

Bias reduction of a conditional maximum likelihood estimator for a Gaussian second-order moving average model

Fumiaki Honda, Takeshi Kurosawa*

Tokyo University of Science, 1-3, Kagurazaka, Shinjuku-ku, Tokyo, 162-8601, Japan

fh0890.e13@gmail.com (F. Honda), tkuro@rs.tus.ac.jp (T. Kurosawa)

Received: 8 December 2020, Revised: 8 July 2021, Accepted: 8 July 2021,
Published online: 4 August 2021

Abstract In this study, we consider a bias reduction of the conditional maximum likelihood estimators for the unknown parameters of a Gaussian second-order moving average (MA(2)) model. In many cases, we use the maximum likelihood estimator because the estimator is consistent. However, when the sample size n is small, the error is large because it has a bias of $O(n^{-1})$. Furthermore, the exact form of the maximum likelihood estimator for moving average models is slightly complicated even for Gaussian models. We sometimes rely on simpler maximum likelihood estimation methods. As one of the methods, we focus on the conditional maximum likelihood estimator and examine the bias of the conditional maximum likelihood estimator for a Gaussian MA(2) model. Moreover, we propose new estimators for the unknown parameters of the Gaussian MA(2) model based on the bias of the conditional maximum likelihood estimators. By performing simulations, we investigate properties of this bias, as well as the asymptotic variance of the conditional maximum likelihood estimators for the unknown parameters. Finally, we confirm the validity of the new estimators through this simulation study.

Keywords Gaussian second-order moving average model, conditional maximum likelihood estimators, bias reduction, asymptotic expansion

2010 MSC 62M15

*Corresponding author.

1 Introduction

Estimators of unknown parameters must be consistent. The consistency is ensured when we have large samples. The estimators might have a bias if the sample size is small. In recent years, some computational methods has been developed to compute estimates for unknown parameters. However, an analytical solution for the bias enables us to know a relationship between unknown parameter and the bias. It means that the bias changes depending on unknown parameters. Many analytical evaluations for the bias for a class of nonlinear estimators in models with i.i.d. samples have been conducted for many years. Tanaka (1983) [14] provided asymptotic expansions of the least square estimator for the first-order autoregressive process AR(1) and computed its bias. Tanaka (1984) [15] also gave asymptotic expansions of the maximum likelihood estimators for autoregressive moving average (ARMA) models, including AR(1), AR(2), MA(1), and MA(2), and also computed their bias. Cordeiro and Klein (1994) [8] derived the bias of the maximum likelihood estimators for ARMA models in another way although the result for MA(2) was not shown. Cheang and Reinsel (2000) [7] developed a way to reduce the bias of AR models using the restricted maximum likelihood estimation.

Practically, we often rely on the *conditional* maximum likelihood estimation for reducing the computational cost of the maximum likelihood estimation and for a prediction of an unobserved variable which is the next value of the observed data (see Section 2 for the definition of the conditional maximum likelihood estimation). The conditional maximum likelihood estimation is often referred as the quasi-maximum likelihood estimation (QMLE). Statistical properties of the conditional maximum likelihood estimation have been discussed in some literatures (see [3] and [4] by Bao and Ullah, for example). Giummolè and Vidoni (2010) [9] showed the bias of the conditional maximum likelihood estimator for a Gaussian MA(1) model in a process of obtaining improved coverage probabilities for ARMA models. However, the bias of the estimator for a Gaussian first-order moving average (MA(1)) model was slightly strange. Hence, Kurosawa, Noguchi, and Honda (2017) [12] corrected the bias and deduced a simple expression for the bias using a method by Barndorff-Nielsen and Cox (1994) [5]. We also should note the recent remarkable results by Y. Bao (2016) [1] and (2018) [2]. We shall discuss his results in Remark 3.4 below.

In this study, we show the bias of the conditional maximum likelihood estimators of unknown parameters for a Gaussian second-order moving average (MA(2)) model followed by the method in [12]. In Section 2, we introduce a Gaussian MA(2) model and the conditional maximum likelihood function. In Section 3, we derive both the bias and the mean squared errors (MSEs) of the conditional maximum likelihood estimator for a Gaussian MA(2) model, and then propose new estimators based on the $O(n^{-1})$ term in the bias of the conditional maximum likelihood estimators. Moreover, we show that the proposed estimators are less biased and have the lower MSEs than those of the conditional maximum likelihood estimators. In Section 4, we conduct a simple simulation study to verify our results. Furthermore, we apply our method to GNP in United States of America as an illustrative example of our method in Section 6.

2 A Gaussian MA(2) model and the conditional maximum likelihood estimator

Let $\{Y_t\}$ be a Gaussian MA(1) model (see, e.g., [6, 10]) defined by

$$Y_t = \mu + \varepsilon_t + \rho\varepsilon_{t-1}, \quad \varepsilon_t \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma^2) \quad (t \geq 1), \tag{1}$$

where $|\rho| < 1$. Kurosawa, Noguchi, and Honda (2017) [12] computed the bias of the conditional maximum likelihood estimator under the condition that

$$\varepsilon_0 = 0 \tag{2}$$

for the Gaussian MA(1) model. Assumption (2) is a useful condition for not only an estimation problem but also a prediction problem, since ε_T ($T \geq 1$) can be written by a linear combination of Y_1, \dots, Y_T . Then, the best linear unbiased estimator \widehat{Y}_{T+h} of Y_{T+h} ($h > 0$) given $S = \{Y_1 = y_1, \dots, Y_T = y_T\}$ is described using a finite linear combination of $\varepsilon_1, \dots, \varepsilon_T$ (see, e.g., [12]). They gave the following:

Theorem 2.1 ([12]). *The bias of the conditional maximum likelihood estimators of the unknown parameters given (2) for the Gaussian MA(1) model is*

$$\begin{aligned} E[\widehat{\mu} - \mu] &= o(n^{-1}), \\ E[\widehat{\sigma} - \sigma] &= -\frac{5\sigma}{4n} + o(n^{-1}), \\ E[\widehat{\rho} - \rho] &= \frac{2\rho - 1}{n} + o(n^{-1}). \end{aligned}$$

In this study, we consider a Gaussian MA(2) model defined by

$$Y_t = \mu + \varepsilon_t + \rho_1\varepsilon_{t-1} + \rho_2\varepsilon_{t-2} \quad (t \geq 1), \tag{3}$$

where $\varepsilon_t \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma^2)$ and $\boldsymbol{\theta} = (\mu, \sigma, \rho_1, \rho_2)^\top$ is a vector consisting of the unknown parameters. Although the MA(2) model has the property of *stationarity* regardless of the values of ρ_1 and ρ_2 , we assume the *invertibility*, which means

$$\rho_1 - \rho_2 < 1, \quad \rho_1 + \rho_2 > -1, \quad -1 < \rho_2 < 1,$$

to identify the model uniquely. Otherwise, the maximum likelihood function takes the same value at different points.

If we consider an estimation problem for the maximum likelihood function with Y_1, \dots, Y_n , then the likelihood function can be expressed using the infinite number of ε_s . To avoid the problem of infinite number of ε_s , we solve the conditional maximum likelihood function given

$$\varepsilon_0 = \varepsilon_{-1} = 0 \tag{4}$$

for the Gaussian MA(2) model. In this case, Y_1, \dots, Y_n can be transformed using the finite number of ε_s . The conditional log-likelihood function using the finite number of ε_s for the Gaussian MA(2) model given (4) is expressed as

$$\mathcal{L}(\boldsymbol{\theta}; \mathbf{y}) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \sum_{t=1}^n \frac{\{\varepsilon_t(\boldsymbol{\theta}; \mathbf{y})\}^2}{2\sigma^2}. \tag{5}$$

The likelihood function with (4) is referred as the *conditional* likelihood function (see [11, p. 653] and [17]). We use the following lemma to compute the bias of the unknown parameters in the Gaussian MA (2) model.

Lemma 2.2. *Let $\mathbf{Y} = (Y_1, \dots, Y_n)^\top$ be a vector of random variables generated by a Gaussian MA(2) model in (3). Assume that (4) holds. Then, we have*

$$\varepsilon_t(\boldsymbol{\theta}; \mathbf{Y}) = \varepsilon_t = \sum_{k=0}^{t-1} \left(\sum_{l=0}^{t-k-1} \lambda_1^l \lambda_2^{t-k-l-1} \right) (Y_{k+1} - \mu) \quad (t \geq 1), \tag{6}$$

where

$$\lambda_1 = \frac{-\rho_1 + \Delta}{2}, \quad \lambda_2 = \frac{-\rho_1 - \Delta}{2}, \quad \Delta = \sqrt{\rho_1^2 - 4\rho_2}. \tag{7}$$

The proof will be in the Appendix. We know that

$$\begin{aligned} \varepsilon_t &= \frac{1}{(1 - \lambda_1 L)(1 - \lambda_2 L)} (Y_t - \mu) = \sum_{k=0}^{\infty} \sum_{l=0}^k \lambda_1^l \lambda_2^{k-l} (Y_{t-k} - \mu) \\ &= \sum_{k=-\infty}^{t-1} \left(\sum_{l=0}^{t-k-1} \lambda_1^l \lambda_2^{t-k-l-1} \right) (Y_{k+1} - \mu) \end{aligned}$$

since the process is invertible. The lemma suggests that the coefficients of Y_t ($t \leq 0$) are zero when $\varepsilon_0 = \varepsilon_{-1} = 0$.

Since the conditional likelihood function is expressed as a function of independent samples $\varepsilon_1, \dots, \varepsilon_n$, we can apply the following lemma in [5] to the conditional log-likelihood function. We apply Lemma 2.3 for i.i.d. random variables $\boldsymbol{\varepsilon}$, not \mathbf{Y} . The high-order differentiations of the conditional log-likelihood function (5) are required to obtain the bias and the MSEs of the maximum likelihood estimators and Lemma 2.2 will be used for the calculations. An asymptotic expansion of the bias of the maximum likelihood estimator is given by the following:

Lemma 2.3 (See Barndorff-Nielsen and Cox (1994) [5, p. 150]). *Let $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)^\top$ be a vector of unknown parameters for a random variable Z , and $\widehat{\boldsymbol{\theta}} = (\widehat{\theta}_1, \dots, \widehat{\theta}_d)^\top$ be a vector of the maximum likelihood estimators of $\boldsymbol{\theta}$ for a vector of random samples $\mathbf{Z} = (Z_1, \dots, Z_n)^\top$. Then, the bias of $\widehat{\theta}_r$ ($1 \leq r \leq d$) is given by*

$$\begin{aligned} E_Z[\widehat{\theta}_r - \theta_r] &= \frac{1}{2} \sum_{s=1}^d \sum_{t=1}^d \sum_{u=1}^d i^{\theta_r \theta_s} i^{\theta_t \theta_u} (v_{\theta_s \theta_t \theta_u} + 2v_{\theta_s \theta_t, \theta_u}) + O(n^{-3/2}) \\ &\quad (r = 1, \dots, d), \end{aligned}$$

where $\mathcal{L}(\boldsymbol{\theta}; \mathbf{Z})$ is the log-likelihood function for $\mathbf{Z} = (Z_1, \dots, Z_n)^\top$,

$$\begin{aligned} l_{\theta_s} &= \frac{\partial \mathcal{L}(\boldsymbol{\theta}; \mathbf{Z})}{\partial \theta_s}, \quad l_{\theta_s \theta_t} = \frac{\partial^2 \mathcal{L}(\boldsymbol{\theta}; \mathbf{Z})}{\partial \theta_s \partial \theta_t}, \quad l_{\theta_s \theta_t \theta_u} = \frac{\partial^3 \mathcal{L}(\boldsymbol{\theta}; \mathbf{Z})}{\partial \theta_s \partial \theta_t \partial \theta_u}, \\ i^{\theta_r \theta_s} &= \left(I_n(\boldsymbol{\theta})^{-1} \right)_{r,s}, \quad v_{\theta_s \theta_t \theta_u} = E_Z[l_{\theta_s \theta_t \theta_u}], \quad v_{\theta_s \theta_t, \theta_u} = E_Z[l_{\theta_s \theta_t} l_{\theta_u}], \end{aligned}$$

and $I_n(\boldsymbol{\theta})$ is the Fisher information matrix for \mathbf{Z} .

3 Main results

In this section, we compute the biases of estimators for the unknown parameters of the Gaussian MA(2) model using the conditional maximum likelihood estimate, and also propose new estimators for these parameters. Before obtaining the results on bias, we observe the MSEs. The MSEs appear in the diagonal elements in the covariance matrix by

$$E[(\hat{\theta} - \theta)(\hat{\theta} - \theta)^\top] = I_n(\theta)^{-1} + o(n^{-1}),$$

where $I_n(\theta)$ is the Fisher's information matrix. It can be simplified by applying asymptotic properties

$$\lim_{n \rightarrow \infty} nI_n(\theta)^{-1} = J(\theta).$$

Therefore,

$$E[(\hat{\theta} - \theta)(\hat{\theta} - \theta)^\top] = \frac{J(\theta)}{n} + o(n^{-1}). \tag{8}$$

Theorem 3.1. *The elements in the asymptotic covariance matrix of the conditional maximum likelihood estimators of the unknown parameters under (4) for the Gaussian MA(2) model in (3) is given by $E[(\hat{\mu} - \mu)(\hat{\sigma} - \sigma)] = E[(\hat{\mu} - \mu)(\hat{\rho}_1 - \rho_1)] = E[(\hat{\mu} - \mu)(\hat{\rho}_2 - \rho_2)] = E[(\hat{\sigma} - \sigma)(\hat{\rho}_1 - \rho_1)] = E[(\hat{\sigma} - \sigma)(\hat{\rho}_2 - \rho_2)] = o(n^{-1})$ and*

$$\begin{aligned} E[(\hat{\mu} - \mu)^2] &= \frac{\sigma^2(\rho_1 + \rho_2 + 1)^2}{n} + o(n^{-1}), & E[(\hat{\sigma} - \sigma)^2] &= \frac{\sigma^2}{2n} + o(n^{-1}), \\ E[(\hat{\rho}_1 - \rho_1)^2] &= \frac{1 - \rho_2^2}{n} + o(n^{-1}), & E[(\hat{\rho}_2 - \rho_2)^2] &= \frac{1 - \rho_2^2}{n} + o(n^{-1}), \\ E[(\hat{\rho}_1 - \rho_1)(\hat{\rho}_2 - \rho_2)] &= \frac{\rho_1(1 - \rho_2)}{n} + o(n^{-1}). \end{aligned}$$

The proof is given in Section 5.1. By applying Lemma 2.3 to (5), we obtain the bias of the conditional maximum likelihood estimators.

Theorem 3.2. *The bias of the conditional maximum likelihood estimators of the unknown parameters under (4) for the Gaussian MA(2) model in (3) is given by*

$$E[\hat{\mu} - \mu] = o(n^{-1}), \tag{9}$$

$$E[\hat{\sigma} - \sigma] = -\frac{7\sigma}{4n} + o(n^{-1}), \tag{10}$$

$$E[\hat{\rho}_1 - \rho_1] = \frac{\rho_1 + \rho_2 - 1}{n} + o(n^{-1}), \tag{11}$$

$$E[\hat{\rho}_2 - \rho_2] = \frac{3\rho_2 - 1}{n} + o(n^{-1}). \tag{12}$$

The proof is given in Section 5.2. We observe that the bias of the conditional maximum likelihood estimators for the Gaussian MA(2) model is the same as that for the full maximum likelihood estimators for a Gaussian MA(2) model (see Tanaka (1984) [15]). Although (11) looks different from the result by Tanaka, we can deduce the same result for the bias of $\hat{\sigma}^2$ (see (14)).

Remark 3.3. We note that MA(2) is reduced to MA(1) if we put $\rho_2 = 0$ in (3). However, the bias for $\rho (= \rho_1)$ in Theorem 2.1 is not obtained even if we put $\rho_2 = 0$ in (11). The results in Theorem 3.2 with $\rho_2 = 0$ are obtained from a solution of the maximum likelihood estimate in four dimensions. The bias of MA(1) in Theorem 2.1 can be considered as the bias of MA(2) given $\rho_2 = 0$ in Theorem 3.2. This implies that Theorem 3.2 is not result under the assumption with algebraic relationships among unknown parameters such as $\rho_2 = 0, \rho_2 = 1, \Delta = \rho_1^2 - 4\rho_2 = 0$, and so on. Namely, we do not assume any algebraic relationships among the unknown parameters in advance. Thus, $\Delta \neq 0$ which is equivalent to $\lambda_1 \neq \lambda_2$ is used in the proof of Theorem 3.2 and Propositions and Lemmas in the appendix except for Lemma 2.2.

Remark 3.4. We have recently found the notable results by Y. Bao (2016) [1] although we originally proved Theorem 3.2. The abstract in [1] claims that the bias of the conditional Gaussian likelihood estimation with nonnormal errors is derived. The gap between the ‘‘Gaussian’’ and the ‘‘nonnormal errors’’ imply that he used likelihood function (5) even if the errors follow a nonnormal distribution. For the calculation of the bias, we require the values of the skewness γ_1 and the kurtosis γ_2 . Namely, he derived the bias regarding the likelihood function as the Gaussian likelihood function without the conditions $\gamma_1 = 0$ and $\gamma_2 = 3$ for a normal distribution. He gave the bias of various models including MA(2) with a matrix representation which was originally studied by Corderio and Klein (1994) [8], while we use the roots of the characteristic function for the derivation of the bias. We purely focus on the Gaussian MA(2) model with the conditional likelihood function and evaluate the corrected bias. Furthermore, we propose a new estimator below based on the corrected bias and discuss the corrected estimators under a pure Gaussian MA(2) model in detail using the (estimated) bias and the MSE in the simulation study.

Using (10), (11), and (12), we propose the following new estimators for the Gaussian MA(2) model:

$$\tilde{\sigma} = \hat{\sigma} + \frac{7\hat{\sigma}}{4n}, \quad \tilde{\rho}_1 = \hat{\rho}_1 - \frac{\hat{\rho}_1 + \hat{\rho}_2 - 1}{n}, \quad \tilde{\rho}_2 = \hat{\rho}_2 - \frac{3\hat{\rho}_2 - 1}{n}. \tag{13}$$

As we can see, the proposed estimators are asymptotically equal to the usual estimators. We consider the MSEs of the new estimators:

$$E[\tilde{\sigma}] = \left(1 - \frac{7}{4n}\right)\sigma + o(n^{-1}), \quad E[\tilde{\sigma}^2] = \left(1 - \frac{3}{n}\right)\sigma^2 + o(n^{-1}). \tag{14}$$

Thus, we have

$$E[(\tilde{\sigma} - \sigma)^2] = \frac{\sigma^2}{2n} + o(n^{-1}).$$

In other words, the MSEs of $\tilde{\sigma}$ and $\hat{\sigma}$ are asymptotically the same as well:

$$\begin{aligned} E[\hat{\rho}_1^2] &= \rho_1^2 + \frac{1 - \rho_2^2}{n} + 2\rho_1 \frac{\rho_1 + \rho_2 - 1}{n} + o(n^{-1}), \\ E[\tilde{\rho}_2^2] &= \rho_2^2 + \frac{1 - \rho_2^2}{n} + 2\rho_2 \frac{3\rho_2 - 1}{n} + o(n^{-1}), \end{aligned}$$

$$E[\widehat{\rho}_1 \widehat{\rho}_2] = \rho_1 \rho_2 + \rho_2 \frac{3\rho_1 + \rho_2 - 1}{n} + o(n^{-1}).$$

Therefore, we have

$$E[(\widetilde{\rho}_1 - \rho_1)^2] = \frac{1 - \rho_2^2}{n} + o(n^{-1}), \quad E[(\widetilde{\rho}_2 - \rho_2)^2] = \frac{1 - \rho_2^2}{n} + o(n^{-1}).$$

In other words, the MSEs of $\widetilde{\rho}_1$ and $\widehat{\rho}_1$ are asymptotically the same, as are the MSEs of $\widetilde{\rho}_2$ and $\widehat{\rho}_2$.

4 Simulation study

In this section, we conduct a simulation study in order to verify Theorems 3.1 and 3.2 and evaluate the validity of the new estimators. Let $\mu = 1$ and $\sigma = 1$. For a fixed n and a fixed ρ , we generate $\mathbf{y} = (y_1, \dots, y_n)^\top$ 30,000 times from the Gaussian MA(2) model. For each \mathbf{y} , we calculate a vector of the conditional maximum likelihood estimators $\widehat{\boldsymbol{\theta}} = (\widehat{\mu}, \widehat{\sigma}, \widehat{\rho}_1, \widehat{\rho}_2)^\top$. Using the 30,000 replications, we calculate the estimated bias and MSEs using Monte Carlo simulations. In Subsections 4.2 and 4.3, we also compute the full maximum likelihood estimators $\widehat{\boldsymbol{\theta}}^{\text{MLE}}$ to compare it with our estimators.

4.1 Evaluation of asymptotic variances

We evaluate how much the MSEs of the conditional maximum likelihood estimators change depending on the true values of the unknown parameters. Table 1 shows the estimated MSEs and the values of $J(\boldsymbol{\theta})/n$ obtained in Theorem 3.1 of the conditional maximum likelihood estimators for each unknown parameter.

Table 1. Comparisons of the estimated MSEs and $J(\boldsymbol{\theta})/n$ for each unknown parameter (upper: the estimated MSE, lower: $J(\boldsymbol{\theta})/n$)

$(\rho_1, \rho_2) = (0.25, -0.25)$					$(\rho_1, \rho_2) = (-0.40, -0.59)$				
n	$\widehat{\mu}$	$\widehat{\sigma}$	$\widehat{\rho}_1$	$\widehat{\rho}_2$	n	$\widehat{\mu}$	$\widehat{\sigma}$	$\widehat{\rho}_1$	$\widehat{\rho}_2$
50	0.02046	0.01133	0.03102	0.03684	50	0.01412	0.01146	0.04705	0.02783
	0.02000	0.01000	0.01875	0.01875		0.01312	0.01000	0.01304	0.01304
100	0.01020	0.00526	0.01136	0.01233	100	0.00691	0.00578	0.01566	0.00944
	0.01000	0.00500	0.00938	0.00938		0.00656	0.00500	0.00652	0.00652
150	0.00675	0.00346	0.00706	0.00739	150	0.00453	0.00386	0.00921	0.00564
	0.00667	0.00333	0.00625	0.00625		0.00437	0.00333	0.00435	0.00435

$(\rho_1, \rho_2) = (0.15, -0.55)$					$(\rho_1, \rho_2) = (0.15, 0.55)$				
n	$\widehat{\mu}$	$\widehat{\sigma}$	$\widehat{\rho}_1$	$\widehat{\rho}_2$	n	$\widehat{\mu}$	$\widehat{\sigma}$	$\widehat{\rho}_1$	$\widehat{\rho}_2$
50	0.00786	0.01148	0.03108	0.03334	50	0.05849	0.01099	0.01943	0.02612
	0.00720	0.01000	0.01395	0.01395		0.05780	0.01000	0.01395	0.01395
100	0.00379	0.00526	0.01051	0.01123	100	0.02907	0.00520	0.00814	0.00937
	0.00360	0.00500	0.00698	0.00698		0.02890	0.00500	0.00698	0.00698
150	0.00248	0.00346	0.00597	0.00620	150	0.01932	0.00343	0.00511	0.00559
	0.00240	0.00333	0.00465	0.00465		0.01927	0.00333	0.00465	0.00465

We conducted simulations under the four settings. The top-left table is prepared for checking of performance under the condition that the true parameters are within

the invertibility condition. On the other hand, the top-right table is close to the boundary of the invertibility condition. The two bottom tables are made for checking the symmetry of ρ_2 . It is clearly observed from Table 1 for all the settings that the estimated MSEs decrease when the sample size n is large. The estimated MSE of $\hat{\sigma}$ does not depend on the values of ρ_1 and ρ_2 , which coincides with the result that $J(\theta)/n$ of $\hat{\sigma}$ in Theorem 3.1 does not include ρ_1 and ρ_2 in the expression. Since $J(\theta)/ns$ of $\hat{\rho}_1$ and $\hat{\rho}_2$ depend on the value of ρ_2 but are independent of the value of ρ_1 , we expect that the estimated MSEs of $\hat{\rho}_1$ and $\hat{\rho}_2$ on $(\rho_1, \rho_2) = (0.15, -0.55)$ and $(\rho_1, \rho_2) = (0.15, 0.55)$ are close, but the results show different values in the small sample size $n = 50$. This result may be the influence of $o(n^{-1})$. We compare n times the estimated MSEs and $J(\theta)$ on $(\rho_1, \rho_2) = (0.15, -0.55)$ and $(\rho_1, \rho_2) = (0.15, 0.55)$ of $n = 50$ and $n = 1000$ to verify the influence by $o(n^{-1})$ in Table 2.

Table 2. Comparisons of n times the estimated MSEs and $J(\theta)$ of $n = 50$ and $n = 1000$

		$(\rho_1, \rho_2) = (0.15, -0.55)$			
		$\hat{\mu}$	$\hat{\sigma}$	$\hat{\rho}_1$	$\hat{\rho}_2$
$n \times$ the estimated MSE	$n = 50$	0.39307	0.57401	1.55412	1.66715
	$n = 1000$	0.36061	0.51210	0.72833	0.71769
$J(\theta)$		0.36000	0.50000	0.69750	0.69750
		$(\rho_1, \rho_2) = (0.15, 0.55)$			
		$\hat{\mu}$	$\hat{\sigma}$	$\hat{\rho}_1$	$\hat{\rho}_2$
$n \times$ the estimated MSE	$n = 50$	2.92462	0.54941	0.97164	1.30611
	$n = 1000$	2.88205	0.51144	0.69979	0.71517
$J(\theta)$		2.89000	0.50000	0.69750	0.69750

Generally, n times the estimated MSEs converge to $J(\theta)$ if n is large. The result shows that the estimated MSE of $\hat{\rho}_1$ is close to $\hat{\rho}_2$ for $n = 1000$. Next, we present the behavior of the estimated MSEs for $\rho_1 = 0.25$ and $\rho_2 = -0.25$ when the sample size is small in Table 3.

Table 3. Behavior of the estimated MSEs when the sample size is small

n	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\rho}_1$	$\hat{\rho}_2$
10	0.14485	0.09400	0.41133	0.36819
11	0.12914	0.08262	0.34886	0.33276
12	0.11453	0.07419	0.29700	0.29798
13	0.10000	0.06659	0.26646	0.27767
14	0.08955	0.06034	0.23022	0.25188
15	0.08262	0.05502	0.21125	0.23577
16	0.07426	0.05051	0.18974	0.21314
17	0.06889	0.04616	0.17351	0.19723
18	0.06413	0.04295	0.15869	0.18569
19	0.05969	0.04028	0.14624	0.17146
20	0.05529	0.03762	0.13552	0.16067

The estimated MSE becomes smaller as the sample size becomes larger.

4.2 Estimator of σ

We express the bias of $\widehat{\sigma}$ as

$$E[\widehat{\sigma} - \sigma] = e_1 + e_2, \quad e_1 = -\frac{7\sigma}{4n}, \quad e_2 = o(n^{-1}),$$

which implies that

$e_1 + e_2$ is the bias of the conditional maximum likelihood estimator,

e_2 is the bias of the conditional maximum likelihood estimator without the term $O(n^{-1})$.

Table 4. Evaluation of the estimated bias of $\widehat{\sigma}$

$(\rho_1, \rho_2) = (0.25, -0.25)$			$(\rho_1, \rho_2) = (0.15, 0.55)$		
n	$e_1 + e_2$	e_2	n	$e_1 + e_2$	e_2
50	-0.03579	-0.00079	50	-0.02814	0.00686
100	-0.01635	0.00115	100	-0.01322	0.00428
150	-0.01035	0.00132	150	-0.00825	0.00342

The bias of $\widehat{\sigma}$ does not depend on the value of ρ_1 and ρ_2 , which coincide with (10). Moreover, $|e_2|$ is smaller than $|e_1 + e_2|$ because of the exclusion of the term $O(n^{-1})$. Next, we compare the bias and MSEs of $\widehat{\sigma}$ and the proposed estimator $\widetilde{\sigma}$ for $\rho_1 = 0.25$ and $\rho_2 = -0.25$. We also compute the full maximum likelihood estimator $\widehat{\sigma}^{\text{MLE}}$.

Table 5. Comparison of the bias and MSEs of $\widehat{\sigma}$ and $\widetilde{\sigma}$

Bias				MSE			
n	$\widehat{\sigma}$	$\widehat{\sigma}^{\text{MLE}}$	$\widetilde{\sigma}$	n	$\widehat{\sigma}$	$\widehat{\sigma}^{\text{MLE}}$	$\widetilde{\sigma}$
50	-0.03579	-0.04193	-0.00204	50	0.01133	0.01195	0.01077
100	-0.01635	-0.01820	0.00087	100	0.00526	0.00531	0.00517
150	-0.01035	-0.01142	0.00120	150	0.00346	0.00347	0.00343

The estimated bias of $\widetilde{\sigma}$ is lower than those of $\widehat{\sigma}$ and $\widehat{\sigma}^{\text{MLE}}$. On the other hand, there is no difference among the three estimated MSEs. This result certainly is in accordance with the discussion in Section 3.

4.3 Estimators of ρ_1 and ρ_2

We express the bias of $\widehat{\theta}_i$ ($i = 3, 4$) as

$$E[\widehat{\theta}_i - \theta_i] = e_1 + e_2, \quad e_1 = O(n^{-1}), \quad e_2 = o(n^{-1}),$$

where $e_1 = (\rho_1 + \rho_2 - 1)/n$ for $i = 3$, $e_1 = (3\rho_2 - 1)/n$ for $i = 4$. Then, $e_1 + e_2$ and e_2 imply the bias of the conditional maximum likelihood estimator and the bias of the conditional maximum likelihood estimator without the term $O(n^{-1})$, respectively.

Except for the case where ρ_1 and ρ_2 are close to the boundary condition $(\rho_1, \rho_2) = (0.40, -0.59)$, $|e_2|$ is smaller than $|e_1 + e_2|$. The bias of $\widehat{\rho}_1$ depends on the value of ρ_1 and ρ_2 . The bias of $\widehat{\rho}_2$ depends on the value of ρ_2 but not ρ_1 , which coincides with (12). Next, we compare the bias between $\widehat{\rho}_1$ and $\widetilde{\rho}_1$ and between $\widehat{\rho}_2$ and $\widetilde{\rho}_2$. We also compute the full maximum likelihood estimators $\widehat{\rho}_i^{\text{MLE}}$.

Table 6. Evaluation of the estimated bias of $\widehat{\rho}_1$ and $\widehat{\rho}_2$

$(\rho_1, \rho_2) = (0.25, -0.25)$						$(\rho_1, \rho_2) = (0.40, -0.59)$					
$\widehat{\rho}_1$			$\widehat{\rho}_2$			$\widehat{\rho}_1$			$\widehat{\rho}_2$		
n	$e_1 + e_2$	e_2	$e_1 + e_2$	e_2		n	$e_1 + e_2$	e_2	$e_1 + e_2$	e_2	
50	-0.03591	-0.01591	-0.04913	-0.01413		50	-0.13333	-0.10953	-0.00976	0.04564	
100	-0.01319	-0.00319	-0.01969	-0.00219		100	-0.07309	-0.06119	0.01046	0.03816	
150	-0.00847	-0.00181	-0.01271	-0.00104		150	-0.05414	-0.04621	0.01191	0.03037	

$(\rho_1, \rho_2) = (0.15, -0.55)$						$(\rho_1, \rho_2) = (0.15, 0.55)$					
$\widehat{\rho}_1$			$\widehat{\rho}_2$			$\widehat{\rho}_1$			$\widehat{\rho}_2$		
n	$e_1 + e_2$	e_2	$e_1 + e_2$	e_2		n	$e_1 + e_2$	e_2	$e_1 + e_2$	e_2	
50	-0.06690	-0.03890	-0.06690	-0.01390		50	-0.00862	-0.00262	0.00769	-0.00531	
100	-0.02493	-0.01093	-0.02765	-0.00115		100	-0.00305	-0.00005	0.00247	-0.00403	
150	-0.01475	-0.00542	-0.01663	0.00104		150	-0.00250	-0.00050	0.00095	-0.00338	

Table 7. Comparisons of estimated bias for the estimator of ρ_1 and ρ_2

$(\rho_1, \rho_2) = (0.25, -0.25)$						
n	$\widehat{\rho}_1$	$\widehat{\rho}_1^{MLE}$	$\widetilde{\rho}_1$	$\widehat{\rho}_2$	$\widehat{\rho}_2^{MLE}$	$\widetilde{\rho}_2$
50	-0.03591	-0.03166	-0.01421	-0.04913	-0.05512	-0.01118
100	-0.01319	-0.01087	-0.00286	-0.01969	-0.02115	-0.00160
150	-0.00847	-0.00716	-0.00166	-0.01271	-0.01338	-0.00079

$(\rho_1, \rho_2) = (0.40, -0.59)$						
n	$\widehat{\rho}_1$	$\widehat{\rho}_1^{MLE}$	$\widetilde{\rho}_1$	$\widehat{\rho}_2$	$\widehat{\rho}_2^{MLE}$	$\widetilde{\rho}_2$
50	-0.13333	-0.08952	-0.10667	-0.00976	-0.05341	0.04623
100	-0.07309	-0.03134	-0.06057	0.01046	-0.01752	0.03784
150	-0.05414	-0.01796	-0.04592	0.01191	-0.01095	0.03014

$(\rho_1, \rho_2) = (0.15, -0.55)$						
n	$\widehat{\rho}_1$	$\widehat{\rho}_1^{MLE}$	$\widetilde{\rho}_1$	$\widehat{\rho}_2$	$\widehat{\rho}_2^{MLE}$	$\widetilde{\rho}_2$
50	-0.06690	-0.06346	-0.03622	-0.06690	-0.09484	-0.00988
100	-0.02493	-0.01971	-0.01041	-0.02765	-0.03755	-0.00032
150	-0.01475	-0.01109	-0.00521	-0.01663	-0.02199	0.00137

$(\rho_1, \rho_2) = (0.15, 0.55)$						
n	$\widehat{\rho}_1$	$\widehat{\rho}_1^{MLE}$	$\widetilde{\rho}_1$	$\widehat{\rho}_2$	$\widehat{\rho}_2^{MLE}$	$\widetilde{\rho}_2$
50	-0.00862	-0.00733	-0.00260	0.00769	0.02588	-0.00577
100	-0.00305	-0.00265	-0.00004	0.00247	0.00854	-0.00410
150	-0.00250	-0.00223	-0.00049	0.00095	0.00473	-0.00340

The biases of the proposed estimators $\widetilde{\rho}_1$ and $\widetilde{\rho}_2$ are less than those of $\widehat{\rho}_1$ and $\widehat{\rho}_2$, respectively, except for the case as mentioned the above. Moreover, the calibration of $\widetilde{\rho}_1$ and $\widetilde{\rho}_2$ depends on the values of $\widehat{\rho}_1$ and $\widehat{\rho}_2$. Next, we compare the MSEs between $\widehat{\rho}_1$ and $\widetilde{\rho}_1$, and between $\widehat{\rho}_2$ and $\widetilde{\rho}_2$.

There is no difference between the estimated MSEs of $\widehat{\rho}_i$ and $\widetilde{\rho}_i$ ($i = 1, 2$), which certainly coincides with the discussion in Section 3.

5 Proof of theorems

We show our main Theorems 3.1 and 3.2 using lemmas and propositions in Appendix A.

Table 8. Comparisons of the estimated MSEs for the estimators of ρ_1 and ρ_2

$(\rho_1, \rho_2) = (0.25, -0.25)$						
n	$\hat{\rho}_1$	$\hat{\rho}_1^{\text{MLE}}$	$\tilde{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_2^{\text{MLE}}$	$\tilde{\rho}_2$
50	0.03102	0.03332	0.02822	0.03684	0.04005	0.03055
100	0.01136	0.01157	0.01090	0.01233	0.01264	0.01124
150	0.00706	0.00711	0.00686	0.00739	0.00748	0.00695
$(\rho_1, \rho_2) = (0.40, -0.59)$						
n	$\hat{\rho}_1$	$\hat{\rho}_1^{\text{MLE}}$	$\tilde{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_2^{\text{MLE}}$	$\tilde{\rho}_2$
50	0.04705	0.03905	0.03866	0.02783	0.03581	0.02665
100	0.01566	0.01050	0.01364	0.00944	0.01016	0.01021
150	0.00921	0.00579	0.00826	0.00564	0.00569	0.00619
$(\rho_1, \rho_2) = (0.15, -0.55)$						
n	$\hat{\rho}_1$	$\hat{\rho}_1^{\text{MLE}}$	$\tilde{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_2^{\text{MLE}}$	$\tilde{\rho}_2$
50	0.03108	0.03492	0.02635	0.03334	0.04263	0.02561
100	0.01051	0.01124	0.00972	0.01123	0.01288	0.00985
150	0.00597	0.00613	0.00568	0.00620	0.00668	0.00569
$(\rho_1, \rho_2) = (0.15, 0.55)$						
n	$\hat{\rho}_1$	$\hat{\rho}_1^{\text{MLE}}$	$\tilde{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_2^{\text{MLE}}$	$\tilde{\rho}_2$
50	0.01943	0.01993	0.01848	0.02612	0.03126	0.02306
100	0.00814	0.00818	0.00795	0.00937	0.00993	0.00883
150	0.00511	0.00513	0.00503	0.00559	0.00577	0.00538

5.1 Proof of Theorem 3.1

The second derivatives are given by the inverse of the Fisher information matrix and the components of the matrix can be obtained by the expectation of the second derivative of the log-likelihood functions. The expectations of the components are given in Proposition A.4, and then the Fisher information matrix is

$$I_n(\theta) = \begin{bmatrix} i_{\mu\mu} & 0 & 0 & 0 \\ 0 & i_{\sigma\sigma} & 0 & 0 \\ 0 & 0 & i_{\rho_1\rho_1} & i_{\rho_1\rho_2} \\ 0 & 0 & i_{\rho_1\rho_2} & i_{\rho_2\rho_2} \end{bmatrix},$$

where

$$i_{\mu\mu} = -E[l_{\mu\mu}] = \frac{1}{\sigma^2} \sum_{t=1}^n d_{t-1}^2, \quad i_{\sigma\sigma} = -E[l_{\sigma\sigma}] = \frac{2n}{\sigma^2},$$

$$i_{\rho_1\rho_1} = -E[l_{\rho_1\rho_1}] = \sum_{t=1}^n \sum_{k=1}^{t-1} (\varphi_1(k))^2,$$

$$i_{\rho_2\rho_2} = -E[l_{\rho_2\rho_2}] = \sum_{t=1}^n \sum_{k=1}^{t-2} (\varphi_1(k))^2, \quad i_{\rho_1\rho_2} = -E[l_{\rho_1\rho_2}] = \sum_{t=1}^n \sum_{k=2}^{t-1} \varphi_1(k)\varphi_1(k-1).$$

The functions φ_1 and d_t are defined in (A.3) and (A.5), respectively. Thus, the inverse matrix is given by

$$I_n(\boldsymbol{\theta})^{-1} = \begin{bmatrix} i^{\mu\mu} & 0 & 0 & 0 \\ 0 & i^{\sigma\sigma} & 0 & 0 \\ 0 & 0 & i^{\rho_1\rho_1} & i^{\rho_1\rho_2} \\ 0 & 0 & i^{\rho_1\rho_2} & i^{\rho_2\rho_2} \end{bmatrix}, \tag{15}$$

where

$$\begin{aligned} i^{\mu\mu} &= \frac{1}{i_{\mu\mu}} = \frac{\sigma^2}{\sum_{t=1}^n d_{t-1}^2}, & i^{\sigma\sigma} &= \frac{1}{i_{\sigma\sigma}} = \frac{\sigma^2}{2n}, \\ i^{\rho_1\rho_1} &= M^{-1}i_{\rho_2\rho_2}, & i^{\rho_1\rho_2} &= -M^{-1}i_{\rho_1\rho_2}, & i^{\rho_2\rho_2} &= M^{-1}i_{\rho_1\rho_1}, \\ & & M &= i_{\rho_1\rho_1}i_{\rho_2\rho_2} - (i_{\rho_1\rho_2})^2. \end{aligned}$$

We shall compute the limiting values for $i^{\mu\mu}$, $i^{\sigma\sigma}$, $i^{\rho_1\rho_1}$, $i^{\rho_1\rho_2}$, and $i^{\rho_2\rho_2}$. Using the different expression (B.4) for φ_1 , we have

$$d_{t-1} = \sum_{k=1}^t \varphi_1(k) = \frac{1}{1 + \rho_1 + \rho_2} + \frac{\rho_2\varphi_1(t) - \varphi_1(t + 1)}{1 + \rho_1 + \rho_2}. \tag{16}$$

We consider the summation from $t = 1$ to n of d_{t-1}^2 to get $i^{\mu\mu}$. The second term in d_{t-1} is not a main term for the summation as $n \rightarrow \infty$. Therefore,

$$i^{\mu\mu} = \frac{\sigma^2(1 + \rho_1 + \rho_2)^2}{n} + o(n^{-1}).$$

Similarly, we have

$$\sum_{k=1}^{t-1} (\varphi_1(k))^2 = \frac{1}{\Delta^2} \left(\frac{\lambda_1^2 - \lambda_1^{2t}}{1 - \lambda_1^2} - 2\frac{\rho_2 - \rho_2^t}{1 - \rho_2} + \frac{\lambda_2^2 - \lambda_2^{2t}}{1 - \lambda_2^2} \right)$$

and

$$\sum_{k=2}^{t-1} \varphi_1(k)\varphi_1(k - 1) = \frac{1}{\Delta^2} \left(\frac{\lambda_1^2 - \lambda_1^{2t-1}}{1 - \lambda_1^2} + \rho_1\frac{1 - \rho_2^t}{1 - \rho_2} + \frac{\lambda_2^2 - \lambda_2^{2t-1}}{1 - \lambda_2^2} \right),$$

and then

$$\begin{aligned} i_{\rho_1\rho_1} = i_{\rho_2\rho_2} &= \frac{1}{\Delta^2} \left(\frac{\lambda_1^2}{1 - \lambda_1^2} - 2\frac{\rho_2}{1 - \rho_2} + \frac{\lambda_2^2}{1 - \lambda_2^2} \right) n + o(n) \\ &= -\frac{1 + \rho_2}{(1 - \rho_1^2 + 2\rho_2 + \rho_2^2)(1 - \rho_2)} n + o(n) \end{aligned}$$

and

$$i_{\rho_1\rho_2} = \frac{1}{\Delta^2} \left(\frac{\lambda_1}{1 - \lambda_1^2} + \frac{\rho_1}{1 - \rho_2} + \frac{\lambda_2}{1 - \lambda_2^2} \right) n + o(n)$$

$$= -\frac{\rho_1}{(1 - \rho_1^2 + 2\rho_2 + \rho_2^2)(1 - \rho_2)}n + o(n).$$

Thus,

$$M = i_{\rho_1\rho_1}i_{\rho_2\rho_2} - (i_{\rho_1\rho_2})^2 = -\frac{1}{(1 - \rho_1^2 + 2\rho_2 + \rho_2^2)(1 - \rho_2)^2}n^2 + o(n^2).$$

Therefore, we obtain

$$i^{\rho_1\rho_1} = i^{\rho_2\rho_2} = \frac{1 - \rho_2^2}{n} + o(n^{-1}), \tag{17}$$

$$i^{\rho_1\rho_2} = \frac{\rho_1(1 - \rho_2)}{n} + o(n^{-1}). \tag{18}$$

5.2 Proof of Theorem 3.2

Eq. (9) is trivial by Lemma 2.3 and Proposition A.6. We show (10) using Lemma A.7. The lemma can be reduced to

$$E[\widehat{\theta}_r - \theta_r] = \frac{1}{2}i^{\theta_r\theta_r} \sum_{t=1}^d \sum_{u=1}^d i^{\theta_t\theta_u} (v_{\theta_r\theta_t\theta_u} + 2v_{\theta_r\theta_t,\theta_u}) + O(n^{-3/2}) \tag{19}$$

if $i^{\theta_r\theta_s} = 0$ for all $s \neq r$. We see $i^{\sigma\mu} = i^{\sigma\rho_1} = i^{\sigma\rho_1} = 0$ from (15), and then we can apply (19) for the bias $\widehat{\sigma}$. The components in the sum (19) are

$$i^{\mu\mu} (v_{\sigma\mu\mu} + 2v_{\sigma\mu,\mu}) = \frac{\sigma^2}{\sum_{t=1}^n d_{t-1}^2} \left(-\frac{2}{\sigma^3} \sum_{t=1}^n d_{t-1}^2 \right) = -\frac{2}{\sigma}, \tag{20}$$

$$i^{\sigma\sigma} (v_{\sigma\sigma\sigma} + 2v_{\sigma\sigma,\sigma}) = \frac{\sigma^2}{2n} \left(-\frac{2n}{\sigma^3} \right) = -\frac{1}{\sigma}, \tag{21}$$

$$i^{\rho_1\rho_1} (v_{\sigma\rho_1\rho_1} + 2v_{\sigma\rho_1,\rho_1}) = M^{-1}i_{\rho_2\rho_2} \left(-\frac{2}{\sigma}i_{\rho_1\rho_1} \right) = -\frac{2i_{\rho_1\rho_1}i_{\rho_2\rho_2}}{M\sigma}, \tag{22}$$

$$i^{\rho_2\rho_2} (v_{\sigma\rho_2\rho_2} + 2v_{\sigma\rho_2,\rho_2}) = M^{-1}i_{\rho_1\rho_1} \left(-\frac{2}{\sigma}i_{\rho_2\rho_2} \right) = -\frac{2i_{\rho_1\rho_1}i_{\rho_2\rho_2}}{M\sigma}, \tag{23}$$

$$\begin{aligned} & i^{\rho_1\rho_2} (v_{\sigma\rho_1\rho_2} + 2v_{\sigma\rho_1,\rho_2}) + i^{\rho_2\rho_1} (v_{\sigma\rho_2\rho_1} + 2v_{\sigma\rho_2,\rho_1}) \\ &= 2i^{\rho_1\rho_2} (v_{\sigma\rho_1\rho_2} + v_{\sigma\rho_1,\rho_2} + v_{\sigma\rho_2,\rho_1}) \\ &= -2M^{-1}i_{\rho_1\rho_2} \left(-\frac{2}{\sigma}i_{\rho_1\rho_2} \right) = \frac{4(i_{\rho_1\rho_2})^2}{M\sigma}. \end{aligned} \tag{24}$$

The summation of (22), (23), and (24) is

$$-\frac{4i_{\rho_1\rho_1}i_{\rho_2\rho_2}}{M\sigma} + \frac{4(i_{\rho_1\rho_2})^2}{M\sigma} = -\frac{4}{\sigma}.$$

By adding (20) and (21) on the above, we obtain (10).

Next, we show (11) and (12) using Proposition A.8. The 10 components of the equation in Lemma 2.3 which are used for the calculation of the biases of ρ_1 and ρ_2 are given by merely substituting the results in Proposition A.8:

$$i^{\mu\mu}(v_{\rho_1\mu\mu} + 2v_{\rho_1\mu,\mu}) = \frac{-2 \sum_{t=1}^n \sum_{k=1}^{t-1} d_{t-1}d_{t-k-1}\varphi_1(k)}{\sum_{s=1}^n d_{s-1}^2},$$

$$i^{\sigma\sigma}(v_{\rho_1\sigma\sigma} + 2v_{\rho_1\sigma,\sigma}) = 0,$$

$$i^{\rho_1\rho_1}(v_{\rho_1\rho_1\rho_1} + 2v_{\rho_1\rho_1,\rho_1}) = M^{-1}i_{\rho_2\rho_2} \sum_{t=1}^n (S_{1,t} + 2T_{0,0,t}),$$

$$2i^{\rho_1\rho_2}(v_{\rho_1\rho_1\rho_2} + v_{\rho_1\rho_1,\rho_2} + v_{\rho_1\rho_2,\rho_1}) = -2M^{-1}i_{\rho_1\rho_2} \sum_{t=1}^n (S_{2,t} + T_{0,1,t} + T_{1,0,t}),$$

$$i^{\rho_2\rho_2}(v_{\rho_1\rho_2\rho_2} + 2v_{\rho_1\rho_2,\rho_2}) = M^{-1}i_{\rho_1\rho_1} \sum_{t=1}^n (S_{3,t} + 2T_{1,1,t}),$$

and

$$i^{\mu\mu}(v_{\rho_2\mu\mu} + 2v_{\rho_2\mu,\mu}) = \frac{-2 \sum_{t=1}^n \sum_{k=1}^{t-2} d_{t-1}d_{t-k-2}\varphi_1(k)}{\sum_{s=1}^n d_{s-1}^2},$$

$$i^{\sigma\sigma}(v_{\rho_2\sigma\sigma} + 2v_{\rho_2\sigma,\sigma}) = 0,$$

$$i^{\rho_1\rho_1}(v_{\rho_2\rho_1\rho_1} + 2v_{\rho_2\rho_1,\rho_1}) = M^{-1}i_{\rho_2\rho_2} \sum_{t=1}^n (S_{0,t-1} + 2T_{1,0,t}),$$

$$2i^{\rho_1\rho_2}(v_{\rho_1\rho_2\rho_2} + v_{\rho_1\rho_2,\rho_2} + v_{\rho_2\rho_2,\rho_1}) = -2M^{-1}i_{\rho_1\rho_2} \sum_{t=1}^n (S_{1,t-1} + T_{1,1,t} + T_{0,0,t-1}),$$

$$i^{\rho_2\rho_2}(v_{\rho_2\rho_2\rho_2} + 2v_{\rho_2\rho_2,\rho_2}) = M^{-1}i_{\rho_1\rho_1} \sum_{t=1}^n (S_{2,t-1} + 2T_{0,1,t}),$$

where $S_{p,q}$ and $T_{p,q,t}$ are defined in (A.26) and (A.27), respectively. Now, we want to know the limiting values of the right-hand side of the above equations and to evaluate the sum of the first 5 equations and the remaining 5 equations. Let U and V be the sum of the first 5 equations and that of the remaining 5 equations, respectively. We see that

$$\sum_{t=1}^n \sum_{k=1}^{t-1} d_{t-1}d_{t-k-1}\varphi_1(k) = \frac{1}{(\rho_1 + \rho_2 + 1)^3}n + o(n),$$

$$\sum_{t=1}^n \sum_{k=1}^{t-2} d_{t-1}d_{t-k-2}\varphi_1(k) = \frac{1}{(\rho_1 + \rho_2 + 1)^3}n + o(n)$$

by (16). What regards the other summations, the values are constructed by $S_{p,q}$ and $T_{p,q,t}$. The components of $S_{p,q}$ and $T_{p,q,t}$ are given by using (B.4) and (B.5) which are a sort of geometric series. We consider the summations of $S_{p,q}$ and $T_{p,q,t}$ with respect to t from 1 to n as $n \rightarrow \infty$. Since the components are expressed as a sort of geometric series, the following limiting evaluations are useful.

$$\sum_{t=1}^n \sum_{k=1}^t \alpha^k = \frac{\alpha}{1-\alpha}n + o(n), \quad \sum_{t=1}^n \sum_{k=1}^t (k+1)\alpha^{k+1} = \frac{\alpha^2(2-\alpha)}{(1-\alpha)^2}n + o(n),$$

$$\sum_{t=1}^n \sum_{k=1}^t \sum_{m=1}^{t-k} \alpha^k \beta^m = \frac{\alpha\beta}{(1-\alpha)(1-\beta)}n + o(n)$$

for any $|\alpha| < 1$ and $|\beta| < 1$. Then, the summations of $S_{p,t}$ and $T_{p,q,t}$ with respect to t are

$$\begin{aligned} \sum_{t=1}^n S_{1,t} &= \frac{2\rho_1(\rho_2+1)}{(\rho_2-1)\{\rho_1^2-(\rho_2+1)^2\}^2}n + o(n), \\ \sum_{t=1}^n S_{2,t} &= \frac{2\{\rho_1^2-\rho_2(\rho_2+1)^2\}}{(\rho_2-1)^2\{\rho_1^2-(\rho_2+1)^2\}^2}n + o(n), \\ \sum_{t=1}^n S_{3,t} &= \frac{2\rho_1\{-\rho_1^2+2\rho_2(\rho_2+1)\}}{(\rho_2-1)^2\{\rho_1^2-(\rho_2+1)^2\}^2}n + o(n), \\ \sum_{t=1}^n S_{0,t-1} &= \frac{2\{1+\rho_2(2-\rho_1^2+\rho_2)\}}{(\rho_2-1)^2\{\rho_1^2-(\rho_2+1)^2\}^2}n + o(n), \\ \sum_{t=1}^n S_{1,t-1} &= \frac{2\rho_1(\rho_2+1)}{(\rho_2-1)^2\{\rho_1^2-(\rho_2+1)^2\}^2}n + o(n), \\ \sum_{t=1}^n S_{2,t-1} &= \frac{2\{\rho_1^2-\rho_2(\rho_2+1)^2\}}{(\rho_2-1)^2\{\rho_1^2-(\rho_2+1)^2\}^2}n + o(n), \\ \sum_{t=1}^n T_{0,0,t} &= \frac{-2\rho_1(\rho_2+1)}{(\rho_2-1)\{\rho_1^2-(\rho_2+1)^2\}^2}n + o(n), \\ \sum_{t=1}^n T_{0,1,t} &= \frac{-2\{\rho_1^2-\rho_2(\rho_2+1)^2\}}{(\rho_2-1)^2\{\rho_1^2-(\rho_2+1)^2\}^2}n + o(n), \\ \sum_{t=1}^n T_{1,0,t} &= \frac{\rho_1^2+(\rho_2+1)^2}{(\rho_2-1)\{\rho_1^2-(\rho_2+1)^2\}^2}n + o(n), \\ \sum_{t=1}^n T_{1,1,t} &= \frac{\rho_1(1+\rho_1^2-2\rho_2-3\rho_2^2)}{(\rho_2-1)^2\{\rho_1^2-(\rho_2+1)^2\}^2}n + o(n), \\ \sum_{t=1}^n T_{0,0,t-1} &= \frac{-2\rho_1(\rho_2+1)}{(\rho_2-1)\{\rho_1^2-(\rho_2+1)^2\}^2}n + o(n). \end{aligned}$$

Now, we are ready to use Lemma 2.3. Using the above evaluations, we have

$$U = \frac{-2}{\rho_1 + \rho_2 + 1} + \frac{4\rho_1(\rho_2 + 1)^2 - 2\rho_1\{\rho_1^2 + (\rho_2 + 1)^2\}}{\{\rho_1^2 - (\rho_2 + 1)^2\}^2} + o(1),$$

$$V = \frac{-2}{\rho_1 + \rho_2 + 1} + \frac{4\rho_1(\rho_2 + 1)\{\rho_1^2 - \rho_2(\rho_2 + 1)^2\} + 2\rho_1^2(1 + \rho_1^2 - 2\rho_2 - 3\rho_2^2)}{(1 - \rho_2)\{\rho_1^2 - (\rho_2 + 1)^2\}^2} + o(1).$$

By (17) and (18), the biases of the unknown parameters $\widehat{\rho}_1$ and $\widehat{\rho}_2$ are

$$E[\widehat{\rho}_1 - \rho_1] = \frac{1}{2}i^{\rho_1\rho_1}U + \frac{1}{2}i^{\rho_1\rho_2}V = \frac{\rho_1 + \rho_2 - 1}{n} + o(n^{-1})$$

and

$$E[\widehat{\rho}_2 - \rho_2] = \frac{1}{2}i^{\rho_1\rho_2}U + \frac{1}{2}i^{\rho_2\rho_2}V = \frac{3\rho_2 - 1}{n} + o(n^{-1}).$$

Therefore, we obtain (11) and (12).

6 Practical examples

We provide a practical example using quarterly U.S. GNP from 1947(1) to 2002(3), $n = 223$ observations. The original data which are provided by the Federal Reserve Bank of St. Louis [16] are introduced after being adjusted as a good example for MA(2) in [13] when the data are transformed to compute the GNP rate from X_t by

$$Y_t = \nabla \log(X_t), \tag{25}$$

where X_t is the U.S. GNP. The adjusted data which are different from the original can be obtained from the web site by the author of the book.

We do not know true values for unknown parameters in the MA(2) but we assume that the GNP rate follows

$$Y_t = 0.0083 + \varepsilon_t + 0.3028\varepsilon_{t-1} + 0.2036\varepsilon_{t-2}, \tag{26}$$

where the coefficients are calculated by the (full) maximum likelihood estimation using $n = 223 - 1 = 222$ observations because we need to take the difference by (25) and confirm whether the bias using the model with the last 20 samples is reduced by our method or not.

For an unknown parameter θ ($\theta = \mu, \sigma, \rho_1, \rho_2$), MLE and QMLE in Table 9 correspond to the maximum likelihood estimate $\widehat{\theta}^{\text{MLE}}$ and the conditional maximum likelihood estimate (quasi-maximum likelihood estimate) $\widehat{\theta}$ using $n = 20$ observations. The corrected MLE and corrected QMLE correspond to the result by Tanaka (1984) [15] and $\tilde{\theta}$ defined in (13), respectively. The reason why there are dashes (-) in the table for μ is that the correction is not required for μ . The four biases in the table from the bottom correspond to those for MLE, QMLE, corrected MLE, and corrected QMLE. We note that the corrected MLE is expected to be the best when we use $n = 20$ observations because the estimate uses the full maximum likelihood estimation. The both corrections work well, namely the bias for true model (26) become small against the models without the corrections.

Table 9. MLE and conditional MLE with a correction for U.S. GNP using MA(2)

	μ	σ	ρ_1	ρ_2
TRUE ($n = 222$)	0.0083	0.0094	0.3028	0.2036
MLE ($n = 20$)	0.0071	0.0060	0.1446	0.1370
QMLE ($n = 20$)	0.0072	0.0060	0.1564	0.1321
corrected MLE	–	0.0065	0.1824	0.1680
corrected QMLE	–	0.0065	0.1939	0.1639
Bias for MLE	–0.0012	–0.0035	–0.1582	–0.0666
Bias for QMLE	–0.0012	–0.0035	–0.1464	–0.0714
Bias for corrected MLE	–	–0.0029	–0.1204	–0.0356
Bias for corrected QMLE	–	–0.0029	–0.1089	–0.0397

A Lemmas

We show the derivatives of the conditional log-likelihood function and expectations. Let

$$\lambda_1 + \lambda_2 = -\rho_1, \quad \lambda_1 - \lambda_2 = \Delta, \quad \lambda_1 \lambda_2 = \rho_2 \tag{A.1}$$

and

$$\lambda_i^2 + \rho_1 \lambda_i + \rho_2 = 0 \tag{A.2}$$

for $i = 1, 2$. Furthermore, we define the following functions:

$$\varphi_1(k) = \sum_{l=0}^{k-1} \lambda_1^l \lambda_2^{k-1-l}, \tag{A.3}$$

$$\varphi_2(k) = -2 \frac{\partial \varphi_1(k)}{\partial \rho_1}, \tag{A.4}$$

and

$$d_t = \sum_{k=1}^{t+1} \varphi_1(k). \tag{A.5}$$

These functions reduce calculation costs. We list the lemmas and propositions in the following. Their proofs are provided in Appendix B except for Lemmas A.3 and A.5 because Lemmas A.3 and A.5 just list the derivatives of the log-likelihood function and they are easily obtained by the chain rule for (A.6) and (A.7).

The first derivatives of the conditional log-likelihood function (5) for the Gaussian MA(2) model are given by

$$l_\mu = \frac{1}{\sigma^2} \sum_{t=1}^n d_{t-1} \varepsilon_t, \quad l_\sigma = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{t=1}^n \varepsilon_t^2, \tag{A.6}$$

$$l_{\rho_1} = -\frac{1}{\sigma^2} \sum_{t=1}^n \varepsilon_t \frac{\partial \varepsilon_t}{\partial \rho_1}, \quad l_{\rho_2} = -\frac{1}{\sigma^2} \sum_{t=1}^n \varepsilon_t \frac{\partial \varepsilon_t}{\partial \rho_2}. \tag{A.7}$$

Lemma A.1. *The first derivatives of ε_t in (6) are given by*

$$\frac{\partial \varepsilon_t}{\partial \mu} = -d_{t-1}, \tag{A.8}$$

$$\frac{\partial \varepsilon_t}{\partial \rho_1} = -\sum_{k=1}^{t-1} \varphi_1(k) \varepsilon_{t-k} = -\sum_{k=1}^{t-1} \varphi_1(t-k) \varepsilon_k, \tag{A.9}$$

$$\frac{\partial \varepsilon_t}{\partial \rho_2} = \frac{\partial \varepsilon_{t-1}}{\partial \rho_1} = -\sum_{k=1}^{t-2} \varphi_1(k) \varepsilon_{t-k-1} = -\sum_{k=1}^{t-2} \varphi_1(t-k-1) \varepsilon_k. \tag{A.10}$$

Lemma A.1 implies that $\partial \varepsilon_t / \partial \rho_1$ does not depend on ε_t , but is expressed as a linear combination of $\varepsilon_1, \dots, \varepsilon_{t-1}$. Furthermore, $\partial \varepsilon_t / \partial \rho_2$ does not depend on ε_{t-1} and ε_t , but depends on $\varepsilon_1, \dots, \varepsilon_{t-2}$.

Lemma A.2. *The second derivatives of ε_t in (6) are given by*

$$\frac{\partial^2 \varepsilon_t}{\partial \mu \partial \rho_1} = \frac{\partial d_{t-1}}{\partial \rho_1} = -\sum_{k=1}^{t-1} \varphi_1(k) d_{t-k-1}, \tag{A.11}$$

$$\frac{\partial^2 \varepsilon_t}{\partial \mu \partial \rho_2} = \frac{\partial^2 \varepsilon_{t-1}}{\partial \mu \partial \rho_1} = \frac{\partial d_{t-1}}{\partial \rho_2} = -\sum_{k=1}^{t-2} \varphi_1(k) d_{t-k-2}, \tag{A.12}$$

$$\frac{\partial^2 \varepsilon_t}{\partial \rho_1^2} = \sum_{k=1}^{t-2} \varphi_2(k+1) \varepsilon_{t-k-1} = \sum_{k=1}^{t-2} \varphi_2(t-k) \varepsilon_k, \tag{A.13}$$

$$\frac{\partial^2 \varepsilon_t}{\partial \rho_1 \partial \rho_2} = \frac{\partial^2 \varepsilon_{t-1}}{\partial \rho_1^2} = \sum_{k=1}^{t-3} \varphi_2(k+1) \varepsilon_{t-k-2} = \sum_{k=1}^{t-3} \varphi_2(t-k-1) \varepsilon_k, \tag{A.14}$$

$$\frac{\partial^2 \varepsilon_t}{\partial \rho_2^2} = \frac{\partial^2 \varepsilon_{t-2}}{\partial \rho_1^2} = \sum_{k=1}^{t-4} \varphi_2(k+1) \varepsilon_{t-k-3} = \sum_{k=1}^{t-4} \varphi_2(t-k-2) \varepsilon_k. \tag{A.15}$$

Similarly, $\partial^2 \varepsilon_t / \partial \rho_1^2$ is expressed as a linear combination of $\varepsilon_1, \dots, \varepsilon_{t-2}$, $\partial^2 \varepsilon_t / \partial \rho_1 \partial \rho_2$ is a linear combination of $\varepsilon_1, \dots, \varepsilon_{t-3}$, and $\partial^2 \varepsilon_t / \partial \rho_2^2$ is that of $\varepsilon_1, \dots, \varepsilon_{t-4}$.

Lemma A.3. *The second derivatives of the conditional log-likelihood function (5) for the Gaussian MA(2) model are given by*

$$\begin{aligned} l_{\mu\mu} &= -\frac{1}{\sigma^2} \sum_{t=1}^n d_{t-1}^2, \quad l_{\mu\sigma} = -\frac{2}{\sigma^3} \sum_{t=1}^n d_{t-1} \varepsilon_t, \quad l_{\mu\rho_1} = \frac{1}{\sigma^2} \sum_{t=1}^n \left(\varepsilon_t \frac{\partial d_{t-1}}{\partial \rho_1} + d_{t-1} \frac{\partial \varepsilon_t}{\partial \rho_1} \right), \\ l_{\mu\rho_2} &= \frac{1}{\sigma^2} \sum_{t=1}^n \left(\varepsilon_t \frac{\partial d_{t-1}}{\partial \rho_2} + d_{t-1} \frac{\partial \varepsilon_t}{\partial \rho_2} \right), \quad l_{\sigma\sigma} = \frac{n}{\sigma^2} - \frac{3}{\sigma^4} \sum_{t=1}^n \varepsilon_t^2, \quad l_{\sigma\rho_1} = \frac{2}{\sigma^3} \sum_{t=1}^n \varepsilon_t \frac{\partial \varepsilon_t}{\partial \rho_1}, \\ l_{\sigma\rho_2} &= \frac{2}{\sigma^3} \sum_{t=1}^n \varepsilon_t \frac{\partial \varepsilon_t}{\partial \rho_2}, \quad l_{\rho_1\rho_1} = -\frac{1}{\sigma^2} \sum_{t=1}^n \left\{ \left(\frac{\partial \varepsilon_t}{\partial \rho_1} \right)^2 + \varepsilon_t \frac{\partial^2 \varepsilon_t}{\partial \rho_1^2} \right\}, \\ l_{\rho_1\rho_2} &= -\frac{1}{\sigma^2} \sum_{t=1}^n \left(\frac{\partial \varepsilon_t}{\partial \rho_1} \frac{\partial \varepsilon_t}{\partial \rho_2} + \varepsilon_t \frac{\partial^2 \varepsilon_t}{\partial \rho_1 \partial \rho_2} \right), \quad l_{\rho_2\rho_2} = -\frac{1}{\sigma^2} \sum_{t=1}^n \left\{ \left(\frac{\partial \varepsilon_t}{\partial \rho_2} \right)^2 + \varepsilon_t \frac{\partial^2 \varepsilon_t}{\partial \rho_2^2} \right\}. \end{aligned}$$

Since $\partial\varepsilon_t/\partial\rho_1$ is a linear combination of ε_k ($k = 1, \dots, t - 1$), and $\partial\varepsilon_t/\partial\rho_2$ is a linear combination of ε_l ($l = 1, \dots, t - 2$) by Lemma A.1, their expectations equal 0. Lemma A.3 is easily obtained by the first derivatives (A.6) and (A.7) with the chain rule.

Proposition A.4. *The expectations of the second derivatives of the conditional log-likelihood function (5) for the Gaussian MA(2) model are given by*

$$\begin{aligned} E[l_{\mu\mu}] &= -\frac{1}{\sigma^2} \sum_{t=1}^n d_{t-1}^2, & E[l_{\sigma\sigma}] &= -\frac{2n}{\sigma^2}, \\ E[l_{\mu\sigma}] &= E[l_{\mu\rho_1}] = E[l_{\mu\rho_2}] = E[l_{\sigma\rho_1}] = E[l_{\sigma\rho_2}] = 0, \\ E[l_{\rho_1\rho_1}] &= -\sum_{t=1}^n \sum_{k=1}^{t-1} (\varphi_1(k))^2, \\ E[l_{\rho_1\rho_2}] &= -\sum_{t=1}^n \sum_{k=2}^{t-1} \varphi_1(k)\varphi_1(k-1), \\ E[l_{\rho_2\rho_2}] &= -\sum_{t=1}^n \sum_{k=1}^{t-2} (\varphi_1(k))^2. \end{aligned} \tag{A.16}$$

Lemma A.5. *The third derivatives of the conditional log-likelihood function (5) for the Gaussian MA(2) model are given by*

$$\begin{aligned} l_{\mu\mu\mu} &= 0, & l_{\mu\sigma\sigma} &= \frac{6}{\sigma^4} \sum_{t=1}^n d_{t-1}\varepsilon_t, & l_{\mu\rho_1\rho_1} &= \frac{1}{\sigma^2} \sum_{t=1}^n \left(d_{t-1} \frac{\partial^2\varepsilon_t}{\partial\rho_1^2} + \varepsilon_t \frac{\partial^2 d_{t-1}}{\partial\rho_1^2} + 2 \frac{\partial\varepsilon_t}{\partial\rho_1} \frac{\partial d_{t-1}}{\partial\rho_1} \right), \\ l_{\mu\rho_2\rho_2} &= \frac{1}{\sigma^2} \sum_{t=1}^n \left(d_{t-1} \frac{\partial^2\varepsilon_t}{\partial\rho_2^2} + \varepsilon_t \frac{\partial^2 d_{t-1}}{\partial\rho_2^2} + 2 \frac{\partial\varepsilon_t}{\partial\rho_2} \frac{\partial d_{t-1}}{\partial\rho_2} \right), \\ l_{\mu\rho_1\rho_2} &= \frac{1}{\sigma^2} \sum_{t=1}^n \left(\frac{\partial d_{t-1}}{\partial\rho_1} \frac{\partial\varepsilon_t}{\partial\rho_2} + \frac{\partial\varepsilon_t}{\partial\rho_1} \frac{\partial d_{t-1}}{\partial\rho_2} + \varepsilon_t \frac{\partial^2 d_{t-1}}{\partial\rho_1\partial\rho_2} + d_{t-1} \frac{\partial^2\varepsilon_t}{\partial\rho_1\partial\rho_2} \right), \\ l_{\sigma\mu\mu} &= \frac{2}{\sigma^3} \sum_{t=1}^n d_{t-1}^2, & l_{\sigma\sigma\sigma} &= -\frac{2n}{\sigma^3} + \frac{12}{\sigma^5} \sum_{t=1}^n \varepsilon_t^2, & l_{\sigma\rho_1\rho_1} &= \frac{2}{\sigma^3} \sum_{t=1}^n \left\{ \left(\frac{\partial\varepsilon_t}{\partial\rho_1} \right)^2 + \varepsilon_t \frac{\partial^2\varepsilon_t}{\partial\rho_1^2} \right\}, \\ l_{\sigma\rho_2\rho_2} &= \frac{2}{\sigma^3} \sum_{t=1}^n \left\{ \left(\frac{\partial\varepsilon_t}{\partial\rho_2} \right)^2 + \varepsilon_t \frac{\partial^2\varepsilon_t}{\partial\rho_2^2} \right\}, & l_{\sigma\rho_1\rho_2} &= \frac{2}{\sigma^3} \sum_{t=1}^n \left(\frac{\partial\varepsilon_t}{\partial\rho_1} \frac{\partial\varepsilon_t}{\partial\rho_2} + \varepsilon_t \frac{\partial^2\varepsilon_t}{\partial\rho_1\partial\rho_2} \right), \\ l_{\rho_1\mu\mu} &= -\frac{2}{\sigma^2} \sum_{t=1}^n d_{t-1} \frac{\partial d_{t-1}}{\partial\rho_1} = \frac{2}{\sigma^2} \sum_{t=1}^n \sum_{k=1}^{t-1} d_{t-1} d_{t-k-1} \varphi_1(k), \\ l_{\rho_1\sigma\sigma} &= -\frac{6}{\sigma^4} \sum_{t=1}^n \varepsilon_t \frac{\partial\varepsilon_t}{\partial\rho_1}, & l_{\rho_1\rho_1\rho_1} &= -\frac{1}{\sigma^2} \sum_{t=1}^n \left(3 \frac{\partial^2\varepsilon_t}{\partial\rho_1^2} \frac{\partial\varepsilon_t}{\partial\rho_1} + \varepsilon_t \frac{\partial^3\varepsilon_t}{\partial\rho_1^3} \right), \\ l_{\rho_1\rho_2\rho_2} &= -\frac{1}{\sigma^2} \sum_{t=1}^n \left(\frac{\partial\varepsilon_t}{\partial\rho_1} \frac{\partial^2\varepsilon_t}{\partial\rho_2^2} + 2 \frac{\partial\varepsilon_t}{\partial\rho_2} \frac{\partial^2\varepsilon_t}{\partial\rho_1\partial\rho_2} + \varepsilon_t \frac{\partial^3\varepsilon_t}{\partial\rho_1\partial\rho_2^2} \right), \\ l_{\rho_1\rho_1\rho_2} &= -\frac{1}{\sigma^2} \sum_{t=1}^n \left(\frac{\partial\varepsilon_t}{\partial\rho_2} \frac{\partial^2\varepsilon_t}{\partial\rho_1^2} + 2 \frac{\partial\varepsilon_t}{\partial\rho_1} \frac{\partial^2\varepsilon_t}{\partial\rho_1\partial\rho_2} + \varepsilon_t \frac{\partial^3\varepsilon_t}{\partial\rho_1^2\partial\rho_2} \right), \end{aligned}$$

$$l_{\rho_2\mu\mu} = -\frac{2}{\sigma^2} \sum_{t=1}^n d_{t-1} \frac{\partial d_{t-1}}{\partial \rho_2} = \frac{2}{\sigma^2} \sum_{t=1}^n \sum_{k=1}^{t-2} d_{t-1} d_{t-k-2} \varphi_1(k),$$

$$l_{\rho_2\sigma\sigma} = -\frac{6}{\sigma^4} \sum_{t=1}^n \varepsilon_t \frac{\partial \varepsilon_t}{\partial \rho_2}, \quad l_{\rho_2\rho_2\rho_2} = -\frac{1}{\sigma^2} \sum_{t=1}^n \left(3 \frac{\partial^2 \varepsilon_t}{\partial \rho_2^2} \frac{\partial \varepsilon_t}{\partial \rho_2} + \varepsilon_t \frac{\partial^3 \varepsilon_t}{\partial \rho_2^3} \right).$$

We easily obtain the third derivatives in Lemma A.5 by Lemma A.3 and the chain rule. The following three propositions are derived from the expectations of the third derivatives and the products of the first and second derivatives of the conditional log-likelihood function (5). The three propositions correspond to the unknown parameters μ , σ , ρ_1 and ρ_2 , respectively.

Proposition A.6. *The expectations of the third derivatives and the products of the first and second derivatives of the conditional log-likelihood function (5) for the Gaussian MA(2) model are given by*

$$v_{\mu\mu\mu} = v_{\mu\sigma\sigma} = v_{\mu\rho_1\rho_1} = v_{\mu\rho_2\rho_2} = v_{\mu\rho_1\rho_2} = 0,$$

$$v_{\mu\mu,\mu} = v_{\mu\sigma,\sigma} = v_{\mu\rho_1,\rho_1} = v_{\mu\rho_1,\rho_2} = v_{\mu\rho_2,\rho_1} = v_{\mu\rho_2,\rho_2} = 0.$$

Proposition A.7. *The expectations of the third derivatives and the products of the first and second derivatives of the conditional log-likelihood function (5) for the Gaussian MA(2) model are given by*

$$v_{\sigma\mu\mu} = \frac{2}{\sigma^3} \sum_{t=1}^n d_{t-1}^2, \quad v_{\sigma\sigma\sigma} = \frac{10n}{\sigma^3},$$

$$v_{\sigma\rho_1\rho_1} = \frac{2}{\sigma} i_{\rho_1\rho_1}, \quad v_{\sigma\rho_2\rho_2} = \frac{2}{\sigma} i_{\rho_2\rho_2}, \quad v_{\sigma\rho_1\rho_2} = \frac{2}{\sigma} i_{\rho_1\rho_2},$$

$$v_{\sigma\mu,\mu} = -\frac{2}{\sigma^3} \sum_{t=1}^n d_{t-1}^2, \quad v_{\sigma\sigma,\sigma} = -\frac{6n}{\sigma^3}, \quad v_{\sigma\rho_1,\rho_1} = -\frac{2}{\sigma} i_{\rho_1\rho_1},$$

$$v_{\sigma\rho_2,\rho_2} = -\frac{2}{\sigma} i_{\rho_2\rho_2}, \quad v_{\sigma\rho_1,\rho_2} = -\frac{2}{\sigma} i_{\rho_1\rho_2}, \quad v_{\sigma\rho_2,\rho_1} = -\frac{2}{\sigma} i_{\rho_1\rho_2}.$$

Proposition A.8. *The expectations of the third derivatives and the products of the first and second derivatives of the conditional log-likelihood function (5) for the Gaussian MA(2) model are given by*

$$v_{\rho_1\mu\mu} = \frac{2}{\sigma^2} \sum_{t=1}^n \sum_{k=1}^{t-1} d_{t-1} d_{t-k-1} \varphi_1(k), \quad v_{\rho_2\mu\mu} = \frac{2}{\sigma^2} \sum_{t=1}^n \sum_{k=1}^{t-2} d_{t-1} d_{t-k-2} \varphi_1(k), \tag{A.17}$$

$$v_{\rho_1\sigma\sigma} = 0, \quad v_{\rho_2\sigma\sigma} = 0, \tag{A.18}$$

$$v_{\rho_1\rho_1\rho_1} = 3 \sum_{t=1}^n S_{1,t}, \quad v_{\rho_2\rho_1\rho_1} = \sum_{t=1}^n (S_{0,t-1} + 2S_{2,t}), \tag{A.19}$$

$$v_{\rho_1\rho_2\rho_2} = \sum_{t=1}^n (S_{3,t} + 2S_{1,t-1}), \quad v_{\rho_2\rho_2\rho_2} = 3 \sum_{t=1}^n S_{2,t-1}, \tag{A.20}$$

$$v_{\rho_1\mu,\mu} = -\frac{2}{\sigma^2} \sum_{t=1}^n \sum_{k=1}^{t-1} d_{t-1}d_{t-k-1}\varphi_1(k), \quad v_{\rho_2\mu,\mu} = -\frac{2}{\sigma^2} \sum_{t=1}^n \sum_{k=1}^{t-2} d_{t-1}d_{t-k-2}\varphi_1(k), \tag{A.21}$$

$$v_{\rho_1\sigma,\sigma} = 0, \quad v_{\rho_2\sigma,\sigma} = 0, \tag{A.22}$$

$$v_{\rho_1\rho_1,\rho_1} = \sum_{t=1}^n (T_{0,0,t} - S_{1,t}), \quad v_{\rho_2\rho_1,\rho_1} = \sum_{t=1}^n (T_{1,0,t} - S_{2,t}), \tag{A.23}$$

$$v_{\rho_1\rho_2,\rho_2} = \sum_{t=1}^n (T_{1,1,t} - S_{1,t-1}), \quad v_{\rho_2\rho_2,\rho_2} = \sum_{t=1}^n (T_{0,1,t-1} - S_{2,t-1}), \tag{A.24}$$

$$v_{\rho_1\rho_1,\rho_2} = \sum_{t=1}^n (T_{0,1,t} - S_{0,t-1}), \quad v_{\rho_2\rho_2,\rho_1} = \sum_{t=1}^n (T_{0,0,t-1} - S_{3,t}), \tag{A.25}$$

where

$$S_{p,t} = \sum_{k=1}^{t-p-1} \varphi_1(k+p)\varphi_2(k+1), \tag{A.26}$$

$$T_{p,q,t} = -\sum_{k=1}^{t-p-1} \sum_{m=1}^{t-k-1-p-q} \varphi_1(k)\varphi_1(k+m+p+q)\varphi_1(m) - \sum_{k=1}^{t-1} \sum_{m=1}^{t-k-q-1} \varphi_1(k)\varphi_1(k+m-p+q)\varphi_1(m). \tag{A.27}$$

B Proof of lemmas and propositions

B.1 Proof of Lemma 2.2

We show that the solution of

$$\begin{cases} \varepsilon_t = Y_t - \mu - \rho_1\varepsilon_{t-1} - \rho_2\varepsilon_{t-2} & (t \geq 1), \\ \varepsilon_{-1} = \varepsilon_0 = 0 \end{cases}$$

coincides with (6) by a mathematical induction. For $t = -1$ and 0 , the values of the both sides of (6) are 0. Then, (6) holds when $t = -1, 0$.

We assume that

$$\begin{aligned} \varepsilon_{s-2} &= \sum_{k=0}^{s-3} \left(\sum_{l=0}^{s-k-3} \lambda_1^l \lambda_2^{s-k-l-3} \right) (Y_{k+1} - \mu), \\ \varepsilon_{s-1} &= \sum_{k=0}^{s-2} \left(\sum_{l=0}^{s-k-2} \lambda_1^l \lambda_2^{s-k-l-2} \right) (Y_{k+1} - \mu) \end{aligned}$$

hold for $s \geq 1$. By (A.1) and (A.2), we have

$$\begin{aligned} \varepsilon_s &= (Y_s - \mu) - \rho_1 \varepsilon_{s-1} - \rho_2 \varepsilon_{s-2} \\ &= (Y_s - \mu) - \rho_1 \sum_{k=0}^{s-2} \left(\sum_{l=0}^{s-k-2} \lambda_1^l \lambda_2^{s-k-l-2} \right) (Y_{k+1} - \mu) \\ &\quad - \rho_2 \sum_{k=0}^{s-2} \left(\sum_{l=0}^{s-k-3} \lambda_1^l \lambda_2^{s-k-l-3} \right) (Y_{k+1} - \mu) \\ &= (Y_s - \mu) + \sum_{k=0}^{s-2} \left\{ \sum_{l=0}^{s-k-3} \lambda_1^l \lambda_2^{s-k-l-1} - \rho_1 \lambda_1^{s-k-2} \right\} (Y_{k+1} - \mu) \\ &= (Y_s - \mu) + \sum_{k=0}^{s-2} \left(\sum_{l=0}^{s-k-1} \lambda_1^l \lambda_2^{s-k-l-1} \right) (Y_{k+1} - \mu). \end{aligned}$$

Therefore, (6) holds for all $t \geq -1$. □

B.2 Proof of Lemma A.1

First, we show (A.8). We see that ε_t can be written using φ_1 as

$$\varepsilon_t = \sum_{k=0}^{t-1} \varphi_1(t-k)(Y_{k+1} - \mu). \tag{B.1}$$

Thus,

$$\frac{\partial \varepsilon_t}{\partial \mu} = - \sum_{k=0}^{t-1} \varphi_1(t-k).$$

Therefore, (A.8) follows by (A.5).

Next, we show (A.9). We have, by (A.4) and (B.1),

$$\begin{aligned} -2 \frac{\partial \varepsilon_t}{\partial \rho_1} &= \sum_{k=0}^{t-1} \varphi_2(t-k)(Y_{k+1} - \mu) \\ &= \sum_{k=0}^{t-1} \varphi_2(t-k)(\varepsilon_{k+1} + \rho_1 \varepsilon_k + \rho_2 \varepsilon_{k-1}) \\ &= \sum_{k=1}^{t-2} (\varphi_2(t-k+1) + \rho_1 \varphi_2(t-k) + \rho_2 \varphi_2(t-k-1)) \varepsilon_k \\ &\quad + (\rho_1 \varphi_2(1) + \varphi_2(2)) \varepsilon_{t-1} + \varphi_2(1) \varepsilon_t \end{aligned} \tag{B.2}$$

since $\varepsilon_{-1} = \varepsilon_0 = 0$. We note that $\varphi_2(2) = -2(\frac{\partial \lambda_1}{\partial \rho_1} + \frac{\partial \lambda_2}{\partial \rho_1}) = 2$ and $\varphi_2(1) = 0$. Thus, the function does not contain ε_t in the linear combination. We know by (A.2) that

$$\varphi_1(k+1) + \rho_1 \varphi_1(k) + \rho_2 \varphi_1(k-1) = 0. \tag{B.3}$$

Therefore,

$$\varphi_2(k + 1) + \rho_1\varphi_2(k) + \rho_2\varphi_2(k - 1) = 2\varphi_1(k),$$

and then the coefficient of ε_k ($1 \leq k \leq t - 1$) in (B.2) is $2\varphi_1(t - k)$.

Finally, we prove (A.10). By (B.3), we have

$$\frac{\partial\varphi_1(k + 1)}{\partial\rho_2} + \rho_1\frac{\partial\varphi_1(k)}{\partial\rho_2} + \rho_2\frac{\partial\varphi_1(k - 1)}{\partial\rho_2} = -\varphi_1(k - 1),$$

and then

$$\begin{aligned} \frac{\partial\varepsilon_t}{\partial\rho_2} &= \sum_{k=1}^{t-2} \left(\frac{\partial\varphi_1(t - k + 1)}{\partial\rho_2} + \rho_1\frac{\partial\varphi_1(t - k)}{\partial\rho_2} + \rho_2\frac{\partial\varphi_1(t - k - 1)}{\partial\rho_2} \right) \varepsilon_k \\ &+ \left(\rho_1\frac{\partial\varphi_1(1)}{\partial\rho_2} + \frac{\partial\varphi_1(2)}{\partial\rho_2} \right) \varepsilon_{t-1} + \frac{\partial\varphi_1(1)}{\partial\rho_2} \varepsilon_t = -\sum_{k=1}^{t-2} \varphi_1(t - k - 1) \varepsilon_k \end{aligned}$$

by the same way as in (B.2).

B.3 Proof of Lemma A.2

First, we show (A.11). Eqs. (A.8) and (A.9) yield

$$\frac{\partial d_{t-1}}{\partial\rho_1} = -\frac{\partial^2\varepsilon_t}{\partial\rho_1\partial\mu} = \sum_{k=1}^{t-1} \varphi_1(k) \frac{\partial\varepsilon_{t-k}}{\partial\mu} = -\sum_{k=1}^{t-1} \varphi_1(k) d_{t-k-1}.$$

By applying the same way to (A.10), we have

$$\frac{\partial d_{t-1}}{\partial\rho_2} = -\frac{\partial^2\varepsilon_t}{\partial\rho_2\partial\mu} = -\frac{\partial}{\partial\mu} \frac{\partial\varepsilon_{t-1}}{\partial\rho_1} = \frac{\partial d_{t-2}}{\partial\rho_1}.$$

Then, we obtain (A.12).

Next, we show (A.13). First, we express φ_2 in a different form. By (A.1), we have

$$\frac{\partial\Delta}{\partial\rho_1} = \frac{\rho_1}{\sqrt{\rho_1^2 - 4\rho_2}} = \frac{\rho_1}{\Delta}, \quad \frac{\partial\lambda_1}{\partial\rho_1} = \frac{-\Delta + \rho_1}{2\Delta} = -\frac{\lambda_1}{\Delta}, \quad \frac{\partial\lambda_2}{\partial\rho_1} = -\frac{\Delta + \rho_1}{2\Delta} = \frac{\lambda_2}{\Delta}.$$

We note that we do not have any algebraic relationship among unknown parameters $(\mu, \sigma, \rho_1, \rho_2)$ such as $\Delta = \rho_1^2 - 4\rho_2 = 0$, $\mu = 0$, etc., which makes a different problem. Using these derivatives, we have

$$\varphi_1(k) = \frac{\lambda_1^k - \lambda_2^k}{\Delta}, \tag{B.4}$$

$$\varphi_2(k) = \frac{2}{\Delta^2} \left\{ \rho_1\varphi_1(k) + k(\lambda_1^k + \lambda_2^k) \right\}. \tag{B.5}$$

This implies

$$2 \sum_{l=1}^k \varphi_1(l)\varphi_1(k + 1 - l) = \varphi_2(k + 1)$$

and then, we get with (A.4)

$$\begin{aligned} \frac{\partial}{\partial \rho_1} \frac{\partial \varepsilon_t}{\partial \rho_1} &= - \sum_{k=1}^{t-1} \left\{ \frac{\partial \varphi_1(k)}{\partial \rho_1} \varepsilon_{t-k} - \varphi_1(k) \sum_{l=1}^{t-k-1} \varphi_1(l) \varepsilon_{t-k-l} \right\} \\ &= \sum_{s=1}^{t-2} \left\{ - \frac{\partial \varphi_1(s+1)}{\partial \rho_1} + \sum_{k=1}^s \varphi_1(k) \varphi_1(s+1-k) \right\} \varepsilon_{t-s-1} \\ &= \sum_{s=1}^{t-2} \varphi_2(s+1) \varepsilon_{t-s-1}. \end{aligned}$$

This gives (A.13), and we also obtain (A.14) and (A.15) by (A.10).

B.4 Proof of Proposition A.4

We show only (A.16) because the remaining equations are trivial by Lemma A.3. The reason why they trivial is that $\frac{\partial d_{t-1}}{\partial \rho_1}$ in $l_{\mu\rho_1}$ is not a function of ε_s ($1 \leq s \leq t$). As for $l_{\sigma\rho_1}$, it contains $\frac{\partial \varepsilon_t}{\partial \rho_1}$, but $\frac{\partial \varepsilon_t}{\partial \rho_1}$ is a linear combination of $\varepsilon_1, \dots, \varepsilon_{t-1}$ by (A.9). The expectation of $\varepsilon_t \varepsilon_s$ ($1 \leq s \leq t-1$) vanishes. So, we only consider (A.16) carefully.

$$\begin{aligned} E[l_{\rho_1\rho_2}] &= -\frac{1}{\sigma^2} E \left[\sum_{t=1}^n \frac{\partial \varepsilon_t}{\partial \rho_1} \frac{\partial \varepsilon_t}{\partial \rho_2} \right] \\ &= -\frac{1}{\sigma^2} E \left[\sum_{t=1}^n \sum_{k=1}^{t-1} \sum_{m=1}^{t-2} \varphi_1(k) \varepsilon_{t-k} \varphi_1(m) \varepsilon_{t-m-1} \right] \\ &= -\frac{1}{\sigma^2} E \left[\sum_{t=1}^n \sum_{k=2}^{t-1} \varphi_1(k) \varphi_1(k-1) \varepsilon_{t-k}^2 \right] = - \sum_{t=1}^n \sum_{k=2}^{t-1} \varphi_1(k) \varphi_1(k-1). \end{aligned}$$

B.5 Proof of Proposition A.6

The log-likelihood functions appearing in the first line in Proposition A.6 are linear functions of ε_k . Thus, the expectations vanish because the expectation for ε_k is 0. For the expectations in the second line, we see that the first or the cubic power of ε_k appears in the product of the first and the second derivatives of the log-likelihood functions as in

$$\begin{aligned} v_{\mu\mu,\mu} &= E[l_{\mu\mu}l_{\mu}] = -\frac{1}{\sigma^4} E \left[\sum_{t=1}^n \sum_{s=1}^n d_{t-1}^2 \varepsilon_s d_{s-1} \right] = 0, \\ v_{\mu\sigma,\sigma} &= -\frac{2}{\sigma^6} E \left[\sum_{t=1}^n \sum_{s=1}^n d_{t-1} \varepsilon_t \varepsilon_s^2 \right] = 0. \end{aligned}$$

We also see that the functions in the expectations of the following values are expressed as the product of three linear combinations of ε_k . At least, the first power or the cubic power of ε_k is included in the expression. Therefore, we have

$$\begin{aligned} v_{\mu\rho_1,\rho_1} &= -\frac{1}{\sigma^4} E \left[\sum_{t=1}^n \sum_{s=1}^n \varepsilon_t \frac{\partial d_{t-1}}{\partial \rho_1} \varepsilon_s \frac{\partial \varepsilon_s}{\partial \rho_1} + \sum_{t=1}^n \sum_{s=1}^n d_{t-1} \frac{\partial \varepsilon_t}{\partial \rho_1} \varepsilon_s \frac{\partial \varepsilon_s}{\partial \rho_1} \right] = 0, \\ v_{\mu\rho_1,\rho_2} &= v_{\mu\rho_2,\rho_1} = v_{\mu\rho_2,\rho_2} = 0. \end{aligned}$$

B.6 Proof of Proposition A.7

Unlike Proposition A.6, the product of the first and the second derivative of the log-likelihood function is expressed as the sum of the second or the fourth power of ε_k . The expectations are obtained by presenting courteously the power of ε_k as

$$\begin{aligned}
 v_{\sigma\mu,\mu} &= -\frac{2}{\sigma^5} E \left[\sum_{t=1}^n \sum_{s=1}^n d_{t-1} \varepsilon_t d_{s-1} \varepsilon_s \right] = -\frac{2}{\sigma^5} E \left[\sum_{t=1}^n d_{t-1}^2 \varepsilon_t^2 \right] = -\frac{2}{\sigma^3} \sum_{t=1}^n d_{t-1}^2, \\
 v_{\sigma\sigma,\sigma} &= -\frac{n^2}{\sigma^3} + \frac{n \cdot n\sigma^2}{\sigma^5} + \frac{3n \cdot n\sigma^2}{\sigma^5} - \frac{3}{\sigma^7} E \left[\sum_{t=1}^n \sum_{s=1}^n \varepsilon_t^2 \varepsilon_s^2 \right] \\
 &= \frac{3n^2}{\sigma^3} - \frac{3}{\sigma^3} n(n+2) = -\frac{6n}{\sigma^3}.
 \end{aligned}$$

Similarly, the following expectations are given by the second and the fourth moments. The resulting values are expressed using $i_{\rho_1\rho_1}$, $i_{\rho_1\rho_2}$, and $i_{\rho_2\rho_2}$ which are components in the Fisher information matrix.

$$\begin{aligned}
 v_{\sigma\rho_1,\rho_1} &= -\frac{2}{\sigma^5} E \left[\sum_{t=1}^n \sum_{s=1}^n \varepsilon_t \frac{\partial \varepsilon_t}{\partial \rho_1} \varepsilon_s \frac{\partial \varepsilon_s}{\partial \rho_1} \right] = -\frac{2}{\sigma^5} E \left[\sum_{t=1}^n \varepsilon_t^2 \left(\frac{\partial \varepsilon_t}{\partial \rho_1} \right)^2 \right] \\
 &= -\frac{2}{\sigma^5} E \left[\sum_{t=1}^n \sum_{k=1}^{t-1} \sum_{m=1}^{t-1} \varepsilon_t^2 \varphi_1(t-k) \varepsilon_k \varphi_1(t-m) \varepsilon_m \right] \\
 &= -\frac{2}{\sigma^5} \sum_{t=1}^n \sum_{k=1}^{t-1} (\varphi_1(t-k))^2 \sigma^4 \\
 &= -\frac{2}{\sigma} \sum_{t=1}^n \sum_{k=1}^{t-1} (\varphi_1(k))^2 = -\frac{2}{\sigma} i_{\rho_1\rho_1}, \\
 v_{\sigma\rho_2,\rho_2} &= -\frac{2}{\sigma} \sum_{t=1}^n \sum_{k=1}^{t-2} (\varphi_1(k))^2 = -\frac{2}{\sigma} i_{\rho_2\rho_2}, \\
 v_{\sigma\rho_1,\rho_2} &= -\frac{2}{\sigma^5} E \left[\sum_{t=1}^n \sum_{s=1}^n \varepsilon_t \frac{\partial \varepsilon_t}{\partial \rho_1} \varepsilon_s \frac{\partial \varepsilon_s}{\partial \rho_2} \right] = -\frac{2}{\sigma^5} E \left[\sum_{t=1}^n \varepsilon_t^2 \frac{\partial \varepsilon_t}{\partial \rho_1} \frac{\partial \varepsilon_t}{\partial \rho_2} \right] \\
 &= -\frac{2}{\sigma^5} E \left[\sum_{t=1}^n \sum_{k=1}^{t-1} \sum_{m=1}^{t-2} \varepsilon_t^2 \varphi_1(t-k) \varepsilon_k \varphi_1(t-m-1) \varepsilon_m \right] \\
 &= -\frac{2}{\sigma^5} E \left[\sum_{t=1}^n \sum_{k=1}^{t-2} \varepsilon_t^2 \varphi_1(t-k) \varphi_1(t-k-1) \varepsilon_k^2 \right] \\
 &= -\frac{2}{\sigma} \sum_{t=1}^n \sum_{k=2}^{t-1} \varphi_1(k) \varphi_1(k-1) = -\frac{2}{\sigma} i_{\rho_1\rho_2}, \\
 v_{\sigma\rho_2,\rho_1} &= -\frac{2}{\sigma^5} E \left[\sum_{t=1}^n \sum_{s=1}^n \varepsilon_t \frac{\partial \varepsilon_t}{\partial \rho_2} \varepsilon_s \frac{\partial \varepsilon_s}{\partial \rho_1} \right] = -\frac{2}{\sigma} i_{\rho_1\rho_2}.
 \end{aligned}$$

B.7 Proof of Proposition A.8

Eqs. (A.17) and (A.18) are trivial by the expressions of $l_{\rho_1\mu\mu}$, $l_{\rho_2\mu\mu}$, $l_{\rho_1\sigma\sigma}$, and $l_{\rho_2\sigma\sigma}$. First, we show (A.21). The expectation of $l_{\rho_1\mu,\mu}$ is expressed as

$$v_{\rho_1\mu,\mu} = \frac{1}{\sigma^4} E \left[\sum_{t=1}^n \sum_{s=1}^n \varepsilon_t \frac{\partial d_{t-1}}{\partial \rho_1} d_{s-1} \varepsilon_s \right] + \frac{1}{\sigma^4} E \left[\sum_{t=1}^n \sum_{s=1}^n d_{t-1} \frac{\partial \varepsilon_t}{\partial \rho_1} d_{s-1} \varepsilon_s \right]. \tag{B.6}$$

The first term of the right-hand side in (B.6) is given by

$$\begin{aligned} \frac{1}{\sigma^4} E \left[\sum_{t=1}^n \sum_{s=1}^n \varepsilon_t \frac{\partial d_{t-1}}{\partial \rho_1} d_{s-1} \varepsilon_s \right] &= \frac{1}{\sigma^2} \sum_{t=1}^n d_{t-1} \frac{\partial d_{t-1}}{\partial \rho_1} \\ &= -\frac{1}{\sigma^2} \sum_{t=1}^n \sum_{k=1}^{t-1} d_{t-1} d_{t-k-1} \varphi_1(k). \end{aligned}$$

Similarly, the second term in (B.6) is given by

$$\begin{aligned} &\frac{1}{\sigma^4} E \left[\sum_{t=1}^n \sum_{s=1}^n d_{t-1} \frac{\partial \varepsilon_t}{\partial \rho_1} d_{s-1} \varepsilon_s \right] \\ &= -\frac{1}{\sigma^4} E \left[\sum_{t=1}^n \sum_{s=1}^n \sum_{k=1}^{t-1} d_{t-1} \varphi_1(t-k) \varepsilon_k d_{s-1} \varepsilon_s \right] \\ &= -\frac{1}{\sigma^4} E \left[\sum_{t=1}^n \sum_{k=1}^{t-1} d_{t-1} d_{t-k-1} \varphi_1(t-k) \varepsilon_k^2 \right] = -\frac{1}{\sigma^2} \sum_{t=1}^n \sum_{k=1}^{t-1} d_{t-1} d_{t-k-1} \varphi_1(k). \end{aligned}$$

Hence, we obtain

$$v_{\rho_1\mu,\mu} = -\frac{2}{\sigma^2} \sum_{t=1}^n \sum_{k=1}^{t-1} d_{t-1} d_{t-k-1} \varphi_1(k).$$

By the same process, we have

$$v_{\rho_2\mu,\mu} = -\frac{2}{\sigma^2} \sum_{t=1}^n \sum_{k=1}^{t-2} d_{t-1} d_{t-k-2} \varphi_1(k).$$

Eq. (A.22) is trivial. Next, we show (B.7), (B.9), and (B.10) to prove (A.19), (A.20), (A.23), (A.24), and (A.25). We have

$$\begin{aligned} E \left[\frac{\partial^2 \varepsilon_{t-p+1}}{\partial \rho_1^2} \frac{\partial \varepsilon_t}{\partial \rho_1} \right] &= -E \left[\sum_{k=1}^{t-p-1} \sum_{m=1}^{t-1} \varphi_2(t-p-k+1) \varepsilon_k \varphi_1(t-m) \varepsilon_m \right] \\ &= -\sigma^2 \sum_{k=1}^{t-p-1} \varphi_1(t-k) \varphi_2(t-p-k+1) \end{aligned}$$

$$= -\sigma^2 \sum_{k=1}^{t-p-1} \varphi_1(k+p)\varphi_2(k+1) = -\sigma^2 S_{p,t}. \tag{B.7}$$

Moreover, we have

$$E \left[\sum_{s=1}^n \frac{\partial \varepsilon_{t-p}}{\partial \rho_1} \frac{\partial \varepsilon_t}{\partial \rho_1} \varepsilon_s \frac{\partial \varepsilon_{s-q}}{\partial \rho_1} \right] \\ = - E \left[\sum_{s=1}^n \sum_{k=1}^{t-p-1} \sum_{m=1}^{t-1} \sum_{l=1}^{s-q-1} \varphi_1(t-p-k)\varepsilon_k \varphi_1(t-m)\varepsilon_m \varepsilon_s \varphi_1(s-q-l)\varepsilon_l \right] \tag{B.8}$$

by (A.9) for $p, q \geq 0$. We consider the right-hand side of (B.8). First, we fix k and m . For l ($1 \leq l \leq s - q - 1$), s must not be l . The summation appears only when $s = k$ and $l = m$, or $l = k$ and $s = m$. When $s = k$ and $l = m$, we have

$$E \left[\sum_{k=1}^{t-p-1} \sum_{m=1}^{k-q-1} \varphi_1(t-p-k)\varphi_1(t-m)\varphi_1(k-q-m)\varepsilon_k^2 \varepsilon_m^2 \right] \\ = \sigma^4 \sum_{k=1}^{t-p-1} \sum_{m=1}^{k-q-1} \varphi_1(t-p-k)\varphi_1(t-k+q+m)\varphi_1(m) \\ = \sigma^4 \sum_{k=1}^{t-p-1} \sum_{m=1}^{t-k-1-p-q} \varphi_1(k)\varphi_1(k+m+p+q)\varphi_1(m).$$

On the other hand, when $l = k$ and $s = m$, we have

$$E \left[\sum_{m=1}^{t-1} \sum_{k=1}^{m-q-1} \varphi_1(t-p-k)\varphi_1(t-m)\varphi_1(m-q-k)\varepsilon_m^2 \varepsilon_k^2 \right] \\ = \sigma^4 \sum_{m=1}^{t-1} \sum_{k=1}^{m-q-1} \varphi_1(t-m)\varphi_1(t-m+k-p+q)\varphi_1(k) \\ = \sigma^4 \sum_{k=1}^{t-1} \sum_{m=1}^{t-k-q-1} \varphi_1(k)\varphi_1(k+m-p+q)\varphi_1(m).$$

Therefore,

$$E \left[\sum_{t=1}^n \sum_{s=1}^n \frac{\partial \varepsilon_{t-p}}{\partial \rho_1} \frac{\partial \varepsilon_t}{\partial \rho_1} \varepsilon_s \frac{\partial \varepsilon_{s-q}}{\partial \rho_1} \right] = \sigma^4 \sum_{t=1}^n T_{p,q,t}. \tag{B.9}$$

Moreover, we consider the following expectation:

$$E \left[\sum_{t=1}^n \sum_{s=1}^n \varepsilon_t \frac{\partial^2 \varepsilon_{t-p}}{\partial \rho_1} \varepsilon_s \frac{\partial \varepsilon_{s-q}}{\partial \rho_1} \right]$$

$$= -E \left[\sum_{t=1}^n \sum_{s=1}^n \sum_{k=1}^{t-p-2} \sum_{m=1}^{s-q-1} \varepsilon_t \varphi_2(t-p-k) \varepsilon_k \varepsilon_s \varphi_1(s-q-m) \varepsilon_m \right].$$

Similarly, we fix t and k . Then, the summation appears only when $s = t$ and $m = k$.

$$E \left[\sum_{t=1}^n \sum_{k=1}^{t-p-2} \varphi_2(t-p-k) \varphi_1(t-q-k) \varepsilon_t^2 \varepsilon_k^2 \right]$$

$$= \sigma^4 \sum_{t=1}^n \sum_{k=1}^{t-p-2} \varphi_1(k+p-q+1) \varphi_2(k+1).$$

Therefore,

$$E \left[\sum_{t=1}^n \sum_{s=1}^n \varepsilon_t \frac{\partial^2 \varepsilon_{t-p}}{\partial \rho_1} \varepsilon_s \frac{\partial \varepsilon_{s-q}}{\partial \rho_1} \right] = -\sigma^4 \sum_{t=1}^n S_{p-q+1, t-q}. \tag{B.10}$$

We complete the proof of the lemma by (B.7), (B.9), and (B.10).

Acknowledgement

The authors would like to thank anonymous referees for giving us suitable suggestions and comments. They gave us not only technical comments, but also practical insights. By their comments, we revised a lot of equations and added Remark 3.3 and Section 6. The paper was dramatically improved by their comments. The author also thanks Dr. Tao Zou (ANU) for giving us advices about the data generation process we use and for having discussions with us. This work was supported by JSPS KAKENHI Grant Numbers JP18K11200.

References

[1] Bao, Y.: Finite-sample bias of the conditional Gaussian maximum likelihood estimator in ARMA models. *Adv. Econom.* **36**, 207–244 (2016)

[2] Bao, Y.: A general result on the estimation bias of ARMA models. *J. Stat. Plan. Inference* **197**, 107–125 (2018). [MR3799026. https://doi.org/10.1016/j.jspi.2018.01.001](https://doi.org/10.1016/j.jspi.2018.01.001)

[3] Bao, Y., Ullah, A.: The second-order bias and mean squared error of estimators in time-series models. *J. Econom.* **140**, 650–669 (2007). [MR2408921. https://doi.org/10.1016/j.jeconom.2006.07.007](https://doi.org/10.1016/j.jeconom.2006.07.007)

[4] Bao, Y., Ullah, A.: Expectation of quadratic forms in normal and nonnormal variables with applications. *J. Stat. Plan. Inference* **140**, 1193–1205 (2010). [MR2581122. https://doi.org/10.1016/j.jspi.2009.11.002](https://doi.org/10.1016/j.jspi.2009.11.002)

[5] Barndorff-Nielsen, O.E., Cox, D.R.: *Inference and Asymptotics*. Chapman & Hall/CRC (1994). [MR1317097. https://doi.org/10.1007/978-1-4899-3210-5](https://doi.org/10.1007/978-1-4899-3210-5)

[6] Brockwell, P.J., Davis, R.A.: *Time Series: Theory and Methods*. Springer (1991). [MR2839251](https://doi.org/10.1007/978-1-4899-3210-5)

- [7] Cheang, W.K., Reinsel, G.C.: Bias reduction of autoregressive estimates in time series regression model through restricted maximum likelihood. *J. Am. Stat. Assoc.* **95**, 1173–1184 (2000). [MR1804241](#). <https://doi.org/10.2307/2669758>
- [8] Cordeiro, G.M., Klein, R.: Bias correction in ARMA models. *Stat. Probab. Lett.* **19**, 169–176 (1994). [MR1278646](#). [https://doi.org/10.1016/0167-7152\(94\)90100-7](https://doi.org/10.1016/0167-7152(94)90100-7)
- [9] Giummolé, F., Vidoni, P.: Improved prediction limits for a general class of Gaussian models. *J. Time Ser. Anal.* **31**, 483–493 (2010). [MR2732602](#). <https://doi.org/10.1111/j.1467-9892.2010.00680.x>
- [10] Hamilton, J.D.: *Time Series Analysis*. Princeton University Press (1994). [MR1278033](#)
- [11] Hillmer, S.C., Tiao, G.C.: Likelihood function of stationary multiple autoregressive moving average models. *J. Am. Stat. Assoc.* **74**(367), 652–660 (1979). [MR0548261](#)
- [12] Kurosawa, T., Noguchi, K., Honda, F.: Bias reduction of the maximum likelihood estimator for a conditional Gaussian MA(1) model. *Commun. Stat., Theory Methods* **31**(17), 8588–8602 (2017). [MR3680779](#). <https://doi.org/10.1080/03610926.2016.1185119>
- [13] Shumway, R.H., Stoffer, D.S.: *Time Series Analysis and Its Applications With R Examples*, 4th edn. Springer (2017)
- [14] Tanaka, K.: Asymptotic expansions associated with the AR(1) model with unknown mean. *Econometrica* **51**, 1221–1231 (1983). [MR0710228](#). <https://doi.org/10.2307/1912060>
- [15] Tanaka, K.: An asymptotic expansion associated with the maximum likelihood estimators in ARMA models. *J. R. Stat. Soc. B* **46**, 58–67 (1984). [MR0745216](#)
- [16] The Federal Reserve Bank of St. Louis. <http://research.stlouisfed.org/>
- [17] Wilson, G.T.: The estimation of parameters in multivariate time series models. *J. R. Stat. Soc.* **35**(367), 76–85 (1973). [MR0339421](#)