STATISTICA, anno LXXXII, n. 1, 2022

COMPARISON BETWEEN THE EXACT LIKELIHOOD AND WHITTLE LIKELIHOOD FOR MOVING AVERAGE PROCESSES

Xiaofei Xu [1]
Research Institute for Science and Engineering, Waseda University, Tokyo, Japan
Zhengze Li
Graduate School of Fundamental Science and Engineering, Waseda University, Tokyo, Japan
Masanobu Taniguchi
Research Institute for Science and Engineering, Waseda University, Tokyo, Japan

1. Introduction

Suppose that $\{X_t\}$ is a Gaussian stationary process, and $\mathbf{X}_T = (X_1, X_2, \ldots, X_T)'$ is an observed stretch from $\{X_t\}$. The likelihood function based on $\mathbf{X}_T$ is expressed in terms of the inverse matrix and the determinant of the variance-covariance matrix $\mathbf{V}$ of $\mathbf{X}_T$. If T is large, the calculation for the exact likelihood and the maximum likelihood estimator (MLE) becomes very difficult. Whittle (1961) introduced a feasible approximation of the log-likelihood for avoiding expensive matrix inversions in terms of the spectral density f_θ, which is defined by

$$W(\theta) \equiv -\frac{1}{2}\sum_{s=1}^{T}\left\{\log f_\theta(\lambda_s) + \frac{I_T(\lambda_s)}{f_\theta(\lambda_s)}\right\} \tag{1}$$

where $I_T(\lambda_s) = \frac{1}{2\pi T}\left|\sum_{t=1}^{T} X_t e^{it\lambda_s}\right|^2$ is the periodogram and $\lambda_s = \frac{2\pi s}{T}$.

The Whittle estimator is defined by the maximiser of $W(\theta)$ with respect to θ. The inference by $W(\theta)$ has been developed for variety of directions. For example, among many others, Hosoya and Taniguchi (1982) elucidated the asymptotics of Whittle estimator for vector-valued non-Gaussian processes permitted to the case when the spectral model is misspecified. Taniguchi and Kakizawa (2000) investigated the higher-order efficiency of Whittle estimators in autoregressive moving average process. Giraitis and Taqqu (1999) derived the properties of Whittle estimation in long-memory Gaussian

[1] Corresponding Author. E-mail: xu.xiaofei@aoni.waseda.jp

processes and nonlinear processes. It is known that MLE and Whittle estimator are asymptotically efficient. However, evidently, for finite sample size, the exact likelihood and Whittle likelihood are different, and so are MLE and Whittle estimator. In this paper we aim to evaluate these differences numerically. For moving average process of first order (MA(1)), Anderson (1971) derived a useful explicit form of $\mathbf{V}^{-1}$, see also Anderson (1977). Using this we evaluated the differences of exact likelihood and Whittle likelihood as well as the asymptotics of the corresponding estimators by simulation for MA(1) with coefficient θ. In fact, for autoregressive process of first order (AR(1)), Anderson (1971) derived a third-order polynomial equation which describes the exact MLE. However, the equation does not have a simple solution. Anderson (1971) then proposed an estimator which is the third-order approximation. Fujikoshi and Ochi (1984) investigated the third-order asymptotics of the estimator and compared it with the Whittle estimator, which becomes the Yule-Walker estimator (see Taniguchi, 1983). For MA(1), Anderson (1971, pp. 292-293) derived the exact form of the inverse matrix of the covariance matrix of an observed stretch. Based on this, in this paper, we investigate some behaviors of the exact likelihood and MLE of MA(1), and compare with the Whittle likelihood and Whittle estimator respectively.

Researchers have noticed and exploited the unique properties of estimators or test statistics when the true value of parameter vector nears or is at the boundary compared with the interior of the parameter space (Ketz, 2018). For example, Monti and Taniguchi (2018) developed the higher-order asymptotics for a class of test statistics when the parameter of interest is on the boundary of the parameter space. In this work, we also compare the performance of MLE and Whittle estimator when θ is close to the boundary, i.e. ± 1. We observed definite difference in the two likelihood functions and estimators, especially if θ nears 1 (unit root case) or -1, Whittle estimator deteriorates in comparison with MLE. Hence this is an important warning for using Whittle estimator when the parameter of moving average process nears the boundary of space.

This paper is organized as follows. Section 2 introduces the exact log likelihood and Whittle likelihood for MA(1), based on Anderson's explicit expression for variance matrix and the periodogram, respectively. In Section 3 we evaluate the expectation of the exact likelihood and Whittle likelihood. Section 4 discusses the numerical comparisons among the exact value, expectation, and estimators of the two likelihood functions. The results show that, if the parameter of moving average process nears the boundary of parameter space, the Whittle estimator becomes worse in comparison with MLE. Section 5 concludes.

2. Likelihood Function

Suppose that $\{X_t\}$ is generated by

$$X_t = \varepsilon_t + \theta\varepsilon_{t-1}, \quad (|\theta| < 1), \tag{2}$$

where ε_t follows independent standard normal distribution. Let $\mathbf{X}_T = (X_1, \ldots, X_T)'$ denote an observed stretch from $\{X_t\}$, and denote the covariance matrix by $\mathbf{V} \equiv \mathrm{Var}(\mathbf{X}_T)$. The exact likelihood function for $\mathbf{X}_T$, denoted by $L(\theta)$, is

$$L(\theta) = -\frac{1}{2}\log\det\mathbf{V} - \frac{1}{2}\mathbf{X}_T'\mathbf{V}^{-1}\mathbf{X}_T - \frac{T}{2}\log 2\pi, \tag{3}$$

where

$$\mathbf{V} = \begin{bmatrix} 1+\theta^2 & \theta & & \\ \theta & 1+\theta^2 & \theta & \\ & \theta & \ddots & \theta \\ & & \theta & 1+\theta^2 \end{bmatrix} = (1+\theta^2)\mathbf{I}_T + \theta \underbrace{\begin{pmatrix} 0 & 1 & \cdots & 0 \\ 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & 0 & 1 \\ 0 & \cdots & 1 & 0 \end{pmatrix}}_{L_T},$$

and $\mathbf{I}_T$ is the $T \times T$ identity matrix. Let

$$\mathbf{P}_T = \sqrt{\frac{2}{T+1}} \begin{bmatrix} \sin\frac{\pi}{T+1} & \cdots & \sin\frac{T\pi}{T+1} \\ \vdots & \ddots & \vdots \\ \sin\frac{T\pi}{T+1} & \cdots & \sin\frac{T^2\pi}{T+1} \end{bmatrix},$$

which is an orthogonal matrix. Anderson (1971) evaluated the form of covariance matrix as

$$\mathrm{L}_T = \mathbf{P}_T \begin{bmatrix} 2\cos\frac{\pi}{T+1} & & 0 \\ & \ddots & \\ 0 & & 2\cos\frac{\pi T}{T+1} \end{bmatrix} \mathbf{P}_T'.$$

Hence

$$\mathbf{V} = \mathbf{P}_T \begin{bmatrix} 1+\theta^2+2\theta\cdot\cos\frac{\pi}{T+1} & & 0 \\ & \ddots & \\ 0 & & 1+\theta^2+2\theta\cdot\cos\frac{\pi T}{T+1} \end{bmatrix} \mathbf{P}_T',$$

which implies

$$\mathbf{V}^{-1} = \mathbf{P}_T \begin{bmatrix} \{1+\theta^2+2\theta\cdot\cos\frac{\pi}{T+1}\}^{-1} & & 0 \\ & \ddots & \\ 0 & & \{1+\theta^2+2\theta\cdot\cos\frac{\pi T}{T+1}\}^{-1} \end{bmatrix} \mathbf{P}_T'.$$

PROPOSITION 1. *The exact log likelihood in (3) is equal to*

$$L(\theta) = -\frac{1}{2}\sum_{s=1}^{T}\log\left\{1+\theta^2+2\theta\cos\frac{\pi s}{T+1}\right\} - \frac{1}{2}\sum_{s=1}^{T}\frac{\xi(s)^2}{1+\theta^2+2\theta\cos\frac{\pi s}{T+1}} - \frac{T}{2}\log 2\pi, \tag{4}$$

where

$$\xi(s) = \sqrt{\frac{2}{T+1}}\sum_{t=1}^{T}X_t\sin\frac{\pi t s}{T+1}.$$

Whittle likelihood (Whittle, 1961) is known as a frequency-domain approximation for the actual likelihood in parametric spectral analysis. Given the observed time series $\mathbf{X}_T$, the Whittle likelihood, denoted by $L_W(\theta)$, is

$$L_W(\theta) = -\frac{1}{2}\sum_{s=1}^{T}\left\{\log f_\theta(\lambda_s) + \frac{I_T(\lambda_s)}{f_\theta(\lambda_s)} + \log(2\pi)\right\}, \tag{5}$$

where $f_\theta(\lambda_s) = \frac{1}{2\pi}\left|1+\theta e^{i\lambda_s}\right|^2$, $I_T(\lambda_s) = \frac{1}{2\pi T}\left|\sum_{t=1}^{T}X_t e^{it\lambda_s}\right|^2$, and $\lambda_s = \frac{2\pi s}{T}$.

3. EXPECTATION OF THE LOG-LIKELIHOOD

In this section we evaluate the expectation of the two log-likelihood $L(\theta)$ and $L_W(\theta)$, denoted by $E(L)$ and $E(L_W)$ respectively. To evaluate $E(L)$ we have

$$\begin{aligned} E\left\{\xi(s)^2\right\} &= E\left[\frac{2}{T+1}\sum_{t=1}^{T}\sum_{t'=1}^{T}X_tX_{t'}\sin\frac{\pi t s}{T+1}\sin\frac{\pi t' s}{T+1}\right] \\ &= \frac{2}{T+1}\left\{\sum_{t=1}^{T}E(X_t^2)(\sin\frac{\pi t s}{T+1})^2\right. \\ &\quad + \sum_{t=1}^{T-1}E\left(X_tX_{t-1}\right)\sin\frac{\pi t s}{T+1}\sin\frac{\pi(t+1)s}{T+1} \\ &\quad \left. + \sum_{t=2}^{T}E\left(X_tX_{t-1}\right)\sin\frac{\pi + s}{T+1}\sin\frac{\pi(t-1)s}{T+1}\right\}, \end{aligned} \tag{6}$$

where $E\left(X_t^2\right) = 1+\theta^2$ and $E\left(X_tX_{t-1}\right) = \theta$.
From (6) we can obtain the following result.

PROPOSITION 2. *The expectation of the exact log-likelihood $L(\theta)$ is*

$$
\begin{aligned}
E(L) = -\frac{1}{2}\sum_{s=1}^{T}\Big[&\log\Big(1+\theta^2+2\theta\cos\frac{\pi s}{T+1}\Big) \\
&+\frac{1+\theta^2+2\frac{T}{T+1}\theta\cos\frac{\pi s}{T+1}}{1+\theta^2+2\theta\cos\frac{\pi s}{T+1}}+\log 2\pi\Big].
\end{aligned} \tag{7}
$$

Next we evaluate $E(L_W)$. For this, note that

$$
E(I_T(\lambda_s)) = \frac{1}{2\pi}\sum_{\ell=-T+1}^{T-1} E(\hat{R}(\ell))e^{-i\ell\lambda_s},
$$

where $\hat{R}(\ell) = \frac{1}{T}\sum_{t=1}^{T-|\ell|} X_t X_{t+|\ell|}$.
Because

$$
E(\hat{R}(0)) = 1+\theta^2,\ E(\hat{R}(\pm 1)) = \theta,\ \text{and } E(\hat{R}(\ell)) = 0 \text{ for } |\ell| \geq 2,
$$

we then obtain

$$
E(I_T(\lambda_s)) = \frac{1}{2\pi}\big(1+\theta^2+2\theta\cos\lambda_s\big).
$$

Hence, we have the following result.

PROPOSITION 3. *The expectation of the Whittle likelihood $L_W(\theta)$ is*

$$
E(L_W) = -\frac{1}{2}\sum_{s=1}^{T}\Big[\log\left|1+\theta e^{i\lambda_s}\right|^2 + 1 + \log 2\pi\Big]. \tag{8}
$$

4. NUMERICAL STUDIES

As stated before, the exact likelihood includes the determinant and inverse of covariance matrix $\mathbf{V} \in \mathbb{R}^T \times \mathbb{R}^T$. If T is large, the exact likelihood is intractable. Whittle likelihood defined in Eq.(5) is a feasible approximation in a frequency-domain. However, for finite T, Whittle likelihood and the exact likelihood are different. In this section, we illustrate this difference numerically. We firstly consider various MA(1) processes with the true values of the parameter setting as $\theta \in \{\pm 0.1, \pm 0.2, \cdots, \pm 0.9\}$. We also investigate the performance when θ nears the boundary (± 1) and set the true values of the parameter as $\theta \in \{\pm 0.91, \pm 0.92 \cdots, \pm 0.99\}$. We repeat the experiments 500 times with various sample sizes $T = \{30, 50, 100\}$.

TABLE 1
The values of exact likelihood and Whittle likelihood functions and their difference under sample size 30, and the "Diff" column refers to $L(\theta)-L_W(\theta)$.

	$T=30$			$T=50$			$T=100$		
θ	$L(\theta)$	$L_W(\theta)$	Diff	$L(\theta)$	$L_W(\theta)$	Diff	$L(\theta)$	$L_W(\theta)$	Diff
0.1	-14.891	-14.893	0.002	-24.844	-24.850	0.006	-49.736	-49.743	0.007
0.2	-14.769	-14.794	0.025	-24.916	-24.944	0.028	-50.359	-50.387	0.028
0.3	-15.045	-15.097	0.052	-24.882	-24.923	0.041	-49.909	-49.982	0.073
0.4	-15.049	-15.146	0.097	-25.598	-25.751	0.154	-50.451	-50.564	0.113
0.5	-15.089	-15.298	0.209	-24.821	-25.032	0.210	-50.084	-50.281	0.196
0.6	-15.582	-15.905	0.324	-25.403	-25.763	0.360	-50.211	-50.568	0.356
0.7	-15.258	-15.935	0.676	-25.558	-26.142	0.585	-50.758	-51.373	0.615
0.8	-15.217	-16.490	1.273	-25.760	-27.205	1.445	-50.814	-52.241	1.427
0.9	-16.022	-20.240	4.218	-25.668	-29.132	3.463	-51.694	-54.667	2.973
0.91	-15.877	-20.339	4.462	-25.825	-29.643	3.818	-51.588	-55.669	4.081
0.92	-16.034	-21.655	5.620	-25.904	-30.803	4.900	-51.009	-55.317	4.308
0.93	-16.221	-23.637	7.416	-26.045	-30.943	4.898	-50.972	-56.242	5.270
0.94	-16.098	-24.141	8.043	-26.278	-33.740	7.462	-51.230	-58.103	6.873
0.95	-16.230	-30.256	14.026	-26.600	-38.328	11.727	-51.066	-60.315	9.249
0.96	-16.232	-36.165	19.933	-26.260	-41.032	14.772	-51.530	-62.931	11.401
0.97	-16.213	-51.962	35.749	-26.641	-49.572	22.930	-50.831	-66.675	15.844
0.98	-16.613	-92.643	76.030	-26.754	-77.134	50.380	-51.115	-85.095	33.980
0.99	-16.360	-380.252	363.892	-27.485	-222.605	195.120	-52.158	-146.903	94.745
-0.1	-14.890	-14.899	0.009	-24.792	-24.785	-0.007	-50.076	-50.077	0.002
-0.2	-14.785	-14.817	0.032	-25.313	-25.330	0.017	-49.914	-49.924	0.010
-0.3	-15.031	-15.071	0.040	-24.915	-24.956	0.041	-49.301	-49.395	0.094
-0.4	-15.060	-15.146	0.086	-25.031	-25.159	0.127	-49.807	-49.885	0.079
-0.5	-15.144	-15.313	0.169	-24.832	-25.009	0.177	-50.687	-50.881	0.194
-0.6	-15.598	-15.956	0.357	-24.878	-25.140	0.263	-50.920	-51.289	0.369
-0.7	-15.302	-15.962	0.660	-25.524	-26.129	0.604	-50.312	-50.916	0.605
-0.8	-15.216	-16.600	1.384	-25.002	-26.462	1.459	-51.004	-52.269	1.265
-0.9	-16.016	-20.159	4.143	-26.368	-30.162	3.794	-51.162	-54.653	3.491
-0.91	-15.873	-20.641	4.768	-25.944	-29.700	3.756	-50.674	-54.199	3.525
-0.92	-16.098	-21.504	5.406	-26.255	-31.589	5.334	-51.353	-55.907	4.554
-0.93	-16.212	-23.603	7.391	-26.322	-32.154	5.832	-50.809	-55.660	4.851
-0.94	-16.107	-24.460	8.353	-26.079	-33.312	7.233	-51.310	-57.683	6.373
-0.95	-16.164	-30.325	14.161	-26.131	-34.757	8.626	-50.865	-59.987	9.122
-0.96	-16.238	-35.990	19.753	-26.487	-40.409	13.923	-51.050	-60.834	9.783
-0.97	-16.244	-52.579	36.335	-26.697	-49.849	23.152	-51.740	-66.725	14.985
-0.98	-16.626	-92.263	75.637	-26.806	-76.500	49.694	-51.301	-85.415	34.114
-0.99	-16.303	-380.636	364.333	-26.711	-234.113	207.403	-52.065	-143.982	91.917

4.1. Likelihood function comparison

We compare the value of exact likelihood $L(\theta)$ in (4) and Whittle likelihood function $L_W(\theta)$ in (5) and the corresponding expectation defined in (7) and (8) respectively.

Table 1 reports the average values of $L(\theta)$ and $L_W(\theta)$ over 500 repetitions with various parameters and sample sizes. It is shown that $L(\theta)$ is larger than $L_W(\theta)$ among different values of θ and various sample sizes. When the absolute value of θ is not too large, e.g., smaller than 0.7, the difference of $L(\theta)$ and $L_W(\theta)$ is smaller than 1 for all the three T cases, which is not very significant. While when θ increases from 0.9 to 0.99 or decreases from -0.9 to -0.99, that is $|\theta|$ approaches to 1, the deviation of $L(\theta)$ to $L_W(\theta)$ dramatically increases from around 4 to 360 at T=30, from about 3.5 to 200 at T =50, and from about 3 to 91 when T=100. That is, the difference greatly increases when the absolute value of θ becomes closer to 1 and larger sample size shows some benefits for reducing such difference. The sign of θ does not have significant effect to the performance of likelihoods. We conclude that Whittle likelihood is a good approximate to the exact likelihood with similar performance when the absolute value of θ is not too large, while if a moving average parameter is close to ± 1, this difference becomes dramatic. This is a big warning when we use Whittle likelihood and estimator for the MA process.

Table 2 reports the comparison of performance between the expectations $E(L)$ and $E(L_W)$ under different parameter values of θ, and sample size T equals 30, 50, and 100 respectively. From Eq.(7) and Eq.(8), it is not difficult to prove that the sign of θ does not affect the value of $E(L)$ and $E(L_W)$, thus in Table 2 we report a single result for $\pm\theta$ in one row. It is shown that $E(L_W)$ is always larger than $E(L)$ for all the scenarios. When the absolute value of θ increases from 0.8 to 0.99, the deviation of $E(L)$ and $E(L_W)$ also exhibits large increments from around $1.97 \sim 2.19$ to $7.56 \sim 17.63$ among different sample sizes. That is, the difference greatly increases when the MA(1) process has near unit root ($\theta \to -1$) or $\theta \to 1$. It also shows that the value of $E(L_W)$ is quite stable for all θ, which differs from the poor performance of $L_W(\theta)$ at $\theta > 0.90$ shown in Table 1. The stable performance of $E(L_W)$ is because that $E(L_W)$ is less sensitive to the changing of θ than $E(L)$.

4.2. Estimation accuracy comparison

In this part we compare the parameter estimation accuracy of MLE denoted by $\hat{\theta}$ and the Whittle likelihood estimator denoted with $\tilde{\theta}_W$ for the MA(1) process in model (2). We evaluate the performance of two estimators in terms of the estimation bias (Bias), the mean of absolute error (MAE) and mean squared error (MSE), which are respectively

TABLE 2
Comparison between the expectations of exact likelihood $E(L)$ and Whittle likelihood $E(L_W)$, and the "Diff" column refers to $E(L)-E(L_W)$.

	T=30			T=50			T=100		
θ	$E(L)$	$E(L_W)$	Diff	$E(L)$	$E(L_W)$	Diff	$E(L)$	$E(L_W)$	Diff
±0.1	-15.014	-15.000	-0.014	-25.015	-25.000	-0.015	-50.015	-50.000	-0.015
±0.2	-15.059	-15.000	-0.059	-25.060	-25.000	-0.060	-50.061	-50.000	-0.061
±0.3	-15.139	-15.000	-0.139	-25.142	-25.000	-0.142	-50.144	-50.000	-0.144
±0.4	-15.263	-15.000	-0.263	-25.269	-25.000	-0.269	-50.273	-50.000	-0.273
±0.5	-15.449	-15.000	-0.449	-25.460	-25.000	-0.460	-50.468	-50.000	-0.468
±0.6	-15.729	-15.000	-0.729	-25.751	-25.000	-0.751	-50.768	-50.000	-0.768
±0.7	-16.176	-15.000	-1.176	-26.224	-25.000	-1.224	-51.260	-50.000	-1.260
±0.8	-16.970	-14.999	-1.971	-27.095	-25.000	-2.095	-52.191	-50.000	-2.191
±0.9	-18.659	-14.957	-3.702	-29.214	-24.995	-4.219	-54.649	-50.000	-4.649
±0.91	-18.919	-14.939	-3.980	-29.599	-24.991	-4.608	-55.143	-50.000	-5.143
±0.92	-19.198	-14.914	-4.284	-30.043	-24.984	-5.058	-55.737	-50.000	-5.737
±0.93	-19.495	-14.880	-4.615	-30.553	-24.973	-5.580	-56.464	-49.999	-6.465
±0.94	-19.805	-14.830	-4.975	-31.137	-24.954	-6.184	-57.375	-49.998	-7.377
±0.95	-20.121	-14.758	-5.362	-31.799	-24.920	-6.880	-58.541	-49.994	-8.547
±0.96	-20.432	-14.652	-5.780	-32.527	-24.861	-7.667	-60.065	-49.983	-10.082
±0.97	-20.728	-14.487	-6.241	-33.286	-24.754	-8.532	-62.069	-49.951	-12.118
±0.98	-20.994	-14.211	-6.783	-34.007	-24.547	-9.460	-64.582	-49.858	-14.724
±0.99	-21.219	-13.654	-7.565	-34.598	-24.071	-10.527	-67.172	-49.544	-17.628

defined as:

$$\text{Bias} = \frac{1}{S}\sum_{i=1}^{S}\left(\hat{\theta}_T^{(i)} - \theta\right),$$

$$\text{MAE} = \frac{1}{S}\sum_{i=1}^{S}\left|\hat{\theta}_T^{(i)} - \theta\right|,$$

$$\text{MSE} = \frac{1}{S}\sum_{i=1}^{S}\left(\hat{\theta}_T^{(i)} - \theta\right)^2,$$

where $\hat{\theta}_T^{(i)}$ denotes the estimator $\hat{\theta}$ or $\tilde{\theta}_W$ in the i-th simulation. And θ is the true value of the parameter, which varies from $\{\pm 0.1, \pm 0.2, \cdots, \pm 0.9, \pm 0.91, \cdots, \pm 0.99\}$, and sample size $T = \{30, 50, 100\}$. We repeat the experiment $S = 500$ times.

Table 3 reports the accuracy comparison results of the MLE and Whittle estimator with T=50, and the results of T=30 and 100 are reported at Supplementary Materials.

It shows that $\tilde{\theta}_W$ has very small advantage when the absolute value of true parameter θ is smaller than 0.5, which exhibits slightly smaller values of Bias, MAE and MSE. However, the MLE $\hat{\theta}$ performs much better than $\tilde{\theta}_W$ when the absolute θ is larger than 0.7, which presents much smaller values of Bias, MAE and MSE. Especially $\tilde{\theta}_W$ performs very poorly when the absolute θ is larger than 0.9. This indicates that when the parameter of MA(1) nears the boundary (±1), the difference between MLE and Whittle

TABLE 3
Accuracy performance of MLE and Whittle likelihood estimator for T =50. The better performances with smaller absolute value of Bias, smaller MAE and MSE are marked in bold.

	Estimate		Bias		MAE		MSE	
θ	$\hat{\theta}$	$\tilde{\theta}_W$	$\hat{\theta}$	$\tilde{\theta}_W$	$\hat{\theta}$	$\tilde{\theta}_W$	$\hat{\theta}$	$\tilde{\theta}_W$
0.1	0.128	0.126	0.028	**0.026**	0.101	**0.099**	**0.017**	0.017
0.2	0.222	0.217	0.022	**0.017**	0.116	**0.115**	0.020	**0.020**
0.3	0.305	0.296	0.005	**-0.004**	0.113	**0.110**	0.020	**0.019**
0.4	0.415	0.403	0.015	**0.003**	0.120	**0.116**	0.023	**0.022**
0.5	0.513	0.496	0.013	**-0.004**	0.116	**0.111**	0.022	**0.020**
0.6	0.608	0.587	**0.008**	-0.013	**0.101**	0.102	0.017	**0.016**
0.7	0.709	0.679	**0.009**	-0.021	**0.094**	0.100	**0.016**	0.017
0.8	0.824	0.769	**0.024**	-0.031	**0.086**	0.096	**0.011**	0.016
0.9	0.914	0.831	**0.014**	-0.070	**0.062**	0.099	**0.005**	0.019
0.91	0.919	0.849	**0.009**	-0.061	**0.061**	0.088	**0.005**	0.012
0.92	0.929	0.840	**0.009**	-0.080	**0.059**	0.099	**0.005**	0.015
0.93	0.936	0.852	**0.006**	-0.078	**0.055**	0.098	**0.005**	0.015
0.94	0.938	0.845	**-0.002**	-0.095	**0.054**	0.105	**0.005**	0.017
0.95	0.945	0.848	**-0.006**	-0.102	**0.049**	0.110	**0.004**	0.019
0.96	0.954	0.851	**-0.006**	-0.109	**0.043**	0.116	**0.003**	0.021
0.97	0.962	0.854	**-0.008**	-0.116	**0.036**	0.119	**0.003**	0.022
0.98	0.961	0.861	**-0.019**	-0.119	**0.037**	0.121	**0.004**	0.023
0.99	0.969	0.855	**-0.021**	-0.135	**0.030**	0.135	**0.003**	0.027
-0.1	-0.099	-0.100	0.001	**0.000**	0.114	**0.114**	0.022	**0.022**
-0.2	-0.209	-0.204	-0.009	**-0.004**	0.119	**0.117**	0.023	**0.022**
-0.3	-0.306	-0.300	-0.006	**0.000**	**0.119**	0.119	**0.023**	0.024
-0.4	-0.411	-0.396	-0.011	**0.004**	0.121	**0.121**	**0.025**	0.025
-0.5	-0.509	-0.492	-0.009	**0.008**	0.111	**0.108**	0.021	**0.020**
-0.6	-0.616	-0.594	-0.016	**0.006**	0.101	**0.101**	0.017	**0.017**
-0.7	-0.717	-0.681	**-0.017**	0.019	**0.094**	0.096	**0.015**	0.015
-0.8	-0.816	-0.761	**-0.016**	0.039	**0.084**	0.106	**0.012**	0.019
-0.9	-0.918	-0.830	**-0.018**	0.070	**0.067**	0.102	**0.006**	0.019
-0.91	-0.927	-0.836	**-0.017**	0.074	**0.062**	0.100	**0.005**	0.018
-0.92	-0.928	-0.827	**-0.008**	0.093	**0.059**	0.113	**0.005**	0.022
-0.93	-0.938	-0.836	**-0.008**	0.094	**0.058**	0.114	**0.005**	0.024
-0.94	-0.947	-0.845	**-0.007**	0.095	**0.050**	0.110	**0.004**	0.022
-0.95	-0.945	-0.843	**0.005**	0.107	**0.049**	0.116	**0.005**	0.026
-0.96	-0.956	-0.843	**0.004**	0.117	**0.043**	0.124	**0.004**	0.027
-0.97	-0.958	-0.843	**0.012**	0.127	**0.040**	0.130	**0.004**	0.030
-0.98	-0.964	-0.846	**0.016**	0.134	**0.035**	0.135	**0.003**	0.030
-0.99	-0.970	-0.843	**0.020**	0.147	**0.029**	0.148	**0.004**	0.036

estimator becomes large, and the performance of Whittle becomes very poor compared with MLE. Again these findings indicate that it is a big warning to use the Whittle likelihood and estimator when MA(1) has $\theta \to \pm 1$. In fact, Taniguchi (1983) evaluated the expectation of MLE and Whittle likelihood in higher-order cases, and showed that the expected Whittle likelihood diverges to infinity as the true value of MA parameter goes to one, while it is not so for MLE method. This helps explain the poor performance of Whittle estimator when the process is close to the non-invertibility region.

5. Conclusion

This paper investigates the difference between the exact likelihood and Whittle likelihood for moving average process of order one. We elucidate the theoretical expressions of two likelihood functions and their expectations. We conduct numerical simulation to compare the finite sample performance of exact likelihood and Whittle likelihood in terms of actual likelihood function value, expectation, and parameter estimation accuracy, respectively. It shows that the exact likelihood and Whittle likelihood deliver similar performance in the likelihood value and parameter estimation accuracy when the true value of parameter is small (i.e. close to 0), while the difference of two likelihoods becomes large and Whittle estimator performs poorly when parameter nears the boundary (± 1). The results indicate a big warning for using the Whittle likelihood and estimator when moving average process is close to the non-invertibility region.

Acknowledgements

We thank the Editor, AE and anonymous referees for their valuable time and insightful comments, which have helped us to improve our study. Especially, one of them provided some nice information on explicit inverses of Toeplitz and associated matrices (Dow, 2002), which is of interest for the future study. The work was supported by JSPS Grant-in-Aid for Scientific Research (S) 18H05290.

Supplementary information

Supplementary materials available online include the simulation results of parameter estimation accuracy for MLE and Whittle estimator under sample size $T = 30$ and 100.

References

T. Anderson (1977). *Estimation for autoregressive moving average models in the time and frequency domains*. The Annals of Statistics, 5, no. 5, pp. 842–865.

T. W. Anderson (1971). *The statistical analysis of time series*, vol. 19. John Wiley & Sons.

M. DOW (2002). *Explicit inverses of toeplitz and associated matrices*. ANZIAM Journal, 44, pp. E185–E215.

Y. FUJIKOSHI, Y. OCHI (1984). *Asymptotic properties of the maximum likelihood estimate in the first order autoregressive process*. Annals of the Institute of Statistical Mathematics, 36, no. 1, pp. 119–128.

L. GIRAITIS, M. S. TAQQU (1999). *Whittle estimator for finite-variance non-gaussian time series with long memory*. The Annals of Statistics, 27, no. 1, pp. 178–203.

Y. HOSOYA, M. TANIGUCHI (1982). *A central limit theorem for stationary processes and the parameter estimation of linear processes*. The Annals of Statistics, pp. 132–153.

P. KETZ (2018). *Subvector inference when the true parameter vector may be near or at the boundary*. Journal of Econometrics, 207, no. 2, pp. 285–306.

A. C. MONTI, M. TANIGUCHI (2018). *Adjustments for a class of tests under nonstandard conditions*. Statistica Sinica, 28, no. 3, pp. 1437–1458.

M. TANIGUCHI (1983). *On the second order asymptotic efficiency of estimators of Gaussian ARMA processes*. The Annals of Statistics, pp. 157–169.

M. TANIGUCHI, Y. KAKIZAWA (2000). *Asymptotic theory of statistical inference for time series*. Springer Science & Business Media.

P. WHITTLE (1961). *Gaussian estimation in stationary time series*. Bull. Internat. Statist. Inst., 39, pp. 105–129.

SUMMARY

For Gaussian stationary processes, the likelihood functions include the inverse and determinant of the covariance matrices, and Whittle likelihood is considered as a standard technique to avoid expensive matrix determinant and inversions under large sample size. In this paper, we investigate the difference between the exact likelihood and Whittle likelihood with finite sample size for moving average processes of order one. We elucidate the theoretical expressions of two likelihood functions and their expectations and evaluate the performance between exact likelihood and Whittle likelihood numerically. We find that the exact likelihood and Whittle likelihood perform similarly when the true value of parameter is close to zero, while the difference becomes large and Whittle estimator performs poorly when absolute value of parameter gets close to one. This is an important warning when we use the Whittle likelihood and estimator if the parameter of moving average process nears the boundary of parameter space.

Keywords: Gaussian stationary process; Spectral density; Likelihood function; Whittle likelihood; Moving average process.