THE EXACT GAUSSIAN LIKELIHOOD
ESTIMATION OF TIME-DEPENDENT VARMA MODELS

Abdelkamel Alj*, Kristján Jónasson† and Guy Mélard‡.

June 13, 2014

Abstract

An algorithm for the evaluation of the exact Gaussian likelihood of an \( r \)-dimensional vector autoregressive-moving average (VARMA) process of order \((p, q)\), with time-dependent coefficients, including a time dependent innovation covariance matrix, is proposed. The elements of the matrices of coefficients and those of the innovation covariance matrix are deterministic functions of time and assumed to depend on a finite number of parameters. These parameters are estimated by maximizing the Gaussian likelihood function. The advantages of that approach is that the Gaussian likelihood function can be computed exactly and efficiently. The algorithm is based on the Cholesky decomposition method for block-band matrices. It is shown that the number of operations as a function of \( p, q \) and \( n \), the size of the series, is barely doubled with respect to a VARMA model with constant coefficients. A detailed description of the algorithm followed by a data example is provided.

Key words and phrases: Vector ARMA, Time-varying models, Cholesky decomposition method.

*Dept of Mathematics, Université libre de Bruxelles CP112, Biv du Triomphe, B-1050 Bruxelles, Belgium (e-mail: abdelalj@ulb.ac.be).
†Department of Computer Science, School of Engineering and Natural Sciences, University of Iceland (e-mail: Jonasson@hi.is).
‡ECARES, Université libre de Bruxelles CP114/4, Avenue Franklin Roosevelt 50, B-1050 Bruxelles, Belgium (e-mail: gmelard@ulb.ac.be).
1 Introduction

Let \( \{x_t : t \in \mathbb{Z}\} \) be an \( r \)-vector valued time series. We assume that \( x_t \) follows the time dependent vector autoregressive moving average process of order \((p, q)\), in short \( \text{tdVARMA}(p, q) \), defined by the following equations:

\[
x_t = \sum_{i=1}^{p} A_{t,i} x_{t-i} + y_t, \quad (1.1)
\]

\[
y_t = g_t \epsilon_t + \sum_{j=1}^{q} B_{t,j} g_{t-j} \epsilon_{t-j}, \quad (1.2)
\]

where \( p \) are \( q \) are positive integer constants, the coefficients \( A_{t,i}, i = 1, \ldots, p \) and \( B_{t,j}, j = 1, \ldots, q \), and \( g_t \) are \( r \times r \) matrices and their elements are deterministic functions of time and assumed to depend on a finite number of parameters, and the \( \epsilon_t \)'s are independent Gaussian \( r \)-dimensional random variables, with zero mean and a positive definite covariance matrix \( \Sigma \). A stochastic equation like (1.1)-(1.2) requires initial conditions. They will be presented later in Section 3. The vector of parameters of interest \( \theta = (\theta_1, \ldots, \theta_m) \) belongs to an open set \( \Theta \subset \mathbb{R}^m \). \( \Sigma \) is considered as a nuisance parameter. In the following we use \( T \) and \( \det(\cdot) \) as the transpose and the determinant, respectively. If the \( A_{t,i} \)'s, \( B_{t,j} \)'s, and \( g_t \) do not depend on \( t \), then \( g_t \) can be omitted and the data generating process is the classical VARMA process of order \((p, q)\), e.g. Reinsel (1998) or Lütkepohl (2005), where the standard stationarity, invertibility and identifiability conditions are stated. By the way, Lütkepohl (2005, Chapter 17) covers the \( \text{tdVAR}(p) \) processes. When \( r = 1 \), we are in the case of classical ARMA processes. Note that \( \text{tdARMA} \) processes have been more deeply studied in the literature, see references in Azrak and Mélard (2006), than \( \text{tdVARMA} \) processes. Presence of \( g_t \) allows for (unconditional) heteroscedasticity. Note that some recent papers deal with time-varying models using a Bayesian approach (Triantafyllopoulos and Nason, 2007), periodic coefficients (Hindrayanto, Koopman and Ooms, 2010), or focussing on tests (Chen and Hsu, 2014).

In principle, equation (1.1) should contain an intercept term or a deterministic trend. This has not been done here in order to simplify the related estimation algorithms. It is assumed that the possibly time-dependent mean has been subtracted prior to estimation. There would be no problem to estimate these trend parameters simultaneously with the \( \text{tdVARMA} \) parameters.

Now consider a sample size \( n \) and let the \( nr \times 1 \) vector \( x = (x_1^T, \ldots, x_n^T)^T \) be an observed time series of the process defined in (1.1)-(1.2) from which we want to estimate \( \theta \). The likelihood function is a density function evaluated as a function of the parameters and \( x \). It is given by

\[
L(\theta; x) = (2\pi)^{-nr/2} \det(S)^{-1/2} \exp\left(\frac{1}{2} x^T S^{-1} x \right), \quad (1.3)
\]

with \( S = \text{cov}_\theta(x) \) the \( nr \times nr \) covariance matrix of \( x \). Mélard (1982) showed in the scalar case, \( r = 1 \), how to use the Cholesky decomposition method to obtain the
exact likelihood of an ARMA process with time dependent coefficients. Azrak and Mélard (1998) developed an algorithm based on the Kalman filter. In this article we show in detail how to compute efficiently the exact likelihood (1.3) by a generalization of Mélard (1982) to the multivariate case. The idea is to transform \( x \) into another vector \( w \) such that the matrix \( S \) is transformed into a block band matrix depicted in equations (3.2) and (3.3). The block band structure allows for an efficient evaluation of the likelihood. This is done by modifying an algorithm for standard VARMA models, i.e. with constant coefficients, developed by Jónasson and Ferrando (2008). The submatrices in (3.2) and (3.3) were used in that article without the first subscript. Consequently our paper is also a generalization of that algorithm to models with time-dependent coefficients using the Matlab implementation of Jónasson (2008).

The asymptotic properties of the estimation methods are the subject of ongoing work. The Gaussian specification is not a restriction. Indeed, the properties are valid for other distributions provided some assumptions are satisfied, such as existence of 8th-order moments. Note also that identifiability conditions should be verified on \( \Sigma_t = g_t \Sigma g_t^T \), the covariance matrix of \( g_t \epsilon_t \). Here we impose that the non-zero elements \((i, j)\) of \( g_t \) are such that \( \prod_{t=1}^{n} (g_{ij}) = 1 \). Alternatively, a parametrization of \( \Sigma_t \) can be used.

**Example 1.1** For simplicity, we consider the following tdVMA(1) model with \( r = 2 \)

\[
x_t = B_{t,1} g_{t-1} \epsilon_{t-1} + g_t \epsilon_t,
\]

where \( B_{t,1} \), \( g_t \) and \( \Sigma \) are given by

\[
\begin{align*}
B_{t,1} &= \left( \begin{array}{cc}
B'_{11} & B'_{12} \\
B'_{21} & B'_{22}
\end{array} \right) + \frac{1}{n-1} (t - \frac{n+1}{2}) \left( \begin{array}{cc}
B''_{11} & B''_{12} \\
B''_{21} & B''_{22}
\end{array} \right), \\
g_t &= \left( \begin{array}{cc}
ex^{\eta_{11}(t - \frac{n+1}{2})} & 0 \\
0 & ex^{\eta_{22}(t - \frac{n+1}{2})}
\end{array} \right)
\end{align*}
\text{and} \quad \Sigma = \left( \begin{array}{cc}
1 & 0 \\
0 & 1
\end{array} \right).
\]

Then, the vector of the parameters to be estimated is

\[
\theta = (B'_{11}, B'_{12}, B'_{21}, B'_{22}, B''_{11}, B''_{12}, B''_{21}, B''_{22}, \eta_{11}, \eta_{22})^T.
\]

Indeed, a linear or an exponential function of time are the simplest generalizations after a constant. We prefer an exponential function for \( g_t \) to keep its elements positive. Otherwise, polynomial functions of time can be considered, if necessary. The expression is based on orthogonal polynomials to reduce quasi multicollinearity problems so we use \((t - \frac{n+1}{2})\) instead of just \( t \), see Azrak and Mélard (2006). The factor \( 1/(n-1) \) (it might be \( 1/n \)) guarantees that the elements of \( B_{t,1} \) and \( g_t \) are bounded. This is needed for the asymptotic theory but it is usually omitted in practical estimation, like in Section 6. Then, the coefficients \( B''_{ij} \) and \( \eta_{ij} \) will often be very small. Other parametrizations are of course possible in other contexts. For example, it is possible to provide a generalization of Kwoun and Yajima (1986), where \( B_t(\theta) = B' \sin(\alpha_t t + B'') \) to the bivariate case so we can have for \( i, j = 1, 2 \)

\[
B''_{ij}(\theta) = B'_{ij} \sin(\alpha_{ij} t + B''_{ij}),
\]
with fixed $\alpha_{ij}$. Of course, periodic coefficients like Hindrayanto et al. (2010) is also a trivial special case.

Example 1.1 can be considered as a vector generalization of the tdMA(1) univariate process given in Azrak and Mélard (2006, Example 4). Here the coefficients depend also on the length $n$ of the series but this has no effect on the algorithm.

2 Exact likelihood function of VARMA models

2.1 Assumptions

We consider a zero mean $r$-dimensional Gaussian VARMA($p, q$) process, which satisfies the following equation and can be considered as a special case of (1.1) and (1.2):

$$x_t - \sum_{i=1}^{p} A_i x_{t-i} = \epsilon_t + \sum_{j=1}^{q} B_j \epsilon_{t-j},$$

(2.1)

where the matrices $A_1, \ldots, A_p$ and $B_1, \ldots, B_q$ are of dimension $r \times r$ and the $\epsilon_t$'s are like in Section 1. Let us denote by $\theta$ the $(p + q)r^2$-dimensional vector of all the parameters to be estimated except for $\Sigma$. In our context, we consider $\Sigma$ as a nuisance parameter, not as a parameter of interest. We assume stationarity and invertibility, which require that

$$\det \{ A(z) \} \neq 0, \quad \det \{ B(z) \} \neq 0, \quad \forall z \in \mathbb{C}, \quad \text{such that} \quad |z| \leq 1,$$

(2.2)

where $A(z) = I_r - A_1 z - A_2 z^2 - \ldots - A_p z^p$, $B(z) = I_r + B_1 z + B_2 z^2 + \ldots + B_q z^q$, $I_r$ is the identity $r \times r$ matrix. There are also conditions for identifiability, see e.g. Hannan and Deistler (1988) and Caines (1988). Under these conditions a single model is selected from a set of models that are likelihood equivalent. The likelihood function is given by an expression similar to (1.3), where $S$ has a simpler form than in Section 1, to be discussed below, and the log-likelihood function of (2.1) is given by

$$l(\theta; x) = -\frac{1}{2} \left[ nr \log(2\pi) + \log \{ \det(S) \} + x^T S^{-1} x \right],$$

(2.3)

The evaluation of (2.3) requires computing both the inverse and the determinant of the symmetric matrix $S$, hence the number of multiplications by direct evaluation is roughly proportional to $(nr^2)^3$. Some algorithms were developed to reduce substantially the computational effort, e.g. Shea (1984, 1989), Penzer and Shea (1997) via a Kalman filter and Mauricio (2002), Jónasson and Ferrando (2008) via the Cholesky decomposition method. We present an algorithm based on a Cholesky factorization of a block band matrix instead of the block matrix $S$, like in Jónasson and Ferrando (2008). For other references, see Jónasson and Ferrando (2008). Related references are also Mauricio (2006) and Mélard et al. (2006). In the latter paper, VARMA models in echelon form are handled by maximum likelihood. In principle, this can also be done in the present context.
2.2 Cholesky decomposition method

Ansley (1979) has obtained a band covariance matrix for a univariate ARMA model, which enables a simple Cholesky decomposition and computationally efficient evaluation of the likelihood. Mauricio (2002) has adapted Ansley’s algorithm to the multivariate case and argues that it is more efficient for likelihood evaluation than the Kalman filter methods. The fundamental device used by Ansley is a transformation from $x$ to a vector of random variables $z_t$ defined as follows

$$z_t = \begin{cases} 
  x_t & \text{for } t \leq s^* \\
  y_t = x_t - \sum_{i=1}^{p} A_i x_{t-i} & \text{for } t = s^* + 1, ..., n,
\end{cases}$$

(2.4)

where $s^* = \max(p, q)$. In our method, we prefer to use the slightly more efficient variant where $s^* = p$, first used in the univariate case by Mélard (1982) and Penzer and Shea (1997), and in the multivariate case by Jónasson and Ferrando (2008).

The idea behind it is to transform $x^T S^{-1} x$ to $w^T w$, with $w$ a vector of uncorrelated random variables. Let $z = (z_1^T, ..., z_n^T)^T$, then the transformation from $x$ to $z$ has unit Jacobian so that (2.3) can be expressed as:

$$-\frac{1}{2} \left[(nr \log(2\pi) + \log \{\det(\Omega)\} + z^T \Omega^{-1} z)\right],$$

(2.5)

where $\Omega$ is a symmetric block-band matrix with blocks $\Omega(t, s)_{r \times r} = \text{cov}(z_t, z_s)$. That matrix $\Omega$ contains the following covariances matrices:

$$W_j = \text{cov}(y_t, y_{t-j}) = \sum_{i=j}^{q} B_i \Sigma B_i^T, \quad j = 0, 1, ..., q,$$

(2.6)

$$G_j = \text{cov}(y_t, x_{t-j}) = W_j + \sum_{i=1}^{\min(p, q-j)} G_{t+i-j} A_i^T, \quad j = q, q-1, ..., 0.$$

(2.7)

Let $B_0 = I_r$, $W_j = 0$ and $G_j = 0$ for $j < 0$ and $j > q$. We have

$$S_j = \text{cov}(x_t, x_{t-j}), \quad \text{for } j = 0, 1, ..., p,$$

which satisfy the following linear system (the vector Yule-Walker equations)

$$S_0 - A_1 S_1^T - A_2 S_2^T - ... - A_p S_p^T = G_0$$

$$S_1 - A_1 S_0 - A_2 S_1^T - ... - A_p S_{p-1}^T = G_1$$

$$S_2 - A_1 S_1 - A_2 S_0 - ... - A_p S_{p-2}^T = G_2$$

$$...$$

$$S_p - A_1 S_{p-1} - A_2 S_{p-2} - ... - A_p S_0 = G_p.$$ 

(2.8)

We have a linear system with the $r(r+1)/2 + pr^2$ elements of $S_0, ..., S_p$ (noting that $S_0$ is symmetric). The solution of this linear system is dealt with in Jónasson and Ferrando (2008, appendix B) which is close to that given by Mauricio (1997). The computation
of the theoretical autocovariance of a VARMA process has been considered by Ansley (1980), Hall and Nicholls (1980), Kohn and Ansley (1982) and Mauricio (1995).

The matrix $\Omega$ is a block band matrix, with $\max(p, q + 1)$ diagonal blocks along each side of the main diagonal blocks. Its precise form in the case $p \geq q$ and $p < q$ can be seen in Section 3.2 by omitting the first subscript everywhere. Now taking the Cholesky decomposition $\Omega = LL^T$, where $L$ is a block band lower triangular matrix, the rows of $L$ are evaluated recursively using a standard algorithm for block matrices and the entries in $w = L^{-1}z$ can be calculated by block forward substitution. Then $\det(\Omega) = \prod_{i=1}^{nr} l_{ii}^2$, where $l_{ii}$ is $(i, i)$-th element of $L$ and the log-likelihood function is given by

$$-\frac{1}{2} \left\{ nr \log(2\pi) + 2 \sum_{i=1}^{nr} \log(l_{ii}) + w^T w \right\}. \quad (2.9)$$

**Remark 2.1** The method used by Jónasson and Ferrando (2008) is slightly different from ours in the calculation of $G_j$. Their method is based on the calculations of a set of covariances $C_j$ defined by

$$C_j = \text{cov} (x_t, \epsilon_{t-j}) = B_j \Sigma + \sum_{i=1}^{\min(p, j)} A_i C_{j-i},$$

where $B_0 = I_r$ and $C_j = 0$ for $j < 0$ and $j > q$. Then from (2.7) we get

$$G_j = \sum_{i=j}^{q} B_i C_{i-j}^T. \quad (2.10)$$

Determining the $W_j$’s, $C_j$’s and $G_j$’s by using (2.6), (2.7) and (2.10) costs about $r^3(s^2/2 + q^2)$ multiplications (Jónasson and Ferrando, 2008) where $s = \min(p, q)$. However, by using our algorithm the determination of $G_j$’s and $W_j$’s costs about $r^3q^2$ multiplications if $p \geq q$ and otherwise about $r^3(q^2/2 + qp - p^2/2)$ multiplications. Thus, our variant has two advantages, since it is slightly faster, and also simpler by calculating the covariances $G_j$ by a simple recurrence formula.

### 2.3 The algorithm

Thus an efficient algorithm for computing the exact likelihood function defined in (2.3) for a standard VARMA($p, q$) process exploiting the full symmetric block-band property of $\Omega$ is as follows. For a given value of $\theta$, the log-likelihood is computed through the following stages.

**Stage 1:** Compute the covariances $W_j$ and $G_j$ for $j = 0, 1, ..., q$ by using (2.6)-(2.7), and compute the covariances $S_j$ for $j = 0, 1, ..., p - 1$, by solving the linear system of vector-Yule-Walker equations defined in (2.8).

**Stage 2:** Compute $z_t$ for $t = p + 1, ..., n$ by (2.4).
Stage 3: Compute the matrix $L$ by a block triangularization algorithm for the block band matrix $\Omega$.

Stage 4: Determine $w = L^{-1} z$ using block forward substitution.

Stage 5: Compute $w^T w$, $\log \{ \det(\Omega) \} = 2 \sum_{i=1}^{nr} \log(l_{ii})$ and (2.9).

3 The exact likelihood function of tdVARMA($p$, $q$) models

3.1 Initialization

As explained in Section 1, the tdVARMA($p$, $q$) model requires initial condition. For startup reasons, as we have no information on the past behavior of the process, we assume

$$
\begin{align*}
A_{t,i} &= A_i := A_{0,i} \quad \text{for} \quad i = 1, \ldots, p \quad \text{and} \quad t = 0, -1, \ldots, -p + 1, \\
B_{t,j} &= B_j := B_{0,j} \quad \text{for} \quad j = 1, \ldots, q \quad \text{and} \quad t = 0, -1, \ldots, -q + 1, \\
g_t &= g_1 \quad \text{for} \quad t = 0, -1, \ldots, -q + 1,
\end{align*}
$$

(3.1)

the last line for compatibility with the LIKAMT program (Mélard, 1982), instead of $g_0$ which would have been more coherent. We suppose in addition (2.2) so that the process defined in (1.1) and (1.2) is causal and invertible in the past ($t \leq 0$). The assumption of causality in the past implies that we can calculate the covariance of the stationary process at $t = 0$. For $t \leq 0$ the covariances $S_{t,j} = S_j := S_{0,j}$ for $0 \leq j \leq p$ are obtained like in the stationary VARMA($p$, $q$) models by solving the linear system of equations given in (2.8), but with $\Sigma$ in (2.6) replaced by $g_1 \Sigma g_1^T$. We refer the reader to e.g. Lanne and Saikkonen (2010) for the studies of non-causal VAR models, and for (univariate) AR models, see e.g. Brockwell and Davis (1991, Chapter 3). The assumption of invertibility in the past implies that the non-stationary process is invertible (Hallin, 1978), i.e. for all finite $t$, $\epsilon_t$ can be expressed as a mean-square convergent limit of linear combinations of $(x_t, x_{t-1}, x_{t-2}, \ldots)$. Without that property, the process would remain unidentifiable and inefficient forecasts would be produced by the model, see Hallin (1986).

3.2 The likelihood function

Using a transformation like in (2.4) for the model defined in (1.1)-(1.2) but with $A_i$ replaced by $A_{t,i}$, we obtain $z = \Lambda x$ where $\Lambda$ is the $nr \times nr$ lower triangular block-band
matrix given by:

\[
\Lambda = \begin{bmatrix}
I_r & 0 & \ldots & 0 \\
0 & \ddots & & \\
-A_{p+1,p} & \ldots & -A_{p+1,1} & I_r \\
& \ddots & \ddots & \ddots & 0 \\
& & -A_{n,p} & \ldots & -A_{n,1} & I_r
\end{bmatrix}.
\]

Since \( \Lambda \) has unit diagonal, the log-likelihood (1.3) can be expressed as:

\[
I(\theta; z) = -\frac{1}{2}[nr \log(2\pi) + \log\{\det(\Omega)\} + z^T \Omega^{-1} z],
\]

but the blocks of \( \Omega = \text{cov}_\theta(z) \) are now the \( r \times r \) covariances matrices:

\[
\begin{align*}
W_{t,j} &= \text{cov}(y_t, y_{t-j}), \quad j = 0, 1, \ldots, q, \\
G_{t,j} &= \text{cov}(y_t, x_{t-j}), \quad j = 0, 1, \ldots, q, \\
S_{t,j} &= \text{cov}(x_t, x_{t-j}), \quad j = 0, 1, \ldots, p,
\end{align*}
\]

which depend on \( t \). We will provide relations to evaluate these matrices in Section 3.3. More precisely, if \( p \geq q \),

\[
\Omega = \begin{bmatrix}
S_{1,0} & S_{2,1}^T & S_{p,p-1}^T & 0 & \ldots & 0 \\
S_{2,1} & S_{2,0} & \ddots & G_{p+1,q}^T & 0 & \\
& \ddots & \ddots & \ddots & \ddots & \ddots \\
& & S_{p,1} & S_{p,0}^T & G_{p+1,1}^T & G_{p+q,q}^T & 0 & \ldots & 0 \\
0 & G_{p+1,q} & \ldots & G_{p+1,1} & W_{p+1,0} & W_{p+2,1} & W_{p+q+1,q} & 0 & \ldots & 0 \\
0 & & \ddots & \ddots & W_{p+2,1} & W_{p+2,0} & W_{p+3,1} & \ddots & \ddots & \ddots \\
& \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
& & & & & & & & & & \ddots \\
0 & & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
0 & \ldots & 0 & 0 & 0 & \ldots & 0 & W_{n,q} & W_{n-1,1} & W_{n-1,0} & W_{n-1,1}^T & W_{n,1} & W_{n,0}
\end{bmatrix},
\]

(3.2)
whereas if $p < q$, the depiction is slightly different with

$$
\begin{bmatrix}
S_{1,0} & S_{2,1}^T & S_{p,p-1}^T & G_{p+1,p}^T & G_{q+1,q}^T & 0 & \cdots & 0 \\
S_{2,1} & S_{2,0} & \cdots & G_{p+1,2}^T & \cdots & \cdots & \cdots & \cdots \\
S_{p,p-1} & S_{p,1} & S_{p,0} & G_{p+1,1}^T & G_{p+2,2}^T & \cdots & \cdots & \cdots \\
G_{p+1,p} & G_{p+1,2} & G_{p+1,1} & G_{p+2,2} & \cdots & \cdots & \cdots & \cdots \\
G_{q+1,q} & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
G_{p+q,q} & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & \cdots & 0 & 0 & 0 & W_{n,q} & \cdots & \cdots \\
0 & \cdots & 0 & 0 & 0 & \cdots & W_{n-1,0} & W_{n-1,1} \\
\end{bmatrix}
$$

(3.3)

Let us now detail the computation of these blocks.

### 3.3 Computation of the covariance matrix $\Omega$

The following relations are a generalization of Mélard (1982) to the vector case.

- The covariances $W_{t,j}$

For $t = 1, \ldots, n$ and $j = 0, 1, \ldots, q$,

$$
W_{t,j} = \sum_{k=0}^{q} \sum_{l=0}^{q} B_{t,k} g_{l-k} \text{cov} (\epsilon_{t-k}, \epsilon_{t-l}) g_{t-j}^T B_{l-j,k}^T
$$

and since $\text{cov} (\epsilon_{l-k}, \epsilon_{t-l}) = \Sigma \delta_{k,j+l}$, with $\delta_{k,j} = 1$ if $k = j$ and 0 otherwise, letting $B_{t,0} = I_r$, we have

$$
W_{t,j} = \sum_{k=j}^{q} B_{t,k} g_{l-k} \Sigma g_{l-k}^T B_{l-j,k-j}^T, \quad j = 0, 1, \ldots, q, \quad (3.4)
$$

with $W_{t,k} = 0$ for $k > q$.

- The covariances $G_{t,j}$

For $t = 1, \ldots, p+q$ and $j = q, q-1, \ldots, 0$, the covariances $G_{t,j}$ can be obtained by the
recurrence formula:

\[ G_{t,j} = W_{t,j} + \sum_{i=1}^{\min(p,q-j)} G_{t,i+j} A_{t,j,i}^T, \]  

(3.5)

and \( G_{t,j} = 0 \) for \( j > q \), because \( y_t \) is composed of \( g_{t-1} \epsilon_{t-1}, \ldots, g_{t-q} \epsilon_{t-q} \) which are independent from \( x_{t-j} \).

- The covariances \( S_{t,j} \)

For \( t = 1, \ldots, p \) and \( j = 0, 1, \ldots, p \), the autocovariances \( S_{t,j} \) of the process are obtained by the recurrences

\[ S_{t,j} = G_{t,j} + \sum_{i=1}^{p} A_{t,i} S_{t-i,j-i}, \]  

(3.6)

recalling that \( S_{t,j} = S_{t,0} \) for \( t \leq 0 \) and noting that for \( j - i < 0 \)

\[ S_{t-j,i-j} = \text{cov}(x_{t-i}, x_{t-j}) = \text{cov}(x_{t-j}, x_{t-i})^T = S_{t-j,i-j}^T. \]  

(3.7)

4 The algorithm

To our knowledge, this is the first algorithm which extends the Cholesky decomposition method to the tdVARMA case. In scalar tdARMA models there exist only two algorithms to compute the exact likelihood function. The first one denoted LIKAMT (see Mélard, 1982) was also based on a Cholesky decomposition. The second one TKALMAR is based on the Kalman filter (see Azrak and Mélard, 1998). In addition Azrak and Mélard (1998) have compared the two algorithms in the univariate case. Their conclusions are that LIKAMT is faster for some models. On the contrary, for models with order \( p \) and \( q \) equal or larger than 13, they can observe a superiority of TKALMAR.

Now we can summarize the algorithm in the case of tdVARMA models in the 6 stages below:

**Stage 1:** Starting conditions:

- The coefficients \( A_{t,i} \) and \( B_{t,j} \) for \( t \leq 0, \ i = 1, \ldots, p, \ j = 1, \ldots, q \) and the values of \( g_t \) for \( t \leq 1 \) are determined by using (3.1);
- The covariances \( W_{0,j}, G_{0,j} \) for \( j = 0, \ldots, q \) are determined using (2.6) and (2.7), and the covariances \( S_{0,j} \) for \( j = 1, \ldots, p \) are obtained by solving (2.8).

**Stage 2:** We compute:

- The covariances \( W_{t,j} \) by using (3.4) for \( j = 0, 1, \ldots, q \) and \( t = 1, \ldots, n \);
- The covariances \( G_{t,j} \) by using (3.5) for \( j = q, q-1, \ldots, 0 \) and \( t = p+1, \ldots, p+q \).
• The covariances $S_{t,j}$ by using (3.6) for $j = 0, 1, ..., p$ and $t = 1, ..., p$.

They are stored in $\Omega$ as in Section 3.2.

**Stage 3:** $z_t = y_t$ is computed for $t = p + 1, ..., n$.

**Stage 4:** Compute the matrix $L$ by a triangularization algorithm by blocks of the block band matrix $\Omega$.

**Stage 5:** Determine $w = L^{-1}z$ using block forward substitution.

**Stage 6:** Compute $w^T w$, $\log \{\det(\Omega)\} = 2 \sum_{i=1}^{n_p} \log(I_{i})$ and (2.9).

**Remark 4.1** In Mélard (1982) and its implementation for the univariate case $r = 1$ only four square matrices of size $\max(p,q) + 1$ are needed.

**Remark 4.2** In the case where there is a deterministic trend, as discussed in the Introduction, none of the matrix calculations in the Algorithm of Section 4 would change; only the vector computations in stages 3, 5 and 6 would be affected.

**Remark 4.3** It could be argued that it would be more numerically stable to use QR rather than Cholesky factorization. However, extensive numerical experiments done in Jónasson (2008), and following that work, indicate that the Cholesky method is unlikely to suffer instabilities for practical VARMA time series modelling. We see no reason why adding time dependence should change that.

### 4.1 An illustration

We consider the following mixed $r$-variate tdVARMA(2,1) process,

$$x_t = A_{t,1}x_{t-1} + A_{t,2}x_{t-2} + B_{t,1}g_{t-1}\epsilon_{t-1} + g_t\epsilon_t,$$

where the coefficients $A_{t,1}, A_{t,2}, B_{t,1}$ and $g_t$ are deterministic functions of time, then

$$z_t = \begin{cases} 
  x_t & \text{if } t = 1, 2, \\
  y_t & \text{if } t = 3, ..., n,
\end{cases}$$

and

$$\Omega = \begin{pmatrix}
  S_{1,0} & S_{1}^T \\
  S_{2,1} & S_{2,0} & G_{3,1}^T \\
  G_{3,1} & W_{3,0} & W_{4,1}^T \\
  & W_{4,1} & W_{4,0} & \ddots \\
  & & \ddots & \ddots & \ddots \\
  & & & \ddots & \ddots & \ddots \\
  & & & & \ddots & \ddots & \ddots & \ddots \\
  & & & & & W_{n-1,0} & W_{n,1}^T \\
  & & & & & W_{n,1} & W_{n,0}
\end{pmatrix}.$$
\begin{itemize}
\item **Stage 1:**
\[
\begin{aligned}
A_{t,i} &= A_{0,i} & t &= -1, 0, \quad i = 1, 2, \\
B_{t,j} &= B_{0,j} & t &= 0, \quad j = 1,
\end{aligned}
\]
and the covariances $W_{0,0}, W_{0,1}, G_{0,0}$ and $G_{0,1}$ like in Section 2.2, but with $\Sigma$ in (2.6) replaced by $g_1 \Sigma g_1^T$. Then the $S_{0,j}$’s for $j = 0, 1, 2$ are determined by solving (2.8).

\item **Stage 2:** For $t = 1, 2$, we should start by calculating the $W_{t,j}$, by using (3.4):
\[
W_{1,0} = g_1 \Sigma g_1^T + B_{1,1} g_0 \Sigma g_0^T B_{1,1}^T, \quad W_{1,1} = B_{1,1} g_0 \Sigma g_0^T,
\]
\[
W_{2,0} = g_2 \Sigma g_2^T + B_{2,1} g_1 \Sigma g_1^T B_{2,1}^T, \quad W_{2,1} = B_{2,1} g_1 \Sigma g_1^T.
\]
Note that $g_1 = g_0$. Similarly
\[
W_{t,0} = g_t \Sigma g_t^T + B_{t,1} g_{t-1} \Sigma g_{t-1}^T B_{t,1}^T \quad \text{and} \quad W_{t,1} = B_{t,1} g_{t-1} \Sigma g_{t-1}^T
\]
for $t = 3, \ldots, n$. For the covariances $G_{t,k}$ we have by using (3.5):
\[
G_{1,1} = W_{1,1}, \quad G_{2,1} = W_{2,1}, \quad G_{3,1} = W_{3,1},
\]
\[
G_{1,0} = W_{1,0} + G_{1,1} A_{1,1}^T, \quad G_{2,0} = W_{2,0} + G_{2,1} A_{2,1}^T.
\]
Taking care of the covariances $S_{t,j}$ by using the covariances computed before, we have for $t = 1$:
\[
S_{1,2} = A_{1,1} S_{0,1} + A_{1,2} S_{0,0},
\]
\[
S_{1,1} = G_{1,1} + A_{1,1} S_{0,0} + A_{1,2} S_{0,1}^T.
\]
Knowing that by (3.7) $S_{0,-1} = S_{1,1}^T$ and $S_{-1,-2} = S_{1,2}^T$ we have
\[
S_{1,0} = G_{1,0} + A_{1,1} S_{1,1}^T + A_{1,2} S_{1,2}^T.
\]
Then, for $t = 2$:
\[
S_{2,2} = A_{2,1} S_{1,1} + A_{2,2} S_{0,0},
\]
\[
S_{2,1} = G_{2,1} + A_{2,1} S_{1,0} + A_{2,2} S_{1,1}^T,
\]
and, finally knowing that $S_{0,-2} = S_{2,2}^T$ and $S_{1,-1} = S_{2,1}^T$ we have
\[
S_{2,0} = G_{2,0} + A_{2,1} S_{2,1}^T + A_{2,2} S_{2,2}^T.
\]

\item By using the remaining stages of the above algorithm we can evaluate the log-likelihood function of this tdVARMA($2,1$) process.

**Remark 4.4** We can notice that $\Omega$ does not contain some of the computed covariances like $S_{1,1}, S_{1,2}, \ldots$. However they are involved in the calculation of the other covariances.
\end{itemize}
5 Complexity

5.1 Operation count: tdVARMA\((p, q)\) vs VARMA\((p, q)\)

Table 1: Number of multiplications for computing the likelihood of \(r\)-dimensional tdVARMA\((p, q)\) and VARMA\((p, q)\) model, using \(h = \max(p, q)\).

<table>
<thead>
<tr>
<th>(t = 0)</th>
<th>(W_j)</th>
<th>(G_j)</th>
<th>(S_j)</th>
<th>(W_{1,j})</th>
<th>(G_{1,j})</th>
<th>(S_{1,j})</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tdVARMA((p, q))</td>
<td>(r^{3}{q(q+1)/2 + q})</td>
<td>(r^{3}{q(q+1)/2})</td>
<td>(r_{p}^{3})</td>
<td>(nr^{3}{q(q+1)/2 + q})</td>
<td>(r^{3}{p+q}{q+1}/2)</td>
<td>(r_{p}^{3}{p+1}/2)</td>
</tr>
<tr>
<td>VARMA((p, q))</td>
<td>(r^{3}{q(q+1)/2 + q})</td>
<td>(r^{3}{q(q+1)/2})</td>
<td>(r_{p}^{3})</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For \(t \geq 1\)

<table>
<thead>
<tr>
<th>(z) for (t = p+1, \ldots, n)</th>
<th>((n-p)r_{p}^{2})</th>
<th>((n-p)r_{p}^{2})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(L)</td>
<td>((n-h)r^{3}{q^{2}/2 + 7/6} + r^{3}h^{3}/6)</td>
<td>((n-h)r^{3}{q^{2}/2 + 7/6} + r^{3}h^{3}/6)</td>
</tr>
<tr>
<td>(w)</td>
<td>((n-h)r^{2}(p/2 + q) + r^{2}h^{2}/2)</td>
<td>((n-h)r^{2}(p/2 + q) + r^{2}h^{2}/2)</td>
</tr>
</tbody>
</table>

Table 1 gives the total number of multiplications required by our algorithm. If \(p\) is small compared to \(q\) and if \(q\) is small enough with respect to \(n\), the stages 4, 5 and 6 will take the most of the computation time and the total number of operations to compute \(z^{T}\Omega^{-1}z\) is \(O(n)\).

Finally, the total number of multiplications is of order \(n[r^{3}q^{2} + r^{2}(3p/2 + q)] + r^{6}p^{3}/3\) for a tdVARMA\((p, q)\) and \(n[r^{3}q^{2}/2 + r^{2}(3p/2 + q)] + r^{6}p^{3}/3\) in the case of a VARMA\((p, q)\), by neglecting \(r\), \(p\) and \(q\) with respect to \(n\) but not their powers or products. In addition, for the term proportional to \(n\), the total number of multiplications for tdVARMA models is about less than twice what is needed for VARMA models of the same order \((p, q)\) (not counting operations for the computation of the coefficients).

6 Numerical results

In this part we illustrate the use of the algorithm and its implementation in Matlab, called AJM, which represents a generalisation of Jónasson (2008) Matlab’s program to tdVARMA models. The estimation of the parameters by AJM is based on the maximisation of the exact Gaussian likelihood function using the fminunc Matlab function.
belonging to the optimization toolbox, see Coleman, Branch and Grace (1999). This function uses the BFGS quasi-Newton method with a mixed quadratic and cubic line search procedure. AJM can be summarized in three steps. In Step 1 we estimate the mean and the observation covariance matrix which will be used as initial values for $\mu$ and $\Sigma$ in Step 2. Step 2 estimates all the parameters of the model. Finally Step 3 estimates only the parameters of interest in order to derive the covariance matrix of the estimates. This is an alternative to the Tunnicliffe Wilson (1973) method for standard VARMA models.

6.1 Data base and VMA(3) models

The series (Example 8.1 in Tsay, 2005) considered here is a bivariate ($r=2$) time series $x_t = (x_{t,1}, x_{t,2})^T$ where $x_{t,1}$ and $x_{t,2}$ represent respectively the monthly log returns of IBM stock and S&P500 index from January 1926 to December 1999 with $n=888$ observations. Figure 1 shows the time plots of $x_t$ using the same scale.

![Figure 1](image)

Figure 1: Time plots of monthly log returns in percentages for (a) IBM stocks ($x_{t,1}$) and (b) the S&P500 index ($x_{t,2}$) from January 1926 to December 1999.

Tsay (2005) showed, by using the cross-correlation matrices and other techniques
Table 2: Estimation results of VMA(3) models for the IBM and S&P500 series.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Full model</th>
<th>Simplified model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(± Std Error)</td>
<td>(± Std Error)</td>
</tr>
<tr>
<td></td>
<td>(1.24 (±0.24)</td>
<td>0.54 (±0.18)</td>
</tr>
<tr>
<td></td>
<td>(±0.043)</td>
<td>(±0.051)</td>
</tr>
<tr>
<td></td>
<td>-0.121 (±1.01)</td>
<td>0.020 (±0.01)</td>
</tr>
<tr>
<td></td>
<td>-0.038 (±0.44)</td>
<td>0.018 (±0.01)</td>
</tr>
<tr>
<td></td>
<td>0.108 (±0.062)</td>
<td>0.105 (±0.01)</td>
</tr>
<tr>
<td></td>
<td>44.48</td>
<td>23.52</td>
</tr>
<tr>
<td></td>
<td>23.52</td>
<td>31.20</td>
</tr>
<tr>
<td></td>
<td>1.24</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>0.0127</td>
<td>0.040</td>
</tr>
<tr>
<td></td>
<td>0.1299</td>
<td>0.1013</td>
</tr>
<tr>
<td></td>
<td>-0.0198</td>
<td>-0.0036</td>
</tr>
<tr>
<td></td>
<td>2.6567</td>
<td>2.3471</td>
</tr>
<tr>
<td></td>
<td>0.0381</td>
<td>-0.1083</td>
</tr>
<tr>
<td></td>
<td>0.0441</td>
<td>-0.0274</td>
</tr>
<tr>
<td></td>
<td>-0.034</td>
<td>-0.1046</td>
</tr>
<tr>
<td></td>
<td>-0.3666</td>
<td>-2.4561</td>
</tr>
<tr>
<td></td>
<td>44.4775</td>
<td>25.5195</td>
</tr>
<tr>
<td></td>
<td>23.5195</td>
<td>31.1968</td>
</tr>
<tr>
<td></td>
<td>0.54</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.126</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.084</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.082</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.114</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.046</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.374</td>
<td></td>
</tr>
<tr>
<td></td>
<td>44.54</td>
<td></td>
</tr>
<tr>
<td></td>
<td>23.51</td>
<td></td>
</tr>
<tr>
<td></td>
<td>31.21</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.54</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.084</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.084</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.113</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.046</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.374</td>
<td></td>
</tr>
<tr>
<td></td>
<td>44.5412</td>
<td></td>
</tr>
<tr>
<td></td>
<td>23.5103</td>
<td></td>
</tr>
<tr>
<td></td>
<td>31.2102</td>
<td></td>
</tr>
</tbody>
</table>

(selection criteria,...), that a VMA(3) should be adequate. It is given by

\[
\begin{pmatrix}
  x_{t,1} \\
  x_{t,2}
\end{pmatrix}
= \mu + \begin{pmatrix}
  \epsilon_{1,t} \\
  \epsilon_{2,t}
\end{pmatrix} + B_1 \begin{pmatrix}
  \epsilon_{1,t-1} \\
  \epsilon_{2,t-1}
\end{pmatrix} + B_3 \begin{pmatrix}
  \epsilon_{1,t-3} \\
  \epsilon_{2,t-3}
\end{pmatrix},
\]

(6.1)

where \( \mu \) and \( \Sigma \) represent respectively the mean of \( x_t \) and the covariance matrix of \( \epsilon_t \). \( B_1 \) and \( B_3 \) are the coefficients. The estimation results of complete and simplified models of (6.1) are given in Table 2. The simplified models were obtained by removing one by one the non-significant parameters at the 5\% significant level, starting with the least significant, until all the parameters are significantly different from zero. The justification is to reduce the number of parameters to the minimum.

We can see that the estimation results found by our program and those given in Tsay (2005, Table 8.6, p. 370) and obtained by using SCA version 8.1 are very close.

**6.2 Building tdVMA(3) models**

Let us now assume that the bivariate series follows a tdVMA(3) model, a special case of (1.1)-(1.2). Instead of being constant like in (6.1) we assume that the coefficients are slowly varying deterministic functions of time. We use linear functions of time.
except for the scale $g_t$ which is supposed to be exponential to avoid negative values:

$$
\begin{pmatrix}
  x_{t,1} \\
  x_{t,2}
\end{pmatrix}
= \mu + g_t \begin{pmatrix}
  \epsilon_{1,t} \\
  \epsilon_{2,t}
\end{pmatrix}
+ B_{t,1} g_{t-1} \begin{pmatrix}
  \epsilon_{1,t-1} \\
  \epsilon_{2,t-1}
\end{pmatrix}
+ B_{t,3} g_{t-3} \begin{pmatrix}
  \epsilon_{1,t-3} \\
  \epsilon_{2,t-3}
\end{pmatrix},
$$

(6.2)

with

\[
\begin{cases}
  B_{t,1} = B'_1 + (t - \frac{n+1}{2})B''_1, & B'_1 = \begin{pmatrix} B'^{1}_{11} & B'^{1}_{12} \\ B'^{2}_{11} & B'^{2}_{12} \end{pmatrix} \quad \text{and} \quad B''_1 = \begin{pmatrix} B''^{1}_{11} & B''^{1}_{12} \\ B''^{2}_{11} & B''^{2}_{12} \end{pmatrix}; \\
  B_{t,3} = B'_3 + (t - \frac{n+1}{2})B''_3, & B'_3 = \begin{pmatrix} B'^{3}_{11} & B'^{3}_{12} \\ B'^{2}_{11} & B'^{2}_{12} \end{pmatrix} \quad \text{and} \quad B''_3 = \begin{pmatrix} B''^{3}_{11} & B''^{3}_{12} \\ B''^{2}_{11} & B''^{2}_{12} \end{pmatrix}; \\
  g_t = \begin{pmatrix} \exp \{ \eta_1 (t - \frac{n+1}{2}) \} & 0 \\ 0 & \exp \{ \eta_2 (t - \frac{n+1}{2}) \} \end{pmatrix}.
\end{cases}
\]

(6.3)

We have first considered a model without heteroscedasticity i.e. without $g_t$, and then a model with heteroscedasticity i.e. with $g_t$. Table 4 shows the estimation results of the parameters of interest of (6.2)-(6.3) using the program based on the algorithm described in this paper. For example, for the full model without heteroscedasticity 14 iterations were needed and the computation time on a computer with a 3GHz Core Duo E8400 processor was around 385 seconds. The simplified model was obtained again by removing one by one the least non-significant parameters. The theory behind the use of the standard errors and $t$-statistics for tdVARMA models is based on Alj et al. (2014), a generalization of Azrak and Mélard (2006) to multivariate models. Table 3 gives the information criteria of each model, where AICc is the corrected Akaike information criterion (Hurvich and Tsai, 1989), SBC is the Schwarz Bayesian criterion (Rissanen, 1978; Schwarz, 1978), HQC is the Hannan-Quinn criterion (Hannan and Quinn, 1979) and finally FPE represents the final prediction error criterion. Note that

i. Most of the parameters in $B''_1$ and $B''_3$ were significantly different from 0.
ii. It even happens that the final simplified models involve no $B'_1$ and $B'_3$ parameters, showing that the trend in the coefficients is more important than the intercept.
iii. The criteria in Table 3 agree in favour of the simplified model with heteroscedasticity.
iv. Consequently, the suggested kind of models seems useful for at least one pair of series and the proposed algorithm seems justified.

Here we have modelled the time dependence with a linear function. There would be no problem to use polynomials in $t$. For example, using a quadratic function of time would produce an increase of about 50% for the number of parameters in the
Table 3: Specification statistics for the IBM and S&P\textsubscript{500} series. The numbers in bold represent the minimum values in each column.

<table>
<thead>
<tr>
<th>Model</th>
<th>Nbp(^1)</th>
<th>AIC</th>
<th>AICC</th>
<th>FPE</th>
<th>HQC</th>
<th>SBC</th>
</tr>
</thead>
<tbody>
<tr>
<td>VMA(3) Full Model</td>
<td>8</td>
<td>6.7447</td>
<td>6.7448</td>
<td>849.5600</td>
<td>6.7612</td>
<td>6.7879</td>
</tr>
<tr>
<td>Simplified Model</td>
<td>4</td>
<td>6.7393</td>
<td>6.7393</td>
<td>844.9808</td>
<td>6.7476</td>
<td>6.7609</td>
</tr>
<tr>
<td>tdVMA(3) Full Model(^2)</td>
<td>16</td>
<td>6.7599</td>
<td>6.7600</td>
<td>862.5443</td>
<td>6.7764</td>
<td>6.8030</td>
</tr>
<tr>
<td>Simplified Model(^2)</td>
<td>2</td>
<td>6.7493</td>
<td>6.7494</td>
<td>853.5014</td>
<td>6.7535</td>
<td>6.7601</td>
</tr>
<tr>
<td>Full Model(^3)</td>
<td>18</td>
<td>6.6082</td>
<td>6.6086</td>
<td>741.1335</td>
<td>6.6453</td>
<td>6.7053</td>
</tr>
<tr>
<td>Simplified Model(^3)</td>
<td>5</td>
<td>6.5995</td>
<td>6.5996</td>
<td>734.7519</td>
<td>6.6140</td>
<td>6.6373</td>
</tr>
</tbody>
</table>

1 Nbp: the number of parameters.
2 Without heteroscedasticity.
3 With heteroscedasticity.
Table 4: Estimation results by AJM of tdVMA(3) models for the IBM and S&P500 series.

Full model without heteroscedasticity

<table>
<thead>
<tr>
<th>Parameters</th>
<th>$\mu$</th>
<th>$B_1^\tau \times 10^7$</th>
<th>$B_2^\tau \times 10^3$</th>
<th>$B_3^\tau \times 10^9$</th>
<th>$B_4^\tau \times 10^7$</th>
<th>$\eta \times 10^3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
<td>1.24</td>
<td>0.0077 ± 0.0002</td>
<td>0.2015 ± 0.0013</td>
<td>0.0025 ± 0.0001</td>
<td>0.2868 ± 0.0002</td>
<td>- -</td>
</tr>
<tr>
<td>(± Std Error)</td>
<td>0.53</td>
<td>±42.919 ± 0.0002</td>
<td>±0.1360 ± 0.0013</td>
<td>±0.0013 ± 0.0001</td>
<td>±0.1990 ± 0.0002</td>
<td>±0.0380 ± 0.0013</td>
</tr>
<tr>
<td>t-statistic</td>
<td></td>
<td>±59.571 ± 0.0002</td>
<td>±0.1990 ± 0.0013</td>
<td>±0.0013 ± 0.0002</td>
<td>±0.1990 ± 0.0002</td>
<td>±0.0380 ± 0.0013</td>
</tr>
</tbody>
</table>

Simplified model without heteroscedasticity

<table>
<thead>
<tr>
<th>Parameters</th>
<th>$\mu$</th>
<th>$B_1^\tau \times 10^7$</th>
<th>$B_2^\tau \times 10^3$</th>
<th>$B_3^\tau \times 10^9$</th>
<th>$B_4^\tau \times 10^7$</th>
<th>$\eta \times 10^3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
<td>1.24</td>
<td>0 0</td>
<td>0 -0.3150 ± 0.1341</td>
<td>0 0</td>
<td>0.1941 ± 0.0942</td>
<td>- -</td>
</tr>
<tr>
<td>(± Std Error)</td>
<td>0.54</td>
<td>(±0.0081)</td>
<td>(±0.0000)</td>
<td>(±0.0000)</td>
<td>(±0.0000)</td>
<td>(±0.0000)</td>
</tr>
<tr>
<td>t-statistic</td>
<td></td>
<td>(±0.5435)</td>
<td>(±0.394)</td>
<td>(±0.1057)</td>
<td>(±0.3807)</td>
<td>(±0.0996)</td>
</tr>
</tbody>
</table>

Full model with heteroscedasticity

<table>
<thead>
<tr>
<th>Parameters</th>
<th>$\mu$</th>
<th>$B_1^\tau \times 10^7$</th>
<th>$B_2^\tau \times 10^3$</th>
<th>$B_3^\tau \times 10^9$</th>
<th>$B_4^\tau \times 10^7$</th>
<th>$\eta \times 10^3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
<td>1.24</td>
<td>49.370 ± 1.1779</td>
<td>0.1685 ± 0.0081</td>
<td>-8.9076 ± 1.2826</td>
<td>0.3232 ± 0.1688</td>
<td>- -</td>
</tr>
<tr>
<td>(± Std Error)</td>
<td>0.54</td>
<td>(±49.915)</td>
<td>(±0.1329)</td>
<td>(±1.2682)</td>
<td>(±0.1532)</td>
<td>±0.7264 ± 0.1540</td>
</tr>
<tr>
<td>t-statistic</td>
<td></td>
<td>±50.414 ± 0.8497</td>
<td>±0.1387 ± 0.0084</td>
<td>±1.6555 ± 0.1342</td>
<td>±0.7727 ± 0.1342</td>
<td>±0.7751 ± 0.0794</td>
</tr>
</tbody>
</table>

Simplified model with heteroscedasticity

<table>
<thead>
<tr>
<th>Parameters</th>
<th>$\mu$</th>
<th>$B_1^\tau \times 10^7$</th>
<th>$B_2^\tau \times 10^3$</th>
<th>$B_3^\tau \times 10^9$</th>
<th>$B_4^\tau \times 10^7$</th>
<th>$\eta \times 10^3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
<td>1.24</td>
<td>0 0</td>
<td>0.1746 ± 0.1032</td>
<td>0 0</td>
<td>0.2081 ± 0.0953</td>
<td>- -</td>
</tr>
<tr>
<td>(± Std Error)</td>
<td>0.54</td>
<td>(±0.0074)</td>
<td>(±0.1345)</td>
<td>(±0.3084)</td>
<td>(±0.2417)</td>
<td>±0.7237 ± 0.1044</td>
</tr>
<tr>
<td>t-statistic</td>
<td></td>
<td>(±0.0007)</td>
<td>(±0.3179)</td>
<td>(±0.0007)</td>
<td>(±0.0007)</td>
<td>±0.7237 ± 0.0794</td>
</tr>
</tbody>
</table>

18
Acknowledgements

The authors would like to thank the referees, for their careful reading of the manuscript, their comments and their pertinent suggestions, and also the associate editor and the co-editor Erricos Kontoghiorghes.

References


