IJST (2012) A2: 137-145
Iranian Journal of Science & Technology

http://www.shirazu.ac.ir/en

A weighted pairwise likelihood approach to multivariate AR(1) models

M. R. Kazemi and A. R. Nematollahi*

Department of Statistics, College of Sciences, Shiraz University, Shiraz 71454, Iran
E-mails: kazemimr88@hotmail.com & nematollahi@susc.ac.ir

Abstract

In this paper, the use of weighted pairwise likelihood instead of the full likelihood in estimating the parameters of the multivariate AR(1) is investigated. A closed formula for typical elements of the Godambe information (sandwich information) is presented. Some efficiency calculations are also given to discuss the feasibility and computational advantages of the weighted pairwise likelihood approach relative to the full likelihood approach.

Keywords*:* Weighted pairwise likelihood; multivariate time series; Godambe information

1. Introduction

The multivariate autoregression model is an essential and informative tool for describing the dynamic behavior of the applied sciences, e.g. neurosciences, biomedical research, business, economic, environmental, industrial and financial time series, see Schelter et al. and therein references for more examples and applications [1]. In multivariate models, the standard likelihood-based inference often involves high-dimensional integrals which become more difficult to evaluate when the sample size increases, Xu [2] and Parner [3]. In the case that the likelihood function is difficult to evaluate or cannot be solved, certain methods based on modifications of the likelihood are used by several authors, e.g. Besag [4, 5], Cox [6] and Lindsay [7]. The conditional Gaussian likelihood procedure proposed by Hannan and Rissanen [8] has been complemented by Salau [9] to investigate the effects of different choices of order for autoregressive approximation on the fully efficient parameter estimates for ARMA models. Asymptotic quasi-likelihood has been advanced by Heyde and Lin [10], Heyde [11] and Jung et al. [12]. A quasi-maximum likelihood unit root test is studied by Rothenberg and Stock [13] and Carstensen [14], and the quasi-maximum likelihood theory to simulate the likelihood ratio test is given by Lee [15] and Ziegler [16]. Parameters estimation by quasimaximum likelihood through the optimization of a Gaussian log-likelihood function is discussed in Bollerslev and Wooldridge [17], Newey and Steigerwald [18] and Rodrigues and Rubia [19]. Modified versions of the likelihood function to find an approximate sufficient statistic for the parameters of the vector-valued (multivariate) ARMA models, in terms of the periodogram are applied by Kharrati-Kopaei et al. [20]. The pseudo-likelihood based estimation procedure is also used by many authors. The reader is referred to Hu and Zhang [21] and the references therein for a short discussion on this subject.

In recent years, the composite likelihood approach (based on the likelihoods of low-dimensional marginal distributions) has received much attention in statistical models. The pairwise likelihood, given by Cox and Reid [22], takes the bivariate margins to produce the pseudo-likelihood. An excellent review on the composite likelihoods with emphasis on some applications in genetic, longitudinal data, survival analysis and spatial statistics can be found in Varin [23]. Recently, Ng et al. [24] have established the consistency and asymptotic normality properties in a time series model with a latent Gaussian autoregressive process for various composite likelihood estimators. The asymptotic properties of pairwise likelihood estimation procedures for linear time series models are also studied by Davis and Yau [25].

Pairwise likelihood approach in estimating the parameters of a multivariate AR(1) is studied by Nematollahi and Kazemi [26]. They concentrate on the maximum pairwise likelihood estimators (MPLE) and specify its efficiency for the arbitrary matrix of parameters of multivariate AR(1) model, along the line of the equi-correlation normal matrix example in Cox and Reid [22]. They have used unweighted pairwise likelihood and by some

*Corresponding author
Received: 26 June 2011 / Accepted: 9 October 2011

tedious computations, it is shown that the efficiency can be low, so that using weighted pairwise likelihood-based inferences is recommended. In this paper, we discuss the feasibility of the weighted pairwise likelihood approach (Joe and Lee [27]) and discuss the computational advantages of weighted pairwise likelihood relative to the full likelihood approach. Practical advantages of the weighted MPLE (WMPLE) to MLE are brought to light and some efficiency calculations are presented.

This paper is organized as follows: The pairwise likelihood function for m-dimensional random vector is introduced Section 2. In Section 3, we consider the multivariate AR(1) models and compute asymptotic variance of the parameters based on the pairwise likelihood. Numerical studies in Section 4 are in support of our derivations.

2. Preliminaries

Let $\mathrm{Y} = (\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_m)^{\boldsymbol{T}}$ be a m-dimensional random vector with joint density $f(\mathbf{y},\boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbf{R}^m$ is unknown parameter. There are a number of situations where it is difficult to specify the full m-dimensional distribution in simple form but it may be possible to specify some low-dimensional distributions. Here we consider the two-dimensional one, i.e., we concentrate on the weighted bivariate densities $f(y_i, y_j, \boldsymbol{\theta})^{\pi_{ij}}$ for all $i, j = 1,\ldots,m, i \neq j,$ where $\pi_{ij}, i, j = 1,\ldots,m,$ are the suitable nonnegative weights.

The weighted pairwise likelihood (WPL) from a single m-dimensional random vector Y is constructed by

$$L_{wpl}(\mathbf{y},\boldsymbol{\theta}) = \prod_{i=1}^{m-1}\prod_{j=i+1}^{m} f(y_i, y_j, \boldsymbol{\theta})^{\pi_{ij}},$$

see e.g. Cox and Ried [22]. If the true parameter value $\boldsymbol{\theta}_0$ belongs to the interior of the compact parameter space, then the weighted maximum pairwise likelihood estimate (WMPLE) of $\boldsymbol{\theta}$, denoted by $\tilde{\theta}$ is the solution of the composite score function,

$$\mathbf{S}_{wpl}(\mathbf{y},\boldsymbol{\theta}) = \sum_{i=1}^{m-1}\sum_{j=i+1}^{m} \frac{\partial \log\left(f\left(y_i, y_j, \boldsymbol{\theta}\right)^{\pi_{ij}}\right)}{\partial \boldsymbol{\theta}} = \mathbf{0}.$$

Since the composite score is a linear combination of valid likelihood score functions, then its unbiasedness follows under usual regularity conditions.

We will present the weighted pairwise likelihood (WPL) based on n (m-dimensional) random vectors $\mathbf{Y}_1,\ldots,\mathbf{Y}_n$ in Section 3. The information in the composite score equation provided by n observations is given by

$$\mathbf{G}(\boldsymbol{\theta}) = \mathbf{J}_{wpl}(\boldsymbol{\theta})\mathbf{K}_{wpl}^{-1}(\boldsymbol{\theta})\mathbf{J}_{wpl}(\boldsymbol{\theta}),$$

where $\mathbf{J}_{wpl}(\boldsymbol{\theta}) = E\left(-\frac{\partial \mathbf{S}_{wpl}(\boldsymbol{\theta},\mathbf{y})}{\partial \boldsymbol{\theta}}\right)$, and $\mathbf{K}_{wpl}(\boldsymbol{\theta})$ is the variance of the composite score matrix given by $\mathbf{K}_{wpl}(\boldsymbol{\theta}) = E[\mathbf{S}_{wpl}(\boldsymbol{\theta},\mathbf{y})\mathbf{S}_{wpl}(\boldsymbol{\theta},\mathbf{y})^T]$. The standard theory for inference functions [28] can be applied to derive the general asymptotic properties of $\tilde{\theta}$ Under the theory of estimating equations and the regularity conditions on the log-likelihood, it can be shown that $\tilde{\theta}$ is consistent and asymptotically normal distributed with asymptotic mean $\boldsymbol{\theta}$, and variance matrix

$$\mathbf{G}^{-1}(\boldsymbol{\theta}) = \mathbf{J}_{wpl}^{-1}(\boldsymbol{\theta})\mathbf{K}_{wpl}(\boldsymbol{\theta})\mathbf{J}_{wpl}^{-1}(\boldsymbol{\theta}), \tag{1}$$

which follows from the Taylor expansion to second order. A proof can be found in Godambe [28]. See also Zhao [29], Zhao and Joe [30] and Joe and Lee [27] for more details.

In this paper, we consider a multivariate AR (1) model which satisfies the difference equation

$$\mathbf{X}_t = \boldsymbol{\Phi}\mathbf{X}_{t-1} + \mathbf{Z}_t, \tag{2}$$

where $\boldsymbol{\Phi}$ is $m \times m$ matrix and $\mathbf{Z}_t$ is a sequence of independent multivariate normal with zero mean vector and covariance matrix $\boldsymbol{\Sigma}$. We assume that the process is causal in the sense that all eigenvalues of $\boldsymbol{\Phi}$ are less than 1 in absolute value. From the well-known Yule-Walker equations, we have $\boldsymbol{\Gamma}(h) = \boldsymbol{\Phi}\boldsymbol{\Gamma}(h-1)$ for $h = 1,2,3,\ldots$ and $vec(\boldsymbol{\Gamma}(0)) = (\mathbf{I} - (\boldsymbol{\Phi}\otimes\boldsymbol{\Phi}))^{-1} vec(\boldsymbol{\Sigma})$ where the " vec" notation for a matrix $A = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n]$ is defined by $vec(A) = [\mathbf{a}_1^T, \mathbf{a}_2^T, \ldots, \mathbf{a}_n^T]^T$ and operator $\otimes$ is the Kronecker product.

Joe and Lee [27] studied the m-dimensional Gaussian AR(1) model with mean $\mu\mathbf{1}_m$ and covariance matrix $\eta^2\mathbf{R}(\rho)$, where $\mathbf{R}(\rho) = [\rho^{|j-k|}]_{j,k=1}^m$ with $-1 < \rho < 1$. They obtained some efficiency results for the estimate of one of the parameters assuming the other two to be known, using the weighted pairwise likelihoods with weights depending on lag. In this paper, we study weighted pairwise likelihoods in a more general case with the weights constructed based on the autoregressive property of the time series and considering the serial dependence within each component series and interdependence between the different component series. Using this kind of weights, we will show that the loss of efficiency compared to maximum likelihood estimation is acceptable and balanced by the computational ease.

3. The weighted pairwise likelihood inferences

Let $\{\mathbf{X}_1,...,\mathbf{X}_n\}$ be a set of m-dimensional observations, where $\mathbf{X}_t = \left(X_{t1}, X_{t2}, \ldots, X_{tm}\right)^T$, $t = 1,...,n$. Following Nematollahi and Kazemi [26], consider a $mn \times 1$ vector $\mathbf{X}$ constructed by

$$\mathbf{X} = \left(\mathbf{X}_1^T, \mathbf{X}_2^T, \ldots, \mathbf{X}_n^T\right)^T$$
$$= (X_{11}, X_{12}, \ldots, X_{1m,} X_{21}, X_{22}, \ldots, X_{2m,} \ldots, X_{n1}, X_{n2}, \ldots, X_{nm}^T). \quad (3)$$

The vector-valued (multivariate) time series $\{\mathbf{X}_t\}$ having not only serial dependence within each component series $\{X_{ti}\}$, but also interdependence between the different component series $\{X_{ti}\}$ and $\{X_{tj}\}$, $i \neq j$. According to the autoregressive property, each $\{\mathbf{X}_t\}$ is also dependent on $\{\mathbf{X}_{t-1}\}$. So, we will consider the weighted pairwise likelihood for $\mathbf{X}$ as

$$L_{wpl}(\mathbf{x},\boldsymbol{\theta}) = \prod_i \prod_j \prod_b \prod_c f_{\left(X_{ib}, X_{jc}\right)}\left(x_{ib}, x_{jc}, \boldsymbol{\theta}\right)^{\pi_{ijbc}} \quad (4)$$
$$= \left[\prod_{a=1}^{n} \prod_{1 \le b < c \le m} f_{\left(X_{ab}, X_{ac}\right)}\left(x_{ab}, x_{ac}, \boldsymbol{\theta}\right)\right]\left[\prod_{i=1}^{n-1} \prod_{b,c=1}^{m} f_{\left(X_{ib}, X_{(i+1)c}\right)}\left(x_{ib}, x_{(i+1)c}, \boldsymbol{\theta}\right)\right],$$

where

$$\pi_{ijbc} = \begin{cases} 1, & (i = j, 1 \le b < c \le m) \text{or} (j = i+1, 1 \le b, c \le m) \\ 0, & \text{otherwise.} \end{cases}$$

Remark 1. Note that in order to use all of the information in the time series, we have to write $\mathbf{X}$ and $L_{wpl}\left(\mathbf{x},\boldsymbol{\theta}\right)$ as above, since in the multivariate time series the vector of observations at different times are dependent and it is also dependent with other observations. In (4), the first term is the likelihood function of the observations within the $\{\mathbf{X}_n\}$ and the second term is the counterpart between groups $\{\mathbf{X}_n\}$ and $\{\mathbf{X}_{n+1}\}$, this is one justification for considering the weights π_{ijbc} as given above. Another justification is the autoregressive property.

Remark 2. The pairwise likelihood function considered by Nematollahi and Kazemi [26] can be easily derived by setting $\pi_{ijbc} = 1$, for all i, j, b and c.

Now, suppose that $\{\mathbf{X}_t, t = 1,2,\ldots,n\}$ is a set of n mean-zero m-dimensional Gaussian time series with covariance matrix $\mathbf{K}(i, j) = E\left(\mathbf{X}_i \mathbf{X}_j^T\right)$, then for a nm-dimensional Gaussian vector $\mathbf{X}$, we can say

$$\begin{bmatrix} X_{ib} \\ X_{jc} \end{bmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \gamma_{bb}(0) & \gamma_{bc}(i-j) \\ \gamma_{bc}(i-j) & \gamma_{cc}(0) \end{bmatrix}\right). \quad (5)$$

Let

$$\mathbf{M}_{bc}(h) = \begin{bmatrix} \gamma_{bb}(0) & \gamma_{bc}(h) \\ \gamma_{bc}(h) & \gamma_{cc}(0) \end{bmatrix},$$

then (5) can be written as

$$\begin{pmatrix} X_{ib} \\ X_{jc} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \mathrm{M}_{bc}(i - j)\right]. \quad (6)$$

Now, using (6) the logarithm of the weighted pairwise likelihood function $L_{wpl}(\mathbf{x},\boldsymbol{\theta})$ reduces to

$$l_{wpl}(\boldsymbol{\theta}) = \log(L_{wpl}(\mathbf{x},\boldsymbol{\theta}))$$
$$= -\frac{1}{2}\sum_{a=1}^{n}\sum_{1 \le b < c \le m}\left[\log\left|\mathbf{M}_{bc}(0)\right| + \left\{\mathbf{X}_{bc}^T(a)\mathbf{M}_{bc}^{-1}(0)\mathbf{X}_{bc}(a)\right\}\right]$$
$$-\frac{1}{2}\sum_{i=1}^{n-1}\sum_{b,c=1}^{m}\log\left|\mathbf{M}_{bc}(-1)\right| + \left\{\mathbf{X}_{bc}^T(i)\mathbf{M}_{bc}^{-1}(-1)\mathbf{X}_{bc}(i)\right\}, \quad (7)$$

where $\boldsymbol{\theta} = \left(\boldsymbol{\Phi}, \boldsymbol{\Sigma}\right)$ is the unknown parameter of model, $\mathbf{X}_{bc}(a) = \begin{pmatrix} X_{ab} \\ X_{ac} \end{pmatrix}$ and $\mathbf{X}_{bc}(i) = \begin{pmatrix} X_{ib} \\ X_{(i+1)c} \end{pmatrix}$. Now let $\boldsymbol{\Phi} = \left[\phi_{ij}\right]_{i,j=1}^{m}$ and $\boldsymbol{\Sigma} = \left[\sigma_{ij}\right]_{i,j=1}^{m}$ and take

$$w_1 = \phi_{11}, w_2 = \phi_{21}, ..., w_m = \phi_{m1}, ..., w_q = \sigma_{mm},$$

where $q = mm + (m(m+1)/2)$ is the number of unknown parameters. Then using (7), we have

$$\mathbf{S}_{wpl}(\mathbf{x},\boldsymbol{\theta}) = \left[\frac{\partial l_{wpl}(\boldsymbol{\theta})}{\partial w_i}\right]_{i=1}^{q},$$

$$\mathbf{J}_{wpl}(\boldsymbol{\theta}) = \left[E\left(-\frac{\partial^2 l_{wpl}(\boldsymbol{\theta})}{\partial w_k \partial w_l}\right)\right]_{k,l=1}^{q},$$

and finally

$$\mathbf{K}_{wpl}(\boldsymbol{\theta}) = \left[E\left(\frac{\partial l_{wpl}(\boldsymbol{\theta})}{\partial w_k}\frac{\partial l_{wpl}(\boldsymbol{\theta})}{\partial w_l}\right)\right]_{k,l=1}^{q}.$$

To compute the above quantities, first note that

$$\frac{\partial l_{wpl}(\boldsymbol{\theta})}{\partial w_k} = -\frac{1}{2}\sum_{a=1}^{n}\sum_{1\le b<c\le m}\left[\left\{\frac{\partial \log|\mathbf{M}_{bc}(0)|}{\partial w_k}\right\} + \left\{\mathbf{X}_{bc}^{T}(a)\left\{\frac{\partial \mathbf{M}_{bc}^{-1}(0)}{\partial w_k}\right\}\mathbf{X}_{bc}(a)\right\}\right]$$
$$-\frac{1}{2}\sum_{i=1}^{n-1}\sum_{b,c=1}^{m}\left[\left\{\frac{\partial \log|\mathbf{M}_{bc}(-1)|}{\partial w_k}\right\} + \left\{\mathbf{X}_{bc}^{T}(i)\left\{\frac{\partial \mathbf{M}_{bc}^{-1}(-1)}{\partial w_k}\right\}\mathbf{X}_{bc}(i)\right\}\right], \quad (8)$$

and then we have,

$$\frac{\partial^2 l_{wpl}(\boldsymbol{\theta})}{\partial w_k \partial w_l} = -\frac{1}{2}\sum_{a=1}^{n}\sum_{1\le b<c\le m}\left[\left\{\frac{\partial^2 \log|\mathbf{M}_{bc}(0)|}{\partial w_k \partial w_l}\right\} + \left\{\mathbf{X}_{bc}^{T}(a)\left\{\frac{\partial^2 \mathbf{M}_{bc}^{-1}(0)}{\partial w_k \partial w_l}\right\}\mathbf{X}_{bc}(a)\right\}\right]$$
$$-\frac{1}{2}\sum_{i=1}^{n-1}\sum_{b,c=1}^{m}\left[\left\{\frac{\partial^2 \log|\mathbf{M}_{bc}(-1)|}{\partial w_k \partial w_l}\right\} + \left\{\mathbf{X}_{bc}^{T}(i)\left\{\frac{\partial^2 \mathbf{M}_{bc}^{-1}(-1)}{\partial w_k \partial w_l}\right\}\mathbf{X}_{bc}(i)\right\}\right]. \quad (9)$$

Now we can state the following theorem which plays the main role in computing the asymptotic variance-covariance matrix of the WMPLE.

Theorem 1. The typical (k,l)th element of the matrices $\mathbf{J}_{wpl}(\boldsymbol{\theta})$ and $\mathbf{K}_{wpl}(\boldsymbol{\theta})$ have the form

$$\mathbf{J}_{\mathrm{wpl}}(\boldsymbol{\theta})_{kl} = E\left(-\frac{\partial^2 l_{wpl}(\boldsymbol{\theta})}{\partial w_k \partial w_l}\right)$$
$$= \frac{1}{2}\sum_{a=1}^{n}\sum_{1\le b<c\le m}\left[\left\{\frac{\partial^2 \log|\mathbf{M}_{bc}(0)|}{\partial w_k \partial w_l}\right\} + tr\left(\left\{\frac{\partial^2 \mathbf{M}_{bc}^{-1}(0)}{\partial w_k \partial w_l}\right\}\mathbf{M}_{bc}(0)\right)\right]$$
$$+\frac{1}{2}\sum_{i=1}^{n-1}\sum_{b,c=1}^{m}\left[\left\{\frac{\partial^2 \log|\mathbf{M}_{bc}(-1)|}{\partial w_k \partial w_l}\right\} + \mathrm{tr}\left(\left\{\frac{\partial^2 \mathbf{M}_{bc}^{-1}(-1)}{\partial w_k \partial w_l}\right\}\mathbf{M}_{bc}(-1)\right)\right], \quad (10)$$

and

$$\mathbf{K}_{\mathrm{wpl}}(\boldsymbol{\theta})_{kl} = E\left(\frac{\partial l_{wpl}(\boldsymbol{\theta})}{\partial w_k}\frac{\partial l_{wpl}(\boldsymbol{\theta})}{\partial w_l}\right)$$
$$= \frac{1}{2}\sum_{a=1}^{n}\sum_{1\le b<c\le m}\sum_{a'=1}^{n}\sum_{1\le b'<c'\le m} tr\left[\mathbf{N}_{bc,k}(0)\mathbf{M}^{(0)}_{bb'cc'}(a,a')\mathbf{N}_{b'c',l}(0)\mathbf{M}^{(0)T}_{bb'cc'}(a,a')\right]$$
$$+\frac{1}{2}\sum_{a=1}^{n}\sum_{1\le b<c\le m}\sum_{i'=1}^{n-1}\sum_{b',c'=1}^{m} tr\left[\mathbf{N}_{bc,k}(0)\mathbf{M}^{(1)}_{bb'cc'}(a,i')\mathbf{N}_{b'c',l}(-1)\mathbf{M}^{(1)T}_{bb'cc'}(a,i')\right]$$
$$+\frac{1}{2}\sum_{i=1}^{n-1}\sum_{b,c=1}^{m}\sum_{a'=1}^{n}\sum_{1\le b'<c'\le m} tr[\mathbf{N}_{b'c',l}(-1)\mathbf{M}^{(2)}_{bb'cc'}(i,a')\mathbf{N}_{bc,k}(0)\mathbf{M}^{(2)T}_{bb'cc'}(i,a')]$$
$$+\frac{1}{2}\sum_{i=1}^{n-1}\sum_{b,c=1}^{m}\sum_{i'=1}^{n-1}\sum_{b',c'=1}^{m} \mathrm{tr}[\mathbf{N}_{bc,k}(-1)\mathbf{M}^{(3)}_{bb'cc'}(i,i')\mathbf{N}_{b'c',l}(-1)\mathbf{M}^{(3)T}_{bb'cc'}(i,i')], \quad (11)$$

respectively, where

$$\mathbf{N}_{bc,k}(h) = \partial \mathbf{M}_{bc}^{-1}(h)/\partial w_k,$$

$$\mathbf{M}^{(0)}_{bb'cc'}(h,h') = \begin{bmatrix} \gamma_{bb'}(h-h') & \gamma_{bc'}(h-h') \\ \gamma_{cb'}(h-h') & \gamma_{cc'}(h-h') \end{bmatrix},$$

$$\mathbf{M}^{(1)}_{bb'cc'}(h,h') = \begin{bmatrix} \gamma_{bb'}(h-h') & \gamma_{bc'}(h-h'-1) \\ \gamma_{cb'}(h-h') & \gamma_{cc'}(h-h'-1) \end{bmatrix},$$

$$\mathbf{M}^{(2)}_{bb'cc'}(h,h') = \begin{bmatrix} \gamma_{bb'}(h-h') & \gamma_{bc'}(h-h') \\ \gamma_{cb'}(h+1-h') & \gamma_{cc'}(h+1-h') \end{bmatrix},$$

$$\mathbf{M}^{(3)}_{bb'cc'}(h,h') = \begin{bmatrix} \gamma_{bb'}(h-h') & \gamma_{bc'}(h-h'-1) \\ \gamma_{cb'}(h+1-h') & \gamma_{cc'}(h-h') \end{bmatrix}.$$

Proof: The proof of (10) is simply done by using the following fact. If $E(\mathbf{X}) = \mathbf{0}$, then $E(\mathbf{X}^T\mathbf{A}\mathbf{X}) = tr(\mathbf{A}\boldsymbol{\Omega})$, where $\boldsymbol{\Omega} = E(\mathbf{X}\mathbf{X}^T)$.

To prove (11), at first note that,

$$E[(\partial l_2(\boldsymbol{\theta})/\partial w_k)(\partial l_2(\boldsymbol{\theta})/\partial w_l)] = \pi_1 + \pi_2 + \pi_3 + \cdots + \pi_{16},$$

where the equations π_1 up to π_{16} can be divided to 4 groups as below. Note that for the four following groups (q,r) varies in $\{(0,0),(0,-1),(-1,0),(-1,-1)\}$ according to s, varying from 1 up to 4 for the first group, from 5 up to 8 for the second group, from 9 up to 12 for the third group and from 13 up to 16 for the forth group, respectively.

Group1: for π_s , $s = 1,2,3,4$ we have,

$$\pi_s = \frac{1}{4}\sum_{i_q}\sum_{b,c}\sum_{i'_r}\sum_{b',c'}\left[\frac{\partial \log|\mathbf{M}_{bc}(q)|}{\partial w_k}\right]\left[\frac{\partial \log|\mathbf{M}_{b'c'}(r)|}{\partial w_l}\right].$$

Group2: for π_s , $s = 5,6,7,8$ we have,

$$\pi_s = \frac{1}{4}\sum_{i_q}\sum_{b,c}\sum_{i'_r}\sum_{b',c'}\left[\frac{\partial \log|\mathbf{M}_{bc}(q)|}{\partial w_k}\right] tr\left[\left(\frac{\partial \mathbf{M}^{-1}_{b'c'}(r)}{\partial w_l}\right)\mathbf{M}_{b'c'}(r)\right].$$

Group3: for π_s , $s = 9,10,11,12$ we have,

$$\pi s = \frac{1}{4}\sum_{i_q}\sum_{b,c}\sum_{i'_r}\sum_{b',c'} tr\left[\left(\frac{\partial \mathbf{M}_{bc}^{-1}(q)}{\partial w_k}\right)\mathbf{M}_{bc}(q)\right]\left[\frac{\partial \log|\mathbf{M}_{b'c'}(r)|}{\partial w_l}\right].$$

Group4: for π_s , $s = 13,14,15,16$ we have,

$$\pi_s = \frac{1}{4}\sum_{i_q}\sum_{b,c}\sum_{i_r'}\sum_{b',c'}[cov\left(\mathbf{X}_{bc}^T(i_q)(\tfrac{\partial \mathbf{M}_{bc}^{-1}(q)}{\partial w_k})\mathbf{X}_{bc}(i_q), \mathbf{X}_{b'c'}^T(i_r')(\tfrac{\partial \mathbf{M}_{b'c'}^{-1}(r)}{\partial w_l})\mathbf{X}_{b'c'}(i_r')\right)$$

$$+ tr\left[(\frac{\partial \mathbf{M}_{bc}^{-1}(q)}{\partial w_k})\mathbf{M}_{bc}(q)\right] tr\left[(\frac{\partial \mathbf{M}_{b'c'}^{-1}(r)}{\partial w_l})\mathbf{M}_{b'c'}(r)\right]].$$

Note that in the all of the above groups, if i_q varies from 1 up to n-1 then both b,c vary from 1 up to m and if i_q varies from 1 to n then b,c vary from $1 \le b < c \le m.$

For simplifying these expressions, we have to use the following useful facts.

(I) If $E(\mathbf{X}) = \mathbf{0}$ then $E(\mathbf{X}^T\mathbf{A}\mathbf{X}) = tr(\mathbf{A}\mathbf{\Omega})$ where $\mathbf{\Omega} = E(\mathbf{X}\mathbf{X}^T)$.

(II) $\frac{\partial \log|\mathbf{V}|}{\partial w} = tr\left(\mathbf{V}^{-1}(\frac{\partial \mathbf{V}}{\partial w})\right) = tr\left((\frac{\partial \mathbf{V}}{\partial w})\mathbf{V}^{-1}\right)$ and so $\frac{\partial \log|\mathbf{V}^{-1}|}{\partial w} = tr\left(\mathbf{V}(\frac{\partial \mathbf{V}^{-1}}{\partial w})\right) = tr\left((\frac{\partial \mathbf{V}^{-1}}{\partial w})\mathbf{V}\right)$.

Also note that

$$\frac{\partial \log|\mathbf{V}^{-1}|}{\partial w} = -\frac{\partial \log|\mathbf{V}|}{\partial w}.$$

(III) For simplifying $\pi_{13},\ldots\pi_{16}$ we have used the following well-known result (Searle, [31]). Let $\mathbf{X_i} \sim \mathbf{N}(\mathbf{0}, \mathbf{C_{ii}})$ and $\mathbf{C_{ij}} = \mathbf{E}(\mathbf{X_i}\mathbf{X_j^T})$ then,

$$\mathbf{Cov}(\mathbf{X_1^T A_{12} X_2}, \mathbf{X_3^T A_{34} X_4^T}) = \mathbf{tr}(\mathbf{A_{12}C_{23}A_{34}C_{41}} + \mathbf{A_{12}C_{24}A_{34}C_{31}}).$$

Using (I) and (II) it can be easily shown that,

$$\pi_1 + \pi_5 = \pi_2 + \pi_6 = \pi_3 + \pi_7 = \pi_4 + \pi_8 = 0.$$

And also by using (I) and (II) we can have,

$$\pi_s = -\frac{f(n)}{4}\sum_{b,c}\sum_{b',c'}\left[\frac{\partial \log|\mathbf{M}_{bc}(q)|}{\partial w_k}\right]\left[\frac{\partial \log|\mathbf{M}_{b'c'}(r)|}{\partial w_l}\right],$$

for $s = 9,10,11,12$. If $s = 9$ and $(q,r) = (0,0)$ then $\mathrm{f(n)} = \mathrm{n}^2,$ and if $s = 10,11$ and $(q,r) \in \{(0,-1),(-1,0)\}$ then $f(n) = n(n-1),$ and if $s = 12$ and $(q,r) = (-1,-1)$ then $f(n) = (n-1)^2$.

Recall $\mathbf{N}_{bc,k}(h)$, $\mathbf{M}_{bb'cc'}^{(0)}(h,h')$, $\mathbf{M}_{bb'cc'}^{(1)}(h,h'), \mathbf{M}_{bb'cc'}^{(2)}(h,h')$ and $\mathbf{M}_{bb'cc'}^{(3)}(h,h')$ as in Theorem 1 and using results (I), (II) and (III) we have,

$$\pi_{13} = \sum_{a=1}^{n}\sum_{1\le b<c\le m}\sum_{a'=1}^{n}\sum_{1\le b'<c'\le m}$$

$$\mathbf{\frac{1}{2}tr}\left[\left[\frac{\partial \mathrm{M}_{\mathbf{bc}}^{-1}(0)}{\partial \mathbf{w_k}}\right]\mathbf{cov}\left(\mathbf{X_{bc}(a)}, \mathbf{X_{b'c'}(a')}\right)\left[\frac{\partial \mathrm{M}_{\mathbf{b'c'}}^{-1}(0)}{\partial \mathbf{w_l}}\right]\mathbf{cov}\left(\mathbf{X_{b'c'}(a')}, \mathbf{X_{bc}(a)}\right)\right]$$

$$+\mathbf{\frac{1}{4}}\sum_{\mathbf{a=1}}^{\mathbf{n}}\sum_{\mathbf{1\le b<c\le m}}\sum_{\mathbf{a'=1}}^{\mathbf{n}}\sum_{\mathbf{1\le b'<c'\le m}}\left[\frac{\partial \log|\mathrm{M}_{\mathbf{bc}}(0)|}{\partial \mathbf{w_k}}\right]\left[\frac{\partial \log|\mathrm{M}_{\mathbf{b'c'}}(0)|}{\partial \mathbf{w_l}}\right]$$

$$= \frac{1}{2}\sum_{a=1}^{n}\sum_{1\le b<c\le m}\sum_{a'=1}^{n}\sum_{1\le b'<c'\le m} tr\left[\mathbf{N}_{bc,k}(0)\mathbf{M}_{bb'cc'}^{(0)}(a,a')\mathbf{N}_{b'c',l}(0)\mathbf{M}_{bb'cc'}^{(0)T}(a,a')\right] - \pi_9,$$

Similarly for $\pi_{14}, \pi_{15}, \pi_{16}$ we have,

$$\pi_{14} = \frac{1}{2}\sum_{a=1}^{n}\sum_{1\le b<c\le m}\sum_{i'=1}^{n-1}\sum_{b',c'=1}^{m} tr\left[\mathbf{N}_{bc,k}(0)\mathbf{M}_{bb'cc'}^{(1)}(a,i')\mathbf{N}_{b'c',l}(-1)\mathbf{M}_{bb'cc'}^{(1)T}(a,i')\right] - \pi_{10},$$

$$\pi_{15} = \frac{1}{2}\sum_{i=1}^{n-1}\sum_{b,c=1}^{m}\sum_{a'=1}^{n}\sum_{1\le b'<c'\le m} tr\left[\mathbf{N}_{b'c',l}(-1)\mathbf{M}_{bb'cc'}^{(2)}(i,a')\mathbf{N}_{bc,k}(0)\mathbf{M}_{bb'cc'}^{(2)T}(i,a')\right] - \pi_{11},$$

$$\pi_{16} = \frac{1}{2}\sum_{i=1}^{n-1}\sum_{b,c=1}^{m}\sum_{i'=1}^{n-1}\sum_{b',c'=1}^{m} tr[\mathbf{N}_{bc,k}(-1)\mathbf{M}_{bb'cc'}^{(3)}(i,i')\mathbf{N}_{b'c',l}(-1)\mathbf{M}_{bb'cc'}^{(3)T}(i,i')] - \pi_{12}.$$

By simplifying the quantities $\pi_1, \pi_2, \ldots, \pi_{16}$ then (11) holds.

Remark 3. The asymptotic variance-covariance matrix can now be derived, by using the relations (10) and (11) via (1).

Remark 4. For $\pi_{ijbc} = 1,$ for all i, j, c and $b,$ (10) and (11) reduce to the formulas obtained by Nematollahi and Kazemi [26].

4. Simulation studies

In the following section we have conducted some simulation studies to study the asymptotic relative efficiency of the WMPLE with respect to the MLE (ARE). The optimization method for the weighted pairwise likelihood function is the quasi-Newton method, with a relative tolerance of 10^{-8}, implemented by the *fmincon* function of the software package Matlab 2009a, v7.8. The *fmincon* function is a numerical optimization routine, including sequential quadratic programming algorithm to solve for constrained optima.

This section consists of three parts. In the first part, we assume that $m = 2$ and the variance-covariance matrix $\mathbf{\Sigma}$ is known and is given by $\mathbf{\Sigma} = \begin{bmatrix} 1 & .71 \\ .71 & 2 \end{bmatrix}$. The model (2) is then written as

$$\begin{bmatrix} X_{t1} \\ X_{t2} \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix}\begin{bmatrix} X_{t-1,1} \\ X_{t-1,2} \end{bmatrix} + \begin{bmatrix} Z_{t1} \\ Z_{t2} \end{bmatrix}, t = 1,\ldots,n.$$

We compare the variance of WMPLE computed by (3.8) and (11) with respect to MLE. We will study the relative efficiency of WPMLE with respect to MLE in terms of ϕ_{11} by letting $\phi_{12} = \phi_{21} = 0, \phi_{22} = 0.9$ and ϕ_{11} varying in the set $0.1, 0.2, ..., 0.9$.

The condition required for the existence of a causal stationary solution of the multivariate AR(1) is all the eigenvalues of $\mathbf{\Phi}$ are less than 1 in absolute value, i.e., provided that

$$det(\mathbf{I} - z\mathbf{\Phi}) \neq 0, \tag{12}$$

for all complex values of z, such that $|z| \leq 1$. For the case $m = 2$, this causality condition, which ensure us the existence of a stationary solution of the considered model can be written as

$$\det\begin{pmatrix} 1 - z\phi_{11} & -z\phi_{12} \\ -z\phi_{21} & 1 - z\phi_{22} \end{pmatrix} \neq 0,$$

for all complex values of z, such that $|z| \leq 1$. In the first example, this condition reduces to $|\phi_{11}| < 1$ and $|\phi_{22}| < 1$, which are satisfied for $\phi_{12} = \phi_{21} = 0, \phi_{22} = 0.9$ and ϕ_{11} varying in the set $0.1, 0.2, ..., 0.9$.

To compute the asymptotic variance of the MLE of matrix of coefficients $\mathbf{\Phi}$ we use the matrix $1/n(\mathbf{\Sigma} \otimes \mathbf{\Gamma}(0)^{-1})$, see Reinsel [32] and Shumway and Stoffer [33] for more details.

In Table 1, for various values of ϕ_{11}, the asymptotic variance of the WMPLE (avar($\tilde{\phi}_{11}$)), MLE (avar($\hat{\phi}_{11}$)), and ARE of the WMPLE with respect to the MLE for n=100 are reported.

Table 1. The avar($\tilde{\phi}_{11}$), the avar($\hat{\phi}_{11}$) and the ARE for n=100

ϕ_{11}	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
avar($\hat{\phi}_{11}$)	0.0175	0.0171	0.0165	0.0155	0.0141	0.0124	0.0103	0.0076	0.0042
avar($\tilde{\phi}_{11}$)	0.0226	0.0312	0.0339	0.0321	0.0268	0.0195	0.0143	0.0093	0.0044
ARE	0.7716	0.5494	0.4864	0.4827	0.5291	0.6372	0.7205	0.8233	0.9613

The asymptotic variances of the WMPLE ($\tilde{\phi}_{11}$), MLE ($\hat{\phi}_{11}$), and ARE of the WMPLE with respect to the MLE for different values of ϕ_{11} are plotted in Fig. 1.

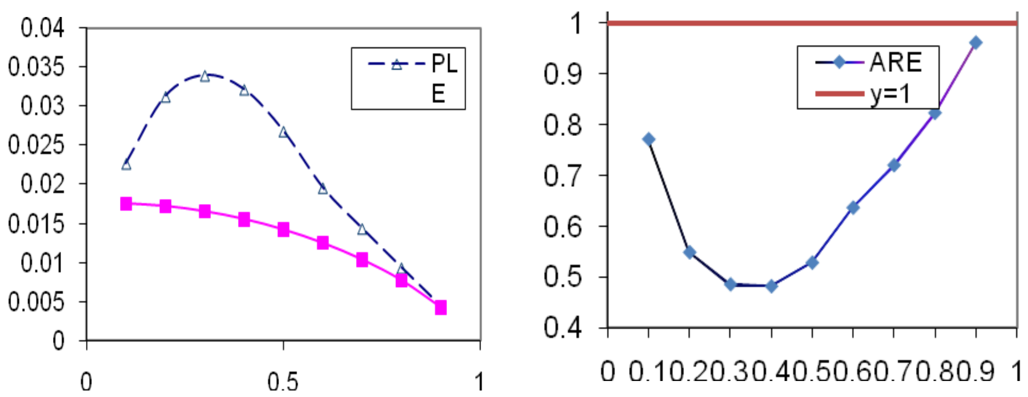

Fig. 1. The asymptotic variances of the WMPLE ($\tilde{\phi}_{11}$), MLE ($\hat{\phi}_{11}$) (left) and the relative efficiency of the WMPLE with respect to the MLE (ARE) (right)for different values of ϕ_{11}

As seen, the loss of efficiency is acceptable and balanced by the computational ease, however, the role of the values of n, m and ϕ_{11} in the value of efficiency is crucial.

In the second part, given $m=2$ and the variance-covariance matrix $\mathbf{\Sigma}$ considered in the first part, we present a preliminary study on performance of WMPLE in which all parameters in the matrix of coefficients are unknown but we limit attention to the case $\phi_{11} = \phi_{22}$ and $\phi_{12} = \phi_{21}$. The causality condition given by (4.1) of the above model, in this case, reduces to the following constraints

$$|\phi_{11} - \phi_{12}| < 1, |\phi_{11} + \phi_{12}| < 1.$$

We simulate 500 multivariate time series of the above AR(1) model with length $n = 1000$. For evaluating our estimator, we have reported the asymptotic variance of WMPLE (avar($\tilde{\phi}_{11}$)), and the asymptotic variance of the MLE (avar($\tilde{\phi}_{11}$)).

In Table 2, for various values of ϕ_{11} and ϕ_{12}, the sample means and standard errors of the WMPLE and MLE are reported.

Table 2. Sample means and standard errors of the WMPLE and MLE of the parameters ϕ_{11} and ϕ_{12}

Parameters	True values	Mean of WMPLE	Mean of MLE	SE of WMPLE	SE of MLE
(ϕ_{11}, ϕ_{12})	(0,0)	(.0063,.0114)	(.0091,.0088)	(.0352,.03)	(.0346,0.0283)
(ϕ_{11}, ϕ_{12})	(.75,.05)	(.7325,.0624)	(.7417,.0503)	(.0279,.02)	(.02,.02)
(ϕ_{11}, ϕ_{12})	(.75,.15)	(.7250,.1661)	(.7391,.1503)	(.03,.02)	(.0173,.0173)
(ϕ_{11}, ϕ_{12})	(.95,0)	(.9417,.0004)	(.9415,-.0014)	(.0084,.01)	(.0082,.01)
(ϕ_{11}, ϕ_{12})	(-.95,0)	(-.9386,-.0012)	(-.9388,.0013)	(.0173,.0079)	(.0173,.0079)

As shown in Table 2, the WMPLE performs very well.

In the third part, we simulate 500 multivariate time series of the considered AR(1) model with $m = 3$ and length $n = 500$. We assume that the variance-covariance matrix $\mathbf{\Sigma}$ is known and is given by $\mathbf{\Sigma} = \begin{bmatrix} 1 & 0.7 & 0.6 \\ 0.7 & 1 & 0.5 \\ 0.6 & 0.5 & 1 \end{bmatrix}$. The ARE is studied by letting the matrix of parameters be

$$\mathbf{\Phi} = \begin{pmatrix} 0.500 & 0.030 & 0.070 \\ 0.030 & 0.600 & 0.030 \\ 0.070 & 0.030 & 0.500 \end{pmatrix}.$$

Note that the condition required for the existence of a causal stationary solution of the multivariate AR(1) given by (4.1) is again satisfied for this proposed $\mathbf{\Phi}$. In the following table, the sample means and standard errors of the WMPLE and MLE for all parameters in $\mathbf{\Phi}$ including the ARE of the WMPLE with respect to the MLE are reported.

As Table 3 shows, the loss of efficiency is roughly more than one-half and acceptable.

As shown in Tables 1, 4.2 and 4.3 the WMPLE has a good practical performance. It is consistent, sufficiently accurate and acts as good as the MLE for the large sample size.

5. Discussion

In this paper, we have shown how weighting can be considered for inferences about the parameters in the multivariate AR(1) time series. By using the analytical and numerical computations, we showed that where the unweighted pairwise likelihood does poorly in efficiency and may be computationally very slow, difficult and complicated, the weighted pairwise likelihood is suggested as a practically good choice.

It also seems that dimension parameter m is a key variable here, which should be brought into play, as m is the key variable in increasing the computational complexity that might encourage one to use pairwise.

Table 3. Sample means, standard errors and ARE of the WMPLE with respect to the MLE of the parameters of the matrix $\mathbf{\Phi}$

Parameters	True values	Mean of WMPLE	Mean of MLE	SE of WMPLE	SE of MLE	ARE
ϕ_{11}	0.500	0.415	0.474	0.2079	0.1400	0.4386
ϕ_{12}	0.030	0.053	0.025	0.1810	0.1237	0.6830
ϕ_{13}	0.070	0.106	0.069	0.1799	0.1164	0.6540
ϕ_{21}	0.030	0.051	0.020	0.2145	0.1444	0.659
ϕ_{22}	0.600	0.551	0.584	0.2331	0.1217	0.5220
ϕ_{23}	0.030	0.054	0.025	0.2154	0.1118	0.5190
ϕ_{31}	0.070	0.084	0.078	0.2006	0.1382	0.6889
ϕ_{32}	0.030	0.015	0.019	0.1756	0.1248	0.7107
ϕ_{33}	0.500	0.404	0.459	0.1917	0.1200	0.6259

References

[1] Schelter, B., Winterhalder, M. & Timmer, J., (2006). *Handbook of Time Series Analysis*: Recent Theoretical Developments and Applications. Berlin, Wiley-VCH.

[2] Xu, J. J. (1996). *Statistical modelling and inference for multivariate and longitudinal discrete response data, Unpublished doctoral dissertation, Department of Statistics*. Vancouver University of British Columbia.

[3] Parner, E. T. (2001). A composite likelihood approach to multivariate survival data. *Scand. J. Stat. 28*, 295-302.

[4] Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society B, 36*, 192-236.

[5] Besag, J. (1977). Efficiency of pseudo-likelihood estimation for simple Gaussian fields. *Biometrika*, *64*, 616-618.

[6] Cox, D. R. (1975). Partial likelihood. *Biometrika*, *62*, 269-276.

[7] Lindsay, B. G. (1988). Composite likelihood methods. *Contemp. Math*, *80*, 221-240.

[8] Hannan, E. J. & Rissanen, J. (1982). Recursive estimation of mixed autoregressive moving average order. *Biometrika*, *69*, 81-94.

[9] Salau, M. O. (2003). The effects of different choices of order for autoregressive approximation on the Gaussian likelihood estimates for ARMA models. *Statistical papers*, *44*, 89-105.

[10] Heyde, C. C. & Lin, Y-X. (1992). On quasi-likelihood methods and estimation for branching processes and heteroscedastic regression models. *Aust. J. Stat*. *34*, 199-206.

[11] Heyde, C. C. (1997). *Quasi-likelihood and its applications*. New York, Springer.

[12] Jung, R. C., Ronning, G. & Tremayne, A. R. (2005). Estimation in conditional first order autoregression with discrete support. *Statistical papers*, *46*, 195-224.

[13] Rothenberg, T. J. & Stock, J. H. (1997). Inference in a nearly integrated autoregressive model with nonnormal innovations. *Journal of Econometrics*, *80*, 269-286.

[14] Carstensen, K. (2003). The nite-sample performance of robust unit root tests. *Statistical papers*, *44*, 469-482.

[15] Lee, L. F. (1999). Statistical Inference with Simulated Likelihood Functions. *Econometric Theory, 15*, 337-360.

[16] Ziegler, A. (2007). Simulated classical tests in multinomial probit models. *Statistical papers*, *48*, 655-681.

[17] Bollerslev, T. & Wooldridge, J. (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Economet Rev*., *11*, 143-172.

[18] Newey, W. & Steigerwald, D. (1997). Asymptotic bias for quasi maximum likelihood estimators in conditional heteroskedasticity models. *Econometrica*, *65*, 587-599.

[19] Rodrigues, P. M. M & Rubia, A. (2008). A note on testing for nonstationarity in autoregressive processes with level dependent conditional heteroskedasticity. *Statistical papers*, *49*, 581-593.

[20] Kharrati-Kopaei, M., Nematollahi, A. R. & Shishebor, Z. (2009). On the sufficient statistics for multivariate ARMA models: approximate approach. *Statistical papers*, *50*(2), 261-276.

[21] Hu, X. J. & Zhang, B. (2010). Pseudolikelihood ratio test with biased observations, Statistical papers, in press.

[22] Cox, D. R. & Ried, N. (2004). A note on pseudo likelihood constructed from marginal densities. *Biometrika*, *91*, 729-737.

[23] Varin, C. (2008). On composite marginal likelihoods. *Advances in statistical Analysis*, *92*, 1-28.

[24] Ng, C. T., Joe, H., Karlis, D. & Liu, J. (2011). Composite likelihood for time series models with a latent autoregressive process. *Statistica Sinica*, *21*, 279-305.

[25] Davis, R. A. & Yau, C. Y. (2011). Comments on

pairwise likelihood in time series models, Statistica Sinica, *21*, 255-277.

[26] Nematollahi, A. R. & Kazemi, M. R. (2010). Pairwise likelihood inferences for the multivariate AR(1) models, Technical Report. Department of Statistics, Shiraz University.

[27] Joe, H. & Lee, Y.(2009). On weighting of bivariate margins in pairwise likelihood. *Journal of Multivariate Analysis*, *100*, 670-685.

[28] Godambe, V. (1991). *Estimating Functions*. New York, Oxford University Press

[29] Zhao, Y. (2004). *Statistical modelling and inference for discrete and censored familial data, Unpublished doctoral dissertation*. Department of Statistics, University of British Columbia, Vancouver.

[30] Zhao, Y. & Joe, H. (2005). Composite likelihood estimation in multivariate data analysis. *The Canadian Journal of Statistics*, *33*(3), 335-356.

[31] Searle, S. R. (1971). *Linear models*. New York, John-Wiley.

[32] Reinsel, G. C. (1997). *Elements of multivariate time series analysis*. second edition, Springer-Verlag, New York.

[33] Shumway, R. H. & Stoffer, D. S. (2005). *Time series analysis and its applications with R example.* second edition, Springer Science, New York.