# Long memory in the marginalized time series of a VAR revisited.

Tomás del Barrio Castro (UIB)*    Andreu Sansó Rosselló (UIB)
Philipp Sibbertsen (Leibniz University)

December 1, 2025

**Abstract**

In this paper we find an alternative explanation for the presence of long memory in marginalized time series of an autoregressive system in situations earlier explored by Bauwens, Chevillon and Laurent (2023) and Chevillon, Hecq, and Laurent (2018) which is the near cancellation of the damped trend shared by all the time series of the VAR(1) used in these papers and the MA(1) part of the data generating process followed by the marginalized time series. For a given time dimension $T$ the long memory observed in the marginalized time series will depend on the number of time series in the VAR(1) system but not on the specific value of the main diagonal associated with the matrix of coefficients of the VAR(1) as stated in Chevillon, Hecq, and Laurent (2018) and Bauwens, Chevillon and Laurent (2023). Our results are based on the properties of circulant matrices and the Vector Moving Average representation of the VAR(1) model proposed in the previous two papers. Finally a Monte-Carlo experiment supports our analytical findings.

**Keywords:** Long Memory, Marginalized Time Series, Damped Trend.

**JEL codes:** C3

# 1 Introduction

In the recent literature at least three relevant papers focused on situations where it is possible to find long memory in the individual or marginalized time series that belong to a high dimensional VAR(1) system, see Chevillon, Hecq, and Laurent (2018), Schennach (2018) and Bauwens, Chevillon and Laurent (2023). Bauwens, Chevillon and Laurent (2023) is of special interest as it provides a simple set-up to understand how long-memory could arise or be detected in a marginalized time series which is part of a high dimesional VAR(1) system.

In this paper we revisit the Bauwens, Chevillon and Laurent (2023) approach, in particular their section 2, where a simple and easy to handle definition of the $n \times n$ matrix of coefficients $\mathbf{A}_n$ of the VAR(1) system with $n$ time series is proposed. Bauwens, Chevillon and Laurent (2023) claim that their set-up is general enough to cover the previous works of Chevillon, Hecq, and Laurent (2018) and Schennach (2018). Using their framework, we are able to show that contrary to this the long memory behavior is not connected to the main diagonal value of the matrix $\mathbf{A}_n$ (denoted by $d_0$ in these two previous works) but on the relation between the number of time series $n$ in the VAR(1) system and the time dimension of the VAR(1) system $T$. That is, for a given $T$ the larger the number of time series in the VAR(1) $n$ the lower is the value of the long memory detected in the marginalized time series. In addition we show that $d_0$ (the main diagonal element of the matrix $\mathbf{A}_n$) is not connected to the long memory observed in each marginalize time series of the VAR(1) but with the AR(1) part of the data generating process associated with the marginalized time series, see (16) below for more details. Finally, from the set-up proposed in Bauwens, Chevillon and Laurent (2023) it is possible to show that the long memory associated to each time series of the VAR(1) system is caused by the presence of a damped trend shared by all the time series of the VAR(1) system, and hence it is possible to see that there is cointegration between every two time series of the VAR(1) system with cointegration vector $[1, -1]$.

The organization of the paper is as follows, in the following section 2 we introduce the set-up of Bauwens, Chevillon and Laurent (2023) and Chevillon, Hecq, and Laurent (2018), then present our analytical results

*We are in dept for constructive comments and discussions from Guillaume Chevillon and Alain Hecq. Financial support by Grant "PID2023-150095NB-C43" by MICIU/AEI/ 10.13039/501100011033 and, as appropriate, by "ERDF A way of making Europe", by "ERDF/EU", by the "European Union" or by the "European Union NextGenerationEU/PRTR. AEI is gratefully acknowledged.

based on the properties of circulant matrices and the resulting Vector Moving Average representation of the VAR(1) model. In this section we also pay attention to example 1 in Chevillon, Hecq and Laurent (2018) where the $n \times n$ matrix $\mathbf{A}_n$ of coefficients of the VAR(1) model is defined as a Toeplitz matrix. In a Monte-Carlo section 3 we illustrate the findings of section 2 in an extensive simulation experiment. Finally section 4 concludes. All proofs are gathered in the Appendix.

# 2 Zero Frequency Long-Memory with a VAR(1).

In Chevillon, Hecq, and Laurent (2018) –CHL hereafter– and Bauwens, Chevillon and Laurent (2023) –BCL hereafter– (see also Schennach (2018)) it was shown that the zero frequency long-memory observed in a univariate time series could be the result of the marginalization of a large VAR(1) system that satisfies certain specific assumptions. Denote by $n$ the number of time series in the VAR(1) system and $t$ is the time dimension of the system, hence $T$ is the total number of time observations. The $n$ time series in the VAR(1) system are collected in the $n \times 1$ vector $\mathbf{y}_{n,t} = [y_{1,t}, y_{2,t}, \ldots, y_{n,t}]'$ so that we have the following VAR(1) model:

$$\begin{aligned} \mathbf{y}_{n,t} &= \mathbf{A}_n \mathbf{y}_{n,t-1} + \epsilon_{n,t} \\ (I_n - \mathbf{A}_n L)\, \mathbf{y}_{n,t} &= \epsilon_{n,t} \\ t &= 1, 2, \ldots, T, \end{aligned} \tag{1}$$

where $\epsilon_{n,t}$ is a $n \times 1$ vector of identically and independently distributed innovations with zero expectation and variance–covariance matrix $\Sigma_n$. This variance-covariance matrix $\Sigma_n$ could be diagonal but this is not a necessary condition. We denote by $L$ the lag operator[1] and the matrix polynomial in the lag operator $(I_n - \mathbf{A}_n L)$ as $(I_n - \mathbf{A}_n L) = \mathbf{A}_n^0 (L)$. Finally $\mathbf{A}_n$ is a $n \times n$ matrix of coefficients. This matrix was defined by CHL as a square matrix $\mathbf{A}_n = \mathbf{T}_n^*$ where $\mathbf{T}_n^*$ is a Toeplitz matrix:

$$\mathbf{T}_n^* = \begin{bmatrix} t_0^n & t_1^n & t_2^n & \cdots & t_{n-1}^n \\ t_{-1}^n & t_0^n & t_1^n & \cdots & t_{n-2}^n \\ t_{-2}^n & t_{-1}^n & t_0^n & \cdots & t_{n-3}^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t_{-(n-1)}^n & t_{-(n-2)}^n & t_{-(n-3)}^n & \cdots & t_0^n \end{bmatrix}. \tag{2}$$

Toeplitz matrices are analyzed in detail by Grenander and Szegö (1958). A more accessible approach to the connection between Toeplitz matrices and forms to circulant matrices can be found in Gray (2006).

The proof of CHL relies on the use of the Final Equation Representation (FER) of (1), which is (see Zellner and Palm (1974)):

$$\det\left((I_n - \mathbf{A}_n L)\right) \mathbf{y}_{n,t} = adj\,(I_n - \mathbf{A}_n L)\, \epsilon_{n,t},$$

where $\det(\mathbf{X})$ and $adj\,(\mathbf{X})$ denote the determinant and the adjugate of the square matrix $\mathbf{X}$. For a marginalized time series of the VAR(1) system, say $y_{i,t}$, we have:

$$\begin{aligned} \det(I_n - \mathbf{A}_n L)\, y_{i,t} &= \sum_{j=1}^{n} adj\,(I_n - \mathbf{A}_n L)_{ij}\, \epsilon_{j,t} \\ y_{i,t} &= \det(I_n - \mathbf{A}_n L)^{-1} \sum_{j=1}^{n} adj\,(I_n - \mathbf{A}_n L)_{ij}\, \epsilon_{j,t}, \end{aligned} \tag{3}$$

where $adj\,(I_n - \mathbf{A}_n L)_{ij}$ denotes the element in the $ith$ row and $jth$ column of the adjugate matrix $adj\,(I_n - \mathbf{A}_n L)$. At this point, we can re-write (3) as:

$$\begin{aligned} y_{i,t} &= \det(I_n - \mathbf{A}_n L)^{-1}\, adj\,(I_n - \mathbf{A}_n L)_{ii}\, \epsilon_{it} \\ &+ \det(I_n - \mathbf{A}_n L)^{-1} \sum_{k=1,\ k\neq i}^{n} adj\,(I_n - \mathbf{A}_n L)_{ik}\, \epsilon_{k,t}. \end{aligned} \tag{4}$$

[1]Here and all over the paper, the lag operator $L$ will work for vectors and also for scalars.

Hence, we have two parts, one associated with the innovation of the marginalized time series $\det\left(I_n - \mathbf{A}_n L\right)^{-1}$ $adj\left(I_n - \mathbf{A}_n L\right)_{ii}\epsilon_{it}$ and the other part associated with the innovations associated with the rest of the time series included in the VAR(1) system. In this step CHL introduce their assumption P. This assumption has four parts, but the more relevant part for the present paper is P(i) (part (i) of the assumption)[2]. Basically, as CHL stated "P(i) ensures that in the moving average representation of $y_{i,t}$ the only innovation that retains a contribution as $n \to \infty$ is $\epsilon_{i,t}$, the others play no role asymptotically" and hence have $y_{i,t} = \det\left(I_n - \mathbf{A}_n L\right)^{-1} adj\left(I_n - \mathbf{A}_n L\right)_{ii}\epsilon_{i,t} + o_p(1)$[3].

Once CHL move from $y_{i,t} = \det\left(I_n - \mathbf{A}_n L\right)^{-1} \sum_{j=1}^{n} adj\left(I_n - \mathbf{A}_n L\right)_{ij}\epsilon_{j,t}$ to $y_{i,t} = \det\left(I_n - \mathbf{A}_n L\right)^{-1} adj\left(I_n - \mathbf{A}_n L\right)_{ii}\epsilon_{i,t} + o_p(1)$ as the representation of the marginalized time series of the VAR(1) system (1), the following step is to use Szegö's Theorem (see Grenander and Szegö (1958) and Gray (2006) for more details). This theorem requires the existence of the Fourier series with coefficients $t_k^n$ related to each other by:

$$f(\lambda) = \sum_{k=-\infty}^{\infty} t_k^n e^{ik\lambda} \quad \lambda \in [0, 2\pi]$$

$$t_k^n = \frac{1}{2\pi}\int_0^{2\pi} f(\lambda)\, e^{-ik\lambda} d\lambda.$$

Hence, the sequence $t_k^n$ determines the function $f(\lambda)$ and vice versa. The coefficients $t_k^n$are the elements of the matrix $\mathbf{A}_n = \mathbf{T}_n^*$ defined as (2) for the VAR(1) model (1). CHL adapt Szegö's Theorem to their case (Lemma 1 in CHL) as:

$$\frac{\det\left(I_n - \mathbf{A}_{n-1}L\right)}{\det\left(I_n - \mathbf{A}_n L\right)} \to \frac{1}{\exp\left\{\frac{1}{2\pi}\int_0^{2\pi}\log\left(1 - f\left(e^{i\omega}\right)z\right)d\omega\right\}}. \tag{5}$$

In their Assumption T part (i.c) they claim that:

$$\frac{1}{2\pi}\int_0^{2\pi}\log\left(1 - f\left(e^{i\omega}\right)z\right)d\omega = \delta\log\left(1 - z\right), \tag{6}$$

where $\delta$ is defined in (19) below, but basically is connected to the element in the main diagonal of the matrix $\mathbf{A}_n = \mathbf{T}_n^*$. Hence, assuming (6) they were able to write:

$$\frac{\det\left(I_n - \mathbf{A}_{n-1}L\right)}{\det\left(I_n - \mathbf{A}_n L\right)} \to (1 - z)^{-\delta}. \tag{7}$$

Note that in order to use the equation (7) in the relation $y_{i,t} = \det\left(I_n - \mathbf{A}_n L\right)^{-1} adj\left(I_n - \mathbf{A}_n L\right)_{ii}\epsilon_{i,t} + o_p(1)$ it is sufficient to assume that $adj\left(I_n - \mathbf{A}_n L\right)_{ii}$ is equivalent to $\det\left(I_n - \mathbf{A}_{n-1}L\right)$ which is assumption P(ii) in CHL. This is intuitive as the $ii$th element $adj\left(I_n - \mathbf{A}_n L\right)_{ii}$ of the global adjugate matrix $adj\left(I_n - \mathbf{A}_n L\right)$ is obtained as a determinant of the matrix based on $\left(I_n - \mathbf{A}_n L\right)$ without the $ith$ column and the $ith$ row.

In this paper we show that assumption P(i) in CHL and T(i.c) does not hold for the approach of BCL and we propose an alternative approach to show how long-memory arises in the marginalized time series of a VAR(1) system. We also show that our approach also holds for the definition of $\mathbf{A}_n$ in the VAR(1) model (1) as in example 1 in CHL (see also (19) below). Below we present the procedure of BCL, followed by a section presenting our findings for the set-up in BCL and CHL.

In BCL an alternative and simpler definition of the matrix $\mathbf{A}_n$ is proposed so that they claim it to be compatible with the theoretical results in CHL and Schennach (2018). Their approach is based on the fact that the generic entries of the matrix $\mathbf{A}_n$, that is, $a_{ij}^n$ are defined in such a way that the following 3 conditions hold, for "small" $\varepsilon > 0$ and $\varepsilon' > 0$:

[2] As we show later, P(i) does not hold in the case of the VAR(1) model (1) with the matrix $\mathbf{A}_n$ as defined in BCL (see (8) below)

[3] In particular, assumption P(i) in CHL assumes as $n \to \infty$ that $E\left[\det\left(I_n - \mathbf{A}_n L\right)^{-1}\sum_{k=1,\ k\neq j}^{n} adj\left(I_n - \mathbf{A}_n L\right)_{ik}\epsilon_{k,t}\right] \to 0$

1. The elements of the main diagonal of $\mathbf{A}_n$, that is, $a_{ii}^n$ for $i = 1, 2, \ldots, n$ are close to 1/2 ($a_{ii}^n \in (1/2 - \varepsilon, 1/2]$)

2. The elements outside the main diagonal of $\mathbf{A}_n$, that is, $a_{ij}^n$ for $i \neq j$ are non-negative and close to zero and of order $O\left(n^{-1}\right)$ ($0 \leq n a_{ij}^n < \varepsilon'$).

3. The sums of the elements of a row and of a column of the matrix $\mathbf{A}_n$ are equal to 1 ($\sum_{i=1}^n a_{ij}^n = 1$ and $\sum_{j=1}^n a_{ij}^n = 1$).

Under these conditions the VAR(1) model (1) is compatible with the presence of long memory in the marginalized time series ($y_{i,t}$ for $i = 1, 2, \ldots, n$) and also compatible with the theoretical results of both CHL and Schennach (2018). In particular, BCL propose the following definition for $\mathbf{A}_n$ that guarantees the previous 3 conditions:

$$\begin{aligned}
\mathbf{A}_n &= d_0 \mathbf{I}_n + \frac{1 - d_0}{n - 1} (\mathbf{J}_n - \mathbf{I}_n) \\
&= \frac{d_0 n - 1}{n - 1} \mathbf{I}_n + \frac{1 - d_0}{n - 1} \mathbf{J}_n \\
&= \frac{d_0 n - 1}{n - 1} \mathbf{I}_n + \frac{1 - d_0}{n - 1} 1_n \times 1_n',
\end{aligned} \tag{8}$$

where $d_0$ is close to 1/2, $\mathbf{I}_n$ and $\mathbf{J}_n$ are the identity matrix and a square matrix of ones respectively. Note that $\mathbf{J}_n$ is a circulant matrix of rank one and it is possible to write $\mathbf{J}_n = 1_n \times 1_n'$, where $1_n$ is a $n \times 1$ vector of ones.

Hence, for a generic variable of the VAR(1) system $y_{i,t}$ applying (1)-(8) it is possible to write:

$$\begin{aligned}
y_{i,t} &= \frac{(d_0 n - 1)}{(n - 1)} y_{i,t-1} + \frac{(1 - d_0)}{(n - 1)} 1_n \times 1_n' \mathbf{y}_{n,t-1} + \epsilon_{i,t} \\
&= \frac{(d_0 n - 1)}{(n - 1)} y_{i,t-1} + \frac{(1 - d_0)\, n}{(n - 1)} \bar{y}_{t-1} + \epsilon_{i,t} \\
\bar{y}_t &= \frac{1}{n} 1_n' \mathbf{y}_{n,t} = \frac{1}{n} \sum_{i=1}^n y_{i,t}.
\end{aligned} \tag{9}$$

Thus, as stated by BCL, $y_{i,t}$ can be expressed as the weighted average of the idiosyncratic innovation $\epsilon_{i,t}$, the lagged value $y_{i,t-1}$ and the lagged cross-section average $\bar{y}_{t-1}$. This cross-section average or common factor behaves like a random walk for finite $n$. This fact was shown by BCL as the columns of $\mathbf{A}_n$ sum to unity. Then, we have that $1_n' \times \mathbf{A}_n^0(1) = 1_n' \times (I_n - \mathbf{A}_n) = 0$ and $1_n' \times \mathbf{A}_n^0(L) = 1_n' \times (I_n - \mathbf{A}_n L) = (1 - L) \times 1_n'$, and hence we can write for $\bar{y}_t$:

$$\begin{aligned}
\bar{y}_t &= \bar{y}_{t-1} + \bar{\epsilon}_t \\
\bar{\epsilon}_t &= \frac{1}{n} \sum_{i=1}^n \epsilon_{i,t}.
\end{aligned} \tag{10}$$

As shown in BCL, the variance of the average innovation in (10) $\bar{\epsilon}_t$ decreases as $n$ gets larger, and hence $\bar{y}_t$ corresponds to a "damped trend" process $\bar{y}_t = O_p\left(\sqrt{T/n}\right) = o_p(1)$.

In the following two sub-sections we analyze the approach of BCL and CHL respectively and show that assumptions P(i) and T(i.c) do not hold for BCL and that for a given time dimension $T$ the long memory observed in the marginalized time series depends on the number of time series in the VAR(1) system but not in the specific value of the main diagonal associated with the matrix of coefficients of the VAR(1) model.

## 2.1 Long memory in the set-up in Bauwens, Chevillon and Laurent (2023)

In this section, we first obtain the Moving Average representation of one of the marginalized time series in the VAR(1) system (1) with $\mathbf{A}_n$ defined in (8). We use the fact that both $\mathbf{A}_n = \frac{d_0 n - 1}{n - 1} \mathbf{I}_n + \frac{1 - d_0}{n - 1} 1_n \times 1_n'$ and the polynomial matrix $(I_n - \mathbf{A}_n L) = \mathbf{A}_n^0(L)$ are circulant matrices. Note that $\mathbf{A}_n^0(L)$ has the following form:

$$\mathbf{A}_n^0(L)=\begin{bmatrix}(1-d_0L) & -\frac{1-d_0}{n-1}L & -\frac{1-d_0}{n-1}L & \cdots & -\frac{1-d_0}{n-1}L\\ -\frac{1-d_0}{n-1}L & (1-d_0L) & -\frac{1-d_0}{n-1}L & \cdots & -\frac{1-d_0}{n-1}L\\ -\frac{1-d_0}{n-1}L & -\frac{1-d_0}{n-1}L & (1-d_0L) & \cdots & -\frac{1-d_0}{n-1}L\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ -\frac{1-d_0}{n-1}L & -\frac{1-d_0}{n-1}L & -\frac{1-d_0}{n-1}L & \cdots & (1-d_0L)\end{bmatrix}. \tag{11}$$

Hence, for $(I_n-\mathbf{A}_nL)\,\mathbf{y}_{n,t}=\epsilon_{n,t}$ we have the following Vector Moving Average representation:

$$\mathbf{y}_{n,t}=(I_n-\mathbf{A}_nL)^{-1}\,\epsilon_{n,t}. \tag{12}$$

We know from Davis (1979) and Gray (2006) that, if $\mathbf{C}$ is an $n\times n$ circulant matrix written as $\mathbf{C}=\mathrm{Ci}\,rc\,[c_1,c_2,c_3,\cdots,c_n]$, the first row of $(c_1,c_2,c_3,\cdots,c_n)$, defines all elements of the matrix[4]. In addition we know that for circulant matrices we can write $\mathbf{C}=\mathbf{F}^*\Lambda\mathbf{F}$, where $\mathbf{F}$[5] is an $n\times n$ complex-valued matrix associated with the eigenvectors of $\mathbf{C}$, $\mathbf{F}^*$ is the conjugate transpose of $\mathbf{F}$ and $\Lambda=\mathbf{diag}\,[\lambda_1,\lambda_2,\lambda_3,\cdots,\lambda_n]$, where $\lambda_j$ for $j=1,2,3,\cdots,n$, are the eigenvalues of $\mathbf{C}$. Indeed, $\mathbf{F}$ has the same (known) elements across all circulant matrices, so that such matrices differ only in their eigenvalues. The eigenvalues of a circulant matrix can be obtained based on elements of the first row $(c_1,c_2,c_3,\cdots,c_n)$ (call it $\mathbf{c}_1=[c_1,c_2,c_3,\cdots,c_n]$) and the eigenvectors of the circulant matrix collected in $\mathbf{F}$ as:

$$\begin{aligned}\mathbf{Fc}_1'&=\frac{1}{\sqrt{n}}\begin{bmatrix}1 & 1 & 1 & \cdots & 1\\ 1 & e^{-i\frac{2\pi}{n}} & e^{-i\frac{4\pi}{n}} & \cdots & e^{-i\frac{2(n-1)\pi}{n}}\\ 1 & e^{-i\frac{4\pi}{n}} & e^{-i\frac{8\pi}{n}} & \cdots & e^{-i\frac{4(n-1)\pi}{n}}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 1 & e^{-i\frac{2(n-1)\pi}{n}} & e^{-i\frac{4(n-1)\pi}{n}} & \cdots & e^{-i\frac{2(n-1)^2\pi}{n}}\end{bmatrix}\mathbf{c}_1'\\ &=\frac{1}{\sqrt{n}}\begin{bmatrix}\sum_{j=0}^{n-1}c_{j+1}\\ \sum_{j=0}^{n-1}e^{-i\frac{2\pi}{n}j}c_{j+1}\\ \sum_{j=0}^{n-1}e^{-i\frac{4\pi}{n}j}c_{j+1}\\ \vdots\\ \sum_{j=0}^{n-1}e^{-i\frac{2(n-1)\pi}{n}j}c_{j+1}\end{bmatrix}=\begin{bmatrix}\lambda_1\\ \lambda_2\\ \lambda_3\\ \vdots\\ \lambda_n\end{bmatrix}.\end{aligned} \tag{13}$$

Circulant matrices have useful properties, are computationally simpler to work with and can be used to approximate Toeplitz matrices or forms (see Gray (2006) and Grenander and Szegö (1958) for more details). Here we use the fact that for the inverse of a circulant matrix it holds:

$$\begin{aligned}\mathbf{C}&=\mathbf{F}^*\Lambda\mathbf{F}\\ \mathbf{C}^{-1}&=\mathbf{F}^*\Lambda^{-1}\mathbf{F}.\end{aligned}$$

In the following lemma using results from the inverse of a circulant matrix we present the Moving Average representation of the first time series in the VAR(1) model (1) with $\mathbf{A}_n$ defined in (8). Note that we focus on the first time series of the $n\times 1$ vector $\mathbf{y}_{n,t}$, as it is less tedious than working with any generic time series of the vector $\mathbf{y}_{n,t}$, but it is straightforward to show that the results of the following lemma hold for any time series of the VAR(1) model.

[4] The identity matrix $\mathbf{I}_n$ and a matrix in which all elements are equal to unity are trivial examples of circulant matrices.

[5] $\mathbf{F}$ is defined for all the circulant matrices of dimesion $n\times n$ as:

$$\mathbf{F}=\frac{1}{\sqrt{n}}\begin{bmatrix}1 & 1 & 1 & \cdots & 1\\ 1 & e^{-i\frac{2\pi}{n}} & e^{-i\frac{4\pi}{n}} & \cdots & e^{-i\frac{2(n-1)\pi}{n}}\\ 1 & e^{-i\frac{4\pi}{n}} & e^{-i\frac{8\pi}{n}} & \cdots & e^{-i\frac{4(n-1)\pi}{n}}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 1 & e^{-i\frac{2(n-1)\pi}{n}} & e^{-i\frac{4(n-1)\pi}{n}} & \cdots & e^{-i\frac{2(n-1)^2\pi}{n}}\end{bmatrix}.$$

Note that the matrices associated with the eigenvector of the circulant matrix, that is, $\mathbf{F}$ and $\mathbf{F}^*$, each form an orthogonal basis in the Fourier analysis.

**Lemma 1** *Let $\mathbf{y}_{1,t}$ be the first time series in the vector $n \times 1$ vector $\mathbf{y}_{n,t}$ associate to the VAR(1) model (1) with $\mathbf{A}_n$ defined as (8), then we can write:*

$$\mathbf{y}_{1,t} = \frac{\epsilon_{1,t}}{\left(1 - \frac{nd_0-1}{n-1}L\right)} - \frac{\frac{(d_0-1)}{n-1}\sum_{j=1}^{n}\epsilon_{j,t-1}}{(1-L)\left(1 - \frac{nd_0-1}{n-1}L\right)}$$

$$(1-L)\left(1 - \frac{nd_0-1}{n-1}L\right)\mathbf{y}_{1,t} = (1-L)\,\epsilon_{1,t} - \frac{(d_0-1)}{n-1}\sum_{j=1}^{n}\epsilon_{j,t-1}$$

$$(1-L)\left(1 - \frac{nd_0-1}{n-1}L\right)\mathbf{y}_{1,t} = (1-L)\,\epsilon_{1,t} + \frac{n\,(1-d_0)}{n-1}\frac{1}{n}\sum_{j=1}^{n}\epsilon_{j,t-1}$$

$$(1-L)\left(1 - \frac{nd_0-1}{n-1}L\right)\mathbf{y}_{1,t} = (1-L)\,\epsilon_{1,t} + \frac{n\,(1-d_0)}{n-1}\frac{1}{n}\overline{\epsilon}_{t-1}$$

$$\overline{\epsilon}_{t-1} = \frac{1}{n}\sum_{j=1}^{n}\epsilon_{j,t-1}. \tag{14}$$

Note that expression (14) is exactly the same as the one reported in equation (29) in the appendix associated with expression (9).

In Chevillon, Hecq and Laurent (2018) (see their section 2.3) a similar approach to (14) is used. In particular CHL use the Final Equation Representation of the VAR model (see expressions (6) and (7) in section (2.3) of CHL). Note that expression (7) in CHL is basically the same as our expression (12). Also note that for $A_n$ defined as in (8) and based on (14) we can write for the first time series of $\mathbf{y}_{n,t} = (I_n - \mathbf{A}_n L)^{-1}\epsilon_{n,t}$:

$$\mathbf{y}_{1,t} = \frac{\epsilon_{1,t}}{\left(1 - \frac{nd_0-1}{n-1}L\right)} - \frac{\frac{(d_0-1)}{n-1}\sum_{j=1}^{n}\epsilon_{j,t-1}}{(1-L)\left(1 - \frac{nd_0-1}{n-1}L\right)}. \tag{15}$$

Following assumption P(i) of CHL, the equivalent terms to $\left[(1-L)\left(1 - \frac{nd_0-1}{n-1}L\right)\right]^{-1}\frac{(d_0-1)}{n-1}\sum_{j=1}^{n}\epsilon_{j,t-1}$ in (7) of CHL will be negligible. In our case this situation does not apply. Note also that in CHL assumption P(i) is crucial for the use of the Szegö Theorem (see Lemma 1 and Theorem 1 in CHL). But in our case, for the VAR(1) model (1) with $A_n$ defined as in (8), the terms $\left[(1-L)\left(1 - \frac{nd_0-1}{n-1}L\right)\right]^{-1}\frac{(d_0-1)}{n-1}\sum_{j=1}^{n}\epsilon_{j,t-1}$ are not negligible, and also the weight associated to $\epsilon_{1,t}$ is $\left(1 - \frac{nd_0-1}{n-1}L\right)^{-1}$ which differs from $(1-L)^{-d_0}$[6] as stated in CHL in their Lemma 1 and Theorem 1.

In the following Proposition we show that in (9)-(10) we have long-memory in the marginalized time series of the VAR(1) model (1) with $\mathbf{A}_n$ defined as in (8). Note also that in our case the intensity of the long memory depends on the relation between $T$ and $n$. That is, we obtain longer memory if $n$ is smaller for a given value of $T$.

**Proposition 2** *For a marginalized time series $y_{i,t}$ of the VAR(1) model (1) such that the matrix $\mathbf{A}_n$ is defined as in (8) such that (9)-(10) holds for $y_{i,t}$ we have that:*

$$(1-L)\left(1 - \frac{d_0 n-1}{n-1}L\right)y_{i,t} = (1-\theta_n L)\,\eta_t, \tag{16}$$

*with $\eta_t$ being a white noise process and the moving average parameter $\theta_n$ being a function of n (see the appendix for details). If the Variance-Covariance matrix $\Sigma_n$ of $\epsilon_{n,t}$ in (1) is diagonal $\theta_n = 1 - O\left(n^{-1/2}\right)$ and (16) becomes:*

$$(1-L)\left(1 - \frac{d_0 n-1}{n-1}L\right)y_{i,t} = \left(1 - \left[1 - O\left(n^{-1/2}\right)\right]L\right)\eta_t. \tag{17}$$

*If the Variance-Covariance matrix $\Sigma_n$ of $\epsilon_{n,t}$ in (1) is not diagonal $\theta_n = 1 - O(1)$ and (16) becomes:*

$$(1-L)\left(1 - \frac{d_0 n-1}{n-1}L\right)y_{i,t} = (1 - [1 - O(1)]\,L)\,\eta_t. \tag{18}$$

[6]Note that $\delta$ and $d_0$ are playing the same role here.

*See the appendix for the details about (17) and (18).*

**Remark 3** *The resulting model for $y_{i,t}$ in (17) and (18) is an ARIMA(1,1,1) with autoregressive coefficient equal to $(d_0 n - 1) / (n-1)$ and the difference operator $(1-L)$ almost canceles out with the MA(1) polynomial $\left(1 - \left[1 - O\left(n^{-1/2}\right)\right] L\right)$ or $(1 - [1 - O(1)] L)$. Hence, we see that the intensity of the observed long memory in $y_{i,t}$ depends on $n$. As $n$ increases the MA(1) is close to cancellation with the difference operator $(1-L)$ and the intensity of the long memory decreases.*

The situation reported in Proposition 2 and Remark 3 is illustrated in the Monte-Carlo experiment of section 3.

**Remark 4** *Note also that based on Proposition 2, Remark 3 and the expressions (9) and (10) we can say that there is a common behavior in all of the marginalized time series of the VAR(1) model 1 and that, as shown in Proposition 2 and Remark 3 this common behavior is manifested as a long memory behavior in each of the marginalized time series. Hence, we expect to have cointegration between the marginalized time series of the VAR(1) system.*

**Corollary 5** *If $n$ grows with $T$, such that $n = O\left(T^{\delta}\right)$, and if the Variance-Covariance matrix $\Sigma_n$ of $\epsilon_{n,t}$ in (1) is diagonal, then $\theta_n = 1 - O\left(T^{-\delta/2}\right)$. If $\Sigma_n$ is not diagonal, then $\theta_n = 1 - O(1)$.*

Therefore, following the previous analysis we can say that the long memory observed in the marginalized time series of the VAR(1) model (1) with the matrix of coefficients $A_n$ defined as in (8) is not explained by the proof given in Chevillon, Hecq, and Laurent (2018) as claimed by Bauwens, Chevillon and Laurent (2023). Furthermore, in this paper we also show that the damped trend associated with the VAR(1) model in the simplified model proposed by Bauwens, Chevillon and Laurent (2023) also explains that the long-run behavior of the time series of the VAR(1) model is governed by this common stochastic trend (damped trend). This explains the fact that each pair of time series of the VAR(1) model (1) is cointegrated with cointegration vector $[1, -1]$.

Finally, the results about the VAR(1) model are extended to a VAR(2) model associated with harmonic frequencies in a recent work by del Barrio Castro, Escribano and Sibbertsen (2024). They show that it is possible to have long memory associated to a harmonic frequency (cyclical behavior) in the marginalized time series of a VAR(2) system and also that this non-stationary behavior is common to all time series of the VAR(2) model. Hence, we have fractional cointegration between the time series of this VAR(2) system. The results of Proposition 2 in this section also apply to the situation analyzed by del Barrio Castro, Escribano and Sibbertsen (2024).

## 2.2 Long memory in the set-up of Chevillon, Hecq, and Laurent (2018)

In Chevillon, Hecq and Laurent (2018) the matrix $\mathbf{A}_n$ for $(I_n - \mathbf{A}_n L)\,\mathbf{y}_{n,t} = \epsilon_{n,t}$ is a Toeplitz matrix like in (2). We focus on example 1 (see section 2.1 in CHL and also the Monte-Carlo section in BCL). For their example 1 the elements of matrix $\mathbf{A}_n = \mathbf{T}_n^*$ are determined based on the following expression:

$$t_k^n = \operatorname{Re}\left[\frac{1}{n}\sum_{j=0}^{n-1} g\left(\delta_n, e^{i\frac{2\pi j}{n}}\right) e^{-i\frac{2\pi jk}{n}}\right] \tag{19}$$

$$g\left(\delta, e^{i\omega}\right) = \mathbf{1}_{[0\leq\omega\leq\delta\pi]} + \mathbf{1}_{\left[\pi\left(\frac{3}{2}-\delta\right)\leq\omega\leq\frac{3\pi}{2}\right]} \quad \delta \in (0, 1/2) \quad \omega \in [0, 2\pi).$$

Note that (19) can be also expressed as[7]:

$$\begin{aligned} t_k^n &= \frac{1}{n}\sum_{j=0}^{\delta n/2}\cos\left(\frac{2\pi jk}{n}\right) + \frac{1}{n}\sum_{j=\left(\frac{3}{2}-\delta\right)/2}^{3/4n}\cos\left(\frac{2\pi jk}{n}\right) \\ &= \frac{1}{n}\sum_{j=0}^{\delta n/2}\cos\left(\frac{2\pi jk}{n}\right) + \frac{1}{n}\sum_{j=0}^{3/4n}\cos\left(\frac{2\pi jk}{n}\right) \\ &\quad - \frac{1}{n}\sum_{j=0}^{\left(\frac{3}{2}-\delta\right)/2}\cos\left(\frac{2\pi jk}{n}\right). \end{aligned} \tag{20}$$

[7] As we have a real part affecting this expression and the function is a combination of two indicators functions.

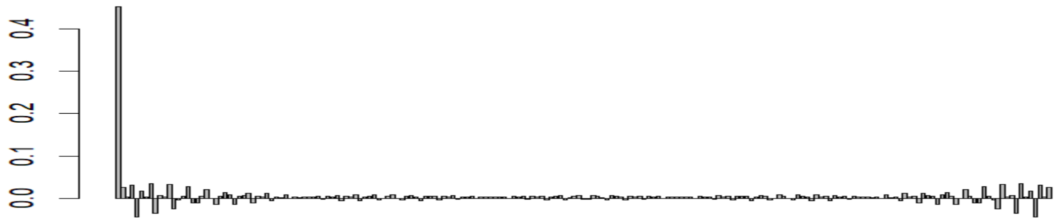


Figure 1: Values of first row of matrix $\mathbf{T}_n^*$, that is, $t_0^n, t_1^n, t_2^n, \cdots, t_{n-1}^n$ for $n = 201$.

Based on the expressions for the cumulate sum of a sequence of cosines[8] we can re-write expression (20) as:

$$t_k^n = \frac{1}{n}\left[1 + \cos\left(\frac{\pi\left[\frac{\delta n}{2}+1\right]k}{n}\right)\frac{\sin\left(\frac{\pi\left[\frac{\delta n}{2}\right]k}{n}\right)}{\sin\left(\frac{\pi k}{n}\right)} + \cos\left(\frac{\pi\left[\frac{3n}{4}+1\right]k}{n}\right)\frac{\sin\left(\frac{\pi\left[\frac{3n}{4}\right]k}{n}\right)}{\sin\left(\frac{\pi k}{n}\right)}\right. \tag{21}$$
$$\left. - \cos\left(\frac{\pi\left[\left(\frac{3}{4}-\frac{\delta}{2}\right)+1\right]k}{n}\right)\frac{\sin\left(\frac{\pi\left(\frac{3}{4}-\frac{\delta}{2}\right)k}{n}\right)}{\sin\left(\frac{\pi k}{n}\right)}\right].$$

Given (20) and (21) it can be seen that $t_k^n = t_{n-k}^n$ for $k = 1, 2, \ldots, (n-1)/2$. That is, $t_1^n = t_{n-1}^n$, $t_2^n = t_{n-2}^n, \ldots, t_{(n-1)/2}^n = t_{n-(n-1)/2}^n$. Further, from (20) and (21) it can be derived that $t_k^n = t_{-k}^n$ for $k = 1, 2, \ldots, n-1$. Hence, the Toeplitz matrix $\mathbf{T}_n^*$ (2) with the elements $t_k^n$ defined as in (19) is also symmetric and circulant. In figures 1 and 2 we report the values of the first row of the matrix $\mathbf{T}_n^*$, that is, $t_0^n, t_1^n, t_2^n, \cdots, t_{n-1}^n$ (figure 1) and the set of values $t_1^n, t_2^n, \cdots, t_{n-1}^n$ (figure 2) for $n = 201$ and it shows that there is symmetry in the values. In addition, panel a of figure 5 in CHL shows the same pattern as the one observed in figure 3 below.

Hence, we can proceed as in the previous subsection and obtain the VMA($\infty$) representation of the VAR(1) model $(I_n - \mathbf{A}_n L)\,\mathbf{y}_{n,t} = \epsilon_{n,t}$ when $\mathbf{A}_n = \mathbf{T}_n^*$ and the elements of $\mathbf{T}_n^*$ defined as in (20) or (21). First note that now we have:

[8] Note that we can write:

$$\begin{aligned}\sum_{j=0}^{\delta n/2}\cos\left(\frac{2\pi jk}{n}\right) &= \mathrm{Re}\left[\sum_{j=0}^{\delta n/2} e^{i\frac{2\pi jk}{n}}\right] = \mathrm{Re}\left[1 + \sum_{j=1}^{\delta n/2} e^{i\frac{2\pi jk}{n}}\right] \\ &= \mathrm{Re}\left[1 + \frac{e^{i\frac{2\pi(\delta n/2+1)k}{n}} - e^{i\frac{2\pi k}{n}}}{e^{i\frac{2\pi k}{n}} - 1}\right] \\ &= \mathrm{Re}\left[1 + e^{i\frac{2\pi k}{n}}\frac{e^{i\frac{\pi(\delta n/2)k}{n}}}{e^{i\frac{\pi k}{n}}}\frac{e^{i\frac{\pi(\delta n/2)k}{n}} - e^{-i\frac{\pi(\delta n/2)k}{n}}}{e^{i\frac{\pi k}{n}} - e^{-i\frac{\pi k}{n}}}\right] \\ &= \mathrm{Re}\left[1 + e^{i\frac{\pi(\delta n/2+1)k}{n}}\frac{\sin\left(\frac{\pi(\delta n/2)k}{n}\right)}{\sin\left(\frac{\pi k}{n}\right)}\right] \\ &= 1 + \cos\left(\frac{\pi(\delta n/2+1)k}{n}\right)\frac{\sin\left(\frac{\pi(\delta n/2)k}{n}\right)}{\sin\left(\frac{\pi k}{n}\right)},\end{aligned}$$

where $\sin\left(\frac{\pi(\delta n/2)k}{n}\right)/\sin\left(\frac{\pi k}{n}\right)$ is $n$ times the Dirichlet kernel, in Bloomfield (2000) (exercice 2.2 part ii) you can find a similar expression that also involves the Dirichlet kernel.

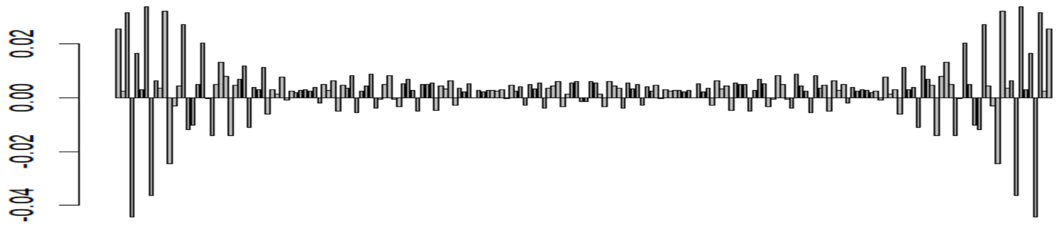


Figure 2: As figure 1 but the set of values $t_1^n, t_2^n, \cdots, t_{n-1}^n$.

$$\mathbf{A}_n^0(L) = \begin{bmatrix} (1-t_0^n L) & -t_1^n L & -t_2^n L & \cdots & -t_{n-1}^n L \\ -t_{-1}^n L & (1-t_0^n L) & -t_1^n L & \cdots & -t_{n-2}^n L \\ -t_{-2}^n L & -t_{-1}^n L & (1-t_0^n L) & \cdots & -t_{n-2}^n L \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -t_{-(n-1)}^n L & -t_{-(n-2)}^n L & -t_{-(n-3)}^n L & \cdots & (1-t_0^n L) \end{bmatrix} \tag{22}$$
$$= \begin{bmatrix} (1-t_0^n L) & -t_1^n L & -t_2^n L & \cdots & -t_1^n L \\ -t_1^n L & (1-t_0^n L) & -t_1^n L & \cdots & -t_2^n L \\ -t_1^n L & -t_1^n L & (1-t_0^n L) & \cdots & -t_3^n L \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -t_1^n L & -t_2^n L & -t_3^n L & \cdots & (1-t_0^n L) \end{bmatrix}.$$

In the second line of expression (22), we use the fact that (22) is a symmetric Toeplitz matrix and a circulant matrix as well. Then we are able to proceed as in the case of (11) and use the fact that (22) is circulant and hence the eigenvalues of $\mathbf{A}_n^0(L)$ (22) are as follows:

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \vdots \\ \lambda_n \end{bmatrix} = \frac{1}{\sqrt{n}} \begin{bmatrix} \left(1-\sum_{j=0}^{n-1} t_j^n L\right) = (1-L) \\ \left(1-\sum_{j=0}^{n-1} e^{-i\frac{2\pi}{n}j} t_j^n L\right) \\ \left(1-\sum_{j=0}^{n-1} e^{-i\frac{4\pi}{n}j} t_j^n L\right) \\ \vdots \\ \left(1-\sum_{j=0}^{n-1} e^{-i\frac{2(n-1)\pi}{n}j} t_j^n L\right) \end{bmatrix}. \tag{23}$$

For the first eigenvalue $\lambda_1$ we use the fact that we have $\sum_{j=0}^{n-1} t_j^n = 1$. Note also that for the remaining eigenvalues $\lambda_k$ for $k = 2, 3, \ldots, n$ it is possible to write based on (23) and (21) that:

$$\lambda_k = \frac{1}{\sqrt{n}}\left(1-\sum_{j=0}^{n-1} e^{-i\frac{2(k-1)\pi}{n}j} t_j^n L\right) \tag{24}$$

$$= \frac{1}{\sqrt{n}}\left(1-\frac{1}{n}\sum_{j=0}^{n-1} e^{-i\frac{2(k-1)\pi}{n}j} L - \frac{1}{n}\sum_{j=0}^{n-1} e^{-i\frac{2(k-1)\pi}{n}j} \cos\left(\frac{\pi\left[\frac{\delta n}{2}+1\right]j}{n}\right)\frac{\sin\left(\frac{\pi\left[\frac{\delta n}{2}\right]j}{n}\right)}{\sin\left(\frac{\pi j}{n}\right)} L\right.$$

$$\left. -\frac{1}{n}\sum_{j=0}^{n-1} e^{-i\frac{2(k-1)\pi}{n}j} \cos\left(\frac{\pi\left[\frac{3n}{4}+1\right]j}{n}\right)\frac{\sin\left(\frac{\pi\left[\frac{3n}{4}\right]j}{n}\right)}{\sin\left(\frac{\pi j}{n}\right)} L + \frac{1}{n}\sum_{j=0}^{n-1} e^{-i\frac{2(k-1)\pi}{n}j} \cos\left(\frac{\pi\left[\left(\frac{3}{4}-\frac{\delta}{2}\right)+1\right]j}{n}\right)\frac{\sin\left(\frac{\pi\left(\frac{3}{4}-\frac{\delta}{2}\right)j}{n}\right)}{\sin\left(\frac{\pi j}{n}\right)}\right)$$

$$= \frac{1}{\sqrt{n}}\left(1-\frac{1}{n}\sum_{j=0}^{n-1} e^{-i\frac{2(k-1)\pi}{n}j} \cos\left(\frac{\pi\left[\frac{\delta n}{2}+1\right]j}{n}\right)\frac{\sin\left(\frac{\pi\left[\frac{\delta n}{2}\right]j}{n}\right)}{\sin\left(\frac{\pi j}{n}\right)} L\right.$$

$$\left. -\frac{1}{n}\sum_{j=0}^{n-1} e^{-i\frac{2(k-1)\pi}{n}j} \cos\left(\frac{\pi\left[\frac{3n}{4}+1\right]j}{n}\right)\frac{\sin\left(\frac{\pi\left[\frac{3n}{4}\right]j}{n}\right)}{\sin\left(\frac{\pi j}{n}\right)} L + \frac{1}{n}\sum_{j=0}^{n-1} e^{-i\frac{2(k-1)\pi}{n}j} \cos\left(\frac{\pi\left[\left(\frac{3}{4}-\frac{\delta}{2}\right)+1\right]j}{n}\right)\frac{\sin\left(\frac{\pi\left(\frac{3}{4}-\frac{\delta}{2}\right)j}{n}\right)}{\sin\left(\frac{\pi j}{n}\right)}\right).$$

After the third equality in the previous expression we use the fact that $\sum_{j=0}^{n-1} e^{-i\frac{2(k-1)\pi}{n}j} = 0$. Note that the eigenvalues (24), are a linear combination of the Discrete Fourier Transforms (DFT) of three different cosine waves. Also, each of these three DFT´s are weighted by ratios similar to the Dirichlet kernel. From picture A and B it can be seen that the sequence of values $t_j^n$ for $j = 0, 1, \ldots, n-1$ is quite similar to a combination of a single impulse ($t_0^n$) and a step function ($t_1^n, t_2^n, \ldots, t_{n-1}^n$). We know from Bloomfield (2000) that the Discrete Fourier Transform (DFT) of a single impulse is a constant and that the DFT of a step function has a maximum at the zero frequency and decays for the remaining frequencies (see Bloomfield (2000) section 4.4 parts (ii) and (iii)). Finally, we further know from Bloomfield (2000) (see section 4.4 part i) that the DFT of a cosine wave is not zero but vanishes when we have leakage, that is, when the $f \pm f_0$ is a Fourier frequency, where $f$ is the frequency associated to the DFT and $f_0$ the frequency associated to the cosine wave. In table 1 below we report the eigenvalues $\lambda_k^{t_j^n} = \sum_{j=0}^{n-1} e^{-i\frac{2(k-1)\pi}{n}j} t_j^n$ for $k = 1, 2, \ldots, n$ in the case of $n = 201$ as in pictures A and B.

Table 1: Eigenvalues of the sequence $t_j^n$ for $j = 0, 1, \ldots, 200$

| $k =$ | 1 | $2, \ldots, 45$ | $46, \ldots, 50$ | $51, \ldots, 95$ | $96, \ldots, 105$ | $106, \ldots, 150$ | $151, \ldots, 155$ | $156, \ldots, 201$ |
|---|---|---|---|---|---|---|---|---|
| $\lambda_k^{t_j^n} =$ | 1 | 0.5 | 0 | 0.5 | 0 | 0.5 | 0 | 0.5 |

Hence, based on table 1 we can see that the eigenvalues of $\mathbf{A}_n^0(L)$ with $\mathbf{A}_n = \mathbf{T}_n^*$ and $\mathbf{T}_n^*$ defined in (19) are very similar to those reported in (23). If the first eigenvalue $\lambda_1^{t_j^n} = 1$, then the first eigenvalue of $(I_n - T_n^* L)$ always is $\frac{1}{\sqrt{n}}(1-L)$ as can be seen in the first row of (23). If $\lambda_k^{t_j^n} = 0.5$ in table 1 the eigenvalues of $(I_n - T_n^* L)$ are going to be $\frac{1}{\sqrt{n}}(1-0.5L)$ for 181 cases and, finally if $\lambda_k^{t_j^n} = 0$ the eigenvalues will be $\frac{1}{\sqrt{n}}$ for the remaining 20 cases again in accordance with (23). So we obtain a similar pattern for the inverse matrix $(I_n - \mathbf{T}_n^* L)^{-1}$ leading to the ones reported in (26) for $(I_n - \mathbf{A}_n L)^{-1}$ in the situation discussed in BCL. Therefore, a similar expression as the one reported in (15) is obtained here for the marginalized time series of the VAR(1) model.

# 3 Monte Carlo Experiment

In this section we run a Monte Carlo experiment with $n = \{101, 201, 401, 601, 1201\}$ and $T = 4000$. With the number of replications set equal to 4000, we generate the data based on the VAR(1) system (1) with $\mathbf{A}_n$ as in (8) with $d_0 = 0.499$ following the approach in BCL. In addition we also generate data as in example 1 in CHL with $d_0 = 0.45$ (see page 56 and the footnote of figure 2 page 59, see also figure 1 in BCL, pages 8 and 9). In figure 3 the density of the empirical distribution of the estimators of the long-memory parameter can be seen obtained by applying the Whittle estimator proposed by Shimotsu and Phillips (2005),[9] as well as the Pseudo Maximum Likelihood estimator by Beran (1995) and the Minimum Distance estimator by Mayoral (2007) of $d$ for an ARFIMA(0,d,0)[10].

Figure 3 supports the findings of Proposition 2 and Remark 3. We observe the decay of the fitted values of the long-memory parameter in the marginalized time series in the case of using the semiparametric Whittle estimator as it is not influenced by the autoregressive part. But for the parametric Pseudo Maximum Likelihood estimator and the Minimum Distance estimator applied to a wrongly specified ARFIMA(0,d,0) model the decay of the fitted values is not so pronounced.

From Proposition 2 and Remark 3 it follows that the marginalized time series of the VAR(1) model is an ARIMA(1,1,1) process with near cancellation of the first difference and the MA(1) part. Hence, we expect that we obtain a more accurate estimation of $d$ fitting an ARFIMA(1,d,0) model. Figure 4 presents equivalent results to those reported in figure 3 but reporting the estimation of $d$ applying the Pseudo Maximum Likelihood estimator and the Minimum Distance estimator now based on fitting an ARFIMA(1,d,0) model. In the figure 4 we now observe a decay of the estimated values of the long-memory parameter as $n$ increases for the three different method used to fit this parameter.

Also from the results in Proposition 2 and Remark 3 and in particular expressions (16), (17) and (18) it can be observed that the value in the main diagonal of matrix $\mathbf{A}_n$ $d_0$ is not connected to the memory observed in the marginalized time series of the VAR(1) model. But, on the other hand it shows that the coefficient associated with the AR(1) process is connected to $d_0$. That is, for example in (16) we have $(1-L)\left(1-\frac{d_0 n-1}{n-1}L\right)y_{i,t}=(1-\theta_n L)\,\eta_t$ and therefore the AR(1) coefficient equals $(d_0 n-1)\,/\,(n-1)$. Hence, as $n$ increases the AR1 coefficient tends to $d_0$.

With the same number of replications (4000) $n=101$ and $T=4000$ we simulated the VAR(1) model with $\mathbf{A}_n$ as in (8) for $d_0=\{0.5,0.4,0,3,0.2\}$ and also following example 1 in CHL but with $d_0=\{0.5,0.4,0,3,0.2\}$. First, in figure 5 we present the results of the estimation of the long memory parameter $d$ with the same three methods described before. As expected, figure 5 shows no difference in the estimations of the long memory parameter as there is no connection between $d_0$ and the long memory parameter.

We report in figure 6 the empirical distribution of the estimation of the AR(1) coefficient by on the one hand applying the two parametric estimators, namely the Pseudo Maximum Likelihood and the Minimum Distance estimator, to the marginalized time series and on the other hand after the common damped trend is extracted of the marginalized time series of the VAR(1) model by using two different methods. The damped trend is extracted either by taking the difference of the first and second time series $y_{1,t}-y_{2,t}$ or by running a regression of $y_{1,t}$ on $y_{2,t}$ and considering the residuals. Afterwards an AR(1) model is fitted to the respective series. We report the empirical distribution of the respective estimators of the AR(1) coefficient applying these four methods. It can be seen that with decreasing $d_0$ the estimators of the autoregressive coefficient also decrease as is expected from Proposition 2 and Remark 3. This effect is though more clear for estimation after removing the damped trend as the parametric estimators face a big downward bias.

Without loss of generality and only in the case of $n=101$ we test for cointegration between the first and second marginalized time series using the fractional cointegration rank tests by Chen and Hurvich (2003) and Zhang, Robinson and Yao (2019). We find that the cointegration rank is one in all the 4000 replications[11]. Finally, for $n=201$ we report in figures 7 and 8 the average ACF and PACF of the first marginalized time series $y_{1,t}$ of the VAR(1) model, the average ACF and PACF of the residuals of $y_{1,t}$ on $y_{2,t}$ and finally the average ACF and PACF of $y_{1,t}-y_{2,t}$. In figure 7 we collect these results for the VAR(1) model (1) with $\mathbf{A}_n$ as in (8) with $d_0=0.499$. In figure 8 we display the results for a VAR(1) model as in example 1 of CHL with $d_0=0.45$ (see page 56 and the footnote of figure 2 page 59, see also figure 1 in BCL pages 8 and 9). After taking into account the common damped trend by either considering the residuals of the regression of $y_{1,t}$ on $y_{2,t}$ or the difference $y_{1,t}-y_{2,t}$, there is no long memory behavior left but only the AR(1) part of the data generation process associated with the marginalized time series $(1-L)\left(1-\frac{d_0 n-1}{n-1}L\right)y_{i,t}=(1-\theta_n L)\,\eta_t$. This is the part associated with the polynomial $\left(1-\frac{d_0 n-1}{n-1}L\right)$.

# 4 Conclusion

In this paper we propose a modified approach to obtaining long memory in marginalized time series in a vector autoregressive system compared to the earlier results of Chevillon, Hecq and Laurent (2018) and

[9] We use the function **ELW** from the R package Maechler (2024). .

[10] We use the function **mle.arfima** and **mde.arfima** of the R package **nsarfima** by Groebe (2020).

[11] We use the functions **FCI_CH03** and **FCI_ZRY18** of the R package library **LongMemoryTS** by Leschinski (2022).

Bauwens, Chevillon and Laurant (2023). Based on results for Toeplitz and circulant matrices we show that the observed memory after marginalization is due to the near cancellation of a damped trend (common to all series in the system) and the moving average part of the data generating process of the marginalized time series. It is, therefore, rather related to near cancellation in an ARIMA process than a long-memory behavior of the marginalized time series. The memory decreases with an increasing size of the vector autoregressive system in relation to the time dimension. The theoretical results are underpinned by an extensive Monte Carlo exercise showing in addition that pairs of marginalized time series of the system are fractionally cointegrated with cointegration vector $(1, -1)$ which strongly supports our results.

# References


[1] Bauwens, L. & Chevillon, G. & Laurent S. (2023) We modeled long memory with just one lag!, Journal of Econometrics, 236, 105467.

[2] Beran, J. (1995) Maximum Likelihood Estimation of the Differencing Parameter for Short and Long Memory Autoregressive Integrated Moving Average Models. Journal of the Royal Statistical Society. Series B (Methodological), 57, No. 4, 659-672.

[3] Bloomfield P. (2000) Fourier Analysis of Time Series: An Introduction, Wiley.

[4] Chen, W. W. and Hurvich, C. M. (2003) Semiparametric estimation of multivariate fractional cointegration. Journal of the American Statistical Association, Vol. 98, No. 463, pp. 629- 642.

[5] Chevillon, G., Hecq, A. and Laurent, S. (2018) Generating univariate fractional integration within a large VAR(1), Journal of Econometrics 204, 54–65.

[6] del Barrio Castro, T., Escribano, A. and Sibbertsen, P. (2024) Modeling and Forecasting the Long Memory of Cyclical Trends in Paleoclimate Data, Hannover Economic Papers (HEP) dp-722.

[7] Gray, R. M. (2006) Toeplitz and Circulant Matrices. A Review. Foundation and Trends(R) in Communications and Information Theory, Now Publishers Inc.

[8] Grenander U. and Szegö G. (1958) Toeplitz Forms and their Applications, AMS ChelseaPublishing.

[9] Groebe, B. (2020). nsarfima: Methods for Fitting and Simulating Non-Stationary ARFIMA Models. https://CRAN.R-project.org/package=nsarfima.

[10] Leschinski, C. (2019). LongMemoryTS: Long Memory Time Series. https://CRAN.R-project.org/package=LongMemoryTS.

[11] Maechler, M. (2024). fracdiff: Fractionally Differenced ARIMA aka ARFIMA(P,d,q) Models. https://CRAN.R-project.org/package=fracdiff.

[12] Mayoral, L. (2007) Minimum distance estimation of stationary and non-stationary ARFIMA processes. The Econometrics Journal, 10, 124-148

[13] Schennach, S.M., (2018) Long memory via networking. Econometrica 86, 2221–2248.

[14] Shimotsu, K. and Phillips, P. C. B. (2005) Exact Local Whittle Estimation Of Fractional Integration. The Annals of Statistics, Vol. 33, No. 4, pp. 1890 - 1933

[15] Zellner A. and Palm F. (1974) Time series analysis and simultaneous equation econometrics models, Journal of Econometrics, 2. 17-54.

[16] Zhang, R., Robinson, P. and Yao, Q. (2019) Identifying Cointegration by Eigenanalysis. Journal of the American Statistical Association, 114, 526, 916-927.

# 5 Appendix

**Proof of Lemma 1:**

Note first, that the eigenvalues of $\mathbf{A}_n^0(L)$ (11) $\lambda_j^{\mathbf{A}_n^0(L)}$ for $j = 1, 2, \ldots, n$ are as follows:

$$\begin{bmatrix} \lambda_1^{\mathbf{A}_n^0(L)} \\ \lambda_2^{\mathbf{A}_n^0(L)} \\ \lambda_3^{\mathbf{A}_n^0(L)} \\ \vdots \\ \lambda_n^{\mathbf{A}_n^0(L)} \end{bmatrix} = \frac{1}{\sqrt{n}} \begin{bmatrix} (1-d_0L) - \frac{1-d_0}{n-1}Ln \\ (1-d_0L) - \frac{1-d_0}{n-1}L\sum_{j=1}^{n-1} e^{-i\frac{2\pi}{n}j} \\ (1-d_0L) - \frac{1-d_0}{n-1}L\sum_{j=1}^{n-1} e^{-i\frac{4\pi}{n}j} \\ \vdots \\ (1-d_0L) - \frac{1-d_0}{n-1}L\sum_{j=1}^{n-1} e^{-i\frac{2(n-1)\pi}{n}j} \end{bmatrix}$$
$$= \frac{1}{\sqrt{n}} \begin{bmatrix} (1-d_0L) - \frac{1-d_0}{n-1}L(n-1) \\ (1-d_0L) + \frac{1-d_0}{n-1}L \\ (1-d_0L) + \frac{1-d_0}{n-1}L \\ \vdots \\ (1-d_0L) + \frac{1-d_0}{n-1}L \end{bmatrix} \tag{25}$$
$$= \frac{1}{\sqrt{n}} \begin{bmatrix} (1-L) \\ \left(1 - \frac{nd_0-1}{n-1}L\right) \\ \left(1 - \frac{nd_0-1}{n-1}L\right) \\ \vdots \\ \left(1 - \frac{d_0n-1}{n-1}L\right) \end{bmatrix}.$$

In (25) we use the fact that $\sum_{j=0}^{n-1} e^{-i\frac{2\pi k}{n}j} = 0$ $\sum_{j=0}^{n-1} e^{-i\frac{2\pi k}{n}j} = 1 + \sum_{j=1}^{n-1} e^{-i\frac{2\pi k}{n}j}$ which implies that $\sum_{j=1}^{n-1} e^{-i\frac{2\pi k}{n}j} = -1$.

Hence, we can write for $(I_n - \mathbf{A}_nL)^{-1}$:

$$(I_n - \mathbf{A}_nL)^{-1} = \mathbf{F}^* diag \begin{bmatrix} (1-L)^{-1} & \left(1 - \frac{nd_0-1}{n-1}L\right)^{-1} & \left(1 - \frac{nd_0-1}{n-1}L\right)^{-1} & \cdots & \left(1 - \frac{nd_0-1}{n-1}L\right)^{-1} \end{bmatrix} \mathbf{F}. \tag{26}$$

If we focus on the first row of $(I_n - \mathbf{A}_nL)^{-1}$ knowing that the first row of $\mathbf{F}^*$ is a vector of ones, we can see that the first row of $(I_n - \mathbf{A}_nL)^{-1}$ is as follows:

$$\begin{bmatrix} \frac{1}{\sqrt{n}}\frac{1}{(1-L)} & \frac{1}{\sqrt{n}}\frac{1}{\left(1-\frac{nd_0-1}{n-1}L\right)} & \frac{1}{\sqrt{n}}\frac{1}{\left(1-\frac{nd_0-1}{n-1}L\right)} & \cdots & \frac{1}{\sqrt{n}}\frac{1}{\left(1-\frac{nd_0-1}{n-1}L\right)} \end{bmatrix} \mathbf{F} \tag{27}$$
$$= \begin{bmatrix} \frac{1}{n}\left(\frac{1}{(1-L)} + \frac{(n-1)}{\left(1-\frac{nd_0-1}{n-1}L\right)}\right) & \frac{1}{n}\left(\frac{1}{(1-L)} + \frac{\sum_{j=1}^{n-1} e^{i\frac{2\pi}{n}j}}{\left(1-\frac{nd_0-1}{n-1}L\right)}\right) & \frac{1}{n}\left(\frac{1}{(1-L)} + \frac{\sum_{j=1}^{n-1} e^{i\frac{4\pi}{n}j}}{\left(1-\frac{nd_0-1}{n-1}L\right)}\right) & \cdots & \frac{1}{n}\left(\frac{1}{(1-L)} + \frac{\sum_{j=1}^{n-1} e^{-i\frac{2(n-1)\pi}{n}j}}{\left(1-\frac{nd_0-1}{n-1}L\right)}\right) \end{bmatrix}$$
$$= \begin{bmatrix} \frac{1}{n}\left(\frac{1}{(1-L)} + \frac{(n-1)^2}{\left(1-\frac{nd_0-1}{n-1}L\right)}\right) & \frac{1}{n}\left(\frac{1}{(1-L)} - \frac{1}{\left(1-\frac{nd_0-1}{n-1}L\right)}\right) & \frac{1}{n}\left(\frac{1}{(1-L)} - \frac{1}{\left(1-\frac{nd_0-1}{n-1}L\right)}\right) & \cdots & \frac{1}{n}\left(\frac{1}{(1-L)} - \frac{1}{\left(1-\frac{nd_0-1}{n-1}L\right)}\right) \end{bmatrix}$$
$$= \begin{bmatrix} \frac{1}{n}\left(\frac{n}{\left(1-\frac{nd_0-1}{n-1}L\right)} - \frac{\frac{n(d_0-1)}{n-1}L}{(1-L)\left(1-\frac{nd_0-1}{n-1}L\right)}\right) & \frac{1}{n}\left(-\frac{\frac{n(d_0-1)}{n-1}L + \frac{L}{n-1}}{(1-L)\left(1-\frac{nd_0-1}{n-1}L\right)}\right) & \frac{1}{n}\left(-\frac{\frac{n(d_0-1)}{n-1}L + \frac{L}{n-1}}{(1-L)\left(1-\frac{nd_0-1}{n-1}L\right)}\right) & \cdots & \frac{1}{n}\left(-\frac{\frac{n(d_0-1)}{n-1}L + \frac{L}{n-1}}{(1-L)\left(1-\frac{nd_0-1}{n-1}L\right)}\right) \end{bmatrix}$$
$$= \begin{bmatrix} \frac{1}{\left(1-\frac{nd_0-1}{n-1}L\right)} - \frac{\frac{(d_0-1)}{n-1}L}{(1-L)\left(1-\frac{nd_0-1}{n-1}L\right)} & -\frac{\frac{(d_0-1)}{n-1}L + \frac{L}{(n-1)n}}{(1-L)\left(1-\frac{nd_0-1}{n-1}L\right)} & -\frac{\frac{(d_0-1)}{n-1}L + \frac{L}{(n-1)n}}{(1-L)\left(1-\frac{nd_0-1}{n-1}L\right)} & \cdots & -\frac{\frac{(d_0-1)}{n-1}L + \frac{L}{(n-1)n}}{(1-L)\left(1-\frac{nd_0-1}{n-1}L\right)} \end{bmatrix}.$$

Note that the term $\frac{L}{(n-1)n}$ is negligible so combining $\mathbf{y}_{n,t} = (I_n - \mathbf{A}_nL)^{-1}\epsilon_{n,t}$ with (27) we can write for

$\mathbf{y}_{1,t}$:

$$\mathbf{y}_{1,t} = \frac{\epsilon_{1,t}}{\left(1-\frac{nd_0-1}{n-1}L\right)} - \frac{\frac{(d_0-1)}{n-1}\sum_{j=1}^{n}\epsilon_{j,t-1}}{(1-L)\left(1-\frac{nd_0-1}{n-1}L\right)}$$

$$(1-L)\left(1-\frac{nd_0-1}{n-1}L\right)\mathbf{y}_{1,t} = (1-L)\,\epsilon_{1,t} - \frac{(d_0-1)}{n-1}\sum_{j=1}^{n}\epsilon_{j,t-1}$$

$$(1-L)\left(1-\frac{nd_0-1}{n-1}L\right)\mathbf{y}_{1,t} = (1-L)\,\epsilon_{1,t} + \frac{n\,(1-d_0)}{n-1}\frac{1}{n}\sum_{j=1}^{n}\epsilon_{j,t-1}$$

$$(1-L)\left(1-\frac{nd_0-1}{n-1}L\right)\mathbf{y}_{1,t} = (1-L)\,\epsilon_{1,t} + \frac{n\,(1-d_0)}{n-1}\frac{1}{n}\overline{\epsilon}_{t-1}$$

$$\overline{\epsilon}_{t-1} = \frac{1}{n}\sum_{j=1}^{n}\epsilon_{j,t-1}. \tag{28}$$

■

**Proof of Proposition 2:**

First, note that from (9)-(10) we can write:

$$y_{i,t} = \frac{d_0 n-1}{n-1}y_{i,t-1} + \frac{(1-d_0)\,n}{n-1}\overline{y}_{t-1} + \epsilon_{i,t}$$
$$\overline{y}_t = \overline{y}_{t-1} + \overline{\epsilon}_t$$
$$\overline{\epsilon}_t = \frac{1}{n}\sum_{i=1}^{n}\epsilon_{i,t}$$

with $\epsilon_t = (\epsilon_{1,t}, ..., \epsilon_{n,t})' \sim iid\,(0, \Sigma_n)$, $\sigma_{\overline{\epsilon}}^2 = n^{-2}1_n'\Sigma_n 1_n$. Then,

$$\begin{aligned}(1-L)\left(1-\frac{d_0 n-1}{n-1}L\right)y_{i,t} &= \frac{(1-d_0)\,n}{n-1}\overline{\epsilon}_{t-1} + (1-L)\,\epsilon_{i,t} \\ &= \epsilon_{i,t} - \epsilon_{i,t-1} + v_{n,t-1} \\ &= (1-\theta_n L)\,\eta_t\end{aligned} \tag{29}$$

where $v_{n,t} = \varphi_n\overline{\epsilon}_t$ is a white noise with variance $\sigma_v^2 = \varphi_n^2\sigma_{\overline{\epsilon}}^2$, $\varphi_n = \frac{(1-d_0)n}{n-1} > 0$. In what follows, we obtain an expression of $\theta_n$ and show that $\eta_t$ is a white noise with variance

$$\sigma_{\eta,i}^2 = \frac{1-\varphi_n/n}{1+\frac{1}{2}\left(\psi_n - \sqrt{\psi_n^2+2\psi_n}\right)}\sigma_i^2$$

where $\psi_n = \frac{\varphi_n^2 r_n}{1-\varphi_n/n} > 0$ and $r_n = \sigma_{\overline{\epsilon}}^2/\sigma_i^2 = \sigma_i^{-2}n^{-2}1_n'\Sigma_n 1_n$ is a signal-to-noise ratio.

Let us define the MA process $x_t = (1-L)\left(1-\frac{d_0 n-1}{n-1}L\right)y_{i,t} = v_{n,t-1} + (1-L)\,\epsilon_{i,t} = (1-\theta_n L)\,\eta_t$, which has an autocorrelation function given by:

$$\gamma_\tau = \begin{cases}\left(1+\theta_n^2\right)\sigma_{\eta,i}^2 & \tau = 0 \\ -\theta_n\sigma_{\eta,i}^2 & \tau = 1 \\ 0 & \tau \geq 2\end{cases},$$

Then, provided that $E\,(v_{n,t}\epsilon_{i,t}) = \varphi_n\sigma_i^2/n$ and $E\,(v_{n,t}v_{n,t-j}) = E\,(v_{n,t}\epsilon_{i,t-j}) = E\,(v_{n,t-j}\epsilon_{i,t}) = E\,(\epsilon_{i,t-j}\epsilon_{i,t}) = 0$ for $j \geq 1$

$$\begin{cases}E\left[(v_{n,t-1}+\epsilon_{i,t}-\epsilon_{i,t-1})^2\right] = \varphi_n^2\sigma_{\overline{\epsilon}}^2 + 2\,(1-\varphi_n/n)\,\sigma_i^2 \\ E\left[(v_{n,t-1}+\epsilon_{i,t}-\epsilon_{i,t-1})\,(v_{n,t-2}+\epsilon_{i,t-1}-\epsilon_{i,t-2})\right] = -\,(1-\varphi_n/n)\,\sigma_i^2 \\ E\left[(v_{n,t-1}+\epsilon_{i,t}-\epsilon_{i,t-1})\,(v_{n,t-j}+\epsilon_{i,t-j}-\epsilon_{i,t-j-1})\right] = 0 \text{ for } j \geq 2\end{cases},$$

so that,

$$\begin{cases}\varphi_n^2\sigma_{\overline{\epsilon}}^2 + 2\,(1-\varphi_n/n)\,\sigma_i^2 = \left(1+\theta_n^2\right)\sigma_{\eta,i}^2 \\ -\,(1-\varphi_n/n)\,\sigma_i^2 = -\theta_n\sigma_{\eta,i}^2\end{cases}$$

so $\sigma^2_{\eta,i} = (1 - \varphi_n/n)\,\sigma^2_i/\theta_n$, and adding the two equations

$$\begin{aligned}
\varphi_n^2\sigma_\epsilon^2 + 2\sigma_i^2 - 2\varphi_n\sigma_i^2/n - \sigma_i^2 + \varphi_n\sigma_i^2/n &= \sigma^2_{\eta,i} + \theta_n^2\sigma^2_{\eta,i} - \theta_n\sigma^2_{\eta,i} \\
\varphi_n^2\sigma_\epsilon^2 + \sigma_i^2 - \varphi_n\sigma_i^2/n &= \left(1 - \theta_n + \theta_n^2\right)\sigma^2_{\eta,i} \\
\frac{\varphi_n^2\sigma_\epsilon^2 + \sigma_i^2 - \varphi_n\sigma_i^2/n}{(1 - \varphi_n/n)\,\sigma_i^2}\theta_n &= 1 - \theta_n + \theta_n^2
\end{aligned}$$

so that we have to solve the quadratic equation $1 + b\theta_n + \theta_n^2 = 0$ where, defining $r_n = \sigma_\epsilon^2/\sigma_i^2 = \sigma_i^{-2}n^{-2}1_n'\Sigma_n 1_n > 0$ we may write,

$$\begin{aligned}
b &= -1 - \frac{\varphi_n^2\sigma_\epsilon^2 + \sigma_i^2 - \varphi_n\sigma_i^2/n}{(1 - \varphi_n/n)\,\sigma_i^2} \\
&= -\left(1 + \frac{1 - \varphi_n/n + \varphi_n^2\sigma_\epsilon^2/\sigma_i^2}{(1 - \varphi_n/n)}\right) \\
&= -\left(2 + \frac{\varphi_n^2 r_n}{1 - \varphi_n/n}\right) \\
&= -\left(2 + \psi_n\right),
\end{aligned}$$

where $\psi_n = \frac{\varphi_n^2 r_n}{1 - \varphi_n/n}$. Then

$$\begin{aligned}
\theta_n &= \frac{-b \pm \sqrt{b^2 - 4}}{2} \\
&= \frac{1}{2}\left(2 + \psi_n\right) \pm \frac{1}{2}\sqrt{\left(2 + \psi_n\right)^2 - 4} \\
&= 1 + \frac{1}{2}\left(\psi_n \pm \sqrt{\psi_n^2 + 4\psi_n}\right).
\end{aligned}$$

Note that $\varphi_n = \frac{(1-d_0)n}{n-1} \approx O(1)$, $\varphi_n/n \approx O\left(n^{-1}\right)$. Moreover, $r_n = \sigma_i^{-2}n^{-2}1_n'\Sigma_n 1_n \approx O\left(n^{-1}\right)$ if $\Sigma_n$ is diagonal or $r_n = \sigma_i^{-2}n^{-2}1_n'\Sigma_n 1_n \approx O(1)$ if not, so that $\varphi_n^2 r_n \approx O\left(n^{-1}\right)$ if $\Sigma_n$ is diagonal and $\varphi_n^2 r_n \approx O(1)$ if not. Then, when $\Sigma_n$ is diagonal

$$\psi_n = \frac{\varphi_n^2 r_n}{1 - \varphi_n/n} = O\left(n^{-1}\right)$$

and

$$\sqrt{\psi_n^2 - 2\psi_n} = \sqrt{O\left(n^{-2}\right) + O\left(n^{-1}\right)} = O\left(n^{-1/2}\right)$$

so that, discarting the non-invertible solution of the quadratic equation, we get

$$\theta_n = 1 - O\left(n^{-1/2}\right).$$

Moreover,

$$\begin{aligned}
\sigma^2_{\eta,i} &= \frac{1 - \varphi_n/n}{\theta_n}\sigma_i^2 \\
&= \frac{1 - \varphi_n/n}{1 + \frac{1}{2}\psi_n - \frac{1}{2}\sqrt{\psi_n^2 + 4\psi_n}}\sigma_i^2
\end{aligned}$$

and when $\Sigma_n$ is diagonal we may write,

$$\sigma^2_{\eta,i} = \sigma_i^2 + O\left(n^{-1}\right).$$

If $\Sigma_n$ is not diagonal, so that $r_n = \sigma_i^{-2}n^{-2}1_n'\Sigma_n 1_n \approx O(1)$, $\varphi_n^2 r_n \approx O(1)$ and $\psi_n = O(1)$, then $\theta_n = 1 - O(1)$.■

**Proof of the Corollary:**

If $n = O\left(T^\delta\right)$ and $\Sigma_n$ is diagonal, we have that $\varphi_n = \frac{(1-d_0)n}{n-1} \approx O(1)$, $\varphi_n/n \approx O\left(T^{-\delta}\right)$, $r_n = \sigma_i^{-2}n^{-2}1_n'\Sigma_n 1_n \approx O\left(T^{-\delta}\right)$, $\varphi_n^2 r_n \approx O\left(T^{-\delta}\right)$, and $\psi_n = \frac{\varphi_n^2 r_n}{1-\varphi_n/n} = O\left(T^{-\delta}\right)$, so that

$$\theta_n = 1 + \frac{1}{2}\left(\psi_n \pm \sqrt{\psi_n^2 + 4\psi_n}\right) = 1 - O\left(T^{-\delta/2}\right)$$

and

$$\sigma_{\eta,i}^2 = \frac{1-\varphi_n/n}{1+\frac{1}{2}\psi_n - \frac{1}{2}\sqrt{\psi_n^2+4\psi_n}}\sigma_i^2 = \sigma_i^2 + O\left(T^{-\delta}\right).$$

If $\Sigma_n$ is not diagonal, we have that $\varphi_n = \frac{(1-d_0)n}{n-1} \approx O(1)$, $\varphi_n/n \approx O\left(T^{-\delta}\right)$, $r_n = \sigma_i^{-2} n^{-2} 1_n' \Sigma_n 1_n \approx O(1)$, $\varphi_n^2 r_n \approx O(1)$, $\psi_n = \frac{\varphi_n^2 r_n}{1-\varphi_n/n} = O(1)$, so that

$$\theta_n = 1 + \frac{1}{2}\left(\psi_n \pm \sqrt{\psi_n^2+4\psi_n}\right) = 1 - O(1).$$

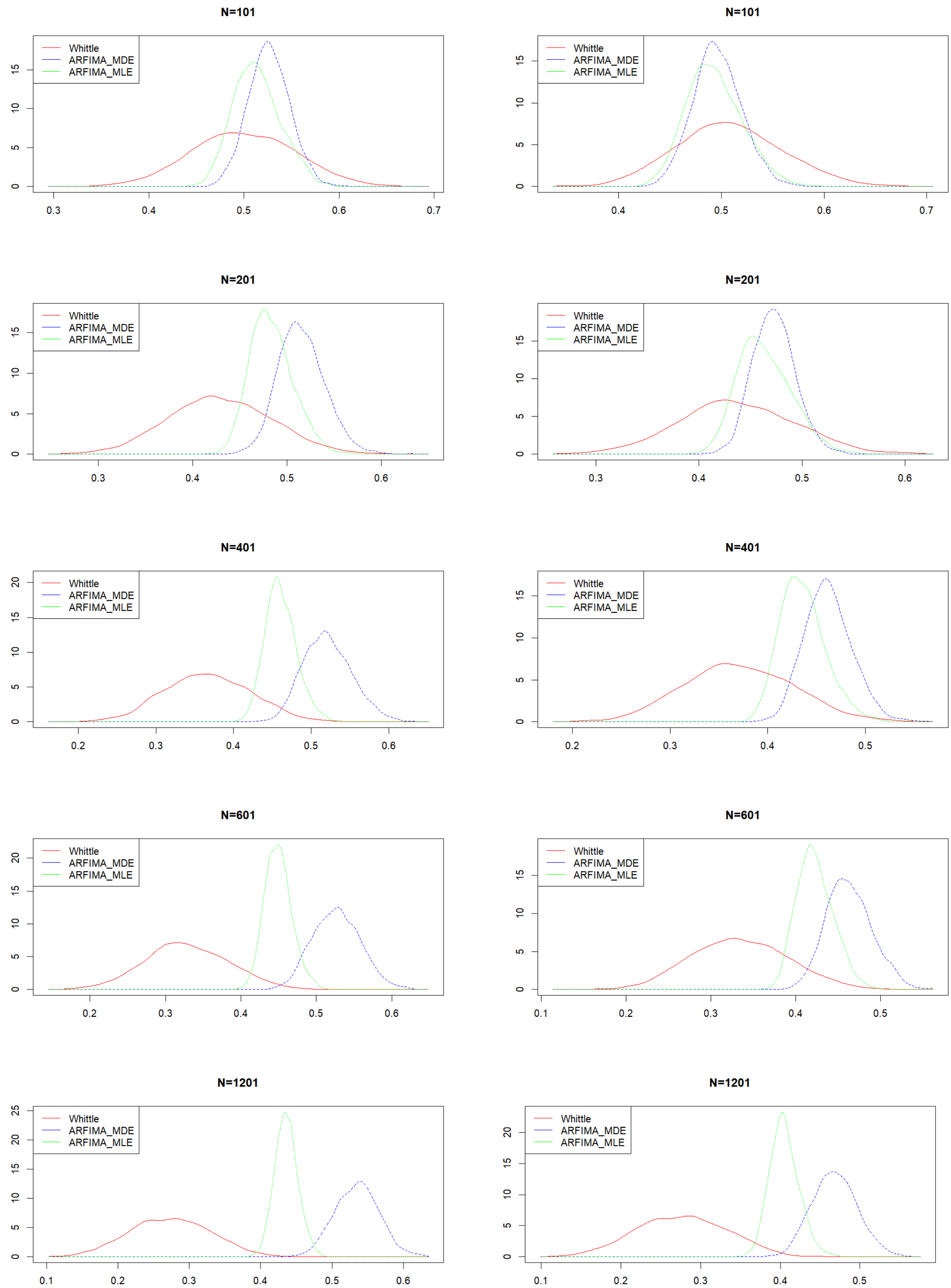


Figure 3: Simulation results for a $n$-dimensional VAR(1) model (1) with $\mathbf{A}_n$ as in (8) with $d_0 = 0.499$ in the left panel and $\mathbf{A}_n$ as in example 1 in CHL $d_0 = 0.45$ in the right panel. The empirical distribution of the local Whittle, the Pseudo Maximum Likelihood and the Minimum Distance estimator for an ARFIMA(0,d,0) model for $d$ with 4000 replications, $T = 4000$ and $n = \{101, 201, 401, 601, 1201\}$ is given.

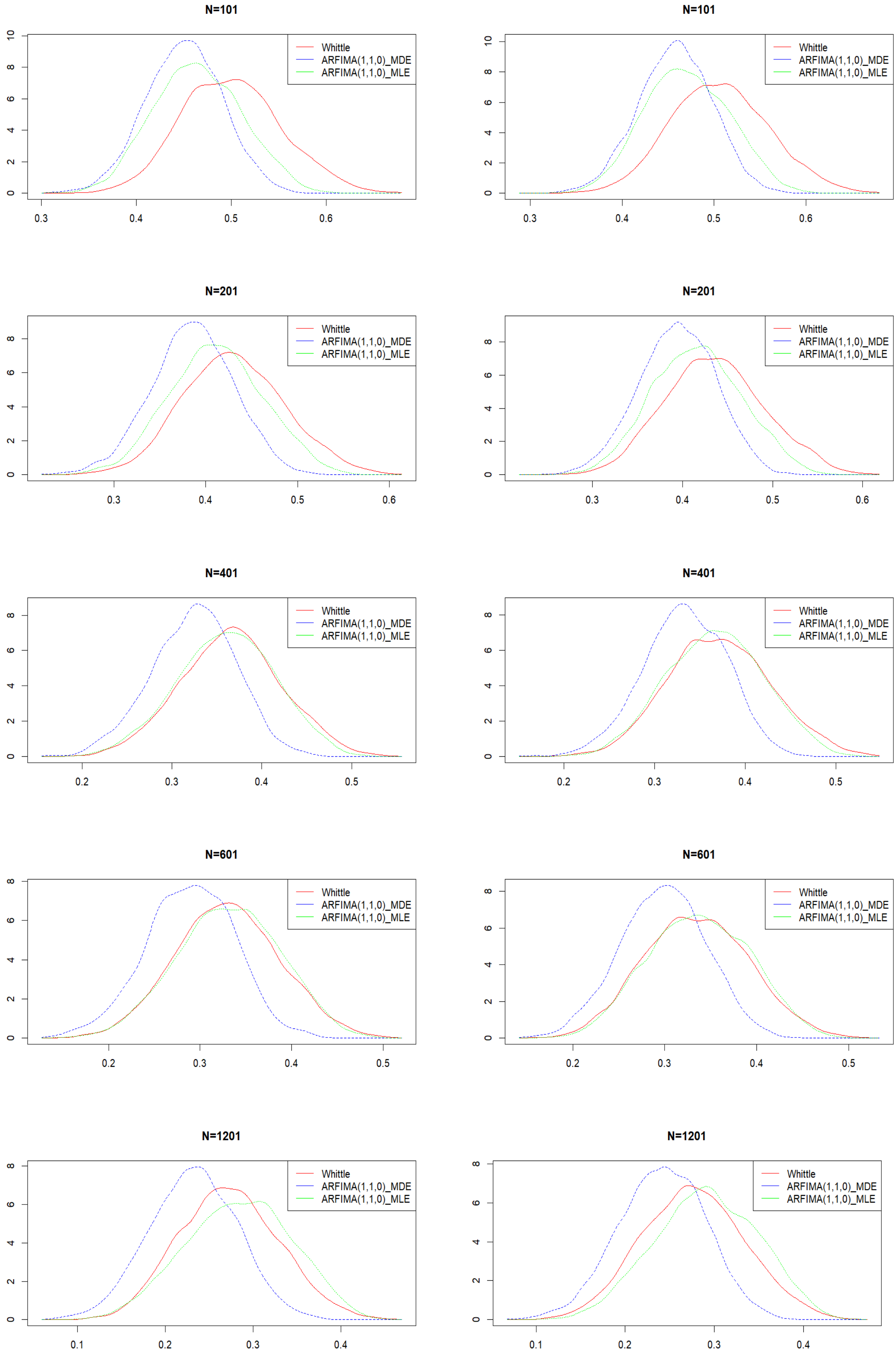


Figure 4: Simulation results as in figure 3 but for an ARFIMA(1,d,0) model.

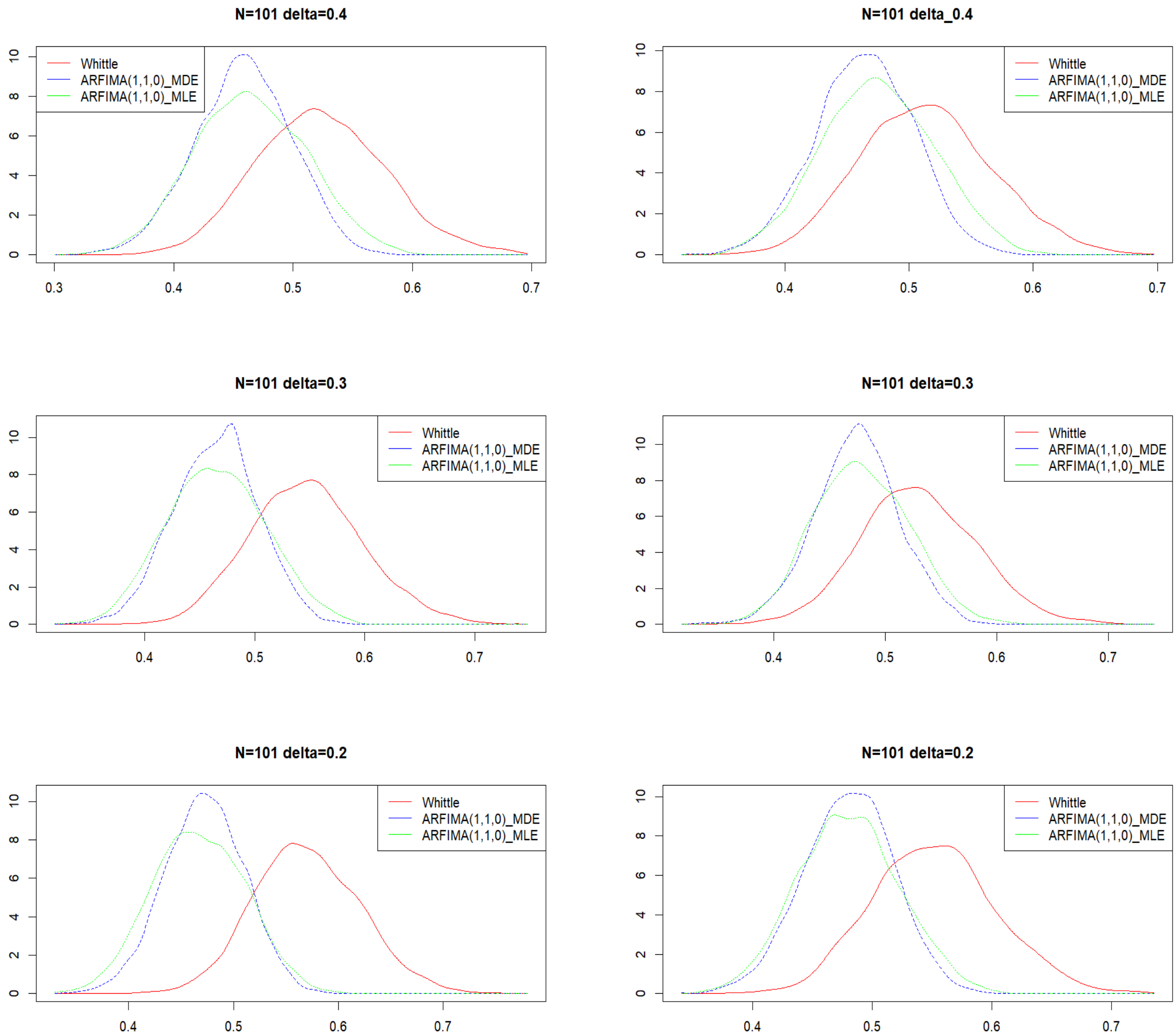


Figure 5: Simulation results as figure 4 with $n = 101$ and $d_0 = \{0.4, 0.3, 0.3\}$.

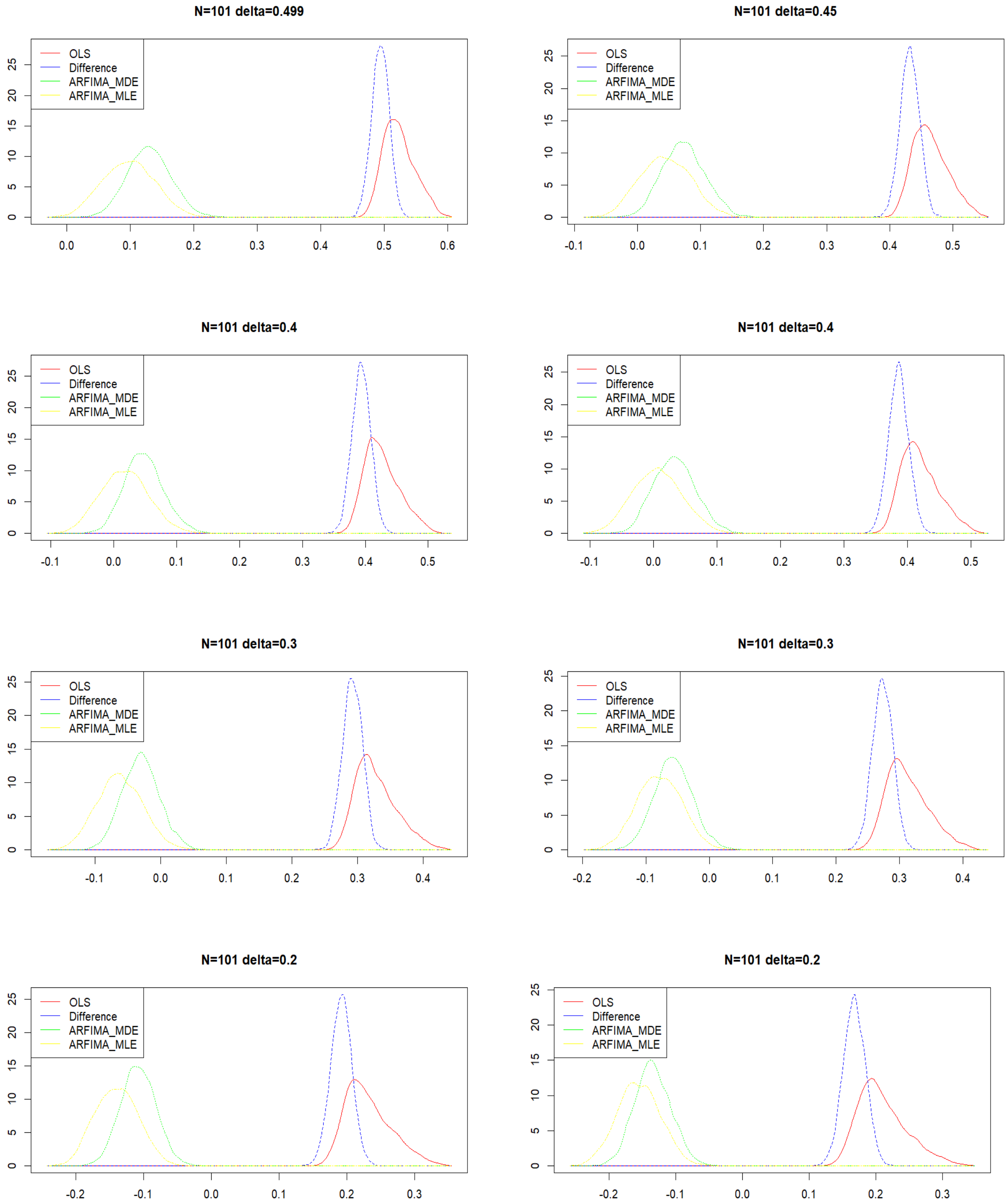


Figure 6: Simulation results for the empirical distribution of the AR(1) coefficient as in figure 5 but in the left panel with $d_0 = \{0.499, 0.4, 0.3, 0.3\}$ and in the right panel with $d_0 = \{0.499, 0.4, 0.3, 0.3\}$.

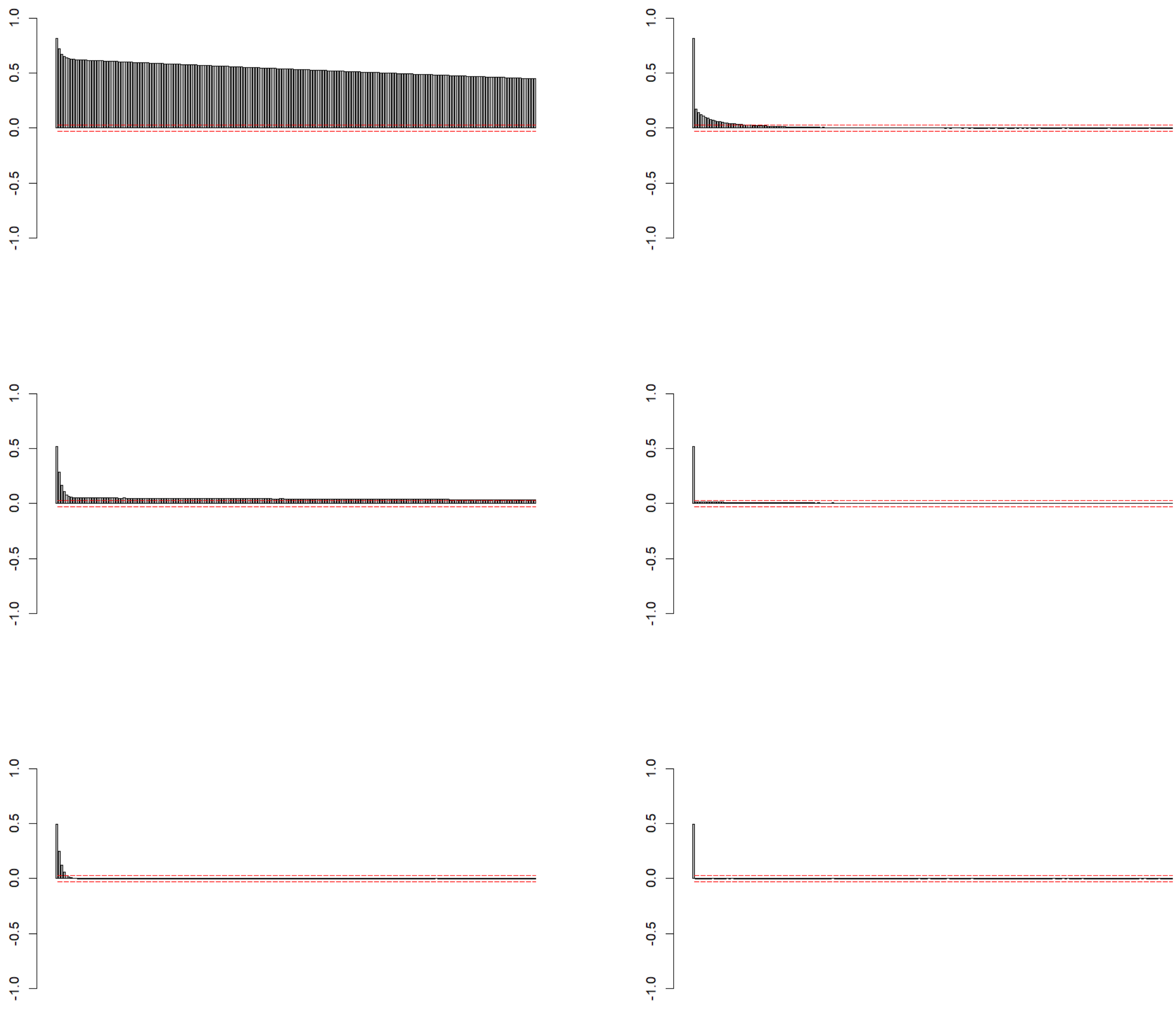


Figure 7: Simulation results for the averaged ACF and PACF of the marginalized time series in the first row and the residuals of the regression of the first on the second series in th esecond row and of the difference of the first and second series in the last row as in the left panel of figure 3 and $n = 201$.

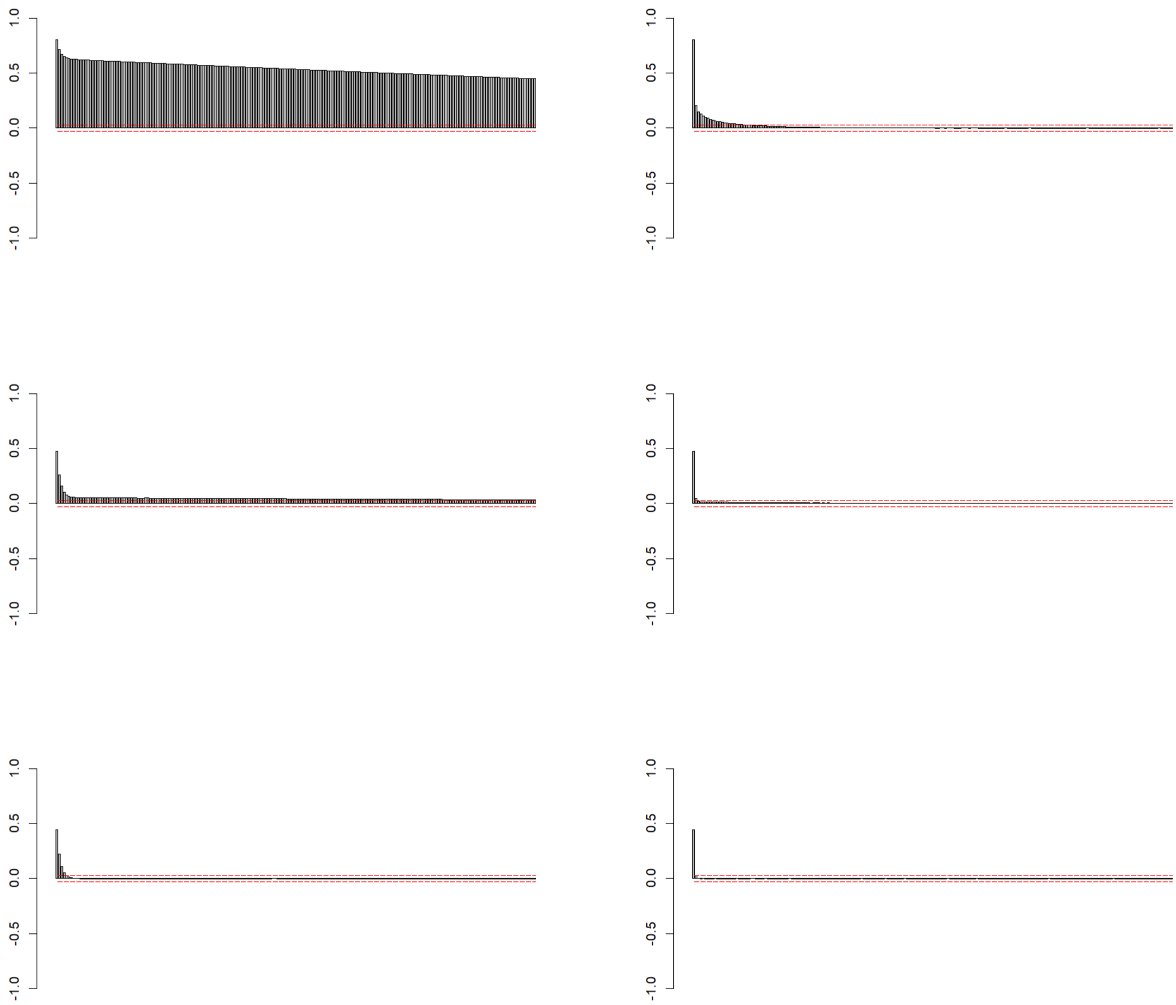

Figure 8: Simulation results for the averaged ACF and PACF of the marginalized time series in the first row and the residuals of the regression of the first on the second series in the second row and of the difference of the first and second series in the last row as in the right panel of figure 3 and $n = 201$.