# Bootstrap inference for network vector autoregression in large-scale social network

Manho Hong and Eunju Hwang[1]

Department of Applied Statistics, Gachon University

**Abstract:** A large amount of online social network data such as Facebook or Twitter are extensively generated by the growth of social network flatforms in recent years. Developing a time series model and its statistical inference in the large-scale social network are as important as the rapid progress on the social network technology and evolution. In this work we consider a network vector autoregression for large-scale social network and study the bootstrap estimation applying both the stationary bootstrap and the classical residual bootstrap. In the network vector autoregression with $N$ users tending to infinity, the two kinds of bootstrap methods are employed to construct the bootstrap version of the least sqaures estimator. The notion of the strict stationarity of the model with $N$ users tending to infinity is used to generate the stationary bootstrap sample with sample size $T$, while both $N$ and $T$ are adopted to obtain classical i.i.d. residual bootstrap sample. Consistency of the bootstrap estimators is established and bootstrap confidence intervals are discussed. A simulation study is given to demonstrate the better performance of the bootstrap inference than the existing ordinary least squares type estimators.

## 1 Introduction

A large-scale social network such as Facebook or Twitter has been immensely used in recent years and a huge amount of online social network data are created, which have attention from researchers and are used for the big-data analysis. For example, worldwide there are over 2.50 billion monthly active users for December 2019 on Facebook, and their online data provide a marketing source or a poll survey. Due to the popular social network systems along with rapid developments of network technology, meaningful analysis on vast sets of information should be done well in academia and practical areas. As a statistical aspect, it is very significant that one should develop suitably time series models for representing the large-scale social network and establish optimally their statistical inference in order to predict and offer beneficial influences to the future.

For this purpose many researchers have studied network models for the large-scale social network system. Zhu et al. (2017) proposed a network vector autoregressive model with connected network

[1] Corresponding author. Mailing Address: Dept. of Applied Statistics, Gachon University, South Korea. Tel: +82-31-750-5373; *E-mail address*: ehwang@gachon.ac.kr

structure. Wei and Tian (2018) considered multiple connections among $N$ nodes as an extension of Zhu et al. (2017) and studied the heterogeneous effects in the network regression. As seen in Zhu et al. (2017) and Wei and Tian (2018), to describe relational data in the social network connection, it is often to use a binary matrix whose entries represent link between pairs of nodes. An adjacency binary matrix $A = (a_{ij}) \in \mathbb{R}^{N \times N}$ describes the social relationship: $a_{ij} = 1$ if there exists a social (directed) relation from $i$ to $j$ and $a_{ij} = 0$ otherwise, where $i, j \in \{1, 2, \dots, N\}$ are nodes in the network of the size $N$. If undirected relations are considered, the matrix is symmetric, i.e., $A = A^\top$. For symmetric binary matrices with time varying relational structure, we refer to Durante and Dunson (2014) who proposed a nonparametric Bayesian dynamical model along with similarity measure to reduce dimensionality. As a recent work we refer to Zhu et al. (2020), who studied a multivariate spatial autoregressive model for large social networks. Can and Alatas (2019) have presented a new direction of online social network analysis problems as well as related applications.

In this paper we consider the network vector autoregression of Zhu et al. (2017), where continuous responses are observed at equally spaced time points from different $N$ nodes, (i.e., users in the network), and the response at a given time point comprises of a linear combination of four components: its previous value, the average of its connected neighbors, a set of node-specific covariates and an independent noise. Zhu et al. (2017) investigated the strict stationarity of the model as the number of nodes tends to infinity and developed asymptotic properties of ordinary least squares type estimator.

As an alternative of the estimator in Zhu et al. (2017), we propose a bootstrap least squares estimator by adopting the stationary bootstrap method of Politis and Romano (1994). More particularly, the stationary bootstrap and the i.i.d. residual bootstrap are combined to construct the bootstrap least squares estimator and its asymptotics are established as both network size $N$ and sample size $T$ tend to infinity. The stationary bootstrap is a very powerful nonparametric technique with random block-length, applicable to stationary weakly-dependent time series data. See Politis and Romano (1994) and Hwang and Shin (2012), references therein. Hwang and Shin (2013) also have applied both the stationary bootstrap and the classical residual bootstrap to construct a bootstrap estimator for the $p$th order autoregressive models. However, the present work following the model of Zhu et al. (2017) takes $N$-dimensional time series with $N \to \infty$ into consideration unlike Hwang and Shin (2013). The $N$-dimensional time series with tending to infinity is applied to bootstrapping in the network vector autoregression by using the strictly stationarity which has been defined by Zhu et al. (2017) for the $N$-dimensional time series with $N \to \infty$. The classical residual bootstrap is computed from the empirical distribution function of residuals with both the network size and the sample size adopted, while the stationary bootstrap sample is obtained only with the sample size $T$ in the stationary bootstrap procedure, using the notion of the strictly stationarity. A main result is that the consistency of the bootstrap least squares estimator is proven and the proposed bootstrap method has better finite sample performance than the ordinary least squares estimator, which is shown in a simulation study.

The remainder of the paper is organized as follows: In Section 2 the network vector autoregression is described and in Section 3 the bootstrap algorithm is stated and the main theory is presented. Section 4 gives a simulation study while Section 5 provide technical lemmas and proofs.

## 2 Network vector autoregression

We consider a network vector autoregressive model $\{Y_{it}\}$ of Zhu et al. (2017) for a large-scale social network with $N$ nodes, given by

$$Y_{it} = \beta_0 + Z_i^\top \gamma + \beta_1 n_i^{-1} \sum_{j=1}^{N} a_{ij} Y_{j(t-1)} + \beta_2 Y_{i(t-1)} + \varepsilon_{it}, \quad 1 \le i \le N,\ 1 \le t \le T \tag{1}$$

where $Z_i = (Z_{i1}, \dots, Z_{iq})^\top \in \mathbb{R}^q$ is a $q$-dimensional node-specific random vector, $\gamma = (\gamma_1, \dots, \gamma_q)^\top \in \mathbb{R}^q$ is nodal-effect associated coefficient, $\beta_0, \beta_1, \beta_2$ are parameters, $n_i = \sum_{j \ne i} a_{ij}$ is the total number of nodes that $i$ follows, and $\{\varepsilon_{it}\}$ are i.i.d. error terms with mean zero and variance $\sigma_\varepsilon^2$. Response $Y_{it}$ at time $t$ is a linear combination of four components: its previous value, the average of its connected neighbors, a set of node-specific covariates and an independent noise. The corresponding coefficients are referred to as the momentum effect, the network effect and the nodal effect, respectively.

In this work we consider ultra-high dimensional time series $\mathbb{Y}_t := (Y_{1t}, \dots, Y_{Nt})^\top$ of $N$-dimensional vector, $1 \le t \le T$ with $T \to \infty$ and $N \to \infty$. As mentioned in Zhu et al. (2017), it is remarkable that the dimension of $\mathbb{Y}_t$ is diverging and so it needs to define clearly the strictly stationarity of $\mathbb{Y}_t$. According to Zhu et al. (2017), $\{\mathbb{Y}_t\}$ is said to be strictly stationary if it satisfies the followings: for any $\omega \in \mathcal{W} := \{\omega \in \mathbb{R}^\infty : \omega = (\omega_1, \omega_2, \dots)^\top, \sum |\omega_i| < \infty\}$, letting $\mathbf{w}_N = (\omega_1, \dots, \omega_N)^\top \in \mathbb{R}^N$ be the truncated $N$-dimensional vector, (i) $Y_t^\omega = \lim_{N\to\infty} \mathbf{w}_N^\top \mathbb{Y}_t$ exists in the almost sure sense; and (ii) $\{Y_t^\omega\}$ is strictly stationary.

To describe the social network structure, an adjacency matrix $A = (a_{ij}) \in \mathbb{R}^{N\times N}$ is adopted with $a_{ij} = 1$ if there exists a social relationship from $i$ to $j$, (that is, user $i$ follows user $j$ on a social network system such as Twitter), and $a_{ij} = 0$ otherwise. We assume $a_{ii} = 0$ and the adjacency matrix $A$ is nonrandom. Recently, Zhu et al. (2017) developed an ordinary least squares type estimator in model (1) and investigated its asymptotic properties under some technical conditions such as moment conditions and Markov chain properties. Here we omit the technical conditions, which are needed in proving the asymptotic normality of the ordinary least squares estimator. We defer to (C1)–(C3) on p. 1101–1102 of Zhu et al. (2017) for detailed technical conditions because of the lengthy.

We rewrite model (1) as follows:

$$Y_{it} = \beta_0 + \beta_1 w_i^\top \mathbb{Y}_{t-1} + \beta_2 Y_{i(t-1)} + Z_i^\top \gamma + \varepsilon_{it} = X_{i(t-1)}^\top \theta + \varepsilon_{it}$$

where

$$w_i = \left(\frac{a_{i1}}{n_i}, \frac{a_{i2}}{n_i}, \ldots, \frac{a_{iN}}{n_i}\right)^\top \in \mathbb{R}^N, \qquad \mathbb{Y}_{t-1} = \big(Y_{1(t-1)}, Y_{2(t-1)}, \ldots, Y_{N(t-1)}\big)^\top \in \mathbb{R}^N$$

$$X_{i(t-1)} = (1, w_i^\top \mathbb{Y}_{t-1}, Y_{i(t-1)}, Z_i^\top)^\top \in \mathbb{R}^{q+3}, \qquad \theta = (\beta_0, \beta_1, \beta_2, \gamma^\top)^\top \in \mathbb{R}^{q+3}.$$

Denote

$$\mathbb{X}_t = (X_{1t}, X_{2t}, \ldots, X_{Nt})^\top \in \mathbb{R}^{N\times(q+3)}, \qquad \mathcal{E}_t = (\varepsilon_{1t}, \varepsilon_{2t}, \ldots, \varepsilon_{Nt})^\top \in \mathbb{R}^N$$

and write the matrix form with $i = 1, 2, \ldots, N$ in rows as $\mathbb{Y}_t = \mathbb{X}_{t-1}\theta + \mathcal{E}_t$. The OLSE is given by

$$\widehat{\theta} = \left(\sum_{t=1}^{T} \mathbb{X}_{t-1}^\top \mathbb{X}_{t-1}\right)^{-1} \left(\sum_{t=1}^{T} \mathbb{X}_{t-1}^\top \mathbb{Y}_t\right) = \left(\sum_{t=1}^{T}\sum_{i=1}^{N} X_{i(t-1)} X_{i(t-1)}^\top\right)^{-1} \left(\sum_{t=1}^{T}\sum_{i=1}^{N} X_{i(t-1)} Y_{it}\right) \tag{2}$$

Under the stationary condition $|\beta_1| + |\beta_2| < 1$ and some technical conditions, Zhu et al. (2017) addressed the asymptotic normality in the following:

$$\sqrt{NT}(\widehat{\theta} - \theta) \xrightarrow{\mathrm{d}} N(0, \sigma_\varepsilon^2 \Sigma^{-1})$$

as $\min\{N, T\} \to \infty$ where $\Sigma$ is given by (2.10) on p.1103 of Zhu et al. (2017)

$$\Sigma = \begin{pmatrix} 1 & c_\beta & c_\beta & \mathbf{0}^\top \\ c_\beta & \Sigma_1 & \Sigma_2 & \kappa_8 \gamma^\top \Sigma_z \\ c_\beta & \Sigma_2 & \Sigma_3 & \kappa_3 \gamma^\top \Sigma_z \\ \mathbf{0} & \kappa_8 \gamma^\top \Sigma_z & \kappa_3 \gamma^\top \Sigma_z & \Sigma_z \end{pmatrix}, \tag{3}$$

$c_\beta = \beta_0(1-\beta_1-\beta_2)^{-1}, \Sigma_1 = c_\beta^2 + \kappa_5\gamma^\top\Sigma_z\gamma + \kappa_6, \Sigma_2 = c_\beta^2 + \kappa_7\gamma^\top\Sigma_z\gamma + \kappa_2, \Sigma_3 = c_\beta^2 + \kappa_4\gamma^\top\Sigma_z\gamma + \kappa_1$, and $\mathbf{0} = (0, ..., 0)^\top$ is a vector with compatible dimension.

In the next section, its bootstrap version is constructed and its consistency is established.

# 3 Stationary bootstrap and main result

In this section we apply the stationary bootstrap method to stationary weakly-dependent time series $\{\mathbb{Y}_t : 0 \le t \le T-1\}$, (or $\{Y_t^\omega : 0 \le t \le T-1\}$ for any, but fixed, $\omega \in \mathcal{W}$).

In order to generate the stationary bootstrap sample, first we define an extension $\{\mathbb{Y}_{T,j} : j \ge 0\}$ by a periodic extension of the observed data set: for each $j \ge 0$, define $\mathbb{Y}_{T,j} := \mathbb{Y}_t$ where $t$ is such that $j = qT + t$ for some $q$ and $0 \le t < T$. In this way we write $\mathbb{Y}_j$ for all $j \in \{0, 1, 2, \dots\}$ by omitting the subscript $T$ in $\mathbb{Y}_{T,j}$ for just notational simplicity, and so do other vectors such as $\mathbb{X}_j$ and $Y_j^\omega$.

Next, for $\ell \ge 0$ and $0 \le t \le T-1$, define the blocks $B(t,\ell) = \{\mathbb{Y}_t, \dots, \mathbb{Y}_{t+\ell-1}\}$ consisting of $\ell$ observations starting from $\mathbb{Y}_t = \mathbb{Y}_{T,t}$. Bootstrap observations under the stationary bootstrap method are obtained by selecting a random number of blocks from collection $\{B(t,\ell) : 0 \le t \le T-1,\, \ell \ge 0\}$. To do this, we generate random variables $I_1, I_2, \dots$ and $L_1, L_2, \dots$ as follows: (i) $I_1, I_2, \dots$ are i.i.d. discrete uniform on $\{0, \dots, T-1\}$: $P(I_1 = t) = \frac{1}{T}$, for $t = 0, 1, \dots, T-1$, (ii) $L_1, L_2, \dots$ are i.i.d. random variables having the geometric distribution with a parameter $\rho \in (0,1)$: $P(L_1 = \ell) = \rho(1-\rho)^{\ell-1}$ for $\ell = 1, 2, \dots$, where $\rho = \rho_T$ depends on the sample size $T$ and (iii) the collections $\{I_1, I_2, \dots\}$ and $\{L_1, L_2, \dots\}$ are independent.

Now, let $\kappa = \inf\{k \ge 1 : L_1 + \cdots + L_k \ge T\}$ and select $\kappa$ blocks $B(I_1, L_1), \dots, B(I_\kappa, L_\kappa)$. Note that there are $L_1 + \cdots + L_\kappa$ elements in the resampled blocks $B(I_1, L_1), \dots, B(I_\kappa, L_\kappa)$. Arranging these elements in a series and deleting the last $L_1 + \cdots + L_\kappa - T$ elements, we get the stationary bootstrap observations $\mathbb{Y}_1^*, \dots, \mathbb{Y}_T^*$. Conditional on $\{\mathbb{Y}_1, \dots, \mathbb{Y}_T\}$, the process $\{\mathbb{Y}_t^* : t = 0, 1, 2, \dots\}$ is stationary.

The following algorithm gives some steps for computing the bootstrap least squares estimator. Note that applying the stationary bootstrap to $\{\mathbb{Y}_t\}$ of $N$-dimensional vectors with $N \to \infty$ is the same as applying to $\{Y_t^\omega\}$, for any (but fixed) $\omega \in \mathcal{W}$, which is strictly stationary in Section 2, because the same uniform and the same geometric random variables in the blocks are used to construct $\{\mathbb{Y}_t^*\}$ and $\{Y_t^{\omega *}\}$ in the bootstrap procedure.

## Algorithm.

Step 1. We compute the OLSE $\hat\theta$ in (2)

Step 2. The OLSE-residuals $\hat\varepsilon_{it}$ are computed as follows:

$$\hat\varepsilon_{it} = Y_{it} - X_{i(t-1)}^\top \hat\theta, \qquad 1 \le i \le N,\ 1 \le t \le T$$

Let $\hat{F}_\varepsilon(\cdot)$ be the empirical distribution function of the residuals, i.e.,

$$\hat{F}_\varepsilon(x) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{I}_{(-\infty,\hat{\varepsilon}_{it}]}(x).$$

We draw an i.i.d array of $\{\hat{\varepsilon}^*_{it} : 1 \le i \le N,\ 1 \le t \le T\}$ from the empirical distribution $\hat{F}_\varepsilon$.

Step 3. We apply the stationary bootstrap method to $\{\mathbb{Y}_t : 0 \le t \le T-1\}$ to obtain stationary bootstrap observations $\{\mathbb{Y}^*_t : 0 \le t \le T-1\}$ with $\mathbb{Y}^*_t = (Y^*_{1t}, \ldots, Y^*_{Nt})^\top$. From these, for $t = 0, 1, \ldots, T-1$ we define $\mathbb{X}^*_t = (X^*_{1t}, \ldots, X^*_{Nt})^\top$ with $X^*_{it} = (1, w_i^\top \mathbb{Y}^*_t, Y^*_{it}, Z_i^\top)^\top$, $i = 1, 2, \ldots, N$.

Step 4. We use $\{\hat{\varepsilon}^*_{it} : 1 \le i \le N,\ 1 \le t \le T\}$ in Step 2 and $\{X^{*\top}_{i(t-1)} : 1 \le i \le N,\ 1 \le t \le T\}$ in Step 3 to compute $\tilde{Y}^*_{it} = X^{*\top}_{i(t-1)}\hat{\theta} + \hat{\varepsilon}^*_{it}$ for $i = 1, 2, \ldots, N;\ t = 1, 2, \ldots, T$.

Step 5. We construct the bootstrap estimator $\hat{\theta}^*$ of $\theta$ by

$$\widehat{\theta}^* = \left(\sum_{t=1}^{T}\mathbb{X}^{*\top}_{t-1}\mathbb{X}^*_{t-1}\right)^{-1}\left(\sum_{t=1}^{T}\mathbb{X}^{*\top}_{t-1}\tilde{\mathbb{Y}}^*_t\right) = \left(\sum_{t=1}^{T}\sum_{i=1}^{N}X^*_{i(t-1)}X^{*\top}_{i(t-1)}\right)^{-1}\left(\sum_{t=1}^{T}\sum_{i=1}^{N}X^*_{i(t-1)}\tilde{Y}^*_{it}\right)$$

where $\tilde{\mathbb{Y}}^*_t = (\tilde{Y}^*_{1t}, \ldots, \tilde{Y}^*_{Nt})^\top$.

In the following, $P^*, E^*, Var^*$ denote the conditional probability, the conditional expectation, and the conditional variance respectively, given $\{\mathbb{Y}_0, \ldots, \mathbb{Y}_{T-1}\}$. Note that the classical residual bootstrap is applied to $\{\varepsilon_{it}\}$ for all $i$ and $t$, whereas the stationary bootstrap to $\{\mathbb{Y}_t\}$ for $t$, but it is independent of user $i$. For this reason we have $Z_i^\top$ itself in the fourth component of $X^*_{it}$ in Step 3, rather than the bootstrap version of $Z_i^\top$.

**Theorem 3.1** *Model (1) is considered with* $|\beta_1| + |\beta_2| < 1$. *Assume that technical conditions (C1)–(C3) on p. 1101–1102 of Zhu et al. (2017) are fulfilled. If* $\min\{N, T\} \to \infty$ *and* $T\rho \to \infty$ *then we have*

$$\sup_{\mathbf{x}}\left|P^*\left(\sqrt{NT}[\hat{\theta}^* - \hat{\theta}] \le \mathbf{x}\right) - P\left(\sqrt{NT}[\hat{\theta} - \theta] \le \mathbf{x}\right)\right| \xrightarrow{\text{p}} 0.$$

According to Theorem 3.1, an $(1-\alpha)\%$ bootstrap confidence interval (CI) for each parameter $\theta$ is constructed as follows: $1 - \alpha =$

$$P^*(q^*_{\alpha/2} - \hat{\theta} \le \hat{\theta}^* - \hat{\theta} \le q^*_{1-\alpha/2} - \hat{\theta}) \approx P(q^*_{\alpha/2} - \hat{\theta} \le \hat{\theta} - \theta \le q^*_{1-\alpha/2} - \hat{\theta}) = P(2\hat{\theta} - q^*_{1-\alpha/2} \le \theta \le 2\hat{\theta} - q^*_{\alpha/2})$$

where $q^*_{\alpha/2},\ q^*_{1-\alpha/2}$ are, respectively, the $\alpha/2$-th, $(1-\alpha/2)$-th quantiles of the bootstrap estimates $\hat{\theta}^*$ of $\theta$.

## 4 A Monte-Carlo Study

We conduct a simulation study to illustrate the performance of the proposed bootstrap estimator. In order to compare the bootstrap estimates with the existing ones, we adopt the simulation-model of Zhu et al. (2017). Example 2 with stochastic block model of Nowicki and Snijders (2001) in Zhu et al. (2017) is used here to see the comparison. Performances such as the empirical coverage probability and the average length of bootstrap CIs are compared with those using the normal approximation of the OLSE in Zhu et al. (2017).

The stochastic block model of Nowicki and Snijders (2001) is of particular interest for community detection as seen in Zhao et al. (2012). In the network model of Nowicki and Snijders (2001), $K$ blocks are considered for community connections and each node is randomly assigned to a block label $k \in \{1, 2, \ldots, K\}$ with equal probability. If user $i$ and user $j$ belong to the same block, then set $P(a_{ij} = 1) = 0.3N^{-0.3}$; otherwise $P(a_{ij} = 1) = 0.3N^{-1}$. It indicates that more likely to be connected if nodes are within the same block.

We adopt $\gamma = (-0.5, 0.3, 0.8, 0, 0)^\top$, $\beta = (\beta_0, \beta_1, \beta_2)^\top = (0, 0.1, -0.2)^\top$, and $\mathbb{Z} = (Z_1, Z_2, \ldots, Z_N)^\top \in \mathbb{R}^{N\times q}$ where $Z_i = (Z_{i1}, Z_{i2}, \ldots, Z_{iq})^\top \in \mathbb{R}^q$ with $q = 5$ is from a multivariate normal distribution with mean $\mathbf{0}$ and covariance $\Sigma_Z = (\sigma_{j_1,j_2})$ where $\sigma_{j_1,j_2} = 0.5^{|j_1-j_2|}$. An initial value $\mathbb{Y}_0$ is randomly simulated according to Proposition 1 of Zhu et al. (2017), following a normal distribution with mean and covariance given by $\mu = (I-G)^{-1}\mathcal{B}_0$, $\mathrm{vec}\{\Gamma(0)\} = \sigma^2(I-G\otimes G)^{-1}\mathrm{vec}(I)$ where $\mathcal{B}_0 = \beta_0\mathbf{1}+\mathbb{Z}\gamma$, $G = \beta_1 W + \beta_2 I$, $W = \mathrm{diag}(n_1^{-1}, \ldots, n_N^{-1})A$. For the stationary bootstrap setting, we use the geometric parameter $\rho = 0.05(T/100)^{-1/3}$. $N = 100, 200$; $T = 50, 100, 200$; $K = 5, 10, 20$ are used in the experiment. In Tables 1 and 2, which have network size 100, 200, respectively, empirical coverage probability and average length of 90% CIs are displayed. 1000 independent replications are used.

**Table 1.** Coverage probability(%) and average length of 90% CIs for $N = 100$

| K | T | $\beta_0$ | | $\beta_1$ | | $\beta_2$ | | $\gamma_1$ | | $\gamma_2$ | | $\gamma_3$ | | $\gamma_4$ | | $\gamma_5$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 0 | | 0.1 | | $-0.2$ | | $-0.5$ | | 0.3 | | 0.8 | | 0 | | 0 | |
| | | NA | SB | NA | SB | NA | SB | NA | SB | NA | SB | NA | SB | NA | SB | NA | SB |
| 5 | 50 | 94.1 | 70.8 | 94.3 | 93.6 | 93.7 | 90.2 | 95.0 | 93.8 | 94.3 | 94.3 | 93.9 | 93.1 | 94.8 | 94.2 | 95.5 | 95.5 |
| | | .051 | 3.49 | .051 | .055 | .044 | .070 | .060 | .062 | .063 | .065 | .068 | .078 | .059 | .061 | .055 | .056 |
| 5 | 100 | 94.7 | 94.5 | 94.4 | 94.0 | 94.1 | 92.6 | 95.2 | 95.0 | 93.9 | 93.9 | 93.5 | 93.2 | 95.6 | 95.3 | 95.3 | 95.4 |
| | | .033 | .034 | .034 | .038 | .031 | .050 | .041 | .046 | .046 | .047 | .047 | .055 | .043 | .044 | .039 | .039 |
| 5 | 200 | 95.2 | 95.3 | 94.9 | 95.2 | 94.7 | 94.5 | 95.7 | 95.8 | 94.3 | 94.5 | 94.0 | 94.2 | 96.1 | 96.3 | 95.3 | 95.7 |
| | | .025 | .022 | .024 | .020 | .020 | .028 | .031 | .028 | .031 | .030 | .038 | .033 | .031 | .031 | .028 | .028 |
| 10 | 50 | 95.2 | 69.2 | 95.3 | 95.0 | 93.9 | 92.2 | 94.0 | 93.6 | 95.0 | 95.0 | 95.2 | 94.7 | 96.0 | 96.1 | 93.7 | 93.9 |
| | | .048 | 3.70 | .046 | .049 | .046 | .070 | .059 | .064 | .065 | .066 | .068 | .077 | .062 | .063 | .056 | .054 |
| 10 | 100 | 95.1 | 94.6 | 95.9 | 95.5 | 95.2 | 94.8 | 94.3 | 95.6 | 95.3 | 95.5 | 94.2 | 94.0 | 94.9 | 94.1 | 95.0 | 95.2 |
| | | .034 | .036 | .031 | .034 | .051 | .041 | .044 | .046 | .046 | .046 | .050 | .057 | .044 | .045 | .041 | .040 |
| 10 | 200 | 95.3 | 95.7 | 95.4 | 95.6 | 94.9 | 94.9 | 94.7 | 95.0 | 96.0 | 96.1 | 94.6 | 94.5 | 95.1 | 95.5 | 95.7 | 95.7 |
| | | .024 | .023 | .025 | .022 | .035 | .022 | .032 | .030 | .034 | .033 | .040 | .035 | .032 | .031 | .028 | .028 |
| 20 | 50 | 94.2 | 69.9 | 95.3 | 94.9 | 94.9 | 93.3 | 94.6 | 93.9 | 94.1 | 93.1 | 94.6 | 92.9 | 95.4 | 95.5 | 94.7 | 94.5 |
| | | .048 | 3.66 | .045 | .050 | .045 | .072 | .058 | .061 | .065 | .066 | .067 | .080 | .066 | .066 | .056 | .057 |
| 20 | 100 | 95.4 | 95.5 | 94.1 | 94.0 | 95.9 | 95.8 | 95.0 | 95.2 | 94.1 | 94.2 | 96.8 | 97.7 | 95.0 | 95.2 | 95.2 | 95.3 |
| | | .033 | .031 | .031 | .034 | .032 | .048 | .041 | .046 | .044 | .047 | .048 | .057 | .044 | .044 | .040 | .041 |
| 20 | 200 | 95.4 | 95.5 | 95.1 | 95.2 | 95.7 | 97.7 | 95.4 | 98.0 | 94.7 | 95.3 | 96.8 | 97.0 | 96.6 | 97.0 | 95.5 | 95.9 |
| | | .024 | .022 | .026 | .023 | .035 | .023 | .032 | .030 | .033 | .032 | .038 | .034 | .032 | .032 | .030 | .029 |

**Table 2.** Coverage probability(%) and average length of 90% CIs for $N = 200$

| K | T | $\beta_0$ | | $\beta_1$ | | $\beta_2$ | | $\gamma_1$ | | $\gamma_2$ | | $\gamma_3$ | | $\gamma_4$ | | $\gamma_5$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 0 | | 0.1 | | $-0.2$ | | $-0.5$ | | 0.3 | | 0.8 | | 0 | | 0 | |
| | | NA | SB | NA | SB | NA | SB | NA | SB | NA | SB | NA | SB | NA | SB | NA | SB |
| 5 | 50 | 94.8 | 70.1 | 94.3 | 93.6 | 93.7 | 90.2 | 95.0 | 93.8 | 94.3 | 94.3 | 93.9 | 93.1 | 94.8 | 94.2 | 95.5 | 95.5 |
| | | .033 | 3.67 | .037 | .041 | .032 | .055 | .042 | .046 | .043 | .044 | .047 | .056 | .041 | .042 | .039 | .039 |
| 5 | 100 | 95.1 | 93.8 | 94.9 | 94.6 | 94.3 | 92.9 | 93.6 | 94.2 | 95.1 | 95.3 | 94.6 | 94.8 | 95.2 | 95.4 | 96.3 | 95.9 |
| | | .025 | .030 | .027 | .030 | .023 | .037 | .029 | .030 | .031 | .032 | .035 | .041 | .031 | .032 | .028 | .028 |
| 5 | 200 | 96.3 | 96.1 | 95.2 | 95.6 | 95.7 | 96.0 | 95.2 | 95.4 | 96.1 | 96.3 | 95.1 | 95.8 | 96.3 | 96.7 | 97.2 | 97.5 |
| | | .016 | .015 | .020 | .016 | .026 | .016 | .023 | .022 | .023 | .022 | .028 | .025 | .022 | .021 | .019 | .019 |
| 10 | 50 | 93.9 | 72.9 | 94.2 | 92.1 | 94.1 | 92.9 | 94.7 | 94.1 | 93.8 | 93.5 | 95.2 | 94.8 | 94.3 | 92.9 | 95.9 | 95.0 |
| | | .033 | 3.62 | .034 | .037 | .031 | .056 | .042 | .046 | .043 | .046 | .048 | .055 | .043 | .043 | .039 | .040 |
| 10 | 100 | 94.3 | 94.0 | 95.2 | 93.9 | 95.2 | 94.8 | 95.1 | 95.0 | 94.7 | 94.1 | 95.3 | 95.6 | 95.2 | 95.0 | 97.0 | 96.9 |
| | | .023 | .026 | .022 | .037 | .030 | .035 | .031 | .032 | .034 | .038 | .031 | .031 | .027 | .028 | .031 | .033 |
| 10 | 200 | 94.5 | 94.9 | 96.8 | 95.5 | 95.8 | 96.2 | 94.7 | 95.6 | 95.2 | 95.2 | 94.6 | 97.0 | 95.3 | 95.4 | 95.5 | 95.3 |
| | | 017 | .015 | .018 | .016 | .027 | .016 | .022 | .020 | .023 | .023 | .028 | .024 | .023 | .023 | .021 | .021 |
| 20 | 50 | 94.3 | 66.6 | 94.3 | 94.0 | 95.2 | 95.0 | 93.6 | 92.8 | 93.2 | 92.9 | 93.5 | 93.4 | 93.7 | 93.2 | 94.1 | 93.8 |
| | | .034 | 3.51 | .032 | .033 | .033 | .055 | .040 | .045 | .046 | .046 | .046 | .055 | .044 | .045 | .039 | .039 |
| 20 | 100 | 95.2 | 95.0 | 95.1 | 94.8 | 93.8 | 92.8 | 94.9 | 94.2 | 93.9 | 93.5 | 95.2 | 95.0 | 94.8 | 94.7 | 93.8 | 93.7 |
| | | .021 | .023 | .024 | .022 | .023 | .038 | .030 | .032 | .032 | .033 | .033 | .038 | .031 | .032 | .027 | .028 |
| 20 | 200 | 96.7 | 97.7 | 95.8 | 96.7 | 96.8 | 97.1 | 95.9 | 96.7 | 97.1 | 98.0 | 97.2 | 98.1 | 96.8 | 97.7 | 95.8 | 97.1 |
| | | .017 | .015 | .017 | .015 | .025 | .016 | .022 | .020 | .022 | .021 | .028 | .024 | .022 | .021 | .020 | .020 |

## 5 Proofs

**Proof.** According to Theorem 3 of Zhu et al. (2017), it follows that

$$\sqrt{NT}[\hat{\theta}-\theta] \xrightarrow{\text{d}} \text{N}(0, \sigma_\varepsilon^2 \Sigma^{-1})$$

where $\Sigma$ is as in (3). It suffices to show that $\sqrt{NT}[\hat{\theta}^* - \hat{\theta}]$ has the same limiting distribution. By the proof of Theorem 3 by Zhu et al. (2017), we see $\hat{\theta} = \theta + \hat{\Sigma}^{-1}\hat{\Sigma}_{\text{xe}}$ where $\hat{\Sigma} = (NT)^{-1}\sum_{t=1}^T \mathbb{X}_{t-1}^\top \mathbb{X}_{t-1}$ and $\hat{\Sigma}_{\text{xe}} = (NT)^{-1}\sum_{t=1}^T \mathbb{X}_{t-1}^\top \mathcal{E}_t$ and as a result, it was proven that

$$\hat{\Sigma} \to \Sigma \quad \text{and} \quad \sqrt{NT}\hat{\Sigma}_{\text{xe}} \xrightarrow{\text{d}} N(0, \sigma_\varepsilon^2 \Sigma) \quad \text{as} \quad \min\{N, T\} \to \infty. \tag{4}$$

Now we observe

$$\sqrt{NT}[\hat{\theta}^* - \hat{\theta}] = \left( \frac{\sum_{t=1}^T \mathbb{X}_{t-1}^{*\top} \mathbb{X}_{t-1}^*}{NT} \right)^{-1} \left( \frac{1}{\sqrt{NT}} \sum_{t=1}^T \mathbb{X}_{t-1}^{*\top} \mathcal{E}_t^* \right)$$

where $\mathcal{E}_t^* = (\hat{\varepsilon}_{1t}^*, \ldots, \hat{\varepsilon}_{Nt}^*)^\top$ and will first show that $(NT)^{-1}\sum_{t=1}^T \mathbb{X}_{t-1}^\top \mathbb{X}_{t-1}$ and $(NT)^{-1}\sum_{t=1}^T \mathbb{X}_{t-1}^{*\top} \mathbb{X}_{t-1}^*$ have the same limiting in probability and second that

$$\frac{1}{\sqrt{NT}} \sum_{t=1}^T \mathbb{X}_{t-1}^{*\top} \mathcal{E}_t^* \xrightarrow{\text{d}^*} \text{N}(0, \sigma_\varepsilon^2 \Sigma)$$

For the first desired result, noting that

$$\sum_{t=1}^T \mathbb{X}_{t-1}^\top \mathbb{X}_{t-1} = \sum_{t=1}^T \sum_{i=1}^N X_{i(t-1)} X_{i(t-1)}^\top, \quad \sum_{t=1}^T \mathbb{X}_{t-1}^{*\top} \mathbb{X}_{t-1}^* = \sum_{t=1}^T \sum_{i=1}^N X_{i(t-1)}^* X_{i(t-1)}^{*\top}$$

we show that $(NT)^{-1}\sum_{t=1}^T \sum_{i=1}^N X_{i(t-1)}^* X_{i(t-1)}^{*\top}$ and $(NT)^{-1}\sum_{t=1}^T \sum_{i=1}^N X_{i(t-1)} X_{i(t-1)}^\top$ have the same limiting in probability.

Let $s_\tau = L_1 + L_2 + \cdots + L_\tau$ and we write

$$\left| \frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N X_{i(t-1)}^* X_{i(t-1)}^{*\top} - \frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N X_{i(t-1)} X_{i(t-1)}^\top \right| \le$$

$$\left| \frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N X_{i(t-1)}^* X_{i(t-1)}^{*\top} - \frac{1}{NT} \sum_{t=1}^{s_\tau} \sum_{i=1}^N X_{i(t-1)}^* X_{i(t-1)}^{*\top} \right| + \left| \frac{1}{NT} \sum_{t=1}^{s_\tau} \sum_{i=1}^N X_{i(t-1)}^* X_{i(t-1)}^{*\top} - \frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N X_{i(t-1)} X_{i(t-1)}^\top \right| \tag{5}$$

In the following lemma, it is shown that the first absolute value above in (5) tends to zero in (conditional) probability.

**Lemma 5.1**

$$\frac{1}{NT} \sum_{t=T+1}^{s_\tau} \sum_{i=1}^N X_{i(t-1)}^* X_{i(t-1)}^{*\top} \xrightarrow{\text{p}^*} 0$$

**Proof:** Note that $\sum_{t=T+1}^{s_\tau}\sum_{i=1}^{N} X^*_{i(t-1)}X^{*\top}_{i(t-1)}$ is the sum of observations in block $B(I_\tau, L_\tau)$ after deleting the first $T - s_{\tau-1}$ elements, where $s_{\tau-1} = L_1 + \cdots + L_{\tau-1}$. Let $R_1 = T - s_{\tau-1}$ and $R = L_\tau - R_1$. Note that $R$, conditional on $(s_{\tau-1}, R_1)$, has a geometric distribution with mean $1/\rho$ by the memoryless property of the geometric distribution. Thus $(NT)^{-1}\sum_{t=T+1}^{s_\tau}\sum_{i=1}^{N} X^*_{i(t-1)}X^{*\top}_{i(t-1)}$ is equal in distribution to $(NT)^{-1}\sum_{j=I}^{I+R-1}\sum_{i=1}^{N} X_{ij}X_{ij}^\top$, which is related with block $B(I,R)$, where $I$ is uniform on $\{0,1,\ldots,T-1\}$. Now we show that $(NT)^{-1}U_{I,R} \xrightarrow{\mathrm{p}^*} 0$ where $U_{I,R} = \sum_{j=I}^{I+R-1}\mathbb{X}_j^\top\mathbb{X}_j = \sum_{j=I}^{I+R-1}\sum_{i=1}^{N} X_{ij}X_{ij}^\top$ with $\mathbb{X}_j = (X_{1j},\ldots,X_{Nj})^\top$ and $X_{ij} = (1, w_i^\top\mathbb{Y}_j, Y_{ij}, Z_i^\top)^\top$, $i = 1,2,\ldots,N$.

To do this, first we observe the (conditional) mean $(NT)^{-1}E^*[U_{I,R}]$:

$$E^*[U_{I,R}|R] = E_I[E^*[U_{I,R}|R,I]] = \frac{1}{T}\sum_{t=0}^{T-1}\left[\sum_{j=t}^{t+R-1}\sum_{i=1}^{N} X_{ij}X_{ij}^\top\right] = R\left[\frac{1}{T}\sum_{t=0}^{T-1}\sum_{i=1}^{N} X_{it}X_{it}^\top\right]$$

$$\frac{1}{NT}E^*[U_{I,R}] = \frac{1}{NT}E_R[E^*[U_{I,R}|R]] = \frac{1}{T\rho}\left[\frac{1}{NT}\sum_{t=0}^{T-1}\sum_{i=1}^{N} X_{it}X_{it}^\top\right] = O_p\left(\frac{1}{T\rho}\right)$$

Secondly we observe the (conditional) variance $(NT)^{-2}Var^*[U_{I,R}]$: we use the identity $Var^*[U_{I,R}] = E_R[Var^*[U_{I,R}|R]] + Var_R[E^*[U_{I,R}|R]]$ and $Var^*[U_{I,R}|R] = E^*[U_{I,R}U_{I,R}^\top|R] + E^*[U_{I,R}|R](E^*[U_{I,R}|R])^\top$ and can get $(NT)^{-2}Var^*[U_{I,R}] = O_p(1/(T^2\rho^2))$. Thus $(NT)^{-1}U_{I,R} \xrightarrow{\mathrm{p}^*} 0$ and the desired result in Lemma 5.1 is obtained. $\square$

Now we show in Lemma 5.2 that the second absolute value above in (5) tends to zero in (conditional) probability.

**Lemma 5.2**
$$\left|\frac{1}{NT}\sum_{t=1}^{s_\tau}\sum_{i=1}^{N} X^*_{i(t-1)}X^{*\top}_{i(t-1)} - \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N} X_{i(t-1)}X_{i(t-1)}^\top\right| \xrightarrow{\mathrm{p}^*} 0$$

**Proof:** Recall $U_{t,\ell} = \sum_{j=t}^{t+\ell-1}\mathbb{X}_j^\top\mathbb{X}_j = \sum_{j=t}^{t+\ell-1}\sum_{i=1}^{N} X_{ij}X_{ij}^\top$, the sum of related observations in block $B(t,\ell) = \{\mathbb{Y}_t,\ldots,\mathbb{Y}_{t+\ell-1}\}$. We may write

$$\frac{1}{NT}\sum_{t=1}^{s_\tau}\sum_{i=1}^{N} X^*_{i(t-1)}X^{*\top}_{i(t-1)} = \frac{1}{NT}\sum_{l=1}^{\tau} U_{I_l.L_l}$$

Since $\tau = T\rho + O_p(\sqrt{T\rho})$, we may consider a sequence $m = m_T$ with $m \to \infty$ and $m/(T\rho) \to 1$, and show that

$$\frac{1}{NT}\sum_{l=1}^{m} U_{I_l.L_l} - \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N} X_{i(t-1)}X_{i(t-1)}^\top \xrightarrow{\mathrm{p}^*} 0$$

The left-hand side is decomposed as

$$\left(\frac{1}{NT}\sum_{l=1}^{m} U_{I_l.L_l} - \frac{\rho}{Nm}\sum_{l=1}^{m} U_{I_l.L_l}\right) + \left(\frac{\rho}{Nm}\sum_{l=1}^{m} U_{I_l.L_l} - \frac{\rho}{N}E^*[U_{I_1,L_1}]\right) + \left(\frac{\rho}{N}E^*[U_{I_1,L_1}] - \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N} X_{i(t-1)}X_{i(t-1)}^\top\right)$$

For the first term, we see that it is equal to $\frac{1}{NT}\left(\frac{1}{m}\sum_{l=1}^{m} U_{I_l.L_l}\right)(m - T\rho)$, which is $o_p(1)$, and for the second term, it is clear that the second term is $o_p(\rho/N)$ by the weak law of large numbers of i.i.d sequence $\{U_{I_l,L_l} : l = 1, 2, \dots\}$. For the third term, by the same argument as in above we have $E^*[U_{I_1,L_1}] = \frac{1}{T\rho}\sum_{t=0}^{T-1}\sum_{i=1}^{N} X_{it}X_{it}^\top$ and thus the third term is zero. Hence the desired convergence holds. $\square$

Finally we verify the asymptotic multivariate normality as follows:

**Lemma 5.3**

$$\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\mathbb{X}_{t-1}^{*\top}\mathcal{E}_t^* \xrightarrow{\text{d}^*} N(0, \sigma_\varepsilon^2\Sigma)$$

**Proof:** Write

$$\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\mathbb{X}_{t-1}^{*\top}\mathcal{E}_t^* = \frac{1}{\sqrt{NT}}\sum_{t=1}^{s_\tau}\mathbb{X}_{t-1}^{*\top}\mathcal{E}_t^* - \frac{1}{\sqrt{NT}}\sum_{t=T+1}^{s_\tau}\mathbb{X}_{t-1}^{*\top}\mathcal{E}_t^*.$$

We show the asymptotic normality of the first term of the right-hand side and the convergence of the second to zero in probability. The second term, $\frac{1}{\sqrt{NT}}\sum_{t=T+1}^{s_\tau}\mathbb{X}_{t-1}^{*\top}\mathcal{E}_t^*$, can be written as $\frac{1}{\sqrt{NT}}V_{I,R}$ for the same reason of memoryless property of geometric distribution as above, where

$$V_{I,R} = \sum_{t=I}^{I+R-1}\mathbb{X}_{t-1}^{\top}\mathcal{E}_t^* = \sum_{t=I}^{I+R-1}\sum_{i=1}^{N} X_{i(t-1)}\hat{\varepsilon}_{it}^* \tag{6}$$

Noting that $E^*[V_{I,R}] = 0$ and $\{\mathcal{E}_t^* : t = 1, 2, \dots, T\}$ are i.i.d., observe

$$\frac{1}{r}Var^*[V_{I,R}|R = r] = \frac{1}{r}Var^*\left[\sum_{t=I}^{I+r-1}\mathbb{X}_{t-1}^{\top}\mathcal{E}_t^* | R = r\right]$$

$$= \frac{1}{rT}\sum_{j=1}^{T} Var^*\left[\sum_{t=j}^{j+r-1}\mathbb{X}_{t-1}^{\top}\mathcal{E}_t^*\middle| R = r, I = j\right] = \frac{1}{rT}\sum_{j=1}^{T}\sum_{t=j}^{j+r-1} Var^*\left[\mathbb{X}_{t-1}^{\top}\mathcal{E}_t^*\middle| R = r, I = j\right]$$

$$= \frac{1}{rT}\sum_{j=1}^{T}\sum_{t=j}^{j+r-1} E^*[(\mathbb{X}_{t-1}^{\top}\mathcal{E}_t^*)(\mathbb{X}_{t-1}^{\top}\mathcal{E}_t^*)^\top] = \frac{1}{T}\sum_{t=1}^{T} E^*[\mathbb{X}_{t-1}^{\top}\mathcal{E}_t^*\mathcal{E}_t^{*\top}\mathbb{X}_{t-1}] = \frac{\hat{\sigma}_\varepsilon^2}{T}\sum_{t=1}^{T}\mathbb{X}_{t-1}^{\top}\mathbb{X}_{t-1}$$

where the last equality holds since $E[\hat{\varepsilon}_{it}^*\hat{\varepsilon}_{jt}^*] = 0$ if $i \neq j$ and $E[\hat{\varepsilon}_{it}^{*2}] = \hat{\sigma}_\varepsilon^2$ and thus $E^*[\mathcal{E}_t^*\mathcal{E}_t^{*\top}] = \hat{\sigma}_\varepsilon^2 I_N$ where $I_N$ is $N \times N$ identity matrix and $\hat{\sigma}_\varepsilon^2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^2$. Thus we have

$$\frac{1}{NT}Var^*[V_{I,R}|R] = \frac{\hat{\sigma}_\varepsilon^2 R}{T}\frac{1}{NT}\sum_{t=1}^{T}\mathbb{X}_{t-1}^{\top}\mathbb{X}_{t-1}$$

and thus

$$\frac{1}{NT}Var^*[V_{I,R}] = \frac{\hat{\sigma}_\varepsilon^2}{T\rho}\frac{1}{NT}\sum_{t=1}^{T}\mathbb{X}_{t-1}^{\top}\mathbb{X}_{t-1}$$

which is $O_p(1/(T\rho)) \xrightarrow{\text{p}} 0$ by (4) and by the fact $\hat{\sigma}_\varepsilon^2 \xrightarrow{\text{p}} \sigma_\varepsilon^2$. Therefore, the convergence of the second term $\frac{1}{\sqrt{NT}}\sum_{t=T+1}^{s_\tau}\mathbb{X}_{t-1}^{*\top}\mathcal{E}_t^*$ in (conditional) probability to zero is obtained.

The first term $\frac{1}{\sqrt{NT}}\sum_{t=1}^{s_\tau}\mathbb{X}_{t-1}^{*\top}\mathcal{E}_t^*$ can be written as $\frac{1}{\sqrt{NT}}\sum_{l=1}^{\tau}V_{I_l,L_l}$ where $V_{I,L}$ is given in (6). Similarly above we may show that

$$\frac{1}{\sqrt{NT}}\sum_{l=1}^{m}V_{I_l,L_l}\xrightarrow{\mathrm{d}^*}N(0,\sigma_\varepsilon^2\Sigma)$$

for any sequence $m=m_T$ with $m\to\infty$ and $m/(T\rho)\to 1$. Since $\{(I_l,L_l):l=1,2,\ldots,m\}$ are i.i.d. we have i.i.d. sequence $\{V_{I_l,L_l}:1,2,\ldots,m\}$ and thus we apply the CLT of i.i.d. sequence for the sum $\frac{1}{\sqrt{NT}}\sum_{l=1}^{m}V_{I_l,L_l}$ with asymptotic variance, that is given by $\lim_{\min\{N,T\}\to\infty}\frac{1}{NT}Var^*\left(\sum_{l=1}^{m}V_{I_l,L_l}\right)$

$$=\lim_{\min\{N,T\}\to\infty}\frac{1}{NT}\sum_{l=1}^{m}Var^*(V_{I_l,L_l})=\lim_{\min\{N,T\}\to\infty}m\frac{\hat{\sigma}_\varepsilon^2}{T\rho}\left(\frac{1}{NT}\sum_{j=1}^{T}\mathbb{X}_{j-1}^\top\mathbb{X}_{j-1}\right)=\sigma_\varepsilon^2\Sigma$$

by (4) and by the fact $\hat{\sigma}_\varepsilon^2\xrightarrow{\mathrm{p}}\sigma_\varepsilon^2$. The details of asymptotic normality for the i.i.d. sequence are omitted here and we refer to the proof of Hwang and Shin (2013).

**Acknowledgement:** This work was supported by Research Fund of Gachon University (GCU-2019-0376).

## References

Can, U. Alatas, B. 2019. A new direction in social network analysis: Online social network analysis problems and applications. Physica A: Statistical Mechanics and its Applications, 535, 122372.

Durante, D. Dunson, D.B. 2014. Nonparametric bayes dynamic modelling of relational data. Biometrika 101 (4), 883–898.

Hwang, E. Shin, D.W. 2012. Strong consistency of the stationary bootstrap under $\psi$-weak dependence. Statistics and Probability Letters 82 (3), 488–495.

Hwang, E. Shin, D.W. 2013. New bootstrap method for autoregressive models. Communications for Statistical Applications and Methods 20 (1), 85–96.

Nowicki, K. Snijders, T.A.B. 2001. Estimation and prediction for stochastic block structures. Journal of American Statistical Association 96, 1077–1087.

Politis, D.N., Romano, J.P., 1994. The stationary bootstrap. Journal of the American Statistical Association 89, 1303–1313.

Wei, F. Tian, W. 2018. Heterogeneous connection effects. Statistics and Probability Letters 133, 9–14.

Zhu, X. Pan, R. Liu, Y. Wang, H. 2017. Network vector autoregression. Annals of Statistics 45 (3), 1096–1123.

Zhu, X. Huang, D. Pan, Wang, H. 2020. Multivariate spatial autoregressive model for large scale social networks. Journal of Econometrics 215, 591–606.