Research Article

Maximum Likelihood Estimation of the VAR(1) Model Parameters with Missing Observations

Helena Mouriño and Maria Isabel Barão

Departamento de Estatística e Investigação Operacional, Faculdade de Ciências, Universidade de Lisboa, Edifício C6, Piso 4, Campo Grande, 1749-016 Lisboa, Portugal

Correspondence should be addressed to Helena Mouriño; mhnunes@fc.ul.pt

Received 4 January 2013; Revised 29 March 2013; Accepted 8 April 2013

Academic Editor: Xuejun Xie

Copyright © 2013 H. Mouriño and M. I. Barão. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Missing-data problems are extremely common in practice. To achieve reliable inferential results, we need to take into account this feature of the data. Suppose that the univariate data set under analysis has missing observations. This paper examines the impact of selecting an auxiliary complete data set—whose underlying stochastic process is to some extent interdependent with the former—to improve the efficiency of the estimators for the relevant parameters of the model. The Vector AutoRegressive (VAR) Model has revealed to be an extremely useful tool in capturing the dynamics of bivariate time series. We propose maximum likelihood estimators for the parameters of the VAR(1) Model based on monotone missing data pattern. Estimators' precision is also derived. Afterwards, we compare the bivariate modelling scheme with its univariate counterpart. More precisely, the univariate data set with missing observations will be modelled by an AutoRegressive Moving Average (ARMA(2,1)) Model. We will also analyse the behaviour of the AutoRegressive Model of order one, AR(1), due to its practical importance. We focus on the mean value of the main stochastic process. By simulation studies, we conclude that the estimator based on the VAR(1) Model is preferable to those derived from the univariate context.

1. Introduction

Statistical analyses of data sets with missing observations have long been addressed in the literature. For instance, Morrison [1] deduced the maximum likelihood estimators of the parameters of the multinormal mean vector and covariance matrix for the monotonic pattern with only a single incomplete variate. The exact expectations and variances of the estimators were also deduced. Dahiya and Korwar [2] obtained the maximum likelihood estimators for a bivariate normal distribution with missing data. They focused on estimating the correlation coefficient as well as the difference of the two means. Following this line of research and having in mind that the majority of the empirical studies are characterised by temporal dependence between observations, we will try to generalise the previous study by introducing

a bivariate time series model to describe the relationship between the processes under consideration.

The literature on missing data has expanded in the last decades focusing mainly on univariate time series models [3–7], but there is still a lack of developments in the vectorial context.

This paper aims at analysing the main properties of the estimators from data generated by one of the most influential models in empirical studies, that is, the first-order Vector AutoRegressive (VAR(1)) Model, when the data set from the main stochastic process, designated by $\{Y_t\}_{t \in \mathbb{Z}}$, has missing observations. Therefore, we assume that there is also available a suitable auxiliary stochastic process, denoted by $\{X_t\}_{t \in \mathbb{Z}}$, which is to some extent interdependent with the main stochastic process. Additionally, the data set obtained from this process is complete. In this context, a natural question

arises: is it possible to exchange information between the two data sets to increase knowledge about the process whose data set has missing observations, or should we analyse the univariate stochastic process by itself? The goal of this paper is to answer this question.

Throughout this paper, we assume that the incomplete data set has a monotone missing data pattern. We follow a likelihood-based approach to estimate the parameters of the model. It is worth pointing out that, in the literature, likelihood-based estimation is largely used to manage the problem of missing data [3, 8, 9]. The precision of the maximum likelihood estimators is also derived.

In order to answer the question raised above, we must verify if the introduction of an auxiliary variable for estimating the parameters of the model increases the accuracy of the estimators. To accomplish this goal, we compare the precision of the estimators just cited with those obtained from modelling the dynamics of the univariate stochastic process $\{Y_t\}_{t \in \mathbb{Z}}$ by an AutoRegressive Moving Average (ARMA(2,1)) Model, which corresponds to the marginal model of the bivariate VAR(1) Model [10, 11]. The behaviour of the AutoRegressive Model of order one, AR(1), is also analysed due to its practical importance in time series modelling. Simulation studies allow us to assess the relative efficiency of the different approaches. Special attention is paid to the estimator for the mean value of the stochastic process about which information available is scarce. This is a reasonable choice given the importance of the mean function of a stochastic process in understanding the behaviour of the time series under consideration.

The paper is organised as follows. In Section 2, we review the VAR(1) Model and highlight a few statistical properties that will be used in the remaining sections. In Section 3, we establish the monotone pattern of missing data and factorise the likelihood function of the VAR(1) Model. The maximum likelihood estimators of the parameters are obtained in Section 4. Their precision is also deduced. Section 5 reports the simulation studies in evaluating different approaches to estimate the mean value of the stochastic process $\{Y_t\}_{t \in \mathbb{Z}}$. The main conclusions are summarised in Section 6.

2. Brief Description of the VAR(1) Model

In this section, a few properties of the Vectorial Autoregressive Model of order one are analysed. These features will play an important role in determining the estimators for the parameters when there are missing observations, as we will see in Section 4.

Hereafter, the stochastic process underlying the complete data set is denoted by $\{X_t\}_{t \in \mathbb{Z}}$, while the other one is represented by $\{Y_t\}_{t \in \mathbb{Z}}$. The VAR(1) Model under consideration takes the form

$$\begin{aligned} X_t &= \alpha_0 + \alpha_1 X_{t-1} + \epsilon_t, \\ Y_t &= \beta_0 + \beta_1 Y_{t-1} + \beta_2 X_{t-1} + \xi_t, \end{aligned} \quad t = 0, \pm 1, \pm 2, \dots, \quad (1)$$

where ϵ_t and ξ_t are Gaussian white noise processes with zero mean and variances σ_ϵ^2 and σ_ξ^2 , respectively. The structure

of correlation between the error terms is different from zero only at the same date t , that is, $\text{Cov}(\epsilon_{t-i}, \xi_{t-j}) = \sigma_\epsilon \sigma_\xi$, for $i = j$; $\text{Cov}(\epsilon_{t-i}, \xi_{t-j}) = 0$, for $i \neq j$, $i, j \in \mathbb{Z}$. Exchanging information between both time series might introduce some noise in the overall process. Therefore, transfer of information from the smallest series to the largest one is not allowed here.

We have to introduce the restrictions $|\alpha_1| < 1$ and $|\beta_1| < 1$. They ensure not only that the underlying processes are ergodic for the respective means but also that the stochastic processes are covariance stationary (see Nunes [12, ch.3]). Hereafter, we assume that these restrictions are satisfied.

Next, we overview some relevant properties of the VAR(1) Model (1). Theoretical details can be found in Nunes [12, ch.3].

The mean values of X_t and Y_t are, respectively, given by

$$E(X_t) = \frac{\alpha_0}{1 - \alpha_1}, \quad E(Y_t) = \frac{\alpha_0 \beta_2 + \beta_0 (1 - \alpha_1)}{(1 - \alpha_1)(1 - \beta_1)}. \quad (2)$$

Concerning the covariance structure of the process X_t ,

$$\text{Cov}(X_{t-i}, X_{t-j}) = \sigma_\epsilon^2 \frac{\alpha_1^{|i-j|}}{1 - \alpha_1^2}, \quad \forall i, j \in \mathbb{Z}. \quad (3)$$

For $\alpha_1 \neq \beta_1$, the covariance of the stochastic process Y_t is given by

$$\begin{aligned} \text{Cov}(Y_{t-i}, Y_{t-j}) &= \frac{\sigma_\xi^2 \beta_1^{|i-j|}}{1 - \beta_1^2} \\ &+ \sigma_{\epsilon\xi} \beta_2 \left\{ \frac{1}{\beta_1 - \alpha_1} \right. \\ &\quad \times \left(\frac{\beta_1^{|i-j|}}{1 - \beta_1^2} - \frac{\alpha_1^{|i-j|}}{1 - \alpha_1 \beta_1} \right) \\ &\quad \left. + \frac{\beta_1 \beta_1^{|i-j|}}{(1 - \beta_1^2)(1 - \alpha_1 \beta_1)} \right\} \\ &+ \sigma_\epsilon^2 \frac{\beta_2^2}{(\beta_1 - \alpha_1)(1 - \alpha_1 \beta_1)} \\ &\times \left(\frac{\beta_1 \beta_1^{|i-j|}}{1 - \beta_1^2} - \frac{\alpha_1 \alpha_1^{|i-j|}}{1 - \alpha_1^2} \right), \quad \forall i, j \in \mathbb{Z}. \end{aligned} \quad (4)$$

Considering that $\alpha_1 = \beta_1$, we have

$$\begin{aligned} \text{Cov}(Y_{t-i}, Y_{t-j}) &= \frac{\beta_1^{|i-j|}}{1 - \beta_1^2} \left\{ \sigma_\xi^2 + 2 \sigma_{\epsilon\xi} \beta_2 \right. \\ &\quad \times \left(\frac{\beta_1}{1 - \beta_1^2} + \frac{|i-j|}{\beta_1} \right) \\ &\quad \left. + \sigma_\epsilon^2 \frac{\beta_2^2}{1 - \beta_1^2} \left(|i-j| + \frac{1 + \beta_1^2}{1 - \beta_1^2} \right) \right\}, \end{aligned}$$

for $i, j \in \mathbb{Z}$. (5)

In regard to the structure of covariance between the stochastic processes X_t and Y_t , for $\alpha_1 \neq \beta_1$, we have

$$\begin{aligned} \text{Cov}(Y_{t-i}, X_{t-j}) \\ = \sigma_{\epsilon\xi} \frac{\alpha_1^{|i-j|}}{1 - \alpha_1\beta_1} + \sigma_\epsilon^2 \frac{\alpha_1 \beta_2 \alpha_1^{|i-j|}}{(1 - \alpha_1\beta_1)(1 - \alpha_1^2)}, \quad \forall i, j \in \mathbb{Z}. \end{aligned} \quad (6)$$

When $\alpha_1 = \beta_1$, the covariance function under study takes the form

$$\text{Cov}(Y_{t-i}, X_{t-j}) = \sigma_\epsilon^2 \frac{\beta_2 \alpha_1 \alpha_1^{|i-j|}}{(1 - \alpha_1^2)^2} + \sigma_{\epsilon\xi} \frac{\alpha_1^{|i-j|}}{1 - \alpha_1^2}, \quad \forall i, j \in \mathbb{Z}. \quad (7)$$

By writing out the stochastic system of (1) in matrix notation, the bivariate stochastic process $\mathbf{Z}_t = [X_t \ Y_t]'$ can be expressed as

$$\begin{aligned} \mathbf{Z}_t &= \begin{bmatrix} \alpha_0 \\ \beta_0 \end{bmatrix} + \begin{bmatrix} \alpha_1 & 0 \\ \beta_2 & \beta_1 \end{bmatrix} \begin{bmatrix} X_{t-1} \\ Y_{t-1} \end{bmatrix} \\ &+ \begin{bmatrix} \epsilon_t \\ \xi_t \end{bmatrix} = \mathbf{c} + \Phi_1 \mathbf{Z}_{t-1} + \boldsymbol{\epsilon}_t, \quad t \in \mathbb{Z}, \end{aligned} \quad (8)$$

where $\boldsymbol{\epsilon}_t = [\epsilon_t \ \xi_t]'$ is the 2-dimensional Gaussian white noise random vector.

Hence, at each date $t, t \in \mathbb{Z}$, the conditional stochastic process $\mathbf{Z}_t | \mathbf{Z}_{t-1} = \mathbf{z}_{t-1}$ follows a bivariate Gaussian distribution, $\mathbf{Z}_t | \mathbf{Z}_{t-1} = \mathbf{z}_{t-1} \sim \mathcal{N}_2(\boldsymbol{\mu}_{t|t-1}, \boldsymbol{\Omega}_{t|t-1})$, where the two-dimensional conditional mean value vector and the variance-covariance matrix are, respectively, given by

$$\begin{aligned} \boldsymbol{\mu}_{t|t-1} &= \mathbf{c} + \Phi_1 \mathbf{Z}_{t-1}, \\ \boldsymbol{\Omega}_{t|t-1} &\equiv \boldsymbol{\Omega} = \begin{bmatrix} \sigma_\epsilon^2 & \sigma_{\epsilon\xi} \\ \sigma_{\epsilon\xi} & \sigma_\xi^2 \end{bmatrix}. \end{aligned} \quad (9)$$

Straightforward computations lead us to the following factoring of the probability density function of $\mathbf{Z}_t$ conditional to $\mathbf{Z}_{t-1} = \mathbf{z}_{t-1}$:

$$\begin{aligned} f_{\mathbf{Z}_t | \mathbf{Z}_{t-1}}(\mathbf{z}_t | \mathbf{z}_{t-1}) &= f_{X_t | \mathbf{Z}_{t-1}}(x_t | \mathbf{z}_{t-1}) \\ &\times f_{Y_t | X_t, \mathbf{Z}_{t-1}}(y_t | x_t, \mathbf{z}_{t-1}). \end{aligned} \quad (10)$$

Thus, the joint distribution of the pair X_t and Y_t conditional to the values of the process at the previous date $t-1$, $\mathbf{Z}_{t-1}$, can be decomposed into the product of the marginal distribution of $X_t | \mathbf{Z}_{t-1}$ and the conditional distribution of $Y_t | X_t, \mathbf{Z}_{t-1}$. Both densities follow univariate Gaussian probability laws:

$$X_t | \mathbf{Z}_{t-1} = \mathbf{z}_{t-1} \sim \mathcal{N}(\alpha_0 + \alpha_1 x_{t-1}, \sigma_\epsilon^2), \quad \text{for each date } t, \quad t \in \mathbb{Z}. \quad (11)$$

Also, $Y_t | X_t = x_t, \mathbf{Z}_{t-1} = \mathbf{z}_{t-1}$ follows a Gaussian distribution with

$$\begin{aligned} E(Y_t | X_t = x_t, \mathbf{Z}_{t-1} = \mathbf{z}_{t-1}) \\ = \beta_0 + \beta_1 y_{t-1} + \beta_2 x_{t-1} + \frac{\sigma_{\epsilon\xi}}{\sigma_\epsilon^2} (x_t - \alpha_0 - \alpha_1 x_{t-1}) \\ = \psi_0 + \psi_1 x_t + \psi_2 x_{t-1} + \beta_1 y_{t-1}, \end{aligned} \quad (12)$$

where $\psi_1 = (\sigma_{\epsilon\xi}/\sigma_\epsilon^2)$ or, for interpretive purposes, $\psi_1 = (\sigma_\xi/\sigma_\epsilon)\rho_{\epsilon\xi}$. The parameter ψ_1 describes, thus, a weighted correlation between the error terms ϵ_t and ξ_t . The weight corresponds to the ratio of their standard deviations. Moreover, $\psi_0 = \beta_0 - \psi_1, \psi_2 = \beta_2 - \psi_1\alpha_1$.

The variance has the following structure:

$$\begin{aligned} \text{Var}(Y_t | X_t = x_t, \mathbf{Z}_{t-1} = \mathbf{z}_{t-1}) &= \sigma_\xi^2 - \frac{\sigma_{\epsilon\xi}^2}{\sigma_\epsilon^2} \\ &= \sigma_\xi^2 (1 - \rho_{\epsilon\xi}^2) \equiv \psi_3. \end{aligned} \quad (13)$$

The conditional distribution of $Y_t | X_t, \mathbf{Z}_{t-1}$ can be interpreted as a straight-line relationship between Y_t and X_t, X_{t-1} , and Y_{t-1} . Additionally, it is worth mentioning that if $\rho_{\epsilon\xi} = \pm 1$ or $\sigma_\xi^2 = 0$, the above conditional distribution degenerates into its mean value. Henceforth, we will discard these particular cases, which means that $\psi_3 \neq 0$.

3. Factoring the Likelihood Based on Monotone Missing Data Pattern

We focus here on theoretical background for factoring the likelihood function from the VAR(1) Model when there are missing values in the data. Suppose that we have the following monotone pattern of missing data:

$$\begin{array}{ccccccc} x_0 & x_2 & \cdots & x_{m-1} & x_m & \cdots & x_{n-1} \\ y_0 & y_2 & \cdots & y_{m-1} & \boxed{} & \cdots & \end{array} \quad (14)$$

That is, there are n observations available from the stochastic process $\{X_t\}_{t \in \mathbb{Z}}$, whereas due to some uncontrolled factors it was only possible to record $m(m < n)$ observations from the stochastic process $\{Y_t\}_{t \in \mathbb{Z}}$. In other words, there are $n - m$ missing observations from Y_t .

Let the observed bivariate sample of size n with missing values:

$$\{(x_0, y_0), (x_1, y_1), \dots, (x_{m-1}, y_{m-1}), x_m, \dots, x_{n-1}\}; \quad (15)$$

denote a realisation of the random process $\mathbf{Z}_t = [X_t \ Y_t]'$, $t \in \mathbb{Z}$, which follows a vectorial autoregressive model of order one. The likelihood function, $L(\boldsymbol{\theta})$, is given by

$$\begin{aligned} L(\boldsymbol{\theta}) &\equiv f_{\mathbf{Z}_0, \mathbf{Z}_1, \dots, \mathbf{Z}_{m-1}, X_m, \dots, X_{n-1}} \\ &\times (\mathbf{z}_0, \mathbf{z}_1, \dots, \mathbf{z}_{m-1}, x_m, \dots, x_{n-1}) \\ &= f_{\mathbf{Z}_0}(\mathbf{z}_0) \prod_{t=1}^{m-1} f_{\mathbf{Z}_t | \mathbf{Z}_{t-1}}(\mathbf{z}_t | \mathbf{z}_{t-1}) f_{X_m | \mathbf{Z}_{m-1}}(x_m | \mathbf{z}_{m-1}) \\ &\times \prod_{t=m+1}^{n-1} f_{X_t | X_{t-1}}(x_t | x_{t-1}; \boldsymbol{\theta}) \end{aligned}$$

$$\begin{aligned}
&= f_{\mathbf{z}_0}(\mathbf{z}_0) \prod_{t=1}^{m-1} f_{\mathbf{z}_t|\mathbf{z}_{t-1}}(\mathbf{z}_t | \mathbf{z}_{t-1}) \\
&\quad \times \prod_{t=m}^{n-1} f_{X_t|X_{t-1}}(x_t | x_{t-1}),
\end{aligned} \tag{16}$$

where $\boldsymbol{\theta} = [\alpha_0 \ \alpha_1 \ \sigma_e^2 \ \beta_0 \ \beta_1 \ \beta_2 \ \sigma_\xi^2 \ \sigma_{\varepsilon\xi}]'$ is the 8-dimensional vector of population parameters. To lighten notation, we assume that there is no need for conditioning the arguments of the above probability density functions on the values of the processes at date $t - 1$. The likelihood function becomes

$$L(\boldsymbol{\theta}) = f_{\mathbf{z}_0}(\mathbf{z}_0) \prod_{t=1}^{m-1} f_{\mathbf{z}_t|\mathbf{z}_{t-1}}(\mathbf{z}_t) \prod_{t=m}^{n-1} f_{X_t|X_{t-1}}(x_t). \tag{17}$$

Two points must be emphasised: first, we emphasise that the maximum likelihood estimators (m.l.e.) for the unknown vector of parameters will be obtained by maximising the natural logarithm of the above likelihood function. Second, a worthwhile improvement in reducing the complexity of the function to maximise is to determine the conditional maximum likelihood estimators regarding the first pair of random variables, $\mathbf{Z}_0 = [X_0 \ Y_0]'$, as deterministic and maximising the log-likelihood function conditioned on the values $X_0 = x_0$ and $Y_0 = y_0$. The loss of efficiency of the estimators obtained from such a procedure is negligible when compared with the exact maximum likelihood estimators computed by iterative techniques. Even for moderate sample sizes, the first pair of observations makes a negligible contribution to the total likelihood. Hence, the exact m.l.e. and the conditional m.l.e. turn out to have the same large sample properties, Hamilton [13]. Hereafter, we restrict the study to the conditional loglikelihood function.

Despite the above solutions for reducing the complexity of the problem, some difficulties still remain. The loglikelihood equations are intractable. To go over this problem we have to factorise the conditional likelihood function. From (17) we get

$$\begin{aligned}
L(\boldsymbol{\theta}) &= \prod_{t=1}^{m-1} f_{(X_t, Y_t)|(X_{t-1}, Y_{t-1})}(x_t, y_t) \prod_{t=m}^{n-1} f_{X_t|X_{t-1}}(x_t) \\
&= \prod_{t=1}^{m-1} (f_{X_t|(X_{t-1}, Y_{t-1})}(x_t) \times f_{Y_t|X_t, (X_{t-1}, Y_{t-1})}(y_t)) \\
&\quad \times \prod_{t=m}^{n-1} f_{X_t|X_{t-1}}(x_t) \\
&= \prod_{t=1}^{m-1} f_{X_t|X_{t-1}}(x_t) \prod_{t=1}^{m-1} f_{Y_t|X_t, (X_{t-1}, Y_{t-1})}(y_t).
\end{aligned} \tag{18}$$

So as to work out the analytical expressions for the unknown parameters under study, we have to decompose

the entire likelihood function (18) into easily manipulated components.

For the Gaussian VAR processes, the conditional maximum likelihood estimators coincide with the least squares estimators [13]. Therefore, we may find a solution to the problem just raised in the geometrical context. The identification of such components relies on two of the most famous theorems in the Euclidean space: the Orthogonal Decomposition Theorem and the Approximation Theorem [14, Volume I, pages 572–575]. Based on these tools it is straightforward to establish that the estimation subspaces associated with the conditional distributions $X_t|X_{t-1}$ and $Y_t|X_t, X_{t-1}, Y_{t-1}$ are, by construction, orthogonal to each other. This means that each element belonging to one of those subspaces is uncorrelated with each element that pertains to their orthogonal complement. Hence, events that happen on one subspace provide no information about events on the other subspace.

The aforementioned arguments guarantee that the decomposition of the joint likelihood in two components can be carried out with no loss of information for the whole estimation procedure. From (18) we can, thus, decompose the conditional loglikelihood function as follows:

$$\begin{aligned}
l \equiv l(\boldsymbol{\theta}) &= \log L(\boldsymbol{\theta}) = \sum_{t=1}^{m-1} \log f_{X_t|X_{t-1}}(x_t) \\
&\quad + \sum_{t=1}^{m-1} \log f_{Y_t|X_t, (X_{t-1}, Y_{t-1})}(y_t) = l_1 + l_2.
\end{aligned} \tag{19}$$

Henceforth, l_1 denotes the loglikelihood from the marginal distribution of X_t , based on the whole sampled data with dimension n , that is, $x_0, x_1, \dots, x_{n-1}$. The function l_2 represents the loglikelihood from the conditional density of $Y_t|X_t, \mathbf{z}_{t-1}$ computed by the bivariate sample of size m :

$$(x_0, y_0), (x_1, y_1), \dots, (x_{m-1}, y_{m-1}). \tag{20}$$

The components l_1 and l_2 of (19) will be maximised separately in Section 4.1.

4. Maximum Likelihood Estimators for the Parameters

In Section 4.1 the m.l.e. of the parameters from the fragmentary VAR(1) Model are deduced. The precision of the estimators is examined in Section 4.2.

4.1. Analytical Expressions. Theoretical developments carried out in this section rely on solving the loglikelihood equations obtained from the factored loglikelihood given by (19). Before proceeding with theoretical matters, we introduce some relevant notation in the ensuing paragraphs.

Let $\bar{X}_k^{(l)} = (1/k) \sum_{t=1}^k X_{t-l}$ represent the sample mean lagged l time units, $l = 0, 1$. The subscript $k, k = 1, \dots, n-1$, allows us to identify the number of observations that takes part in the computation of the sample mean. A similar notation is used for denoting the sample mean of the random

sample $Y_0, \dots, Y_k$, for $k = 1, \dots, m-1$, $\bar{Y}_k^{(l)}$. According to this new definition, the sample variance of each univariate random variable based on k observations and lagged l time units is denoted by

$$\begin{aligned}\hat{\gamma}_{X,k}^{(l)} &= \frac{1}{k} \sum_{t=1}^k \left(X_{t-l} - \bar{X}_k^{(l)} \right)^2, \\ \hat{\gamma}_{Y,k}^{(l)} &= \frac{1}{k} \sum_{t=1}^k \left(Y_{t-l} - \bar{Y}_k^{(l)} \right)^2, \quad l = 0, 1.\end{aligned}\quad (21)$$

Let $\hat{\gamma}_{X,k}^*(1) = (1/k) \sum_{t=1}^k (X_t - \bar{X}_k^{(0)})(X_{t-1} - \bar{X}_k^{(1)})$ describe the sample autocovariance coefficient at lag one for the stochastic process X_t , based on k observations. Its counterpart for the stochastic process Y_t , $\hat{\gamma}_{Y,k}^*(1)$, is obtained by changing notation accordingly. The sample autocorrelation coefficient of the random process X_t at lag one is denoted by $\hat{\rho}_{X,k}^{(1)} = \hat{\gamma}_{X,k}^*(1) / \hat{\gamma}_{X,k}^{(0)}$. The empirical covariance between the random processes X_t and Y_t lagged one time unit is represented by

$$\begin{aligned}\hat{\gamma}_{XY}^*(1) &= \frac{1}{m-1} \sum_{t=1}^{m-1} \left(X_t - \bar{X}_{m-1}^{(0)} \right) \left(Y_{t-1} - \bar{Y}_{m-1}^{(1)} \right), \\ &\quad \text{for lagged values on } Y, \\ \hat{\gamma}_{YX}^*(1) &= \frac{1}{m-1} \sum_{t=1}^{m-1} \left(X_{t-1} - \bar{X}_{m-1}^{(1)} \right) \left(Y_t - \bar{Y}_{m-1}^{(0)} \right), \\ &\quad \text{for lagged values on } X.\end{aligned}\quad (22)$$

The sample covariance coefficient of X_t and Y_t computed from l time units lag for each series is given by

$$\begin{aligned}\hat{\gamma}_{XY}^{(l)} &= \frac{1}{m-1} \sum_{t=1}^{m-1} \left(X_{t-l} - \bar{X}_{m-1}^{(l)} \right) \left(Y_{t-l} - \bar{Y}_{m-1}^{(l)} \right), \\ &\quad \text{with } l = 0, 1.\end{aligned}\quad (23)$$

(i) *Maximising the loglikelihood function l_1* : Using the results (11) and (19), we readily find the following m.l.e.

$$\hat{\alpha}_0 = \bar{X}_{n-1}^{(0)} - \hat{\alpha}_1 \bar{X}_{n-1}^{(1)}, \quad \hat{\alpha}_1 = \frac{\hat{\gamma}_{X,n-1}^*(1)}{\hat{\gamma}_{X,n-1}^{(1)}}, \quad (24)$$

$$\hat{\sigma}_\epsilon^2 = \frac{SS_R}{n-1},$$

where SS_R is the respective residual sum of squares.

(ii) *Maximising the loglikelihood function l_2* : Based on (12) and (13) we get the loglikelihood function l_2

$$\begin{aligned}l_2 &= \sum_{t=1}^{m-1} \log f_{Y_t|X_t, X_{t-1}, Y_{t-1}}(y_t) = -\frac{m-1}{2} \log(2\pi) \\ &\quad - \frac{m-1}{2} \log \psi_3 - \frac{1}{2\psi_3} \\ &\quad \times \sum_{t=1}^{m-1} (y_t - \psi_0 - \psi_1 x_t - \psi_2 x_{t-1} - \beta_1 y_{t-1})^2.\end{aligned}\quad (25)$$

We readily find out that the m.l.e. for the parameters under study are given by

$$\begin{aligned}\hat{\psi}_0 &= \bar{Y}_{m-1}^{(0)} - \hat{\psi}_1 \bar{X}_{m-1}^{(0)} - \hat{\psi}_2 \bar{X}_{m-1}^{(1)} - \hat{\beta}_1 \bar{Y}_{m-1}^{(1)}, \\ \hat{\psi}_1 &= \frac{1}{\hat{\gamma}_{X,m-1}^{(0)}} \left\{ \hat{\gamma}_{XY}^{(0)} - \hat{\psi}_2 \hat{\gamma}_{X,m-1}^*(1) - \hat{\beta}_1 \hat{\gamma}_{XY}^*(1) \right\}, \\ \hat{\psi}_2 &= \frac{1}{\left(1 - \left(\hat{\rho}_{X,m-1}^{(1)} \right)^2 \right) \hat{\gamma}_{X,m-1}^{(1)}} \\ &\quad \times \left\{ \hat{\gamma}_{YX}^*(1) - \frac{\hat{\gamma}_{XY}^{(0)} \hat{\gamma}_{X,m-1}^*(1)}{\hat{\gamma}_{X,m-1}^{(0)}} - \hat{\beta}_1 \hat{\gamma}_{XY}^*(1) \right. \\ &\quad \left. + \frac{\hat{\beta}_1 \hat{\gamma}_{XY}^*(1) \hat{\gamma}_{X,m-1}^*(1)}{\hat{\gamma}_{X,m-1}^{(0)}} \right\}, \\ \hat{\psi}_3 &= \frac{SS_R^*}{m-1}, \\ \hat{\beta}_1 &= \frac{1}{\left(\hat{\gamma}_{Y,m-1}^{(1)} + \left(\bar{Y}_{m-1}^{(1)} \right)^2 \right)} \\ &\quad \times \left\{ \hat{\gamma}_{Y,m-1}^*(1) + \bar{Y}_{m-1}^{(0)} \bar{Y}_{m-1}^{(1)} \right. \\ &\quad - \hat{\psi}_0 \bar{Y}_{m-1}^{(1)} - \hat{\psi}_1 \hat{\gamma}_{XY}^*(1) - \hat{\psi}_1 \bar{X}_{m-1}^{(0)} \bar{Y}_{m-1}^{(1)} \\ &\quad \left. - \hat{\psi}_2 \left(\hat{\gamma}_{XY}^{(1)} - \bar{X}_{m-1}^{(1)} \bar{Y}_{m-1}^{(1)} \right) \right\},\end{aligned}\quad (26)$$

where SS_R^* denotes the corresponding residual sums of squares.

Using the results from Section 2 we get the following estimators for the original parameters:

$$\begin{aligned}\hat{\beta}_0 &= \hat{\psi}_0 + \hat{\psi}_1 \hat{\alpha}_0, & \hat{\beta}_2 &= \hat{\psi}_2 + \hat{\psi}_1 \hat{\alpha}_1, \\ \hat{\sigma}_{\epsilon\xi} &= \hat{\psi}_1 \hat{\sigma}_\epsilon^2, & \hat{\sigma}_\xi^2 &= \hat{\psi}_3 + \frac{\hat{\sigma}_{\epsilon\xi}^2}{\hat{\sigma}_\epsilon^2}.\end{aligned}\quad (27)$$

Thus, the analytical expressions for the estimators of the mean values, variances, and covariances of the VAR(1) Model are given by

$$\begin{aligned}
 \hat{\mu}_X &= \frac{\hat{\alpha}_0}{1 - \hat{\alpha}_1}, \\
 \hat{\mu}_Y &= \frac{\hat{\alpha}_0 \hat{\beta}_2 + \hat{\beta}_0 (1 - \hat{\alpha}_1)}{(1 - \hat{\alpha}_1)(1 - \hat{\beta}_1)}, \\
 \hat{\sigma}_X^2 &= \frac{\hat{\sigma}_\epsilon^2}{1 - \hat{\alpha}_1^2}, \\
 \hat{\sigma}_Y^2 &= \frac{\hat{\sigma}_\epsilon^2}{1 - \hat{\beta}_1^2} + \frac{2\hat{\sigma}_{\epsilon\xi}\hat{\beta}_1\hat{\beta}_2}{(1 - \hat{\alpha}_1\hat{\beta}_1)(1 - \hat{\beta}_1^2)} \\
 &\quad + \frac{\hat{\sigma}_\epsilon^2 \hat{\beta}_2^2 (1 + \hat{\alpha}_1\hat{\beta}_1)}{(1 - \hat{\alpha}_1^2)(1 - \hat{\beta}_1^2)(1 - \hat{\alpha}_1\hat{\beta}_1)} \quad (\alpha_1 \neq \beta_1), \\
 \hat{\sigma}_Y^2 &= \frac{\hat{\sigma}_\epsilon^2}{1 - \hat{\alpha}_1^2} + 2 \frac{\hat{\sigma}_{\epsilon\xi}\hat{\alpha}_1\hat{\beta}_2}{(1 - \hat{\alpha}_1^2)^2} + \hat{\sigma}_\epsilon^2 \hat{\beta}_2^2 \frac{1 + \hat{\alpha}_1^2}{(1 - \hat{\alpha}_1^2)^3}, \\
 \hat{\sigma}_{XY} &= \frac{\hat{\sigma}_{\epsilon\xi}}{1 - \hat{\alpha}_1\hat{\beta}_1} + \frac{\hat{\alpha}_1\hat{\beta}_2\hat{\sigma}_\epsilon^2}{(1 - \hat{\alpha}_1\hat{\beta}_1)(1 - \hat{\alpha}_1^2)} \\
 &\quad (\text{at the same date } t, t \in \mathbb{Z}).
 \end{aligned} \tag{28}$$

These estimators will play a central role in the following sections.

4.2. Precision of the Estimators. In the section, the precision of the maximum likelihood estimators underlying equations (28) is derived. The whole analysis will be separated in three stages. First, we study the statistical properties of the vector $\hat{\Theta}$, where $\hat{\Theta} = [\hat{\Theta}_1' \hat{\Theta}_2']'$, with $\hat{\Theta}_1 = [\hat{\alpha}_0 \hat{\alpha}_1 \hat{\sigma}_\epsilon^2]'$ and $\hat{\Theta}_2 = [\hat{\psi}_0 \hat{\psi}_1 \hat{\psi}_2 \hat{\beta}_1 \hat{\psi}_3]'$. For notation consistency, the unknown parameter β_1 is either denoted by β_1 or ψ_4 . That is, $\psi_4 \equiv \beta_1$. Secondly, we derive the precision of the m.l.e. of the original parameters of the VAR(1) Model (see (1)). Finally, we will focus our attention on the estimators for the mean vector and the variance-covariance matrix at lag zero of the VAR(1) model with a monotone pattern of missingness.

There are a few points worth mentioning. From Section 3 we know that there is no loss of information in maximising

separately the loglikelihood functions l_1 and l_2 (19). As a consequence, the variance-covariance matrix associated with the whole set of estimated parameters is a block diagonal matrix. For sufficiently large sample size, the distribution of the maximum likelihood estimator is accurately approximated by the following multivariate Gaussian distribution:

$$\hat{\Theta} \approx \mathcal{N}_8 \left(\begin{bmatrix} \hat{\Theta}_1 \\ \hat{\Theta}_2 \end{bmatrix}, \begin{bmatrix} \mathbf{I}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_2^{-1} \end{bmatrix} \right), \tag{29}$$

where $\mathbf{I}_1$ and $\mathbf{I}_2$ denote the Fisher information matrices, respectively, from the components l_1 and l_2 of the loglikelihood function (see (19)). There is an asymptotic equivalence between the Fisher information matrix and the Hessian matrix (see [8, ch.2]). Moreover, as long as $\hat{\Theta} \rightarrow \Theta$ there is also an asymptotic equivalence between the Hessian matrix computed at the points $\hat{\Theta}$ and Θ . Henceforth, the Fisher information matrices from (29) are estimated, respectively, by

$$\hat{\mathbf{I}}_1 = - \left(\frac{\partial^2 l_1}{\partial \Theta_1 \partial \Theta_1'} \right) \Big|_{\Theta_1 = \hat{\Theta}_1}, \quad \hat{\mathbf{I}}_2 = - \left(\frac{\partial^2 l_1}{\partial \Theta_2 \partial \Theta_2'} \right) \Big|_{\Theta_2 = \hat{\Theta}_2}. \tag{30}$$

To lighten notation, from now on we suppress the “hat” from the consistent estimators of the information matrices.

The variance-covariance matrix for $\hat{\Theta}_1$ takes the following form:

$$\begin{aligned}
 \mathbf{I}_1^{-1} &= \frac{\hat{\sigma}_\epsilon^2}{(n-1) \hat{\gamma}_{X,n-1}^{(1)}(0)} \\
 &\times \begin{bmatrix} \hat{\gamma}_{X,n-1}^{(1)}(0) + \left(\overline{X}_{n-1}^{(1)} \right)^2 - \overline{X}_{n-1}^{(1)} & 0 \\ -\overline{X}_{n-1}^{(1)} & 1 \\ 0 & 0 \end{bmatrix} \Big|_{\hat{\gamma}_{X,n-1}^{(1)}(0)}.
 \end{aligned} \tag{31}$$

We stress that there is orthogonality between the error and the estimation subspaces underlying the loglikelihood function l_1 .

Calculating the second derivatives of the loglikelihood function l_2 results in the following approximate information matrix:

$$\mathbf{I}_2 = \frac{1}{\hat{\psi}_3} \begin{bmatrix} m-1 & (m-1) \overline{X}_{m-1} & (m-1) \overline{X}_{m-1}^{(1)} & (m-1) \overline{Y}_{m-1}^{(1)} & 0 \\ (m-1) \overline{X}_{m-1} & \sum_{t=1}^{m-1} X_t^2 & \sum_{t=1}^{m-1} X_t X_{t-1} & \sum_{t=1}^{m-1} X_t Y_{t-1} & 0 \\ (m-1) \overline{X}_{m-1}^{(1)} & \sum_{t=1}^{m-1} X_t X_{t-1} & \sum_{t=1}^{m-1} X_{t-1}^2 & \sum_{t=1}^{m-1} X_{t-1} Y_{t-1} & 0 \\ (m-1) \overline{Y}_{m-1}^{(1)} & \sum_{t=1}^{m-1} X_t Y_{t-1} & \sum_{t=1}^{m-1} X_{t-1} Y_{t-1} & \sum_{t=1}^{m-1} Y_{t-1}^2 & 0 \\ 0 & 0 & 0 & 0 & \frac{m-1}{2 \hat{\psi}_3} \end{bmatrix}. \tag{32}$$

Once again, we mention that there is orthogonality between the error and the estimation subspaces underlying the loglikelihood function l_2 . The matrix $\mathbf{I}_2$ can be written in a compact form:

$$\mathbf{I}_2 = \frac{1}{\hat{\psi}_3} \left[\begin{array}{c|c} \mathbf{I}_{21} & \mathbf{0} \\ \hline \mathbf{0} & I_{22} \end{array} \right], \quad (33)$$

where the (4×4) submatrix $\mathbf{I}_{21}$ and the scalar I_{22} are, respectively, defined as

$$\mathbf{I}_{21} = \mathbf{U}'\mathbf{U}, \quad I_{22} = \frac{m-1}{2(\hat{\psi}_3)^2}, \quad (34)$$

with

$$\mathbf{U} = \begin{bmatrix} 1 & X_1 & X_0 & Y_0 \\ 1 & X_2 & X_1 & Y_1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{m-1} & X_{m-2} & Y_{m-2} \end{bmatrix}. \quad (35)$$

Using the above partition of $\mathbf{I}_2$ it is rather simple to compute the inverse matrix. In fact,

$$\mathbf{I}_2^{-1} = \hat{\psi}_3 \left[\begin{array}{c|c} \mathbf{I}_{21}^{-1} & \mathbf{0} \\ \hline \mathbf{0} & I_{22}^{-1} \end{array} \right], \quad (36)$$

with $\mathbf{I}_{21}^{-1} = (\mathbf{U}'\mathbf{U})^{-1}$ and $I_{22}^{-1} = (2/(m-1))(\hat{\psi}_3)^2$.

Unfortunately, there is no explicit expression for the inverse matrix $\mathbf{I}_{21}^{-1}$. As a result, there are no explicit expressions for the approximate variance-covariance of the m.l.e. for the vector of unknown parameters $\hat{\Theta}_2$.

Now, we have to analyse the precision of the m.l.e. of the original parameters of the VAR(1) Model, that is, $\mathbf{Y} = [\alpha_0 \ \alpha_1 \ \sigma_\epsilon^2 \ \beta_0 \ \beta_1 \ \beta_2 \ \sigma_\xi^2 \ \sigma_{\epsilon\xi}]'$.

Recalling from Section 2, the one-one monotone functions that relate the vector of parameters under consideration, that is,

$$\Theta_2 = [\psi_0 \ \psi_1 \ \psi_2 \ \psi_4 \ \psi_3]'$$

$$\mathbf{Y}_2 = [\beta_0 \ \beta_1 \ \beta_2 \ \sigma_\xi^2 \ \sigma_{\epsilon\xi}]', \quad \text{are} \quad (37)$$

$$\psi_0 = \beta_0 - \alpha_0\psi_1, \quad \psi_1 = \frac{\sigma_{\epsilon\xi}}{\sigma_\epsilon^2}, \quad \psi_2 = \beta_2 - \psi_1\alpha_1$$

$$\psi_3 = \sigma_\xi^2 - \sigma_\epsilon^2\psi_1^2, \quad \psi_4 = \beta_1.$$

The parameters α_0, α_1 , and σ_ϵ^2 remain unchanged. A key assumption in the following developments is that neither the estimates of the unknown parameters nor the true values fall on the boundary of the allowable parameter space.

The variance-covariance matrix of the m.l.e. for the vector of parameters $\mathbf{Y}$ is obtained by the first-order Taylor expansion at $\mathbf{Y}$. We also use the chain rule for derivatives of vector fields ([for details, see [14, Volume II, pages 269–275]).

Writing the vector of parameters $\mathbf{Y}$ as a function of the vector Θ , the respective first-order partial derivatives can be joined together in the following partitioned matrix:

$$\mathbf{D} = \left[\begin{array}{c|c} \mathbf{D}_1 & \mathbf{D}_2 \\ \hline \mathbf{D}_3 & \mathbf{D}_4 \end{array} \right], \quad (38)$$

where the (3×3) submatrix $\mathbf{D}_1$ corresponds to the first-order partial derivatives of the vector $\mathbf{Y}_1 \equiv \Theta_1 = [\alpha_0 \ \alpha_1 \ \sigma_\epsilon^2]'$ with respect to itself, which means that $\mathbf{D}_1$ is nothing but the identity matrix of order 3, $\mathbf{D}_1 = \mathbf{I}_3$. On the other hand, this statement also means that the derivatives of the parameters under consideration with respect to either $\psi_0, \psi_1, \psi_2, \psi_3$, or ψ_4 are zero. In other words, the (3×5) submatrix $\mathbf{D}_2$ is equal to the null vector, that is, $\mathbf{D}_2 = \mathbf{0}$.

The (5×3) submatrix $\mathbf{D}_3$ and the (5×5) submatrix $\mathbf{D}_4$ are composed by the first-order partial derivatives of each component of the vector of parameters $\mathbf{Y}_2 = [\beta_0 \ \beta_1 \ \beta_2 \ \sigma_\xi^2 \ \sigma_{\epsilon\xi}]'$ with respect to, respectively, $\alpha_0, \alpha_1, \sigma_\epsilon^2$ and $\psi_0, \psi_1, \psi_2, \psi_4, \psi_3$. Their structures are, thus, given by

$$\mathbf{D}_3 = \begin{bmatrix} \psi_1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \psi_1 & 0 \\ 0 & 0 & \psi_1^2 \\ 0 & 0 & \psi_1 \end{bmatrix}, \quad (39)$$

$$\mathbf{D}_4 = \left[\begin{array}{ccc|ccc} 1 & \alpha_0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & \alpha_1 & 1 & 0 & 0 & 0 \\ 0 & 2\psi_1\sigma_\epsilon^2 & 0 & 0 & 1 & 0 \\ 0 & \sigma_\epsilon^2 & 0 & 0 & 0 & 1 \end{array} \right] = \left[\begin{array}{c|c} \mathbf{D}_{41} & \mathbf{D}_{42} \\ \hline \mathbf{D}_{43} & \mathbf{D}_{44} \end{array} \right].$$

For finding out the approximate variance-covariance matrix of the maximum likelihood estimators for the unknown vector of parameters $\mathbf{Y}$, it is only necessary to pre- and postmultiply the variance-covariance matrix arising from expressions (29), (31), and (36) by, respectively, the matrix $\mathbf{D}$ and its transpose, $\mathbf{D}'$. More precisely,

$$\Sigma_{\mathbf{Y}} \approx \mathbf{D}\mathbf{I}^{-1}\mathbf{D}' = \left[\begin{array}{c|c} \mathbf{I}_3 & \mathbf{0}_{3 \times 5} \\ \hline \mathbf{D}_3 & \mathbf{D}_4 \end{array} \right] \left[\begin{array}{c|c} \mathbf{I}_1^{-1} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{I}_2^{-1} \end{array} \right] \left[\begin{array}{c|c} \mathbf{I}_3 & \mathbf{D}_3' \\ \hline \mathbf{0}_{5 \times 3} & \mathbf{D}_4' \end{array} \right]. \quad (40)$$

Hence,

$$\Sigma_{\mathbf{Y}} \approx \left[\begin{array}{c|c} \mathbf{I}_1^{-1} & \mathbf{I}_1^{-1}\mathbf{D}_3' \\ \hline (\mathbf{I}_1^{-1}\mathbf{D}_3')' & \mathbf{D}_3\mathbf{I}_1^{-1}\mathbf{D}_3' + \mathbf{D}_4\mathbf{I}_2^{-1}\mathbf{D}_4' \end{array} \right], \quad (41)$$

with $\Sigma_{\mathbf{Y}}$ denoting the variance-covariance matrix of the m.l.e. for the vector of unknown parameters $\mathbf{Y}$. A more detailed analysis of the variance-covariance matrix (41) can be found in Nunes [12, ch.3, p.91-92].

We can now deduce the approximate variance-covariance matrix of the maximum likelihood estimators for the mean vector and the variance-covariance matrix at lag zero of the VAR(1) Model with a monotone pattern of missingness, represented by $\Xi = [\alpha_0 \ \alpha_1 \ \sigma_\epsilon^2 \ \mu_X \ \mu_Y \ \sigma_X^2 \ \sigma_Y^2 \ \sigma_{XY}]'$. The first-order partial derivatives of the vector Ξ with respect to the vector $\mathbf{Y}$ are placed in a matrix that is denoted by $\mathbf{F}$. It takes the following form:

$$\mathbf{F} = \left[\begin{array}{c|c} \mathbf{F}_1 & \mathbf{F}_2 \\ \hline \mathbf{F}_3 & \mathbf{F}_4 \end{array} \right]. \quad (42)$$

According to the partition of the matrix $\mathbf{D}$ into four blocks—expression (38)—we partition the matrix $\mathbf{F}$ into the

following blocks: the (3×3) submatrix $\mathbf{F}_1$ corresponds to the partial derivatives of α_0, α_1 , and σ_ϵ^2 with respect to themselves. As a consequence, $\mathbf{F}_1$ is the identity matrix of order 3, that is, $\mathbf{F}_1 = \mathbf{I}_3$. Regards to the (3×5) sub-matrix $\mathbf{F}_2$, its elements correspond to the partial derivatives of α_0, α_1 , and σ_ϵ^2 with respect to $\beta_0, \beta_1, \beta_2, \sigma_\xi^2$, and $\sigma_{\epsilon\xi}$. Therefore, $\mathbf{F}_2 = \mathbf{0}$. The partial derivatives of $\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2$, and σ_{XY} with respect to α_0, α_1 , and σ_ϵ^2 are gathered together in the (5×3) sub-matrix $\mathbf{F}_3$:

$$\mathbf{F}_3 = \begin{bmatrix} f_{11}^3 & f_{12}^3 & 0 \\ f_{21}^3 & f_{22}^3 & 0 \\ 0 & f_{32}^3 & f_{33}^3 \\ 0 & f_{42}^3 & f_{43}^3 \\ 0 & f_{52}^3 & f_{53}^3 \end{bmatrix}, \quad (43)$$

where

$$\begin{aligned} f_{11}^3 &= \frac{1}{1 - \alpha_1}, & f_{12}^3 &= \frac{\alpha_0}{(1 - \alpha_1)^2}, \\ f_{21}^3 &= \frac{\beta_2}{(1 - \beta_1)(1 - \alpha_1)}, \\ f_{22}^3 &= \frac{\alpha_0\beta_2}{(1 - \beta_1)(1 - \alpha_1)^2}, & f_{32}^3 &= \frac{2\alpha_1\sigma_\epsilon^2}{(1 - \alpha_1^2)^2}, \\ f_{33}^3 &= \frac{1}{1 - \alpha_1^2}, \\ f_{42}^3 &= 2\beta_2 \left(\sigma_{\epsilon\xi}\beta_1^2(1 - \alpha_1^2)^2 \right. \\ &\quad \left. + \sigma_\epsilon^2\beta_2(\beta_1(1 - \alpha_1^2) + \alpha_1(1 - \alpha_1^2\beta_1^2)) \right) \\ &\quad \times \left(((1 - \alpha_1\beta_1)(1 - \alpha_1^2))^2(1 - \beta_1^2) \right)^{-1}, \\ f_{43}^3 &= \frac{\beta_2(1 + \alpha_1\beta_1)}{(1 - \alpha_1^2)(1 - \beta_1^2)(1 - \alpha_1\beta_1)}, \\ f_{52}^3 &= \frac{\sigma_{\epsilon\xi}\beta_1}{(1 - \alpha_1\beta_1)^2} + \sigma_\epsilon^2\beta_2 \frac{1 + \alpha_1^2(1 - 2\alpha_1\beta_1)}{(1 - \alpha_1\beta_1)^2(1 - \alpha_1^2)^2}, \\ f_{53}^3 &= \frac{\alpha_1\beta_2}{(1 - \alpha_1\beta_1)(1 - \alpha_1^2)}. \end{aligned} \quad (44)$$

The 5-dimensional square sub-matrix $\mathbf{F}_4$ corresponds to the partial derivatives of $\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2$, and σ_{XY} with respect to $\beta_0, \beta_1, \beta_2, \sigma_\xi^2, \sigma_{\epsilon\xi}$:

$$\mathbf{F}_4 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ f_{21}^4 & f_{22}^4 & f_{23}^4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & f_{42}^4 & f_{43}^4 & f_{44}^4 & f_{45}^4 \\ 0 & f_{52}^4 & f_{53}^4 & 0 & f_{55}^4 \end{bmatrix}, \quad (45)$$

with its nonnull elements taking the following analytical expressions:

$$\begin{aligned} f_{21}^4 &= \frac{1}{1 - \beta_1}, & f_{22}^4 &= \frac{\beta_0}{(1 - \beta_1)^2} + \frac{\alpha_0\beta_2}{(1 - \alpha_1)(1 - \beta_1)^2}, \\ f_{23}^4 &= \frac{\alpha_0}{(1 - \alpha_1)(1 - \beta_1)}, \\ f_{42}^4 &= \frac{2\sigma_\xi^2\beta_1}{(1 - \beta_1^2)^2} + \frac{2\sigma_{\epsilon\xi}\beta_2(1 + \beta_1^2(1 - 2\alpha_1\beta_1))}{(1 - \alpha_1\beta_1)^2(1 - \beta_1^2)^2} \\ &\quad + \frac{2\sigma_\epsilon^2\beta_2^2(\alpha_1(1 - \beta_1^2) + \beta_1(1 - \alpha_1^2\beta_1^2))}{(1 - \alpha_1^2)(1 - \beta_1^2)^2(1 - \alpha_1\beta_1)^2}, \\ f_{43}^4 &= \frac{2}{(1 - \alpha_1\beta_1)(1 - \beta_1^2)} \left\{ \beta_1\sigma_{\epsilon\xi} + \frac{\sigma_\epsilon^2\beta_2(1 + \alpha_1\beta_1)}{1 - \alpha_1^2} \right\}, \\ f_{44}^4 &= \frac{1}{1 - \beta_1^2}, & f_{45}^4 &= \frac{2\beta_1\beta_2}{(1 - \alpha_1\beta_1)(1 - \beta_1^2)}, \\ f_{52}^4 &= \frac{(1 - \alpha_1^2)\sigma_{\epsilon\xi}\alpha_1 + \sigma_\epsilon^2\alpha_1^2\beta_2}{(1 - \alpha_1^2)(1 - \alpha_1\beta_1)^2}, \\ f_{53}^4 &= \frac{\sigma_\epsilon^2\alpha_1}{(1 - \alpha_1\beta_1)(1 - \alpha_1^2)}, & f_{55}^4 &= \frac{1}{1 - \alpha_1\beta_1}. \end{aligned} \quad (46)$$

Straightforward calculations have paved the way to the desired partitioned variance-covariance matrix, called here Σ_Ξ ,

$$\Sigma_\Xi \approx \mathbf{F}\Sigma_Y\mathbf{F}' \approx \mathbf{F}\mathbf{D}\mathbf{I}^{-1}\mathbf{D}'\mathbf{F}' = \begin{bmatrix} \Sigma_\Xi^{11} & \Sigma_\Xi^{12} \\ \Sigma_\Xi^{21} & \Sigma_\Xi^{22} \end{bmatrix}, \quad (47)$$

with its submatrices defined by

$$\begin{aligned} \Sigma_\Xi^{11} &= \mathbf{I}_1^{-1}, & \Sigma_\Xi^{12} &= \mathbf{I}_1^{-1}(\mathbf{F}_3 + \mathbf{F}_4\mathbf{D}_3)' = \mathbf{I}_1^{-1}(\mathbf{F}_3' + \mathbf{D}_3'\mathbf{F}_4'), \\ \Sigma_\Xi^{21} &= (\mathbf{F}_3 + \mathbf{F}_4\mathbf{D}_3)\mathbf{I}_1^{-1} = (\Sigma_F^{12})', \\ \Sigma_\Xi^{22} &= \mathbf{F}_3\mathbf{I}_1^{-1}\mathbf{F}_3' + \mathbf{F}_4\mathbf{D}_3\mathbf{I}_1^{-1}\mathbf{F}_3' + \mathbf{F}_3\mathbf{I}_1^{-1}\mathbf{D}_3'\mathbf{F}_4' \\ &\quad + \mathbf{F}_4(\mathbf{D}_3\mathbf{I}_1^{-1}\mathbf{D}_3' + \mathbf{D}_4\mathbf{I}_2^{-1}\mathbf{D}_4')\mathbf{F}_4' \\ &= (\mathbf{F}_3 + \mathbf{F}_4\mathbf{D}_3)\mathbf{I}_1^{-1}(\mathbf{F}_3 + \mathbf{F}_4\mathbf{D}_3)' + \mathbf{F}_4\mathbf{D}_4\mathbf{I}_2^{-1}(\mathbf{F}_4\mathbf{D}_4)' \\ &= \mathbf{G}\mathbf{I}_1^{-1}\mathbf{G}' + \mathbf{H}\mathbf{I}_2^{-1}\mathbf{H}'. \end{aligned} \quad (48)$$

The matrix $\mathbf{G}$ that has just been defined as $\mathbf{G} = \mathbf{F}_3 + \mathbf{F}_4\mathbf{D}_3$ corresponds to the first-order partial derivatives from the composite functions that relate $\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2$, and σ_{XY} with the vector of parameters Θ_1 . The elements of the matrix $\mathbf{H} = \mathbf{F}_4\mathbf{D}_4$ are the first-order partial derivatives from the composite functions that relate $\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2$, and σ_{XY} with the vector of unknown parameters Θ_2 .

The 3-dimensional square sub-matrix Σ_{Ξ}^{11} corresponds to the approximate covariance structure between the m.l.e. of the parameters α_0, α_1 , and σ_{ϵ}^2 . The (3×5) sub-matrix Σ_{Ξ}^{12} is composed of the approximate covariances between the m.l.e. that have just been cited and $\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2$, and σ_{XY} ; its transpose is denoted by Σ_{Ξ}^{21} . This is the reason why Σ_{Ξ}^{12} , or Σ_{Ξ}^{21} , results from the product of the variance-covariance matrix I_1^{-1} and $\mathbf{G}$. The 5-dimensional square sub-matrix Σ_{Ξ}^{22} is formed by the covariances between the m.l.e. for $\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2$, and σ_{XY} .

The main point of the section is to study the variances and covariances that take part of the sub-matrix Σ_{Ξ}^{22} . Thus, it

is of interest to further explore its analytical expression. The matrix $\mathbf{G}$ takes a cumbersome form. The most efficient way to deal with it is to consider its partition rather than the whole matrix at once.

Let

$$\mathbf{G} = \mathbf{F}_3 + \mathbf{F}_4 \mathbf{D}_3 = \left[\begin{array}{c|c} \mathbf{G}_{11} & \mathbf{G}_{12} \\ \hline \mathbf{G}_{21} & \mathbf{G}_{22} \end{array} \right], \quad (49)$$

where the (4×2) sub-matrix $\mathbf{G}_{11}$ takes the form

$$\mathbf{G}_{11} = \begin{bmatrix} \frac{1}{1 - \alpha_1} & \frac{\alpha_0}{(1 - \alpha_1)^2} \\ \frac{\psi_1}{1 - \beta_1} + \frac{\beta_2}{(1 - \alpha_1)(1 - \beta_1)} & \frac{\alpha_0}{(1 - \alpha_1)(1 - \beta_1)} \left(\psi_1 + \frac{\beta_2}{1 - \alpha_1} \right) \\ 0 & \frac{2\alpha_1\sigma_{\epsilon}^2}{(1 - \alpha_1^2)^2} \\ 0 & g_{42}^{11} \end{bmatrix}, \quad (50)$$

with

$$\begin{aligned} g_{42}^{11} = & \psi_1 f_{43}^4 + \frac{2\beta_2}{(1 - \beta_1^2)(1 - \alpha_1\beta_1)^2(1 - \alpha_1^2)^2} \\ & \times \left(\sigma_{\epsilon\xi}\beta_1^2(1 - \alpha_1^2)^2 + \sigma_{\epsilon}^2\beta_2(\beta_1(1 - \alpha_1^2) \right. \\ & \left. + \alpha_1(1 - \alpha_1^2\beta_1^2)) \right), \end{aligned} \quad (51)$$

where f_{43}^4 is defined by (45).

The 4-dimensional column vector $\mathbf{G}_{12}$, the 2-dimensional row vector $\mathbf{G}_{21}$ and the scalar G_{22} are, respectively, given by

$$\begin{aligned} \mathbf{G}_{12} = & \left[\begin{array}{cc} 0 & 0 \end{array} \frac{1}{1 - \alpha_1^2} \frac{\psi_1^2}{1 - \beta_1^2} + \frac{\beta_2}{(1 - \alpha_1\beta_1)(1 - \beta_1^2)} \right. \\ & \left. \times \left(2\psi_1\beta_1 + \frac{\beta_2(1 + \alpha_1\beta_1)}{1 - \alpha_1^2} \right) \right]', \\ \mathbf{G}_{21} = & \left[\begin{array}{c} 0 \\ \frac{1}{(1 - \alpha_1\beta_1)} \end{array} \right. \\ & \times \left\{ \frac{\sigma_{\epsilon\xi}\beta_1}{1 - \alpha_1\beta_1} + \frac{\sigma_{\epsilon}^2}{1 - \alpha_1^2} \right. \\ & \left. \times \left(\psi_1\alpha_1 + \frac{\beta_2(1 + \alpha_1^2(1 - 2\alpha_1\beta_1))}{(1 - \alpha_1^2)(1 - \alpha_1\beta_1)} \right) \right\} \left. \right], \\ G_{22} = & \frac{1}{1 - \alpha_1\beta_1} \left(\psi_1 + \frac{\alpha_1\beta_2}{1 - \alpha_1^2} \right). \end{aligned} \quad (52)$$

On the other hand, we can also make the following partition of the matrix $\mathbf{H} = \mathbf{F}_4 \mathbf{D}_4$:

$$\mathbf{H} = \left[\begin{array}{c|c} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \hline \mathbf{H}_{21} & H_{22} \end{array} \right], \quad (53)$$

where the sub-matrix $\mathbf{H}_{11}$ corresponds to the first order partial derivatives of the vector $[\mu_X \ \mu_Y \ \sigma_X^2 \ \sigma_Y^2]'$ with respect to the vector $[\psi_0 \ \psi_1 \ \psi_2 \ \psi_4]'$, whereas their derivatives with respect to the parameter ψ_3 constitute the sub-matrix $\mathbf{H}_{12}$. The sub-matrix $\mathbf{H}_{21}$ is composed of the first order partial derivatives of σ_{XY} with respect to each component of the vector $[\psi_0 \ \psi_1 \ \psi_2 \ \psi_4]'$. Finally, the scalar $H_{22} = \partial\sigma_{XY}/\partial\psi_3 = 0$.

The desired variance-covariance matrix can therefore be written in the following partitioned form:

$$\Sigma_{\Xi}^{22} = \left[\begin{array}{c|c} \Sigma_1^{22} & \Sigma_2^{22} \\ \hline \Sigma_3^{22} & \Sigma_4^{22} \end{array} \right], \quad (54)$$

with

$$\begin{aligned} \Sigma_1^{22} = & \hat{\sigma}_{\epsilon}^2 \mathbf{G}_{11} (\mathbf{U}'_R \mathbf{U}_R)^{-1} \mathbf{G}'_{11} + \frac{2\hat{\sigma}_{\epsilon}^4}{n - 1} \mathbf{G}_{12} \mathbf{G}'_{12} \\ & + \psi_1 \mathbf{H}_{11} (\mathbf{U}' \mathbf{U})^{-1} \mathbf{H}'_{11} \\ & + \frac{2\psi_3^2}{m - 1} \mathbf{H}_{12} \mathbf{H}'_{12}, \\ \Sigma_2^{22} = & \hat{\sigma}_{\epsilon}^2 \mathbf{G}_{11} (\mathbf{U}'_R \mathbf{U}_R)^{-1} \mathbf{G}'_{21} + \frac{2\hat{\sigma}_{\epsilon}^4}{n - 1} \mathbf{G}_{12} \mathbf{G}'_{22} \\ & + \psi_3 \mathbf{H}_{11} (\mathbf{U}' \mathbf{U})^{-1} \mathbf{H}'_{21}, \end{aligned}$$

$$\begin{aligned}
\Sigma_3^{22} &= \hat{\sigma}_\epsilon^2 \mathbf{G}_{21} (\mathbf{U}'_R \mathbf{U}_R)^{-1} \mathbf{G}'_{11} \\
&\quad + \frac{2\hat{\sigma}_\epsilon^4}{n-1} \mathbf{G}_{22} \mathbf{G}'_{12} + \psi_3 \mathbf{H}_{21} (\mathbf{U}' \mathbf{U})^{-1} \mathbf{H}'_{11}, \\
\Sigma_4^{22} &= \hat{\sigma}_\epsilon^2 \mathbf{G}_{21} (\mathbf{U}'_R \mathbf{U}_R)^{-1} \mathbf{G}'_{21} \\
&\quad + \frac{2\hat{\sigma}_\epsilon^4}{n-1} \mathbf{G}_{22} \mathbf{G}'_{22} + \psi_3 \mathbf{H}_{21} (\mathbf{U}' \mathbf{U})^{-1} \mathbf{H}'_{21},
\end{aligned} \tag{55}$$

where the matrix $\mathbf{U}$ is defined by (35). The matrix $\mathbf{U}_R$ takes the form

$$\mathbf{U}_R = \begin{bmatrix} 1 & X_0 \\ 1 & X_1 \\ \vdots & \vdots \\ 1 & X_{m-2} \end{bmatrix}. \tag{56}$$

In short, the matrix defined by (54) corresponds to the approximate variance-covariance matrix of the m.l.e. for the mean vector and variance-covariance matrix at lag zero for the VAR(1) Model with missing data. We cannot write down explicit expressions for those variances and covariances. The limitation arises from the inability to invert the matrix product $\mathbf{U}' \mathbf{U}$ in analytical terms (see (36)). Hence, its inverse can only be accomplished by numerical techniques using the observed sampled data. This point will be pursued further in Section 5.

Despite the above restrictions, several investigations can be done regarding the amount of additional information obtained by making full use of the fragmentary data available. The strength of the correlation between the stochastic processes here plays a crucial role. These ideas will be developed in Section 5.

5. Simulation Studies

In this section, we analyse the effects of using different strategies to estimate the mean value of the stochastic process $\{Y_t, t \in \mathbb{Z}\}$, denoted by μ_Y . More precisely, the bivariate modelling scheme and its univariate counterparts are compared. Simulation studies are carried out to evaluate the relative efficiency of the estimators with interest.

The m.l.e. of the mean value of the stochastic process $\{Y_t, t \in \mathbb{Z}\}$ based on the VAR(1) Model is obtained by the second equation of the system of (28). We need to compare this estimator to those obtained by considering the univariate stochastic process $\{Y_t, t \in \mathbb{Z}\}$ itself. More precisely, having in mind that we are handling a bivariate VAR(1) Model, the corresponding marginal model is the ARMA(2,1) [10, 11]. On the other hand, the AR(1) Model is one of the most popular models due to its practical importance in time series modelling. Therefore, the behaviour of the AR(1) Model will be also evaluated. In short, we will compare the performance of the VAR(1) Model with both the ARMA(2,1) and the AR(1) Models.

To avoid any confusion between the parameters coming from the bivariate and the univariate modelling strategies,

from now on we denote the parameter from the VAR(1) Model by μ_{VAR} , whereas those from the ARMA(2,1) and the AR(1) Models are represented by μ_{ARMA} and μ_{AR} , respectively.

The bivariate VAR(1) Model is described by the system of (1). Thus, the univariate stochastic process $\{Y_t, t \in \mathbb{Z}\}$ follows an ARMA(2,1) Model, and the m.l.e. of the mean value are given by

$$\hat{\mu}_{\text{ARMA}} = \frac{\hat{\beta}_0}{1 - \hat{\alpha}_1 (1 - \hat{\beta}_1) - \hat{\beta}_1}. \tag{57}$$

On the other hand, if we assumed that $\{Y_t, t \in \mathbb{Z}\}$ followed an AR(1) Model, the m.l.e. of the mean value would be given by

$$\hat{\mu}_{\text{AR}} = \frac{\hat{\beta}_0}{1 - \hat{\beta}_1}. \tag{58}$$

Next, we will compare the performance of the estimators (57) and (58) with the m.l.e. based on the VAR(1) Model (second equation of the system (28)). It is important to stress that the strategy behind the AR(1) Model has not taken into account the relationship between the stochastic processes $\{X_t, t \in \mathbb{Z}\}$ and $\{Y_t, t \in \mathbb{Z}\}$. This feature will certainly introduce an additional noise in the overall estimation procedure.

Following the techniques used in Section 4.2 for determining the precision of the estimators under consideration, here we also have used the first-order Taylor expansion at the mean value μ_Y for computing the estimate of the variance of μ_Y .

Considering the ARMA(2,1) Model, let $\boldsymbol{\theta} = [\beta_0 \ \beta_1 \ \alpha_1]'$ be the vector of the unknown parameters. Then,

$$\begin{aligned}
\text{Var}(\hat{\mu}_{\text{ARMA}}) &\approx \sum_{i=1}^3 \left(\frac{\partial \hat{\mu}_{\text{ARMA}}}{\partial \theta_i} \Big|_{\theta_i = \hat{\theta}_i} \right)^2 \text{Var}(\hat{\theta}_i) \\
&\quad + 2 \sum_{i=1}^3 \sum_{j=i+1}^3 \frac{\partial \hat{\mu}_{\text{ARMA}}}{\partial \theta_i} \Big|_{\theta_i = \hat{\theta}_i} \\
&\quad \times \frac{\partial \hat{\mu}_{\text{ARMA}}}{\partial \theta_j} \Big|_{\theta_j = \hat{\theta}_j} \text{Cov}(\hat{\theta}_i, \hat{\theta}_j).
\end{aligned} \tag{59}$$

In regard to the AR(1) Model, $\hat{\mu}_{\text{AR}}$ is given by (58) and

$$\begin{aligned}
\text{Var}(\hat{\mu}_{\text{AR}}) &\approx \frac{2\hat{\beta}_0}{(1 - \hat{\beta}_1)^3} \text{Cov}(\hat{\beta}_0, \hat{\beta}_1) \\
&\quad + \frac{\text{Var}(\hat{\beta}_0)}{(1 - \hat{\beta}_1)^2} + \frac{\hat{\beta}_0^2 \text{Var}(\hat{\beta}_1)}{(1 - \hat{\beta}_1)^4}.
\end{aligned} \tag{60}$$

Improvements in choosing the sophisticated m.l.e. for μ_Y based on the VAR(1) Model rather than considering

FIGURE 1: Graphical representation of GI_1 . The data were obtained from a VAR(1) Model, with $\alpha_0 = \beta_0 = 0$, $\alpha_1 = 0.6$, $\beta_1 = 0.7$, and $\beta_2 = 0.8$.

its univariate counterparts are next discussed. Simulation studies are carried out to evaluate the relative efficiency of the estimators under consideration.

The data were generated by the VAR(1) Model (system of (1)). In order to make comparisons on the same basis, a few assumptions to the parameters of the VAR(1) Model are made. We consider that $\mu_X = \mu_Y = 0$. These restrictions have no influence on the results because they are equivalent to $\alpha_0 = \beta_0 = 0$, that is, the constant terms of the VAR(1) Model are equal to zero (system of (1)). Additionally, we introduce the restriction $\sigma_\epsilon^2 = \sigma_\xi^2 = 1$.

Since the correlation coefficient regulates the supply of information between the stochastic processes $\{X_t\}_{t \in \mathbb{Z}}$ and $\{Y_t\}_{t \in \mathbb{Z}}$, particular emphasis is given to this parameter. Using the grid of points $\rho_{\epsilon\xi} = 0.1, 0.5, 0.75, 0.9$, the Gain index is computed. We stress that the value $\rho_{\epsilon\xi} = 1$ is not allowable in this context (see Section 2 for the details).

We analyse the performance of the estimators based on different sample sizes, $n = 50, 100, 250$, and 500 . The simulations reported next are based on different percentages of missing observations referred to the dimension of the sampled data from the auxiliary random process $\{X_t\}_{t \in \mathbb{Z}}$.

Simulation runs for each combination of the parameters are based on 1000 replicates.

It is worth emphasising that the estimates of the covariance terms that take part of the variances given by (59) and (60) were computed by the R package tseries [15].

The simulation goes as follows: after each simulation run, the relative efficiency of $\hat{\mu}_{VAR}$ with respect to each estimator $\hat{\mu}_{ARMA}$ and $\hat{\mu}_{AR}$ is quantified by the Gain index, GI_1 and GI_2 , respectively, expressed as percentage:

$$GI_1 = \frac{\text{Var}(\hat{\mu}_{ARMA}) - \text{Var}(\hat{\mu}_{VAR})}{\text{Var}(\hat{\mu}_{ARMA})} \times 100\%,$$

$$GI_2 = \frac{\text{Var}(\hat{\mu}_{AR}) - \text{Var}(\hat{\mu}_{VAR})}{\text{Var}(\hat{\mu}_{AR})} \times 100\%.$$
(61)

A word of notation: the above quantities, that is, GI_1 and GI_2 , were computed from the estimates of the corresponding variances. To lighten the notation, we skipped the conventional nomenclature used to represent the estimates.

If $GI_1 > 0$, then $\hat{\mu}_{VAR}$ is more precise than $\hat{\mu}_{ARMA}$. Otherwise, $\hat{\mu}_{VAR}$ loses precision, and $\hat{\mu}_{ARMA}$ becomes a better

FIGURE 2: Graphical representation of GI_2 . The data were obtained from a VAR(1) Model, with $\alpha_0 = \beta_0 = 0$, $\alpha_1 = 0.6$, $\beta_1 = 0.7$, and $\beta_2 = 0.8$.

estimator for the mean value of $\{Y_t\}_{t \in \mathbb{Z}}$. A similar reasoning applies to the comparison between $\hat{\mu}_{VAR}$ and $\hat{\mu}_{AR}$.

Figures 1 and 2 display the main results from the simulation studies. The estimators $\hat{\mu}_{VAR}$ and $\hat{\mu}_{ARMA}$ are compared in Figure 1, whereas Figure 2 exhibits the comparison between $\hat{\mu}_{VAR}$ and $\hat{\mu}_{AR}$. For each combination of the parameters of the model, we represent graphically the gain indexes as functions of the percentage of missing data in the sampled data from the stochastic process $\{Y_t\}_{t \in \mathbb{Z}}$.

Either Figures 1 or 2 shows that the plot of the gain index against the percentage of missing data in the sample from the stochastic process $\{Y_t\}_{t \in \mathbb{Z}}$ behaves roughly as a linear function, regardless of the combination of the parameters. In outline, the more the percentage of missing values in the sampled data is, the more precise is the estimator $\hat{\mu}_{VAR}$ when compared with the univariate context, that is, $\hat{\mu}_{ARMA}$ or $\hat{\mu}_{AR}$ (see Figures 1 and 2).

Further, the gain in precision by using the sophisticated estimator $\hat{\mu}_{VAR}$ rather than $\hat{\mu}_{ARMA}$ or $\hat{\mu}_{AR}$ increases as the strength of the linear relationship between the processes $\{X_t\}_{t \in \mathbb{Z}}$ and $\{Y_t\}_{t \in \mathbb{Z}}$ (described by the correlation coefficient) rises from $\rho = 0.1$ to $\rho = 0.9$. This statement is true for both the ARMA(2,1) and AR(1) modelling schemes (see Figures 1 and 2).

A final point to highlight from the comparison between Figures 1 and 2 is that the increase in precision obtained by using the estimator for the mean value of $\{Y_t\}_{t \in \mathbb{Z}}$ based on the VAR(1) Model is higher when we compare its performance with the results from the AR(1) Model than when we compare the VAR(1) Model with the ARMA(2,1) Model. This feature emphasises the idea that has already been raised that the ARMA(2,1) Model describes more accurately the dynamics of the stochastic process $\{Y_t\}_{t \in \mathbb{Z}}$ than the AR(1) Model does. In short, it seems that the AR(1) Model is not a good approach in this context because it incorporates a noise term related to the simulation scheme that we cannot control.

Summing up, the estimator $\hat{\mu}_{VAR}$ is preferable to those explored in the univariate context, that is, either $\hat{\mu}_{ARMA}$ or $\hat{\mu}_{AR}$.

6. Conclusions

This article deals with the problem of missing data in an univariate sample. We have considered an auxiliary complete data set, whose underlying stochastic process is serially correlated with the former by the VAR(1) Model structure. We have proposed maximum likelihood estimators for the relevant parameters of the model based on a monotone

missing data pattern. The precision of the estimators has also been derived. Special attention has been given to the estimator for the mean value of the stochastic process whose sampled data has missing values, μ_Y .

We have compared the performance of the estimator for μ_Y based on the VAR(1) Model with a monotone pattern of missing data with those obtained from both the ARMA(2,1) Model and the AR(1) Model. By simulation studies, we have showed that the estimator derived in this article based on the VAR(1) Model performs better than those derived from the univariate context. It is essential to emphasise that, even numerically, it was quite difficult to compute the precision of the later estimators as we have shown in Section 4.2.

A compelling question remains unresolved. From an applied point of view, it would be extremely useful to develop estimators for the dynamics of the stochastic processes. More precisely, we would like to get estimators for the correlation and cross-correlation matrices as well as their precision when there are missing observations in one of the data sets. It was not possible to achieve this goal based on maximum likelihood principles. As we have shown in Section 4.2, we have only developed estimators for the covariance matrix at lag zero. In future research, we will try to solve this problem in the framework of Kalman filter.

Acknowledgments

This work was financed by the Portuguese Foundation for Science and Technology (FCT), Projecto Estratégico PEst-OE/MAT/UI0209/2011. The authors are also thankful for the comments of the two anonymous referees.

References

- [1] D. F. Morrison, "Expectations and variances of maximum likelihood estimates of the multivariate normal distribution parameters with missing data," *Journal of the American Statistical Association*, vol. 66, no. 335, pp. 602–604, 1971.
- [2] R. C. Dahiya and R. M. Korwar, "Maximum likelihood estimates for a bivariate normal distribution with missing data," *The Annals of Statistics*, vol. 8, no. 3, pp. 687–692, 1980.
- [3] V. Gómez and A. Maravall, "Estimation, prediction, and interpolation for nonstationary series with the Kalman filter," *Journal of the American Statistical Association*, vol. 89, no. 426, pp. 611–624, 1994.
- [4] R. H. Jones, "Maximum likelihood fitting of ARMA models to time series with missing observations," *Technometrics*, vol. 22, no. 3, pp. 389–395, 1980.
- [5] R. Kohn and C. F. Ansley, "Estimation, prediction, and interpolation for ARIMA models with missing data," *Journal of the American Statistical Association*, vol. 81, no. 395, pp. 751–761, 1986.
- [6] M. Pourahmadi, "Estimation and interpolation of missing values of a stationary time series," *Journal of Time Series Analysis*, vol. 10, no. 2, pp. 149–169, 1989.
- [7] V. Gómez, A. Maravall, and D. Peña, "Missing observations in ARIMA models: skipping approach versus additive outlier approach," *Journal of Econometrics*, vol. 88, no. 2, pp. 341–363, 1999.
- [8] R. J. A. Little and D. B. Rubin, *Statistical Analysis with Missing Data*, John Wiley & Sons, New York, NY, USA, 1987.
- [9] R. Sparks, "SUR models applied to an environmental situation with missing data and censored values," *Journal of Applied Mathematics and Decision Sciences*, vol. 8, no. 1, pp. 15–32, 2004.
- [10] C. Heij, P. De Boer, P. H. Franses, T. Kloek, and H. K. van Dijk, *Econometric Methods with Applications in Business and Economics*, Oxford University Press, New York, NY, USA, 2004.
- [11] R. S. Tsay, *Analysis of Financial Time Series*, John Wiley & Sons, Hoboken, NJ, USA, 3rd edition, 2010.
- [12] M. H. Nunes, *Dynamics relating phytoplankton abundance with upwelling events. An approach to the problem of missing data in the gaussian context [Ph.D. thesis]*, University of Lisbon, Lisbon, Portugal, 2006.
- [13] J. D. Hamilton, *Time Series Analysis*, Princeton University Press, Princeton, NJ, USA, 1994.
- [14] T. M. Apostol, *Calculus*, John Wiley & Sons, Singapore, 2nd edition, 1969.
- [15] A. Trapletti and K. Hornik, "tseries: Time series analysis and computational finance," R package version 0.10-25, 2011, <http://CRAN.R-project.org/package=tseries>.