Statistics and Its Interface Volume 16 (2023) 593–615

Network vector autoregressive moving average model*

Xiao Chen, Yu Chen†, and Xixu Hu

Modeling a continuous response of a large-scale network is an important task and it has become prevailing in practice at present. This paper proposes a novel network vector autoregressive moving average (NARMA) model which considers the responses from both an ultra-high dimension vector and the network structure effects. Compared with the network vector autoregressive (NAR, [26]) model, we take into account the lagged innovations and corresponding network effect in our proposed model. With more parameters considered and a moving average term incorporated, the proposed NARMA model can fit the data more closely and accurately, thus has a better performance than the NAR model. A modified least square estimation for the NARMA model is introduced, and the consistency properties are fully investigated. Finally, we demonstrate the superiority of the proposed NARMA model by investigating the financial contagions of S&P500 index constituents.

Keywords and phrases: Network data, Modified least square estimator, Vector autoregressive moving average, High dimensional time series.

1. INTRODUCTION

High-dimension network structure is common in social networks and financial networks, such as the interpersonal relationships on Facebook, Twitter and the interplay between stock prices which are included in the same index or market. In this article, we concentrate our interest on financial data. On the one hand, different companies in the market are no longer independent of each other due to globalization and an increasing number of upstream and downstream enterprises. On the other hand, a huge number of covariates can be collected for each stock, such as market cap, PE ratio, EPS, and so on. As a result, it indicates that network data plays an important role in many fields, such as being used to provide site user portraits ([11]), characterize social capital flow patterns ([3]), and analyze consumer behaviors ([9]).

*We thank the AE and the two referees for their helpful comments and suggestions which greatly improved our paper.

†Corresponding author.

Consider a large-scale network with N nodes (i.e., users, indexes) indexed by $1 \leq i \leq N$. To describe the network structure, we define an adjacency matrix $A = (a_{ij}) \in \mathbb{R}^{N\times N}$, where $a_{ij} = 1$ if there exists a relationship (i.e., a directed edge) from i to j, and $a_{ij} = 0$ otherwise ([24]). Throughout this paper we assume that the adjacency matrix A is non-random. By convention, let $a_{ii} = 0$ for any $1 \leq i \leq N$. Let $Y_{it} \in \mathbb{R}$ be a continuous response obtained from node i at time point $t, 1 \leq t \leq T$. Denoting $\mathbb{Y}_t = (Y_{1t}, \ldots, Y_{Nt})^\top \in \mathbb{R}^N$, we are interested in studying the dynamic pattern of $\mathbb{Y}_t$. In the past literatures, vector autoregression (VAR) and the corresponding dimension reduction methods have been extensively studied, especially the factor models ([17], [10], [2], [18]).

Zhu et al. [26] proposed a network vector autoregression (NAR) model. The NAR model assumes that each node's response at a given time point is a linear combination of (a) its previous value, (b) the average value of its connected neighbors, (c) a set of node-specific covariates, and (d) an independent noise. The advantages of the NAR model are tow-fold. First, as a variant of the vector autoregression (VAR) model, it captures the relationship between multivariate time series, thus fully utilizing information of all time series. ([2], [14]). Second, by taking network structures into consideration, the NAR model significantly reduces the computation time complexity $O(N^2)$ of the VAR model, to an acceptable level, while keeping the high prediction performance.

Taking network structure into time series modeling has shown its effectiveness in many scenarios. For example, Zhu and Pan [25] propose the grouped network vector autoregression (GNAR) model classifying individuals in the network into different groups to express heterogeneity. Zhu et al. [28] further design the network quantile autoregression (NQAR) model accounting the network information into a quantile regression framework and concentrated on the tail dependency. Chen, Fan and Zhu [4] propose the community network autoregression (CNAR) model, allowing heterogeneous network effects across different network communities with unknown cross-sectional dependence accounted by the non-community-related latent factors. Last but not least, Armillotta and Fokianos [1] propose the Poisson network autoregression to link multivariate observation-driven count time series models with time-varying network data.

Despite its simple form and easy interpretation, the only two essential parameters that the NAR model considers, the coefficients of its previous values and the average of its connected neighbors, strictly restrict the scenarios it can be applied and may cause model misspecification. To address this issue, some flexible extensions to the NAR model have been considered in the literature. For instance, Dou, Parrella and Yao [5] and Zhu et al. [29] implement a node-specific network effect to characterize different inferential powers of different nodes. Sun [21], Wang, Lin and Wang [23], and Sun and Malikov [22] investigate the nonlinear and nonparametric extensions. Sojourner [20], Liu, Patacchini and Rainone [13] and Zhu et al. [27] consider the multivariate responses. However, almost all of these articles concentrate on the impact of previous responses while ignoring the influence of the previous noises.

In this article, we propose a novel network autoregressive moving average (NARMA) model which considers both the influence of the lagged responses and the influence of the lagged noises. The response Y_{it} is influenced by six factors, (a) its own lagged value $Y_{i(t-1)}$, (b) a set of node-specific covariates Z_i, (c) its connected neighbors $n_i^{-1}\sum_{j=1}^N a_{ij}Y_{j(t-1)}$, (d) its lagged noise $\varepsilon_{i(t-1)}$, (e) the lagged noise of its connected neighbors $n_i^{-1}\sum_{j=1}^N a_{ij}\varepsilon_{j(t-1)}$, and (f) an independent noise ε_{it}.

Our paper contributes to the literature in the following four aspects. Firstly, we propose a network vector autoregressive moving average model by considering the network effect of the lagged error terms. Meanwhile, compared with the ordinary ARMA model, our model significantly reduces the dimensions while keeping high prediction accuracy when the data dimension is exceptionally high. Secondly, to investigate its theoretical properties, we first propose a specific definition of invertibility in high-dimensional time series modeling. Combined with the high-dimensional stationarity definition proposed by Zhu et al. [26], we provide an estimator for the proposed NARMA model based on a modified least square estimation and obtain its consistency. Thirdly, to tackle the unobservable problem before time $T = 0$, we innovatively use the truncated residual error terms to approximate the whole ones with consistency guarantees. Fourthly, we demonstrate the superiority of our proposed model through extensive simulation studies and a real-world example. The NARMA model fits better with lower root mean square error in the high-dimensional case than the NAR model. We also illustrate the asymptotic properties through simulation results. In the S&P500 index constituents price modeling, the NARMA model predicts with higher accuracy than the NAR model, thus holding outstanding advantages in ultra-high-dimensional network structure data modeling.

The rest of the article is organized as follows. Section 2 introduces the NARMA model with its stationarity, invertibility, consistency and a p-lag extension. Section 3 gives three kinds of numerical studies and analyses. A case study about the prices of S&P 500 constituents is given in Section 4. A brief discussion is given to conclude the article in Section 5. All technical details can be found in the Appendix.

2. NETWORK VECTOR AUTOREGRESSIVE MOVING AVERAGE MODEL

2.1 Model and notations

Let N be the network size and Y_{it} be the response collected from the ith subject at time t. We also assume a p-dimensional node-specific random vector $Z_i = (Z_{i1}, \ldots, Z_{ip})^\top \in \mathbb{R}^p$ can be observed. We propose the NARMA model as follows:

$$
\begin{aligned}
Y_{it} &= \lambda + Z_i^\top\gamma + \beta_1 n_i^{-1}\sum_{j=1}^N a_{ij}Y_{j(t-1)} \\
&\quad + \alpha_1 n_i^{-1}\sum_{j=1}^N a_{ij}\varepsilon_{j(t-1)} + \beta_0 Y_{i(t-1)} \\
&\quad + \alpha_0\varepsilon_{i(t-1)} + \varepsilon_{it},
\end{aligned}
\tag{1}
$$

where $n_i = \sum_{j\neq i} a_{ij}$ (out-degree of the node) is the number of nodes that node i follows. We set $a_{ij} = 1$ if i follows j with the meaning that follower i can be affected by its leader j. Specifically, we assume $\frac{1}{N}\sum_{i=1}^N \frac{1}{n_i} > 0$, which means most rows of the adjacency matrix are sparse.

The term $\lambda + Z_i^\top\gamma$ characterizes the nodal impact of the ith node, where $\lambda \in \mathbb{R}$ is the intercept and $\gamma = (\gamma_1, \ldots, \gamma_p)^\top \in \mathbb{R}^p$ is the associated coefficient vector which is invariant over time t. The quantity $n_i^{-1}\sum_{j=1}^N a_{ij}Y_{j(t-1)}$ represents the average impact of the leaders that the ith node follows. Its associated parameter β_1 is referred to as the network effect. Similarly, $n_i^{-1}\sum_{j=1}^N a_{ij}\varepsilon_{j(t-1)}$ is the average impact of the error terms from the nodes that node i follows. α_1 is referred to as the network effect of the error term. The term $Y_{i(t-1)}$ is the standard autoregressive impact with β_0 referred to as the momentum effect. Compared with the NAR model([26]), $\mathcal{E}_{t-1}$ appears directly for $\mathbb{Y}_t$ rather than an indirect effect.

Now we introduce some notations used in this paper. The parameter space is established as $\Theta = \{(\lambda, \gamma, \alpha_0, \alpha_1, \beta_0, \beta_1) : \alpha_0 + \beta_0 \neq 0, \alpha_1 + \beta_1 \neq 0\}$ (the condition exists as the coefficients cannot be zero). Moreover, ε_{it} is the error term following a normal distribution with $E(\varepsilon_{it}) = 0$ and $\mathrm{Var}(\varepsilon_{it}) = \sigma^2$. When one considers the multivariate distribution of $\mathcal{E}_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Nt})^\top$, it is more realistic to take the covariance matrix as a non-diagonal one. However, this might lead to a large number of extra unknown parameters, which adds additional complexity to the model. For convenience, we assume a diagonal covariance structure for $\mathcal{E}_t$. Lastly, $\|\cdot\|$ without subscript means ℓ_2-norm in the following.

Remark 2.1 (Relation to the ARMA model)**.** *The* NARMA *model is a simplification of the ordinary* ARMA *model. It reduces the number of coefficients by only taking influential nodes into consideration. Hence it is much more convenient when we consider a large-scale network. Taking the network structure into consideration saves lots of time while also retaining great accuracy.*

Remark 2.2 (Advantages compared with the NAR model)**.** *Compared with an* NAR *model, the proposed* NARMA *model considers the influence of the past errors, which can obtain more accurate predicted values, especially when the time dimension is large. Though the parameter estimation is a bit harder than the* NAR *model, we can still give solutions and the efforts are worthwhile.*

For simplicity, we define $\mathbb{Z} = (Z_1, \ldots, Z_N)^\top \in \mathbb{R}^{N \times p}$ and $\mathcal{B}_0 = \lambda \mathbf{1} + \mathbb{Z}\gamma \in \mathbb{R}^N$, where $\mathbf{1} = (1, \ldots, 1)^\top$ is a compatible vector with all elements of 1. We collect all the unknown parameters and denote them by $\boldsymbol{\theta} = (\lambda, \gamma, \alpha_0, \alpha_1, \beta_0, \beta_1)$. Recall that $\mathbb{Y}_t = (Y_{1t}, \ldots, Y_{Nt})^\top \in \mathbb{R}^N$. Then we can rewrite Eq. (1) in a vector form:

$$\mathbb{Y}_t = \mathcal{B}_0 + \mathcal{G}\mathbb{Y}_{t-1} + \mathcal{K}\mathcal{E}_{t-1} + \mathcal{E}_t. \tag{2}$$

In Eq. (2), $\mathcal{G} = \beta_1 W + \beta_0 I$ and $\mathcal{K} = \alpha_1 W + \alpha_0 I$, W is defined as $\text{diag}\{n_1^{-1}, \ldots, n_N^{-1}\}A$ which is a row-normalized adjacency matrix whose rows are not null and I is an identity matrix with a compatible dimension. $\mathcal{E}_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Nt})^\top \in \mathbb{R}^N$ is the innovation vector. Since the adjacency matrix A is assumed to be a non-stochastic matrix as before, matrices $\mathcal{G}$, $\mathcal{K}$ and W are all non-stochastic. However, $\mathcal{B}_0$ is a random matrix due to the assumption that $\mathbb{Z}$ is random.

2.2 Strict stationarity

Since $\mathbb{Y}_t$ is a time series, it is of interest to study its stationarity. In the following, we derive the conditions for strict stationarity in the non-asymptotic and the asymptotic cases respectively: N is a finite number and $N \to \infty$. We follow the definitions in both cases given in Zhu et al. [26]. We start with the non-asymptotic case first.

Since the definition of strictly stationary when N is fixed is well-known, we give no more detailed description here. Under the definition, we can give the following theorem.

Theorem 2.1. *Suppose that $E\|Z_i\| < \infty$ and N is fixed. If $|\beta_0| + |\beta_1| < 1$, then there exists a unique strictly stationary solution with a finite first-order moment to the NARMA model (2). The solution takes the form of*

$$\mathbb{Y}_t = (I - \mathcal{G})^{-1}\mathcal{B}_0 + \mathcal{E}_t + \sum_{j=1}^{\infty} \mathcal{G}^{j-1}(\mathcal{K} + \mathcal{G})\mathcal{E}_{t-j}. \tag{3}$$

The proof of Theorem 2.1 is given in the Appendix.

Next, we study the strict stationarity under the condition that $N \to \infty$. In classical time series theory, the dimension of the model is fixed. However, when incorporating the network structure, the number of nodes (i.e, the dimension of the model, N) could be really large. Therefore, it is interesting to investigate whether the strict stationarity is still satisfied when $N \to \infty$. Zhu et al. [26] first states the definition of strict stationarity in the extreme case:

Definition 1. *Let $\{\mathbb{Y}_t \in \mathbb{R}^N\}$ be a N-dimensional time series with $N \to \infty$. Define $\mathcal{W} = \{\omega \in \mathbb{R}^\infty : \sum |\omega_i| < \infty\}$, where $\omega = (\omega_i \in \mathbb{R} : 1 \leq i \leq \infty)^\top \in \mathbb{R}^\infty$. For each $\omega \in \mathcal{W}$, let $\boldsymbol{w}_N = (\omega_1, \ldots, \omega_N)^\top \in \mathbb{R}^N$ be a truncated N-dimensional vector. Then $\{\mathbb{Y}_t\}$ is said to be strictly stationary, if it satisfies the following conditions for any $\omega \in \mathcal{W}$,*

(1) $Y_t^\omega = \lim_{N\to\infty} \boldsymbol{w}_N^\top \mathbb{Y}_t$ exists in the almost sure sense;
(2) $\{Y_t^\omega\}$ is strictly stationary.

Moreover, $\mathbb{Y}_t$ is said to have a finite m-th order moment if $\max_{1\leq i<\infty} E|Y_{it}|^m < \infty$.

In the classical settings when N is fixed, one can verify that $\{\mathbb{Y}_t\}$ is strictly stationary if and only if $\{\mathbf{w}_N^\top \mathbb{Y}_t\}$ is strictly stationary for any $\mathbf{w}_N \in \mathbb{R}^N$. As a result, Definition 1 can be viewed as an extension of the classical stationarity with fixed N to the diverging case. Using the definition above, we can have the following theorem for the NARMA model:

Theorem 2.2. *Under the same conditions as in Theorem 2.1 with $N \to \infty$. Then the solution defined as Eq. (3) is a unique strictly stationary solution (in the sense of Definition 1) to the* NARMA *model with a finite first-order moment.*

The proof is also given in the Appendix. Remarkably, the strict stationary solution when $N \to \infty$ shares the same form when N is fixed. Theorem 2.2 shows the stationarity of the high-dimensional NARMA model.

2.3 Invertibility

Compared with the seminal NAR model, with the addition of moving average part, the response vector $\mathbb{Y}_t$ satisfies an ARMA-structure. Therefore, it is very important to investigate the invertibility of NARMA model. Similar to the discussion in stationarity above, we can define two sets of invertibility according to whether N is a finite value or $N \to \infty$. First, we start with the classical settings when N is fixed. We adopt the classical definition for the invertibility of ARMA model given in Pham and Tran [19]. Thus we have the following Theorem.

Theorem 2.3. *Suppose that $E\|Z_i\| < \infty$ and N is fixed. If $(|\alpha_0| + |\alpha_1|) < 1$, then there exists a unique invertible solution with a finite first-order moment to the NARMA model (2). And the solution takes the form of*

$$\begin{aligned}\mathcal{E}_t := \mathcal{E}_t(\boldsymbol{\theta}) &= -(I + \mathcal{K})^{-1}\mathcal{B}_0 \\ &\quad + \mathbb{Y}_t - \sum_{j=1}^{\infty}(-\mathcal{K})^{j-1}(\mathcal{K} + \mathcal{G})\mathbb{Y}_{t-j}.\end{aligned} \tag{4}$$

The proof of Theorem 2.3 is similar to the proof of stationarity. From Theorem 2.3 we can get an inverse representation of $\mathcal{E}_t$. Eq. (4) can be further simplified by combining responses $\mathbb{Y}_t$ with the constant terms. Without loss of generality, we use the following form in the rest of the article:

$$\mathcal{E}_t = \mathbb{Y}_t - \sum_{j=1}^{\infty}(-\mathcal{K})^{j-1}(\mathcal{K} + \mathcal{G})\mathbb{Y}_{t-j}. \tag{5}$$

From Eq. (5) we can see that the inverse representation relies on the values of $\mathbb{Y}_t$ and the values of $\mathcal{K}$ and $\mathcal{G}$. Recall the formulae of $\mathcal{K}$ and $\mathcal{G}$, we can further notice that the inverse representation relies on the parameters $\boldsymbol{\theta} = (\lambda, \gamma, \alpha_0, \alpha_1, \beta_0, \beta_1)$. Actually, the response vector $\mathbb{Y}_t$ before time 0 are unobservable and the parameter vector is unknown. To tackle these problems, by convention, we always initialize $\mathbb{Y}_0$ as a arbitrary value $\mathbf{y}_0$ and replace the parameter $\boldsymbol{\theta}$ with an estimated version $\hat{\boldsymbol{\theta}} = (\hat{\lambda}, \hat{\gamma}, \hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_0, \hat{\beta}_1)$. Therefore, we need to extend the conditions for invertibility into the practical scenarios. Denote the value of $\mathcal{E}_t$ given $\mathbf{y}_0$ by $\mathcal{E}_{(t|\mathbf{y}_0)}$ and the value of $\mathcal{E}_t$ given $(\mathbf{y}_0, \hat{\boldsymbol{\theta}})$ by $\mathcal{E}_{\hat{\boldsymbol{\theta}}(t|\mathbf{y}_0)}$. We show the extended conditions for invertibility as the following remarks.

Remark 2.3. *Following Pham and Tran [19], we say that the process $\mathbb{Y}_t$ is invertible if $\mathcal{E}_{(t|\mathbf{y}_0)} - \mathcal{E}_t$ converges to 0 in some sense, as $t \to \infty$, regardless of $\mathbf{y}_0$.*

Remark 2.4. *We say that the model (2) is invertible at $\hat{\boldsymbol{\theta}}$ relative to the observation process $\mathbb{Y}_t$ if there exists a stationary process $\mathcal{E}_{\hat{\boldsymbol{\theta}}t} := \mathcal{E}_t(\hat{\boldsymbol{\theta}})$ such that $\mathcal{E}_{\hat{\boldsymbol{\theta}}(t|\mathbf{y}_0)} - \mathcal{E}_{\hat{\boldsymbol{\theta}}t}$ converges to 0 in some sense as $t \to \infty$, regardless of $\mathbf{y}_0$.*

Next we address the situation when $N \to \infty$. To our best knowledge, there exists no widely accepted general definition. As a possible attempt, we give definitions in several aspects.

Definition 2. *Let $\{\mathbb{Y}_t \in \mathbb{R}^N\}$ be a N-dimensional time series with $N \to \infty$.*

(a) $\{\mathbb{Y}_t\}$ is said to be strongly invertible under ℓ_1-norm if it satisfies that

$$\|\mathcal{E}_{\boldsymbol{\theta}(T|\mathbf{y}_0)} - \mathcal{E}_T\|_1 \to 0 \qquad a.s.,$$

when $N, T \to \infty$.

(b) $\{\mathbb{Y}_t\}$ is said to be strongly invertible under ℓ_∞-norm if it satisfies that

$$\|\mathcal{E}_{\boldsymbol{\theta}(T|\mathbf{y}_0)} - \mathcal{E}_T\|_\infty \to 0 \qquad a.s.$$

when $N, T \to \infty$.

(c) $\{\mathbb{Y}_t\}$ is said to be weak invertible in mean if it satisfies that

$$\lim_{\min\{N,T\}\to\infty} \frac{1}{NT}\sum_{t=1}^{T} \mathbf{1}^\top(\mathcal{E}_{\boldsymbol{\theta}(t|\mathbf{y}_0)} - \mathcal{E}_t) \to 0.$$

(d) $\{\mathbb{Y}_t\}$ is said to be weak invertible in mean square if it satisfies that

$$\lim_{\min\{N,T\}\to\infty} \frac{1}{NT}\sum_{t=1}^{T} \mathcal{E}^\top_{\boldsymbol{\theta}(t|\mathbf{y}_0)}\mathcal{E}_{\boldsymbol{\theta}(t|\mathbf{y}_0)} \to \sigma^2_{\mathcal{E}_t}.$$

Based on the definitions $(a), (b)$ and (d), we have the following results:

Theorem 2.4. *Suppose that $(|\alpha_0| + |\alpha_1|) < 1$ and $(|\beta_0| + |\beta_1|) < 1$, the solutions defined in (5) satisfy the extended invertible definitions under l_1-norm, l_∞-norm, and mean square sense.*

(1) When $\min\{N, T\} \to \infty$ and $\log N = o(T)$,

$$\|\mathcal{E}_{\boldsymbol{\theta}(T|\mathbf{y}_0)} - \mathcal{E}_T\|_1 \to 0 \qquad a.s.$$

(2) When $\min\{N, T\} \to \infty$,

$$\|\mathcal{E}_{\boldsymbol{\theta}(T|\mathbf{y}_0)} - \mathcal{E}_T\|_\infty \to 0 \qquad a.s.$$

(3) When $\min\{N, T\} \to \infty$,

$$\frac{1}{NT}\sum_{t=1}^{T} \mathcal{E}^\top_{\boldsymbol{\theta}(t|\mathbf{y}_0)}\mathcal{E}_{\boldsymbol{\theta}(t|\mathbf{y}_0)} - \frac{1}{NT}\sum_{t=1}^{T}\mathcal{E}_t^\top\mathcal{E}_t \to 0 \qquad a.s.$$

The details of the proof are included in the Appendix. Based on Theorem 2.4, we prove the consistency of the $\mathcal{E}_{\boldsymbol{\theta}(t|\mathbf{y}_0)}$ and the consistency of the variance. Combining with Remarks 2.3 and 2.4, we can say that the NARMA model is invertible even in the diverging cases.

From the Appendix, it can be noted that the existence of $\mathcal{B}_0$ makes no difference in the estimation. Thus, we can omit it here. Define $\mathcal{E}_{\boldsymbol{\theta}(t|\mathbf{y}_0)}$ as follows,

$$\mathcal{E}_{\boldsymbol{\theta}(t|\mathbf{y}_0)} = \mathbb{Y}_t - \sum_{j=1}^{t-1}(-\mathcal{K})^{j-1}(\mathcal{K}+\mathcal{G})\mathbb{Y}_{t-j} - (-\mathcal{K})^{t-1}(\mathcal{K}+\mathcal{G})\mathbf{y}_0, \tag{6}$$

where $\mathbf{y}_0$ is the starting value. Then invertibility in variance can be written as

$$\frac{1}{NT}\sum_{t=1}^{T} \mathcal{E}^\top_{\boldsymbol{\theta}(t|\mathbf{y}_0)}\mathcal{E}_{\boldsymbol{\theta}(t|\mathbf{y}_0)} - \frac{1}{NT}\sum_{t=1}^{T}\mathcal{E}_{\boldsymbol{\theta}t}^\top\mathcal{E}_{\boldsymbol{\theta}t} \to 0, \qquad a.s.,$$

when $\min\{N, T\} \to \infty$. The proof of the invertibility of the NARMA model in variance is given in the Appendix.

2.4 Parameters estimation

Theorem 2.4(1)(invertibility in variance) shows that if we use the true parameters to get noises with an arbitrary $\mathbf{y}_0$, the estimated noises will be close to the real ones. Therefore, we can derive a least square estimation for the parameter vector $\boldsymbol{\theta}$. Define the sum of squares of "errors" as

$$V_T(\tilde{\boldsymbol{\theta}}) = \frac{1}{NT}\sum_{t=1}^{T} \mathcal{E}^\top_{\tilde{\boldsymbol{\theta}}(t|\mathbf{y}_0)}\mathcal{E}_{\tilde{\boldsymbol{\theta}}(t|\mathbf{y}_0)}, \tag{7}$$

where $\tilde{\boldsymbol{\theta}}$ is defined on a given set $\widetilde{\Theta}$. Following the idea of the least square estimation, we can minimize $V_T(\tilde{\boldsymbol{\theta}})$ to obtain an estimate of $\boldsymbol{\theta}$:

$$\hat{\boldsymbol{\theta}}_T = \arg\min_{\tilde{\boldsymbol{\theta}}\in\widetilde{\Theta}} V_T(\tilde{\boldsymbol{\theta}}). \tag{8}$$

Next, we discuss the asymptotic properties of $\hat{\boldsymbol{\theta}}_T$.

Theorem 2.5. *Let $\delta > 0$, if $(|\alpha_0| + |\alpha_1|) < 1 - \delta$ and $(|\beta_0|+|\beta_1|) < 1-\delta$, then the modified least squares estimator $\hat{\boldsymbol{\theta}}_T$ on $\widetilde{\Theta} = \{(\lambda, \gamma, \alpha_0, \alpha_1, \beta_0, \beta_1) : (|\alpha_0|+|\alpha_1|) \leq 1-\delta, (|\beta_0|+|\beta_1|) \leq 1-\delta\}$ is strongly consistent.*

Theorem 2.5 shows the consistency of the estimated $\hat{\boldsymbol{\theta}}_T$, which is obtained by minimizing $V_T(\tilde{\boldsymbol{\theta}})$. It provides an estimation together with theoretical guarantees for estimating the parameters of the NARMA model. Since the items containing λ and γ have no connection with the network, for simplicity, we omit them in the following. To prove Theorem 2.5, we first provide some useful lemmas in the Appendix. Theorem 2.5 can be easily proved by using these lemmas.

Remark 2.5. *Maximum likelihood estimation is not worth recommending here since this method is relatively complex with large estimation error in high-order case. Parameter estimation methods for ARMA model are given in some literatures([8], [15]) and their methods are interesting for our further studies.*

2.5 General NARMA(p) model

Note that the proposed model (1) only considers one lag. For simplicity, we refer to it as a NARMA (1) model. As a flexible extension, one could consider the NARMA (p) model as follows:

$$
\begin{aligned}
Y_{it} &= \lambda + Z_i^\top \gamma + \sum_{m=1}^{p} \frac{\beta_m}{n_i} \sum_{j=1}^{N} a_{ij} Y_{j(t-m)} \\
&\quad + \sum_{m=1}^{p} \frac{\alpha_m}{n_i} \sum_{j=1}^{N} a_{ij} \varepsilon_{j(t-m)} + \sum_{m=1}^{p} \phi_m Y_{i(t-m)} \\
&\quad + \sum_{m=1}^{p} \rho_m \varepsilon_{i(t-m)} + \varepsilon_{it}.
\end{aligned} \tag{9}
$$

Let $\mathbb{Y}_t^* = (\mathbb{Y}_t^\top, \mathbb{Y}_{t-1}^\top, \ldots, \mathbb{Y}_{t-p+1}^\top)^\top \in \mathbb{R}^{Np}$. Then the NARMA($p$) model (9) can be expressed in vector form as

$$
\mathbb{Y}_t^* = \mathcal{B}_0^* + \mathcal{G}^* \mathbb{Y}_{t-1}^* + \mathcal{K}^* \mathcal{E}_{t-1}^* + \mathcal{E}_t^*, \tag{10}
$$

with $\mathcal{B}_0^* = (\mathcal{B}_0^\top, \mathbf{0}_{N(p-1)}^\top)^\top \in \mathbb{R}^{Np}, \mathcal{E}_t^* = (\mathcal{E}_t^\top, \mathcal{E}_{t-1}^\top, \ldots, \mathcal{E}_{t-p+1}^\top)^\top \in \mathbb{R}^{Np}$ and

$$
\mathcal{G}^* = \begin{pmatrix} \Im & \beta_p W + \phi_p I_N \\ I_{N(p-1)} & O_{N(p-1),N} \end{pmatrix},
$$
$$
\mathcal{K}^* = \begin{pmatrix} \aleph & \alpha_p W + \rho_p I_N \\ I_{N(p-1)} & O_{N(p-1),N} \end{pmatrix},
$$

where $\Im = (\beta_1 W + \phi_1 I_N, \ldots, \beta_{p-1} W + \phi_{p-1} I_N) \in \mathbb{R}^{N \times N(p-1)}$, $\aleph = (\alpha_1 W + \rho_1 I_N, \ldots, \alpha_{p-1} W + \rho_{p-1} I_N) \in \mathbb{R}^{N \times N(p-1)}$, $\mathbf{0}_n$ is the n-dimensional zero vector, O_{n_1,n_2} is the $n_1 \times n_2$ dimensional zero matrix, and I_n is the $n \times n$ dimensional identity matrix. The NARMA(p) model has the same properties just like the NARMA(1) model and we give no more detailed description here.

3. NUMERICAL STUDIES

3.1 Simulation models

We demonstrate the finite sample performance of our proposed methodology by considering the three kinds of network model same as Zhu et al. [26]. Specifically, for each example, the random error ϵ_{it} is simulated from a standard normal distribution $N(0,1)$. We simulate a three dimensional covariate vector $Z_i = (Z_{i1}, Z_{i2}, Z_{i3}) \in \mathbb{R}^3$ from a multivariate normal distribution $MVN(\mathbf{0}, \Sigma_z)$, where the (i^{th}, j^{th}) element $\sigma_{ij} = 0.5^{|i-j|}$. For the nodal effect parameter γ, we set it as a fixed value $\gamma = (-0.5, 0.3, 0.8)^\top$. Lastly, we simulate the response vector at time 0 (i.e.$\mathbb{Y}_0$) from $N(0,1)$.

Example 1. (*Dyad Independence Model*). In the first example, we consider a simple case where the network structure is built on the relationships between pairs. We call the relationship between two nodes a dyad, $D_{ij} = (a_{ij}, a_{ji}), \forall 1 \leq i < j \leq N$, and the relationships between different dyads are independent. For the simulation probability of the mutually connected dyads (i.e., $D_{ij} = (1,1)$), we set it as $P(D_{ij} = (1,1)) = 20N^{-1}$ following Zhu et al. [26]. For the one-direction connected dyads (i.e., $D_{ij} = (1,0)$ or $D_{ij} = (0,1)$), we set the probability as $P(D_{ij} = (1,0)) = P(D_{ij} = (0,1)) = 0.5N^{-0.8}$. For dyad without connections (i.e., $D_{ij} = (0,0)$), the probability is $P(D_{ij} = (0,0)) = 1 - 20N^{-1} - N^{-0.8}$, which is very close to 1 for large N. These settings ensure the network sparsity. To avoid degenerate cases, we always set one element of the row equal to one if the elements of this row are all zero. Lastly, we investigate the cases that $T = 10, 50, 100$ and $(\lambda, \beta_0, \beta_1, \alpha_0, \alpha_1)^\top = (0.3, 0.5, 0.3, 0.5, 0.2)^\top$.

Example 2. (*Stochastic Block Model*). The stochastic block model is considered to be the major cause of financial contagion. In the stochastic block model, nodes within the same block have higher probability to be connected, which corresponds to the phenomenon in stock market that stocks in the same industry tend to connect more closely and form the block structure. For the simulation settings, we follow ([16]) and randomly assign a block label ($k = 1, \ldots, K$) to each node. We consider the cases where $K \in \{5, 10, 20\}$. For the connection probabilities within blocks and outside of blocks, we set them as $P(a_{ij} = 1) = 0.3N^{-0.3}$ and $P(a_{ij} = 1) = 0.3N^{-1}$ respectively, making the connections much denser between nodes with the same block labels. Lastly, we fix $T = 30$ and $(\lambda, \beta_0, \beta_1, \alpha_0, \alpha_1)^\top = (0.0, -0.2, 0.1, -0.1, 0.2)^\top$.

Example 3. (*Power-Law Distribution Model*). In reality, there is a common phenomenon that a few nodes have a great amount of connections while the rest of the majority only have a few connections. This phenomenon is modeled as the power-law distribution model in network science, which is also quite popular in financial studies. To simulate such network structure, we first generate in-degrees $d_i = \sum_j a_{ji}$ following a discrete power-law distribution $P(d_i = k) = ck^{-\eta}$, where c is a normalizing constant and we take $\eta \in \{1.2, 2.5, 5.0\}$. According to Gabaix [6], η equals to 3 fits the stock market really well. Therefore, we take the values of η around 3, with smaller values implying a heavier distribution tail. Then we randomly sample d_i nodes to be the ith node's followers. Lastly, we fix $T = 30$ and $(\lambda, \beta_0, \beta_1, \alpha_0, \alpha_1)^\top = (0.3, 0.5, -0.1, 0.4, 0.3)^\top$.

3.2 Performance measurements and simulation results

For each simulation example, a series of network sizes is considered (i.e., $N = 100, 500, 1000$). To obtain the modified least squares estimator described above, we use an optimal algorithm for implementation: we generate 100 random vectors and choose the one with the least residual. We repeat the generation and selection for $R = 300$ times and take the mean as our final estimation. We denote the optimal estimator in the rth replication as $\hat{\boldsymbol{\theta}}^{(r)} = (\hat{\lambda}^{(r)}, \hat{\gamma}^{(r)\top}, \hat{\beta}_0^{(r)}, \hat{\beta}_1^{(r)}, \hat{\alpha}_0^{(r)}, \hat{\alpha}_1^{(r)})^\top$, then the final estimation is given by $\hat{\boldsymbol{\theta}} = R^{-1} \sum_{r=1}^{R} \hat{\boldsymbol{\theta}}^{(r)}$. To gauge the performances of these estimators, we use the root mean square error (RMSE). The RMSE for the j^{th} estimator in $\boldsymbol{\theta}$ is given by $RMSE_j = \{R^{-1} \sum_{r=1}^{R} (\hat{\boldsymbol{\theta}}_j^{(r)} - \boldsymbol{\theta}_j)^2\}^{1/2}$. The total number of observed edges (i.e., $\sum_{i,j} a_{ij}$) and the network density(i.e., $\{N(N-1)\}^{-1} \sum_{i,j} a_{ij}$) are also reported.

The detailed simulation results are summarized in Table 1. For the first example (i.e., Table 1), we find that, if T is fixed, the RMSE values for all $\hat{\boldsymbol{\theta}}_j$'s decrease towards 0 as N increases. For example, $\hat{\beta}_0$(i.e., the estimated momentum effect) with $T = 30$, the RMSE value drops from 5.0% to 3.1%, as N increases from 100 to 1000. In the meanwhile, the network density drops from 21.13% to 2.32%, which means that the network structure is increasingly sparse. The $RMSE$ values are the same order compared with the simulation results of the NAR model. The main possible reason may be the optimization method could produce some error.

Remark 3.1. *From Table 2, we find that when $N = 1000$, the $RMSE$ of β_1 and α_1 are larger than those when $N = 500$. The $RMSE$s of λ in Table 1 has the same problem. This maybe incredible but can be explained. The network density rapidly decreases when the dimension N increases. This leads to fewer information and the increasement of the $RMSE$s.*

4. CASE STUDIES

In this section, we illustrate the performance of the proposed NARMA model on real data set. We focus on the application in financial data and consider the data from the Standard & Poor's 500 Index (S&P500) from January 2, 2018 to March 13, 2019. The data contains $N = 477$ stocks during the $T = 300$ trading days. The response Y_{it} is the daily stock return. We select three covariates which are related to a corporation's fundamentals: SIZE (i.e., the market cap), PE (i.e., price earnings ratio), and Beta (β or beta coefficient, a measure of the risk arising from exposure to general market movements as opposed to idiosyncratic factors).

We first conduct a preliminary descriptive analysis of the data. First, the average return during the analysis period, i.e., $T^{-1} \sum_t Y_{it}$ is calculated for each stock. This leads to the histogram in the left panel of Figure 1. The mean daily return of all stocks is given by -0.0035%. Next, the average daily return over all stocks, i.e., $N^{-1} \sum_i Y_{it}$ are calculated for each day. This yields the line chart in the right panel of Figure 1. A higher volatility level can be captured at the beginning and the end of 2018.

The network relationship can be constructed by using the top ten shareholder information of stocks ([28]). However, when we study the stocks in S&P500, unlike the Chinese stock market, some of the investment institutions and fund companies in the U.S. hold the stock of almost every large company, such as the Vanguard Group Inc., the Blackrock Inc., the State Street Corporation, the FMR, LLC and so on. Here in our study, we let $a_{ij} = 1$ if the ith and jth stock share at least five shareholders in top ten, otherwise $a_{ij} = 0$. The histogram of the degree of the adjacency matrix is given in the left panel of Figure 2. Thereby, the network structure graph can be obtained. This yields the right panel of Figure 2.

We next fit the data to the NARMA model. We adopt a sliding window approach to evaluate the model performance. The time window for model training is set to be $T_{train} = 300$ days. The subsequent $T_{test} = 5$ days are employed for model testing.

The parameter estimation results are given in Table 4. Firstly, it is obviously that the lagged error term is in the same order as the lagged value term which means the error term should not be cancelled. Secondly, the network effects β_1, α_1 are both significant which states that the network structure exists in financial data.

Then we compare the NARMA model with the NAR model ([26]) in terms of the prediction accuracy. The results are given in Figure 3. From this figure, we could find that the NARMA model is better than the NAR model most of the time. However, the performance of the NARMA model is not significantly better than the NAR model. This might due to the unperfect fit of the stocks' return data to the ARMA structure. Next we pay attention to the estimated parameters. For the selected three covariates, except for the coefficient of Beta, the other two coefficients in the NARMA model and the NAR model are both close to zero, which

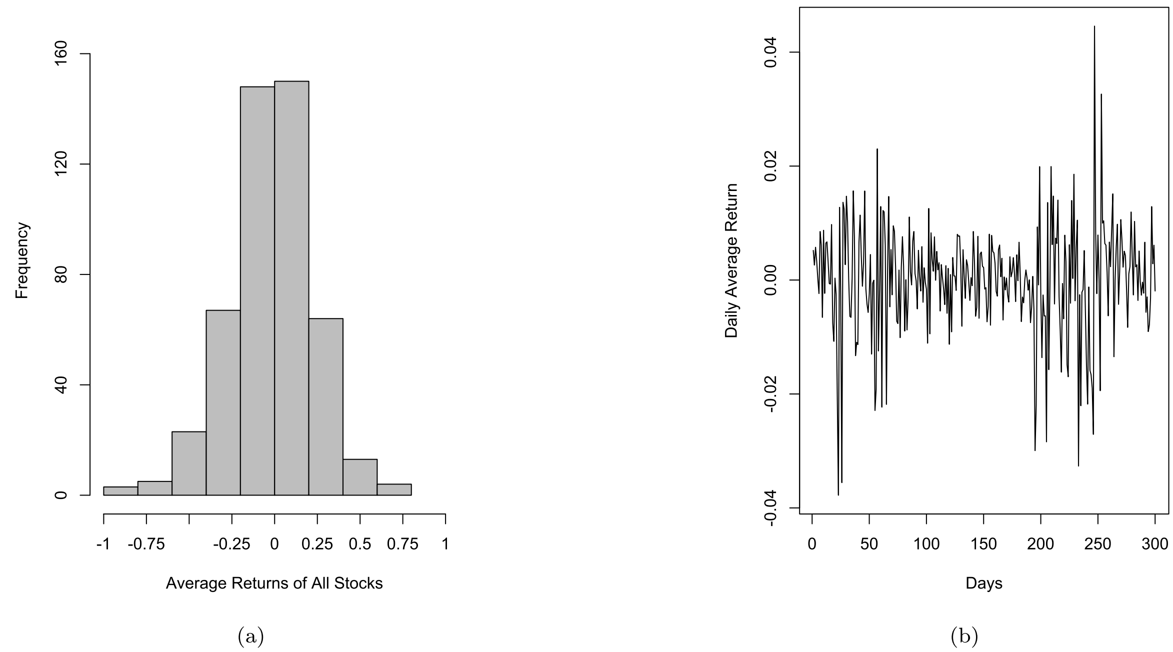

Figure 1. Left panel: the histogram of the average return of $N = 477$ stocks; right panel: the line plot of daily average returns of $N = 477$ stocks.

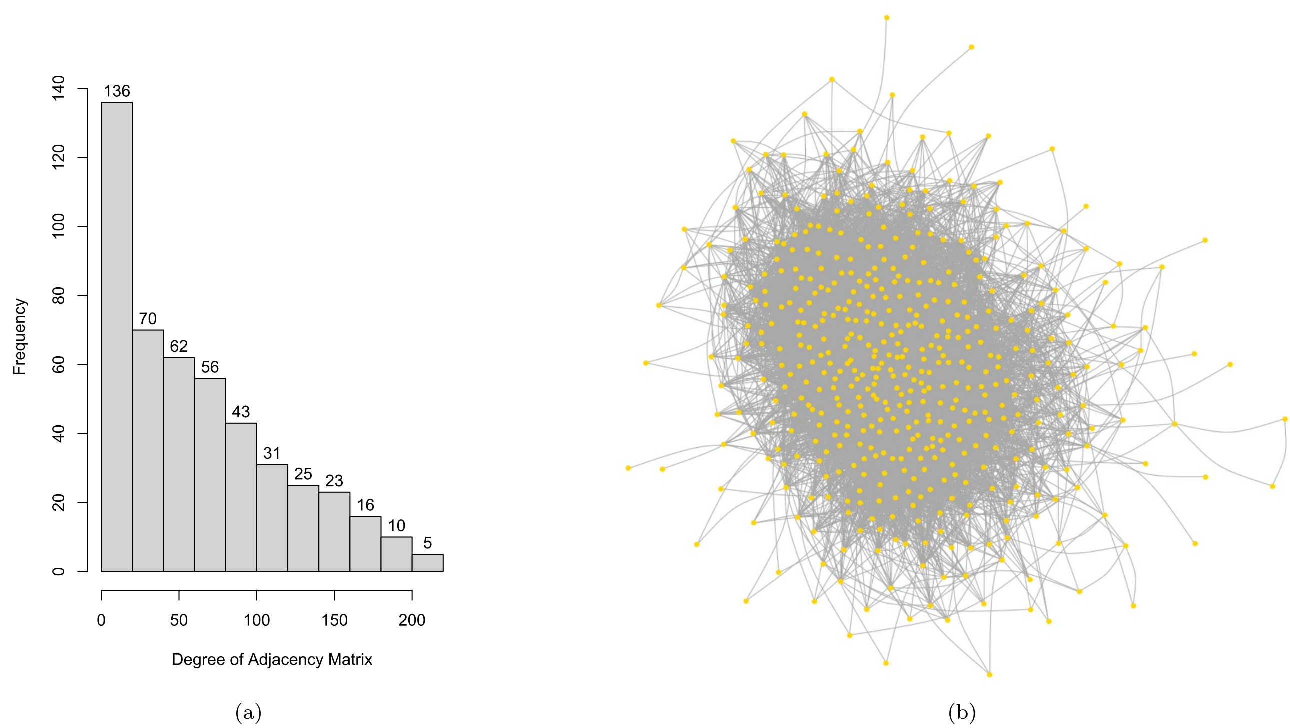

Figure 2. Left panel: the histogram of the degree of adjacency matrix; right panel: the graph of the network structure.

Table 1. Simulation results for Example 1. The parameter estimation is given. The RMSE values ($\times 10^{-2}$) are reported for every parameter estimates in parentheses. Total number of observed edges (TNOE) and network density (ND) are also reported

	T=10			T=30			T=100		
	N=100	N=500	N=1000	N=100	N=500	N=1000	N=100	N=500	N=1000
λ	0.281(15.4)	0.291(10.8)	0.323(12.3)	0.278(15.4)	0.298(11.4)	0.332(11.1)	0.343(15.3)	0.286(9.0)	0.336(10.2)
γ_1	−0.373(12.0)	−0.376(9.8)	−0.486(8.9)	−0.421(9.0)	−0.439(6.6)	−0.506(5.4)	−0.498(5.5)	−0.487(4.1)	−0.494(3.9)
γ_2	0.321(14.7)	0.277(11.0)	0.222(9.9)	0.363(11.1)	0.275(6.9)	0.275(6.4)	0.320(6.0)	0.295(4.7)	0.287(4.1)
γ_3	0.701(17.4)	0.729(11.1)	0.700(11.1)	0.745(12.1)	0.783(8.9)	0.766(7.5)	0.783(7.3)	0.799(5.4)	0.790(4.8)
β_0	0.599(6.7)	0.549(5.0)	0.532(4.9)	0.522(5.0)	0.511(3.8)	0.510(3.1)	0.502(3.3)	0.500(2.4)	0.504(2.0)
β_1	0.370(20.0)	0.348(12.8)	0.305(12.0)	0.325(12.8)	0.302(8.5)	0.275(7.8)	0.271(9.5)	0.310(6.2)	0.278(6.2)
α_0	0.454(8.0)	0.408(5.8)	0.420(5.0)	0.488(4.7)	0.474(3.4)	0.472(3.1)	0.489(3.1)	0.491(2.3)	0.490(1.9)
α_1	0.338(21.3)	0.328(18.9)	0.331(17.3)	0.305(14.0)	0.277(14.2)	0.267(12.6)	0.278(10.9)	0.257(10.4)	0.251(9.1)
TNOE	2113	11547	23155	2113	11547	23155	2113	11547	23155
ND(%)	21.13	4.42	2.32	21.13	4.42	2.32	21.13	4.42	2.32

Table 2. Simulation results for Example 2. The parameter estimation is given. The RMSE values ($\times 10^{-2}$) are reported for every parameter estimates in parentheses. Total number of observed edges (TNOE) and network density (ND) are also reported

	K=5			K=10			K=20		
	N=100	N=500	N=1000	N=100	N=500	N=1000	N=100	N=500	N=1000
λ	0.024(8.5)	0.001(4.2)	0.005(3.9)	−0.014(6.6)	0.006(4.5)	0.011(4.1)	0.032(8.7)	−0.001(4.6)	0.019(4.1)
γ_1	−0.534(8.5)	−0.509(6.0)	−0.544(5.4)	−0.548(9.5)	−0.505(6.6)	−0.536(5.6)	−0.498(9.7)	−0.526(6.1)	−0.527(5.2)
γ_2	0.387(9.1)	0.309(5.5)	0.301(5.2)	0.338(10.4)	0.305(5.9)	0.327(5.3)	0.327(9.4)	0.325(6.0)	0.319(5.1)
γ_3	0.899(12.6)	0.858(8.6)	0.843(7.9)	0.887(12.9)	0.856(9.3)	0.842(7.8)	0.834(12.1)	0.867(8.9)	0.862(7.5)
β_0	−0.338(13.7)	−0.285(11.3)	−0.283(10.1)	−0.321(13.8)	−0.279(11.7)	−0.269(10.2)	−0.293(13.8)	−0.295(10.6)	−0.290(9.3)
β_1	0.095(12.8)	0.102(12.6)	0.108(15.2)	0.146(10.6)	0.100(9.1)	0.086(9.9)	0.105(8.6)	0.114(7.6)	0.112(7.8)
α_0	0.088(15.1)	0.010(12.2)	0.002(11.0)	0.029(15.1)	0.004(12.7)	−0.015(10.7)	0.018(14.0)	0.008(11.8)	−0.002(9.8)
α_1	0.210(16.2)	0.196(17.5)	0.195(19.4)	0.173(13.4)	0.208(13.2)	0.242(13.9)	0.241(11.8)	0.195(10.1)	0.207(10.4)
TNOE	291	2901	8770	195	1772	5094	159	1244	3226
ND(%)	2.91	1.16	0.88	1.95	0.71	0.51	1.59	0.50	0.32

Table 3. Simulation results for Example 3. The parameter estimation is given. The RMSE values ($\times 10^{-2}$) are reported for every parameter estimates in parentheses. Total number of observed edges (TNOE) and network density (ND) are also reported

	η=1.2			η=2.5			η=5		
	N=100	N=500	N=1000	N=100	N=500	N=1000	N=100	N=500	N=1000
λ	0.344(11.5)	0.316(7.5)	0.329(6.0)	0.388(13.3)	0.327(7.9)	0.327(7.0)	0.382(13.0)	0.320(9.0)	0.330(7.6)
γ_1	−0.461(10.9)	−0.505(7.0)	−0.495(5.7)	−0.402(12.0)	−0.444(7.8)	−0.464(6.4)	−0.402(12.0)	−0.463(8.6)	−0.469(7.1)
γ_2	0.394(11.3)	0.300(7.9)	0.283(6.1)	0.361(13.1)	0.279(8.9)	0.291(7.2)	0.354(12.3)	0.277(9.0)	0.294(7.8)
γ_3	0.803(14.6)	0.805(9.6)	0.803(7.9)	0.801(15.9)	0.794(10.8)	0.805(9.5)	0.779(16.5)	0.801(11.4)	0.788(9.9)
β_0	0.497(6.2)	0.508(4.2)	0.497(3.7)	0.507(6.5)	0.507(4.5)	0.503(4.0)	0.516(6.6)	0.513(5.0)	0.513(4.3)
β_1	−0.082(6.4)	−0.093(3.8)	−0.104(3.5)	−0.072(4.2)	−0.088(2.4)	−0.095(1.9)	−0.111(4.1)	−0.093(2.3)	−0.095(1.9)
α_0	0.414(6.1)	0.369(4.5)	0.386(3.8)	0.415(6.5)	0.384(4.7)	0.380(4.0)	0.406(6.4)	0.381(4.7)	0.377(4.2)
α_1	0.293(9.4)	0.293(6.2)	0.301(5.7)	0.262(6.3)	0.268(4.6)	0.280(3.5)	0.271(6.3)	0.284(4.3)	0.282(3.6)
TNOE	1276	21111	72177	173	902	1835	104	519	1038
ND(%)	12.76	8.44	7.21	1.73	0.36	0.18	1.04	0.21	0.10

Table 4. Parameter estimation for the S&P500 data

λ	β_0	β_1	α_0	α_1
0.0136	0.9628	0.9119	0.1199	0.3842

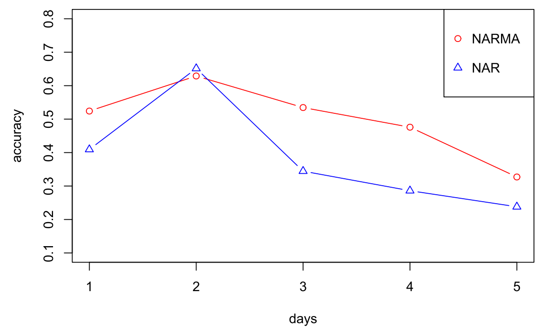

Figure 3. The accuracy of the NARMA model(red line) and the NAR model(blue line). The NARMA model could achieve higher accuracy compared with the NAR model most of the time.

means the market cap and the price earning ratios have little influence on the prediction. This may indicate that during the period we choose, the trend of the returns is mainly driven by the general market movements.

5. CONCLUDING REMARKS

In this article, we propose a network ARMA model which considers the network structure information. Compared with the classical ARMA model, NARMA model decreases the computational complexity significantly due to the dimension reduction benefited from the incorporation of network structure. We provide theoretical guarantees for the consistency of the proposed estimators, which have been confirmed by extensive numerical studies. We further illustrate the usefulness of our model using a real data set from S&P500 index. A significant network structure term is detected.

To conclude this work, we discuss some potential extensions for future study. First, the NARMA model proposed here applies to the case where the responses are continuous. However, discrete responses are ubiquitous in real practice. Thus, it is of great significance to extend the NARMA model into noncontinuous cases. Second, the network structure is assumed to be static through out the model specification. This assumption is not so valid in reality, in which the network changes and evolves as time progress. Hence it is worth studying how to model a dynamic network structure in the time series models. Third, researchers often use optimization methods to estimate an ARMA-like model. Fitting a model in such a way increases the computational cost and induces slow convergence. Therefore, a lightweight and fast estimation method specifically designed for ARMA-like models is in great need. Lastly, in this article, we only prove the consistency of the NARMA model, while the asymptotic properties are yet to be confirmed. The main difficulties lie in the complex solutions of $\mathbb{Y}_t$ in our model. How to give a brief expression of the asymptotic properties is awaited for deeper investigation in the future.

APPENDIX: SUPPLEMENTARY MATERIAL

This is a supplementary material that contains the proofs of Theorem 1, Theorem 2, Theorem 4 and some lemmas.

A.1 Proof of Theorem 2.1

Proof of Theorem 2.1. Denote $\lambda_i(M)$ by the ith eigenvalue of any arbitrary matrix $M \in \mathbb{R}^{N\times N}$. We first verify that the solution satisfies strict stationarity. To this end, note that since the absolute sums of W's rows are equal to 1, then $\max_i |\lambda_i(W)| \le 1$. Also we have

$$\rho = \max_{1\le i<N} |\lambda_i(\mathcal{G})| \le |\beta_1| \max_{1\le i<N} |\lambda_i(W)| + |\beta_0| < 1. \tag{11}$$

It holds that $\lim_{m\to\infty} \sum_{j=1}^m \mathcal{G}^{j-1}(\mathcal{K}+\mathcal{G})\mathcal{E}_{t-j}$ exists, and then $\{\mathbb{Y}_t\}$ is a strictly stationary process. It is straightforward to verify that $\{\mathbb{Y}_t\}$ satisfies the NARMA model.

Next, we verify the uniqueness of the strictly stationarity solution. Assume that $\{\widetilde{\mathbb{Y}}_t\}$ is another strictly stationary solution to the NARMA model with $E\|\widetilde{\mathbb{Y}}_t\| < \infty$. Then

$$\widetilde{\mathbb{Y}}_t = \sum_{j=1}^{m-1} \mathcal{G}^{j-1}(\mathcal{K}+\mathcal{G})\Big(\mathcal{B}_0 + \mathcal{E}_{t-j}\Big) + \mathcal{G}^{m-1}(\mathcal{K}+\mathcal{G})\widetilde{\mathbb{Y}}_{t-m},$$

for any positive integer m. Hence by (11),

$$E\|\mathbb{Y}_t - \widetilde{\mathbb{Y}}_t\| = E\Big\| \sum_{j=m}^{\infty} \mathcal{G}^{j-1}(\mathcal{K}+\mathcal{G})\Big(\mathcal{B}_0 + \mathcal{E}_{t-j}\Big) - \mathcal{G}^{m-1}(\mathcal{K}+\mathcal{G})\widetilde{\mathbb{Y}}_{t-m}\Big\| \le C\rho^m,$$

where C is a constant independent of t and m. Note that m is chosen arbitrarily. Hence, we have that $E\|\mathbb{Y}_t - \widetilde{\mathbb{Y}}_t\| = 0$, *i.e.*, so $\mathbb{Y}_t = \widetilde{\mathbb{Y}}_t$ with probability one. The proof of Theorem 1 is completed. □

A.2 Proof of Theorem 2.2

Proof of Theorem 2.2. Note that $\{\mathbb{Y}_t\}$ satisfies the NARMA model for any N. To prove the existence of a stationary solution, it is sufficient to show that $\{\mathbb{Y}_t\}$ is strictly stationary according to the definition.

Define $|M|_e$ as $|M|_e = (|m_{ij}|) \in \mathbb{R}^{n\times p}$ for any arbitrary matrix $M = (m_{ij}) \in \mathbb{R}^{n\times p}$. Moreover, for matrices $M_1 = (m_{ij}^{(1)}) \in \mathbb{R}^{n\times p}$ and $M_2 = (m_{ij}^{(2)}) \in \mathbb{R}^{n\times p}$, define $M_1 \preccurlyeq M_2$ as $m_{ij}^{(1)} \le m_{ij}^{(2)}$ for $1 \le i \le n$ and $1 \le j \le p$. Noting that

$$E|\mathcal{B}_0 + (\mathcal{K}+\mathcal{G})\mathcal{E}_{t-j}|_e \preccurlyeq (|\lambda| + E|Z_I^\top \gamma| + |\mathcal{K}+\mathcal{G}|_e E|\varepsilon_{it}|)\mathbf{1}$$

and

$$|\mathcal{G}|_e^j \mathbf{1} = (|\beta_1|W + |\beta_0|I)^j \mathbf{1} \preccurlyeq (|\beta_0| + |\beta_1|)^j \mathbf{1},$$

we have,

$$\begin{aligned} &E|\mathbf{w}_N^\top(\sum_{j=1}^\infty \mathcal{G}^j(\mathcal{B}_0 + (\mathcal{K}+\mathcal{G})\mathcal{E}_{t-j}) + \mathcal{E}_t)| \\ &\le \sum_{i=1}^\infty |\omega_i| \sum_{j=1}^\infty (|\beta_0| + |\beta_1|)^j (|\lambda| + E|Z_i^\top\gamma| + |\mathcal{K}+\mathcal{G}|_e E|\varepsilon_{it}|) \\ &< \infty, \end{aligned}$$

which implies that

$$\lim_{N\to\infty} \mathbf{w}_N^\top \mathbb{Y}_t = \lim_{N\to\infty} \mathbf{w}_N^\top(\sum_{j=1}^\infty \mathcal{G}^j(\mathcal{B}_0 + (\mathcal{K}+\mathcal{G})\mathcal{E}_{t-j}) + \mathcal{E}_t)$$

exists with probability one. Let $Y_t^\omega = \lim_{N\to\infty} \mathbf{w}_N^\top \mathbb{Y}_t$, and it is obvious that $\{Y_t^\omega\}$ is strictly stationary. Hence, $\{\mathbb{Y}_t\}$ is strictly stationary according to the definition.

Next, we verify the uniqueness of the strictly stationary solution. Assume that $\{\widetilde{\mathbb{Y}}_t\}$ is another strictly stationary solution to the NARMA model with finite first order moment. Therefore, $E|\widetilde{\mathbb{Y}}_t|_e \preccurlyeq C_1\mathbf{1}$ for some constant C_1. Then for any N and weight ω, we have

$$\begin{aligned} &E|\mathbf{w}_N^\top \mathbb{Y}_t - \mathbf{w}_N^\top \widetilde{\mathbb{Y}}_t| \\ &= E|\sum_{j=m}^\infty \mathbf{w}_N^\top \mathcal{G}^j(\mathcal{B}_0 + (\mathcal{K}+\mathcal{G})\mathcal{E}_{t-j}) - \mathbf{w}_N^\top \mathcal{G}^m \widetilde{\mathbb{Y}}_{t-m}| \\ &\le C_2 \sum_{j=m}^\infty \{(|\beta_0| + |\beta_1|)^j + C_1(|\beta_0| + |\beta_1|)^m\} \sum_{i=1}^\infty |\omega_i|, \end{aligned}$$

where C_2 is a constant. Consequently, by the arbitrary specification of m, we have $Y_t^\omega = \widetilde{Y}_t^\omega$ with probability one. This completes the proof of Theorem 2.2. □

A.3 Proof of Theorem 2.3

Proof of Theorem 2.3. Let $\hat{\mathcal{E}}_t$ be an estimate of $\mathcal{E}_t$ satisfying the equation

$$\mathbb{Y}_t = \mathcal{B}_0 + \mathcal{G}\mathbb{Y}_{t-1} + \mathcal{K}\hat{\mathcal{E}}_{t-1} + \hat{\mathcal{E}}_t,$$

where the initial value $\hat{\mathcal{E}}_t, t \le 0$ can be arbitrarily chosen. Let $\widetilde{\mathcal{E}}_t = \mathcal{E}_t - \hat{\mathcal{E}}_t$, we have

$$\widetilde{\mathcal{E}}_t = -\mathcal{K}\widetilde{\mathcal{E}}_{t-1},$$

where $\widetilde{\mathcal{E}}_t$ is the estimated error. Following Granger and Andersen [7], we say that model 2 is invertible if $\widetilde{\mathcal{E}}_t$ converges to 0 in mean square as t tends to infinity for any initial values. Then

$$\begin{aligned} E\left|\widetilde{\mathcal{E}}_t\widetilde{\mathcal{E}}_t^\top\right| &= E\left|\mathcal{K}\widetilde{\mathcal{E}}_{t-1}\widetilde{\mathcal{E}}_{t-1}^\top\mathcal{K}^\top\right| \\ &\le \mathcal{K}E\left|\widetilde{\mathcal{E}}_{t-1}\widetilde{\mathcal{E}}_{t-1}^\top\right|\mathcal{K}^\top. \end{aligned}$$

Setting $M_t = \text{Vec}\left(E\left|\widetilde{\xi}_t\widetilde{\xi}_t^T\right|\right)$, we have

$$|M_t| \leq (|\mathcal{K}| \otimes |\mathcal{K}|)\, |M_{t-1}|\,.$$

Since $\max_i |\lambda_i(W)| \leq 1$. Also we have

$$\phi = \max_{1\leq i<N} |\lambda_i(\mathcal{K})| \leq |\alpha_1| \max_{1\leq i<N} |\lambda_i(W)| + |\alpha_0| < 1. \tag{12}$$

It follows the sufficient condition of invertibility that the maximum eigenvalue in absolute value of $\mathcal{K}$ is less than unity(See Lemma 2.3(a) of Ling [12]). Thus the invertibility is proved. It is straightforward to verify that the $\{\mathcal{E}_t\}$ given in Eq. 4 satisfies the NARMA model.

Next, we verify the uniqueness of the invertible solution. Assume that $\{\mathcal{E}_t^*\}$ is another invertible solution to the NARMA model with $E\|\mathcal{E}_t^*\| < \infty$. Then

$$\mathcal{E}_t^* = \sum_{j=1}^{m-1} (-\mathcal{K})^{j-1}(\mathcal{K}+\mathcal{G})\Big(\mathcal{B}_0 + \mathbb{Y}_{t-j}\Big) + (-\mathcal{K})^{m-1}(\mathcal{K}+\mathcal{G})\mathcal{E}_{t-m}^*,$$

for any positive integer m. Hence by (12),

$$E\|\mathcal{E}_t - \mathcal{E}_t^*\| = E\Big\| \sum_{j=m}^{\infty} (-\mathcal{K})^{j-1}(\mathcal{K}+\mathcal{G})\Big(\mathcal{B}_0 + \mathbb{Y}_{t-j}\Big) - (-\mathcal{K})^{m-1}(\mathcal{K}+\mathcal{G})\mathcal{E}_{t-m}^*\Big\| \leq C\phi^m,$$

where C is a constant independent of t and m. Note that m is chosen arbitrarily. Hence, we have that $E\|\mathcal{E}_t - \mathcal{E}_t^*\| = 0$, *i.e.*, so $\mathcal{E}_t = \mathcal{E}_t^*$ with probability one. The proof of Theorem 2.3 is completed. □

A.4 Proof of the Theorem 4

To prove the invertibility, it is obvious that $\mathcal{B}_0$ does not influence our work, so for simplification, $\mathcal{B}_0$ is omitted here.

Proof of Theorem 2.4 (1). Note that $\mathcal{E}_{\boldsymbol{\theta}(T|\mathbf{y}_0)} - \mathcal{E}_T = \sum_{j=T}^{\infty}(-\mathcal{K})^{j-1}(\mathcal{K}+\mathcal{G})\mathbb{Y}_{T-j}$. Since $\mathbb{Y}_{T-j}$ is the return in our article, so there exists $|\mathbb{Y}_{T-j}| \leq M$. From Lemma 7, it shows that $|(-\mathcal{K})^{j-1}(\mathcal{K}+\mathcal{G})\mathbb{Y}_{T-j}| \leq CM\rho^{j-1}\mathbf{1}$, where $0<\rho<1$. Therefore

$$\begin{aligned}\|\mathcal{E}_{\boldsymbol{\theta}(T|\mathbf{y}_0)} - \mathcal{E}_T\|_1 &= \Big\| \sum_{j=T}^{\infty} (-\mathcal{K})^{j-1}(\mathcal{K}+\mathcal{G})\mathbb{Y}_{T-j}\Big\|_1 \\ &\leq \Big\| \sum_{j=T}^{\infty} CM\rho^{j-1}\mathbf{1}\Big\|_1 = N\Big|\sum_{j=T}^{\infty} CM\rho^{j-1}\Big| \\ &= CMN\frac{\rho^{T-1}}{1-\rho}.\end{aligned}$$

Since $N, T \to \infty$ and $\log N = o(T)$, thus

$$\|\mathcal{E}_{\boldsymbol{\theta}(T|\mathbf{y}_0)} - \mathcal{E}_T\|_1 = CMN\frac{\rho^{T-1}}{1-\rho} \sim o(1).$$

So the proof has been completed. □

Proof of Theorem 2.4 (2). From Theorem 2.4 (1), there exists $\mathcal{E}_{\boldsymbol{\theta}(T|\mathbf{y}_0)} - \mathcal{E}_T = \sum_{j=T}^{\infty}(-\mathcal{K})^{j-1}(\mathcal{K}+\mathcal{G})\mathbb{Y}_{T-j}$ and $|(-\mathcal{K})^{j-1}(\mathcal{K}+\mathcal{G})\mathbb{Y}_{T-j}| \leq CM\rho^{j-1}\mathbf{1}$, thus

$$\begin{aligned}\|\mathcal{E}_{\boldsymbol{\theta}(T|\mathbf{y}_0)} - \mathcal{E}_T\|_\infty &= \Big\| \sum_{j=T}^{\infty} (-\mathcal{K})^{j-1}(\mathcal{K}+\mathcal{G})\mathbb{Y}_{T-j}\Big\|_\infty \\ &\leq \Big\| \sum_{j=T}^{\infty} CM\rho^{j-1}\mathbf{1}\Big\|_\infty \\ &= CM\frac{\rho^{T-1}}{1-\rho} \longrightarrow 0,\end{aligned}$$

when $N, T \to \infty$. So the proof has been completed. □

Proof of Theorem 2.4 (3). To prove this result, first we should calculate each term of $\frac{1}{NT}\sum_{t=1}^{T}\mathcal{E}_{\boldsymbol{\theta}(t|\mathbf{y}_0)}^\top\mathcal{E}_{\boldsymbol{\theta}(t|\mathbf{y}_0)}$ and $\frac{1}{NT}\sum_{t=1}^{T}\mathcal{E}_t^\top\mathcal{E}_t$ respectively in the following:

$$\begin{aligned}&\frac{1}{NT}\sum_{t=1}^{T}\mathcal{E}_{\boldsymbol{\theta}(t|\mathbf{y}_0)}^\top\mathcal{E}_{\boldsymbol{\theta}(t|\mathbf{y}_0)} \\ &= \frac{1}{NT}\sum_{t=1}^{T}\Bigg\{\Big(\mathbb{Y}_t - \sum_{j=1}^{t-1}(-\mathcal{K})^{j-1}(\mathcal{K}+\mathcal{G})\mathbb{Y}_{t-j}\Big)^\top \Big(\mathbb{Y}_t - \sum_{j=1}^{t-1}(-\mathcal{K})^{j-1}(\mathcal{K}+\mathcal{G})\mathbb{Y}_{t-j}\Big) \\ &\quad - \Big(\mathbb{Y}_t - \sum_{j=1}^{t-1}(-\mathcal{K})^{j-1}(\mathcal{K}+\mathcal{G})\mathbb{Y}_{t-j}\Big)^\top \Big((-\mathcal{K})^{t-1}(\mathcal{K}+\mathcal{G})\mathbf{y}_0\Big) \\ &\quad - \Big((-\mathcal{K})^{t-1}(\mathcal{K}+\mathcal{G})\mathbf{y}_0\Big)^\top \Big(\mathbb{Y}_t - \sum_{j=1}^{t-1}(-\mathcal{K})^{j-1}(\mathcal{K}+\mathcal{G})\mathbb{Y}_{t-j}\Big) \\ &\quad + \Big((-\mathcal{K})^{t-1}(\mathcal{K}+\mathcal{G})\mathbf{y}_0\Big)^\top \Big((-\mathcal{K})^{t-1}(\mathcal{K}+\mathcal{G})\mathbf{y}_0\Big)\Bigg\},\end{aligned}$$

and

$$\begin{aligned}&\frac{1}{NT}\sum_{t=1}^{T}\mathcal{E}_t^\top\mathcal{E}_t \\ &= \frac{1}{NT}\sum_{t=1}^{T}\Bigg\{\Big(\mathbb{Y}_t - \sum_{j=1}^{t-1}(-\mathcal{K})^{j-1}(\mathcal{K}+\mathcal{G})\mathbb{Y}_{t-j}\Big)^\top \Big(\mathbb{Y}_t - \sum_{j=1}^{t-1}(-\mathcal{K})^{j-1}(\mathcal{K}+\mathcal{G})\mathbb{Y}_{t-j}\Big)\end{aligned}$$

$$-\Big(\mathbb{Y}_t-\sum_{j=1}^{t-1}(-\mathcal{K})^{j-1}(\mathcal{K}+\mathcal{G})\mathbb{Y}_{t-j}\Big)^\top \Big(\sum_{j=t}^{\infty}(-\mathcal{K})^{j-1}(\mathcal{K}+\mathcal{G})\mathbb{Y}_{t-j}\Big)$$
$$-\Big(\sum_{j=t}^{\infty}(-\mathcal{K})^{j-1}(\mathcal{K}+\mathcal{G})\mathbb{Y}_{t-j}\Big)^\top \Big(\mathbb{Y}_t-\sum_{j=1}^{t-1}(-\mathcal{K})^{j-1}(\mathcal{K}+\mathcal{G})\mathbb{Y}_{t-j}\Big)$$
$$+\Big(\sum_{j=t}^{\infty}(-\mathcal{K})^{j-1}(\mathcal{K}+\mathcal{G})\mathbb{Y}_{t-j}\Big)^\top \Big(\sum_{j=t}^{\infty}(-\mathcal{K})^{j-1}(\mathcal{K}+\mathcal{G})\mathbb{Y}_{t-j}\Big)\Bigg\}.$$

In order to prove is $\frac{1}{NT}\sum_{t=1}^{T}\mathcal{E}_{\boldsymbol{\theta}(t|\mathbf{y}_0)}^\top\mathcal{E}_{\boldsymbol{\theta}(t|\mathbf{y}_0)}-\frac{1}{NT}\sum_{t=1}^{T}\mathcal{E}_t^\top\mathcal{E}_t\to 0\ a.s.$, we introduce some notations for convenience:

$$A:=\frac{1}{NT}\sum_{t=1}^{T}\Big(\mathbb{Y}_t-\sum_{j=1}^{t-1}(-\mathcal{K})^{j-1}(\mathcal{K}+\mathcal{G})\mathbb{Y}_{t-j}\Big)^\top \Big((-\mathcal{K})^{t-1}(\mathcal{K}+\mathcal{G})\mathbf{y}_0\Big),$$
$$B:=\frac{1}{NT}\sum_{t=1}^{T}\Big((-\mathcal{K})^{t-1}(\mathcal{K}+\mathcal{G})\mathbf{y}_0\Big)^\top \Big(\mathbb{Y}_t-\sum_{j=1}^{t-1}(-\mathcal{K})^{j-1}(\mathcal{K}+\mathcal{G})\mathbb{Y}_{t-j}\Big),$$
$$C:=\frac{1}{NT}\sum_{t=1}^{T}\Big((-\mathcal{K})^{t-1}(\mathcal{K}+\mathcal{G})\mathbf{y}_0\Big)^\top \Big((-\mathcal{K})^{t-1}(\mathcal{K}+\mathcal{G})\mathbf{y}_0\Big),$$
$$A':=\frac{1}{NT}\sum_{t=1}^{T}\Big(\mathbb{Y}_t-\sum_{j=1}^{t-1}(-\mathcal{K})^{j-1}(\mathcal{K}+\mathcal{G})\mathbb{Y}_{t-j}\Big)^\top \Big(\sum_{j=t}^{\infty}(-\mathcal{K})^{j-1}(\mathcal{K}+\mathcal{G})\mathbb{Y}_{t-j}\Big),$$
$$B':=\frac{1}{NT}\sum_{t=1}^{T}\Big(\sum_{j=t}^{\infty}(-\mathcal{K})^{j-1}(\mathcal{K}+\mathcal{G})\mathbb{Y}_{t-j}\Big)^\top \Big(\mathbb{Y}_t-\sum_{j=1}^{t-1}(-\mathcal{K})^{j-1}(\mathcal{K}+\mathcal{G})\mathbb{Y}_{t-j}\Big),$$
$$C':=\frac{1}{NT}\sum_{t=1}^{T}\Big(\sum_{j=t}^{\infty}(-\mathcal{K})^{j-1}(\mathcal{K}+\mathcal{G})\mathbb{Y}_{t-j}\Big)^\top \Big(\sum_{j=t}^{\infty}(-\mathcal{K})^{j-1}(\mathcal{K}+\mathcal{G})\mathbb{Y}_{t-j}\Big).$$

So after eliminating the same term, what need to be proved is

$$(A-A')+(B-B')-(C-C')\longrightarrow 0,\qquad a.s.$$

To prove the left formula converges to zero, it is sufficient to prove that the three terms of the left formula respectively converge to zero. So we only give the proof of the first term of the left converges to zero and the rest two are just the same.

From the above, we need to prove $A-A'\to 0\ a.s.$ In other words, we should prove

(13)
$$\frac{1}{NT}\sum_{t=1}^{T}\Big(\mathbb{Y}_t-\sum_{j=1}^{t-1}(-\mathcal{K})^{j-1}(\mathcal{K}+\mathcal{G})\mathbb{Y}_{t-j}\Big)^\top \Big((-\mathcal{K})^{t-1}(\mathcal{K}+\mathcal{G})\mathbf{y}_0-\sum_{j=t}^{\infty}(-\mathcal{K})^{j-1}(\mathcal{K}+\mathcal{G})\mathbb{Y}_{t-j}\Big) \longrightarrow 0,\ a.s.$$

Without loss of generality, we can assume $\mathbf{y}_0=\mathbf{0}$. Noting that

$$\Big(\mathbb{Y}_t-\sum_{j=1}^{t-1}(-\mathcal{K})^{j-1}(\mathcal{K}+\mathcal{G})\mathbb{Y}_{t-j}\Big) -\sum_{j=t}^{\infty}(-\mathcal{K})^{j-1}(\mathcal{K}+\mathcal{G})\mathbb{Y}_{t-j}=\mathcal{E}_t$$

and

$$\mathbb{Y}_t=\mathcal{E}_t+\sum_{j=1}^{\infty}\mathcal{G}^{j-1}(\mathcal{K}+\mathcal{G})\mathcal{E}_{t-j},$$

substituting them into Eq. (13), then what we need to prove can be written as

$$\frac{1}{NT}\sum_{t=1}^{T}\Big(\mathcal{E}_t+(-\mathcal{K})^{t-1}\sum_{m=1}^{\infty}\mathcal{G}^{m-1}(\mathcal{K}+\mathcal{G})\mathcal{E}_{1-m}\Big)^\top \Big((-\mathcal{K})^{t-1}\sum_{m=1}^{\infty}\mathcal{G}^{m-1}(\mathcal{K}+\mathcal{G})\mathcal{E}_{1-m}\Big)\longrightarrow 0,\quad a.s.$$

One part of the above equation is

$$\frac{1}{NT}\sum_{t=1}^{T}\Big((-\mathcal{K})^{t-1}\sum_{m=1}^{\infty}\mathcal{G}^{m-1}(\mathcal{K}+\mathcal{G})\mathcal{E}_{1-m}\Big)^\top \Big((-\mathcal{K})^{t-1}\sum_{m=1}^{\infty}\mathcal{G}^{m-1}(\mathcal{K}+\mathcal{G})\mathcal{E}_{1-m}\Big).$$

From (b) of Lemma 3, we can easily find that this part converges to 0. So we only need to consider the rest

$$\frac{1}{NT}\sum_{t=1}^{T}\mathcal{E}_{\theta t}^\top\cdot\Big((-\mathcal{K})^{t-1}\sum_{m=1}^{\infty}\mathcal{G}^{m-1}(\mathcal{K}+\mathcal{G})\mathcal{E}_{1-m}\Big)$$

$$\longrightarrow 0 \qquad a.s.$$

which is equivalent to

$$\frac{1}{NT}\sum_{t=1}^{T}\Big(\mathcal{E}_{\theta t}^{\top}(-\mathcal{K})^{t-1}\Big)\Big(\sum_{m=1}^{\infty}\mathcal{G}^{m-1}(\mathcal{K}+\mathcal{G})\mathcal{E}_{1-m}\Big) \longrightarrow 0 \qquad a.s. \tag{14}$$

Let $M(N) = \sum_{m=1}^{\infty}\mathcal{G}^{m-1}(\mathcal{K}+\mathcal{G})\mathcal{E}_{1-m}$, $\mathcal{E}(N,T) = \sum_{t=1}^{T}(-\mathcal{K}^{\top})^{t-1}\mathcal{E}_t$. Noting that

$$\frac{\mathcal{E}(N,T)^{\top}M(N)}{NT} \to 0 \;\; a.s. \;\text{ when } \;\; \min\{N,T\}\to\infty$$

is equivalent to

$$P\Big(\lim_{\min\{N,T\}\to\infty}\frac{\mathcal{E}(N,T)^{\top}M(N)}{NT}\neq 0\Big)=0,$$

then we have

$$\begin{aligned}
&P\Big(\lim_{\min\{N,T\}\to\infty}\frac{\mathcal{E}(N,T)^{\top}M(N)}{NT}\neq 0\Big)\\
=&P\Big(\omega:\exists n,\forall m,\exists N,T\geq m, s.t.\Big|\frac{\mathcal{E}(N,T)^{\top}M(N)}{NT}\Big|>\frac{1}{n}\Big)\\
=&P\Big(\bigcup_{n=1}^{\infty}\bigcap_{m=1}^{\infty}\bigcup_{N\geq m}\bigcup_{T\geq m}\Big|\frac{\mathcal{E}(N,T)^{\top}M(N)}{NT}\Big|>\frac{1}{n}\Big).
\end{aligned}$$

Also due to

$$\begin{aligned}
&P\Big(\bigcap_{m=1}^{\infty}\bigcup_{N\geq m}\bigcup_{T\geq m}\Big|\frac{\mathcal{E}(N,T)^{\top}M(N)}{NT}\Big|>\frac{1}{n}\Big)\\
&=\lim_{m\to\infty}P\Big(\bigcup_{N\geq m}\bigcup_{T\geq m}\Big|\frac{\mathcal{E}(N,T)^{\top}M(N)}{NT}\Big|>\frac{1}{n}\Big)\\
&\leq\lim_{m\to\infty}\sum_{N\geq m}P\Big(\bigcup_{T\geq m}\Big|\frac{\mathcal{E}(N,T)^{\top}M(N)}{NT}\Big|>\frac{1}{n}\Big)\\
&\leq\lim_{m\to\infty}\sum_{N\geq m}\sum_{T\geq m}P\Big(\Big|\frac{\mathcal{E}(N,T)^{\top}M(N)}{NT}\Big|>\frac{1}{n}\Big),
\end{aligned}$$

then for each n, if we can prove

$$\sum_{N=1}^{\infty}\sum_{T=1}^{\infty}P\Big(\Big|\frac{\mathcal{E}(N,T)^{\top}M(N)}{NT}\Big|>\frac{1}{n}\Big)<\infty,$$

we have

$$P\Big(\bigcap_{m=1}^{\infty}\bigcup_{N\geq m}\bigcup_{T\geq m}\Big|\frac{\mathcal{E}(N,T)^{\top}M(N)}{NT}\Big|>\frac{1}{n}\Big)=0. \tag{15}$$

In fact, by using the Chebyshev's Theorem, we find that

$$\begin{aligned}
&\sum_{N=1}^{\infty}\sum_{T=1}^{\infty}P\Big(\Big|\frac{\mathcal{E}(N,T)^{\top}M(N)}{NT}\Big|>\frac{1}{n}\Big)\\
&\leq\sum_{N=1}^{\infty}\sum_{T=1}^{\infty}\frac{E\Big(\mathcal{E}(N,T)^{\top}M(N)\Big)^4}{N^4T^4(\frac{1}{n})^4}.
\end{aligned}$$

Let $\Sigma_{\mathcal{E}}$ and Σ_M be the covariance matrices of $\mathcal{E}(N,T)$ and $M(N)$ respectively. Noting that $\Sigma_{\mathcal{E}}$ and Σ_M are both symmetric matrices which can be written as

$$\Sigma_{\mathcal{E}}=P\Sigma_{\mathcal{E}'}P^{-1} \;\; and \;\; \Sigma_M=Q\Sigma_{M'}Q^{-1}$$

where P, Q are both orthogonal matrices and $\Sigma_{\mathcal{E}'}$, $\Sigma_{M'}$ are corresponding diagonal matrices. So $\mathcal{E}(N,T)$ and $M(N)$ can be written as $\mathcal{E}(N,T)=P\mathcal{E}^{'}(N,T)$, $M(N)=QM^{'}(N)$, where the covariance matrices of $\mathcal{E}^{'}(N,T)$ and $M^{'}(N)$ are $\Sigma_{\mathcal{E}'}$ and $\Sigma_{M'}$ respectively, so we have

$$\frac{E\Big(\mathcal{E}(N,T)^{\top}M(N)\Big)^4}{N^4T^4\varepsilon^4}=\frac{E\Big(\mathcal{E}^{'}(N,T)^{\top}P^{\top}QM^{'}(N)\Big)^4}{N^4T^4\varepsilon^4}.$$

Since $P^{\top}Q$ is also an orthogonal matrix for both P and Q are orthogonal matrices, then from Lemma 2, we have that if the elements of $\Sigma_{\mathcal{E}'}$ and $\Sigma_{M'}$ are uniformly bounded with N and T, then

$$\begin{aligned}
&\sum_{N=1}^{\infty}\sum_{T=1}^{\infty}\frac{E\Big(\mathcal{E}^{'}(N,T)^{\top}P^{\top}QM^{'}(N)\Big)^4}{N^4T^4}\\
&\leq C_1\sum_{N=1}^{\infty}\sum_{T=1}^{\infty}\frac{1}{N^2T^4}<\infty,
\end{aligned}$$

which is

$$\sum_{N=1}^{\infty}\sum_{T=1}^{\infty}\frac{E\Big(\mathcal{E}(N,T)^{\top}M(N)\Big)^4}{N^4T^4\varepsilon^4}<\infty.$$

Thus Eq. (15) holds and then Eq. (14) holds.

Now we only have to prove that the elements of $\Sigma_{\mathcal{E}'}$ and $\Sigma_{M'}$ are uniformly bounded with N and T. Since

$$\mathcal{E}_t\sim N(\mathbf{0},I), \;\; and \;\; \mathcal{E}(N,T)=\sum_{t=1}^{T}(-\mathcal{K}^{\top})^{t-1}\mathcal{E}_t,$$

then

$$\Sigma_{\mathcal{E}}=\sum_{T=1}^{\infty}(\mathcal{K}^{\top})^{t-1}(\mathcal{K})^{t-1}.$$

From Lemma 3 (a), we know that the largest eigenvalue of $\Sigma_{\mathcal{E}}$ is bounded, which is equivalent to that the elements of $\Sigma_{\mathcal{E}'}$ are uniformly bounded with N and T. From Lemma 3(b), we also have that the largest eigenvalue of Σ_M is bounded, which is equivalent to that the elements of $\Sigma_{M'}$ are uniformly bounded with N and T. The proof is completed. □

A.5 Proof of Theorem 2.5

To prove Theorem 5, we begin with a lemma given by Pham and Tran [19].

Lemma 1. *Let $V_T(\tilde{\boldsymbol{\theta}})$ be a sequence of continuous random functions defined on a compact subset $\widetilde{\Theta}$ of a k-dimensional Euclidean space $\mathbb{R}^k$, $\boldsymbol{\theta}$ be a point of $\widetilde{\Theta}$. Thus for any $\hat{\boldsymbol{\theta}} \in \widetilde{\Theta}, \hat{\boldsymbol{\theta}} \neq \boldsymbol{\theta}$, there exists a neighborhood $U(\hat{\boldsymbol{\theta}})$ of $\hat{\boldsymbol{\theta}}$ such that*

$$\liminf_{N,T\to\infty} \left\{ \inf_{\tilde{\boldsymbol{\theta}}\in U(\hat{\boldsymbol{\theta}})} V_T(\tilde{\boldsymbol{\theta}}) - V_T(\theta) \right\} > 0$$

Then $\hat{\boldsymbol{\theta}}$, solution of the minimization of V_T on Θ, converges to $\boldsymbol{\theta}$.

Lemma 1 gives a method to verify whether the obtained estimator has the property of consistency. Thus to prove Theorem 5, we only need to check whether the $V_T(\tilde{\boldsymbol{\theta}})$ we choose satisfies the conditions in Lemma 1.

Proof of 2.5. Noting that

$$V_T(\boldsymbol{\theta}) = \frac{1}{NT}\sum_{t=1}^{T} \mathcal{E}_{\boldsymbol{\theta}(t|\mathbf{y}_0)}^\top \mathcal{E}_{\boldsymbol{\theta}(t|\mathbf{y}_0)},$$

we have

$$\liminf_{N,T\to\infty} V_T(\boldsymbol{\theta}) = \liminf_{N,T\to\infty} \frac{1}{NT}\sum_{t=1}^{T} \mathcal{E}_{\boldsymbol{\theta}(t|\mathbf{y}_0)}^\top \mathcal{E}_{\boldsymbol{\theta}(t|\mathbf{y}_0)}$$
$$\to \liminf_{N,T\to\infty} \frac{1}{NT}\sum_{t=1}^{T} \mathcal{E}_{\boldsymbol{\theta}t}^\top \mathcal{E}_{\boldsymbol{\theta}t} = \sigma^2,$$

where $\sigma^2 = \frac{1}{N}\mathbb{E}[\mathcal{E}_t^\top \mathcal{E}_t]$. From Lemma 8 we get that

$$\liminf_{N,T\to\infty} \inf_{\tilde{\boldsymbol{\theta}}\in U(\hat{\boldsymbol{\theta}})} V_T(\tilde{\boldsymbol{\theta}}) > \sigma^2$$

if $m \to \infty$, which means if the neighborhood $U(\hat{\boldsymbol{\theta}})$ is small enough, then $\hat{\boldsymbol{\theta}}$, solution of the minimization of V_T on Θ, is strongly consistent. The proof of Theorem 2.5 is completed. □

A.6 Some technical lemmas

Lemma 2. *Let $X = (X_1, \dots, X_N) \sim N(\mathbf{0}, \Sigma_1)$, $Z = (Z_1, \dots, Z_N) \sim N(\mathbf{0}, \Sigma_2)$, $P = (P)_{ij} \in \mathbb{R}^{N\times N}$, where Σ_1, Σ_2 are diagonal matrices with bounded elements and P is an orthogonal matrix. Then we have the following conclusion*

$$\frac{E[X^\top PZ]^4}{N^4} \le \frac{C_1}{N^2}, \tag{16}$$

where C_1 is a constant which has no connection with P, N and T.

Proof. Since $X = (X_1, \dots, X_N)$, $Z = (Z_1, \dots, Z_N)$, $P = (P)_{ij} \in \mathbb{R}^{N\times N}$, then

$$X^\top PZ = \sum_{i=1}^{N}\sum_{j=1}^{N} P_{ij}X_iZ_j.$$

Considering $(X^\top PZ)^4$, we know that its expansion is the linear combination of $(P_{ij}X_iZ_j)^l, 1 \le l \le 4$. We just consider the items whose powers are quadratic or quadruplicate since $E[X_i] = 0, 1 \le i \le N$ and $E[Z_j] = 0, 1 \le j \le N$. Then

$$\begin{aligned}
E[X^\top PZ]^4 = E\Big[&P_{11}X_1Z_1 + P_{21}X_2Z_1 + \dots \\
&+ P_{(N-1)N}X_{N-1}Z_N + P_{NN}X_NZ_N\Big]^4 \\
=& E\Big[\sum_{i=1}^{N}\sum_{j=1}^{N}\Big(P_{ij}X_iZ_j\Big)^4 + 3\sum_{i=1}^{N}\sum_{j=1}^{N}\Big(P_{ij}X_iZ_j\Big)^2 \\
&\sum_{m\neq i, n\neq j}\Big(P_{mn}X_mZ_n\Big)^2\Big] \\
&+ E\Big[6\sum_{i=1}^{N}\sum_{j=1}^{N}(P_{ij}X_iZ_j)\sum_{p=1}^{N}P_{pj}X_pZ_j \\
&\sum_{q=1}^{N}P_{iq}X_iZ_q(P_{pq}X_pZ_q)\Big] \\
&- E\Big[3\sum_{i=1}^{N}\sum_{j=1}^{N}\Big(P_{ij}X_iZ_j\Big)^2\Big(\sum_{p=1}^{N}\Big(P_{pj}X_pZ_j\Big)^2 \\
&+ \sum_{q=1}^{N}\Big(P_{iq}X_iZ_q\Big)^2\Big)\Big] \\
\le& 3E\Big\{\sum_{i=1}^{N}\sum_{j=1}^{N}\Big[\Big(P_{ij}X_iZ_j\Big)^2\sum_{p=1}^{N}\sum_{q=1}^{N}\Big(P_{pq}X_pZ_q\Big)^2\Big]\Big\} \\
&+ 6E\Big[\sum_{i=1}^{N}\sum_{j=1}^{N}(P_{ij}X_iZ_j)\sum_{p=1}^{N}P_{pj}X_pZ_j \\
&\sum_{q=1}^{N}P_{iq}X_iZ_q(P_{pq}X_pZ_q)\Big].
\end{aligned} \tag{17}$$

Recalling that the variances of X_i and Z_j are bounded, we can give a constant $M > 0$ such that $E[X_i^2] \le M$ and $E[Z_j^2] \le M$, then $\mathrm{Var}[X_i] \le M$ and $\mathrm{Var}[Z_j] \le M$. Furthermore, after simple calculations, we have

$$\begin{aligned}
&\mathrm{Var}[X_i^2] \le 2M, \\
&\mathrm{Var}[Z_j^2] \le 2M, \\
&E[X_i^4] \le M^2 + 2M, \\
&E[Z_j^4] \le M^2 + 2M.
\end{aligned}$$

Taking these conclusions into the formula above,

$$\begin{aligned}
E[X^\top PZ]^4 &\le 3\Big(M^2+2M\Big)^2 \sum_{i=1}^N\sum_{j=1}^N \Big[P_{ij}^2 \sum_{p=1}^N\sum_{q=1}^N P_{pq}^2\Big] \\
&\quad + 6\Big(M^2+2M\Big)^2 \sum_{i=1}^N\sum_{j=1}^N |P_{ij}| \\
&\quad \cdot \sum_{p=1}^N |P_{pj}| \sum_{q=1}^N |P_{iq}P_{pq}| \\
&\le 3\Big(M^2+2M\Big)^2 N^2 + 6\Big(M^2+2M\Big)^2 N^2 \\
&= 9\Big(M^2+2M\Big)^2 N^2.
\end{aligned}$$

The first inequal sign is established for the orthogonal matrix P with $\sum_{j=1}^N P_{ij}^2 = 1$ and $\sum_{q=1}^N P_{iq}P_{pq} = 0$ if $i \ne p$. The second inequality holds by using Cauchy-Schwarz inequality, i.e.,

$$\sum_{q=1}^N |P_{iq}P_{pq}| \le \sqrt{\sum_{q=1}^N P_{iq}^2 \sum_{q=1}^N P_{pq}^2} = 1 \quad \textit{and} \quad \sum_{j=1}^N |P_{ij}P_{pj}| \le 1.$$

Then

$$\frac{E[X^\top PZ]^4}{N^4} \le \frac{9\Big(M^2+2M\Big)^2}{N^2} = \frac{C_1(M)}{N^2}.$$

The proof is completed. □

Lemma 3. *(a) For the matrix $\mathcal{K}$, define a matrix*

$$\Sigma_{\mathcal{K}} = \sum_{t=1}^T (\mathcal{K}^\top)^t \mathcal{K}^t.$$

Then the largest eigenvalue of $\Sigma_{\mathcal{K}}$ is bounded when $T \to \infty$.(Here N (the dimension of $\mathcal{K}$) is not fixed.)

(b) For the matrix $\mathcal{K}$, $\mathcal{G}$, $\tilde{\mathcal{K}} = \tilde{\alpha}_1 W + \tilde{\alpha}_0 I$, $\tilde{\mathcal{G}} = \tilde{\beta}_1 W + \tilde{\beta}_0 I$, define

$$Q = \sum_{m=1}^\infty \sum_{j=1}^m (-\widetilde{\mathcal{K}})^{m-j-1}(\widetilde{\mathcal{K}}+\widetilde{\mathcal{G}})\mathcal{G}^{j-1}(\mathcal{K}+\mathcal{G})\mathcal{E}_{1-m},$$

where $\{\mathcal{E}_t\}$ are independent identically distributed random vectors obeying standard multivariate normal distribution. If $(|\alpha_0|+|\alpha_1|) < 1$, $(|\beta_0|+|\beta_1|) < 1$, $(|\tilde{\alpha}_0|+|\tilde{\alpha}|_1) < 1$ and $(|\tilde{\beta}_0+|\tilde{\beta}|_1) < 1$, then all eigenvalues of the covariance matrix of Q are bounded.

Proof. (a) Define $|M|_e$ as $|M|_e = (|m_{ij}|) \in \mathbb{R}^{n\times p}$ for any arbitrary matrix $M = (m_{ij}) \in \mathbb{R}^{n\times p}$. From the Weyl Theorem we know that $\lambda_{\max}(\Sigma_{\mathcal{K}}) \le \sum_{t=1}^T \lambda_{\max}((\mathcal{K}^\top)^t\mathcal{K}^t)$ (for arbitrary matrix M, define $\lambda_{\max}(M)$ to be the largest eigenvalue of M). From the definition of $\lambda_{\max}(M)$, then $\lambda_{\max}(M) = \max_{x^\top x=1} |x^\top M x|$ when M is a positive semidefinite matrix or a positive definite matrix. Thus we have

$$\begin{aligned}
\lambda_{\max}(\Sigma_{\mathcal{K}}) &\le \sum_{t=1}^T \lambda_{\max}((\mathcal{K}^\top)^t\mathcal{K}^t) \\
&= \sum_{t=1}^T \max_{x_t^\top x_t=1} |x_t^\top (\mathcal{K}^\top)^t \mathcal{K}^t x_t| \\
&= \sum_{t=1}^T \max_{x_t^\top x_t=1} |x_t^\top \mathcal{K}^t (\mathcal{K}^\top)^t x_t| \\
&\le \sum_{t=1}^T \max_{x_t^\top x_t=1} \|(\mathcal{K}^\top)^t x_t\|_{1,e}^2 \\
&= \sum_{t=1}^T \max_{x_t^\top x_t=1} (|x_t^\top|_e \; |(\mathcal{K}^\top)^t|_e \; \mathbf{1})^2 \\
&\le \sum_{t=1}^T (|\alpha_0|+|\alpha_1|)^{2t} \\
&\le \frac{(|\alpha_0|+|\alpha_1|)^2}{1-(|\alpha_0|+|\alpha_1|)^2} \\
&< \infty,
\end{aligned}$$

where the second equality is established since the eigenvalues of $(\mathcal{K}^\top)^t\mathcal{K}^t$ and $\mathcal{K}^t(\mathcal{K}^\top)^t$ are the same, and $\|\mathbf{L}\|_{1,e} = \sum_{(i,t)\in[N]\times[T]} |L_{it}|$ is the element-wise l_1 norm.

(b) Let Σ_Q be the covariance matrix of Q. Then we have

$$\begin{aligned}
\Sigma_Q = \sum_{m=1}^\infty \Bigg\{ &\Big(\sum_{j=1}^m (-\widetilde{\mathcal{K}})^{m-j-1}(\widetilde{\mathcal{K}}+\widetilde{\mathcal{G}})\mathcal{G}^{j-1}(\mathcal{K}+\mathcal{G})\Big) \\
&\Big(\sum_{j=1}^m (-\widetilde{\mathcal{K}})^{m-j-1}(\widetilde{\mathcal{K}}+\widetilde{\mathcal{G}})\mathcal{G}^{j-1}(\mathcal{K}+\mathcal{G})\Big)^\top \Bigg\}.
\end{aligned}$$

Similar to the proof of (a), we have

$$\begin{aligned}
\lambda_{\max}(\Sigma_Q) &\le \sum_{m=1}^\infty \max_{x_m^\top x_m=1} \Big(|x_m^\top|_e \,\Big|\Big(\sum_{j=1}^m (-\widetilde{\mathcal{K}})^{m-j-1} \\
&\qquad (\widetilde{\mathcal{K}}+\widetilde{\mathcal{G}})\mathcal{G}^{j-1}(\mathcal{K}+\mathcal{G})\Big)^\top\Big|_e \mathbf{1}\Big)^2 \\
&\le \sum_{m=1}^\infty \sum_{j=1}^m \Big((|\tilde{\alpha}_0|+|\tilde{\alpha}_1|)^{m-j-1}(|\beta_0|+|\beta_1|)^{j-1} \\
&\qquad \cdot (|\alpha_0|+|\alpha_1|+|\beta_0|+|\beta_1|) \\
&\qquad \cdot (|\tilde{\alpha}_0|+|\tilde{\alpha}_1|+|\tilde{\beta}_0|+|\tilde{\beta}_1|)\Big)^2 \\
&= \frac{(|\alpha_0|+|\alpha_1|+|\beta_0|+|\beta_1|)^2}{(|\tilde{\alpha}_0|+|\tilde{\alpha}_1|)^2\Big(1-(|\tilde{\alpha}_0|+|\tilde{\alpha}_1|)^2\Big)} \\
&\qquad \cdot \frac{(|\tilde{\alpha}_0|+|\tilde{\alpha}_1|+|\tilde{\beta}_0|+|\tilde{\beta}_1|)^2}{\Big(1-(|\beta_0|+|\beta_1|)^2\Big)} < \infty.
\end{aligned}$$

Thus we complete this proof. $\square$

Lemma 4. *If* $(|\alpha_0| + |\alpha_1|) < 1$, $(|\beta_0| + |\beta_1|) < 1$, $(|\tilde{\alpha}_0| + |\tilde{\alpha}_1| < 1)$ *and* $(|\tilde{\beta}_0| + |\tilde{\beta}_1| < 1)$, *then the recursive estimation*

$$\mathcal{E}_{\tilde{\boldsymbol{\theta}}t} = \mathbb{Y}_t - \sum_{j=1}^{\infty} (-\tilde{\mathcal{K}})^{j-1} (\tilde{\mathcal{K}} + \tilde{\mathcal{G}}) \mathbb{Y}_{t-j}$$

has a stationary solution and it satisfies that

$$\frac{1}{NT} \sum_{t=1}^{T} \mathcal{E}_{\tilde{\boldsymbol{\theta}}(t|\mathbf{y}_0)}^{\top} \mathcal{E}_{\tilde{\boldsymbol{\theta}}(t|\mathbf{y}_0)} - \frac{1}{NT} \sum_{t=1}^{T} \mathcal{E}_{\tilde{\boldsymbol{\theta}}t}^{\top} \mathcal{E}_{\tilde{\boldsymbol{\theta}}t} \to 0 \quad a.s.$$

Proof. The proof of the stationarity of $\mathcal{E}_{\tilde{\boldsymbol{\theta}}t}$ is just like that of $\mathcal{E}_t$, so we give no more detailed description here.

To consider the convergence of $\mathcal{E}_{\tilde{\boldsymbol{\theta}}t}$, we take the expression

$$\mathbb{Y}_t = \mathcal{E}_t + \sum_{j=1}^{\infty} \mathcal{G}^{j-1} (\mathcal{K} + \mathcal{G}) \mathcal{E}_{t-j}$$

into account to the following equation

$$\mathcal{E}_{\tilde{\boldsymbol{\theta}}t} = \mathbb{Y}_t - \sum_{j=1}^{\infty} (-\tilde{\mathcal{K}})^{j-1} (\tilde{\mathcal{K}} + \tilde{\mathcal{G}}) \mathbb{Y}_{t-j}.$$

Define $\mathcal{G}^{-1} = 0$ by convention, then we have

$$\begin{aligned} \mathcal{E}_{\tilde{\boldsymbol{\theta}}t} = \mathcal{E}_t + \sum_{m=1}^{\infty} \Big(& \mathcal{G}^{m-1} (\mathcal{K} + \mathcal{G}) \\ & - (-\tilde{\mathcal{K}})^{m-1} (\tilde{\mathcal{K}} + \tilde{\mathcal{G}}) \\ & - \sum_{j=1}^{m} (-\tilde{\mathcal{K}})^{j-1} (\tilde{\mathcal{K}} + \tilde{\mathcal{G}}) \mathcal{G}^{m-j-1} (\mathcal{K} + \mathcal{G}) \Big) \mathcal{E}_{t-m}, \end{aligned} \tag{18}$$

where $\{\mathcal{E}_t\}$ are independent identically distributed. Thus it suffices to prove for all N, the largest eigenvalue of the variance matrix $\Sigma_{\mathcal{E}_{\tilde{\boldsymbol{\theta}}t}}$ of $\mathcal{E}_{\tilde{\boldsymbol{\theta}}t}$ is finite, where

$$\begin{aligned} \Sigma_{\mathcal{E}_{\tilde{\boldsymbol{\theta}}t}} = I + \sum_{m=1}^{\infty} \Big\{ & \Big(\mathcal{G}^{m-1} (\mathcal{K} + \mathcal{G}) - (-\tilde{\mathcal{K}})^{m-1} (\tilde{\mathcal{K}} + \tilde{\mathcal{G}}) \\ & - \sum_{j=1}^{m} (-\tilde{\mathcal{K}})^{j-1} (\tilde{\mathcal{K}} + \tilde{\mathcal{G}}) \mathcal{G}^{m-j-1} (\mathcal{K} + \mathcal{G}) \Big)^{\top} \\ & \Big(\mathcal{G}^{m-1} (\mathcal{K} + \mathcal{G}) - (-\tilde{\mathcal{K}})^{m-1} (\tilde{\mathcal{K}} + \tilde{\mathcal{G}}) \\ & - \sum_{j=1}^{m} (-\tilde{\mathcal{K}})^{j-1} (\tilde{\mathcal{K}} + \tilde{\mathcal{G}}) \mathcal{G}^{m-j-1} (\mathcal{K} + \mathcal{G}) \Big) \Big\} \\ =: I + \Sigma^{'}. \end{aligned} \tag{19}$$

By using the same method as Lemma 3, we can easily get the conclusion.

To prove

$$\frac{1}{NT} \sum_{t=1}^{T} \mathcal{E}_{\tilde{\boldsymbol{\theta}}(t|\mathbf{y}_0)}^{\top} \mathcal{E}_{\tilde{\boldsymbol{\theta}}(t|\mathbf{y}_0)} - \frac{1}{NT} \sum_{t=1}^{T} \mathcal{E}_{\tilde{\boldsymbol{\theta}}t}^{\top} \mathcal{E}_{\tilde{\boldsymbol{\theta}}t} \to 0 \ a.s.,$$

we just need to use the same steps as the proof of Theorem 2.4 (3).

Define

$$\begin{aligned} A := \frac{1}{NT} \sum_{t=1}^{T} & \Big(\mathbb{Y}_t - \sum_{j=1}^{t-1} (-\widetilde{\mathcal{K}})^{j-1} (\widetilde{\mathcal{K}} + \widetilde{\mathcal{G}}) \mathbb{Y}_{t-j} \Big)^{\top} \\ & \Big((-\widetilde{\mathcal{K}})^{t-1} (\widetilde{\mathcal{K}} + \widetilde{\mathcal{G}}) \mathbf{y}_0 \Big), \\ B := \frac{1}{NT} \sum_{t=1}^{T} & \Big((-\widetilde{\mathcal{K}})^{t-1} (\widetilde{\mathcal{K}} + \widetilde{\mathcal{G}}) \mathbf{y}_0 \Big)^{\top} \\ & \Big(\mathbb{Y}_t - \sum_{j=1}^{t-1} (-\widetilde{\mathcal{K}})^{j-1} (\widetilde{\mathcal{K}} + \widetilde{\mathcal{G}}) \mathbb{Y}_{t-j} \Big), \\ C := \frac{1}{NT} \sum_{t=1}^{T} & \Big((-\widetilde{\mathcal{K}})^{t-1} (\widetilde{\mathcal{K}} + \widetilde{\mathcal{G}}) \mathbf{y}_0 \Big)^{\top} \\ & \Big((-\widetilde{\mathcal{K}})^{t-1} (\widetilde{\mathcal{K}} + \widetilde{\mathcal{G}}) \mathbf{y}_0 \Big), \\ A' := \frac{1}{NT} \sum_{t=1}^{T} & \Big(\mathbb{Y}_t - \sum_{j=1}^{t-1} (-\widetilde{\mathcal{K}})^{j-1} (\widetilde{\mathcal{K}} + \widetilde{\mathcal{G}}) \mathbb{Y}_{t-j} \Big)^{\top} \\ & \Big(\sum_{j=t}^{\infty} (-\widetilde{\mathcal{K}})^{j-1} (\widetilde{\mathcal{K}} + \widetilde{\mathcal{G}}) \mathbb{Y}_{t-j} \Big), \\ B' := \frac{1}{NT} \sum_{t=1}^{T} & \Big(\sum_{j=t}^{\infty} (-\widetilde{\mathcal{K}})^{j-1} (\widetilde{\mathcal{K}} + \widetilde{\mathcal{G}}) \mathbb{Y}_{t-j} \Big)^{\top} \\ & \Big(\mathbb{Y}_t - \sum_{j=1}^{t-1} (-\widetilde{\mathcal{K}})^{j-1} (\widetilde{\mathcal{K}} + \widetilde{\mathcal{G}}) \mathbb{Y}_{t-j} \Big), \\ C' := \frac{1}{NT} \sum_{t=1}^{T} & \Big(\sum_{j=t}^{\infty} (-\widetilde{\mathcal{K}})^{j-1} (\widetilde{\mathcal{K}} + \widetilde{\mathcal{G}}) \mathbb{Y}_{t-j} \Big)^{\top} \\ & \Big(\sum_{j=t}^{\infty} (-\widetilde{\mathcal{K}})^{j-1} (\widetilde{\mathcal{K}} + \widetilde{\mathcal{G}}) \mathbb{Y}_{t-j} \Big). \end{aligned}$$

After eliminating the same term, we still need to prove that

$$(A - A') + (B - B') - (C - C') \longrightarrow 0 \qquad a.s.$$

We only give the detailed proof of the result $A - A' \to 0$ $a.s.$ and the other are similar and omitted. In other words, it suffices to prove that the following equation

$$\begin{aligned} \frac{1}{NT} \sum_{t=1}^{T} & \Big(\mathbb{Y}_t - \sum_{j=1}^{t-1} (-\tilde{\mathcal{K}})^{j-1} (\tilde{\mathcal{K}} + \tilde{\mathcal{G}}) \mathbb{Y}_{t-j} \Big)^{\top} \\ & \Big((-\tilde{\mathcal{K}})^{t-1} (\tilde{\mathcal{K}} + \tilde{\mathcal{G}}) \mathbf{y}_0 - \sum_{j=t}^{\infty} (-\tilde{\mathcal{K}})^{j-1} (\tilde{\mathcal{K}} + \tilde{\mathcal{G}}) \mathbb{Y}_{t-j} \Big) \end{aligned}$$

$$\longrightarrow 0 \quad a.s.$$

holds.

Without loss of generality, assume that $\mathbf{y}_0 = \mathbf{0}$ and $(-\tilde{\mathcal{K}})^{-1} = 0$ by convention. Noting that

$$\Big(\mathbb{Y}_t - \sum_{j=1}^{t-1}(-\tilde{\mathcal{K}})^{j-1}(\tilde{\mathcal{K}}+\tilde{\mathcal{G}})\mathbb{Y}_{t-j}\Big) - \sum_{j=t}^{\infty}(-\tilde{\mathcal{K}})^{j-1}(\tilde{\mathcal{K}}+\tilde{\mathcal{G}})\mathbb{Y}_{t-j} = \mathcal{E}_{\tilde{\boldsymbol{\theta}}t}$$

and

$$\mathbb{Y}_t = \mathcal{E}_t + \sum_{j=1}^{\infty}\mathcal{G}^{j-1}(\mathcal{K}+\mathcal{G})\mathcal{E}_{t-j},$$

then what we need to prove can be written as

$$\begin{aligned}\frac{1}{NT}\sum_{t=1}^{T}\Big(\mathcal{E}_{\tilde{\boldsymbol{\theta}}t} + (-\tilde{\mathcal{K}})^{t-1}\sum_{m=1}^{\infty}\Big[(-\tilde{\mathcal{K}})^{m-1}(\tilde{\mathcal{K}}+\tilde{\mathcal{G}}) \\ + \sum_{j=1}^{m}(-\tilde{\mathcal{K}})^{m-j-1}(\tilde{\mathcal{K}}+\tilde{\mathcal{G}})\mathcal{G}^{j-1}(\mathcal{K}+\mathcal{G})\Big]\mathcal{E}_{1-m}\Big)^{\top} \\ \cdot\Big((-\tilde{\mathcal{K}})^{t-1}\sum_{m=1}^{\infty}\Big[(-\tilde{\mathcal{K}})^{m-1}(\tilde{\mathcal{K}}+\tilde{\mathcal{G}}) \\ + \sum_{j=1}^{m}(-\tilde{\mathcal{K}})^{m-j-1}(\tilde{\mathcal{K}}+\tilde{\mathcal{G}})\mathcal{G}^{j-1}(\mathcal{K}+\mathcal{G})\Big]\mathcal{E}_{1-m}\Big) \\ \longrightarrow 0 \qquad a.s.\end{aligned}$$

Also we just consider

$$\begin{aligned}\frac{1}{NT}\sum_{t=1}^{T}\mathcal{E}_{\tilde{\boldsymbol{\theta}}t}^{\top}\cdot\Big((-\tilde{\mathcal{K}})^{t-1}\sum_{m=1}^{\infty}\Big[(-\tilde{\mathcal{K}})^{m-1}(\tilde{\mathcal{K}}+\tilde{\mathcal{G}}) \\ + \sum_{j=1}^{m}(-\tilde{\mathcal{K}})^{m-j-1}(\tilde{\mathcal{K}}+\tilde{\mathcal{G}})\mathcal{G}^{j-1}(\mathcal{K}+\mathcal{G})\Big]\mathcal{E}_{1-m}\Big) \\ \longrightarrow 0 \quad a.s.,\end{aligned}$$

which is equivalent to

$$\begin{aligned}\frac{1}{NT}\sum_{t=1}^{T}\Big(\mathcal{E}_{\tilde{\boldsymbol{\theta}}t}^{\top}(-\tilde{\mathcal{K}})^{t-1}\Big)\cdot\Big(\sum_{m=1}^{\infty}\Big[(-\tilde{\mathcal{K}})^{m-1}(\tilde{\mathcal{K}}+\tilde{\mathcal{G}}) \\ + \sum_{j=1}^{m}(-\tilde{\mathcal{K}})^{m-j-1}(\tilde{\mathcal{K}}+\tilde{\mathcal{G}})\mathcal{G}^{j-1}(\mathcal{K}+\mathcal{G})\Big]\mathcal{E}_{1-m}\Big) \\ \longrightarrow 0 \quad a.s.\end{aligned}$$

Let

$$\begin{aligned}\widetilde{M}(N) = \sum_{m=1}^{\infty}\Big[(-\tilde{\mathcal{K}})^{m-1}(\tilde{\mathcal{K}}+\tilde{\mathcal{G}}) \\ + \sum_{j=1}^{m}(-\tilde{\mathcal{K}})^{m-j-1}(\tilde{\mathcal{K}}+\tilde{\mathcal{G}})\mathcal{G}^{j-1}(\mathcal{K}+\mathcal{G})\Big]\mathcal{E}_{1-m},\end{aligned}$$

and

$$\widetilde{\mathcal{E}}(N,T) = \sum_{t=1}^{T}(-\tilde{\mathcal{K}}^{\top})^{t-1}\mathcal{E}_{\tilde{\boldsymbol{\theta}}t}.$$

Noting that

$$\frac{\widetilde{\mathcal{E}}(N,T)^{\top}\widetilde{M}(N)}{NT} \to 0 \;\; a.s. \;\; when \quad \min\{N,T\}\to\infty$$

is equivalent to

$$P\Big(\lim_{\min\{N,T\}\to\infty}\frac{\widetilde{\mathcal{E}}(N,T)^{\top}\widetilde{M}(N)}{NT} \neq 0\Big) = 0,$$

then similar to the proof of Theorem 2.4(3), we only need to prove for all $\varepsilon > 0$,

$$\sum_{N=1}^{\infty}\sum_{T=1}^{\infty}\frac{E\Big(\widetilde{\mathcal{E}}(N,T)^{\top}\widetilde{M}(N)\Big)^4}{N^4T^4\varepsilon^4} < \infty.$$

Let $\Sigma_{\widetilde{\mathcal{E}}}$ and $\Sigma_{\widetilde{M}}$ be the covariance matrices of $\widetilde{\mathcal{E}}(N,T)$ and $\widetilde{M}(N)$. Noting that $\Sigma_{\widetilde{\mathcal{E}}}$ and $\Sigma_{\widetilde{M}}$ are both symmetric matrices and they can be written as

$$\Sigma_{\widetilde{\mathcal{E}}} = P\Sigma_{\widetilde{\mathcal{E}}'}P^{-1} \quad and \quad \Sigma_{\widetilde{M}} = Q\Sigma_{\widetilde{M}'}Q^{-1},$$

where P, Q are both orthogonal matrices and $\Sigma_{\widetilde{\mathcal{E}}'}$, $\Sigma_{\widetilde{M}'}$ are diagonal matrices, then $\mathcal{E}(N,T)$ and $M(N)$ can be written as $\widetilde{\mathcal{E}}(N,T) = P\widetilde{\mathcal{E}}'(N,T)$, $\widetilde{M}(N) = Q\widetilde{M}'(N)$, where the covariance matrices of $\widetilde{\mathcal{E}}'(N,T)$ and $\widetilde{M}'(N)$ are $\Sigma_{\widetilde{\mathcal{E}}'}$ and $\Sigma_{\widetilde{M}'}$ respectively. Thus

$$\frac{E\Big(\widetilde{\mathcal{E}}(N,T)^{\top}\widetilde{M}(N)\Big)^4}{N^4T^4\varepsilon^4} = \frac{E\Big(\widetilde{\mathcal{E}}'(N,T)^{\top}P^{\top}Q\widetilde{M}'(N)\Big)^4}{N^4T^4\varepsilon^4}.$$

$P^{\top}Q$ is an orthogonal matrix since both P and Q are orthogonal matrices. From Lemma 2 we can konw that if the elements of $\Sigma_{\widetilde{\mathcal{E}}'}$ and $\Sigma_{\widetilde{M}'}$ are bounded, then

$$\begin{aligned}\sum_{N=1}^{\infty}\sum_{T=1}^{\infty}\frac{E\Big(\widetilde{\mathcal{E}}'(N,T)^{\top}P^{\top}Q\widetilde{M}'(N)\Big)^4}{N^4T^4} \\ \le C_1\sum_{N=1}^{\infty}\sum_{T=1}^{\infty}\frac{1}{N^2T^4} < \infty,\end{aligned}$$

and it is equivalent to

$$\sum_{N=1}^{\infty}\sum_{T=1}^{\infty}\frac{E\Big(\widetilde{\mathcal{E}}(N,T)^{\top}\widetilde{M}(N)\Big)^4}{N^4T^4\varepsilon^4}<\infty.$$

Then we only have to prove the elements of $\Sigma_{\widetilde{\mathcal{E}}'}$ and $\Sigma_{\widetilde{M}'}$ are bounded. Since

$$\mathcal{E}_{\tilde{\boldsymbol{\theta}}t}\sim N(\mathbf{0},I+\Sigma^{'})\quad and\quad \widetilde{\mathcal{E}}(N,T)=\sum_{t=1}^{T}(-\tilde{\mathcal{K}}^{\top})^{t-1}\mathcal{E}_{\tilde{\boldsymbol{\theta}}t},$$

then we have

$$\Sigma_{\widetilde{\mathcal{E}}}=\sum_{t=1}^{\infty}(\tilde{\mathcal{K}}^{\top})^{t-1}(\tilde{\mathcal{K}})^{t-1}+\sum_{t=1}^{\infty}(\tilde{\mathcal{K}}^{\top})^{t-1}\Sigma^{'}(\tilde{\mathcal{K}})^{t-1},$$

where $\Sigma^{'}$ is defined as (19). It is sufficient to consider the latter part. Define

$$L_m=\mathcal{G}^{m-1}(\mathcal{K}+\mathcal{G})-(-\tilde{\mathcal{K}})^{m-1}(\tilde{\mathcal{K}}+\tilde{\mathcal{G}})-\sum_{j=1}^{m}(-\tilde{\mathcal{K}})^{j-1}(\tilde{\mathcal{K}}+\tilde{\mathcal{G}})\mathcal{G}^{m-j-1}(\mathcal{K}+\mathcal{G}),$$

thus

$$\begin{aligned}&\lambda_{\max}\Big(\sum_{t=1}^{\infty}(\tilde{\mathcal{K}}^{\top})^{t-1}\Sigma^{'}(\tilde{\mathcal{K}})^{t-1}\Big)\\&=\lambda_{\max}\Big(\sum_{t=1}^{\infty}\sum_{m=1}^{\infty}(\tilde{\mathcal{K}}^{\top})^{t-1}L_m^{\top}L_m(\tilde{\mathcal{K}})^{t-1}\Big)\\&\le\sum_{t=1}^{\infty}\max_{x_m^{\top}x_m=1}\left(|x_m^{\top}|_e\sum_{m=1}^{\infty}\Big|(\tilde{\mathcal{K}}^{\top})^{t-1}L_m^{\top}\Big|_e\mathbf{1}\right)^2\end{aligned}$$

Just like Lemma 3, it can be easily proved that the largest eigenvalue of $\sum_{t=1}^{\infty}(\tilde{\mathcal{K}}^{\top})^{t-1}\Sigma^{'}(\tilde{\mathcal{K}})^{t-1}$ is finite. Using Weyl Theorem, we know that the largest eigenvalue of $\Sigma_{\widetilde{\mathcal{E}}}$ is bounded and it is equivalent to that the elements of $\Sigma_{\widetilde{\mathcal{E}}'}$ are uniformly bounded with N and T. Also from Lemma 3(b) we know that the largest eigenvalue of $\Sigma_{\widetilde{M}}$ is bounded and it is equivalent to that the elements of $\Sigma_{\widetilde{M}'}$ are uniformly bounded with N and T. Thus we complete this proof. □

Lemma 5. *For arbitrary matrices $A=(a_{ij})_{N\times N}$ and $B=(b_{ij})_{N\times N}$, define $(A*B)_{ij}=a_{ij}b_{ij}$ is the Hadamard product of matrices. Then for arbitrary matrices $A,B,C,D\in\mathbb{R}^{N\times N}$ and N-dimension random vector $\{\mathcal{E}_t:\mathcal{E}_t\sim N(\mathbf{0},I_N),t=\dots,-2,-1,0,1,2,\dots\}$, we have*

$$\operatorname{Cov}\Big(\mathcal{E}_t^{\top}A^{\top}B\mathcal{E}_{t-m},\mathcal{E}_t^{\top}C^{\top}D\mathcal{E}_{t-m}\Big)=\mathbf{1}^{\top}\Big[(AC)^{\top}*(BD^{\top})\Big]\mathbf{1},\tag{20}$$

$$\operatorname{Cov}\Big(\mathcal{E}_t^{\top}A\mathcal{E}_{t-m},\mathcal{E}_t^{\top}B^{\top}C\mathcal{E}_{t-m}\Big)=\mathbf{1}^{\top}\Big[(BA^{\top})*C\Big]\mathbf{1},\tag{21}$$

$$\operatorname{Cov}\Big(\mathcal{E}_t^{\top}A^{\top}A\mathcal{E}_t,\mathcal{E}_t^{\top}B^{\top}B\mathcal{E}_t\Big)=2\cdot\mathbf{1}^{\top}\Big[(AC^{\top})*(BD^{\top})\Big]\mathbf{1},\tag{22}$$

where m is an arbitrary positive integer.

Proof. Before proving the above equations, first we give some conclusions. Since $\mathcal{E}_t\in\mathbb{R}^N$ and $\mathcal{E}_t\sim N(\mathbf{0},I_N)$, let $\mathcal{E}_t=(\varepsilon_{t1},\dots,\varepsilon_{tN})^{\top}$, then $E[\varepsilon_{tk}^2]=1$. Also, since ε_{tk}^2 obeys Chi-square distribution, then $\operatorname{Var}(\varepsilon_{tk}^2)=2$, and $E[\varepsilon_{tk}^4]=\operatorname{Var}(\varepsilon_{tk}^2)+(E[\varepsilon_{tk}^2])^2=3$. Therefore,

$$E[\mathcal{E}_t^{\top}\mathcal{E}_t]=\sum_{k=1}^{N}E[\varepsilon_{tk}^2]=N$$

and

$$\begin{aligned}E[\mathcal{E}_t^{\top}A^{\top}A\mathcal{E}_t]&=E\Big[\sum_{i=1}^{N}(\sum_{j=1}^{N}a_{ij}\varepsilon_{tj})^2\Big]\\&=\sum_{i=1}^{N}\sum_{j=1}^{N}a_{ij}^2=tr(AA^{\top}).\end{aligned}$$

Then we turn back to the proof. Firstly,

$$\begin{aligned}&\operatorname{Cov}\Big(\mathcal{E}_t^{\top}A^{\top}B\mathcal{E}_{t-m},\mathcal{E}_t^{\top}C^{\top}D\mathcal{E}_{t-m}\Big)\\&=E\Big[\Big(\mathcal{E}_t^{\top}A^{\top}B\mathcal{E}_{t-m}-E\Big[\mathcal{E}_t^{\top}A^{\top}B\mathcal{E}_{t-m}\Big]\Big)\\&\quad\Big(\mathcal{E}_t^{\top}C^{\top}D\mathcal{E}_{t-m}-E\Big[\mathcal{E}_t^{\top}C^{\top}D\mathcal{E}_{t-m}\Big]\Big)\Big]\\&=E\Big\{\Big(\mathcal{E}_t^{\top}A^{\top}B\mathcal{E}_{t-m}\Big)\Big(\mathcal{E}_t^{\top}C^{\top}D\mathcal{E}_{t-m}\Big)\Big\}\\&=E\Big\{\Big[\sum_{i=1}^{N}\Big(\sum_{j=1}^{N}a_{ij}\varepsilon_{t,j}\Big)\Big(\sum_{q=1}^{N}b_{iq}\varepsilon_{t-m,q}\Big)\Big]\\&\quad\Big[\sum_{i=1}^{N}\Big(\sum_{j=1}^{N}c_{ij}\varepsilon_{t,j}\Big)\Big(\sum_{q=1}^{N}d_{iq}\varepsilon_{t-m,q}\Big)\Big]\Big\}\\&=E\Big\{\sum_{i=1}^{N}\Big[\sum_{j=1}^{N}a_{ij}\varepsilon_{t,j}\sum_{q=1}^{N}b_{iq}\varepsilon_{t-m,q}\sum_{m=1}^{N}c_{mj}\varepsilon_{t,j}\cdot d_{mq}\varepsilon_{t-m,q}\Big]\Big\}\\&=\sum_{i=1}^{N}\sum_{m=1}^{N}\Big[\sum_{j=1}^{N}a_{ij}c_{mj}\sum_{q=1}^{N}b_{iq}d_{mq}\Big]\\&=\sum_{i=1}^{N}\sum_{m=1}^{N}\Big[(AC^{\top})_{i\times m}(BD^{\top})_{i\times m}\Big]\\&=\mathbf{1}^{\top}\Big[(AC^{\top})*(BD^{\top})\Big]\mathbf{1},\end{aligned}$$

where $A=(a_{ij})_{N\times N}$, $B=(b_{ij})_{N\times N}$, $C=(c_{ij})_{N\times N}$, $D=(d_{ij})_{N\times N}$, $(AC^{\top})_{i\times m}$ means the ith row, mth column element of $AC^{\top}$ and the second equal sign is established since $\mathcal{E}_t$ and $\mathcal{E}_{t-m}$ are independent. The proof of (20) is completed. The proof of (21) is similar to that of Eq. (20) thus we omit it here.

As for Eq. (22),

$$
\begin{aligned}
&\operatorname{Cov}\Big(\mathcal{E}_t^\top A^\top A\mathcal{E}_t, \mathcal{E}_t^\top B^\top B\mathcal{E}_t\Big)\\
&= E\Big[\Big(\mathcal{E}_t^\top A^\top A\mathcal{E}_t - E\Big[\mathcal{E}_t^\top A^\top A\mathcal{E}_t\Big]\Big)\\
&\qquad\Big(\mathcal{E}_t^\top B^\top B\mathcal{E}_t - E\Big[\mathcal{E}_t^\top B^\top B\mathcal{E}_t\Big]\Big)\Big]\\
&= E\Big[\Big(\mathcal{E}_t^\top A^\top A\mathcal{E}_t - tr(AA^\top)\Big)\Big(\mathcal{E}_t^\top B^\top B\mathcal{E}_t - tr(BB^\top)\Big)\Big]\\
&= E\Big[\mathcal{E}_t^\top A^\top A\mathcal{E}_t\mathcal{E}_t^\top B^\top B\mathcal{E}_t\Big] - tr(AA^\top)tr(BB^\top)\\
&= E\Big\{\Big[\sum_{i=1}^N\Big(\sum_{j=1}^N a_{ij}\varepsilon_{t,j}\Big)^2\Big]\Big[\sum_{i=1}^N\Big(\sum_{j=1}^N b_{ij}\varepsilon_{t,j}\Big)^2\Big]\Big\}\\
&\qquad - tr(AA^\top)tr(BB^\top).
\end{aligned}
\tag{23}
$$

Then we have

$$
\begin{aligned}
&E\Big\{\Big[\sum_{i=1}^N\Big(\sum_{j=1}^N a_{ij}\varepsilon_{t,j}\Big)^2\Big]\Big[\sum_{i=1}^N\Big(\sum_{j=1}^N b_{ij}\varepsilon_{t,j}\Big)^2\Big]\Big\}\\
&= E\Big[\sum_{i=1}^N\sum_{j=1}^N a_{ij}^2\varepsilon_{t,j}^2\cdot\sum_{i=1}^N\sum_{j=1}^N b_{ij}^2\varepsilon_{t,j}^2\Big]\\
&\quad + E\Big\{\sum_{i=1}^N\Big[\sum_{j=1}^N a_{ij}\varepsilon_{t,j}\sum_{q=1,q\neq j}^N a_{iq}\varepsilon_{t,q}\\
&\qquad \cdot 2\sum_{m=1}^N b_{mj}\varepsilon_{t,j}b_{mq}\varepsilon_{t,q}\Big]\Big\}.
\end{aligned}
$$

Then let the former part be written as κ_1 and the latter be written as κ_2. Considering κ_1 and κ_2,

$$
\begin{aligned}
\kappa_1 &= E\Big\{\sum_{i=1}^N\sum_{j=1}^N a_{ij}^2\varepsilon_{t,j}^2\sum_{m=1}^N\sum_{q=1,q\neq j}^N b_{mq}^2\varepsilon_{t,q}^2\Big\}\\
&\quad + E\Big\{\sum_{i=1}^N\sum_{j=1}^N a_{ij}^2\sum_{m=1}^N b_{mj}^2\varepsilon_{t,j}^4\Big\}\\
&= \sum_{i=1}^N\sum_{j=1}^N a_{ij}^2\sum_{m=1}^N\sum_{q=1,q\neq j}^N b_{mq}^2 + 3\sum_{i=1}^N\sum_{j=1}^N a_{ij}^2\sum_{m=1}^N b_{mj}^2\\
&= \sum_{i=1}^N\sum_{j=1}^N a_{ij}^2\sum_{m=1}^N\sum_{q=1}^N b_{mq}^2 + 2\sum_{i=1}^N\sum_{j=1}^N a_{ij}^2\sum_{m=1}^N b_{mj}^2\\
&= tr(AA^\top)tr(BB^\top) + 2\sum_{j=1}^N(A^\top A)_{j\times j}(B^\top B)_{j\times j},
\end{aligned}
$$

and

$$
\begin{aligned}
\kappa_2 &= 2E\left\{\sum_{i=1}^N\sum_{m=1}^N\Big[\sum_{j=1}^N a_{ij}\sum_{q=1,q\neq j}^N a_{iq}b_{mj}b_{mq}\Big]\varepsilon_{t,j}^2\varepsilon_{t,q}^2\right\}\\
&= 2\sum_{i=1}^N\sum_{m=1}^N\Big[\sum_{j=1}^N a_{ij}\sum_{q=1,q\neq j}^N a_{iq}b_{mj}b_{mq}\Big]\\
&= \sum_{i=1}^N\sum_{m=1}^N\Big[\sum_{j=1}^N a_{ij}b_{mj}\sum_{q=1,q\neq j}^N a_{iq}b_{mq}\Big]\\
&=2\sum_{i=1}^N\sum_{m=1}^N\Big[\sum_{j=1}^N a_{ij}b_{mj}\sum_{q=1}^N a_{iq}b_{mq}\Big] - 2\sum_{i=1}^N\sum_{j=1}^N a_{ij}^2\sum_{m=1}^N b_{mj}^2\\
&= 2\sum_{i=1}^N\sum_{m=1}^N\Big((AB^\top)_{i\times m}(AB^\top)_{i\times m}\Big)\\
&\qquad - 2\sum_{j=1}^N(A^\top A)_{j\times j}(B^\top B)_{j\times j}\\
&= 2\cdot\mathbf{1}^\top\Big[(AB^\top)*(AB^\top)\Big]\mathbf{1} - 2\sum_{j=1}^N(A^\top A)_{j\times j}(B^\top B)_{j\times j}.
\end{aligned}
$$

Thus, substituting κ_1 and κ_2 into (23),

$$
\begin{aligned}
&\operatorname{Cov}\Big(\mathcal{E}_t^\top A^\top A\mathcal{E}_t, \mathcal{E}_t^\top B^\top B\mathcal{E}_t\Big)\\
&= E\left\{\Big[\sum_{i=1}^N\Big(\sum_{j=1}^N a_{ij}\varepsilon_{t,j}\Big)^2\Big]\Big[\sum_{i=1}^N\Big(\sum_{j=1}^N b_{ij}\varepsilon_{t,j}\Big)^2\Big]\right\}\\
&\quad - tr(AA^\top)tr(BB^\top)\\
&= \kappa_1 + \kappa_2 - tr(AA^\top)tr(BB^\top)\\
&= 2\cdot\mathbf{1}^\top\Big[(AB^\top)*(AB^\top)\Big]\mathbf{1}.
\end{aligned}
$$

The proof of Eq. (22) is completed. □

Lemma 6. *Let two matrices $A, B \in \mathbb{R}^{N\times N}$ and $*$ is the Hadamard product defined in Lemma 5, then it holds that*

$$
\mathbf{1}^\top|A*B|_e\mathbf{1} \le \mathbf{1}^\top|A^T|_e|B|_e\mathbf{1}. \tag{24}
$$

Proof. Let $A = (a_{ij})_{N\times N}$, $B = (b_{ij})_{N\times N}$, then

$$
\mathbf{1}^\top|A*B|_e\mathbf{1} = \sum_{i=1}^N\sum_{j=1}^N|a_{ij}b_{ij}|,
$$

$$
\mathbf{1}^\top|A|_e|B|_e\mathbf{1} = \sum_{i=1}^N\Big[\sum_{j=1}^N|a_{ij}|\sum_{j=1}^N|b_{ij}|\Big].
$$

It is obviously that

$$
\mathbf{1}^\top|A*B|_e\mathbf{1} \le \mathbf{1}^\top|A^T|_e|B|_e\mathbf{1}
$$

The proof is completed. □

Lemma 7. *For $m = 1, 2, \ldots,$ denote*

$$
\begin{aligned}
A_m &= \mathcal{G}^{m-1}(\mathcal{K}+\mathcal{G}) - (-\tilde{\mathcal{K}})^{m-1}(\tilde{\mathcal{K}}+\tilde{\mathcal{G}})\\
&\quad - \sum_{j=1}^m(-\tilde{\mathcal{K}})^{j-1}(\tilde{\mathcal{K}}+\tilde{\mathcal{G}})\mathcal{G}^{m-j-1}(\mathcal{K}+\mathcal{G}),
\end{aligned}
$$

and $A_0 = 1$, then there exists a constant $\rho < 1$ such that

$$\sup_{\boldsymbol{\theta}\in\Theta} |A_m|_e \mathbf{1} \le C\rho^{m-1}\mathbf{1},$$

where C is a constant.

Proof. For simplicity, we omit the subscript of sup in the following. By the definition of A_m, we have

$$\begin{aligned}\sup |A_m|_e\mathbf{1} = \sup \Big|&\mathcal{G}^{m-1}(\mathcal{K}+\mathcal{G}) - (-\tilde{\mathcal{K}})^{m-1}(\tilde{\mathcal{K}}+\tilde{\mathcal{G}}) \\ &- \sum_{j=1}^{m}(-\tilde{\mathcal{K}})^{j-1}(\tilde{\mathcal{K}}+\tilde{\mathcal{G}})\mathcal{G}^{m-j-1}(\mathcal{K}+\mathcal{G})\Big|_e\mathbf{1} \\ \preccurlyeq \sup &\Big|\mathcal{G}^{m-1}(\mathcal{K}+\mathcal{G})\Big|_e\mathbf{1} \\ &+ \sup\Big|(-\tilde{\mathcal{K}})^{m-1}(\tilde{\mathcal{K}}+\tilde{\mathcal{G}})\Big|_e\mathbf{1} \\ &+ \sup\Big|\sum_{j=1}^{m}(-\tilde{\mathcal{K}})^{j-1}(\tilde{\mathcal{K}}+\tilde{\mathcal{G}})\mathcal{G}^{m-j-1}(\mathcal{K}+\mathcal{G})\Big|_e\mathbf{1}.\end{aligned}$$

We just need to prove the first part satisfies the lemma since the others are similar. Consider $\sup\Big|\mathcal{G}^{m-1}(\mathcal{K}+\mathcal{G})\Big|\mathbf{1}$, we have

$$\begin{aligned}&\sup\Big|\mathcal{G}^{m-1}(\mathcal{K}+\mathcal{G})\Big|_e\mathbf{1} \\ \preccurlyeq\ &\Big(|\alpha_0|+|\alpha_1|+|\beta_0|+|\beta_1|\Big)\sup|\mathcal{G}^{m-1}|_e\mathbf{1} \\ \preccurlyeq\ &\Big(|\alpha_0|+|\alpha_1|+|\beta_0|+|\beta_1|\Big)\Big(|\beta_0|+|\beta_1|\Big)^{m-1}\mathbf{1}.\end{aligned}$$

So $\sup\Big|\mathcal{G}^{m-1}(\mathcal{K}+\mathcal{G})\Big|\mathbf{1} \le C_1\rho_1^{m-1}\mathbf{1}$ and $\rho_1 = |\beta_0|+|\beta_1| < 1$. Therefore, the proof is completed. □

Lemma 8. *For $\hat{\boldsymbol{\theta}} \in \Theta$ and its neighborhood $U_m(\hat{\boldsymbol{\theta}})$ which shrinks to $\hat{\boldsymbol{\theta}}$, we have*

$$\lim_{m\to\infty}\liminf_{N,T\to\infty}\inf_{\tilde{\boldsymbol{\theta}}\in U_m(\hat{\boldsymbol{\theta}})}\left\{(NT)^{-1}\sum_{t=1}^{T}\mathcal{E}_{\tilde{\boldsymbol{\theta}}t}^\top\mathcal{E}_{\tilde{\boldsymbol{\theta}}t}\right\} > \sigma^2,$$

where $\sigma^2 = \frac{1}{N}E\Big[\mathcal{E}_t^\top\mathcal{E}_t\Big]$.

Proof. Recalling Eq. (18) and the definition of A_m, $\mathcal{E}_{\tilde{\boldsymbol{\theta}}t}$ can be written as

$$\mathcal{E}_{\tilde{\boldsymbol{\theta}}t} = \mathcal{E}_t + \sum_{i=1}^{\infty}A_i\mathcal{E}_{t-i}.$$

Define $A_0 = 1$, then $\mathcal{E}_{\tilde{\boldsymbol{\theta}}t} = \sum_{i=0}^{\infty}A_i\mathcal{E}_{t-i}$. Taking this expression into the following equation, we have

$$\begin{aligned}&\lim_{m\to\infty}\liminf_{N,T\to\infty}\inf_{\tilde{\boldsymbol{\theta}}\in U_m(\hat{\boldsymbol{\theta}})}\left\{(NT)^{-1}\sum_{t=1}^{T}\mathcal{E}_{\tilde{\boldsymbol{\theta}}t}^\top\mathcal{E}_{\tilde{\boldsymbol{\theta}}t}\right\} \\ \ge &\lim_{m\to\infty}\liminf_{N,T\to\infty}\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\left\{\inf_{\tilde{\boldsymbol{\theta}}\in U_m(\hat{\boldsymbol{\theta}})}\mathcal{E}_{t-i}^\top A_i^\top A_j\mathcal{E}_{t-j}\right\}.\end{aligned}$$

Firstly, we prove the existence of right part of the above inequality. Define

$$\left\{V_t : V_t = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\Big[\inf_{\tilde{\boldsymbol{\theta}}\in U_m(\hat{\boldsymbol{\theta}})}\mathcal{E}_{t-i}^\top A_i^\top A_j\mathcal{E}_{t-j}\Big]\right\}.$$

To prove the existence of the mean, it suffices to prove

$$\lim_{m\to\infty}\liminf_{N,T\to\infty}P\Big(\Big|\frac{1}{NT}\sum_{t=1}^{T}V_t - \frac{1}{N}\mathbb{E}V_t\Big| > \delta\Big) \to 0.$$

In fact, using the Chebyshev's inequality,

$$\begin{aligned}&\lim_{m\to\infty}\liminf_{N,T\to\infty}P\Big(\Big|\frac{1}{NT}\sum_{t=1}^{T}V_t - \frac{1}{N}\mathbb{E}V_t\Big| > \delta\Big) \\ \le &\lim_{m\to\infty}\liminf_{N,T\to\infty}\frac{\mathrm{Var}\Big(\sum_{t=1}^{T}V_t\Big)}{N^2T^2\delta^2} \\ = &\lim_{m\to\infty}\liminf_{N,T\to\infty}\frac{T\mathrm{Cov}(V_T,V_T)+\ldots+\mathrm{Cov}(X_T,X_1)}{N^2T^2\delta^2} \\ \le &\lim_{m\to\infty}\liminf_{N,T\to\infty}\frac{|\mathrm{Cov}(V_T,V_T)|+\ldots+|\mathrm{Cov}(X_T,X_1)|}{N^2T\delta^2}.\end{aligned}$$

Define $\eta_i = \mathrm{Cov}(V_T, V_{T-i})$, then we only need to prove

$$\lim_{m\to\infty}\liminf_{N,T\to\infty}\frac{|\eta_0|+|\eta_1|+\ldots+|\eta_\infty|}{N^2T} \to 0.$$

Consider η_t, we have

$$\begin{aligned}\eta_t &= \mathrm{Cov}(V_t, V_0) \\ &= \mathrm{Cov}\Big(\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\inf_{\tilde{\boldsymbol{\theta}}\in U_m(\hat{\boldsymbol{\theta}})}\mathcal{E}_{t-i}^\top A_i^\top A_j\mathcal{E}_{t-j}, \\ &\qquad \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\inf_{\tilde{\boldsymbol{\theta}}\in U_m(\hat{\boldsymbol{\theta}})}\mathcal{E}_{-i}^\top A_i^\top A_j\mathcal{E}_{-j}\Big) \\ &= \sum_{i=0}^{\infty}\sum_{j=-t}^{\infty}\sum_{m=0}^{\infty}\mathrm{Cov}\Big(\inf_{\tilde{\boldsymbol{\theta}}\in U_m(\hat{\boldsymbol{\theta}})}\mathcal{E}_{-i}^\top A_{t+i}^\top A_{t+j}\mathcal{E}_{-j}, \\ &\qquad \inf_{\tilde{\boldsymbol{\theta}}\in U_m(\hat{\boldsymbol{\theta}})}\mathcal{E}_{-i}^\top A_i^\top A_m\mathcal{E}_{-m}\Big) \\ &\le \sum_{i=0}^{\infty}\sum_{j=-t}^{\infty}\sum_{m=0}^{\infty}\Big[\mathrm{Var}\Big(\inf_{\tilde{\boldsymbol{\theta}}\in U_m(\hat{\boldsymbol{\theta}})}\mathcal{E}_{-i}^\top A_{t+i}^\top A_{t+j}\mathcal{E}_{-j}\Big)\Big]^{1/2} \\ &\qquad \Big[\mathrm{Var}\Big(\inf_{\tilde{\boldsymbol{\theta}}\in U_m(\hat{\boldsymbol{\theta}})}\mathcal{E}_{-i}^\top A_i^\top A_m\mathcal{E}_{-m}\Big)\Big]^{1/2} \\ &\le \sum_{i=0}^{\infty}\sum_{j=-t}^{\infty}\sum_{m=0}^{\infty}\Big[\mathbb{E}\Big(\inf_{\tilde{\boldsymbol{\theta}}\in U_m(\hat{\boldsymbol{\theta}})}\mathcal{E}_{-i}^\top A_{t+i}^\top A_{t+j}\mathcal{E}_{-j}\Big)^2\Big]^{1/2} \\ &\qquad \Big[\mathbb{E}\Big(\inf_{\tilde{\boldsymbol{\theta}}\in U_m(\hat{\boldsymbol{\theta}})}\mathcal{E}_{-i}^\top A_i^\top A_m\mathcal{E}_{-m}\Big)^2\Big]^{1/2}.\end{aligned}$$

Considering $\mathbb{E}\Big(\inf_{\tilde{\boldsymbol{\theta}}\in U_m(\hat{\boldsymbol{\theta}})} \mathcal{E}_{-i}^\top A_i^\top A_m \mathcal{E}_{-m}\Big)^2$, we have

$$\begin{aligned}
&\mathbb{E}\Big(\inf_{\tilde{\boldsymbol{\theta}}\in U_m(\hat{\boldsymbol{\theta}})} \mathcal{E}_{-i}^\top A_i^\top A_m \mathcal{E}_{-m}\Big)^2 \\
&\leq \mathbb{E}\Big(\sup|\mathcal{E}_{-i}^\top||A_i^\top||A_j||\mathcal{E}_{-j}|\Big)^2 \\
&= \mathbb{E}\Big(|\mathcal{E}_{-i}^\top|\sup|A_i^\top|\sup|A_j||\mathcal{E}_{-j}|\Big)^2 \\
&= \mathbb{E}\Bigg\{\Big[\sum_{m=1}^N\Big(\sum_{n=1}^N a_{mn}|\varepsilon_{-i,n}|\Big)\Big(\sum_{q=1}^N b_{mq}|\varepsilon_{-j,q}|\Big)\Big] \\
&\quad \Big[\sum_{p=1}^N\Big(\sum_{n=1}^N a_{pn}|\varepsilon_{-i,n}|\Big)\Big(\sum_{q=1}^N b_{pq}|\varepsilon_{-j,q}|\Big)\Big]\Bigg\} \\
&= \sum_{m=1}^N\sum_{p=1}^N \mathbb{E}\Bigg\{\Big[\Big(\sum_{n=1}^N a_{mn}|\varepsilon_{-i,n}|\Big)\Big(\sum_{q=1}^N b_{mq}|\varepsilon_{-j,q}|\Big)\Big] \\
&\quad \Big[\Big(\sum_{n=1}^N a_{pn}|\varepsilon_{-i,n}|\Big)\Big(\sum_{q=1}^N b_{pq}|\varepsilon_{-j,q}|\Big)\Big]\Bigg\} \\
&= \sum_{m=1}^N\sum_{p=1}^N \mathbb{E}\Big[\Big(\sum_{n=1}^N a_{mn}|\varepsilon_{-i,n}|\Big)\Big(\sum_{n=1}^N a_{pn}|\varepsilon_{-i,n}|\Big)\Big] \\
&\quad \cdot \mathbb{E}\Big[\Big(\sum_{q=1}^N b_{mq}|\varepsilon_{-j,q}|\Big)\Big(\sum_{q=1}^N b_{pq}|\varepsilon_{-j,q}|\Big)\Big],
\end{aligned}$$

where $a_{mn} = (\sup|A_i|)_{mn}$ and $b_{pq} = (\sup|A_j|)_{pq}$. Using Lemma 7,

$$\begin{aligned}
&\mathbb{E}\Big[\Big(\sum_{n=1}^N a_{mn}|\varepsilon_{-i,n}|\Big)\Big(\sum_{n=1}^N a_{pn}|\varepsilon_{-i,n}|\Big)\Big] \\
=&\mathbb{E}\Big[\sum_{n=1}^N a_{mn}a_{pn}|\varepsilon_{-i,n}|^2\Big] \\
&+\sum_{n=1}^N \mathbb{E}\Big[\Big(a_{mn}|\varepsilon_{-i,n}|\Big)\Big(\sum_{l=1,l\neq n}^N a_{pl}|\varepsilon_{-i,l}|\Big)\Big] \\
=&\sum_{n=1}^N a_{mn}a_{pn} + \sum_{n=1}^N \mathbb{E}\Big(a_{mn}|\varepsilon_{-i,n}|\Big)\mathbb{E}\Big(\sum_{l=1,l\neq n}^N a_{pl}|\varepsilon_{-i,l}|\Big) \\
\leq&\sum_{n=1}^N a_{mn}a_{pn} + \mathbb{E}\Big(\sum_{n=1}^N a_{mn}|\varepsilon_{-i,n}|\Big)\mathbb{E}\Big(\sum_{n=1}^N a_{pn}|\varepsilon_{-i,n}|\Big) \\
\leq&\sum_{n=1}^N a_{mn}a_{pn} + \frac{2C^2}{\pi}\rho^{2i-2}.
\end{aligned}$$

Therefore

$$\begin{aligned}
&\mathbb{E}\Big(\inf_{\tilde{\boldsymbol{\theta}}\in U_m(\hat{\boldsymbol{\theta}})} \mathcal{E}_{-i}^\top A_i^\top A_m \mathcal{E}_{-m}\Big)^2 \\
\leq&\sum_{m=1}^N\sum_{p=1}^N\Big[\sum_{n=1}^N a_{mn}a_{pn}+\frac{2C^2}{\pi}\rho^{2i-2}\Big]\Big[\sum_{q=1}^N b_{mq}b_{pq}+\frac{2C^2}{\pi}\rho^{2j-2}\Big] \\
&= \mathbf{1}^\top\Big[\Big(\sup|A_i|\sup|A_i|^\top\Big) * \Big(\sup|A_j|\sup|A_j|^\top\Big)\Big]\mathbf{1} \\
&\quad + \frac{4C^4}{\pi^2}\rho^{2i+2j-4}N^2 + \frac{2C^2}{\pi}\rho^{2j-2}\mathbf{1}^\top\Big(\sup|A_i|\sup|A_i|^\top\Big)\mathbf{1} \\
&\quad + \frac{2C^2}{\pi}\rho^{2i-2}\mathbf{1}^\top\Big(\sup|A_j|\sup|A_j|^\top\Big)\mathbf{1} \\
&\leq C^4\rho^{2i+2j-4}N + \frac{4C^4}{\pi^2}\rho^{2i+2j-4}N^2 \\
&\quad + \frac{2C^4}{\pi}\rho^{2i+2j-4}N + \frac{2C^4}{\pi}\rho^{2i+2j-4}N \\
&= \frac{4C^4}{\pi^2}\rho^{2i+2j-4}N^2 + \Big(C^4 + \frac{4C^4}{\pi}\Big)\rho^{2i+2j-4}N \\
&\leq C_1\rho^{2i+2j-4}N^2,
\end{aligned}$$

where $C_1 = \frac{4C^4}{\pi^2} + \Big(C^4 + \frac{4C^4}{\pi}\Big)$ is a constant. Thus

$$\begin{aligned}
&\lim_{m\to\infty}\liminf_{N,T\to\infty}\frac{|\eta_0| + |\eta_1| + \ldots + |\eta_\infty|}{N^2T} \\
&\leq \lim_{m\to\infty}\liminf_{N,T\to\infty}\frac{1}{N^2T}\sum_{t=0}^\infty\sum_{i=0}^\infty\sum_{j=-t}^\infty\sum_{m=0}^\infty \\
&\quad \Big[\mathbb{E}\Big(\inf_{\tilde{\boldsymbol{\theta}}\in U_m(\hat{\boldsymbol{\theta}})} \mathcal{E}_{-i}^\top A_{t+i}^\top A_{t+j}\mathcal{E}_{-j}\Big)^2\Big]^{1/2} \\
&\quad \cdot\Big[\mathbb{E}\Big(\inf_{\tilde{\boldsymbol{\theta}}\in U_m(\hat{\boldsymbol{\theta}})} \mathcal{E}_{-i}^\top A_i^\top A_m\mathcal{E}_{-m}\Big)^2\Big]^{1/2} \\
&\leq \liminf_{N,T\to\infty}\frac{C_1}{T}\sum_{t=0}^\infty\sum_{i=0}^\infty\sum_{j=-t}^\infty\sum_{m=0}^\infty\rho^{2t+2i+j+m-4} \\
&= \liminf_{N,T\to\infty}\frac{C_1}{T}\frac{1}{\rho^4(1-\rho)^3(1-\rho^2)} = 0.
\end{aligned}$$

Therefore, the mean of

$$\lim_{m\to\infty}\liminf_{N,T\to\infty}\frac{1}{NT}\sum_{t=1}^T\sum_{i=0}^\infty\sum_{j=0}^\infty\Big\{\inf_{\tilde{\boldsymbol{\theta}}\in U_m(\hat{\boldsymbol{\theta}})} \mathcal{E}_{t-i}^\top A_i^\top A_j\mathcal{E}_{t-j}\Big\}$$

exists.

Secondly, we begin to prove the results. According to the Fatou's lemma, it can be easily found that

$$\begin{aligned}
&\lim_{m\to\infty}\liminf_{N\to\infty}\frac{1}{N}\sum_{i=0}^\infty\sum_{j=0}^\infty\mathbb{E}\Big\{\inf_{\tilde{\boldsymbol{\theta}}\in U_m(\hat{\boldsymbol{\theta}})} \mathcal{E}_{t-i}^\top A_i^\top A_j\mathcal{E}_{t-j}\Big\} \\
&\geq \lim_{m\to\infty}\mathbb{E}\Big\{\liminf_{N\to\infty}\frac{1}{N}\sum_{i=0}^\infty\sum_{j=0}^\infty\inf_{\tilde{\boldsymbol{\theta}}\in U_m(\hat{\boldsymbol{\theta}})} \mathcal{E}_{t-i}^\top A_i^\top A_j\mathcal{E}_{t-j}\Big\}.
\end{aligned}$$

Denote

$$D_m = \liminf_{N\to\infty}\frac{1}{N}\sum_{i=0}^\infty\sum_{j=0}^\infty\inf_{\tilde{\boldsymbol{\theta}}\in U_m(\hat{\boldsymbol{\theta}})} \mathcal{E}_{t-i}^\top A_i^\top A_j\mathcal{E}_{t-j}.$$

Then we aim to prove $\lim_{m\to\infty}\mathbb{E}[D_m] > \sigma^2$, just like the proof above, we have

$$|D_m| \le \liminf_{N\to\infty} \frac{1}{N}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \sup |\mathcal{E}_{t-i}^\top||A_i^\top||A_j||\mathcal{E}_{t-j}| = H_m.$$

From the above, it shows that $E[H_m] \to E[H]$. By the Dominated Convergence Theorem, it can be proved that

$$E[D_m] \to E[\liminf_{N\to\infty} \frac{1}{N}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \mathcal{E}_{t-i}^\top A_i^\top A_j \mathcal{E}_{t-j}].$$

Note that

$$\begin{aligned}
&E\Big[\liminf_{N\to\infty} \frac{1}{N}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \mathcal{E}_{t-i}^\top A_i^\top A_j \mathcal{E}_{t-j}\Big] \\
&= E\Big[\liminf_{N\to\infty} \frac{1}{N}\sum_{i=0}^{\infty} \mathcal{E}_{t-i}^\top A_i^\top A_i \mathcal{E}_{t-i}\Big] \\
&= E\Big[\liminf_{N\to\infty} \frac{1}{N}\Big\{\mathcal{E}_t^\top \mathcal{E}_t + \sum_{i=1}^{\infty} \mathcal{E}_{t-i}^\top A_i^\top A_i \mathcal{E}_{t-i}\Big\}\Big] \\
&= \sigma^2 + E\Big[\liminf_{N\to\infty} \frac{1}{N}\sum_{i=1}^{\infty} \mathcal{E}_{t-i}^\top A_i^\top A_i \mathcal{E}_{t-i}\Big].
\end{aligned}$$

So it suffices to prove $\mathbb{E}\Big[\liminf_{N\to\infty} \frac{1}{N}\sum_{i=1}^{\infty} \mathcal{E}_{t-i}^\top A_i^\top A_i \mathcal{E}_{t-i}\Big] > 0$. In addition, we observe that

$$\begin{aligned}
&E\Big[\liminf_{N\to\infty} \frac{1}{N}\sum_{i=1}^{\infty} \mathcal{E}_{t-i}^\top A_i^\top A_i \mathcal{E}_{t-i}\Big] \\
&= E\Big[\liminf_{N\to\infty} \frac{1}{N}\sum_{i=1}^{\infty}\sum_{m=1}^{N}\Big(\sum_{n=1}^{N} a_{i,mn}\mathcal{E}_{t-i,n}\Big)^2\Big] \\
&\ge \liminf_{N\to\infty} \frac{1}{N}\sum_{i=1}^{\infty}\sum_{m=1}^{N}\sum_{n=1}^{N} a_{i,mn}^2 \\
&= \liminf_{N\to\infty} \frac{1}{N}\sum_{i=1}^{\infty} tr(A_i A_i^\top),
\end{aligned}$$

where $a_{i,mn} = (A_i)_{mn}$. First consider A_1. Since

$$\begin{aligned}
A_1 &= (\mathcal{K}+\mathcal{G}) - (\widetilde{\mathcal{K}}+\widetilde{\mathcal{G}}) \\
&= (\alpha_0+\beta_0-\tilde{\alpha}_0-\tilde{\beta}_0)I + (\alpha_1+\beta_1-\tilde{\alpha}_1-\tilde{\beta}_1)W,
\end{aligned}$$

if $\alpha_0+\beta_0 \ne \tilde{\alpha}_0+\tilde{\beta}_0$ or $\alpha_1+\beta_1 \ne \tilde{\alpha}_1+\tilde{\beta}_1$, then

$$\liminf_{N\to\infty} \frac{1}{N} tr(A_1 A_1^\top) > 0$$

since $\sum_{i=1}^N n_i/N > 0$ in the definition, which means what we want to prove is established. When we turn to consider A_2, if $\alpha_0+\beta_0 = \tilde{\alpha}_0+\tilde{\beta}_0$ and $\alpha_1+\beta_1 = \tilde{\alpha}_1+\tilde{\beta}_1$,

$$\begin{aligned}
A_2 &= \mathcal{G}(\mathcal{K}+\mathcal{G}) + \widetilde{\mathcal{K}}(\widetilde{\mathcal{K}}+\widetilde{\mathcal{G}}) - (\widetilde{\mathcal{K}}+\widetilde{\mathcal{G}})(\mathcal{K}+\mathcal{G}) \\
&= (\tilde{\alpha}_0-\alpha_0)(\tilde{\alpha}_0+\tilde{\beta}_0)I \\
&\quad + [(\tilde{\alpha}_0-\alpha_0)(\tilde{\alpha}_1+\tilde{\beta}_1) + (\tilde{\alpha}_1-\alpha_1)(\tilde{\alpha}_0+\tilde{\beta}_0)]W \\
&\quad + (\tilde{\alpha}_1-\alpha_1)(\tilde{\alpha}_1+\tilde{\beta}_1)W^2.
\end{aligned}$$

From the definition, $\alpha_0+\beta_0 \ne 0$ and $\alpha_1+\beta_1 \ne 0$,which means $\tilde{\alpha}_0+\tilde{\beta}_0 \ne 0$ and $\tilde{\alpha}_1+\tilde{\beta}_1 \ne 0$. Note that the diagonal elements of A_2 are

$$(\tilde{\alpha}_0-\alpha_0)(\tilde{\alpha}_0+\tilde{\beta}_0) + (\tilde{\alpha}_1-\alpha_1)(\tilde{\alpha}_1+\tilde{\beta}_1)W_{mm}^2.$$

If $\tilde{\alpha}_0 = \alpha_0$ or $\tilde{\alpha}_1 = \alpha_1$, then just the diagonal elements satisfies

$$\liminf_{N\to\infty} \frac{1}{N} tr(A_2 A_2^\top) > 0,$$

so we just consider other conditions. Let $A_2 = (M_1, \dots, M_N) = (K_1, \dots, K_N)^\top$ where M_i is the column vector of A_2 and K_i is the row vector. It is apparently that

$$P\Big(K_i M_i = \frac{(\tilde{\alpha}_0-\alpha_0)(\tilde{\alpha}_0+\tilde{\beta}_0)}{(\tilde{\alpha}_1-\alpha_1)(\tilde{\alpha}_1+\tilde{\beta}_1)}\Big) = o(N),$$

which means that the probability of the diagonal elements of A_2 equals to 0 is $o(N)$. Therefore we can easily get that

$$\liminf_{N\to\infty} \frac{1}{N} tr(A_2 A_2^\top) > 0.$$

Thus

$$\mathbb{E}\Big[\liminf_{N\to\infty} \frac{1}{N}\sum_{i=1}^{\infty} \mathcal{E}_{t-i}^\top A_i^\top A_i \mathcal{E}_{t-i}\Big] > 0.$$

From the above, we can prove that

$$\lim_{m\to\infty} \liminf_{T\to\infty} \inf_{\tilde{\boldsymbol{\theta}}\in U_m(\hat{\boldsymbol{\theta}})} \Big\{(NT)^{-1}\sum_{t=1}^{T} \mathcal{E}_{\tilde{\boldsymbol{\theta}}t}^\top \mathcal{E}_{\tilde{\boldsymbol{\theta}}t}\Big\} > \sigma^2.$$

The proof is completed. □

ACKNOWLEDGEMENTS

This paper is partially supported by NSFC (12101585, 71771203) and Natural Science Foundation of Anhui. We also thank Zhouyu Shen for valuable suggestions for our work.

Received 8 April 2022

REFERENCES

[1] ARMILLOTTA, M. and FOKIANOS, K. (2021). Poisson Network Autoregression. *arXiv preprint arXiv:2104.06296*.

[2] BEDRICK, E. J. and TSAI, C.-L. (1994). Model selection for multivariate regression in small samples. *Biometrics* 226–231.

[3] BOHN, A., BUCHTA, C., HORNIK, K. and MAIR, P. (2014). Making friends and communicating on Facebook: Implications for the access to social capital. *Social Networks* **37** 29–41.

[4] Chen, E. Y., Fan, J. and Zhu, X. (2020). Community Network Auto-Regression for High-Dimensional Time Series. *arXiv preprint arXiv:2007.05521*.

[5] Dou, B., Parrella, M. L. and Yao, Q. (2016). Generalized Yule–Walker estimation for spatio-temporal models with unknown diagonal coefficients. *Journal of Econometrics* **194** 369–382. MR3536988

[6] Gabaix, X. (2016). Power Laws in Economics: An Introduction. *Journal of Economic Perspectives* **30** 185–206.

[7] Granger, C. W. and Andersen, A. (1978). On the invertibility of time series models. *Stochastic Processes and their Applications* **8** 87–92. MR0511877

[8] Graupe, D., Krause, D. and Moore, J. (1975). Identification of autoregressive moving-average parameters of time series. *IEEE Transactions on Automatic Control* **20** 104–107.

[9] Hofstra, B., Corten, R. and Buskens, V. (2015). Learning in social networks: Selecting profitable choices among alternatives of uncertain profitability in various networks. *Social Networks* **43** 100–112.

[10] Lam, C., Yao, Q. et al. (2012). Factor modeling for high-dimensional time series: inference for the number of factors. *The Annals of Statistics* **40** 694–726. MR2933663

[11] Lewis, K., Kaufman, J., Gonzalez, M., Wimmer, A. and Christakis, N. (2008). Tastes, ties, and time: A new social network dataset using Facebook. com. *Social Networks* **30** 330–342.

[12] Ling, S. (1999). On the probabilistic properties of a double threshold ARMA conditional heteroskedastic model. *Journal of Applied probability* **36** 688–705. MR1737046

[13] Liu, X., Patacchini, E. and Rainone, E. (2017). Peer effects in bedtime decisions among adolescents: a social network model with sampled data. *The Econometrics Journal* **20** S103–S125.

[14] Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer Science & Business Media. MR2172368

[15] Marin, J.-M., Pudlo, P., Robert, C. P. and Ryder, R. J. (2012). Approximate Bayesian computational methods. *Statistics and Computing* **22** 1167–1180. MR2992292

[16] Nowicki, K. and Snijders, T. A. B. (2001). Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association* **96** 1077–1087. MR1947255

[17] Pan, J. and Yao, Q. (2008). Modelling multiple time series via common factors. *Biometrika* **95** 365–379. MR2521589

[18] Park, B. U., Mammen, E., Härdle, W. and Borak, S. (2009). Time series modelling with semiparametric factor dynamics. *Journal of the American Statistical Association* **104** 284–298. MR2504378

[19] Pham, T. D. and Tran, L. T. (1981). On the first-order bilinear time series model. *Journal of Applied Probability* 617–627. MR0621225

[20] Sojourner, A. (2013). Identification of peer effects with missing peer data: Evidence from Project STAR. *The Economic Journal* **123** 574–605.

[21] Sun, Y. (2016). Functional-coefficient spatial autoregressive models with nonparametric spatial weights. *Journal of Econometrics* **195** 134–153. MR3545298

[22] Sun, Y. and Malikov, E. (2018). Estimation and inference in functional-coefficient spatial autoregressive panel data models with fixed effects. *Journal of Econometrics* **203** 359–378. MR3770832

[23] Wang, H., Lin, J. and Wang, J. (2016). Nonparametric spatial regression with spatial autoregressive error structure. *Statistics* **50** 60–75. MR3435169

[24] Wasserman, S., Faust, K. et al. (1994). *Social network analysis: Methods and applications*. Cambridge University Press.

[25] Zhu, X. and Pan, R. (2018). Grouped network vector autoregression. *Statistica Sinca*. MR4257540

[26] Zhu, X., Pan, R., Li, G., Liu, Y., Wang, H. et al. (2017). Network vector autoregression. *The Annals of Statistics* **45** 1096–1123. MR3662449

[27] Zhu, X., Huang, D., Pan, R. and Wang, H. (2018). Multivariate spatial autoregression for large scale social networks. *Journal of Econometrics*. MR4076946

[28] Zhu, X., Wang, W., Wang, H. and Härdle, W. K. (2019a). Network quantile autoregression. *Journal of Econometrics* **212** 345–358. MR3994021

[29] Zhu, X., Chang, X., Li, R. and Wang, H. (2019b). Portal nodes screening for large scale social networks. *Journal of Econometrics* **209** 145–157. MR3944747

Xiao Chen
Department of Statistics and Finance
School of Management
University of Science and Technology of China
China
E-mail address: cx990621@mail.ustc.edu.cn

Yu Chen
Department of Statistics and Finance
School of Management
University of Science and Technology of China
China
E-mail address: cyu@ustc.edu.cn

Xixu Hu
School of Data Science
City University of Hong Kong
Hong Kong
E-mail address: xixuhu2-c@my.cityu.edu.hk