared-for application of conditional methods when estimating partially nonstationary models in practice. In order to support this claim, it is shown first, in close relation with the developments outlined by Mélard et al. (2006, Section 3), that EML estimation of partially nonstationary VARMA models is possible and easy to implement in practice. In this regard, a simpler approach than that of Mélard et al. (2006, Section 3) is proposed and developed in detail. Additionally, the proposed approach does not suffer from nonuniqueness issues and is not tied to the state-space framework. Then, in-depth analysis of an example illustrates that real situations can occur in which EML delivers more reliable estimated structures than conditional methods.

Specifically, the theoretical part of this article develops one feasible way of estimating jointly through EML all the parameters of the VEC representation of a useful class of partially nonstationary VARMA models. That class of models is briefly reviewed in Section 2. In Section 3, it is shown that a simple and uniquely identified transformation of a VEC model results in a fully-equivalent VARMA model with the important features that (i) it is a stationary model, and (ii) its parameters are explicit functions of the parameters in the corresponding VEC model. Hence, EML estimation of such parameters can be carried out by means of estimating the equivalent, stationary VARMA model through EML. The applied part of the article is given in Section 4, which provides an illustration of the feasibility of the EML estimation approach and of its potential practical advantages over conditional methods. Finally, in Section 5 conclusions are summarized.

2. ERROR-CORRECTION REPRESENTATION OF PARTIALLY NONSTATIONARY VECTOR ARMA MODELS

A useful class of partially nonstationary VARMA models is introduced below, and some of its well-established properties (see, for example, Watson, 1994; Reinsel, 1997, Section 6.3; and Pham et al., 2003, Section 2) are described. In particular, the error-correction representation of such class of models is reviewed, especially with regard to parameter estimation. This section also serves to introduce the notation that will be used throughout this article.

Consider a vector time series process $\{\mathbf{Y}_t\}$ of dimension $M \geq 2$, with $\mathbf{Y}_t = [Y_{t1}, \ldots, Y_{tM}]^{\mathrm{T}}$, following the vector ARMA(p, q) model

$$\boldsymbol{\Phi}(L)\mathbf{Y}_t = \boldsymbol{\Theta}(L)\mathbf{A}_t, \tag{1}$$

where $\boldsymbol{\Phi}(L) = \mathbf{I}_M - \sum_{i=1}^{p} \boldsymbol{\Phi}_i L^i$ and $\boldsymbol{\Theta}(L) = \mathbf{I}_M - \sum_{i=1}^{q} \boldsymbol{\Theta}_i L^i$ are matrix polynomials in L of degrees p and q, L is the lag (backshift) operator, $\boldsymbol{\Phi}_i$ ($i = 1, \ldots, p$) and $\boldsymbol{\Theta}_i$ ($i = 1, \ldots, q$) are $M \times M$ parameter matrices, and $\{\mathbf{A}_t\}$ is a sequence of Gaussian IID$(\mathbf{0}, \boldsymbol{\Sigma})$ $M \times 1$ random vectors. For model (1), it is assumed that the polynomial equations $|\boldsymbol{\Phi}(x)| = 0$ and $|\boldsymbol{\Theta}(x)| = 0$ have no roots inside the unit circle.

Additional conditions on $\boldsymbol{\Phi}(L)$ and $\boldsymbol{\Theta}(L)$ for parameter identifiability, such as those considered by Yap and Reinsel (1995), are also assumed.

The VARMA model (1) is said to be partially nonstationary when $|\boldsymbol{\Phi}(x)| = 0$ has D roots equal to one (i.e., D unit roots), with $0 < D < M$, and all other roots outside the unit circle. If it is further assumed that $P = \text{rank}[\boldsymbol{\Phi}(1)] = M - D$, with $\boldsymbol{\Phi}(1) = \mathbf{I}_M - \sum_{i=1}^{p} \boldsymbol{\Phi}_i$ having D zero eigenvalues, then it can be shown that $\{\mathbf{Y}_t\}$ is a nonstationary process such that $\{\nabla\mathbf{Y}_t\} = \{\mathbf{Y}_t - \mathbf{Y}_{t-1}\}$ is stationary, and that there exist P distinct linear combinations of $\{\mathbf{Y}_t\}$ that are stationary. In this case, the M components of $\{\mathbf{Y}_t\}$ are said to be cointegrated with cointegrating rank P. Many nonstationary multiple time series found in practice (especially for economic time series) seem to be compatible with these two properties, which means that the class of models introduced above represents a valuable tool in describing dynamic relationships among nonstationary time series and in forecasting them.

When the M components of $\{\mathbf{Y}_t\}$ are cointegrated, direct estimation of (1) without further restrictions is frequently possible (especially through conditional methods), although it is never optimal in the sense given by Phillips (1991). Hence, a great deal of research on partially nonstationary models has concentrated on representing (1) as a model on which asymptotically optimal inference can be carried out and on developing methods for making such inference in practice. Perhaps the most useful representation of (1) to this end is the VEC model

$$\nabla\mathbf{Y}_t = -\boldsymbol{\Pi}\mathbf{Y}_{t-1} + \sum_{i=1}^{p-1} \mathbf{F}_i \nabla\mathbf{Y}_{t-i} + \boldsymbol{\Theta}(L)\mathbf{A}_t\,, \tag{2}$$

where

$$\boldsymbol{\Pi} = \boldsymbol{\Phi}(1) = \mathbf{I}_M - \sum_{i=1}^{p} \boldsymbol{\Phi}_i\,, \text{ and } \mathbf{F}_i = -\sum_{j=i+1}^{p} \boldsymbol{\Phi}_j \quad (i = 1,\ldots, p-1)\,. \tag{3}$$

Note that (2) and (3) are just a convenient rearrangement of (1), so that the VEC model (2)-(3) and the VARMA model (1) are equivalent representations for $\{\mathbf{Y}_t\}$. Note also that this equivalence holds irrespective of whether (1) is a stationary model or not. However, when model (1) satisfies the assumptions for partial nonstationarity, it can be shown (Reinsel, 1997, Section 6.3; and Yap and Reinsel, 1995) that (i) $\mathbf{F}(L) = \mathbf{I}_M - \sum_{i=1}^{p-1} \mathbf{F}_i L^i$ is a stationary operator having all roots of $|\mathbf{F}(x)| = 0$ outside the unit circle, (ii) the $M \times M$ matrix $\boldsymbol{\Pi} = \boldsymbol{\Phi}(1)$, which is of reduced rank $P = M - D$, can be written as

$$\boldsymbol{\Pi} = \boldsymbol{\Lambda}\mathbf{B}^{\mathrm{T}}\,, \tag{4}$$

where $\boldsymbol{\Lambda}$ and $\mathbf{B}$ are $M \times P$ matrices of full column rank, and (iii) the process $\{\mathbf{W}_t\} = \{\mathbf{B}^{\mathrm{T}}\mathbf{Y}_t\}$ is stationary for a suitably chosen matrix $\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_P]$, so that $\mathbf{B}^{\mathrm{T}}\mathbf{Y}^* = \mathbf{0}$ (with $\mathbf{Y}^* = [Y_1^*, \ldots, Y_M^*]^{\mathrm{T}}$) represents P cointegrating or long-term equilibrium relations of the type $\mathbf{b}_i^{\mathrm{T}}\mathbf{Y}^* = 0$ $(i = 1,\ldots, P)$ among the M components of $\{\mathbf{Y}_t\}$. The P linearly independent columns of $\mathbf{B}$ are usually referred to as cointegrating vectors, in the sense that $\{W_{ti}\} = \{\mathbf{b}_i^{\mathrm{T}}\mathbf{Y}_t\}$ is a scalar stationary process for every $i = 1,\ldots, P$, representing short-term deviations of $\{\mathbf{Y}_t\}$ from a cointegrating or long-term equilibrium relation among its components.

Hence, under the partial nonstationarity assumptions, (2) can be written in full as

$$\left[\mathbf{I}_M - \sum_{i=1}^{p-1} \mathbf{F}_i L^i\right]\nabla\mathbf{Y}_t = -\boldsymbol{\Lambda}\mathbf{B}^{\mathrm{T}}\mathbf{Y}_{t-1} + \left[\mathbf{I}_M - \sum_{i=1}^{q} \boldsymbol{\Theta}_i L^i\right]\mathbf{A}_t\,. \tag{5}$$

Equation (5) is a particularly convenient representation for estimation purposes, due to the fact that the nonstationarity of $\mathbf{\Phi}(L)$ in (1) is concentrated in (5) in the full column rank parameter matrices $\mathbf{\Lambda}$ and $\mathbf{B}$. In this regard, it is also convenient for parameter identifiability to normalize $\mathbf{B}$ so that

$$\mathbf{B} = \begin{bmatrix} \mathbf{I}_P \\ \mathbf{B}_2 \end{bmatrix}, \tag{6}$$

where $\mathbf{B}_2$ is an $(M-P)\times P$ parameter matrix which, along with $\mathbf{\Lambda}$, $\mathbf{F}_i$, $\mathbf{\Theta}_i$, and $\mathbf{\Sigma} = \mathrm{Var}[\mathbf{A}_t]$ in (5), must be estimated. Note that the elements of the $M\times 1$ vector $\mathbf{Y}_t$ can always be arranged so that the normalization given in (6) is possible, and implicit in this arrangement is that if $\mathbf{Y}_t$ is partitioned as

$$\mathbf{Y}_t = \begin{bmatrix} \mathbf{Y}_{t1} \\ \mathbf{Y}_{t2} \end{bmatrix}, \tag{7}$$

where $\mathbf{Y}_{t1}$ is $P\times 1$ and $\mathbf{Y}_{t2}$ is $(M-P)\times 1$, then there is no cointegration among the components of $\{\mathbf{Y}_{t2}\}$. In order to avoid potential difficulties associated with a wrong normalization of $\mathbf{B}$, the procedures developed, for example, by Luukkonen et al. (1999) and Kurozumi (2005) might be used in practice to check whether the applied normalization and the corresponding arrangement of $\mathbf{Y}_t$ are appropriate.

Note finally that for any $M\times(M-P)$ full column rank matrix $\mathbf{P}$ such that the columns of $\mathbf{P}$ are orthogonal to the column space of $\mathbf{\Lambda}$, an integrated process $\{\mathbf{V}_t\} = \{\mathbf{P}^{\mathrm{T}}\mathbf{Y}_t\}$ of dimension $M-P$ can be obtained so that $\{\mathbf{Y}_t\}$ may be expressed as a linear combination of the "transitory disequilibrium" (stationary) process $\{\mathbf{W}_t\}$ and the "common-trend" (purely nonstationary) process $\{\mathbf{V}_t\}$.

Estimation of the VEC model given in (5)-(6) through conditional methods has been considered by several authors, as noted in Section 1. In the next section, it is shown that EML is also possible by means of expressing the VEC model (5)-(6) as a fully-equivalent stationary VARMA model. It may be noted that the computation of accurate approximations to the exact likelihood function for partially nonstationary VARMA models, has been considered previously by Luceño (1994) and Ma (1997), although their approaches are based directly on the representation given in (1) without considering the restrictions given in (4) on the autoregressive operator. This implies that inference on parameters in the VEC model (5), especially on the elements of $\mathbf{\Lambda}$ and $\mathbf{B}$ in (4)-(6), is quite complicated, if not impossible at all, in practice, and that such inference, when possible, might not be optimal in the sense given by Phillips (1991).

3. STATIONARY VARMA REPRESENTATION AND ESTIMATION OF VECTOR ERROR-CORRECTION MODELS

Apparent reasons for the current limited availability of EML estimation methods for partially nonstationary models, include (i) the impossibility of computing through standard methods the exact likelihood associated with model (1) when $|\mathbf{\Phi}(x)| = 0$ has unit roots, and (ii) the presence of the error-correction term $-\mathbf{\Lambda}\mathbf{B}^{\mathrm{T}}\mathbf{Y}_{t-1}$ in the right side of (5), which does not allow for considering (5) as a standard, stationary VARMA model with exogenous variables. In this section it is shown that such reasons are, in fact, only apparent in the sense that it is possible to apply a simple and uniquely identified transformation to the VEC model given in (5)-(6) in order to obtain a standard, stationary VARMA model. That model preserves not only the full dynamic structure implicit in (5), but also, in a one-to-one correspondence, the full set of parameter matrices $\mathbf{B}_2$, $\mathbf{\Lambda}$, $\mathbf{F}_i$, $\mathbf{\Theta}_i$, and $\mathbf{\Sigma} = \mathrm{Var}[\mathbf{A}_t]$ appearing in (5)-(6). A closely related transformation has been outlined by Mélard et al. (2006, Section 3), although it is more complicated than the one presented below, and it is not uniquely identified. Also, a somewhat related transformation for pure autoregressive models can be found in Watson (1994), although the procedure given therein omits the fundamental step for EML estimation purposes of obtaining a stationary VARMA representation from a VEC model.

The general transformation of the VEC model (5) into a stationary VARMA model is described now. First, rewrite (5) as in (2), i.e.,

$$\nabla \mathbf{Y}_t = -\mathbf{\Pi}\mathbf{Y}_{t-1} + \sum_{i=1}^{p-1} \mathbf{F}_i \nabla \mathbf{Y}_{t-i} + \mathbf{\Theta}(L)\mathbf{A}_t \,, \tag{8}$$

and note that (4), (6) and (7) imply that

$$\mathbf{\Pi}\mathbf{Y}_{t-1} = \mathbf{\Lambda}[\mathbf{I}_P, \mathbf{B}_2^{\mathrm{T}}]\begin{bmatrix} \mathbf{Y}_{t-1,1} \\ \mathbf{Y}_{t-1,2} \end{bmatrix} = \mathbf{\Lambda}\mathbf{W}_{t-1} = [\mathbf{0}, \mathbf{\Lambda}]\begin{bmatrix} \nabla\mathbf{Y}_{t-1,2} \\ \mathbf{W}_{t-1} \end{bmatrix} = \bar{\mathbf{\Lambda}}\bar{\mathbf{Y}}_{t-1}, \tag{9}$$

where $\mathbf{W}_{t-1} = \mathbf{Y}_{t-1,1} + \mathbf{B}_2^{\mathrm{T}}\mathbf{Y}_{t-1,2}$, $\bar{\mathbf{Y}}_{t-1} = [\nabla\mathbf{Y}_{t-1,2}^{\mathrm{T}}, \mathbf{W}_{t-1}^{\mathrm{T}}]^{\mathrm{T}}$, and the $M \times M$ matrix $\bar{\mathbf{\Lambda}}$ is given by

$$\bar{\mathbf{\Lambda}} = [\mathbf{0}, \mathbf{\Lambda}], \tag{10}$$

with $\mathbf{0}$ representing an $M \times (M-P)$ zero matrix, and $\mathbf{\Lambda}$ being the $M \times P$ matrix introduced in (4).

Define now the $M \times M$ matrix

$$\bar{\mathbf{C}} = \begin{bmatrix} \mathbf{0} & \mathbf{I}_{M-P} \\ \mathbf{I}_P & \mathbf{B}_2^{\mathrm{T}} \end{bmatrix}, \tag{11}$$

where $\mathbf{0}$ represents an $(M-P) \times P$ zero matrix, and $\mathbf{B}_2$ is the $(M-P) \times P$ matrix given implicitly in (4), and explicitly in (6). Hence, premultiplying $\nabla\mathbf{Y}_t = [\nabla\mathbf{Y}_{t1}^{\mathrm{T}}, \nabla\mathbf{Y}_{t2}^{\mathrm{T}}]^{\mathrm{T}}$ by $\bar{\mathbf{C}}$ in (11) gives

$$\bar{\mathbf{C}}\nabla\mathbf{Y}_t = \begin{bmatrix} \nabla\mathbf{Y}_{t2} \\ \nabla\mathbf{Y}_{t1} + \mathbf{B}_2^{\mathrm{T}}\nabla\mathbf{Y}_{t2} \end{bmatrix} = \begin{bmatrix} \nabla\mathbf{Y}_{t2} \\ \mathbf{W}_t - \mathbf{W}_{t-1} \end{bmatrix},$$

or, equivalently,

$$\bar{\mathbf{C}}\nabla\mathbf{Y}_t = \begin{bmatrix} \nabla\mathbf{Y}_{t2} \\ \mathbf{W}_t \end{bmatrix} - \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_P \end{bmatrix}\begin{bmatrix} \nabla\mathbf{Y}_{t-1,2} \\ \mathbf{W}_{t-1} \end{bmatrix} = \bar{\mathbf{Y}}_t - \bar{\mathbf{H}}\bar{\mathbf{Y}}_{t-1}, \tag{12}$$

where the $M \times M$ matrix $\bar{\mathbf{H}}$ is given by

$$\bar{\mathbf{H}} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_P \end{bmatrix}. \tag{13}$$

Note carefully that the stationary process $\{\bar{\mathbf{Y}}_t\}$, appearing in (9) and (12), has been defined so that its first $M-P$ components (those in $\nabla\mathbf{Y}_{t2}$) refer to the elements of $\mathbf{Y}_t$ that are not associated with the normalized columns of the cointegrating matrix (6). This ensures that the vector ARMA model for $\{\bar{\mathbf{Y}}_t\}$ obtained below is uniquely identified.

Now, premultiplying (8) by (11), and writing each $\mathbf{F}_i\nabla\mathbf{Y}_{t-i}$ in (8) as $\mathbf{F}_i\bar{\mathbf{C}}^{-1}\bar{\mathbf{C}}\nabla\mathbf{Y}_{t-i}$, gives $\bar{\mathbf{C}}\nabla\mathbf{Y}_t = -\bar{\mathbf{C}}\mathbf{\Pi}\mathbf{Y}_{t-1} + \sum_{i=1}^{p-1}\bar{\mathbf{C}}\mathbf{F}_i\bar{\mathbf{C}}^{-1}\bar{\mathbf{C}}\nabla\mathbf{Y}_{t-i} + \bar{\mathbf{C}}\mathbf{\Theta}(L)\mathbf{A}_t$, or using (9) and (12),

$$\bar{\mathbf{Y}}_t - \bar{\mathbf{H}}\bar{\mathbf{Y}}_{t-1} = -\bar{\mathbf{C}}\bar{\mathbf{\Lambda}}\bar{\mathbf{Y}}_{t-1} + \sum_{i=1}^{p-1}\left[\,\bar{\mathbf{C}}\mathbf{F}_i\bar{\mathbf{C}}^{-1}(\bar{\mathbf{Y}}_{t-i} - \bar{\mathbf{H}}\bar{\mathbf{Y}}_{t-i-1})\,\right] + \bar{\mathbf{C}}\mathbf{\Theta}(L)\mathbf{A}_t\,,$$

which, on premultiplying by

$$\bar{\mathbf{C}}^{-1} = \begin{bmatrix} -\mathbf{B}_2^{\mathrm{T}} & \mathbf{I}_P \\ \mathbf{I}_{M-P} & \mathbf{0} \end{bmatrix}, \tag{14}$$

can be written equivalently as

$$\bar{\mathbf{C}}^{-1}\bar{\mathbf{Y}}_t = (\bar{\mathbf{C}}^{-1}\bar{\mathbf{H}} - \bar{\mathbf{\Lambda}})\bar{\mathbf{Y}}_{t-1} + \sum_{i=1}^{p-1}\left[\,\mathbf{F}_i\bar{\mathbf{C}}^{-1}\bar{\mathbf{Y}}_{t-i} - \mathbf{F}_i\bar{\mathbf{C}}^{-1}\bar{\mathbf{H}}\bar{\mathbf{Y}}_{t-i-1}\,\right] + \mathbf{\Theta}(L)\mathbf{A}_t\,.$$

Finally, this last expression can also be written as

$$\bar{\mathbf{\Phi}}_0\bar{\mathbf{Y}}_t = \sum_{i=1}^{p}\bar{\mathbf{\Phi}}_i\bar{\mathbf{Y}}_{t-i} + \mathbf{\Theta}(L)\mathbf{A}_t\,, \tag{15}$$

where

$$\begin{gathered} \bar{\mathbf{\Phi}}_0 = \bar{\mathbf{C}}^{-1},\quad \bar{\mathbf{\Phi}}_1 = \bar{\mathbf{C}}^{-1}\bar{\mathbf{H}} - \bar{\mathbf{\Lambda}} + \mathbf{F}_1\bar{\mathbf{C}}^{-1}, \\ \bar{\mathbf{\Phi}}_i = \mathbf{F}_i\bar{\mathbf{C}}^{-1} - \mathbf{F}_{i-1}\bar{\mathbf{C}}^{-1}\bar{\mathbf{H}}\quad (i = 2,\ldots,p-1),\quad \bar{\mathbf{\Phi}}_p = -\mathbf{F}_{p-1}\bar{\mathbf{C}}^{-1}\bar{\mathbf{H}}. \end{gathered} \tag{16}$$

Equations (15) and (16), together with

$$\bar{\mathbf{Y}}_t = \begin{bmatrix} \nabla\mathbf{Y}_{t2} \\ \mathbf{W}_t \end{bmatrix} = \begin{bmatrix} \nabla\mathbf{Y}_{t2} \\ \mathbf{Y}_{t1} + \mathbf{B}_2^{\mathrm{T}}\mathbf{Y}_{t2} \end{bmatrix}, \tag{17}$$

and equations (10), (13) and (14) for $\bar{\mathbf{\Lambda}}$, $\bar{\mathbf{H}}$ and $\bar{\mathbf{C}}^{-1}$, respectively, constitute a stationary VARMA representation which is equivalent to the VEC model given in (5)-(6), or in (8), (4) and (6). An analytical proof of stationarity for the VARMA model (15)-(17) might be possible in line with results outlined by Johansen (2001, 2003) for pure autoregressive models. However, a general proof would certainly not be more illuminating than the proof for a particularly simple case (e.g., a partially nonstationary VAR(1) model with $M = 2$; see Mauricio, 2006). In any case, a general proof seems clearly unnecessary because it can be replaced by the following simple, yet both general and rigorous, argument:

1. As explained in Section 2, the partial nonstationarity assumptions imply that the two components of $\{\bar{\mathbf{Y}}_t\}$ in (17) are stationary, so that the general stochastic process $\{\bar{\mathbf{Y}}_t\}$ is certainly stationary.
2. The derivations leading to equations (15) and (16) show that the stationary process $\{\bar{\mathbf{Y}}_t\}$ follows a

uniquely identified vector ARMA model.

3. If $\{\bar{\mathbf{Y}}_t\}$ is a stationary process following a uniquely identified vector ARMA model, then such model is necessarily stationary.

Some remarks on EML estimation of the VEC model (5)-(6), or, equivalently, the stationary VARMA model (15)-(17), are given now.

Remark 1. The following procedure can be used to obtain EML estimates of the parameter matrices $\mathbf{B}_2$, $\mathbf{\Lambda}$, $\mathbf{F}_i$, $\mathbf{\Theta}_i$, and $\mathbf{\Sigma}$ in the VEC model (5)-(6):

1. Choose a suitable initial guess (perhaps through some conditional method) for every parameter to be estimated, and set up a parameter vector $\mathbf{x}$ containing the full set of initial estimates.
2. Update $\mathbf{x}$ numerically through nonlinear optimization of the exact log-likelihood function of the stationary VARMA model (15)-(17). To this end, note (i) that whenever a computation of the exact log-likelihood is called for, the parameter matrices in (15)-(17) are explicit functions of the elements of $\mathbf{x}$, and (ii) that (15) can be written as

$$\left(\mathbf{I}_M - \textstyle\sum_{i=1}^{p} \mathbf{\Phi}_i^* L^i\right) \bar{\mathbf{Y}}_t = \left(\mathbf{I}_M - \textstyle\sum_{i=1}^{q} \mathbf{\Theta}_i^* L^i\right) \mathbf{A}_t^*, \tag{18}$$

   where $\mathbf{\Phi}_i^* = \bar{\mathbf{\Phi}}_0^{-1}\bar{\mathbf{\Phi}}_i$, $\mathbf{\Theta}_i^* = \bar{\mathbf{\Phi}}_0^{-1}\mathbf{\Theta}_i\bar{\mathbf{\Phi}}_0$, and $\mathbf{A}_t^* = \bar{\mathbf{\Phi}}_0^{-1}\mathbf{A}_t$. Equation (18) represents a stationary, standard VARMA model for which well-established and efficient algorithms exist in order to compute its associated exact log-likelihood within several operational frameworks (such as those given in Shea, 1989, within the state-space framework; and Mauricio, 1997, 2002, outside that framework).

Remark 2. The parameter vector $\mathbf{x}$ of the preceding remark can be defined so it contains any coefficients on which the researcher is ultimately interested, as long as all the elements of both $\bar{\mathbf{Y}}_t$ in (17) and the parameter matrices in (15)-(17) can be written as explicit functions of $\mathbf{x}$ and of available data on $\{\mathbf{Y}_t\}$. This means that (i) the EML estimation procedure given in Remark 1 can accommodate a wide range of restrictions (not only exclusion restrictions) on the parameters of the VEC model (5)-(6), and (ii) that intervention effects on components of $\{\mathbf{Y}_t\}$ can be estimated jointly with the VEC model parameters, which can be of great importance in many empirical applications.

Remark 3. From a practical point of view, implementation of the procedures described in the previous remarks requires some programming in the context of (sufficiently flexible) numerical linear algebra and nonlinear optimization software packages, or, alternatively, moderate work on coding in some high-level language. In particular, both EML and conditional maximum likelihood (CML) estimation of models in the form given in (15)-(17), have been implemented as a set of programs written in the C programming language, whose full source code is freely available upon request for non-commercial purposes.

Remark 4. The numerical algorithms implemented in the C programs mentioned above follow the guidelines given in Mauricio (1995, 1996, 1997) and in Shea (1984, 1989). Specifically, following the general design presented in Mauricio (1996), a "user" C function (subroutine) has been coded implementing the operations described in (15)-(18), which cast the VEC model (5)-(6) into its standard VARMA representation. For the sake of computational accuracy and efficiency, the coding takes full account of the zero restrictions contained in the autoregressive matrices displayed in (16). This user function is called prior to every computation of the associated log-likelihood, which is carried out for the exact case through the methods of Mauricio (1997) or Shea (1989) (both options are available). For the conditional case, a simple function has been coded expressly. The objective function which has to be minimized in order to obtain parameter estimates, has been coded as explained in Mauricio (1995) and Shea (1984). Finally, numerical minimization of the objective function is carried out as described in detail in Mauricio (1995).

Remark 5. Equation (18) can be written as $\boldsymbol{\Phi}^*(L)\bar{\mathbf{Y}}_t = \boldsymbol{\Theta}^*(L)\mathbf{A}_t^*$, where $\boldsymbol{\Phi}^*(L) = \mathbf{I}_M - \sum_{i=1}^{p} \boldsymbol{\Phi}_i^* L^i$ is a standard, stationary autoregressive operator having all roots of $|\boldsymbol{\Phi}^*(x)| = 0$ outside the unit circle, and $\boldsymbol{\Theta}^*(L) = \mathbf{I}_M - \sum_{i=1}^{q} \boldsymbol{\Theta}_i^* L^i$ is a standard moving average operator. Letting $\boldsymbol{\Psi}^*(L) = \boldsymbol{\Phi}^*(L)^{-1}\boldsymbol{\Theta}^*(L)$, it follows that (18) can also be written as $\bar{\mathbf{Y}}_t = \mathbf{U}_t$, with $\mathbf{U}_t = \boldsymbol{\Psi}^*(L)\mathbf{A}_t^* = [\mathbf{U}_{t1}^{\mathrm{T}}, \mathbf{U}_{t2}^{\mathrm{T}}]^{\mathrm{T}}$, or, using (17), as

$$\begin{bmatrix} \nabla \mathbf{Y}_{t2} \\ \mathbf{Y}_{t1} \end{bmatrix} = \begin{bmatrix} \mathbf{U}_{t1} \\ -\mathbf{B}_2^{\mathrm{T}}\mathbf{Y}_{t2} + \mathbf{U}_{t2} \end{bmatrix}. \tag{19}$$

Noting that (19) is a triangular representation of the type introduced by Phillips (1991), it follows that asymptotic optimal inference applies to full-system EML estimation (as described in the previous remarks) of the stationary VARMA representation given in (15)-(17). From a practical point of view, this means that (i) testing for partial nonstationarity can be done using exactly the same procedures as those already well-established in previous literature (see, for example, Yap and Reinsel, 1995; Reinsel, 1997, Section 6.3; and Johansen, 2001), and (ii) that general hypothesis tests on parameters of the VEC model given in (5)-(6) can be conducted using standard (e.g., Wald or likelihood ratio) asymptotic chi-squared tests. Asymptotic optimality also applies to conditional estimation of the VEC model given in (5)-(6); however, as in the case of stationary VARMA models, it is not difficult to find practical situations (always involving finite samples, which are often far from large) where the more efficient use of available information associated with EML delivers more reliable inferences than those obtained through conditional methods. This possibility is illustrated with an example using actual data in Section 4.

Remark 6. The stationary VARMA model given in (15)-(17) for the zero-mean process $\{\bar{\mathbf{Y}}_t\}$, can be readily extended to incorporate the possibility that different components of the original nonstationary process $\{\mathbf{Y}_t\}$ have different local levels and, perhaps, deterministic drifts. This can be accomplished by extending (15) as follows:

$$\bar{\boldsymbol{\Phi}}(L)(\bar{\mathbf{Y}}_t - \mathrm{E}[\bar{\mathbf{Y}}_t]) = \boldsymbol{\Theta}(L)\mathbf{A}_t, \tag{20}$$

where $\bar{\boldsymbol{\Phi}}(L) = \bar{\boldsymbol{\Phi}}_0 - \sum_{i=1}^{p} \bar{\boldsymbol{\Phi}}_i L^i$ is given implicitly in (15) and (16), and

$$\mathrm{E}[\bar{\mathbf{Y}}_t] = \begin{bmatrix} \mathrm{E}[\nabla \mathbf{Y}_{t2}] \\ \mathrm{E}[\mathbf{W}_t] \end{bmatrix}. \tag{21}$$

By appropriately constraining the two unconditional mean vectors in (21), at least the following cases are possible within the stationary VARMA model (20):

Case 1: If $\mathrm{E}[\nabla \mathbf{Y}_{t2}] = \mathbf{0}$ and $\mathrm{E}[\mathbf{W}_t] = \mathbf{0}$, then the components of $\{\mathbf{Y}_t\}$ have no drifts and the equilibrium-error $\{\mathbf{W}_t\}$ is a zero-mean process.

Case 2: If $\mathrm{E}[\nabla \mathbf{Y}_{t2}] = \mathbf{0}$ and $\mathrm{E}[\mathbf{W}_t] \neq \mathbf{0}$, then the components of $\{\mathbf{Y}_t\}$ have no drifts and the equilibrium-error $\{\mathbf{W}_t\}$ is a nonzero-mean process.

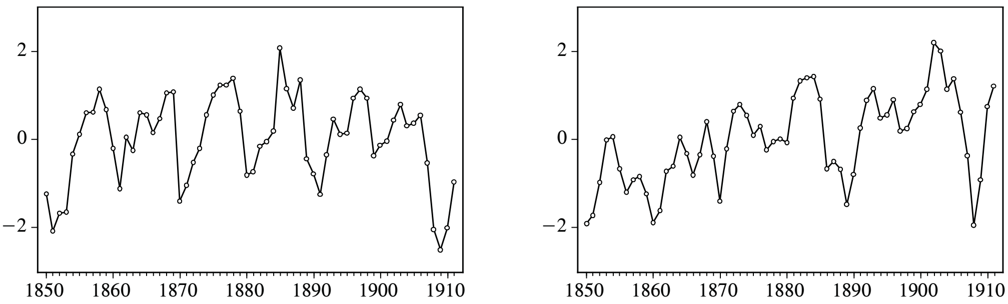
Case 3: If $\mathrm{E}[\nabla \mathbf{Y}_{t2}] \neq \mathbf{0}$ and $\mathrm{E}[\mathbf{W}_t] \neq \mathbf{0}$, then the components of $\{\mathbf{Y}_t\}$ have restricted drifts such that $\mathrm{E}[\nabla \mathbf{Y}_{t1}] = -\mathbf{B}_2^{\mathrm{T}} \mathrm{E}[\nabla \mathbf{Y}_{t2}]$, and the equilibrium-error $\{\mathbf{W}_t\}$ is a nonzero-mean process.

These three cases have been studied extensively in previous literature, due to that they represent adequately most situations occurring in applied research. In particular, it has been shown (see, for example, Yap and Reinsel, 1995; Reinsel, 1997, Section 6.3; and Johansen, 2001) that likelihood ratio test statistics for determining the cointegrating rank among components of $\{\mathbf{Y}_t\}$, follow different asymptotic distributions in each case. Thus, EML estimation of (20) under several specifications for $\mathrm{E}[\bar{\mathbf{Y}}_t]$ in (21), can be used in testing for partial nonstationarity in the same manner as when employing conditional methods, with the additional benefit of any of the three cases considered requiring only a trivial and easily interpretable implementation in (20)-(21).

4. AN EXAMPLE

This section summarizes some practical results on estimating VEC models using a popular data set available from several public sources. Both exact and conditional maximum likelihood estimations have been carried out using the programs mentioned in Remark 3 of Section 3, which also incorporate the possibilities mentioned in Remark 6. Asymptotic p-values for partial nonstationarity tests have been calculated using the programs provided by MacKinnon et al. (1999), noting, as shown, for example, by Yap and Reinsel (1995, Theorem 3), that moving average terms do not affect the asymptotic distributions of usual likelihood ratio test statistics for partial nonstationarity. For the purpose of model comparisons, information criteria (see, for example, Reinsel, 1997, Section 4.5.1) have been calculated as $\mathrm{AIC} = -(2L^* + 2K)/N$ (Akaike criterion) and $\mathrm{BIC} = -(2L^* + K\log N)/N$ (Schwarz Bayesian criterion), where L^* is the maximized log-likelihood, K is the number of freely estimated parameters, and N is the effective length of the multiple time series used for model estimation.

Examples on the advantages of EML over CML when estimating VARMA models for nonstationary but cointegrated time series are not difficult to come across (see Mauricio, 2006, for an example regarding the "Census Housing Data" considered by Mélard et al., 2006, Example 4; Reinsel,



(a) Logarithms of Mink Fur Series

(b) Logarithms of Muskrat Fur Series

Figure 1. Standardized Plots of the Mink-Muskrat Data for the Years 1850 through 1911.

1997, Examples 6.4 and 6.6; and Tiao, 2001). Additionally, partial nonstationary models of the type considered in Section 2 can be useful representations not only for cointegrated nonstationary time series processes. The example developed below illustrates the following important points: (i) In the case of multiple time series processes whose individual components may be both integrated as well as stationary processes, EML estimation of a suitable partially nonstationary model may help in clarifying the nature of such individual components, and (ii) although EML estimation in this case may not reveal anything really surprising, it shows what is reasonably there and, as opposed to CML, does not mislead.

4.1. The Mink-Muskrat Data

The data used in this example consist of annual sales of mink furs (x_{t1}) and muskrat furs (x_{t2}) by The Hudson's Bay Company for the years 1850 through 1911. These data (usually referred to as the "Mink-Muskrat Data") have been analyzed previously by many authors (see, for example, Mélard et al., 2006, Example 2; Reinsel, 1997, Examples 4.1 and 5.1; and the references cited therein), and constitute an interesting example on how CML estimation of partially nonstationary VARMA models can lead to serious conceptual difficulties, whereas EML estimation does not.

The logarithms of the original data shown in Figure 1 suggest that the series $y_{t1} = \log x_{t1}$ (the logarithms of the mink series) is plausibly stationary, and that the series $y_{t2} = \log x_{t2}$ (the logarithms of the muskrat series) is probably nonstationary because of a possible meandering behavior.

Formal unit root tests seem to confirm stationarity for the first series, but give conflicting results for the second series. In this regard, Table 1 summarizes the results of several unit root tests produced by the popular EViews software (Version 4.1, Standard Edition, July 29 2004 build).

4.2. Estimated Models for the Mink-Muskrat Data

In order to clarify the nature of the bivariate time series $\mathbf{y}_t = [y_{t1}, y_{t2}]^{\mathrm{T}}$ displayed in Figure 1, a vector ARMA(2,1) model

$$(\mathbf{I} - \mathbf{\Phi}_1 L - \mathbf{\Phi}_2 L^2)(\mathbf{Y}_t - \mathbf{c}) = (\mathbf{I} - \mathbf{\Theta}_1 L)\mathbf{A}_t \tag{22}$$

Table 1. Unit Root Tests for the Series y_{t1} and y_{t2} Displayed in Figure 1.

ADF Test [†] (H_0: Series has a unit root)	ADF-ERS Test [†] (H_0: Series has a unit root)	KPSS Test [†] (H_0: Series is stationary)
	Series y_{t1}: Logarithms of mink fur sales	
p-value $< 1\%$	p-value $< 1\%$	p-value $> 10\%$
	Series y_{t2}: Logarithms of muskrat fur sales	
p-value $< 1\%$	p-value $> 10\%$	$1\% < p$-value $< 5\%$

[†] ADF: Augmented Dickey-Fuller. ADF-ERS: Elliott-Rothenberg-Stock modification of ADF. KPSS: Kwiatkowski-Phillips-Schmidt-Shin. See Q.M.S. (2002, pp. 329-337) for the details. In all cases, the auxiliary regressions were run with an intercept and without a linear trend.

has been estimated (see Reinsel, 1997, Example 5.1), and the possibility that the 2×2 matrix $\mathbf{\Pi} = \mathbf{I} - \mathbf{\Phi}_1 - \mathbf{\Phi}_2$ in the corresponding VEC model

$$(\mathbf{I} - \mathbf{F}_1 L)\nabla \mathbf{Y}_t = -\mathbf{\Pi}(\mathbf{Y}_{t-1} - \mathbf{c}) + (\mathbf{I} - \mathbf{\Theta}_1 L)\mathbf{A}_t\,, \tag{23}$$

where $\mathbf{F}_1 = -\mathbf{\Phi}_2$, be of reduced rank (especially $P = 1$) has been considered.

Table 2 summarizes the estimation results for model (22) obtained through EML and CML. A few parameters have been set to zero because they were clearly insignificant in a previous estimation run. Both EML and CML results in Table 2 suggest that $\mathbf{\Pi}$ in (23) has a single zero eigenvalue (i.e., that the AR characteristic equation $|\,\mathbf{I} - \mathbf{\Phi}_1 x - \mathbf{\Phi}_2 x^2| = 0$ in (22) has a single unit root). To explore this possibility formally, tests based on likelihood ratio statistics have been considered. The tests for the various hypotheses are displayed in Table 3.

Both EML and CML results in Table 3 indicate that the hypothesis of $P = 0$ (or $D = 2$ unit roots) is strongly rejected in favor of $P = 1$, and that the hypothesis of $P = 1$ (or $D = 1$ unit root) cannot be rejected in favor of $P = 2$. Hence, both EML and CML lead clearly to the same conclusion that there is a single unit root in the autoregressive part of model (22). Note that if the two series displayed in Figure 1 were nonstationary (which does not seem to be the case; recall Table 1), then the presence of a single unit root would indicate that such series would be cointegrated. However, EML results displayed in Table 2 indicate that the estimated autoregressive part of model (22), $\hat{\mathbf{\Phi}}(L) = \mathbf{I} - \hat{\mathbf{\Phi}}_1 L - \hat{\mathbf{\Phi}}_2 L^2$, can be written as

$$\begin{aligned}\hat{\mathbf{\Phi}}(L) &= \begin{bmatrix} 1 - 0.8746L + 0.9263L^2 & 0.9191L - 0.9045L^2 \\ 1.0049L - 0.4191L^2 & 1 - 0.9502L \end{bmatrix} \\ &= \begin{bmatrix} 1 - 0.8746L + 0.9263L^2 & 0.9191L(1 - 0.9841L) \\ 1.0049L - 0.4191L^2 & 1 - 0.9502L \end{bmatrix},\end{aligned}$$

so that $\mathbf{\Phi}(L) = \mathbf{I} - \mathbf{\Phi}_1 L - \mathbf{\Phi}_2 L^2$ in (22) might have the special structure

$$\mathbf{\Phi}(L) = \begin{bmatrix} 1 - \phi_1 L - \phi_2 L^2 & -\phi_3 L(1 - L) \\ -\phi_4 L - \phi_5 L^2 & 1 - L \end{bmatrix}, \tag{24}$$

Table 2. Estimation Results for Model (22).

	Exact Maximum Likelihood [†]	Conditional Maximum Likelihood [†]
$\hat{\mathbf{c}}$	$\begin{bmatrix} 10.7976 \\ (0.0522) \\ 13.0080 \\ (0.2293) \end{bmatrix}$	$\begin{bmatrix} 10.5332 \\ (0.1350) \\ 12.4491 \\ (0.1897) \end{bmatrix}$
$\hat{\mathbf{\Phi}}_1$	$\begin{bmatrix} 0.8746 & -0.9191 \\ (0.1042) & (0.3798) \\ -1.0049 & 0.9502 \\ (0.1537) & (0.0851) \end{bmatrix}$	$\begin{bmatrix} 0.6912 & -0.1790 \\ (0.0890) & (0.1707) \\ -1.6126 & 1.1845 \\ (0.2812) & (0.0861) \end{bmatrix}$
$\hat{\mathbf{\Phi}}_2$	$\begin{bmatrix} -0.9263 & 0.9045 \\ (0.2626) & (0.3480) \\ 0.4191 & 0.0000 \\ (0.1335) & \text{——} \end{bmatrix}$	$\begin{bmatrix} -0.3194 & 0.3761 \\ (0.1210) & (0.1714) \\ 0.9971 & 0.0000 \\ (0.2360) & \text{——} \end{bmatrix}$
$\hat{\mathbf{\Theta}}_1$	$\begin{bmatrix} 0.0000 & -1.4828 \\ \text{——} & (0.4002) \\ -0.5742 & -0.1602 \\ (0.1640) & (0.1009) \end{bmatrix}$	$\begin{bmatrix} 0.0000 & -0.7671 \\ \text{——} & (0.2059) \\ -1.1642 & 0.0000 \\ (0.2847) & \text{——} \end{bmatrix}$
$\hat{\mathbf{\Sigma}}$	$\begin{bmatrix} 0.0371 & \\ 0.0168 & 0.0558 \end{bmatrix}$	$\begin{bmatrix} 0.0523 & \\ 0.0231 & 0.0544 \end{bmatrix}$
Eigenvalues of $\hat{\mathbf{\Pi}} = \mathbf{I} - \hat{\mathbf{\Phi}}_1 - \hat{\mathbf{\Phi}}_2$	0.0413, 1.0602	0.0126, 0.4311
Log-likelihood	15.6116	12.0470
AIC, BIC	−0.0201, 0.4990	0.0640, 0.5485

[†] Standard errors in parentheses.

implying that (22) might be reformulated as a special vector ARMA(2,1) model in terms of $\{Y_{t1}\}$ and $\{(1-L)Y_{t2}\}$, with only $\{Y_{t2}\}$ (the logarithm of the muskrat fur sales) being a nonstationary process. Note that this appears to be in close agreement with the information displayed in Figure 1 and in Table 1, and that this possibility cannot be seen (not even approximately) from CML estimation results given in Table 2. Furthermore, (24) implies a special reduced rank structure for $\mathbf{\Phi}(L)$ in (22) that can be described by imposing on $\mathbf{\Pi} = \mathbf{\Phi}(1)$ in (23) two restrictions, namely (i) that

$$\mathbf{\Pi} = \mathbf{\Lambda}\mathbf{B}^{\mathrm{T}} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} [1, \beta_2],$$

so that $P = \operatorname{rank}(\mathbf{\Pi}) = 1$, and (ii) that $\mathbf{B} = [1, 0]^{\mathrm{T}}$ (i.e., that $\beta_2 = 0$), so that the single unit root in $|\,\mathbf{\Phi}(x)\,| = 0$ is associated with $\{Y_{t2}\}$ only.

To explore this possibility further, the VEC model (23) has been constrained under the restriction that $P = \operatorname{rank}(\mathbf{\Pi}) = 1$ (i.e., that $\mathbf{\Pi} = \mathbf{\Lambda}\mathbf{B}^{\mathrm{T}}$), so that (23) has been conveniently rewritten (see Section 2 and Remark 6 in Section 3) as

Table 3. Tests on the Rank P of $\mathbf{\Pi}$ in Model (23) [i.e., on the Number D of Unit Roots of $|\mathbf{I}-\mathbf{\Phi}_1 x-\mathbf{\Phi}_2 x^2|=0$ in Model (22)] Based on Likelihood Ratio Test Statistics.

Hypotheses	Likelihood Ratio Test Statistic [†]	Asymptotic p-value
Exact Maximum Likelihood:		
H_0: $P=0$ $(D=2)$ H_1: $P=1$ $(D=1)$	$2\times[L_E^*(1)-L_E^*(0)]=35.5308$	Less than 0.01 %
H_0: $P=1$ $(D=1)$ H_1: $P=2$ $(D=0)$	$2\times[L_E^*(2)-L_E^*(1)]=0.9718$	95.51 %
Conditional Maximum Likelihood:		
H_0: $P=0$ $(D=2)$ H_1: $P=1$ $(D=1)$	$2\times[L_C^*(1)-L_C^*(0)]=36.5676$	Less than 0.01 %
H_0: $P=1$ $(D=1)$ H_1: $P=2$ $(D=0)$	$2\times[L_C^*(2)-L_C^*(1)]=3.7228$	45.50 %

[†] $L_E^*(P)$ represents the exact log-likelihood computed at the EML estimates of model (23) for the three different possible values of $P=\text{rank}(\mathbf{\Pi})$ $(P = 0, 1, 2)$. $L_C^*(P)$ represents the conditional log-likelihood computed at the CML estimates of model (23) for the three different possible values of P.

$$(\mathbf{I}-\mathbf{F}_1 L)\nabla\mathbf{Y}_t = -\mathbf{\Lambda}(\mathbf{B}'\mathbf{Y}_{t-1} - \text{E}[W_t]) + (\mathbf{I}-\mathbf{\Theta}_1 L)\mathbf{A}_t, \tag{25}$$

where $\mathbf{\Lambda}=[\lambda_1,\lambda_2]^{\text{T}}$, $\mathbf{B}=[1,\beta_2]^{\text{T}}$, and $W_t=\mathbf{B}^{\text{T}}\mathbf{Y}_t$. Then, (25) has been estimated twice, firstly without imposing the restriction that $\mathbf{B}=[1,0]^{\text{T}}$, and then with such restriction imposed. Tables 4 and 5 summarize the corresponding estimation results obtained through both EML and CML (where the same insignificant parameters than those in Table 2 have been set to zero); see also Figure 2 for a brief diagnostic of the estimated model given in Table 5. Comparisons among Tables 2, 4 and 5 highlight the points considered below.

4.3. Comparisons among Estimated Models

1. Simple visual inspection indicates that EML estimation of model (22) (Table 2) gives very similar results to those of EML estimation of model (25) under the restriction that $\mathbf{B}=[1,0]^{\text{T}}$ (Table 5). However, CML estimation of model (22) gives very similar results to those of CML estimation of model (25) without the restriction that $\mathbf{B}=[1,0]^{\text{T}}$ (Table 4). Hence, only EML gives some (informal) evidence in favor of the special reduced rank structure given in (24).
2. Likelihood ratio test statistics for the hypothesis of $\mathbf{B}=[1,0]^{\text{T}}$ in (25) are given by 5.4512 (EML) and 14.5320 (CML) (see Tables 4 and 5), with corresponding asymptotic p-values from a chi-squared distribution with one degree of freedom being equal to 1.96% (EML) and 0.01% (CML). Again, only EML gives some (formal) evidence in favor of the reduced rank structure given in (24) (which cannot be rejected, for example, at the 1% significance level), whereas CML strongly rejects such possibility.

Table 4. Estimation Results for Model (25).

	Exact Maximum Likelihood [†]	Conditional Maximum Likelihood [†]
$\hat{\mathrm{E}}[W_t]$	8.1345 (0.9484)	5.2124 (1.6849)
$\hat{\boldsymbol{\Lambda}}$	$\begin{bmatrix} 0.8392 \\ (0.2246) \\ 0.5881 \\ (0.1191) \end{bmatrix}$	$\begin{bmatrix} 0.5635 \\ (0.1647) \\ 0.5634 \\ (0.1503) \end{bmatrix}$
$\hat{\mathbf{B}}$	$\begin{bmatrix} 1\ [*] \\ -0.2042 \\ (0.0715) \end{bmatrix}$	$\begin{bmatrix} 1\ [*] \\ -0.4226 \\ (0.1262) \end{bmatrix}$
$\hat{\mathbf{F}}_1$	$\begin{bmatrix} 0.5848 & -0.6458 \\ (0.2273) & (0.3131) \\ -0.6621 & 0.0000 \\ (0.2151) & \text{——} \end{bmatrix}$	$\begin{bmatrix} 0.2946 & -0.3745 \\ (0.1253) & (0.1744) \\ -0.9719 & 0.0000 \\ (0.2443) & \text{——} \end{bmatrix}$
$\hat{\boldsymbol{\Theta}}_1$	$\begin{bmatrix} 0.0000 & -1.1148 \\ \text{——} & (0.3243) \\ -0.8953 & -0.0174 \\ (0.2678) & (0.0745) \end{bmatrix}$	$\begin{bmatrix} 0.0000 & -0.7397 \\ \text{——} & (0.1982) \\ -1.1538 & 0.0000 \\ (0.2969) & \text{——} \end{bmatrix}$
$\hat{\boldsymbol{\Sigma}}$	$\begin{bmatrix} 0.0385 & \\ 0.0181 & 0.0549 \end{bmatrix}$	$\begin{bmatrix} 0.0466 & \\ 0.0186 & 0.0600 \end{bmatrix}$
Eigenvalues of $\hat{\boldsymbol{\Pi}} = \hat{\boldsymbol{\Lambda}}\hat{\mathbf{B}}^{\mathrm{T}}$	0 [*], 0.7191	0 [*], 0.3254
Log-likelihood	15.1257	10.1856
AIC, BIC	−0.0697, 0.3802	0.0595, 0.4747

[†] Standard errors in parentheses.
[*] Normalized or implied parameter value.

3. Additional evidence in favor of (24) can be obtained from the fact that the moving average matrix $\hat{\boldsymbol{\Theta}}_1$ estimated through EML in Table 4 has one positive eigenvalue equal to $+0.9904$ (which is not the case in Tables 2 and 5). This fact might indicate some sort of multivariate overdifferencing in model (25) with no further restrictions, implying that the possibility that the two series displayed in Figure 1 are nonstationary (but cointegrated) is quite implausible.
4. Both EML and CML estimation of model (25) under the restriction that $\mathbf{B} = [1, 0]^{\mathrm{T}}$ (Table 5) give similar results, except for a substantial increase in the estimated variance of the first error process and a considerable loss of fit (in terms of log-likelihood and information criteria values) implied by CML with respect to EML. Additionally, the moving average matrix $\hat{\boldsymbol{\Theta}}_1$ estimated through EML in Table 5 has one negative eigenvalue equal to -0.9971. Although this negative eigenvalue close to -1 has nothing to do with the unit root structure imposed on the autoregressive part of model (25), it might indicate some special feature of the dynamic relationship between the two series considered that might deserve further attention for structural analysis.

Table 5. Estimation Results for Model (25) under the Restriction that $\mathbf{B} = [1, 0]^{\mathrm{T}}$.

	Exact Maximum Likelihood [†]	Conditional Maximum Likelihood [†]
$\hat{\mathrm{E}}[W_t]$	10.8161 (0.0358)	10.7677 (0.0515)
$\hat{\boldsymbol{\Lambda}}$	$\begin{bmatrix} 0.9382 \\ (0.2463) \\ 0.5929 \\ (0.1113) \end{bmatrix}$	$\begin{bmatrix} 0.9137 \\ (0.2746) \\ 0.6031 \\ (0.1138) \end{bmatrix}$
$\hat{\mathbf{F}}_1$	$\begin{bmatrix} 0.8357 & -0.7803 \\ (0.2384) & (0.3409) \\ -0.4501 & 0.0000 \\ (0.1359) & \text{——} \end{bmatrix}$	$\begin{bmatrix} 0.6390 & -0.7042 \\ (0.2659) & (0.4086) \\ -0.5491 & 0.0000 \\ (0.1862) & \text{——} \end{bmatrix}$
$\hat{\boldsymbol{\Theta}}_1$	$\begin{bmatrix} 0.0000 & -1.3429 \\ \text{——} & (0.3643) \\ -0.6039 & -0.1837 \\ (0.1601) & (0.0900) \end{bmatrix}$	$\begin{bmatrix} 0.0000 & -1.1620 \\ \text{——} & (0.4450) \\ -0.5987 & 0.0000 \\ (0.2372) & \text{——} \end{bmatrix}$
$\hat{\boldsymbol{\Sigma}}$	$\begin{bmatrix} 0.0382 & \\ 0.0138 & 0.0589 \end{bmatrix}$	$\begin{bmatrix} 0.0546 & \\ 0.0141 & 0.0608 \end{bmatrix}$
Eigenvalues of $\hat{\boldsymbol{\Pi}} = \hat{\boldsymbol{\Lambda}}\hat{\mathbf{B}}^{\mathrm{T}}$	0 [*], 0.9382	0 [*], 0.9137
Log-likelihood	12.4001	2.9196
AIC, BIC	−0.0131, 0.4021	0.2649, 0.6456

[†] Standard errors in parentheses.
[*] Implied parameter value.

Hence, EML estimation suggests for this example that the model considered in Table 5 is an adequate model, implying that only the series in Figure 1(b) is nonstationary. In contrast, CML estimation clearly selects the model considered in Table 4, implying (i) that there is one common nonstationary component shared by the two series considered (something quite hard to justify on the basis of Figure 1 and Table 1), and (ii) that there exists a conceptually implausible cointegrating relation between a stationary series and a nonstationary series.

As a further remark, it should be noted that the model considered in Table 5 estimated through EML compares favorably, in terms of information criteria and other diagnostic checking tools (as those given, for example, in Figure 2), with alternative models considered for the Mink-Muskrat Data in previous literature (see, for example, Mélard et al., 2006, Example 2; Reinsel, 1997, Examples 4.1 and 5.1; and the references cited therein), including bivariate pure autoregressive models, mixed ARMA models with diagonal autoregressive structure, and mixed ARMA models with echelon form structure.

In summary, the present example has shown that EML estimation of partially nonstationary models can reveal important features of the data considered that cannot be seen when CML is used instead. Note finally that EML estimation in this example does not reveal anything really surprising. However, as opposed to CML, EML simply shows what is reasonably there and does not mislead.

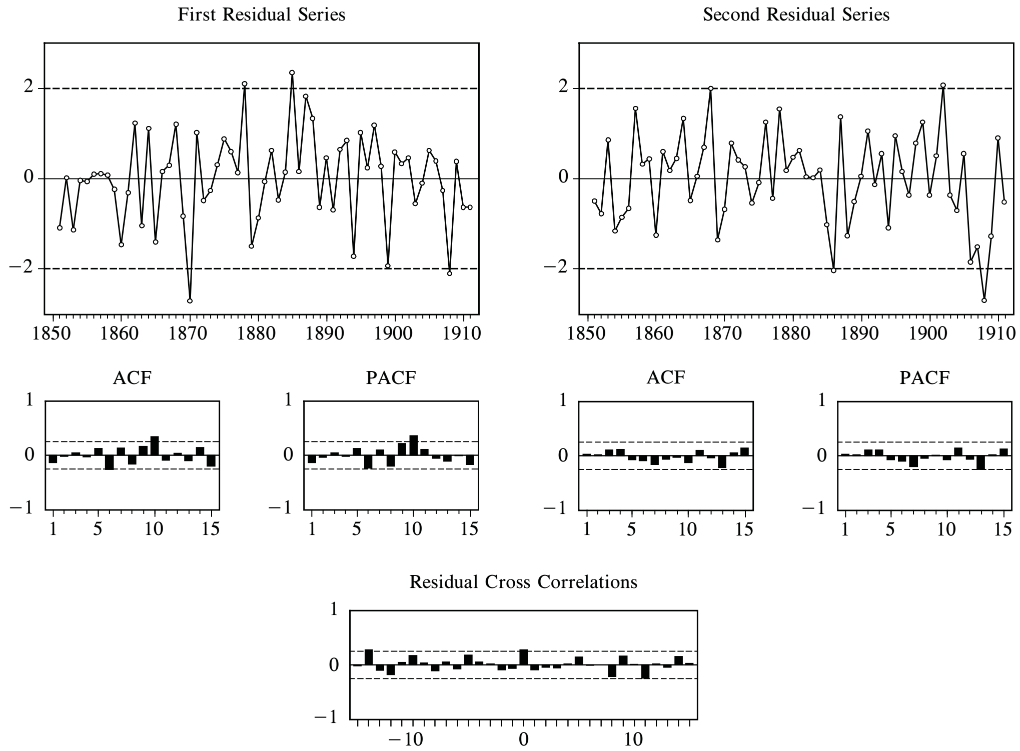


Figure 2. Unconditional Residuals from Exact Maximum Likelihood Estimation of Model (25) under the Restriction that $\mathbf{B} = [1, 0]^{\mathrm{T}}$ (See Table 5). (Note: Residual plots are standardized.) When residual simple (ACF) and partial (PACF) autocorrelations, as well as residual cross correlations, are compared to the limits of $\pm 2/\sqrt{N} = \pm 0.256$ (with $N = 61$ effective observations), there is no indication of misspecification in the estimated model (except perhaps for a single autocorrelation at lag 10 in the first residual series, which is probably due to the presence of a few outliers).

5. CONCLUSIONS

Both the theoretical developments and the example on unconditional, exact maximum likelihood estimation of partially nonstationary vector ARMA models presented in this article have shown the following important points:

1. Joint and reliable estimation of all the parameters in the error-correction form of partially nonstationary models is possible through the exact maximum likelihood approach developed in Section 3.
2. This approach resolves several problems related to the current limited availability of exact maximum likelihood estimation methods for partially nonstationary vector ARMA models, including (i) the impossibility of computing the exact log-likelihood for models with unit-root autoregressive structures through standard methods, and (ii) the inadequacy of considering lagged endogenous variables as exogenous variables in the error-correction form of such models. Additionally, this approach is simpler than the one outlined by Mélard et al. (2006), it is uniquely and unambiguously identified, it can provide EML estimates of common trends as a byproduct in case they are required

(Mauricio, 2006), and it can be used within the context of every available operational framework for stationary VARMA models.

3. Although exact maximum likelihood methods for partially nonstationary vector ARMA have received very limited attention in previous literature, they are usually preferable to conditional methods as in the case of stationary models. Practical benefits from using exact maximum likelihood instead of other popular conditional methods include (i) more accurately estimated moving average and error covariance structures (especially for possibly noninvertible models), and (ii) more reliable inferences on partially nonstationary structures when the possibly nonstationary nature of the data considered is not very clear.

It is true that simplicity, ease of implementation and low computational cost associated with many conditional estimation methods, have played a fundamental role in popularizing the use of such methods in empirical research on partially nonstationary systems. However, the extra computational complexities implied by the use of unconditional estimation methods, as well as the additional care and effort required on the applied researcher's part, may give clear benefits in cases of practical interest which are easily come across.

Acknowledgements

I am grateful to Arthur B. Treadway for invaluable guidelines and discussion on the general subject of this work, and to José M. Casals, Miguel Jerez, Francisco J. Samaniego, Sonia Sotoca, Andrew T.A. Wood, the editors of CSDA and the referees for useful comments that significantly improved its overall presentation.

References

Ahn, S.K., Reinsel, G.C., 1990. Estimation for Partially Nonstationary Multivariate Autoregressive Models. Journal of the American Statistical Association 85, 813-823.

Bauer, D., Wagner, M., 2002. Estimating Cointegrated Systems Using Subspace Algorithms. Journal of Econometrics 111, 47-84.

Box, G.E.P., Tiao, G.C., 1977. A Canonical Analysis of Multiple Time Series. Biometrika 64, 355-365.

Engle, R.F., Granger, C.W.J., 1987. Co-Integration and Error Correction: Representation, Estimation, and Testing. Econometrica 55, 251-276.

Johansen, S., 1988. Statistical Analysis of Cointegration 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., 2001. Cointegration in the VAR Model. In: Peña, D., Tiao, G.C., Tsay, R.S. (Eds.), A Course in Time Series Analysis. Wiley, New York, pp. 408-435.

Johansen, S., 2003. The Asymptotic Variance of the Estimated Roots in a Cointegrated Vector Autoregressive Model. Journal of Time Series Analysis 24, 663-678.

Kurozumi, E., 2005. The Rank of a Submatrix of Cointegration. Econometric Theory 21, 299-325.

Luceño, A., 1994. A Fast Algorithm for the Exact Likelihood of Stationary and Partially Nonstationary Vector Autoregressive Moving Average Processes. Biometrika 81, 551-565.

Lütkepohl, H., Claessen, H., 1997. Analysis of Cointegrated VARMA Processes. Journal of Econometrics 80, 223-239.

Luukkonen, R., Ripatti, A., Saikkonen, P., 1999. Testing for a Valid Normalization of Cointegrating Vectors in Vector Autoregressive Processes. Journal of Business and Economic Statistics 17, 195-204.

Ma, C., 1997. On the Exact Likelihood Function of a Multivariate Autoregressive Moving Average Model. Biometrika 84, 957-964.

MacKinnon, J.G., Haug, A.A., Michelis, L., 1999. Numerical Distribution Functions of Likelihood Ratio Tests for Cointegration. Journal of Applied Econometrics 14, 563-577.

Mauricio, J.A., 1995. Exact Maximum Likelihood Estimation of Stationary Vector Autoregressive Moving Average Models. Journal of the American Statistical Association 90, 282-291.

Mauricio, J.A., 1996. Some Computational Aspects of Exact Maximum Likelihood Estimation of Time Series Models. In: Prat, A. (Ed.), COMPSTAT 1996 Proceedings on Computational Statistics. Physica-Verlag, Heidelberg, pp. 361-366.

Mauricio, J.A., 1997. Algorithm AS 311: The Exact Likelihood Function of a Vector Autoregressive Moving Average Model. Applied Statistics 46, 157-171.

Mauricio, J.A., 2002. An Algorithm for the Exact Likelihood of a Stationary Vector Autoregressive Moving Average Model. Journal of Time Series Analysis 23, 473-486.

Mauricio, J.A., 2006. Additional Material on Exact Maximum Likelihood Estimation of Partially Nonstationary Vector ARMA Models. (Available upon Request.)

Mélard, G., Roy, R., Saidi, A., 2006. Exact Maximum Likelihood Estimation of Structured or Unit Root Multivariate Time Series Models. Computational Statistics and Data Analysis 50, 2958-2986.

Pham, D.T., Roy, R., Cédras, L., 2003. Tests for Non-correlation of Two Cointegrated ARMA Time Series. Journal of Time Series Analysis 24, 553-577.

Phillips, P.C.B., 1991. Optimal Inference in Cointegrated Systems. Econometrica 59, 283-306.

Q.M.S., 2002. EViews 4 User's Guide (revised for EViews 4.1, march 11 2002). Quantitative Micro Software, Irvine.

Reinsel, G.C., 1997. Elements of Multivariate Time Series Analysis (second edition). Springer-Verlag, New York.

Reinsel, G.C., Ahn, S.K., 1992. Vector AR Models With Unit Roots and Reduced Rank Structure: Estimation, Likelihood Ratio Test, and Forecasting. Journal of Time Series Analysis 13, 353-375.

Saikkonen, P., Lütkepohl, H., 1996. Infinite Order Cointegrated Vector Autoregressive Processes: Estimation and Inference. Econometric Theory 12, 814-844.

Saikkonen, P., Luukkonen, R., 1997. Testing Cointegration in Infinite Order Vector Autoregressive Processes. Journal of Econometrics 81, 93-126.

Shea, B.L., 1984. Maximum Likelihood Estimation of Multivariate ARMA Processes via the Kalman Filter. In: Anderson, O.D. (Ed.), Time Series Analysis: Theory and Practice, Vol. 5. North-Holland, Amsterdam, pp. 91-101.

Shea, B.L., 1989. Algorithm AS 242: The Exact Likelihood of a Vector Autoregressive Moving Average Model. Applied Statistics 38, 161-204.

Tanaka, K., 1996. Time Series Analysis: Nonstationary and Noninvertible Distribution Theory. Wiley, New York.

Tiao, G.C., 2001. Vector ARMA Models. In: Peña, D., Tiao, G.C., Tsay, R.S. (Eds.), A Course in Time Series Analysis. Wiley, New York, pp. 365-407.

Yap, S.F., Reinsel, G.C., 1995. Estimation and Testing for Unit Roots in a Partially Nonstationary Vector Autoregressive Moving Average Model. Journal of the American Statistical Association 90, 253-267.

Watson, M.W., 1994. Vector Autoregressions and Cointegration. In: Engle, R.F., McFadden, D.L. (Eds.), Handbook of Econometrics, Vol. IV. Elsevier, Amsterdam, pp. 2844-2915.

White, H., 2001. Asymptotic Theory for Econometricians (revised edition). Academic Press, San Diego.