Maximum Likelihood Estimation of Stationary Multivariate ARFIMA Processes

Wen-Jen Tsay*

(Received 00 Month 200x; In final form 00 Month 200x)

*The Institute of Economics, Academia Sinica, Taiwan

This paper considers the maximum likelihood estimation (MLE) of a class of stationary and invertible vector autoregressive fractionally integrated moving-average (VARFIMA) processes considered in (26) of Luceño [1] or Model A of Lobato [2] where each component $y_{i,t}$ is a fractionally integrated process of order d_i, $i = 1, \ldots, r$. Under the conditions outlined in Assumption 1 of this paper, the conditional likelihood function of this class of VARFIMA models can be efficiently and *exactly* calculated with a conditional likelihood Durbin-Levinson (CLDL) algorithm proposed herein. This CLDL algorithm is based on the multivariate Durbin-Levinson algorithm of Whittle [3] and the conditional likelihood principle of Box and Jenkins [4]. Furthermore, the conditions in the aforementioned Assumption 1 are general enough to include the model considered in Andersen et al. [5] for describing the behavior of realized volatility and the model studied in Haslett and Raftery [6] for spatial data as its special cases. As the computational cost of implementing the CLDL algorithm is much lower than that of using the algorithms proposed in Sowell [7], we are thus able to conduct a Monte Carlo experiment to investigate the finite sample performance of the CLDL algorithm for the 3-dimensional VARFIMA processes with the sample size of 400. The simulation results are very satisfactory and reveal the great potentials of using the CLDL method for empirical applications.

Key words: Durbin-Levinson algorithm; Long memory; Maximum likelihood estimation; Multivariate time series

AMS 2000 Subject classification: 60G18; 65C20

1 Introduction

Consider the maximum likelihood estimation (MLE) of a class of stationary and invertible vector autoregressive fractionally integrated moving-average (VARFIMA) processes:

$$\Phi(B)\mathrm{diag}(\nabla^d)Y_t = \Theta(B)Z_t, \tag{1}$$

where $Y_t = (y_{1,t}, \ldots, y_{r,t})^\top$, $t = 1, 2, \ldots, T$, is an $r-$dimensional vector of observations of interest, and $\Phi(B)$ and $\Theta(B)$ are finite order matrix polynomials in B (usual lag operator), such that:

$$\Phi(B) = \Phi_0 - \Phi_1 B - \ldots - \Phi_p B^p, \qquad \Theta(B) = \Theta_0 + \Theta_1 B + \ldots + \Theta_q B^q, \quad (2)$$

$\Phi_0 = \Theta_0 = I_r$, I_r is an $(r \times r)$ identity matrix, and the diagonal matrix $\mathrm{diag}(\nabla^d)$ is defined as:

$$\mathrm{diag}(\nabla^d) = \begin{bmatrix} \nabla^{d_1} & 0 & \ldots & 0 \\ 0 & \nabla^{d_2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \nabla^{d_r} \end{bmatrix}, \quad (3)$$

with $\nabla = 1 - B$, and $d_i \in (-1/2, 1/2)$ for all $i = 1, 2, \ldots, r$. Here, $Z_t = (z_{1,t}, \ldots, z_{r,t})^\top$ in (1) is an $r-$dimensional independent and identically distributed (i.i.d.) white noise process with a non-singular covariance matrix Σ.

Assuming that $\Phi(B)$ and $\Theta(B)$ satisfy the usual stationary and invertible conditions, respectively, the multivariate process Y_t can be represented with the following MA representation:

$$Y_t = \sum_{j=0}^{\infty} \Psi_j Z_{t-j} = \sum_{j=0}^{\infty} \Psi_j B^j Z_t, \quad (4)$$

where $\Psi_0 = I_r$, and the $(r \times r)$ coefficient matrices Ψ_j are often referred to as impulse response functions. Note that Ψ_j decays at a slow hyperbolic rate when the differencing parameters are not equal to 0. When $r = 1$, this process was first introduced by Granger [8], Granger and Joyeux [9], and Hosking [10]. They show that Y_t is stationary if $d < 1/2$. As $d > 0$, Y_t is said to have a long memory, because the autocovariance functions of the univariate process Y_t are not absolutely summable. See Beran [11] for the overviews of long memory processes.

The model in (1) has been considered in (26) of Luceño [1] and is exactly the Model A investigated in Lobato [2]. Each component $y_{i,t}$ in (1) is a fractionally integrated process of order d_i, $i = 1, \ldots, r$, denoted in (3). Sowell [7] is the pioneer to calculate the likelihood function of the VARFIMA model in (1). Nevertheless, the presence of AR parameters greatly complicates the computation of the corresponding autocovariance functions, because the algorithms of Sowell [7] [12] involve hypergeometric functions which need to be evaluated with a truncated infinite sum, no matter whether the data are univariate or multivariate. On the other hand, Luceño [1] suggests an *approximate*

log-likelihood function based on the autocovariance function of the 'inverse-transpose' model. As a consequence, a rounding error is inevitable from using the approximate procedure of Luceño [1] and those of Sowell [7] [12]. The impacts of rounding errors generated from these algorithms on the estimation results might not be trivial, especially when the dimensionality of the data series is relatively large.

The contribution of this paper is to show that the conditional likelihood function of the VARFIMA process in (1) can be evaluated *exactly* and efficiently if the model in (1) can be represented as:

$$\text{diag}(\nabla^d)\Phi(B)Y_t = \Theta(B)Z_t. \tag{5}$$

For example, the distinction between the models in (1) and (5) is irrelevant if the data series is univariate, i.e., $r = 1$. Under the multivariate time series scenario, the models in (1) and (5) are identical if either $\Phi(B)$ is diagonal or the values of differencing parameters remain intact across $i = 1, \ldots, r$ as imposed in Assumption 1 in the next section. Thus, each component $y_{i,t}$ in (5) is still a fractionally integrated process of order d_i defined in (3) given that the above-mentioned two conditions are satisfied. In fact, these conditions have been employed in the multivariate long memory literature, e.g., Haslett and Raftery [6] impose a homogeneous structure on the fractional differencing and ARMA parameters across meteorological stations to describe the wind speeds recorded at 12 synoptic meterorological stations in Ireland when using the VARFIMA model for their spatial data. Moreover, Andersen et al. [5] employ the long-memory Gaussian trivariate VAR model in their equation (15) to describe the realized logarithm volatilities of exchange rates where the values of differencing parameters are also identical across $i = 1, \ldots, r$. The reason why they impose such a homogenous structure on the value of fractional differencing parameters is to reduce the computational burdens of estimating the VARFIMA model whose dimensionality and span of data series are both large. However, the methodology developed in this paper allows us to deal with the VARFIMA model in (1) where the integration of each component $y_{i,t}$ is different across, $i = 1, \ldots, r$.

The model in (5) has been discussed in (27) of Luceño [1]. It is also equivalent to Model B considered in Lobato [2] and can be estimated with Luceño's [1] *approximate* log-likelihood function. As argued previously, a rounding error is inevitable when using the approximate procedure of Luceño [1]. On the contrary, given that the model in (1) can be transformed as the one in (5), we show that the conditional likelihood function of the VARFIMA process in (1) can be evaluated *exactly* and efficiently by combining the multivariate Durbin-Levinson algorithm of Whittle [3] with the conditional likelihood principle of Box and Jenkins [4] even though the dimensionality and span of data

series are relatively large. We name this procedure as the conditional likelihood Durbin-Levinson (CLDL) algorithm. In the literature, Deriche and Tewfik [13] and Tsay and Härdle [14] estimate the univariate ARFIMA process with the univariate Durbin-Levinson algorithm, while Doornik and Ooms [15] suggest using Whittle's [3] method for VARFIMA models but with no implementation.

Because the VARFIMA processes considered in Andersen et al. [5] belong to the subcases of models in (1), the proposed CLDL algorithm is readily applied to the data of Andersen et al. [5]. Most importantly, the CLDL algorithm will provide a much more efficient estimate than the equation-by-equation ordinary least squares (OLS) method of Andersen et al. [5] who fix the value of d in (1) with a common estimate 0.401 across three data series.

Another valuable feature of the CLDL algorithm is its computational gain over the original algorithm of Sowell [7]. It is well known that the computational burdens of applying Sowell's [7] algorithm to the VARFIMA data are tremendous. Particularly, when studying the joint behavior of U.S. and Canadian bond rates, Dueker and Startz [16] demonstrate that it takes about 35 minutes on a 200-MHz PC for each iteration of the MLE of a bivariate VARFIMA process with 121 observations and 18 parameters when implementing Sowell's [7] algorithm. However, it takes about less than 40 seconds on a 1066-MHz PC for each iteration of the conditional MLE of a bivariate VARFIMA$(1, d, 1)$ process with 400 observations and 11 parameters when implementing the proposed CLDL algorithms. This clearly demonstrates the power of the CLDL algorithm for many potential empirical applications. It also explains why we can conduct a Monte Carlo experiment to investigate the finite sample performance of the CLDL algorithm for 3-dimensional VARFIMA processes under the sample size up to 400.

The remaining parts of this paper are arranged as follows: Section 2 presents the autocovariance functions of a VARFIMA$(0, d, q)$ process and the implementation of the multivariate Durbin-Levinson algorithm. Section 3 considers a class of VARFIMA(p, d, q) processes with which we can compute their corresponding conditional likelihood function exactly with the CLDL algorithm. The finite sample performance of the CLDL algorithm is investigated through a Monte Carlo experiment in Section 4. Section 5 provides a conclusion.

2   Autocovariances of VARFIMA$(0, d, q)$ Processes

The model in (5) is essentially different from the one in (1) on the ordering of $\Phi(B)$ and diag(∇^d). The major idea of this paper is to show that the conditional likelihood function of the VARFIMA process in (1) can be evaluated *exactly* and efficiently if the model in (1) can be transformed as the one in (5). Assumption 1 summarizes these conditions.

Assumption 1. Given that the data is generated by (1), we assume (i) $\Phi(B)$ is diagonal, or (ii) the values of differencing parameters d_i remain intact across $i = 1, \ldots, r$.

Provided that Assumption 1 holds, the difference in choosing between the models in (1) and (5) no longer exists. Consequently, we impose the conditions in Assumption 1 throughout this paper, and each component $y_{i,t}$ in model (1) or (5) is a fractionally integrated process of order d_i, $i = 1, \ldots, r$, as defined in (3). Item (ii) of Assumption 1 allows us to estimate the data of Andersen et al. [5] and those of Haslett and Raftery [6] with the proposed CLDL algorithm explained more clearly later. The advantages of using item (i) of Assumption 1 are more evident when the dimensionality of the data series is relatively large. For example, when $r = 7$, we may require that $\Phi(B)$ is diagonal, and the associated number of parameters for such a 7-dimensional VARFIMA$(1, d, 1)$ process is only 91 as compared to 133 parameters within a VARFIMA$(1, d, 1)$ process without being imposed any restriction. In so doing, the resulting simplified model is still quite flexible, because we do not impose any constraint on the structure of MA parameters, however, the computational burden of estimating such a simplified model is much mitigated based on the CLDL algorithm as has been documented in the introduction section.

Before implementing the MLE for the model in (1), we need to compute the autovariance functions of the VARFIMA models. We first consider the simplest case where Y_t is generated as a VARFIMA$(0, d, 0)$ model:

$$Y_t = \mathrm{diag}(\nabla^{-d})Z_t, \tag{6}$$

or

$$y_{i,t} = \sum_{j=0}^{\infty} \psi_{i,j} z_{i,t-j}, \quad \psi_{i,j} = \frac{\Gamma(j + d_i)}{\Gamma(j+1)\Gamma(d_i)}, \quad i = 1, 2, \ldots, r, \tag{7}$$

where $\Gamma(.)$ is the gamma function. Denoting $\Omega(h) = E(Y_{t+h}Y_t')$, $h = -T + 1, -T + 2, \ldots, 0, 1, 2, \ldots, T - 2, T - 1$, the autocovariance function of Y_t in (6)

at lag h is:

$$\Omega(h) \equiv \begin{bmatrix} \sigma_{11} \sum_{j=0}^{\infty} \psi_{1,j}\psi_{1,j+h} & \cdots & \sigma_{1r} \sum_{j=0}^{\infty} \psi_{1,j}\psi_{r,j+h} \\[2mm] \sigma_{21} \sum_{j=0}^{\infty} \psi_{2,j}\psi_{1,j+h} & \cdots & \sigma_{2r} \sum_{j=0}^{\infty} \psi_{2,j}\psi_{r,j+h} \\[2mm] \vdots & \ddots & \vdots \\[2mm] \sigma_{r1} \sum_{j=0}^{\infty} \psi_{r,j}\psi_{1,j+h} & \cdots & \sigma_{rr} \sum_{j=0}^{\infty} \psi_{r,j}\psi_{r,j+h} \end{bmatrix}, \tag{8}$$

where σ_{mn} represents the (m,n)th element of Σ.

From Abramowitz and Stegun [17] (p. 556, §15.1.1, §15.1.20), we note that the hypergeometric function is defined as:

$$F(a,b;c;z) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)} \sum_{n=0}^{\infty} \frac{\Gamma(a+n)\Gamma(b+n)}{\Gamma(c+n)\Gamma(n+1)} z^n,$$

and it fulfills the following relationship:

$$F(a,b;c;1) = \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)}, \qquad c \neq 0, -1, -2, \ldots, c-a-b > 0.$$

Using the above results, we derive the (m,n)th element of $\Omega(h)$ in (8) as:

$$\Omega_{m,n}(h) \equiv \sigma_{mn} \sum_{j=0}^{\infty} \psi_{m,j}\psi_{n,j+h} = \frac{\sigma_{mn}\Gamma(1-d_m-d_n)}{\Gamma(d_n)\Gamma(1-d_n)} \frac{\Gamma(h+d_n)}{\Gamma(h+1-d_m)}. \tag{9}$$

Luceño [1] (p. 611) independently presents the formula for $\Omega_{m,n}(h)$ in his equation (8), but indirectly.

For computational efficiency, it is well known in the literature that $\Omega_{m,n}(h)$ in (9) can be calculated as follows:

$$\sigma_{mn} \sum_{j=0}^{\infty} \psi_{m,j}\psi_{n,j+h} = \frac{\sigma_{mn}\Gamma(1-d_m-d_n)}{\Gamma(1-d_m)\Gamma(1-d_n)} \prod_{0<k\leq h} \frac{k-1+d_n}{k-d_m}, \tag{10}$$

where $m, n = 1, 2, \ldots, r$, and $h = 1, 2, \ldots$.

Combining the results in (9) and (10) with Whittle's [3] multivariate Durbin-Levinson algorithm, we evaluate the exact unconditional likelihood function of a VARFIMA$(0, d, 0)$ model, $L(\Phi, \Theta, d, \Sigma, Y)$, as:

$$(2\pi)^{-rT/2} \left\{ \prod_{j=1}^{T} \det(V_{j-1}) \right\}^{-1/2} \exp \left\{ -\frac{1}{2} \sum_{j=1}^{T} \left(Y_j - \widehat{Y}_j\right)^{\top} V_{j-1}^{-1} \left(Y_j - \widehat{Y}_j\right) \right\},$$

(11)

where $\widehat{Y}_j$ denotes the one-step ahead predictor of Y_j with the observation $Y(j-1) = (Y_1^{\top}, Y_2^{\top}, \ldots, Y_{j-1}^{\top})^{\top}$ as $j \geq 2$. Here, V_{j-1} is the corresponding one-step ahead prediction error matrix. As $j = 1$, $\widehat{Y}_1 = 0$, and $V_0 = \Omega(0)$. For the definition and computation of $\widehat{Y}_j$ and those of V_{j-1}, see Whittle [3] or Proposition 11.4.1 of Brockwell and Davis [18] for details.

When Y_t is a VARFIMA$(0, d, q)$ process, i.e.:

$$\mathrm{diag}(\nabla^d)Y_t = \Theta(B)Z_t,$$

(12)

the (m, n)th element of its corresponding autocovariance function, $\Omega(h)$, is:

$$
\begin{aligned}
\Omega_{m,n}(h) \equiv \quad & \Omega^* \left(\sigma_{mn}\Theta_{mm,0}\Theta_{nn,0} \right) \\
+ \Omega^* & \left\{ \sum_{f=1}^{q} \sum_{g=1}^{q} \sum_{u=1}^{r} \sum_{v=1}^{r} \sigma_{uv}\Theta_{mu,f}\Theta_{nv,g} \frac{\Gamma(h + d_n + g - f)\Gamma(h + 1 - d_m)}{\Gamma(h + d_n)\Gamma(h + 1 - d_m + g - f)} \right\} \\
+ \Omega^* & \left\{ \sum_{f=1}^{q} \sum_{u=1}^{r} \sigma_{nu}\Theta_{mu,f} \frac{\Gamma(h + d_n - f)}{\Gamma(h + d_n)} \frac{\Gamma(h + 1 - d_m)}{\Gamma(h + 1 - d_m - f)} \right\} \\
+ \Omega^* & \left\{ \sum_{f=1}^{q} \sum_{u=1}^{r} \sigma_{mu}\Theta_{nu,f} \frac{\Gamma(h + d_n + f)}{\Gamma(h + d_n)} \frac{\Gamma(h + 1 - d_m)}{\Gamma(h + 1 - d_m + f)} \right\},
\end{aligned}
$$

(13)

where

$$\Omega^* = \frac{\Gamma(1 - d_m - d_n)}{\Gamma(d_n)\Gamma(1 - d_n)} \frac{\Gamma(h + d_n)}{\Gamma(h + 1 - d_m)},$$

(14)

and $\Theta_{mn,k}$ denotes the (m, n)th element of Θ_k in (1). Note that there are $(rq + 1)^2$ terms at the right-hand side of (13). With the autocovariance functions in (13), we apply the multivariate Durbin-Levinson algorithm of Whittle [3] to the VARFIMA$(0, d, q)$ processes as we have suggested for the VARFIMA$(0, d, 0)$ processes.

Sowell [7] provides a formula to compute the autocovariance functions of the VARFIMA$(0, d, q)$ process in page 12 and page 14 of his manuscript. In this

paper, we present the formula in (13) more explicitly, thus providing an easy access to empirical applications. Furthermore, similar to the results in (10), the ratios of gamma functions in (13) can be computed with the following relationship:

$$\frac{\Gamma(h + d_n + l)}{\Gamma(h + d_n)} \frac{\Gamma(h + 1 - d_m)}{\Gamma(h + 1 - d_m + l)} = \begin{cases} \prod_{k=1}^{l} \dfrac{h + d_n + k - 1}{h - d_m + k}, & \text{if} \quad l > 0, \\[2ex] 1, & \text{if} \quad l = 0, \\[2ex] \prod_{k=1}^{|l|} \dfrac{h - d_m - k + 1}{h + d_n - k}, & \text{if} \quad l < 0, \end{cases} \quad (15)$$

indicating that the unconditional likelihood function of an $r-$dimensional VARFIMA$(0, d, q)$ process can be evaluated efficiently and without a rounding error.

For the more restrictive VARFIMA$(0, d, q)$ process where $\Theta(B)$ is diagonal, the value of $\Omega_{m,n}(h)$ in (13) becomes:

$$\Omega_{m,n}(h) \equiv \sigma_{mn} \, \Omega^* \left\{ \sum_{l=-q}^{q} \Psi(l) \frac{\Gamma(h + d_n + l)}{\Gamma(h + d_n)} \frac{\Gamma(h + 1 - d_m)}{\Gamma(h + 1 - d_m + l)} \right\}, \quad (16)$$

where

$$\Psi(l) = \begin{cases} \sum_{k=0}^{q-|l|} \Theta_{mm,k}\Theta_{nn,k+|l|}, & \text{if} \quad l > 0, \\[2ex] \sum_{k=0}^{q} \Theta_{mm,k}\Theta_{nn,k}, & \text{if} \quad l = 0, \\[2ex] \sum_{k=0}^{q-|l|} \Theta_{nn,k}\Theta_{mm,k+|l|}, & \text{if} \quad l < 0. \end{cases} \quad (17)$$

Now there are only $(q+1)^2$ terms at the right-hand side of (16). As $r = 1$, the autocovariance function in (16) is equivalent to Sowell's [12] (p. 173) formula.

3   CLDL Algorithm for VARFIMA(p, d, q) Processes

We now show that the conditional likelihood function of the model in (1) under Assumption 1 can be evaluated exactly. From (1) or (5), we have

$\Phi(B)Y_t = \text{diag}(\nabla^{-d})\Theta(B)Z_t$, and the term $\text{diag}(\nabla^{-d})\Theta(B)Z_t$ at the right-hand side of the equality sign corresponds to being a VARFIMA$(0, d, q)$ process whose autocovariances are displayed in (13). Adopting the idea of the conditional likelihood function in Box and Jenkins [4] (Chapter 7), we transform Y_t in (1) into $\text{diag}(\nabla^{-d})\Theta(B)Z_t$ for a given choice of parameters in $\Phi(B)$ and of suitable starting values. Particularly, if $p = 1$, then conditional on Φ_1 and Y_1, $Y_t - \Phi_1 Y_{t-1}$, $t = 2, 3, \ldots, T$, is a VARFIMA$(0, d, q)$ process, and we denote its associated conditional likelihood function as:

$$L\left(\Phi, \Theta, d, \Sigma, Y \mid Y_1\right) \equiv L\left(0, \Theta, d, \Sigma, Y_t - \Phi_1 Y_{t-1}\right), \quad \text{for} \quad p = 1, \, t = 2, \ldots, T. \tag{18}$$

Applying the multivariate Durbin-Levinson algorithm of Whittle [3] to the transformed data, $Y_t - \Phi_1 Y_{t-1}$, we simultaneously estimate all the parameters Φ, Θ, d, and Σ of the VARFIMA(p, d, q) process in (1) with the numerical optimization method. No rounding error occurs during the evaluation of the conditional likelihood function of the VARFIMA(p, d, q) processes in (18).

There are three remarks to be addressed here. First, similar to Li and McLeod [19], the CLDL algorithm is a one-step procedure, because we estimate all the parameters Φ, Θ, d, and Σ in (1) simultaneously. Second, if $\Theta(B)$ in (1) is diagonal, then we observe:

$$\Phi(B)Y_t = \text{diag}(\nabla^{-d})\Theta(B)Z_t = \Theta(B)\text{diag}(\nabla^{-d})Z_t = \Theta(B)\widetilde{Y}_t, \tag{19}$$

where $\widetilde{Y}_t = \text{diag}(\nabla^{-d})Z_t$ is a VARFIMA$(0, d, 0)$ process. Again, conditional on a set of $\Phi(B)$ and $\Theta(B)$ and of a suitable starting value, we can evaluate the conditional likelihood function of the process in (19) based on a VARFIMA$(0, d, 0)$ process, just like when we evaluate the conditional likelihood function of a VARMA$(p, 0, q)$ process based on a vector white noise process. The same transformation method for the univariate ARFIMA process has been adopted in Dueker and Serletis [20] and Tsay and Härdle [14].

Third, in contrast to Li and McLeod [19] who first apply a binominal transformation to the data, Y_t, and then calculate the likelihood function for the remaining AR and MA parameters, we initially transform Y_t with a set of initial values of AR parameters and then compute the exact likelihood function of the remaining VARFIMA$(0, d, q)$ process. In so doing, we avoid the bias inherent in many estimation methods for stationary univariate ARFIMA processes which truncate the infinite sum that defines $(1 - B)^d$, including both time-domain and frequency-domain methods. This truncated infinite sum problem was first pointed out by Sowell [12], who argues that the transformed series from a truncated infinite fractional differencing transformation is the sum of an ARMA model and a linear combination of an infinite number of unobserved terms. Thus, ARMA parameters can never be correctly estimated from

the estimators based on a truncated infinite fractional differencing transformation. Therefore, the estimate from the equation-by-equation OLS method of Andersen et al. [5] by fixing the value of d in (1) with a common estimate 0.401 across three data series may be problematic. See Sowell [12] (p. 170) for a detailed discussion about this truncation problem.

4   Monte Carlo Experiment

This section considers the finite sample performance of the MLE of VARFIMA processes with the proposed CLDL algorithm. To our knowledge, this is the first simulation study pertaining to the full time-domain MLE of the VARFIMA processes. Without loss of generality, we focus on the process where all the fractional differencing parameters are positive in that the long memory process is the major concern of the fractionally integrated literature. Moreover, we divide our studies into two parts. The first part deals with the 3-dimensional VARFIMA$(0, d, 1)$ process, while the second part considers the 2-dimensional VARFIMA$(1, d, 1)$ one.

4.1 Three-dimensional VARFIMA$(0, d, 1)$

We first consider the following 3-dimensional VARFIMA$(0, d, 1)$ process for the Monte Carlo experiment:

$$
\begin{bmatrix} \nabla^{d_1} y_{1,t} \\ \nabla^{d_2} y_{2,t} \\ \nabla^{d_3} y_{3,t} \end{bmatrix} = \begin{bmatrix} 1 + \Theta_{11,1}B & 0 & 0 \\ 0 & 1 + \Theta_{22,1}B & 0 \\ 0 & 0 & 1 + \Theta_{33,1}B \end{bmatrix} \begin{bmatrix} z_{1,t} \\ z_{2,t} \\ z_{3,t} \end{bmatrix}. \tag{20}
$$

Note that AR parameters are not included in (20). Moreover, the specification in (20) reveals that $y_{i,t}$ is a fractionally integrated process of order d_i.

The parameters considered for the simulations are:

$$
(d_1, d_2, d_3)^\top = (0.4, 0.3, 0.2)^\top, \quad (\Theta_{11,1}, \Theta_{22,1}, \Theta_{33,1})^\top = (-0.7, -0.5, -0.3)^\top, \tag{21}
$$

$$
(z_{1,t}, z_{2,t}, z_{3,t})^\top = N(0, \Sigma), \qquad \Sigma = \begin{bmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{bmatrix}, \qquad \text{and} \qquad \rho = \{0.2, 0.5\}. \tag{22}
$$

Given the specification in (20), we adopt the Cholesky decomposition algorithm suggested by McLeod and Hipel [21] and Hosking [22] to simulate three $(T \times 1)$ ARFIMA processes $z_{i,t}$, $i = 1, 2, 3$.

We estimate the parameter:

$$\xi = (d_1, d_2, d_3, \Theta_{11,1}, \Theta_{22,1}, \Theta_{33,1})^\top, \tag{23}$$

with the following transformation function:

$$\xi = \kappa(\widetilde{\xi}), \tag{24}$$

where

$$\kappa(\widetilde{\xi}) = \left(\frac{\widetilde{d}_1}{1 + 2|\widetilde{d}_1|}, \frac{\widetilde{d}_2}{1 + 2|\widetilde{d}_2|}, \frac{\widetilde{d}_3}{1 + 2|\widetilde{d}_3|}, \frac{e^{\widetilde{\Theta}_{11,1}} - 1}{e^{\widetilde{\Theta}_{11,1}} + 1}, \frac{e^{\widetilde{\Theta}_{22,1}} - 1}{e^{\widetilde{\Theta}_{22,1}} + 1}, \frac{e^{\widetilde{\Theta}_{33,1}} - 1}{e^{\widetilde{\Theta}_{33,1}} + 1} \right)^\top, \tag{25}$$

and

$$\widetilde{\xi} = (\widetilde{d}_1, \widetilde{d}_2, \widetilde{d}_3, \widetilde{\Theta}_{11,1}, \widetilde{\Theta}_{22,1}, \widetilde{\Theta}_{33,1})^\top, \tag{26}$$

are the parameters really estimated with the proposed algorithm. In order to create a realistic simulation scheme, we do not use the inverse function of the preceding transformation function calculated at the true parameter value of ξ as the initial value for estimation procedure. Instead, we employ the true value of ξ as the initial value for $\widetilde{\xi}$. In other words, in the simulation we use:

$$\widetilde{\xi}_0 = (d_1, d_2, d_3, \Theta_{11,1}, \Theta_{22,1}, \Theta_{33,1})^\top, \tag{27}$$

as the initial value for $\widetilde{\xi}$ in (24). This is the first mechanism to create a realistic simulation scheme.

For the estimation of Σ in (22), we employ a (3×3) upper triangular matrix U to conduct the following transformation:

$$\Sigma = U^\top U. \tag{28}$$

The parameters contained in U are the parameters really estimated with the proposed algorithm. However, the starting values for each element of U are all set to be 1. Thus, the initial value of U for starting the procedure is not set to be the true value of U. This is the second mechanism to create a realistic simulation environment.

All the programs are written in GAUSS evaluated at three sample sizes, $T = 200, 300, 400$. The choice of these sample sizes strongly reveals that the

computational burden of implementing the CLDL algorithm is mild. Moreover, these sample sizes are frequently encountered with the usual macroeconomic time series.

Two hundred additional values are generated in order to obtain random starting values. The optimization algorithm used to implement the CLDL algorithm is the quasi-Newton algorithm of Broyden, Fletcher, Goldfarb, and Shanno (BFGS) contained in the GAUSS MAXLIK library. The maximum number of iterations for each replication is 100. The first 250 replications of normal convergence are recorded for the subsequent data analysis.

Define bias as the true parameter values minus the corresponding average estimated values. The simulation results in Table 1 for $\rho = 0.2$ and Table 2 for $\rho = 0.5$ reveal that the bias performance from using the unconditional maximum likelihood estimator is satisfactory and the associated root-mean-squared error (RMSE) decreases with the increase of sample size for all configurations considered in the experiment.

< Insert Table 1 here >

< Insert Table 2 here >

In order to give a clearer picture about the finite sample performance of the CLDL algorithm, we also report the simulation results with box-plots in Figures 1, 2, and 3 for the case $\rho = 0.2$, and in Figures 4, 5, and 6 for the case $\rho = 0.5$. Figure 1 and Figure 4 show the box-plots of the estimated d, while Figures 2 and 5 illustrate the box-plots of the estimated Θ. The box-plots of the estimated Σ are displayed in Figures 3 and 6. All these figures demonstrate that the performance of the CLDL algorithm performs well at a larger sample size as shown in Table 1 and Table 2, supporting the usefulness of the CLDL algorithm in estimating the VARFIMA processes displayed in (20).

< Insert Figure 1 here >

< Insert Figure 2 here >

< Insert Figure 3 here >

< Insert Figure 4 here >

< Insert Figure 5 here >

< Insert Figure 6 here >

We also compare the performance of the CLDL algorithm in estimating the model in (20) with that of the semiparametric QMLE in (8) of Lobato [23]. The distinguished feature of the QMLE is that it is designed to be robust to the presence of AR and MA parameters. Thus, it can generate robust estimates of fractional differencing parameters in (1) without worrying about the order of the AR and MA polynomials. However, the implementation of Lobato's [23] QMLE requires selecting a number of Fourier frequencies $\lambda_j = 2\pi j/T$ with $j = 1, \ldots, m$ for estimation. The choice of m is crucial on the performance of the semiparametric QMLE. In this section we select two values of $m = [T^{0.5}]$ and $m = [T^{0.65}]$, where $[X]$ is the largest integer less than or equal to X.

We employ a sequential quadratic programming algorithm "sqpSolve" contained in the GAUSS library to implement the QMLE of Lobato [23]. Similar to the experimental design for the CLDL algorithm, two hundred additional values are generated in order to obtain random starting values. Again, the first 250 replications of normal convergence are recorded for the subsequent data analysis as we have done for the CLDL algorithm.

In contrast with the CLDL algorithm, where the maximum number of iterations for each replication is 100, we do not impose a restriction on the maximum number of iteration when conducting the QMLE. This is the first design which favors the implementation of QMLE relative to the CLDL algorithm. Furthermore, the true parameter values of d_i in (20) are set as the initial value for estimation procedure. This is the second design which favors the implementation of QMLE relative to the CLDL algorithm. If the simulation results show that the performance of the CLDL algorithm in estimating the fractional differencing parameters is still better than that of the QMLE of Lobato [23] under the aforementioned experimental design, then we can accrue this finding to the outstanding performance of MLE over its semiparametric counterpart under correct model specification scheme.

For ease of comparison, the simulation results of the QMLE are also displayed in Table 1 for $\rho = 0.2$ and Table 2 for $\rho = 0.5$. Tables 1 and 2 clearly show that performance of the QMLE improves when $m = [T^{0.5}]$ is used as compared to the other choice $m = [T^{0.65}]$. This finding is reasonable, because the design of QMLE is motivated by examining the spectral density function close to the zero frequency of a long memory process. These tables also reveal

that, for all 3 differencing parameters considered in the experment, the perfor-
mance of the CLDL algorithm is much better than that of QMLE even when
$m = [T^{0.5}]$ is chosen, because the RMSE of the CLDL algorithm is less than
that of the QMLE, respectively. This is as expected in that the CLDL algo-
rithm is a conditional maximum likelihood estimator which should be more
efficient than its semiparametric counterpart under correct model specifica-
tion.

The simulation results from the CLDL and QMLE are also graphed with
box-plots for clarity of exposition. Figure 7 displays the results for the case
$\rho = 0.2$, while Figure 8 contains the findings for the case $\rho = 0.5$. To save space,
we only present the simulations concerning the parameter $d_1 = 0.4$. Note that
the results from QMLE are based on the choice $m = [T^{0.5}]$ for the reason
outlined previously. It is clear from these two figures that the performance of
the CLDL algorithm is much better than that of the QMLE.

< Insert Figure 7 here >

< Insert Figure 8 here >

4.2 Two-dimensional VARFIMA$(1, d, 1)$

This subsection considers the impacts of AR parameters on the performance
of the CLDL algorithms when the off-diagonal elements of $\Phi(B)$ in (5) are not
equal to zero. To save the loadings of estimation, we focus on the following
2-dimensional VARFIMA$(1, d, 1)$ model:

$$\begin{bmatrix} \nabla^{d_1} & 0 \\ 0 & \nabla^{d_2} \end{bmatrix} \begin{bmatrix} 1 - \Phi_{11,1}B & -\Phi_{12,1}B \\ -\Phi_{21,1}B & 1 - \Phi_{22,1}B \end{bmatrix} \begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \widetilde{Z}_t, \qquad (29)$$

where

$$\widetilde{Z}_t = \begin{bmatrix} 1 + \Theta_{11,1}B & 0 \\ 0 & 1 + \Theta_{22,1}B \end{bmatrix} \begin{bmatrix} z_{1,t} \\ z_{2,t} \end{bmatrix}. \qquad (30)$$

Note that AR parameters are included in (29) as compared to the model in
(20). The presence of non-zero off-diagonal elements in $\Phi(B)$ indicates that
the conditions imposed in Assumption 1 are no longer binding for the model
defined in (29) and (30) which is a special case of the model in (5). Thus,

the order of integration of $y_{i,t}$ is not equal to d_i any more. This implies that the fractional differencing parameters d_1 and d_2 in (29) cannot be consistently estimated with the QMLE of Lobato [23], because it is the linear combination of $y_{i,t}$, $y_{i,t-1}$, and $y_{j,t-1}$ with $j \neq i$ is a fractionally integrated process of order d_i. However, the CLDL algorithm still can deal with the model defined in (29) and (30) easily.

The parameters considered for the simulations are:

$$(d_1, d_2)^\top = (0.4, 0.2)^\top, \quad (\Theta_{11,1}, \Theta_{22,1})^\top = (-0.7, -0.3)^\top, \qquad (31)$$

$$(\Phi_{11,1}, \Phi_{12,1}, \Phi_{21,1}, \Phi_{22,1})^\top = (-0.5, -0.6, 0.7, -0.8)^\top, \qquad (32)$$

$$(z_{1,t}, z_{2,t})^\top = N(0, \Sigma), \qquad \Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}, \qquad \text{and} \qquad \rho = \{0.2, 0.5\}. \quad (33)$$

The principle of setting initial values for the CLDL algorithms and the QMLE of Lobato [23] is identical to that outlined in subsection 4.1, respectively. In addition, the ways of choosing initial values for the AR parameters are identical to those of choosing initial values for the MA parameters outlined in subsection 4.1 for the CLDL algorithm.

The simulations are graphed with box-plots in Figures 9, 10, and 11 for the case $\rho = 0.2$, and in Figures 12, 13, and 14 for the case $\rho = 0.5$. Figure 9 and Figure 12 exhibit the box-plots of the estimated d, Figures 10 and 13 demonstrate the box-plots of the estimated Θ, and Figure 11 and Figure 14 present the box-plots of the estimated Φ. To save space, we do not present the box-plots of the estimated Σ in this paper, but point out that they are similar to those shown in Figures 3 and 6. All these figures demonstrate that the performance of the CLDL algorithm improves with the increasing sample sizes and the bias from using the CLDL is very small as we find in the studies concerning the 3-dimensional VARFIMA$(0, d, 1)$ process in subsection 4.1. These simulation results explicitly support the usefulness of the CLDL algorithm in estimating the VARFIMA$(1, d, 1)$ processes displayed in (29) and (30).

< Insert Figure 9 here >

< Insert Figure 10 here >

< Insert Figure 11 here >

< Insert Figure 12 here >

< Insert Figure 13 here >

< Insert Figure 14 here >

We also prepare a comparison between the QMLE and the CDLD algorithm in estimating the fractional differencing parameter under the model in (29) and (30). Before discussing the findings, we emphasize here that the order of integration of $y_{1,t}$ in (29) is not 0.4 due to the presence of a non-zero off-diagonal element shown in $\Phi(B)$ of (29). Thus, the estimates for d_1 from using the QMLE of course will not be close to the true parameter value 0.4. Indeed, in Figure 15 for the case $\rho = 0.2$ and in Figure 16 for the case $\rho = 0.5$, respectively, we find the estimates of d_1 based on Lobato's [23] QMLE are far away from the true parameter value $d_1 = 0.4$. On the contrary, the conditional MLE based on the CLDL algorithm provides a promising estimate of d_1 no matter whether the value of ρ is 0.2 or 0.5, indicating the excellent ability of the proposed CLDL algorithm in estimating all the AR, MA, and fractional differencing parameters under the model in (5).

< Insert Figure 15 here >

< Insert Figure 16 here >

5   Conclusion

We explicitly illustrate the analytic formulae of the autocovariance functions of an $r-$dimensional VARFIMA$(0, d, q)$ process and numerically demonstrate that the unconditional MLE of the VARFIMA$(0, d, q)$ processes can be implemented efficiently and exactly with the multivariate Durbin-Levinson algorithm of Whittle [3] via a Monte Carlo experiment. In the presence of AR parameters, we further demonstrate that an exact conditional MLE can be implemented with the proposed CLDL algorithm for the VARFIMA processes displayed in (1) under Assumption 1. As a consequence, the problems of rounding error generated from Sowell's [7] [12] and Luceño's [1] algorithms are com-

pletely avoided with the joint use of the CLDL algorithm and the conditions in Assumption 1. In fact, the CLDL algorithm is the only procedure which is free of rounding error in the multivariate long memory literature. This feature is important when we do not have a clue to evaluate the impacts of rounding errors on the precision of estimation results, especially when the dimensionality and sample length of the data series are relatively large as we find in many meterorological and spatial data.

The coverage of the methodology based on Assumption 1 and the CLDL algorithm is general enough to include the model considered in Andersen et al. [5] and the model studied in Haslett and Raftery [6]. Given the good performance of the long-memory Gaussian VAR model proposed by Andersen et al. [5], the advantages of applying the CLDL algorithm to the time-varying return volatility environment are more evident, because we now are in a much better position to deal with this kind of financial data whose dimensionality and span of data series are both large. Furthermore, it is well known that Haslett and Raftery [6] impose a homogeneous structure on the fractional differencing and ARMA parameters in order to describe the wind speeds recorded at 12 synoptic meterorological stations in Ireland due to the tremendous computational burden of using the unconstrained VARFIMA model. Essentially, the development of the CLDL algorithm greatly relaxes the restrictions imposed in the contemporaneous ARFIMA model of Haslett and Raftery [6] and sheds more lights on the modeling of long memory space-time data.

Reference

[1] Luceño, A., 1996, A fast likelihood approximation for vector general linear processes with long series: Application to fractional differencing. *Biometrika*, **83**, 603-614.

[2] Lobato, I. N., 1997, Consistency of the averaged cross-periodogram in long memory series. *Journal of Time Series Analysis*, **18**, 137-155.

[3] Whittle, P., 1963, On the fitting of multivariate autoregressions and the approximate canonical factorization of a spectral density matrix. *Biometrika*, **50**, 129-134.

[4] Box, G. E. P. and Jenkins, G. M., 1976, *Time Series Analysis, Forecasting, and Control*, 2nd Ed. Holden-Day, San Francisco.

[5] Andersen, T. G., Bollserslev, T., Diebold, F.X. and Labys, P., 2003, Modeling and forecasting realized volatility. *Econometrica*, **71**, 579-625.

[6] Haslett, J. and Raftery, A. E., 1989, Space-time modelling with long-memory dependence: Assessing Ireland's wind power resource. *Applied Statistics*, **38**, 1-50.

[7] Sowell, F., 1989, Maximum likelihood estimation of fractionally integrated time series models. Working paper, Carnegie-Mellon University.

[8] Granger, C. W. J., 1980, Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics,* **14**, 227–238.

[9] Granger, C. W. J. and Joyeux, R., 1980, An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis,* **1**, 15–29.

[10] Hosking, J. R. M., 1981, Fractional differencing. *Biometrika,* **68**, 165–176.

[11] Beran, J., 1994, *Statistics for Long-Memory Processes.* Chapman and Hall, New York.

[12] Sowell, F., 1992, Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics*, **53**, 165-188.

[13] Deriche, J. A. and Tewfik, A. H., 1993, Maximum likelihood estimation of the parameters of discrete fractionally differenced Gaussian noise process. *IEEE Transactions on Signal Processing,* **41**, 2977–2989.

[14] Tsay, W. J. and Härdle, W. K., 2008, A generalized ARFIMA process with Markov-switching fractional differencing parameter. *Journal of Statistical Computation and Simulation*, forthcoming.

[15] Doornik, J. A. and Ooms, M., 2003, Computational aspects of maximum likelihood estimation of autoregressive fractionally integrated moving average models. *Computational Statistics & Data Analysis*, **42**, 333-348.

[16] Dueker, M. and Startz, R., 1998, Maximum-likelihood estimation of fractionally cointegration with an application to U.S. and Canadian Bond Rates. *Review of Economics and Statistics*, **80**, 420-426.

[17] Abramowitz, M. and Stegun, I. A., 1970, *Handbook of Mathematical Functions*. Dover, New York.

[18] Brockwell, P. J. and Davis, R. A., 1991, *Time Series: Theory and Methods*, 2nd Ed. Springer-Verlag, New York.

[19] Li, W. K. and McLeod, A. I., 1986, Fractional time series modelling. *Biometrika*, **73**, 217-221.

[20] Dueker, M. and Serletis, A., 2000, Do real exchange rates have autoregressive unit roots? A test under the alternative of long memory and breaks. Working Paper 2000-016A, Federal Reserve Bank of St. Louis.

[21] McLeod, A. I. and Hipel, K. W., 1978, Preservation of the rescaled adjusted range 1: A reassessment of the Hurst phenomenon. *Water Resources Research*, **14**, 491-508.

[22] Hosking, J. R. M., 1984, Modeling persistence in hydrological time series using fractional differencing. *Water Resources Research*, **20**, 1898-908.

[23] Lobato, I. N., 1999, A semiparametric two-step estimator in a multivariate long memory model. *Journal of Econometrics*, **90**, 129-153.

Table 1.   Bias and RMSE from Estimating a 3-dimensional VARFIMA$(0, d, 1)$ Process

		$d_1 = 0.4$	$d_2 = 0.3$	$d_3 = 0.2$	$\Theta_{11,1} = -0.7$	$\Theta_{22,1} = -0.5$	$\Theta_{33,1} = -0.3$
				$\rho = 0.2$			
				CLDL			
$T = 200$	Bias	0.0926	0.0289	-0.0193	-0.0782	-0.0223	0.0357
	RMSE	0.1681	0.1060	0.1114	0.1739	0.1268	0.1434
$T = 300$	Bias	0.0757	0.0205	-0.0247	-0.0685	-0.0174	0.0383
	RMSE	0.1327	0.0899	0.0903	0.1391	0.1005	0.1147
$T = 400$	Bias	0.0723	0.0181	-0.0304	-0.0688	-0.0159	0.0441
	RMSE	0.1218	0.0819	0.0816	0.1290	0.0915	0.1023
				QMLE			
$m = [T^{0.50}]$							
$T = 200$	Bias	0.2012	0.0847	0.0419	NA	NA	NA
	RMSE	0.2814	0.2078	0.1920	NA	NA	NA
$T = 300$	Bias	0.1403	0.0614	0.0208	NA	NA	NA
	RMSE	0.2126	0.1760	0.1682	NA	NA	NA
$T = 400$	Bias	0.1158	0.0525	0.0097	NA	NA	NA
	RMSE	0.1874	0.1581	0.1379	NA	NA	NA
$m = [T^{0.65}]$							
$T = 200$	Bias	0.3538	0.1716	0.0799	NA	NA	NA
	RMSE	0.3743	0.2077	0.1430	NA	NA	NA
$T = 300$	Bias	0.2948	0.1397	0.0620	NA	NA	NA
	RMSE	0.3102	0.1714	0.1131	NA	NA	NA
$T = 400$	Bias	0.2679	0.1206	0.0461	NA	NA	NA
	RMSE	0.2818	0.1484	0.0920	NA	NA	NA

Notes: The results are all based on 250 replications. Details of the experimental designs are given in (20), (21), and (22). CLDL denotes the conditional likelihood Durbin-Levinson algorithm proposed in this paper. QMLE denotes the QMLE method considered in Lobato [23], and m is the maximum number of Fourier frequencies $\lambda_j = 2\pi j/T$ with $j = 1, \ldots, m$ used for implementing QMLE. NA denotes not available.

Table 2.   Bias and RMSE from Estimating a 3-dimensional VARFIMA$(0, d, 1)$ Process

		$d_1 = 0.4$	$d_2 = 0.3$	$d_3 = 0.2$	$\Theta_{11,1} = -0.7$	$\Theta_{22,1} = -0.5$	$\Theta_{33,1} = -0.3$
				$\rho = 0.5$			
				CLDL			
$T = 200$	Bias	0.1027	0.0157	-0.0537	-0.0915	-0.0069	0.0782
	RMSE	0.1447	0.0837	0.1066	0.1449	0.1024	0.1349
$T = 300$	Bias	0.0925	0.0072	-0.0614	-0.0865	-0.0011	0.0853
	RMSE	0.1264	0.0709	0.0961	0.1275	0.0819	0.1240
$T = 400$	Bias	0.0888	0.0058	-0.0659	-0.0859	-0.0004	0.0889
	RMSE	0.1171	0.0651	0.0913	0.1186	0.0737	0.1181
				QMLE			
$m = [T^{0.50}]$							
$T = 200$	Bias	0.1947	0.0805	0.0466	NA	NA	NA
	RMSE	0.2630	0.1909	0.1862	NA	NA	NA
$T = 300$	Bias	0.1383	0.0592	0.0344	NA	NA	NA
	RMSE	0.2048	0.1672	0.1650	NA	NA	NA
$T = 400$	Bias	0.1120	0.0488	0.0194	NA	NA	NA
	RMSE	0.1765	0.1497	0.1357	NA	NA	NA
$m = [T^{0.65}]$							
$T = 200$	Bias	0.3541	0.1722	0.0826	NA	NA	NA
	RMSE	0.3700	0.2010	0.1379	NA	NA	NA
$T = 300$	Bias	0.2964	0.1407	0.0652	NA	NA	NA
	RMSE	0.3106	0.1663	0.1104	NA	NA	NA
$T = 400$	Bias	0.2676	0.1217	0.0492	NA	NA	NA
	RMSE	0.2803	0.1440	0.0893	NA	NA	NA

Notes: The results are all based on 250 replications. Details of the experimental designs are given in (20), (21), and (22). CLDL denotes the conditional likelihood Durbin-Levinson algorithm proposed in this paper. QMLE denotes the QMLE method considered in Lobato [23], and m is the maximum number of Fourier frequencies $\lambda_j = 2\pi j/T$ with $j = 1, \ldots, m$ used for implementing QMLE. NA denotes not available.
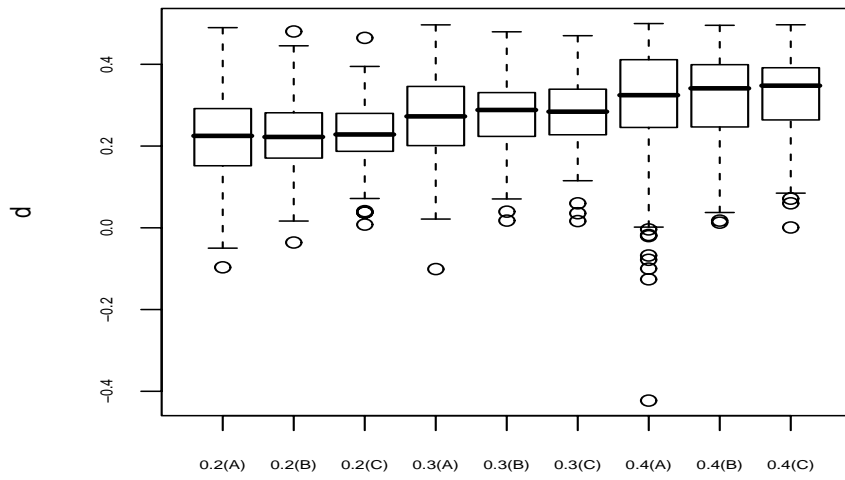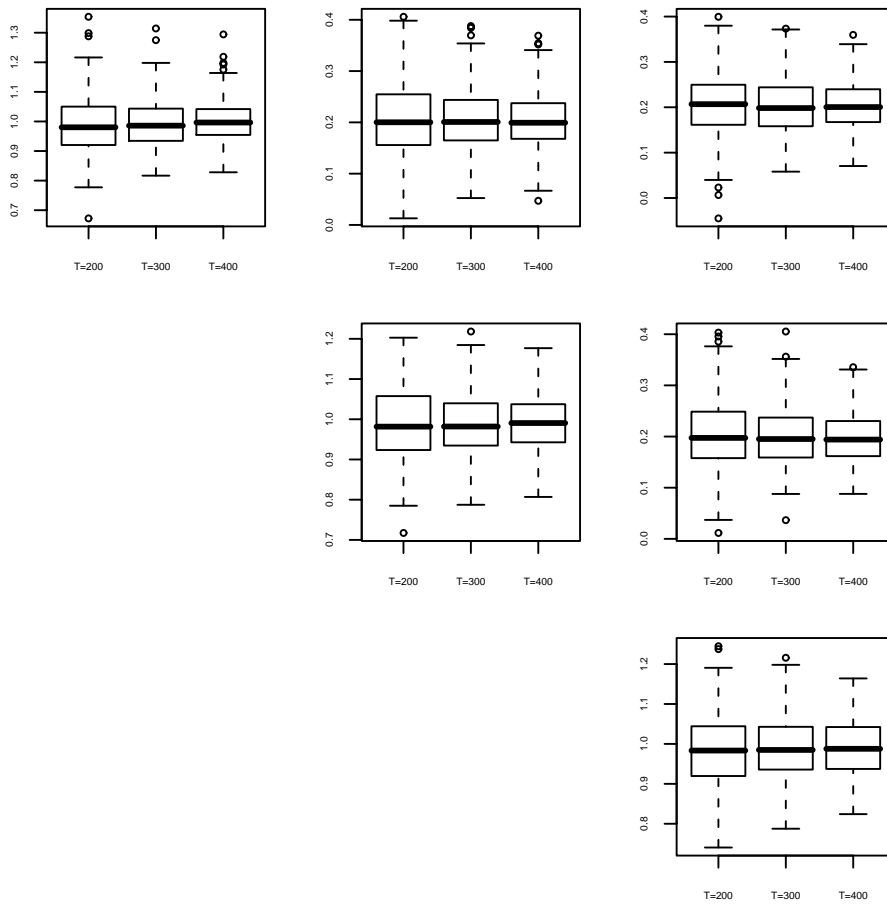
Figure 1.   Box-plots of the estimated d from the 3-dimensional VARFIMA$(0, d, 1)$ model defined in (20), (21), (22), and $\r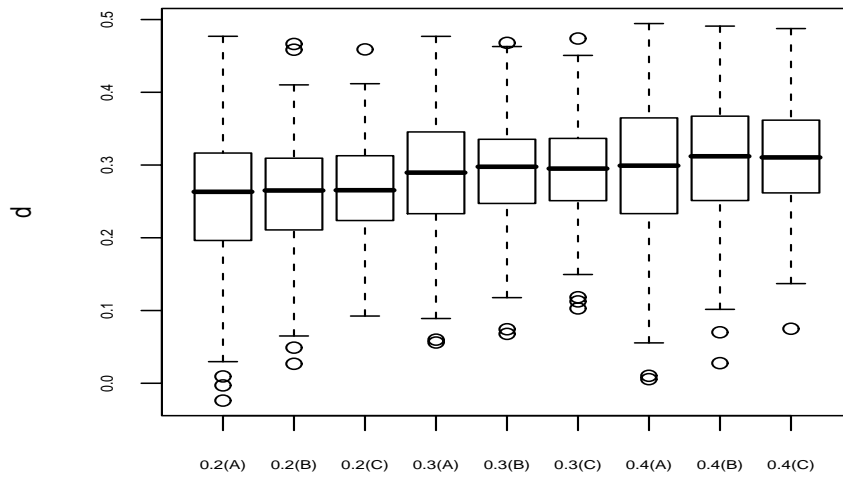ho = 0.2$ based on the CLDL algorithm with 250 replications. The value $f(g)$ denotes the model specification where $f = d$, g denotes the sample size, such that g=A=200, g=B=300, and g=C=400.
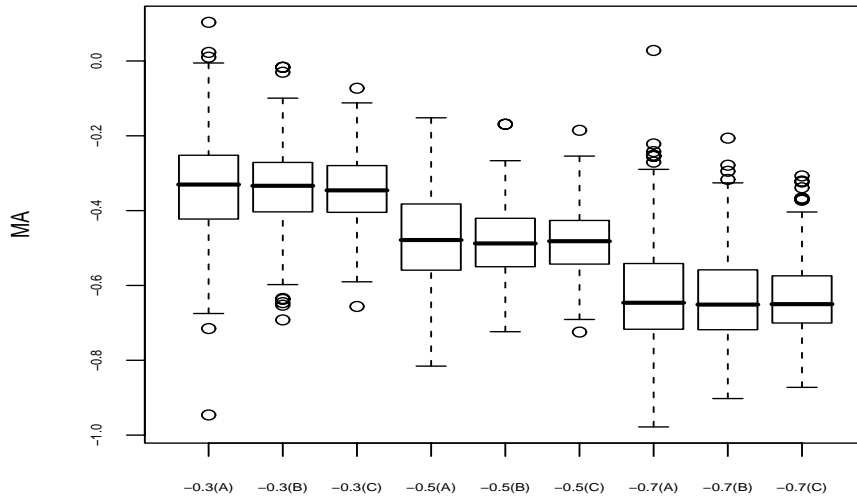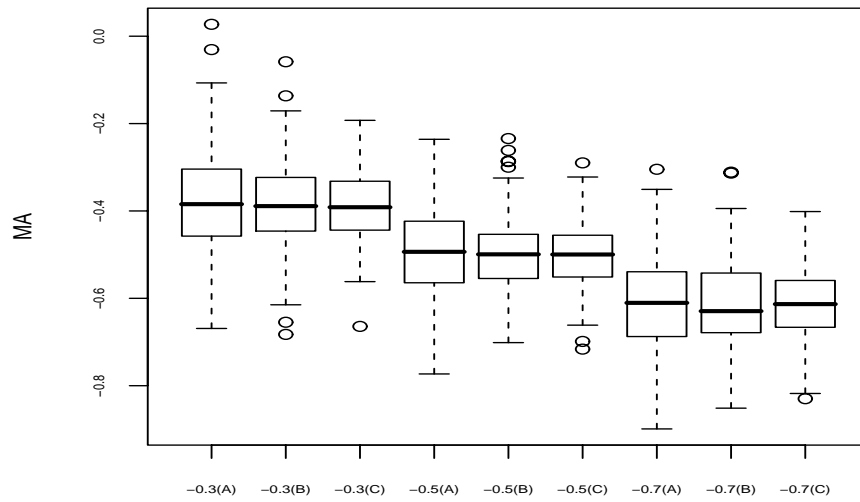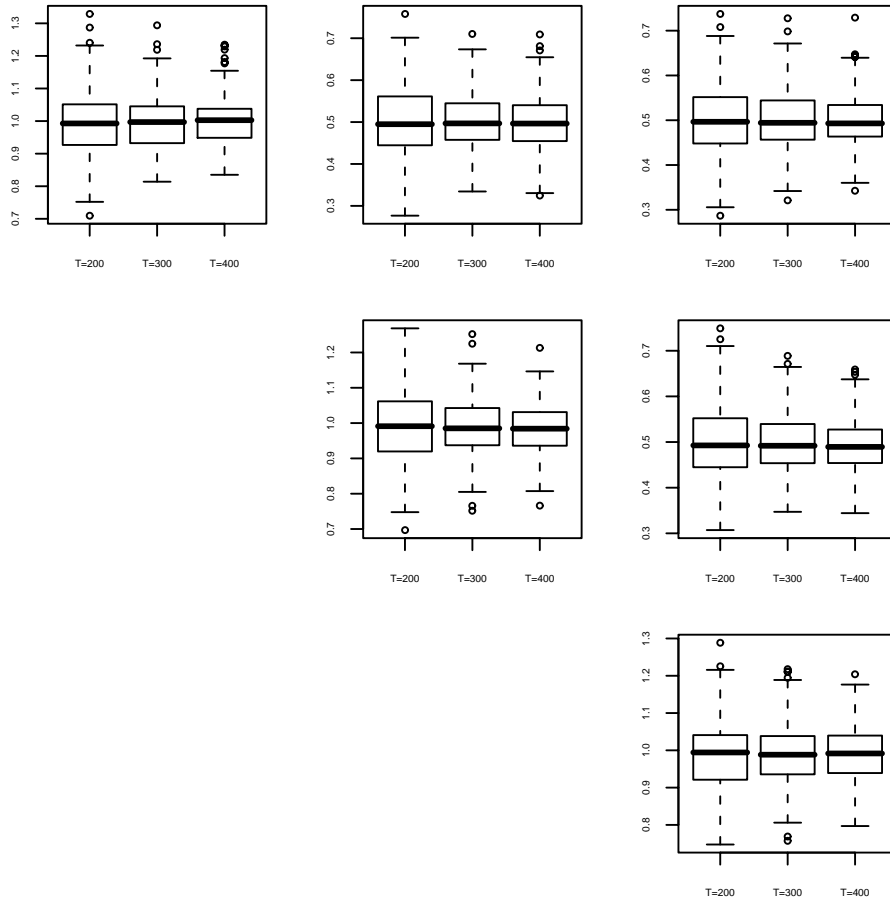
Figure 2. Box-plots of the estimated MA parameter from the 3-dimensional VARFIMA$(0, d, 1)$ model defined in (20), (21), (22), and $\rho = 0.2$ based on the CLDL algorithm with 250 replications. The value $f(g)$ denotes the model specification where f represents the value of MA parameter, and g denotes the sample size, such that g=A=200, g=B=300, and g=C=400.

Figure 3. Box-plots of the estimated Σ from the 3-dimensional VARFIMA$(0, d, 1)$ model defined in (20), (21), (22), and $\rho = 0.2$ based on the CLDL algorithm with 250 replications.

Figure 4. Box-plots of the estimated d from the 3-dimensional VARFIMA$(0, d, 1)$ model defined in (20), (21), (22), and $\rho = 0.5$ based on the CLDL algorithm with 250 replications. The value $f(g)$ denotes the model specification where $f = d$, g denotes the sample size, such that g=A=200, g=B=300, and g=C=400.

Figure 5.   Box-plots of the estimated MA parameter from the 3-dimensional VARFIMA$(0, d, 1)$ model defined in (20), (21), (22), and $\rho = 0.5$ based on the CLDL algorithm with 250 replications. The value $f(g)$ denotes the model specification where f represents the value of MA parameter, and g denotes the sample size, such that g=A=200, g=B=300, and g=C=400.

Figure 6.   Box-plots of the estimated Σ from the 3-dimensional VARFIMA$(0, d, 1)$ model defined in (20), (21), (22), and $\rho = 0.5$ based on the CLDL algorithm with 250 replications.
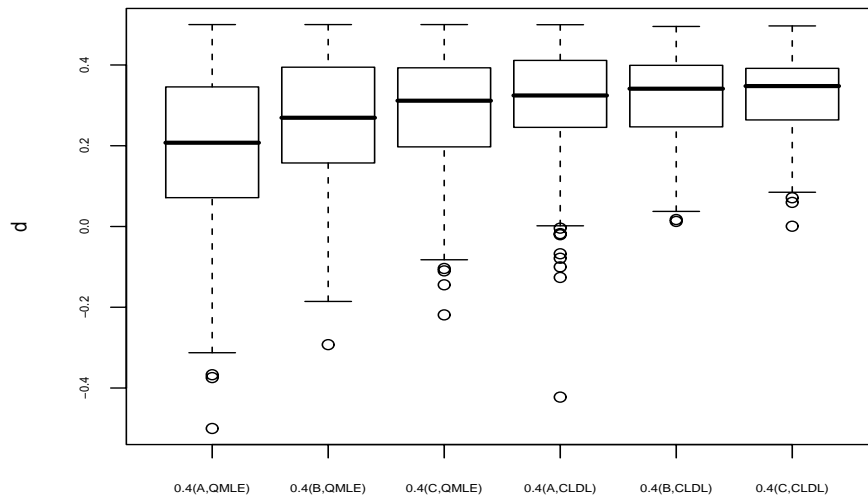
Figure 7. Box-plots of the estimated d from the 3-dimensional VARFIMA$(0, d, 1)$ model defined in (20), (21), and (22) with 250 replications and $\rho = 0.2$. The value $f(g, \text{QMLE})$ denotes the estimation results when the estimator is the QMLE of Lobato [23] and $m = [T^{0.50}]$, f denotes the true value of $d = f$, and g denotes the sample size, such that g=A=200, g=B=300, g=C=400. Likewise, the value $f(g, \text{CLDL})$ denotes the estimation results when the estimator is the proposed CLDL algorithm, f denotes the true value of $d = f$, and g denotes the sample size, such that g=A=200, g=B=300, g=C=400.
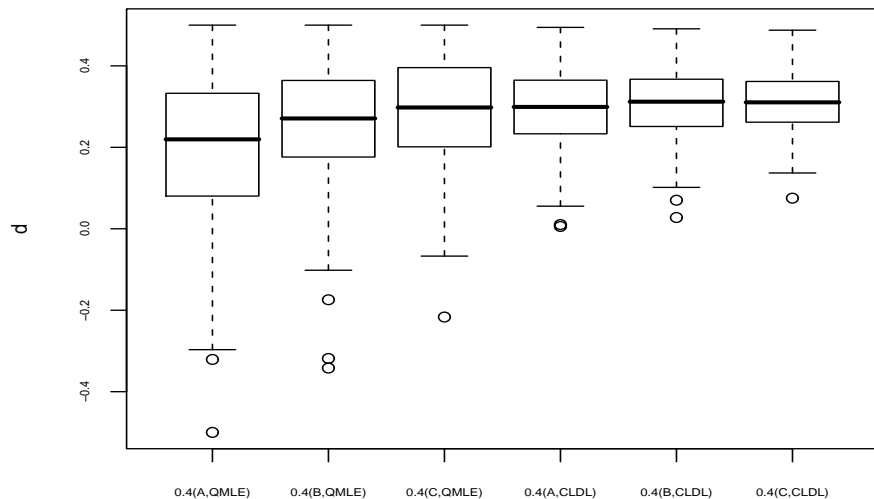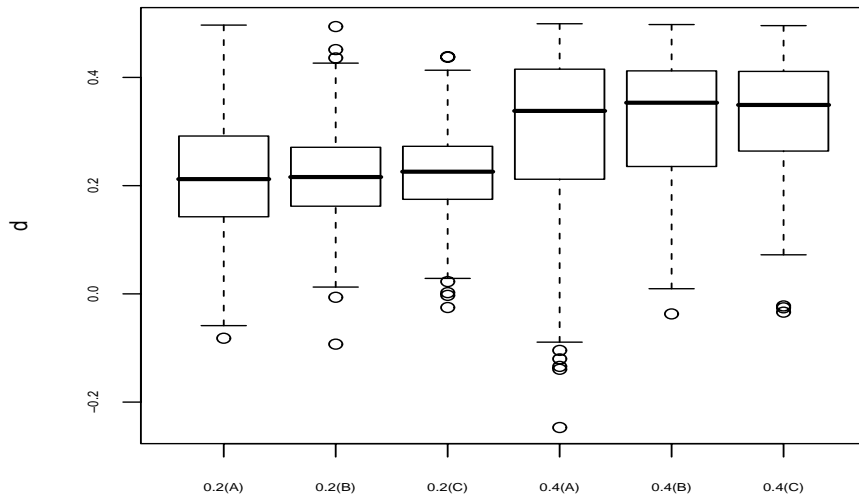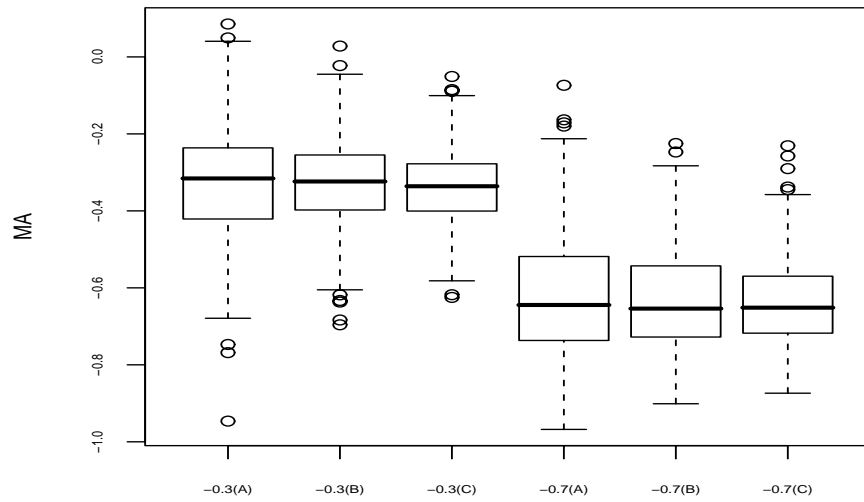
Figure 8.   Box-plots of the estimated d from the 3-dimensional VARFIMA$(0, d, 1)$ model defined in (20), (21), and (22) with 250 replications and $\rho = 0.5$. The value $f(g, \text{QMLE})$ denotes the estimation results when the estimator is the QMLE of Lobato [23] and $m = [T^{0.50}]$, f denotes the true value of $d = f$, and g denotes the sample size, such that g=A=200, g=B=300, g=C=400. Likewise, the value $f(g, \text{CLDL})$ denotes the estimation results when the estimator is the proposed CLDL algorithm, f denotes the true value of $d = f$, and g denotes the sample size, such that g=A=200, g=B=300, g=C=400.
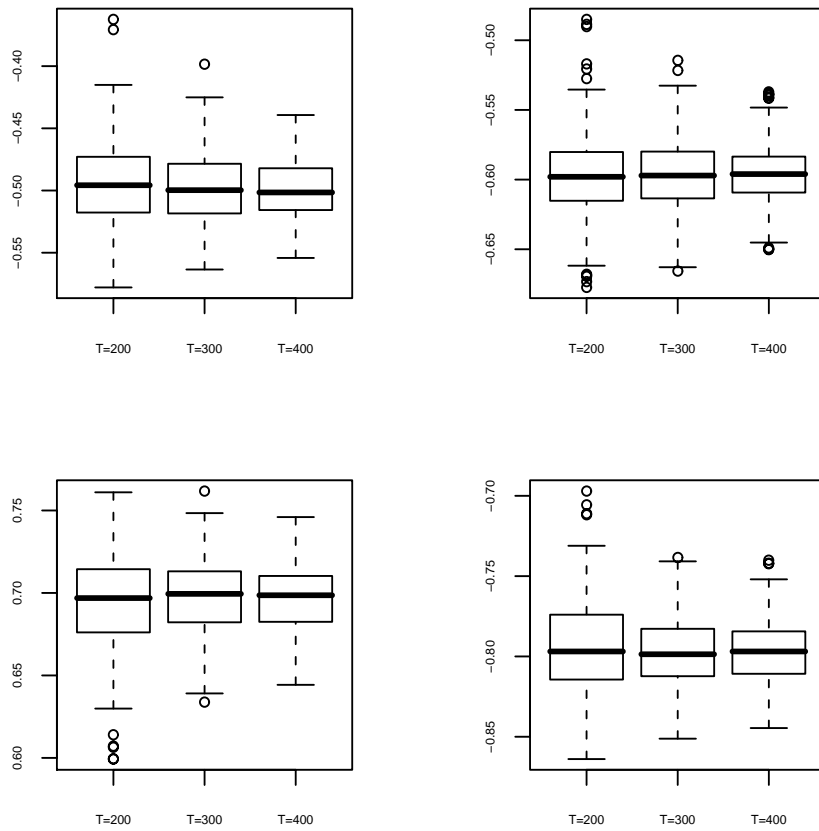
Figure 9.   Box-plots of the estimated d from the 2-dimensional VARFIMA$(1, d, 1)$ model defined in (29), (30), and $\rho = 0.2$ based on the CLDL algorithm with 250 replications. The value $f(g)$ denotes the model specification where $f = d$, and g denotes the sample size, such that g=A=200, g=B=300, and g=C=400.

Figure 10. Box-plots of the estimated MA parameter from the 2-dimensional VARFIMA$(1, d, 1)$ model defined in (29), (30), and $\rho = 0.2$ based on the CLDL algorithm with 250 replications. The value $f(g)$ denotes the model specification where f represents the value of MA parameter, and g denotes the sample size, such that g=A=200, g=B=300, and g=C=400.

Figure 11. Box-plots of the estimated AR parameter from the 2-dimensional VARFIMA$(1, d, 1)$ model defined in (29), (30), and $\rho = 0.2$ based on the CLDL algorithm with 250 replications. The top-left panel depicts the estimates for $\Phi_{11,1}$, the top-right panel depicts the estimates for $\Phi_{12,1}$, the low-left panel depicts the estimates for $\Phi_{21,1}$, and the low-right panel depicts the estimates for $\Phi_{22,1}$.
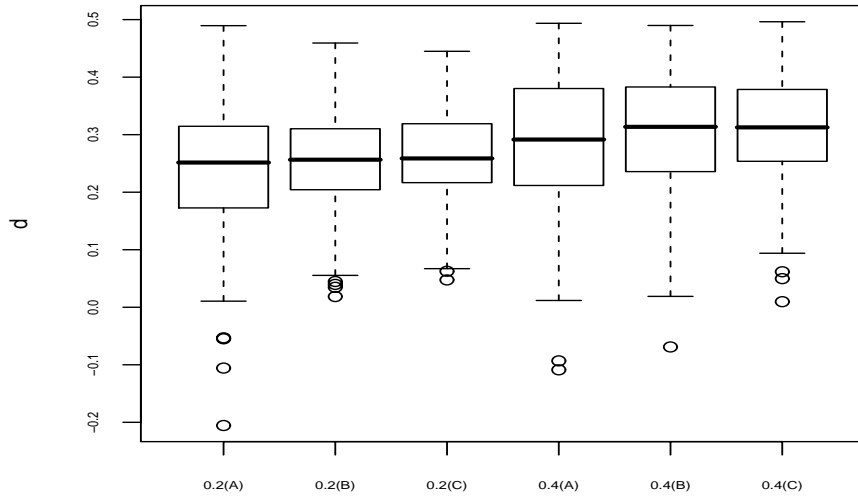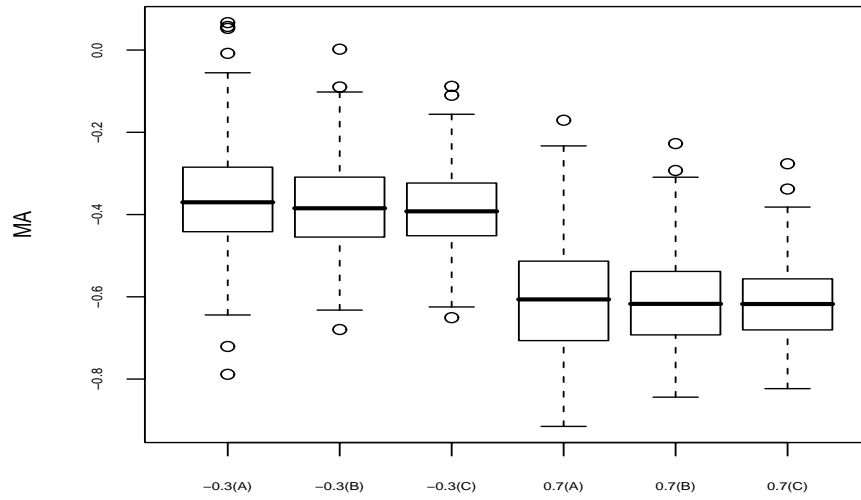
Figure 12. Box-plots of the estimated d from the 2-dimensional VARFIMA$(1, d, 1)$ model defined in (29), (30), and $\rho = 0.5$ based on the CLDL algorithm with 250 replications. The value $f(g)$ denotes the model specification where $f = d$, and g denotes the sample size, such that g=A=200, g=B=300, and g=C=400.
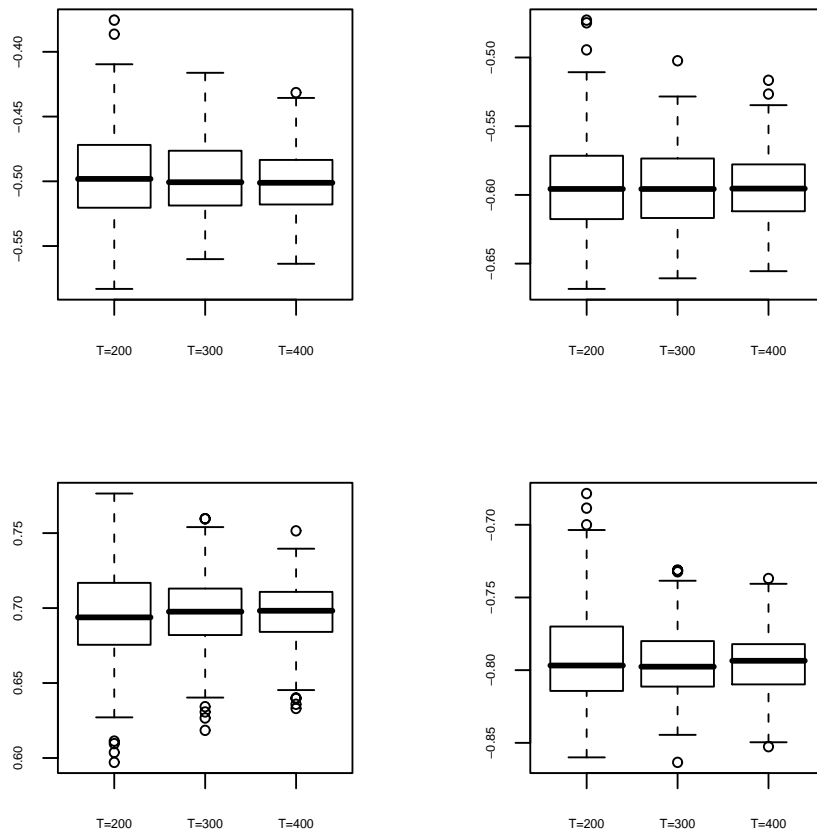
Figure 13.   Box-plots of the estimated MA parameter from the 2-dimensional VARFIMA$(1, d, 1)$ model defined in (29), (30), and $\rho = 0.5$ based on the CLDL algorithm with 250 replications. The value $f(g)$ denotes the model specification where f represents the value of MA parameter, and g denotes the sample size, such that g=A=200, g=B=300, and g=C=400.

Figure 14.   Box-plots of the estimated AR parameter from the 2-dimensional VARFIMA$(1, d, 1)$
model defined in (29), (30), and $\rho = 0.5$ based on the CLDL algorithm with 250 replications. The
top-left panel depicts the estimates for $\Phi_{11,1}$, the top-right panel depicts the estimates for $\Phi_{12,1}$,
the low-left panel depicts the estimates for $\Phi_{21,1}$, and the low-right panel depicts the estimates for
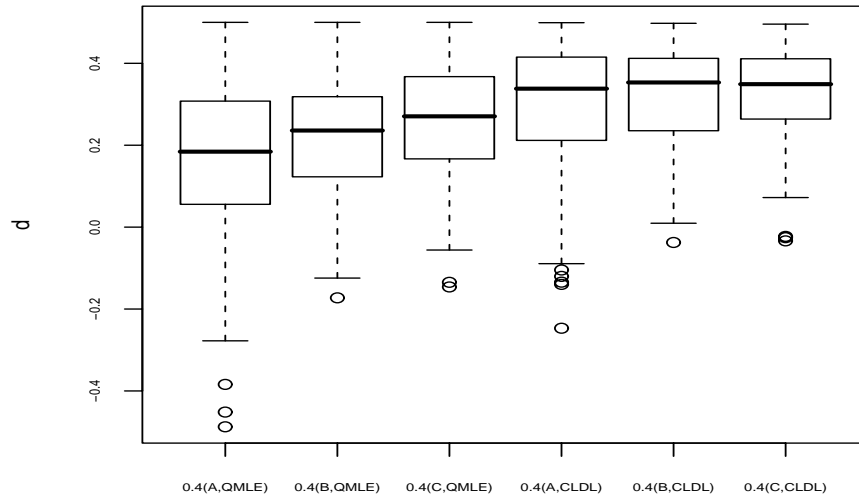$\Phi_{22,1}$.

Figure 15. Box-plots of the estimated d from the 2-dimensional VARFIMA$(1, d, 1)$ model defined in (29) and (30) with 250 replications and $\rho = 0.2$. The value $f(g, \text{QMLE})$ denotes the estimation results when the estimator is the QMLE of Lobato [23] and $m = [T^{0.50}]$, f denotes the true value of $d = f$, and g denotes the sample size, such that g=A=200, g=B=300, g=C=400. Likewise, the value $f(g, \text{CLDL})$ denotes the estimation results when the estimator is the proposed CLDL algorithm, f denotes the true value of $d = f$, and g denotes the sample size, such that g=A=200, g=B=300, g=C=400.
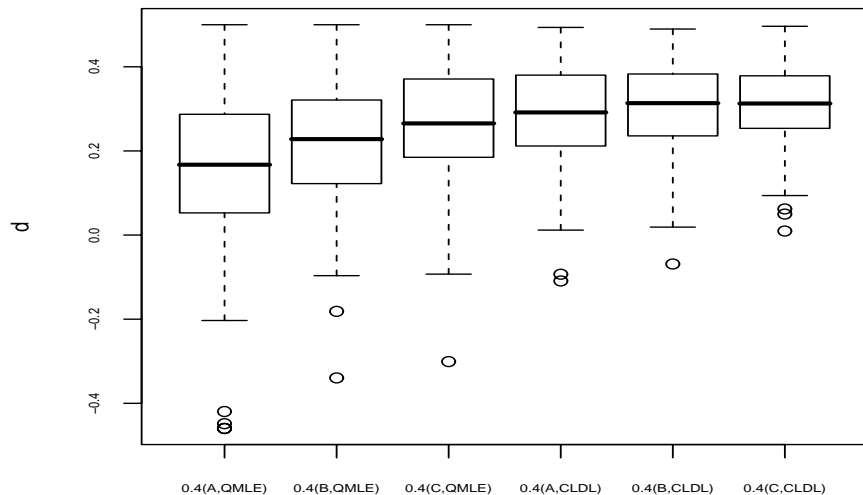
Figure 16.   Box-plots of the estimated d from the 2-dimensional VARFIMA$(1, d, 1)$ model defined in (29) and (30) with 250 replications and $\rho = 0.5$. The value $f(g, \mathrm{QMLE})$ denotes the estimation results when the estimator is the QMLE of Lobato [23] and $m = [T^{0.50}]$, f denotes the true value of $d = f$, and g denotes the sample size, such that g=A=200, g=B=300, g=C=400. Likewise, the value $f(g, \mathrm{CLDL})$ denotes the estimation results when the estimator is the proposed CLDL algorithm, f denotes the true value of $d = f$, and g denotes the sample size, such that g=A=200, g=B=300, g=C=400.