Estimation and Inference on Granger Causality in a Latent High-dimensional Gaussian Vector Autoregressive Model*

Yanqin Fan[†], Fang Han[‡], and Hyeonseok Park[§]

August 28, 2020

Abstract

This paper develops estimation and inference methods for the transition matrices of a latent high-dimensional stationary Gaussian vector autoregressive process when the observed process is an increasing but otherwise unknown transformation of the latent process. Our estimator is based on rank estimators of the large variance and auto-covariance matrices of the latent process. We derive rates of convergence of our estimator based on which we develop inference for Granger causality. Numerical results demonstrate the efficacy of the proposed methods. Although our focus is on the latent process, by the nature of rank estimators, all the methods developed directly apply to the observable process which is a stationary semiparametric high-dimensional Gaussian copula process. In technical terms, our analysis relies heavily on newly developed exponential inequalities for (degenerate) U-statistics under α-mixing condition.

Keywords: High-dimensional time series; Sparse transition matrix; α-mixing; Semiparametric Gaussian copula process; De-biasing inference; Kendall's tau.

JEL Codes: C32, C51

*We thank Yandi Shen, Kevin Song, participants of the Sixth Seattle-Vancouver Econometrics Conference, and seminar participants at Penn State University for helpful discussions. Numerical computation in this paper is done via CSDE Simulation Cluster, which has received partial support from an Eunice Kennedy Shriver National Institute of Child Health and Human Development research infrastructure grant, P2C HD042828.

[†]Corresponding author: Department of Economics, University of Washington, Seattle, WA 98195, USA; email: fany88@uw.edu.

[‡]Department of Statistics, University of Washington, Seattle, WA 98195, USA; e-mail: fanghan@uw.edu.

[§]Department of Economics, University of Washington, Seattle, WA 98195, USA; email: parkh27@uw.edu

1 Introduction

Estimation and inference in high-dimensional stationary Gaussian vector autoregressive (VAR) models have been considered in recent works such as Negahban and Wainwright (2011), Han et al. (2015), and Basu and Michailidis (2015) under the commonly adopted assumption that the underlying process is observable. In many applications in economics and finance, however, the process following a structural model is unobservable. Instead, a process that is deterministically related to the latent process is observable. For example, in structural asset pricing models, the observed equity price is an increasing transformation (unknown) of the latent firm's market value; in structural auction models, the observed bid is an increasing transformation (unknown) of the bidder's private value. This paper considers such a scenario where the latent process follows a high-dimensional stationary Gaussian VAR(p) model for a finite (fixed) p and the observed process is an increasing but otherwise unknown transformation of the latent process. We develop simple methods for the estimation and inference of the transition matrices characterizing the latent process.

To simplify notation and exposition, we focus on developing our methods and theory for the latent VAR(1) model, to be described below, and refer to it as the latent Gaussian VAR model. We will present extensions to general VAR(p) model for any finite (fixed) p in Section 5. In detail, let $\{\boldsymbol{Z}_t\}_{t\in\mathbb{Z}}$, with $\mathbb{Z}$ standing for the set of integers and $\boldsymbol{Z}_t$ in the d-dimensional real space $\mathbb{R}^d$, denote a high-dimensional *latent process* following a stationary Gaussian VAR model with transition matrix $\mathbf{A}$:

$$\boldsymbol{Z}_t = \mathbf{A}\boldsymbol{Z}_{t-1} + \boldsymbol{E}_t,\ \ \boldsymbol{Z}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_0). \tag{1.1}$$

Further, let $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$, with $\boldsymbol{X}_t \in \mathbb{R}^d$, denote an *observable process* such that

$$\boldsymbol{X}_t := \boldsymbol{f}(\boldsymbol{Z}_t) = (f_1(Z_{t,1}), \ldots, f_d(Z_{t,d}))^\top,$$

where $\boldsymbol{f} := \{f_1, \ldots, f_d\}$ consists of univariate strictly increasing *unknown* functions. The dimension $d \equiv d_n$, the parameters $\mathbf{A}$ and $\boldsymbol{\Sigma}_0$ are all allowed to change with the sample size n and d_n may even be larger than n. The scenario thus falls into the application regime of a high-dimensional triangular array framework, which has been explicitly discussed recently in Han and Wu (2019). For notational compactness, we omit the dependence of all model parameters on the sample size n in the rest of the paper.

Following Negahban and Wainwright (2011), Han et al. (2015), and Basu and Michailidis (2015), we impose sparsity constraint on the transition matrix $\mathbf{A}$ and aim at estimating, and also recovering the sparsity pattern of $\mathbf{A}$ in model (1.1) based only on observations $\{\boldsymbol{X}_t\}_{t=1}^n$. Existing methods are now inapplicable because we don't have access to the latent process $\{\boldsymbol{Z}_t\}_{t\in\mathbb{Z}}$. Instead, we propose a new estimator of $\mathbf{A}$ based on $\{\boldsymbol{X}_t\}_{t=1}^n$, bypassing the unknown transformations $\{f_1, \ldots, f_d\}$ without estimating them. It is formulated in two steps. In the first step, we construct rank-based estimators of the variance and first autocovariance matrices of $\{\boldsymbol{Z}_t\}_{t\in\mathbb{Z}}$ from $\{\boldsymbol{X}_t\}_{t=1}^n$ avoiding the estimation of unknown transformations $\{f_1, \ldots, f_d\}$. This is made possible by the nature of rank estimators, which are invariant to increasing transformations, and the relation between the (Pearson) covariance of a bivariate Gaussian distribution and its Kendall's tau. In the second step, we construct a penalized estimator of $\mathbf{A}$ from rank estimators of the variance and first autocovariance matrices of $\{\boldsymbol{Z}_t\}_{t\in\mathbb{Z}}$ using the relation between $\mathbf{A}$ and population variance-autocovariance matrices, known as the Yule-Walker equation.

To justify the proposed estimator, we first derive its rate of convergence and show its consistency to the truth in a high-dimensional asymptotic regime. Under the adopted assumptions, the latent process $\{\boldsymbol{Z}_t\}_{t\in\mathbb{Z}}$ and the observable process $\{\boldsymbol{X}_t\}_{t=1}^n$ are both α-mixing with geometric decaying rates. This paves the way for theoretical studies. However, compared with regression-based estimators of the variance and first autocovariance matrices in Negahban and Wainwright (2011), Han et al. (2015), and Basu and Michailidis (2015), establishing the asymptotic properties of our rank estimators appears to be much more challenging due to their dependence on large random matrices with elements given by U-statistics of α-mixing processes. The analysis is now possible thanks to a recent progress made in Shen et al. (2019), who laid out the path to establishing exponential inequalities for (degenerate) U-statistics of α-mixing processes.

Building on a recently established de-biasing inference framework (Neykov et al., 2018a), we further develop tests for non-Granger causality in model (1.1), which is equivalent to testing non-zeroness of entries of $\mathbf{A}$ (Lütkepohl, 2005). This is via combining the de-biasing strategy with a circular-block-bootstrap (CBB) based variance estimator. The latter has been studied in, e.g, Dehling and Wendler (2010) and Fan et al. (2016), although the analysis here is more demanding due to the complicated form of the target of interest. In addition, when the presence of Granger causality is detected, a sign-consistent estimator of Granger causality is immediate following the strategy in Qiu et al. (2015). Numerical results are presented to illustrate the finite sample behavior of the estimator of $\mathbf{A}$ and also the proposed tests. Extensions of the methods and theory to latent VAR(p) process for any finite and fixed p are also provided.

It is worthwhile pointing out that, although our focus is on the latent process $\{\boldsymbol{Z}_t\}_{t\in\mathbb{Z}}$, the estimation and inference methods we develop for the latent process $\{\boldsymbol{Z}_t\}_{t\in\mathbb{Z}}$ are also valid for the observable process $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$. Model (1.1) and the relation $\boldsymbol{X}_t = \boldsymbol{f}(\boldsymbol{Z}_t)$ imply that the observable process $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$ follows a stationary semiparametric Gaussian copula VAR model with unknown marginal distributions. Unlike the Gaussian VAR model, the Gaussian copula VAR model does not impose any finite moment restrictions on the process $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$ and hence is more suitable for modeling potentially fat-tailed time series. Since the observable process $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$ is nonlinear, this paper also contributes to the growing literature on modelling the possibly nonlinear temporal dependence (Hsing and Wu, 2004; Beare, 2010; Patton, 2012, 2013; Fan and Patton, 2014; Wang and

Xia, 2015), and in particular, the literature on copula-based nonlinear time series models including both the univariate Gaussian copula Markov chains introduced in Chen and Fan (2006b) and studied further in Chen et al. (2009a) and Chen et al. (2009b), and copula-based multivariate time series models studied in Chen and Fan (2006a), Lee and Long (2009), Min and Czado (2014), Rémillard et al. (2012), and Chen et al. (2018). In contrast to the current paper, the model of $\boldsymbol{X}_t$ (and hence also the dimension d) in all these works is assumed to be fixed, while the high-dimensional triangular array framework in this paper induces more challenges.

The rest of this paper is organized as follows. Section 2 specifies the latent model and the observable process, introduces our estimator, and establishes its rates of convergence. Section 3 develops two de-biasing tests for non-Granger causality and introduces a sign consistent estimator of Granger causality. Section 4 presents numerical results. Section 5 extends the methods to stationary latent VAR(p) processes for any finite and fixed p. The last section concludes and discusses extensions. Technical proofs are relegated to a series of appendices.

We now introduce notations used in this paper. For any positive integer d, let $[d] := \{1, 2, \ldots, d\}$. Let $\boldsymbol{v} = (v_1, \ldots, v_d)^\top \in \mathbb{R}^d$ and $\mathbf{M} = [M_{jk}] \in \mathbb{R}^{d\times d}$ be a vector and a matrix of interest. We denote $\boldsymbol{v}_I$ and $\mathbf{M}_{I,J}$ to be the subvector of $\boldsymbol{v}$ and submatrix of $\mathbf{M}$ whose entries are indexed by a set $I \subset [d]$ and sets $I, J \subset [d]$. We denote $\mathbf{M}_{I*}$ and $\mathbf{M}_{*J}$ to be submatrices of $\mathbf{M}$ whose rows are indexed by I and whose columns are indexed by J. For any $p \in (0, \infty]$, we define $\|\boldsymbol{v}\|_p$ to be the vector L_p norm of $\boldsymbol{v}$. Let $\|\mathbf{M}\|_\infty := \max_j \sum_{k=1}^d |M_{jk}|$ be the matrix operator ∞-norm, $\|\mathbf{M}\|_2$ be the matrix spectral norm, and $\|\mathbf{M}\|_{\max}$ be the elementwise maximum norm. For a symmetric matrix $\mathbf{M}$, let $\lambda_{\max}(\mathbf{M})$ and $\lambda_{\min}(\mathbf{M})$ denote the largest and smallest eigenvalues of $\mathbf{M}$. For any univariate function $f(\cdot) : \mathbb{R} \to \mathbb{R}$, define $f(\boldsymbol{v}) = (f(v_1), \ldots, f(v_d))^\top$ and $f(\mathbf{M}) := [f(M_{jk})]$. Let $\mathrm{diag}(\mathbf{M})$ be the diagonal matrix with diagonals $M_{11}, \ldots, M_{dd}$. Let $\mathbf{I}_d$ be the d-dimensional identity matrix. For a random vector $\boldsymbol{X}_t \in \mathbb{R}^d$, let $X_{t,j}$ denote the j-th element of $\boldsymbol{X}_t$. For any $x \in \mathbb{R}$, denote $\mathrm{sign}(x) = \mathbb{1}(x > 0) - \mathbb{1}(x < 0)$, with $\mathbb{1}(\cdot)$ standing for the indicator function. For any two real sequences $\{a_n\}$ and $\{b_n\}$, we write $a_n = O(b_n)$ if there exists an absolute positive constant C such that $|a_n| \le C|b_n|$ for any large enough n. We write $a_n \asymp b_n$ if both $a_n = O(b_n)$ and $b_n = O(a_n)$ hold. We write $a_n = o(b_n)$ if for any absolute positive constant C, we have $|a_n| \le C|b_n|$ for any large enough n. We write $a_n = O_{\mathbb{P}}(b_n)$ and $a_n = o_{\mathbb{P}}(b_n)$ if $a_n = O(b_n)$ and $a_n = o(b_n)$ hold stochastically.

2 Estimation Method and Asymptotic Properties

2.1 The Model, Granger Causality, and α-mixing Property

The latent VAR model we study is characterized by (1.1) and Assumption M below.

Assumption M (Latent Process). (i) $\mathrm{diag}(\boldsymbol{\Sigma}_0) = \mathbf{I}_d$; (ii) $\{\boldsymbol{E}_t\}_{t\in\mathbb{Z}} \overset{\text{i.i.d}}{\sim} N(0, \boldsymbol{\Sigma}_{\boldsymbol{E}})$, where $\boldsymbol{\Sigma}_{\boldsymbol{E}}$ satisfies: $0 < c_{\boldsymbol{E}} < \lambda_{\min}(\boldsymbol{\Sigma}_{\boldsymbol{E}}) \le \lambda_{\max}(\boldsymbol{\Sigma}_{\boldsymbol{E}}) < C_{\boldsymbol{E}} < \infty$ for some absolute constants $c_{\boldsymbol{E}}$ and $C_{\boldsymbol{E}}$; (iii)

The transition matrix $\mathbf{A}$ satisfies: $\|\mathbf{A}\|_2 \le C_{\mathbf{A}} < 1$ for some absolute constant $C_{\mathbf{A}}$ and the following sparsity constraint:

$$\mathbf{A} \in \mathcal{M}(s, M) \quad \text{with } \mathcal{M}(s, M) := \left\{ \mathbf{M} \in \mathbb{R}^{d\times d} : \max_{1\le j\le d} \sum_{k=1}^{d} \mathbb{1}(M_{jk} \neq 0) \le s, \|\mathbf{M}\|_\infty \le M \right\}. \quad (2.1)$$

Here s is a positive integer and M is a positive constant, both of which may depend on d and hence also on n.

Assumption M(i) is a normalization condition ensuring identifiability of $\mathbf{A}$ and $\boldsymbol{f}$. It is not essential and is removable if only the sparsity pattern of $\mathbf{A}$ is of interest. Elementary matrix inequality gives $C_{\mathbf{A}} \le M$. Although it is assumed that $C_{\mathbf{A}} < 1$, there is no such restriction on M, which is allowed to increase to infinity along with d and n.

Under Assumption M, it holds that

$$\mathrm{Var}(\boldsymbol{Z}_t) = \boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}_{\boldsymbol{E}} + \mathbf{A}\boldsymbol{\Sigma}_{\boldsymbol{E}}\mathbf{A}^\top + \mathbf{A}^2\boldsymbol{\Sigma}_{\boldsymbol{E}}(\mathbf{A}^\top)^2 + \cdots,$$

which is also positive definite and the eigenvalues of $\boldsymbol{\Sigma}_0$ are bounded away from zero and infinity by absolute constants. We will show in Proposition 2.2 below that Assumption M yields that the latent process $\{\boldsymbol{Z}_t\}_{t\in\mathbb{Z}}$ is geometrically α-mixing, which implies weak temporal dependence even in high dimensions and is crucial in justifying our proposed methods.

The observable process and the sample information are characterized by Assumption S below.

Assumption S (Observable Process and Sample Information). (i) Let $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$ denote the observable process satisfying $\boldsymbol{X}_t = \boldsymbol{f}(\boldsymbol{Z}_t)$, where $\boldsymbol{f} := \{f_1, \ldots, f_d\}$ consists of univariate strictly increasing *unknown* functions; (ii) Let $\{\boldsymbol{X}_t : t = 1, 2, \ldots, n\}$ denote a length-n segment of $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$ and constitute the data accessible.

Viewed as a model for the observable process $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$, Assumptions M and S(i) define a stationary semiparametric Gaussian copula VAR model. Although the Gaussian process $\{\boldsymbol{Z}_t\}_{t\in\mathbb{Z}}$ has finite moments of all orders, the observable process $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$ may not have any finite moments and is thus more suitable for modeling financial time series than a Gaussian VAR model.

For the Gaussian VAR process $\{\boldsymbol{Z}_t\}_{t\in\mathbb{Z}}$, it is a well-known fact that the sequence $\{Z_{t,k}\}_{t\in\mathbb{Z}}$ *Granger causes* $\{Z_{t,j}\}_{t\in\mathbb{Z}}$ if and only if the (j,k)-th entry of the transition matrix is nonzero (cf. Corollary 2.2.1 in Lütkepohl (2005)). The following proposition, which is a direct consequence of Theorem 6.2 in Qiu et al. (2015), shows that this fact also applies to the observable process $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$.

Proposition 2.1. Under Model (1.1) with Assumption M(ii) and (iii), and Assumption S(i), we have that the sequence $\{X_{t,k}\}_{t\in\mathbb{Z}}$ Granger causes $\{X_{t,j}\}_{t\in\mathbb{Z}}$ if and only if $A_{jk} \neq 0$.

We then proceed to characterize an important property of the observable process $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$, its weak (temporal) dependence. To this end, let's first introduce the α-mixing measure of dependence.

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, for any two σ-algebras $\mathcal{A}$ and $\mathcal{B}$ in $\mathcal{F}$, define the α-mixing measure of dependence between $\mathcal{A}$ and $\mathcal{B}$ as

$$\alpha(\mathcal{A}, \mathcal{B}) := \sup_{A \in \mathcal{A}, B \in \mathcal{B}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|.$$

For any measurable process denoted as $\{\boldsymbol{W}_t\}_{t \in \mathbb{Z}}$, we define its α-mixing coefficient as

$$\alpha(n; \{\boldsymbol{W}_t\}_{t \in \mathbb{Z}}) := \sup_{t \in \mathbb{Z}} \alpha(\mathcal{G}_{-\infty}^{t}, \mathcal{G}_{t+n}^{\infty}), \quad \text{for each } n = 1, 2, \ldots,$$

where for arbitrary $j \in \mathbb{Z}$, $\mathcal{G}_{-\infty}^{j} := \sigma(\boldsymbol{W}_t, t \leq j)$ and $\mathcal{G}_{j}^{\infty} := \sigma(\boldsymbol{W}_t, t \geq j)$ stand for the sigma fields generated by $\{\boldsymbol{W}_t\}_{t \leq j}$ and $\{\boldsymbol{W}_t\}_{t \geq j}$ respectively.

With these concepts introduced, we are now ready to present the weak dependence property of $\{\boldsymbol{X}_t\}_{t \in \mathbb{Z}}$, i.e., its α-mixing coefficient is uniformly geometrically decaying to 0, which holds even if the dimension d explodes as $n \to \infty$. This proposition hinges on the Gaussian assumption and will be heavily used in building up our theory in the follow-up sections.

Proposition 2.2 (Kolmogorov and Rozanov (1960), explicitly stated as Theorem 3.1 in Han and Wu (2019))**.** Suppose that the latent process $\{\boldsymbol{Z}_t\}_{t \in \mathbb{Z}}$ follows the latent Gaussian VAR model (1.1) with Assumption M and the observable process $\{\boldsymbol{X}_t\}_{t \in \mathbb{Z}}$ satisfies Assumption S(i). Then both $\{\boldsymbol{X}_t\}_{t \in \mathbb{Z}}$ and $\{\boldsymbol{Z}_t\}_{t \in \mathbb{Z}}$ are geometrically α-mixing such that

$$\alpha(n; \{\boldsymbol{X}_t\}_{t \in \mathbb{Z}}) = \alpha(n; \{\boldsymbol{Z}_t\}_{t \in \mathbb{Z}}) \leq \kappa_1 \exp(-\kappa_2 n), \quad \text{for all } n \geq 1.$$

Here $\kappa_1 = \sqrt{\frac{C_{\boldsymbol{E}}}{c_{\boldsymbol{E}}(1 - C_{\mathbf{A}}^2)}}$ and $\kappa_2 = -\ln(C_{\mathbf{A}})$ are two absolute positive constants, where $C_{\boldsymbol{E}}$, $c_{\boldsymbol{E}}$, and $C_{\mathbf{A}}$ are defined in Assumption M (ii) and (iii).

2.2 Estimators

A commonly adopted method to estimate parameters in a latent model is to use the equivalent model for the observable process. In our context, the observable process $\{\boldsymbol{X}_t\}_{t \in \mathbb{Z}}$ is a stationary semiparametric Gaussian copula VAR process with unknown marginal distributions. One seemingly plausible approach is to estimate the transition matrix $\mathbf{A}$ and the unknown functions $\{f_1, \ldots, f_d\}$ jointly by a penalized maximum likelihood estimator (MLE). However, in high dimensions this penalized MLE is theoretically intractable. For example, jointly controlling deviations of the estimator $\widehat{f}_j$ from f_j for all $j \in [d]$, even assuming the data are i.i.d., is known to be challenging (cf. the discussions in Sections 2.3 and 4.3 in Liu et al. (2012)). Instead, we propose a simple rank-based estimator of the transition matrix $\mathbf{A}$ without estimating the unknown functions $\{f_1, \ldots, f_d\}$. From a more practical point of view, the Kendall's-tau-type estimators are also known to be more robust to outliers/noises than likelihood based estimators (cf. the simulations conducted in Section 5.2 in Liu et al. (2012)).

In detail, the celebrated Yule-Walker formula yields that $\mathbf{A} = \mathbf{\Sigma}_1^\top \mathbf{\Sigma}_0^{-1}$, where $\mathbf{\Sigma}_1 := \mathbb{E}[\boldsymbol{Z}_t \boldsymbol{Z}_{t+1}^\top]$. This suggests a two-step approach for estimating $\mathbf{A}$, as presented below.

Step 1. We construct "consistent estimators" of $\mathbf{\Sigma}_0$ and $\mathbf{\Sigma}_1$. Notice that $(\boldsymbol{Z}_1^\top, \boldsymbol{Z}_2^\top)^\top$ has the following covariance matrix,

$$\mathbf{\Omega} = \begin{pmatrix} \mathbf{\Sigma}_0 & \mathbf{\Sigma}_1 \\ \mathbf{\Sigma}_1^\top & \mathbf{\Sigma}_0 \end{pmatrix}.$$

For estimating $\mathbf{\Sigma}_0$ and $\mathbf{\Sigma}_1$, it is then sufficient to estimate $\mathbf{\Omega}$. It is well-known – cf. Kruskal (1958, p. 823) – that

$$\mathbf{\Omega} = \sin\left(\underbrace{\frac{\pi}{2} \cdot \mathbb{E}\left[\mathrm{sign}\begin{pmatrix} \boldsymbol{X}_1 - \widetilde{\boldsymbol{X}}_1 \\ \boldsymbol{X}_2 - \widetilde{\boldsymbol{X}}_2 \end{pmatrix} \mathrm{sign}\begin{pmatrix} \boldsymbol{X}_1 - \widetilde{\boldsymbol{X}}_1 \\ \boldsymbol{X}_2 - \widetilde{\boldsymbol{X}}_2 \end{pmatrix}^\top \right]}_{\mathbf{T}} \right),$$

where $(\widetilde{\boldsymbol{X}}_1^\top, \widetilde{\boldsymbol{X}}_2^\top)^\top$ is an independent copy of $(\boldsymbol{X}_1^\top, \boldsymbol{X}_2^\top)^\top$. We propose the following estimator of $\mathbf{\Omega}$:

$$\widehat{\mathbf{\Omega}} = \sin\left(\frac{\pi}{2}\widehat{\mathbf{T}}\right),$$

where

$$\widehat{\mathbf{T}} = \frac{2}{(n-1)(n-2)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n-1} \left[\mathrm{sign}\begin{pmatrix} \boldsymbol{X}_i - \boldsymbol{X}_j \\ \boldsymbol{X}_{i+1} - \boldsymbol{X}_{j+1} \end{pmatrix} \mathrm{sign}\begin{pmatrix} \boldsymbol{X}_i - \boldsymbol{X}_j \\ \boldsymbol{X}_{i+1} - \boldsymbol{X}_{j+1} \end{pmatrix}^\top \right] \tag{2.2}$$

is a high-dimensional matrix of entries formulated as second-order U statistics. We then denote

$$\widehat{\mathbf{\Sigma}}_0 = \widehat{\mathbf{\Omega}}_{[d],[d]} \quad \text{and} \quad \widehat{\mathbf{\Sigma}}_1 = \widehat{\mathbf{\Omega}}_{[d],d+[d]}$$

to be the estimators of variance and auto-covariance matrices $\mathbf{\Sigma}_0$ and $\mathbf{\Sigma}_1$.

Step 2. Using the formula $\mathbf{A} = \mathbf{\Sigma}_1^\top \mathbf{\Sigma}_0^{-1}$, we propose to estimate $\mathbf{A}$ by the following Dantzig-type linear programming estimator,

$$\widehat{\mathbf{A}} := \underset{\mathbf{M} \in \mathbb{R}^{d\times d}}{\mathrm{argmin}} \sum_{j=1}^{d} \sum_{k=1}^{d} |M_{jk}| \quad \text{s.t. } \|\widehat{\mathbf{\Sigma}}_0 \mathbf{M}^\top - \widehat{\mathbf{\Sigma}}_1\|_{\max} \le \lambda. \tag{2.3}$$

Here λ is a tuning parameter to encourage sparsity of the output.

2.3 Rates of Convergence

Estimating matrices $\mathbf{\Sigma}_1$, $\mathbf{\Sigma}_0$ via the Kendal's tau matrix $\widehat{\mathbf{T}}$ in (2.2) allows us to estimate $\mathbf{A}$ without estimating the unknown functions $\{f_1, \ldots, f_d\}$. However, in high dimensions, the asymptotic properties of $\widehat{\mathbf{T}}$ are much harder to establish than the sample covariance matrix used in Negahban and Wainwright (2011), Han et al. (2015), Basu and Michailidis (2015), and more recently in Hecq et al. (2019). Although the asymptotic properties of $\widehat{\mathbf{T}}$ in a finite-dimensional time series model have been well understood (see, for example, Dehling et al. (2017) and the references therein), to our knowledge, Fan et al. (2016) is the only paper in the literature exploring $\widehat{\mathbf{T}}$ in a high-dimensional time series triangular array set-up. Their analysis is, however, based on the assumption that the underlying process is ϕ-mixing. Assuming ϕ-mixing in our case is unrealistic, as a ϕ-mixing Gaussian VAR process is m-dependent (cf. Proposition 1 in Section 2.1 of Doukhan (1994)). Instead, our analysis is built upon Proposition 2.2 that a Gaussian VAR process is α-mixing under conditions therein, and uses heavily the newly developed probability tools in Shen et al. (2019) for α-mixing processes.

In addition to Assumptions M and S, we impose the following scaling condition on (d, n).

Assumption E (i) $\log(d)\log^4(n)/n = O(1)$ so that there exists an absolute positive constant K_0 satisfying $\log(d)\log^4 n \leq K_0 n$ for all large enough n; (ii) $\log(d)/n = o(1)$.

Assumption E is very mild, allowing the dimension d to be even exponentially larger than the sample size n. It is added mainly for the purpose of presentational simplicity. With Assumptions M, S, and E, we are now ready to introduce our first theorem, which is nonasymptotic and characterizes the deviation probability of $\widehat{\mathbf{T}}$ from $\mathbf{T}$.

Theorem 2.1 (Stochastic bound for $\|\widehat{\mathbf{T}} - \mathbf{T}\|_{\max}$)**.** Suppose Assumptions M, S, and E hold. Then, for all sufficiently large n, we have

$$\mathbb{P}\left(\|\widehat{\mathbf{T}} - \mathbf{T}\|_{\max} \geq 2z + \frac{1}{n-2}\right) \leq 16e^{-4}d^{-2} + K_2\{(n-1)d\}^{-2},$$

for any z such that

$$\begin{aligned} z \geq\ & Q\sqrt{\frac{\log(ed)}{n-1}} + \frac{K_3}{4\sqrt{K_0}}\frac{\log\{(n-1)d\}}{n-1} + \frac{K_4}{4\sqrt{K_0}}\frac{1}{\sqrt{(n-1)\log(ed)}} \\ & + \frac{C}{4\sqrt{K_0}}\sqrt{\frac{\log(ed)}{n-1}} + 2C\sqrt{K_2}\frac{1}{(n-1)^2d^2} \\ & + \frac{C\sqrt{K_2}K_3}{2\sqrt{K_0}}\frac{\log\{(n-1)d\}}{(n-1)^2d^2} + \frac{C\sqrt{K_2}K_4}{2\sqrt{K_0}}\frac{1}{d^2(n-1)\sqrt{(n-1)\log(ed)}}, \end{aligned} \tag{2.4}$$

in which Q only depends on κ_1, κ_1 and K_0 introduced in Proposition 2.2 and Assumption E, respectively; K_2 only depends on $c_{\mathbf{A}}$, $c_{\boldsymbol{E}}$ and $C_{\boldsymbol{E}}$; K_3 only depends on K_0; C and K_4 are absolute constants. In particular, we have

$$\|\widehat{\mathbf{T}} - \mathbf{T}\|_{\max} = O_{\mathbb{P}}\left(\sqrt{\frac{\log(ed)}{n}}\right).$$

Remark 2.2. Because $\sin(x)$ is a Lipshitz continuous function, Theorem 2.1 implies that, with probability no smaller than $1 - \epsilon_0$, where

$$\epsilon_0 := 16e^{-4}d^{-2} + K_2\{(n-1)d\}^{-2}, \tag{2.5}$$

we have

$$\|\widehat{\mathbf{\Sigma}}_0 - \mathbf{\Sigma}_0\|_{\max} \le \frac{\pi}{2}\left[2z + \frac{1}{n-1}\right] \quad \text{and} \quad \|\widehat{\mathbf{\Sigma}}_1 - \mathbf{\Sigma}_1\|_{\max} \le \frac{\pi}{2}\left[2z + \frac{1}{n-1}\right]$$

in which z is chosen to satisfy (2.4). Also, we have

$$\|\widehat{\mathbf{\Sigma}}_0 - \mathbf{\Sigma}_0\|_{\max} = O_{\mathbb{P}}\left(\sqrt{\frac{\log(ed)}{n}}\right) \quad \text{and} \quad \|\widehat{\mathbf{\Sigma}}_1 - \mathbf{\Sigma}_1\|_{\max} = O_{\mathbb{P}}\left(\sqrt{\frac{\log(ed)}{n}}\right).$$

Exploiting the proof of Theorem 1 in Han et al. (2015) with Lemmas 1 and 2 therein replaced by the corresponding bounds in Remark 2.2, we immediately have the following theorem that quantifies the deviation of $\widehat{\mathbf{A}}$ from $\mathbf{A}$.

Theorem 2.3 (Stochastic bounds for $\|\widehat{\mathbf{A}} - \mathbf{A}\|_{\max}$ and $\|\widehat{\mathbf{A}} - \mathbf{A}\|_{\infty}$)**.** Suppose Assumptions M, S, and E hold. Let the tuning parameter λ in equation (2.3) be chosen such that

$$\lambda \ge \frac{\pi}{2}(M+1)\left[2z + \frac{1}{n-2}\right],$$

where z is defined in equation (2.4). Then, for any sufficiently large n, it holds that

$$\mathbb{P}(\|\widehat{\mathbf{A}} - \mathbf{A}\|_{\max} \ge 2\|\mathbf{\Sigma}_0^{-1}\|_{\infty}\lambda) \le 16e^{-4}d^{-2} + K_2\{(n-1)d\}^{-2} \quad \text{and}$$
$$\mathbb{P}(\|\widehat{\mathbf{A}} - \mathbf{A}\|_{\infty} \ge 4s\|\mathbf{\Sigma}_0^{-1}\|_{\infty}\lambda) \le 16e^{-4}d^{-2} + K_2\{(n-1)d\}^{-2}.$$

In particular, when λ is picked to be $C\sqrt{\frac{\log(ed)}{n-1}}$ for some large enough absolute constant C, we obtain that

$$\|\widehat{\mathbf{A}} - \mathbf{A}\|_{\max} = O_{\mathbb{P}}\left(\|\mathbf{\Sigma}_0^{-1}\|_{\infty}M\sqrt{\frac{\log(ed)}{n}}\right) \quad \text{and} \quad \|\widehat{\mathbf{A}} - \mathbf{A}\|_{\infty} = O_{\mathbb{P}}\left(s\|\mathbf{\Sigma}_0^{-1}\|_{\infty}M\sqrt{\frac{\log(ed)}{n}}\right).$$

To further illustrate consequences of Theorem 2.3, one can now show that $\|\widehat{\mathbf{A}} - \mathbf{A}\|_{\max} = o_{\mathbb{P}}(1)$ provided that $\|\mathbf{\Sigma}_0^{-1}\|_{\infty}M\sqrt{\frac{\log(ed)}{n}} = o(1)$ and $\lambda = C\sqrt{\log(ed)/n}$ for some pre-specified large enough constant C. Similarly, one has $\|\widehat{\mathbf{A}} - \mathbf{A}\|_{\infty} = o_{\mathbb{P}}(1)$ when $s\|\Sigma_0^{-1}\|_{\infty}M\sqrt{\frac{\log(ed)}{n}} = o(1)$.

Remark 2.4. Consider a particular case, Example 1 in Han et al. (2015). There, the authors proved that $\|\mathbf{\Sigma}_0^{-1}\|_{\infty} = O(1)$ as long as $\mathbf{\Sigma}_0$ is strictly diagonal dominant (Horn and Johnson (2012)), satisfying that, for all $j \in [d]$,

$$\delta_j := |\mathbf{\Sigma}_{0,jj}| - \sum_{j \neq k}|\mathbf{\Sigma}_{0,jk}| \ge \underline{\delta} > 0,$$

where $\underline{\delta}$ is some positive absolute constant. Then, the assumption that $\|\mathbf{\Sigma}_0^{-1}\|_{\infty}M\sqrt{\log(ed)/n} = o(1)$ is equivalent to $M = o\left(\sqrt{n/\log(ed)}\right)$. Similarly, the assumption that $s\|\mathbf{\Sigma}_0^{-1}\|_{\infty}M\sqrt{\log(ed)/n} = o(1)$ implies that $sM = o\left(\sqrt{n/\log(ed)}\right)$. Both allow the parameters s and M to diverge with n.

3 Inference on Granger Causality

In this section, we first construct a test for non-Granger causality of one individual series on another individual series in $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$ using the fact that for any $k \in [d]$, $j \in [d]$, and $j \neq k$, $\{X_{t,k}\}_{t\in\mathbb{Z}}$ *Granger causes* $\{X_{t,j}\}_{t\in\mathbb{Z}}$ if and only if the (j,k)-th entry of the transition matrix, namely A_{jk}, is nonzero. In the presence of Granger causality, we then construct a sign-consistent estimator of the causality relation.

Without loss of generality and also in line with Neykov et al. (2018a), we focus on inference for A_{m1} for some $m \in [d]$. Inference for A_{mj} for $j = 2, ..., d$ is obtained by changing $\mathbf{e}_1 := (1, 0, \dots, 0)^\top$ to $\mathbf{e}_j := (\underbrace{0, \dots, 0}_{j-1}, 1, 0, \dots, 0)^\top$ below. First, we develop a test for non-Granger causality, i.e., testing

$$H_0 : A_{m1} = 0 \text{ against } H_1 : A_{m1} \neq 0. \tag{3.1}$$

The test is constructed in two steps. In the first step, we follow Neykov et al. (2018a) to construct a de-biased estimator of A_{m1} and establish its asymptotic normality. In the second step, we construct a consistent estimator of the asymptotic variance of the de-biased estimator under the null hypothesis. This is done via CBB, and requires more delicate analysis given the strategies paved by Dehling and Wendler (2010) and Fan et al. (2016).

3.1 A De-biased Estimator and its Asymptotic Normality

For simplicity of notation, we write $\boldsymbol{\beta} := \mathbf{A}_{m*} \in \mathbb{R}^d$, the m-th row of $\mathbf{A}$. Decompose $\boldsymbol{\beta}$ as $\boldsymbol{\beta} = (\theta, \boldsymbol{\gamma}^\top)^\top$, where θ is the first element of $\boldsymbol{\beta}$, i.e., $\theta = A_{m1}$. We construct a de-biased estimator of $\theta = A_{m1}$ via the application of Algorithm 1 in Neykov et al. (2018a). It consists of the following three steps.

Step 1 Estimate $\boldsymbol{\beta}$ by $\widehat{\boldsymbol{\beta}} = (\widehat{\theta}, \widehat{\boldsymbol{\gamma}}^\top)^\top$ with

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{v}\in\mathbb{R}^d}{\operatorname{argmin}} \|\boldsymbol{v}\|_1 \text{ such that } \|\widehat{\boldsymbol{\Sigma}}_0\boldsymbol{v} - \widehat{\boldsymbol{\Sigma}}_{1,*m}\|_\infty \leq \lambda,$$

where λ is a tuning parameter. It is easy to see that $\widehat{\boldsymbol{\beta}} = \widehat{\mathbf{A}}_{m*}$, where $\widehat{\mathbf{A}}$ is defined in (2.3).

Step 2 Estimate $\boldsymbol{w} := [\boldsymbol{\Sigma}_0^{-1}]_{*1}$ by

$$\widehat{\boldsymbol{w}} := \underset{\boldsymbol{v}\in\mathbb{R}^d}{\operatorname{argmin}} \|\boldsymbol{v}\|_1 \text{ such that } \|\widehat{\boldsymbol{\Sigma}}_0\boldsymbol{v} - \mathbf{e}_1\|_\infty \leq \lambda',$$

where λ' is another tuning parameter.

Step 3 Estimate θ by solving the estimating equation $\widehat{S}((x, \widehat{\boldsymbol{\gamma}}^\top)^\top) = 0$ with

$$\widehat{S}(\boldsymbol{v}) = \widehat{\boldsymbol{w}}^\top(\widehat{\boldsymbol{\Sigma}}_0\boldsymbol{v} - \widehat{\boldsymbol{\Sigma}}_{1,*m}).$$

It is easy to see that the solution $\widetilde{\theta}$ to the above equation has the following closed form:

$$\widetilde{\theta} = -\frac{\widehat{\boldsymbol{w}}^\top(\widehat{\boldsymbol{\Sigma}}_0\widehat{\boldsymbol{\beta}}(0) - \widehat{\boldsymbol{\Sigma}}_{1,*m})}{\widehat{\boldsymbol{w}}^\top[\widehat{\boldsymbol{\Sigma}}_0]_{*1}} = \widehat{\theta} - \frac{\widehat{\boldsymbol{w}}^\top(\widehat{\boldsymbol{\Sigma}}_0\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\Sigma}}_{1,*m})}{\widehat{\boldsymbol{w}}^\top[\widehat{\boldsymbol{\Sigma}}_0]_{*1}}, \tag{3.2}$$

where $\widehat{\boldsymbol{\beta}}(0) := (0, \widehat{\boldsymbol{\gamma}}^\top)^\top$.

Consistency and asymptotic normality of the above de-biased estimator $\widetilde{\theta}$ are now stated in the following two theorems, with assumptions posed similarly to Neykov et al. (2018a).

Theorem 3.1. Suppose Assumptions M, S, and E hold. Further, suppose that

$$M = O(1), \quad \|\boldsymbol{\Sigma}_0^{-1}\|_\infty = O(1), \quad \text{and} \quad \max\{s_{\boldsymbol{w}}, s\}\frac{\log d}{\sqrt{n}} = o(1),$$

where $s_{\boldsymbol{w}} := \sum_{k=1}^d \mathbb{1}(w_k \neq 0)$. Let $\lambda \asymp \sqrt{\log(ed)/n}$ and $\lambda' \asymp \sqrt{\log(ed)/n}$ and $\lambda \geq C\sqrt{\log(ed)/n}$ and $\lambda' \geq C'\sqrt{\log(ed)/n}$ for some sufficiently large C, C'. Then, we have $\widetilde{\theta} - \theta = o_{\mathbb{P}}(1)$.

Theorem 3.2. Suppose all conditions in Theorem 3.1 hold. Let

$$\sigma_n^2 := (n-1)\mathrm{Var}\Big\{\boldsymbol{w}^\top\Big(\widehat{\boldsymbol{\Sigma}}_0\boldsymbol{\beta} - \widehat{\boldsymbol{\Sigma}}_{1,*m}\Big)\Big\}. \tag{3.3}$$

Assume that $\sigma_n^2 \geq C > 0$ is bounded away from zero by some absolute constant $C > 0$ and for all sufficiently large n. Let

$$U_n = \frac{\sqrt{n-1}}{\sigma_n}(\widetilde{\theta} - \theta).$$

Then $\lim_{n\to\infty}|\mathbb{P}(U_n \leq t) - \Phi(t)| = 0$ for any fixed $t \in \mathbb{R}$. Here $\Phi(\cdot)$ represents the cumulative distribution function of the standard Gaussian.

3.2 Bootstrap Estimation of the Asymptotic Variance

In this section, we use CBB to estimate the asymptotic variance σ_n^2 in (3.3) under H_0. Because our estimator contains $\widehat{\boldsymbol{\Sigma}}_1$, our bootstrap method is based on $\boldsymbol{Y}_t := (\boldsymbol{X}_t^\top, \boldsymbol{X}_{t+1}^\top)^\top$. Note that under H_0, we have $\theta = 0$.

Step 1 Construct $\{\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{n-1}\}$, where $\boldsymbol{Y}_t = (\boldsymbol{X}_t^\top, \boldsymbol{X}_{t+1}^\top)^\top$. Then, draw CBB samples from $\{\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{n-1}\}$. In detail, following Fan et al. (2016), setting ℓ to be some positive integer, we first extend $\{\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{n-1}\}$ by defining $\boldsymbol{Y}_{i+(n-1)} = \boldsymbol{Y}_i$ for $i \geq 1$, randomly draw a block of ℓ consecutive observations, and then obtain a bootstrap sample $\{\boldsymbol{Y}_t^*\}_{t=1}^{b\ell}$ by repeating this $b = \lfloor (n-1)/\ell \rceil$ times so that for each $k = 0, \ldots, b-1$,

$$\mathbb{P}^*(\boldsymbol{Y}_{k\ell+1}^* = \boldsymbol{Y}_j, \ldots, \boldsymbol{Y}_{(k+1)\ell}^* = \boldsymbol{Y}_{j+\ell-1}) = \frac{1}{n-1}$$

for any $j \in [n-1]$. Here $\mathbb{P}^*$ is the bootstrap distribution conditional on $\{\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{n-1}\}$ and $\lfloor x \rceil$ stands for the nearest integer value of $x \in \mathbb{R}$.

Step 2 Construct the bootstrap version of $\widehat{\mathbf{\Omega}}$,

$$\widehat{\mathbf{\Omega}}^* = \sin\left(\underbrace{\frac{\pi}{2}\cdot\frac{2}{b\ell(b\ell-1)}\sum_{i<j}\text{sign}(\boldsymbol{Y}_i^*-\boldsymbol{Y}_j^*)\text{sign}(\boldsymbol{Y}_i^*-\boldsymbol{Y}_j^*)^\top}_{\widehat{\mathbf{T}}^*}\right),$$

and let $\widehat{\mathbf{\Sigma}}_0^* = \widehat{\mathbf{\Omega}}^*_{[d],[d]}$ and $\widehat{\mathbf{\Sigma}}_1^* = \widehat{\mathbf{\Omega}}^*_{[d],d+[d]}$.

Step 3 Use the bootstrap variance defined as

$$\widehat{\sigma}_n^{2*} = (b\ell)\text{Var}^*\left\{\widehat{\boldsymbol{w}}^\top\left(\widehat{\mathbf{\Sigma}}_0^*\widehat{\boldsymbol{\beta}}(0) - \widehat{\mathbf{\Sigma}}^*_{1,*m}\right)\right\}$$

for the consistent estimator of σ_n^2. Here $\text{Var}^*(\cdot)$ stands for the variance operator under $\mathbb{P}^*$.

Theorem 3.3. In addition to assumptions in Theorem 3.2, assume further that $\ell\to\infty$, $\ell^2/n = o(1)$, and $\max\{s_{\boldsymbol{w}}^2, s^2\} = o(\sqrt{n}/\log d)$. Then under H_0,

$$|\widehat{\sigma}_n^{2*} - \sigma_n^2| = o_{\mathbb{P}}(1).$$

3.3 Tests for Granger non-Causality

This section develops two tests for Granger non-causality, i.e., for testing H_0 in (3.1) based on the following two test statistics, where the latter is a modification to the original Wald-type statistic,

$$T_n = \frac{\sqrt{n-1}}{\widehat{\sigma}_n^*}\widetilde{\theta} \quad \text{and} \quad T_n^{\text{adj}} := (\widehat{\boldsymbol{w}}^\top\widehat{\mathbf{\Sigma}}_{0,*1})T_n. \tag{3.4}$$

Theorems 3.2 and 3.3 combined together yield that T_n is asymptotically standard normal under H_0. Similar conclusion also applies to the adjusted statistic T_n^{adj} because $\widehat{\boldsymbol{w}} = 1 + o_{\mathbb{P}}(1)$. However, in finite samples, $\widehat{\boldsymbol{w}}^\top\widehat{\mathbf{\Sigma}}_{0,*1}$ is usually not strictly equal to 1. This motivates the adjusted test statistic that is shown to enjoy better finite sample performance in our simulation setups in Section 4.

Based on T_n and T_n^{adj}, we define the following two tests,

$$\mathbb{T}_{n,\alpha} = \mathbb{1}\left\{|T_n| > z_{1-\alpha/2}\right\} \text{ and } \mathbb{T}^{\text{adj}}_{n,\alpha} = \mathbb{1}\left\{\left|T_n^{\text{adj}}\right| > z_{1-\alpha/2}\right\},$$

where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$ quantile of the standard normal distribution. We then have the following corollary, which justifies the validity of the tests, and is also a direct consequence of Theorems 3.2 and 3.3.

Corollary 3.1 (by Theorems 3.2 and 3.3)**.** Suppose all conditions in Theorems 3.2 and 3.3 hold. We then have

$$\mathbb{P}(\mathbb{T}_{n,\alpha} = 1|H_0) = \alpha + o(1) \text{ and } \mathbb{P}(\mathbb{T}^{\text{adj}}_{n,\alpha} = 1|H_0) = \alpha + o(1).$$

3.4 Estimation of the Sign of Granger Causality

When a test detects Granger causality, it may be of interest to further infer the sign of A_{m1}. Following Qiu et al. (2015), define a new estimator of $\mathbf{A}$ as $\widetilde{\mathbf{A}} = [\widetilde{\mathbf{A}}_{jk}]$, where

$$\widetilde{\mathbf{A}}_{jk} := \widehat{\mathbf{A}}_{jk} \mathbb{1}(|\widehat{\mathbf{A}}_{jk}| \geq \gamma)$$

for some threshold level γ. Then, similar to Theorem 6.4 in Qiu et al. (2015), we have the following sign-consistency theorem.

Theorem 3.4 (Theorem 6.4 in Qiu et al. (2015))**.** Assume that the conditions in Theorem 2.1 hold. Let $\gamma = 2\|\mathbf{\Sigma}_0^{-1}\|_\infty (M+1)z$. Then, with probability no smaller than $1 - \epsilon_0$, where z is defined in equation (2.4) and ϵ_0 is defined in Remark 2.2, we have $\text{sign}(\widetilde{\mathbf{A}}) = \text{sign}(\mathbf{A})$, provided that $\min_{\{(j,k):\mathbf{A}_{jk}>0\}} |\mathbf{A}_{jk}| \geq 2\gamma$.

4 Monte-Carlo Simulation

In this section, we investigate the finite sample performance of the proposed estimator $\widehat{\mathbf{A}}$ and tests for non-Granger causality based on T_n and T_n^{adj}. The data generating process (DGP) used in the simulation is characterized by model (1.1) for the latent process and the transformation $\boldsymbol{f}$ generating the observed process, where the following $\boldsymbol{f}$ is taken from Xue and Zou (2012),

$$\boldsymbol{f} = \{f_1, f_2, f_3, f_4, f_5, f_1, f_2, f_3, f_4, f_5, \dots\},$$

in which

$$\begin{aligned} &f_1(x) = x, \quad f_2(x) = \exp(x), \quad f_3(x) = x^3, \quad f_4(x) = \frac{1}{1+\exp(-x)}, \quad \text{and} \\ &f_5(x) = f_2(x)\mathbb{I}(x < -1) + f_1(x)\mathbb{1}(-1 \leq x \leq 1) + (f_4(x-1)+1)\mathbb{1}(x > 1). \end{aligned} \tag{4.1}$$

The matrices $\mathbf{A}$ and $\mathbf{\Sigma}_0^{-1}$ are tri-diagonal matrices such that

- the diagonal elements of $\mathbf{A}$ are $|\rho_A|$ and sub-diagonal and superdiagonal elements are $\rho_A/3$, where ρ_A is chosen so that $\mathbf{\Sigma}_E = \mathbf{\Sigma}_0 - \mathbf{A}\mathbf{\Sigma}_0\mathbf{A}^\top$ is a positive-definite matrix;
- $\mathbf{\Sigma}_0$ is the normalized $\mathbf{\Xi}^{-1}$ with diagonal elements taking the value 1, where $\mathbf{\Xi}$ is a tri-diagonal matrix with diagonal elements 1 and sub-diagonal and superdiagonal elements 1/3.

Given sample size n, dimension d, and $\rho_A = 0.54$, we generate data in two steps:

DGP Step 1 Generate $\{\boldsymbol{Z}_1, \dots, \boldsymbol{Z}_n\}$ recursively. That is, generate $\boldsymbol{Z}_1$ from $N(\mathbf{0}, \mathbf{\Sigma}_0)$ and given $\boldsymbol{Z}_{t-1}$, generate $\boldsymbol{Z}_t$ from $\boldsymbol{Z}_t = \mathbf{A}\boldsymbol{Z}_{t-1} + \boldsymbol{E}_t$, where $\boldsymbol{E}_t \sim N(\mathbf{0}, \mathbf{\Sigma}_{\boldsymbol{E}})$;

DGP Step 2 Transform $(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n)$ to $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ using $\boldsymbol{X}_t = \boldsymbol{f}(\boldsymbol{Z}_t)$, for $t = 1, \ldots, n$.

Once the data $\{\boldsymbol{X}_t\}_{t=1}^{n}$ is generated, we construct $\{\boldsymbol{Y}_t\}_{t=1}^{n-1}$, where $\boldsymbol{Y}_t = (\boldsymbol{X}_t^\top, \boldsymbol{X}_{t+1}^\top)^\top$ and compute $\widehat{\mathbf{A}}$ using the tuning parameter $\lambda = C\sqrt{\log d/(n-1)}$. Following Neykov et al. (2018a), we choose $d = 60, 80$. The sample sizes are $n = 251, 501, 751$. All the results are obtained from 5000 simulation repetitions. To increase computational speed, we used the fast algorithms for the calculation of Kendall's tau in Pozzi et al. (2012). Especially, we used the implementation by Eleutherios (2012).

We report two sets of results below. The first set of results measures the performance of the estimator $\widehat{\mathbf{A}}$ in terms of its ability to detect the true sparsity pattern in $\mathbf{A}$. The second set of results reports the size and power performance of the two proposed tests for non-Granger causality.

4.1 The Performance of $\widehat{\mathbf{A}}$

Following Liu et al. (2012), we use the receiver operating characteristic (ROC) curves to illustrate the ability of the estimator $\widehat{\mathbf{A}}$ in detecting the sparsity pattern of A. For this, we define the sparsity sets $S_{\mathbf{A}}$ and $S_{\widehat{\mathbf{A}}}$ as follows:

- the pair (i, j) is not an element of the set $S_{\mathbf{A}}$ if and only if $A_{ij} = 0$;
- the pair (i, j) is not an element of the set $S_{\widehat{\mathbf{A}}}$ if and only if $\widehat{A}_{ij} = 0$.

Then, following equations (5.2) and (5.3) in Liu et al. (2012), we define the false positive rate (FPR) and the false negative rate (FNR) given the tuning parameter λ as follows. We first calculate the false positive and false negative numbers.

- $\mathrm{FP}(\lambda)$: the number of pairs in $S_{\widehat{\mathbf{A}}}$ not in $S_{\mathbf{A}}$;
- $\mathrm{FN}(\lambda)$: the number of pairs in $S_{\mathbf{A}}$ not in $S_{\widehat{\mathbf{A}}}$.

Similar to equation (5.4) in Liu et al. (2012), we calculate the FNR, the FPR, and true positive rate (TPR) as follows,

$$\mathrm{FNR}(\lambda) = \frac{\mathrm{FN}(\lambda)}{\mathrm{card}\,(S_{\mathbf{A}})}, \quad \mathrm{FPR}(\lambda) = \frac{\mathrm{FP}(\lambda)}{d^2 - \mathrm{card}\,(S_{\mathbf{A}})}, \quad \mathrm{TPR}(\lambda) = 1 - \mathrm{FNR}(\lambda),$$

where $\mathrm{card}(\cdot)$ outputs the number of elements in the input. Then, we plot the averages of $\mathrm{FPR}(\lambda)$ and $\mathrm{TPR}(\lambda)$ over 5,000 repetitions for given n and d. Here λ is set to be $C\sqrt{\log d/(n-1)}$, where $C \in \{0.10, 0.11, 0.12, ..., 1.00\}$.

Figure 1 shows the ROC plots for $n = 251, 501, 751$ and $d = 80$.

Several observations are in-line. First, for each pair of (n, d), the true positive rate increases along with the false positive rate. This is as expected since, as more edges are selected, it is also

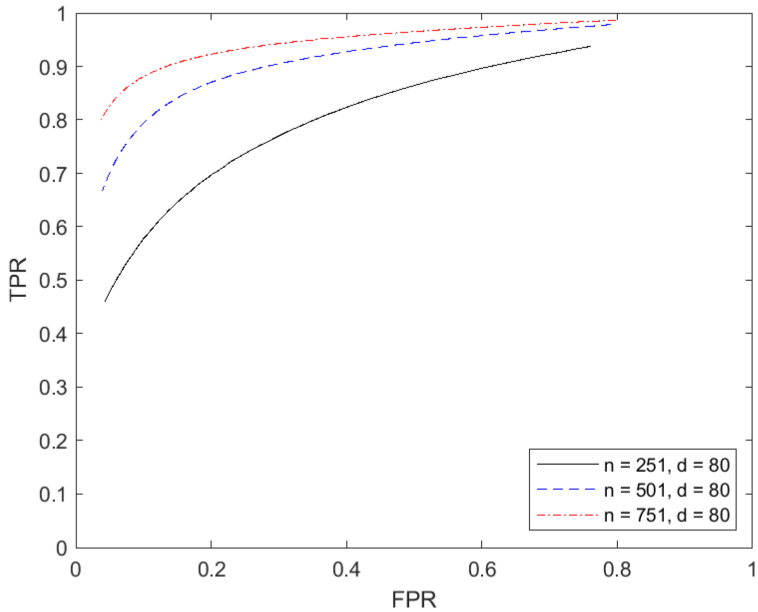

Figure 1: ROC plots when $n = 251$ and $d = 80$ (black solid line); $n = 501$ and $d = 80$ (blue dashed line); and $n = 751$ and $d = 80$ (red dash-dot line).

likely that we will pick more non-zero A_{ij}'s. Secondly, as the sample size increases to 751, the true positive rate has been above 0.9 while the false positive rate is lower than 0.2. This demonstrates the powerfulness of the proposed method in variable selection. Lastly, comparing the three different pairs of (n, d), it could be observed that, as the sample size alone increases, the ROC curves grow higher. This indicates that the proposed method is more capable of correctly selecting the non-zero coefficients as more information is obtained, which is as expected.

4.2 The Size and Power Performance

To examine the size and power performance of T_n and T_n^{adj}, we test the null hypothesis that $A_{1,j} = 0$ for $j = 2, 3, 10, 20, 30, 40, d$. Since $A_{1,2} \neq 0$ and $A_{1,j} = 0$ for $j = 3, 10, 20, 30, 40, d$, rejection probabilities for $A_{1,2}$ give power and those for $A_{1,3}, A_{1,10}, A_{1,20}, A_{1,30}, A_{1,40}, A_{1,d}$ give size. Following Section 5.2 of Neykov et al. (2018a), we compute the debiased estimator using the tuning parameters $\lambda = \lambda' = 0.5\sqrt{\frac{\log d}{n-1}}$. For CBB, we set $\ell = [(n-1)^{1/3}]$ following the discussions in Dehling and Wendler (2010) and compute the bootstrap variance based on $2,000$ bootstrap samples.

Tables 1 to 6 report the results for nominal sizes $1\%, 5\%, 10\%, 15\%, 20\%$. Several conclusions emerge. First, for all settings, the adjusted test has better size than the unadjusted test; second, the sizes of both tests improve and approach the nominal size as n gets larger; third, the power of both tests increases quickly with the sample size; and lastly the results for $d = 60$ and $d = 80$ are qualitatively comparable.

5 Estimation and Inference in Latent VAR(p) Model

Let $\boldsymbol{Z}_t \in \mathbb{R}^d$ follow the VAR(p) process below for a finite and fixed p:

$$\boldsymbol{Z}_t = \mathbf{A}_1 \boldsymbol{Z}_{t-1} + \mathbf{A}_2 \boldsymbol{Z}_{t-2} + \ldots + \mathbf{A}_p \boldsymbol{Z}_{t-p} + \boldsymbol{E}_t, \quad \boldsymbol{Z}_t \sim N(0, \boldsymbol{\Sigma}_0), \tag{5.1}$$

where $\mathbf{A}_1, ..., \mathbf{A}_p$ are unknown transition matrices.

In this section, we assume that Assumption M-p below and Assumption S in Section 2.1 are satisfied. We extend the estimation and inference methods developed in the previous sections to the VAR(p) process for a finite and fixed p. To save space, we present our rank-based estimators of $\mathbf{A}_1, ..., \mathbf{A}_p$ and tests for Granger causality without providing the corresponding rate results for the estimators.

Assumption M-p: i) $\mathrm{diag}(\boldsymbol{\Sigma}_0) = \mathbf{I}_d$; ii) $\{\boldsymbol{E}_t\}_{t\in\mathbb{Z}} \overset{\text{i.i.d.}}{\sim} N(0, \boldsymbol{\Sigma}_{\boldsymbol{E}})$, where $0 < c_{\boldsymbol{E}} < \lambda_{\min}(\boldsymbol{\Sigma}_{\boldsymbol{E}}) \leq \lambda_{\max}(\boldsymbol{\Sigma}_{\boldsymbol{E}}) < C_{\boldsymbol{E}} < \infty$ for some absolute constants $c_{\boldsymbol{E}}$ and $C_{\boldsymbol{E}}$; iii) $\|\mathbf{A}_i\|_2 \leq a_i < 1$ by an absolute constant a_i for all $i = 1, \ldots, p$, and $\sum_{i=1}^{p} a_i < 1$; and iv) the matrix $\mathbf{A} := (\mathbf{A}_1, \ldots, \mathbf{A}_p) \in \mathbb{R}^{d\times pd}$ satisfies:

$$\mathbf{A} \in \mathcal{M}(s, M) \text{ with } \mathcal{M}(s, M) := \left\{ \mathbf{M} \in \mathbb{R}^{d\times pd} : \max_{1\leq j\leq d} \sum_{k=1}^{pd} \mathbb{1}(M_{jk} \neq 0) \leq s, \|\mathbf{M}\|_\infty \leq M \right\}, \tag{5.2}$$

where s is a positive integer which may depend on d and p, and M is a positive constant which may also depend on d and p.

Assumption M-p reduces to Assumption M in Section 2.1 when $p = 1$.

For a Gaussian VAR(p) process $\{\boldsymbol{Z}_t\}_{t\in\mathbb{Z}}$, it is well-known that the sequence $\{Z_{t,k}\}_{t\in\mathbb{Z}}$ *Granger causes* $\{Z_{t,j}\}_{t\in\mathbb{Z}}$ if and only if there exists $i \in [p]$ such that the (j, k)-th element of $\mathbf{A}_i$, $A_{i,jk}$, is non-zero (cf. Corollary 2.2.1 in Lütkepohl (2005)). The following proposition holds for the observable process $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$.

Proposition 5.1. Suppose $\{\boldsymbol{Z}_t\}_{t\in\mathbb{Z}}$ follows VAR(p) model (5.1) with finite and fixed p. Under Assumptions M-p(ii), M-p(iii), and S(i), we obtain that the sequence $\{X_{t,k}\}_{t\in\mathbb{Z}}$ does not Granger cause $\{X_{t,j}\}_{t\in\mathbb{Z}}$ if and only if $A_{i,jk} = 0$ for all $i \in [p]$.

We establish α-mixing property of $\{\boldsymbol{Z}_t\}_{t\in\mathbb{Z}}$ stated below in Appendix A.3.

Theorem 5.1. Suppose $\{\boldsymbol{Z}_t\}_{t\in\mathbb{Z}}$ follows VAR(p) model (5.1) with finite and fixed p. Under Assumptions M-p and S, $\{\boldsymbol{Z}_t\}_{t\in\mathbb{Z}}$ is geometrically α-mixing with mixing coefficient satisfying

$$\alpha(n; \{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}) = \alpha(n; \{\boldsymbol{Z}_t\}_{t\in\mathbb{Z}}) \leq \gamma_1 \exp(-\gamma_2 n), \text{ for all } n \geq 1.$$

Here γ_1 and γ_2 are positive constants that only depend on $c_{\boldsymbol{E}}, C_{\boldsymbol{E}}$, $a_1, \ldots, a_p$, and p.

5.1 Estimators of $\mathbf{A}_1, ..., \mathbf{A}_p$

Let $\boldsymbol{\Omega} = \mathbb{E}\left[\mathfrak{U}_t \mathfrak{U}_t^\top\right]$, where $\mathfrak{U}_t := (\boldsymbol{Z}_{t+p}^\top, \ldots, \boldsymbol{Z}_t^\top)^\top \in \mathbb{R}^{(p+1)d}$. We set $\boldsymbol{\Sigma}_0 := \boldsymbol{\Omega}_{[pd]+d,[pd]+d}$ and $\boldsymbol{\Sigma}_1 := \boldsymbol{\Omega}_{[pd]+d,[d]}$. Then, we estimate $\mathbf{A}_1, \ldots, \mathbf{A}_p$ via the following two steps.

Step 1. Estimate $\boldsymbol{\Omega}$ by $\widehat{\boldsymbol{\Omega}} = \sin\left(0.5\pi\widehat{\mathfrak{T}}\right)$, where

$$\widehat{\mathfrak{T}} = \frac{2}{(n-p)(n-p-1)} \sum_{i=1}^{n-p} \sum_{j=i+1}^{n-p} \left[\sin(\mathfrak{Y}_i - \mathfrak{Y}_j) \sin(\mathfrak{Y}_i - \mathfrak{Y}_j)^\top\right],$$

in which $\mathfrak{Y}_t := (\boldsymbol{X}_{t+p}^\top, \ldots, \boldsymbol{X}_t^\top)^\top \in \mathbb{R}^{(p+1)d}$. Let $\widehat{\boldsymbol{\Sigma}}_0 := \widehat{\boldsymbol{\Omega}}_{[pd]+d,[pd]+d}$ and $\widehat{\boldsymbol{\Sigma}}_1 := \widehat{\boldsymbol{\Omega}}_{[pd]+d,[d]}$.

Step 2. Estimate the stacked matrix $\mathbf{A} = (\mathbf{A}_1, \ldots, \mathbf{A}_p) \in \mathbb{R}^{d \times pd}$ by

$$\widehat{\mathbf{A}} := \underset{\mathbf{M} \in \mathbb{R}^{d \times pd}}{\operatorname{argmin}} \sum_{j=1}^{d} \sum_{k=1}^{pd} |M_{jk}| \text{ such that } \|\widehat{\boldsymbol{\Sigma}}_0 \mathbf{M}^\top - \widehat{\boldsymbol{\Sigma}}_1\|_{\max} \leq \lambda,$$

where λ is a tuning parameter.

5.2 Tests for Granger Causality

We present two tests for Granger non-causality of an individual series $\{X_{t,k}\}_{t\in\mathbb{Z}}$ on another individual series $\{X_{t,m}\}_{t\in\mathbb{Z}}$ under the latent VAR(p) model (5.1) for $m \neq k$, $m \in [d]$, and $k \in [d]$. From Proposition 5.1, it is sufficient to consider

$$H_0 : A_{i,mk} = 0 \text{ for all } i \in [p] \text{ against } H_1 : A_{i,mk} \neq 0 \text{ for some } i \in [p]. \tag{5.3}$$

Following Section 3, we first construct a de-biased estimator of $(A_{1,mk}, \ldots, A_{p,mk})^\top \in \mathbb{R}^p$, then provide a consistent estimator of the asymptotic variance of the de-biased estimator under the null hypothesis via CBB, and finally develops two tests.

5.2.1 De-biased Estimation

For notational compactness, we let

$$\begin{aligned} G \ &: \ = \{k, k+d, k+2d, \ldots, k+(p-1)d\}, \boldsymbol{e}_j := (\underbrace{0, \ldots, 0}_{j-1}, 1, 0, \ldots, 0)^\top \in \mathbb{R}^{pd}, \\ \boldsymbol{\beta} \ &: \ = (\mathbf{A}_{1,m*}, \ldots, \mathbf{A}_{p,m*})^\top \in \mathbb{R}^{pd}, \text{ and } \boldsymbol{\theta} := \boldsymbol{\beta}_G = (A_{1,mk}, \ldots, A_{p,mk})^\top \in \mathbb{R}^p. \end{aligned}$$

For any generic vectors $\boldsymbol{v} = (v_1, \ldots, v_{pd})^\top \in \mathbb{R}^{pd}$ and $\boldsymbol{x} = (x_1, \ldots, x_p)^\top \in \mathbb{R}^p$, we define $\boldsymbol{v}(\boldsymbol{x}) \in \mathbb{R}^{pd}$ as follows,

$$[\boldsymbol{v}(x)]_j = \begin{cases} x_{(j-k)/d+1} & \text{when } j \in G, \\ v_j & \text{otherwise.} \end{cases}$$

That is, $\boldsymbol{v}(\boldsymbol{x})$ is a vector such that the values of $\boldsymbol{v}$ indexed by G are replaced by $\boldsymbol{x}$, keeping the remaining parts intact.

The de-biased estimator $\widetilde{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is constructed via the following three steps.

Step 1. Estimate $\boldsymbol{\beta}$ by

$$\widehat{\boldsymbol{\beta}} := \underset{\boldsymbol{v}\in\mathbb{R}^{pd}}{\operatorname{argmin}}\|\boldsymbol{v}\|_1 \text{ such that } \|\widehat{\boldsymbol{\Sigma}}_0\boldsymbol{v} - \widehat{\boldsymbol{\Sigma}}_{1,*m}\|_\infty \le \lambda,$$

where λ is a tuning parameter.

Step 2. Estimate $\mathbf{W} := [\boldsymbol{\Sigma}_0^{-1}]_{*G} \in \mathbb{R}^{pd\times p}$ by

$$\widehat{\mathbf{W}} := \underset{\mathbf{M}\in\mathbb{R}^{pd\times p}}{\operatorname{argmin}} \sum_{i=1}^{pd}\sum_{j=1}^{p}|M_{ij}| \text{ such that } \|\widehat{\boldsymbol{\Sigma}}_0\mathbf{M} - [\mathbf{I}_{pd}]_{*G}\|_{\max} \le \lambda',$$

where λ' is another tuning parameter. Equivalently, $\widehat{\mathbf{W}} = (\widehat{\boldsymbol{w}}_1, \ldots, \widehat{\boldsymbol{w}}_p) \in \mathbb{R}^{pd\times p}$, where

$$\widehat{\boldsymbol{w}}_j := \underset{\boldsymbol{v}\in\mathbb{R}^{pd}}{\operatorname{argmin}}\|\boldsymbol{v}\|_1 \text{ such that } \|\widehat{\boldsymbol{\Sigma}}_0\boldsymbol{v} - \mathbf{e}_{k+(j-1)p}\|_\infty \le \lambda' \text{ for each } j\in[p].$$

Step 3. Estimate $\boldsymbol{\theta}$ by solving the estimating equation $\widehat{\boldsymbol{S}}(\widehat{\boldsymbol{\beta}}(\boldsymbol{x})) = \mathbf{0}_{p\times 1}$, where $\boldsymbol{x}\in\mathbb{R}^p$ and

$$\widehat{\boldsymbol{S}}(\boldsymbol{v}) = \widehat{\mathbf{W}}^\top(\widehat{\boldsymbol{\Sigma}}_0\boldsymbol{v} - \widehat{\boldsymbol{\Sigma}}_{1,*m}).$$

The solution $\widetilde{\boldsymbol{\theta}}$ to the above equation has the following closed form:

$$\widetilde{\boldsymbol{\theta}} = -\left(\widehat{\mathbf{W}}^\top\widehat{\boldsymbol{\Sigma}}_{0,*G}\right)^{-1}\widehat{\mathbf{W}}^\top\left(\widehat{\boldsymbol{\Sigma}}_0\widehat{\boldsymbol{\beta}}(\mathbf{0}_{p\times 1}) - \widehat{\boldsymbol{\Sigma}}_{1,*m}\right) = \widehat{\boldsymbol{\theta}} - \left(\widehat{\mathbf{W}}^\top\widehat{\boldsymbol{\Sigma}}_{0,*G}\right)^{-1}\widehat{\mathbf{W}}^\top\left(\widehat{\boldsymbol{\Sigma}}_0\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\Sigma}}_{1,*m}\right),$$

where $\widehat{\boldsymbol{\theta}} := \widehat{\boldsymbol{\beta}}_G = (\widehat{A}_{1,mk}, \ldots, \widehat{A}_{p,mk})^\top \in \mathbb{R}^p$.

5.2.2 Bootstrap Estimation of the Asymptotic Variance

Similar to Theorem 3.2, we obtain that the de-biased estimator $\widetilde{\boldsymbol{\theta}}$ is asymptotically normally distributed with the asymptotic variance given by

$$\boldsymbol{\Sigma}_{\boldsymbol{\theta},n} := (n-p)\mathrm{Var}\left\{\mathbf{W}^\top\left(\widehat{\boldsymbol{\Sigma}}_0\boldsymbol{\beta} - \widehat{\boldsymbol{\Sigma}}_{1,*m}\right)\right\}. \tag{5.4}$$

We provide the following estimator of the asymptotic variance $\boldsymbol{\Sigma}_{\boldsymbol{\theta},n}$ in equation (5.4) under $H_0 : \boldsymbol{\theta} = \mathbf{0}_{p\times 1}$.

Step 1. Construct $\{\mathfrak{Y}_t\}_{t=1}^{n-p}$, and draw CBB samples $\{\mathfrak{Y}_t^*\}_{t=1}^{b\ell}$ from $\{\mathfrak{Y}_t\}_{t=1}^{n-p}$, where $b = [(n-p)/\ell]$.

Step 2. Construct $\widehat{\boldsymbol{\Omega}}^* := \sin\left(0.5\pi\widehat{\mathfrak{T}}\right)$, where

$$\widehat{\mathfrak{T}} = \frac{2}{(b\ell)(b\ell-1)}\sum_{i=1}^{b\ell}\sum_{j=i+1}^{b\ell}\left[\sin(\mathfrak{Y}_i^* - \mathfrak{Y}_j^*)\sin(\mathfrak{Y}_i^* - \mathfrak{Y}_j^*)^\top\right].$$

Then, we set $\widehat{\boldsymbol{\Sigma}}_0^* := \widehat{\boldsymbol{\Omega}}^*_{[pd]+d,[pd]+d}$, and $\widehat{\boldsymbol{\Sigma}}_1^* := \widehat{\boldsymbol{\Omega}}^*_{[pd]+d,[d]}$.

Step 3. Calculate the bootstrap variance as

$$\widehat{\boldsymbol{\Sigma}}^*_{\boldsymbol{\theta},n} := (b\ell)\mathrm{Var}^*\left\{\widehat{\mathbf{W}}^\top\left(\widehat{\boldsymbol{\Sigma}}^*_0\widehat{\boldsymbol{\beta}}(\mathbf{0}_{p\times 1}) - \widehat{\boldsymbol{\Sigma}}^*_{1,*m}\right)\right\}.$$

5.2.3 Tests for Granger non-causality

Define the following test statistics,

$$T_n := (n-p)\widetilde{\boldsymbol{\theta}}^\top\left(\widehat{\boldsymbol{\Sigma}}^*_{\boldsymbol{\theta},n}\right)^{-1}\widetilde{\boldsymbol{\theta}} \text{ and}$$
$$T_n^{\mathrm{adj}} := (n-p)\left(\widehat{\mathbf{W}}^\top\widehat{\boldsymbol{\Sigma}}_{0,*G}\widetilde{\boldsymbol{\theta}}\right)^\top\left(\widehat{\boldsymbol{\Sigma}}^*_{\boldsymbol{\theta},n}\right)^{-1}\left(\widehat{\mathbf{W}}^\top\widehat{\Sigma}_{0,*G}\widetilde{\boldsymbol{\theta}}\right).$$

Based on T_n and T_n^{adj}, we define the following two tests,

$$\mathbb{T}_{n,\alpha} = \mathbb{1}\left\{T_n > \chi^2_{p,1-\alpha}\right\} \text{ and } \mathbb{T}^{\mathrm{adj}}_{n,\alpha} = \mathbb{1}\left\{T_n^{\mathrm{adj}} > \chi^2_{p,1-\alpha}\right\},$$

where $\chi^2_{p,1-\alpha}$ is the $(1-\alpha)$ quantile of the chi-square distribution with the degree of the freedom p.

Theorem 5.2. For the latent VAR(p) model (5.1) with finite and fixed p, suppose that Assumptions M-p, S, and E hold. We assume that $\lambda_{\min}(\boldsymbol{\Sigma}_{\boldsymbol{\theta},n}) \geq C_{\boldsymbol{\theta}} > 0$ by some absolute constant $C_{\boldsymbol{\theta}}$. In addition, we assume that $\ell \to \infty$, $\ell^2/n = o(1)$,

$$M = O(1), \ \|\boldsymbol{\Sigma}_0^{-1}\|_\infty = O(1), \text{ and } \max\{s^2_{\mathbf{W}}, s^2\}\frac{\log(d)}{\sqrt{n}} = o(1),$$

where $s_{\mathbf{W}} := \max_{1\leq k\leq p}\sum_{j=1}^{pd}\mathbb{1}(W_{jk} \neq 0)$. Let $\lambda \asymp \sqrt{\log(epd)/n}$, $\lambda' \asymp \sqrt{\log(epd)/n}$, $\lambda \geq C\sqrt{\log(epd)/n}$, and $\lambda' \geq C'\sqrt{\log(epd)/n}$ for some sufficiently large C and C'. We then have

$$\mathbb{P}(\mathbb{T}_{n,\alpha} = 1|H_0) = \alpha + o(1) \text{ and } \mathbb{P}(\mathbb{T}^{\mathrm{adj}}_{n,\alpha} = 1|H_0) = \alpha + o(1).$$

6 Concluding Remarks

In this paper, we have constructed a simple rank-based estimator of the transition matrix in a high dimension latent VAR(p) model for any finite order p and established its rates of convergence. We have also developed tests for Granger non-causality of one individual series on another series in the high-dimensional latent model and provided numerical evidence on their finite sample performance.

It is worth mentioning that all the methods developed in this paper apply directly to the corresponding high dimension latent VAR(p) model with non-zero mean, although we are unable to estimate the mean. Several extensions are possible. First, if it is desirable to test Granger non-causality of an increasing number of individual series on another series in the high-dimensional latent model, the idea underlying multiple testing and the method of multiplier bootstrap in Chernozhukov et al. (2013) may be adopted. Second, allowing for both increasing d and increasing p is interesting but technically challenging.

References

Basu, S. and Michailidis, G. (2015). Regularized estimation in sparse high-dimensional time series models. *The Annals of Statistics*, 43(4):1535–1567.

Beare, B. K. (2010). Copulas and temporal dependence. *Econometrica*, 78(1):395–410.

Bradley, R. C. (1983). Approximation theorems for strongly mixing random variables. *The Michigan Mathematical Journal*, 30(1):69–81.

Bradley, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probability Surveys*, 2(1):107–144.

Chen, X. and Fan, Y. (2006a). Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. *Journal of Econometrics*, 135(1-2):125–154.

Chen, X. and Fan, Y. (2006b). Estimation of copula-based semiparametric time series models. *Journal of Econometrics*, 130(2):307–335.

Chen, X., Huang, Z., and Yi, Y. (2018). On estimation of multivariate semiparametric GARCH filtered copula models. China Center for Economic Research Working Paper Series E2018025. http://img.bimba.pku.edu.cn/resources/file/13/2018/12/31/201812311606304.pdf.

Chen, X., Koenker, R., and Xiao, Z. (2009a). Copula-based nonlinear quantile autoregression. *Econometrics Journal*, 12(SUPPL. 1):S50–S67.

Chen, X., Wu, W. B., and Yi, Y. (2009b). Efficient estimation of copula-based semiparametric Markov models. *The Annals of Statistics*, 37(6 B):4214–4253.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics*, 41(6):2786–2819.

Dedecker, J., Doukhan, P., Lang, G., Leon, J. R., Louhichi, S., and Prieur, C. (2007). *Weak Dependence: With Examples and Applications*. Springer-Verlag, New York.

Dehling, H., Vogel, D., Wendler, M., and Wied, D. (2017). Testing for changes in Kendall's tau. *Econometric Theory*, 33(6):1352–1386.

Dehling, H. and Wendler, M. (2010). Central limit theorem and the bootstrap for U-statistics of strongly mixing data. *Journal of Multivariate Analysis*, 101(1):126–137.

Doukhan, P. (1994). *Mixing: Properties and Examples*. Number 85 in Lecture Notes in Statistics. Springer-Verlag, New York.

Ekström, M. (2014). A general central limit theorem for strong mixing sequences. *Statistics and Probability Letters*, 94:236–238.

Eleutherios, L. (2012). Weighted Kendall rank correlation matrix. MATLAB Central File Exchange. Retrieved Jan 12, 2015. https://www.mathworks.com/matlabcentral/fileexchange/27361-weighted-kendall-rank-correlation-matrix.

Fan, J., Han, F., Liu, H., and Vickers, B. (2016). Robust inference of risks of large portfolios. *Journal of Econometrics*, 194(2):298–308.

Fan, Y. and Patton, A. J. (2014). Copulas in econometrics. *Annual Review of Economics*, 6(1):179–200.

Fang, K. T., Kotz, S., and Ng, K. W. (2017). *Symmetric Multivariate and Related Distributions*. CRC Press.

Gómez, V. (2016). *Multivariate Time Series with Linear State Space Structure*. Springer International Publishing.

Gonçalves, S. and White, H. (2002). The bootstrap of the mean for dependent heterogeneous arrays. *Econometric Theory*, 18(6):1367–1384.

Han, F. and Li, Y. (2019). Moment bounds for Large autocovariance matrices under dependence. *Journal of Theoretical Probability* (in press).

Han, F., Lu, H., and Liu, H. (2015). A direct estimation of high dimensional stationary vector autoregressions. *Journal of Machine Learning Research*, 16:3115–3150.

Han, F. and Wu, W. B. (2019). Probability inequalities for high dimensional time series under a triangular array framework. *arXiv preprint arXiv:1907.06577 [math.ST]*, pages 1–21.

Hecq, A., Margaritella, L., and Smeekes, S. (2019). Granger causality testing in high-dimensional VARs: a post-double-selection procedure. *arXiv preprint arXiv:1902.10991 [econ.EM]*, pages 1–43.

Horn, R. A. and Johnson, C. R. (2012). *Matrix Analysis*. Cambridge University Press, second edition.

Hsing, T. and Wu, W. B. (2004). On weighted U-statistics for stationary processes. *The Annals of Probability*, 32(2):1600–1631.

Kolmogorov, A. N. and Rozanov, Y. A. (1960). On strong mixing conditions for stationary Gaussian processes. *Theory of Probability and Its Applications*, 5(2):204–208.

Kruskal, W. H. (1958). Ordinal Measures of Association. *Journal of the American Statistical Association*, 53(284):814–861.

Lahiri, S. N. (2003). *Resampling Methods for Dependent Data*. Springer-Verlag, New York.

Lee, T. H. and Long, X. (2009). Copula-based multivariate GARCH model with uncorrelated dependent errors. *Journal of Econometrics*, 150(2):207–218.

Liu, H., Han, F., Yuan, M., Lafferty, J., and Wasserman, L. (2012). High-dimensional semiparametric Gaussian copula graphical models. *The Annals of Statistics*, 40(4):2293–2326.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer-Verlag, Berlin Heidelberg.

Merlevède, F., Peligrad, M., and Rio, E. (2009). Bernstein inequality and moderate deviations under strong mixing conditions. In Houdre, C., Koltchinskii, V., Mason, D. M., and Peligrad, M., editors, *High Dimensional Probability V: The Luminy Volume*, volume 5 of *Institute of Mathematical Statistics (IMS) Collections*, pages 273–292. Institute of Mathematical Statistics, Beachwood, Ohio, USA.

Min, A. and Czado, C. (2014). SCOMDY models based on pair-copula constructions with application to exchange rates. *Computational Statistics and Data Analysis*, 76:523–535.

Negahban, S. and Wainwright, M. J. (2011). Estimation of (near) low-rank matrices with noise and high-dimensional scaling. *The Annals of Statistics*, 39(2):1069–1097.

Neykov, M., Ning, Y., Liu, J. S., and Liu, H. (2018a). A unified theory of confidence regions and testing for high-dimensional estimating equations. *Statistical Science*, 33(3):427–443.

Neykov, M., Ning, Y., Liu, J. S., and Liu, H. (2018b). Supplement to "A unified theory of confidence regions and testing for high-dimensional estimating equations".

Patton, A. (2013). *Chapter 16 - Copula Methods for Forecasting Multivariate Time Series*, volume 2 of *Handbook of Economic Forecasting*. Elsevier.

Patton, A. J. (2012). A review of copula models for economic time series. *Journal of Multivariate Analysis*, 110:4–18.

Politis, D. and Romano, J. (1992). A circular block-resampling procedure for stationary data. In LePage, R. and Billard, L., editors, *Exploring the Limits of Bootstrap*, pages 263–270. John Wiley and Sons.

Pötscher, B. M. and Prucha, I. R. (1997). *Dynamic Nonlinear Econometric Models: Asymptotic Theory*. Springer-Verlag, Berlin Heidelberg.

Pozzi, F., Di Matteo, T., and Aste, T. (2012). Exponential smoothing weighted correlations. *The European Physical Journal B*, 85(6):175.

Qiu, H., Xu, S., Han, F., Liu, H., and Caffo, B. (2015). Robust estimation of transition matrices in high dimensional heavy-tailed vector autoregressive processes. *32nd International Conference on Machine Learning, ICML 2015*, 3:1843–1851.

Rémillard, B., Papageorgiou, N., and Soustra, F. (2012). Copula-based semiparametric models for multivariate time series. *Journal of Multivariate Analysis*, 110:30–42.

Shen, Y., Han, F., and Witten, D. (2019). Tail behavior of dependent V-statistics and its applications. *arXiv preprint arXiv:1902.02761 [math.ST]*.

Wang, T. and Xia, Y. (2015). Whittle likelihood estimation of nonlinear autoregressive models with moving average residuals. *Journal of the American Statistical Association*, 110(511):1083–1099.

Xue, L. and Zou, H. (2012). Regularized rank-based estimation of high-dimensional nonparanormal graphical models. *The Annals of Statistics*, 40(5):2541–2571.

Yoshihara, K.-i. (1976). Limiting behavior of U-statistics for stationary, absolutely regular processes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 35(3):237–252.

Table 1: The power and size when $n = 251$, $d = 80$ and $\lambda = 0.5\sqrt{\log d/(n-1)}$

	1%		5%		10%		15%		20%	
Index	T_n	T_n^{adj}	T_n	T_n^{adj}	T_n	T_n^{adj}	T_n	T_n^{adj}	T_n	T_n^{adj}
(1, 2)	0.4892	0.4300	0.6796	0.6368	0.7632	0.7298	0.8186	0.7934	0.8552	0.8350
(1, 3)	0.0316	0.0202	0.1006	0.0764	0.1634	0.1370	0.2188	0.1920	0.2724	0.2386
(1, 10)	0.0202	0.0122	0.0792	0.0618	0.1386	0.1140	0.1942	0.1628	0.2526	0.2204
(1, 20)	0.0228	0.0164	0.0794	0.0632	0.1400	0.1162	0.1928	0.1658	0.2492	0.2150
(1, 30)	0.0262	0.0180	0.0852	0.0648	0.1476	0.1174	0.2068	0.1754	0.2620	0.2296
(1, 40)	0.0250	0.0174	0.0854	0.0640	0.1534	0.1244	0.2086	0.1802	0.2572	0.2294
(1, 80)	0.0256	0.0180	0.0826	0.0646	0.1436	0.1178	0.1982	0.1712	0.2468	0.2182

Table 2: The power and size when $n = 501$, $d = 80$ and $\lambda = 0.5\sqrt{\log d/(n-1)}$

	1%		5%		10%		15%		20%	
Index	T_n	T_n^{adj}	T_n	T_n^{adj}	T_n	T_n^{adj}	T_n	T_n^{adj}	T_n	T_n^{adj}
(1, 2)	0.8454	0.8214	0.9412	0.9302	0.9692	0.9632	0.9816	0.9788	0.9882	0.9856
(1, 3)	0.0202	0.0148	0.0748	0.0604	0.1370	0.1164	0.1922	0.1748	0.2480	0.2244
(1, 10)	0.0212	0.0168	0.0742	0.0610	0.1264	0.1122	0.1778	0.1592	0.2284	0.2064
(1, 20)	0.0192	0.0154	0.0732	0.0602	0.1344	0.1150	0.1862	0.1676	0.2424	0.2176
(1, 30)	0.0180	0.0146	0.0714	0.0580	0.1292	0.1132	0.1824	0.1592	0.2376	0.2160
(1, 40)	0.0196	0.0140	0.0750	0.0616	0.1296	0.1144	0.1824	0.1642	0.2336	0.2120
(1, 80)	0.0192	0.0160	0.0646	0.0550	0.1272	0.1064	0.1856	0.1604	0.2368	0.2152

Table 3: The power and size when $n = 751$, $d = 80$ and $\lambda = 0.5\sqrt{\log d/(n-1)}$

	1%		5%		10%		15%		20%	
Index	T_n	T_n^{adj}	T_n	T_n^{adj}	T_n	T_n^{adj}	T_n	T_n^{adj}	T_n	T_n^{adj}
(1, 2)	0.9686	0.9620	0.9914	0.9900	0.9962	0.9956	0.9982	0.9976	0.9988	0.9988
(1, 3)	0.0218	0.0166	0.0698	0.0606	0.1276	0.1132	0.1802	0.1610	0.2308	0.2136
(1, 10)	0.0172	0.0148	0.0626	0.0532	0.1168	0.1056	0.1652	0.1504	0.2134	0.1956
(1, 20)	0.0144	0.0110	0.0630	0.0528	0.1184	0.1028	0.1714	0.1570	0.2258	0.2048
(1, 30)	0.0178	0.0134	0.0658	0.0568	0.1222	0.113	0.1758	0.1602	0.2342	0.2138
(1, 40)	0.0184	0.0132	0.0746	0.0638	0.1304	0.1186	0.1776	0.1614	0.2274	0.2108
(1, 80)	0.0150	0.0122	0.0638	0.0564	0.1268	0.1076	0.1766	0.1626	0.2294	0.2126

Table 4: The power and size when $n = 251$, $d = 60$ and $\lambda = 0.5\sqrt{\log d/(n-1)}$

	1%		5%		10%		15%		20%	
Index	T_n	T_n^{adj}	T_n	T_n^{adj}	T_n	T_n^{adj}	T_n	T_n^{adj}	T_n	T_n^{adj}
(1, 2)	0.4970	0.4412	0.6958	0.6554	0.7822	0.7516	0.8300	0.8080	0.8646	0.8466
(1, 3)	0.0306	0.0218	0.0970	0.0764	0.1560	0.1330	0.2096	0.1826	0.2626	0.2334
(1, 10)	0.0254	0.0174	0.0800	0.0624	0.1398	0.1152	0.1906	0.1662	0.2390	0.2102
(1, 20)	0.0208	0.0148	0.0782	0.0610	0.1410	0.1150	0.1934	0.1674	0.2496	0.2170
(1, 30)	0.0250	0.0154	0.0792	0.0640	0.1358	0.1144	0.1914	0.1622	0.2428	0.2138
(1, 40)	0.0276	0.0198	0.0826	0.0660	0.1430	0.1214	0.1948	0.1672	0.2474	0.2156
(1, 80)	0.0246	0.0174	0.0804	0.0646	0.1338	0.1106	0.1902	0.1616	0.2402	0.2088

Table 5: The power and size when $n = 501$, $d = 60$ and $\lambda = 0.5\sqrt{\log d/(n-1)}$

	1%		5%		10%		15%		20%	
Index	T_n	T_n^{adj}	T_n	T_n^{adj}	T_n	T_n^{adj}	T_n	T_n^{adj}	T_n	T_n^{adj}
(1, 2)	0.8544	0.8300	0.9436	0.9346	0.9674	0.9640	0.9800	0.9754	0.9866	0.9844
(1, 3)	0.0238	0.0184	0.0784	0.0644	0.1382	0.1196	0.1924	0.1730	0.2518	0.2260
(1, 10)	0.0186	0.0134	0.0696	0.0560	0.1226	0.1076	0.1776	0.1550	0.2334	0.2082
(1, 20)	0.0234	0.0176	0.0736	0.0602	0.1316	0.1130	0.1816	0.1630	0.2296	0.2110
(1, 30)	0.0182	0.0136	0.0704	0.0594	0.1262	0.1110	0.1842	0.1624	0.2348	0.2154
(1, 40)	0.0192	0.0162	0.0732	0.0624	0.1344	0.1168	0.1844	0.1660	0.2312	0.2108
(1, 80)	0.0198	0.0142	0.0694	0.0612	0.1222	0.1072	0.1760	0.1552	0.2254	0.2062

Table 6: The power and size when $n = 751$, $d = 60$ and $\lambda = 0.5\sqrt{\log d/(n-1)}$

	1%		5%		10%		15%		20%	
Index	T_n	T_n^{adj}	T_n	T_n^{adj}	T_n	T_n^{adj}	T_n	T_n^{adj}	T_n	T_n^{adj}
(1, 2)	0.9740	0.9672	0.9940	0.9926	0.9986	0.9970	0.9994	0.9994	0.9996	0.9996
(1, 3)	0.0174	0.0146	0.0666	0.0578	0.1234	0.1116	0.1770	0.1616	0.2300	0.2136
(1, 10)	0.0164	0.0132	0.0670	0.0570	0.1238	0.1110	0.1776	0.1654	0.2290	0.2108
(1, 20)	0.0154	0.0118	0.0634	0.0546	0.1212	0.1098	0.1746	0.1576	0.2280	0.2114
(1, 30)	0.0172	0.0134	0.0640	0.0550	0.1160	0.1030	0.1720	0.1560	0.2276	0.2078
(1, 40)	0.0164	0.0138	0.0636	0.0536	0.1200	0.1078	0.1780	0.1608	0.2248	0.2118
(1, 80)	0.0142	0.0124	0.0690	0.0584	0.1244	0.1122	0.1768	0.1620	0.2284	0.2116

A Proofs

In the sequel, for any random variable $X \in \mathbb{R}$ and any $p \in (0, \infty]$, we will use $\|X\|_p$ to denote its L_p norm. Let $\mathbb{N}$ denote the set of all positive integers.

A.1 Proofs in Section 2

A.1.1 Notation

Some additional notations are introduced below. For a random variable $\boldsymbol{X}$, we denote $\widetilde{\boldsymbol{X}}$ as an independent copy of $\boldsymbol{X}$. Then, for $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^2$, we define the following Hoeffding decomposition components,

- $f(\boldsymbol{x}, \boldsymbol{y}) = \text{sign}(x_1 - y_1)\text{sign}(x_2 - y_2)$, $g(\boldsymbol{x}, \boldsymbol{y}) = f(\boldsymbol{x}, \boldsymbol{y}) - \mathbb{E}[f(\boldsymbol{X}, \widetilde{\boldsymbol{X}})]$;
- $g_1(\boldsymbol{x}) = \mathbb{E}[g(\boldsymbol{x}, \boldsymbol{X})]$;
- $g_2(\boldsymbol{x}, \boldsymbol{y}) = f(\boldsymbol{x}, \boldsymbol{y}) - \mathbb{E}[f(\boldsymbol{x}, \boldsymbol{X})] - \mathbb{E}[f(\boldsymbol{X}, \boldsymbol{y})] + \mathbb{E}[f(\boldsymbol{X}, \widetilde{\boldsymbol{X}})]$.

For a random process $\{\boldsymbol{Z}_t\}_{t\in\mathbb{Z}}$ following VAR(1) model (1.1), we define

$$\boldsymbol{U}_i = (\boldsymbol{Z}_i^\top, \boldsymbol{Z}_{i+1}^\top)^\top \in \mathbb{R}^{2d}.$$

Let $U_{i,k}$ denote the k-th element of the random vector $\boldsymbol{U}_i$. Based on the observation $\{\boldsymbol{U}_t\}_{t=1}^{n-1}$, we define the following V-statistics,

$$\widehat{\mathbf{V}} = \frac{1}{(n-1)^2} \sum_{i,j=1}^{n-1} \text{sign}(\boldsymbol{U}_i - \boldsymbol{U}_j)\text{sign}(\boldsymbol{U}_i - \boldsymbol{U}_j)^\top;$$
$$\mathbf{V} = \int\int \text{sign}(\boldsymbol{U}_i - \boldsymbol{U}_j)\text{sign}(\boldsymbol{U}_i - \boldsymbol{U}_j)^\top dF_{\boldsymbol{U}_i} dF_{\boldsymbol{U}_j},$$

where $F_{\boldsymbol{U}_i}$ is the cumulative distribution function of $\boldsymbol{U}_i$. Then, we can represent the (k, m)-th element $\widehat{T}_{km}$ of $\widehat{\mathbf{T}}$ and the (k, m)-th element $\widehat{V}_{km}$ of $\widehat{\mathbf{V}}$ as

$$\widehat{T}_{km} = \frac{2}{(n-1)(n-2)} \sum_{i<j} \text{sign}(U_{i,k} - U_{j,k})\text{sign}(U_{i,m} - U_{j,m}) = \frac{2}{(n-1)(n-2)} \sum_{i<j} f\left(\begin{pmatrix} U_{i,k} \\ U_{i,m} \end{pmatrix}, \begin{pmatrix} U_{j,k} \\ U_{j,m} \end{pmatrix} \right),$$
$$\widehat{V}_{km} = \frac{1}{(n-1)^2} \sum_{i,j=1}^{n-1} \text{sign}(U_{i,k} - U_{j,k})\text{sign}(U_{i,m} - U_{j,m}) = \frac{1}{(n-1)^2} \sum_{i,j=1}^{n-1} f\left(\begin{pmatrix} U_{i,k} \\ U_{i,m} \end{pmatrix}, \begin{pmatrix} U_{j,k} \\ U_{j,m} \end{pmatrix} \right).$$

A.1.2 Proof of Theorem 2.1

The proof technique is based on the proofs of Theorem 4 and Proposition 5 in Shen et al. (2019). Because $\mathbf{T} = \mathbf{V}$, we have the following relationship when $n \geq 3$,

$$\|\widehat{\mathbf{T}} - \mathbf{T}\|_{\max} \leq \frac{n-1}{n-2} \|\widehat{\mathbf{V}} - \mathbf{V}\|_{\max} + \frac{1}{n-2} \|\mathbf{V}\|_{\max} \leq 2\|\widehat{\mathbf{V}} - \mathbf{V}\|_{\max} + \frac{1}{n-2}.$$

This implies that, when $n \geq 3$,

$$\mathbb{P}\left(\|\widehat{\mathbf{T}} - \mathbf{T}\|_{\max} \geq 2z + \frac{1}{n-2} \right) \leq \mathbb{P}\left(\|\widehat{\mathbf{V}} - \mathbf{V}\|_{\max} \geq z \right).$$

Similar to the proofs of Theorem 4 and Proposition 5 in Shen et al. (2019), we can conduct Hoeffding decomposition for the (k,m)-element $\widehat{V}_{km} - V_{km}$ of $\widehat{\mathbf{V}} - \mathbf{V}$ as follows,

$$\widehat{V}_{km} - V_{km} = \frac{2}{n-1}\sum_{i=1}^{n-1} g_1\left(\begin{pmatrix} U_{i,k} \\ U_{i,m}\end{pmatrix}\right) + \frac{1}{(n-1)^2}\sum_{i,j=1}^{n-1} g_2\left(\begin{pmatrix} U_{i,k} \\ U_{i,m}\end{pmatrix}, \begin{pmatrix} U_{j,k} \\ U_{j,m}\end{pmatrix}\right).$$

Then, we have

$$\mathbb{P}(|\widehat{V}_{km}-V_{km}| \geq z) \leq \mathbb{P}\left(\left|\frac{1}{n-1}\sum_{i=1}^{n-1} g_1\left(\begin{pmatrix} U_{i,k} \\ U_{i,m}\end{pmatrix}\right)\right| \geq \frac{z}{4}\right) + \mathbb{P}\left(\left|\frac{1}{(n-1)^2}\sum_{i,j=1}^{n-1} g_2\left(\begin{pmatrix} U_{i,k} \\ U_{i,m}\end{pmatrix}, \begin{pmatrix} U_{j,k} \\ U_{j,m}\end{pmatrix}\right)\right| \geq \frac{z}{2}\right).$$

It implies that

$$\begin{aligned} &\mathbb{P}\left(\|\widehat{\mathbf{V}} - \mathbf{V}\|_{\max} \geq z\right) \leq \sum_{k,m=1}^{2d} \mathbb{P}\left(|\widehat{V}_{km} - V_{km}| \geq z\right) \\ &\leq \sum_{k,m=1}^{2d} \mathbb{P}\left(\left|\frac{1}{n-1}\sum_{i=1}^{n-1} g_1\left(\begin{pmatrix} U_{i,k} \\ U_{i,m}\end{pmatrix}\right)\right| \geq \frac{z}{4}\right) + \sum_{k,m=1}^{2d} \mathbb{P}\left(\left|\frac{1}{(n-1)^2}\sum_{i,j=1}^{n-1} g_2\left(\begin{pmatrix} U_{i,k} \\ U_{i,m}\end{pmatrix}, \begin{pmatrix} U_{j,k} \\ U_{j,m}\end{pmatrix}\right)\right| \geq \frac{z}{2}\right). \quad \text{(A.1)} \end{aligned}$$

The proof consists of the following five steps. Step 1 will analyze the first part of the right-hand side of inequality (A.1) and Steps 2 to 5 will analyze the second part of the right-hand side of inequality (A.1). The exact form of z in inequality (A.1) will be provided later after analyzing each term.

Step 1. Because $\{\boldsymbol{U}_i = (\boldsymbol{Z}_i^\top, \boldsymbol{Z}_{i+1}^\top)^\top\}_{t\in\mathbb{Z}}$ is geometrically α-mixing, and $g_1(\boldsymbol{U}_{i,\{k,m\}})$ has mean zero with $\sup_i \|g_1(B_{i,\{k,m\}})\|_\infty \leq 2$, we have the following relationship using Lemma A.2 in Section A.1.4, that is, for any $n \geq 5$, $d \geq 1$ and $c_1 > 0$,

$$\begin{aligned} &\mathbb{P}\left(\left|\frac{1}{n-1}\sum_{i=1}^{n-1} g_1\left(\begin{pmatrix} U_{i,k} \\ U_{i,m}\end{pmatrix}\right)\right| \geq c_1\sqrt{\frac{\log(ed)}{n-1}}\right) \\ &\leq 2\exp\left(-\frac{C_3(n-1)c_1^2\log(ed)}{4(n-1) + 2c_1(n-1)\sqrt{\frac{\log(ed)}{n-1}}(\log(n-1))(\log(\log(n-1)))}\right) \\ &= 2\exp\left(-\frac{C_3 c_1^2 \log(ed)}{4 + 2c_1\sqrt{\frac{\log(ed)(\log^2(n-1))(\log^2(\log(n-1)))}{n-1}}}\right) \\ &\leq 2\exp\left(-K_1\frac{c_1^2}{1+c_1}\log(ed)\right), \end{aligned}$$

where the constant C_3 only depends on (κ_1, κ_2) and K_1 only depends on $(K_0, \kappa_1, \kappa_2)$. Here, κ_1 and κ_2 are the α-mixing coefficients in Proposition 2.2 and K_0 is the absolute constant defined in Assumption E. The last inequality holds because of Assumption E (i). Therefore, for $n \geq 5$ and $c_1 \geq 1$,

$$\sum_{k,m=1}^{2d} \mathbb{P}\left(\left|\frac{1}{n-1}\sum_{i=1}^{n-1} g_1\left(\begin{pmatrix} U_{i,k} \\ U_{i,m}\end{pmatrix}\right)\right| \geq c_1\sqrt{\frac{\log(ed)}{n-1}}\right) \leq 8d^2\exp\left(-K_1\frac{c_1^2}{1+c_1}\log(ed)\right).$$

Step 2. Because $g_2(\boldsymbol{U}_{i,\{k,m\}}, \boldsymbol{U}_{j,\{k,m\}})$ is degenerate and centered, using the proof in Proposition 1 and Corollary 7(a) in Shen et al. (2019), we have

$$\mathbb{P}\left(\left|\frac{1}{(n-1)^2}\sum_{i,j=1}^{n-1} g_2\left(\begin{pmatrix}U_{i,k}\\U_{i,m}\end{pmatrix}, \begin{pmatrix}U_{j,k}\\U_{j,m}\end{pmatrix}\right)\right| \geq x + C_4 t'\right)$$
$$\leq 2\exp\left(-\frac{C_5(n-1)y}{A_2^{1/2} + y^{1/2}M_2^{1/2}}\right) + (n-1)^2\left(\sum_{\ell=1}^{2} J_\ell\right) D\widetilde{M}_2 + (n-1)\left[\mathbb{P}(|U_{1,k}| \geq \widetilde{M}_1) + \mathbb{P}(|U_{1,m}| \geq \widetilde{M}_1)\right],$$

where C_4 is an absolute positive constant; and C_5 only depends on (κ_1, κ_2); and all other parameters are defined in Shen et al. (2019).

According to the discussion on page 5, Table 1, and the proof of Proposition 5 in Shen et al. (2019), we have

$$D \coloneqq \text{ the uniform upper bound of density functions of } U_{i,k} - U_{j,k},$$
$$A_2^{1/2} = F(t)\left\{\sigma^2 + \frac{(\log(n-1))^4}{n-1}\right\}, \quad M_2^{1/2} = F(t)^{1/2}(\log(n-1))^2, \quad J_\ell = 3,$$
$$\sigma^2 = \frac{64\kappa_1^{\delta/(2+\delta)}}{1-\exp(-\kappa_2\delta/(2+\delta))}, \quad y = x - \frac{F(t)}{n-1}, \quad t' = C_6\left(t + \sum_{i=0}^{1} v_i + F(t)\sum_{i=0}^{1} v_i\right),$$

where C_6 is an absolute constant, $F(t) \asymp \log^2\log(\widetilde{M}_1/\widetilde{M}_2) + \log^2\log(1/t)$ for a sufficiently large value of $\widetilde{M}_1$ and sufficiently small values of $\widetilde{M}_2$ and t, and v_0 and v_1 satisfy

$$v_0^2 \leq 2\left[\mathbb{P}(|U_{1,j}| \geq \widetilde{M}_1) + \mathbb{P}(|U_{1,k}| \geq \widetilde{M}_1)\right] + 12\widetilde{M}_2 D,$$
$$v_1^2 \leq \left[\mathbb{P}(|U_{1,j}| \geq \widetilde{M}_1) + \mathbb{P}(|U_{1,k}| \geq \widetilde{M}_1)\right] + 12\widetilde{M}_2 D.$$

Note that $F(t) \geq 0$ by construction. Then, we have

$$\sum_{k,m=1}^{2d}\mathbb{P}\left(\left|\frac{1}{(n-1)^2}\sum_{i,j=1}^{n-1} g_2\left(\begin{pmatrix}U_{i,k}\\U_{i,m}\end{pmatrix}, \begin{pmatrix}U_{j,k}\\U_{j,m}\end{pmatrix}\right)\right| \geq y + \frac{F(t)}{n-1} + C_4 t'\right)$$
$$\leq 8d^2\exp\left(-\frac{C_5(n-1)y}{A_2^{1/2} + y^{1/2}M_2^{1/2}}\right) + 24(n-1)^2 d^2 D\widetilde{M}_2 + 4(n-1)d^2\left[\mathbb{P}(|U_{1,k}| \geq \widetilde{M}_1) + \mathbb{P}(|U_{1,m}| \geq \widetilde{M}_1)\right]. \quad \text{(A.2)}$$

Let $y = c_2\sqrt{\frac{\log(ed)}{n-1}}$ for some positive absolute constant c_2, and $t = \frac{1}{4\sqrt{K_0}}\sqrt{\frac{\log(ed)}{n-1}}$. Note that $t \leq \frac{1}{2}$ for sufficiently large n, and $t = o(1)$ under Assumption E. Then, the remaining steps will be as follows. Step 3 will bound the second and third parts of the right-hand side of inequality (A.2). Step 4 will bound the first part of the right-hand side of inequality (A.2). Step 5 will upper bound t', for sufficiently large n and sufficiently small t.

Step 3. Step 3 will bound

$$24(n-1)^2 d^2 D\widetilde{M}_2 + 4(n-1)d^2\left[\mathbb{P}(|U_{1,k}| \geq \widetilde{M}_1) + \mathbb{P}(|U_{1,m}| \geq \widetilde{M}_1)\right].$$

Using the Markov inequality and the fact that $U_{i,k} - U_{j,k}$ follows normal distribution, we have

$$\begin{aligned}
&24(n-1)^2d^2D\widetilde{M}_2 + 4(n-1)d^2\left[\mathbb{P}(|U_{1,k}| \geq \widetilde{M}_1) + \mathbb{P}(|U_{1,m}| \geq \widetilde{M}_1)\right] \\
\leq& 24(n-1)^2d^2D\widetilde{M}_2 + 4(n-1)d^2\left[\frac{\mathbb{E}[|U_{1,k}|]}{\widetilde{M}_1} + \frac{\mathbb{E}[|U_{1,m}|]}{\widetilde{M}_1}\right] \\
\leq& 24(n-1)^2d^2D\widetilde{M}_2 + 4(n-1)d^2\left[\frac{\mathbb{E}[|U_{1,k}|^2]^{1/2}}{\widetilde{M}_1} + \frac{\mathbb{E}[|U_{1,m}|^2]^{1/2}}{\widetilde{M}_1}\right] \\
\leq& 24((n-1)d)^2\frac{1}{\sqrt{2\pi\underline{\sigma}^2}}\widetilde{M}_2 + 8((n-1)d)^2\frac{\sqrt{\overline{\sigma}^2}}{\widetilde{M}_1},
\end{aligned}$$

where $\underline{\sigma}^2$ is the uniform lower bound of the variance of $U_{i,k}-U_{j,k}$, and $\overline{\sigma}^2$ is the uniform upper bound of the variance of $U_{i,k}$. Because Lemma A.3 implies $0 < c_{\mathbf{\Omega}} \leq \lambda_{\min}(\mathbf{\Omega}) \leq \lambda_{\max}(\mathbf{\Omega}) \leq C_{\mathbf{\Omega}} < \infty$, where $c_{\mathbf{\Omega}}$ and $C_{\mathbf{\Omega}}$ only depend on $(C_{\mathbf{A}}, c_{\boldsymbol{E}}, C_{\boldsymbol{E}})$, $\underline{\sigma}^2$ and $\overline{\sigma}^2$ are lower and upper bounded by some positive constants which only depend on $(C_{\mathbf{A}}, c_{\boldsymbol{E}}, C_{\boldsymbol{E}})$. Therefore,

$$24(n-1)^2d^2D\widetilde{M}_2 + 4(n-1)d^2\left[\mathbb{P}(|U_{1,k}| \geq \widetilde{M}_1) + \mathbb{P}(|U_{1,m}| \geq \widetilde{M}_1)\right] \leq \frac{1}{2}K_2((n-1)d)^2\left[\widetilde{M}_2 + \frac{1}{\widetilde{M}_1}\right],$$

where K_2 only depends on $(C_{\mathbf{A}}, c_{\boldsymbol{E}}, C_{\boldsymbol{E}})$. Then, by choosing $\widetilde{M}_2 = \frac{1}{\widetilde{M}_1} = ((n-1)d)^{-4}$, we have

$$24(n-1)^2d^2D\widetilde{M}_2 + 4(n-1)d^2\left[\mathbb{P}(|U_{1,k}| \geq \widetilde{M}_1) + \mathbb{P}(|U_{1,m}| \geq \widetilde{M}_1)\right] \leq K_2((n-1)d)^{-2}.$$

Step 4. Step 4 will bound the first part of the right-hand side of inequality (A.2) when $y = c_2\sqrt{\frac{\log(ed)}{n-1}}$ for some positive constant c_2, $t = \frac{1}{4\sqrt{K_0}}\sqrt{\frac{\log(ed)}{n-1}}$, and n is sufficiently large so that t is sufficiently small under Assumption E.

Step 4-1. We will show that $\frac{F(t)}{n-1} = o(t)$ when $t = m_1\sqrt{\frac{\log(ed)}{n-1}}$ and $\widetilde{M}_2 = \frac{1}{\widetilde{M}_1} = ((n-1)d)^{-m_2}$ for some positive numbers m_1 and m_2. Note that we set $m_1 = \frac{1}{4\sqrt{K_0}}$ and $m_2 = 4$ in Step 3. For a sufficiently large n and sufficiently small t, we have

$$\begin{aligned}
\frac{F(t)}{t(n-1)} &\asymp \frac{\log^2(2m_2\log((n-1)d)) + \log^2\left(\log\left(m_1\sqrt{\frac{n-1}{\log(ed)}}\right)\right)}{m_1\sqrt{(n-1)\log(ed)}} \\
&= \frac{\log^2(2m_2\log((n-1)d))}{m_1\sqrt{(n-1)\log(ed)}} + \frac{\log^2\left(\log\left(m_1\sqrt{\frac{n-1}{\log(ed)}}\right)\right)}{m_1\sqrt{(n-1)\log(ed)}} \\
&= \frac{\log^2(2m_2\log((n-1)d))}{2m_2\log((n-1)d)}\frac{2m_2\log((n-1)d)}{m_1\sqrt{(n-1)\log(ed)}} + \frac{\log^2\left(\log\left(m_1\sqrt{\frac{n-1}{\log(ed)}}\right)\right)}{m_1\sqrt{\frac{n-1}{\log(ed)}}}\frac{1}{\log(ed)} \\
&= o(1)o(1) + o(1)O(1) = o(1).
\end{aligned}$$

Therefore, we have $\frac{F(t)}{n-1} = o(t)$, and for sufficiently large n and sufficiently small $t = \frac{1}{4\sqrt{K_0}}\sqrt{\frac{\log(ed)}{n-1}}$,

$$\frac{F(t)}{t(n-1)} \asymp \frac{\log^2(2m_2\log((n-1)d))}{m_1\sqrt{(n-1)\log(ed)}} + \frac{\log^2\left(\log\left(m_1\sqrt{\frac{n-1}{\log(ed)}}\right)\right)}{m_1\sqrt{(n-1)\log(ed)}} \leq K_3\frac{\log((n-1)d)}{\sqrt{(n-1)\log(ed)}} + K_4\frac{1}{\log(ed)},$$

where K_3 only depends on K_0, and K_4 is an absolute positive constant.

Step 4-2. In this step, we will bound the first part of the right-hand side of inequality (A.2) with $y = c_2\sqrt{\frac{\log(ed)}{n-1}}$ for some positive constant c_2; and $t = \frac{1}{4\sqrt{K_0}}\sqrt{\frac{\log(ed)}{n-1}}$ when n is sufficiently large and $\frac{\log ed}{n-1}$ is sufficiently small, which is true under Assumption E. For the first part, we have

$$\exp\left(-\frac{C_5(n-1)y}{A_2^{1/2}+y^{1/2}M_2^{1/2}}\right) = \exp\left(-\frac{C_5c_2\sqrt{n-1}\sqrt{\log(ed)}}{A_2^{1/2}+c_2^{1/2}\left(\frac{\log(ed)}{n-1}\right)^{1/4}M_2^{1/2}}\right) = \exp\left(-\frac{C_5c_2\log(ed)}{\left(\frac{\log(ed)}{n-1}\right)^{1/2}A_2^{1/2}+c_2^{1/2}\left(\frac{\log(ed)}{n-1}\right)^{3/4}M_2^{1/2}}\right).$$

Then, Steps 4-2-1 and 4-2-2 combined will show $\left(\frac{\log(ed)}{n-1}\right)^{1/2}A_2^{1/2} = O(1)$ and $\left(\frac{\log(ed)}{n-1}\right)^{3/4}M_2^{1/2} = O(1)$.

Step 4-2-1. This step shows that $\left(\frac{\log(ed)}{n-1}\right)^{1/2}A_2^{1/2} = O(1)$.

Under Assumption E, for sufficiently large n and sufficiently small $t \asymp \sqrt{\frac{\log(ed)}{n-1}}$, we have

$$\begin{aligned}(\log(ed)/n)^{1/2}A_2^{1/2} \asymp tF(t) &\asymp t\log^2(\log(\widetilde{M}_1)/\log(\widetilde{M}_2)) + t\log^2(\log(1/t))\\ &= t\log^2(2m_2\log((n-1)d)) + t\log^2(\log(1/t))\\ &= m_1\frac{\sqrt{\log ed}\log^2(2m_2\log((n-1)d))}{\sqrt{n-1}} + \frac{\log^2(\log(1/t))}{1/t}\\ &\le K_{5A},\end{aligned}$$

where $m_1 := \frac{1}{4\sqrt{K_0}}$, $m_2 := 4$, and K_{5A} only depends on K_0. Therefore, $\left(\frac{\log(ed)}{n-1}\right)^{1/2}A_2^{1/2} = O(1)$.

Step 4-2-2. This step shows that $\left(\frac{\log(ed)}{n-1}\right)^{3/4}M_2^{1/2} = O(1)$. Under Assumption E, for sufficiently large n and sufficiently small $t \asymp \sqrt{\frac{\log(ed)}{n-1}}$, we have

$$\left(\left(\frac{\log(ed)}{n-1}\right)^{3/4}M_2^{1/2}\right)^2 \asymp t^3F(t)\log^4(n) = tF(t)(t^2\log^4(n)) \le K_{5A}\widetilde{K}_0,$$

where $\widetilde{K}_0$ only depends on K_0. Therefore, for sufficiently large n and sufficiently small t, we have $\left(\frac{\log(ed)}{n-1}\right)^{3/4}M_2^{1/2} \le K_{5M}$, where $K_{5M} = \sqrt{K_{5A}\widetilde{K}_0}$ only depends on K_0.

According to Steps 4-2-1 and 4-2-2, we have

$$\left(\frac{\log(ed)}{n-1}\right)^{1/2}A_2^{1/2} \le K_{5A} \text{ and } ((\log d/n)^{3/4}M_2^{1/2}) \le K_{5M}$$

for sufficiently large n and sufficiently small $t \asymp \sqrt{\frac{\log(ed)}{n-1}}$. Therefore, Steps 4-2-1 and 4-2-2 combined show that, for sufficiently large n, sufficiently small $t \asymp \sqrt{\frac{\log(ed)}{n-1}}$, and $y = c_2\sqrt{\frac{\log(ed)}{n-1}}$ where $c_2 > 0$, we have

$$8d^2\exp\left(-\frac{C_5(n-1)y}{A_2^{1/2}+y^{1/2}M_2^{1/2}}\right) \le 8d^2\exp\left(-\frac{K_5c_2\log(ed)}{1+c_2^{1/2}}\right),$$

where K_5 only depends on $(K_0, \kappa_1, \kappa_2)$.

Step 5. Step 5 will upper bound

$$t' = C_6 \left(t + \sum_{i=0}^{1} v_i + F(t) \sum_{i=0}^{1} v_i \right).$$

Following the proof of Proposition 5 in Shen et al. (2019), we have

$$\begin{aligned} v_0^2 &\le 2\left[\mathbb{P}(|U_{1,j}| \ge \widetilde{M}_1) + \mathbb{P}(|U_{1,k}| \ge \widetilde{M}_1)\right] + 12\widetilde{M}_2 D, \\ v_1^2 &\le \left[\mathbb{P}(|U_{1,j}| \ge \widetilde{M}_1) + \mathbb{P}(|U_{1,k}| \ge \widetilde{M}_1)\right] + 12\widetilde{M}_2 D. \end{aligned}$$

Step 3 has shown that

$$\begin{aligned} \{(n-1)d\}^2 (v_0 + v_1)^2 &\le 2\{(n-1)d\}^2 (v_1^2 + v_2^2) \\ &\le 8\{(n-1)d\}^2 \left[\left[\mathbb{P}(|U_{1,j}| \ge \widetilde{M}_1) + \mathbb{P}(|U_{1,k}| \ge \widetilde{M}_1)\right] + 6\widetilde{M}_2 D\right] \\ &\le 4K_2 \{(n-1)d\}^{-2}. \end{aligned}$$

Therefore,

$$\sum_{i=0}^{1} v_i = \{(n-1)d\}^{-1} \{(n-1)d\} \sum_{i=0}^{1} v_i \le 2\sqrt{K_2}\{(n-1)d\}^{-2}.$$

Also, for sufficiently large n and the sufficiently small $t = \frac{1}{4\sqrt{K_0}}\sqrt{\frac{\log(ed)}{n-1}}$,

$$\begin{aligned} F(t) \sum_{i=0}^{1} v_i &= \frac{1}{d} \frac{F(t)}{t(n-1)} t\{(n-1)d\} \sum_{i=0}^{1} v_i \\ &\le \frac{1}{d}\left[K_3 \frac{\log((n-1)d)}{\sqrt{(n-1)\log(ed)}} + K_4 \frac{1}{\log(ed)}\right] 2\sqrt{K_2}\{(n-1)d\}^{-1} t \\ &= \frac{\sqrt{K_2}K_3}{2\sqrt{K_0}} \frac{\log((n-1)d)}{(n-1)^2 d^2} + \frac{\sqrt{K_2}K_4}{2\sqrt{K_0}} \frac{1}{d^2(n-1)\sqrt{(n-1)\log(ed)}}. \end{aligned}$$

Therefore, for sufficiently large n and sufficiently small $t = \frac{1}{4\sqrt{K_0}}\sqrt{\frac{\log(ed)}{n-1}}$,

$$\begin{aligned} t' = C_6 \left(t + \sum_{i=0}^{1} v_i + F(t) \sum_{i=0}^{1} v_i \right) &\le C_6 \frac{1}{4\sqrt{K_0}} \sqrt{\frac{\log(ed)}{n-1}} + 2C_6\sqrt{K_2}((n-1)d)^{-2} \\ &\quad + \frac{C_6\sqrt{K_2}K_3}{2\sqrt{K_0}} \frac{\log((n-1)d)}{(n-1)^2 d^2} + \frac{C_6\sqrt{K_2}K_4}{2\sqrt{K_0}} \frac{1}{d^2(n-1)\sqrt{(n-1)\log(ed)}}. \end{aligned}$$

From Steps 3 to 5, when $y = c_2\sqrt{\frac{\log(ed)}{n-1}}$ for some positive constant c_2 and $t = \frac{1}{4\sqrt{K_0}}\sqrt{\frac{\log(ed)}{n-1}}$, for sufficiently large n and sufficiently small t,

$$\sum_{k,m=1}^{2d} \mathbb{P}\left(\left| \frac{1}{(n-1)^2} \sum_{i,j=1}^{n-1} g_2 \left(\begin{pmatrix} U_{i,k} \\ U_{i,m} \end{pmatrix}, \begin{pmatrix} U_{j,k} \\ U_{j,m} \end{pmatrix} \right) \right| \ge c_2 \sqrt{\frac{\log(ed)}{n-1}} + \frac{K_3}{4\sqrt{K_0}} \frac{\log((n-1)d)}{(n-1)}\right.$$

$$+\frac{K_4}{4\sqrt{K_0}}\frac{1}{\sqrt{(n-1)\log(ed)}}+\frac{C_4C_6}{4\sqrt{K_0}}\sqrt{\frac{\log(ed)}{n-1}}+2C_4C_6\sqrt{K_2}((n-1)d)^{-2}$$
$$\left.+\frac{C_4C_6\sqrt{K_2}K_3}{2\sqrt{K_0}}\frac{\log((n-1)d)}{(n-1)^2d^2}+\frac{C_4C_6\sqrt{K_2}K_4}{2\sqrt{K_0}}\frac{1}{d^2(n-1)\sqrt{(n-1)\log(ed)}}\right)$$
$$\leq \sum_{k,m=1}^{2d}\mathbb{P}\left(\left|\frac{1}{(n-1)^2}\sum_{i,j=1}^{n-1}g_2\left(\begin{pmatrix}U_{i,k}\\U_{i,m}\end{pmatrix},\begin{pmatrix}U_{j,k}\\U_{j,m}\end{pmatrix}\right)\right|\geq c_2\sqrt{\frac{\log(ed)}{n-1}}+\frac{F(t)}{n-1}+C_1t'\right)$$
$$\leq 8d^2\exp\left(-K_5\frac{c^2}{1+c_2^{1/2}}\log(ed)\right)+K_2\{(n-1)d\}^{-2}.$$

For some positive constant Q to be specified later, let's set

$$z=Q\sqrt{\frac{\log(ed)}{n-1}}+\frac{K_3}{4\sqrt{K_0}}\frac{\log((n-1)d)}{n-1}+\frac{K_4}{4\sqrt{K_0}}\frac{1}{\sqrt{(n-1)\log(ed)}}+\frac{C_4C_6}{4\sqrt{K_0}}\sqrt{\frac{\log(ed)}{n-1}}$$
$$+2C_4C_6\sqrt{K_2}((n-1)d)^{-2}+\frac{C_4C_6\sqrt{K_2}K_3}{2\sqrt{K_0}}\frac{\log((n-1)d)}{(n-1)^2d^2}+\frac{C_4C_6\sqrt{K_2}K_4}{2\sqrt{K_0}}\frac{1}{d^2(n-1)\sqrt{(n-1)\log(ed)}}.$$

Then, for sufficiently large n and sufficiently small $\frac{\log(ed)}{n-1}$,

$$\mathbb{P}\left(\|\widehat{\mathbf{V}}-\mathbf{V}\|_{\max}\geq z\right)\leq\sum_{k,m=1}^{2d}\mathbb{P}\left(\|\widehat{V}_{km}-V_{km}\|\geq z\right)$$
$$\leq\sum_{k,m=1}^{2d}\mathbb{P}\left(\left|\frac{1}{n-1}\sum_{i=1}^{n-1}g_1\left(\begin{pmatrix}U_{i,k}\\U_{i,m}\end{pmatrix}\right)\right|\geq\frac{z}{4}\right)+\sum_{k,m=1}^{2d}\mathbb{P}\left(\left|\frac{1}{(n-1)^2}\sum_{i,j=1}^{n-1}g_2\left(\begin{pmatrix}U_{i,k}\\U_{i,m}\end{pmatrix},\begin{pmatrix}U_{j,k}\\U_{j,m}\end{pmatrix}\right)\right|\geq\frac{z}{2}\right)$$
$$\leq\sum_{k,m=1}^{2d}\mathbb{P}\left(\left|\frac{1}{n-1}\sum_{i=1}^{n-1}g_1\left(\begin{pmatrix}U_{i,k}\\U_{i,m}\end{pmatrix}\right)\right|\geq\frac{Q}{4}\sqrt{\frac{\log(ed)}{n-1}}\right)+\sum_{k,m=1}^{2d}\mathbb{P}\left(\left|\frac{1}{(n-1)^2}\sum_{i,j=1}^{n-1}g_2\left(\begin{pmatrix}U_{i,k}\\U_{i,m}\end{pmatrix},\begin{pmatrix}U_{j,k}\\U_{j,m}\end{pmatrix}\right)\right|\geq\frac{z}{2}\right)$$
$$\leq 8d^2\exp\left(-K_1\frac{(Q/4)^2}{1+(Q/4)}\log(ed)\right)+8d^2\exp\left(-K_5\frac{(Q/2)}{1+(Q/2)^{1/2}}\log(ed)\right)+K_2\{(n-1)d\}^{-2}.$$

Therefore, by choosing sufficiently large Q, we have $\|\widehat{\mathbf{V}}-\mathbf{V}\|_{\max}=O_{\mathbb{P}}(\sqrt{\log(ed)/n})$. By picking the value of Q large enough so that

$$K_1\frac{(Q/4)^2}{1+(Q/4)}\geq 4\quad\text{and}\quad K_5\frac{(Q/2)}{1+(Q/2)^{1/2}}\geq 4,$$

we have

$$\mathbb{P}\left(\|\widehat{\mathbf{V}}-\mathbf{V}\|_{\max}\geq z\right)\leq 16e^{-4}d^{-2}+K_2\{(n-1)d\}^{-2}.$$

Lastly, noticing the following relationship

$$\|\widehat{\mathbf{T}}-\mathbf{T}\|_{\max}\leq\frac{n-1}{n-2}\|\widehat{\mathbf{V}}-\mathbf{V}\|_{\max}+\frac{1}{n-2}\|\mathbf{V}\|_{\max}\leq\frac{n-1}{n-2}\|\widehat{\mathbf{V}}-\mathbf{V}\|_{\max}+\frac{1}{n-2},$$

the proof is finished.

A.1.3 Proof of Theorem 2.3

The proof of Theorem 2.3 immediately follows from the lemma below, which is in the proof of Theorem 5.3 in Qiu et al. (2015).

Lemma A.1 (Master theorem)**.** Suppose $A \in \mathcal{M}(s, M)$. Further suppose $\|\widehat{\mathbf{\Sigma}}_0 - \mathbf{\Sigma}_0\|_{\max} \le \epsilon_1$ and $\|\widehat{\mathbf{\Sigma}}_1 - \mathbf{\Sigma}_1\|_{\max} \le \epsilon_2$ with probability no less than $1 - \epsilon_0$. Let λ in equation (2.3) be $\lambda \ge \epsilon_1 M + \epsilon_2$. Then with probability no less than $1 - \epsilon_0$, it holds that

$$\|\widehat{\mathbf{A}} - \mathbf{A}\|_{\max} \le 2\|\mathbf{\Sigma}_0^{-1}\|_\infty[\epsilon_1 M + \epsilon_2] = 2\|\mathbf{\Sigma}^{-1}\|_\infty \lambda \quad \text{and}$$
$$\|\widehat{\mathbf{A}} - \mathbf{A}\|_\infty \le 4s\|\mathbf{\Sigma}_0^{-1}\|_\infty[\epsilon_1 M + \epsilon_2] = 4s\|\mathbf{\Sigma}^{-1}\|_\infty \lambda.$$

A.1.4 Technical lemmas used in Section 2

Lemma A.2 (Theorem 1 in Merlevède et al. (2009))**.** Let $\{X_t\}_{t\in\mathbb{Z}}$ be a sequence of centered real-valued random variables. Suppose that the sequence satisfies that

$$\alpha(m) \le \kappa_1 \exp(-\kappa_2 m) \ \text{ for all } m \ge 1,$$

and that there exists a positive M such that $\sup_{i\ge 1}\|X_i\|_\infty \le M$. Then there exist positive constants C_1 and C_2 depending only on (κ_1, κ_2) such that for all $n \ge 4$ and t satisfying that for any $0 < t < \{C_1 M(\log n)(\log\log n)\}^{-1}$, we have

$$\log \mathbb{E}\left(\exp\left(t\sum_{i=1}^n X_i\right)\right) \le \frac{C_2 t^2 n M^2}{1 - C_1 t M(\log n)(\log\log n)}.$$

In terms of probability, there exists a constant C_3 only depending on (κ_1, κ_2) such that for all $n \ge 4$ and $x \ge 0$,

$$\mathbb{P}\left(\left|\sum_{i=1}^n X_i\right| \ge x\right) \le \exp\left(-\frac{C_3 x^2}{nM^2 + Mx(\log n)(\log\log n)}\right). \tag{A.3}$$

Lemma A.3. Suppose that the latent process $\{\boldsymbol{Z}_t\}_{t\in\mathbb{Z}}$ follows VAR(1) model (1.1) with Assumption M(ii) and (iii). Let $\mathbf{\Omega}_j := \text{Var}\left((\boldsymbol{Z}_t^\top, \dots, \boldsymbol{Z}_{t+j}^\top)^\top\right)$ for any $j \in \{0, 1, 2, \cdots\}$. Then, $\mathbf{\Omega}_j$ is positive definite and $0 < c_{\mathbf{\Omega}} < \lambda_{\min}(\mathbf{\Omega}_j) \le \lambda_{\max}(\mathbf{\Omega}_j) < C_{\mathbf{\Omega}} < \infty$ for some constants $c_{\mathbf{\Omega}}$ and $C_{\mathbf{\Omega}}$ that only depend on $(C_{\mathbf{A}}, c_{\boldsymbol{E}}, C_{\boldsymbol{E}})$.

Proof. Let $\boldsymbol{q}_{j,\min}$ be the eigenvector corresponding to $\lambda_{\min}(\mathbf{\Omega}_j)$. We will prove the lemma in two steps: first for the case that $j = 0$ and then the case that $j > 0$.

Step 1. When $j = 0$, $\mathbf{\Omega}_0 = \mathbf{\Sigma}_0 = \sum_{k=0}^\infty \mathbf{A}^k \mathbf{\Sigma}_{\boldsymbol{E}}(\mathbf{A}^\top)^k$. Then,

$$\lambda_{\min}(\mathbf{\Omega}_j) = \sum_{k=0}^\infty \boldsymbol{q}_{j,\min}^\top \mathbf{A}^k \mathbf{\Sigma}_{\boldsymbol{E}}(\mathbf{A}^\top)^k \boldsymbol{q}_{j,\min} \ge \boldsymbol{q}_{j,\min}^\top \mathbf{\Sigma}_{\boldsymbol{E}} \boldsymbol{q}_{j,\min} \ge \lambda_{\min}(\mathbf{\Sigma}_{\boldsymbol{E}}), \text{ and}$$
$$\lambda_{\max}(\mathbf{\Omega}_j) \le \sum_{k=0}^\infty \lambda_{\max}(\mathbf{\Sigma}_{\boldsymbol{E}})\|(\mathbf{A}^\top)^k\|_2^2 \le \frac{\lambda_{\max}(\mathbf{\Sigma}_E)}{1 - \|\mathbf{A}\|_2^2} \le \frac{\lambda_{\max}(\mathbf{\Sigma}_E)}{1 - C_{\mathbf{A}}^2}.$$

Step 2. We investigate the case when $j > 0$. We can represent $\mathbf{\Omega}_j$ as follows (cf. Section 1.4 in Gómez (2016)).

$$\boldsymbol{\Omega}_j = \mathrm{Var}\left(\underbrace{\begin{pmatrix} \mathbf{I}_d & \mathbf{0}_{d\times d} & \cdots & \cdots & \mathbf{0}_{d\times d} \\ \mathbf{A} & \mathbf{I}_d & \mathbf{0}_{d\times d} & \mathbf{0}_{d\times d} & \cdots \\ \mathbf{A}^2 & \mathbf{A} & \mathbf{I}_d & \mathbf{0}_{d\times d} & \cdots \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ \mathbf{A}^j & \mathbf{A}^{j-1} & \ddots & \mathbf{A} & \mathbf{I}_d \end{pmatrix}}_{\mathbf{L}} \begin{pmatrix} \boldsymbol{Z}_t \\ \boldsymbol{E}_{t+1} \\ \boldsymbol{E}_{t+2} \\ \vdots \\ \boldsymbol{E}_{t+j} \end{pmatrix}\right) = \mathbf{L} \begin{pmatrix} \boldsymbol{\Omega}_0 & \mathbf{0}_{d\times d} & \cdots & \cdots & \mathbf{0}_{d\times d} \\ \mathbf{0}_{d\times d} & \boldsymbol{\Sigma}_{\boldsymbol{E}} & \mathbf{0}_{d\times d} & \cdots & \cdots \\ \mathbf{0}_{d\times d} & \mathbf{0}_{d\times d} & \boldsymbol{\Sigma}_{\boldsymbol{E}} & \mathbf{0}_{d\times d} & \cdots \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ \mathbf{0}_{d\times d} & \vdots & \vdots & \mathbf{0}_{d\times d} & \boldsymbol{\Sigma}_{\boldsymbol{E}} \end{pmatrix} \mathbf{L}^\top,$$

with

$$\mathbf{L} = \begin{pmatrix} \mathbf{I}_d & \mathbf{0}_{d\times d} & \cdots & \mathbf{0}_{d\times d} \\ \mathbf{0}_{d\times d} & \mathbf{I}_d & \mathbf{0}_{d\times d} & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0}_{d\times d} \\ \mathbf{0}_{d\times d} & \ddots & \mathbf{0}_{d\times d} & \mathbf{I}_d \end{pmatrix} + \begin{pmatrix} \mathbf{0}_{d\times d} & \mathbf{0}_{d\times d} & \cdots & \mathbf{0}_{d\times d} \\ \mathbf{A} & \mathbf{0}_{d\times d} & \ddots & \vdots \\ \mathbf{0}_{d\times d} & \ddots & \ddots & \vdots \\ \mathbf{0}_{d\times d} & \ddots & \mathbf{A} & \mathbf{0}_{d\times d} \end{pmatrix} + \cdots + \begin{pmatrix} \mathbf{0}_{d\times d} & \mathbf{0}_{d\times d} & \cdots & \mathbf{0}_{d\times d} \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0}_{d\times d} & \mathbf{0}_{d\times d} & \ddots & \mathbf{0}_{d\times d} \\ \mathbf{A}^j & \mathbf{0}_{d\times d} & \ddots & \mathbf{0}_{d\times d} \end{pmatrix} \text{ and}$$

$$\mathbf{L}^{-1} = \begin{pmatrix} \mathbf{I}_d & \mathbf{0}_{d\times d} & \cdots & \mathbf{0}_{d\times d} \\ -\mathbf{A} & \mathbf{I}_d & \mathbf{0}_{d\times d} & \ddots \\ \mathbf{0}_{d\times d} & \ddots & \ddots & \ddots \\ \vdots & \mathbf{0}_{d\times d} & -\mathbf{A} & \mathbf{I}_d \end{pmatrix}.$$

We have

$$\|\mathbf{L}\|_2 = \sum_{k=0}^{j} \|\mathbf{A}^k\|_2 \le \frac{1}{1-\|\mathbf{A}\|_2} \text{ and } \|\mathbf{L}^{-1}\|_2 \le 1 + \|\mathbf{A}\|_2.$$

Therefore,

$$\lambda_{\max}(\boldsymbol{\Omega}_j) \le \lambda_{\max}(\boldsymbol{\Omega}_0)\|\mathbf{L}^\top\|_2^2 \le \frac{\lambda_{\max}(\boldsymbol{\Omega}_0)}{(1-\|\mathbf{A}\|_2)^2}, \text{ and}$$
$$\frac{1}{\lambda_{\min}(\boldsymbol{\Omega}_j)} = \lambda_{\max}(\boldsymbol{\Omega}_j^{-1}) \le \lambda_{\max}(\boldsymbol{\Sigma}_{\boldsymbol{E}}^{-1})\|\mathbf{L}^{-1}\|_2^2 \le \frac{\lambda_{\max}(\boldsymbol{\Sigma}_{\boldsymbol{E}}^{-1})}{(1+\|\mathbf{A}\|_2)^2}.$$

Combining Steps 1 and 2 concludes. □

A.2 Proofs in Section 3

A.2.1 Notation

In addition to the notations introduced in the main text and Section A.1, more are defined as follows. For a generic vector $\boldsymbol{v} \in \mathbb{R}^d$, let $S(\boldsymbol{v}) := \boldsymbol{w}^\top \widehat{\boldsymbol{h}}(\boldsymbol{v}) = \boldsymbol{w}^\top(\widehat{\boldsymbol{\Sigma}}_0\boldsymbol{v} - \widehat{\boldsymbol{\Sigma}}_{1,*m})$. We denote $\boldsymbol{\beta}(c) := (c, \boldsymbol{\gamma}^\top)^\top$ when we use c in the position of θ. Similarly, we denote $\widehat{\boldsymbol{\beta}}(c) := (c, \widehat{\boldsymbol{\gamma}}^\top)^\top$.

For Kendall's tau $\mathbf{T}$, estimator $\widehat{\mathbf{T}}$, and bootstrap version $\widehat{\mathbf{T}}^*$, which are defined in subsections 2.2 and 3.2, we denote

- $\mathbf{T}_0 = \mathbf{T}_{[d],[d]}$, $\mathbf{T}_1 = \mathbf{T}_{[d],d+[d]}$, $\widehat{\mathbf{T}}_0 = \widehat{\mathbf{T}}_{[d],[d]}$, $\widehat{\mathbf{T}}_1 = \widehat{\mathbf{T}}_{[d],d+[d]}$, $\widehat{\mathbf{T}}_0^* = \widehat{\mathbf{T}}^*_{[d],[d]}$, and $\widehat{\mathbf{T}}_1^* = \widehat{\mathbf{T}}^*_{[d],d+[d]}$;

- $\tau_{0,jk}$, $\widehat{\tau}_{0,jk}$, and $\widehat{\tau}^*_{0,jk}$ are the (j,k)-th element of $\mathbf{T}_0$, $\widehat{\mathbf{T}}_0$, and $\widehat{\mathbf{T}}^*_0$, respectively;
- $\tau_{1,jk}$, $\widehat{\tau}_{1,jk}$, and $\widehat{\tau}^*_{1,jk}$ are the (j,k)-th element of $\mathbf{T}_1$, $\widehat{\mathbf{T}}_1$, and $\widehat{\mathbf{T}}_1$, respectively.

Then, $\mathbf{\Sigma}_0 = \sin\left(\frac{\pi}{2}\mathbf{T}_0\right)$, $\mathbf{\Sigma}_1 = \sin\left(\frac{\pi}{2}\mathbf{T}_1\right)$, $\widehat{\mathbf{\Sigma}}_0 = \sin\left(\frac{\pi}{2}\widehat{\mathbf{T}}_0\right)$, $\widehat{\mathbf{\Sigma}}_1 = \sin\left(\frac{\pi}{2}\widehat{\mathbf{T}}_1\right)$, $\widehat{\mathbf{\Sigma}}^*_0 = \sin\left(\frac{\pi}{2}\widehat{\mathbf{T}}^*_0\right)$, and $\widehat{\mathbf{\Sigma}}^*_1 = \sin\left(\frac{\pi}{2}\widehat{\mathbf{T}}^*_1\right)$. We define bootstrap version V-statistic $\widehat{\mathbf{V}}^*$ as

$$\widehat{\mathbf{V}}^* := \frac{1}{(b\ell)^2}\sum_{i,j=1}^{b\ell}\text{sign}(\boldsymbol{Y}^*_i - \boldsymbol{Y}^*_j)\text{sign}(\boldsymbol{Y}^*_i - \boldsymbol{Y}^*_j)^\top = \frac{1}{(b\ell)^2}\sum_{i,j=1}^{b\ell}\text{sign}(\boldsymbol{U}^*_i - \boldsymbol{U}^*_j)\text{sign}(\boldsymbol{U}^*_i - \boldsymbol{U}^*_j)^\top,$$

where $\boldsymbol{Y}^*_t$ is the bootstrap sample from $\boldsymbol{Y}_t = (\boldsymbol{X}_t^\top, \boldsymbol{X}_{t+1}^\top)^\top$ and $\boldsymbol{U}^*_t$ is the latent variable corresponding to $\boldsymbol{Y}^*_t$ via $\boldsymbol{Y}^*_t = \boldsymbol{f}(\boldsymbol{U}^*_t)$.

For $\mathbf{V}$, its estimator $\widehat{\mathbf{V}}$, and bootstrap estimator $\widehat{\mathbf{V}}^*$, we denote

- $\mathbf{V}_0 = \mathbf{V}_{[d],[d]}$, $\mathbf{V}_1 = \mathbf{V}_{[d],d+[d]}$, $\widehat{\mathbf{V}}_0 = \widehat{\mathbf{V}}_{[d],[d]}$, $\widehat{\mathbf{V}}_1 = \widehat{\mathbf{V}}_{[d],d+[d]}$, $\widehat{\mathbf{V}}^*_0 = \widehat{\mathbf{V}}^*_{[d],[d]}$, and $\widehat{\mathbf{V}}^*_1 = \widehat{\mathbf{V}}^*_{[d],d+[d]}$;
- $V_{0,jk}$, $\widehat{V}_{0,jk}$, and $\widehat{V}^*_{0,jk}$ are the (j,k)-th element of $\mathbf{V}_0$, $\widehat{\mathbf{V}}_0$, and $\widehat{\mathbf{V}}^*_0$, respectively;
- $V_{1,jk}$, $\widehat{V}_{1,jk}$, and $\widehat{V}^*_{1,jk}$ are the (j,k)-th element of $\mathbf{V}_1$, $\widehat{\mathbf{V}}_1$, and $\widehat{\mathbf{V}}^*_1$, respectively.

Then, we have the following relationships:

$$\widehat{V}_{0,jk} - V_{0,jk} = \frac{1}{(n-1)^2}\sum_{t,t'=1}^{n-1} g\left(\begin{pmatrix}X_{t,j}\\X_{t,k}\end{pmatrix}, \begin{pmatrix}X_{t',j}\\X_{t',k}\end{pmatrix}\right) = \frac{1}{(n-1)^2}\sum_{t,t'=1}^{n-1} g\left(\begin{pmatrix}Z_{t,j}\\Z_{t,k}\end{pmatrix}, \begin{pmatrix}Z_{t',j}\\Z_{t',k}\end{pmatrix}\right),$$

$$\widehat{V}_{1,jk} - V_{1,jk} = \frac{1}{(n-1)^2}\sum_{t,t'=1}^{n-1} g\left(\begin{pmatrix}X_{t,j}\\X_{t+1,k}\end{pmatrix}, \begin{pmatrix}X_{t',j}\\X_{t'+1,k}\end{pmatrix}\right) = \frac{1}{(n-1)^2}\sum_{t,t'=1}^{n-1} g\left(\begin{pmatrix}Z_{t,j}\\Z_{t+1,k}\end{pmatrix}, \begin{pmatrix}Z_{t',j}\\Z_{t'+1,k}\end{pmatrix}\right),$$

$$V^*_{0,jk} - V_{0,jk} = \frac{1}{(b\ell)^2}\sum_{1\le t,t'\le b\ell}^{b\ell} g\left(\begin{pmatrix}X^*_{t,j}\\X^*_{t,k}\end{pmatrix}, \begin{pmatrix}X^*_{t',j}\\X^*_{t',k}\end{pmatrix}\right) = \frac{1}{(b\ell)^2}\sum_{1\le t,t'\le b\ell} g\left(\begin{pmatrix}Z^*_{t,j}\\Z^*_{t,k}\end{pmatrix}, \begin{pmatrix}Z^*_{t',j}\\Z^*_{t',k}\end{pmatrix}\right), \text{ and}$$

$$V^*_{1,jk} - V_{1,jk} = \frac{1}{(b\ell)^2}\sum_{1\le t,t'\le b\ell}^{b\ell} g\left(\begin{pmatrix}X^*_{t,j}\\X^*_{t+1,k}\end{pmatrix}, \begin{pmatrix}X^*_{t',j}\\X^*_{t'+1,k}\end{pmatrix}\right) = \frac{1}{(b\ell)^2}\sum_{1\le t,t'\le b\ell}^{b\ell} g\left(\begin{pmatrix}Z^*_{t,j}\\Z^*_{t+1,k}\end{pmatrix}, \begin{pmatrix}Z^*_{t',j}\\Z^*_{t'+1,k}\end{pmatrix}\right),$$

where $(X^*_{t,j}, X^*_{t,k})^\top$ and $(X^*_{t,j}, X^*_{t+1,k})^\top$ are bootstrap sample points from $\boldsymbol{Y}_t = (\boldsymbol{X}_t^\top, \boldsymbol{X}_{t+1}^\top)^\top$ and $(Z^*_{t,j}, Z^*_{t,k})^\top$ and $(Z^*_{t,j}, Z^*_{t+1,k})^\top$ are bootstrap sample points corresponding to $(X^*_{t,j}, X^*_{t,k})^\top$ and $(X^*_{t,j}, X^*_{t+1,k})^\top$, respectively, via $\boldsymbol{X}_t = \boldsymbol{f}(\boldsymbol{Z}_t)$. Note that $\max_{j,k}|\widehat{V}_{0,jk} - V_{0,jk}| = O_{\mathbb{P}}\left(\sqrt{\frac{\log(ed)}{n}}\right)$ and $\max_{j,k}|\widehat{V}_{1,jk} - V_{1,jk}| = O_{\mathbb{P}}\left(\sqrt{\frac{\log(ed)}{n}}\right)$ by the proof of Theorem 2.1.

A.2.2 Verification of Assumptions 1-5 in Neykov et al. (2018a)

In this subsection, we will verify Assumptions 1 to 5 in Neykov et al. (2018a). Assumptions 1, 2, 4, and 5 have been verified in Appendix J in Neykov et al. (2018b). We will verify Assumption 3 by following the proofs of Lemma H.4 in Neykov et al. (2018b), Lemmas 3.3, 3.6 to 3.8 in Dehling and Wendler (2010), and Theorem 3.1 in Fan et al. (2016).

Lemma A.4 (Assumption 3 in Neykov et al. (2018a))**.** Suppose all assumptions in Theorem 3.1 and Theorem 3.2 hold. Then, Assumption 3 in Neykov et al. (2018a) holds, that is,

$$\frac{\sqrt{n-1}}{\sigma_n} S(\boldsymbol{\beta}) \xrightarrow{d} N(0,1).$$

Proof. In this part, we will follow the proofs of Lemma H.4 in Neykov et al. (2018b), Lemmas 3.6 and 3.8 in Dehling and Wendler (2010), and Theorem 3.1 in Fan et al. (2016).

Following the proof of Lemma H.4 in Neykov et al. (2018b), we have

$$\begin{aligned}
\sqrt{n-1}S(\boldsymbol{\beta}) &= \sqrt{n-1}\boldsymbol{w}^\top(\widehat{\boldsymbol{\Sigma}}_0\boldsymbol{\beta} - \widehat{\boldsymbol{\Sigma}}_{1,*m}) \\
&= \sqrt{n-1}\boldsymbol{w}^\top((\widehat{\boldsymbol{\Sigma}}_0 - \boldsymbol{\Sigma}_0)\boldsymbol{\beta} - (\widehat{\boldsymbol{\Sigma}}_{1,*m} - \boldsymbol{\Sigma}_{1,*m})) \\
&= \sqrt{n-1}\boldsymbol{w}^\top\left\{\left(\sin\left(\frac{\pi}{2}\widehat{\mathbf{T}}_0\right) - \sin\left(\frac{\pi}{2}\mathbf{T}_0\right)\right)\boldsymbol{\beta} - \left(\sin\left(\frac{\pi}{2}\widehat{\mathbf{T}}_{1,*m}\right) - \sin\left(\frac{\pi}{2}\mathbf{T}_{1,*m}\right)\right)\right\} \\
&= \sqrt{n-1}\left\{\sum_{j\in S_{\boldsymbol{w}},k\in S} w_j\beta_k\left(\sin\left(\frac{\pi}{2}\widehat{\tau}_{0,jk}\right) - \sin\left(\frac{\pi}{2}\tau_{0,jk}\right)\right) + \sum_{j\in S_{\boldsymbol{w}}} w_j\left(\sin\left(\frac{\pi}{2}\widehat{\tau}_{1,jm}\right) - \sin\left(\frac{\pi}{2}\tau_{1,jm}\right)\right)\right\} \\
&= \underbrace{\sqrt{n-1}\left\{\sum_{j\in S_{\boldsymbol{w}},k\in S} w_j\beta_k\cos\left(\frac{\pi}{2}\tau_{0,jk}\right)\frac{\pi}{2}(\widehat{\tau}_{0,jk}-\tau_{0,jk}) - \sum_{j\in S_{\boldsymbol{w}}} w_j\cos\left(\frac{\pi}{2}\tau_{1,jm}\right)\frac{\pi}{2}(\widehat{\tau}_{1,jm}-\tau_{1,jm})\right\}}_{K_1} \\
&\quad \underbrace{-\sqrt{n-1}\left\{\frac{1}{2}\sum_{j\in S_{\boldsymbol{w}},k\in S} w_j\beta_k\sin\left(\frac{\pi}{2}\widetilde{\tau}_{0,jk}\right)\left(\frac{\pi}{2}(\widehat{\tau}_{0,jk}-\tau_{0,jk})\right)^2 - \frac{1}{2}\sum_{j\in S_{\boldsymbol{w}}} w_j\sin\left(\frac{\pi}{2}\widetilde{\tau}_{1,jm}\right)\left(\frac{\pi}{2}(\widehat{\tau}_{1,jm}-\tau_{1,jm})\right)^2\right\}}_{K_2},
\end{aligned}$$

where $\widetilde{\tau}_{0,jk}$ is between $\widehat{\tau}_{0,jk}$ and $\tau_{0,jk}$, and $\widetilde{\tau}_{1,jm}$ is between $\widehat{\tau}_{1,jm}$ and $\tau_{1,jm}$. The second equality holds because $\boldsymbol{\Sigma}_0\boldsymbol{\beta} - \boldsymbol{\Sigma}_{1,*m} = \mathbf{0}$. Steps 1 and 2 show $K_2 = o_{\mathbb{P}}(1)$. We also have $K_1/\nu_n \xrightarrow{d} N(0,1)$ for an appropriate value of ν_n, which will be presented in Step 2-2.

Step 1. We will show $K_2 = o_{\mathbb{P}}(1)$ by proving $\mathbb{E}[K_2^2] = o(1)$ using Lemma A.10 in Section A.2.5, which follows the proof of Lemma 3.6 in Dehling and Wendler (2010).

Because $(a+b)^2 \le 2(a^2+b^2)$, it is sufficient to show that

$$(n-1)\mathbb{E}\left[\left(\sum_{j\in S_{\boldsymbol{w}},k\in S} w_j\beta_k\sin\left(\frac{\pi}{2}\widetilde{\tau}_{0,jk}\right)(\widehat{\tau}_{0,jk}-\tau_{0,jk})^2\right)^2\right] = o(1) \text{ and}$$

$$(n-1)\mathbb{E}\left[\left(\sum_{j\in S_{\boldsymbol{w}}} w_j\sin\left(\frac{\pi}{2}\widetilde{\tau}_{1,jm}\right)(\widehat{\tau}_{1,jm}-\tau_{1,jm})^2\right)^2\right] = o(1).$$

For the first one, we have

$$\begin{aligned}
&\left|(n-1)\mathbb{E}\left[\left(\sum_{j\in S_{\boldsymbol{w}},k\in S} w_j\beta_k\sin\left(\frac{\pi}{2}\widetilde{\tau}_{0,jk}\right)(\widehat{\tau}_{0,jk}-\tau_{0,jk})^2\right)^2\right]\right| \\
&= \left|(n-1)\mathbb{E}\left[\sum_{j\in S_{\boldsymbol{w}},k\in S}\sum_{j'\in S_{\boldsymbol{w}},k'\in S} w_j\beta_k w_{j'}\beta_{k'}\sin\left(\frac{\pi}{2}\widetilde{\tau}_{0,jk}\right)\sin\left(\frac{\pi}{2}\widetilde{\tau}_{0,j'k'}\right)(\widehat{\tau}_{0,jk}-\tau_{0,jk})^2(\widehat{\tau}_{0,j'k'}-\tau_{0,j'k'})^2\right]\right|
\end{aligned}$$

$$\le (n-1)\mathbb{E}\left[\sum_{j\in S_{\boldsymbol{w}},k\in S}\sum_{j'\in S_{\boldsymbol{w}},k'\in S}|w_j||\beta_k||w_{j'}||\beta_{k'}|(\widehat{\tau}_{0,jk}-\tau_{0,jk})^2(\widehat{\tau}_{0,j'k'}-\tau_{0,j'k'})^2\right]$$
$$\le \sum_{j\in S_{\boldsymbol{w}},k\in S}\sum_{j'\in S_{\boldsymbol{w}},k'\in S}|w_j||\beta_k||w_{j'}||\beta_{k'}|(n-1)\mathbb{E}[(\widehat{\tau}_{0,jk}-\tau_{0,jk})^2(\widehat{\tau}_{0,j'k'}-\tau_{0,j'k'})^2]$$
$$\le \frac{1}{2}\sum_{j\in S_{\boldsymbol{w}},k\in S}\sum_{j'\in S_{\boldsymbol{w}},k'\in S}|w_j||\beta_k||w_{j'}||\beta_{k'}|(n-1)\left\{\mathbb{E}[(\widehat{\tau}_{0,jk}-\tau_{0,jk})^4]+\mathbb{E}[(\widehat{\tau}_{0,j'k'}-\tau_{0,j'k'})^4]\right\}$$
$$= \|\boldsymbol{w}\|_1\|\boldsymbol{\beta}\|_1\sum_{j\in S_{\boldsymbol{w}},k\in S}|w_j||\beta_k|(n-1)\mathbb{E}[(\widehat{\tau}_{0,jk}-\tau_{0,jk})^4],$$

where the last inequality comes from $|ab|\le(a^2+b^2)/2$. Because $(a+b)^4\le 8a^4+8a^4$, we have

$$\mathbb{E}[(\widehat{\tau}_{0,jk}-\tau_{0,jk})^4]$$
$$=\mathbb{E}\left[\left\{\frac{2}{n-1}\sum_{t=1}^{n-1}g_1(\boldsymbol{Z}_{t,\{j,k\}})+\frac{2}{(n-1)(n-2)}\sum_{t<t'}g_2(\boldsymbol{Z}_{t,\{j,k\}},\boldsymbol{Z}_{t',\{j,k\}})\right\}^4\right]$$
$$\le\frac{128}{(n-1)^4}\mathbb{E}\left[\left(\sum_{t=1}^{n-1}g_1(\boldsymbol{Z}_{t,\{j,k\}})\right)^4\right]+\frac{128}{(n-1)^4(n-2)^4}\mathbb{E}\left[\left(\sum_{t<t'}g_2(\boldsymbol{Z}_{t,\{j,k\}},\boldsymbol{Z}_{t',\{j,k\}})\right)^4\right]$$
$$\le\frac{128}{(n-1)^4}\sum_{i_1,i_2,i_3,i_4=1}^{n-1}\left|\mathbb{E}[g_1(\boldsymbol{Z}_{i_1,\{j,k\}})g_1(\boldsymbol{Z}_{i_2,\{j,k\}})g_1(\boldsymbol{Z}_{i_3,\{j,k\}})g_1(\boldsymbol{Z}_{i_4,\{j,k\}})]\right|$$
$$+\frac{8}{(n-1)^4(n-2)^4}\sum_{i_1,\dots,i_8=1}^{n-1}\left|\mathbb{E}[g_2(\boldsymbol{Z}_{i_1,\{j,k\}},\boldsymbol{Z}_{i_2,\{j,k\}})g_2(\boldsymbol{Z}_{i_3,\{j,k\}},\boldsymbol{Z}_{i_4,\{j,k\}})g_2(\boldsymbol{Z}_{i_5,\{j,k\}},\boldsymbol{Z}_{i_6,\{j,k\}})g_2(\boldsymbol{Z}_{i_7,\{j,k\}},\boldsymbol{Z}_{i_8,\{j,k\}})]\right|.$$

Because $\{\boldsymbol{U}_t\}_{t\in\mathbb{Z}}$ is geometrically α-mixing, Lemma A.10 in Section A.2.5 implies $\mathbb{E}[(\widehat{\tau}_{0,jk}-\tau_{0,jk})^4]\le C(n-1)^{-2}$, where C is a constant which only depends on ($C_{\mathbf{A}}$, $c_{\boldsymbol{E}}$, $C_{\boldsymbol{E}}$, κ_1, κ_2). Therefore, one has

$$\left|(n-1)\mathbb{E}\left[\left(\sum_{j\in S_{\boldsymbol{w}},k\in S}w_j\beta_k\sin\left(\frac{\pi}{2}\widetilde{\tau}_{0,jk}\right)(\widehat{\tau}_{0,jk}-\tau_{0,jk})^2\right)^2\right]\right|\le C\|\boldsymbol{w}\|_1^2\|\boldsymbol{\beta}\|_1^2O(n^{-1})=o(1).$$

Similarly, we have

$$(n-1)\mathbb{E}\left[\left(\sum_{j\in S_{\boldsymbol{w}}}w_j\sin\left(\frac{\pi}{2}\widetilde{\tau}_{1,jm}\right)(\widehat{\tau}_{1,jm}-\tau_{1,jm})^2\right)^2\right]=o(1).$$

In conclusion, we have $\mathbb{E}[K_2^2]=o(1)$, $\mathrm{Var}(K_2)=o(1)$, and $K_2=o_{\mathbb{P}}(1)$.

Step 2. We will investigate the asymptotic distribution of K_1.

Using Hoeffding decomposition of $\widehat{\tau}_{0,jm}-\tau_{0,jm}$ and $\widehat{\tau}_{1,jm}-\tau_{1,jm}$, we can decompose K_1 as

$$K_1=\sqrt{n-1}\left\{\sum_{j\in S_{\boldsymbol{w}},k\in S}w_j\beta_k\cos\left(\frac{\pi}{2}\tau_{0,jk}\right)\frac{\pi}{2}(\widehat{\tau}_{0,jk}-\tau_{0,jk})-\sum_{j\in S_w}w_j\cos\left(\frac{\pi}{2}\tau_{1,jm}\right)\frac{\pi}{2}(\widehat{\tau}_{1,jm}-\tau_{1,jm})\right\}$$
$$=K_{11}+K_{12},\ \text{where}$$

$$K_{11}=\sqrt{n-1}\left\{\sum_{j\in S_{\boldsymbol{w}},k\in S}w_j\beta_k\cos\left(\frac{\pi}{2}\tau_{0,jk}\right)\frac{\pi}{n-1}\sum_{t=1}^{n-1}g_1\left(\begin{pmatrix}Z_{t,j}\\Z_{t,k}\end{pmatrix}\right)-\sum_{j\in S_{\boldsymbol{w}}}w_j\cos\left(\frac{\pi}{2}\tau_{1,jm}\right)\frac{\pi}{n-1}\sum_{t=1}^{n-1}g_1\left(\begin{pmatrix}Z_{t,j}\\Z_{t+1,m}\end{pmatrix}\right)\right\},$$

$$K_{12} = \sqrt{n-1}\Bigg\{\sum_{j\in S_{\boldsymbol{w}},k\in S} w_j\beta_k\cos\left(\frac{\pi}{2}\tau_{0,jk}\right)\frac{\pi}{(n-1)(n-2)}\sum_{t<t'} g_2\left(\begin{pmatrix} Z_{t,j}\\ Z_{t,k}\end{pmatrix},\begin{pmatrix} Z_{t',j}\\ Z_{t',k}\end{pmatrix}\right)$$
$$- \sum_{j\in S_{\boldsymbol{w}}} w_j\cos\left(\frac{\pi}{2}\tau_{1,jm}\right)\frac{\pi}{(n-1)(n-2)}\sum_{t<t'} g_2\left(\begin{pmatrix} Z_{t,j}\\ Z_{t+1,m}\end{pmatrix},\begin{pmatrix} Z_{t',j}\\ Z_{t'+1,m}\end{pmatrix}\right)\Bigg\}.$$

Steps 2-1 and 2-2 combined show that $K_{12} = o_{\mathbb{P}}(1)$ and $K_{11}/\nu_n \xrightarrow{d} N(0,1)$, where $\nu_n^2 = \mathbb{E}[K_{11}^2]$.

Step 2-1. In this part, we will show $K_{12} = o_{\mathbb{P}}(1)$ by proving $\mathbb{E}[K_{12}^2] = o(1)$. Similar to Step 1, Lemma A.10 in Section A.2.5 implies

$$(n-1)\mathbb{E}\left[\left(\sum_{j\in S_{\boldsymbol{w}},k\in S} w_j\beta_k\cos\left(\frac{\pi}{2}\tau_{0,jk}\right)\frac{\pi}{(n-1)(n-2)}\sum_{t<t'} g_2\left(\begin{pmatrix} Z_{t,j}\\ Z_{t,k}\end{pmatrix},\begin{pmatrix} Z_{t',j}\\ Z_{t',k}\end{pmatrix}\right)\right)^2\right] \le C\|\boldsymbol{w}\|_1^2\|\boldsymbol{\beta}\|_1^2 O(n^{-1})$$

and

$$(n-1)\mathbb{E}\left[\left(\sum_{j\in S_{\boldsymbol{w}}} w_j\cos\left(\frac{\pi}{2}\tau_{1,jm}\right)\frac{\pi}{(n-1)(n-2)}\sum_{t<t'} g_2\left(\begin{pmatrix} Z_{t,j}\\ Z_{t+1,m}\end{pmatrix},\begin{pmatrix} Z_{t',j}\\ Z_{t'+1,m}\end{pmatrix}\right)\right)^2\right] \le C'\|\boldsymbol{w}\|_1^2 O(n^{-1}),$$

where C and C' are some constants which only depend on $(C_{\mathbf{A}}, c_{\boldsymbol{E}}, C_{\boldsymbol{E}}, \kappa_1, \kappa_2)$. Therefore, we have $\mathbb{E}[K_{12}^2] = o(1)$, $\mathrm{Var}(K_{12}) = o(1)$, and $K_{12} = o_{\mathbb{P}}(1)$.

Step 2-2. We show $K_{11}/\nu_n \xrightarrow{d} N(0,1)$, where $\nu_n^2 = \mathbb{E}[K_{11}^2]$. First, we can rewrite K_{11} as

$$K_{11} = \frac{1}{\sqrt{n-1}}\sum_{t=1}^{n-1}\left[\underbrace{\sum_{j\in S_{\boldsymbol{w}},k\in S} w_j\beta_k\pi\cos\left(\frac{\pi}{2}\tau_{0,jk}\right)g_1\left(\begin{pmatrix} Z_{t,j}\\ Z_{t,k}\end{pmatrix}\right) - \sum_{j\in S_{\boldsymbol{w}}} w_j\pi\cos\left(\frac{\pi}{2}\tau_{1,jm}\right)g_1\left(\begin{pmatrix} Z_{t,j}\\ Z_{t+1,m}\end{pmatrix}\right)}_{G_{n1}(\boldsymbol{U}_t)}\right].$$

Then, $\{G_{n1}(\boldsymbol{U}_t)\}_{t\in\mathbb{Z}}$ is also geometrically α-mixing because $G_{n1}(\boldsymbol{U}_t)$ is the function of $\boldsymbol{U}_t = (\boldsymbol{Z}_t^\top, \boldsymbol{Z}_{t+1}^\top)^\top$ and $\boldsymbol{U}_t$ is geometrically α-mixing. Also, $G_{n1}(\boldsymbol{U}_t)$ is bounded under our assumption because

$$|G_{n1}(\boldsymbol{U}_t)| \le 2\pi\|\boldsymbol{w}\|_1(\|\boldsymbol{\beta}\|_1 + 1) \le 2\pi\|\boldsymbol{\Sigma}_0^{-1}\|_1(M+1) = O(1).$$

It implies that $\nu_n^2 < \widetilde{C}$ by some absolute positive constant $\widetilde{C}$. Also, similar to the proof of Theorem 3.1 in Fan et al. (2016), we have

$$|\sigma_n^2 - \mathbb{E}[K_{11}^2]| \le \mathrm{Var}(K_{12} + K_2) + \sqrt{\mathbb{E}[K_{11}^2]}\sqrt{\mathrm{Var}(K_{12} + K_2)} = o(1) + O(1)o(1), \qquad \text{(A.4)}$$

which implies $\sigma_n^2 - \mathbb{E}[K_{11}^2] = o(1)$. Assumptions in Theorem 3.2 imply that we $\sigma_n^2 = \mathrm{Var}(K_{11} + K_{12} + K_2) > C > 0$ by some absolute constant C. This implies that there exist absolute positive constants N, $\widetilde{c}$, and $\widetilde{C}$ such that we have $\widetilde{c} < \nu_n^2 = \mathbb{E}[K_{11}^2] < \widetilde{C}$ for all $n \ge N$. Therefore, we can use triangular array CLT and get the following result (see Theorem 10.2 in Pötscher and Prucha (1997), and also Ekström (2014)),

$$K_{11}/\nu_n \xrightarrow{d} N(0,1) \text{ where } \nu_n^2 = \mathbb{E}[K_{11}^2].$$

Because Steps 1 and 2-1 imply $\text{Var}(K_{12}) = o(1)$, $\text{Var}(K_2) = o(1)$, $K_{12} = o_{\mathbb{P}}(1)$, and $K_2 = o_{\mathbb{P}}(1)$, we have the desired result. □

A.2.3 Proofs of Theorems 3.1 and 3.2

Given Lemma A.4, Theorem 3.1 now holds by checking all conditions in Theorem 1 in Neykov et al. (2018a), and Theorem 3.2 now holds by Theorem 2 in Neykov et al. (2018a).

A.2.4 Proof of Theorem 3.3

Note that under assumptions in Theorem 3.3, Theorem 2.3 implies $\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\|_1 = O_{\mathbb{P}}(s\sqrt{\log(ed)/n})$, and the proof of Theorem 3.1 implies $\|\widehat{\boldsymbol{w}}-\boldsymbol{w}\|_1 = O_{\mathbb{P}}(s_{\boldsymbol{w}}\sqrt{\log(ed)/n})$ by verifying Assumption 2 in Neykov et al. (2018a). Given this information, we will prove Theorem 3.3 in two steps. Step 1 will show $(b\ell)\text{Var}^*(\boldsymbol{w}^\top(\widehat{\boldsymbol{\Sigma}}_0^*\boldsymbol{\beta}-\widehat{\boldsymbol{\Sigma}}_{1,*m}^*))-\sigma_n^2 = o_{\mathbb{P}}(1)$. Step 2 will show $\widehat{\sigma}_n^{2*}-(b\ell)\text{Var}^*(\boldsymbol{w}^\top(\widehat{\boldsymbol{\Sigma}}_0^*\boldsymbol{\beta}-\widehat{\boldsymbol{\Sigma}}_{1,*m}^*)) = o_{\mathbb{P}}(1)$.

Step 1. In this step, we will show $(b\ell)\text{Var}^*(\boldsymbol{w}^\top(\widehat{\boldsymbol{\Sigma}}_0^*\boldsymbol{\beta}-\widehat{\boldsymbol{\Sigma}}_{1,*m}^*)) - \sigma_n^2 = o_{\mathbb{P}}(1)$. Let $S^*(\boldsymbol{\beta}) = \boldsymbol{w}^\top(\widehat{\boldsymbol{\Sigma}}_0^*\boldsymbol{\beta}-\widehat{\boldsymbol{\Sigma}}_{1,*m}^*)$. Then

$$\begin{aligned}\sqrt{b\ell}S^*(\boldsymbol{\beta}) = &\underbrace{\sqrt{b\ell}\left\{\sum_{j\in S_{\boldsymbol{w}},k\in S} w_j\beta_k \cos\left(\frac{\pi}{2}\tau_{0,jk}\right)\frac{\pi}{2}(\widehat{\tau}_{0,jk}^*-\tau_{0,jk}) - \sum_{j\in S_{\boldsymbol{w}}} w_j \cos\left(\frac{\pi}{2}\tau_{1,jm}\right)\frac{\pi}{2}(\widehat{\tau}_{1,jm}^*-\tau_{1,jm})\right\}}_{K_1^*} \\ &\underbrace{-\frac{\sqrt{b\ell}}{2}\left\{\sum_{j\in S_{\boldsymbol{w}},k\in S} w_j\beta_k \sin\left(\frac{\pi}{2}\widetilde{\tau}_{jk}^*\right)\left[\frac{\pi}{2}(\widehat{\tau}_{0,jk}^*-\tau_{0,jk})\right]^2 - \sum_{j\in S_{\boldsymbol{w}}} w_j \sin\left(\frac{\pi}{2}\widetilde{\tau}_{1,jm}^*\right)\left[\frac{\pi}{2}(\widehat{\tau}_{1,jm}^*-\tau_{1,jm})\right]^2\right\}}_{K_2^*},\end{aligned}$$

where $\widetilde{\tau}_{0,jk}^*$ is between $\widehat{\tau}_{0,jk}^*$ and $\tau_{0,jk}$, and $\widetilde{\tau}_{1,jm}^*$ is between $\widehat{\tau}_{1,jm}^*$ and $\tau_{1,jm}$. We have

$$\begin{aligned}K_1^* &= \sqrt{b\ell}\left\{\sum_{j\in S_{\boldsymbol{w}},k\in S} w_j\beta_k \cos\left(\frac{\pi}{2}\tau_{0,jk}\right)\frac{\pi}{2}(\widehat{\tau}_{0,jk}^*-\tau_{0,jk}) - \sum_{j\in S_{\boldsymbol{w}}} w_j \cos\left(\frac{\pi}{2}\tau_{1,jm}\right)\frac{\pi}{2}(\widehat{\tau}_{1,jm}^*-\tau_{1,jm})\right\} \\ &= \underbrace{\frac{1}{\sqrt{b\ell}}\sum_{t=1}^{n}\left\{\sum_{j\in S_{\boldsymbol{w}},k\in S} w_j\beta_k \cos\left(\frac{\pi}{2}\tau_{0,jk}\right)\pi g_1\begin{pmatrix}Z_{t,j}^*\\ Z_{t,k}^*\end{pmatrix} - \sum_{j\in S_{\boldsymbol{w}}} w_j \cos\left(\frac{\pi}{2}\tau_{1,jm}\right)\pi g_1\begin{pmatrix}Z_{t,j}^*\\ Z_{t+1,m}^*\end{pmatrix}\right\}}_{K_{11}^*} + K_{12}^*,\end{aligned}$$

where

$$K_{12}^* := \sum_{j\in S_{\boldsymbol{w}},k\in S} w_j\beta_k \cos\left(\frac{\pi}{2}\tau_{0,jk}\right)\pi\frac{\sqrt{b\ell}}{(b\ell)(b\ell-1)}\sum_{t<t'} g_2\left(\begin{pmatrix}Z_{t,j}^*\\ Z_{t,k}^*\end{pmatrix}, \begin{pmatrix}Z_{t',j}^*\\ Z_{t',k}^*\end{pmatrix}\right)$$

$$- \sum_{j \in S_{\boldsymbol{w}}} w_j \cos\left(\frac{\pi}{2}\tau_{0,jm}\right) \pi \frac{\sqrt{b\ell}}{(b\ell)(b\ell-1)} \sum_{t<t'} g_2 \left(\begin{pmatrix} Z^*_{t,j} \\ Z^*_{t+1,m} \end{pmatrix}, \begin{pmatrix} Z^*_{t',j} \\ Z^*_{t+1',m} \end{pmatrix} \right).$$

Steps 1-1 to 1-3 below combined will show $\mathbb{E}^*[(K_2^*)^2] = o_{\mathbb{P}}(1)$, $\mathrm{Var}^*(K_{12}^*) = o_{\mathbb{P}}(1)$, and $\mathrm{Var}^*(K_{11}^*) - \mathbb{E}[K_{11}^2] = o_{\mathbb{P}}(1)$ by following the proofs of Theorem 2.1 in Dehling and Wendler (2010), Theorem 2.1 in Gonçalves and White (2002), and Theorem 3.1 in Lahiri (2003). Note that Steps 1-1 to 1-3 imply $\mathrm{Var}^*(K_{11}^*) - (b\ell)\mathrm{Var}^*(\boldsymbol{w}^\top(\widehat{\boldsymbol{\Sigma}}_0^*\boldsymbol{\beta} - \widehat{\boldsymbol{\Sigma}}_{1,*m}^*)) = o_{\mathbb{P}}(1)$ because, similar to the proof of Theorem 3.2 in Fan et al. (2016), we have

$$|\mathrm{Var}^*(K_{11}^*) - (b\ell)\mathrm{Var}^*(\boldsymbol{w}^\top(\widehat{\boldsymbol{\Sigma}}_0^*\boldsymbol{\beta} - \widehat{\boldsymbol{\Sigma}}_{1,*m}^*))| \le \mathrm{Var}^*(K_{12}^* + K_2^*) + \sqrt{\mathrm{Var}^*(K_{11}^*)}\sqrt{\mathrm{Var}^*(K_{12}^* + K_2^*)} = o_{\mathbb{P}}(1).$$

Step 1-1. We show that $\mathbb{E}^*[(K_2^*)^2] = o_{\mathbb{P}}(1)$. Note that

$$K_2^* = \frac{\sqrt{b\ell}}{2} \left\{ \sum_{j \in S_{\boldsymbol{w}}, k \in S} w_j \beta_k \sin\left(\frac{\pi}{2}\widetilde{\tau}_{jk}\right) \left[\frac{\pi}{2}(\widehat{\tau}^*_{0,jk} - \tau_{0,jk})\right]^2 - \sum_{j \in S_{\boldsymbol{w}}} w_j \sin\left(\frac{\pi}{2}\widetilde{\tau}_{1,jm}\right) \left[\frac{\pi}{2}(\widehat{\tau}^*_{1,jm} - \tau_{1,jm})\right]^2 \right\}.$$

Then, similar to the proof of Lemma A.4, it is sufficient to bound $(b\ell)\mathbb{E}[\mathbb{E}^*[(\widehat{\tau}^*_{0,jk} - \tau_{0,jk})^4]$ and $(b\ell)\mathbb{E}[\mathbb{E}^*[(\widehat{\tau}^*_{1,jm} - \tau_{1,jm})^4]]$ uniformly. Because $(a+b)^4 \le 8a^4 + 8a^4$, we have

$$\begin{aligned}
&\mathbb{E}[\mathbb{E}^*[(\widehat{\tau}^*_{0,jk} - \tau_{0,jk})^4]] \\
=&\mathbb{E}\left[\mathbb{E}^*\left[\left\{\frac{2}{b\ell}\sum_{t=1}^{b\ell} g_1\left(\begin{pmatrix} Z^*_{t,j} \\ Z^*_{t,k}\end{pmatrix}\right) + \frac{2}{b\ell(b\ell-1)}\sum_{t<t'} g_2\left(\begin{pmatrix} Z^*_{t,j} \\ Z^*_{t,k}\end{pmatrix}, \begin{pmatrix} Z^*_{t',j} \\ Z^*_{t',k}\end{pmatrix}\right)\right\}^4\right]\right] \\
\le&\frac{128}{(b\ell)^4}\mathbb{E}\left[\mathbb{E}^*\left[\left\{\sum_{t=1}^{b\ell} g_1\left(\begin{pmatrix} Z^*_{t,j} \\ Z^*_{t,k}\end{pmatrix}\right)\right\}^4\right]\right] + \frac{128}{(b\ell)^4(b\ell-1)^4}\mathbb{E}\left[\mathbb{E}^*\left[\left\{\sum_{t<t'} g_2\left(\begin{pmatrix} Z^*_{t,j} \\ Z^*_{t,k}\end{pmatrix}, \begin{pmatrix} Z^*_{t',j} \\ Z^*_{t',k}\end{pmatrix}\right)\right\}^4\right]\right] \\
\le&\frac{128}{(b\ell)^4}\sum_{i_1,i_2,i_3,i_4=1}^{n-1}\left|\mathbb{E}\left[g_1\left(\begin{pmatrix} Z_{i_1,j} \\ Z_{i_1,k}\end{pmatrix}\right) g_1\left(\begin{pmatrix} Z_{i_2,j} \\ Z_{i_2,k}\end{pmatrix}\right) g_1\left(\begin{pmatrix} Z_{i_3,j} \\ Z_{i_3,k}\end{pmatrix}\right) g_1\left(\begin{pmatrix} Z_{i_4,j} \\ Z_{i_4,k}\end{pmatrix}\right)\right]\right| \\
&+ \frac{8}{(b\ell)^4(b\ell-1)^4}\sum_{i_1,\dots,i_8=1}^{n-1}\left|\mathbb{E}\left[g_2\left(\begin{pmatrix} Z_{i_1,j} \\ Z_{i_1,k}\end{pmatrix}, \begin{pmatrix} Z_{i_2,j} \\ Z_{i_2,k}\end{pmatrix}\right) g_2\left(\begin{pmatrix} Z_{i_3,j} \\ Z_{i_3,k}\end{pmatrix}, \begin{pmatrix} Z_{i_4,j} \\ Z_{i_4,k}\end{pmatrix}\right) g_2\left(\begin{pmatrix} Z_{i_5,j} \\ Z_{i_5,k}\end{pmatrix}, \begin{pmatrix} Z_{i_6,j} \\ Z_{i_6,k}\end{pmatrix}\right) g_2\left(\begin{pmatrix} Z_{i_7,j} \\ Z_{i_7,k}\end{pmatrix}, \begin{pmatrix} Z_{i_8,j} \\ Z_{i_8,k}\end{pmatrix}\right)\right]\right| \\
&\le \widetilde{C}'^{-2},
\end{aligned}$$

where $\widetilde{C}$ is some constant which only depends on $(C_{\mathbf{A}}, c_{\boldsymbol{E}}, C_{\boldsymbol{E}}, \kappa_1, \kappa_2)$. The last two inequalities come from the proofs of Lemmas 3.7 and 3.8 in Dehling and Wendler (2010) and Lemmas A.8 to A.10 in Section A.2.5. Similarly, one can show that $\mathbb{E}[\mathbb{E}^*[(\widehat{\tau}^*_{1,jk} - \tau_{1,jk})^4]] \le \widetilde{C}'(n-1)^{-2}$, where $\widetilde{C}'$ is some constant which only depends on $(C_{\mathbf{A}}, c_{\boldsymbol{E}}, C_{\boldsymbol{E}}, \kappa_1, \kappa_2)$. Then, similar to the proof of Lemma A.4, we have $\mathbb{E}[\mathbb{E}^*[(K_2^*)^2]] \le C\|\boldsymbol{w}\|_1^2(\|\boldsymbol{\beta}\|_1^2 + 1)O(n^{-1}) = o(1)$, where C is some constant which only depends on $(C_{\mathbf{A}}, c_{\boldsymbol{E}}, C_{\boldsymbol{E}}, \kappa_1, \kappa_2)$. Therefore, we have $\mathbb{E}^*[(K_2^*)^2] = o_{\mathbb{P}}(1)$.

Step 1-2. We show that $\mathrm{Var}^*(K_{12}^*) = o_{\mathbb{P}}(1)$ by proving $\mathbb{E}[\mathbb{E}^*[(K_{12}^*)^2]] = o(1)$. For simplicity, let's denote

$$U^*_{0,jk}(g_2) = \frac{2}{(b\ell)(b\ell-1)}\sum_{t<t'} g_2\left(\begin{pmatrix} Z^*_{t,j} \\ Z^*_{t,k}\end{pmatrix}, \begin{pmatrix} Z^*_{t',j} \\ Z^*_{t',k}\end{pmatrix}\right) \text{ and } U^*_{1,jk}(g_2) = \frac{2}{(b\ell)(b\ell-1)}\sum_{t<t'} g_2\left(\begin{pmatrix} Z^*_{t,j} \\ Z^*_{t+1,k}\end{pmatrix}, \begin{pmatrix} Z^*_{t',j} \\ Z^*_{t'+1,k}\end{pmatrix}\right).$$

Then, we can rewrite K_{12}^* as

$$K_{12}^* = \frac{\pi}{2}\sum_{j\in S_{\boldsymbol{w}},k\in S} w_j\beta_k \cos\left(\frac{\pi}{2}\tau_{0,jk}\right)\sqrt{b\ell}U_{0,jk}^*(g_2) - \frac{\pi}{2}\sum_{j\in S_{\boldsymbol{w}}} w_j \cos\left(\frac{\pi}{2}\tau_{0,jm}\right)\sqrt{b\ell}U_{1,jk}^*(g_2).$$

Because $(a+b)^2 \le 2(a^2+b^2)$, it is sufficient to show

$$\mathbb{E}\left[\mathbb{E}^*\left[\left(\sum_{j\in S_{\boldsymbol{w}},k\in S} w_j\beta_k \cos\left(\frac{\pi}{2}\tau_{0,jk}\right)\sqrt{b\ell}U_{0,jk}^*(g_2)\right)^2\right]\right] = o(1) \text{ and } \mathbb{E}\left[\mathbb{E}^*\left[\left(\sum_{j\in S_{\boldsymbol{w}}} w_j \cos\left(\frac{\pi}{2}\tau_{0,jm}\right)\sqrt{b\ell}U_{1,jm}^*(g_2)\right)^2\right]\right] = o(1).$$

From proofs of Lemmas 3.6 and 3.7 in Dehling and Wendler (2010) and Lemma A.8 in Section A.2.5, we have

$$\begin{aligned}
(b\ell)\mathbb{E}[\mathbb{E}^*[(U_{0,jk}^*(g_2))^2]] &\le \frac{C}{(b\ell)(b\ell-1)^2}\sum_{i_1,i_2,i_3,i_4}\left|\mathbb{E}\left[g_2\left(\begin{pmatrix}Z_{i_1,j}\\Z_{i_1,k}\end{pmatrix},\begin{pmatrix}Z_{i_2,j}\\Z_{i_2,k}\end{pmatrix}\right)g_2\left(\begin{pmatrix}Z_{i_3,j}\\Z_{i_3,k}\end{pmatrix},\begin{pmatrix}Z_{i_4,j}\\Z_{i_4,k}\end{pmatrix}\right)\right]\right| \\
&\le \frac{C'}{(b\ell)(b\ell-1)^2}n^2\sum_{m=1}^{n} m\alpha_{\boldsymbol{U}}(m)^{2/5} \\
&\le C''^{-1},
\end{aligned}$$

where C is an absolute constant, C' only depends on $(C_{\mathbf{A}}, c_{\boldsymbol{E}}, C_{\boldsymbol{E}})$, and C'' only depends on $(C_{\mathbf{A}}, c_{\boldsymbol{E}}, C_{\boldsymbol{E}}, \kappa_1, \kappa_2)$. Therefore, by doing the same calculation as in Lemma A.4, we have

$$\mathbb{E}\left[\mathbb{E}^*\left[\left(\sum_{j\in S_{\boldsymbol{w}},k\in S} w_j\beta_k \cos\left(\frac{\pi}{2}\tau_{0,jk}\right)\sqrt{b\ell}U_{0,jk}^*(g_2)\right)^2\right]\right] \le 2C''\|\boldsymbol{w}\|_1^2\|\boldsymbol{\beta}\|_1^2 O(n^{-1}) = o(1).$$

Using the same approach, we have

$$\mathbb{E}\left[\mathbb{E}^*\left[\left(\sum_{j\in S_{\boldsymbol{w}}} w_j \cos\left(\frac{\pi}{2}\tau_{0,jm}\right)\sqrt{b\ell}U_{1,jm}^*(g_2)\right)^2\right]\right] = o(1).$$

Therefore, we have $\mathbb{E}[\mathbb{E}^*[(K_{12}^*)^2]] = o(1)$ and it implies that $\mathrm{Var}^*(K_{12}^*) \le \mathbb{E}^*(K_{12}^*) = o_{\mathbb{P}}(1)$.

Step 1-3. We will show $\mathrm{Var}^*(K_{11}^*) - \mathbb{E}^*[K_{11}^2] = o_{\mathbb{P}}(1)$. Define

$$\begin{aligned}
G_{n1}(\boldsymbol{U}_t) &= \left\{\sum_{j\in S_{\boldsymbol{w}},k\in S} w_j\beta_k\cos\left(\frac{\pi}{2}\tau_{0,jk}\right)\pi g_1\begin{pmatrix}Z_{t,j}\\Z_{t,k}\end{pmatrix} - \sum_{j\in S_{\boldsymbol{w}}} w_j\cos\left(\frac{\pi}{2}\tau_{1,jm}\right)\pi g_1\begin{pmatrix}Z_{t,j}\\Z_{t+1,m}\end{pmatrix}\right\}, \\
G_{n1}^*(\boldsymbol{U}_t^*) &= \left\{\sum_{j\in S_{\boldsymbol{w}},k\in S} w_j\beta_k\cos\left(\frac{\pi}{2}\tau_{0,jk}\right)\pi g_1\begin{pmatrix}Z_{t,j}^*\\Z_{t,k}^*\end{pmatrix} - \sum_{j\in S_{\boldsymbol{w}}} w_j\cos\left(\frac{\pi}{2}\tau_{1,jm}\right)\pi g_1\begin{pmatrix}Z_{t,j}^*\\Z_{t+1,m}^*\end{pmatrix}\right\},
\end{aligned}$$

where $\boldsymbol{U}_t^* = \boldsymbol{f}^{-1}(\boldsymbol{Y}_t^*)$. Then, we can rewrite K_{11} and K_{11}^* as

$$K_{11} = \frac{1}{\sqrt{n-1}}\sum_{t=1}^{n-1} G_{n1}(\boldsymbol{U}_t) \quad \text{and} \quad K_{11}^* = \frac{1}{\sqrt{b\ell}}\sum_{t=1}^{b\ell} G_{n1}^*(\boldsymbol{U}_t^*).$$

Note that $\{G_{n1}(\boldsymbol{U}_t)\}_{t\in\mathbb{Z}}$ is a triangular array with geometrically α-mixing process because $\{\boldsymbol{U}_t\}_{t\in\mathbb{Z}}$ is a geometrically α-mixing process and $G_{n1}(\cdot)$ is continuous non-random function. Then, we can show the consistency of bootstrap version of variance following the proofs of Theorem 2.1 in Gonçalves and White (2002) and Theorem 3.1 in Lahiri (2003).

Let $\widehat{\sigma}^{2*}_{n,\mathrm{MBB}}$ be the bootstrap variance of K^*_{11} when $G^*_{n1}(U^*_t)$ is constructed from the moving block bootstrap (MBB) sample, not the circular block bootstrap (CBB) sample. Let $\widehat{\sigma}^{2*}_{n,\mathrm{CBB}} := \mathrm{Var}^*(K^*_{11})$ be the bootstrap variance of K^*_{11} when $G^*_{n1}(U^*_t)$ is constructed from the circular block bootstrap (CBB) sample. We refer to Chapter 3 of Lahiri (2003) for definitions of MBB and CBB. Then, we finish Step 1-3 by doing the following procedure.

Step 1-3-1. We show that $\widehat{\sigma}^{2*}_{n,\mathrm{MBB}} - \widehat{\sigma}^{2*}_{n,\mathrm{CBB}} = o_{\mathbb{P}}(1)$ by following the proof of Theorem 3.1 in Lahiri (2003) even though we are dealing with the triangular array because Assumption M implies $G_{n1}(\cdot)$ is bounded above by some absolute constant.

Step 1-3-2. Because an α-mixing process is the special case of an NED process, we can show $\widehat{\sigma}^{2*}_{n,\mathrm{MBB}} - \mathbb{E}[K^2_{11}] = o_{\mathbb{P}}(1)$ by following the proof of Theorem 2.1 in Gonçalves and White (2002).

We skip the proofs of Steps 1-3-1 and 1-3-2 because it can be easily verified by following the proofs of Theorem 3.1 in Lahiri (2003) and Theorem 2.1 in Gonçalves and White (2002) step by step. In conclusion, the proof of Lemma A.4 and Steps 1-1 to 1-3 imply that $(b\ell)\mathrm{Var}^*(\boldsymbol{w}^\top(\widehat{\boldsymbol{\Sigma}}^*_0\boldsymbol{\beta} - \widehat{\boldsymbol{\Sigma}}^*_{1,*m})) - \sigma^2_n = o_{\mathbb{P}}(1)$.

Step 2. we will show that $\widehat{\sigma}^{2*}_n - (b\ell)\mathrm{Var}^*(\boldsymbol{w}^\top(\widehat{\boldsymbol{\Sigma}}^*_0\boldsymbol{\beta} - \widehat{\boldsymbol{\Sigma}}^*_{1,*m})) = o_{\mathbb{P}}(1)$. Note that

$$\sqrt{b\ell}\widehat{\boldsymbol{w}}^\top(\widehat{\boldsymbol{\Sigma}}^*_0\widehat{\boldsymbol{\beta}}(\theta)-\widehat{\boldsymbol{\Sigma}}^*_{1,*m})-\sqrt{b\ell}\boldsymbol{w}^\top(\widehat{\boldsymbol{\Sigma}}^*_0\boldsymbol{\beta}-\widehat{\boldsymbol{\Sigma}}^*_{1,*m}) = \sqrt{b\ell}(\widehat{\boldsymbol{w}}-\boldsymbol{w})^\top(\widehat{\boldsymbol{\Sigma}}^*_0\boldsymbol{\beta}-\widehat{\boldsymbol{\Sigma}}^*_{1,*m})+\sqrt{b\ell}\widehat{\boldsymbol{w}}^\top\widehat{\boldsymbol{\Sigma}}^*_0(\widehat{\boldsymbol{\beta}}(\theta)-\boldsymbol{\beta}).$$

Steps 2-1 and 2-2 below combined will show that

$$\mathbb{E}^*[|\sqrt{b\ell}(\widehat{\boldsymbol{w}} - \boldsymbol{w})^\top(\widehat{\boldsymbol{\Sigma}}^*_0\boldsymbol{\beta} - \widehat{\boldsymbol{\Sigma}}^*_{1,*m})|^2] = o_{\mathbb{P}}(1) \text{ and } \mathbb{E}^*[|\sqrt{b\ell}\widehat{\boldsymbol{w}}^\top\widehat{\boldsymbol{\Sigma}}^*_0(\widehat{\boldsymbol{\beta}}(\theta) - \boldsymbol{\beta})|^2] = o_{\mathbb{P}}(1).$$

Step 2-1. We show that $\mathbb{E}^*[|\sqrt{b\ell}(\widehat{\boldsymbol{w}}^\top - \boldsymbol{w})^\top(\widehat{\boldsymbol{\Sigma}}^*_0\boldsymbol{\beta} - \widehat{\boldsymbol{\Sigma}}^*_{1,*m})|^2] = o_{\mathbb{P}}(1)$.

Because $\boldsymbol{\Sigma}_0\boldsymbol{\beta} - \boldsymbol{\Sigma}_{1,*m} = \boldsymbol{0}$, we have

$$\sqrt{b\ell}(\widehat{\boldsymbol{w}} - \boldsymbol{w})^\top(\widehat{\boldsymbol{\Sigma}}^*_0\boldsymbol{\beta} - \widehat{\boldsymbol{\Sigma}}^*_{1,*m}) = \sqrt{b\ell}(\widehat{\boldsymbol{w}} - \boldsymbol{w})^\top((\widehat{\boldsymbol{\Sigma}}^*_0 - \boldsymbol{\Sigma}_0)\boldsymbol{\beta} - (\widehat{\boldsymbol{\Sigma}}^*_{1,*m} - \boldsymbol{\Sigma}_{1,*m})).$$

Therefore, it is sufficient to show that

$$\mathbb{E}^*[|\sqrt{b\ell}(\widehat{\boldsymbol{w}} - \boldsymbol{w})^\top(\widehat{\boldsymbol{\Sigma}}^*_0 - \boldsymbol{\Sigma}_0)\boldsymbol{\beta}|^2] = o_{\mathbb{P}}(1) \text{ and } \mathbb{E}^*[(\sqrt{b\ell}(\widehat{\boldsymbol{w}} - \boldsymbol{w})^\top(\widehat{\boldsymbol{\Sigma}}^*_{1,*m} - \boldsymbol{\Sigma}_{1,*m}))^2] = o_{\mathbb{P}}(1).$$

Because $|\sin(x) - \sin(y)| \le |x - y|$ for any $x, y \in \mathbb{R}$, and

$$|\widehat{\tau}^*_{0,jk} - \tau_{0,jk}| \le \frac{b\ell}{(b\ell-1)}|V^*_{0,jk} - V_{0,jk}| + \frac{1}{(b\ell-1)}|V_{0,jk}| \le \frac{b\ell}{(b\ell-1)}|V^*_{0,jk} - V_{0,jk}| + \frac{1}{(b\ell-1)},$$

we have

$$|\sqrt{b\ell}(\widehat{\boldsymbol{w}} - \boldsymbol{w})^\top(\widehat{\boldsymbol{\Sigma}}^*_0 - \boldsymbol{\Sigma}_0)\boldsymbol{\beta}|$$

$$
\begin{aligned}
=&\left|\sqrt{b\ell}\sum_{j\in S_{\boldsymbol{w}},k\in S}(\widehat{w}_j-w_j)\left(\sin\left(\frac{\pi}{2}\widehat{\tau}^*_{0,jk}\right)-\sin\left(\frac{\pi}{2}\tau_{0,jk}\right)\right)\beta_k\right|\\
\leq&\frac{\pi}{2}\sqrt{b\ell}\sum_{j\in S_{\boldsymbol{w}},k\in S}|\widehat{w}_j-w_j||\beta_w||\widehat{\tau}^*_{0,jk}-\tau_{0,jk}|\\
\leq&\frac{\pi}{2}\sqrt{b\ell}\frac{b\ell}{(b\ell-1)}\sum_{j\in S_{\boldsymbol{w}},k\in S}|\widehat{w}_j-w_j||\beta_k||V^*_{0,jk}-V_{0,jk}|+\frac{\pi}{2}\sqrt{b\ell}\frac{1}{(b\ell-1)}\sum_{j\in S_{\boldsymbol{w}},k\in S}|\widehat{w}_j-w_j||\beta_k|\\
\leq&\frac{\pi}{2}\sqrt{b\ell}\frac{b\ell}{(b\ell-1)}\sum_{j\in S_{\boldsymbol{w}},k\in S}|\widehat{w}_j-w_j||\beta_k||V^*_{0,jk}-V_{0,jk}|+\frac{\pi}{2}\sqrt{b\ell}\frac{1}{(b\ell-1)}\|\widehat{\boldsymbol{w}}-\boldsymbol{w}\|_1\|\boldsymbol{\beta}\|_1.
\end{aligned}
$$

This implies

$$
\begin{aligned}
&\mathbb{E}^*[|\sqrt{b\ell}(\widehat{\boldsymbol{w}}-\boldsymbol{w})^\top(\widehat{\boldsymbol{\Sigma}}^*_0-\boldsymbol{\Sigma}_0)\boldsymbol{\beta}|^2]\\
\leq&\frac{\pi^2}{2}(b\ell)\left(\frac{b\ell}{(b\ell-1)}\right)^2\mathbb{E}^*\left[\left(\sum_{j\in S_{\boldsymbol{w}},k\in S}|\widehat{\boldsymbol{w}}_j-\boldsymbol{w}_j||\boldsymbol{\beta}_k||V^*_{0,jk}-V_{0,jk}|\right)^2\right]+\frac{\pi^2}{2}(b\ell)\frac{1}{(b\ell-1)^2}\|\widehat{\boldsymbol{w}}-\boldsymbol{w}\|_1^2\|\boldsymbol{\beta}\|_1^2.
\end{aligned}
$$

First, we have

$$
(b\ell)\frac{1}{(b\ell-1)^2}\|\widehat{\boldsymbol{w}}-\boldsymbol{w}\|_1^2\|\boldsymbol{\beta}\|_1^2=\|\boldsymbol{\beta}\|_1^2O_{\mathbb{P}}\left(\frac{s_{\boldsymbol{w}}^2\log(ed)}{n^2}\right)=o_{\mathbb{P}}(1)
$$

Second, by doing a similar calculation as in Step 1 in Lemma A.4, one has

$$
\begin{aligned}
&(b\ell)\left(\frac{b\ell}{(b\ell-1)}\right)^2\mathbb{E}^*\left[\left(\sum_{j\in S_{\boldsymbol{w}},k\in S}|\widehat{\boldsymbol{w}}_j-\boldsymbol{w}_j||\boldsymbol{\beta}_k||V^*_{0,jk}-V_{0,jk}|\right)^2\right]\\
\leq&(b\ell)\left(\frac{b\ell}{(b\ell-1)}\right)^2\|\widehat{\boldsymbol{w}}-\boldsymbol{w}\|_1\|\boldsymbol{\beta}\|_1\left\{\sum_{j\in S_{\boldsymbol{w}},k\in S}|\widehat{\boldsymbol{w}}_j-\boldsymbol{w}_j||\boldsymbol{\beta}_k|\mathbb{E}^*[(V^*_{0,jk}-V_{0,jk})^2]\right\}.
\end{aligned}
$$

Note that Lemma A.11 in Section A.2.5 implies

$$
\left|\mathbb{E}^*[(V^*_{0,jk}-V_{0,jk})^2]-\frac{b\ell(b\ell-\ell)(b\ell-2\ell)(b\ell-3\ell)}{(b\ell)^4}(\widehat{V}_{0,jk}-V_{0,jk})^2\right|\leq\frac{50\ell}{b\ell}.
$$

Therefore, by combining the above three equations, we have

$$
\begin{aligned}
&\mathbb{E}^*[|\sqrt{b\ell}(\widehat{\boldsymbol{w}}-\boldsymbol{w})^\top(\widehat{\boldsymbol{\Sigma}}^*_0-\boldsymbol{\Sigma}_0)\boldsymbol{\beta}|^2]\\
\leq&(b\ell)\left(\frac{b\ell}{(b\ell-1)}\right)^2\|\widehat{\boldsymbol{w}}-\boldsymbol{w}\|_1\|\boldsymbol{\beta}\|_1\left\{\sum_{j\in S_{\boldsymbol{w}},k\in S}|\widehat{\boldsymbol{w}}_j-\boldsymbol{w}_j||\boldsymbol{\beta}_k|\mathbb{E}^*[(V^*_{0,jk}-V_{0,jk})^2]\right\}+o_{\mathbb{P}}(1)\\
\leq&(b\ell)\left(\frac{b\ell}{(b\ell-1)}\right)^2\|\widehat{\boldsymbol{w}}-\boldsymbol{w}\|_1\|\boldsymbol{\beta}\|_1\left\{\sum_{j\in S_{\boldsymbol{w}},k\in S}|\widehat{\boldsymbol{w}}_j-\boldsymbol{w}_j||\boldsymbol{\beta}_k|\frac{b(b-1)(b-2)(b-3)\ell^4}{(b\ell)^4}(\widehat{V}_{0,jk}-V_{0,jk})^2\right\}\\
&+(b\ell)\left(\frac{b\ell}{(b\ell-1)}\right)^2\|\widehat{\boldsymbol{w}}-\boldsymbol{w}\|_1\|\boldsymbol{\beta}\|_1\left\{\sum_{j\in S_{\boldsymbol{w}},k\in S}|\widehat{\boldsymbol{w}}_j-\boldsymbol{w}_j||\boldsymbol{\beta}_k|50\frac{\ell}{b\ell}\right\}+o_{\mathbb{P}}(1)
\end{aligned}
$$

$$
\begin{aligned}
=&\frac{b(b-1)(b-2)(b-3)\ell^4}{(b\ell)^4}\left(\frac{b\ell}{(b\ell-1)}\right)^2 (b\ell)\|\widehat{\boldsymbol{w}}-\boldsymbol{w}\|_1^2\|\boldsymbol{\beta}\|_1^2(\max_{j,k}\{|\widehat{V}_{0,jk}-V_{0,jk}|\})^2 \\
&+50\left(\frac{b\ell}{(b\ell-1)}\right)^2 \ell\|\widehat{\boldsymbol{w}}-\boldsymbol{w}\|_1^2\|\boldsymbol{\beta}\|_1^2+o_{\mathbb{P}}(1) \\
=&O(n)O_p\left(\frac{s_{\boldsymbol{w}}^2\log(ed)}{n}\right)\|\boldsymbol{\beta}\|_1^2 O_{\mathbb{P}}\left(\frac{\log(ed)}{n}\right)+O(\sqrt{n})O_{\mathbb{P}}\left(\frac{s_{\boldsymbol{w}}^2\log(ed)}{n}\right)\|\boldsymbol{\beta}\|_1^2+o_{\mathbb{P}}(1) \\
=&\|\boldsymbol{\beta}\|_1^2 O_{\mathbb{P}}\left(\frac{s_{\boldsymbol{w}}^2(\log(ed))^2}{n}\right)+\|\boldsymbol{\beta}\|_1^2 O_{\mathbb{P}}\left(\frac{s_{\boldsymbol{w}}^2\log(ed)}{\sqrt{n}}\right)+o_{\mathbb{P}}(1) \\
=&o_{\mathbb{P}}(1),
\end{aligned}
$$

because $b\ell \asymp n$ and $\ell^{-1}+\ell^2/n=o(1)$ imply

$$
\frac{b(b-1)(b-2)(b-3)\ell^4}{(b\ell)^4}=O(1),\ \frac{\ell}{\sqrt{n}}=o(1),\ \text{and}\ \left(\frac{b\ell}{(b\ell-1)}\right)^2=O(1).
$$

Similarly, we can show $\mathbb{E}^*[(\sqrt{b\ell}(\widehat{\boldsymbol{w}}-\boldsymbol{w})^\top(\widehat{\boldsymbol{\Sigma}}_{1,*m}^*-\boldsymbol{\Sigma}_{1,*m}))^2]=o_{\mathbb{P}}(1)$. Therefore, we have $\mathbb{E}^*[|\sqrt{b\ell}(\widehat{\boldsymbol{w}}^\top-\boldsymbol{w})^\top(\widehat{\boldsymbol{\Sigma}}_0^*\boldsymbol{\beta}-\widehat{\boldsymbol{\Sigma}}_{1,*m}^*)|^2]=o_{\mathbb{P}}(1)$.

Step 2-2. We show $\mathbb{E}^*[|\sqrt{b\ell}\widehat{\boldsymbol{w}}^\top\widehat{\boldsymbol{\Sigma}}_0^*(\widehat{\boldsymbol{\beta}}(\theta)-\boldsymbol{\beta})|^2]=o_{\mathbb{P}}(1)$.

Because $\boldsymbol{w}^\top\boldsymbol{\Sigma}_0(\widehat{\boldsymbol{\beta}}(\theta)-\boldsymbol{\beta})=0$, we have

$$
\begin{aligned}
\sqrt{b\ell}\widehat{\boldsymbol{w}}^\top\widehat{\boldsymbol{\Sigma}}_0^*(\widehat{\boldsymbol{\beta}}(\theta)-\boldsymbol{\beta}) &= \sqrt{b\ell}(\widehat{\boldsymbol{w}}^\top\widehat{\boldsymbol{\Sigma}}_0^*-\boldsymbol{w}^\top\boldsymbol{\Sigma}_0)(\widehat{\boldsymbol{\beta}}(\theta)-\boldsymbol{\beta}) \\
&= \sqrt{b\ell}\widehat{\boldsymbol{w}}^\top(\widehat{\boldsymbol{\Sigma}}_0^*-\boldsymbol{\Sigma}_0)(\widehat{\boldsymbol{\beta}}(\theta)-\boldsymbol{\beta})+\sqrt{b\ell}(\widehat{\boldsymbol{w}}-\boldsymbol{w})^\top\boldsymbol{\Sigma}_0(\widehat{\boldsymbol{\beta}}(\theta)-\boldsymbol{\beta}).
\end{aligned}
$$

Then, Steps 2-2-1 and 2-2-2 below combined will show $\mathbb{E}^*[(\sqrt{b\ell}\widehat{\boldsymbol{w}}^\top(\widehat{\boldsymbol{\Sigma}}_0^*-\boldsymbol{\Sigma}_0)(\widehat{\boldsymbol{\beta}}(\theta)-\boldsymbol{\beta}))^2]=o_{\mathbb{P}}(1)$ and $E^*[(\sqrt{b\ell}(\widehat{\boldsymbol{w}}-\boldsymbol{w})^\top\boldsymbol{\Sigma}_0(\widehat{\boldsymbol{\beta}}(\theta)-\boldsymbol{\beta}))^2]=o_{\mathbb{P}}(1)$.

Step 2-2-1. We show that $\mathbb{E}^*[(\sqrt{b\ell}\widehat{\boldsymbol{w}}^\top(\widehat{\boldsymbol{\Sigma}}_0^*-\boldsymbol{\Sigma}_0)(\widehat{\boldsymbol{\beta}}(\theta)-\boldsymbol{\beta}))^2]=o_{\mathbb{P}}(1)$. We can do the same calculation as in Step 2-1 and obtain

$$
\begin{aligned}
&\mathbb{E}^*[(\sqrt{b\ell}\widehat{\boldsymbol{w}}^\top(\widehat{\boldsymbol{\Sigma}}_0^*-\boldsymbol{\Sigma}_0)(\widehat{\boldsymbol{\beta}}(\theta)-\boldsymbol{\beta}))^2] \\
\leq&\frac{b(b-1)(b-2)(b-3)\ell^4}{(b\ell)^4}\left(\frac{b\ell}{(b\ell-1)}\right)^2 (b\ell)\|\widehat{\boldsymbol{w}}\|_1^2\|\widehat{\boldsymbol{\beta}}(\theta)-\boldsymbol{\beta}\|_1^2(\max_{j,k}\{|\widehat{V}_{0,jk}-V_{0,jk}|\})^2 \\
&+50\left(\frac{b\ell}{(b\ell-1)}\right)^2\ell\|\widehat{\boldsymbol{w}}\|_1^2\|\widehat{\boldsymbol{\beta}}(\theta)-\boldsymbol{\beta}\|_1^2+o_{\mathbb{P}}(1) \\
=&O(1)O(n)\left(\|\boldsymbol{w}\|_1^2+O_{\mathbb{P}}\left(\frac{s_{\boldsymbol{w}}^2\log(ed)}{n}\right)\right)O_{\mathbb{P}}\left(\frac{s^2\log(ed)}{n}\right)O_{\mathbb{P}}\left(\frac{\log(ed)}{n}\right) \\
&+O(1)O(\sqrt{n})\left(\|\boldsymbol{w}\|_1^2+O_{\mathbb{P}}\left(\frac{s_{\boldsymbol{w}}^2\log(ed)}{n}\right)\right)O_{\mathbb{P}}\left(\frac{s^2\log(ed)}{n}\right)+o_{\mathbb{P}}(1) \\
=&o_{\mathbb{P}}(1).
\end{aligned}
$$

Step 2-2-2, We show that $\mathbb{E}^*[(\sqrt{b\ell}(\widehat{\boldsymbol{w}}-\boldsymbol{w})^\top\boldsymbol{\Sigma}_0(\widehat{\boldsymbol{\beta}}(\theta)-\boldsymbol{\beta}))^2]=o_{\mathbb{P}}(1)$. This holds because

$$
|\mathbb{E}^*[(\sqrt{b\ell}(\widehat{\boldsymbol{w}}-\boldsymbol{w})^\top\boldsymbol{\Sigma}_0(\widehat{\boldsymbol{\beta}}(\theta)-\boldsymbol{\beta}))^2]|\leq(b\ell)\|\widehat{\boldsymbol{w}}-\boldsymbol{w}\|_1^2\|(\widehat{\boldsymbol{\beta}}(\theta)-\boldsymbol{\beta})\|_1^2=O_{\mathbb{P}}\left(\frac{s_{\boldsymbol{w}}^2 s^2(\log(ed))^2}{n}\right)=o_{\mathbb{P}}(1).
$$

By combining Steps 1 and 2, we have the desired result.

A.2.5 Technical Lemmas used in Section 3

The following lemmas apply Dehling and Wendler (2010) to our case. In the following we use $\mathbb{1}_A$ as a shorthand of $\mathbb{1}(x \in A)$.

Lemma A.5 (Theorem 2 in Bradley (1983))**.** Suppose $\boldsymbol{R}$ and $\boldsymbol{W}$ are random variables taking their values in Borel spaces S_1 and S_2 respectively, and suppose U is a uniform-$[0,1]$ random variable independent of $(\boldsymbol{R}, \boldsymbol{W})$. Suppose N is a positive integer and $H = \{H_1, H_2, \ldots, H_N\}$ is a measurable partition of S_2. Then, there exists an S_2-valued random variable $\breve{\boldsymbol{W}} = f(\boldsymbol{R}, \boldsymbol{W}, U)$, where f is a measurable function from $S_1 \times S_2 \times [0,1]$ into S_2, such that

1. $\breve{\boldsymbol{W}}$ is independent of $\boldsymbol{W}$;
2. the probability distributions of $\breve{\boldsymbol{W}}$ and W on S_2 are identical;
3. $\mathbb{P}(\breve{\boldsymbol{W}}$ and $\boldsymbol{W}$ are not elements of the same $H_i \in H) \leq (8N)^{1/2}\alpha(\sigma(\boldsymbol{R}), \sigma(\boldsymbol{W}))$.

The following is Bradley's lemma with $S_2 = \mathbb{R}^2$ by following the proof of Theorem 3 in Bradley (1983), which presents Bradley's lemma when $S_2 = \mathbb{R}$.

Lemma A.6. Suppose $\boldsymbol{R}$ and $\boldsymbol{W}$ are random variables and $\boldsymbol{W} = (W_1, W_2)^\top \in \mathbb{R}^2$. Let's denote $M = \max\{\|W_1\|_\gamma, \|W_2\|_\gamma\}$. Then there exists a random variable $\breve{\boldsymbol{W}}$ such that

1. $\breve{\boldsymbol{W}}$ is independent of $\boldsymbol{R}$;
2. the probability distributions of $\breve{\boldsymbol{W}}$ and $\boldsymbol{W}$ on $\mathbb{R}^2$ are identical;
3. For any $0 < q \leq M$,

$$\mathbb{P}(\|\boldsymbol{W} - \breve{\boldsymbol{W}}\|_\infty \geq q) \leq 50 \left(\frac{M}{q}\right)^{\frac{\gamma}{\gamma+1}} \alpha(\sigma(\boldsymbol{R}), \sigma(\boldsymbol{W}))^{\frac{\gamma}{\gamma+1}}.$$

Proof. Following the proof of Theorem 3 in Bradley (1983), let $\alpha = \alpha(\sigma(\boldsymbol{R}), \sigma(\boldsymbol{W}))$, $M = \max\{\|W_1\|_\gamma, \|W_2\|_\gamma\}$, and $x = [(1/\alpha)(M/q)^\gamma, \max\{\gamma, 1\}]^{1/(\gamma+1)}$. Note that $x \geq 1$ by construction. Further let m be such that $x \leq m \leq 2x$. We construct the rectangular H_{ij} for $-m \leq i, j \leq m$ and $H_{m+1,m+1}$ as follows:

$$H_{ij} = [xi - q/2, xi + q/2) \times [xj - q/2, xj + q/2)$$

for $-m \leq i, j \leq m$, and $H_{m+1,m+1} = \left(\bigcup_{-m \leq i,j \leq m} H_{ij}\right)^c$. Then, we can set H as the union of $H_{m+1,m+1}$ and $\bigcup_{-m \leq i,j \leq m} H_{ij}$. From Lemma A.5 , we have that the random variable $\breve{\boldsymbol{W}} = (\breve{W}_1, \breve{W}_2)^\top$ is independent of $\boldsymbol{R}$, has the same distribution as $\boldsymbol{W}$, and

$$\mathbb{P}(\breve{\boldsymbol{W}} \text{ and } \boldsymbol{W} \text{ are not elements of the same } H_{ij} \in H) \leq 4((2m+1)^2 + 1)^{1/2}\alpha.$$

Therefore,

$$
\begin{aligned}
\mathbb{P}(\|\boldsymbol{W}-\breve{\boldsymbol{W}}\|_\infty \geq q) &\leq 4((2m+1)^2+1)^{1/2}\alpha + P(|W_1| \geq mq + q/2) + \mathbb{P}(|Y_2| \geq mq + q/2) \\
&\quad + \mathbb{P}(|\breve{W}_1| \geq mq + q/2) + \mathbb{P}(|\breve{W}_2| \geq mq + q/2) \\
&= 4((2m+1)^2+1)^{1/2}\alpha + 2\mathbb{P}(|W_1| \geq mq + q/2) + 2\mathbb{P}(|W_2| \geq mq + q/2) \\
&\leq 24x\alpha + 2\mathbb{P}(|W_1| \geq xq) + 2\mathbb{P}(|W_2| \geq xq) \\
&\leq 24x\alpha + 2\left(\frac{\|W_1\|_\gamma}{q}\right)^\gamma x^{-\gamma} + 2\left(\frac{\|W_2\|_\gamma}{q}\right)^\gamma x^{-\gamma} \\
&\leq 24x\alpha + 4\left(\frac{M}{q}\right)^\gamma x^{-\gamma} \leq C_\gamma \left(\frac{M}{q}\right)^{\frac{\gamma}{\gamma+1}} \alpha^{\frac{\gamma}{\gamma+1}} \\
&\leq 50\left(\frac{M}{q}\right)^{\frac{\gamma}{\gamma+1}} \alpha^{\frac{\gamma}{\gamma+1}},
\end{aligned}
$$

where $C_\gamma := 2 + 24[\max\{\gamma, 1\}]^{1/(\gamma+1)}$. The last inequality comes from the last sentence of the proof of Theorem 3 in Bradley (1983). □

Lemma A.7 ($\mathbb{P}$-Lipschitz continuity of $f(\cdot,\cdot)$)**.** Suppose all assumptions in Proposition 2.2 hold. Then, there exists a constant L only depending on $(C_{\mathbf{A}}, c_{\boldsymbol{E}}, C_{\boldsymbol{E}})$ such that

$$\mathbb{E}[|f(\boldsymbol{W},\boldsymbol{R}) - f(\boldsymbol{W}',\boldsymbol{R})|\mathbb{1}_{\|\boldsymbol{W}-\boldsymbol{W}'\|_\infty \leq \epsilon}] \leq L\epsilon$$

for every $\epsilon > 0$, every pair $(\boldsymbol{W}^\top, \boldsymbol{R}^\top)^\top$ with the common distribution $\mathbb{P}_{\boldsymbol{U}_{1,\{j,k\}}, \boldsymbol{U}_{t,\{j,k\}}}$ for a $t \in \mathbb{N}$ or the product measure $\mathbb{P}_{\boldsymbol{U}_{1,jk}} \times \mathbb{P}_{\boldsymbol{U}_{1,jk}}$, and $(\boldsymbol{W}'^\top, \boldsymbol{R}^\top)^\top$ also with one of these common distributions. Here, $\boldsymbol{W}$, $\boldsymbol{W}'$, and $\boldsymbol{R}$ are in $\mathbb{R}^2$.

Proof. Because $|\text{sign}(\cdot)| \leq 1$, we have

$$|f(\boldsymbol{W},\boldsymbol{R}) - f(\boldsymbol{W}',\boldsymbol{R})| \leq |\operatorname{sign}(W_1 - R_1) - \operatorname{sign}(W_1' - R_1)| + |\operatorname{sign}(W_2 - R_2) - \operatorname{sign}(W_2' - R_2)|.$$

We then have

$$
\begin{aligned}
&\mathbb{E}[|f(\boldsymbol{W},\boldsymbol{R}) - f(\boldsymbol{W}',\boldsymbol{R})|\mathbb{1}_{\|\boldsymbol{W}-\boldsymbol{W}'\|_\infty \leq \epsilon}] \\
\leq& \mathbb{E}[|\operatorname{sign}(W_1 - R_1) - \operatorname{sign}(W_1' - R_1)|\mathbb{1}_{\|\boldsymbol{W}-\boldsymbol{W}'\|_\infty \leq \epsilon}] + \mathbb{E}[|\operatorname{sign}(W_2 - R_2) - \operatorname{sign}(X_2' - R_2)|\mathbb{1}_{\|\boldsymbol{W}-\boldsymbol{W}'\|_\infty \leq \epsilon}] \\
\leq& \mathbb{E}[|\operatorname{sign}(W_1 - R_1) - \operatorname{sign}(W_1' - R_1)|\mathbb{1}_{|W_1 - W_1'| \leq \epsilon}] + \mathbb{E}[|\operatorname{sign}(W_2 - R_2) - \operatorname{sign}(W_2' - R_2)|\mathbb{1}_{|W_2 - R_2'| \leq \epsilon}] \\
\leq& 2\mathbb{P}(|W_1 - R_1| \leq \epsilon) + 2\mathbb{P}(|W_2 - R_2| \leq \epsilon).
\end{aligned}
$$

The last inequality holds because $|x - x'| \leq \epsilon$ and $|x - y| > \epsilon$ imply $|\text{sign}(x-y) - \text{sign}(x'-y)| = 0$ for any real numbers x, x', and y . Under the assumptions in Proposition 2.2, $\mathbb{P}(|W_1 - R_1| \leq \epsilon) \leq D_1\epsilon$ for some absolute $D_1 > 0$. This is because of the following reasons:

- Note that the density of $W_1 - R_1$ is the density of $U_{1,j} - U_{t,j}$. When the distribution of $(\boldsymbol{W}^\top, \boldsymbol{R}^\top)^\top$ is from $\mathbb{P}_{\boldsymbol{U}_{1,\{j,k\}}} \times \mathbb{P}_{\boldsymbol{U}_{1,\{j,k\}}}$, W_1 and R_1 are independent of each other, so that the density function of $W_1 - R_1$ is bounded above under the assumptions. Therefore, let's consider the case when $(\boldsymbol{W}^\top, \boldsymbol{R}^\top)^\top$ is from $\mathbb{P}_{\boldsymbol{U}_{1,\{j,k\}}, \boldsymbol{U}_{t,\{j,k\}}}$;
- Because $\{\boldsymbol{Z}_t\}_{t=1}^n$ are jointly Gaussian, the maximum of the values of the density functions of $Z_{1,j} - Z_{t,j}$ is determined by the variance of $Z_{1,j} - Z_{t,j}$;
- Under the assumptions in Proposition 2.2, we can show that the variance of $Z_{1,j} - Z_{t,j}$ is uniformly bounded below using Lemma A.3. This implies that the density functions are bounded above by a constant D_1 which only depends on $(C_{\mathbf{A}}, c_{\boldsymbol{E}}, C_{\boldsymbol{E}})$;
- Therefore, $\mathbb{P}(|W_1 - R_1| \leq \epsilon) \leq D_1\epsilon$.

By the same logic, we have $\mathbb{P}(|W_2 - R_2| \leq \epsilon) \leq D_1\epsilon$. Therefore,

$$\mathbb{E}[|f(\boldsymbol{W}, \boldsymbol{R}) - f(\boldsymbol{W}', \boldsymbol{R})| \mathbb{1}_{\|\boldsymbol{W} - \boldsymbol{W}'\|_\infty \leq \epsilon}] \leq 2\mathbb{P}(|W_1 - R_1| \leq \epsilon) + 2\mathbb{P}(|W_2 - R_2| \leq \epsilon) \leq 4D_1\epsilon = L\epsilon,$$

where $L := 4D_1$. That is, $f(\cdot, \cdot)$ is $\mathbb{P}$ -Lipschitz-continuous with constant $L = 4D_1$. □

Note A.1. The proof of Lemma 3.3 in Dehling and Wendler (2010) implies that $g_1(\cdot)$ is $\mathbb{P}$-Lipschitz continuous with constant L and $g_2(\cdot, \cdot)$ is also $\mathbb{P}$-Lipschitz-continuous with constant $2L$.

In the following, we will use $\boldsymbol{W}_t$ to represent $(U_{t,j}, U_{t,k})^\top$, and $\alpha_{\boldsymbol{U}}(\cdot)$ as the α-mixing coefficient of the process $\{\boldsymbol{U}_t = (\boldsymbol{Z}_t^\top, \boldsymbol{Z}_{t+1}^\top)^\top\}_{t \in \mathbb{Z}}$.

The second lemma is to check if Lemma 3.3 in Dehling and Wendler (2010) holds.

Lemma A.8. Suppose all assumptions in Proposition 2.2 hold. Then, for some constants C_7 and C_8 only depending on $(C_{\mathbf{A}}, c_{\boldsymbol{E}}, C_{\boldsymbol{E}})$, we have

$$|\mathbb{E}[g_1(\boldsymbol{W}_{i_1}) g_1(\boldsymbol{W}_{i_2}) g_1(\boldsymbol{W}_{i_3}) g_1(\boldsymbol{W}_{i_4})]| \leq C_7 \alpha_{\boldsymbol{U}}(m)^{2/5},$$
$$\mathbb{E}[g_2(\boldsymbol{W}_{i_1}, \boldsymbol{W}_{i_2}) g_2(\boldsymbol{W}_{i_3}, \boldsymbol{W}_{i_4})]| \leq C_8 \alpha_{\boldsymbol{U}}(m)^{2/5},$$

where $m := \max\{a_2 - a_1, a_4 - a_3\}$ and $a_1, \ldots, a_4$ are the increasingly rearranged numbers of $i_1, \ldots, i_4$.

Proof. The proof follows the proof of Lemma 3.3 in Dehling and Wendler (2010) using Lemma A.6. In this proof, we will only look at the case where $i_1 < i_2 \leq i_3 < i_4$ with $m_1 = i_2 - i_1 = \max\{i_2 - i_1, i_4 - i_3\}$. Other cases could be handled similarly.

By Lemma A.6, there exists a random vector $\breve{\boldsymbol{W}}_{i_1} \in \mathbb{R}^2$ such that (a) $\breve{\boldsymbol{W}}_{i_1}$ has the same distribution as $\boldsymbol{W}_{i_1}$, (b) $\breve{\boldsymbol{W}}_{i_1}$ is independent of $\boldsymbol{W}_{i_2} \ldots \boldsymbol{W}_{i_4}$, and (c) when $0 \leq \epsilon \leq 1$, we have $\mathbb{P}(\|\boldsymbol{W}_{i_1} - \breve{\boldsymbol{W}}_{i_1}\|_\infty \geq \epsilon) \leq 50\epsilon^{-2/3}\alpha_{\boldsymbol{U}}(m)^{2/3}$. Then, $\mathbb{E}[g_1(\breve{\boldsymbol{W}}_{i_1}) g_1(\boldsymbol{W}_{i_2}) g_1(\boldsymbol{W}_{i_3}) g_1(\boldsymbol{W}_{i_4})] = 0$. Because $|g_1(\cdot)| \leq 2$ and Lemma A.7 implies $g_1(\cdot)$ is $\mathbb{P}$-Lipchitz continuous with L, we have

$$|\mathbb{E}[g_1(\boldsymbol{W}_{i_1}) g_1(\boldsymbol{W}_{i_2}) g_1(\boldsymbol{W}_{i_3}) g_1(\boldsymbol{W}_{i_4})]|$$

$$
\begin{aligned}
&= \left| \mathbb{E}\left[[g_1(\boldsymbol{W}_{i_1}) - g_1(\breve{\boldsymbol{W}}_{i_1})] g_1(\boldsymbol{W}_{i_2}) g_1(\boldsymbol{W}_{i_3}) g_1(\boldsymbol{W}_{i_4}) \right] \right| \\
&\leq 8\mathbb{E}[|g_1(\boldsymbol{W}_{i_1}) - g_1(\breve{\boldsymbol{W}}_{i_1})|] \\
&= 8\mathbb{E}[|g_1(\boldsymbol{W}_{i_1}) - g_1(\breve{\boldsymbol{W}}_{i_1})| \mathbb{1}_{\|\boldsymbol{W}_{i_1} - \breve{\boldsymbol{W}}_{i_1}\|_\infty \geq \epsilon}] + 8\mathbb{E}[|g_1(\boldsymbol{W}_{i_1}) - g_1(\breve{\boldsymbol{W}}_{i_1})| \mathbb{1}_{\|\boldsymbol{W}_{i_1} - \breve{\boldsymbol{W}}_{i_1}\|_\infty \leq \epsilon}] \\
&\leq 32\mathbb{P}(\|\boldsymbol{W}_{i_1} - \breve{\boldsymbol{W}}_{i_1}\|_\infty \geq \epsilon) + 8L\epsilon \\
&\leq 1600\epsilon^{-2/3}\alpha_{\boldsymbol{U}}(m)^{2/3} + 8\widetilde{L}\epsilon,
\end{aligned}
$$

where $\widetilde{L} := \max\{L, 1\}$. Picking $\epsilon = \widetilde{L}^{-3/5}\alpha_{\boldsymbol{U}}(m)^{2/5}$, we have

$$
|\mathbb{E}[g_1(\boldsymbol{W}_{i_1}) g_1(\boldsymbol{W}_{i_2}) g_1(\boldsymbol{W}_{i_3}) g_1(\boldsymbol{W}_{i_4})]| \leq \widetilde{C}_7 \widetilde{L}^{2/5} \alpha_{\boldsymbol{U}}(m)^{2/5},
$$

where $\widetilde{C}_7$ is some absolute constant. Using the same approach, one could obtain

$$
|\mathbb{E}[g_2(\boldsymbol{W}_{i_1}, \boldsymbol{W}_{i_2}) g_2(\boldsymbol{W}_{i_3}, \boldsymbol{W}_{i_4})]| \leq \widetilde{C}_8 \widetilde{L}^{2/5} \alpha_{\boldsymbol{U}}(m)^{2/5},
$$

where $\widetilde{C}_8$ is some absolute constant. These conclude the proof. □

Lemma A.9. Suppose all assumptions in Proposition 2.2 hold. Suppose that among $i_1, i_2, \ldots, i_8$, at most two of them are identical. Then, for some constant C which only depends on $(C_{\mathbf{A}}, c_{\boldsymbol{E}}, C_{\boldsymbol{E}})$, we have

$$
|\mathbb{E}[g_2(\boldsymbol{W}_{i_1}, \boldsymbol{W}_{i_2}) g_2(\boldsymbol{W}_{i_3}, \boldsymbol{W}_{i_4}) g_2(\boldsymbol{W}_{i_5}, \boldsymbol{W}_{i_6}) g_2(\boldsymbol{W}_{i_7}, \boldsymbol{W}_{i_8})]| \leq C\alpha_{\boldsymbol{U}}(m)^{2/5},
$$

where $m := \max\{a_2 - a_1, a_8 - a_7\}$ and $a_1, \ldots, a_8$ are the increasingly rearranged numbers of $i_1, \ldots, i_8$.

Proof. The proof is the same as Lemma A.8, and Lemma 3.3 in Dehling and Wendler (2010).

Note that $m \geq 1$ because we assume that at most two of i's are identical.

Without loss of generality, we can assume that $a_1 = i_1$; $a_8 = i_8$; $i_1 \leq i_2$; $i_3 \leq i_4$; $i_5 \leq i_6$; and $i_7 \leq i_8$ because $g_2(\boldsymbol{W}_1, \boldsymbol{W}_2) = g_2(\boldsymbol{W}_2, \boldsymbol{W}_1)$ for any generic vectors $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$.

Suppose $m = a_2 - a_1 \geq a_8 - a_7$. From Lemma A.6, we have $\breve{\boldsymbol{W}}_{i_1} \in \mathbb{R}^2$ such that (a) $\breve{\boldsymbol{W}}_{i_1}$ has the same distribution as $\boldsymbol{W}_{i_1}$; (b) $\breve{\boldsymbol{W}}_{i_1}$ is independent of $\boldsymbol{W}_{i_2} \ldots \boldsymbol{W}_{i_8}$; and (c) when $0 \leq \epsilon \leq 1$, $\mathbb{P}(\|\boldsymbol{W}_{i_1} - \breve{\boldsymbol{W}}_{i_1}\|_\infty \geq \epsilon) \leq 50\epsilon^{-2/3}\alpha_{\boldsymbol{U}}(m)^{2/3}$. Then, we have $\mathbb{E}[g_2(\breve{\boldsymbol{W}}_{i_1}, \boldsymbol{W}_{i_2}) g_2(\boldsymbol{W}_{i_3}, \boldsymbol{W}_{i_4}) g_2(\boldsymbol{W}_{i_5}, \boldsymbol{W}_{i_6}) g_2(\boldsymbol{W}_{i_7}, \boldsymbol{W}_{i_8})] = 0$. Because $|g_2(\cdot)| \leq 4$, and Lemma A.7 implies that $g_2(\cdot)$ is $\mathbb{P}$ -Lipchitz continuous with $2L$, we have

$$
\begin{aligned}
&|\mathbb{E}[g_2(\boldsymbol{W}_{i_1}, \boldsymbol{W}_{i_2}) g_2(\boldsymbol{W}_{i_3}, \boldsymbol{W}_{i_4}) g_2(\boldsymbol{W}_{i_5}, \boldsymbol{W}_{i_6}) g_2(\boldsymbol{W}_{i_7}, \boldsymbol{W}_{i_8})]| \\
=& \left| \mathbb{E}\left[[g_2(\boldsymbol{W}_{i_1}, \boldsymbol{W}_{i_2}) - g_2(\breve{\boldsymbol{W}}_{i_1}, \boldsymbol{W}_{i_2})] g_2(\boldsymbol{W}_{i_3}, \boldsymbol{W}_{i_4}) g_2(\boldsymbol{W}_{i_5}, \boldsymbol{W}_{i_6}) g_2(\boldsymbol{W}_{i_7}, \boldsymbol{W}_{i_8}) \right] \right| \\
\leq& 64\mathbb{E}[|g_2(\boldsymbol{W}_{i_1}, \boldsymbol{W}_{i_2}) - g_2(\breve{\boldsymbol{W}}_{i_1}, \boldsymbol{W}_{i_2})|] \\
=& 64\mathbb{E}[|g_2(\boldsymbol{W}_{i_1}, \boldsymbol{W}_{i_2}) - g_2(\breve{\boldsymbol{W}}_{i_1}, \boldsymbol{W}_{i_2})| \mathbb{1}_{\|\boldsymbol{W}_{i_1} - \breve{\boldsymbol{W}}_{i_1}\|_\infty \geq \epsilon}] + 64\mathbb{E}[|g_2(\boldsymbol{W}_{i_1}, \boldsymbol{W}_{i_2}) - g_2(\breve{\boldsymbol{W}}_{i_1}, \boldsymbol{W}_{i_2})| \mathbb{1}_{\|\boldsymbol{W}_{i_1} - \breve{\boldsymbol{W}}_{i_1}\|_\infty \leq \epsilon}] \\
\leq& 512\mathbb{P}(\|\boldsymbol{W}_{i_1} - \breve{\boldsymbol{W}}_{i_1}\|_\infty \geq \epsilon) + 128L\epsilon \\
\leq& 25600\epsilon^{-2/3}\alpha_{\boldsymbol{U}}(m)^{2/3} + 128\widetilde{L}\epsilon,
\end{aligned}
$$

where $\widetilde{L} := \max\{L, 1\}$. When we pick $\epsilon = \widetilde{L}^{-3/5}\alpha_{\boldsymbol{U}}(m)^{2/5}$, we have

$$|\mathbb{E}[g_2(\boldsymbol{W}_{i_1}, \boldsymbol{W}_{i_2})g_2(\boldsymbol{W}_{i_3}, \boldsymbol{W}_{i_4})g_2(\boldsymbol{W}_{i_5}, \boldsymbol{W}_{i_6})g_2(\boldsymbol{W}_{i_7}, \boldsymbol{W}_{i_8})]| \leq 12928\widetilde{L}^{2/5}\alpha_{\boldsymbol{u}}(m)^{2/5}.$$

This completes the proof. □

Lemma A.10. Suppose all assumptions in Proposition 2.2 hold. Let's denote

$$U(g_1) = \frac{2}{n}\sum_{i=1}^{n} g_1(\boldsymbol{W}_i) \text{ and } U(g_2) = \frac{2}{n(n-1)}\sum_{i<j} g_2(\boldsymbol{W}_i, \boldsymbol{W}_j).$$

Then, for some constants C, C', and C'' which only depend on $(C_{\mathbf{A}}, c_{\boldsymbol{E}}, C_{\boldsymbol{E}}, \kappa_1, \kappa_2)$, we have $|\mathbb{E}[U^2(g_1)]| \leq Cn^{-2}$, $|\mathbb{E}[U^2(g_2)]| \leq C'^{-2}$, and

$$\begin{aligned}|\mathbb{E}[U^4(g_2)]| &= \left|\frac{16}{n^4(n-1)^4}\sum_{i_1<i_2}\sum_{i_3<i_4}\sum_{i_5<i_6}\sum_{i_7<i_8} E[g_2(\boldsymbol{W}_{i_1}, \boldsymbol{W}_{i_2})g_2(\boldsymbol{W}_{i_3}, \boldsymbol{W}_{i_4})g_2(\boldsymbol{W}_{i_5}, \boldsymbol{W}_{i_6})g_2(\boldsymbol{W}_{i_7}, \boldsymbol{W}_{i_8})]\right| \\ &\leq C''^{-2}.\end{aligned}$$

Proof. We only have to present the proof for $\mathbb{E}[U^4(g_2)] \leq C''^{-2}$ because we can prove $|\mathbb{E}[U^2(g_1)]| \leq Cn^{-2}$ and $|\mathbb{E}[U^2(g_2)]| \leq C'^{-2}$ by following the proof of Lemma 3.6 in Dehling and Wendler (2010). The proof is similar to Lemma 3 in Yoshihara (1976). Note that

$$|\mathbb{E}[U^2(g_2)]| \leq \frac{16}{n^4(n-1)^4}\sum_{1\leq i_1,\ldots,i_8\leq n}|\mathbb{E}[g_2(\boldsymbol{W}_{i_1}, \boldsymbol{W}_{i_2})g_2(\boldsymbol{W}_{i_3}, \boldsymbol{W}_{i_4})g_2(\boldsymbol{W}_{i_5}, \boldsymbol{W}_{i_6})g_2(\boldsymbol{W}_{i_7}, \boldsymbol{W}_{i_8})]|.$$

Case 1. Suppose more than two indices among $i_1, \ldots, i_8$ are identical. Then, we have at most n^6 possible cases for $i_1, \ldots, i_8$. Therefore, for some constant C''_1 which only depends on $(C_{\mathbf{A}}, c_{\boldsymbol{E}}, C_{\boldsymbol{E}})$,

$$\frac{16}{n^4(n-1)^4}\sum_{\substack{0\leq i_1,\ldots,i_8\leq n,\\ \text{at least three of } i\text{'s are identical}}}|\mathbb{E}[g_2(\boldsymbol{W}_{i_1}, \boldsymbol{W}_{i_2})g_2(\boldsymbol{W}_{i_3}, \boldsymbol{W}_{i_4})g_2(\boldsymbol{W}_{i_5}, \boldsymbol{W}_{i_6})g_2(\boldsymbol{W}_{i_7}, \boldsymbol{W}_{i_8})]| \leq C''_1 n^{-2},$$

because $|g_2(\cdot,\cdot)| \leq 4$.

Case 2. Suppose at most two indices are identical. Let $a_1 \leq a_2 \leq \ldots \leq a_8$ be ordered indices of $i_1, \ldots, i_8$ and let's denote $b_1 := a_2 - a_1$, $b_3 := a_4 - a_3$, $b_5 := a_6 - a_5$, and $b_7 := a_8 - a_7$. Let's set $m = \max\{b_1, b_7\}$. Note that $m \geq 1$ by construction. For a fixed m, we have at most $2 \times n^6 m$ number of combinations for the values of $a_1, \ldots, a_8$. This implies that for a fixed m, we have at most $2 \times 8! \times n^6 m$ number of combinations for the values of $i_1, \ldots i_8$. Here $k!$ means the factorial of a number $k \in \mathbb{N}$. Therefore, using Lemma A.9 , we have

$$\begin{aligned}&\frac{16}{n^4(n-1)^4}\sum_{\substack{0\leq i_1,\ldots,i_8\leq n,\\ \text{at most two of } i\text{'s are identical}}}|\mathbb{E}[g_2(\boldsymbol{W}_{i_1}, \boldsymbol{W}_{i_2})g_2(\boldsymbol{W}_{i_3}, \boldsymbol{W}_{i_4})g_2(\boldsymbol{W}_{i_5}, \boldsymbol{W}_{i_6})g_2(\boldsymbol{W}_{i_7}, \boldsymbol{W}_{i_8})]| \\ &\leq \frac{16}{n^4(n-1)^4}\widetilde{C}''_2\sum_{m=1}^{n} n^6 m\alpha_{\boldsymbol{U}}(m)^{2/5} \leq C''_2 n^{-2},\end{aligned}$$

for some constants $\widetilde{C}''_2$ and C''_2 where $\widetilde{C}''_2$ only depends on $(C_{\mathbf{A}}, c_{\boldsymbol{E}}, C_{\boldsymbol{E}})$, and C''_2 only depends on $(C_{\mathbf{A}}, c_{\boldsymbol{E}}, C_{\boldsymbol{E}}, \kappa_1, \kappa_2)$. Therefore, Cases 1 and 2 combined imply the desired result. □

Lemma A.11. The CBB bootstrap V-statistics $V^*_{0,jk}$ and $V^*_{1,jk}$, and V-statistics $V_{0,jk}$ and $V_{1,jk}$ using $\{\mathbf{Y}_t\}_{t=1}^{n-1}$ have the following relationships:

$$\left|\mathbb{E}^*[(V^*_{0,jk} - V_{0,jk})^2] - \frac{b\ell(b\ell-\ell)(b\ell-2\ell)(b\ell-3\ell)}{(b\ell)^4}(\widehat{V}_{0,jk} - V_{0,jk})^2\right| \le \frac{50\ell}{b\ell},$$

$$\left|\mathbb{E}^*[(V^*_{1,jk} - V_{1,jk})^2] - \frac{b\ell(b\ell-\ell)(b\ell-2\ell)(b\ell-3\ell)}{(b\ell)^4}(\widehat{V}_{1,jk} - V_{1,jk})^2\right| \le \frac{50\ell}{b\ell},$$

where $\mathbb{E}^*(\cdot)$ stands for the expectation operator under $\mathbb{P}^*$.

Proof. Because the proof is identical for $V^*_{0,jk}$ and $V^*_{1,jk}$, we will examine the case of $V^*_{0,jk}$ only. One can decompose $\mathbb{E}^*[(V^*_{0,jk} - V_{0,jk})^2]$ as

$$\begin{aligned}
&\mathbb{E}^*[(V^*_{0,jk} - V_{0,jk})^2] \\
=&\frac{1}{(b\ell)^4}\sum_{1\le i_1,i_2,i_3,i_4\le b\ell}\mathbb{E}^*\left[g\left(\begin{pmatrix}Z^*_{i_1,j}\\Z^*_{i_1,k}\end{pmatrix},\begin{pmatrix}Z^*_{i_2,j}\\Z^*_{i_2,k}\end{pmatrix}\right)g\left(\begin{pmatrix}Z^*_{i_3,j}\\Z^*_{i_3,k}\end{pmatrix},\begin{pmatrix}Z^*_{i_4,j}\\Z^*_{i_4,k}\end{pmatrix}\right)\right] \\
=&\frac{1}{(b\ell)^4}\sum_{\substack{1\le i_1,i_2,i_3,i_4\le b\ell,\\ \text{all } i\text{'s are in different blocks}}}\mathbb{E}^*\left[g\left(\begin{pmatrix}Z^*_{i_1,j}\\Z^*_{i_1,k}\end{pmatrix},\begin{pmatrix}Z^*_{i_2,j}\\Z^*_{i_2,k}\end{pmatrix}\right)g\left(\begin{pmatrix}Z^*_{i_3,j}\\Z^*_{i_3,k}\end{pmatrix},\begin{pmatrix}X^*_{i_4,j}\\X^*_{i_4,k}\end{pmatrix}\right)\right] \\
&+\frac{1}{(b\ell)^4}\sum_{\substack{1\le i_1,i_2,i_3,i_4\le b\ell,\\ \text{at least two of } i\text{'s is in the same block}}}\mathbb{E}^*\left[g\left(\begin{pmatrix}Z^*_{i_1,j}\\Z^*_{i_1,k}\end{pmatrix},\begin{pmatrix}Z^*_{i_2,j}\\Z^*_{i_2,k}\end{pmatrix}\right)g\left(\begin{pmatrix}Z^*_{i_3,j}\\Z^*_{i_3,k}\end{pmatrix},\begin{pmatrix}Z^*_{i_4,j}\\Z^*_{i_4,k}\end{pmatrix}\right)\right] \\
=&\frac{b(b-1)(b-2)(b-3)\ell^4}{(b\ell)^4}(\widehat{V}_{0,jk} - V_{0,jk})^2 \\
&+\frac{1}{(b\ell)^4}\sum_{\substack{1\le i_1,i_2,i_3,i_4\le b\ell,\\ \text{at least two of } i\text{'s is in the same block}}}\mathbb{E}^*\left[g\left(\begin{pmatrix}Z^*_{i_1,j}\\Z^*_{i_1,k}\end{pmatrix},\begin{pmatrix}Z^*_{i_2,j}\\Z^*_{i_2,k}\end{pmatrix}\right)g\left(\begin{pmatrix}Z^*_{i_3,j}\\Z^*_{i_3,k}\end{pmatrix},\begin{pmatrix}Z^*_{i_4,j}\\Z^*_{i_4,k}\end{pmatrix}\right)\right].
\end{aligned}$$

The last equality holds because there are $b(b-1)(b-2)(b-3)\ell^4$ possibilities that i_1, i_2, i_3, and i_4 are in different blocks; and when i_1, i_2, i_3, and i_4 are in different blocks, similar to the proof of Lemma 3.7 in Dehling and Wendler (2010), we have

$$\begin{aligned}
&\mathbb{E}^*\left[g\left(\begin{pmatrix}Z^*_{i_1,j}\\Z^*_{i_1,k}\end{pmatrix},\begin{pmatrix}Z^*_{i_2,j}\\Z^*_{i_2,k}\end{pmatrix}\right)g\left(\begin{pmatrix}Z^*_{i_3,j}\\Z^*_{i_3,k}\end{pmatrix},\begin{pmatrix}Z^*_{i_4,j}\\Z^*_{i_4,k}\end{pmatrix}\right)\right] \\
=&\frac{1}{(n-1)^4}\sum_{1\le i_1,i_2,i_3,i_4\le n-1}\left[g\left(\begin{pmatrix}Z_{i_1,j}\\Z_{i_1,k}\end{pmatrix},\begin{pmatrix}Z_{i_2,j}\\Z_{i_2,k}\end{pmatrix}\right)g\left(\begin{pmatrix}Z_{i_3,j}\\Z_{i_3,k}\end{pmatrix},\begin{pmatrix}Z_{i_4,j}\\Z_{i_4,k}\end{pmatrix}\right)\right] \\
=&(\widehat{V}_{0,jk} - V_{0,jk})^2,
\end{aligned}$$

because $\begin{pmatrix}Z^*_{i_1,j}\\Z^*_{i_1,k}\end{pmatrix}, \begin{pmatrix}Z^*_{i_2,j}\\Z^*_{i_2,k}\end{pmatrix}, \begin{pmatrix}Z^*_{i_3,j}\\Z^*_{i_3,k}\end{pmatrix}$, and $\begin{pmatrix}Z^*_{i_4,j}\\Z^*_{i_4,k}\end{pmatrix}$ are independent under $\mathbb{P}^*$. Therefore, we have

$$\left|\mathbb{E}^*[(V^*_{0,jk} - V_{0,jk})^2] - \frac{b(b-1)(b-2)(b-3)\ell^4}{(b\ell)^4}(\widehat{V}_{0,jk} - V_{0,jk})^2\right|$$

$$= \left| \frac{1}{(b\ell)^4} \sum_{\substack{1 \le i_1, i_2, i_3, i_4 \le b\ell, \\ \text{at least one of } i\text{'s in the same block}}} \mathbb{E}^* \left[g\left(\begin{pmatrix} Z^*_{i_1,j} \\ Z^*_{i_1,k} \end{pmatrix}, \begin{pmatrix} Z^*_{i_2,j} \\ Z^*_{i_2,k} \end{pmatrix} \right) g\left(\begin{pmatrix} Z^*_{i_3,j} \\ Z^*_{i_3,k} \end{pmatrix}, \begin{pmatrix} Z^*_{i_4,j} \\ Z^*_{i_4,k} \end{pmatrix} \right) \right] \right|$$

$$\le 4 \frac{(b\ell)^4 - b(b-1)(b-2)(b-3)\ell^4}{(b\ell)^4} = 4 \frac{6b^3\ell^4 - 11b^2\ell^4 + 6b^2\ell^4}{(b\ell)^4} \le \frac{50b^3\ell^4}{(b\ell)^4} = \frac{50\ell}{b\ell},$$

because $|g(\cdot)| \le 2$. □

A.3 Proof in Section 5

A.3.1 Notation

For any square matrix $\mathbf{M}$, let $\rho(\mathbf{M})$ denote its spectral radius. Let

$$\overline{\mathbf{e}}_p := \begin{pmatrix} \mathbf{I}_d \\ \mathbf{0}_{(p-1)d\times d} \end{pmatrix}, \ \underline{\mathbf{e}}_p := \begin{pmatrix} \mathbf{0}_{(p-1)d\times d} \\ \mathbf{I}_d \end{pmatrix}, \ \mathbf{C}_p := \begin{pmatrix} \mathbf{A}_p \\ \mathbf{A}_{p-1} \\ \vdots \\ \mathbf{A}_1 \end{pmatrix}, \ \mathbf{Q}_p := \begin{pmatrix} \mathbf{I}_d & -\mathbf{A}_1 & -\mathbf{A}_2 & \dots & -\mathbf{A}_{p-1} \\ \mathbf{0}_{d\times d} & \mathbf{I}_d & -\mathbf{A}_1 & \ddots & -\mathbf{A}_{p-2} \\ \mathbf{0}_{d\times d} & \mathbf{0}_{d\times d} & \mathbf{I}_d & \ddots & -\mathbf{A}_{p-3} \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ \mathbf{0}_{d\times d} & \cdots & \mathbf{0}_{d\times d} & \mathbf{0}_{d\times d} & \mathbf{I}_d \end{pmatrix}$$

with $\overline{\mathbf{e}}_1 = \underline{\mathbf{e}}_1 = \mathbf{I}_d$ and $\mathbf{Q}_1 := \mathbf{I}_d$. We set $\mathfrak{Z}_t \equiv (\boldsymbol{Z}_t^\top, \boldsymbol{Z}_{t-1}^\top, \dots, \boldsymbol{Z}_{t-p+1}^\top)^\top \in \mathbb{R}^{pd}$, $\mathfrak{E}_t = (\boldsymbol{E}_t^\top, \mathbf{0}_{d\times(p-1)d})^\top \in \mathbb{R}^{pd}$,

$$\mathfrak{A}_p = \begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \mathbf{A}_3 & \dots & \mathbf{A}_p \\ \mathbf{I}_d & \mathbf{0}_{d\times d} & \mathbf{0}_{d\times d} & \dots & \mathbf{0}_{d\times d} \\ \mathbf{0}_{d\times d} & \mathbf{I}_d & \mathbf{0}_{d\times d} & \dots & \mathbf{0}_{d\times d} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0}_{d\times d} & \ddots & \ddots & \mathbf{I}_d & \mathbf{0}_{d\times d} \end{pmatrix} \in \mathbb{R}^{pd\times pd}, \ \text{and } \overline{\mathfrak{A}_p} := \begin{pmatrix} a_1 & a_2 & a_3 & \dots & a_p \\ 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & 1 & 0 \end{pmatrix} \in \mathbb{R}^{p\times p}.$$

For a pd by pd matrix $\mathfrak{A}_p$, we partition $\mathfrak{A}_p^k$ by p^2 blocks where each block is a d by d matrix. We denote by $[\mathfrak{A}_p^k]_{jm}$ the (j,m)-th block of $\mathfrak{A}_p^k$. Let $\overline{\mathfrak{A}_p^k}$ denote a p by p matrix where the (j,m)-th element $a_{k,jm}$ of $\overline{\mathfrak{A}_p^k}$ is defined as $a_{k,jm} = \|[\mathfrak{A}_p^k]_{jm}\|_2$. Then, we can write $\mathfrak{A}_p^k$ and $\overline{\mathfrak{A}_p^k}$ as

$$\mathfrak{A}_p^k = \begin{pmatrix} [\mathfrak{A}_p^k]_{11} & \cdots & [\mathfrak{A}_p^k]_{1p} \\ \vdots & \ddots & \vdots \\ [\mathfrak{A}_p^k]_{p1} & \cdots & [\mathfrak{A}_p^k]_{pp} \end{pmatrix} \quad \text{and} \quad \overline{\mathfrak{A}_p^k} = \begin{pmatrix} a_{k,11} & \cdots & a_{k,1p} \\ \vdots & \ddots & \vdots \\ a_{k,p1} & \cdots & a_{k,pp} \end{pmatrix}.$$

We have $\|\mathfrak{A}_p^k\|_2 \le \|\overline{\mathfrak{A}_p^k}\|_2 \le \|\overline{\mathfrak{A}_p}^k\|_2$ because for any vectors $\boldsymbol{v} = (\boldsymbol{v}_1^\top, \boldsymbol{v}_2^\top, \dots, \boldsymbol{v}_p^\top)^\top \in \mathbb{R}^{pd}$ and $\overline{\boldsymbol{v}} = (\|\boldsymbol{v}_1\|_2, \dots, \|\boldsymbol{v}_p\|_2)^\top \in \mathbb{R}^p$ where $\{\boldsymbol{v}_j\}_{j=1}^p$ is a $d \times 1$ vector, we have

$$\|\mathfrak{A}_p^k \boldsymbol{v}\|_2 \le \|\overline{\mathfrak{A}_p^k}\overline{\boldsymbol{v}}\|_2 \le \|\overline{\mathfrak{A}_p^{k-1}}\,\overline{\mathfrak{A}_p}\overline{\boldsymbol{v}}\|_2 \le \ldots \le \|\overline{\mathfrak{A}_p}^k \overline{\boldsymbol{v}}\|_2, \ \text{and } \|\boldsymbol{v}\|_2 = \|\overline{\boldsymbol{v}}\|_2.$$

A.3.2 Proof of Theorem 5.1

When $\{\boldsymbol{Z}_t\}_{t\in\mathbb{Z}}$ follows VAR(p) model (5.1), we know that $\{\mathfrak{Z}_t\}_{t\in\mathbb{Z}}$ follows VAR(1) model: $\mathfrak{Z}_t = \mathfrak{A}_p\mathfrak{Z}_{t-1} + \mathfrak{E}_t$ with $\boldsymbol{\Sigma}_0 = \mathrm{Var}(\mathfrak{Z}_t)$. Under Assumption M-p, the spectral radius of $\overline{\mathfrak{A}_p}$, $\rho(\overline{\mathfrak{A}_p})$, satisfies: $\rho(\overline{\mathfrak{A}_p}) \le C_{\overline{\mathfrak{A}_p}} < 1$ for some absolute constant $C_{\overline{\mathfrak{A}_p}}$.

By following the proof of Theorem 3.1 in Han and Wu (2019), we know the ϱ-mixing coefficient

$$\varrho(\sigma(\mathfrak{Z}_0), \sigma(\mathfrak{Z}_k)) \begin{cases} = 1 & \text{when } k < p, \\ \le \sqrt{(\lambda_{\max}(\boldsymbol{\Sigma}_0)/\lambda_{\min}(\boldsymbol{\Sigma}_0))}\|\mathfrak{A}_p^k\|_2 & \text{when } k \ge p. \end{cases}$$

Here $\lambda_{\max}(\boldsymbol{\Sigma}_0)/\lambda_{\min}(\boldsymbol{\Sigma}_0)$ is upper bounded by a positive constant that only depends on $(c_{\boldsymbol{E}}, C_{\boldsymbol{E}}, a_1, \ldots, a_p, p)$ by the following lemma.

Lemma A.12. Suppose $\{\boldsymbol{Z}_t\}_{t\in\mathbb{Z}}$ follows VAR(p) model (5.1) with finite and fixed p. Under Assumptions M-p and S, the following results hold: For the covariance matrix $\boldsymbol{\Omega}_j := \mathrm{Var}((\boldsymbol{Z}_t^\top, \ldots, \boldsymbol{Z}_{t+j}^\top)^\top)$ with $j \in \{0, 1, 2, \cdots\}$, we have $0 < c_p \le \lambda_{\min}(\mathrm{Var}(\boldsymbol{\Omega}_j)) \le \lambda_{\max}(\mathrm{Var}(\boldsymbol{\Omega}_j)) \le C_p < \infty$ for some constants c_p and C_p that only depend on $(c_{\boldsymbol{E}}, C_{\boldsymbol{E}}, a_1, \ldots, a_p, p)$.

Because p is finite and fixed, we only have to consider the case that $k \ge p$. Then,

$$\varrho(\sigma(\mathfrak{Z}_0), \sigma(\mathfrak{Z}_k)) \le \sqrt{\frac{\lambda_{\max}(\boldsymbol{\Sigma}_0)}{\lambda_{\min}(\boldsymbol{\Sigma}_0)}}\|\mathfrak{A}^k\|_2 \le \sqrt{\frac{\lambda_{\max}(\boldsymbol{\Sigma}_0)}{\lambda_{\min}(\boldsymbol{\Sigma}_0)}}\|\overline{\mathfrak{A}_p}^k\|_2$$

because $\|\mathfrak{A}_p^k\|_2 \le \|\overline{\mathfrak{A}_p^k}\|_2 \le \|\overline{\mathfrak{A}_p}^k\|_2$. When p is finite and fixed, Gelfand's formula implies that there exists a K_ϵ such that $\|\overline{\mathfrak{A}_p}^k\|_2 < (\rho(\overline{\mathfrak{A}_p}) + \epsilon)^k$ for all $k \ge K_\epsilon$, by choosing some $\epsilon > 0$ such that $\rho(\overline{\mathfrak{A}_p}) + \epsilon < C < 1$ in which C is a positive constant that only depends on $(a_1, \ldots, a_p)$. Then, when $k > \overline{K} = \max\{K_\epsilon, p\}$, we have

$$\varrho(\sigma(\mathfrak{Z}_0), \sigma(\mathfrak{Z}_k)) \le \sqrt{\frac{\lambda_{\max}(\boldsymbol{\Sigma}_0)}{\lambda_{\min}(\boldsymbol{\Sigma}_0)}}\|\overline{\mathfrak{A}_p}^k\|_2 < \sqrt{\frac{\lambda_{\max}(\boldsymbol{\Sigma}_0)}{\lambda_{\min}(\boldsymbol{\Sigma}_0)}}(\rho(\overline{\mathfrak{A}_p}) + \epsilon)^k.$$

Therefore, $\{\boldsymbol{Z}_t\}_{t\in\mathbb{Z}}$ is geometrically mixing with $\alpha(k; \{\boldsymbol{Z}_t\}_{t\in\mathbb{Z}}) \le \gamma_1 \exp(-\gamma_2 k)$, where γ_1 and γ_2 are positive constants that only depend on $(c_{\boldsymbol{E}}, C_{\boldsymbol{E}}, a_1, \ldots, a_p, p)$. This completes the proof.

A.3.3 Proof of Theorem 5.2

Because the proof of this theorem is similar to VAR(1) case, we will just sketch the proof. First, we replace Lemma A.3 and Theorem 2.2 with Lemma A.12 and Theorem 5.1, respectively. Then, we have the conclusion that is similar to Section 2 because $\{\mathfrak{Z}_t\}_{t\in\mathbb{Z}}$ follows a VAR(1) model when $\{\boldsymbol{Z}_t\}_{t\in\mathbb{Z}}$ follows VAR(p) model (5.1).

Regarding the inference, our estimator $\widetilde{\boldsymbol{\theta}}$ and the related equation $\widehat{\boldsymbol{S}}(\boldsymbol{v}) = \widehat{\mathbf{W}}^\top(\widehat{\boldsymbol{\Sigma}}_0\boldsymbol{v} - \widehat{\boldsymbol{\Sigma}}_1)$ are p-dimensional vectors, not scalar values. Therefore, we need appropriate norms and additional inequalities in order to prove the theorem. Regarding this issue, we have the following properties.

(i) Because p is finite and fixed, the choice of norm is not important. In particular, for a square matrix $\mathbf{M} \in \mathbb{R}^{p\times p}$, we have $\|\mathbf{M}\|_{\max} \leq \|\mathbf{M}\|_2 \leq p\|\mathbf{M}\|_{\max}$.

(ii) For any random variables $\boldsymbol{R}_t \in \mathbb{R}^p$ and $\boldsymbol{F}_t \in \mathbb{R}^p$, Cauchy-Schwartz theorem implies

$$\|\mathrm{Cov}(\boldsymbol{R}_t, \boldsymbol{F}_t)\|_{\max} \leq \sqrt{\|\mathrm{Var}(\boldsymbol{R}_t)\|_{\max}\|\mathrm{Var}(\boldsymbol{F}_t)\|_{\max}} \text{ and } \|\mathrm{Var}(\boldsymbol{R}_t)\|_{\max} = \max_{j\in p}|\mathrm{Var}(\boldsymbol{R}_{t,j})|.$$

(iii) For positive semi-definite matrices $\mathbf{M}, \mathbf{N} \in \mathbb{R}^{p,p}$, we have the following result (c.f. Wielandt-Hoffman inequality).

$$\sum_{j=1}^{p}|\sigma_i(\mathbf{M}) - \sigma_i(\mathbf{N})|^2 \leq \|\mathbf{M} - \mathbf{N}\|_2^2.$$

These properties imply that we can prove the theorem by following the proofs in VAR(1) case under assumption in Theorem 5.2. First, properties (i) and (ii) imply that under the max norm, we can apply lemmas and proofs in VAR(1) case to VAR(p) case. Second, property (iii) helps to present the line of reasoning which is similar to the logic below equation (A.4). Therefore, we will skip the proof.

A.3.4 Proof of Lemma A.12

We will prove this lemma as follows. First, we will analyze the case where $j = p - 1$ in Case 1. Then, we will consider the case where $j \geq p$ in Case 2. We only need to consider these two cases to prove this lemma because Cauchy interlacing theorem implies that this lemma also holds when $0 \leq j < p - 1$ if it is true when $j \geq p - 1$.

Case 1. We analyze the case where $j = p - 1$. Let $\mathbf{P}$ be a pd by pd permutation matrix with

$$\mathbf{P} = \begin{pmatrix} \mathbf{0}_{d\times d} & \mathbf{0}_{d\times d} & \cdots & \mathbf{0}_{d\times d} & \mathbf{I}_d \\ \mathbf{0}_{d\times d} & \cdots & \mathbf{0}_{d\times d} & \mathbf{I}_d & \mathbf{0}_{0\times d} \\ \vdots & \mathbf{0}_{0\times d} & \mathbf{I}_d & \mathbf{0}_{0\times d} & \mathbf{0}_{0\times d} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{I}_d & \mathbf{0}_{d\times d} & \cdots & \mathbf{0}_{d\times d} & \mathbf{0}_{d\times d} \end{pmatrix}.$$

Then, we have $(\boldsymbol{Z}_t^\top, \ldots, \boldsymbol{Z}_{t+p-1}^\top)^\top = \mathbf{P}(\boldsymbol{Z}_{t+p-1}^\top, \ldots, \boldsymbol{Z}_t^\top)^\top = \mathbf{P}\mathfrak{Z}_{t+p-1}$ and $\boldsymbol{\Omega}_{p-1} = \mathbf{P}\boldsymbol{\Sigma}_0\mathbf{P}^\top$. Because $\mathbf{P}^\top = \mathbf{P} = \mathbf{P}^{-1}$, the matrices $\boldsymbol{\Omega}_{p-1}$ and $\boldsymbol{\Sigma}_0$ are similar. (cf. two matrices $\mathbf{C}$ and $\mathbf{D}$ are called similar if there exists an invertible matrix $\mathbf{H}$ such that $\mathbf{C} = \mathbf{H}^{-1}\mathbf{D}\mathbf{H}$.) Therefore, we will focus on the eigenvalues of $\boldsymbol{\Sigma}_0$ instead.

Because $\boldsymbol{\Sigma}_0 = \sum_{k=0}^{\infty} \mathfrak{A}_p^k \boldsymbol{\Sigma}_{\mathfrak{E}}(\mathfrak{A}_p^\top)^k = \sum_{k=0}^{\infty}(\mathfrak{A}_p^k\bar{\mathbf{e}}_p)\ \boldsymbol{\Sigma}_{\boldsymbol{E}}(\mathfrak{A}_p^k\bar{\mathbf{e}}_p)^\top$, we have

$$\lambda_{\max}(\boldsymbol{\Sigma}_0) \leq \lambda_{\max}(\boldsymbol{\Sigma}_{\boldsymbol{E}})\sum_{k=0}^{\infty}\|\mathfrak{A}_p^k\bar{\mathbf{e}}_p\|_2^2 \leq \lambda_{\max}(\boldsymbol{\Sigma}_{\boldsymbol{E}})\sum_{k=0}^{\infty}\|\overline{\mathfrak{A}_p}^k\|_2^2 \leq C < \infty,$$

where C is a positive constant which only depends on $(C_{\boldsymbol{E}}, a_1, \ldots, a_p, p)$. The second last inequality holds because of Assumption M-p(ii) and (iii). Let $\boldsymbol{v}_{\min}$ be the eigenvector of $\boldsymbol{\Sigma}_0$ corresponding to $\lambda_{\min}(\boldsymbol{\Sigma}_0)$. We then have

$$\begin{aligned}\lambda_{\min}(\boldsymbol{\Sigma}_0) &= \sum_{k=0}^{\infty} \boldsymbol{v}_{\min}^{\top}(\mathfrak{A}_p^k \overline{\mathbf{e}}_p)\boldsymbol{\Sigma}_{\boldsymbol{E}}(\mathfrak{A}_p^k \overline{\mathbf{e}}_p)^{\top}\boldsymbol{v}_{\min} \\ &\geq \boldsymbol{v}_{\min}^{\top}\left(\sum_{k=0}^{p-1}(\mathfrak{A}_p^k \overline{\mathbf{e}}_p)\boldsymbol{\Sigma}_{\boldsymbol{E}}(\mathfrak{A}_p^k \overline{\mathbf{e}}_p)^{\top}\right)\boldsymbol{v}_{\min} \geq \frac{1}{\lambda_{\max}\left(\left(\sum_{k=0}^{p-1}(\mathfrak{A}_p^k \overline{\mathbf{e}}_p)\boldsymbol{\Sigma}_{\boldsymbol{E}}(\mathfrak{A}_p^k \overline{\mathbf{e}}_p)^{\top}\right)^{-1}\right)}.\end{aligned}$$

Note that $\sum_{k=0}^{p-1}(\mathfrak{A}_p^k \overline{\mathbf{e}}_p)\ \boldsymbol{\Sigma}_{\boldsymbol{E}}(\mathfrak{A}_p^k \overline{\mathbf{e}}_p)^{\top} = \mathbf{G}_p(\mathbf{I}_p \otimes \boldsymbol{\Sigma}_{\boldsymbol{E}})\mathbf{G}_p^{\top}$, where $\mathbf{G}_p = \begin{pmatrix}\mathbf{I}_{pd}\overline{\mathbf{e}}_p & \mathfrak{A}_p \overline{\mathbf{e}}_p & \cdots & \mathfrak{A}_p^{p-1}\overline{\mathbf{e}}_p\end{pmatrix}$, and $\otimes$ is the Kronecker product. The eigenvalues of $(\mathbf{I}_p \otimes\ \boldsymbol{\Sigma}_{\boldsymbol{E}})$ are lower bounded by some positive constant which only depends on $c_{\boldsymbol{E}}$. Lemma A.13 implies

$$\mathbf{G}_p^{-1} = \mathbf{Q}_p = \begin{pmatrix}1 & 0 & \cdots & 0\\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1\end{pmatrix} \otimes \mathbf{I}_d - \begin{pmatrix}0 & 1 & \ddots & 0\\ 0 & 0 & \ddots & 0 \\ \vdots & \ddots & 0 & 1 \\ 0 & \cdots & 0 & 0\end{pmatrix} \otimes \mathbf{A}_1 - \ldots - \begin{pmatrix}0 & \cdots & 0 & 1\\ 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots \\ 0 & \cdots & 0 & 0\end{pmatrix} \otimes \mathbf{A}_{p-1},$$

and $\|\mathbf{G}_p^{-1}\|_2 \leq 1 + \sum_{k=1}^{p-1}\|\mathbf{A}_k\|_2 \leq 1 + \sum_{k=1}^{p-1} a_k < \infty$. Therefore, $\lambda_{\max}\left(\left(\sum_{k=0}^{p-1}(\mathfrak{A}_p^k \overline{\mathbf{e}}_p)\boldsymbol{\Sigma}_{\boldsymbol{E}}(\mathfrak{A}_p^k \overline{\mathbf{e}}_p)^{\top}\right)^{-1}\right)$ is upper bounded by some positive constant that only depends on $(c_{\boldsymbol{E}}, a_1, \ldots, a_p)$, and we have $\lambda_{\min}(\boldsymbol{\Sigma}_0) \geq c > 0$ by some constant c that only depends on $(c_{\boldsymbol{E}}, a_1, \ldots, a_p)$.

Case 2. We analyze the case where $j \geq p$. From

$$\begin{pmatrix}\boldsymbol{Z}_{t+p-1+k}\\ \boldsymbol{Z}_{t+p-2+k} \\ \vdots \\ \boldsymbol{Z}_{t+k}\end{pmatrix} = \begin{pmatrix}\boldsymbol{E}_{t+p-1+k}\\ \mathbf{0}_{d\times d} \\ \vdots \\ \mathbf{0}_{d\times d}\end{pmatrix} + \mathfrak{A}_p \begin{pmatrix}\boldsymbol{E}_{t+p-2+k}\\ \mathbf{0}_{d\times d} \\ \vdots \\ \mathbf{0}_{d\times d}\end{pmatrix} + \cdots + \mathfrak{A}_p^{k-1}\begin{pmatrix}\boldsymbol{E}_{t+p}\\ \mathbf{0}_{d\times d} \\ \vdots \\ \mathbf{0}_{d\times d}\end{pmatrix} + \mathfrak{A}_p^k \begin{pmatrix}\boldsymbol{Z}_{t+p-1}\\ \boldsymbol{Z}_{t+p-2} \\ \vdots \\ \boldsymbol{Z}_{t}\end{pmatrix} \quad \text{for } k \in \mathbb{N},$$

we have

$$\begin{pmatrix}\boldsymbol{Z}_t \\ \boldsymbol{Z}_{t+1} \\ \vdots \\ \boldsymbol{Z}_{t+p-1} \\ \boldsymbol{Z}_{t+p} \\ \boldsymbol{Z}_{t+p+1} \\ \vdots \\ \boldsymbol{Z}_{t+j}\end{pmatrix} = \underbrace{\begin{pmatrix}\mathbf{I}_d & \mathbf{0}_{d\times d} & \mathbf{0}_{d\times d} & \ldots & \cdots & \cdots & \cdots & \mathbf{0}_{d\times d} \\ \mathbf{0}_{d\times d} & \mathbf{I}_d & \mathbf{0}_{d\times d} & \mathbf{0}_{d\times d} & \cdots & \cdots & \cdots & \mathbf{0}_{d\times d} \\ \mathbf{0}_{d\times d} & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0}_{d\times d} & \vdots & \mathbf{0}_{d\times d} & \mathbf{I}_d & \mathbf{0}_{d\times d} & \ddots & \ddots & \mathbf{0}_{d\times d} \\ [\mathfrak{A}_p]_{1p} & \cdots & [\mathfrak{A}_p]_{12} & [\mathfrak{A}_p]_{11} & \mathbf{I}_d & \mathbf{0}_{d\times d} & \ddots & \ddots \\ [\mathfrak{A}_p^2]_{1p} & \cdots & [\mathfrak{A}_p^2]_{12} & [\mathfrak{A}_p^2]_{11} & [\mathfrak{A}_p]_{11} & \mathbf{I}_d & \mathbf{0}_{d\times d} & \ddots \\ \vdots & \cdots & \cdots & \ddots & \ddots & \ddots & \ddots & \ddots \\ [\mathfrak{A}_p^{j-p}]_{1p} & \cdots & [\mathfrak{A}_p^{j-p}]_{12} & [\mathfrak{A}_p^{j-p}]_{11} & \cdots & [\mathfrak{A}_p^2]_{11} & [\mathfrak{A}_p^1]_{11} & \mathbf{I}_d\end{pmatrix}}_{\mathbf{L}} \begin{pmatrix}\boldsymbol{Z}_t \\ \boldsymbol{Z}_{t+1} \\ \vdots \\ \boldsymbol{Z}_{t+p-1} \\ \boldsymbol{E}_{t+p} \\ \boldsymbol{E}_{t+p+1} \\ \vdots \\ \boldsymbol{E}_{j}\end{pmatrix}.$$

From $\boldsymbol{Z}_t - \mathbf{A}_1\boldsymbol{Z}_{t-1} - \cdots - \mathbf{A}_p\boldsymbol{Z}_{t-p} = \boldsymbol{E}_t$, we have

$$\begin{pmatrix} \boldsymbol{Z}_t \\ \boldsymbol{Z}_{t+1} \\ \vdots \\ \boldsymbol{Z}_{t+p-1} \\ \boldsymbol{E}_{t+p} \\ \boldsymbol{E}_{t+p+1} \\ \vdots \\ \boldsymbol{E}_j \end{pmatrix} = \underbrace{\begin{pmatrix} \mathbf{I}_d & \mathbf{0}_{d\times d} & \mathbf{0}_{d\times d} & \cdots & \cdots & \cdots & \cdots & \mathbf{0}_{d\times d} \\ \mathbf{0}_{d\times d} & \mathbf{I}_d & \mathbf{0}_{d\times d} & \ddots & \ddots & \ddots & \ddots & \ddots \\ \mathbf{0}_{d\times d} & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ \mathbf{0}_{d\times d} & \vdots & \mathbf{0}_{d\times d} & \mathbf{I}_d & \mathbf{0}_{d\times d} & \ddots & \ddots & \ddots \\ -\mathbf{A}_p & -\mathbf{A}_{p-1} & \cdots & -\mathbf{A}_1 & \mathbf{I}_d & \mathbf{0}_{d\times d} & \ddots & \ddots \\ \mathbf{0}_{d\times d} & -\mathbf{A}_p & -\mathbf{A}_{p-1} & \cdots & -\mathbf{A}_1 & \mathbf{I}_d & \mathbf{0}_{d\times d} & \ddots \\ \mathbf{0}_{d\times d} & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ \mathbf{0}_{d\times d} & \ddots & \mathbf{0}_{d\times d} & -\mathbf{A}_p & -\mathbf{A}_{p-1} & \cdots & -\mathbf{A}_1 & \mathbf{I}_d \end{pmatrix}}_{\mathbf{L}^{-1}} \begin{pmatrix} \boldsymbol{Z}_t \\ \boldsymbol{Z}_{t+1} \\ \vdots \\ \boldsymbol{Z}_{t+p-1} \\ \boldsymbol{Z}_{t+p} \\ \boldsymbol{Z}_{t+p+1} \\ \vdots \\ \boldsymbol{Z}_{t+j} \end{pmatrix}.$$

Then, it suffices to show that $\|\mathbf{L}\|_2$ and $\|\mathbf{L}^{-1}\|_2$ are bounded from above by some absolute positive constants because Case 1 implies that the eigenvalues of $\mathrm{Var}((\boldsymbol{Z}_t^\top, \ldots, \boldsymbol{Z}_{t+p-1}^\top, \boldsymbol{E}_{t+p}^\top, \ldots, \boldsymbol{E}_{t+j}^\top)^\top)$ are lower and upper bounded by some positive constants that only depend on $(c_{\boldsymbol{E}}, C_{\boldsymbol{E}}, a_1, \ldots, a_p, p)$. We have

$$\|\mathbf{L}^{-1}\|_2 \le 1 + \sum_{k=1}^{p} \|\mathbf{A}_k\|_2 \le 1 + a_1 + \ldots + a_p < \infty, \text{ and}$$
$$\|\mathbf{L}\|_2 \le 1 + \sum_{k=1}^{j-p} \|\mathfrak{A}_p^k\|_2 + (p-1)\sum_{k=1}^{j-p} \|\mathfrak{A}_p^k\|_2 \le 1 + p\sum_{k=1}^{\infty} \|\overline{\mathfrak{A}_p}^k\|_2 \le C_{\mathbf{L}} < \infty,$$

where $C_{\mathbf{L}}$ is a positive constant that only depends on $(a_1, \ldots, a_p, p)$. Therefore, $0 < c \le \lambda_{\min}(\boldsymbol{\Omega}_j) \le \lambda_{\max}(\boldsymbol{\Omega}_j) \le C < \infty$ by constants c and C which only depend on $(c_{\boldsymbol{E}}, C_{\boldsymbol{E}}, a_1, \ldots, a_p, p)$.

A.3.5 Technical lemmas used in Section 5

Lemma A.13. Let $\mathbf{G}_p = \begin{pmatrix} \bar{\mathbf{e}}_p & \mathfrak{A}_p\bar{\mathbf{e}}_p & \cdots & \mathfrak{A}_p^{p-1}\bar{\mathbf{e}}_p \end{pmatrix} \in \mathbb{R}^{pd\times pd}$. For any $p \in \mathbb{N}$, the inverse $\mathbf{G}_p^{-1}$ exists and $\mathbf{G}_p^{-1} = \mathbf{Q}_p$, where $\mathbf{Q}_p$ is defined in Section A.3.1.

Proof. Throughout the proof, it is noted that we can represent $\mathfrak{A}_p$ and $\mathbf{Q}_p$ recursively as follows,

$$\mathfrak{A}_p = \begin{pmatrix} \mathfrak{A}_{p-1} & \bar{\mathbf{e}}_{p-1}\mathbf{A}_p \\ \underline{\mathbf{e}}_{p-1}^\top & \mathbf{0}_{d\times d} \end{pmatrix} \text{ and } \mathbf{Q}_p = \begin{pmatrix} \mathbf{Q}_{p-1} & -\mathbf{C}_{p-1} \\ \mathbf{0}_{d\times(p-1)d} & \mathbf{I}_d \end{pmatrix} \text{ with } \mathfrak{A}_1 = \mathbf{A}_1 \text{ and } \mathbf{Q}_1 = \mathbf{I}_d.$$

With this information, we will prove this lemma with two steps. In Step 1, we will show that $\mathbf{G}_p$ can be written recursively, and we have $\mathbf{Q}_{p-1}\mathfrak{A}_{p-1}^{p-1}\bar{\mathbf{e}}_{p-1} = \mathbf{C}_{p-1}$. In Step 2, using the result in Step 1 and the information that $\mathbf{Q}_p$ can be represented recursively, we will prove this lemma using mathematical induction.

Step 1. We will show that $\mathbf{G}_p$ can be written as

$$\mathbf{G}_p = \begin{pmatrix} \mathbf{G}_{p-1} & \mathfrak{A}_{p-1}^{p-1}\bar{\mathbf{e}}_{p-1} \\ \mathbf{0}_{d\times(p-1)d} & \mathbf{I}_d \end{pmatrix} \text{ for any } p \ge 2,$$

and we have $\mathbf{Q}_{p-1}\mathfrak{A}_{p-1}^{p-1}\overline{\mathbf{e}}_{p-1} = \mathbf{C}_{p-1}$.

For simplicity of notation, we partition $\mathfrak{A}_p^{k-1}\overline{\mathbf{e}}_p \in \mathbb{R}^{pd\times d}$ with $k \in [p]$ to two blocks so that $\mathfrak{A}_p^{k-1}\overline{\mathbf{e}}_p = \begin{pmatrix}\mathbf{D}_{p,k}\\ \mathbf{F}_{p,k}\end{pmatrix}$, where $\mathbf{D}_{p,k}$ is a $(p-1)d$ by d matrix and $\mathbf{F}_{p,k}$ is a d by d matrix. Then, we have

$$\mathbf{G}_p = \begin{pmatrix}\overline{\mathbf{e}}_p & \mathfrak{A}_p\overline{\mathbf{e}}_p & \cdots & \mathfrak{A}_p^{p-1}\overline{\mathbf{e}}_p\end{pmatrix} = \begin{pmatrix}\mathbf{D}_{p,1} & \cdots & \mathbf{D}_{p,p-1} & \mathbf{D}_{p,p}\\ \mathbf{F}_{p,1} & \cdots & \mathbf{F}_{p,p-1} & \mathbf{F}_{p,p}\end{pmatrix}.$$

In order to finish Step 1, it is sufficient to show that (1) $\mathbf{D}_{p,k} = \mathfrak{A}_{p-1}^{k-1}\overline{\mathbf{e}}_{p-1}$ and $\mathbf{F}_{p,k} = \mathbf{0}_{d\times d}$ for $k \in [p-1]$; and (2) $\mathbf{D}_{p,p} = \mathfrak{A}_{p-1}^{p-1}\overline{\mathbf{e}}_{p-1}$, $\mathbf{F}_{p,p} = \mathbf{I}_d$, and $\mathbf{Q}_{p-1}\mathbf{D}_{p,p} = \mathbf{C}_{p-1}$.

Note that we have $\mathbf{D}_{p,k} = \mathfrak{A}_{p-1}\mathbf{D}_{p,k-1} + \overline{\mathbf{e}}_{p-1}\mathbf{A}_p\mathbf{F}_{p,k-1}$ and $\mathbf{F}_{p,k} = \underline{\mathbf{e}}_{p-1}^\top\mathbf{D}_{p,k-1}$ with $\mathbf{D}_{p,1} = \overline{\mathbf{e}}_{p-1}$ and $\mathbf{F}_{p,1} = \mathbf{0}_{d\times d}$. This is because

$$\mathfrak{A}_p^{k-1}\overline{\mathbf{e}}_p = \mathfrak{A}_p\mathfrak{A}_p^{k-2}\overline{\mathbf{e}}_p = \begin{pmatrix}\mathfrak{A}_{p-1} & \overline{\mathbf{e}}_{p-1}\mathbf{A}_p\\ \underline{\mathbf{e}}_{p-1}^\top & \mathbf{0}_{d\times d}\end{pmatrix}\begin{pmatrix}\mathbf{D}_{p,k-1}\\ \mathbf{F}_{p,k-1}\end{pmatrix} = \begin{pmatrix}\mathfrak{A}_{p-1}\mathbf{D}_{p,k-1} + \overline{\mathbf{e}}_{p-1}\mathbf{A}_p\mathbf{F}_{p,k-1}\\ \underline{\mathbf{e}}_{p-1}^\top\mathbf{D}_{p,k-1}\end{pmatrix}.$$

With this information, it is possible to show that (1) and (2) are true for any $p \geq 2$ using mathematical induction as follows.

Base Step. When $p = 2$, we have $\mathbf{D}_{2,1} = \overline{\mathbf{e}}_1 = \mathfrak{A}_1^0\overline{\mathbf{e}}_1$, $\mathbf{F}_{2,1} = \mathbf{0}_{d\times d}$,

$$\mathbf{D}_{2,2} = \mathfrak{A}_1\mathbf{D}_{2,1} + \overline{\mathbf{e}}_1\mathbf{A}_2\mathbf{F}_{2,1} = \mathfrak{A}_1\overline{\mathbf{e}}_1, \mathbf{F}_{2,2} = \underline{\mathbf{e}}_1^\top\mathbf{D}_{2,1} = \underline{\mathbf{e}}_1^\top\overline{\mathbf{e}}_1 = \mathbf{I}_d, \text{ and } \mathbf{Q}_1\mathbf{D}_{2,2} = \mathbf{I}_d\mathfrak{A}_1\overline{\mathbf{e}}_1 = \mathbf{A}_1 = \mathbf{C}_1.$$

Therefore, (1) and (2) are true when $p = 2$.

Induction Step. Suppose (1) and (2) hold when $p = m$. Then, the following three steps will show that (1) and (2) are true for $p = m + 1$.

(i) We have $\mathbf{D}_{m+1,1} = \overline{\mathbf{e}}_m = \mathfrak{A}_m^0\overline{\mathbf{e}}_m$ and $\mathbf{F}_{m+1,1} = \mathbf{0}_{d\times d}$.

(ii) Suppose we have that $\mathbf{D}_{m+1,j} = \mathfrak{A}_m^{j-1}\overline{\mathbf{e}}_m$ and $\mathbf{F}_{m+1,j} = \mathbf{0}_{d\times d}$ for $j \in [m-1]$. Then, we have

$$\mathbf{D}_{m+1,j+1} = \mathfrak{A}_m\mathbf{D}_{m+1,j} + \overline{\mathbf{e}}_m\mathbf{A}_{m+1}\mathbf{F}_{m+1,j} = \mathfrak{A}_m\mathfrak{A}_m^{j-1}\overline{\mathbf{e}}_m = \mathfrak{A}_m^j\overline{\mathbf{e}}_m \text{ and}$$
$$\mathbf{F}_{m+1,j+1} = \underline{\mathbf{e}}_m^\top\mathbf{D}_{m+1,j} = \underline{\mathbf{e}}_m^\top\mathfrak{A}_m^{j-1}\overline{\mathbf{e}}_m = \begin{pmatrix}\mathbf{0}_{d\times(m-1)d} & \mathbf{I}_{d\times d}\end{pmatrix}\begin{pmatrix}\mathbf{D}_{m,j}\\ \mathbf{F}_{m,j}\end{pmatrix} = \mathbf{F}_{m,j} = \mathbf{0}_{d\times d}.$$

The results of (i) and (ii) combined imply that $\mathbf{D}_{m+1,k} = \mathfrak{A}_m^{k-1}\overline{\mathbf{e}}_m$ and $\mathbf{F}_{m+1,k} = \mathbf{0}_{d\times d}$ for $k \in [m]$.

(iii) From (i),(ii), and the assumption in the induction step, we have

$$\mathbf{D}_{m+1,m+1} = \mathfrak{A}_m\mathbf{D}_{m+1,m} + \overline{\mathbf{e}}_m\mathbf{A}_{m+1}\mathbf{F}_{m+1,m} = \mathfrak{A}_m\mathfrak{A}_m^{m-1}\overline{\mathbf{e}}_m = \mathfrak{A}_m^m\overline{\mathbf{e}}_m,$$
$$\mathbf{F}_{m+1,m+1} = \underline{\mathbf{e}}_m^\top\mathbf{D}_{m+1,m} = \underline{\mathbf{e}}_m^\top\mathfrak{A}_m^{m-1}\overline{\mathbf{e}}_m = \begin{pmatrix}\mathbf{0}_{d\times(m-1)d} & \mathbf{I}_{d\times d}\end{pmatrix}\begin{pmatrix}\mathbf{D}_{m,m}\\ \mathbf{F}_{m,m}\end{pmatrix} = \mathbf{F}_{m,m} = \mathbf{I}_d, \text{ and}$$
$$\begin{aligned}\mathbf{Q}_m\mathbf{D}_{m+1,m+1} &= (\mathbf{Q}_m\mathfrak{A}_m)(\mathfrak{A}_m^{m-1}\overline{\mathbf{e}}_m)\\ &= \begin{pmatrix}\mathbf{0}_{d\times(m-1)d} & \mathbf{A}_m\\ \mathbf{Q}_{m-1} & \mathbf{0}_{(m-1)d\times d}\end{pmatrix}\begin{pmatrix}\mathbf{D}_{m,m}\\ \mathbf{F}_{m,m}\end{pmatrix} = \begin{pmatrix}\mathbf{A}_m\mathbf{F}_{m,m}\\ \mathbf{Q}_{m-1}\mathbf{D}_{m,m}\end{pmatrix} = \begin{pmatrix}\mathbf{A}_m\\ \mathbf{C}_{m-1}\end{pmatrix} = \mathbf{C}_m.\end{aligned}$$

The results of (i), (ii), and (iii) combined imply that (1) and (2) are also true for $p = m + 1$. In conclusion, by mathematical induction, we have the desired result.

Step 2. We prove $\mathbf{G}_p^{-1} = \mathbf{Q}_p$ using mathematical induction as follows.

Base Step. When $p = 1$, we have $\mathbf{G}_p^{-1} = \mathbf{Q}_p$ because $\mathbf{G}_1 = \mathbf{I}_d$ and $\mathbf{Q}_1 = \mathbf{I}_d$.

Induction Step. Suppose $\mathbf{G}_p^{-1} = \mathbf{Q}_p$ is true for $p = m$, where $m \geq 1$. Using the result in Step 1, we have

$$\begin{aligned}
\mathbf{Q}_{m+1}\mathbf{G}_{m+1} &= \begin{pmatrix} \mathbf{Q}_m & -\mathbf{C}_m \\ \mathbf{0}_{d\times md} & \mathbf{I}_d \end{pmatrix} \begin{pmatrix} \mathbf{G}_m & \mathbf{D}_{m+1,m+1} \\ \mathbf{0}_{d\times md} & \mathbf{I}_d \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{Q}_m\mathbf{G}_m & \mathbf{Q}_m\mathbf{D}_{m+1,m+1} - \mathbf{C}_m \\ \mathbf{0}_{d\times md} & \mathbf{I}_d \end{pmatrix} = \begin{pmatrix} \mathbf{I}_{md} & \mathbf{0}_{md\times d} \\ \mathbf{0}_{d\times md} & \mathbf{I}_d \end{pmatrix}.
\end{aligned}$$

The last equality holds because Step 1 implies $\mathbf{Q}_m\mathbf{D}_{m+1,m+1} = \mathbf{C}_m$. This shows that $\mathbf{G}_p^{-1} = \mathbf{Q}_p$ is true for $p = m + 1$.

Therefore, by mathematical induction, we have the desired result.

□