Journal of Data Science **10**(2012), 723-755

Estimation of Linear Regression Models with Serially Correlated Errors

Chiao-Yi Yang
National Central University and Academia Sinica

Abstract: This paper develops a generalized least squares (GLS) estimator in a linear regression model with serially correlated errors. In particular, the asymptotic optimality of the proposed estimator is established. To obtain this result, we use the modified Cholesky decomposition to estimate the inverse of the error covariance matrix based on the ordinary least squares (OLS) residuals. The resulting matrix estimator maintains positive definiteness and converges to the corresponding population matrix at a suitable rate. The outstanding finite sample performance of the proposed GLS estimator is illustrated using simulation studies and two real datasets.

Key words: Asymptotic optimality, generalized least squares estimator, modified Cholesky decomposition, positive definiteness, serially correlated error.

1. Introduction

The ordinary least squares (OLS) method is the best-known method for estimating unknown regression coefficients of a linear regression model. When errors in a linear regression model are assumed to be uncorrelated or have a known correlation structure, the OLS estimator of regression coefficients is consistent and optimal in the class of linear unbiased estimators. However, in practical applications, cases in which errors are serially correlated occur frequently. For example, the observed data of global average temperatures over the past century exhibit a steadily increasing trend plus serially correlated noise (Bloomfield, 1992); and in clinical trails or stock markets, data from each subject are usually serially correlated, especially when subjects are measured repeatedly over time (Chan and Choy, 2008). Therefore, we are led to consider the following linear regression model,

$$y_t = \boldsymbol{x}_t^\top \boldsymbol{\beta} + \varepsilon_t, \quad t = 1, \cdots, n, \tag{1}$$

where y_t is an observed variate, $\boldsymbol{x}_t^\top = (x_{t,1}, \cdots, x_{t,p})$ is a p-dimensional non-stochastic explanatory variable, $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)^\top$ is a vector of the parameters of interest, and ε_t is independent of $\boldsymbol{x}_t$ and is from the stationary infinite-order autoregressive (AR) process,

$$\varepsilon_t + a_1\varepsilon_{t-1} + a_2\varepsilon_{t-2} + \cdots = \eta_t, \tag{2}$$

where $\{\eta_t\}$ is a sequence of independent random disturbances with $E(\eta_t) = 0$ and $E\left(\eta_t^2\right) = \sigma^2 > 0$ for all t. The model coefficients in (2) are assumed to obey the restriction

$$\sum_{i=1}^{\infty} |a_i| < \infty. \tag{3}$$

The aim of this article is to develop an efficient estimator of $\boldsymbol{\beta}$ for such a model as (1).

Over the past half-century, the linear regression model with autocorrelated errors has received increasing attention. Due to a lack of knowledge of the error covariance matrix, $\boldsymbol{\Sigma}_n = E\left((\varepsilon_1, \cdots, \varepsilon_n)^\top(\varepsilon_1, \cdots, \varepsilon_n)\right) = [\gamma_{i-j}]_{i,j=1}^n$, numerous articles reported that the OLS estimator is at least as efficient asymptotically as the best linear unbiased estimator (BLUE, the generalized least squares (GLS) estimator). Grenander (1954) proved that the asymptotic efficiency of the OLS estimator compared to the BLUE is one under the assumptions that ε_t has a continuous and positive spectral density, and the regressors, $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n$, satisfy certain regularity conditions, called "Grenander conditions"; this is discussed in more detail in Section 3.1. The same results are also in Anderson (1994). Some articles examined the efficiency of the GLS estimator with an estimated $\boldsymbol{\Sigma}_n^{-1}$ (the estimated GLS estimator) relative to the OLS estimator. For example, Koreisha and Fang (2001) investigated the properties of the estimated GLS estimator by assuming that the errors obey a finite-order AR model; they further illustrated the finite sample efficiency of the estimated GLS estimator over the OLS estimator. Almost all studies, including those cited above, assumed that errors are observable. However, this is impossible in practical situations. Amemiya (1973) first studied the large sample properties of the GLS estimator with an estimated $\boldsymbol{\Sigma}_n^{-1}$ obtained utilizing the least squares residuals for a linear regression model with stationary and autoregressive-moving average errors. He assumed that the residuals follow a "fixed-order" AR model to calculate an estimate of $\boldsymbol{\Sigma}_n$, and then constructed the estimated GLS estimator. Although he proved the asymptotic normality of his GLS estimator and showed that the limiting distribution is equivalent to that of the BLUE, the assumption that the order of AR process has been kept fixed is unnatural. (see Berk, 1974; Ing and Wei, 2003 for more details). As mentioned above, the estimator of $\boldsymbol{\Sigma}_n^{-1}$ plays a significant role in

studies of the GLS estimator. Thus, finding a consistent estimator of $\mathbf{\Sigma}_n^{-1}$ is the primary task to achieve the goal of the present study.

Although covariance matrix estimates of multivariate data have been developed over several decades, only a few articles discussed this issue for time-series data. For stationary time series data, $\mathbf{\Sigma}_n$ has n distinct elements, $\gamma_k\,(=\gamma_{-k})$, $k = 0, \cdots, n-1$. Reliably estimating all of them based on only one realization is not possible. Wu and Pourahmadi (2009) showed that the sample covariance matrix of the stationary process, which is formed by plugging the sample autocovariance estimate, $\hat{\gamma}_k$, instead of γ_k into $\mathbf{\Sigma}_n$, is not a consistent estimator of $\mathbf{\Sigma}_n$ under the operator norm. This result is to be expected, since $\hat{\gamma}_k$ is unreliable when the lag, k, is large. Xiao and Wu (2011) further attested that under the operator norm the convergence rate of the sample covariance matrix is of order $\log n$. Based on the property that γ_k decreases with k, Wu and Pourahmadi (2009) extended the idea of banding the sample covariance matrix in the multivariate case (see Bickel and Levina, 2008) to the stationary time series case. They established operator norm consistency of the banded sample covariance matrix estimator of $\mathbf{\Sigma}_n$. Since the banded sample covariance matrix is not guaranteed to be positive definite, the estimator of $\mathbf{\Sigma}_n^{-1}$ cannot be obtained. The Schur product theorem in Horn and Johnson (1990, Theorem 7.5.3) says that for any positive definite matrix, the Schur product with a positive definite matrix preserves its positive definiteness. That motivated the development of a tapered covariance matrix, which is a Schur product of the sample covariance matrix and a positive definite matrix (Furrer and Bengtsson, 2007; McMurry and Politis, 2010; Wu, 2011). The Schur product $\boldsymbol{A} \circ \boldsymbol{B}$ is the matrix formed by the coordinate-wise multiplication of the elements of matrices $\boldsymbol{A}$ and $\boldsymbol{B}$.

The present study considers the GLS estimator with the estimated $\mathbf{\Sigma}_n^{-1}$ obtained by a modified Cholesky decomposition. The method for estimating $\mathbf{\Sigma}_n^{-1}$ is explained in Section 2.2. It circumvents the aforementioned problems, including the following two facts: errors are unobservable and only one realization is available for estimating $\mathbf{\Sigma}_n^{-1}$, and including one constraint that the positive definiteness of $\mathbf{\Sigma}_n$ must hold. This paper is organized as follows. The method for constructing the GLS estimator is explained in Section 2. The consistency of the estimated $\mathbf{\Sigma}_n^{-1}$ and the asymptotic efficiency of the estimated GLS estimator are proven in Section 3. In Section 4, simulation studies confirm the theoretical results, and two real datasets are used to present the outstanding performance of the proposed GLS estimator. Finally, we end the paper with a short concluding section. All technical proofs are given in the Appendix.

2. The GLS Estimator of Regression Coefficients

2.1 Model and Notation

Consider the following linear regression model

$$\boldsymbol{y}_n = \boldsymbol{X}_n^\top \boldsymbol{\beta} + \boldsymbol{\epsilon}_n, \tag{4}$$

where $\boldsymbol{y}_n = (y_n, y_{n-1}, \cdots, y_1)^\top$, $\boldsymbol{X}_n = (\boldsymbol{x}_n, \cdots, \boldsymbol{x}_1)^\top$, and $\boldsymbol{\epsilon}_n = (\varepsilon_n, \cdots, \varepsilon_1)^\top$. After observing $\boldsymbol{x}_1, y_1, \boldsymbol{x}_2, y_2, \cdots, \boldsymbol{x}_n, y_n$, $\boldsymbol{\beta}$ can be estimated. The BLUE of $\boldsymbol{\beta}$,

$$\left(\boldsymbol{X}_n^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{X}_n\right)^{-1} \boldsymbol{X}_n^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{y}_n, \tag{5}$$

which is denoted by $\hat{\boldsymbol{\beta}}_{\mathrm{GLS}}$, is studied. However, due to unknown $\boldsymbol{\Sigma}_n^{-1}$, $\hat{\boldsymbol{\beta}}_{\mathrm{GLS}}$ could not be determined. To estimate $\boldsymbol{\Sigma}_n^{-1}$, the unobservable errors are first estimated by the least squares residuals, namely,

$$\hat{\varepsilon}_t = y_t - \boldsymbol{x}_t^\top \hat{\boldsymbol{\beta}}_{\mathrm{OLS}}, \quad t = 1, \cdots, n, \tag{6}$$

where $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}} = (\boldsymbol{X}_n^\top \boldsymbol{X}_n)^{-1} \boldsymbol{X}_n^\top \boldsymbol{y}_n$ is the OLS estimator of $\boldsymbol{\beta}$. The method for estimating $\boldsymbol{\Sigma}_n^{-1}$ is described in the next subsection.

2.2 An Estimator of the $\boldsymbol{\Sigma}_n^{-1}$

To avoid loss of positive definiteness of the estimated $\boldsymbol{\Sigma}_n$, a modified Cholesky decomposition of $\boldsymbol{\Sigma}_n^{-1}$ (Wu and Pourahmadi, 2003) is used to develop an estimator for $\boldsymbol{\Sigma}_n^{-1}$. The key idea is to predict ε_t based on its predecessors. More specifically, $\boldsymbol{\epsilon}_n$ is subjected to the transformation,

$$\boldsymbol{A}_n \boldsymbol{\epsilon}_n = (\eta_{n,n-1}, \eta_{n-1,n-2}, \cdots, \eta_{2,1}, \varepsilon_1)^\top, \tag{7}$$

where

$$\boldsymbol{A}_n = \begin{pmatrix} 1 & a_{n-1,1} & a_{n-1,2} & \cdots & a_{n-1,n-1} \\ 0 & 1 & a_{n-2,1} & \cdots & a_{n-2,n-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & a_{1,1} \\ 0 & \cdots & & 0 & 1 \end{pmatrix},$$

and

$$\eta_{t+1,j} = \varepsilon_{t+1} + \boldsymbol{a}^\top(j) \boldsymbol{\epsilon}_t(j), \quad j = 1, \cdots, t, \ t = 1, \cdots, n-1,$$

with $\boldsymbol{\epsilon}_t(j) = (\varepsilon_t, \varepsilon_{t-1}, \cdots, \varepsilon_{t-j+1})^\top$ and

$$\boldsymbol{a}^\top(j) = (a_{j,1}, \cdots, a_{j,j}) = \arg\min_{\boldsymbol{c} \in \mathcal{R}^j} E\left(\varepsilon_{t+1} + \boldsymbol{c}^\top \boldsymbol{\epsilon}_t(j)\right)^2.$$

The term $\eta_{t+1,\,j}$ is the least squares residual obtained by regressing ε_{t+1} on the past value, $\boldsymbol{\epsilon}_t(j)$. Thus, elements on the right-hand side of (7) are pairwise

uncorrelated. Then, the autocovariance matrix of $\boldsymbol{A}_n\boldsymbol{\epsilon}_n$ is $\boldsymbol{A}_n\boldsymbol{\Sigma}_n\boldsymbol{A}_n^\top = \boldsymbol{D}_n$, where

$$\boldsymbol{D}_n = \mathrm{Diag}\left(E\left(\eta_{n,\,n-1}^2\right),\, E\left(\eta_{n-1,\,n-2}^2\right),\, \cdots,\, E\left(\eta_{2,\,1}^2\right),\, E(\varepsilon_1^2)\right).$$

The notation $\mathrm{Diag}(\boldsymbol{d})$, where $\boldsymbol{d}$ is a vector, stands for the diagonal matrix, whose diagonal elements are the corresponding components of $\boldsymbol{d}$. Thus,

$$\boldsymbol{\Sigma}_n^{-1} = \boldsymbol{A}_n^\top \boldsymbol{D}_n^{-1} \boldsymbol{A}_n, \tag{8}$$

which is called the modified Cholesky decomposition of $\boldsymbol{\Sigma}_n^{-1}$. Define k_n as a prescribed positive integer satisfying that as $n \to \infty$,

$$k_n \to \infty \ \text{ and } \ k_n^3/n \to 0. \tag{9}$$

Due to the stationarity of the error process (2) and the parameters in (2) satisfying (3), Lemma 4 of Berk (1974) yields that for all $m > k_n$,

$$\sum_{i=1}^{k_n} |a_{k_n,i} - a_{m,i}| \le \sum_{i=1}^{k_n} |a_{k_n,i} - a_i| + \sum_{i=1}^{k_n} |a_{m,i} - a_i| \to 0 \ \text{ as } k_n \to \infty,$$

and

$$\sum_{i=k_n}^{m} |a_{m,i}| \le \sum_{i=k_n}^{m} |a_{m,i} - a_i| + \sum_{i=k_n}^{m} |a_i| \to 0 \ \text{ as } k_n \to \infty$$

hold. These results motivate us to further approximate matrices $\boldsymbol{A}_n$ and $\boldsymbol{D}_n$ by

$$\boldsymbol{A}_n(k_n) = \begin{pmatrix}
1 & a_{k_n,1} & \cdots & & a_{k_n,k_n} & 0 & & \cdots & & 0 \\
 & \ddots & \ddots & & & \ddots & & & & \\
\vdots & & 1 & a_{k_n,1} & \cdots & & a_{k_n,k_n} & 0 & & \vdots \\
 & & & \ddots & \ddots & & & \ddots & & \\
0 & & & & 1 & a_{k_n,1} & a_{k_n,2} & \cdots & a_{k_n,k_n} & 0 \\
 & & & & & 1 & a_{k_n,1} & \cdots & & a_{k_n,k_n} \\
\vdots & & & & & & 1 & a_{k_n-1,1} & \cdots & a_{k_n-1,k_n-1} \\
 & & & & & & & \ddots & \ddots & \vdots \\
0 & & & & \cdots & & & 0 & 1 & a_{1,1} \\
0 & & & & \cdots & & & & 0 & 1
\end{pmatrix},$$

and

$$\boldsymbol{D}_n(k_n) = \mathrm{Diag}\left(\underbrace{E\left(\eta_{k_n+1,k_n}^2\right),\, \cdots,\, E\left(\eta_{k_n+1,k_n}^2\right)}_{n-k_n},\, \cdots,\, E\left(\eta_{2,\,1}^2\right),\, E(\varepsilon_1^2)\right).$$

The new matrices $\boldsymbol{A}_n(k_n)$ and $\boldsymbol{D}_n(k_n)$ reduce the number of unknown parameters, since now only $k_n(k_n+1)/2$ parameters need to be estimated. Matrices $\boldsymbol{A}_n(k_n)$ and $\boldsymbol{D}_n(k_n)$ are easy to estimate. In view of (8), the Cholesky-based estimator of $\boldsymbol{\Sigma}_n^{-1}$ is defined as

$$\widehat{\boldsymbol{\Sigma}}_n^{-1}(k_n,\, \hat{\boldsymbol{\epsilon}}_n) = \widehat{\boldsymbol{A}}_n^{\top}(k_n,\, \hat{\boldsymbol{\epsilon}}_n)\widehat{\boldsymbol{D}}_n^{-1}(k_n,\, \hat{\boldsymbol{\epsilon}}_n)\widehat{\boldsymbol{A}}_n(k_n,\, \hat{\boldsymbol{\epsilon}})_n, \tag{10}$$

where $\hat{\boldsymbol{\epsilon}}_n = (\hat{\varepsilon}_n, \cdots, \hat{\varepsilon}_1)^{\top}$, $\widehat{\boldsymbol{A}}_n(k_n,\, \hat{\boldsymbol{\epsilon}}_n) =$

$$\begin{pmatrix}
1 & \hat{a}_{k_n,1}(\hat{\boldsymbol{\epsilon}}_n) & \cdots & \hat{a}_{k_n,k_n}(\hat{\boldsymbol{\epsilon}}_n) & 0 & \cdots & & 0 \\
0 & 1 & \hat{a}_{k_n,1}(\hat{\boldsymbol{\epsilon}}_n) & \cdots & \hat{a}_{k_n,k_n}(\hat{\boldsymbol{\epsilon}}_n) & 0 & \cdots & 0 \\
\vdots & & \ddots & \ddots & & & & \vdots \\
0 & \cdots & 0 & 1 & \hat{a}_{k_n,1}(\hat{\boldsymbol{\epsilon}}_n) & & \cdots & \hat{a}_{k_n,k_n}(\hat{\boldsymbol{\epsilon}}_n) \\
0 & \cdots & & 0 & 1 & \hat{a}_{k_n-1,1}(\hat{\boldsymbol{\epsilon}}_n) & \cdots & \hat{a}_{k_n-1,k_n-1}(\hat{\boldsymbol{\epsilon}}_n) \\
\vdots & & & & & \ddots & \ddots & \vdots \\
0 & & & \cdots & & & 1 & \hat{a}_{1,1}(\hat{\boldsymbol{\epsilon}}_n) \\
0 & & & \cdots & & & 0 & 1
\end{pmatrix},$$

and $\widehat{\boldsymbol{D}}_n(k_n,\, \hat{\boldsymbol{\epsilon}}_n) =$

$$\frac{1}{N_n}\mathrm{Diag}\left(\underbrace{\sum_{t=k_n}^{n-1}\hat{\eta}_{t+1,k_n}^2(\hat{\boldsymbol{\epsilon}}_n),\, \cdots,\, \sum_{t=k_n}^{n-1}\hat{\eta}_{t+1,k_n}^2(\hat{\boldsymbol{\epsilon}}_n)}_{n-k_n},\, \cdots,\, \sum_{t=k_n}^{n-1}\hat{\eta}_{t+1,1}^2(\hat{\boldsymbol{\epsilon}}_n),\, \sum_{t=k_n}^{n-1}\hat{\varepsilon}_t^2\right),$$

with $N_n = n - k_n$, $\hat{\eta}_{t+1,j}(\hat{\boldsymbol{\epsilon}}_n) = \hat{\varepsilon}_{t+1} + \hat{\boldsymbol{a}}^{\top}(j,\, \hat{\boldsymbol{\epsilon}}_n)\hat{\boldsymbol{\epsilon}}_t(j)$, $\hat{\boldsymbol{\epsilon}}_t(j) = (\hat{\varepsilon}_t,\, \hat{\varepsilon}_{t-1} \cdots,\, \hat{\varepsilon}_{t-j+1})^{\top}$, and

$$\begin{aligned}
\hat{\boldsymbol{a}}(j,\, \hat{\boldsymbol{\epsilon}}_n) &= (\hat{a}_{j,1}(\hat{\boldsymbol{\epsilon}}_n),\, \cdots,\, \hat{a}_{j,j}(\hat{\boldsymbol{\epsilon}}_n))^{\top} \\
&= -\left(\sum_{t=k_n}^{n-1}\hat{\boldsymbol{\epsilon}}_t(j)\hat{\boldsymbol{\epsilon}}_t^{\top}(j)\right)^{-1}\sum_{t=k_n}^{n-1}\hat{\boldsymbol{\epsilon}}_t(j)\hat{\varepsilon}_{t+1},
\end{aligned}$$

for $j = 1, \cdots, k_n$. The existence of $(N_n^{-1}\sum_{t=k_n}^{n-1}\hat{\boldsymbol{\epsilon}}_t(k_n)\hat{\boldsymbol{\epsilon}}_t^{\top}(k_n))^{-1}$ is shown in Lemma 1. Based on the estimator of $\boldsymbol{\Sigma}_n^{-1}$ defined in (10), the GLS estimator of $\boldsymbol{\beta}$, called the "Cholesky-based generalized least squares (CGLS) estimator", is obtained by

$$\widehat{\boldsymbol{\beta}}_{\mathrm{CGLS}}(k_n,\, \hat{\boldsymbol{\epsilon}}_n) = \left(\boldsymbol{X}_n^{\top}\widehat{\boldsymbol{\Sigma}}_n^{-1}(k_n,\, \hat{\boldsymbol{\epsilon}}_n)\boldsymbol{X}_n\right)^{-1}\boldsymbol{X}_n^{\top}\widehat{\boldsymbol{\Sigma}}_n^{-1}(k_n,\, \hat{\boldsymbol{\epsilon}}_n)\boldsymbol{y}_n. \tag{11}$$

3. Assumptions and Some Asymptotic Properties

3.1 Technical Conditions

For the sake of establishing consistency of $\widehat{\boldsymbol{\Sigma}}_n^{-1}(k_n, \hat{\boldsymbol{\epsilon}}_n)$ and asymptotic optimality of $\widehat{\boldsymbol{\beta}}_{\mathrm{CGLS}}(k_n, \hat{\boldsymbol{\epsilon}}_n)$, the design matrix, $\boldsymbol{X}_n$, and errors, $\{\varepsilon_t\}$, are assumed to satisfy the following conditions. Let $s_{i,j}^{(n)}(h) = \sum_{t=1-h}^{n-h} x_{t,i}x_{t+h,j}$, $i, j = 1, \cdots, p$, $h = 0, \pm 1, \pm 2 \cdots$, and note that $x_{t,i} = 0$, if $t \notin \{1, \cdots, n\}$. This study assumes $\lim_{n\to\infty} s_{i,i}^{(n)}(0) = \infty$, $i = 1, \cdots, p$, and the following conditions.

(C1) Let $\boldsymbol{X}_n$ meet the following Grenander's conditions:

$$\lim_{n\to\infty} \frac{x_{n+h,i}^2}{s_{i,i}^{(n)}(0)} = 0,$$

for $i = 1, \cdots, p$, and any positive integer h. As $n \to \infty$, the limit of

$$\frac{s_{i,j}^{(n)}(h)}{\sqrt{s_{i,i}^{(n)}(0)s_{j,j}^{(n)}(0)}}$$

is assumed to exist, and is denoted by $\rho_{i,j}(h)$ for all i, j, and h. Define $\boldsymbol{\Lambda}(h) = [\rho_{i,j}(h)]_{i,j=1}^p$, $h = 0, \pm 1, \pm 2, \cdots$. $\{\boldsymbol{\Lambda}(h)\}$ is a sequence of positive definite matrices. It follows that

$$\boldsymbol{\Lambda}(h) = \int_{-\pi}^{\pi} e^{-ih\lambda} d\boldsymbol{M}(\lambda),$$

where $\boldsymbol{M}(\lambda)$ is a nondecreasing matrix-valued function.

(C2) The errors, $\varepsilon_1, \cdots, \varepsilon_n$, are from process (2) with coefficients satisfying (3). Moreover, the roots of the characteristic polynomial, $A(z) = 1 + a_1 z + a_2 z^2 + \cdots$, of the model (2) lie outside the unit circle.

(C3) Disturbances of the error model, $\{\eta_t\}$, satisfy $\sup_{-\infty<t<\infty} E\left(|\eta_t|^4\right) < \infty$.

Remark 1. Grenander's conditions are common and can refer to Grenander and Rosenblatt (1957), Amemiya (1973), and Anderson (1994), among others. Define $\boldsymbol{S}_n = \sum_{t=1}^n \boldsymbol{x}_t \boldsymbol{x}_t^\top$. Using (C1),

$$\frac{1}{n}\boldsymbol{S}_n \longrightarrow \boldsymbol{S} \text{ as } n \to \infty, \tag{12}$$

where $\boldsymbol{S}$ is a positive definite matrix.

Remark 2. Based on Theorem 3.8.4 of Brillinger (2001), (C2) implies that $\varepsilon_t = \sum_{j=0}^{\infty} b_j \eta_{t-j}$, where $b_0 = 1$ and $\{b_j\}$ satisfies $\sum_{j=0}^{\infty} |b_j| < \infty$, and $B(z) = \sum_{j=0}^{\infty} b_j z^j \neq 0$ for all $|z| \leq 1$. These specifications yield the spectral density of ε_t,

$$f_{\varepsilon}(\lambda) = \frac{\sigma^2}{2\pi} \left| \sum_{j=0}^{\infty} b_j e^{-ij\lambda} \right|^2, \tag{13}$$

which is continuous on $[-\pi, \pi]$, and $f_{\min}$ and $f_{\max}$ exist such that

$$0 < f_{\min} \leq f_{\varepsilon}(\lambda) \leq f_{\max} < \infty, \quad -\pi \leq \lambda \leq \pi. \tag{14}$$

Additionally, using (14) yields

$$\sup_{k \geq 1} \|\boldsymbol{\Sigma}_k\| \leq 2\pi f_{\max} \quad \text{and} \quad \sup_{k \geq 1} \|\boldsymbol{\Sigma}_k^{-1}\| \leq (2\pi f_{\min})^{-1}, \tag{15}$$

where $\boldsymbol{\Sigma}_k = E\left((\varepsilon_1, \cdots, \varepsilon_k)^{\top}(\varepsilon_1, \cdots, \varepsilon_k)\right)$ is a k-dimensional covariance matrix.

3.2 The Consistency of $\widehat{\boldsymbol{\Sigma}}_n^{-1}(k_n, \hat{\boldsymbol{\epsilon}}_n)$

In presenting the main result of this section, some auxiliary lemmas must be introduced in advance. In the sequel, C denotes a generic positive constant whose value may vary from place to place but is independent of n and k_n. For any $p \times m$ matrix, $\boldsymbol{Z}$, with real entries, $\{z_{i,j}\}$, its operator norm is defined by $\|\boldsymbol{Z}\| = \sup_{\|\boldsymbol{w}\|=1} \|\boldsymbol{Z}\boldsymbol{w}\|$, where $\|\boldsymbol{w}\|$ is the Euclidean norm of the real vector $\boldsymbol{w} = (w_1, \cdots, w_m)^{\top}$. $\|\boldsymbol{Z}\|_1 = \max_{1 \leqslant j \leqslant m} \sum_{i=1}^{p} |z_{i,j}|$, $\|\boldsymbol{Z}\|_{\infty} = \max_{1 \leqslant i \leqslant p} \sum_{j=1}^{m} |z_{i,j}|$, and $\|\boldsymbol{w}\|_1 = \sum_{i=1}^{m} |w_i|$. Moreover, $\widehat{\boldsymbol{R}}(k_n) = N_n^{-1} \sum_{t=k_n}^{n-1} \boldsymbol{\epsilon}_t(k_n)\boldsymbol{\epsilon}_t^{\top}(k_n) = [\hat{\gamma}_{i,j}]_{i,j=1}^{k_n}$, and $\widehat{\boldsymbol{R}}(k_n, \hat{\boldsymbol{\epsilon}}_n) = N_n^{-1} \sum_{t=k_n}^{n-1} \hat{\boldsymbol{\epsilon}}_t(k_n)\hat{\boldsymbol{\epsilon}}_t^{\top}(k_n)$.

Lemma 1. Assume $\{\boldsymbol{x}_t, y_t\}$ are generated from (1) and conditions (C1)-(C3) hold. Define $O_{1,n} = \{\lambda_{\min}(\widehat{\boldsymbol{R}}(k_n)) < \pi f_{\min}\}$ and $O_{2,n} = \{\lambda_{\min}(\widehat{\boldsymbol{R}}(k_n, \hat{\boldsymbol{\epsilon}}_n) < \pi f_{\min}/2\}$, where k_n satisfies (9), $\lambda_{\min}(\boldsymbol{Z})$ denotes the minimum eigenvalue of the square matrix $\boldsymbol{Z}$, and $O_1 \bigcup O_2$ denotes the union of events O_1 and O_2. Then

$$P(O_{1,n}) = o(1), \tag{16}$$

and

$$P(O_{1,n} \bigcup O_{2,n}) = o(1). \tag{17}$$

Moreover,

$$\max_{1 \leq l \leq k_n} \|\widehat{\boldsymbol{R}}^{-1}(l)\|_1 I_{O_{1,n}^c} = O_p(1), \tag{18}$$

$$\max_{1 \leq l \leq k_n} \|\widehat{\boldsymbol{R}}^{-1}(l) - \widehat{\boldsymbol{R}}^{-1}(l, \hat{\boldsymbol{\epsilon}}_n)\|_1 I_{(O_{1,n} \bigcup O_{2,n})^c} = O_p\left(\frac{k_n}{N_n^{3/4}}\right), \tag{19}$$

and

$$\max_{1\le l\le k_n} \|\widehat{\boldsymbol{R}}^{-1}(l) - \widehat{\boldsymbol{R}}^{-1}(l, \hat{\boldsymbol{\epsilon}}_n)\| I_{(O_{1,n}\bigcup O_{2,n})^c} = O_p\left(\frac{k_n}{N_n}\right), \tag{20}$$

where I_{O^c} stands for the indicator function of event O^c, the complement of event O.

Lemma 2. Assume that (C2) and (C3) hold. Then,

$$\left\| \frac{1}{N_n} \sum_{t=k_n}^{n-1} \boldsymbol{\epsilon}_t(k_n)\varepsilon_{t+1} \right\|_1 = O_p(k_n), \tag{21}$$

and

$$\left\| \frac{1}{N_n} \sum_{t=k_n}^{n-1} \boldsymbol{\epsilon}_t(k_n)\varepsilon_{t+1} \right\| = O_p(k_n^{1/2}). \tag{22}$$

Lemma 3. Assume that the assumptions of Lemma 2 hold. Suppose $k_n = o(n^{1/3})$. Then,

$$\max_{1\le l\le k_n} \left\| \frac{1}{N_n^{5/6}} \sum_{t=k_n}^{n-1} \boldsymbol{\epsilon}_t(l)\eta_{t+1,l} \right\| = O_p(1), \tag{23}$$

and

$$\max_{1\le l\le k_n} \left\| \frac{1}{N_n^{5/6}} \sum_{t=k_n}^{n-1} \boldsymbol{\epsilon}_t(l)\eta_{t+1,l} \right\|_1 = O_p(k_n^{1/2}). \tag{24}$$

Lemma 4. Assume that the assumptions of Lemma 1 hold. Then,

$$\left\| \sum_{t=k_n}^{n-1} \left(\hat{\boldsymbol{\epsilon}}_t(k_n)\hat{\varepsilon}_{t+1} - \boldsymbol{\epsilon}_t(k_n)\varepsilon_{t+1}\right) \right\|_1 = O_p\left(k_n\right), \tag{25}$$

and

$$\left\| \sum_{t=k_n}^{n-1} \left(\hat{\boldsymbol{\epsilon}}_t(k_n)\hat{\varepsilon}_{t+1} - \boldsymbol{\epsilon}_t(k_n)\varepsilon_{t+1}\right) \right\| = O_p\left(k_n^{1/2}\right). \tag{26}$$

The proofs of Lemmas 1-4 are provided in the Appendix.

Theorem 1. Given the same assumptions as those of Lemma 1, then,

$$\|\widehat{\boldsymbol{\Sigma}}_n^{-1}(k_n, \hat{\boldsymbol{\epsilon}}_n) - \boldsymbol{\Sigma}_n^{-1}\| = O_p\left(\left(\frac{k_n}{n^{1/3}}\right)^{1/2} + \sum_{i>k_n} |a_i|\right). \tag{27}$$

Proof. Define $\widehat{\boldsymbol{H}}_n(k_n, \hat{\boldsymbol{\epsilon}}_n) = \widehat{\boldsymbol{A}}_n(k_n, \hat{\boldsymbol{\epsilon}}_n) - \boldsymbol{A}_n$ and $\widehat{\boldsymbol{G}}_n(k_n, \hat{\boldsymbol{\epsilon}}_n) = \widehat{\boldsymbol{D}}_n^{-1}(k_n, \hat{\boldsymbol{\epsilon}}_n) -$

$\boldsymbol{D}_n^{-1}$. Then, the following holds true:

$$\begin{aligned}
&\widehat{\boldsymbol{\Sigma}}_n^{-1}(k_n, \hat{\boldsymbol{\epsilon}}_n) - \boldsymbol{\Sigma}_n^{-1} \\
&= \widehat{\boldsymbol{H}}_n^{\top}(k_n, \hat{\boldsymbol{\epsilon}}_n)\boldsymbol{D}_n^{-1}\boldsymbol{A}_n + \boldsymbol{A}_n^{\top}\widehat{\boldsymbol{G}}_n(k_n, \hat{\boldsymbol{\epsilon}}_n)\boldsymbol{A}_n + \boldsymbol{A}_n^{\top}\boldsymbol{D}_n^{-1}\widehat{\boldsymbol{H}}_n(k_n, \hat{\boldsymbol{\epsilon}}_n) \\
&\quad + \widehat{\boldsymbol{H}}_n^{\top}(k_n, \hat{\boldsymbol{\epsilon}}_n)\widehat{\boldsymbol{G}}_n(k_n, \hat{\boldsymbol{\epsilon}}_n)\boldsymbol{A}_n + \widehat{\boldsymbol{H}}_n^{\top}(k_n, \hat{\boldsymbol{\epsilon}}_n)\boldsymbol{D}_n^{-1}\widehat{\boldsymbol{H}}_n(k_n, \hat{\boldsymbol{\epsilon}}_n) \\
&\quad + \boldsymbol{A}_n^{\top}\widehat{\boldsymbol{G}}_n(k_n, \hat{\boldsymbol{\epsilon}}_n)\widehat{\boldsymbol{H}}_n(k_n, \hat{\boldsymbol{\epsilon}}_n) + \widehat{\boldsymbol{H}}_n^{\top}(k_n, \hat{\boldsymbol{\epsilon}}_n)\widehat{\boldsymbol{G}}_n(k_n, \hat{\boldsymbol{\epsilon}}_n)\widehat{\boldsymbol{H}}_n(k_n, \hat{\boldsymbol{\epsilon}}_n).
\end{aligned}$$

Since $E(\varepsilon_1^2) \geq \sigma^2$ and $\inf_{t\geq 1} E(\eta_{t+1,t}^2) \geq \sigma^2$, then

$$\|\boldsymbol{D}_n^{-1}\| \leq 1/\sigma^2.$$

Condition (3), together with Lemma 4 of Berk (1974), yields that

$$\|\boldsymbol{A}_n\|_1 < \infty \ \text{ and } \ \|\boldsymbol{A}_n\|_\infty < \infty, \tag{28}$$

and hence $\|\boldsymbol{A}_n\| < \infty$ holds.

To investigate the spectral norm of $\widehat{\boldsymbol{H}}_n(k_n, \hat{\boldsymbol{\epsilon}}_n)$, the following is defined first:

$$\boldsymbol{H}_n^{(1)}(k_n) = \begin{pmatrix}
0 & \cdots & 0 & -a_{n-1,k_n+1} & \cdots & & -a_{n-1,n-1} \\
 & & & 0 & -a_{n-2,k_n+1} & \cdots & -a_{n-2,n-2} \\
 & & & & \ddots & \ddots & \vdots \\
\vdots & & & & & 0 & -a_{k_n+1,k_n+1} \\
 & & & & & & 0 \\
 & & & & & & \vdots \\
0 & & & \cdots & & & 0
\end{pmatrix},$$

and $\boldsymbol{H}_n^{(2)}(k_n, \hat{\boldsymbol{\epsilon}}_n) = \widehat{\boldsymbol{H}}_n(k_n, \hat{\boldsymbol{\epsilon}}_n) - \boldsymbol{H}_n^{(1)}(k_n)$. It follows from Lemma 4 of Berk (1974) that $\|\boldsymbol{H}_n^{(1)}(k_n)\|^2 \leq \|\boldsymbol{H}_n^{(1)}(k_n)\|_1\|\boldsymbol{H}_n^{(1)}(k_n)\|_\infty < C(\sum_{i>k_n}^{\infty}|a_i|)^2$ holds. Again applying Lemma 4 of Berk (1974) and the same algebraic manipulations used in dealing with $\boldsymbol{H}_n^{(1)}(k_n)$ to $\boldsymbol{H}_n^{(2)}(k_n, \hat{\boldsymbol{\epsilon}}_n)$, the following is generated:

$$\boldsymbol{H}_n^{(2)}(k_n, \hat{\boldsymbol{\epsilon}}_n) \leq C\left(\max_{1\leq l\leq k_n}\|\hat{\boldsymbol{a}}(l, \hat{\boldsymbol{\epsilon}}_n) - \boldsymbol{a}(l)\|_1 + \sum_{i>k_n}^{\infty}|a_i|\right).$$

This paper shows in the Appendix that

$$\max_{1\leq l\leq k_n}\|\hat{\boldsymbol{a}}(l, \hat{\boldsymbol{\epsilon}}_n) - \boldsymbol{a}(l)\|_1 = O_p\left(\left(\frac{k_n}{n^{1/3}}\right)^{1/2}\right), \tag{29}$$

through which the following is obtained:

$$\|\widehat{\boldsymbol{H}}_n(k_n,\,\hat{\boldsymbol{\epsilon}}_n)\| = O_p\left(\left(\frac{k_n}{n^{1/3}}\right)^{1/2} + \sum_{i>k_n}^{\infty}|a_i|\right).$$

Next, as is also shown in the Appendix,

$$\|\widehat{\boldsymbol{G}}_n(k_n,\,\hat{\boldsymbol{\epsilon}}_n)\| = O_p\left(n^{-1/3} + \sum_{i>k_n}^{\infty}|a_i|\right). \tag{30}$$

Hence, the proof is complete.

□

3.3 The Asymptotic Efficiency of the CGLS Estimator

The following theorem proves the asymptotic normality of the CGLS estimator.

Theorem 2. Define $\boldsymbol{F}_n = (\mathrm{Diag}\,(\boldsymbol{s}))^{1/2}$ and assume that the assumptions of Theorem 1 hold. Then, as $n \to \infty$,

$$\boldsymbol{F}_n(\widehat{\boldsymbol{\beta}}_{\mathrm{CGLS}}(k_n,\,\hat{\boldsymbol{\epsilon}}_n) - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}\left(\mathbf{0},\,\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{f_\varepsilon(\lambda)}d\boldsymbol{M}(\lambda)\right)^{-1}\right), \tag{31}$$

where $\boldsymbol{s}$ is the main diagonal of matrix $\boldsymbol{S}_n$ defined before (12), $\xrightarrow{d}$ denotes convergence in distribution, $f_\varepsilon(\lambda)$ is defined by (13), and $\boldsymbol{M}(\lambda)$ is defined in condition (C1) of Section 3.1.

Proof. First,

$$\widehat{\boldsymbol{\beta}}_{\mathrm{CGLS}}(k_n,\,\hat{\boldsymbol{\epsilon}}_n) - \boldsymbol{\beta} = (\widehat{\boldsymbol{\beta}}_{\mathrm{CGLS}}(k_n,\,\hat{\boldsymbol{\epsilon}}_n) - \widehat{\boldsymbol{\beta}}_{\mathrm{GLS}}) + (\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}} - \boldsymbol{\beta}), \tag{32}$$

where $\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}}$ is defined by (5). It is shown in the Appendix that

$$\|\boldsymbol{F}_n(\widehat{\boldsymbol{\beta}}_{\mathrm{CGLS}}(k_n,\,\hat{\boldsymbol{\epsilon}}_n) - \widehat{\boldsymbol{\beta}}_{\mathrm{GLS}})\| = o_p(1), \tag{33}$$

and

$$\boldsymbol{F}_n(\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}\left(\mathbf{0},\,\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{f_\varepsilon(\lambda)}d\boldsymbol{M}(\lambda)\right)^{-1}\right). \tag{34}$$

Then, Slutsky's theorem can be applied to complete the proof.

□

Note that (34) first appeared in Amemiya (1973). Here, it is shown in a more straightforward manner. Equations (31) and (34) indicate that the CGLS estimator, $\widehat{\boldsymbol{\beta}}_{\mathrm{CGLS}}(k_n, \hat{\boldsymbol{\epsilon}}_n)$, has the same asymptotic distribution as the BLUE, $\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}}$. In conclusion, the CGLS is an asymptotically efficient (optimal) estimator. In the next section, the finite sample performances of $\widehat{\boldsymbol{\Sigma}}_n^{-1}(k_n, \hat{\boldsymbol{\epsilon}}_n)$ and $\widehat{\boldsymbol{\beta}}_{\mathrm{CGLS}}(k_n, \hat{\boldsymbol{\epsilon}}_n)$ are evaluated using some simulated data, and application of the proposed method to two actual datasets is also illustrated.

4. Simulation and Data Analysis

4.1 Finite Sample Performance of $\widehat{\boldsymbol{\Sigma}}_n^{-1}(k_n, \hat{\boldsymbol{\epsilon}}_n)$

In this section, the finite sample performance of the Cholesky-based estimator of $\boldsymbol{\Sigma}_n^{-1}$, given in (10), is compared to those of three other estimators. The estimation error of each estimator relative to the target matrix, $\boldsymbol{\Sigma}_n^{-1}$, is calculated by averaging over 1000 replications. The relative estimation error of the Cholesky-based matrix estimator is defined as

$$r_{\mathrm{C},n} = \frac{1}{1000}\sum_{t=1}^{1000}\frac{\|\widehat{\boldsymbol{\Sigma}}_{n,t}^{-1}(k_{n,t}^*, \hat{\boldsymbol{\epsilon}}_n) - \boldsymbol{\Sigma}_n^{-1}\|}{\|\boldsymbol{\Sigma}_n^{-1}\|}, \quad n = 50,\ 100,\ 200,\ 400, \tag{35}$$

where

$$k_{n,t}^* = \arg\min_{1\le l\le l_n} \|\widehat{\boldsymbol{\Sigma}}_{n,t}^{-1}(l, \hat{\boldsymbol{\epsilon}}_n) - \boldsymbol{\Sigma}_n^{-1}\|,$$

with $l_n = 10\log n$. The other three estimators and their corresponding estimation errors are described as follows. One of the estimators is the sample covariance matrix, $\widehat{\boldsymbol{\Sigma}}_{\mathrm{S},n}$, which is formed by plugging $\hat{\gamma}_k = n^{-1}\sum_{i=1}^{n-k}(\varepsilon_i - \bar{\varepsilon}_{(n)})(\varepsilon_{i+k} - \bar{\varepsilon}_{(n)})$, with $\bar{\varepsilon}_{(n)} = n^{-1}\sum_{i=1}^{n}\varepsilon_i$, instead of γ_k into $\boldsymbol{\Sigma}_n$, $k = 0, 1, \cdots, n-1$. Its relative estimation error, $r_{\mathrm{S},n}$, is defined in the same way as (35) but with $\widehat{\boldsymbol{\Sigma}}_{n,t}^{-1}(k_{n,t}^*, \hat{\boldsymbol{\epsilon}}_n)$ replaced by $\widehat{\boldsymbol{\Sigma}}_{\mathrm{S},n,t}^{-1}(\hat{\boldsymbol{\epsilon}}_n)$. The other two estimators are tapered covariance matrices, which are Schur products of the sample covariance matrix and a positive definite matrix. They consist of a main diagonal and l diagonals above and below the main diagonal of $\widehat{\boldsymbol{\Sigma}}_{\mathrm{S},n}$, like a banded sample covariance matrix, but the off-diagonal entries are tapered off rather than maintained intact. The first tapered covariance matrix is constructed using the Bartlett window function, $b_l(z) = 1 - z/l$ if $0 \le z \le l$ and 0 otherwise. It is defined as

$$\widehat{\boldsymbol{\Sigma}}_{\mathrm{B},n}(l) = [b_l(|i-j|)\hat{\gamma}_{|i-j|}]_{i,j=1}^n. \tag{36}$$

The other one is defined as

$$\widehat{\boldsymbol{\Sigma}}_{\mathrm{P},n}(l) = [p_l(|i-j|)\hat{\gamma}_{|i-j|}]_{i,j=1}^n, \tag{37}$$

where $p_l(z) = 1 - 6(z/l)^2 + 6(z/l)^3$ if $0 \le z \le l/2$, $= 2(1 - z/l)^3$ if $l/2 < z \le l$ and 0 otherwise, which is the Parzen window function. The functions, $b_l(z)$ and $p_l(z)$, decrease in z, since the corresponding sample autocovariance estimate, $\hat{\gamma}_z$, is less reliable when the value of z is large. The estimation errors of these two tapered estimators, $r_{\mathrm{B},n}$ and $r_{\mathrm{P},n}$, are also defined in the same way as (35) but $\widehat{\mathbf{\Sigma}}_{n,t}^{-1}(k_{n,t}^*, \hat{\boldsymbol{\epsilon}}_n)$ is respectively replaced by $\widehat{\mathbf{\Sigma}}_{\mathrm{B},n,t}^{-1}(k_{n,t}^0, \hat{\boldsymbol{\epsilon}}_n)$ and $\widehat{\mathbf{\Sigma}}_{\mathrm{P},n,t}^{-1}(k_{n,t}^1, \hat{\boldsymbol{\epsilon}}_n)$, where

$$k_{n,t}^0 = \arg \min_{1 \le l \le l_n} \|\widehat{\mathbf{\Sigma}}_{\mathrm{B},n,t}(l, \hat{\boldsymbol{\epsilon}}_n) - \mathbf{\Sigma}_n\|,$$

and

$$k_{n,t}^1 = \arg \min_{1 \le l \le l_n} \|\widehat{\mathbf{\Sigma}}_{\mathrm{P},n,t}(l, \hat{\boldsymbol{\epsilon}}_n) - \mathbf{\Sigma}_n\|.$$

Values of $r_{\mathrm{S},n}$, $r_{\mathrm{B},n}$, $r_{\mathrm{P},n}$, and $r_{\mathrm{C},n}$, and average values of $k_{n,t}^0$, $k_{n,t}^1$, and $k_{n,t}^*$ over 1000 replications, respectively denoted by k_n^0, k_n^1, and k_n^*, are shown in Tables 1 and 2. These values are obtained under model (1) with $\boldsymbol{x}_t^\top = (1\ x_t)$, where x_t's obey an i.i.d standard Gaussian distribution, $\boldsymbol{\beta}^\top = (\beta_0\ \beta_1) = (2\ 0.5)$, and ε_t's are generated from one of the following data generating processes (DGPs).

DGP 1: $\varepsilon_t = \phi_1 \varepsilon_{t-1} + \eta_t$ with $\phi_1 \in \{0.2, 0.5, 0.8\}$;

DGP 2: $\varepsilon_t = \eta_t - \theta_1 \eta_{t-1}$ with $\theta_1 \in \{0.2, 0.5, 0.8\}$;

DGP 3: $\varepsilon_t = \phi_1 \varepsilon_{t-1} + \phi_2 \varepsilon_{t-2} + \eta_t$ with $(\phi_1, \phi_2) = (1.60, -0.64)$, $(1.80, -0.90)$ or $(1.43, -0.51)$; and

DGP 4: $\varepsilon_t = \eta_t - \theta_1 \eta_{t-1} - \theta_2 \eta_{t-2}$ with $(\theta_1, \theta_2) = (1.60, -0.64)$, $(1.80, -0.90)$ or $(1.43, -0.51)$;

where η_t's obey an i.i.d standard Gaussian distribution.

Table 1 presents the above-mentioned values under model (1) with errors generated by DGP 1 or 2. To facilitate an explanation of the values of k_n^0, k_n^1 and k_n^*, the finite AR representations of DGP 2 with $\theta_1 \in \{0.2, 0.5, 0.8\}$ are given in the footnotes to Table 1. The results shown in Table 1 reveal that regardless of the sample size, the value of r_{C} is smallest for DGP 1 with $\phi_1 \in \{0.2, 0.5, 0.8\}$ and DGP 2 with $\theta_1 \in \{0.2, 0.5, 0.8\}$. In addition, a similar phenomenon occurs in the cases of DGP 1 with $\phi_1 \in \{-0.2, -0.5, -0.8\}$ and DGP 2 with $\theta_1 \in \{-0.2, -0.5, -0.8\}$. Thus, those results are omitted for simplicity. Results in cases of $\phi_1 < 0$ or $\theta_1 < 0$ and in Table 1 together indicate that the performance of the Cholesky-based estimator is remarkably superior to those of the other three estimators as $|\phi_1|$ or $|\theta_1|$ is close to 1. According to the form of the Cholesky decomposition we used, k_n^* should depend on the magnitude of AR coefficients. Our simulation results confirm this conjecture. Specifically, when errors obey DGP 1 (an AR(1) process), the values of k_n^* are equal to 1 for $n = 50$, 100, 200

and 400. In contrast, k_n^0 and k_n^1 seem to depend on the magnitude of not only the autocovariance but also the AR coefficients. These values of k_n^0 and k_n^1 increase with increasing n in all cases of DGPs 1 and 2. When errors are generated by DGP 3 or 4, Table 2 also shows that the Cholesky-based matrix estimator has the smallest relative estimation error compared to the other three estimators for all considered combinations of (ϕ_1, ϕ_2) or (θ_1, θ_2). Tables 1 and 2 support the deduction that the Cholesky-based estimator for the inverse of the error covariance matrix approaches the target matrix as n tends to infinity at the fastest convergence rate among all matrix estimators considered here.

Table 1: Values of $r_{\mathrm{S},n}$, $r_{\mathrm{B},n}$, $r_{\mathrm{P},n}$, $r_{\mathrm{C},n}$, k_n^0, k_n^1, and k_n^* under model (1) with errors following DGP 1 or 2

		ϕ_1			θ_1		
n		0.2	0.5	0.8	0.2†	0.5‡	0.8§
50	r_{S}	9.838	6.431	4.745	12.347	8.581	2.775
	$r_{\mathrm{B}}(k^0)$	0.329(3)	0.433(7)	0.543(13)	0.286(3)	0.501(4)	0.897(4)
	$r_{\mathrm{P}}(k^1)$	0.335(4)	0.512(9)	0.831(18)	0.297(4)	0.471(5)	0.858(5)
	$r_{\mathrm{C}}(k^*)$	0.318(1)	0.216(1)	0.177(1)	0.290(1)	0.232(3)	0.247(9)
100	r_{S}	10.231	7.764	6.595	13.786	11.781	4.274
	$r_{\mathrm{B}}(k^0)$	0.246(4)	0.352(8)	0.533(19)	0.225(3)	0.477(4)	0.894(5)
	$r_{\mathrm{P}}(k^1)$	0.243(5)	0.387(11)	0.718(24)	0.219(4)	0.419(6)	0.854(6)
	$r_{\mathrm{C}}(k^*)$	0.199(1)	0.138(1)	0.111(1)	0.207(1)	0.195(4)	0.241(12)
200	r_{S}	12.400	9.605	8.793	14.830	11.590	6.072
	$r_{\mathrm{B}}(k^0)$	0.193(4)	0.287(10)	0.460(23)	0.182(4)	0.442(5)	0.881(5)
	$r_{\mathrm{P}}(k^1)$	0.184(6)	0.294(13)	0.561(30)	0.167(5)	0.362(6)	0.828(7)
	$r_{\mathrm{C}}(k^*)$	0.140(1)	0.100(1)	0.079(1)	0.144(2)	0.149(4)	0.207(14)
400	r_{S}	15.373	12.648	11.889	19.838	15.013	8.467
	$r_{\mathrm{B}}(k^0)$	0.150(5)	0.232(12)	0.374(28)	0.155(4)	0.409(6)	0.862(6)
	$r_{\mathrm{P}}(k^1)$	0.138(7)	0.218(14)	0.420(35)	0.136(6)	0.326(7)	0.790(7)
	$r_{\mathrm{C}}(k^*)$	0.095(1)	0.068(1)	0.055(1)	0.111(2)	0.117(5)	0.150(15)

† Coefficients of the AR representation, $\phi_1, \cdots, \phi_{15} = -2 \times 10^{-1}$, -4×10^{-2}, -8×10^{-3}, 1.6×10^{-3}, 3.2×10^{-4}, -6.4×10^{-5}, -1.3×10^{-5}, -2.6×10^{-6}, -5.1×10^{-7}, -1.0×10^{-7}, -2.0×10^{-8}, -4.1×10^{-9}, -8.2×10^{-10}, -1.6×10^{-10} and -3.3×10^{-11}.

‡ Coefficients of the AR representation, $\phi_1, \cdots, \phi_{15} = -5 \times 10^{-1}$, -2.5×10^{-1}, -1.3×10^{-1}, -6.3×10^{-2}, 3.1×10^{-2}, 1.6×10^{-2}, -7.8×10^{-3}, 3.9×10^{-3}, -2.0×10^{-3}, -9.8×10^{-4}, -4.9×10^{-4}, -2.4×10^{-4}, -1.2×10^{-4}, -6.1×10^{-5} and -3.1×10^{-5}.

§ Coefficients of the AR representation, $\phi_1, \cdots, \phi_{15} = -0.800$, -0.640, -0.512, -0.410, -0.328, -0.262, -0.210, -0.168, -0.134, -0.107, -0.086, -0.069, -0.055, -0.044 and -0.035.

Table 2: Values of $r_{\mathrm{S},n}$, $r_{\mathrm{B},n}$, $r_{\mathrm{P},n}$, $r_{\mathrm{C},n}$, k_n^0, k_n^1, and k_n^* under model (1) with errors following DGP 3 or 4

		(ϕ_1, ϕ_2)		
n		$(1.60, -0.64)$	$(1.80, -0.90)$	$(1.43, -0.51)$
50	r_{S}	1.219	1.069	1.773
	$r_{\mathrm{B}}(k^0)$	0.904(18)	0.935(30)	0.809(15)
	$r_{\mathrm{P}}(k^1)$	0.787(24)	0.884(33)	0.671(21)
	$r_{\mathrm{C}}(k^*)$	0.349(2)	0.467(2)	0.255(2)
100	r_{S}	1.301	1.117	2.256
	$r_{\mathrm{B}}(k^0)$	0.911(25)	0.938(37)	0.809(21)
	$r_{\mathrm{P}}(k^1)$	0.786(31)	0.860(40)	0.648(26)
	$r_{\mathrm{C}}(k^*)$	0.501(3)	0.529(2)	0.330(3)
200	r_{S}	2.208	1.631	4.169
	$r_{\mathrm{B}}(k^0)$	0.896(31)	0.921(46)	0.782(24)
	$r_{\mathrm{P}}(k^1)$	0.712(38)	0.784(49)	0.559(31)
	$r_{\mathrm{C}}(k^*)$	0.434(3)	0.446(2)	0.243(3)
400	r_{S}	4.703	3.277	7.315
	$r_{\mathrm{B}}(k^0)$	0.864(37)	0.898(54)	0.725(29)
	$r_{\mathrm{P}}(k^1)$	0.555(45)	0.651(56)	0.403(36)
	$r_{\mathrm{C}}(k^*)$	0.231(3)	0.259(2)	0.123(3)
		(θ_1, θ_2)		
n		$(1.60, -0.64)$†	$(1.80, -0.90)$‡	$(1.43, -0.51)$§
50	r_{S}	0.914	0.972	0.924
	$r_{\mathrm{B}}(k^0)$	0.997(5)	0.997(5)	0.986(5)
	$r_{\mathrm{P}}(k^1)$	0.991(7)	0.994(7)	0.965(7)
	$r_{\mathrm{C}}(k^*)$	0.728(20)	0.751(21)	0.550(16)
100	r_{S}	0.933	0.983	1.365
	$r_{\mathrm{B}}(k^0)$	0.996(6)	0.998(6)	0.984(6)
	$r_{\mathrm{P}}(k^1)$	0.986(8)	0.993(8)	0.948(7)
	$r_{\mathrm{C}}(k^*)$	0.639(28)	0.785(27)	0.385(20)
200	r_{S}	1.024	0.981	2.006
	$r_{\mathrm{B}}(k^0)$	0.996(7)	0.998(7)	0.982(7)
	$r_{\mathrm{P}}(k^1)$	0.982(8)	0.992(8)	0.926(8)
	$r_{\mathrm{C}}(k^*)$	0.583(36)	0.727(42)	0.364(17)
400	r_{S}	1.385	1.050	2.938
	$r_{\mathrm{B}}(k^0)$	0.995(8)	0.998(8)	0.979(8)
	$r_{\mathrm{P}}(k^1)$	0.971(9)	0.990(9)	0.902(9)
	$r_{\mathrm{C}}(k^*)$	0.436(34)	0.716(37)	0.405(25)

† Coefficients of the AR representation, $\phi_1, \cdots, \phi_{15} = -1.600, -1.920, -2.048, -2.048, -1.966, -1.835, -1.678, -1.510, -1.342, -1.181, -1.031, -0.893, -0.770, -0.660$ and -0.563.

‡ Coefficients of the AR representation, $\phi_1, \cdots, \phi_{15} = -1.800, -2.340, -2.592, -2.560, -2.275, -1.790, -1.176, -0.505, 0.149, 0.723, 1.167, 1.450, 1.560, 1.503$ and 1.301.

§ Coefficients of the AR representation, $\phi_1, \cdots, \phi_{15} = -1.430, -1.535, -1.466, -1.313, -1.130, -0.946, -0.777, -0.629, -0.502, -0.398, -0.313, -0.244, -0.190, -0.147$ and -0.113.

4.2 Finite Sample Performance of the CGLS Estimator

In this section, we compare the finite sample efficiency of the CGLS estima-

tor with those of three other GLS estimators. The competitive estimators are described as follows. The first one is the sample generalized least squares (SGLS) estimator, which has the same form as (5) but with $\boldsymbol{\Sigma}_n$ replaced by $\widehat{\boldsymbol{\Sigma}}_{\mathrm{S},n}$; the other two GLS estimators with $\boldsymbol{\Sigma}_n$ estimated by tapered covariance matrices, defined in (36) and (37), are tapered GLS estimators, respectively denoted by $\mathrm{TGLS_B}$ and $\mathrm{TGLS_P}$. Since the OLS method is the most popular tool for regression analysis, the relative efficiencies of these GLS estimates to the OLS estimate are evaluated through the measure expressed as

$$\hat{\varsigma}_{\beta_i} = \frac{\sum_{t=1}^{1000}(\hat{\beta}_{\mathrm{GLS},i} - \beta_i)^2}{\sum_{t=1}^{1000}(\hat{\beta}_{\mathrm{OLS},i} - \beta_i)^2}, \quad i = 0,\, 1.$$

A ratio of > 1 indicates that the OLS estimate is more efficient than the GLS estimate. Tables 3 and 4 summarize values of $\hat{\varsigma}_{\beta_0}$ and $\hat{\varsigma}_{\beta_1}$ for the four GLS estimators, namely, SGLS, $\mathrm{TGLS_B}$, $\mathrm{TGLS_P}$ and CGLS, under the simulation settings of Tables 1 and 2, respectively. Estimates of the relative efficiency of the GLS estimator given in (5) over the OLS estimator are also included in Tables 3 and 4, but only as a bench mark for comparison.

Tables 3 and 4 indicate that regardless of the sample size and the method used for estimating the regression coefficients, values of $\hat{\varsigma}_{\beta_1}$ are < 1 except for the SGLS estimator used in cases where the errors follow DGP 1 with $\phi_1 = 0.2$ or DGP 2 with $\theta_1 = 0.2$. With the exception of errors that obey DGP 4, values of $\hat{\varsigma}_{\beta_1}$ appear to depend on the magnitude of the roots of the characteristic polynomial of the underlying error model. In other words, when the roots of the characteristic polynomial are close to 1, values of $\hat{\varsigma}_{\beta_1}$ are smaller. For example, in the case of $n = 50$, when errors obey DGP 1 with $\phi_1 = 0.2$, $\hat{\varsigma}_{\beta_1}$ falls between 0.94 and 1.10, whereas $\hat{\varsigma}_{\beta_1}$ falls between 0.26 and 0.50 when ϕ_1 is changed to 0.8. Another example is that for $n = 50$ and errors from DGP 3 with $(\phi_1, \phi_2) = (1.80, -0.90)$ (where the roots of its characteristic polynomial are $1 + 0.333i$ and $1 - 0.333i$), values of $\hat{\varsigma}_{\beta_1}$ are between 0.01 and 0.06, but values of $\hat{\varsigma}_{\beta_1}$ are between 0.08 and 0.24 for errors following DGP 3 with $(\phi_1, \phi_2) = (1.43, -0.51)$ (where the roots of its characteristic polynomial are 1.471 and 1.333). Values of $\hat{\varsigma}_{\beta_0}$ shown in Tables 3 and 4 reveal that except in cases where errors are from DGP 1 or 2 and the SGLS is used to estimate the intercept term of the regression model, the OLS estimators are less efficient than all GLS estimators considered. Tables 3 and 4 show that when $n = 50$, the CGLS estimator performs worse than the TGLS estimators in most cases. However, when n is increased to 400, the CGLS estimator outperforms the other three estimators, with the exception of estimating β_0 in cases of DGP 1 with $\phi_1 = 0.5$ and DGP 3 with $(\phi_1, \phi_2) = (1.8, -0.9)$. In addition, from the results shown in Tables 3 and 4, we find that when $n = 400$, values of $\hat{\varsigma}_{\beta_{i,\mathrm{CGLS}}} - \hat{\varsigma}_{\beta_{i,\mathrm{GLS}}}$ are > 0 and < 0.026 for $i = 0,\, 1$. Indeed,

Table 3: Values of $\hat{\varsigma}_{\beta_0}$ and $\hat{\varsigma}_{\beta_1}$ under the setting of Table 1

		ϕ_1					
		0.2		0.5		0.8	
n	Method	$\hat{\varsigma}_{\beta_0}$	$\hat{\varsigma}_{\beta_1}$	$\hat{\varsigma}_{\beta_0}$	$\hat{\varsigma}_{\beta_1}$	$\hat{\varsigma}_{\beta_0}$	$\hat{\varsigma}_{\beta_1}$
50	GLS	0.998	0.941	0.984	0.631	0.932	0.269
	SGLS	1.040	1.093	1.030	0.837	1.002	0.499
	TGLS_B	1.000	0.971	0.993	0.683	0.963	0.309
	TGLS_P	1.000	0.964	0.993	0.675	0.962	0.312
	CGLS	0.998	0.984	0.985	0.656	0.948	0.274
100	GLS	1.000	0.929	0.997	0.612	0.980	0.286
	SGLS	1.031	1.040	1.031	0.819	1.025	0.564
	TGLS_B	1.001	0.944	0.998	0.634	0.984	0.319
	TGLS_P	1.000	0.941	0.998	0.627	0.984	0.324
	CGLS	1.002	0.947	1.000	0.619	0.986	0.288
200	GLS	1.001	0.927	1.004	0.614	1.004	0.253
	SGLS	1.031	1.120	1.029	0.865	1.024	0.549
	TGLS_B	1.001	0.947	1.002	0.636	1.000	0.269
	TGLS_P	1.001	0.941	1.002	0.633	1.000	0.271
	CGLS	1.002	0.943	1.005	0.620	1.005	0.253
400	GLS	1.000	0.937	1.004	0.605	0.993	0.283
	SGLS	1.021	1.108	1.027	0.835	1.029	0.562
	TGLS_B	1.000	0.946	1.003	0.619	0.993	0.290
	TGLS_P	1.000	0.946	1.003	0.616	0.993	0.294
	CGLS	1.000	0.946	1.004	0.605	0.992	0.283
		θ_1					
		0.2		0.5		0.8	
n	Method	$\hat{\varsigma}_{\beta_0}$	$\hat{\varsigma}_{\beta_1}$	$\hat{\varsigma}_{\beta_0}$	$\hat{\varsigma}_{\beta_1}$	$\hat{\varsigma}_{\beta_0}$	$\hat{\varsigma}_{\beta_1}$
50	GLS	0.975	0.920	0.824	0.689	0.521	0.511
	SGLS	1.022	1.108	0.938	0.927	0.733	0.773
	TGLS_B	0.984	0.949	0.876	0.769	0.651	0.642
	TGLS_P	0.983	0.947	0.867	0.750	0.620	0.609
	CGLS	0.980	0.964	0.862	0.777	0.664	0.695
100	GLS	1.001	0.911	0.984	0.611	0.761	0.240
	SGLS	1.027	1.102	1.026	0.870	0.878	0.550
	TGLS_B	1.002	0.945	0.988	0.690	0.832	0.479
	TGLS_P	1.001	0.936	0.987	0.668	0.816	0.428
	CGLS	1.003	0.942	0.987	0.670	0.800	0.344
200	GLS	0.997	0.936	0.968	0.603	0.706	0.195
	SGLS	1.049	1.130	1.032	0.856	0.847	0.481
	TGLS_B	0.998	0.936	0.971	0.639	0.756	0.356
	TGLS_P	0.998	0.935	0.969	0.626	0.737	0.306
	CGLS	0.997	0.940	0.970	0.618	0.712	0.232
400	GLS	1.000	0.920	0.996	0.633	0.861	0.250
	SGLS	1.026	1.101	1.022	0.887	0.951	0.547
	TGLS_B	1.000	0.940	0.996	0.679	0.888	0.395
	TGLS_P	1.000	0.935	0.996	0.664	0.879	0.348
	CGLS	1.000	0.930	0.996	0.652	0.867	0.272

Table 4: Values of $\hat{\varsigma}_{\beta_0}$ and $\hat{\varsigma}_{\beta_1}$ under the setting of Table 2

		(ϕ_1, ϕ_2)					
		$(1.60, -0.64)$		$(1.80, -0.90)$		$(1.43, -0.51)$	
n	Method	$\hat{\varsigma}_{\beta_0}$	$\hat{\varsigma}_{\beta_1}$	$\hat{\varsigma}_{\beta_0}$	$\hat{\varsigma}_{\beta_1}$	$\hat{\varsigma}_{\beta_0}$	$\hat{\varsigma}_{\beta_1}$
50	GLS	0.890	0.037	0.853	0.013	0.932	0.087
	SGLS	0.960	0.119	0.891	0.056	0.987	0.236
	TGLS_B	0.939	0.047	0.911	0.021	0.957	0.105
	TGLS_P	0.932	0.050	0.908	0.023	0.952	0.105
	CGLS	0.974	0.038	0.914	0.014	0.972	0.088
100	GLS	0.892	0.006	0.886	0.004	0.924	0.019
	SGLS	0.978	0.068	0.955	0.042	1.001	0.162
	TGLS_B	0.935	0.010	0.929	0.007	0.951	0.029
	TGLS_P	0.928	0.011	0.919	0.008	0.946	0.030
	CGLS	0.936	0.008	0.934	0.004	0.946	0.023
200	GLS	0.971	0.004	0.944	0.003	0.963	0.019
	SGLS	0.997	0.087	0.971	0.059	1.012	0.200
	TGLS_B	0.975	0.006	0.946	0.005	0.974	0.022
	TGLS_P	0.975	0.007	0.944	0.006	0.971	0.023
	CGLS	0.986	0.005	0.955	0.003	0.968	0.019
400	GLS	0.964	0.009	0.976	0.005	0.980	0.027
	SGLS	1.026	0.169	1.024	0.118	1.036	0.278
	TGLS_B	0.972	0.010	0.978	0.006	0.984	0.029
	TGLS_P	0.971	0.010	0.979	0.007	0.983	0.028
	CGLS	0.968	0.009	0.980	0.005	0.982	0.027
		(θ_1, θ_2)					
		$(1.60, -0.64)$		$(1.80, -0.90)$		$(1.43, -0.51)$	
n	Method	$\hat{\varsigma}_{\beta_0}$	$\hat{\varsigma}_{\beta_1}$	$\hat{\varsigma}_{\beta_0}$	$\hat{\varsigma}_{\beta_1}$	$\hat{\varsigma}_{\beta_0}$	$\hat{\varsigma}_{\beta_1}$
50	GLS	0.042	0.013	0.141	0.007	0.156	0.032
	SGLS	0.178	0.220	0.240	0.185	0.280	0.249
	TGLS_B	0.240	0.192	0.287	0.163	0.327	0.209
	TGLS_P	0.158	0.122	0.217	0.095	0.257	0.140
	CGLS	0.486	0.736	0.585	0.718	0.326	0.552
100	GLS	0.036	0.018	0.078	0.005	0.197	0.043
	SGLS	0.226	0.199	0.224	0.153	0.321	0.208
	TGLS_B	0.229	0.200	0.241	0.166	0.341	0.162
	TGLS_P	0.138	0.112	0.154	0.079	0.264	0.100
	CGLS	0.559	0.430	0.429	0.301	0.225	0.174
200	GLS	0.052	0.010	0.182	0.005	0.203	0.035
	SGLS	0.143	0.163	0.299	0.111	0.378	0.208
	TGLS_B	0.154	0.114	0.300	0.108	0.334	0.152
	TGLS_P	0.086	0.047	0.221	0.041	0.253	0.081
	CGLS	0.063	0.080	0.238	0.053	0.216	0.052
400	GLS	0.182	0.013	0.388	0.003	0.405	0.018
	SGLS	0.251	0.183	0.455	0.112	0.507	0.222
	TGLS_B	0.264	0.135	0.445	0.077	0.474	0.122
	TGLS_P	0.202	0.051	0.402	0.019	0.421	0.045
	CGLS	0.183	0.039	0.394	0.010	0.407	0.034

the CGLS estimator is expected to be as efficient as the GLS estimator (BLUE) for large n. In conclusion, the simulation results reported in this section agree with the theoretical results established in Theorem 2. At the end of this section, we note that although the performance of the CGLS estimator is sometimes worse than those of the TGLS_{B} and TGLS_{P} in situations with small sample sizes, the Cholesky-based estimator of the inverse covariance matrix outperforms the other estimators even when $n = 50$. This finding suggests that the Cholesky-based matrix estimator is often a good choice if one is only interested in estimating the covariance matrix of a serial process.

4.3 Real Data Analysis

This section illustrates practical applications of the CGLS estimator using two datasets. The most interesting question is whether the CGLS estimator can yield a predictor, $\hat{y}_t = \hat{\boldsymbol{\beta}}_{\text{CGLS}}^{\top}\boldsymbol{x}_t$, with the smallest prediction error among all candidate predictors built by the SGLS, TGLS_{B}, TGLS_{P} and CGLS estimators. To answer this question, each dataset is split into two parts. The first part contains the first 90% of the observations, $\{(\boldsymbol{x}_1, y_1), \cdots, (\boldsymbol{x}_T, y_T)\}$, used to obtain the four GLS estimates, where $T = \lfloor 0.9 \times n \rfloor$ with n being the sample size and $\lfloor a \rfloor$ denoting the largest integer $\leq a$. The second part contains the remaining observations, $\{(\boldsymbol{x}_{T+1}, y_{T+1}), \cdots, (\boldsymbol{x}_n, y_n)\}$. Based on the four GLS estimates, we can obtain the predicted values, $\hat{y}_t$, $t = T+1, \cdots, n$ and calculate the average squared prediction error (ASPE) defined by

$$\text{ASPE} = \frac{1}{n-T}\sum_{t=T+1}^{n}(y_t - \hat{y}_t)^2. \tag{38}$$

Table 5 presents values of ASPE based on the four GLS estimators. Note that the values of k^0, k^1, and k^* in Table 5 are obtained by minimizing (38).

Now we introduce the datasets. The first dataset given in Durbin and Watson (1951) reports the annual consumption of spirits in the United Kingdom from 1870 to 1938. The sample size, n, is equal to 69. The response variable, y_i, is the annual per capita consumption of spirits, and the explanatory variables, $x_{i,1}$ and $x_{i,2}$, are per capita real income and price of spirits deflated by a general price index, respectively. All observations are recorded in logarithmic form. Fuller (1996, Example 9.7.1) analyzed the data using the following regression model

$$y_t = \beta_0 + \beta_1 x_{t,1} + \beta_2 x_{t,2} + \beta_3 t + \beta_4 (t-35)^2 + \varepsilon_t,$$

where 1869 is the origin for t, and ε_t is assumed to be from a stationary process. The OLS predictor based on all observations is

$$\hat{y}_{t,\text{OLS}} = 2.14 + 0.69x_{t,1} - 0.63x_{t,2} - 0.0095t - 0.00011(t-35)^2. \tag{39}$$

Fuller (1996, Example 9.7.1) showed that the errors are serially correlated by the Durbin-Watson test, and that the errors obey a first-order AR model. The second dataset given in Bryant and Smith (1995) pertains to spot prices of natural gas in Louisiana and Oklahoma from January 1988 to October 1991. The value of n is 46. They described that the spot price for Louisiana natural gas was known on or before the first day of trading on the Oklahoma market, and the Oklahoma spot price typically ran about 85% of the Louisiana spot price, even though seasonal variations existed in these two markets. The following regression model,

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t, \;\; t = 1, \cdots, 46,$$

was used to fit this dataset by Wei (2006, Example 15.1), where y_t and x_t denote the spot prices for Oklahoma and Louisiana, respectively. The OLS predictor is

$$\hat{y}_{t,\mathrm{OLS}} = 0.12 + 0.8 x_t. \tag{40}$$

Wei (2006, Example 15.1) showed that the errors follow a first-order AR model with autoregressive conditionally heteroscedastic (ARCH) disturbances. These findings from Fuller (1996, Example 9.7.1) for the first dataset and from Wei (2006, Example 15.1) for the second one that the errors are serially correlated allow us to apply the proposed estimator to these two datasets. In Table 5, values of ASPE show that the CGLS predictor produces the smallest prediction error among the predictors considered so far in each of the datasets. From Sections 4.1 to 4.3, results reveal that the CGLS method can be a good alternative for both estimation and prediction.

Table 5: Values of the ASPE for the OLS and four GLS predictors

	ASPE	
	Annual consumption of spirits dataset	Spot prices of natural gas dataset
OLS	0.3506	19.3710
SGLS	0.3332	10.1860
$\mathrm{TGLS_B}(k^0)$	0.2947(24)	12.9870(37)
$\mathrm{TGLS_P}(k^1)$	0.3055(1)	12.9870(1)
CGLS(k^*)	0.2633(1)	7.0244(17)

5. Conclusions

In this article, we develop a method for estimating regression models with time series errors. The proposed method evades three common problems with

this research topic. The first one is that the unobservable errors are estimated by the least squares residuals, which allows the proposed method to be used in practice. The second one is that a reduction in the number of parameters needed to calculate the inverse of the error covariance matrix is achieved. This enhances estimation reliability. The third one is that the proposed inverse covariance matrix estimator retains positive definiteness. This study establishes the convergence for the proposed estimator of the inverse covariance matrix and the asymptotic normality of the Cholesky-based generalized least squares estimator. Both the simulation study and real data analysis illustrate the outperformance of the proposed matrix estimator and its corresponding GLS estimator.

Appendix

Proof of Lemma 1. We begin by proving (16). It is easy to see that $O_{1,n} \subseteq \{\|\widehat{\boldsymbol{R}}(k_n) - \boldsymbol{\Sigma}_{k_n}\| > \pi f_{\min}\}$. This, together with (14), Chebyshev's inequality, and the following result that $E(\|\widehat{\boldsymbol{R}}(k_n) - \boldsymbol{\Sigma}_{k_n}\|^2) = o(1)$, finishes the proof of (16). By an argument used in the proof of Lemma 2 of Ing and Wei (2003), one obtains

$$E\left(\|\widehat{\boldsymbol{R}}(k_n) - \boldsymbol{\Sigma}_{k_n}\|^2\right) \le \sum_{i=1}^{k_n}\sum_{j=1}^{k_n} E\left((\hat{\gamma}_{i,j} - \gamma_{i-j})^2\right) \le C\left(\frac{k_n^2}{N_n}\right), \tag{A.1}$$

where γ_{i-j} denotes the (i,j)th element of $\boldsymbol{\Sigma}_{k_n}$. Thus, (16) is proven. To show (17), we start by noting that

$$\begin{aligned}
&\left(E\left(\sum_{t=k_n}^{n-1}(\hat{\varepsilon}_{t-i}\hat{\varepsilon}_{t-j} - \varepsilon_{t-i}\varepsilon_{t-j})\right)^2\right)^{1/2} \\
&\le \left\|\left(\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{x}_t\boldsymbol{x}_t^\top\right)^{-1}\right\| \left\{E\left(\left\|\frac{1}{n^{1/2}}\sum_{t=1}^{n}\boldsymbol{x}_t\varepsilon_t\right\|^4\right) E\left(\left\|\frac{1}{n^{1/2}}\sum_{t=k_n}^{n-1}\boldsymbol{x}_{t-i}\varepsilon_{t-j}\right\|^4\right)\right\}^{1/4} \\
&+ \left\|\left(\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{x}_t\boldsymbol{x}_t^\top\right)^{-1}\right\| \left\{E\left(\left\|\frac{1}{n^{1/2}}\sum_{t=1}^{n}\boldsymbol{x}_t\varepsilon_t\right\|^4\right) E\left(\left\|\frac{1}{n^{1/2}}\sum_{t=k_n}^{n-1}\boldsymbol{x}_{t-j}\varepsilon_{t-i}\right\|^4\right)\right\}^{1/4} \\
&+ \left\|\left(\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{x}_t\boldsymbol{x}_t^\top\right)^{-1}\right\|^2 \left\|\frac{1}{n}\sum_{t=k_n}^{n-1}\boldsymbol{x}_{t-i}\boldsymbol{x}_{t-j}^\top\right\| \left\{E\left(\left\|\frac{1}{n^{1/2}}\sum_{t=1}^{n}\boldsymbol{x}_t\varepsilon_t\right\|^4\right)\right\}^{1/2},
\end{aligned} \tag{A.2}$$

for $0 \le i, j < k_n$. This follows from Minkowski's and Hölder's inequalities. By

Jensen's inequality, Lemma 2 of Wei (1987), and $\sum_{i=1}^{\infty}|b_i|<\infty$, one obtains

$$E\left(\left\|\frac{1}{n^{1/2}}\sum_{t=1}^{n}\boldsymbol{x}_t\varepsilon_t\right\|^4\right)\le p\sum_{i=1}^{p}\frac{1}{n^2}E\left(\sum_{t=1}^{n}x_{t,i}\varepsilon_t\right)^4\le C\frac{1}{n^2}\left(\sum_{t=1}^{n}\left(\sum_{j=0}^{\infty}b_j\right)^2\right)^2<C.$$

Combining this with (A.2) and (12) yields

$$E\left(\sum_{t=k_n}^{n-1}(\hat{\varepsilon}_{t-i}\hat{\varepsilon}_{t-j}-\varepsilon_{t-i}\varepsilon_{t-j})\right)^2=O(1),\ 0\le i,j<k_n, \tag{A.3}$$

which implies that the following result holds true:

$$E\left(\|\widehat{\boldsymbol{R}}(k_n,\hat{\boldsymbol{\epsilon}}_n)-\widehat{\boldsymbol{R}}(k_n)\|^2\right)\le C\left(\frac{k_n}{N_n}\right)^2. \tag{A.4}$$

It is easy to see that $O_{2,n}\subseteq\left\{\|\widehat{\boldsymbol{R}}(k_n,\hat{\boldsymbol{\epsilon}}_n)-\widehat{\boldsymbol{R}}(k_n)\|>\pi f_{\min}/2\right\}$. This, together with Chebyshev's inequality and (A.4), yields

$$P(O_{2,n})=o(1). \tag{A.5}$$

Then, (17) is an immediate consequence of (16) and (A.5).

To prove (18), we first note that for any $M>0$,

$$\begin{aligned}P\left(n^{1/3}\max_{1\le j\le l}|\hat{\gamma}_{i,j}-\gamma_{i-j}|>M\right)&\le\sum_{j=1}^{l}P\left(n^{1/3}|\hat{\gamma}_{i,j}-\gamma_{i-j}|>M\right)\\&=o(1),\ l=1,\cdots,k_n,\end{aligned}$$

by Chebyshev's inequality and an argument used in the proof of Lemma 2 of Ing and Wei (2003). Then, it implies that

$$\max_{1\le l\le k_n}\|\widehat{\boldsymbol{R}}(l)-\boldsymbol{\Sigma}_l\|_1=\max_{1\le l\le k_n}\sum_{i=1}^{l}\max_{1\le j\le l}|\hat{\gamma}_{i,j}-\gamma_{i-j}|=O_p\left(\frac{k_n}{n^{1/3}}\right). \tag{A.6}$$

In addition, since $\|\boldsymbol{D}^{-1}(n)\|_1<1/\sigma^2$ and (28), $\|\boldsymbol{\Sigma}_n^{-1}\|_1=O(1)$. Then, the following result is true:

$$\max_{1\le l\le k_n}\|\boldsymbol{\Sigma}_l^{-1}\|_1=O(1). \tag{A.7}$$

Combining (A.7) with (A.6) yields $\max_{1\le l\le k_n}\left(\|\boldsymbol{\Sigma}_l^{-1}\|_1\|\widehat{\boldsymbol{R}}(l)-\boldsymbol{\Sigma}_l\|_1\right)=o_p(1)$. This, together with (A.7), implies that

$$\begin{aligned}\max_{1\le l\le k_n}\|\widehat{\boldsymbol{R}}^{-1}(l)\|_1I_{O_{1,n}^c}&\le\left\{1-\max_{1\le l\le k_n}\left(\|\boldsymbol{\Sigma}_l^{-1}\|_1\|\widehat{\boldsymbol{R}}(l)-\boldsymbol{\Sigma}_l\|_1\right)\right\}^{-1}\max_{1\le l\le k_n}\|\boldsymbol{\Sigma}_l^{-1}\|_1\\&=O_p(1).\end{aligned}$$

Thus, (18) is proven.

Next, to show (19), we start by observing that

$$\max_{1\le l\le k_n}\|\widehat{\boldsymbol{R}}(l,\hat{\boldsymbol{\epsilon}}_n)-\widehat{\boldsymbol{R}}(l)\|_1\le\max_{0\le j\le k_n-1}\sum_{i=0}^{k_n-1}\frac{1}{N_n}\left|\sum_{t=k_n}^{n-1}(\hat{\varepsilon}_{t-i}\hat{\varepsilon}_{t-j}-\varepsilon_{t-i}\varepsilon_{t-j})\right|. \quad \text{(A.8)}$$

By Chebyshev's inequality and (A.3), it holds true that for any $M>0$,

$$\begin{aligned}
&P\left(\max_{0\le j\le k_n-1}n^{-1/4}\left|\sum_{t=k_n}^{n-1}(\hat{\varepsilon}_{t-i}\hat{\varepsilon}_{t-j}-\varepsilon_{t-i}\varepsilon_{t-j})\right|>M\right)\\
&\le\sum_{j=0}^{k_n-1}P\left(n^{-1/4}\left|\sum_{t=k_n}^{n-1}(\hat{\varepsilon}_{t-i}\hat{\varepsilon}_{t-j}-\varepsilon_{t-i}\varepsilon_{t-j})\right|>M\right)\\
&\le C\left(\frac{k_n}{n^{1/2}}\right)=o(1). \qquad \text{(A.9)}
\end{aligned}$$

Combining (A.8) with (A.9) yields

$$\max_{1\le l\le k_n}\|\widehat{\boldsymbol{R}}(l,\hat{\boldsymbol{\epsilon}}_n)-\widehat{\boldsymbol{R}}(l)\|_1\le C\left(\frac{k_n}{n^{3/4}}\right).$$

Then,

$$\begin{aligned}
&\max_{1\le l\le k_n}\|\widehat{\boldsymbol{R}}^{-1}(l)-\widehat{\boldsymbol{R}}^{-1}(l,\hat{\boldsymbol{\epsilon}}_n)\|_1 I_{(O_{1,n}\bigcup O_{2,n})^c}\\
&\le\left\{1-\max_{1\le l\le k_n}\left(\|\widehat{\boldsymbol{R}}^{-1}(l)\|_1 I_{O_{1,n}^c}\|\widehat{\boldsymbol{R}}(l,\hat{\boldsymbol{\epsilon}}_n)-\widehat{\boldsymbol{R}}(l)\|_1\right)\right\}^{-1}\\
&\max_{1\le l\le k_n}\left(\|\widehat{\boldsymbol{R}}^{-1}(l)\|_1^2 I_{O_{1,n}^c}\|\widehat{\boldsymbol{R}}(l,\hat{\boldsymbol{\epsilon}}_n)-\widehat{\boldsymbol{R}}(l)\|_1\right)=O_p\left(\frac{k_n}{n^{3/4}}\right).
\end{aligned}$$

Finally, to show (20), we use an argument similar to that used to prove (19), and hence it suffices for (20) to show that

$$\max_{1\le l\le k_n}\|\widehat{\boldsymbol{R}}(l,\hat{\boldsymbol{\epsilon}}_n)-\widehat{\boldsymbol{R}}(l)\|=O_p\left(\frac{k_n}{N_n}\right). \quad \text{(A.10)}$$

By (A.3), one gets

$$\begin{aligned}
&E\left(\max_{1\le l\le k_n}\|\widehat{\boldsymbol{R}}(l,\hat{\boldsymbol{\epsilon}}_n)-\widehat{\boldsymbol{R}}(l)\|^2\right)\\
&\le\sum_{i=1}^{k_n}\sum_{j=1}^{k_n}\frac{1}{N_n^2}E\left(\sum_{t=k_n}^{n-1}(\hat{\varepsilon}_{t+1-i}\hat{\varepsilon}_{t+1-j}-\varepsilon_{t+1-i}\varepsilon_{t+1-j})\right)^2\le C\left(\frac{k_n}{N_n}\right)^2.
\end{aligned}$$

This, together with Chebyshev's inequality, yields (A.10). Thus, (20) is proven.

Proof of Lemma 2. It follows from (15) and an argument used in the proof of Lemma 2 of Ing and Wei (2003) that

$$E\left(\frac{1}{N_n}\sum_{t=k_n}^{n-1}\varepsilon_{t+1-i}\varepsilon_{t+1}\right)^2 \le C\left(\gamma_i^2 + E\left((\hat{\gamma}_{i,0}-\gamma_i)^2\right)\right) = O(1),\ i=1,\cdots,k_n. \tag{A.11}$$

By Chebyshev's inequality and Jensen's inequality, one has that for any $\delta > 0$, there exists $M > \delta^{-1/2}$ such that

$$P\left(\frac{1}{k_n}\left\|\frac{1}{N_n}\sum_{t=k_n}^{n-1}\boldsymbol{\epsilon}_t(k_n)\varepsilon_{t+1}\right\|_1 > M\right) \le \frac{1}{M^2 k_n}\sum_{i=1}^{k_n}E\left(\frac{1}{N_n}\sum_{t=k_n}^{n-1}\varepsilon_{t+1-i}\varepsilon_{t+1}\right)^2 < \delta.$$

Thus, the proof of (21) is complete. In addition, (A.11) also implies

$$E\left(\left\|\frac{1}{N_n}\sum_{t=k_n}^{n-1}\boldsymbol{\epsilon}_t(k_n)\varepsilon_{t+1}\right\|^2\right) \le Ck_n.$$

Combining this and Chebyshev's inequality yields (22).

Proof of Lemma 3. Lemmas 3 and 4 of Ing and Wei (2003) yield that

$$E\left(\left\|\frac{1}{N_n^{1/2}}\sum_{t=k_n}^{n-1}\boldsymbol{\epsilon}_t(l)\eta_{t+1,l}\right\|^2\right) \le Cl\left(\|\boldsymbol{a}-\boldsymbol{a}(l)\|_{\boldsymbol{R}}^2 + O(1)\right),\ 1\le l\le k_n, \tag{A.12}$$

where $\boldsymbol{a} = (a_1, a_2, \cdots)^\top$. For an infinite-dimensional vector $\boldsymbol{w} = (w_1, w_2, \cdots)^\top$ with $\|\boldsymbol{w}\| < \infty$,

$$\|\boldsymbol{w}\|_{\boldsymbol{R}}^2 = \boldsymbol{w}^\top\boldsymbol{R}\boldsymbol{w} = E\left(\sum_{j=1}^{\infty}w_j\varepsilon_{t+1-j}\right)^2 = \int_{-\pi}^{\pi}|\sum_{j=1}^{\infty}w_j e^{-ij\lambda}|^2 f_\varepsilon(\lambda)d\lambda.$$

By (3) and (14), one obtains

$$\|\boldsymbol{a}-\boldsymbol{a}(l)\|_{\boldsymbol{R}}^2 \le E\left(\sum_{j>l}a_j\varepsilon_{t+1-j}\right)^2 \le 2\pi f_{\max}\sum_{j>l}a_j^2 < \infty,\ 1\le l\le k_n. \tag{A.13}$$

It follows from (A.12), (A.13), and Chebyshev's inequality that for any $M > 0$,

$$P\left(\max_{1\le l\le k_n}\left\|\frac{1}{N_n^{5/6}}\sum_{t=k_n}^{n-1}\boldsymbol{\epsilon}_t(l)\eta_{t+1,l}\right\| > M\right) \le \sum_{l=1}^{k_n} P\left(\left\|\frac{1}{N_n^{5/6}}\sum_{t=k_n}^{n-1}\boldsymbol{\epsilon}_t(l)\eta_{t+1,l}\right\| > M\right)$$
$$= o(1).$$

Thus, (23) holds true. The second statement of Lemma 3 can be shown by (23) and some algebraic manipulations. The details are thus omitted.

Proof of Lemma 4. By Chebyshev's inequality, Jensen's inequality, and (A.3), we can show that for any $\delta > 0$, there exists $M > \delta^{-1/2}$ such that

$$P\left(\frac{1}{k_n}\left\|\sum_{t=k_n}^{n-1}\left(\hat{\boldsymbol{\epsilon}}_t(k_n)\hat{\varepsilon}_{t+1} - \boldsymbol{\epsilon}_t(k_n)\varepsilon_{t+1}\right)\right\|_1 > M\right)$$
$$\le \frac{1}{M^2k_n^2}E\left(\left\|\sum_{t=k_n}^{n-1}\left(\hat{\boldsymbol{\epsilon}}_t(k_n)\hat{\varepsilon}_{t+1} - \boldsymbol{\epsilon}_t(k_n)\varepsilon_{t+1}\right)\right\|_1^2\right)$$
$$\le \frac{1}{M^2k_n}\sum_{i=1}^{k_n}E\left(\sum_{t=k_n}^{n-1}(\hat{\varepsilon}_{t-i}\hat{\varepsilon}_{t+1} - \varepsilon_{t-i}\varepsilon_{t+1})\right)^2 < \delta.$$

Then, (25) holds. Moreover, by an similar argument used in proving (25), the proof of (26) is complete. The details are omitted.

We are now ready to prove (29), (30), (33) and (34).

Proof of (29). First we define

$$\boldsymbol{B}(l,\hat{\boldsymbol{\epsilon}}_n) = \left(\widehat{\boldsymbol{R}}^{-1}(l) - \widehat{\boldsymbol{R}}^{-1}(l,\hat{\boldsymbol{\epsilon}}_n)\right) I_{(O_{1,n}\bigcup O_{2,n})^c},\ l = 1, \cdots, k_n,$$

and

$$\boldsymbol{c}(l,\hat{\boldsymbol{\epsilon}}_n) = \frac{1}{N_n}\sum_{t=k_n}^{n-1}\left(\hat{\boldsymbol{\epsilon}}_t(k_n)\hat{\varepsilon}_{t+1} - \boldsymbol{\epsilon}_t(k_n)\varepsilon_{t+1}\right),\ l = 1, \cdots, k_n,$$

where $\widehat{\boldsymbol{R}}(l)$, $\widehat{\boldsymbol{R}}(l,\hat{\boldsymbol{\epsilon}}_n)$, $O_{1,n}$, and $O_{2,n}$ are as defined previously. Then, one can write

$$\hat{\boldsymbol{a}}(l,\hat{\boldsymbol{\epsilon}}_n) - \boldsymbol{a}(l) = \boldsymbol{B}(l,\hat{\boldsymbol{\epsilon}}_n)\left(\boldsymbol{c}(l,\hat{\boldsymbol{\epsilon}}_n) + \frac{1}{N_n}\sum_{t=k_n}^{n-1}\boldsymbol{\epsilon}_t(l)\varepsilon_{t+1}\right)$$
$$-\widehat{\boldsymbol{R}}^{-1}(l)I_{O_{1,n}^c}\left(\boldsymbol{c}(l,\hat{\boldsymbol{\epsilon}}_n) + \frac{1}{N_n}\sum_{t=k_n}^{n-1}\boldsymbol{\epsilon}_t(l)\eta_{t+1,l}\right).$$

Thus, it follows from Lemmas 1-4 that

$$\begin{aligned}
&\max_{1\le l\le k_n}\|\hat{\boldsymbol{a}}(l,\hat{\boldsymbol{\epsilon}}_n)-\boldsymbol{a}(l)\|_1\\
&=\max_{1\le l\le k_n}\|\boldsymbol{B}(l,\hat{\boldsymbol{\epsilon}}_n)\|_1\left(\|\boldsymbol{c}(l,\hat{\boldsymbol{\epsilon}}_n)\|_1+\left\|\frac{1}{N_n}\sum_{t=k_n}^{n-1}\boldsymbol{\epsilon}_t(l)\varepsilon_{t+1}\right\|_1\right)\\
&\quad+\max_{1\le l\le k_n}\|\widehat{\boldsymbol{R}}^{-1}(l)I_{O^c_{1,n}}\|_1\left(\|\boldsymbol{c}(l,\hat{\boldsymbol{\epsilon}}_n)\|_1+\max_{1\le l\le k_n}\left\|\frac{1}{N_n}\sum_{t=k_n}^{n-1}\boldsymbol{\epsilon}_t(l)\eta_{t+1,l}\right\|_1\right)\\
&=O_p\left(\left(\frac{k_n}{n^{1/3}}\right)^2\right).
\end{aligned}$$

Proof of (30). We begin the proof by noting that

$$\begin{aligned}
&\|\widehat{\boldsymbol{G}}_{k_n}(n,\,\hat{\boldsymbol{\epsilon}}_n)\|\\
&\le\left|\frac{1}{N_n}\sum_{t=k_n}^{n-1}(\hat{\varepsilon}_t^2-E(\varepsilon_1^2))\right|+\max_{1\le j\le k_n}\left|\frac{1}{N_n}\sum_{t=k_n}^{n-1}(\hat{\eta}^2_{t+1,j}(\hat{\boldsymbol{\epsilon}}_n)-\hat{\eta}^2_{t+1,j})\right|\\
&\quad+\max_{1\le j\le k_n}\left|\frac{1}{N_n}\sum_{t=k_n}^{n-1}(\hat{\eta}^2_{t+1,j}-\eta^2_{t+1,j})\right|+\max_{1\le j\le k_n}\left|\frac{1}{N_n}\sum_{t=k_n}^{n-1}\eta^2_{t+1,j}-E(\eta^2_{j+1,j})\right|\\
&\quad+\max_{k_n<j<n}\left|E(\eta^2_{k_n+1,k_n})-E(\eta^2_{j+1,j})\right|:=\mathrm{I}+\mathrm{II}+\mathrm{III}+\mathrm{IV}+\mathrm{V},
\end{aligned}\tag{A.14}$$

where $\hat{\eta}_{t+1,j}=\varepsilon_{t+1}+\hat{\boldsymbol{a}}^\top(j)\boldsymbol{\epsilon}_t(j)$ with

$$\hat{\boldsymbol{a}}(j)=-\left(\sum_{t=k_n}^{n-1}\boldsymbol{\epsilon}_t(j)\boldsymbol{\epsilon}_t^\top(j)\right)^{-1}\sum_{t=k_n}^{n-1}\boldsymbol{\epsilon}_t(j)\varepsilon_{t+1}.$$

(A.3) implies that

$$E\left(\frac{1}{N_n}\sum_{t=k_n}^{n-1}(\hat{\varepsilon}^2_{t+1}-\varepsilon^2_{t+1})\right)^2\le CN_n^{-2}.\tag{A.15}$$

This, together with an argument used in the proof of Lemma 2 of Ing and Wei (2003), yields that

$$E(\mathrm{I}^2)\le C\left\{E\left(\frac{1}{N_n}\sum_{t=k_n}^{n-1}(\hat{\varepsilon}_t^2-\varepsilon_t^2)\right)^2+E\left(\frac{1}{N_n}\sum_{t=k_n}^{n-1}(\varepsilon_t^2-E(\varepsilon_1^2))\right)^2\right\}\le CN_n^{-1}.$$

Thus, by Markov's inequality, one gets

$$I = O_p(N_n^{-1/2}). \tag{A.16}$$

With some algebraic calculations, one has

$$\begin{aligned}
\mathrm{II} \le & \left| \frac{1}{N_n} \sum_{t=k_n}^{n-1} (\hat{\varepsilon}_{t+1}^2 - \varepsilon_{t+1}^2) \right| \\
& + \max_{1 \le j \le k_n} \left| \left(\frac{1}{N_n} \sum_{t=k_n}^{n-1} \hat{\boldsymbol{\epsilon}}_t(j) \hat{\varepsilon}_{t+1} \right)^\top \widehat{\boldsymbol{R}}^{-1}(k_n, \hat{\boldsymbol{\epsilon}}_n) I_{O_{2,n}^c} \left(\frac{1}{N_n} \sum_{t=k_n}^{n-1} \hat{\boldsymbol{\epsilon}}_t(j) \hat{\varepsilon}_{t+1} \right) \right. \\
& \left. - \left(\frac{1}{N_n} \sum_{t=k_n}^{n-1} \boldsymbol{\epsilon}_t(j) \varepsilon_{t+1} \right)^\top \widehat{\boldsymbol{R}}^{-1}(k_n) I_{O_{1,n}^c} \left(\frac{1}{N_n} \sum_{t=k_n}^{n-1} \boldsymbol{\epsilon}_t(j) \varepsilon_{t+1} \right) \right| \\
\le & \left| \frac{1}{N_n} \sum_{t=k_n}^{n-1} (\hat{\varepsilon}_{t+1}^2 - \varepsilon_{t+1}^2) \right| + \left\| \frac{1}{N_n} \sum_{t=k_n}^{n-1} \boldsymbol{\epsilon}_t(k_n) \varepsilon_{t+1} \right\|^2 \max_{1 \le j \le k_n} \|\boldsymbol{B}(j, \hat{\boldsymbol{\epsilon}}_n)\| \\
& + \|\boldsymbol{c}(j, \hat{\boldsymbol{\epsilon}}_n)\| \left(\max_{1 \le j \le k_n} \|\boldsymbol{B}(j, \hat{\boldsymbol{\epsilon}}_n)\| + \max_{1 \le j \le k_n} \|\widehat{\boldsymbol{R}}^{-1}(j)\| I_{O_{1,n}^c} \right) \\
& \left(\|\boldsymbol{c}(j, \hat{\boldsymbol{\epsilon}}_n)\| + 2 \left\| \frac{1}{N_n} \sum_{t=k_n}^{n-1} \boldsymbol{\epsilon}_t(k_n) \varepsilon_{t+1} \right\| \right),
\end{aligned}$$

where $\boldsymbol{B}(j, \hat{\boldsymbol{\epsilon}}_n)$ and $\boldsymbol{c}(j, \hat{\boldsymbol{\epsilon}}_n)$ are defined in the proof of (29). Then, it follows from Lemmas 1, 2 and 4 and (A.15) that

$$\mathrm{II} = O_p\left(\frac{k_n^2}{N_n} \right). \tag{A.17}$$

To deal with III, we first note that

$$\mathrm{III} \le C \max_{1 \le j \le k_n} \left| \left(\frac{1}{N_n} \sum_{t=k_n}^{n-1} \boldsymbol{\epsilon}_t(j) \eta_{t+1,j} \right)^\top \widehat{\boldsymbol{R}}^{-1}(k_n) I_{O_{1,n}^c} \left(\frac{1}{N_n} \sum_{t=k_n}^{n-1} \boldsymbol{\epsilon}_t(j) \eta_{t+1,j} \right) \right|.$$

By Lemmas 1 and 3, one obtains

$$\mathrm{III} = O_p(N_n^{-1/3}). \tag{A.18}$$

To deal with IV, by Theorem 1 of Ing and Wei (2003), one shows that

$$E\left(\left| \frac{1}{N_n} \sum_{t=k_n}^{n-1} \eta_{t+1,j}^2 - E(\eta_{j+1,j}^2) \right|^2 \right) \le C N_n^{-1}, \quad j = 1, \cdots, k_n.$$

Combining this with Chebyshev's inequality and $k_n = o(n^{1/3})$, one can assert that for any $\delta > 0$, there exists $M > 0$ such that

$$P\left(\max_{1\leq j\leq k_n} N_n^{1/3}\left|\frac{1}{N_n}\sum_{t=k_n}^{n-1}\eta_{t+1,j}^2 - E(\eta_{j+1,j}^2)\right| > M\right) = o(1).$$

Thus, the following result holds true:

$$\mathrm{IV} = O_p\left(N_n^{-1/3}\right). \tag{A.19}$$

Finally, to deal with V, the fact that $\|\boldsymbol{a}(j) - \boldsymbol{a}\|_{\boldsymbol{R}} \leq \|\boldsymbol{a}(k_n) - \boldsymbol{a}\|_{\boldsymbol{R}}$, $j \geqslant k_n$ and some algebraic manipulations imply that

$$\mathrm{V} = O_p\left(\|\boldsymbol{a}(k_n) - \boldsymbol{a}\|_{\boldsymbol{R}}\right). \tag{A.20}$$

Combining (A.14) with (A.16)-(A.18), (A.19), and (A.20) gives (30).

Proof of (33). One writes

$$\boldsymbol{F}_n(\widehat{\boldsymbol{\beta}}_{\mathrm{CGLS}}(k_n,\, \hat{\boldsymbol{\epsilon}}_n) - \widehat{\boldsymbol{\beta}}_{\mathrm{GLS}}) = (\boldsymbol{L}_1 + \boldsymbol{L}_2)^{-1}\boldsymbol{L}_3 - (\boldsymbol{L}_1 + \boldsymbol{L}_2)^{-1}\boldsymbol{L}_1\boldsymbol{L}_2^{-1}\boldsymbol{L}_4, \tag{A.21}$$

where

$$\boldsymbol{L}_1 = \boldsymbol{F}_n^{-1}\boldsymbol{X}_n^{\top}\left(\widehat{\boldsymbol{\Sigma}}_n^{-1}(k_n,\, \hat{\boldsymbol{\epsilon}}_n) - \boldsymbol{\Sigma}_n^{-1}\right)\boldsymbol{X}_n\boldsymbol{F}_n^{-1},\ \boldsymbol{L}_2 = \boldsymbol{F}_n^{-1}\boldsymbol{X}_n^{\top}\boldsymbol{\Sigma}_n^{-1}\boldsymbol{X}_n\boldsymbol{F}_n^{-1},$$
$$\boldsymbol{L}_3 = \boldsymbol{F}_n^{-1}\boldsymbol{X}_n^{\top}\left(\widehat{\boldsymbol{\Sigma}}_n^{-1}(k_n,\, \hat{\boldsymbol{\epsilon}}_n) - \boldsymbol{\Sigma}_n^{-1}\right)\boldsymbol{\epsilon}_n, \text{ and } \boldsymbol{L}_4 = \boldsymbol{F}_n^{-1}\boldsymbol{X}_n^{\top}\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\epsilon}_n.$$

By (12) and the existence of $\widehat{\boldsymbol{\Sigma}}_{k_n}^{-1}(n,\, \hat{\boldsymbol{\epsilon}}_n)$, one shows that

$$\|\,(\boldsymbol{L}_1 + \boldsymbol{L}_2)^{-1}\,\| = (\lambda_{\min}(\boldsymbol{L}_1 + \boldsymbol{L}_2))^{-1} < C. \tag{A.22}$$

By an argument similar to that used in the previous statement, (12) and (15) yield

$$\|\boldsymbol{L}_2^{-1}\| < C. \tag{A.23}$$

In addition, it follows from (3), (12), Theorem 1, and $k_n = o(n^{1/3})$ that

$$\|\boldsymbol{L}_1\| \leq \left\|\widehat{\boldsymbol{\Sigma}}_n^{-1}(k_n,\, \hat{\boldsymbol{\epsilon}}_n) - \boldsymbol{\Sigma}_n^{-1}\right\|\left\|\frac{1}{n}\boldsymbol{S}_n\right\|\|n^{1/2}\boldsymbol{F}_n^{-1}\|^2 = o_p(1). \tag{A.24}$$

In view of (A.21)-(A.24), it suffices for (33) to show that

$$\|\boldsymbol{L}_3\| = o_p(1), \tag{A.25}$$

and

$$\|\boldsymbol{L}_4\| = O_p(1). \tag{A.26}$$

To prove (A.25), we first let $\boldsymbol{F}_n^{-1}\boldsymbol{X}_n^\top = \left(\boldsymbol{x}_1^*, \boldsymbol{x}_2^*, \cdots, \boldsymbol{x}_p^*\right)^\top$. Then, one has

$$\begin{aligned}\|\boldsymbol{L}_3\|^2 &= \sum_{i=1}^{p}\left(\boldsymbol{x}_i^{*\top}\left(\widehat{\boldsymbol{\Sigma}}_n^{-1}(k_n, \hat{\boldsymbol{\epsilon}}_n) - \boldsymbol{R}^{-1}(n)\right)\boldsymbol{\epsilon}_n\right)^2 \\ &\le p\|\widehat{\boldsymbol{\Sigma}}_n^{-1}(k_n, \hat{\boldsymbol{\epsilon}}_n) - \boldsymbol{\Sigma}_n^{-1}\|^2\|\boldsymbol{\epsilon}_n\boldsymbol{\epsilon}_n^\top\| \max_{1\le i\le p}\|\boldsymbol{x}_i^*\|^2.\end{aligned}$$

This, together with (15), Theorem 1, and the fact that $\|\boldsymbol{x}_i^*\| = 1$, $1 \le i \le p$ gives (A.25). By an argument similar to that used to prove (A.25), the proof of (A.26) is complete. The details are thus omitted.

Proof of (34). Define $\boldsymbol{e}_n = \boldsymbol{F}_n^{-1}\boldsymbol{X}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\epsilon}_n$. Then, one has

$$\boldsymbol{F}_n(\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}} - \boldsymbol{\beta}) = \left(\boldsymbol{F}_n^{-1}\boldsymbol{X}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{X}_n\boldsymbol{F}_n^{-1}\right)^{-1}\boldsymbol{e}_n, \tag{A.27}$$

This proof can be completed by combining (A.27) and the following two results:

$$\lim_{n\to\infty}\boldsymbol{F}_n^{-1}\boldsymbol{X}^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{X}\boldsymbol{F}_n^{-1} = \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{f_\varepsilon(\lambda)}d\boldsymbol{M}(\lambda), \tag{A.28}$$

and

$$\boldsymbol{e}_n \xrightarrow{d} \mathcal{N}\left(\boldsymbol{0}, \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{f_\varepsilon(\lambda)}d\boldsymbol{M}(\lambda)\right) \text{ as } n\to\infty. \tag{A.29}$$

(A.28) follows from Theorem 10.2.7 of Anderson (1994). To show (A.29), we define $\boldsymbol{e}_{n,m} = \boldsymbol{F}_n^{-1}\boldsymbol{X}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\epsilon}_{n,m}$ with $\boldsymbol{\epsilon}_{n,m}$ denoted by

$$\boldsymbol{\epsilon}_{n,m} = \left(\sum_{j=0}^{m} b_j\eta_{n-j}, \sum_{j=0}^{m} b_j\eta_{n-1-j}, \cdots, \sum_{j=0}^{m} b_j\eta_{1-j}\right)^\top,$$

where m is a positive integer dependent on n and increasing to infinity. Note that $\boldsymbol{\epsilon}_{n,\,m}$ can be decomposed as $\boldsymbol{\epsilon}_{n,m} = \boldsymbol{\Gamma}_{n,m}\boldsymbol{\eta}_n(n+m)$, where

$$\boldsymbol{\Gamma}_{n,m} = \begin{pmatrix} 1 & b_1 & b_2 & \cdots & b_m & 0 & & \cdots & 0 & \cdots & 0 \\ 0 & 1 & b_1 & \cdots & b_{m-1} & b_m & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & b_{m-2} & b_{m-1} & b_m & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & & \ddots & \ddots & \ddots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & \cdots & b_{m-2} & b_{m-1} & b_m & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & & \ddots & & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & \cdots & 1 & \cdots & b_{m-2} & b_{m-1} & b_m \end{pmatrix},$$

and $\boldsymbol{\eta}_n(n+m) = \left(\eta_n, \eta_{n-1}, \cdots, \eta_0, \eta_{-1}, \cdots, \eta_{-(m-1)}\right)^\top$. If we can show that for each m,

$$\boldsymbol{e}_{n,m} \xrightarrow{d} \boldsymbol{e}_m^* \text{ as } n \to \infty, \tag{A.30}$$

and

$$\boldsymbol{e}_m^* \xrightarrow{d} \mathcal{N}\left(\mathbf{0},\ \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1}{f_\varepsilon(\lambda)} d\boldsymbol{M}(\lambda)\right) \text{ as } m \to \infty. \tag{A.31}$$

In addition, for any $\delta > 0$, the following convergence result is proven:

$$\lim_{m\to\infty} \limsup_{n\to\infty} P\left(\|\boldsymbol{e}_{n,m} - \boldsymbol{e}_n\| \geq \delta\right) = 0. \tag{A.32}$$

Then, Theorem 4.2 of Billingsley (1968) enables us to conclude that (A.29) holds true.

To show (A.30) and (A.31), we define $\boldsymbol{X}_n^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\Gamma}_{n,m} = (\boldsymbol{u}_n, \boldsymbol{u}_{n-1}, \cdots, \boldsymbol{u}_0, \cdots, \boldsymbol{u}_{-(m-1)})$ and let $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_p)^\top$ satisfy $\|\boldsymbol{\alpha}\| = 1$. Then, one has

$$\boldsymbol{\alpha}^\top \boldsymbol{e}_{n,m} = \boldsymbol{\alpha}^\top \boldsymbol{F}_n^{-1} \sum_{j=-(m-1)}^{n} \boldsymbol{u}_j \eta_j := \sigma \left(\sum_{j=-(m-1)}^{n} \left(\boldsymbol{\alpha}^\top \boldsymbol{F}_n^{-1} \boldsymbol{u}_j\right)^2\right)^{1/2} \sum_{j=-(m-1)}^{n} W_j,$$

where

$$W_j = \frac{\boldsymbol{\alpha}^\top \boldsymbol{F}_n^{-1} \boldsymbol{u}_j}{\sigma \left(\sum\limits_{j=-(m-1)}^{n} \left(\boldsymbol{\alpha}^\top \boldsymbol{F}_n^{-1} \boldsymbol{u}_j\right)^2\right)^{1/2}} \eta_j.$$

It follows from (12), (15) and $\sum_{i=1}^{\infty} |b_i| < \infty$ that

$$0 < \sum_{j=-(m-1)}^{n} \left(\boldsymbol{\alpha}^\top \boldsymbol{F}_n^{-1} \boldsymbol{u}_j\right)^2 \leq \left(\|\boldsymbol{X}\boldsymbol{F}_n^{-1}\| \, \|\boldsymbol{\Sigma}_n^{-1}\| \, \|\boldsymbol{\Gamma}_{n,m}^\top\|\right)^2 < C. \tag{A.33}$$

It is not difficult to see that $\{W_j\}$ is a sequence of independent random variables with $E(W_j) = 0$ and $\sum_{j=-(m-1)}^{n} Var\,(W_j) = 1$. Define F_j and F_j^* as the distribution functions of η_j and W_j. Then, for any $\delta > 0$,

$$\sum_{j=-(m-1)}^{n} \int_{|w|>\delta} w^2 dF_j^*(w) \leq \frac{1}{\sigma^2} \sup_j \int_{|\eta|>\delta\eta_n^*} \eta^2 dF_j(\eta),$$

where $\eta_n^* = \sigma(\sum\limits_{j=-(m-1)}^{n} (\boldsymbol{\alpha}^\top \boldsymbol{F}_n^{-1} \boldsymbol{u}_j)^2)^{1/2} / \sup\limits_j |\boldsymbol{\alpha}^\top \boldsymbol{F}_n^{-1} \boldsymbol{u}_j|$. Letting $\boldsymbol{A}(n)\boldsymbol{\Gamma}_{n,m} =$

$\left\{v_{i,j}^*\right\}$, one has

$$\sup_j |\alpha^\top \boldsymbol{F}_n^{-1}\boldsymbol{u}_j| \le \frac{1}{\sigma^2}\sup_j \left| \sum_{m=1}^{p} \frac{\alpha_m}{\sqrt{\sum_{i=1}^n x_{i,m}^2}} \sum_{i=1}^{n} \frac{v_{n-i+1,j}^*}{d_i} \sum_{l=0}^{i-1} a_{i-1,l}x_{i-l,m} \right|$$
$$\le C \max_{1\leqslant m\leqslant p} \frac{\max_{1\leqslant l\leqslant n}|x_{l,m}|}{\sqrt{\sum_{i=1}^n x_{i,m}^2}} \sup_{0\le i<n} \sum_{l=0}^{i} |a_{i,l}| \sup_j \sum_{i=1}^{n} |v_{i,j}^*|, \quad \text{(A.34)}$$

where d_i is the ith component on the main diagonal of $\boldsymbol{D}_n$. By (C1), Lemma 4 of Berk (1974), (3), (28), and $\sum_{i=1}^{\infty}|b_i|<\infty$, one obtains

$$\lim_{n\to\infty}\max_{1\leqslant m\leqslant p} \frac{\max_{1\leqslant l\leqslant n}|x_{l,m}|}{\sqrt{\sum_{i=1}^n x_{i,m}^2}} = 0, \quad \sup_{0\le i<n}\sum_{l=0}^{i}|a_{i,l}|<\infty, \quad \text{(A.35)}$$

and

$$\sup_j \sum_{i=1}^{n} |v_{i,j}^*| \le \|\boldsymbol{A}(n)\|_1 \|\boldsymbol{\Gamma}_{n,m}\|_1 \le \infty. \quad \text{(A.36)}$$

Then, it follows from (A.33)-(A.36) that $\lim_{n\to\infty}\eta_n^* = \infty$. By the Lindeberg-Feller central limit theorem, one gets $\sum_{j=-(m-1)}^{n} W_j \xrightarrow{d} \mathcal{N}(0,1)$ as $n\to\infty$ for each m. Thus, $\boldsymbol{\alpha}^\top \boldsymbol{e}_{n,m}$ has a limiting normal distribution with zero mean and variance

$$\lim_{n\to\infty}\boldsymbol{\alpha}^\top E\left(\boldsymbol{e}_{n,m}\boldsymbol{e}_{n,m}^\top\right)\boldsymbol{\alpha} = \sigma^2 \lim_{n\to\infty}\sum_{j=-(m-1)}^{n}\left(\boldsymbol{\alpha}^\top \boldsymbol{F}_n^{-1}\boldsymbol{u}_j\right)^2 < \infty.$$

Note that $\lim_{n\to\infty} E\left(\boldsymbol{\epsilon}_{n,m}\boldsymbol{\epsilon}_{n,m}^\top\right) = \boldsymbol{\Sigma}_n$. This, together with (A.28), implies that

$$\lim_{m\to\infty}\lim_{n\to\infty} E\left(\boldsymbol{e}_{n,m}\boldsymbol{e}_{n,m}^\top\right) = \lim_{n\to\infty}\lim_{m\to\infty} E\left(\boldsymbol{e}_{n,m}\boldsymbol{e}_{n,m}^\top\right) = \lim_{n\to\infty}\boldsymbol{F}_n^{-1}\boldsymbol{X}_n^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{X}_n\boldsymbol{F}_n^{-1}$$
$$= \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{f_\varepsilon(\lambda)}d\boldsymbol{M}(\lambda).$$

Consequently, the proofs of (A.30) and (A.31) are complete.

To show (A.32), Chebyshev's inequality tells us that for $\delta>0$,

$$P\left(\|\boldsymbol{e}_{n,m}-\boldsymbol{e}_n\|\ge\delta\right) \le \frac{E\left(\|\boldsymbol{e}_{n,m}-\boldsymbol{e}_n\|^2\right)}{\delta^2}. \quad \text{(A.37)}$$

Since (A.28) and (A.37) imply that as m and n increase to infinity,

$$E\left(\|\boldsymbol{e}_{n,m}-\boldsymbol{e}_n\|^2\right) = \operatorname{tr}\left\{\boldsymbol{F}_n^{-1}\boldsymbol{X}_n^\top\boldsymbol{\Sigma}_n^{-1}E\left(\left(\boldsymbol{\epsilon}_n-\boldsymbol{\epsilon}_{n,m}\right)\left(\boldsymbol{\epsilon}_n-\boldsymbol{\epsilon}_{n,m}\right)^\top\right)\boldsymbol{\Sigma}_n^{-1}\boldsymbol{X}_n\boldsymbol{F}_n^{-1}\right\}$$
$$= \operatorname{tr}\left\{\boldsymbol{F}_n^{-1}\boldsymbol{X}_n^\top\boldsymbol{\Sigma}_n^{-1}\left(\boldsymbol{\Sigma}_n - E\left(\boldsymbol{\epsilon}_{n,m}\boldsymbol{\epsilon}_{n,m}^\top\right)\right)\boldsymbol{\Sigma}_n^{-1}\boldsymbol{X}_n\boldsymbol{F}_n^{-1}\right\} = o(1),$$

the proof is complete.

Acknowledgements

The author is grateful to Prof. Ching-Kang Ing for stimulating discussions and many helpful suggestions, and thanks the anonymous referee for very helpful comments.

References

Amemiya, T. (1973). Generalized least squares with an estimated autocovariance matrix. *Econometrica* **41**, 723-732.

Anderson, T. W. (1994). *The Statistical Analysis of Time Series.* Wiley Classics Library Edition Published, New York.

Berk, K. (1974). Consistent autoregressive spectral estimates. *The Annals of Statistics* **2**, 489-502.

Bickel, P. J. and Levina, E. (2008). Regularized estimation of large covariance matrices. *The Annals of Statistics* **36**, 199-227.

Bloomfield, P. (1992). Trends in global temperature. *Climatic Change* **21**, 1-16.

Brillinger, D. R. (2001). *Time Series*: *Data Analysis and Theory.* Society for Industrial and Applied Mathematics, Philadelphia.

Bryant, P. G. and Smith, M. A. (1995). *Practical Data Analysis*: *Case Studies in Business Statistics*, 2nd edition. Volume 1. Richard D. Irwin, Inc., Chicago.

Chan, J. S. K. and Choy, S. T. B. (2008). Analysis of Covariance Structures in Time Series. *Journal of Data Science* **6**, 573-589.

Durbin, J. and Watson, G. S. (1951). Testing for serial correlation in least squares regression. II. *Biometrika* **38**, 159-177.

Fuller, W. A. (1996). *Introduction to Statistical Time Series*, 2nd edition. Wiley-Interscience, New York.

Furrer, R. and Bengtsson, T. (2007). Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants. *Journal of Multivariate Analysis* **98**, 227-255.

Grenander, U. (1954). On the estimation of regression coefficients in the case of an autocorrelated disturbance. *The Annals of Mathematical Statistics* **25**, 252-272.

Grenander, U. and Rosenblatt, M. (1957). Rrgression analysis of time series with stationary residuals. *Proceedings of the National Academy of Sciences of the United States of America* **40**, 812-816.

Horn, R. A. and Johnson, C. R. (1990). *Matrix Analysis.* Corrected reprint of the 1985 original. Cambridge University Press, Cambridge.

Ing, C. K. and Wei, C. Z. (2003). On same-realization prediction in an infinite-order autoregressive process. *Journal of Multivariate Analysis* **85**, 130-155.

Koreisha, S. G. and Fang, Y. (2001). Generalized least squares with misspecified serial correlation structures. *Journal of the Royal Statistical Society, Series B* **63**, 515-531.

McMurry, T. L. and Politis, D. N. (2010). Banded and tapered estimates for autocovariance matrices and the linear process bootstrap. *Journal of Time Series Analysis* **31**, 471-482.

Wei, C. Z. (1987). Adaptive prediction by least squares predictors in stochastic regression models with application to time series. *The Annals of Statistics* **15**, 1667-1682.

Wei, W. S. W. (2006). *Time Series Analysis*: *Univariate and Multivariate Methods*, 2nd edition. Pearson Addison-Wesley, New York.

Wu, W. B. (2011). Asymptotic theory for stationary processes. *Statistics and Its Interface* **4**, 207-226.

Wu, W. B. and Pourahmadi, M. (2003). Nonparametric etimation of large covariance matrices of longitudinal data. *Biometrika* **90**, 831-844.

Wu, W. B. and Pourahmadi, M. (2009). Banding sample autocovariance matrices of stationary processes. *Statistica Sinica* **19**, 1755-1768.

Xiao, H. and Wu, W. B. (2012). Covariance matrix estimation for stationary time series. *The Annals of Statistics* **40**, 466-493.

Received March 4, 2012; accepted June 19, 2012.

Chiao-Yi Yang
Institute of Statistical Science
Academia Sinica
No. 128, Sec. 2, Academia Rd., Nankang, Taipei 115, Taiwan
chiaoyiy@stat.sinica.edu.tw