Computation of the Exact Fisher Information Matrix of a Multiple Input Single Output Time Series Models

Emilie EPEKA MBAMBE, Angèle YULE SOTAZO, Jacques SABITI KISETA*
Bel Campus Technological University of Kinshasa, Democratic Republic of Congo
*Corresponding Author: jacquessabiti38@gmail.com

Doi: https://doi.org/10.47285/isr.v2i2.93

Citation: MBAMBE, E.E., SOTAZO, A.Y., & KISETA, J.S. (2021). Computation of the Exact Fisher Information Matrix of a Multiple Input Single Output Time Series Models, International Science Review, 2(1), 1-11. DOI: https://doi.org/10.47285/isr.v2i2.93

Research Article

Abstract

Klein, Mélard, and Zahaf (1998) have proposed the computation of the exact Fisher information matrix of a large class of Gaussian time series models called the single-input-single-output (SISO) model, includes dynamic regression with autocorrelated errors and the transfer function model, with autoregressive moving average errors. For computing the Fisher information matrix of a SISO model, they introduced an algorithm based on a combination of two computational procedures: recursions for the covariance matrix of the derivatives of the state vector with respect to the parameters and the fast Kalman filter recursions used in the evaluation of the likelihood function. In this paper, we propose a generalization of this method for computing the Fisher information matrix of a MISO model.

Keywords: Fisher information matrix, MISO models, Chandrasekhar equations, Kalman filter

1. Introduction

We consider a MISO model given by the equation

\[
\frac{\alpha(B)}{\lambda(B)} y(t) = \sum_{i=1}^{k} \frac{\omega_i(B)}{\delta_i(B)} x_i(t-b_i) + \frac{\theta(B)}{\phi(B)} \varepsilon(t)
\]

where \( y(t) \) is the endogenous variable, \( x_i(t) \) \( (i = 1, ..., k) \) are the exogenous variables, \( B \) is the backshift operator such that \( B^i y(t) = y(t-j) \), \( \varepsilon(t) \) are normally and independently random variables with mean zero and constant variance \( \sigma^2_\varepsilon \), \( b_i \) is the delay of transmission of influence between the \( i \) exogenous variable and the endogenous variable, or the delay parameter which represents the number of complete-time intervals before a change in \( x_i(t) \) begins to have an effect on \( y(t) \) and \( \beta = (\alpha^T, \lambda^T, \omega^T_1, ..., \omega^T_k, \delta^T_1, ..., \delta^T_k, \phi^T, \theta^T) \) is the vector of parameters where :
\[
\alpha = (\alpha_1, ..., \alpha_z)^T, \quad \lambda = (\lambda_1, ..., \lambda_z)^T,
\omega_i = (\omega_{i0}, \omega_{i1}, ..., \omega_{ik})^T \quad (2)\]
\[
\delta_i = (\delta_{i1}, ..., \delta_{in})^T, \quad \phi = (\phi_1, ..., \phi_p)^T, \quad \theta = (\theta_1, ..., \theta_q)^T \quad (2b)
\]

and \( l = \bar{s} + \bar{r} + s_1 + ... + s_k + r_1 + ... + r_k + p + q \) is the number of parameters. The different polynomials in (1) are given by
\[
\alpha(B) = 1 - \sum_{j=1}^{r} \alpha_j B^j, \quad \lambda(B) = 1 - \sum_{j=1}^{r} \lambda_j B^j, \quad \alpha_0 = \lambda_0 = 1 \quad (3a)
\]
\[
\phi(B) = 1 - \sum_{j=1}^{p} \phi_j B^j, \quad \theta(B) = 1 - \sum_{j=1}^{q} \theta_j B^j, \quad \phi_0 = \theta_0 = 1 \quad (3b)
\]
\[
\omega_i(B) = \omega_{i0} + \sum_{j=1}^{k} \omega_{ij} B^j, \quad \delta_i(B) = 1 - \sum_{j=1}^{n} \delta_{ij} B^j, \quad \delta_{i0} = 1, \quad i = 1, ..., k. \quad (3c)
\]

The assumptions (1, 2a, and 2b) made on these six polynomials are the regularity conditions to ensure the stability of the model (1).

2. The likelihood function and partial derivatives

We consider the conditional likelihood function since it is simple and asymptotically, provides similar results as those of the exact maximum likelihood (EML) method. If in (1), the \( \varepsilon(t) \) are independent and identically normally distributed with mean zero and variance \( \sigma^2_\varepsilon \), the conditional log-likelihood function of the parameters \( (\beta, \sigma^2_\varepsilon) \) can be written under the form
\[
\log L = \frac{-N}{2} \log(2\pi) - \frac{N}{2} \log(\sigma^2_\varepsilon) - \frac{1}{2\sigma^2_\varepsilon} \sum_{t=1}^{N} \varepsilon^2(t) \quad (4)
\]

where the error term \( \varepsilon(t) \) is given by
\[
\varepsilon(t) = \frac{\theta(B)}{\phi(B)} \left[ \alpha(B) y(t) - \sum_{i=1}^{k} \omega_i(B) x_i(t-b_i) \right]. \quad (5)
\]

Assuming that the \( \varepsilon(t) \) are normally distributed, the maximization of the log-likelihood function (4) yields the same estimators of \( \beta \) which minimize the sum of squares \( S(\beta) = \sum_{t=1}^{N} \varepsilon^2(t) \) as a function of \( \beta \) and an estimator \( \hat{\sigma}^2_\varepsilon = (1/N) \sum_{t=1}^{N} \hat{\varepsilon}^2(t) \) \( \sigma^2_\varepsilon \) of. Since \( \varepsilon(t) \) it is not a linear function \( \beta \), these estimators are computed in practice by using the optimization algorithms (Newton-Raphson or Levenberg-Marquardt) for obtaining, for \( N \) enough large, the estimators which will be close to those provided by maximizing an EML function.

We give the partial derivatives of the log-likelihood function for all parameters. To evaluate these partial derivatives with respect to all parameters, we will use the following results
\[
\frac{\partial S(\beta)}{\partial \beta} = 2 \sum_{t=1}^{N} \varepsilon(t) \frac{\partial \varepsilon(t)}{\partial \beta}, \quad \frac{\partial^2 S(\beta)}{\partial \beta^2} = 2 \left( \frac{\partial \varepsilon(t)}{\partial \beta} \right) \left( \frac{\partial \varepsilon(t)}{\partial \beta} \right) + \varepsilon(t) \frac{\partial^2 \varepsilon(t)}{\partial \beta^2} \quad (6)
\]

where \( \beta^T \) is the transpose of \( \beta \) and
\[ S(\beta) = \sum_{t=1}^{N} \varepsilon^2(t). \]  

From (6), we can write the log-likelihood function as
\[ \log L = -\frac{N}{2} \log(2\pi) - \frac{N}{2} \log(\sigma^2_e) - \frac{1}{2\sigma^2_e} S(\beta) \]  

where the partial derivatives of (7b) concerning the vector \( \beta \) are given respectively by
\[ \frac{\partial \log L}{\partial \beta} = \frac{1}{2\sigma^2_e} \frac{\partial S(\beta)}{\partial \beta} = -\frac{1}{2\sigma^2_e} \sum_{t=1}^{N} \varepsilon(t) \frac{\partial \varepsilon(t)}{\partial \beta} \]  

\[ \frac{\partial^2 \log L}{\partial \beta \partial \beta^T} = -\frac{1}{\sigma^2_e} \frac{\partial^2 S(\beta)}{\partial \beta \partial \beta^T} = -\frac{1}{\sigma^2_e} \sum_{t=1}^{N} \left( \frac{\partial \varepsilon(t)}{\partial \beta} \frac{\partial \varepsilon(t)}{\partial \beta^T} + \varepsilon(t) \frac{\partial^2 \varepsilon(t)}{\partial \beta \partial \beta^T} \right). \]  

We see that the obtaining of the partial derivatives of the log-likelihood function for all parameters, i.e. \( \alpha_j \), \( \lambda_j \), \( \omega_j \), \( \delta_j \), \( \omega_j \), \( \phi_j \) and \( \theta_j \) needs the knowledge of the first and second derivatives of \( \varepsilon(t) \) all parameters.

According to Klein and Mélard (1994 a-b), these derivatives are respectively given by
\[ \frac{\partial \varepsilon(t)}{\partial \alpha_j} = -\frac{\phi(B)}{\lambda(B)\theta(B)} y(t-j), \quad \frac{\partial \varepsilon(t)}{\partial \lambda_j} = -\frac{\phi(B)\alpha(B)}{\lambda^2(B)\theta(B)} y(t-j) \]  

\[ \frac{\partial \varepsilon(t)}{\partial \omega_j} = -\frac{\phi(B)}{\delta_i(B)\theta(B)} x_i(t-b_i-j), \quad \frac{\partial \varepsilon(t)}{\partial \delta_j} = -\frac{\omega_i(B)\phi(B)}{\delta_i^2(B)\theta(B)} x_i(t-b_i-j) \]  

\[ \frac{\partial \varepsilon(t)}{\partial \phi_j} = -\frac{1}{\phi(B)} \varepsilon(t-j), \quad \frac{\partial \varepsilon(t)}{\partial \theta_j} = -\frac{1}{\theta(B)} \varepsilon(t-j) \]  

where the backshift operator plays only at its right hand side.

Following (8a), the first-order partial derivatives of the log-likelihood function with respect to \( \alpha_j \), \( \lambda_j \), \( \omega_j \), \( \delta_j \), \( \omega_j \), \( \phi_j \) and \( \theta_j \) are respectively given by
\[ \frac{\partial \log L}{\partial \alpha_j} = \frac{1}{\sigma^2_e} \sum_{t=1}^{N} \varepsilon(t) \left[ \frac{\phi(B)}{\lambda(B)\theta(B)} y(t-j) \right] \]  

\[ \frac{\partial \log L}{\partial \lambda_j} = -\frac{1}{\sigma^2_e} \sum_{t=1}^{N} \varepsilon(t) \left[ \frac{\phi(B)\alpha(B)}{\lambda^2(B)\theta(B)} y(t-j) \right] \]  

\[ \frac{\partial \log L}{\partial \omega_j} = \frac{1}{\sigma^2_e} \sum_{t=1}^{N} \varepsilon(t) \left[ \frac{\phi(B)}{\delta_i(B)\theta(B)} x_i(t-b_i-j) \right], \quad i=1,\ldots,k \]  

\[ \frac{\partial \log L}{\partial \delta_j} = -\frac{1}{\sigma^2_e} \sum_{t=1}^{N} \varepsilon(t) \left[ \frac{\omega_i(B)\phi(B)}{\delta_i^2(B)\theta(B)} x_i(t-b_i-j) \right], \quad i=1,\ldots,k \]
\[ \frac{\partial \log L}{\partial \phi_j} = \frac{1}{\sigma_x^2} \sum_{t=1}^{N} \varepsilon(t) \left[ \frac{\varepsilon(t-j)}{\phi(B)} \right] \]  
(10e)

\[ \frac{\partial \log L}{\partial \theta_j} = -\frac{1}{\sigma_x^2} \sum_{t=1}^{N} \varepsilon(t) \left[ \frac{\varepsilon(t-j)}{\theta(B)} \right]. \]  
(10f)

We can also write these first-order partial derivatives as

\[ \frac{\partial \log L}{\partial \alpha_j} = \frac{1}{\sigma_x^2} \sum_{t=1}^{N} \varepsilon(t) \xi(t-j), \quad \frac{\partial \log L}{\partial \lambda_j} = -\frac{1}{\sigma_x^2} \sum_{t=1}^{N} \varepsilon(t) \zeta(t-j) \]  
(11a)

\[ \frac{\partial \log L}{\partial \omega_{ij}} = \frac{1}{\sigma_x^2} \sum_{t=1}^{N} \varepsilon(t) \eta_{ij}(t-j), \quad \frac{\partial \log L}{\partial \delta_{ij}} = \frac{1}{\sigma_x^2} \sum_{t=1}^{N} \varepsilon(t) \nu_{ij}(t-j) \]  
(11b)

\[ \frac{\partial \log L}{\partial \phi_j} = \frac{1}{\sigma_x^2} \sum_{t=1}^{N} \varepsilon(t) \xi(t-j), \quad \frac{\partial \log L}{\partial \theta_j} = -\frac{1}{\sigma_x^2} \sum_{t=1}^{N} \varepsilon(t) \tau(t-j) \]  
(11c)

where

\[ \xi(t) = \frac{\phi(B)}{\lambda(B) \theta(B)} y(t), \quad \zeta(t) = -\frac{\phi(B) \alpha(B)}{\lambda^2(B) \theta(B)} y(t) \]  
(12a)

\[ \eta_{ij}(t) = \frac{\phi(B)}{\delta_{ij}(B) \theta(B)} x_{ij}(t-j), \quad \nu_{ij}(t) = \frac{\omega_{i}(B) \phi(B)}{\delta_{ij}^2(B) \theta(B)} x_{ij}(t-j) \]  
(12b)

\[ \xi(t) = \frac{\varepsilon(t)}{\phi(B)}, \quad \tau(t) = -\frac{\varepsilon(t)}{\theta(B)}. \]  
(12c)

The partial derivatives with respect to \( \sigma_x^2 \) are respectively given by

\[ \frac{\partial \log L}{\partial \sigma_x^2} = -\frac{N}{2\sigma_x^2} + \frac{1}{2\sigma_x^4} \sum_{t=1}^{N} (\varepsilon(t))^2, \quad \frac{\partial \log L}{\partial \sigma_x^2} = \frac{N}{2\sigma_x^4} + \frac{1}{\sigma_x^2} \sum_{t=1}^{N} (\varepsilon(t))^2. \]  
(13)

Since the first partial derivatives \( \sigma_x^2 \) must be equal to zero, i.e. \( \frac{\partial \log L}{\partial \sigma_x^2} = 0 \) has a solution \( \sigma_x^2 = S(\beta) / N \) that suggests that \( \hat{\sigma}_x^2 = S(\beta) / N \) is the maximum likelihood estimator.

Following (8b), the different partial second derivatives are respectively given by

\[ \frac{\partial^2 \log L}{\partial \alpha_j \partial \alpha_i} = -\frac{1}{\sigma_x^2} \sum_{t=1}^{N} \left( \frac{\partial \varepsilon(t)}{\partial \alpha_j} \frac{\partial \varepsilon(t)}{\partial \alpha_i} + \varepsilon(t) \frac{\partial^2 \varepsilon(t)}{\partial \alpha_j \partial \alpha_i} \right) \]  
(14a)

\[ \frac{\partial^2 \log L}{\partial \lambda_j \partial \lambda_i} = -\frac{1}{\sigma_x^2} \sum_{t=1}^{N} \left( \frac{\phi(B)}{\lambda(B) \theta(B)} y(t-j) \left[ \frac{\phi(B)}{\lambda(B) \theta(B)} y(t-l) \right] + \varepsilon(t) \frac{\partial^2 \varepsilon(t)}{\partial \lambda_j \partial \lambda_i} \right) \]
\[
\frac{\partial^2 \log L}{\partial \omega_y \partial \omega_d} = -\frac{1}{\sigma^2} \sum_{i=1}^{N} \left(\frac{\partial \varepsilon(t)}{\partial \omega_y} \frac{\partial \varepsilon(t)}{\partial \omega_d} + \varepsilon(t) \frac{\partial^2 \varepsilon(t)}{\partial \omega_y \partial \omega_d}\right)
\]

\[
\frac{\partial^2 \log L}{\partial \phi_j \partial \phi_i} = -\frac{1}{\sigma^2} \sum_{i=1}^{N} \left(\frac{\partial \varepsilon(t)}{\partial \phi_{j'}} \frac{\partial \varepsilon(t)}{\partial \phi_i} + \varepsilon(t) \frac{\partial^2 \varepsilon(t)}{\partial \phi_{j'} \partial \phi_i}\right)
\]

\[
\frac{\partial^2 \log L}{\partial \theta_j \partial \theta_i} = -\frac{1}{\sigma^2} \sum_{i=1}^{N} \left(\frac{\partial \varepsilon(t)}{\partial \theta_{j'}} \frac{\partial \varepsilon(t)}{\partial \theta_i} + \varepsilon(t) \frac{\partial^2 \varepsilon(t)}{\partial \theta_{j'} \partial \theta_i}\right)
\]

The partial derivatives with respect to $\alpha_j$ and $\lambda_i$ are respectively given by

\[
\frac{\partial^2 \log L}{\partial \alpha_j \partial \lambda_i} = -\frac{1}{\sigma^2} \sum_{i=1}^{N} \left(\frac{\partial \varepsilon(t)}{\partial \alpha_j} \frac{\partial \varepsilon(t)}{\partial \lambda_i} + \varepsilon(t) \frac{\partial^2 \varepsilon(t)}{\partial \alpha_j \partial \lambda_i}\right)
\]

The partial derivatives with respect to $\alpha_j$ and $\omega_{i\ell}$ are respectively given by

\[
\frac{\partial^2 \log L}{\partial \alpha_j \partial \omega_{i\ell}} = -\frac{1}{\sigma^2} \sum_{i=1}^{N} \left(\frac{\partial \varepsilon(t)}{\partial \alpha_j} \frac{\partial \varepsilon(t)}{\partial \omega_{i\ell}} + \varepsilon(t) \frac{\partial^2 \varepsilon(t)}{\partial \alpha_j \partial \omega_{i\ell}}\right)
\]
\[
= -\frac{1}{\sigma^2} \sum_{i=1}^{N} \left[ \frac{\phi(B)}{\lambda(B)\theta(B)} y(t - j) \left[ \frac{\phi(B)}{\delta_i(B)\theta(B)} x_i(t - b_i - 1) \right] + \varepsilon(t) \frac{\partial^2 \varepsilon(t)}{\partial \alpha_j \partial \omega_i} \right]. \tag{14h}
\]

The partial derivatives with respect to \( \alpha_j \) and \( \delta_i \) are respectively given by
\[
\frac{\partial^2 \log L}{\partial \alpha_j \partial \delta_i} = -\frac{1}{\sigma^2} \sum_{i=1}^{N} \left( \frac{\partial \varepsilon(t)}{\partial \alpha_j} \frac{\partial \varepsilon(t)}{\partial \delta_i} + \varepsilon(t) \frac{\partial^2 \varepsilon(t)}{\partial \alpha_j \partial \delta_i} \right) \]
\[
= -\frac{1}{\sigma^2} \sum_{i=1}^{N} \left[ \frac{\phi(B)}{\lambda(B)\theta(B)} y(t - j) \left[ \frac{\partial \varepsilon(t)}{\partial \alpha_j} \frac{\partial \varepsilon(t)}{\partial \delta_i} + \varepsilon(t) \frac{\partial^2 \varepsilon(t)}{\partial \alpha_j \partial \delta_i} \right] \right]. \tag{14i}
\]

The partial derivatives with respect to \( \alpha_j \) and \( \phi_i \) are respectively given by
\[
\frac{\partial^2 \log L}{\partial \alpha_j \partial \phi_i} = -\frac{1}{\sigma^2} \sum_{i=1}^{N} \left( \frac{\partial \varepsilon(t)}{\partial \alpha_j} \frac{\partial \varepsilon(t)}{\partial \phi_i} + \varepsilon(t) \frac{\partial^2 \varepsilon(t)}{\partial \alpha_j \partial \phi_i} \right) \]
\[
= -\frac{1}{\sigma^2} \sum_{i=1}^{N} \left[ \frac{\phi(B)}{\lambda(B)\theta(B)} y(t - j) \left[ \frac{\partial \varepsilon(t)}{\partial \alpha_j} \frac{\partial \varepsilon(t)}{\partial \phi_i} + \varepsilon(t) \frac{\partial^2 \varepsilon(t)}{\partial \alpha_j \partial \phi_i} \right] \right]. \tag{14j}
\]

The partial derivatives with respect to \( \alpha_j \) and \( \theta_i \) are respectively given by
\[
\frac{\partial^2 \log L}{\partial \alpha_j \partial \theta_i} = -\frac{1}{\sigma^2} \sum_{i=1}^{N} \left( \frac{\partial \varepsilon(t)}{\partial \alpha_j} \frac{\partial \varepsilon(t)}{\partial \theta_i} + \varepsilon(t) \frac{\partial^2 \varepsilon(t)}{\partial \alpha_j \partial \theta_i} \right) \]
\[
= \frac{1}{\sigma^2} \sum_{i=1}^{N} \left[ \frac{\phi(B)}{\lambda(B)\theta(B)} y(t - j) \left[ \frac{\partial \varepsilon(t)}{\partial \alpha_j} \frac{\partial \varepsilon(t)}{\partial \theta_i} + \varepsilon(t) \frac{\partial^2 \varepsilon(t)}{\partial \alpha_j \partial \theta_i} \right] \right]. \tag{14k}
\]

The partial derivatives with respect to \( \lambda_j \) and \( \omega_i \) are respectively given by
\[
\frac{\partial^2 \log L}{\partial \lambda_j \partial \omega_i} = -\frac{1}{\sigma^2} \sum_{i=1}^{N} \left( \frac{\partial \varepsilon(t)}{\partial \lambda_j} \frac{\partial \varepsilon(t)}{\partial \omega_i} + \varepsilon(t) \frac{\partial^2 \varepsilon(t)}{\partial \lambda_j \partial \omega_i} \right) \]
\[
= \frac{1}{\sigma^2} \sum_{i=1}^{N} \left[ \frac{\phi(B)}{\lambda^2(B)\theta(B)} y(t - j) \left[ \frac{\partial \varepsilon(t)}{\partial \lambda_j} \frac{\partial \varepsilon(t)}{\partial \omega_i} + \varepsilon(t) \frac{\partial^2 \varepsilon(t)}{\partial \lambda_j \partial \omega_i} \right] \right]. \tag{14l}
\]

The partial derivatives with respect to \( \lambda_j \) and \( \delta_i \) are respectively given by
\[
\frac{\partial^2 \log L}{\partial \lambda_j \partial \delta_i} = -\frac{1}{\sigma^2} \sum_{i=1}^{N} \left( \frac{\partial \varepsilon(t)}{\partial \lambda_j} \frac{\partial \varepsilon(t)}{\partial \delta_i} + \varepsilon(t) \frac{\partial^2 \varepsilon(t)}{\partial \lambda_j \partial \delta_i} \right) \]
\[
\frac{\partial^2 \log L}{\partial \lambda_j \partial \phi_i} = -\frac{1}{\sigma^2_n} \sum_{i=1}^N \left( \frac{\partial \varepsilon(t)}{\partial \lambda_j} \frac{\partial \varepsilon(t)}{\partial \phi_i} + \varepsilon(t) \frac{\partial^2 \varepsilon(t)}{\partial \lambda_j \partial \phi_i} \right)
\]

The partial derivatives with respect to \( \lambda_j \) and \( \phi_i \) are respectively given by

\[
\frac{\partial^2 \log L}{\partial \lambda_j \partial \phi_i} = \frac{1}{\sigma^2_n} \sum_{i=1}^N \left( \frac{\phi(B)\alpha(B)}{\lambda^2(B)} y(t-j) \left[ \frac{\varepsilon(t-l)}{\phi(B)} \right] - \varepsilon(t) \frac{\partial^2 \varepsilon(t)}{\partial \lambda_j \partial \phi_i} \right).
\]  \( (14n) \)

The partial derivatives with respect to \( \lambda_j \) and \( \theta_i \) are respectively given by

\[
\frac{\partial^2 \log L}{\partial \lambda_j \partial \theta_i} = -\frac{1}{\sigma^2_n} \sum_{i=1}^N \left( \frac{\partial \varepsilon(t)}{\partial \lambda_j} \frac{\partial \varepsilon(t)}{\partial \theta_i} + \varepsilon(t) \frac{\partial^2 \varepsilon(t)}{\partial \lambda_j \partial \theta_i} \right)
\]

The partial derivatives with respect to \( \omega_j \) and \( \delta_i \) are respectively given by

\[
\frac{\partial^2 \log L}{\partial \omega_j \partial \theta_i} = \frac{1}{\sigma^2_n} \sum_{i=1}^N \left( \frac{\phi(B)}{\delta_i(B)\theta(B)} x_i(t-b_i-j) \left[ \frac{\omega_i(B)\phi(B)}{\delta^2_i(B)\theta(B)} x_i(t-b_i-l) \right] + \varepsilon(t) \frac{\partial^2 \varepsilon(t)}{\partial \omega_j \partial \theta_i} \right).
\]  \( (14p) \)

The partial derivatives with respect to \( \omega_j \) and \( \phi_i \) are respectively given by

\[
\frac{\partial^2 \log L}{\partial \omega_j \partial \phi_i} = \frac{1}{\sigma^2_n} \sum_{i=1}^N \left( \frac{\phi(B)}{\delta_i(B)\theta(B)} x_i(t-b_i-j) \left[ \frac{\varepsilon(t-l)}{\phi(B)} \right] - \varepsilon(t) \frac{\partial^2 \varepsilon(t)}{\partial \omega_j \partial \phi_i} \right).
\]  \( (14q) \)

The partial derivatives with respect to \( \omega_j \) and \( \phi_i \) are respectively given by

\[
\frac{\partial^2 \log L}{\partial \omega_j \partial \phi_i} = \frac{1}{\sigma^2_n} \sum_{i=1}^N \left( \frac{\partial \varepsilon(t)}{\partial \omega_j} \frac{\partial \varepsilon(t)}{\partial \phi_i} + \varepsilon(t) \frac{\partial^2 \varepsilon(t)}{\partial \omega_j \partial \phi_i} \right)
\]

The partial derivatives with respect to \( \omega_j \) and \( \phi_i \) are respectively given by

\[
\frac{\partial^2 \log L}{\partial \omega_j \partial \phi_i} = \frac{1}{\sigma^2_n} \sum_{i=1}^N \left( \frac{\phi(B)}{\delta_i(B)\theta(B)} x_i(t-b_i-j) \left[ \frac{\varepsilon(t-l)}{\phi(B)} \right] - \varepsilon(t) \frac{\partial^2 \varepsilon(t)}{\partial \omega_j \partial \phi_i} \right).
\]  \( (14r) \)

The partial derivatives with respect to \( \delta_j \) and \( \phi_i \) are respectively given by
\[
\frac{\partial^2 \log L}{\partial \delta_j \partial \phi_l} = -\frac{1}{\sigma^2} \sum_{i=1}^{N} \left( \frac{\partial \varepsilon(t)}{\partial \delta_j} \frac{\partial \varepsilon(t)}{\partial \phi_l} + \varepsilon(t) \frac{\partial^2 \varepsilon(t)}{\partial \delta_j \partial \phi_l} \right) + \varepsilon(t) \frac{\partial^2 \varepsilon(t)}{\partial \delta_j ^2} .
\]

The partial derivatives with respect to \( \delta_j \) and \( \theta_i \) are respectively given by

\[
\frac{\partial^2 \log L}{\partial \delta_j \partial \theta_i} = -\frac{1}{\sigma^2} \sum_{i=1}^{N} \left( \frac{\partial \varepsilon(t)}{\partial \delta_j} \frac{\partial \varepsilon(t)}{\partial \theta_i} + \varepsilon(t) \frac{\partial^2 \varepsilon(t)}{\partial \delta_j \partial \theta_i} \right) + \varepsilon(t) \frac{\partial^2 \varepsilon(t)}{\partial \delta_j \partial \theta_i} ,
\]

and

\[
\frac{\partial^2 \log L}{\partial \phi_j \partial \theta_i} = -\frac{1}{\sigma^2} \sum_{i=1}^{N} \left( \frac{\partial \varepsilon(t)}{\partial \phi_j} \frac{\partial \varepsilon(t)}{\partial \theta_i} + \varepsilon(t) \frac{\partial^2 \varepsilon(t)}{\partial \phi_j \partial \theta_i} \right) + \varepsilon(t) \frac{\partial^2 \varepsilon(t)}{\partial \phi_j \partial \theta_i} .
\]

3. Computation of exact Fisher information matrix

The problem of computing the Fisher information matrix has been treated by several authors. Porat and Friedlander (1986, 1987) have proposed an algorithm for the exact computation of the Fisher information matrix based on the Durbin-Levinson algorithm for computing the orthogonal polynomials of a Toepliz matrix. For the MISO model (1), the Fisher information matrix is given by

\[
\Omega = -E \left( \frac{\partial^2 \log L}{\partial \beta \partial \beta^T} \right)
\]

evaluated at the true unknown value of \( \beta \) is useful for obtaining the Cramer-Rao (lower) bound (CRB) of the estimated parameter vector \( \hat{\beta} \) such that \( \text{cov}(\hat{\beta}) \geq \Omega^{-1} \). A good estimator of its asymptotic covariance matrix is \( \Omega^{-1} \) assuming that the estimation method yields asymptotically efficient estimators.

The CRB is obtained from the Fisher information matrix, given the finite sample size \( N \), the computation of the Fisher information matrix includes matrix inversions and matrix multiplications, and the computation of the derivatives. From the results given by (14a-u), the symmetric matrix of second derivatives is given by
\[
\Omega = \begin{pmatrix}
\frac{\partial^2 \log L}{\partial \alpha_i \partial \alpha_j} & \frac{\partial^2 \log L}{\partial \alpha_i \partial \lambda_i} & \frac{\partial^2 \log L}{\partial \alpha_i \partial \omega_i} \\
\frac{\partial^2 \log L}{\partial \lambda_i \partial \alpha_i} & \frac{\partial^2 \log L}{\partial \lambda_i \partial \lambda_j} & \frac{\partial^2 \log L}{\partial \lambda_i \partial \omega_i} \\
\frac{\partial^2 \log L}{\partial \omega_i \partial \alpha_i} & \frac{\partial^2 \log L}{\partial \omega_i \partial \lambda_j} & \frac{\partial^2 \log L}{\partial \omega_i \partial \omega_i} \\
\end{pmatrix}.
\quad (16)
\]

To derive the Fisher information matrix, we write the MISO model (1) as
\[
\alpha(B)\delta_j(B)y(t) = \sum_{i=1}^{k} \omega_i(B)\lambda(B)\phi_i(t - b_i) + \theta(B)\lambda(B)\delta_j(B)e(t).
\quad (17)
\]

We denote \( \mu(t) \) by
\[
\mu(t) = \sum_{i=1}^{k} \frac{\omega_i(B)\lambda(B)}{\delta_j(B)\alpha(B)} x_i(t - b_i)
\quad (18a)
\]

and \( w(t) \) by
\[
w(t) = y(t) - \mu(t) = \frac{\lambda(B)\theta(B)}{\alpha(B)\phi(B)} e(t).
\quad (18b)
\]

We first suppose that data are available for \( y(t) x_i(t) \) and for \( t = 1, \ldots, N \). Therefore, \( \mu(t) \) can be computed for \( t \geq \max(\bar{r} + s_k, \bar{s} + r_k) \). The log-likelihood from time 1 to time \( N \) can be written in the following form
\[
\log L = \frac{-N}{2} \log(2\pi) - \frac{N}{2} \log(\det \Gamma) - \frac{1}{2\sigma^2} w^T \Gamma^{-1} w
\quad (19)
\]

where \( \Gamma \) is the covariance matrix of the zero mean and vector \( w \). The element \((i, j)\) of the exact information matrix \( \Omega \) can be written as in Porat and Friedlander (1986)
\[
\Omega_{ij} = \frac{1}{2} Tr \left\{ \Gamma^{-1} \frac{\partial \Gamma}{\partial \beta_i} \Gamma^{-1} \frac{\partial \Gamma}{\partial \beta_j} \right\} + \left[ \frac{\partial \mu}{\partial \beta_i} \right] \Gamma^{-1} \left[ \frac{\partial \mu}{\partial \beta_j} \right].
\quad (20)
\]

The algorithm of Porat and Friedlander (1986) makes use of the Durbin-Levinson algorithm for computing the orthogonal polynomials of a Toeplitz matrix. Let \( \hat{a}(t) \) be the difference between \( w(t) \) and \( \hat{w}(t) \), which is called the sample innovation at time \( t \). Let \( h(t) \sigma \) be the standard
deviation of \( \hat{a}(t) \) and let the normalized sample innovation be \( \hat{e}(t) = \hat{a}(t)/h(t) \), with mean zero and variance \( \sigma^2 \). The \( \hat{e}(t) \) and the \( h(t) \) can be obtained by the Gram-Schmidt orthogonalization procedure or any procedure which yields equivalent results. The exact likelihood function is now built as the density of the vector \( w \) or the density of the vector \( \hat{e} \), multiplied by the Jacobian of the transformation, which is \( \prod_{i=1}^{N} h(t) \). Hence the log-likelihood from time 1 to time \( N \) can be written in the form

\[
\log L = -\frac{N}{2} \log(2\pi) - N \log(\sigma) - \sum_{i=1}^{N} \log(h(t)) - \frac{1}{2} \sum_{i=1}^{N} \hat{e}^2(t) \sigma^2. \tag{21}
\]

The information matrix is equal to minus the mathematical expectation of the matrix of second derivatives of the log-likelihood

\[
\frac{\partial^2 \log L}{\partial \beta \partial \beta^t} = -\sum_{i=1}^{N} \frac{1}{h(t)} \frac{\partial^2 h(t)}{\partial \beta \partial \beta^t} + \sum_{i=1}^{N} \frac{1}{h^2(t)} \frac{\partial h(t)}{\partial \beta} \frac{\partial h(t)}{\partial \beta^t} - \frac{1}{\sigma^2} \sum_{i=1}^{N} \hat{e}(t) \frac{\partial^2 \hat{e}(t)}{\partial \beta \partial \beta^t}. \tag{22a}
\]

This exact information matrix can also be written as

\[
-E \left( \frac{\partial^2 \log L}{\partial \beta \partial \beta^t} \right) = -\sum_{i=1}^{N} \frac{1}{h(t)} \frac{\partial^2 h(t)}{\partial \beta \partial \beta^t} + \sum_{i=1}^{N} \frac{1}{h^2(t)} \frac{\partial h(t)}{\partial \beta} \frac{\partial h(t)}{\partial \beta^t}
+ \frac{1}{\sigma^2} \sum_{i=1}^{N} E \left( \frac{\hat{e}(t) \hat{e}(t)}{\partial \beta} \frac{\partial \hat{e}(t)}{\partial \beta^t} \right) - \frac{1}{\sigma^2} \sum_{i=1}^{N} E \left( \hat{e}(t) \frac{\partial^2 \hat{e}(t)}{\partial \beta \partial \beta^t} \right). \tag{22b}
\]

For time \( t \geq 2 \), we have

\[
4. \quad h(t) \hat{e}(t) + \hat{w}(t) = w(t) \tag{23}
\]

whereas \( w(1) = h(1) \hat{e}(1) \). Hence, differentiation of (23) yields

\[
\frac{\partial h(t)}{\partial \beta} + h(t) \frac{\partial \hat{e}(t)}{\partial \beta} + \frac{\partial \hat{w}(t)}{\partial \beta} = \frac{\partial w(t)}{\partial \beta} = -\frac{\partial \mu(t)}{\partial \beta} \tag{24a}
\]

because of (18b). Differentiation a second time gives

\[
\frac{\partial^2 h(t)}{\partial \beta \partial \beta^t} \hat{e}(t) + h(t) \frac{\partial^2 \hat{e}(t)}{\partial \beta \partial \beta^t} + 2 \frac{\partial h(t)}{\partial \beta} \frac{\partial \hat{e}(t)}{\partial \beta^t} + \frac{\partial^2 \hat{w}(t)}{\partial \beta \partial \beta^t} = \frac{\partial^2 w(t)}{\partial \beta \partial \beta^t}. \tag{24b}
\]

Because (18a) is not considered as a random variable and since the normalized sample innovations have zero mean, we deduce from (24) that

\[
\frac{\partial h(t)}{\partial \beta} E(\hat{e}^2(t)) + h(t) E \left( \frac{\partial \hat{e}(t)}{\partial \beta} \hat{e}(t) \right) + E \left( \frac{\partial \hat{w}(t)}{\partial \beta} \hat{e}(t) \right) = 0 \tag{25}.
\]

From (25) we have
and similarly, from (24b) and (26), we have
\[ E\left( \hat{e}(t) \frac{\partial^2 \hat{e}(t)}{\partial \beta \partial \beta^T} \right) = -\frac{1}{h(t)} \frac{\partial^2 \hat{e}(t)}{\partial \beta} \sigma^2 - \frac{2}{h(t)} \frac{\partial h(t)}{\partial \beta} E\left( \frac{\partial h(t)}{\partial \beta^T} \hat{e}(t) \right) \]
\[ = \sigma^2 \left( -\frac{1}{h(t)} \frac{\partial^2 \hat{e}(t)}{\partial \beta} + \frac{2}{h^2(t)} \frac{\partial h(t)}{\partial \beta} \frac{\partial h(t)}{\partial \beta^T} \right). \]  

Summarizing (22b) and (27), the exact Fisher information matrix is now given by
\[ -E\left( \frac{\partial^2 \log L}{\partial \beta \partial \beta^T} \right) = \sum_{t=1}^{N} \frac{1}{h^2(t)} \frac{\partial h(t)}{\partial \beta} \frac{\partial h(t)}{\partial \beta^T} + \frac{1}{\sigma^2} \sum_{t=1}^{N} E\left( \frac{\partial \hat{e}(t)}{\partial \beta} \frac{\partial \hat{e}(t)}{\partial \beta^T} \right). \]  

4. Conclusion
This paper has presented an algorithm for the computation of the exact Fisher information matrix of a MISO model. The algorithm is a generalization of that proposed by Klein, Mélard, and Zahaf (1998). The computation is made following the Porat and Friedlander (1986, 1987) approach for the evaluation of the exact Fisher information matrix of an ARMA model based on the Durbin-Levinson algorithm for computing the orthogonal polynomials of a Toeplitz matrix.

Conflict of Interest: The authors declare no conflict of interest.

REFERENCES
[1] Klein, A. and Mélard, G. (1994a). The information matrix of multiple-input single-output time series models. Journal of Computation and Applied Mathematics, 51, 349-356.
[2] Klein, A. and Mélard, G. (1994b). Computation of the Fisher information matrix for SISO models. IEEE Trans. Acoust., Speech, Signal Processing, 42, 684-688.
[3] Klein, A., Mélard, G. and Zahaf, T. (1998). Computation of the exact information matrix of Gaussian dynamic regression time series models. Annals of Statistics, 26, 1636-1650.
[4] Porat, B. and Friedlander, B. (1986). Computation of the exact information matrix of Gaussian time series with stationary random components. IEEE Trans. Acoust., Speech, Signal Processing, 14, 118-130.
[5] Porat, B. and Friedlander, B. (1987). The exact Cramér-Rao bound for bound for Gaussian autoregressive processes. IEEE Trans. Accrosp., Electron. Syst., 23, 537-542.