Journal of Machine Learning Research 27 (2026) 1-72 Submitted 5/25; Revised 11/25; Published 5/26

Embedding Network Autoregression for Time Series Analysis and Causal Peer Effect Inference

Jae Ho Chang CHANG.2090@OSU.EDU
Department of Statistics
The Ohio State University
Columbus, OH 43210, USA

Subhadeep Paul PAUL.963@OSU.EDU
Department of Statistics
The Ohio State University
Columbus, OH 43210, USA

Editor: Elias Bareinboim

Abstract

We propose an Embedding Network Autoregressive Model for multivariate networked longitudinal data. We assume the network is generated from a latent variable model, and these unobserved variables are included in a structural peer effect model or a time series network autoregressive model. This approach takes a unified view of two related yet different problems: (1) modeling and predicting multivariate networked time series data and (2) causal peer influence estimation in the presence of confounding due to homophily from finite-time longitudinal data. Our estimation strategy comprises estimating latent variables from the observed network, followed by least squares estimation of the network autoregressive model. We show that the momentum and peer effect parameters estimated with our method are consistent and asymptotically normally distributed in setups with a growing number of network vertices (N) while considering both a growing number of time points T (for the time series problem) and finite T cases (for the peer effect problem). We allow the number of latent vectors K to grow at appropriate rates. We also develop a selection criterion when K is unknown that provably does not under-select. We show that the theoretical guarantees hold with the selected number for K, and study the bias rates when K is misspecified. With the new methods, we study peer effects in conflict and school climate perception using data on more than 7000 students from 23 schools.

Keywords: Network time series; social influence; peer effect; social network; latent homophily; network embedding.

1. Introduction

A network of relationships and longitudinal node-level responses commonly appear in research problems in multiple domains, including social sciences, economics, public health, and biomedical sciences. We consider two key statistical problems associated with such data that have been widely investigated in the literature. The first is to causally estimate peer effects or social influence propagating through an observed network when the node level outcome of interest is measured in at least two but finite time points (Shalizi and Thomas, 2011; VanderWeele, 2011; Goldsmith-Pinkham and Imbens, 2013; McFowland III

and Shalizi, 2023; Nath et al., 2025). The second problem is to model and predict a high dimensional time series when network information is also observed (Zhu et al., 2017, 2019; Knight et al., 2020; Zhu and Pan, 2020; Chen et al., 2023). The goals in these two problems are generally different and have been investigated separately in the literature. We take a unified view of these two problems and propose to include latent homophily variables in both of these problems. In the case of estimating peer influence, this approach aids causal identification by controlling for unobserved confounders. On the other hand, in the case of network time series, this approach leads to improved forecasting and model estimation.

It will be convenient to formally introduce the network autoregressive model (NAR or NAM), which has been used in the literature for both problems, to facilitate further discussion on its interpretation. We assume that we have measurements y_{it} for $i = 1, \ldots, N$, and $t = 0, \ldots, T-1$ with $N, T \in \mathbb{N}$ for an univariate outcome measured at N vertices of a network over T time periods. Let ℓ_{ij} denote the (i, j)-th entry of the normalized (symmetric) adjacency matrix. Then, our measurements y_{it} are assumed to be generated via

$$y_{i,t+1} = \alpha y_{it} + \theta \sum_{j \neq i} \ell_{ij} y_{jt} + \mathbf{z}_{it}{}^{\intercal} \gamma + \epsilon_{i,t+1}, \tag{1}$$

where $\mathbf{z}_{it}$ is a vector of (possibly time-varying) covariates and ϵ_{it} is the error term. We differentiate between the utility and the interpretation of this model in terms of whether T is finite or growing. The former is useful as a linear structural model for estimating peer influence, and the latter is useful for multivariate time series modeling. The parameter θ has been termed the "peer effect" parameter in the literature on both problems, while the parameter α is referred to as the "momentum effect."

The above NAR or NAM model with longitudinally measured outcomes in Equation (1) has been widely employed for the causal identification of peer influence (Christakis and Fowler, 2007; Shalizi and Thomas, 2011; VanderWeele et al., 2012; Christakis and Fowler, 2013; O'Malley et al., 2014; McFowland III and Shalizi, 2023; Nath et al., 2025). However, several authors have noted issues with causal identification of the peer effect parameter θ with such models, including confounding due to latent homophily in peer selection and other unobserved omitted variables (Shalizi and Thomas, 2011; Goldsmith-Pinkham and Imbens, 2013; O'Malley et al., 2014; An et al., 2022; Egami and Tchetgen Tchetgen, 2024). The latent homophily can be thought of as unobserved characteristics of individuals that simultaneously may affect the selection of peers or connections in a network and the outcome on which peer effect is being estimated. Since in the vast majority of research on peer effects, the researchers do not have control over how individuals form networks, this is a persistent problem for causal identification of peer effects (Shalizi and Thomas, 2011). The authors in McFowland III and Shalizi (2023); Nath et al. (2025) suggest augmenting the linear peer effects model with latent variables, which are responsible for network formation as a way of controlling for latent homophily. They proceed to show that asymptotic identification of the peer influence parameter is possible with this approach. However, consistency, limiting distribution, and the rate of convergence of the estimator of the peer effect parameter remain unsolved problems.

The NAR model in Equation (1) has also been used in the context of modeling and predicting multivariate high-dimensional network-linked time series in a line of work including Zhu et al. (2017, 2019); Knight et al. (2020); Zhu and Pan (2020); Chen et al. (2023). Those

papers then investigate the stationarity of the model along with consistency and asymptotic normality of the parameter estimates in an asymptotic setup where $T \to \infty$. While several extensions of the baseline NAR model in Equation (1) have been proposed, additional latent information from the network, in addition to the peer effects, has not been incorporated into the model. Therefore, the benefits of including latent variables in improving predictive performance and the conditions under which consistent estimation is possible are also currently unknown.

Our proposal in this paper is to augment the NAR model (Equation (1)) with latent variables that are related to both the outcome and the formation of the network. Accordingly, we assume that the network adjacency matrix is generated from the Random Dot Product Graph (RDPG) model (Athreya et al., 2017). The RDPG model is a general latent variable model for network data whose special cases include several popular latent variable models, namely, the Stochastic Block Model (SBM), Degree-Corrected SBM (DCSBM), Mixed Membership SBM (MMSBM), and DCMMSBM (Athreya et al., 2017; Rubin-Delanchy et al., 2022; Xie and Xu, 2023).

We assume the network adjacency matrix $\mathbf{A}$ is generated from a K-dimensional RDPG model (defined later) with sparsity parameter ρ_N and latent position parameter matrix $\mathbf{X} \in \mathbb{R}^{N\times K}$ such that the probability of connections is $\mathbf{P} = \rho_N \mathbf{X}\mathbf{X}^\intercal$. We let $\mathbf{U} \in \mathbb{R}^{N\times K}$ be a generic notation for a matrix that contains latent homophily vectors for the vertices. This matrix $\mathbf{U}$ is related to the matrix of underlying latent positions $\mathbf{X}$. In this paper, we consider this matrix to be either $\mathbf{V}$, the orthogonal eigenvectors of $\mathbf{P}$ for the K largest-magnitude eigenvalues (all the non-zero eigenvalues of $\mathbf{P}$), or the matrix of corresponding spectral embedding vectors $\mathbf{V}\mathbf{S}^{1/2}$, where $\mathbf{S}$ is the $K \times K$ diagonal matrix containing the K largest-magnitude eigenvalues. Then, our Embedding Network Autoregression (ENAR) model augments the NAR model with these K-dimensional unknown (latent) vectors. Formally, the ENAR model assumes that the measurements y_{it} for $i = 1, \ldots, N$ are generated via

$$y_{i,t+1} = \alpha y_{it} + \theta \sum_{j\neq i} \ell_{ij} y_{jt} + \mathbf{u}_i^\intercal \beta + \mathbf{z}_{it}^\intercal \gamma + \epsilon_{it}. \tag{2}$$

However, the vectors of latent variables $\mathbf{u}_i$'s are not observed. Therefore, we propose to estimate the latent variables from the observed network through spectral embedding and replace $\mathbf{u}_i$ with its estimated version $\hat{\mathbf{u}}_i$ in Equation (2). However, a natural concern is whether the true peer influence parameter θ can be estimated consistently, and the asymptotic variance can be characterized to enable inference when the $\mathbf{u}_i$ is replaced with the estimated $\hat{\mathbf{u}}_i$. McFowland III and Shalizi (2023) considered this issue as trading off omitted variable bias with measurement error bias and remarked that it would only succeed if the measurement error bias is low.

In this article, we show that the answer is affirmative and develop a theory for the consistency and asymptotic normality of the peer influence parameter. In addition, we propose to include estimated latent variables in the context of time series modeling as well to enable accurate inference on the parameters of the model and improve predictive performance. We develop theoretical results under an asymptotic setup where the dimension of the latent space K also grows with the number of nodes N. We explore several questions that naturally arise in modeling two distinct phenomena with a similar model. These questions include differences in modeling latent homophily with eigenvectors $\mathbf{V}$ and with spectral embedding

vectors $\mathbf{VS}^{1/2}$. We discuss asymptotic biases of ignoring latent homophily vectors with both choices and develop theoretical results for the peer effect problem for both choices. We also delve into the challenging problem of obtaining results when the number of latent variables K is unknown, which is often the practical case but has not been explored in the literature before. We study the properties of the estimators when K is misspecified and develop a selection criterion for selecting K. Together, these results greatly expand the scope of the methodology and theory for both the peer effect problem and the time series problem.

1.1 Connection to related literature

Our work is connected to several recent works encompassing network time series, network peer effect estimation, and network regression. In the network time series literature, our approach is related to the CNAR model of Chen et al. (2023). We compare our results theoretically and in simulation extensively with the CNAR model and show the theoretical advantage of our approach as well as superior or comparable prediction performance in simulation. In the peer effect literature, the approach we take is most related to McFowland III and Shalizi (2023); Nath et al. (2025), while the problem we solve is also related to the recent work Egami and Tchetgen Tchetgen (2024). However, the theoretical results for the peer effect problem in this paper go well beyond what was considered in McFowland III and Shalizi (2023); Nath et al. (2025) and delve into consistency and asymptotic normality of peer effect parameters, selection of latent dimension, and differences between modeling choices. For both problems, no earlier work addresses the problem of selecting the number of latent dimensions or communities. Our work is also related to network regression Fosdick and Hoff (2015); He and Hoff (2019); Le and Li (2022); Lunde et al. (2025). For this problem, our corollaries develop consistency and asymptotic normality for growing embedding dimensions, and our model selection procedure provides theoretical guarantees on selecting the number of embedding vectors. Another paper related to ours is the random graph autoregressive model of Wu and Leng (2023). However, the theoretical results in Wu and Leng (2023) are only developed for $T \to \infty$, and the method itself does not work unless we have data available in at least a few time points. This is not appropriate for the peer effect problem, where we often have data on only two or three time points.

1.2 Main results and contributions

We summarize our theoretical results for both of these statistical goals.

1.2.1 Network time series results:

In the case of modeling time series, we assume $T \to \infty$. The true data-generating model is augmented with eigenvectors $\mathbf{V}$ from the RDPG model. We show that under certain regularity conditions on the expected density of the network ρ_N (or equivalently on the eigengap $s_K(\mathbf{P})$) namely, $s_K(\mathbf{P}) \asymp N\rho_N = \omega(T + \log^4 N)$, the estimated parameter vector $\hat{\mu}$, suitably normalized, converges to a multivariate normal distribution with finite variance around the true parameter vector μ^H as long as $K^2 = o(N)$ and $\log K = o(T)$ (Theorems 10 and 2).

To compare with existing results in related literature, the asymptotic growth rates necessary for the CNAR model in Chen et al. (2023) (which albeit is a different but related model)

is $K^4 = o(N)$ and $K^2 \log K = o(T)$ and the eigengap $s_K(\mathbf{P}) = \omega(K\sqrt{NT\rho_N})$. Therefore, our results on the ENAR model require strictly weaker assumptions on the eigengap and the necessary sample size. In the simulation, we see that ENAR can accurately estimate the latent effects and network effects, yet it achieves comparable or better prediction performance than CNAR and NAR.

1.2.2 Peer effect inference results

Second, for the case of finite T (with $T \geq 2$), we are interested in the accurate inference of the peer influence parameter θ. We start with results under the ENAR model that uses eigenvectors $\mathbf{V}$ to model the effects of latent homophily. Our result shows if $N\rho_N = \omega(\log^4 N)$ and $N = \omega(K^2)$, while $\hat{\beta}_v - \mathbf{H}_v^\intercal \beta_v = O_{\mathbb{P}}(1)$, the rest of the parameters including the peer influence parameter is estimated with a $\sqrt{N}$ convergence to a multivariate normal distribution with finite variance (Theorems 11 and 3).

We find that for the peer effect problem, the more interesting case is when the true data-generating model contains the spectral embedding vectors $\mathbf{VS}^{1/2}$ as latent homophily as opposed to eigenvectors. In Theorem 9, we show that *ignoring latent homophily in terms of spectral embedding vectors leads to inconsistent estimation of peer effect* in the case of dense networks, and less than $\sqrt{N}$ rate of consistency even for the sparse case, as $\sqrt{N}(\hat{\theta}_{-,e} - \theta_e) = O_{\mathbb{P}}(\sqrt{N}\rho_N)$. On the other hand, our method that fits least squares with estimated spectral embeddings $\hat{\mathbf{V}}|\hat{\mathbf{S}}|^{1/2}$ leads to consistent and asymptotically normal estimators of peer effect for fixed K (Theorem 4) and growing K (Theorem 5) regimes. For the growing K case, under conditions $N\rho_N = \omega(\log^4 N)$ and $N = \omega(K^2 \log K)$ we have $\frac{\sqrt{N}\log K}{\sqrt{K}}(\hat{\theta}_e - \theta_e) = O_{\mathbb{P}}(1)$. Therefore, we have $\sqrt{N}$ consistency for the fixed K case and a slightly slower rate of convergence for the growing K case, depending upon the growth rate of K.

These results can be compared with the results in McFowland III and Shalizi (2023); Nath et al. (2025) where the authors showed asymptotic unbiasedness of the peer influence parameter under the SBM and the RDPG models, respectively. In contrast, our results hold for RDPG models with growing dimensions of the latent space K and generalize and supplement those results to include consistency and asymptotic normality.

The theoretical results for both problems are succinctly summarized in Table 1.

1.2.3 Selection of K and results with model misspecification

We address the challenging problem of selecting the number of latent vectors to include in the ENAR model (or equivalently, the dimension of the latent homophily vectors). While this is an important problem for practitioners, previous works in both network time series literature (Chen et al. (2023)) and Peer effect literature McFowland III and Shalizi (2023); Nath et al. (2025) did not address this problem. We develop a criterion for the selection of the number of latent vectors that involves a modified residual sum of squares and a penalty term for model complexity. We prove in Theorem 30 that the procedure does not underselect the number of latent vectors K both when the eigenvectors are used or when the spectral embedding vectors are used, i.e., $\mathbb{P}\left(\hat{K} \geq K\right) \to 1$. We complement this result with results that show the same level of theoretical guarantees on the estimators described

above are available when the selected K is higher than the true K, and consequently, we include some additional latent vectors in the model (Theorems 14 and 15).

Model	**Rate of consistency**	**Growing-K conditions**	**Consistency with $\hat{K}$**	**Effect of omitting**
Time-series regime ($T \to \infty$, $N \to \infty$)				
Eigenvectors $\mathbf{U}_v$	$\sqrt{NT}$ (Thm 10)	$N = \omega(K^2)$, $T = \omega(\log K)$, $N\rho_N = \omega(T + \log^4 N)$ (Thm 2)	Yes (Thm 30, 14)	Higher bias rate (Thm 8)
Peer-effect regime ($N \to \infty$, finite T)				
Eigenvectors $\mathbf{U}_v$	$\sqrt{N}$ (Thm 11)	$N = \omega(K^2)$, $N\rho_N = \omega(\log^4 N)$ for $\sqrt{N}$-consistency (Thm 3)	Yes (Thm 30, 14)	Consistent, but higher bias (Thm 8)
Embedding $\mathbf{U}_e$	$\sqrt{N}$ (Thm 4)	$N\rho_N = \omega(\log^4 N)$, $N = \omega(K^2 \log K)$ for $\frac{\sqrt{N}\log K}{\sqrt{K}}$-consistency (Thm 5)	Yes (Thm 30, 15)	Inconsistent (Thm 9)

Table 1: Rates of consistency of peer effect (θ) and momentum effect parameters (α) under various models and asymptotic setups

1.2.4 Omitted variable bias when latent vectors are ignored

In Theorems 8 and 9, we study the omitted variable bias due to under-selecting the latent vectors and, as a special case, the situation of not including the latent vectors at all in the fitted model. First, to study this situation, we derive new results on the concentration of eigenvectors and spectral embedding vectors when the dimension is under-selected in Theorem 7. In Theorem 8 we see that $\sqrt{NT}(\hat{\theta}_{-,v} - \theta_v) = O_{\mathbb{P}}\left(\sqrt{\frac{T}{N}}\right)$, while in Theorem 9, we see that $\sqrt{N}(\hat{\theta}_{-,e} - \theta_e) = O_{\mathbb{P}}\left(\sqrt{N}\rho_N\right)$ when the latent vectors are omitted. This tells us that the model with eigenvectors is unable to induce effective confounding, and the omitted variable bias asymptotically vanishes. On the other hand, the spectral embedding vectors introduce significant omitted variable bias and are therefore well-suited as a model for the peer effect problem. However, in both cases, ignoring the latent vectors leads to higher bias, as can be seen by comparing the rate of convergence of the bias terms between those theorems and the corresponding theorems with latent vectors.

We also consider the conceptual and theoretical differences between using spectral embedding vectors and eigenvectors from the network in the outcome model. As can be seen from Table 1, for the peer effect problem, the true data-generating model should include latent positions or embedding vectors and not eigenvectors to induce the desired confounding. Instead, if the true model contains eigenvectors, then the peer effect parameter is asymptotically identified as the omitted variable bias asymptotically goes away, and one does not need to do anything to remove the confounding. On the other hand, with embedding vectors, the peer effect parameter is inconsistent due to the omitted variable bias, and

trading off the omitted variable bias with measurement error bias is fruitful since that leads to consistent estimation of the peer effect.

Table 2: Notation summary

Symbol	Description
N	Number of nodes (network size)
T	Number of time points (panel length)
K	True Latent dimension
ρ_N	Network sparsity or density scaling parameter
$\mathbf{A} = (a_{ij})$	Network adjacency matrix
$\mathbf{X}$	Latent positions in RDPG model such that, $\mathbb{E}[\mathbf{A}] = \rho_N \mathbf{X}\mathbf{X}^\intercal$
$\mathbf{P} = \mathbb{E}[\mathbf{A}]$	Population network link probability
$\mathbf{V}, \mathbf{S}$	Eigenvector and eigenvalue matrices of $\mathbf{P} = \mathbf{V}\mathbf{S}\mathbf{V}^\intercal$
$\mathbf{U}$	Generic notation for latent homophily variable ($N \times K$)
$\mathbf{y}_t = (y_{1t}, \ldots, y_{Nt})^\top$	Response vector at time t
$\mathbf{Z}_t = (\mathbf{z}_{1t}, \ldots, \mathbf{z}_{Nt})^\top$	Exogenous covariate matrix at time t ($N \times p$)
$\mathcal{E}_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Nt})^\top$	Noise vector at time t
$\mathcal{L}_\mathbf{A} = (\ell_{ij})$	Normalized adjacency matrix
θ	Peer effect parameter
α	Momentum effect parameter
β	K-dimensional coefficient vector for latent variables $\mathbf{u}_i$
γ	p-dimensional coefficient vector for covariates $\mathbf{z}_{it}$
$\mu = (\beta^\top, \alpha, \theta, \gamma^\top)^\top$	Stacked parameter vector in regression form
$\mathbf{W}_t$	Design matrix at time t, stacking regressors for all nodes

2. Embedding Network Autoregressive Model

In this section, we describe the proposed Embedding Network Autoregressive (ENAR) model. Let $\mathcal{O}_{N,K} := \left\{\mathbf{H} \in \mathbb{R}^{N\times K}; \mathbf{H}^\intercal\mathbf{H} = \mathbf{I}_K\right\}$ be a collection of $N \times K$ matrices with real orthonormal columns. Write an m-dimensional vector and $m \times m$ matrix with zeros as $\mathbf{o}_m$ and $\mathbf{O}_m$, and let $\mathbf{I}_m$ denotes an $m \times m$ identity matrix. Let $S^{p-1} = \{\mathbf{h} \in \mathbb{R}^p; \|\mathbf{h}\| = 1\}$ and $B^{p-1} = \{\mathbf{h} \in \mathbb{R}^p; \|\mathbf{h}\| \leq 1\}$ denote p-dimensional unit sphere and ball, respectively. Table 2 summarizes the meaning of major notations. More description on notations can be found in Section A in the Appendix.

2.1 Network Vector Autoregression and Random Dot Product Graph

In this section, we review some background materials related to our methods. As stated in the introduction, we assume a statistical problem where we have longitudinal measurements y_{it} on a univariate outcome over N subjects at T time points. We let $i = 1, ..., N$ and $t = 0, ..., T-1$ with $(N, T \in \mathbb{N})$. We further assume that these N individuals are connected in a network with an (undirected) adjacency matrix $\mathbf{A}$, which is also observed. We define its normalized (symmetric) Laplacian matrix as $\mathcal{L}_{\mathbf{A}} = \mathcal{D}^{-1/2}\mathbf{A}\mathcal{D}^{-1/2}$ where $\mathcal{D}$ is a diagonal matrix containing its degrees. The elements of $\mathcal{L}_{\mathbf{A}}$ are denoted as ℓ_{ij}. For each unit i, we further have measurements on p-dimensional covariates $\mathbf{z}_{it}$, where the subscript t indicates that the covariates may vary over time. Recall the network vector autoregressive model (Zhu et al., 2017; McFowland III and Shalizi, 2023) in Equation (1), $y_{i,t+1} = \alpha y_{it} + \theta \sum_{j \neq i} \ell_{ij} y_{jt} + \mathbf{z}_{it}{}^{\intercal}\gamma + \epsilon_{i,t+1}$. We assume that $\mathbf{z}_{it} \in \mathbb{R}^p$ are i.i.d. Gaussian random vectors with independent coordinates with zero mean and finite fourth-order moments, and Σ_z denotes their common covariance. This assumption is similar to one described in Chen et al. (2023). For the model errors, we assume $\epsilon_{it} \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$ for $\sigma > 0$. Among the parameters, $\alpha \in \mathbb{R}$ denotes the momentum effect, $\theta \in \mathbb{R}$ denotes the peer influence effect, and $\gamma \in \mathbb{R}^p$ denotes the time-invariant covariate effects (Zhu et al., 2017).

Let $\mathcal{E}_t \triangleq (\epsilon_{1t}, \ldots, \epsilon_{Nt})^{\intercal} \in \mathbb{R}^N$, and $\mathbf{y}_t \triangleq (y_{1t}, \ldots, y_{Nt})^{\intercal} \in \mathbb{R}^N$ be the vectorized forms of the error term and the response obtained by stacking the corresponding terms for the N individuals. Similarly, let $\mathbf{Z}_t \triangleq [\mathbf{z}_{1t}, \ldots, \mathbf{z}_{Nt}]^{\intercal} \in \mathbb{R}^{N \times p}$ be the matrix of covariates whose ith row is $\mathbf{z}_{it}$, the covariate for the ith subject. Then the above model can be expressed in the vector and matrix notations as $\mathbf{y}_{t+1} = \alpha \mathbf{y}_t + \theta \mathcal{L}_{\mathbf{A}} \mathbf{y}_t + \mathbf{Z}_t \gamma + \mathcal{E}_{t+1}$.

Latent position random graph models assume that a network is created by random edges independently sampled with probabilities that are functions of distance kernels between latent positions of vertices in some underlying latent space. In a K-dimensional random dot product graph (RDPG), this distance kernel is the dot product of two K-dimensional latent vectors.

Let $\mathcal{M}$ be a subset of $\mathbb{R}^K$ such that $\mathbf{x}_1{}^{\intercal}\mathbf{x}_2 \in [0, 1]$ for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{M}$. Let $K \leq N$ and ρ_N be a sequence such that $\rho_N \in (0, 1]$ for all N. Then, $\mathbf{A}$ is said to follow a random dot product graph with latent positions $\mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_N]^{\intercal} \in \mathcal{M}^N$ and sparsity factor ρ_N, denoted by $\mathbf{A} \sim \text{RDPG}(\rho_N, \mathbf{X})$, if $a_{ij} \overset{ind.}{\sim} \text{Bernoulli}(\rho_N \mathbf{x}_i{}^{\intercal}\mathbf{x}_j)\mathbb{I}_{\{i \neq j\}}$ and $a_{ji} \triangleq a_{ij}$ for $\quad 1 \leq i \leq j \leq N$. Using matrix notations, we have $\mathbb{E}(\mathbf{A}) = \mathbf{P}$ for $\mathbf{P} \triangleq \rho_N \mathbf{X}\mathbf{X}^{\intercal}$. The role of ρ_N is to control the sparsity of the network. For example, the expected degrees are $\sum_{j=1}^{N} p_{ij} \in [0, N\rho_N]$ for every $i = 1, ..., N$. Hence, when $\rho_N = 1 \ \forall N$, the resulting graph is dense in the sense that the expected number of edges $\sum_{i<j} p_{ij} \sim N^2$. If $\rho_N \to 0$ as $N \to \infty$, the graph becomes sparse in the sense that $\sum_{i<j} p_{ij} = o(N^2)$ (Xie and Xu, 2023).

2.2 Embedding Network Autoregression

Now we define our model, which augments the NAR model with latent variables that are common for both the model of the univariate responses and the model for the network. Accordingly, in this section, we further assume the network is generated from an RDPG model, $\mathbf{A} \sim \text{RDPG}(\rho_N, \mathbf{X})$. We assume that every vertex in $\mathbf{A}$ is connected to at least one vertex.

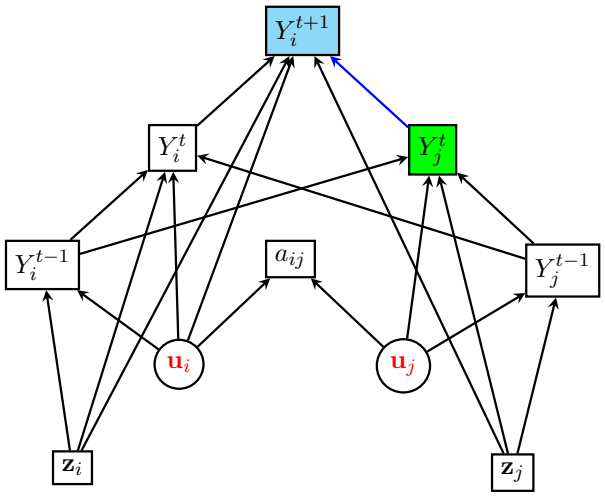
The eigenvalue decomposition of $\mathbf{P}$ can be expressed as $\mathbf{P} = \mathbf{V}\mathbf{S}\mathbf{V}^\intercal$, where $\mathbf{V} \in \mathbb{R}^{N\times K}$ contains orthonormal eigenvectors for K largest magnitude eigenvalues, and $\mathbf{S}$ is a diagonal matrix containing corresponding eigenvalues $s_1(\mathbf{P}) \geq \cdots \geq s_K(\mathbf{P}) > 0$. Let $\mathbf{U}_v \triangleq \mathbf{V}$ and $\mathbf{U}_e \triangleq \mathbf{V}\mathbf{S}^{1/2}$. The generic notation $\mathbf{U}$ then denotes one of these two forms, i.e., $\mathbf{U} \in \{\mathbf{U}_v, \mathbf{U}_e\}$. Each row of $\mathbf{U}$, denoted by $\mathbf{U}_i$, contains information about the associated latent variables of each vertex. Then, we define the ENAR model as a set of two models as follows,

$$y_{i,t+1} = \alpha y_{it} + \theta \sum_{j\neq i} \ell_{ij} y_{jt} + \mathbf{u}_i{}^\intercal \beta + \mathbf{z}_{it}{}^\intercal \gamma + \epsilon_{it}, \quad a_{ij} \overset{ind.}{\sim} \text{Bernoulli}(\rho_N \mathbf{x}_i{}^\intercal \mathbf{x}_j)\mathbb{I}_{\{i\neq j\}}. \tag{3}$$

As before, we can write the model in vector and matrix notation as follows: $\mathbf{y}_{t+1} = \alpha \mathbf{y}_t + \theta \mathcal{L}_\mathbf{A} \mathbf{y}_t + \mathbf{U}\beta + \mathbf{Z}_t\gamma + \mathcal{E}_{t+1}$, where now the parameter β denotes the global effect of latent variables, and $\mathbf{U}$ is either the matrix of eigenvectors $\mathbf{U}_v$ or the matrix of spectral embedding vectors $\mathbf{U}_e$. Finally, assume that $\mathbf{Z}_t$ is independent of $\{\mathcal{E}_{t+1}, \mathcal{E}_t, \ldots\}$ and $\{\mathbf{y}_t, \mathbf{y}_{t-1}, \ldots\}$ for each t and $\mathbf{A}$ is generated independently with the rest of the random components across t.

2.2.1 Finite time model for peer influence

Figure 1: Causal diagram for peer effects



The ENAR model is motivated by the problem of estimating causal peer influence adjusting for latent homophily in the settings of longitudinal data, but with finite time points (e.g., $T = 2$). The model is identical to the ENAR model described in the earlier section, except we do not have a time series, but only a finite number of time points, and the asymptotic setup is for $N \to \infty$.

The problem can be illustrated with a causal diagram similar to McFowland III and Shalizi (2023); Nath et al. (2025); O'Malley et al. (2014). In the causal diagram Pearl (2009) in Figure 1, the observed variables are represented by rectangles while the unobserved or latent variables are represented by circles. Intuitively, the causal peer influence is the causal effect of the outcome of a "peer" who is linked in the network on the outcome of an individual in the next time period. The problem of estimating causal peer influence is then estimating the causal effect corresponding to the path $Y_j^t \to Y_i^{t+1}$, conditioning on the observed network links a_{ij}'s. We can write the causal estimand we are interested in using the notations of do-interventions as follows (Nath et al., 2025; Egami and Tchetgen Tchetgen, 2024; Sridhar et al., 2022),

$$\tau_{ij}(h, h') = \mathbb{E}[Y_i^{t+1} | \operatorname{do}(Y_j^t = h), a_{ij} = 1] - \mathbb{E}[Y_i^{t+1} | \operatorname{do}(Y_j^t = h'), a_{ij} = 1].$$

However, as the causal diagram shows, there are already several backdoor paths open. Moreover, conditioning on a_{ij} opens additional backdoor paths involving $\mathbf{u}_i$ and $\mathbf{u}_j$ since a_{ij} is a collider variable in those paths. Intuitively, we have the same latent variables that are responsible for individuals' network selections (homophily) and also affect the outcomes. It is our conditioning on the observed network that opens those backdoor paths involving the latent variables that lead to confounding. Below, we enumerate all the backdoor paths as follows. These are, (1) $Y_j^t \leftarrow Y_i^{t-1} \rightarrow Y_i^t \rightarrow Y_i^{t+1}$, (2) $Y_j^t \leftarrow \mathbf{u}_j \rightarrow a_{ij} \rightarrow Y_i^{t+1}$, (3) $Y_j^t \leftarrow \mathbf{u}_j \rightarrow a_{ij} \leftarrow \mathbf{u}_i \rightarrow Y_i^{t+1}$, (4) $Y_j^t \leftarrow \mathbf{u}_j \rightarrow a_{ij} \leftarrow \mathbf{u}_i \rightarrow Y_i^t \rightarrow Y_i^{t+1}$, (5) $Y_j^t \leftarrow \mathbf{u}_j \rightarrow a_{ij} \leftarrow \mathbf{u}_i \rightarrow Y_i^{t-l} \ldots \rightarrow Y_i^{t-1} \rightarrow Y_i^t \rightarrow Y_i^{t+1}$, and several paths that involve $\mathbf{z}_i$ and $\mathbf{z}_j$. The first backdoor path can be closed by conditioning on Y_i^t, while the backdoor paths involving $\mathbf{z}_i$ and $\mathbf{z}_j$ can be closed by conditioning on those observed covariates. However, as we can see from the other open backdoor paths, we need to condition on $\mathbf{u}_i$ and $\mathbf{u}_j$ to close all of those backdoor paths. Therefore, the linear structural equation model that we want to estimate is $\mathbf{y}_{t+1} = \alpha \mathbf{y}_t + \theta \mathcal{L}_{\mathbf{A}} \mathbf{y}_t + \mathbf{U}\beta + \mathbf{Z}_t\gamma + \mathcal{E}_{t+1}$.

Here θ is our target peer influence parameter that we want to estimate consistently. We emphasize that $\mathbf{U}$ is a latent variable that is not observed. We theoretically show that we *can* estimate the structural peer effect parameter θ consistently under our modeling assumptions with the methodology described below.

Remark on $\mathcal{L}_{\mathbf{A}}$ Careful readers may note that our definition of $\mathcal{L}_{\mathbf{A}}$ for ENAR differs from that used in Zhu et al. (2017); McFowland III and Shalizi (2023). Specifically, those studies adopt a row-normalized adjacency matrix, $\mathcal{L}_{\mathbf{A}} = \mathcal{D}^{-1}\mathbf{A}$, to incorporate the network (peer) influence, whereas ENAR employs a symmetrically normalized version, $\mathcal{L}_{\mathbf{A}} = \mathcal{D}^{-1/2}\mathbf{A}\mathcal{D}^{-1/2}$, corresponding to the graph Laplacian. Importantly, all procedures and theoretical results established for the ENAR estimators remain valid under the row-normalized formulation as well. This is because both forms preserve two key properties: a spectral radius bounded by one—ensuring the stationarity of the resulting time series, and diagonalizability—guaranteeing a well-defined and stable autocovariance structure. The Table 3 lists the different models we consider in this paper for easy reference of readers.

Table 3: Table of models, $i = 1, ..., N$ and $t = 1, ..., T$

Model	**Full Name**	**Data Generating Model**
ENR	Embedding Network Regression	$y_i = \alpha + \mathbf{u}_i^\intercal \beta + \mathbf{z}_i^\intercal \gamma + \epsilon_i$
NAR	Network Vector Autoregression (Zhu et al., 2017)	$y_{i,t+1} = \alpha y_{it} + \theta \sum_{j \neq i} \ell_{ij} y_{jt} + \mathbf{z}_{it}^\intercal \gamma + \epsilon_{it}$
ENAR	Embedding Network Autoregression	$y_{i,t+1} = \alpha y_{it} + \theta \sum_{j \neq i} \ell_{ij} y_{jt} + \mathbf{u}_i^\intercal \beta + \mathbf{z}_{it}^\intercal \gamma + \epsilon_{it}$
CNAR	Community Network Autoregression	Equation (8) with $\Lambda = (0)$ in Chen et al. (2023)

3. Estimation

3.1 Spectral Embedding under RDPG

The ENAR model embeds the latent variables using the population spectra of an expected adjacency matrix, which represents the connection probability between entities. One challenge in estimating the parameters of the ENAR model is that these latent vectors $\mathbf{U}$ are unobservable. To address this, we can utilize the asymptotic properties governing the differences between population spectra and sample spectra. Specifically, we employ the following adjacency spectral embedding procedure to estimate the unknown latent vectors.

Definition 1 (Xie and Xu, 2023) *The adjacency spectral embedding (ASE) of* $\mathbf{A}$ *into* $\mathbb{R}^K$ *is given by* $\hat{\mathbf{V}}|\hat{\mathbf{S}}|^{1/2}$ *for* $(|\hat{\mathbf{S}}|)_{ij} \triangleq |(\hat{\mathbf{S}})_{ij}|$ *where* $\hat{\mathbf{S}} \in \mathbb{R}^{K\times K}$ *is the diagonal matrix with the* K *largest magnitude eigenvalues of* $\mathbf{A}$ *and* $\hat{\mathbf{V}} \in \mathcal{O}_{N,K}$ *is the corresponding eigenvectors of* $\mathbf{A}$.

In the analysis of ASE, the eigengap $s_K(\mathbf{P})$ plays a crucial role. Intuitively, the eigengap measures how strongly the network's latent structure (e.g., communities or continuous factors) stands out from the stochasticity of an observed network $\mathbf{A}$. A larger eigengap means that the top K eigenvectors of the population connectivity $\mathbf{P}$ are well separated from the rest, ensuring stable recovery of the latent positions.

There are many concentration results available for this subspace approximation problem, and we highlight the works devoted to the analysis of entrywise approximation of adjacency spectral embedding (ASE) in the case of the RDPG model (Tang et al., 2017; Athreya et al., 2017; Tang and Priebe, 2018; Cape et al., 2019b; Xie and Xu, 2023; Rubin-Delanchy et al., 2022; Xie, 2024). In Proposition 23 in Section C.3 of the Appendix, we prove a concentration result for ASE in terms of the spectral norm instead of the more commonly used Frobenius norm by compiling known proof techniques from the literature while tracking the order in terms of K. We propose to use $\hat{\mathbf{U}}_v \triangleq \hat{\mathbf{V}}$ and $\hat{\mathbf{U}}_e \triangleq \hat{\mathbf{V}}|\hat{\mathbf{S}}|^{1/2}$ as the approximations of unknown latent variables modeled by eigenvector and embedding, respectively.

As various notations related to $\mathbf{U}$ are used throughout the paper, Table 4 provides a summary for clarity.

3.2 Least Squares Estimation of Parameters

The stationarity of the ENAR process with growing N is discussed in Appendix C.1. From the equation of the model (3), we obtain the linear regression representation $y_{i,t+1} = \mathbf{w}_{i,t}^\intercal \mu + \epsilon_{i,t+1}$, where $\mathbf{w}_{it}^\intercal \triangleq (\mathbf{u}_i^\intercal, y_{it}, \ell_{i\cdot}^\intercal \mathbf{y}_t, \mathbf{z}_{it}^\intercal)$ and our parameters of interest, $\mu \triangleq (\beta^\intercal, \alpha, \theta, \gamma^\intercal)^\intercal \in \mathbb{R}^{K+p+2}$. Thus, the auto-regression of the networked measurements at time $t+1$ can be written as

$$\mathbf{y}_{t+1} = \mathbf{W}_t \mu + \mathcal{E}_{t+1},$$

where $\mathbf{W}_t = [\mathbf{w}_{1t}, ..., \mathbf{w}_{Nt}]^\intercal$. We can further collect the entire time series as $\mathbf{y} \triangleq (\mathbf{y}_1{}^\intercal, ..., \mathbf{y}_T{}^\intercal)^\intercal$, $\mathbf{W} \triangleq [\mathbf{W}_0{}^\intercal, ..., \mathbf{W}_{T-1}{}^\intercal]^\intercal$, and $\mathcal{E} \triangleq (\mathcal{E}_1{}^\intercal, ..., \mathcal{E}_T{}^\intercal)^\intercal$, thereby obtaining the representation $\mathbf{y} = \mathbf{W}\mu + \mathcal{E}$ in $\mathbb{R}^{NT}$. Note that $\mathbf{W}$ contains the population latent variables $\mathbf{U}$ from the network model. Therefore, it is interpreted as the population design matrix for the ENAR model (Chen et al., 2023). Utilizing the estimated latent variables $\hat{\mathbf{U}}$, we can obtain the approximated version $\hat{\mathbf{W}}_t = [\hat{\mathbf{w}}_{1t}, ..., \hat{\mathbf{w}}_{Nt}]^\intercal$ for $\hat{\mathbf{w}}_{it}^\intercal \triangleq (\hat{\mathbf{u}}_i^\intercal, y_{it}, \ell_{i\cdot}{}^\intercal \mathbf{y}_t, \mathbf{z}_{it}{}^\intercal)$ and $\hat{\mathbf{W}}$ accordingly.

Table 4: Notations for $\mathbf{U}$

Symbol	Meaning
$\mathbf{U}$	A generic $N \times K$ matrix containing K latent variables
$\mathbf{U}_v, \hat{\mathbf{U}}_v$	$\mathbf{V}$ and $\hat{\mathbf{V}}$, respectively
$\mathbf{U}_e, \hat{\mathbf{U}}_e$	$\mathbf{V}\mathbf{S}^{1/2}$ and $\hat{\mathbf{V}}\lvert\hat{\mathbf{S}}\rvert^{1/2}$, respectively
$\mathbf{U}_{:j}$	A generic $N \times j$ matrix containing first j latent variables
$\mathbf{U}_{j:j+k}$	A generic $N \times k$ matrix containing next k latent variables
$\mathbf{V}_{i:j}, \hat{\mathbf{V}}_{i:j}$	$N \times (j-i)$ matrix with $i+1, ..., j$-th leading eigenvectors of $\mathbf{P}, \mathbf{A}$ for $0 \leq i \leq j$
$\mathbf{S}_{i:j}\hat{\mathbf{S}}_{i:j}$	Diagonal matrix with $i+1, ..., j$-th leading eigenvalues of $\mathbf{P}, \mathbf{A}$ for $0 \leq i \leq j$
$\mathbf{U}_{i:j,e}, \hat{\mathbf{U}}_{i:j,e}$	Truncated spectral embeddings $\mathbf{V}_{i:j}\mathbf{S}_{i:j}^{1/2}, \hat{\mathbf{V}}_{i:j}\lvert\hat{\mathbf{S}}_{i:j}\rvert^{1/2}$

This expression naturally motivates the least squares estimation of μ. Therefore, we target the least squares estimator

$$\hat{\mu} = \left(\hat{\mathbf{W}}^\intercal\hat{\mathbf{W}}\right)^{-1}\hat{\mathbf{W}}^\intercal\mathbf{y} = \left(\hat{\beta}^\intercal, \hat{\alpha}, \hat{\theta}, \hat{\gamma}^\intercal\right)^\intercal$$

and study its asymptotic properties as an estimator of μ. We establish the asymptotic distribution of the estimator through non-asymptotic concentration results and the martingale central limit theorem (Hall and Heyde, 2014; Zhu et al., 2017; Chen et al., 2023). To avoid notation abuse, we write $\hat{\mu}_v$ and $\hat{\mu}_e$ to denote the least squares estimators under embedding of $\hat{\mathbf{U}}_v = \hat{\mathbf{V}}$ and $\hat{\mathbf{U}}_e = \hat{\mathbf{V}}\lvert\hat{\mathbf{S}}\rvert^{1/2}$, respectively.

4. Large Sample Results

We next develop a theory on consistency and asymptotic normality of the least squares estimator of the ENAR model under an asymptotic setup where we always assume $N \to \infty$, but consider both the finite T case and the case of $T \to \infty$ separately. We start with the results for the ENAR model with eigenvectors from the RDPG network model, and then discuss the results for the case of spectral embedding vectors. In this section, we write $\mu_l^H \triangleq (\beta_l^\intercal\mathbf{H}_l, \alpha_l, \theta_l, \gamma_l^\intercal)^\intercal$, where $\mathbf{H}_l \in \mathcal{O}_K$ is an arbitrary matrix with orthonormal columns from Proposition 23 and $l \in \{v, e\}$. It is well known that the multiplicative latent variables can only be estimated from a network up to an ambiguity of such a matrix from the class $\mathcal{O}_K$ (Hoff et al., 2002; Athreya et al., 2017). Therefore, the parameter β can only be recovered up to the ambiguity of $\mathbf{H}^\intercal\beta$. We differentiate between μ_v and μ_e as they represent different modeling choices.

For the definitions of asymptotic precision matrices stated in the theorems below, we refer readers to Table 5.

4.1 Growing T results

The first theorem stated below shows the asymptotic normality of the least squares estimator when we embed eigenvectors $\mathbf{U}_v = \mathbf{V}$, both N and T grow, and K also grows along with

Table 5: Asymptotic precision matrices specified in large-sample results of ENAR.

Symbol	**Structure**	**Definitions**
Σ_v	$\begin{bmatrix} \mathbf{I}_K & (0) & (0) \\ & \Sigma_{\alpha\theta} & (0) \\ & & \Sigma_z \end{bmatrix}$	$\Sigma_{\alpha\theta} \triangleq \begin{bmatrix} \tau_2 & \tau_{23} \\ & \tau_3 \end{bmatrix}$, $\tau_2 \triangleq \lim_{N\to\infty} \frac{1}{N}\mathbb{E}[\mathrm{tr}(\Gamma)]$, $\tau_{23} \triangleq \lim_{N\to\infty} \frac{1}{N}\mathbb{E}[\mathrm{tr}(\Gamma\mathcal{L}_{\mathbf{A}})]$, $\tau_3 \triangleq \lim_{N\to\infty} \frac{1}{N}\mathbb{E}[\mathrm{tr}(\Gamma{\mathcal{L}_{\mathbf{A}}}^2)]$; Σ_z: Covariance of $\mathbf{z}$
$\Sigma_{-u,v}$	$\begin{bmatrix} \Sigma_{\alpha\theta} & (0) \\ & \Sigma_z \end{bmatrix}$	Submatrix of Σ_v that omits the covariance of $\hat{\mathbf{U}}_v$
Σ_e	$\begin{bmatrix} \mathbf{E} & \mathbf{B} & (0) \\ & \Sigma_{\alpha\theta,e} + \Sigma_{\alpha\theta} & (0) \\ & & \Sigma_z \end{bmatrix}$	$\mathbf{E} \triangleq \lim_{N\to\infty} \frac{1}{N\rho_N}\mathbf{S}$; $\mathbf{B} = [\mathbf{b}_{12}, \mathbf{b}_{13}] \in \mathbb{R}^{K\times 2}$ where $\Vert\mathbf{b}_{12}\Vert$, $\Vert\mathbf{b}_{13}\Vert$ are $c_j\sqrt{\rho_\infty}$ for constants $c_1, c_2 > 0$; $\Sigma_{\alpha\theta,e} \triangleq \lim_{N\to\infty} \mathbb{E}\begin{bmatrix} \frac{1}{N}{\varphi_e}^\intercal\varphi_e & \frac{1}{N}{\varphi_e}^\intercal\mathcal{L}_{\mathbf{A}}\varphi_e \\ & \frac{1}{N}{\varphi_e}^\intercal\mathcal{L}_{\mathbf{A}}^2\varphi_e \end{bmatrix}$; $\varphi_e \triangleq (\mathbf{I}_N - \mathbf{G})^{-1}\mathbf{U}_e\beta_e$

them. The proof of this theorem and other theorems from this section are in Appendix section C. A similar theorem for the fixed K case is given in Appendix B.1.

Theorem 2 *Assume that* $|\alpha_v| + |\theta_v| < 1$ *and* $\Vert\beta_v\Vert = O(1)$ *as* $K \to \infty$. *Furthermore, assume that* $N\rho_N = \omega\left(T + \log^4 N\right)$, $N = \omega(K^2)$, *and* $T = \omega(\log K)$. *For a positive integer* m, *suppose we have an* $m \times (K + p + 2)$ *matrix* $\mathbf{Q}_K$ *such that* $\Vert\mathbf{Q}_K\Vert = O(1)$ *as* $K \to \infty$. *If we define* $\Lambda_v \triangleq \lim_{K\to\infty} \mathbf{Q}_K{\Sigma_v}^{-1}{\mathbf{Q}_K}^\intercal \in \mathbb{R}^{m\times m}$, *then, for* $\mathbf{D}_v \triangleq \sqrt{T}\,\mathrm{diag}\left(\mathbf{I}_K, \sqrt{N}\mathbf{I}_{p+2}\right)$ *we have, as* $N, T, K \to \infty$,

$$\mathbf{Q}_K\mathbf{D}_v\left(\hat{\mu}_v - \mu_v^H\right) \Rightarrow \mathcal{N}\left(\mathbf{o}_m, \sigma^2\Lambda_v\right).$$

The condition $|\alpha_v| + |\theta_v| < 1$ is required to ensure that $\mathbf{y}_t$ has the stationary distribution discussed in section C.1 as $N, T \to \infty$. It is noteworthy that to consistently estimate the parameters, the population network density $N\rho_N$ should grow faster than T and $\log^4 N$ in Theorem 2. First, the condition $N\rho_N = \omega(\log^4 N)$ ensures that the deviation $\mathbf{A} - \mathbf{P}$ of the observed network from its expectation is well controlled by the overall network density $N\rho_N$. Intuitively, this means each node has enough connections for $\hat{\mathbf{U}}_v$ to reliably recover $\mathbf{U}_v$, making the measurement error negligible. Moreover, the condition $N\rho_N = \omega(T)$ guarantees that the network information grows faster than the time horizon, preventing the estimation bias of β_v from accumulating across time and ensuring stable inference.

The growth rate of network density needed in Theorem 2 matches the rate in Cape et al. (2019a) as long as $T = O(\log^4 N)$. If T grows faster than this, Theorem 10 requires a better concentration of $\hat{\mathbf{U}}$ for ENAR estimation to remain accurate. The convergence rate of $\sqrt{T}$

for the latent position effects $\hat{\beta}_v$ is slower than $\sqrt{NT}$ for the rest, which matches the rates obtained in Chen et al. (2023) for their CNAR model.

Since $\frac{1}{N}\mathbf{X}^\intercal\mathbf{X}$ converges to a finite, positive definite matrix under our RDPG regime (Xie and Xu, 2023), we have $s_K(\mathbf{P}) = \Theta(N\rho_N)$. Therefore, our condition on $N\rho_N$ can be equivalently represented using eigengap $s_K(\mathbf{P})$ to compare with the result in Chen et al. (2023). The condition on $N\rho_N$ in Theorem 2 then is that $s_K(\mathbf{P}) = \omega\left(T + \log^4 N\right)$. This improves upon the rate $s_K(\mathbf{P}) = \omega\left(K\sqrt{TN\rho_N}\right)$, i.e., $N\rho_N = \omega\left(K^2T\right)$ presented in the context of CNAR model in Chen et al. (2023). The improvement by a factor of K^2 is substantial when K is large.

Further, to consistently estimate the parameters, we require $N = \omega(K^2)$ and $T = \omega(\log K)$. These guarantee a consistent estimation of the parameter when its dimension grows as $K \to \infty$, specifying how the amount of network information (N) and temporal information (T) must increase relative to model complexity for asymptotic normality to hold. This rate can be compared to the required rates of $N = \omega(K^4)$ and $T = \omega(K^2 \log K)$ in Chen et al. (2023). Hence, consistent estimation in the CNAR model Chen et al. (2023) requires $O(K^2)$ times more sample size both in terms of N and T. Therefore, this represents a substantial relaxation of conditions. These reductions are largely due to the ENAR model having $\mathbf{U}$ entering the model in an additive form as opposed to a multiplicative form in the CNAR model, and consequently, requiring us to estimate fewer parameters attached to latent variables.

4.2 Finite T peer effect results

Now, we prove asymptotic results for the model in the finite T case, which is appropriate for the problem of causal peer influence estimation. This finite T case was not studied in earlier works of Zhu et al. (2017); Chen et al. (2023) on network autoregressive models. Although motivated by the peer effect literature, the recent work of Wu and Leng (2023) does not study the finite T case either. Note that we no longer require $|\alpha| + |\theta| < 1$ in the finite T case. Distinctively from 2 (and its fixed K version Theorem 10 in the Appendix B.1), the consistency of $\hat{\mu}$ given finite T is different as the estimation error for the latent position effects, $\hat{\beta}_v - \mathbf{H}_v{}^\intercal\beta_v$, will only be bounded in probability. However, our result still shows the peer influence parameter θ_v along with other parameters, namely, α_v and γ_v, which we are typically interested in inferring are all $\sqrt{N}$ consistent. The result for growing K is as follows.

Theorem 3 *Assume that* $\|\beta_v\| = O(1)$ *as* $K \to \infty$, $N\rho_N = \omega(\log^4 N)$, *and* $N = \omega(K^2)$. *Partition the parameter vector as* $\mu_v = (\beta_v^\intercal, \mu_{-\beta,v}{}^\intercal)^\intercal$ *and* $\hat{\mu}_v$ *accordingly as well. For a positive integer* m, *suppose we have an* $m \times (K + p + 2)$ *matrix* $\mathbf{Q}_K$ *such that* $\|\mathbf{Q}_K\| = O(1)$ *as* $K \to \infty$. *Then,* $\mathbf{Q}_K\left(\hat{\beta}_v - \mathbf{H}_v{}^\intercal\beta_v\right) = O_{\mathbb{P}}(1)$ *and we have*

$$\sqrt{N}\left(\hat{\mu}_{-\beta,v} - \mu^H_{-\beta,v}\right) \Rightarrow \mathcal{N}\left(\mathbf{o}_{p+2}, \frac{\sigma^2}{T}{\Sigma_{-u,v}}^{-1}\right)$$

as $N, K \to \infty$.

A similar theorem for the fixed K case is given in Appendix B.1. As we limit T to be finite, $N\rho_N$ is allowed to grow at less restrictive rates compared to the case where T diverges.

However, we still require the number of dimensions to grow at the same rate with the network size N as described in Theorem 2. We discuss the peer effect problem more in the next section with spectral embeddings.

We also consider a regression model for data observed at a single time point ($T = 1$). Here, our goal is to model a response observed only once as a function of several covariates or predictors while controlling for latent homophily variables that may be correlated with the covariates whose effects we want to estimate. Accordingly, we propose the Embedding Network Regression (ENR) model as, $\mathbf{y} = \alpha\mathbf{1}_N + \mathbf{U}\beta + \mathbf{Z}\gamma + \mathcal{E}$. This model has appeared in various forms previously in the literature Fosdick and Hoff (2015); He and Hoff (2019); Le and Li (2022); Lunde et al. (2025); Cai et al. (2021); Wang et al. (2024); Hayes et al. (2025). In addition, this model is a network analog of the popular spatial confounding regression model used in spatial data analysis Guan et al. (2023). While this model has been theoretically studied before under different settings Le and Li (2022); Wang et al. (2024); Lunde et al. (2025), the results we develop in this paper can also be applied to this model as corollaries. In Appendix B.2, we provide two corollaries that obtain results for the embedding network regression (ENR) problem.

4.3 Results with spectral embeddings for the finite T peer effect problem

The previous section comprehensively dealt with consistency and asymptotic normality for the ENAR model with eigenvectors for both the problems of network time series and peer effect. However, as will be shown later in Section 5.2, the model may not be appropriate to induce the confounding necessary for the peer effect problem. Intuitively, because eigenvectors are constrained to have columns whose ℓ_2 norms are 1 by definition, the entries of the design matrix made with them decreases at $O(N^{-1/2})$ rate as N increases. Hence with the typical assumption that coefficients β do not grow with N, the overall contribution of $\mathbf{U}\beta$ to the outcome Y asymptotically vanishes even if β is not 0. Therefore, asymptotically, the effect of ignoring the eigenvectors in the model on the peer effect parameter estimates ("omitted variable bias") is negligible. Hence, this model is not appropriate for the peer effect problem, where we expect non-trivial omitted variable bias due to the omission of latent variables. Therefore, we next develop a theory on consistency and asymptotic normality of the least squares estimator for the ENAR model with spectral embedding vectors. In this case, we consider the finite T case, while $N \to \infty$.

The first two theorems stated below show the asymptotic normality of the least squares estimator in two cases as N grows: first when the number of latent dimensions K is fixed, and second when K also grows with N. Recall from Proposition 23 that $s_K(\mathbf{P})$ denotes the smallest non-zero eigenvalue of $\mathbf{P}$. The proof is given in section C.6.1.

Theorem 4 *Assume that* $N\rho_N = \omega\left(\log^4 N\right)$. *Then, for* $\mathbf{D}_e \triangleq \sqrt{N}\,\mathrm{diag}\left(\sqrt{\rho_N}\mathbf{I}_K, \mathbf{I}_{p+2}\right)$, *we have, as* $N \to \infty$,

$$\mathbf{D}_e\left(\hat{\mu}_e - \mu_e^H\right) \Rightarrow \frac{1}{\sqrt{T}} \cdot \mathcal{N}\left(\mathbf{b}, \sigma^2 {\Sigma_e}^{-1}\right)$$

for $\mathbf{b} \triangleq (\mathbf{b}_1{}^\intercal, \mathbf{b}_2{}^\intercal, \mathbf{o}_p{}^\intercal)^\intercal$ *for some fixed* $\mathbf{b}_1 \in \mathbb{R}^K, \mathbf{b}_2 \in \mathbb{R}^2$ *such that* $\|\mathbf{b}_1\| < c_1$ *and* $\|\mathbf{b}_2\| < \sqrt{c_2\rho_\infty}$ *for some constants* $c_1, c_2 > 0$.

This theorem establishes the consistency and asymptotic normality of the peer effect parameter. This result, therefore, expands the results of McFowland III and Shalizi (2023); Nath et al. (2025) on asymptotically unbiased estimation to $\sqrt{N}$ consistency in the sense that the estimation error is $O_{\mathbb{P}}\left(N^{-1/2}\right)$. We note that there is a small $O\left(N^{-1/2}\rho_\infty^{1/2}\right)$ asymptotic bias in the estimation of the peer effect parameter, which vanishes for sparse networks, but remains for dense networks. As we will show in Section 5.2, the omitted variable bias in this model is substantial, and peer effect parameter estimation is not possible at $\sqrt{N}$-consistent rate (and even maybe inconsistent) if one ignores the latent variable.

Remark on source of bias in Theorem 4 The bias for the peer effect in Theorem 4 originates from two sources: measurement error and mean vector scale under $\mathbf{U} = \mathbf{U}_e$. First, the measurement error $\left\|\hat{\mathbf{U}}_e - \mathbf{U}_e\mathbf{H}_e\right\| = O(1)$ *whp.* obtained from spectral embedding of the random graph. Second, the mean vector

$$\varphi_t \triangleq \mathbb{E}(\mathbf{y}_t) = \mathbb{E}(\mathbf{G}^t\mathbf{y}_0) + \mathbb{E}[(\mathbf{I} - \mathbf{G})^{-1}(\mathbf{I}_N - \mathbf{G}^t)]\mathbf{U}_e\beta_e$$

which has the norm of order $\|\varphi_t\| = O(\sqrt{N\rho_N})$ from the order of $\|\mathbf{U}_e\beta_e\|$. Then, given A, $\sqrt{N}(\hat{\theta}_e - \theta_e)$ approximates a normal distribution with the mean as

$$\frac{1}{\sqrt{N}T}\sum_{t=1}^{T}\varphi_t^\intercal\mathcal{L}_{\mathbf{A}}(\hat{\mathbf{U}}_e - \mathbf{U}_e\mathbf{H}_e)\mathbf{H}_e^\intercal\beta_e.$$

Therefore the asymptotic normality includes a bias component of the order of $O(\sqrt{\rho_N})$ for finite T. In contrast, the bias in Theorem 9 for the case when latent variables are omitted is $O(\sqrt{N}\rho_N)$. Therefore if the network density ρ_N converges to 0 at any rate, then this bias goes away in Theorem 4, but does not go away in Theorem 9 unless the network is ultra-sparse. This asymptotic bias is 0 unless the network is extremely dense, i.e., every vertex is connected to almost every other vertex for $O(N^2)$ connections across the network. This is not typically the case for most real-world networks.

Next, we consider the case where K grows along with N.

Theorem 5 *Assume that* $\|\beta_e\| = O(1)$ *as* $K \to \infty$. *Furthermore, assume that* $N\rho_N = \omega\left(\log^4 N\right)$ *and* $N = \omega\left(K^2 \log K\right)$. *For a positive integer* m, *suppose we have a matrix* $\mathbf{Q}_K \triangleq [\mathbf{Q}_1 \mid \mathbf{Q}_2] \in \mathbb{R}^{m\times(K+p+2)}$ *for* $\mathbf{Q}_1 \in \mathbb{R}^{m\times(K+2)}, \mathbf{Q}_2 \in \mathbb{R}^{m\times p}$ *such that* $\|\mathbf{Q}_1\| = O\left(\frac{\log K}{\sqrt{K}}\right)$ *and* $\|\mathbf{Q}_2\| = O(1)$ *as* $K \to \infty$. *If we define* $\Lambda_e \triangleq \lim_{K\to\infty}\mathbf{Q}_K{\Sigma_e}^{-1}{\mathbf{Q}_K}^\intercal \in \mathbb{R}^{m\times m}$, *then we have, as* $N, K \to \infty$,

$$\mathbf{Q}_K\mathbf{D}_e\left(\hat{\mu}_e - \mu_e^H\right) \Rightarrow \mathcal{N}\left(\mathbf{o}_m, \frac{\sigma^2}{T}\Lambda_e\right).$$

Compared to Theorem 3, we require that N grows faster by a $\log K$ factor. This is for attaining the convergence of the asymptotic covariance (see Lemma 29). As evident from the order of $\|\mathbf{A}_1\|$, now the convergence rates of $\hat{\beta}_e$ and $\hat{\alpha}_e, \hat{\theta}_e$ are slightly slower by a factor of $\frac{\log K}{\sqrt{K}}$ from $\sqrt{N}$-consistency.

5. Selection of K and results under misspecification

Until now, we have assumed the number of latent vectors to include in the ENAR model, K, is known. However, in practice, K is unknown and must be obtained from data. In this section, we provide several results for the selection of K and results for the cases of over-selection and under-selection. We develop a strategy for guaranteed *over-estimating* of K, based on a new model selection criterion developed in this paper. The reason for focusing on guarantee for non-under-selection is that we show the consistency and asymptotic normality of some model parameters, including the peer effect parameter, hold under the model misspecification in terms of overestimating K. Throughout the section, we assume K does not grow with N and T.

5.1 Over-selection via information criterion

Our goal in this section is to develop an information criterion (IC) that can guarantee non-under-selection of K. Since the true latent variables $\mathbf{U}$ and the true latent dimension K are not observable in practice, we frame this as a model selection problem. This problem can be addressed using a criterion, leading to the estimation of the number of latent variables as $\hat{K}$. Then, we obtain an $N \times \hat{K}$ matrix $\hat{\mathbf{U}}_{:\hat{K}}$, which contains $\hat{K}$ latent variables aligned with the leading eigenvalues of the observed network $\mathbf{A}$.

Let K be the true dimension of the RDPG model and consequently, the true number of latent variables included in the ENAR model. Then, given $l = e, v$, consider the following criterion with a penalty function p_{NT} whose order depends on N, T for the dimension determination:

$$Cr_l(k) \triangleq \hat{R}_l^{(k)} + p_{NT}(k),\ k \in \{0, 1, ..., J_N\},\ J_N \in \mathbb{N}$$

where $\hat{R}_l^{(k)} \triangleq \frac{1}{T}\mathbf{y}^\intercal \left(\mathbf{I}_{NT} - \hat{\Pi}_l^{(k)}\right)\mathbf{y}$ for the projection $\hat{\Pi}_l^{(k)}$ onto the column space of $\mathbf{1}_T \otimes \hat{\mathbf{U}}_{:k,l}$. Here $\hat{\mathbf{U}}_{:k,l} \in \mathbb{R}^{N\times k}$ is either the eigenvector matrix $\hat{\mathbf{U}}_{:k,v}$ or the spectral embedding matrix $\hat{\mathbf{U}}_{:k,e}$ corresponding to the k largest-magnitude eigenvalues of $\mathbf{A}$. The term $\hat{R}_l^{(k)}$ is, therefore, a modified residual sum of squares (RSS), namely, the RSS one would obtain by regressing only on the latent variables. The penalty $p_{NT}(k)$ is a positive non-decreasing function in k. Then, our choice of k will be

$$\hat{K} \triangleq \arg\min_{k=0,1,...,J_N} Cr_l(k),$$

where J_N is a large number. We describe a general result on the class of penalty functions p_{NT} for which we have a theoretical guarantee on non-under-selection in Theorem 30 and Remark 31 in section D. Here we present a corollary that specifies a choice of penalty function p_{NT} and a choice of J_N that can be used in practice.

Corollary 6 *Suppose $K \le J_N$, $T = \omega(\log K)$, $K = o(N\rho_N)$, and $N\rho_N = \omega\left(\log^2 N\right)$. Set the penalty as $p_{NT}(k) \triangleq \frac{k^2}{N+T}$ and*

- *Choose $J_N = \lceil \sqrt{N} \rceil$ if $N = o(T)$.*
- *Choose any J_N such that $J_N = o(\sqrt{N})$ if $T = o(N)$.*

Then, we have

$$\mathbb{P}\left[\liminf_{N,K,T\to\infty}\left\{\hat{K}\geq K\right\}\right]=1.$$

While we have assumed a finite time model for the case of spectral embedding when estimating parameters, we do not put such an asymptotic condition when selecting the number of latent variables K, and the above theorem holds for both $\mathbf{U}_v$ and $\mathbf{U}_e$.

Algorithm 1 summarizes the complete estimation procedure for the ENAR model, including the computation of $\widehat{\mathrm{Var}}(\hat{\mu})$ based on the imputed design matrix $\hat{\mathbf{W}}$ and mean squared residuals $\hat{\sigma}^2$ to facilitate confidence interval construction for the peer effect and the rest of the parameters.

Algorithm 1 Embedding Network Autoregression (ENAR)

Require: Adjacency matrix $\mathbf{A}\in\{0,1\}^{N\times N}$ (undirected, hollow), responses $\{\mathbf{y}_t\}_{t=1}^{T}$ with $\mathbf{y}_t\in\mathbb{R}^N$, covariates $\{\mathbf{Z}_t\in\mathbb{R}^{N\times p}\}_{t=0}^{T-1}$, search bound J_N, penalty $p_{NT}(k)$, embedding type $l\in\{e,v\}$

Ensure: Selected dimension $\hat{K}$, parameter estimates $\hat{\mu}=(\hat{\beta}^\intercal,\hat{\alpha},\hat{\theta},\hat{\gamma}^\intercal)^\intercal$, standard errors, confidence intervals

Preprocessing:

Compute $\mathcal{D}=\mathrm{diag}(\mathbf{A}\mathbf{1}_N)$ and normalized Laplacian $\mathcal{L}_{\mathbf{A}}=\mathcal{D}^{-1/2}\mathbf{A}\mathcal{D}^{-1/2}$. Obtain $\mathbf{y}=(\mathbf{y}_1^\intercal,\ldots,\mathbf{y}_T^\intercal)^\intercal\in\mathbb{R}^{NT}$.

Obtain J_N leading eigenvalues/eigenvectors of $\mathbf{A}$: $(\hat{\mathbf{S}}_{:J_N},\hat{\mathbf{V}}_{:J_N})$.

Dimension Selection (for fixed embedding l):

if $l=e$ **then**

  Let $\hat{\mathbf{U}}_{:J_N,l}=\hat{\mathbf{V}}_{:J_N}|\hat{\mathbf{S}}_{:J_N}|^{1/2}$.

else if $l=v$ **then**

  Let $\hat{\mathbf{U}}_{:J_N,l}=\hat{\mathbf{V}}_{:J_N}$.

for $k=0,1,\ldots,J_N$ **do**

  if $k=0$ **then**

    Regress $\mathbf{y}$ on $\mathbf{F}_0\triangleq\mathbf{1}_{NT}$ to obtain $\hat{\beta}^{(0)}$.

  else

    Regress $\mathbf{y}$ on $\mathbf{F}_k\triangleq\mathbf{1}_T\otimes\hat{\mathbf{U}}_{:k,l}$ by OLS to obtain $\hat{\beta}^{(k)}$.

  Compute $\hat{R}_l^{(k)}=\frac{1}{T}\left\|\mathbf{y}-\mathbf{F}_k\hat{\beta}^{(k)}\right\|^2$ and criterion $Cr_l(k)=\hat{R}_l^{(k)}+p_{NT}(k)$.

Select $\hat{K}\leftarrow\arg\min_{0\leq k\leq J_N}Cr_l(k)$ and let $\hat{\mathbf{U}}\leftarrow\hat{\mathbf{U}}_{:\hat{K},l}$.

Estimation of selected model:

Construct $\hat{\mathbf{W}}_t=[\,\hat{\mathbf{U}}\,|\,\mathbf{y}_t\,|\,\mathcal{L}_{\mathbf{A}}\mathbf{y}_t\,|\,\mathbf{Z}_t\,]$ for $t=1,\ldots,T$, and stack $\hat{\mathbf{W}}=[\hat{\mathbf{W}}_1^\intercal,\ldots,\hat{\mathbf{W}}_T^\intercal]^\intercal$.

Regress $\mathbf{y}$ on $\hat{\mathbf{W}}$, including an intercept term, to obtain $\hat{\mu}=(\hat{\beta}^\intercal,\hat{\alpha},\hat{\theta},\hat{\gamma}^\intercal)^\intercal$ and residuals $\mathbf{r}=\mathbf{y}-\hat{\mathbf{W}}\hat{\mu}$.

Center the design matrix by setting $\hat{\mathbf{W}}\leftarrow(\mathbf{I}_{NT}-\frac{1}{NT}\mathbf{1}_{NT}\mathbf{1}_{NT}^\intercal)\hat{\mathbf{W}}$.

Estimate the covariance components as $\hat{\Sigma}_w=\frac{1}{NT}\hat{\mathbf{W}}^\intercal\hat{\mathbf{W}}$ and $\hat{\sigma}^2=\frac{1}{NT-d}\,\mathbf{r}^\intercal\mathbf{r}$, where $d=p+\hat{K}+2$.

Compute a covariance estimate $\widehat{\mathrm{Var}}(\hat{\mu})=\frac{\hat{\sigma}^2}{NT}\hat{\Sigma}_w^{-1}$. Report $\hat{\theta}$ with $\mathrm{SE}(\hat{\theta})=[\widehat{\mathrm{Var}}(\hat{\theta})]^{1/2}$ and asymptotic $(1-\alpha)100\%$ confidence interval $[\hat{\theta}\pm z_{1-\alpha/2}\,\mathrm{SE}(\hat{\theta})]$.

5.2 Asymptotic results under misspecification of K

Here, we discuss the theoretical results when we choose $K_o \neq K$ number of latent variables to embed for ENAR. The theorems in the case of over-selection, i.e., when $K_o > K$, are given in the Appendix section B.3. Those theorems on over-selection (Theorems 14 and 15) show that for the fixed K case, the same theoretical results on consistency and asymptotic normality hold when K is over-selected as well.

5.2.1 When $K_o < K$

Now, suppose we have selected the number of latent variables $K_o \geq 0$ smaller than the true latent dimension K. Note that the ENAR with $K_o = 0$ reduces to the NAR model in the time series context and to the model omitting latent variables in the peer effect context. Using conventional array indexing notations, the true latent position information can now be partitioned as $\mathbf{U} = [\mathbf{U}_{:K_o} | \mathbf{U}_{K_o:K}]$ where $\mathbf{U}_{:K_o} \in \mathbb{R}^{N \times K_o}$ contains first K_o ($< K$) columns of $\mathbf{U}$ and $\mathbf{U}_{K_o:K} \in \mathbb{R}^{N \times (K-K_o)}$ contains the next $K - K_o$ columns of $\mathbf{U}$ which will be missed in the model fit. If $K_o = 0$, then we will have $\mathbf{U}_{K_o:K} = \mathbf{U}$.

To study the under-selection case, we need new results that quantify the deviation of the estimated $N \times K_o$ matrix of eigenvectors and spectral embedding vectors from the true $N \times K$ matrix of eigenvectors and spectral embedding vectors, respectively. The next theorem proves two concentration inequalities, providing results that may be of independent interest to researchers in statistical inference in networks.

Theorem 7 *Assume the network* $\mathbf{A}$ *is generated from a* K*-dimensional RDPG model,* $\mathbf{A} \sim \text{RDPG}(\rho_N, \mathbf{X})$*. Let* $k \in \{1, \ldots, K-1\}$ *for* $K > 1$*. Write a truncated eigenvector of* $\mathbf{A}$ *as* $\hat{\mathbf{V}}_{:k}$ *and truncated spectral embedding as* $\hat{\mathbf{V}}_{:k}|\hat{\mathbf{S}}_{:k}|^{1/2}$ *which contain* k *leading eigenvectors (and eigenvalues). Then, provided that* $N\rho_N = \omega\left(\log^4 N\right)$*, there exists a sequence of* $\mathbf{H}_{k,v}, \mathbf{H}_{k,e} \in \mathcal{O}_{K,k}$ *such that*

$$\left\|\hat{\mathbf{V}}_{:k} - \mathbf{V}\mathbf{H}_{k,v}\right\| = O(1) \ \ whp.$$

$$\left\|\hat{\mathbf{V}}_{:k}|\hat{\mathbf{S}}_{:k}|^{1/2} - \mathbf{V}\mathbf{S}^{1/2}\mathbf{H}_{k,e}\right\| = O\left(\sqrt{N\rho_N}\right) \ \ whp.$$

The proof is given in C.3. Now denote the imputed design matrix with $\hat{\mathbf{U}}_{:K_o}$ by $\hat{\mathbf{W}}_- = \left[\hat{\mathbf{U}}_{:K_o} \,|\, \mathbf{y} \,|\, \mathcal{L}_{\mathbf{A}} \,|\, \mathbf{Z}\right] \in \mathbb{R}^{NT \times (K_o+p+2)}$, which contains an under-selected number of latent variables, either $\hat{\mathbf{U}}_{:K_o,v} = \hat{\mathbf{V}}_{:K_o}$ or $\hat{\mathbf{U}}_{:K_o,e} = \hat{\mathbf{V}}_{:K_o}|\hat{\mathbf{S}}_{:K_o}|^{1/2}$ depending on the embedding method. A least squares estimator $\hat{\mu}_-$ can be obtained as $\left(\hat{\mathbf{W}}_-^\intercal \hat{\mathbf{W}}_-\right)^{-1} \hat{\mathbf{W}}_-^\intercal \mathbf{y}$.

Theorem 8 *Assume that* $N\rho_N = \omega\left(T + \log^4 N\right)$*. Then, for* $K_o \in \{0, 1, \ldots, K-1\}$ *we have*

$$\sqrt{T}\left(\hat{\beta}_{-,v} - \mathbf{H}_{:K_o,v}{}^\intercal \beta_v\right) = \mathbf{b}_{\beta,v} + \mathbf{t}_1, \quad \sqrt{NT}\begin{bmatrix}\hat{\alpha}_{-,v} - \alpha_v \\ \hat{\theta}_{-,v} - \theta_v\end{bmatrix} = \mathbf{b}_{\alpha\theta,v} + \mathbf{t}_2,$$

$$\sqrt{NT}\left(\hat{\gamma}_{-,v} - \gamma_v\right) = \mathbf{t}_3$$

where $\mathbf{h}_1{}^\intercal \mathbf{b}_{\beta,v} = O\left(\sqrt{T}\right)$ *whp.,* $\mathbf{h}_2{}^\intercal \mathbf{b}_{\alpha\theta,v} = O\left(\sqrt{T/N}\right)$ *whp.,* $\mathbf{h}_1{}^\intercal \mathbf{t}_1 = O_{\mathbb{P}}(1), \mathbf{h}_2{}^\intercal \mathbf{t}_2 = O_{\mathbb{P}}(1), \mathbf{h}_3{}^\intercal \mathbf{t}_3 = O_{\mathbb{P}}(1)$*, for any* $\mathbf{h}_1 \in B^{K_o-1}, \mathbf{h}_2 \in B^1, \mathbf{h}_3 \in B^{p-1}$ *as* $N, T \to \infty$*.*

In the above Theorem 8, the limit of suitably normalized estimators is written in terms of two terms. The terms denoted as $\mathbf{t}_1, \mathbf{t}_2, \mathbf{t}_3$ are all $O_{\mathbb{P}}(1)$ and converge to normal distributions as in the correctly selected K case in Theorem 10. The terms denoted by $\mathbf{b}_{\beta,v}, \mathbf{b}_{\alpha\theta,v}$ are asymptotic bias terms. We can compare these limits with the results in Theorem 10, where the orders for the estimation errors of μ_v^H corresponding to $\mathbf{b}_{\beta,v}, \mathbf{b}_{\alpha\theta,v}$ were $O\left(\sqrt{\frac{T}{N\rho_N}}\right)$ *whp.*, and $O\left(\sqrt{\frac{T}{N^2\rho_N}}\right)$ *whp.* respectively. In contrast, in Theorem 8, the bias term corresponding to α, θ parameters $\mathbf{b}_{\alpha\theta,v}$ is $O(\sqrt{T/N})$, which is higher than $O\left(\sqrt{\frac{T}{N^2\rho_N}}\right)$. Therefore, even though $\mathbf{t}_2$ will converge to a normal limiting distribution and we will have $\sqrt{NT}$ consistency for α, θ parameters, the biases in those parameter estimates are of a higher order than in the correct or over-selected K case.

The next theorem is for the case when the latent variables correspond to spectral embedding.

Theorem 9 *Assume that* $N\rho_N = \omega\left(\log^4 N\right)$ *as* $N \to \infty$. *Then, as* $N \to \infty$, *for* $K_o \in \{0, 1, \ldots, K-1\}$ *we have*

$$\sqrt{NT\rho_N}\left(\hat{\beta}_{-,e} - \mathbf{H}_{:K_o,e}{}^{\intercal}\beta_e\right) = \mathbf{b}_{\beta,e} + \mathbf{t}_1, \quad \sqrt{NT}\begin{bmatrix}\hat{\alpha}_{-,e} - \alpha_e \\ \hat{\theta}_{-,e} - \theta_e\end{bmatrix} = \mathbf{b}_{\alpha\theta,e} + \mathbf{t}_2,$$
$$\sqrt{NT}\left(\hat{\gamma}_{-,e} - \gamma_e\right) = \mathbf{b}_{\gamma,e} + \mathbf{t}_3$$

where $\mathbf{h}_1{}^{\intercal}\mathbf{b}_{\beta,e} = O\left(\sqrt{N\rho_N}\right)$ *whp.*, $\mathbf{h}_2{}^{\intercal}\mathbf{b}_{\alpha\theta,e} = O\left(\sqrt{N}\rho_N\right)$ *whp.*, $\mathbf{h}_3{}^{\intercal}\mathbf{b}_{\gamma,e} = O_{\mathbb{P}}\left(\sqrt{\rho_N}\right)$, $\mathbf{h}_1{}^{\intercal}\mathbf{t}_1 = O_{\mathbb{P}}(1)$, $\mathbf{h}_2{}^{\intercal}\mathbf{t}_2 = O_{\mathbb{P}}(1)$, $\mathbf{h}_3{}^{\intercal}\mathbf{t}_3 = O_{\mathbb{P}}(1)$, *for any* $\mathbf{h}_1 \in B^{K_o-1}, \mathbf{h}_2 \in B^1, \mathbf{h}_3 \in B^{p-1}$.

This result can also be compared with the result in Theorem 4, where the orders of estimation error corresponding to the "bias terms" $\mathbf{b}_{\beta,e}, \mathbf{b}_{\alpha\theta,e}$ were $O(1)$ *whp.*, and $O\left(\sqrt{\rho_N}\right)$ *whp.* respectively. In comparison, the bias terms $\mathbf{b}_{\beta,e}, \mathbf{b}_{\alpha\theta,e}$ in the above theorem are $O\left(\sqrt{N\rho_N}\right)$, and $O\left(\sqrt{N}\rho_N\right)$ *whp.* respectively. Theorem 9 shows that when the true data-generating peer effect model contains spectral embedding vectors, ignoring those vectors in the estimation process leads to possibly inconsistent estimation. The estimation error for θ_e is $O_{\mathbb{P}}(\rho_N + N^{-1/2})$, which increases as the network gets denser, and in the limit of dense networks, the estimator of the peer effect parameter becomes inconsistent. In contrast, Theorem 8 showed that for the model with eigenvectors when $K_o = 0$, i.e., all the latent variables are ignored during estimation, the estimation error for θ_v is $O_{\mathbb{P}}\left(1/\sqrt{NT}\right)$. It appears that with eigenvectors, there is no detrimental effect on the estimation of peer effect due to omitting latent variables (higher asymptotic bias, but same rate of consistency). This is contrary to the assumptions from the Directed Acyclic Graph, which says that there is non-trivial confounding due to the latent variables. Since modeling choice should be made based on the DAG (Pearl, 2009), the model with spectral embedding vectors is appropriate for the peer effect problem, while the model with eigenvectors is not. Both theorems also quantify the estimation error when the number of latent vectors selected, K_o, is less than the true number of latent vectors. The asymptotic orders for the parameter estimation errors when $0 < K_o < K$ are the same as when $K_o = 0$, as both theorems show.

6. Simulations

In this section, we use Monte Carlo simulations to illustrate the finite sample performance of the ENAR model estimator with $\mathbf{U} = \mathbf{U}_v$ and compare it with NAR and CNAR. We examine the sensitivity of the considered models under model misspecification in terms of estimations of model parameters and one-step-ahead prediction of $\mathbf{y}_{T+1}$. In this regard, we consider the scenarios where $\{\mathbf{y}_0, ..., \mathbf{y}_T\}$ and $\mathbf{y}_{T+1}$ follow each of ENAR, CNAR, and NAR. In the case where we generate $\mathbf{y}_t$ with CNAR (Chen et al., 2023), we assumed that there is no latent variable structure in the model noise.

We consider the DCSBM and DCMMSBM for generating the network $\mathbf{A}$, using `fastRG` package in R Rohe et al. (2018). First, we used the matrix $2q\mathbf{I}_K + q\mathbf{1}_K\mathbf{1}_K^\intercal$ where $q = \frac{9}{40}$ to generate the $K \times K$ block matrix of connection probabilities. As a result, the ratio of inter-community and between-community connectivity is 3. The maximum expected degree for each graph was set to be $N\rho_N$ where $\rho_N \triangleq N^{-1/2}$, ensuring that the graphs are sparse. Further details of the simulation setup are given in Appendix section E.1.1.

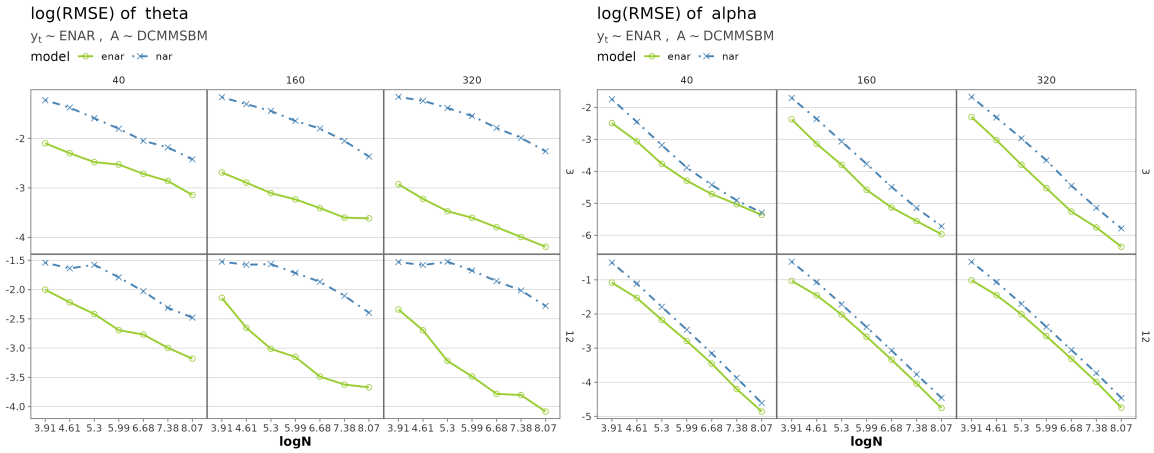


Figure 2: Plot of estimates of θ (left) and α (right) from ENAR and NAR models with increasing N when data are generated from the ENAR model with DCMMSBM. The rows correspond to $K = 3$ and $K = 12$, and the columns correspond to $T = 40, 160, 320$.

To compare estimation performance, we computed relative root mean squared errors (RMSE) as $\frac{\|\mathbf{B}-\hat{\mathbf{B}}\|}{\|\mathbf{B}\|}$ for an arbitrary matrix $\mathbf{B}$ and its estimate $\hat{\mathbf{B}}$. Also, we report one-step-ahead prediction errors as $\frac{\|\mathbf{W}_T(\hat{\mu}-\mu)\|}{\|\mathbf{W}_T\mu\|}$, i.e., root mean squared prediction errors (RMSP), as the systematic noise incurred by $\mathcal{E}_{T+1}$ is difficult to predict. To track the model performance as its dimension grows, we take $N = 50, 100, 200, 400, 800, 1600, 3200$ and $K = 3, 12$. We also consider finite T case where $T = 2$ and growing T case where $T = 40, 160, 320$. For each configuration, we conduct 300 replications to compute RMSEs and RMSPs, and report their averages as the plotted points. To organize the display of the results better, in most cases, we plot log RMSE or RMSP against $\log N$ as N increases from 50 to 3200 for the three values of T separately.

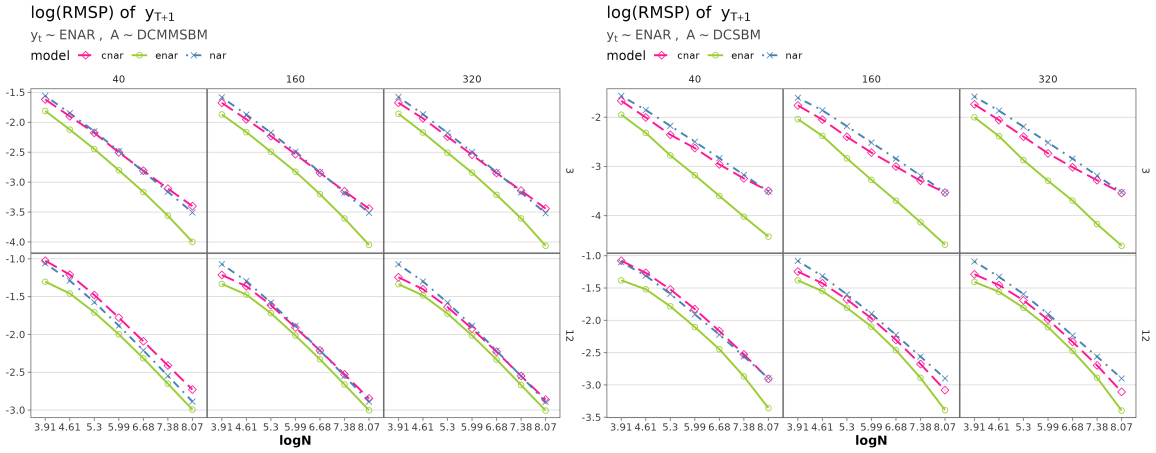


Figure 3: Plot of the prediction error from ENAR, CNAR, and NAR models when data is generated from the ENAR model with DCMMSBM and DCSBM, respectively.

6.1 Generating data from ENAR model

The RMSE plots of peer influence and momentum effects are shown in Figure 2 for data generated from the ENAR model with DCMMSBM. Each column and row of the facet grid corresponds to a different value of T and K, respectively, while the figures within the grids are with increasing N. From Figure 2, we observe that when the true model is ENAR, the RMSE of its estimates of α and θ consistently decreases as N and T grow, which is the expected phenomenon from our asymptotic theories.

For estimates from the NAR model, while the RMSE for the α parameter still decreases with increasing N, the RMSE for estimation of θ continues to remain high even when N and T increase. This is because the NAR fit omits the latent variable effects, and it incurs irreducible bias in parameter estimation.

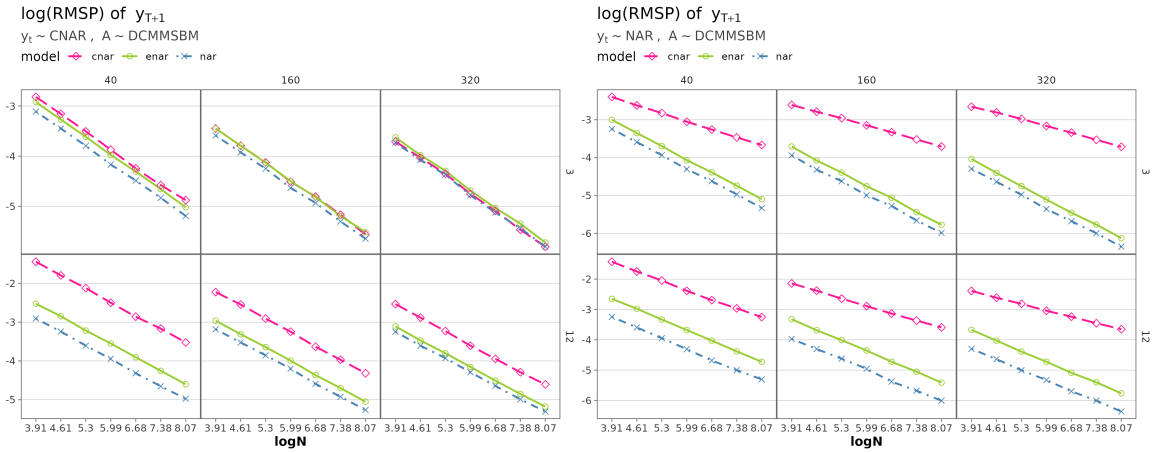


Figure 4: Model misspecification: Plot of prediction error from CNAR, ENAR, and NAR models when data is generated from CNAR and NAR models, respectively.

The predictions of ENAR, as shown in Figure 3 for graphs generated from DCMMSBM and DCSBM, also improve with both N and T. This figure shows that if ENAR is the true data-generating model, then omitting the latent variables from the fitted model (i.e., fitting the NAR model) not only leads to inaccurate parameter estimates but also to higher prediction error. The CNAR model also does not perform well for prediction when ENAR is the true data-generating model.

6.2 Comparison Under Model Misspecification

Next, we generate $\mathbf{y}_t$ from the NAR and CNAR models while generating the underlying networks from the DCMMSBM model. This corresponds to model misspecification for ENAR, and we can compare prediction and parameter-estimation accuracy in this setting. In Figure 4 (left), we see a very good performance of both ENAR and NAR even when the data is generated from the CNAR model. This is especially true for smaller values of N and T (e.g., $N, T \leq 50$). When K is increased to 12, the performance of CNAR is worse than NAR and ENAR, even when the data is generated from CNAR. This is because with increasing K, the performance of the CNAR model estimators becomes worse.

When we assumed the true model NAR, as expected, the predictive ability of NAR was the best overall for all settings of N, T, K (Figure 4 (right)). However, in each case, the ENAR model came close in terms of predictive ability while the CNAR model produced large errors, especially when K was larger and N and T were smaller.

Finally, beyond prediction error, we evaluate the estimation error of the θ and α parameters under model misspecification when the data is generated from the NAR model. We observe that the estimation of the peer influence effect becomes biased for ENAR, especially when T and N are larger (Figure C3 in Appendix). However, their estimation bias is significantly smaller than that of using NAR when data is generated from ENAR, as shown in Figure 2. Moreover, the ENAR model was able to consistently estimate α at comparable rates to NAR, showing robustness under model misspecification. In contrast, the NAR model produced large errors even for estimating α when ENAR was the true data-generating model in Figure 2.

6.3 Finite T case: ENAR model

Next, we investigate the performance of ENAR and NAR peer and momentum effect estimators in the fixed T case. We set $T = 2$, and $K = 3$ and increase N. In Figure 5, it is clear that ENAR can estimate α and θ well even under model misspecification of generating data from NAR. The estimation error for estimating both α and θ from ENAR is comparable to NAR and decreases with increasing N. In contrast, when the data is generated from ENAR, we see that the estimate of θ from NAR is biased and continues to show high error even when N increases, while the estimation error decreases for ENAR.

Combined with our previous observations in growing N and T cases, ENAR shows robust estimation performance under various model misspecifications, while achieving better prediction and parameter estimation performance when the data is generated from ENAR in both growing and fixed K cases.

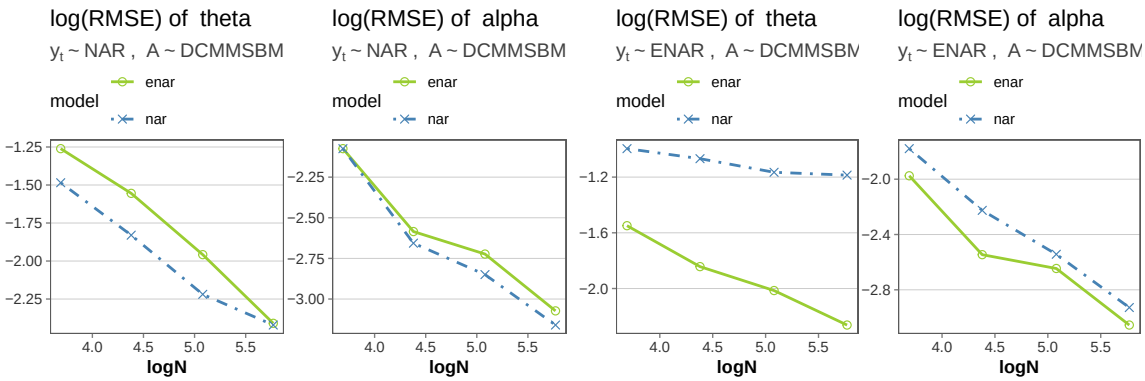


Figure 5: Finite T simulations: Plot of estimates of θ and α from ENAR, and NAR model when data is generated from ENAR and NAR model with DCMMSBM.

6.4 Coverage Rates for θ

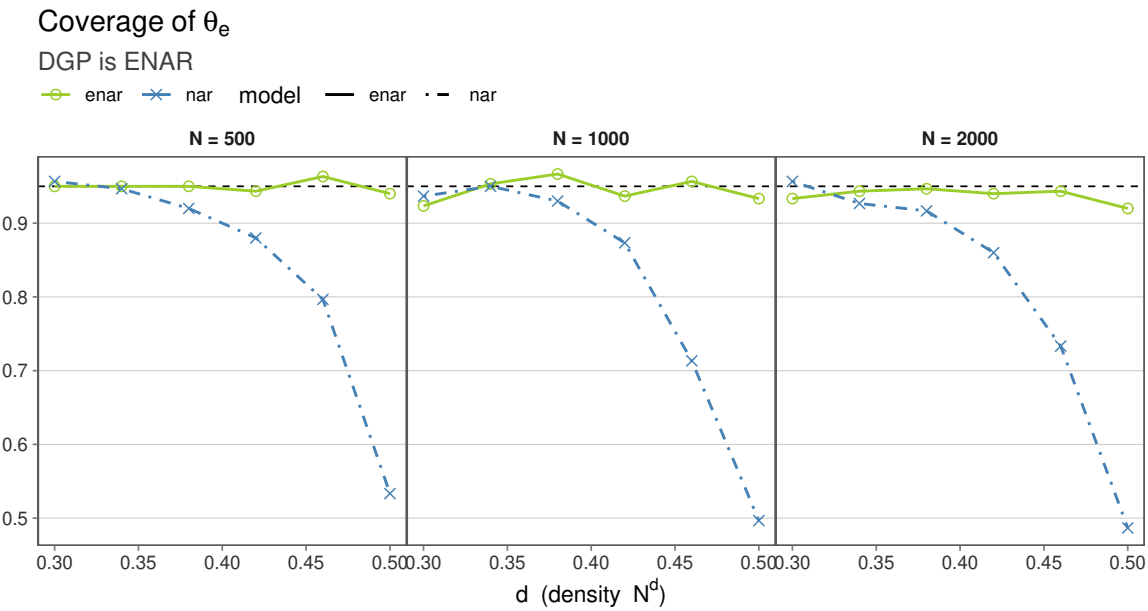


Figure 6: Coverage simulation results

Here, we present simulation results for the interval estimation of the peer-effect parameter θ_e under the ENAR data-generating process with latent embedding $\mathbf{U} = \mathbf{U}_e$. We fix $T = 2$ and $K = 4$, and consider three values of $N \in \{500, 1000, 2000\}$. We vary a parameter $d \in \{0.30, 0.34, 0.38, 0.42, 0.46, 0.50\}$, which controls the network density through the relationship to expected degree per node, $N\rho_N = N^d$. Equivalently, for each d, we set $\rho_N = N^{d-1}$ so that smaller d corresponds to sparser networks. The adjacency matrices $\mathbf{A}$ were generated from the DCMMSBM as before. The 95% confidence intervals were constructed using the standard error estimate $\widehat{\mathrm{Var}}(\hat{\theta})^{1/2}$ obtained from Algorithm 1, assuming a correctly specified embedding dimension K. Figure 6 compares the empirical coverage

rates of the ENAR and NAR estimators for θ_e, averaged over 300 Monte Carlo replications for each configuration. This parameterization allows us to examine how the reliability of peer-effect inference evolves as the network transitions from relatively sparse (expected density $N^{-0.7}$) to moderately dense (expected density $N^{-0.5}$) regimes.

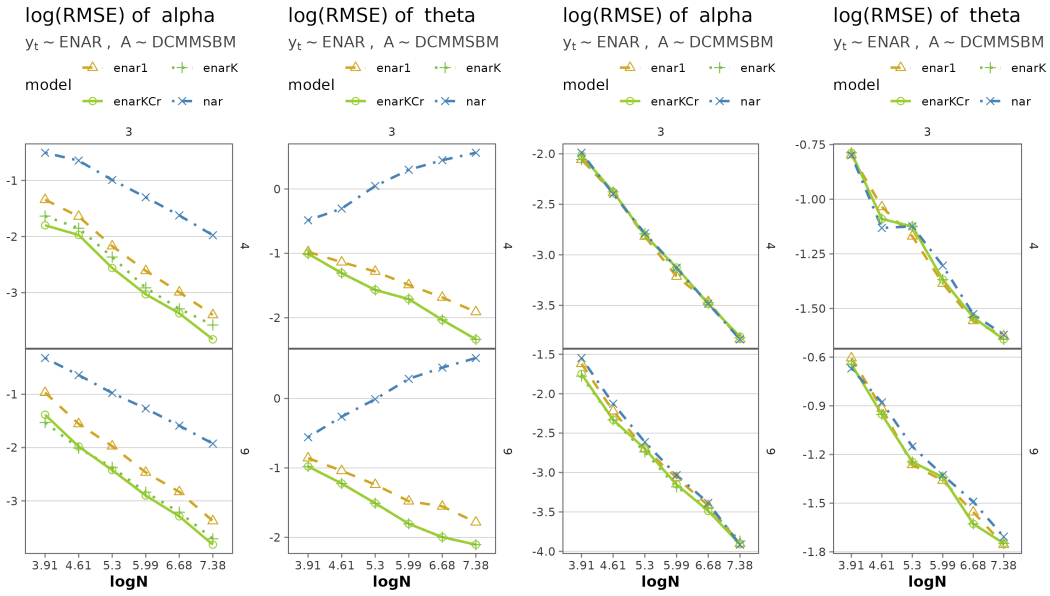


Figure 7: Plot of estimates of θ and α from ENAR with (left) $\mathbf{U} = \mathbf{U}_e$ and (right) $\mathbf{U} = \mathbf{U}_v$ for $K_o \in \left\{1, K, \hat{K}\right\}$, and NAR model when data is generated from ENAR model with DCMMSBM.

When the data are generated from the ENAR model, the ENAR estimator (green solid line) maintains coverage rates that remain close to the nominal 0.95 level for $d \geq 0.3$ and different network sizes. In contrast, the NAR estimator (blue dashed line) shows a clear deterioration in coverage as the network becomes denser (larger d), dropping below 0.6 when the density is $N^{-0.5}$. This pattern illustrates the critical role of the homophily adjustment: as network density increases, omitted-variable bias in NAR invalidates its peer influence inference. The ENAR estimator mitigates this bias by explicitly modeling the latent structure, yielding robust inference even under dense regimes. Results for extreme low density of below $N^{-0.7}$ are omitted, as both NAR and ENAR recover the nominal 95% coverage level in those ultra-sparse settings. Notably, the stability of the ENAR coverage across N indicates that its asymptotic variance approximation remains accurate in finite N, while the NAR estimator's under-coverage grows systematically with both network density and size, reflecting an accumulation of unaccounted latent confounding.

6.5 Misspecification of K in ENAR

Next, we perform a simulation to compare the performance of the estimators when K is selected according to our criterion $Cr(k)$, searched over $k \in \{1, \ldots, J_N = \|\mathbf{A}\|_\infty^{1/2}\}$ using $p_{NT}(k) = \frac{k}{N+T}$, with the estimators when the true K is used and when K is underselected to 1. In Figure 7, we plot the RMSEs for estimation of α and θ for increasing N and

fixed $T = 3$ for the ENAR model with $\mathbf{U} = \mathbf{U}_e$ (left two panels), and $\mathbf{U} = \mathbf{U}_v$ (right two panels) respectively. In Figure C1 in the Appendix section E.1.2, we plot the RMSEs for growing N, T for the model with $\mathbf{U} = \mathbf{U}_v$. In all cases, the estimators with selected $K = \hat{K}$ perform similarly or sometimes even better than the estimators with true K. In contrast, the estimators with $K = 1$ and $K = 0$ (which correspond to the NAR model) underperform. For both modeling setups, our method with a fully data-driven choice of the number of latent vectors provides effective control of confounding due to latent homophily.

Further in section E.1.3 in the Appendix, we plot the empirical frequency of non-under-selection using our IC-based selection criterion. We find in our simulation that the empirical frequency is very close to 1, validating our model selection criterion in finite samples.

7. Real Data Example

In this section, we will analyze two datasets. The first one is a finite-time dataset where the primary goal is to infer causal peer effects and effects of covariates, and the second one is a time series dataset where the goal is both accurate prediction and parameter estimation. For all datasets, we select K using our IC based selection method with $p_{l,NT}(k) = \frac{k^2}{N+T}$, and $J_N = C\sqrt{N}$.

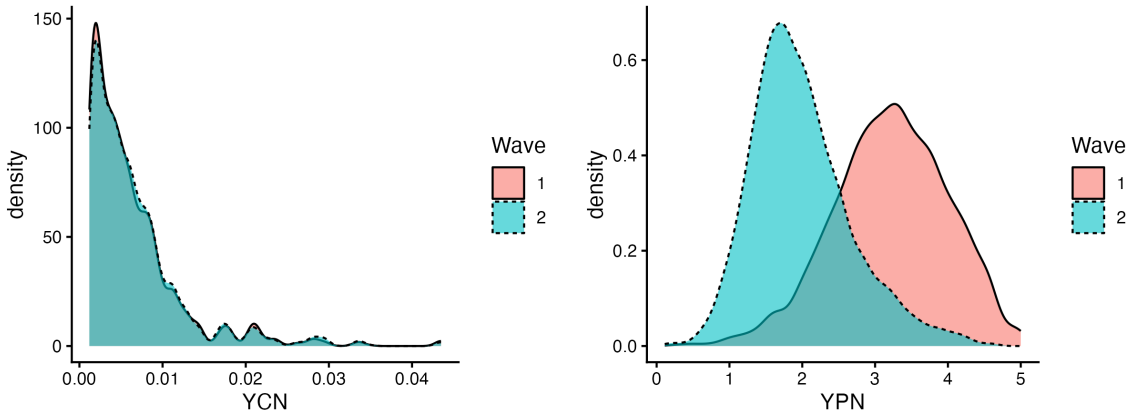


Figure 8: Distributions of two outcome variables across two survey waves.

Peer effects in conflict and school climate perception: We apply the methodology developed in this paper to estimate peer effects on conflict and school climate perception using data from Paluck et al. (2016). This large-scale field experiment was conducted in 56 New Jersey public middle schools to evaluate the effectiveness of a social intervention designed to reduce conflict. The students were given two surveys, one at the beginning of the school year (wave 1) and the other at the end of the school year (wave 2). For this peer effect analysis, we focus only on the schools in the control group that did not receive any treatment. The students were asked several sets of questions in the surveys. We focus on two primary sets of questions to construct our outcome variables.

Conflict Proportion (YCN). We compute the proportion of students that a student has reported having conflicts with in their school. The survey had five questions (CN1 to CN5)

that asked students to list the IDs of other students in their school with whom they had conflicts. We count the number of students they mentioned and divide by the number of students the school had to arrive at this YCN outcome variable. We take the proportions as our outcome variable instead of the actual counts since the number of students in the schools varied substantially from 95 in the smallest school to 685 in the largest one in our sample.

Prescriptive Norms (YPN). We construct our second variable as the average of the 12 Prescriptive norms (PN1 to PN12) questions in the survey (we call this YPN). The prescriptive norm series of questions asked students what percentage of students in their school, according to them, were engaging in friendly and conflict-related behavior using a pictorial scale. For PN3 to PN6 variables we recode them by subtracting the values from 5 since they are about conflict or bad behavior, while the rest of the variables are anti-conflict or good behavior variables.

We restrict attention to respondents with complete information on key covariates, resulting in $N = 7872$ observations for the outcome YCN and $N = 7344$ for YPN. In addition to the surveys at the two time points, the students nominated (up to 10) peers with whom they "spend time," forming directed friendship networks. Following previous work (Le and Li, 2022), we convert the network into an undirected network by considering a friendship link if any of the two individuals in the pair have nominated the other person.

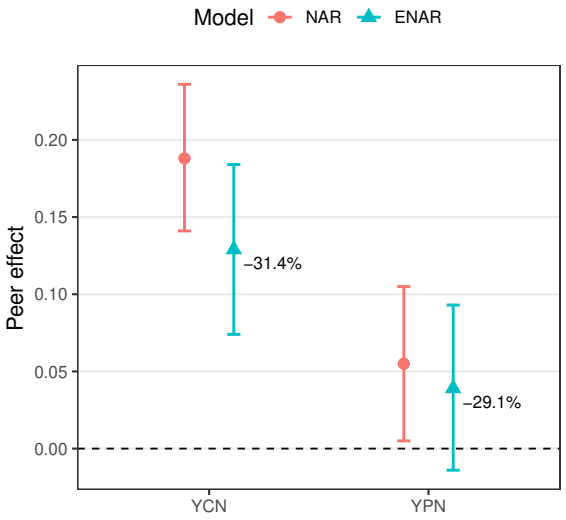


Figure 9: Peer effect estimates with 95% CIs (YCN vs. YPN, NAR vs. ENAR).

Table 6 reports the NAR and ENAR parameter estimates with corresponding 95% confidence intervals for YCN and YPN outcomes. We estimate the embedding dimension $\hat{K}$ using our IC developed in section 5.1. We estimated rather large values of K at 45 and 43 respectively. This is not a surprise, since there are 23 schools and the students come from 4 different grades, leading to strong clustering in the friendship network. We included the intervals for the significant predictors ($p < 0.05$) only. In both specifications, the momentum effects are strongly significant, reflecting the persistence of behavioral norms over time. The peer-effect coefficients, highlighted in bold, change between NAR and ENAR. For YCN, it remains positive and statistically significant under both models, but the magnitude decreases from $\hat{\theta}_e^{\text{NAR}} = 0.188$ to $\hat{\theta}_e^{\text{ENAR}} = 0.129$. For YPN, the peer effect for NAR is $\hat{\theta}_e^{\text{NAR}} = 0.055$ and for ENAR is $\hat{\theta}_e^{\text{ENAR}} = 0.039$. The peer effect, while statistically significant under NAR model, becomes statistically insignificant at the 5% level under ENAR. The ENAR peer-effect estimates are reduced by 31.4% for YCN and 29.1% for YPN relative to those from the NAR model (Figure 9). Together, these results suggest that a substantial portion of the network influence

detected by NAR may have been driven by unobserved homophily rather than true peer effects.

Table 6: Regression table for Conflict and School Climate Perception Data

Dependent variable		YCN (N=7,872)		YPN (N=7,344)
Model	NAR	ENAR ($\hat{K}$=45)	NAR	ENAR ($\hat{K}$=43)
Peer influence	**0.188*****	**0.129*****	**0.055***	**0.039**
	(0.141, 0.236)	**(0.074, 0.184)**	**(0.005, 0.105)**	**(-0.014, 0.093)**
Momentum	0.499***	0.493***	0.246***	0.244***
	(0.479, 0.519)	(0.473, 0.513)	(0.23, 0.27)	(0.223, 0.265)
Age	0.0001	0.0001	0.009	0.01
6th grade	-0.001**	-0.001**	-0.068	-0.072
	(-0.002, -0.001)	(-0.002, -0.001)		
7th grade	-0.002***	-0.002***	-0.128*	-0.129*
	(-0.002, -0.001)	(-0.002, -0.001)	(-0.23, -0.02)	(-0.235, -0.022)
8th grade	-0.002***	-0.002***	-0.144*	-0.15*
	(-0.003, -0.001)	(-0.003, -0.001)	(-0.27, -0.01)	(-0.282, -0.018)
Gender (Boy)	-0.0001	-0.0001	0.129***	0.126***
			(0.096, 0.162)	(0.093, 0.160)
Black ethnicity	0.00001	-0.0001	-0.019	-0.017
Hispanic	0.0001	0.0001	-0.004	-0.013
Mother went to college	-0.00001	0.00001	-0.022	-0.021
Lives with just mom	0.0002	0.0002	0.010	0.005
∟ with just dad	-0.00002	0.00003	0.011	0.008
∟ with parents but separated	0.0002	0.0002	0.042	0.04
Participate in music	0.0005***	0.0005***	-0.024	-0.026
	(0.0002, 0.001)	(0.0002, 0.001)		
∟ in sports at school	0.0001	0.0001	-0.001	0.0001
∟ in sports outside of school	-0.00001	-0.00001	-0.035	-0.035
∟ in theater drama	0.001***	0.001***	-0.006	-0.006
	(0.0003, 0.001)	(0.0002, 0.001)		
∟ in other arts	0.0002	0.0002	-0.040	-0.041
∟ in other school clubs	-0.0001	-0.00004	-0.031	-0.036
Date people at this school	0.0004**	0.0004**	-0.070***	-0.065**
	(0.0002, 0.001)	(0.0002, 0.001)	(-0.11, -0.03)	(-0.105, -0.026)
Do lots of homework	0.0001	0.0001	-0.047**	-0.048**
			(-0.08, -0.01)	(-0.08, -0.015)
Do read books for fun	-0.0001	-0.0001	-0.120***	-0.117***
			(-0.16, -0.09)	(-0.153, -0.082)

Note: $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$. 95% Confidence Interval provided in parenthesis.

In order to interpret the peer effect parameter result, let us recall the causal diagram from Figure 1. For the result on YCN, an increase of 1% in the average peer response (i.e., the proportion of students the peers mention to have conflict with on average) leads to an increase of 0.129 % increase in an individual's own response (proportion of students the individual mentions having conflict with) in the next time period. Similarly, for the result on YPN, a 1 unit increase in the average response across 12 PN categories by the peers (averaged across peers) leads to an increase of 0.039 units (however, it is statistically insignificant) in an individual's average response across PN questions in the next time period.

Wind speed time series data: Next, we apply the ENAR model to a multivariate time series consisting of wind speed measurements recorded over 721 time periods at 102

weather stations across England and Wales. The dataset is available from the `GNAR` package in Knight et al. (2020). This setting represents a high-dimensional regime with both large T and large N.

For comparison, we include the NAR and VAR(1) models, where VAR(1) denotes the vector autoregressive model of order one specified as $\mathbf{y}_t = \Gamma \mathbf{y}_{t-1} + \epsilon_t$, with $\Gamma \in \mathbb{R}^{N \times N}$ being a coefficient matrix. Structurally, the VAR(1) model allows arbitrary pairwise dependencies among all nodes, whereas NAR and ENAR impose network-based parsimony by encoding peer dependence through a scalar parameter θ.

We assess 1-step-ahead prediction accuracy 200 times for the considered models. In the i-th prediction task, we fit the NAR, ENAR, and VAR(1) models to the time series up to time $519 + i$ and predict the response at time point $520 + i$. We continue increasing i and slide the training-data window until $i = 200$. For the ENAR model, we estimate the dimension of the latent variables as $K = 10$ using our IC model-selection criterion.

As shown in Table 7, NAR produces the largest mean of MSPE across 200 prediction windows, indicating the lowest predictive accuracy. Although VAR(1) achieves the smallest average MSPE, it has the largest standard deviation (SD) and widest interquartile range (IQR), implying greater variability in its prediction performance. By contrast, ENAR attains higher accuracy than NAR with the smallest SD and IQR, demonstrating improved stability in predictions over time.

Table 7: MSPE distributions over 200 test time points

Model	mean	SD	IQR
ENAR	0.37740	0.09104	0.11391
NAR	0.37791	0.09146	0.11665
VAR(1)	0.36884	0.09489	0.15266

Next, to compare model fit, we compute Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) for each model as follows:

$$\mathrm{AIC}(k) := 2k + NT \log \frac{\mathrm{RSS}}{NT}$$
$$\mathrm{BIC}(k) := k \log(NT) + NT \log \frac{\mathrm{RSS}}{NT}$$

where RSS denotes the residual sum of squares. In the left panel of Figure 10, we report the AIC and BIC values of NAR, ENAR, and VAR(1) models as the training time window gradually expands. As the training window expands, all models improve monotonically under both AIC and BIC. Yet their rankings diverge: AIC favors VAR(1) for its flexibility, whereas BIC, which imposes a stronger penalty on model size, prefers the more parsimonious NAR and ENAR. The discrepancy reflects VAR(1)'s $O(N^2)$ parameter space, which inflates its BIC penalty.

Due to the vertical scale of the left panel, the NAR and ENAR curves appear nearly indistinguishable. The right panel therefore plots the information-criterion differences, IC(ENAR) − IC(NAR), across 200 evaluation windows for both AIC (solid red) and BIC (solid blue). Negative values indicate that ENAR achieves a lower information criterion than NAR and thus provides a more favorable model fit. Both the AIC and BIC differentials remain persistently negative and diminish slowly over time, indicating that ENAR consistently outperforms NAR in balancing goodness-of-fit with model complexity as more

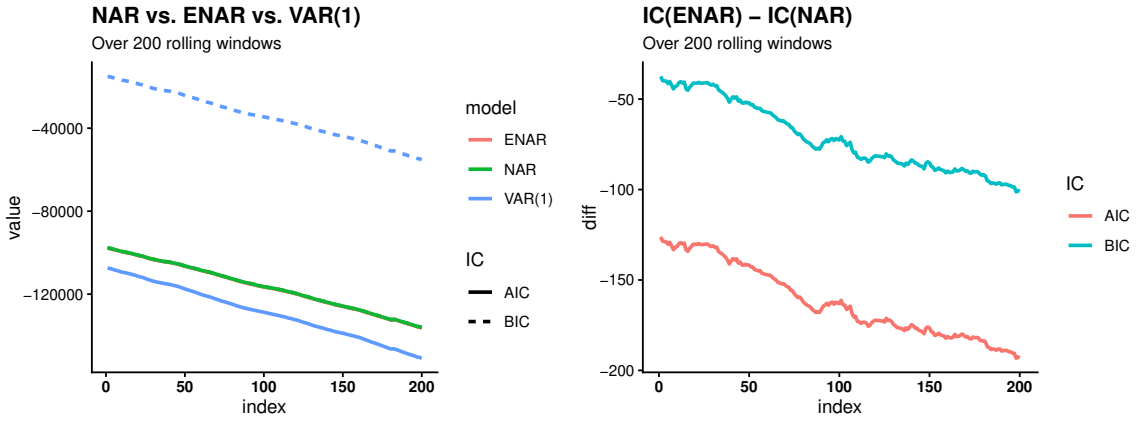


Figure 10: (left) AIC and BIC of all three models with increasing time, (right) Difference in information criteria between ENAR and NAR for the Wind speed data.

data become available. Overall, ENAR attains predictive accuracy comparable to VAR(1) while offering a more stable, parsimonious, and interpretable network-based representation.

8. Conclusion

The ENAR model can successfully address two major statistical problems. It embeds latent homophily effects in the network autoregressive model, tackling both consistent estimation of causal peer effects and enhanced predictive performance for time series. We proved that variants of the ENAR estimators have asymptotic normality in both long-term and finite-time cases. Our theoretical results provide conclusive answers to several previously open questions on peer effects. These questions include how to model peer effects with latent variables in a growing N and K setup, what the trade-offs are between measurement error and omitted-variable bias, and how to select the dimension of latent variables. Further, for the growing T case of the network time-series autoregressive model, our results answer whether it is possible to consistently estimate model parameters when latent variables are included.

Our numerical studies further clarify the practical regimes where each method is most advantageous. For short panels with finite T and networks of moderate size, the ENAR model yields clear gains over NAR and CNAR by effectively adjusting for latent homophily, resulting in lower bias and more accurate inference for the peer-effect parameter θ. These benefits are most evident under sparse or moderately dense networks where $N\rho_N$ grows slowly, consistent with the asymptotic regimes in our theoretical results. Hence, ENAR is most beneficial for causal peer-effect inference in sparse and moderately dense networks.

Taken together, this work advances the methodological and theoretical foundations for both causal peer effect estimation and predictive modeling in networked longitudinal data.

References

Weihua An, Roberson Beauvile, and Benjamin Rosche. Causal network analysis. *Annual Review of Sociology*, 48:23–41, 2022.

Avanti Athreya, Donniell E Fishkind, Minh Tang, Carey E Priebe, Youngser Park, Joshua T Vogelstein, Keith Levin, Vince Lyzinski, and Yichen Qin. Statistical inference on random dot product graphs: a survey. *The Journal of Machine Learning Research*, 18(1):8393–8484, 2017.

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

Junhui Cai, Dan Yang, Wu Zhu, Haipeng Shen, and Linda Zhao. Network regression and supervised centrality estimation. *arXiv preprint arXiv:2111.12921*, 2021.

T. Tony Cai and Anru Zhang. Rate-optimal perturbation bounds for singular subspaces with applications to high-dimensional statistics. *The Annals of Statistics*, 46(1):60 – 89, 2018.

Joshua Cape, Minh Tang, and Carey E. Priebe. The kato–temple inequality and eigenvalue concentration with applications to graph inference. *Electronic Journal of Statistics*, 11 (2):3954–3978, 2017.

Joshua Cape, Minh Tang, and Carey E Priebe. Signal-plus-noise matrix models: eigenvector deviations and fluctuations. *Biometrika*, 106(1):243–250, 2019a.

Joshua Cape, Minh Tang, and Carey E. Priebe. The two-to-infinity norm and singular subspace geometry with applications to high-dimensional statistics. *The Annals of Statistics*, 47(5):2405 – 2439, 2019b.

Elynn Y Chen, Jianqing Fan, and Xuening Zhu. Community network auto-regression for high-dimensional time series. *Journal of Econometrics*, 235(2):1239–1256, 2023.

Nicholas A Christakis and James H Fowler. The spread of obesity in a large social network over 32 years. *New England journal of medicine*, 357(4):370–379, 2007.

Nicholas A Christakis and James H Fowler. Social contagion theory: examining dynamic social networks and human behavior. *Statistics in medicine*, 32(4):556–577, 2013.

Naoki Egami and Eric J Tchetgen Tchetgen. Identification and estimation of causal peer effects using double negative controls for unmeasured network confounding. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 86(2):487–511, 2024.

Bailey K Fosdick and Peter D Hoff. Testing and modeling dependencies between a network and nodal attributes. *Journal of the American Statistical Association*, 110(511):1047–1056, 2015.

Paul Goldsmith-Pinkham and Guido W Imbens. Social networks and the identification of peer effects. *Journal of Business & Economic Statistics*, 31(3):253–264, 2013.

Yawen Guan, Garritt L Page, Brian J Reich, Massimo Ventrucci, and Shu Yang. Spectral adjustment for spatial confounding. *Biometrika*, 110(3):699–719, 2023.

Peter Hall and Christopher C Heyde. *Martingale limit theory and its application*. Academic press, 2014.

Alex Hayes, Mark M Fredrickson, and Keith Levin. Estimating network-mediated causal effects via principal components network regression. *Journal of Machine Learning Research*, 26(13):1–99, 2025.

Yanjun He and Peter D Hoff. Multiplicative coevolution regression models for longitudinal networks and nodal attributes. *Social Networks*, 57:54–62, 2019.

Peter D Hoff, Adrian E Raftery, and Mark S Handcock. Latent space approaches to social network analysis. *Journal of the american Statistical association*, 97(460):1090–1098, 2002.

Marina Knight, Kathryn Leeming, Guy Nason, and Matthew Nunes. Generalized network autoregressive processes and the gnar package. *Journal of Statistical Software*, 96(5): 1–36, 2020. doi: 10.18637/jss.v096.i05.

Can M Le and Tianxi Li. Linear regression and its inference on noisy network-linked data. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(5): 1851–1885, 2022.

Jing Lei and Alessandro Rinaldo. Consistency of spectral clustering in stochastic block models. *The Annals of Statistics*, 43(1), 2015.

Robert Lunde, Elizaveta Levina, and Ji Zhu. Conformal prediction for network-assisted regression. *Journal of the American Statistical Association*, pages 1–22, 2025.

Edward McFowland III and Cosma Rohilla Shalizi. Estimating causal peer influence in homophilous social networks by inferring latent locations. *Journal of the American Statistical Association*, pages 1–12, 2023.

Shanjukta Nath, Keith Warren, and Subhadeep Paul. Identifying peer influence in therapeutic communities adjusting for latent homophily. *The Annals of Applied Statistics*, 19 (1):529–565, 2025.

A James O'Malley, Felix Elwert, J Niels Rosenquist, Alan M Zaslavsky, and Nicholas A Christakis. Estimating peer effects in longitudinal dyadic data using instrumental variables. *Biometrics*, 70(3):506–515, 2014.

Elizabeth Levy Paluck, Hana Shepherd, and Peter M Aronow. Changing climates of conflict: A social network experiment in 56 schools. *Proceedings of the National Academy of Sciences*, 113(3):566–571, 2016.

Judea Pearl. *Causality*. Cambridge university press, 2009.

Karl Rohe, Jun Tao, Xintian Han, and Norbert Binkiewicz. A note on quickly sampling a sparse matrix with low rank expectation. *Journal of Machine Learning Research*, 19(77): 1–13, 2018.

Patrick Rubin-Delanchy, Joshua Cape, Minh Tang, and Carey E Priebe. A statistical interpretation of spectral embedding: The generalised random dot product graph. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(4):1446–1473, 2022.

Mark Rudelson and Roman Vershynin. Hanson-Wright inequality and sub-gaussian concentration. *Electronic Communications in Probability*, 18(none):1 – 9, 2013.

Cosma Rohilla Shalizi and Andrew C Thomas. Homophily and contagion are generically confounded in observational social network studies. *Sociological methods & research*, 40 (2):211–239, 2011.

Dhanya Sridhar, Caterina De Bacco, and David Blei. Estimating social influence from observational data. In *Conference on Causal Learning and Reasoning*, pages 712–733. PMLR, 2022.

Minh Tang and Carey E Priebe. Limit theorems for eigenvectors of the normalized laplacian for random graphs. *The Annals of Statistics*, 46(5):2360–2415, 2018.

Minh Tang, Avanti Athreya, Daniel L Sussman, Vince Lyzinski, Youngser Park, and Carey E Priebe. A semiparametric two-sample hypothesis testing problem for random graphs. *Journal of Computational and Graphical Statistics*, 26(2):344–354, 2017.

Tyler J VanderWeele. Sensitivity analysis for contagion effects in social networks. *Sociological Methods & Research*, 40(2):240–255, 2011.

Tyler J VanderWeele, Elizabeth L Ogburn, and Eric J Tchetgen Tchetgen. Why and when" flawed" social network analyses still yield valid tests of no contagion. *Statistics, Politics and Policy*, 3(1), 2012.

Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

Jianxiang Wang, Can M Le, and Tianxi Li. Perturbation-robust predictive modeling of social effects by network subspace generalized linear models. *arXiv preprint arXiv:2410.01163*, 2024.

Weichi Wu and Chenlei Leng. A random graph-based autoregressive model for networked time series. *arXiv preprint arXiv:2309.08488*, 2023.

Fangzheng Xie. Entrywise limit theorems for eigenvectors of signal-plus-noise matrix models with weak signals. *Bernoulli*, 30(1):388–418, 2024.

Fangzheng Xie and Yanxun Xu. Efficient estimation for random dot product graphs via a one-step procedure. *Journal of the American Statistical Association*, 118:651–664, 2023.

Xuening Zhu and Rui Pan. Grouped network vector autoregression. *Statistica Sinica*, 30 (3):1437–1462, 2020.

Xuening Zhu, Rui Pan, Guodong Li, Yuewen Liu, and Hansheng Wang. Network vector autoregression. *The Annals of Statistics*, 45(3):1096 – 1123, 2017.

Xuening Zhu, Weining Wang, Hansheng Wang, and Wolfgang Karl Härdle. Network quantile autoregression. *Journal of econometrics*, 212(1):345–358, 2019.

Appendix A. Notations

Here, we summarize the notations we frequently use in the main text. Let a, $\mathbf{a}$, and $\mathbf{A}$ be generic notations for scalars, vectors, and matrices, respectively. Let $\mathbb{I}_A$ denotes an indicator with support set A. For a matrix $\mathbf{A} \in \mathbb{R}^{n\times m}$, write its element at (i,j)-th entry as a_{ij}. Let $\mathcal{O}_{N,K} := \left\{\mathbf{H} \in \mathbb{R}^{N\times K}; \mathbf{H}^\intercal\mathbf{H} = \mathbf{I}_K\right\}$ be a collection of $N \times K$ matrices with real orthonormal columns. Write an m-dimensional vector and $m \times m$ matrix with zeros as $\mathbf{o}_m$ and $\mathbf{O}_m$, and let $\mathbf{I}_m$ denotes an $m \times m$ identity matrix. Let $S^{p-1} = \{\mathbf{h} \in \mathbb{R}^p; \|\mathbf{h}\| = 1\}$ and $B^{p-1} = \{\mathbf{h} \in \mathbb{R}^p; \|\mathbf{h}\| \leq 1\}$ denote p-dimensional unit sphere and ball, respectively.

For an n-row square matrix $\mathbf{X}$, let $\mathrm{tr}\,(\mathbf{X}) \triangleq \sum_{i=1}^n x_{ii}$ denote its trace and $s_i(\mathbf{X})$ denote the i-th largest magnitude eigenvalue of $\mathbf{X}$ such that $|s_1(\mathbf{X})| \geq \cdots \geq |s_n(\mathbf{X})|$. The i-th leading eigenvector of $\mathbf{X}$ will mean the eigenvector corresponding to this ith leading eigenvalue. The following are some matrix norms: ℓ_∞-norm $\|\mathbf{X}\|_\infty \triangleq \max_i \sum_j |x_{ij}|$, ℓ_2-norm $\|\mathbf{X}\| \triangleq \sup_{\mathbf{h}\in S^{n-1}} \|\mathbf{X}\mathbf{h}\|$, max-norm $\|\mathbf{X}\|_{\max} \triangleq \max_{i,j} |x_{ij}|$, and Frobenius norm $\|\mathbf{X}\|_F \triangleq \mathrm{tr}\,(\mathbf{X}^\intercal\mathbf{X})^{1/2}$.

For $a_n, b_n \geq 0$, we write $a_n \gtrsim b_n$ if there exists $c > 0$ independent with n such that $a_n \geq cb_n$ for all n. Also, for all $e > 0$, write $a_n = \omega(b_n)$ if there exists $n_0 > 0$ such that $a_n > eb_n$ for any $n > n_0$ and write $a_n = \Omega(b_n)$ if there exists $n_0 > 0$ such that $a_n \geq eb_n$ for any $n > n_0$. The notation $f_n = \Theta(g_n)$ means that both $f_n = O(g_n)$ and $g_n = O(f_n)$ hold. We say $X_n = O(Y_n)$ *whp.* (with high probability) if for any $c > 0$ there exist $C = C(c) > 0, n_0 > 0$ such that $\mathbb{P}\left(|X_n| > C\,|Y_n|\right) < n^{-c}$ for all $n > n_0$. Also, we write $X_n = O(Y_n)$ *as.* if $|X_n| \lesssim |Y_n|$ almost surely, and write $X_n = O(1)$ *as.* if there exists $M > 0$ such that $|X_n| \leq M$ for all n almost surely. Finally, we denote a weak convergence of a sequence of random variables by $X_n \Rightarrow X$.

Appendix B. Additional theoretical results

B.1 ENAR model fixed K asymptotic normality results

In this section, we discuss the asymptotic normality results in the ENAR model when K is fixed. These results complement the growing K results discussed in the main text. The first theorem is for growing T case, while the second one is for fixed T case.

Theorem 10 *Assume that* $|\alpha_v| + |\theta_v| < 1$ *and* $N\rho_N = \omega(T + \log^4 N)$. *Then, for* $\mathbf{D}_v \triangleq \sqrt{T}\,\mathrm{diag}\left(\mathbf{I}_K, \sqrt{N}\mathbf{I}_{p+2}\right)$, *we have, as* $N, T \to \infty$,

$$\mathbf{D}_v\left(\hat{\mu}_v - \mu_v^H\right) \Rightarrow \mathcal{N}\left(\mathbf{o}_{K+p+2}, \sigma^2{\Sigma_v}^{-1}\right).$$

Theorem 11 *Assume that* $N\rho_N = \omega(\log^4 N)$. *Partition the parameter vector as* $\mu_v = (\beta_v^\intercal, \mu_{-\beta,v}{}^\intercal)^\intercal$ *and* $\hat{\mu}_v$ *accordingly as well. Then, we have* $\hat{\beta}_v - \mathbf{H}_v{}^\intercal\beta_v = O_{\mathbb{P}}(1)$ *and*

$$\sqrt{N}\left(\hat{\mu}_{-\beta,v} - \mu_{-\beta,v}^H\right) \Rightarrow \mathcal{N}\left(\mathbf{o}_{p+2}, \frac{\sigma^2}{T}{\Sigma_{-u,v}}^{-1}\right), \quad \text{as } N \to \infty. \tag{4}$$

B.2 Network regression model results

Our next two results are related to the Embedding Network Regression (ENR) model. As aforementioned, ENR is a special case of ENAR with $T = 1 < \infty$. Specifically, it can be

derived from ENAR with finite T: let $\mathbf{y}_0 \triangleq \mathbf{1}_N$, $\mathbf{Z} \triangleq \mathbf{Z}_0$, and assume that there is no peer influence effect, i.e., $\theta = 0$.

Without loss of generality, we may omit the grand mean effect, α. Therefore, we have the reduced data model as $\mathbf{w}_i^{r\intercal} \triangleq (\mathbf{x}_{v,i}^{\intercal}, \mathbf{z}_i{}^{\intercal})$ and $\mu^r \triangleq (\beta_v^{r\intercal}, \gamma_v^{r\intercal})^{\intercal} \in \mathbb{R}^{K+p}$, hence giving the representation $\mathbf{y}_1 = \mathbf{W}^r \mu_v^r + \mathcal{E}_1$ for $\mathcal{E}_1 \sim \mathcal{N}(\mathbf{o}_N, \sigma^2 \mathbf{I}_N)$. With a usual least squares estimator $\hat{\mu}_v^r \triangleq (\mathbf{W}^{r\intercal}\mathbf{W}^r)^{-1}\mathbf{W}^{r\intercal}\mathbf{y}_1$, the following two results are corollaries that follow from Theorems 11 and 3, respectively.

Corollary 12 (Fixed K) *Assume that $N\rho_N = \omega(\log^4 N)$. Then, we have $\hat{\beta}_v^r - \mathbf{H}_v{}^{\intercal}\beta_v^r = O_{\mathbb{P}}(1)$ and $\sqrt{N}\,(\hat{\gamma}_v^r - \gamma_v^r) \Rightarrow \mathcal{N}\left(\mathbf{o}_p, \sigma^2 {\Sigma_z}^{-1}\right)$, as $N \to \infty$.*

Corollary 13 (Growing K) *Further assume that $\|\beta_v^r\| = O(1)$ as $K \to \infty$, $N\rho_N = \omega\left(\log^4 N\right)$, and $N = \omega(K^2)$. For a positive integer m, suppose we have an $m \times (K+p+2)$ matrix $\mathbf{Q}_K$ such that $\|\mathbf{Q}_K\| = O(1)$ as $K \to \infty$. Then, we have $\mathbf{Q}_K\left(\hat{\beta}_v^r - \mathbf{H}_v^r{}^{\intercal}\beta_v^r\right) = O_{\mathbb{P}}(1)$ and $\sqrt{N}\,(\hat{\gamma}_v^r - \gamma_v^r) \Rightarrow \mathcal{N}\left(\mathbf{o}_p, \sigma^2 {\Sigma_z}^{-1}\right)$ as $N, K \to \infty$.*

B.3 Misspecification: when $K_o > K$

Suppose that we have embedded more latent variables than the true K value in ENAR and obtained the over-specified model. This is guaranteed with a high probability according to Theorem 30. Therefore, we obtain an augmented estimation of latent variables as $\hat{\mathbf{U}}_{:K_o} \triangleq \left[\hat{\mathbf{U}}_{:K} | \hat{\mathbf{U}}_{K:K_o}\right] \in \mathbb{R}^{N \times K_o}$ with $\hat{\mathbf{U}}_{K:K_o}$, which contains $K_o - K$ auxiliary latent variables. Then, we can pick its population counterpart, which is partitioned as $\mathbf{U} = [\mathbf{U}|\mathbf{U}_{K:K_o}]$. For the case of spectral embedding, $\mathbf{U}_e = \mathbf{V}\mathbf{S}^{1/2}$ contains original K-dimensional latent variable information while $\mathbf{U}_{K:K_o,e} = \mathbf{O}_{N \times (K_o - K)}$ because, for all $i > K$, $s_i(\mathbf{P}) = 0$ as the rank of the population network connectivity $\mathbf{P}$ is K. For eigenvectors we have $\mathbf{U}_v = [\mathbf{V}|\mathbf{V}_{K:K_o}]$ where $\mathbf{V}_{K:K_o} \in \mathbb{R}^{N \times (K_o - K)}$ contains the eigenvectors corresponding to those trivial eigenvalues.

Now let us augment the estimate of the design matrix $\hat{\mathbf{W}}$ as $\hat{\mathbf{W}}_+ \triangleq \left[\mathbf{1}_T \otimes \hat{\mathbf{U}}_{K:K_o} \,|\, \hat{\mathbf{W}}\right] \in \mathbb{R}^{NT \times (K_o + p + 2)}$. Also, denote the augmented true parameter vector by $\mu_+ \in \mathbb{R}^{K_o + p + 2}$, which puts $K_o - K$ null latent effects on those auxiliary latent variables as the true data-generating model is assumed to regress on $\mathbf{U} \in \mathbb{R}^{N \times K}$ only. Then, naturally we obtain a least squares estimator $\hat{\mu}_+ \triangleq \left(\hat{\mathbf{W}}_+^{\intercal}\hat{\mathbf{W}}_+\right)^{-1}\mathbf{W}_+{}^{\intercal}\mathbf{y}$ which can be decomposed as

$$\hat{\mu}_+ = \left(\hat{\mathbf{W}}_+^{\intercal}\hat{\mathbf{W}}_+\right)^{-1}\hat{\mathbf{W}}_+^{\intercal}\left(\mathbf{W}^H\mu^H + \mathcal{E}\right).$$

To keep the notation consistent, we write $\hat{\mu}_{+,e}, \hat{\mu}_{+,v}$ to denote the estimators with spectral embedding and eigenvectors, respectively. Now we state the central limit theorem under the case of over-selection of K.

Theorem 14 *Assume that $N\rho_N = \omega\left(T + \log^4 N\right)$. Define an augmented asymptotic precision covariance as $\Sigma_{+,v} \triangleq \begin{bmatrix} \mathbf{I}_{K_o - K} & \mathbf{O}_{(K_o-K)\times(K+p+2)} \\ & \Sigma_v \end{bmatrix}$. Then, for an augmented convergence rates $\mathbf{D}_{+,v} \triangleq \sqrt{T}\,\mathrm{diag}\left(\mathbf{I}_{K_o}, \sqrt{N}\mathbf{I}_{p+2}\right)$ we have, as $N, T \to \infty$,*

$$\mathbf{D}_{+,v}\left(\hat{\mu}_{+,v} - \mu_{+,v}^H\right) \Rightarrow \mathcal{N}\left(\mathbf{o}_{K_o+p+2}, \sigma^2 {\Sigma_{+,v}}^{-1}\right).$$

Theorem 15 *Assume that* $N\rho_N = \omega\left(\log^4 N\right)$. *Define an augmented asymptotic precision covariance as* $\Sigma_{+,e} \triangleq \begin{bmatrix} \mathbf{O}_{K_o-K} & \mathbf{O}_{(K_o-K)\times(K+p+2)} \\ & \Sigma_e \end{bmatrix}$. *Then, for an augmented convergence rates* $\mathbf{D}_{+,e} \triangleq \sqrt{N}\,\mathrm{diag}\left(\sqrt{\rho_N}\mathbf{I}_{K_o}, \mathbf{I}_{p+2}\right)$ *and* $\hat{\Sigma}_{+,e}$ *defined in* (12), *we have, as* $N \to \infty$,

$$\hat{\Sigma}_{+,e}\mathbf{D}_{+,e}\left(\hat{\mu}_{+,e} - \mu^H_{+,e}\right) \Rightarrow \frac{1}{\sqrt{T}} \cdot \mathcal{N}\left(\mathbf{b}, \sigma^2\Sigma_{+,e}\right)$$

for $\mathbf{b} \triangleq (\mathbf{o}_{K_o-K}{}^\intercal, \mathbf{b}_1{}^\intercal, \mathbf{b}_2{}^\intercal, \mathbf{o}_p{}^\intercal)^\intercal$ *for some fixed* $\mathbf{b}_1 \in \mathbb{R}^K, \mathbf{b}_2 \in \mathbb{R}^2$ *such that* $\|\mathbf{b}_1\| = c_1$ *and* $\|\mathbf{b}_2\| = \sqrt{c_2\rho_\infty}$ *for some constants* $c_1, c_2 > 0$.

The above theorems show that for the fixed K case, the same theoretical results on consistency and asymptotic normality of parameter estimates hold when K is over-selected as well. The rates of convergence for all parameters, including the β parameters corresponding to the latent factors that are truly in the model, are the same as the case when K is known. Note for Theorem 14, we let both $N, T \to \infty$ since that pertains to the network time series and peer effect models with eigenvectors, while for Theorem 15, we let only $N \to \infty$ and keep T fixed since the model with spectral embeddings is intended to be used for the peer effect problem only.

Appendix C. Proofs and additional results for ENAR model in Sections 3 & 4

C.1 Strict stationarity

We start describing our estimation methodology with a discussion on the stationary distribution of $\mathbf{y}_t$ for ENAR. Given our aim to establish its asymptotic distribution under both finite and diverging T and growing network size N, we first derive a stationary solution for $\mathbf{y}_t$. We denote $\mathbb{P}^* \triangleq \mathbb{P}(\cdot|\mathbf{A})$, $\mathbb{E}^* \triangleq \mathbb{E}(\cdot|\mathbf{A})$, and $\mathrm{Cov}^* \triangleq \mathrm{Cov}(\cdot,\cdot|\mathbf{A})$, as the conditional probability, expectation, and covariance, respectively, conditioning on $\mathbf{A}$. Using notations similar to Zhu et al. (2017), define $\mathbf{G} \triangleq \alpha\mathbf{I}_N + \theta\mathcal{L}_\mathbf{A}$ and $\tilde{\mathcal{E}}_{t+1} \triangleq \mathbf{Z}_t\gamma + \mathcal{E}_{t+1}$. Then, we can rewrite the ENAR model equivalently as

$$\mathbf{y}_{t+1} = \mathbf{U}\beta + \mathbf{G}\mathbf{y}_t + \tilde{\mathcal{E}}_{t+1}. \tag{5}$$

When N is fixed, the following results hold.

Theorem 16 *If* $|\alpha| + |\theta| < 1$, *then there is a unique strictly stationary solution to the ENAR model* (5) *with a finite first moment, and the solution is given by,*

$$\mathbf{y}_t = (\mathbf{I}_N - \mathbf{G})^{-1}\mathbf{U}\beta + \sum_{j=0}^{\infty}\mathbf{G}^j\tilde{\mathcal{E}}_{t-j} \tag{6}$$

where the latent variable $\mathbf{U}$ *is either* $\mathbf{U}_v$ *or* $\mathbf{U}_e$.

Lemma 17 *Define* $\Gamma(h) \triangleq \mathrm{Cov}^*(\mathbf{y}_t, \mathbf{y}_{t-h})$ *for all* t. *Upon the conditions in Theorem 16 and conditional on* $\mathbf{A}$, (6) *follows a normal distribution with the mean* $\varphi \triangleq (\mathbf{I}_N - \mathbf{G})^{-1}\mathbf{U}\beta$ *and* $\mathrm{vec}\,\Gamma(0) = (\mathbf{I}_{N^2} - \mathbf{G}\otimes\mathbf{G})^{-1}\,\mathrm{vec}\left\{\left(\sigma^2 + \gamma^\intercal\Sigma_z\gamma\right)\mathbf{I}_N\right\}$ *and* $\Gamma(h) = \begin{cases}\mathbf{G}^h\Gamma(0) & , h > 0 \\ \Gamma(0)\mathbf{G}^{-h} & , h < 0.\end{cases}$.

The proof of Theorem 16, along with all other theorems and lemmas, is contained in the Appendix. These results closely resemble the stationarity results presented in Zhu et al. (2017) for the model without the latent effects. In addition, we now need to distinguish the mean $\mathbb{E}(\mathbf{y}_t) = \varphi$ between the cases $\mathbf{U} = \mathbf{U}_v$ and $\mathbf{U} = \mathbf{U}_e$ by denoting them as φ_v and φ_e, respectively, since they correspond to different means under different modeling choices. For simplicity, we will denote $\Gamma(0)$ by Γ henceforth. Next, we note that $\mathbf{y}_t$ is asymptotically stationary in the sense of Definition 24 in the Appendix when the network size N grows. The following theorem is also proved in the Appendix.

Theorem 18 *Upon the conditions in Theorem 16 with $N \to \infty$,* (6) *is a unique strictly stationary solution with a finite first moment for either latent variable $\mathbf{U}_v$ or $\mathbf{U}_e$. i.e.,* $\max_{1\le i<\infty} \mathbb{E}\,|y_{it}| < \infty$.

C.2 Technical Results

Here, we list some technical results that are useful in proving the asymptotic properties of our estimators for ENAR and AMNAR.

Proposition 19 *If $X_n = O(Y_n)$ whp. and $Y_n = O(Z_n)$ whp., then $X_n = O(Z_n)$ whp. (or as.) If $X_n = O(Y_n)$ whp., then we have $\mathbb{P}\left(\{|X_n| \le C\,|Y_n|\}\ ev.\right) = 1$ by Borel-Cantelli lemma.*

By the model assumptions, the responses of interest, $\mathbf{y}_t$, follow stationary sub-Gaussian distributions. To establish the concentration of our estimators, we employ sub-Gaussian concentration theories as discussed in Vershynin (2018). However, since both the observed graph $\mathbf{A}$ and the covariates $\mathbf{z}_{it}$ are random, the analysis requires investigating the asymptotic behavior of inner products involving multiple random components. Consequently, classical sub-Gaussian concentration inequalities such as Bernstein's or Hoeffding's cannot be applied directly.

For instance, in the case of the Hanson–Wright inequality (Rudelson and Vershynin, 2013), it remains unclear how the tail behavior of the corresponding quadratic forms changes when the associated matrices themselves are random. Under our assumptions on the random graphs and random predictors, these components are shown to be stochastically bounded, which in turn enables the derivation of analogous concentration results.

Lemma 20 (Hoeffding's Inequality) *Let $\mathbf{y} \in \mathbb{R}^n$ be a sub-Gaussian random vector such that $\sup_{\mathbf{u}\in S^{n-1}} \|\mathbf{u}'\mathbf{y}\|_{\psi_2} < \infty$, with independent coordinates and zero mean. Let $\mathbf{x} \in \mathbb{R}^n$ be a random vector such that $\|\mathbf{x}\| = O(g)$ whp. for some $g > 0$, and assume that $\mathbf{x}$ and $\mathbf{y}$ are independent. Then, there exist constants $c > 0$ and $n_0 > 0$ such that*

$$\mathbb{P}(|\mathbf{x}^\intercal\mathbf{y}| \ge \upsilon) \le 2\exp\left(-\frac{c\upsilon^2}{g^2}\right) + \frac{1}{n}$$

for all $n > n_0$ and $\upsilon > 0$. If g is a constant, then the $\frac{1}{n}$ term on the right-hand side disappears.

Proof By the general Hoeffding inequality, we have for every fixed realization of $\mathbf{x}$,

$$\mathbb{P}\left(|\mathbf{x}^\intercal\mathbf{y}| \ge \upsilon \mid \mathbf{x}\right) \le 2\exp\left(-\frac{c\upsilon^2}{\|\mathbf{x}\|^2}\right).$$

Denote the left- and right-hand sides by $f_n(\mathbf{x})$ and $h_n(\mathbf{x})$, respectively. By assumption, there exist constants $M, n_0 > 0$ such that $\mathbb{P}(\|\mathbf{x}\| > Mg) < 1/n$ for all $n > n_0$. Define the event $A_n \triangleq \{\|\mathbf{x}\| \leq Mg\}$. Then

$$f_n(\mathbf{x})\mathbb{I}_{A_n} \leq h_n(\mathbf{x})\mathbb{I}_{A_n} \leq 2\exp\left(-\frac{c'\upsilon^2}{g^2}\right)\mathbb{I}_{A_n}$$

for some constant $c' > 0$. Since A_n is $\sigma(\mathbf{x})$-measurable, taking expectations yields

$$\mathbb{E}\left[\mathbb{P}\left(A_n \cap \{|\mathbf{x}^\intercal\mathbf{y}| \geq \upsilon\} \mid \mathbf{x}\right)\right] = \mathbb{P}\left(A_n \cap \{|\mathbf{x}^\intercal\mathbf{y}| \geq \upsilon\}\right) \leq 2\exp\left(-\frac{c'\upsilon^2}{g^2}\right)\mathbb{P}(A_n).$$

Hence,

$$\begin{aligned} \mathbb{P}\left(\{|\mathbf{x}^\intercal\mathbf{y}| \geq \upsilon\} \setminus A_n^c\right) &\leq 2\exp\left(-\frac{c'\upsilon^2}{g^2}\right), \\ \therefore \quad \mathbb{P}\left(|\mathbf{x}^\intercal\mathbf{y}| \geq \upsilon\right) &\leq 2\exp\left(-\frac{c'\upsilon^2}{g^2}\right) + \mathbb{P}(A_n^c). \end{aligned}$$

For almost surely bounded $\mathbf{x}$, the term $\mathbb{P}(A_n^c)$ vanishes, yielding the desired result. ■

Lemma 21 *In Lemma 20, now suppose that* $\|\mathbf{x}\| = O_{\mathbb{P}}(g)$. *Then,* $\forall \epsilon, \upsilon > 0$ *there exist some* $c > 0, n_0 > 0$ *such that*

$$\mathbb{P}(|\mathbf{x}^\intercal\mathbf{y}| \geq \upsilon) \leq 2\exp\left(-\frac{c\upsilon^2}{g^2}\right) + \epsilon$$

for all $n > n_0$.

The proof of Lemma 21 coincides with Lemma 20.

Lemma 22 *(Hanson-Wright Inequality) Let* $\mathbf{y} \in \mathbb{R}^n$ *be a random vector with independent coordinates and zero mean such that* $\sup_{\mathbf{u}\in S^{n-1}} \|\mathbf{u}'\mathbf{y}\|_{\psi_2} < \infty$. *Let* $\mathbf{S} \in \mathbb{R}^{n\times n}$ *be a random matrix which is independent with* $\mathbf{y}$ *such that* $\|\mathbf{S}\|_F^2 = O(h)$ *whp. and* $\|\mathbf{S}\| = O(k)$ *whp. for some* $h, k > 0$. *Then, there exists* $n_0 > 0$ *such that*

$$\mathbb{P}\left(|\mathbf{y}^\intercal\mathbf{S}\mathbf{y} - \mathbb{E}(\mathbf{y}^\intercal\mathbf{S}\mathbf{y})| > \upsilon\right) \leq 2\exp\left[-c\min\left(\frac{\upsilon^2}{h}, \frac{\upsilon}{k}\right)\right] + \frac{2}{n}$$

for all $n > n_0$ *and* $\upsilon > 0$. *If* $\|\mathbf{S}\|_F^2 = O(h)$ *as. and* $\|\mathbf{S}\| = O(k)$ *as. , then we have*

$$\mathbb{P}\left(|\mathbf{y}^\intercal\mathbf{S}\mathbf{y} - \mathbb{E}(\mathbf{y}^\intercal\mathbf{S}\mathbf{y})| > \upsilon\right) \leq 2\exp\left[-c\min\left(\frac{\upsilon^2}{h}, \frac{\upsilon}{k}\right)\right]$$

for all $n > 0$.

C.3 Concentration of ASE

In this section, we prove the asymptotic order of the measurement errors when approximating $\mathbf{U}$ by $\hat{\mathbf{U}}$ using either eigenvectors or spectral embedding. First, we have $s_K(\mathbf{P}) = \Theta(N\rho_N)$. By Davis-Kahan Theorem, we have

$$\min_{\mathbf{H}\in\mathcal{O}_K} \left\|\hat{\mathbf{V}} - \mathbf{V}\mathbf{H}\right\| = \mathcal{O}\left(\frac{\|\mathbf{A}-\mathbf{P}\|}{s_K(\mathbf{P})}\right) = \mathcal{O}\left(\frac{\|\mathbf{A}-\mathbf{P}\|}{N\rho_N}\right).$$

That is, a large eigengap reduces the error of approximating $\mathbf{V}$ by $\hat{\mathbf{V}}$, which in turn stabilizes the downstream regression estimates in ENAR.

Proposition 23 *Assume that* $\mathbf{A} \sim \text{RDPG}(\rho_N, \mathbf{X})$ *and there exists constants* $c, N_0 > 0$ *such that* $cN\rho_N \leq s_K(\mathbf{P})$ *for all* $N > N_0$. *Then, if* $N\rho_N = \omega\left(\log^4 N\right)$, *there exists a sequence of matrices* $\mathbf{H}_v, \mathbf{H}_e \in \mathcal{O}_K$ *such that*

$$\left\|\hat{\mathbf{V}} - \mathbf{V}\mathbf{H}_v\right\| = O\left(\frac{1}{\sqrt{N\rho_N}}\right) \ \textit{whp.}$$

$$\left\|\hat{\mathbf{V}}|\hat{\mathbf{S}}|^{1/2} - \mathbf{V}\mathbf{S}^{1/2}\mathbf{H}_e\right\| = O\left(1 + \sqrt{\frac{K\log^2 N}{N\rho_N}}\right) \ \textit{whp.}$$

Proof The spectral norm version of Lemma 5.1 of Lei and Rinaldo (2015) is straightforward as we have $\inf_{\mathbf{H}\in\mathcal{O}_K}\left\|\hat{\mathbf{V}} - \mathbf{V}\mathbf{H}\right\| = \Theta\left\{\left\|\sin\angle(\mathbf{V},\hat{\mathbf{V}})\right\|\right\}$ and:

$$\left\|\sin\angle(\mathbf{V},\hat{\mathbf{V}})\right\| = \Theta\left\{\left\|\mathbf{V}\mathbf{V}^\intercal\left(\mathbf{I} - \hat{\mathbf{V}}\hat{\mathbf{V}}\right)^\intercal\right\|\right\}$$

where $\sin\angle(\mathbf{V},\hat{\mathbf{V}}) \triangleq \text{diag}\left[\sin\circ\cos^{-1} s_1(\mathbf{V}^\intercal\hat{\mathbf{V}}), \ldots, \sin\circ\cos^{-1} s_K(\mathbf{V}^\intercal\hat{\mathbf{V}})\right]$ denotes the distances between the subspaces spanned by $\mathbf{V}$ and $\hat{\mathbf{V}}$, measured in the sin's of canonical angles (Lemma 1 of Cai and Zhang, 2018). By Davis–Kahan Theorem, we have

$$\left\|\mathbf{V}\mathbf{V}^\intercal\left(\mathbf{I} - \hat{\mathbf{V}}\hat{\mathbf{V}}\right)^\intercal\right\| \leq \frac{2\|\mathbf{A}-\mathbf{P}\|}{s_K(\mathbf{P})} = O\left(\frac{1}{\sqrt{N\rho_N}}\right) \ \textit{whp.}$$

so, there exists $\mathbf{H}_v \in \mathcal{O}_K$ such that $\left\|\hat{\mathbf{V}} - \mathbf{V}\mathbf{H}_v\right\| = O\left(\frac{1}{\sqrt{N\rho_N}}\right)$ *whp.*, and this completes the proof for the concentration of $\hat{\mathbf{V}} - \mathbf{V}\mathbf{H}_v$.

Now we proceed to the case of spectral embedding. Let us denote the orthogonal matrices containing left and right singular vectors of $\mathbf{V}^\intercal\hat{\mathbf{V}}$ by $\mathbf{H}_1$ and $\mathbf{H}_2$ respectively, and define an orthogonal matrix $\mathbf{H}^* \triangleq \mathbf{H}_1\mathbf{H}_2^\intercal \in \mathcal{O}_K$. Then, we have

$$\left(\mathbf{H}^*|\hat{\mathbf{S}}|^{1/2} - \mathbf{S}^{1/2}\mathbf{H}^*\right)_{ij} = (\mathbf{H}^*)_{ij}\frac{|s_j(\mathbf{A})| - s_i(\mathbf{P})}{|s_j(\mathbf{A})|^{1/2} + s_i(\mathbf{P})^{1/2}} \leq (\mathbf{H}^*)_{ij}\frac{|s_j(\mathbf{A})| - s_i(\mathbf{P})}{s_K(\mathbf{P})^{1/2}}$$

hence by Kato-Temple inequality (Cape et al., 2017) and $\|\mathbf{H}\|_F = \sqrt{K}$,

$$\left\|\mathbf{H}^*|\hat{\mathbf{S}}|^{1/2} - \mathbf{S}^{1/2}\mathbf{H}^*\right\| \leq \left\|\mathbf{H}^*|\hat{\mathbf{S}}|^{1/2} - \mathbf{S}^{1/2}\mathbf{H}^*\right\|_F = O\left(\sqrt{\frac{K\log^2 N}{N\rho_N}}\right) \ \textit{whp.}$$

Let $\mathbf{R}_1 \triangleq \mathbf{V}\mathbf{V}^\intercal\hat{\mathbf{V}} - \mathbf{V}\mathbf{H}^*$ and $\mathbf{R}_2 \triangleq \mathbf{H}^*|\hat{\mathbf{S}}|^{1/2} - \mathbf{S}^{1/2}\mathbf{H}^*$. Note that $\hat{\mathbf{V}}|\hat{\mathbf{S}}|^{1/2} = \mathbf{A}\hat{\mathbf{V}}|\hat{\mathbf{S}}|^{-1/2}$ and $\mathbf{V}^\intercal\mathbf{P} = \mathbf{S}\mathbf{V}^\intercal$. Then,

$$\begin{aligned}
\hat{\mathbf{V}}|\hat{\mathbf{S}}|^{1/2} - \mathbf{V}\mathbf{S}^{1/2}\mathbf{H}^* &= \hat{\mathbf{V}}|\hat{\mathbf{S}}|^{1/2} - \mathbf{V}\mathbf{H}^*|\hat{\mathbf{S}}|^{1/2} + \mathbf{V}\mathbf{R}_2 \\
&= \left(\hat{\mathbf{V}} - \mathbf{V}\mathbf{V}^\intercal\hat{\mathbf{V}}\right)|\hat{\mathbf{S}}|^{1/2} + \mathbf{R}_1|\hat{\mathbf{S}}|^{1/2} + \mathbf{V}\mathbf{R}_2 \\
&= \left(\mathbf{I}_N - \mathbf{V}\mathbf{V}^\intercal\right)\mathbf{A}\hat{\mathbf{V}}|\hat{\mathbf{S}}|^{-1/2} + \mathbf{R}_1|\hat{\mathbf{S}}|^{1/2} + \mathbf{V}\mathbf{R}_2 \\
&= \left(\mathbf{I}_N - \mathbf{V}\mathbf{V}^\intercal\right)\left(\mathbf{A} - \mathbf{P}\right)\hat{\mathbf{V}}|\hat{\mathbf{S}}|^{-1/2} + \mathbf{R}_1|\hat{\mathbf{S}}|^{1/2} + \mathbf{V}\mathbf{R}_2.
\end{aligned}$$

It is easy to check that $\left\|\mathbf{V}^\intercal\hat{\mathbf{V}} - \mathbf{H}^*\right\| = O\left(\frac{1}{N\rho_N}\right)$ *whp.* as

$$\begin{aligned}
\left\|\mathbf{V}^\intercal\hat{\mathbf{V}} - \mathbf{H}^*\right\| &\leq \max_{i=1,\ldots,K}\left\{1 - s_i(\mathbf{V}^\intercal\hat{\mathbf{V}})\right\} \\
&\leq \max_i\left|1 - s_i(\mathbf{V}^\intercal\hat{\mathbf{V}})^2\right| = \left\|\sin\angle(\mathbf{V},\hat{\mathbf{V}})\right\|^2 \lesssim \frac{\|\mathbf{A}-\mathbf{P}\|^2}{s_K(\mathbf{P})^2}.
\end{aligned}$$

So, we observe that

$$\|\mathbf{R}_1\| \leq \left\|\mathbf{V}^\intercal\hat{\mathbf{V}} - \mathbf{H}^*\right\| = O\left(\frac{1}{N\rho_N}\right), \|\mathbf{R}_2\| = O\left(\sqrt{\frac{K\log^2 N}{N\rho_N}}\right) \text{ whp.}$$

and

$$\left\|\left(\mathbf{I}_N - \mathbf{V}\mathbf{V}^\intercal\right)\left(\mathbf{A} - \mathbf{P}\right)\hat{\mathbf{V}}|\hat{\mathbf{S}}|^{-1/2}\right\| \leq \|\mathbf{A}-\mathbf{P}\|\,|s_K(\hat{\mathbf{S}})|^{-1/2} = O(1) \text{ whp.}$$

by $|s_K(\mathbf{A})| = \Theta(N\rho_N)$ *whp.* Therefore, we have

$$\left\|\hat{\mathbf{V}}|\hat{\mathbf{S}}|^{1/2} - \mathbf{V}\mathbf{S}^{1/2}\mathbf{H}^*\right\| = O\left(1 + \sqrt{\frac{K\log^2 N}{N\rho_N}}\right) \text{ whp.}$$

■

Proof [of Theorem 7] In Theorem 7, we establish the asymptotic orders of ASE concentration under under-embedding, i.e., the case where we embed k eigenvectors of $\mathbf{A}$ while the true latent dimension of the graph regime, $\text{rank}(\mathbf{P})$, is $K > k$.

Basic steps are similar to the proof of Proposition 23. Again we have $\inf_{\mathbf{H}_k\in\mathcal{O}_{K,k}}\left\|\hat{\mathbf{V}}_{:k} - \mathbf{V}\mathbf{H}_k\right\| = \Theta\left\{\left\|\sin\angle(\mathbf{V},\hat{\mathbf{V}}_{:k})\right\|\right\}$ and:

$$\left\|\sin\angle(\mathbf{V},\hat{\mathbf{V}}_{:k})\right\| = \Theta\left\{\left\|\mathbf{V}\mathbf{V}^\intercal\left(\mathbf{I} - \hat{\mathbf{V}}_{:k}\hat{\mathbf{V}}_{:k}\right)^\intercal\right\|\right\}.$$

Now let $\mathbf{A}_k = \hat{\mathbf{V}}_{:k}\hat{\mathbf{S}}_{:k}\hat{\mathbf{V}}_{:k}^\intercal$ be a truncated eigenvalue decomposition of $\mathbf{A}$ where $\hat{\mathbf{S}}_{:k}$ is a diagonal matrix with k ($\leq K$) largest magnitude eigenvalues of $\mathbf{A}$ and $\hat{\mathbf{V}}_{:k} \in \mathcal{O}_{N,k}$ is a matrix with corresponding k eigenvectors. By Weyl's inequality, for $\hat{\mathbf{V}} = \hat{\mathbf{V}}_{:K} = \left[\hat{V}_1 \cdots \hat{V}_K\right]$ we have

$$\|\mathbf{A} - \mathbf{A}_k\| = \left\|\sum_{i=k+1}^{K} s_i(\mathbf{A})\hat{V}_i\hat{V}_i^\intercal\right\| = s_{k+1}(\mathbf{A}) \leq s_{k+1}(\mathbf{P}) + \|\mathbf{A}-\mathbf{P}\|.$$

So, we have $\|\mathbf{A}_k - \mathbf{P}\| \leq s_{k+1}(\mathbf{P}) + 2\|\mathbf{A} - \mathbf{P}\|$. Then, by Davis–Kahan Theorem,

$$\left\|\mathbf{V}\mathbf{V}^\intercal\left(\mathbf{I} - \hat{\mathbf{V}}_{:k}\hat{\mathbf{V}}_{:k}\right)^\intercal\right\| \leq \frac{2\|\mathbf{A}_k - \mathbf{P}\|}{s_K(\mathbf{P})} \leq \frac{2s_{k+1}(\mathbf{P})}{s_K(\mathbf{P})} + O\left(\frac{1}{\sqrt{N\rho_N}}\right) \ whp.$$

provided that $N\rho_N = \omega\left(\log^4 N\right)$. Therefore, there exists $\mathbf{H}_{k,v} \in \mathcal{O}_{K,k}$ such that

$$\left\|\hat{\mathbf{V}}_{:k} - \mathbf{V}\mathbf{H}_{k,v}\right\| = q_{k+1} + O\left(\frac{1}{\sqrt{N\rho_N}}\right) \ whp.$$

where $q_{k+1} \triangleq \frac{2s_{k+1}(\mathbf{P})}{s_K(\mathbf{P})}$. Notice that $q_k \geq q_{k+1}$ for $k = 1, \ldots, K$ and $q_{k+1} = \Theta(1)$ as $s_i(\mathbf{P}) = \Theta\left(N\rho_N\right)$ for all $i = 1, \ldots, K$ and $q_{k+1} = 0$ for all $k \geq K$. This proves the first conclusion.

Now consider the case of spectral embedding. Let us denote the orthogonal matrices containing left and right singular vectors of $\mathbf{V}^\intercal\hat{\mathbf{V}}_{:k}$ by $\mathbf{R}_k$ and $\mathbf{R}'_k$ respectively, and define $\mathbf{H}^*_k \triangleq \mathbf{R}_k\mathbf{R}_k{}^\intercal \in \mathcal{O}_{K,k}$. We have

$$\left(\mathbf{H}^*_k|\hat{\mathbf{S}}_{:k}|^{1/2} - \mathbf{S}^{1/2}\mathbf{H}^*_k\right)_{ij} = (\mathbf{H}^*_k)_{ij}\frac{|s_j(\mathbf{A}_k)| - s_i(\mathbf{P})}{|s_j(\mathbf{A}_k)|^{1/2} + s_i(\mathbf{P})^{1/2}}\begin{cases} i = 1, \ldots, K \\ j = 1, \ldots, k \end{cases}$$

hence by $\|\mathbf{H}^*_k\|_F = \sqrt{k}$ and the increased magnitudes of the elements at $i > k$,

$$\left\|\mathbf{H}^*_k|\hat{\mathbf{S}}_{:k}|^{1/2} - \mathbf{S}^{1/2}\mathbf{H}^*_k\right\| \leq \left\|\mathbf{H}^*_k|\hat{\mathbf{S}}_{:k}|^{1/2} - \mathbf{S}^{1/2}\mathbf{H}^*_k\right\|_F = O\left(\sqrt{kN\rho_N}\right) \ whp.$$

Let $\mathbf{R}_1 \triangleq \mathbf{V}\mathbf{V}^\intercal\hat{\mathbf{V}}_{:k} - \mathbf{V}\mathbf{H}^*_k$ and $\mathbf{R}_2 \triangleq \mathbf{H}^*_k|\hat{\mathbf{S}}_{:k}|^{1/2} - \mathbf{S}^{1/2}\mathbf{H}^*_k$. Since $\left\|\hat{\mathbf{V}}_{:k}|\hat{\mathbf{S}}_{:k}|^{1/2}\right\| = \left\|\mathbf{A}_k\hat{\mathbf{V}}_{:k}|\hat{\mathbf{S}}_{:k}|^{-1/2}\right\|$ and $\mathbf{V}^\intercal\mathbf{P} = \mathbf{S}\mathbf{V}^\intercal$, we have

$$\begin{aligned}
\|\hat{\mathbf{V}}_{:k}|\hat{\mathbf{S}}_{:k}|^{1/2} - \mathbf{V}\mathbf{S}^{1/2}\mathbf{H}^*_k\| &\leq \left\|\hat{\mathbf{V}}_{:k}|\hat{\mathbf{S}}_{:k}|^{1/2} - \mathbf{V}\mathbf{H}^*_k|\hat{\mathbf{S}}_{:k}|^{1/2}\right\| + \|\mathbf{V}\mathbf{R}_2\| \\
&\leq \left\|\left(\hat{\mathbf{V}}_{:k} - \mathbf{V}\mathbf{V}^\intercal\hat{\mathbf{V}}_{:k}\right)|\hat{\mathbf{S}}_{:k}|^{1/2}\right\| + \left\|\mathbf{R}_1|\hat{\mathbf{S}}_{:k}|^{1/2}\right\| + \|\mathbf{R}_2\| \\
&\leq \left\|\left(\mathbf{I}_N - \mathbf{V}\mathbf{V}^\intercal\right)\mathbf{A}_k\hat{\mathbf{V}}_{:k}|\hat{\mathbf{S}}_{:k}|^{-1/2}\right\| + \left\|\mathbf{R}_1|\hat{\mathbf{S}}_{:k}|^{1/2}\right\| + \|\mathbf{R}_2\| \\
&\leq \left\|\left(\mathbf{I}_N - \mathbf{V}\mathbf{V}^\intercal\right)\left(\mathbf{A}_k - \mathbf{P}\right)\hat{\mathbf{V}}_{:k}|\hat{\mathbf{S}}_{:k}|^{-1/2}\right\| + \left\|\mathbf{R}_1|\hat{\mathbf{S}}_{:k}|^{1/2}\right\| + \|\mathbf{R}_2\|.
\end{aligned}$$

Check that $\left\|\mathbf{V}^\intercal\hat{\mathbf{V}}_{:k} - \mathbf{H}^*_k\right\| = O(1) \ whp.$, and we have

$$\|\mathbf{R}_1\| = \left\|\mathbf{V}^\intercal\hat{\mathbf{V}}_{:k} - \mathbf{H}^*_k\right\| = O(1), \|\mathbf{R}_2\| = O\left(\sqrt{N\rho_N}\right) \ whp.$$

and

$$\left\|\left(\mathbf{I}_N - \mathbf{V}\mathbf{V}^\intercal\right)\left(\mathbf{A}_k - \mathbf{P}\right)\hat{\mathbf{V}}_{:k}|\hat{\mathbf{S}}_{:k}|^{-1/2}\right\| \leq \|\mathbf{A}_k - \mathbf{P}\|\left\||\hat{\mathbf{S}}_{:k}|^{-1/2}\right\| = O\left(\sqrt{N\rho_N}\right) \ whp.$$

Therefore, we have

$$\left\|\hat{\mathbf{V}}_{:k}|\hat{\mathbf{S}}_{:k}|^{1/2} - \mathbf{V}\mathbf{S}^{1/2}\mathbf{H}^*_k\right\| = O\left(\sqrt{N\rho_N}\right) \ whp.$$

■

C.4 Proofs of theorems on Stationarity

To show stationarity, we adapt the proofs of Zhu et al. (2017) to the setup of ENAR.

Proof [of Theorem 16] First, note that the maximum absolute eigenvalue of $\mathcal{L}_{\mathbf{A}}$ is strictly less than one. So it is straightforward that the spectral radius of $\mathbf{G}$, denoted by g, satisfies

$$g \leq |\alpha| + |\theta| < 1 \tag{7}$$

with probability one. Therefore, $\sum_{j=0}^{\infty} \mathbf{G}^j \tilde{\mathcal{E}}_{t-j}$ exists $as.$, and $\mathbf{y}_t$ in (6) is a strictly stationary process. It is straightforward that (6) satisfies (3). Next, assume that $\bar{\mathbf{y}}_t$ is another strictly stationary solution with $\mathbb{E}\|\bar{\mathbf{y}}_t\| < \infty$. Then,

$$\bar{\mathbf{y}}_t = \sum_{j=0}^{m-1} \mathbf{G}^j(\mathbf{U}\beta + \tilde{\mathcal{E}}_{t-j}) + \mathbf{G}^m \bar{\mathbf{y}}_{t-m}$$

for any positive integer m. Therefore,

$$\mathbb{E}^*\|\mathbf{y}_t - \bar{\mathbf{y}}_t\| = \mathbb{E}^* \left\| \sum_{j=m}^{\infty} \mathbf{G}^j(\mathbf{U}\beta + \tilde{\mathcal{E}}_{t-j}) - \mathbf{G}^m \bar{\mathbf{y}}_t \right\| \leq C g^m$$

for a constant C independent of t and m. Growing m to infinity, we get $\mathbb{E}\|\mathbf{y}_t - \bar{\mathbf{y}}_t\| = 0$ hence $\mathbf{y}_t = \bar{\mathbf{y}}_t$ almost surely. ■

Therefore, as $\mathbf{G}^j \tilde{\mathcal{E}}_{t-j} \sim \mathcal{N}_N\left[\mathbf{o}, \left(\sigma^2 + \gamma^\intercal \Sigma_z \gamma\right) \mathbf{G}^{2j}\right]$, the movinag average representation in (6) will follow a normal distribution with mean $\mathbb{E}^*(\mathbf{y}_t) = (\mathbf{I}_N - \mathbf{G})^{-1}\mathbf{U}\beta$ and $\mathrm{Var}^*(\mathbf{y}_t) = \left(\sigma^2 + \gamma^\intercal \Sigma_z \gamma\right) \sum_{j=0}^{\infty} \mathbf{G}^{2j}$ for all t. After solving the Yule-Walker equation for determining Γ, we get Lemma 17. Next, we prove Theorem 18 according to the following definition.

Definition 24 (Zhu et al., 2017) *Let $\{\mathbf{y}_t \in \mathbb{R}^N\}$ be an N–dimensional vector with $N \to \infty$. Define $\mathcal{M} \triangleq \{\omega \in \mathbb{R}^\infty : \sum_{i=1}^{\infty} |w_i| < \infty\}$. For each $\omega \in \mathcal{M}$, let $\mathbf{w}_N = (w_1, ..., w_N)^\intercal \in \mathbb{R}^N$ be the truncated N–dimensional process. $\mathbf{y}_t$ is said to be strictly stationary if $\forall \omega \in \mathcal{M}$*

1. *$\mathbf{y}_t^\omega = \lim_{N\to\infty} \mathbf{w}_N{}^\intercal \mathbf{y}_t$ exists almost surely.*
2. *$\mathbf{y}_t^\omega$ is strictly stationary.*

Proof [of Theorem 18] To prove the existence of a stationary solution, it suffices to show that (6) is strictly stationary according to the above definition. Write $|\mathbf{A}|_{\epsilon\ell}$ as a matrix of absolute elements of a matrix $\mathbf{A}$. Moreover, write $\mathbf{A} \preccurlyeq \mathbf{B}$ if $\mathbf{B}$ is not less than $\mathbf{A}$ elementwisely. Recall that $\mathbf{y}_t = \sum_{j=0}^{m-1} \mathbf{G}^j(\mathbf{U}\beta + \tilde{\mathcal{E}}_{t-j}) + \mathbf{G}^m \mathbf{y}_{t-m}$ hence

$$\mathbf{y}_t = \lim_{m\to\infty} \mathbf{y}_t = \sum_{j=0}^{\infty} \mathbf{G}^j(\mathbf{U}\beta + \tilde{\mathcal{E}}_{t-j}).$$

i. For the columns of $\mathbf{U} = \mathbf{V}$, say $V_1, ..., V_K$, We have $\mathbf{U}\beta = \sum_{j=1}^{K} V_j \beta_j$ and its ℓ_∞-norm is bounded as $\|\mathbf{U}\beta\|_\infty \leq \sum_{j=1}^{K} |\beta_j| \, \|V_j\|_\infty \leq \|\beta\|_1 \, \mathbf{1}_N$ hence $|\mathbf{U}\beta|_{\epsilon\ell} \preccurlyeq \|\beta\|_1 \, \mathbf{1}_N$.

ii. For $\mathbf{U} = \mathbf{V}\mathbf{S}^{1/2}$, then $\|\mathbf{U}\beta\|_\infty \lesssim \|\mathbf{V}\|_\infty \, s_1(\mathbf{P})^{1/2} \, \|\beta\|_\infty \leq \sqrt{K} \, \|\mathbf{V}\|_{2,\infty} \, s_1(\mathbf{P})^{1/2} \, \|\beta\|_\infty = O\left(\sqrt{K\rho_N} \, \|\beta\|_\infty\right)$ hence $|\mathbf{U}\beta|_{\epsilon\ell} \preccurlyeq \|\beta\|_\infty \sqrt{K\rho_N} \mathbf{1}_N$. Note that K is fixed in this proof and ρ_N is not greater than one.

So, for both choice of $\mathbf{U}$, we have $\mathbb{E}\left|\mathbf{U}\beta+\tilde{\mathcal{E}}_{t-j}\right|_{\epsilon\ell} \preccurlyeq C\cdot\mathbf{1}_N$ for $C=\sqrt{K}\left\|\beta\right\|_1+\mathbb{E}\left|\mathbf{z}_1^\intercal\gamma\right|+\mathbb{E}\left|\epsilon_{11}\right|$. Since $\mathcal{L}_{\mathbf{A}}$ is normal and symmetric, its spectral decomposition can be given as $\mathcal{L}_{\mathbf{A}}=\mathbf{U}_{\mathcal{L}}\mathbf{S}_{\mathcal{L}}\mathbf{U}_{\mathcal{L}}{}^\intercal$ with orthogonal eigenvectors $\mathbf{U}_{\mathcal{L}}\in\mathbb{R}^{N\times N}$ and $\mathbf{S}_{\mathcal{L}}$ containing corresponding eigenvalues of $\mathcal{L}_{\mathbf{A}}$. So, we have $\left\|\mathcal{L}_{\mathbf{A}}{}^j\right\|_\infty\leq\left\|\mathbf{U}_{\mathcal{L}}\right\|_\infty\left\|\mathbf{U}_{\mathcal{L}}{}^\intercal\right\|_\infty\rho\left(\mathcal{L}_{\mathbf{A}}\right)^j\leq C'$ for a constant $C'>0$ independent of j. Therefore, $|\mathbf{G}|_{\epsilon\ell}^j\mathbf{1}_N=(|\theta|\mathcal{L}_{\mathbf{A}}+|\alpha|\mathbf{I}_N)^j\mathbf{1}_N\preccurlyeq C'(|\theta|+|\alpha|)^j\mathbf{1}_N$. Consequently,

$$\begin{aligned}\mathbb{E}^*\left|\mathbf{w}_N{}^\intercal\mathbf{y}_t\right|\leq\left\|\mathbf{w}_N\right\|_1\mathbb{E}^*\left\|\mathbf{y}_t\right\|_\infty&\leq\sum_{i=1}^\infty|w_i|\sum_{j=0}^\infty\mathbb{E}^*\left\|\left|\mathbf{G}\right|_{\epsilon\ell}^j\left|\mathbf{U}\beta+\tilde{\mathcal{E}}_{t-j}\right|_{\epsilon\ell}\right\|_\infty\\&\lesssim\sum_{i=1}^\infty|w_i|\sum_{j=0}^\infty(|\theta|+|\alpha|)^j\end{aligned}$$

implying that $\lim_{N\to\infty}\mathbf{w}_N{}^\intercal\mathbf{y}_t$ exists almost surely. Next, assume that $\bar{\mathbf{y}}_t$ is another strictly stationary solution with a finite first moment. Then, $\mathbb{E}\left|\bar{\mathbf{y}}_t\right|_{\epsilon\ell}\precsim\mathbf{1}_N$. We have

$$\mathbb{E}^*\left|\mathbf{w}_N{}^\intercal(\mathbf{y}_t-\bar{\mathbf{y}}_t)\right|=\mathbb{E}^*\left|\sum_{j=m}^\infty\mathbf{w}_N{}^\intercal\mathbf{G}^j(\mathbf{U}\beta+\tilde{\mathcal{E}}_{t-j})-\mathbf{w}_N{}^\intercal\mathbf{G}^m\bar{\mathbf{y}}_{t-m}\right|$$

which is bounded above by the product of a constant and

$$\sum_{i=1}^\infty|w_i|\sum_{j=m}^\infty\left\{(|\alpha|+|\theta|)^j+(|\alpha|+|\theta|)^m\right\}$$

for any $\omega\in\mathcal{M}$. Growing $m\to\infty$, we have $\mathbf{w}_N{}^\intercal(\mathbf{y}_t-\bar{\mathbf{y}}_t)=0$ *as.* hence $\mathbf{y}_t=\bar{\mathbf{y}}_t$ *as.* ■

C.5 When ENAR embeds eigenvectors $\mathbf{U}=\mathbf{V}$ under growing N,T

C.5.1 Consistency

Here, we prove the asymptotic normality of $\hat{\mu}$ for both cases where K is fixed and growing. First, we clarify some notations here. Let $\Gamma_{\mathbf{y}}\triangleq\mathrm{Cov}^*(\mathbf{y})$. Write $\Phi_v\triangleq\mathbf{1}_T\otimes\varphi_v$ and $\tilde{\mathbf{y}}\triangleq{\Gamma_{\mathbf{y}}}^{-\frac{1}{2}}(\mathbf{y}-\Phi_v)$ so that $\mathbf{y}={\Gamma_{\mathbf{y}}}^{1/2}\tilde{\mathbf{y}}+\Phi_v$. Note that the entries of $\tilde{\mathbf{y}}$ are independent by the property of the multivariate normal distribution.

By Proposition 2.2 and 2.3 of Basu and Michailidis (2015), with probability one, we have

$$\left\|\Gamma_{\mathbf{y}}\right\|\leq\frac{\sigma^2}{m_{\min}(\mathbf{G})}\leq\frac{\sigma^2\left\|\mathbf{V}_{\mathbf{G}}\right\|^2\left\|{\mathbf{V}_{\mathbf{G}}}^{-1}\right\|^2}{(1-g)^2}$$

where g is the spectral radius of $\mathbf{G}$ in (7), $\mathbf{V}_{\mathbf{G}}$ is an orthogonal matrix that contains eigenvectors of $\mathbf{G}$ and $m_{\min}(\mathbf{G})\triangleq\min_{\{z\in\mathcal{C};|z|=1\}}(\mathbf{I}_N-\mathbf{G}z)^*(\mathbf{I}_N-\mathbf{G}z)$. The last upper bound holds because $\mathbf{G}$ is diagonalizable. So, we have $\left\|\Gamma_{\mathbf{y}}\right\|=O(1)$ *as.*

Proof [of Theorem 10] First, define $\tilde{\mathbf{V}}\triangleq\mathbf{V}\mathbf{H}_v$ and let $\mathbf{W}_t^H\triangleq\left[\tilde{\mathbf{V}}\,|\,\mathbf{y}_t,\mathcal{L}_{\mathbf{A}}\mathbf{y}_t\,|\,\mathbf{Z}_t\right]$, $\mu_{w,v}^H=\left(\frac{\beta_v{}^\intercal\mathbf{H}_v}{\sqrt{N}},\alpha_v,\theta_v,\gamma_v^\intercal\right)^\intercal$, $\mathcal{W}_t\triangleq\left[\sqrt{N}\mathbf{V}\,|\,\mathbf{y}_t,\mathcal{L}_{\mathbf{A}}\mathbf{y}_t\,|\,\mathbf{Z}_t\right]$, $\mathcal{W}_t^H\triangleq\left[\sqrt{N}\tilde{\mathbf{V}}\,|\,\mathbf{y}_t,\mathcal{L}_{\mathbf{A}}\mathbf{y}_t\,|\,\mathbf{Z}_t\right]$, and $\hat{\mathcal{W}}_t=$

$\left[\sqrt{N}\hat{\mathbf{V}} \,|\, \mathbf{y}_t, \mathcal{L}_{\mathbf{A}}\mathbf{y}_t \,|\, \mathbf{Z}_t\right]$. Under this representation, we have $\hat{\mathcal{W}}_t\mathbf{D}_v = \sqrt{NT}\hat{\mathbf{W}}_t$, $\sqrt{NT}{\mathbf{D}_v}^{-1}\mu_{w,v}^H = \mu_v^H$, and $\mathbf{W}_t^H\mu_v^H = \mathcal{W}_t^H\mu_{w,v}^H$. Collect and bind them row-wise for $t = 0, ..., T-1$ to obtain $NT \times (K+p+2)$ matrices $\mathbf{W}^H$, $\mathcal{W}^H$, and define $\hat{\mathcal{W}}$ analogously. For $\hat{\Sigma}_v \triangleq \frac{1}{NT}\hat{\mathcal{W}}^\intercal\hat{\mathcal{W}}$ and $\mathbf{r}_{w\epsilon} \triangleq \frac{1}{NT}\hat{\mathcal{W}}^\intercal\mathcal{E}$, we have

$$
\begin{aligned}
\hat{\mu}_v &= \left(\hat{\mathbf{W}}^\intercal\hat{\mathbf{W}}\right)^{-1}\hat{\mathbf{W}}^\intercal\left(\mathbf{W}^H\mu_v^H + \mathcal{E}\right) \\
&= \left(\frac{\mathbf{D}_v}{NT}\hat{\mathcal{W}}^\intercal\hat{\mathcal{W}}\mathbf{D}_v\right)^{-1}\frac{\mathbf{D}_v}{\sqrt{NT}}\hat{\mathcal{W}}^\intercal\left(\mathcal{W}^H\mu_{w,v}^H + \mathcal{E}\right) \\
&= \left(\hat{\Sigma}_v\mathbf{D}_v\right)^{-1}\frac{1}{\sqrt{NT}}\hat{\mathcal{W}}^\intercal\left\{\left(\mathcal{W}^H - \hat{\mathcal{W}}\right)\mu_{w,v}^H + \hat{\mathcal{W}}\mu_{w,v}^H + \mathcal{E}\right\} \\
&= {\mathbf{D}_v}^{-1}\hat{\Sigma}_v^{-1}\left\{\frac{1}{\sqrt{NT}}\hat{\mathcal{W}}^\intercal\left(\mathcal{W}^H - \hat{\mathcal{W}}\right)\mu_{w,v}^H + \sqrt{NT}\mathbf{r}_{w\epsilon}\right\} + \sqrt{NT}{\mathbf{D}_v}^{-1}\mu_{w,v}^H \\
&\therefore \mathbf{D}_v\left(\hat{\mu}_v - \mu_v^H\right) = \frac{\hat{\Sigma}_v^{-1}}{\sqrt{NT}}\hat{\mathcal{W}}^\intercal\left(\mathcal{W}^H - \hat{\mathcal{W}}\right)\mu_{w,v}^H + \sqrt{NT}\hat{\Sigma}_v^{-1}\mathbf{r}_{w\epsilon}.
\end{aligned}
$$

Therefore, we next show that the first term on the RHS is negligible and $\sqrt{NT}\hat{\Sigma}_v^{-1}\mathbf{r}_{w\epsilon}$ is converging to a multivariate normal distribution. We start by claiming that $\hat{\Sigma}_v$ is converging to a matrix with finite entries as N and T tend to infinity. This will allow us to focus on the behaviors of $\frac{1}{\sqrt{NT}}\hat{\mathcal{W}}^\intercal(\mathcal{W}^H - \hat{\mathcal{W}})\mu_{w,v}^H$ and $\sqrt{NT}\mathbf{r}_{w\epsilon}$, and then apply Slutsky's Theorem.

Claim C.5.1 *$\hat{\Sigma}_v$ converges to Σ_v in probability, i.e., $\hat{\Sigma}_v \Rightarrow \Sigma_v$.*

Proof As consequences of statements 1–9 from Lemma 25, we have

$$
\hat{\Sigma}_v = \frac{1}{NT}\sum_{t=0}^{T-1}\begin{bmatrix} N\mathbf{I}_K & \sqrt{N}\hat{\mathbf{V}}^\intercal\mathbf{y}_t & \sqrt{N}\hat{\mathbf{V}}^\intercal\mathcal{L}_{\mathbf{A}}\mathbf{y}_t & \sqrt{N}\hat{\mathbf{V}}^\intercal\mathbf{Z}_t \\ & \mathbf{y}_t{}^\intercal\mathbf{y}_t & \mathbf{y}_t{}^\intercal\mathcal{L}_{\mathbf{A}}\mathbf{y}_t & \mathbf{y}_t{}^\intercal\mathbf{Z}_t \\ & & \mathbf{y}_t{}^\intercal\mathcal{L}_{\mathbf{A}}^2\mathbf{y}_t & \mathbf{y}_t{}^\intercal\mathcal{L}_{\mathbf{A}}\mathbf{Z}_t \\ & & & \mathbf{Z}_t{}^\intercal\mathbf{Z}_t \end{bmatrix} \tag{8}
$$

$$
\Rightarrow \lim_{N,T\to\infty}\begin{bmatrix} \mathbf{I}_K & \frac{1}{\sqrt{N}}\tilde{\mathbf{V}}^\intercal\varphi_v & \frac{1}{\sqrt{N}}\tilde{\mathbf{V}}^\intercal\mathcal{L}_{\mathbf{A}}\varphi_v & \mathbf{O}_{K\times p} \\ & \frac{1}{N}\varphi_v{}^\intercal\varphi_v + \tau_2 & \frac{1}{N}\varphi_v{}^\intercal\mathcal{L}_{\mathbf{A}}\varphi_v + \tau_{23} & \mathbf{o}_p^\intercal \\ & & \frac{1}{N}\varphi_v{}^\intercal\mathcal{L}_{\mathbf{A}}^2\varphi_v + \tau_3 & \mathbf{o}_p^\intercal \\ & & & \Sigma_z \end{bmatrix} = \Sigma_v \tag{9}
$$

for $N\rho_N = \omega(\log^4 N)$. Existence of τ_2, τ_{23}, and τ_3 come by dominated convergence theorem after noting that $\operatorname{tr}(\mathcal{L}_{\mathbf{A}}\Gamma) \le \|\mathcal{L}_{\mathbf{A}}\|\operatorname{tr}(\Gamma) = O(N)$ *a.s.* and $\operatorname{tr}(\mathcal{L}_{\mathbf{A}}^2\Gamma) = O(N)$ *a.s.* Also noting that $\left\|\tilde{\mathbf{V}}\right\| = 1$ and the asymptotic order of each term found in the proof of Lemma 25, we have

$$
\begin{gathered}
\lim_{N\to\infty}\frac{1}{\sqrt{N}}\mathbf{h}_1{}^\intercal\tilde{\mathbf{V}}^\intercal\varphi_v = 0, \frac{1}{\sqrt{N}}\mathbf{h}_1{}^\intercal\tilde{\mathbf{V}}^\intercal\mathcal{L}_{\mathbf{A}}\varphi_v \Rightarrow 0, \lim_{N\to\infty}\frac{1}{N}\varphi_v{}^\intercal\varphi_v = 0, \\
\frac{1}{N}\varphi_v{}^\intercal\mathcal{L}_{\mathbf{A}}\varphi_v \Rightarrow 0, \frac{1}{N}\varphi_v{}^\intercal\mathcal{L}_{\mathbf{A}}^2\varphi_v \Rightarrow 0
\end{gathered}
$$

for all $\mathbf{h}_1 \in B^{K-1}$.

Claim C.5.2 $\frac{1}{\sqrt{NT}}\hat{\mathcal{W}}^\intercal(\mathcal{W}^H - \hat{\mathcal{W}})\mu_{w,v}^H \Rightarrow \mathbf{o}_{K+p+2}$.

Proof Since $\hat{\mathcal{W}}$ is different from $\mathcal{W}^H$ by $\sqrt{N}\hat{\mathbf{V}}$ only, we have

$$\frac{1}{\sqrt{NT}}\hat{\mathcal{W}}^\intercal\left(\mathcal{W}^H - \hat{\mathcal{W}}\right)\mu_{w,v}^H = \begin{bmatrix} \sqrt{T}\hat{\mathbf{V}}^\intercal \\ \frac{1}{\sqrt{NT}}\sum_{t=0}^{T-1}\mathbf{y}_t{}^\intercal \\ \frac{1}{\sqrt{NT}}\sum_{t=0}^{T-1}\mathbf{y}_t{}^\intercal\mathcal{L}_{\mathbf{A}} \\ \frac{1}{\sqrt{NT}}\sum_{t=0}^{T-1}\mathbf{Z}_t{}^\intercal \end{bmatrix}\left(\hat{\mathbf{V}} - \tilde{\mathbf{V}}\right)\mathbf{H}_v{}^\intercal\beta_v.$$

Then, by the statements 10–13 of Lemma 25, for any $\mathbf{h} \in B^{K+p+1}$, we have

$$\mathbb{P}\left[\frac{1}{\sqrt{NT}}\left|\mathbf{h}^\intercal\hat{\mathcal{W}}^\intercal\left(\mathcal{W}^H - \hat{\mathcal{W}}\right)\mu_{w,v}^H\right| > \upsilon\right] < C_1/N + C_2\exp\left(-cN^2\rho_N\upsilon^2\right)$$

for some constants $c, C_1, C_2 > 0$ and any υ such that $\upsilon \geq C'\sqrt{\frac{T}{N\rho_N}}$ for some $C' > 0$. Provided that $N\rho_N = \omega(T)$, we have the conclusion by Cramér–Wold.

Claim C.5.3 $\sqrt{NT}\mathbf{r}_{w\epsilon} \Rightarrow \mathcal{N}(\mathbf{o}_{K+p+2}, \sigma^2\Sigma_v)$.

Proof It is sufficient to show that for any $\eta \in \mathbb{R}^{K+p+2}$ such that $\|\eta\| \leq 1$, we have $\sqrt{NT}\eta^\intercal\mathbf{r}_{w\epsilon} \Rightarrow \mathcal{N}(0, \sigma^2\eta^\intercal\Sigma_v\eta)$. Denoting $\xi_{Nt} \triangleq (NT)^{-1/2}\eta^\intercal\hat{\mathcal{W}}_{t-1}^\intercal\mathcal{E}_t$ and

$$\mathcal{F}_{Nt} = \sigma(\mathbf{A}, \epsilon_{is}, \mathbf{Z}_{i,(s-1)}; i \leq N, -\infty < s \leq t),$$

$\{\sum_{s=1}^t \xi_{Ns}, \mathcal{F}_{Nt}\}$ constitutes a martingale array for each $N, t \leq T$. Then, we can apply Corollary 3.1 of Hall and Heyde (2014) to $\sqrt{NT}\eta^\intercal\mathbf{r}_{w\epsilon} = \sum_{t=0}^{T-1}\xi_{N,t+1}$ by checking following two conditions:

(1) $\sum_t \mathbb{E}\left(\xi_{N,t+1}^2 1_{\{|\xi_{N,t+1}|>\upsilon\}}|\mathcal{F}_{Nt}\right) = O_{\mathbb{P}}(1)$.

(2) $\sum_t \mathbb{E}\left(\xi_{N,t+1}^2|\mathcal{F}_{Nt}\right) = \eta^\intercal\sigma^2\Sigma_v\eta + O_{\mathbb{P}}(1)$.

Proof [of (1)] First, we have

$$\sum_{t=0}^{T-1}\mathbb{E}\left(\xi_{N,t+1}^2 1_{\{|\xi_{N,t+1}|>\upsilon\}}|\mathcal{F}_{Nt}\right) \leq \upsilon^{-2}\sum_t \mathbb{E}\left(\xi_{N,t+1}^4|\mathcal{F}_{Nt}\right).$$

One can easily verify that

$$\mathbb{E}\left(\xi_{N,t+1}^4|\mathcal{F}_{Nt}\right) \lesssim \sigma^4\left(\frac{1}{NT}\eta^\intercal\hat{\mathcal{W}}_t{}^\intercal\hat{\mathcal{W}}_t\eta\right)^2.$$

So, we only need to show that $\sum_t\left(\frac{1}{NT}\eta^\intercal\hat{\mathcal{W}}_t{}^\intercal\hat{\mathcal{W}}_t\eta\right)^2 = O_{\mathbb{P}}(1)$. First note that

$$(NT)^{-1}\hat{\mathcal{W}}_t{}^\intercal\hat{\mathcal{W}}_t = \begin{bmatrix} \frac{1}{T}\mathbf{I}_K & \frac{1}{\sqrt{NT}}\hat{\mathbf{V}}^\intercal\mathbf{y}_t & \frac{1}{\sqrt{NT}}\hat{\mathbf{V}}^\intercal\mathcal{L}_{\mathbf{A}}\mathbf{y}_t & \frac{1}{\sqrt{NT}}\hat{\mathbf{V}}^\intercal\mathbf{Z}_t \\ & \frac{1}{NT}\mathbf{y}_t{}^\intercal\mathbf{y}_t & \frac{1}{NT}\mathbf{y}_t{}^\intercal\mathcal{L}_{\mathbf{A}}\mathbf{y}_t & \frac{1}{NT}\mathbf{y}_t{}^\intercal\mathbf{Z}_t \\ & & \frac{1}{NT}\mathbf{y}_t{}^\intercal\mathcal{L}_{\mathbf{A}}^2\mathbf{y}_t & \frac{1}{NT}\mathbf{y}_t{}^\intercal\mathcal{L}_{\mathbf{A}}\mathbf{Z}_t \\ & & & \frac{1}{NT}\mathbf{Z}_t^\intercal\mathbf{Z}_t \end{bmatrix}.$$

Since similar arguments can be used to show the convergence of each entry, take

$$\sum_{t=0}^{T-1}\left(\frac{1}{\sqrt{NT}}\eta_1{}^\intercal\hat{\mathbf{V}}^\intercal\mathbf{y}_t\right)^2=\frac{1}{NT^2}\sum_t\eta_1{}^\intercal\hat{\mathbf{V}}^\intercal\mathbf{y}_t\mathbf{y}_t{}^\intercal\hat{\mathbf{V}}\eta_1$$

for example, where $\eta_1\in\mathbb{R}^K;\|\eta_1\|\leq 1$. Then, we have

$$\sum_{t=0}^{T-1}\eta_1{}^\intercal\hat{\mathbf{V}}^\intercal\mathbf{y}_t\mathbf{y}_t{}^\intercal\hat{\mathbf{V}}\eta_1=\mathbf{y}^\intercal\left\{\mathbf{I}_T\otimes\left(\hat{\mathbf{V}}\eta_1\eta_1{}^\intercal\hat{\mathbf{V}}^\intercal\right)\right\}\mathbf{y}.$$

Denoting $\hat{\mathbf{V}}-\tilde{\mathbf{V}}$ by Δ_v, we get

$$\tilde{\mathbf{y}}^\intercal{\Gamma_{\mathbf{y}}}^{1/2}\left\{\mathbf{I}_T\otimes\left(\Delta_v\eta_1\eta_1{}^\intercal\Delta_v^\intercal\right)\right\}{\Gamma_{\mathbf{y}}}^{1/2}\tilde{\mathbf{y}}=\tilde{\mathbf{y}}^\intercal\left\{\mathbf{I}_T\otimes\left(\Gamma^{1/2}\Delta_v\eta_1\eta_1{}^\intercal\Delta_v^\intercal\Gamma^{1/2}\right)\right\}\tilde{\mathbf{y}}.$$

Letting $\hat{\mathbf{S}}\triangleq\mathbf{I}_T\otimes\left(\Gamma^{1/2}\Delta_v\eta_1\eta_1{}^\intercal\Delta_v^\intercal\Gamma^{1/2}\right)$ and $\tilde{\mathbf{C}}\triangleq\mathbf{I}_T\otimes\left(\Gamma^{1/2}\tilde{\mathbf{V}}\eta_1\eta_1{}^\intercal\tilde{\mathbf{V}}^\intercal\Gamma^{1/2}\right)$, we have $\left\|\hat{\mathbf{S}}\right\|_F^2=T\operatorname{tr}\left\{\left(\Delta_v\eta_1\eta_1{}^\intercal\Delta_v^\intercal\Gamma\right)^2\right\}\leq T\left\|\Delta_v\eta_1\eta_1{}^\intercal\Delta_v^\intercal\right\|^2\operatorname{tr}\left(\Gamma^2\right)=O\left(NT\left(N\rho_N\right)^{-2}\right)$ *whp.* and $\left\|\hat{\mathbf{S}}\right\|=\left\|\Gamma^{1/2}\Delta_v\eta_1\eta_1{}^\intercal\Delta_v^\intercal\Gamma^{1/2}\right\|\leq\left\|\Delta_v\right\|^2\|\Gamma\|=O((N\rho_N)^{-1})$ *whp.* By Lemma 22,

$$\mathbb{P}\left[\frac{1}{\sqrt{NT}}\left|\tilde{\mathbf{y}}^\intercal\hat{\mathbf{S}}\tilde{\mathbf{y}}-\mathbb{E}\left(\tilde{\mathbf{y}}^\intercal\hat{\mathbf{S}}\tilde{\mathbf{y}}\right)\right|>\upsilon\right]<2\exp\left(-c\min\left\{\frac{\upsilon^2}{(N\rho_N)^{-2}},\frac{\upsilon\sqrt{NT}}{(N\rho_N)^{-1}}\right\}\right)+\frac{2}{N}.$$

Note that $\mathbb{E}\left(\tilde{\mathbf{y}}^\intercal\hat{\mathbf{S}}\tilde{\mathbf{y}}\right)=T\mathbb{E}\operatorname{tr}\left(\Delta_v\eta_1\eta_1{}^\intercal\Delta_v^\intercal\Gamma\right)$ and $T\operatorname{tr}\left(\Delta_v\eta_1\eta_1{}^\intercal\Delta_v^\intercal\Gamma\right)\leq T\left\|\Delta_v\eta_1\eta_1{}^\intercal\Delta_v^\intercal\right\|\operatorname{tr}\left(\Gamma\right)$ hence is $O\left(T/\rho_N\right)$ *whp.* Therefore, $\lim_{N,T\to\infty}\frac{\tilde{\mathbf{y}}^\intercal\hat{\mathbf{S}}\tilde{\mathbf{y}}}{T/\rho_N}<\infty$ *as.* hence $\tilde{\mathbf{y}}^\intercal\hat{\mathbf{S}}\tilde{\mathbf{y}}=O_{\mathbb{P}}(NT^2)$. So we have $\mathbb{E}\left(\tilde{\mathbf{y}}^\intercal\hat{\mathbf{S}}\tilde{\mathbf{y}}\right)=o\left(NT^2\right)$ by dominated convergence theorem. Next, observe that $\|\mathbf{G}\|\leq|\alpha_1|+|\theta|<1$ hence $\left\|(\mathbf{I}_N-\mathbf{G})^{-1}\right\|\leq(1-\|\mathbf{G}\|)^{-1}<1/(1-|\alpha_1|-|\theta|)$. So, $\|\varphi_v\|=O(\|\mathbf{V}\beta_v\|)=O(1)$ *as.* Using Lemma 20, we have

$$\mathbb{P}\left(\frac{1}{\sqrt{NT}}\left|2\tilde{\mathbf{y}}^\intercal\hat{\mathbf{S}}\Phi_v\right|>\upsilon\right)<2\exp\left(-\frac{cNT\upsilon^2}{T/\left(N\rho_N\right)^2}\right)+\frac{1}{N}$$
$$\mathbb{P}\left(\frac{1}{\sqrt{NT}}\left|2\tilde{\mathbf{y}}^\intercal\tilde{\mathbf{C}}\Phi_v\right|>\upsilon\right)<2\exp\left(-\frac{cNT\upsilon^2}{T}\right)+\frac{1}{N}$$

because $\left\|\hat{\mathbf{S}}\Phi_v\right\|\leq\left\|\hat{\mathbf{S}}\right\|\sqrt{T}\left\|\varphi_v\right\|=O(\sqrt{T}/\left(N\rho_N\right))$ *whp.* and $\left\|\tilde{\mathbf{C}}\Phi_v\right\|=O(\sqrt{T})$ *as.* Also, $\left|\Phi_v{}^\intercal\tilde{\mathbf{C}}\Phi_v\right|=O(T)$ *as.* Therefore, we have $\frac{1}{NT^2}\sum_t\left(\eta_1{}^\intercal\hat{\mathbf{V}}^\intercal\mathbf{y}_t\right)^2=O_{\mathbb{P}}(1)$. Noting that both $\mathbf{y}$ and $\mathbf{Z}_t$ have finite fourth-order moments, one can show that the rest are also $O_{\mathbb{P}}(1)$ similarly.

Proof [of (2)] Since $\sum_t\mathbb{E}\left(\xi_{N,t+1}^2|\mathcal{F}_{Nt}\right)=\frac{\sigma^2}{NT}\eta^\intercal\hat{\mathcal{W}}^\intercal\hat{\mathcal{W}}\eta=\sigma^2\eta^\intercal\hat{\Sigma}_v\eta$, by Lemma 26, we have (2).

Therefore, by the three claims made, we have

$$\mathbf{D}_v\left(\hat{\mu}_v-\mu_v^H\right)\Rightarrow{\Sigma_v}^{-1}\mathcal{N}\left(\mathbf{o}_{K+p+2},\sigma^2\Sigma_v\right)$$

which leads to the desired asymptotic normality. ■

Proof [of Theorem 2] It suffices to show that for any $\mathbf{Q}_K$ in Theorem 2, we have

$$\frac{\mathbf{Q}_K\hat{\Sigma}_v^{-1}}{\sqrt{NT}}\hat{\mathcal{W}}^\intercal(\mathcal{W}^H-\hat{\mathcal{W}})\mu_{w,v}^H=O_{\mathbb{P}}(1).$$

First provided that $\|\beta_v\|=1$ as $K\to\infty$, we have $\frac{1}{\sqrt{NT}}\hat{\mathcal{W}}^\intercal(\mathcal{W}^H-\hat{\mathcal{W}})\mu_{w,v}^H=O_{\mathbb{P}}(1)$ by the statements **10-13**. of Lemma 25. Since $\left\|\mathbf{Q}_K\hat{\Sigma}_v^{-1}\right\|=\left\|\mathbf{Q}_K\left(\hat{\Sigma}_v^{-1}-{\Sigma_v}^{-1}+{\Sigma_v}^{-1}\right)\right\|$, we only need to check if $\left\|\hat{\Sigma}_v^{-1}-{\Sigma_v}^{-1}\right\|=O_{\mathbb{P}}\left(\left\|{\Sigma_v}^{-1}\right\|\right)$. Since $\|\Sigma_v\|=O(1)$ and $\left\|\hat{\Sigma}_v-\Sigma_v\right\|=O_{\mathbb{P}}\left\{\min_i\lambda_i(\Sigma_v)^{-1}\right\}$ by Lemma 26, we have the first claim.

Next, It is sufficient to show that for any $\eta\in\mathbb{R}^m$ such that $\|\eta\|\leq 1$, we have $\eta^\intercal\mathbf{Q}_K\hat{\Sigma}_v^{-1}\sqrt{NT}\mathbf{r}_{w\epsilon}\Rightarrow\mathcal{N}(0,\sigma^2\eta^\intercal\Lambda_v\eta)$. Since

$$\begin{aligned}\eta^\intercal\mathbf{Q}_K\hat{\Sigma}_v^{-1}\sqrt{NT}\mathbf{r}_{w\epsilon}&=\eta^\intercal\mathbf{Q}_K\left(\hat{\Sigma}_v^{-1}-{\Sigma_v}^{-1}\right)\sqrt{NT}\mathbf{r}_{w\epsilon}+\eta^\intercal\mathbf{Q}_K{\Sigma_v}^{-1}\sqrt{NT}\mathbf{r}_{w\epsilon}\\&=O_{\mathbb{P}}(1)+\eta^\intercal\mathbf{Q}_K{\Sigma_v}^{-1}O_{\mathbb{P}}(1),\end{aligned}$$

we focus on the latter. Define $\xi_{Nt}\triangleq(NT)^{-1/2}\,\eta^\intercal\mathbf{Q}_K{\Sigma_v}^{-1}\hat{\mathcal{W}}_{t-1}^\intercal\mathcal{E}_t$ and

$$\mathcal{F}_{Nt}=\sigma(\mathbf{A},\epsilon_{is},\mathbf{Z}_{i,(s-1)};i\leq N,-\infty<s\leq t).$$

Then, a set of pairs $\{\sum_{s=1}^t\xi_{Ns},\mathcal{F}_{Nt}\}$ constitutes a martingale array for each $N,t\leq T$. Then, we repeat the following arguments (Hall and Heyde, 2014):

(1) $\sum_t\mathbb{E}\left(\xi_{N,t+1}^2 1_{\{|\xi_{N,t+1}|>v\}}|\mathcal{F}_{Nt}\right)=O_{\mathbb{P}}(1)$.

(2) $\sum_t\mathbb{E}\left(\xi_{N,t+1}^2|\mathcal{F}_{Nt}\right)=\sigma^2\eta^\intercal\Lambda_v\eta+O_{\mathbb{P}}(1)$.

Proof [of (1)] We only need to show that

$$\sum_t\left(\frac{1}{NT}\eta^\intercal\mathbf{Q}_K{\Sigma_v}^{-1}\hat{\mathcal{W}_t}^\intercal\hat{\mathcal{W}}_t{\Sigma_v}^{-1}{\mathbf{Q}_K}^\intercal\eta\right)^2=O_{\mathbb{P}}(1).$$

Since both $\mathbf{Q}_K$ and Σ_v have spectral norm of $O(1)$ as $K\to\infty$, this can be shown in the same manner as in the Proof of (1) of Claim 3 in Theorem 10.

Proof [of (2)] Since

$$\sum_t\mathbb{E}\left(\xi_{N,t+1}^2|\mathcal{F}_{Nt}\right)=\eta^\intercal\mathbf{Q}_K{\Sigma_v}^{-1}\frac{\sigma^2}{NT}\hat{\mathcal{W}}^\intercal\hat{\mathcal{W}}{\Sigma_v}^{-1}{\mathbf{Q}_K}^\intercal\eta=\sigma^2\eta^\intercal\mathbf{Q}_K{\Sigma_v}^{-1}\hat{\Sigma}_v{\Sigma_v}^{-1}{\mathbf{Q}_K}^\intercal\eta,$$

and noting that $\hat{\Sigma}_v=\Sigma_v+O_{\mathbb{P}}(1)$ for each $K>0$, we have (2) provided that the limit $\lim_{K\to\infty}\mathbf{Q}_K{\Sigma_v}^{-1}{\mathbf{Q}_K}^\intercal$ exists. ■

C.5.2 Consistency under misspecification of K

Proof [of Theorem 14] Now let us augment $\mathbf{D}_v$ as $\mathbf{D}_{+,v} \triangleq \sqrt{T}\,\mathrm{diag}\left(\mathbf{I}_{K_o}, \sqrt{N}\mathbf{I}_{p+2}\right)$ and augment $\hat{\mathcal{W}}$ in the proof of Theorem 10 as $\hat{\mathcal{W}}_+ \triangleq \left[\sqrt{N}\mathbf{1}_T \otimes \hat{\mathbf{V}}_{K:K_o} \,|\, \hat{\mathcal{W}}\right] \in \mathbb{R}^{NT\times(K_o+p+2)}$ so that $\sqrt{NT}\hat{\mathbf{W}}_+ = \hat{\mathcal{W}}_+\mathbf{D}_{+,v}$. For convenience of notation, let us denote $\hat{\mathbf{V}}_{K:K_o}$ by $\hat{\mathbf{V}}_+$. This representation gives

$$\hat{\mathcal{W}}_+\mu^H_{w+,v} = \hat{\mathcal{W}}\mu^H_{w,v}, \quad \sqrt{NT}{\mathbf{D}_{+,v}}^{-1}\mu^H_{w+,v} = \mu^H_{+,v}.$$

Define

$$\hat{\Sigma}_{+,v} \triangleq \frac{1}{NT}\hat{\mathcal{W}}_+^\intercal\hat{\mathcal{W}}_+ \tag{10}$$

and $\mathbf{r}_{w\epsilon,+} \triangleq \frac{1}{NT}\hat{\mathcal{W}}_+^\intercal\mathcal{E}$. Then, recalling the proof in the previous section,

$$\begin{aligned}
\hat{\mu}_{+,v} &= \left(\frac{\mathbf{D}_{+,v}}{NT}\hat{\mathcal{W}}_+^\intercal\hat{\mathcal{W}}_+\mathbf{D}_{+,v}\right)^{-1}\frac{\mathbf{D}_{+,v}}{\sqrt{NT}}\hat{\mathcal{W}}_+^\intercal\left(\mathcal{W}^H\mu^H_{w,v} + \mathcal{E}\right)\\
&= \left(\hat{\Sigma}_{+,v}\mathbf{D}_{+,v}\right)^{-1}\frac{1}{\sqrt{NT}}\hat{\mathcal{W}}_+^\intercal\left\{\left(\mathcal{W}^H - \hat{\mathcal{W}}\right)\mu^H_{w,v} + \hat{\mathcal{W}}_+\mu^H_{w+,v} + \mathcal{E}\right\}\\
&= {\mathbf{D}_{+,v}}^{-1}\hat{\Sigma}^{-1}_{+,v}\left\{\frac{1}{\sqrt{NT}}\hat{\mathcal{W}}_+^\intercal\left(\mathcal{W}^H - \hat{\mathcal{W}}\right)\mu^H_{w,v} + \sqrt{NT}\mathbf{r}_{w\epsilon,+}\right\} + \sqrt{NT}{\mathbf{D}_{+,v}}^{-1}\mu^H_{w+,v}\\
&\therefore \mathbf{D}_{+,v}\left(\hat{\mu}_{+,v} - \mu^H_{+,v}\right) = \frac{\hat{\Sigma}^{-1}_{+,v}}{\sqrt{NT}}\hat{\mathcal{W}}_+^\intercal\left(\mathcal{W}^H - \hat{\mathcal{W}}\right)\mu^H_{w,v} + \hat{\Sigma}^{-1}_{+,v}\sqrt{NT}\mathbf{r}_{w\epsilon,+}.
\end{aligned}$$

Therefore, we can use similar logic as the proof of Theorem 10.

Claim C.5.4 $\hat{\Sigma}_{+,v} \Rightarrow \Sigma_{+,v}$.

Proof Now observe that $\hat{\mathbf{V}}_+^\intercal\hat{\mathbf{V}}_+ = \mathbf{I}_{K_o-K}$ and $\left\|\hat{\mathbf{V}}_+\right\| = 1$ so by a similar result as item 0 of Lemma 25,

$$\hat{\Sigma}_{+,v} = \begin{bmatrix}\mathbf{I}_{K_o-K} & \frac{1}{\sqrt{NT}}\sum_t \hat{\mathbf{V}}_+^\intercal\hat{\mathcal{W}}_t\\ & \hat{\Sigma}_v\end{bmatrix} \overset{N,T\to\infty}{\Rightarrow} \begin{bmatrix}\mathbf{I}_{K_o-K} & \mathbf{O}_{(K_o-K)\times(K+p+2)}\\ & \Sigma_v\end{bmatrix}.$$

This is easy to check as $\hat{\mathbf{V}}_+^\intercal\hat{\mathbf{V}}_{:K} = \mathbf{O}_{(K_o-K)\times K}$ and for any $\mathbf{h} \in B^{K_o-K}$, we have $\|\Phi_v\| = O(1)$ *as.* and $\left\|\mathbf{1}_T \otimes \hat{\mathbf{V}}_+\right\| = O(\sqrt{T})$ *as.* so

$$\begin{aligned}
\frac{1}{\sqrt{NT}}\sum_t\left(\hat{\mathbf{V}}_+\mathbf{h}\right)^\intercal\mathbf{y}_t &= \frac{1}{\sqrt{NT}}\left(\mathbf{1}_T \otimes \hat{\mathbf{V}}_+\mathbf{h}\right)^\intercal{\Gamma_{\mathbf{y}}}^{1/2}\tilde{\mathbf{y}} + O\left(\frac{1}{\sqrt{NT}}\right)\ as.,\\
\mathbb{P}\left[\frac{1}{\sqrt{NT}}\left|\left(\mathbf{1}_T \otimes \hat{\mathbf{V}}_+\mathbf{h}\right)^\intercal{\Gamma_{\mathbf{y}}}^{1/2}\tilde{\mathbf{y}}\right| > \upsilon\right] &< 2\exp\left(-cNT\upsilon^2\right)
\end{aligned}$$

for example.

Claim C.5.5 $\frac{1}{\sqrt{NT}}\hat{\mathcal{W}}_+^\intercal(\mathcal{W}^H - \hat{\mathcal{W}})\mu^H_{w,v} = \mathbf{o}_{K_o+p+2} + O_{\mathbb{P}}(1)$.

Proof Now observe that

$$\frac{1}{\sqrt{NT}}\hat{\mathcal{W}}_+^\intercal\left(\mathcal{W}^H-\hat{\mathcal{W}}\right)\mu_{w,v}^H=-\frac{1}{\sqrt{NT}}\begin{bmatrix}\sqrt{N}\left(\mathbf{1}_T\otimes\hat{\mathbf{V}}_+\right)^\intercal\\ \hat{\mathcal{W}}^\intercal\end{bmatrix}\mathbf{1}_T\otimes\left(\hat{\mathbf{V}}_{:K}-\mathbf{V}\mathbf{H}_v\right)\mathbf{H}_v{}^\intercal\beta_v.$$

Then, our additional bias term of interest is

$$\sqrt{T}\left(\hat{\mathbf{V}}_+\mathbf{h}\right)^\intercal\left(\hat{\mathbf{V}}_{:K}-\mathbf{V}\mathbf{H}_v\right)\mathbf{H}_v{}^\intercal\beta_v=O\left(\sqrt{\frac{T}{N\rho_N}}\right)\ whp.$$

for any $\mathbf{h}\in B^{K_o-K}$, which leads to the conclusion by Lemma 25 provided that $N\rho_N=\omega(T)$.

Claim C.5.6 $\forall\eta\in B^{K_o+p+1}$, $\sqrt{NT}\eta^\intercal\mathbf{r}_{w\epsilon,+}=O_{\mathbb{P}}(1)$.

Proof Using the same steps in the third claim in the proof of Theorem 10, we can derive that $\sqrt{NT}\mathbf{r}_{w\epsilon,+}\Rightarrow\mathcal{N}(\mathbf{o}_{K_o+p+2},\sigma^2\Sigma_{+,v})$.

Therefore, we have the conclusion. ■

Proof [of Theorem 8] Here, we have $K_o<K$. Define a diagonal matrix of convergence rates

$$\mathbf{D}_{-,v}\triangleq\sqrt{T}\,\mathrm{diag}\left(\mathbf{I}_{K_o},\sqrt{N}\mathbf{I}_{p+2}\right)$$

and a truncated design matrix with $\hat{\mathbf{V}}_{:K_o}$ embedded over the entire time,

$$\hat{\mathcal{W}}_-\triangleq\left[\sqrt{N}\mathbf{1}_T\otimes\hat{\mathbf{V}}_{:K_o}\,|\,\mathbf{y},(\mathbf{I}_T\otimes\mathcal{L}_{\mathbf{A}})\,\mathbf{y}\,|\,\mathbf{Z}\right]\in\mathbb{R}^{NT\times(K_o+p+2)}.$$

Let $\hat{\mathcal{W}}_{:K}\in\mathbb{R}^{(K+p+2)\times NT}$ denote the augmented version of $\hat{\mathcal{W}}_-$ which contains zeros between first and second block of $\hat{\mathcal{W}}_-$ above:

$$\hat{\mathcal{W}}_{:K}\triangleq\left[\sqrt{N}\mathbf{1}_T\otimes\hat{\mathbf{V}}_{:K_o}\,|\mathbf{O}_{NT\times(K-K_o)}\,|\,\mathbf{y},(\mathbf{I}_T\otimes\mathcal{L}_{\mathbf{A}})\,\mathbf{y}\,|\,\mathbf{Z}\right]\in\mathbb{R}^{NT\times(K+p+2)}.$$

Partition $\mathbf{H}_v=[\mathbf{H}_{:K_o,v}\,|\,\mathbf{H}_{K_o:K,v}]$ and $\mu_v^{H_-}\triangleq\left(\beta_v^\intercal\mathbf{H}_{:K_o,v},\mu_{-\beta_v}^\intercal\right)^\intercal\in\mathbb{R}^{K_o+p+2}$ and $\mu_{w,v}^{H_-}\triangleq\left(\frac{1}{\sqrt{N}}\beta_v^\intercal\mathbf{H}_{:K_o,v},\mu_{-\beta_v}^\intercal\right)^\intercal$. This expression gives $\sqrt{NT}\hat{\mathbf{W}}_-=\hat{\mathcal{W}}_-\mathbf{D}_{-,v}$ and $\hat{\mathcal{W}}_{:K}\mu_{w,v}^H=\hat{\mathcal{W}}_-\mu_{w,v}^{H_-}$. $\hat{\Sigma}_{-,v}$ and $\mathbf{r}_{w\epsilon,-}$ are defined analogously as the case of over-selection. Then, we can derive the following in the same way as the previous Theorems:

$$\mathbf{D}_{-,v}\left(\hat{\mu}_{-,v}-\mu_v^{H_-}\right)=\frac{\hat{\Sigma}_{-,v}^{-1}}{\sqrt{NT}}\hat{\mathcal{W}}_-^\intercal\left(\mathcal{W}^H-\hat{\mathcal{W}}_{:K}\right)\mu_{w,v}^H+\sqrt{NT}\hat{\Sigma}_{-,v}^{-1}\mathbf{r}_{w\epsilon,-}.$$

Let $\Sigma_{-,v}$ be a version of Σ_v which misses $K-K_o$ rows and columns corresponding to $K_o+1,\ldots,K$-th latent variables.

Next, by Proposition 23 there exists $\mathbf{H}_v\in\mathcal{O}_K$ such that

$$\left\|\hat{\mathbf{V}}_{:K}-\mathbf{V}\mathbf{H}_v\right\|=O\left(1/\sqrt{N\rho_N}\right)\ whp.$$

Therefore, there exists $\mathbf{H}_v'\in\mathcal{O}_{K,K_o}$ such that

$$\left\|\hat{\mathbf{V}}_{:K_o}-\mathbf{V}\mathbf{H}_{:K_o,v}\right\|\le\left\|\hat{\mathbf{V}}_{:K_o}-\mathbf{V}\mathbf{H}_v'\right\|+\left\|\mathbf{I}_{K_o}-\mathbf{H}_v'{}^\intercal\mathbf{H}_{:K_o,v}\right\|=O(1)\ whp.$$

by Theorem 7.

Claim C.5.7 $\hat{\Sigma}_{-,v} \Rightarrow \Sigma_{-,v}$.

Proof Now we only need to show the following:

1. $\mathbb{P}\left(\frac{1}{\sqrt{NT}}\left|\mathbf{h}_1{}^\intercal \sum_t \left\{\hat{\mathbf{V}}_{:K_o}^\intercal \mathbf{y}_t - (\mathbf{V}\mathbf{H}_{:K_o,v})^\intercal \varphi_v\right\}\right| > \upsilon\right) \lesssim e^{-d_2\upsilon^2} + 1/N$
2. $\mathbb{P}\left[\frac{1}{\sqrt{NT}}\left|\mathbf{h}_1{}^\intercal \sum_t \left\{\hat{\mathbf{V}}_{:K_o}^\intercal \mathcal{L}_{\mathbf{A}}\mathbf{y}_t - (\mathbf{V}\mathbf{H}_{:K_o,v})^\intercal \mathcal{L}_{\mathbf{A}}\varphi_v\right\}\right| > \upsilon\right] \lesssim e^{-d_2\upsilon^2} + 1/N$
3. $\mathbb{P}\left(\frac{1}{\sqrt{NT}}\left|\sum_t \mathbf{h}_1{}^\intercal \hat{\mathbf{V}}_{:K_o}^\intercal \mathbf{Z}_t\mathbf{h}_2\right| > \upsilon\right) \lesssim e^{-d_2\upsilon^2} + 1/N$

where $d_2 \triangleq c_2 NT$ for fixed $c_2 > 0$. These can be easily shown by a similar proof to 1–3 of Lemma 25. Also, we have

$$\lim_{N,T\to\infty} \frac{1}{\sqrt{N}} (\mathbf{V}\mathbf{H}_{:K_o,v})^\intercal \varphi_v = \mathbf{o}_{K_o}, \quad \lim_{N,T\to\infty} \frac{1}{\sqrt{N}} (\mathbf{V}\mathbf{H}_{:K_o,v})^\intercal \mathcal{L}_{\mathbf{A}}\varphi_v = \mathbf{o}_{K_o}.$$

If $K_o = 0$, then we have the reduced asymptotic covariance as Σ_{-u}.

Next, we analyze the entry-wise asymptotic orders of $\frac{1}{\sqrt{NT}}\hat{\mathcal{W}}_-^\intercal(\mathcal{W}^H - \hat{\mathcal{W}}_{:K})\mu_{w,v}^H$. First note that $\mathbf{H}_{:K_o,v}\mathbf{H}_{:K_o,v}{}^\intercal + \mathbf{H}_{K_o:K,v}\mathbf{H}_{K_o:K,v}{}^\intercal = \mathbf{I}_K$ as $\mathbf{H}_v \in \mathcal{O}_K$ and

$$\mathcal{W}^H - \hat{\mathcal{W}}_{:K} = \sqrt{N}\mathbf{1}_T \otimes \left[\mathbf{V}\mathbf{H}_{:K_o,v} - \hat{\mathbf{V}}_{:K_o} \,|\, \mathbf{V}\mathbf{H}_{K_o:K,v} \,|\, \mathbf{O}_{N\times(p+2)}\right]$$

Then, $\frac{1}{\sqrt{NT}}\hat{\mathcal{W}}_-^\intercal\left(\mathcal{W}^H - \hat{\mathcal{W}}_{:K}\right)\mu_{w,v}^H$ equals

$$\frac{1}{\sqrt{NT}}\sum_t \hat{\mathcal{W}}_{t,-}^\intercal \left\{\left(\mathbf{V}\mathbf{H}_{:K_o,v} - \hat{\mathbf{V}}_{:K_o}\right)\mathbf{H}_{:K_o,v}{}^\intercal + \mathbf{V}\mathbf{H}_{K_o:K,v}\mathbf{H}_{K_o:K,v}{}^\intercal\right\}\beta_v$$
$$= \begin{bmatrix} \sqrt{T}\hat{\mathbf{V}}_{:K_o}^\intercal \\ \frac{1}{\sqrt{NT}}\sum_t \mathbf{y}_t{}^\intercal \\ \frac{1}{\sqrt{NT}}\sum_t \mathbf{y}_t{}^\intercal\mathcal{L}_{\mathbf{A}} \\ \frac{1}{\sqrt{NT}}\sum_t \mathbf{Z}_t{}^\intercal \end{bmatrix}\left(\mathbf{V} - \hat{\mathbf{V}}_{:K_o}\mathbf{H}_{:K_o,v}{}^\intercal\right)\beta_v.$$

Denoting the RHS by $\mathbf{q}(K_o)$ which depends on K_o, we have $\mathbf{q}(0) = \mathbf{V}\beta_v$ for $K_o = 0$ as $\hat{\mathbf{V}}_{:K_o}\mathbf{H}_{:K_o,v}{}^\intercal$ is removed. So, $\|\mathbf{q}(0)\| = O(1)$. Otherwise, $\|\mathbf{q}(K_o)\| = O(1)$ *whp.* by noting that

$$\left\|\mathbf{V} - \hat{\mathbf{V}}_{:K_o}\mathbf{H}_{:K_o,v}{}^\intercal\right\| = \left\|\hat{\mathbf{V}}_{:K_o} - \mathbf{V}\mathbf{H}_{:K_o,v}\right\|.$$

Then, we can easily show the following by similar arguments used in proving 10–13 of Lemma 25:

1. $\sqrt{T}\left|\left(\hat{\mathbf{V}}_{:K_o}\mathbf{h}_1\right)^\intercal \mathbf{q}(K_o)\right| = O\left(\sqrt{T}\right)$ *whp.*
2. $\mathbb{P}\left[\frac{1}{\sqrt{NT}}\left|\sum_t (\mathbf{y}_t - \varphi_v)^\intercal \mathbf{q}(K_o)\right| > \upsilon\right] < e^{-cN\upsilon^2}$
3. $\mathbb{P}\left[\frac{1}{\sqrt{NT}}\left|\sum_t (\mathbf{y}_t - \varphi_v)^\intercal \mathcal{L}_{\mathbf{A}}\mathbf{q}(K_o)\right| > \upsilon\right] < e^{-cN\upsilon^2}$
4. $\mathbb{P}\left[\frac{1}{\sqrt{NT}}\left|(\sum_t \mathbf{Z}_t\mathbf{h}_2)^\intercal \mathbf{q}(K_o)\right| > \upsilon\right] < e^{-cN\upsilon^2}$

for any $\upsilon > 0$ where $\mathbf{h}_1 \in B^{K_o-1}, \mathbf{h}_2 \in B^{p-1}$. Now partition $\frac{1}{\sqrt{NT}}\hat{\mathcal{W}}_-^\intercal \left(\mathcal{W}^H - \hat{\mathcal{W}}_{:K}\right)\mu_{w,v}^H$ into three sub-vectors of sizes K_o, 2, and p denoted by $\mathbf{b}_1, \mathbf{b}_2$, and $\mathbf{b}_3$ respectively. These imply the following:

$$b_{21} = \sqrt{\frac{T}{N}}\varphi_v{}^\intercal \mathbf{q}(K_o) + O_{\mathbb{P}}(1), \quad b_{22} = \sqrt{\frac{T}{N}}\varphi_v{}^\intercal \mathcal{L}_{\mathbf{A}}\mathbf{q}(K_o) + O_{\mathbb{P}}(1),$$
$$\mathbf{h}_2{}^\intercal \mathbf{b}_3 = O_{\mathbb{P}}(1)$$

These give the specified asymptotic orders of the estimation error terms.

Now, using the same steps in the third claim in the proof of Theorem 10, we can derive that $\sqrt{NT}\mathbf{r}_{w\epsilon,-} \Rightarrow \mathcal{N}(\mathbf{o}_{K_o+p+2}, \sigma^2\Sigma_{-,v})$.

Therefore, recalling the rates of $\mathbf{D}_{-,v}$, we can summarize as follows:

$$\sqrt{T}\left(\hat{\beta}_{-,v} - \mathbf{H}_{:K_o,v}^\intercal \beta_v\right) = \mathbf{b}_1 + \mathbf{t}_1, \quad \sqrt{NT}\begin{bmatrix}\hat{\alpha}_{-,v} - \alpha_v \\ \hat{\theta}_{-,v} - \theta_v\end{bmatrix} = \mathbf{b}_2 + \mathbf{t}_2,$$
$$\sqrt{NT}\left(\hat{\gamma}_{-,v} - \gamma_v\right) = \mathbf{t}_3$$

where

$$\mathbf{h}_1{}^\intercal \mathbf{b}_1 = O\left(\sqrt{\frac{T}{N\rho_N}}\right) \ whp., \quad \mathbf{h}_2{}^\intercal \mathbf{b}_2 = O\left(\sqrt{\frac{T}{N}}\right) \ whp.,$$
$$\mathbf{h}_1{}^\intercal \mathbf{t}_1 = O_{\mathbb{P}}(1), \mathbf{h}_2{}^\intercal \mathbf{t}_2 = O_{\mathbb{P}}(1), \mathbf{h}_3{}^\intercal \mathbf{t}_3 = O_{\mathbb{P}}(1)$$

for any $\mathbf{h}_1 \in B^{K_o-1}, \mathbf{h}_2 \in B^1, \mathbf{h}_3 \in B^{p-1}$. ■

C.5.3 Concentration inequalities

In the following lemma, we state each item for $N > N_0$ with a large enough $N_0 > 0$.

Lemma 25 *Let $\mathbf{h}_i \in B^{j_i-1}$ for $j_1 = K$, $j_2 = p$. Then under the conditions of Theorem 10, there exist $N_0, T_0 > 0$ and $\upsilon_0(N_0, T_0) > 0$ such that the statements **1.**–**13.** hold for all $\upsilon \in (0, \upsilon_0)$ and $N > N_0, T > T_0$:*

1. $\mathbb{P}\left(\frac{1}{\sqrt{NT}}\left|\mathbf{h}_1{}^\intercal \sum_t (\hat{\mathbf{V}}^\intercal \mathbf{y}_t - \tilde{\mathbf{V}}^\intercal \varphi_v)\right| > \upsilon\right) \lesssim e^{-d_1\upsilon^2} + e^{-d_2\upsilon^2} + 1/N$
2. $\mathbb{P}\left[\frac{1}{\sqrt{NT}}\left|\mathbf{h}_1{}^\intercal \sum_t \left(\hat{\mathbf{V}}^\intercal \mathcal{L}_{\mathbf{A}}\mathbf{y}_t - \tilde{\mathbf{V}}^\intercal \mathcal{L}_{\mathbf{A}}\varphi_v\right)\right| > \upsilon\right] \lesssim e^{-d_1\upsilon^2} + e^{-d_2\upsilon^2} + 1/N$
3. $\mathbb{P}\left(\frac{1}{\sqrt{NT}}\left|\sum_t \mathbf{h}_1{}^\intercal \hat{\mathbf{V}}^\intercal \mathbf{Z}_t \mathbf{h}_2\right| > \upsilon\right) \lesssim e^{-d_1\upsilon^2} + e^{-d_2\upsilon^2} + 1/N$
4. $\mathbb{P}\left[\frac{1}{NT}\left|\sum_t \{\mathbf{y}_t{}^\intercal \mathbf{y}_t - \mathbb{E}\,\mathrm{tr}(\Gamma) - \varphi_v{}^\intercal \varphi_v\}\right| > \upsilon\right] \lesssim e^{-d_2\upsilon^2} + e^{-d_3\upsilon^2}$
5. $\mathbb{P}\left[\frac{1}{NT}\left|\sum_t \{\mathbf{y}_t{}^\intercal \mathcal{L}_{\mathbf{A}}\mathbf{y}_t - \mathbb{E}\,\mathrm{tr}(\mathcal{L}_{\mathbf{A}}\Gamma) - \varphi_v^\intercal \mathcal{L}_{\mathbf{A}}\varphi_v\}\right| > \upsilon\right] e^{-d_2\upsilon^2} + e^{-d_3\upsilon^2}$
6. $\mathbb{P}\left(\frac{1}{NT}\left|\sum_t \mathbf{y}_t{}^\intercal \mathbf{Z}_t \mathbf{h}_2\right| > \upsilon\right) \lesssim e^{-d_4\upsilon^2} + e^{-d_3\upsilon^2} + 1/N$

7. $\mathbb{P}\left[\frac{1}{NT}\left|\sum_t \left\{\mathbf{y}_t{}^\mathsf{T}\mathcal{L}_\mathbf{A}^2\mathbf{y}_t - \mathbb{E}\,\mathrm{tr}(\mathcal{L}_\mathbf{A}^2\Gamma) - \varphi_v{}^\mathsf{T}\mathcal{L}_\mathbf{A}^2\varphi_v\right\}\right| > \upsilon\right] \lesssim e^{-d_2\upsilon^2} + e^{-d_3\upsilon^2}$

8. $\mathbb{P}\left(\frac{1}{NT}\left|\sum_t \mathbf{y}_t{}^\mathsf{T}\mathcal{L}_\mathbf{A}\mathbf{Z}_t\mathbf{h}_2\right| > \upsilon\right) \lesssim e^{-d_4\upsilon^2} + 2^{-d_3\upsilon^2} + 1/N$

9. $\mathbb{P}\left[\frac{1}{NT}\left|\mathbf{h}_2{}^\mathsf{T}\sum_t \left(\mathbf{Z}_t{}^\mathsf{T}\mathbf{Z}_t - \Sigma_z\right)\mathbf{h}_2\right| > \upsilon\right] \lesssim e^{-d_2\upsilon^2}$

10. $\mathbb{P}\left[\sqrt{T}\left|\mathbf{h}_1{}^\mathsf{T}\hat{\mathbf{V}}^\mathsf{T}\left(\hat{\mathbf{V}} - \tilde{\mathbf{V}}\right)\mathbf{H}_v^\mathsf{T}\beta_v\right| > \upsilon\right] \lesssim \frac{1}{N}$

11. $\mathbb{P}\left[\frac{1}{\sqrt{NT}}\left|\sum_t \mathbf{y}_t{}^\mathsf{T}\left(\hat{\mathbf{V}} - \tilde{\mathbf{V}}\right)\mathbf{H}_v^\mathsf{T}\beta_v\right| > \upsilon\right] \lesssim e^{-cN^2\rho_N\upsilon^2} + \frac{1}{N}$

12. $\mathbb{P}\left[\frac{1}{\sqrt{NT}}\left|\sum_t \mathbf{y}_t{}^\mathsf{T}\mathcal{L}_\mathbf{A}\left(\hat{\mathbf{V}} - \tilde{\mathbf{V}}\right)\mathbf{H}_v^\mathsf{T}\beta_v\right| > \upsilon\right] \lesssim e^{-cN^2\rho_N\upsilon^2} + \frac{1}{N}$

13. $\mathbb{P}\left[\frac{1}{\sqrt{NT}}\left|\mathbf{h}_2{}^\mathsf{T}\sum_t \mathbf{Z}_t{}^\mathsf{T}\left(\hat{\mathbf{V}} - \tilde{\mathbf{V}}\right)\mathbf{H}_v^\mathsf{T}\beta_v\right| > \upsilon\right] \lesssim e^{-cN^2\rho_N\upsilon^2} + \frac{1}{N}$

where $d_1 \triangleq c_1N^2\rho_NT$, $d_2 \triangleq c_2NT$, $d_3 \triangleq c_3N^2T$, *and* $d_4 \triangleq \frac{c_4NT}{p}$ *and* $c, c_1, \ldots, c_4, b > 0$ *are constants.*

Proof [of 1] Note that

$$\sum_{t=1}^{T}\mathbf{h}_1{}^\mathsf{T}\left(\hat{\mathbf{V}}^\mathsf{T}\mathbf{y}_t - \tilde{\mathbf{V}}^\mathsf{T}\varphi_v\right) = \mathbf{h}_1{}^\mathsf{T}\left(\mathbf{1}_T\otimes\Delta_v\right)^\mathsf{T}{\Gamma_\mathbf{y}}^{1/2}\tilde{\mathbf{y}} + \mathbf{h}_1{}^\mathsf{T}\left(\mathbf{1}_T\otimes\tilde{\mathbf{V}}\right)^\mathsf{T}{\Gamma_\mathbf{y}}^{1/2}\tilde{\mathbf{y}}$$
$$+T\left(\Delta_v\mathbf{h}_1\right)^\mathsf{T}\varphi_v = \left(\mathbf{C}\mathbf{h}_1\right)^\mathsf{T}\tilde{\mathbf{y}} + \left(\tilde{\mathbf{C}}\mathbf{h}_1\right)^\mathsf{T}\tilde{\mathbf{y}} + T\left(\Delta_v\mathbf{h}_1\right)^\mathsf{T}\varphi_v.$$

where $\mathbf{C} \triangleq {\Gamma_\mathbf{y}}^{1/2}\left(\mathbf{1}_T\otimes\Delta_v\right)$ and $\tilde{\mathbf{C}} \triangleq {\Gamma_\mathbf{y}}^{1/2}\left(\mathbf{1}_T\otimes\tilde{\mathbf{V}}\right)$. Since $\|\mathbf{C}\mathbf{h}_1\| = O\left(\sqrt{\frac{T}{N\rho_N}}\right)$ *whp.*, by Lemma 20 we have

$$\mathbb{P}\left(\left|\left(\mathbf{C}\mathbf{h}_1\right)^\mathsf{T}\tilde{\mathbf{y}}\right| > \sqrt{N}T\upsilon\right) < 2\exp\left(\frac{-cN^2\rho_NT^2\upsilon^2}{T}\right) + \frac{1}{N}.$$

Likewise, since $\left\|\tilde{\mathbf{C}}\right\| = O(\sqrt{T})$ *as.* , by Lemma 20 we have $\mathbb{P}\left(\left|\left(\tilde{\mathbf{C}}\mathbf{h}_1\right)^\mathsf{T}\tilde{\mathbf{y}}\right| > \sqrt{N}T\upsilon\right) < 2\exp(-c_2NT^2\upsilon^2/T)$. Finally, $\left|T\left(\Delta_v\mathbf{h}_1\right)^\mathsf{T}\varphi_v\right| = O(T/\sqrt{N\rho_N})$ *whp.*, i.e.,

$$\mathbb{P}\left(\left|\left(\Delta_v\mathbf{h}_1\right)^\mathsf{T}\varphi_v\right| > \frac{C}{\sqrt{N\rho_N}}\right) < 1/N$$

for all large enough N. By selecting υ such that $\sqrt{N}\upsilon = \omega(1/\sqrt{N\rho_N})$, i.e., $\upsilon = \omega(1/\sqrt{N^2\rho_N})$, and noting that $N\rho_N = \omega(\log^4 N)$, we can reduce υ sufficiently small for large enough N. For any such υ, we have

$$\mathbb{P}\left(\frac{1}{\sqrt{N}}\left|\left(\Delta_v\mathbf{h}_1\right)^\mathsf{T}\varphi_v\right| > \upsilon\right) \le \mathbb{P}\left(\left|\left(\Delta_v\mathbf{h}_1\right)^\mathsf{T}\varphi_v\right| > \frac{C}{\sqrt{N\rho_N}}\right) < \frac{1}{N}.$$

This completes the proof. ■

Proof [of 2] First note that

$$\mathbf{h}_1{}^\intercal \sum_t \left(\hat{\mathbf{V}}^\intercal \mathcal{L}_\mathbf{A} \mathbf{y}_t - \tilde{\mathbf{V}}^\intercal \mathcal{L}_\mathbf{A} \varphi_v\right) = \mathbf{h}_1{}^\intercal \left[\mathbf{1}_T \otimes \{\mathcal{L}_\mathbf{A} \Delta_v\}\right]^\intercal \Gamma_\mathbf{y}{}^{1/2} \tilde{\mathbf{y}}$$
$$+\mathbf{h}_1{}^\intercal \left\{\mathbf{1}_T \otimes \mathcal{L}_\mathbf{A} \tilde{\mathbf{V}}\right\}^\intercal \Gamma_\mathbf{y}{}^{1/2} \tilde{\mathbf{y}} + T \left(\Delta_v \mathbf{h}_1\right)^\intercal \mathcal{L}_\mathbf{A} \varphi_v.$$

For $\mathbf{C} \triangleq \Gamma_\mathbf{y}{}^{1/2} \{\mathbf{1}_T \otimes (\mathcal{L}_\mathbf{A} \Delta_v)\}$ and $\tilde{\mathbf{C}} \triangleq \Gamma_\mathbf{y}{}^{1/2} \left\{\mathbf{1}_T \otimes \left(\mathcal{L}_\mathbf{A} \tilde{\mathbf{V}}\right)\right\}$, $\|\mathbf{C}\mathbf{h}_1\| = O\left(\sqrt{\frac{T}{N\rho_N}}\right)$ *whp.* because $\rho\left(\mathcal{L}_\mathbf{A}\right) < 1$ *as.* hence $\mathbb{P}\left(|(\mathbf{C}\mathbf{h}_1)^\intercal \tilde{\mathbf{y}}| > \sqrt{N} T \upsilon\right) < 2\exp\left(-cN^2 \rho_N T \upsilon^2\right) + \frac{1}{N}$. Since $\left\|\tilde{\mathbf{C}}\mathbf{h}_1\right\| = O\left(\sqrt{T}\right)$ *as.* , we have $\mathbb{P}\left(|\left(\tilde{\mathbf{C}}\mathbf{h}_1\right)^\intercal \tilde{\mathbf{y}}| > \sqrt{N} T \upsilon\right) < 2\exp\left(-cNT\upsilon^2\right)$. The tail probability of $|T\left(\Delta_v \mathbf{h}_1\right)^\intercal \mathcal{L}_\mathbf{A} \varphi_v|$ can be bounded similarly to 1. ■

Proof [of 3] Let $\mathbf{Z} \triangleq [\mathbf{Z}_0{}^\intercal, ..., \mathbf{Z}_{T-1}{}^\intercal]^\intercal$ and $\mathbf{C} \triangleq \mathbf{1}_T \otimes \Delta_v$ and $\tilde{\mathbf{C}} \triangleq \mathbf{1}_T \otimes \tilde{\mathbf{V}}$. Then, we can express $\sum_t \mathbf{h}_1{}^\intercal \hat{\mathbf{V}}^\intercal \mathbf{Z}_t \mathbf{h}_2$ as $\operatorname{tr}\left(\mathbf{Z}\mathbf{h}_2 \mathbf{h}_1{}^\intercal \mathbf{C}^\intercal\right) + \operatorname{tr}\left(\mathbf{Z}\mathbf{h}_2 \mathbf{h}_1{}^\intercal \tilde{\mathbf{C}}^\intercal\right)$. This is equal to

$$\operatorname{vec}\left(\mathbf{h}_2 \mathbf{h}_1{}^\intercal\right)^\intercal \left\{\mathbf{I}_p \otimes \left(\mathbf{C} + \tilde{\mathbf{C}}\right)^\intercal\right\} \operatorname{vec}\left(\mathbf{Z}\right).$$

Note that $\|\mathbf{I}_p \otimes \mathbf{C}^\intercal\| = O\left(\sqrt{\frac{T}{N\rho_N}}\right)$ *whp.* and $\left\|\mathbf{I}_p \otimes \tilde{\mathbf{C}}^\intercal\right\| = O\left(\sqrt{T}\right)$. Therefore, by Lemma 20, we have

$$\mathbb{P}\left[\left|\sum_t \mathbf{h}_1{}^\intercal \hat{\mathbf{V}}^\intercal \mathbf{Z}_t \mathbf{h}_2\right| > \sqrt{N} T \upsilon\right] \leq 2\exp\left(-\frac{cNT\upsilon^2}{\left(N\rho_N\right)^{-1}}\right) + 2\exp\left(-c'NT\upsilon^2\right) + \frac{1}{N}.$$

■

Proof [of 4] First, we have

$$\sum_t \left(\mathbf{y}_t{}^\intercal \mathbf{y}_t - \varphi_v^\intercal \varphi_v\right) = \left\|\Gamma_\mathbf{y}{}^{1/2} \tilde{\mathbf{y}} + \Phi_v\right\|^2 - T\varphi_v^\intercal \varphi_v = \tilde{\mathbf{y}}^\intercal \Gamma_\mathbf{y} \tilde{\mathbf{y}} + 2\Phi_v^\intercal \Gamma_\mathbf{y}{}^{1/2} \tilde{\mathbf{y}}.$$

Note that $\|\Gamma_\mathbf{y}\|_F^2 = \sum_{i=1}^{NT} \sigma_i(\Gamma_\mathbf{y})^2 = O(NT)$ *as.* By Lemma 22, we have

$$\mathbb{P}\left(|\tilde{\mathbf{y}}^\intercal \Gamma_\mathbf{y} \tilde{\mathbf{y}} - T \operatorname{tr} \mathbb{E}\Gamma| > \upsilon\right) < 2\exp\left(-c\min\left\{\frac{\upsilon^2}{NT}, \upsilon\right\}\right).$$

By Lemma 20,

$$\mathbb{P}\left(2|\Phi_v^\intercal \Gamma_\mathbf{y}{}^{1/2} \tilde{\mathbf{y}}| > \upsilon\right) < 2\exp\left(-\frac{c'\upsilon^2}{T}\right)$$

since $\left\|\Gamma_\mathbf{y}{}^{1/2} \Phi_v\right\|^2 \leq \|\Gamma_\mathbf{y}\| T\varphi_v^\intercal \varphi_v = O(T)$ *as.* ■

Proof [of 6] We have

$$\sum_t \mathbf{y}_t{}^\intercal \mathbf{Z}_t \mathbf{h}_2 = \left(\Gamma_\mathbf{y}{}^{1/2} \tilde{\mathbf{y}} + \Phi_v\right)^\intercal \mathbf{Z}\mathbf{h}_2 = \operatorname{tr}\left(\mathbf{h}_2 \tilde{\mathbf{y}}^\intercal \Gamma_\mathbf{y}{}^{1/2} \mathbf{Z}\right) + \operatorname{tr}\left(\mathbf{h}_2 \Phi_v{}^\intercal \mathbf{Z}\right)$$
$$= \operatorname{vec}\left(\tilde{\mathbf{y}}\mathbf{h}_2{}^\intercal\right)^\intercal \left(\mathbf{I}_p \otimes \Gamma_\mathbf{y}{}^{1/2}\right) \operatorname{vec}\left(\mathbf{Z}\right) + \operatorname{vec}\left(\Phi_v \mathbf{h}_2{}^\intercal\right)^\intercal \left(\mathbf{I}_p \otimes \mathbf{I}_{NT}\right) \operatorname{vec}\left(\mathbf{Z}\right).$$

Therefore, since $\|\Gamma_{\mathbf{y}}\| = O(1)$ *as.* , $\left\|\left(\mathbf{I}_p \otimes {\Gamma_{\mathbf{y}}}^{1/2}\right) \mathrm{vec}\left(\tilde{\mathbf{y}}\mathbf{h}_2{}^{\intercal}\right)\right\| \lesssim \|\mathrm{vec}\left(\tilde{\mathbf{y}}\mathbf{h}_2{}^{\intercal}\right)\|$ which is concentrated around $\sqrt{NTp}$ (Vershynin, 2018) hence is $O(\sqrt{NTp})$ *whp.* Also, we have $\|(\mathbf{I}_p \otimes \mathbf{I}_{NT}) \mathrm{vec}\left(\Phi_v \mathbf{h}_2{}^{\intercal}\right)\| = O(\sqrt{T})$ *as.* By Lemma 20,

$$\mathbb{P}\left(\left|\sum_t \mathbf{y}_t{}^{\intercal}\mathbf{Z}_t\mathbf{h}_2\right| > \upsilon\right) \leq 2\exp\left(-\frac{c\upsilon^2}{NTp}\right) + 2\exp\left(-\frac{c\upsilon^2}{T}\right) + \frac{1}{N}.$$

Therefore, the conclusion follows after letting $NT\upsilon = \omega(\sqrt{NTp})$. ■

Proof [of 8] The same logic as the proof in 6 applies here. Since $\|\mathcal{L}_{\mathbf{A}}\| \leq 1$, we get the same bound by letting $NT\upsilon = \omega(\sqrt{NTp})$. ■

Proof [of 9] Note that

$$\sum_t \mathbf{h}_2{}^{\intercal}\mathbf{Z}_t{}^{\intercal}\mathbf{Z}_t\mathbf{h}_2 = \mathrm{tr}\left(\mathbf{Z}\mathbf{h}_2\mathbf{h}_2{}^{\intercal}\mathbf{Z}^{\intercal}\right) = \mathrm{vec}\left(\mathbf{Z}^{\intercal}\right)^{\intercal}\left\{\mathbf{I}_{NT} \otimes \left(\mathbf{h}_2\mathbf{h}_2{}^{\intercal}\right)\right\}\mathrm{vec}\left(\mathbf{Z}^{\intercal}\right).$$

Note that $\mathbb{E}\left[\mathrm{vec}\left(\mathbf{Z}^{\intercal}\right)\mathrm{vec}\left(\mathbf{Z}^{\intercal}\right)^{\intercal}\right] = \mathbf{I}_{NT} \otimes \Sigma_z$. Also, by the definition of $\mathbf{h}_2$, we have $\|\mathbf{I}_{NT} \otimes \left(\mathbf{h}_2\mathbf{h}_2^{\intercal}\right)\|_F^2 \leq NT$ and $\|\mathbf{I}_{NT} \otimes \left(\mathbf{h}_2\mathbf{h}_2^{\intercal}\right)\| \leq 1$. Therefore, by Hanson-Wright inequality (Rudelson and Vershynin, 2013),

$$\mathbb{P}\left[\left|\mathbf{h}_2{}^{\intercal}\sum_t \mathbf{Z}_t^{\intercal}\mathbf{Z}_t\mathbf{h}_2 - NT\mathbf{h}_2{}^{\intercal}\Sigma_z\mathbf{h}_2\right| > \upsilon\right] < 2\exp\left[-c\min\left\{\frac{\upsilon^2}{NT}, \upsilon\right\}\right].$$

■

Proof [of 10] Note that $\|\Delta_v\|^2 = O\left(\frac{1}{N\rho_N}\right)$ *whp.* and $\left\|\tilde{\mathbf{V}}\right\| = 1$. Also, $\|\mathbf{H}_v^{\intercal}\beta_v\| = O(1)$. So,

$$\left|\mathbf{h}_1{}^{\intercal}\hat{\mathbf{V}}^{\intercal}\Delta_v\mathbf{H}_v^{\intercal}\beta_v\right| \leq \left(\|\Delta_v\|^2 + \|\Delta_v\|\right)\|\beta_v\| = O\left(1/\sqrt{N\rho_N}\right) \text{ whp.}$$

By letting $\upsilon = \omega\left(\sqrt{\frac{T}{N\rho_N}}\right)$ and noting that $N\rho_N = \omega(T)$, we have the conclusion. ■

Proof [of 11] Similar to the proof of 1., taking $\mathbf{C} \triangleq {\Gamma_{\mathbf{y}}}^{1/2}\left(\mathbf{1}_T \otimes \Delta_v\right)$ which has the spectral norm of $O\left(\sqrt{\frac{T}{N\rho_N}}\right)$ *whp.*, we have

$$\sum_t \mathbf{y}_t{}^{\intercal}\Delta_v\mathbf{H}_v^{\intercal}\beta_v = \left({\Gamma_{\mathbf{y}}}^{1/2}\tilde{\mathbf{y}} + \Phi_v\right)^{\intercal}\left(\mathbf{1}_T \otimes \Delta_v\right)\mathbf{H}_v^{\intercal}\beta_v = \tilde{\mathbf{y}}^{\intercal}\mathbf{C}\mathbf{H}_v^{\intercal}\beta_v + \Phi_v{}^{\intercal}{\Gamma_{\mathbf{y}}}^{-\frac{1}{2}}\mathbf{C}\mathbf{H}_v^{\intercal}\beta_v.$$

First noting that $\|\beta_v\| = O(1)$ as $K \to \infty$, Lemma 20 gives

$$\mathbb{P}\left(|\tilde{\mathbf{y}}^{\intercal}\mathbf{C}\mathbf{H}_v^{\intercal}\beta_v| \geq \sqrt{NT}\upsilon\right) \leq 2\exp\left(-cN^2\rho_N\upsilon^2\right) + \frac{1}{N}.$$

It is straightforward that $\left|\Phi_v{}^{\intercal}{\Gamma_{\mathbf{y}}}^{-\frac{1}{2}}\mathbf{C}\mathbf{H}_v^{\intercal}\beta_v\right| = O(T/\sqrt{N\rho_N})$ *whp.* hence by letting $\upsilon = \omega\left(\sqrt{\frac{T}{N^2\rho_N}}\right)$, we obtain the conclusion. ■

Proof [of 12] Let $\mathbf{C} \triangleq \Gamma_{\mathbf{y}}{}^{1/2} \{\mathbf{1}_T \otimes (\mathcal{L}_{\mathbf{A}} \Delta_v)\}$ which is of $O\left(\sqrt{\frac{T}{N\rho_N}}\right)$ *whp.* The rest of the proof coincides with the proof of 11. ■

Proof [of 13] Note that $\mathbf{h}_2{}^{\intercal} \sum_t \mathbf{Z}_t{}^{\intercal} \Delta_v \mathbf{H}_v^{\intercal} \beta_v = \mathbf{h}_2{}^{\intercal} \mathbf{Z}^{\intercal} \left(\mathbf{1}_T \otimes \Delta_v\right) \mathbf{H}_v^{\intercal} \beta_v$ which equals

$$\operatorname{tr}\left\{\mathbf{Z}\mathbf{h}_2 \beta_v^{\intercal} \mathbf{H}_v \left(\mathbf{1}_T \otimes \Delta_v\right)^{\intercal}\right\} = \operatorname{vec}\left(\mathbf{h}_2 \beta_v^{\intercal} \mathbf{H}_v\right)^{\intercal} \left\{\mathbf{I}_p \otimes \left(\mathbf{1}_T \otimes \Delta_v\right)^{\intercal}\right\} \operatorname{vec}\left(\mathbf{Z}\right).$$

Since the norm of $\operatorname{vec}\left(\mathbf{h}_2 \beta_v^{\intercal} \mathbf{H}_v\right)^{\intercal} \left\{\mathbf{I}_p \otimes \left(\mathbf{1}_T \otimes \Delta_v\right)^{\intercal}\right\}$ is of $O\left(\sqrt{\frac{T}{N\rho_N}}\right)$ *whp.*, we obtain the bound after applying Lemma 20. ■

Lemma 26 *Assume that* $N\rho_N = \omega\left(T + \log^4 N\right)$. *Then, there exist* $N_0, T_0 > 0$ *and* $\upsilon_0(N_0, T_0) \in (0,1)$ *such that for all* $\upsilon \in (0, \upsilon_0)$ *and* $N > N_0, T > T_0$,

$$\mathbb{P}\left[\sup_{\mathbf{h} \in B^{K+p+1}} \left|\mathbf{h}^{\intercal}\left(\hat{\Sigma}_v - \Sigma_v\right)\mathbf{h}\right| > \upsilon\right] \lesssim K^2 \left\{\frac{1}{N} + \exp\left(-\frac{cNT\upsilon^2}{K^2 p}\right)\right\}.$$

Proof Let $\mathbf{h} \in B^{K+p+1}$ and $\mathbf{e}_i \in S^{K+p+1}$ be the i^{th} canonical basis of $\mathbb{R}^{K+p+2}$, i.e., $(\mathbf{e}_i)_j \triangleq \mathbb{I}_{\{i\}}(j)$ for $j = 1, \ldots, K+p+2$. By Lemma 25, we have

$$\max_{i,j} \mathbb{P}\left(\left|\mathbf{e}_i{}^{\intercal}\left(\hat{\Sigma}_v - \Sigma_v\right)\mathbf{e}_j\right| > \upsilon\right) \lesssim \frac{1}{N} + \exp\left(-c_1 \upsilon^2 NT/p\right) + \exp\left(-c_2 \upsilon^2 T N^2 \rho_N\right).$$

Since

$$\sup_{\mathbf{h} \in B^{K+p+1}} \left|\mathbf{h}^{\intercal}\left(\hat{\Sigma}_v - \Sigma_v\right)\mathbf{h}\right| \leq \sup_{\mathbf{h}} \|\mathbf{h}\|_1^2 \max_{i,j} \left|\hat{\sigma}_{v,ij} - \sigma_{v,ij}\right| \lesssim K \max_{i,j} \left|\hat{\sigma}_{v,ij} - \sigma_{v,ij}\right|,$$

we have

$$\begin{aligned}
\mathbb{P}\left[\sup_{\mathbf{h} \in B^{K+p+1}} \left|\mathbf{h}^{\intercal}\left(\hat{\Sigma}_v - \Sigma_v\right)\mathbf{h}\right| > \upsilon\right] &\leq \mathbb{P}\left(\max_{i,j} \left|\hat{\sigma}_{v,ij} - \sigma_{v,ij}\right| > \frac{ce}{K}\right) \\
&\lesssim K^2 \left\{\frac{1}{N} + \exp\left(-c_1{}'\upsilon^2 \frac{NT}{K^2 p}\right) + \exp\left(-c_2{}'\upsilon^2 \frac{N^2 \rho_N T}{K^2}\right)\right\} \\
&\lesssim K^2 \left[\frac{1}{N} + \exp\left\{-\frac{cNT\upsilon^2}{K^2} \min\left\{N\rho_N, \frac{1}{p}\right\}\right\}\right].
\end{aligned}$$

■

Lemma 27 *If* $N\rho_N = \omega(T + \log^4 N)$, *then we have*

$$\sqrt{T}\left(\hat{\mathbf{V}}\hat{\beta}_v - \mathbf{V}\beta_v\right) \Rightarrow \mathcal{N}\left(\mathbf{o}_K, \sigma^2 \mathbf{V}\mathbf{V}^{\intercal}\right)$$

as $\min(N, T) \to \infty$.

Proof Recall that $\sqrt{T}\left(\hat{\beta}_v - \mathbf{H}_v{}^\intercal \beta_v\right) \Rightarrow \mathcal{N}\left(\mathbf{o}_K, \sigma^2 \mathbf{I}_K\right)$ by Theorem 10. By simple algebra, we have

$$\begin{aligned}\sqrt{T}\left(\hat{\mathbf{V}}\hat{\beta}_v - \mathbf{V}\beta_v\right) &= \sqrt{T}\left(\hat{\mathbf{V}}\hat{\beta}_v - \tilde{\mathbf{V}}\mathbf{H}_v{}^\intercal \beta_v\right) \\ &= \Delta_v \sqrt{T}\left(\hat{\beta}_v - \mathbf{H}_v{}^\intercal \beta_v\right) + \sqrt{T}\Delta_v \mathbf{H}_v{}^\intercal \beta_v \\ &\quad + \tilde{\mathbf{V}}\sqrt{T}\left(\hat{\beta}_v - \mathbf{H}_v{}^\intercal \beta_v\right). \end{aligned} \tag{11}$$

By assumption, $\sqrt{T}\Delta_v = O\left(\sqrt{\frac{T}{N\rho_N}}\right)$ *whp.* hence we have

$$\mathbb{P}\left[\left\{\sqrt{T}\Delta_v \leq C\sqrt{\frac{T}{N\rho_N}}\right\} \; ev.\right] = 1$$

for some constant $C > 0$. Provided that $N\rho_N = \omega(T)$, this implies $\sqrt{T}\Delta_v \to 0$ *as.* Therefore, we have (11) $= O_{\mathbb{P}}(1) + \tilde{\mathbf{V}}\sqrt{T}\left(\hat{\beta}_v - \mathbf{H}_v{}^\intercal \beta_v\right)$ and the conclusion by Slutsky. ■

C.6 ENAR with spectral embedding $\mathbf{U}_e = \mathbf{V}\mathbf{S}^{1/2}$ under finite time

Note that under the finite time assumption, we proposed to use spectral embedding $\mathbf{U} = \mathbf{U}_e$ for ENAR instead of eigenvectors. i.e., $\mathbf{U}_e = \mathbf{V}\mathbf{S}^{1/2}$ and $\hat{\mathbf{U}}_e = \hat{\mathbf{V}}|\hat{\mathbf{S}}|^{1/2}$. Here, we present theoretical results for the finite-time corollary models specified in section 2.2.1. Recall ENR:

$$\mathbf{y} = \alpha_e \mathbf{1}_N + \mathbf{U}_e \beta_e + \mathbf{Z}\gamma_e + \mathcal{E}$$

and ENAR (with finite T) models:

$$\mathbf{y}_{t+1} = \alpha_e \mathbf{y}_t + \theta_e \mathcal{L}_{\mathbf{A}} \mathbf{y}_t + \mathbf{U}_e \beta_e + \mathbf{Z}\gamma + \mathcal{E}_{t+1}.$$

Note that ENR can be treated as a subset model of ENAR. Therefore, we only verify the asymptotic properties of ENAR with finite T.

C.6.1 Consistency

As we have $\mathbf{U} = \mathbf{U}_e$, $\|\varphi_e\| = O(\|\mathbf{U}_e\beta_e\|)$ $\mathbb{P}^* - as.$ (hence $\mathbb{P} - as.$) Since $\left\|\mathbf{V}\mathbf{S}^{1/2}\right\| = O(\sqrt{N\rho_N})$, we have $\|\Phi_e\| = O(\sqrt{T}\,\|\varphi_e\|) = O(\sqrt{TN\rho_N})$ *as.* Though we assume that T is finite in this section, our proof of the consistency of $\hat{\mu}$ here keeps track of the order of T. Define $\tilde{\mathbf{U}}_e \triangleq \mathbf{U}_e\mathbf{H}_e$. For convenience, we denote $1 + \frac{K\log N}{\sqrt{N\rho_N}}$ by π_e in this section so that $\|\Delta_e\| \triangleq \left\|\hat{\mathbf{U}}_e - \tilde{\mathbf{U}}_e\right\| = O\left(\pi_e\right)$ *whp.* if $N\rho_N = \omega\left(\log^4 N\right)$ by Proposition 23.

Proof [of Theorem 4] Here, the formulation is very similar to the case of ENAR with eigenvectors. First, let

$$\mathbf{W}_t^H \triangleq \left[\tilde{\mathbf{U}}_e \,|\, \mathbf{y}_t, \mathcal{L}_{\mathbf{A}}\mathbf{y}_t \,|\, \mathbf{Z}_t\right], \qquad \mu_{w,e}^H \triangleq (\sqrt{\rho_N}\beta_e{}^\intercal \mathbf{H}_e, \alpha_e, \theta_e, \gamma_e{}^\intercal)^\intercal,$$

$$\mathcal{W}_t \triangleq \left[\frac{1}{\sqrt{\rho_N}}\mathbf{U}_e \,|\, \mathbf{y}_t, \mathcal{L}_{\mathbf{A}}\mathbf{y}_t \,|\, \mathbf{Z}_t\right], \qquad \mathcal{W}_t^H \triangleq \left[\frac{1}{\sqrt{\rho_N}}\tilde{\mathbf{U}}_e \,|\, \mathbf{y}_t, \mathcal{L}_{\mathbf{A}}\mathbf{y}_t \,|\, \mathbf{Z}_t\right]$$

and $\hat{\mathcal{W}}_t \triangleq \left[\frac{1}{\sqrt{\rho_N}}\hat{\mathbf{U}}_e \,|\, \mathbf{y}_t, \mathcal{L}_{\mathbf{A}}\mathbf{y}_t \,|\, \mathbf{Z}_t\right]$. Denote $\mathbf{D}_e \triangleq \sqrt{NT}\,\mathrm{diag}(\sqrt{\rho_N}\mathbf{I}_K, \mathbf{I}_{p+2})$. Under this representation, we have $\hat{\mathcal{W}}_t\mathbf{D}_e = \sqrt{NT}\hat{\mathbf{W}}_t$, $\sqrt{NT}{\mathbf{D}_e}^{-1}\mu_{w,e}^H = \mu_e^H$, and $\mathbf{W}_t^H\mu_e^H = \mathcal{W}_t^H\mu_{w,e}^H$. Collect and bind them row-wise for $t = 0, ..., T-1$ to obtain $NT \times (K+p+2)$ matrices $\mathbf{W}^H$, $\mathcal{W}^H$, and define $\hat{\mathcal{W}}$ analogously. For $\hat{\Sigma}_e \triangleq \frac{1}{NT}\hat{\mathcal{W}}^\intercal\hat{\mathcal{W}}$ and $\mathbf{r}_{w\epsilon} \triangleq \frac{1}{NT}\hat{\mathcal{W}}^\intercal\mathcal{E}$, we have

$$
\begin{aligned}
\hat{\mu}_e &= \left(\hat{\mathbf{W}}^\intercal\hat{\mathbf{W}}\right)^{-1}\hat{\mathbf{W}}^\intercal\left(\mathbf{W}^H\mu_e^H + \mathcal{E}\right) \\
&= \left(\frac{\mathbf{D}_e}{NT}\hat{\mathcal{W}}^\intercal\hat{\mathcal{W}}\mathbf{D}_e\right)^{-1}\frac{\mathbf{D}_e}{\sqrt{NT}}\hat{\mathcal{W}}^\intercal\left(\mathcal{W}^H\mu_{w,e}^H + \mathcal{E}\right) \\
&= \left(\hat{\Sigma}_e\mathbf{D}_e\right)^{-1}\frac{1}{\sqrt{NT}}\hat{\mathcal{W}}^\intercal\left\{\left(\mathcal{W}^H - \hat{\mathcal{W}}\right)\mu_{w,e}^H + \hat{\mathcal{W}}\mu_{w,e}^H + \mathcal{E}\right\} \\
&= {\mathbf{D}_e}^{-1}\hat{\Sigma}_e^{-1}\left\{\frac{1}{\sqrt{NT}}\hat{\mathcal{W}}^\intercal\left(\mathcal{W}^H - \hat{\mathcal{W}}\right)\mu_{w,e}^H + \sqrt{NT}\mathbf{r}_{w\epsilon}\right\} + \sqrt{NT}{\mathbf{D}_e}^{-1}\mu_{w,e}^H \\
&\therefore \mathbf{D}_e\left(\hat{\mu}_e - \mu_e^H\right) = \frac{\hat{\Sigma}_e^{-1}}{\sqrt{NT}}\hat{\mathcal{W}}^\intercal\left(\mathcal{W}^H - \hat{\mathcal{W}}\right)\mu_{w,e}^H + \sqrt{NT}\hat{\Sigma}_e^{-1}\mathbf{r}_{w\epsilon}.
\end{aligned}
$$

Therefore, we next show that the first term on the RHS is negligible and $\sqrt{NT}\hat{\Sigma}_e^{-1}\mathbf{r}_{w\epsilon}$ is converging to a multivariate normal distribution. We start with claiming that $\hat{\Sigma}_e$ is converging to a matrix with finite entries as N and T tend to infinity. This will allow us to focus on the behaviors of $\frac{1}{\sqrt{NT}}\hat{\mathcal{W}}^\intercal(\mathcal{W}^H - \hat{\mathcal{W}})\mu_{w,e}^H$ and $\sqrt{NT}\mathbf{r}_{w\epsilon}$, and then apply Slutsky's Theorem.

Claim C.6.1 *$\hat{\Sigma}_e$ converges to Σ_e in probability, i.e., $\hat{\Sigma}_e \Rightarrow \Sigma_e$.*

Proof As consequences of Lemma 29, we have

$$
\begin{aligned}
\hat{\Sigma}_e &= \frac{1}{NT}\sum_t \begin{bmatrix} \frac{1}{\rho_N}|\hat{\mathbf{S}}| & \frac{1}{\sqrt{\rho_N}}\hat{\mathbf{U}}_e^\intercal\mathbf{y}_t & \frac{1}{\sqrt{\rho_N}}\hat{\mathbf{U}}_e^\intercal\mathcal{L}_{\mathbf{A}}\mathbf{y}_t & \frac{1}{\sqrt{\rho_N}}\hat{\mathbf{U}}_e^\intercal\mathbf{Z}_t \\ & \mathbf{y}_t^\intercal\mathbf{y}_t & \mathbf{y}_t^\intercal\mathcal{L}_{\mathbf{A}}\mathbf{y}_t & \mathbf{y}_t^\intercal\mathbf{Z}_t \\ & & \mathbf{y}_t^\intercal\mathcal{L}_{\mathbf{A}}^2\mathbf{y}_t & \mathbf{y}_t^\intercal\mathcal{L}_{\mathbf{A}}\mathbf{Z}_t \\ & & & \mathbf{Z}_t^\intercal\mathbf{Z}_t \end{bmatrix} \\
&\overset{N\to\infty}{\Rightarrow} \lim_{N\to\infty} \begin{bmatrix} \frac{1}{N\rho_N}\mathbf{S} & \frac{1}{N\sqrt{\rho_N}}\tilde{\mathbf{U}}_e^\intercal\varphi_e & \frac{1}{N\sqrt{\rho_N}}\tilde{\mathbf{U}}_e^\intercal\mathcal{L}_{\mathbf{A}}\varphi_e & \mathbf{O}_{K\times p} \\ & \frac{1}{N}\varphi_e^\intercal\varphi_e + \tau_2 & \frac{1}{N}\varphi_e^\intercal\mathcal{L}_{\mathbf{A}}\varphi_e + \tau_{23} & \mathbf{o}_p^\intercal \\ & & \frac{1}{N}\varphi_e^\intercal\mathcal{L}_{\mathbf{A}}^2\varphi_e + \tau_3 & \mathbf{o}_p^\intercal \\ & & & \Sigma_z \end{bmatrix}
\end{aligned}
$$

Note that $\|\mathbf{S}\| = O(N\rho_N)$ and $\varphi_e = O(\sqrt{N\rho_N})$ *as*. Existence of τ_2, τ_{23}, and τ_3 come by dominated convergence theorem after noting that $\mathrm{tr}\,(\mathcal{L}_{\mathbf{A}}\Gamma) \leq \|\mathcal{L}_{\mathbf{A}}\|\,\mathrm{tr}\,(\Gamma) = O\,(N)$ *as*. and $\mathrm{tr}\,(\mathcal{L}_{\mathbf{A}}^2\Gamma) = O(N)$ *as*. Also from the assumptions and the proof of Lemma 28, we have

$$
\begin{aligned}
&\frac{1}{N\sqrt{\rho_N}}\mathbf{h}_1^\intercal\tilde{\mathbf{U}}_e^\intercal\varphi_e = O(\sqrt{\rho_N}), \frac{1}{N\sqrt{\rho_N}}\mathbf{h}_1^\intercal\tilde{\mathbf{U}}_e^\intercal\mathcal{L}_{\mathbf{A}}\varphi_e = O(\sqrt{\rho_N}), \\
&\qquad \frac{1}{N}\varphi_e^\intercal\varphi_e = O(\rho_N), \frac{1}{N}\varphi_e^\intercal\mathcal{L}_{\mathbf{A}}\varphi_e = O(\rho_N), \frac{1}{N}\varphi_e^\intercal\mathcal{L}_{\mathbf{A}}^2\varphi_e = O(\rho_N)\ as.
\end{aligned}
$$

for all $\mathbf{h}_1 \in B^{K-1}$. So, we obtain the asymptotic precision matrix.

The following claim clearly shows that the parameter estimation errors, except for $\hat{\gamma}_e$, are not guaranteed to be finite if $T \to \infty$.

Claim C.6.2 *For any* $\mathbf{h} \in B^{K+p+1}$, $\frac{1}{\sqrt{NT}}\left(\hat{\mathcal{W}}\mathbf{h}\right)^{\intercal}\left(\mathcal{W}^{H}-\hat{\mathcal{W}}\right)\mu_{w,e}^{H} = O\left(1\right)$ *whp.*

Proof Since $\hat{\mathcal{W}}$ is different from $\mathcal{W}^{H}$ by latent variables only, we have

$$\frac{1}{\sqrt{NT}}\hat{\mathcal{W}}^{\intercal}\left(\mathcal{W}^{H}-\hat{\mathcal{W}}\right)\mu_{w,e}^{H} = \begin{bmatrix} \sqrt{\frac{T}{N\rho_N}}\hat{\mathbf{U}}_e^{\intercal} \\ \frac{1}{\sqrt{NT}}\sum_t \mathbf{y}_t{}^{\intercal} \\ \frac{1}{\sqrt{NT}}\sum_t \mathbf{y}_t{}^{\intercal}\mathcal{L}_{\mathbf{A}} \\ \frac{1}{\sqrt{NT}}\sum_t \mathbf{Z}_t{}^{\intercal} \end{bmatrix} \Delta_e \mathbf{H}^{\intercal}\beta_e.$$

Then, by the statements 10–13 of Lemma 28, we have $\frac{1}{\sqrt{NT}}\hat{\mathcal{W}}^{\intercal}\left(\mathcal{W}^{H}-\hat{\mathcal{W}}\right)\mu_{w,e}^{H} = \begin{bmatrix}\mathbf{b}_1\\ \mathbf{b}_2\\ \mathbf{b}_3\end{bmatrix}$ where $\mathbf{b}_1 \in \mathbb{R}^K, \mathbf{b}_2 \in \mathbb{R}^2, \mathbf{b}_3 \in \mathbb{R}^p$ are such that $\mathbf{x}^{\intercal}\mathbf{b}_1 = O\left(\sqrt{\pi_e T}\right)$ *whp.*, $\mathbf{y}^{\intercal}\mathbf{b}_2 = O\left(\sqrt{\pi_e T\rho_N}\right)$ *whp.*, and $\mathbf{z}^{\intercal}\mathbf{b}_3 = O(\sqrt{\pi_e/N})$ *whp.* for any $\mathbf{x} \in B^{K-1}, \mathbf{y} \in B^1, \mathbf{z} \in B^{p-1}$. Noting that $\pi_e = O(1)$, we get the conclusion.

Claim C.6.3 $\sqrt{NT}\mathbf{r}_{w\epsilon} \Rightarrow \mathcal{N}(\mathbf{o}_{K+p+2}, \sigma^2\Sigma_e)$.

Proof It is sufficient to show that for any $\eta \in B^{K+p+1}$, we have $\sqrt{NT}\eta^{\intercal}\mathbf{r}_{w\epsilon} \Rightarrow \mathcal{N}(0, \sigma^2\eta^{\intercal}\Sigma_e\eta)$. Denoting $\xi_{Nt} \triangleq (NT)^{-1/2}\eta^{\intercal}\hat{\mathcal{W}}_{t-1}^{\intercal}\mathcal{E}_t$ and $\mathcal{F}_{Nt} = \sigma(\mathbf{A}, \epsilon_{is}, \mathbf{Z}_{i,(s-1)}; i \le N, -\infty < s \le t)$, $\{\sum_{s=1}^{t}\xi_{Ns}, \mathcal{F}_{Nt}\}$ constitutes a martingale array for each $N, t \le T$. Then, we can apply Corollary 3.1 of Hall and Heyde (2014) to $\sqrt{NT}\eta^{\intercal}\mathbf{r}_{w\epsilon} = \sum_t \xi_{N,t+1}$ by checking the following two conditions, (1) and (2).

(1) $\sum_t \mathbb{E}\left(\xi_{N,t+1}^2 1_{\{|\xi_{N,t+1}|>\upsilon\}}|\mathcal{F}_{Nt}\right) = O_{\mathbb{P}}(1)$.

(2) $\sum_t \mathbb{E}\left(\xi_{N,t+1}^2|\mathcal{F}_{Nt}\right) = \eta^{\intercal}\sigma^2\Sigma_e\eta + O_{\mathbb{P}}(1)$.

Proof [of (1)] First, we have

$$\sum_{t=0}^{T-1}\mathbb{E}\left(\xi_{N,t+1}^2 1_{\{|\xi_{N,t+1}|>\upsilon\}}|\mathcal{F}_{Nt}\right) \le \upsilon^{-2}\sum_t \mathbb{E}\left(\xi_{N,t+1}^4|\mathcal{F}_{Nt}\right).$$

One can easily verify that

$$\mathbb{E}\left(\xi_{N,t+1}^4|\mathcal{F}_{Nt}\right) \lesssim \sigma^4\left(\frac{1}{NT}\eta^{\intercal}\hat{\mathcal{W}}_t^{\intercal}\hat{\mathcal{W}}_t\eta\right)^2.$$

So, we only need to show that $\sum_t\left(\frac{1}{NT}\eta^{\intercal}\hat{\mathcal{W}}_t^{\intercal}\hat{\mathcal{W}}_t\eta\right)^2 = O_{\mathbb{P}}(1)$. First note that

$$(NT)^{-1}\hat{\mathcal{W}}_t^{\intercal}\hat{\mathcal{W}}_t = \begin{bmatrix} \frac{1}{NT\rho_N}\hat{\mathbf{S}} & \frac{1}{NT\sqrt{\rho_N}}\hat{\mathbf{U}}_e^{\intercal}\mathbf{y}_t & \frac{1}{NT\sqrt{\rho_N}}\hat{\mathbf{U}}_e^{\intercal}\mathcal{L}_{\mathbf{A}}\mathbf{y}_t & \frac{1}{NT\sqrt{\rho_N}}\hat{\mathbf{U}}_e^{\intercal}\mathbf{Z}_t \\ & \frac{1}{NT}\mathbf{y}_t{}^{\intercal}\mathbf{y}_t & \frac{1}{NT}\mathbf{y}_t{}^{\intercal}\mathcal{L}_{\mathbf{A}}\mathbf{y}_t & \frac{1}{NT}\mathbf{y}_t{}^{\intercal}\mathbf{Z}_t \\ & & \frac{1}{NT}\mathbf{y}_t{}^{\intercal}\mathcal{L}_{\mathbf{A}}^2\mathbf{y}_t & \frac{1}{NT}\mathbf{y}_t{}^{\intercal}\mathcal{L}_{\mathbf{A}}\mathbf{Z}_t \\ & & & \frac{1}{NT}\mathbf{Z}_t^{\intercal}\mathbf{Z}_t \end{bmatrix}.$$

Since similar arguments can be used to show the convergence of each entry, take

$$\sum_t\left(\frac{1}{NT\sqrt{\rho_N}}\eta_1{}^{\intercal}\hat{\mathbf{U}}_e^{\intercal}\mathbf{y}_t\right)^2 = \frac{1}{N^2T^2\rho_N}\sum_t \eta_1{}^{\intercal}\hat{\mathbf{U}}_e^{\intercal}\mathbf{y}_t\mathbf{y}_t{}^{\intercal}\hat{\mathbf{U}}_e\eta_1$$

for example, where $\eta_1 \in B^{K-1}$. Then, we have

$$\sum_t \eta_1{}^\intercal \hat{\mathbf{U}}_e^\intercal \mathbf{y}_t \mathbf{y}_t{}^\intercal \hat{\mathbf{U}}_e \eta_1 = \mathbf{y}^\intercal \left\{ \mathbf{I}_T \otimes \left(\hat{\mathbf{U}}_e \eta_1 \eta_1{}^\intercal \hat{\mathbf{U}}_e^\intercal \right) \right\} \mathbf{y}.$$

First, we have

$$\tilde{\mathbf{y}}^\intercal {\Gamma_{\mathbf{y}}}^{1/2} \left\{ \mathbf{I}_T \otimes \left(\hat{\mathbf{U}}_e \eta_1 \eta_1{}^\intercal \hat{\mathbf{U}}_e^\intercal \right) \right\} {\Gamma_{\mathbf{y}}}^{1/2} \tilde{\mathbf{y}} = \tilde{\mathbf{y}}^\intercal \left\{ \mathbf{I}_T \otimes \left(\Gamma^{1/2} \hat{\mathbf{U}}_e \eta_1 \eta_1{}^\intercal \hat{\mathbf{U}}_e^\intercal \Gamma^{1/2} \right) \right\} \tilde{\mathbf{y}}.$$

Letting $\tilde{\Gamma} \triangleq \mathbf{I}_T \otimes \left(\Gamma^{1/2} \hat{\mathbf{U}}_e \eta_1 \eta_1{}^\intercal \hat{\mathbf{U}}_e^\intercal \Gamma^{1/2} \right)$, we have $\left\| \tilde{\Gamma} \right\|_F^2 = T \operatorname{tr} \left\{ \left(\hat{\mathbf{U}}_e \eta_1 \eta_1{}^\intercal \hat{\mathbf{U}}_e^\intercal \Gamma \right)^2 \right\} \leq T \left\| \hat{\mathbf{U}}_e \eta_1 \eta_1{}^\intercal \hat{\mathbf{U}}_e^\intercal \right\|^2 \operatorname{tr} (\Gamma^2) = O\left(NT \cdot N^2 {\rho_N}^2 \right)$ *whp.* and $\left\| \tilde{\Gamma} \right\| = \left\| \Gamma^{1/2} \hat{\mathbf{U}}_e \eta_1 \eta_1{}^\intercal \hat{\mathbf{U}}_e^\intercal \Gamma^{1/2} \right\| \leq \left\| \hat{\mathbf{U}}_e \right\|^2 \|\Gamma\| = O(N \rho_N)$ *whp.* By Lemma 22,

$$\begin{aligned} &\mathbb{P}\left[\frac{1}{N^2 T^2 \rho_N} \left| \tilde{\mathbf{y}}^\intercal \tilde{\Gamma} \tilde{\mathbf{y}} - \mathbb{E}\left(\tilde{\mathbf{y}}^\intercal \tilde{\Gamma} \tilde{\mathbf{y}} \right) \right| > \upsilon \right] \\ &\quad < 2 \exp\left\{ -c \min\left(\frac{\upsilon^2 N^4 T^4 {\rho_N}^2}{N^3 T {\rho_N}^2}, \frac{\upsilon N^2 T^2 \rho_N}{N \rho_N} \right) \right\} + 1/N. \end{aligned}$$

Note that $\mathbb{E}\left(\tilde{\mathbf{y}}^\intercal \tilde{\Gamma} \tilde{\mathbf{y}} \right) = T \mathbb{E} \operatorname{tr}\left(\hat{\mathbf{U}}_e \eta_1 \eta_1{}^\intercal \hat{\mathbf{U}}_e^\intercal \Gamma \right)$ and

$$T \operatorname{tr}\left(\hat{\mathbf{U}}_e \eta_1 \eta_1{}^\intercal \hat{\mathbf{U}}_e^\intercal \Gamma \right) \leq T \left\| \hat{\mathbf{U}}_e \eta_1 \eta_1{}^\intercal \hat{\mathbf{U}}_e^\intercal \right\| \operatorname{tr}(\Gamma)$$

hence is $O\left(N^2 T \rho_N \right)$ *whp.* Therefore, $\frac{\tilde{\mathbf{y}}^\intercal \tilde{\Gamma} \tilde{\mathbf{y}}}{N^2 T \rho_N} < \infty = O(1)$ *whp.* hence $\tilde{\mathbf{y}}^\intercal \tilde{\Gamma} \tilde{\mathbf{y}} = O_{\mathbb{P}}(N^2 T^2 \rho_N)$. So we have $\mathbb{E}\left(\tilde{\mathbf{y}}^\intercal \tilde{\Gamma} \tilde{\mathbf{y}} \right) = o\left(N^2 T^2 \rho_N \right)$ by dominated convergence theorem. Also, $\left| {\Phi_e}^\intercal \tilde{\Gamma} \Phi_e \right| = O(N^2 T {\rho_N}^2)$ *whp.* Therefore, we have $\frac{1}{N^2 T^2 \rho_N} \sum_t \left(\eta_1{}^\intercal \hat{\mathbf{U}}_e^\intercal \mathbf{y}_t \right)^2 = O_{\mathbb{P}}(1)$. Noting that both $\mathbf{y}$ and $\mathbf{Z}_t$ have finite fourth-order moments, one can show that the rest are also $O_{\mathbb{P}}(1)$ similarly.

Proof [of (2)] Since $\sum_t \mathbb{E}\left(\xi_{N,t+1}^2 | \mathcal{F}_{Nt} \right) = \frac{\sigma^2}{NT} \eta^\intercal \hat{\mathcal{W}}^\intercal \hat{\mathcal{W}} \eta = \sigma^2 \eta^\intercal \hat{\Sigma}_e \eta$, by Lemma 29, we have (2).

Therefore, by the three claims, we have the desired asymptotic normality after dividing $\mathbf{D}_e \left(\hat{\mu}_e - \mu_e^H \right)$ by $\sqrt{T}$. ■

Proof [of Theorem 5] We can follow the steps in the proof of Theorem 4. Since $N \rho_N = \omega(\log^4 N)$, we have $\pi_e = C + o\left(\frac{\sqrt{K}}{\log N} \right)$ for some constant $C > 0$. Therefore, by $N = \omega(K^2 \log K)$ we have $N = \omega(\pi_e)$ and $\log K - N/K^2 \to -\infty$ hence $\hat{\Sigma}_e \Rightarrow \Sigma_e$ by Lemma 28 and Lemma 29. Note that an additional order of $\log K$ is required on $N = \omega(K^2)$ because T is assumed to be finite. Check that $\mathbf{Q}_K \mathbf{D}_e \left(\hat{\mu}_e - \mu_e^H \right) = \frac{\mathbf{Q}_K \hat{\Sigma}_e^{-1}}{\sqrt{NT}} \hat{\mathcal{W}}^\intercal \left(\mathcal{W}^H - \hat{\mathcal{W}} \right) \mu_{w,e}^H + \sqrt{NT} \mathbf{Q}_K \hat{\Sigma}_e^{-1} \mathbf{r}_{w\epsilon}$ and $\|\mathbf{Q}_K\| = O(1)$. For convenience, write $\mathbf{d} \triangleq \frac{1}{\sqrt{NT}} \hat{\mathcal{W}}^\intercal \left(\mathcal{W}^H - \hat{\mathcal{W}} \right) \mu_{w,e}^H$. Since

$$\left\| \mathbf{Q}_K \hat{\Sigma}_e^{-1} \mathbf{d} \right\| \lesssim \left\| \mathbf{Q}_K \left(\hat{\Sigma}_e^{-1} - {\Sigma_e}^{-1} \right) \mathbf{d} \right\| + \left\| \mathbf{Q}_K {\Sigma_e}^{-1} \mathbf{d} \right\|,$$

we check if $\left\|\mathbf{Q}_K\left(\hat{\Sigma}_e^{-1}-{\Sigma_e}^{-1}\right)\mathbf{d}\right\| = O_{\mathbb{P}}\left(\left\|\mathbf{Q}_K{\Sigma_e}^{-1}\mathbf{d}\right\|\right)$. It suffices to check if it is satisfied that $\left\|\hat{\Sigma}_e^{-1}-{\Sigma_e}^{-1}\right\| = O_{\mathbb{P}}\left(\left\|{\Sigma_e}^{-1}\right\|\right)$. First, check that

$$\sup_{\mathbf{h}\in S^{K+p+1}}\left|\mathbf{h}^\intercal\left(\hat{\Sigma}_e-\Sigma_e\right)\mathbf{h}\right| \leq \sup_{\mathbf{h}\in \mathbf{B}^{K+p+1}}\left|\mathbf{h}^\intercal\left(\hat{\Sigma}_e-\Sigma_e\right)\mathbf{h}\right|$$

hence by Lemma 29, we confirm that this is true as

$$\left\|\hat{\Sigma}_e^{-1}-{\Sigma_e}^{-1}\right\| \lesssim s_{K+p+2}\left(\hat{\Sigma}_e\right)^{-1}\left\|\hat{\Sigma}_e-\Sigma_e\right\| s_{K+p+2}(\Sigma_e)^{-1} = s_{K+p+2}(\Sigma_e)^{-1}O_{\mathbb{P}}(1).$$

Therefore, it suffices to check the order of $\mathbf{Q}_K{\Sigma_e}^{-1}\mathbf{d}$. First, since $N=\omega(K^2\log K)$, we have $\pi_e = C + o\left(\frac{\sqrt{K}}{\log K}\right)$. Therefore from Claim C.6.2, we have $\mathbf{x}^\intercal\mathbf{Q}_K{\Sigma_e}^{-1}\mathbf{d} = O(1)$ *whp.* for any $\mathbf{x}\in B^{m-1}$ as $\|\mathbf{A}_1\| = O\left(\frac{\log K}{\sqrt{K}}\right)$.

Next, it is sufficient to show that for any $\eta\in B^{m-1}$, we have $\eta^\intercal\mathbf{Q}_K\hat{\Sigma}_e^{-1}\sqrt{NT}\mathbf{r}_{w\epsilon} \Rightarrow \mathcal{N}(0,\sigma^2\eta^\intercal\Lambda_e\eta)$. Since

$$\begin{aligned}\eta^\intercal\mathbf{Q}_K\hat{\Sigma}_e^{-1}\sqrt{NT}\mathbf{r}_{w\epsilon} &= \eta^\intercal\mathbf{Q}_K\left(\hat{\Sigma}_e^{-1}-{\Sigma_e}^{-1}\right)\sqrt{NT}\mathbf{r}_{w\epsilon} + \eta^\intercal\mathbf{Q}_K{\Sigma_e}^{-1}\sqrt{NT}\mathbf{r}_{w\epsilon}\\ &= O_{\mathbb{P}}(1) + \eta^\intercal\mathbf{Q}_K{\Sigma_e}^{-1}O_{\mathbb{P}}(1),\end{aligned}$$

we focus on the latter. Define $\xi_{Nt} \triangleq (NT)^{-1/2}\eta^\intercal\mathbf{Q}_K{\Sigma_e}^{-1}\hat{\mathcal{W}}_{t-1}^\intercal\mathcal{E}_t$ and a filtration $\mathcal{F}_{Nt} = \sigma(\mathbf{A},\epsilon_{is},\mathbf{Z}_{i,(s-1)}; i\leq N, -\infty<s\leq t)$. Then, a set of pairs $\{\sum_{s=1}^t\xi_{Ns},\mathcal{F}_{Nt}\}$ constitutes a martingale array for each $N,t\leq T$. Then, we repeat the following arguments (Hall and Heyde, 2014):

(1) $\sum_t\mathbb{E}\left(\xi_{N,t+1}^2 1_{\{|\xi_{N,t+1}|>v\}}|\mathcal{F}_{Nt}\right) = O_{\mathbb{P}}(1)$.

Proof We only need to show that

$$\sum_t\left(\frac{1}{NT}\eta^\intercal\mathbf{Q}_K{\Sigma_e}^{-1}\hat{\mathcal{W}}_t^\intercal\hat{\mathcal{W}}_t{\Sigma_e}^{-1}\mathbf{Q}_K{}^\intercal\eta\right)^2 = O_{\mathbb{P}}(1).$$

Since both $\mathbf{Q}_K$ and Σ_e have spectral norm of $O(1)$ as $K\to\infty$, this can be shown in the same manner as in the proof of (1) of the third claim in proving Theorem 4.

(2) $\sum_t\mathbb{E}\left(\xi_{N,t+1}^2|\mathcal{F}_{Nt}\right) = \sigma^2\eta^\intercal\Lambda_e\eta + O_{\mathbb{P}}(1)$.

Proof This can be shown by the arguments in the proof of (2) of the third claim in proving Theorem 4.

This completes the proof of Theorem 5. ■

C.6.2 Consistency under misspecification of K

Proof [of Theorem 15] Now let us augment $\mathbf{D}_e$ as $\mathbf{D}_{+,e}\triangleq\sqrt{NT}\,\mathrm{diag}\left(\sqrt{\rho_N}\mathbf{I}_{K_o},\mathbf{I}_{p+2}\right)$ and augment $\hat{\mathcal{W}}$ in the proof of Theorem 4 as $\hat{\mathcal{W}}_+\triangleq\left[\frac{1}{\sqrt{\rho_N}}\mathbf{1}_T\otimes\hat{\mathbf{U}}_{K:K_o,e}\,|\,\hat{\mathcal{W}}\right]\in\mathbb{R}^{NT\times(K_o+p+2)}$ so that $\sqrt{NT}\mathbf{W}_+ = \hat{\mathcal{W}}_+\mathbf{D}_{+,e}$. Also, we write $\mu_{+,e}^H = \left(\mathbf{o}_{K_o-K}{}^\intercal,\mu_e^{H\intercal}\right)^\intercal$ and $\mu_{w,se+}^H =$

$\left(\mathbf{o}_{K_o-K}{}^\intercal, \mu^H_{w,e}{}^\intercal\right)^\intercal$. To avoid complications, we denote $\mathbf{U}_{K:K_o,e}$ by $\mathbf{U}_{+,e}$. This representation gives $\hat{\mathcal{W}}_+\mu^H_{w,se+} = \hat{\mathcal{W}}\mu^H_{w,e}$ and $\sqrt{NT}\mathbf{D}_{+,e}{}^{-1}\mu^H_{w,se+} = \mu^H_{+,e}$. Define

$$\hat{\Sigma}_{+,e} \triangleq \frac{1}{NT}\hat{\mathcal{W}}_+^\intercal\hat{\mathcal{W}}_+ \tag{12}$$

and $\mathbf{r}_{w\epsilon,+} \triangleq \frac{1}{NT}\hat{\mathcal{W}}_+^\intercal\mathcal{E}$. Then, recalling the proof in the previous section,

$$\begin{aligned}
\hat{\mu}_{+,e} &= \left(\frac{\mathbf{D}_{+,e}}{NT}\hat{\mathcal{W}}_+^\intercal\hat{\mathcal{W}}_+\mathbf{D}_{+,e}\right)^{-1}\frac{\mathbf{D}_{+,e}}{\sqrt{NT}}\hat{\mathcal{W}}_+^\intercal\left(\mathcal{W}^H\mu^H_{w,e}+\mathcal{E}\right)\\
&= \left(\hat{\Sigma}_{+,e}\mathbf{D}_{+,e}\right)^{-1}\frac{1}{\sqrt{NT}}\hat{\mathcal{W}}_+^\intercal\left\{\left(\mathcal{W}^H-\hat{\mathcal{W}}\right)\mu^H_{w,e}+\hat{\mathcal{W}}_+\mu^H_{w,se+}+\mathcal{E}\right\}\\
&= \mathbf{D}_{+,e}{}^{-1}\hat{\Sigma}^{-1}_{+,e}\left\{\frac{1}{\sqrt{NT}}\hat{\mathcal{W}}_+^\intercal\left(\mathcal{W}^H-\hat{\mathcal{W}}\right)\mu^H_{w,e}+\sqrt{NT}\mathbf{r}_{w\epsilon,+}\right\}+\sqrt{NT}\mathbf{D}_{+,e}{}^{-1}\mu^H_{w,se+}\\
&\therefore \hat{\Sigma}_{+,e}\mathbf{D}_{+,e}\left(\hat{\mu}_{+,e}-\mu^H_{+,e}\right) = \frac{1}{\sqrt{NT}}\hat{\mathcal{W}}_+^\intercal\left(\mathcal{W}^H-\hat{\mathcal{W}}\right)\mu^H_{w,e}+\sqrt{NT}\mathbf{r}_{w\epsilon,+}.
\end{aligned}$$

Therefore, we can use similar logic as the proof of Theorem 4.

Claim C.6.4 $\hat{\Sigma}_{+,e} \Rightarrow \Sigma_{+,e}$.

Proof Now for $\hat{\mathbf{U}}_{+,e} = \hat{\mathbf{V}}_{K:K_o}|\hat{\mathbf{S}}_{K:K_o}|^{1/2}$, observe that $\hat{\mathbf{U}}^\intercal_{+,e}\hat{\mathbf{U}}_{+,e} = |\hat{\mathbf{S}}_{K:K_o}|$ which is a diagonal matrix containing $K+1,...,K_o$-th leading eigenvalues of $\mathbf{A}$. Also, by Weyl's inequality and $\mathbf{S}_{K:K_o} = (0)$,

$$\left\|\hat{\mathbf{S}}_{K:K_o}\right\| = \left\|\hat{\mathbf{S}}_{K:K_o}-\mathbf{S}_{K:K_o}\right\| \leq \|\mathbf{A}-\mathbf{P}\| = O(\sqrt{N\rho_N})\ whp.$$

hence $\left\|\hat{\mathbf{U}}_{+,e}\right\| = \left\|\hat{\mathbf{S}}_{K:K_o}\right\|^{1/2} = O(\sqrt[4]{N\rho_N})$ *whp.* so by a similar result as item 0 of Lemma 28,

$$\hat{\Sigma}_{+,e} = \begin{bmatrix}\frac{1}{N\rho_N}|\hat{\mathbf{S}}_{K:K_o}| & \frac{1}{NT\sqrt{\rho_N}}\sum_t\hat{\mathbf{U}}^\intercal_{+,e}\hat{\mathcal{W}}_t\\ & \hat{\Sigma}_e\end{bmatrix}\overset{N\to\infty}{\Rightarrow}\begin{bmatrix}\mathbf{O}_{K_o-K} & (0)\\ & \Sigma_e\end{bmatrix}$$

which is easy to check as $\hat{\mathbf{U}}^\intercal_{+,e}\hat{\mathbf{U}}_{:K,e} = \mathbf{O}_{(K_o-K)\times K}$ and for any $\mathbf{h}\in B^{K_o-K}$, we have $\|\Phi_e\| = O(\sqrt{TN\rho_N})$ *as.* and $\left\|\mathbf{1}_T\otimes\hat{\mathbf{U}}_{+,e}\right\| = O(\sqrt[4]{T^2N\rho_N})$ *whp.* so

$$\begin{aligned}
\frac{1}{NT\sqrt{\rho_N}}\mathbf{h}^\intercal\sum_t\hat{\mathbf{U}}^\intercal_{+,e}\mathbf{y}_t &= \frac{1}{NT\sqrt{\rho_N}}\mathbf{h}^\intercal\left(\mathbf{1}_T\otimes\hat{\mathbf{U}}_{+,e}\right)^\intercal\Gamma_{\mathbf{y}}{}^{1/2}\tilde{\mathbf{y}}+O\left(\sqrt[4]{\frac{\rho_N}{N}}\right)\ whp.,\\
\mathbb{P}\left[\frac{1}{NT\sqrt{\rho_N}}\left|\mathbf{h}^\intercal\left(\mathbf{1}_T\otimes\hat{\mathbf{U}}_{+,e}\right)^\intercal\Gamma_{\mathbf{y}}{}^{1/2}\tilde{\mathbf{y}}\right|>v\right] &< 2\exp\left(-cNT\sqrt{N\rho_N}v^2\right)
\end{aligned}$$

for example.

Claim C.6.5 *For any* $\mathbf{h}_1\in B^{K_o-K-1}, \mathbf{h}_2\in B^{K-1}, \mathbf{h}_3\in B^1, \mathbf{h}_4\in B^{p-1}$,

$$\frac{1}{\sqrt{NT}}\operatorname{diag}\left(\mathbf{h}_1,\ldots,\mathbf{h}_4\right)^\intercal\hat{\mathcal{W}}_+^\intercal(\mathcal{W}^H-\hat{\mathcal{W}})\mu^H_{w,e} = \begin{bmatrix}O\left(\frac{\sqrt{KT}}{\sqrt[4]{N\rho_N}}\right)\ whp.\\ O\left(\sqrt{KT}\right)\ whp.\\ O\left(\sqrt{KT\rho_N}\right)\ whp.\\ O_{\mathbb{P}}(1)\end{bmatrix}.$$

Proof Now observe that

$$\frac{1}{\sqrt{NT}}\hat{\mathcal{W}}_+^\intercal\left(\mathcal{W}^H-\hat{\mathcal{W}}\right)\mu_{w,e}^H=-\frac{1}{\sqrt{NT}}\begin{bmatrix}\frac{1}{\sqrt{\rho_N}}\left(\mathbf{1}_T\otimes\hat{\mathbf{U}}_{+,e}\right)^\intercal\\ \hat{\mathcal{W}}^\intercal\end{bmatrix}\mathbf{1}_T\otimes\left(\hat{\mathbf{U}}_{:K,e}-\mathbf{V}\mathbf{S}^{1/2}\mathbf{H}_e\right)\mathbf{H}_e{}^\intercal\beta_e.$$

Then, our additional bias term of interest is

$$\sqrt{\frac{T}{N\rho_N}}\left(\hat{\mathbf{U}}_{+,e}\mathbf{h}\right)^\intercal\left(\hat{\mathbf{U}}_{:K,e}-\mathbf{V}\mathbf{S}^{1/2}\mathbf{H}_e\right)\mathbf{H}_e{}^\intercal\beta_e=O\left(\sqrt[4]{\frac{K^2T^2}{N\rho_N}}\right)\ whp.$$

for any $\mathbf{h}\in B^{K_o-K-1}$, which leads to the conclusion by the observation made in the proof of the second claim in proving Theorem 4.

Claim C.6.6 $\sqrt{NT}\mathbf{r}_{w\epsilon,+}\Rightarrow\mathcal{N}(\mathbf{o}_{K_o+p+2},\sigma^2\Sigma_{+,e})$.

Proof This is trivial from the third claim in the proof of Theorem 4.

Therefore, we have the conclusion. ■

Proof [of Theorem 9] Note that $K_o<K$. Similarly to the proof of Theorem 8, deinfe a diagonal matrix of convergence rates $\mathbf{D}_{-,e}\triangleq\sqrt{NT}\,\mathrm{diag}\left(\sqrt{\rho_N}\mathbf{I}_{K_o},\mathbf{I}_{p+2}\right)$ and a truncated design matrix with $\hat{\mathbf{U}}_{:K_o,e}$ embedded over the entire time,

$$\hat{\mathcal{W}}_-\triangleq\left[\frac{1}{\sqrt{\rho_N}}\mathbf{1}_T\otimes\hat{\mathbf{U}}_{:K_o,e}\,|\,\mathbf{y},\left(\mathbf{I}_T\otimes\mathcal{L}_{\mathbf{A}}\right)\mathbf{y}\,|\,\mathbf{Z}\right]\in\mathbb{R}^{NT\times(K_o+p+2)}.$$

Let $\hat{\mathcal{W}}_{:K}\in\mathbb{R}^{(K+p+2)\times NT}$ denote the augmented version of $\hat{\mathcal{W}}_-$ which contains zeros between first and second block of $\hat{\mathcal{W}}_-$ above:

$$\hat{\mathcal{W}}_{:K}\triangleq\left[\frac{1}{\sqrt{\rho_N}}\mathbf{1}_T\otimes\hat{\mathbf{U}}_{:K_o,e}\,|\mathbf{O}_{NT\times(K-K_o)}\,|\,\mathbf{y},\left(\mathbf{I}_T\otimes\mathcal{L}_{\mathbf{A}}\right)\mathbf{y}\,|\,\mathbf{Z}\right]\in\mathbb{R}^{NT\times(K+p+2)}.$$

Partition $\mathbf{H}_e=[\mathbf{H}_{:K_o,e}\,|\,\mathbf{H}_{K_o:K,e}]$ and define $\mu_e^{H-}\triangleq\left(\beta_e{}^\intercal\mathbf{H}_{:K_o,e},\mu_{-\beta,e}^\intercal\right)^\intercal\in\mathbb{R}^{K_o+p+2}$ and $\mu_{w,e}^{H-}\triangleq\left(\sqrt{\rho_N}\beta_e{}^\intercal\mathbf{H}_{:K_o,e},\mu_{-\beta,e}^\intercal\right)^\intercal$. This expression gives $\sqrt{NT}\hat{\mathbf{W}}_-=\hat{\mathcal{W}}_-\mathbf{D}_{-,e}$ and $\hat{\mathcal{W}}_{:K}\mu_{w,e}^H=\hat{\mathcal{W}}_-\mu_{w,e}^{H-}$. $\hat{\Sigma}_{-,e}$ and $\mathbf{r}_{w\epsilon,-}$ are defined analogously as the case of over-selection. Then, we can derive the following in the same way as the previous Theorems:

$$\mathbf{D}_{-,e}\left(\hat{\mu}_{-,e}-\mu_e^{H-}\right)=\frac{\hat{\Sigma}_{-,e}^{-1}}{\sqrt{NT}}\hat{\mathcal{W}}_-^\intercal\left(\mathcal{W}^H-\hat{\mathcal{W}}_{:K}\right)\mu_{w,e}^H+\sqrt{NT}\hat{\Sigma}_{-,e}^{-1}\mathbf{r}_{w\epsilon,-}.$$

Let $\Sigma_{-,e}$ be a version of Σ_e which misses $K-K_o$ rows and columns corresponding to $K_o+1,\ldots,K$-th latent variables.

Next, by Proposition 23 there exists $\mathbf{H}_e\in\mathcal{O}_K$ such that

$$\left\|\hat{\mathbf{U}}_{:K,e}-\mathbf{V}\mathbf{S}^{1/2}\mathbf{H}_e\right\|=O\left(1\right)\ whp.$$

Therefore, there exists $\mathbf{H}_{:K_o,e}\in\mathcal{O}_{K,K_o}$ such that

$$\left\|\hat{\mathbf{U}}_{:K_o,e}-\mathbf{V}\mathbf{S}^{1/2}\mathbf{H}_{:K_o,e}\right\|\leq\left\|\hat{\mathbf{U}}_{:K_o,e}-\mathbf{V}\mathbf{S}^{1/2}\mathbf{H}_{:K_o,e}\right\|+\|\mathbf{I}_{K_o}-\mathbf{H}_{:K_o,e}{}^\intercal\mathbf{H}_{:K_o,e}\|=O(\sqrt{N\rho_N})\,whp.$$

by Theorem 7.

Claim C.6.7 *For any* $\mathbf{h} \in B^{K_o+p+1}$, $\mathbf{h}^\intercal \left(\hat{\Sigma}_{-,e} - \Sigma_{-,e}\right) \mathbf{h} = O_{\mathbb{P}}(1)$.

Proof Now we only need to show the following:

1. $\mathbb{P}\left(\frac{1}{NT\sqrt{\rho_N}} \left|\mathbf{h}_1{}^\intercal \sum_t \left\{\hat{\mathbf{U}}_{:K_o,e}^\intercal \mathbf{y}_t - \left(\mathbf{V}\mathbf{S}^{1/2}\mathbf{H}_{:K_o,e}\right)^\intercal \varphi_e\right\}\right| > \upsilon\right) \lesssim e^{-d_0\upsilon^2} + 1/N$
2. $\mathbb{P}\left[\frac{1}{NT\sqrt{\rho_N}} \left|\mathbf{h}_1{}^\intercal \sum_t \left\{\hat{\mathbf{U}}_{:K_o,e}^\intercal \mathcal{L}_{\mathbf{A}} \mathbf{y}_t - \left(\mathbf{V}\mathbf{S}^{1/2}\mathbf{H}_{:K_o,e}\right)^\intercal \mathcal{L}_{\mathbf{A}} \varphi_e\right\}\right| > \upsilon\right] \lesssim e^{-d_0\upsilon^2} + 1/N$
3. $\mathbb{P}\left(\frac{1}{NT\sqrt{\rho_N}} \left|\sum_t \mathbf{h}_1{}^\intercal \hat{\mathbf{U}}_{:K_o,e}^\intercal \mathbf{Z}_t \mathbf{h}_2\right| > \upsilon\right) \lesssim e^{-NT\upsilon^2} + 1/N$

where $d_0 \triangleq c_0/\rho_N$ for a constant $c_0 > 0$ and $\upsilon = \omega\left(\sqrt{\rho_N}\right)$. These can be easily shown by a similar proof to 1–3 of Lemma 28. Also, we have $\frac{1}{N\sqrt{\rho_N}} \left(\mathbf{V}\mathbf{S}^{1/2}\mathbf{H}_{:K_o,e}\mathbf{h}_1\right)^\intercal \varphi_e = \Theta(\sqrt{\rho_N})$ *a.s.* and $\frac{1}{N\sqrt{\rho_N}} \left(\mathbf{V}\mathbf{S}^{1/2}\mathbf{H}_{:K_o,e}\mathbf{h}_1\right)^\intercal \mathcal{L}_{\mathbf{A}}\varphi_e = \Theta\left(\sqrt{\rho_N}\right)$. Therefore, according to the same argument as Lemma 29, due to 1–2, we no longer obtain convergence in probability but boundedness in probability. If $K_o = 0$, then we have the reduced asymptotic covariance as Σ_{-u}.

Next, we analyze the entry-wise asymptotic orders of $\frac{1}{\sqrt{NT}}\hat{\mathcal{W}}_-^\intercal(\mathcal{W}^H - \hat{\mathcal{W}}_{:K})\mu_{w,e}^H$. First note that $\mathbf{H}_{:K_o,e}\mathbf{H}_{:K_o,e}{}^\intercal + \mathbf{H}_{K_o:K,e}\mathbf{H}_{K_o:K,e}{}^\intercal = \mathbf{I}_K$ as $\mathbf{H}_e \in \mathcal{O}_K$ and

$$\mathcal{W}^H - \hat{\mathcal{W}}_{:K} = \frac{1}{\sqrt{\rho_N}} \mathbf{1}_T \otimes \left[\mathbf{V}\mathbf{S}^{1/2}\mathbf{H}_{:K_o,e} - \hat{\mathbf{U}}_{:K_o,e} \,|\, \mathbf{V}\mathbf{S}^{1/2}\mathbf{H}_{K_o:K,e} \,|\, \mathbf{O}_{N\times(p+2)}\right].$$

Then, $\frac{1}{\sqrt{NT}}\hat{\mathcal{W}}_-^\intercal \left(\mathcal{W}^H - \hat{\mathcal{W}}_{:K}\right)\mu_{w,e}^H$ equals

$$\begin{aligned}
&\frac{1}{\sqrt{NT}} \sum_t \hat{\mathcal{W}}_{t,-}^\intercal \left\{\left(\mathbf{V}\mathbf{S}^{1/2}\mathbf{H}_{:K_o,e} - \hat{\mathbf{U}}_{:K_o,e}\right)\mathbf{H}_{:K_o,e}{}^\intercal + \mathbf{V}\mathbf{S}^{1/2}\mathbf{H}_{K_o:K,e}\mathbf{H}_{K_o:K,e}{}^\intercal\right\}\beta_e \\
&\qquad = \begin{bmatrix} \sqrt{\frac{T}{N\rho_N}}\hat{\mathbf{U}}_{:K_o,e}^\intercal \\ \frac{1}{\sqrt{NT}}\sum_t \mathbf{y}_t{}^\intercal \\ \frac{1}{\sqrt{NT}}\sum_t \mathbf{y}_t{}^\intercal \mathcal{L}_{\mathbf{A}} \\ \frac{1}{\sqrt{NT}}\sum_t \mathbf{Z}_t{}^\intercal \end{bmatrix} \left(\mathbf{V}\mathbf{S}^{1/2} - \hat{\mathbf{U}}_{:K_o,e}\mathbf{H}_{:K_o,e}{}^\intercal\right)\beta_e.
\end{aligned}$$

Denoting the RHS by $\mathbf{q}(K_o)$ which depends on K_o, we have $\mathbf{q}(0) = \mathbf{V}\mathbf{S}^{1/2}\beta_e$ for $K_o = 0$ as $\hat{\mathbf{U}}_{:K_o,e}\mathbf{H}_{:K_o,e}{}^\intercal$ is removed. So, $\|\mathbf{q}(K_o)\| = O(\sqrt{N\rho_N})$ *whp.* by noting that

$$\left\|\mathbf{V}\mathbf{S}^{1/2} - \hat{\mathbf{U}}_{:K_o,e}\mathbf{H}_{:K_o,e}{}^\intercal\right\| = \left\|\hat{\mathbf{U}}_{:K_o,e} - \mathbf{V}\mathbf{S}^{1/2}\mathbf{H}_{:K_o,e}\right\|.$$

Then, we can easily show the following by similar arguments used in proving 10–13 of Lemma 28:

1. $\sqrt{\frac{T}{N\rho_N}} \left|\left(\hat{\mathbf{U}}_{:K_o,e}\mathbf{h}_1\right)^\intercal \mathbf{q}(K_o)\right| = O\left(\sqrt{TN\rho_N}\right)$ *whp.*
2. $\frac{1}{\sqrt{NT}} \left|\sum_t \mathbf{y}_t{}^\intercal \mathbf{q}(K_o)\right| = O\left(\sqrt{N}\rho_N\right)$ *whp.*
3. $\frac{1}{\sqrt{NT}} \left|\sum_t \mathbf{y}_t{}^\intercal \mathcal{L}_{\mathbf{A}}\mathbf{q}(K_o)\right| = O\left(\sqrt{N}\rho_N\right)$ *whp.*
4. $\mathbb{P}\left[\frac{1}{\sqrt{NT}} \left|(\sum_t \mathbf{Z}_t\mathbf{h}_2)^\intercal \mathbf{q}(K_o)\right| > \upsilon\right] < e^{-c\upsilon^2/\rho_N}$

for any $\upsilon > 0$ where $\mathbf{h}_1 \in B^{K_o-1}, \mathbf{h}_2 \in B^{p-1}$. Now partition $\frac{1}{\sqrt{NT}}\hat{\mathcal{W}}_-^\intercal \left(\mathcal{W}^H - \hat{\mathcal{W}}_{:K}\right)\mu_{w,e}^H$ into three sub-vectors of sizes K_o, 2, and p denoted by $\mathbf{b}_1, \mathbf{b}_2$, and $\mathbf{b}_3$ respectively. These give the specified asymptotic orders of the estimation error terms.

Now, using the same steps in the third claim in the proof of Theorem 4 and recalling Claim C.6.7, we have $\frac{1}{\sqrt{NT}}\left\|\hat{\mathcal{W}}_-\eta\right\| = O_{\mathbb{P}}(1)$ for any $\eta \in B^{K_o+p+1}$. So, we can derive that $\sqrt{NT}\eta^\intercal \mathbf{r}_{w\epsilon,-} = O_{\mathbb{P}}(1)$ as for any $\epsilon > 0$ we have $N_0, T_0 > 0$ such that if $N > N_0, T > T_0$, by Lemma 21:

$$\mathbb{P}\left(\frac{1}{\sqrt{NT}}\left|\left(\hat{\mathcal{W}}_-\eta\right)^\intercal \mathcal{E}\right| > \upsilon\right) < 2\exp\left(-c\upsilon^2\right) + \epsilon$$

for each $\upsilon > 0$. Therefore, recalling the rates of $\mathbf{D}_{-,e}$, we can summarize as follows:

$$\sqrt{N\rho_N}\left(\hat{\beta}_{-,e} - \mathbf{H}_{:K_o,e}{}^\intercal \beta_e\right) = \mathbf{b}_1 + \mathbf{t}_1, \quad \sqrt{N}\begin{bmatrix}\hat{\alpha}_{-,e} - \alpha_e \\ \hat{\theta}_{-,e} - \theta_e\end{bmatrix} = \mathbf{b}_2 + \mathbf{t}_2,$$
$$\sqrt{N}\left(\hat{\gamma}_{-,e} - \gamma_e\right) = \mathbf{b}_3 + \mathbf{t}_3$$

where

$$\mathbf{h}_1{}^\intercal \mathbf{b}_1 = O\left(\sqrt{N\rho_N}\right) \ whp., \quad \mathbf{h}_2{}^\intercal \mathbf{b}_2 = O\left(\sqrt{N}\rho_N\right) \ whp., \quad \mathbf{h}_3{}^\intercal \mathbf{b}_3 = O_{\mathbb{P}}\left(\sqrt{\rho_N}\right),$$
$$\mathbf{h}_1{}^\intercal \mathbf{t}_1 = O_{\mathbb{P}}(1), \mathbf{h}_2{}^\intercal \mathbf{t}_2 = O_{\mathbb{P}}(1), \mathbf{h}_3{}^\intercal \mathbf{t}_3 = O_{\mathbb{P}}(1)$$

for any $\mathbf{h}_1 \in B^{K_o-1}, \mathbf{h}_2 \in B^1, \mathbf{h}_3 \in B^{p-1}$. ■

C.6.3 Concentration inequalities

This section lists analogous results to the results in the Lemma 25 for the case of ENAR with spectral embedding. Although we keep track of T here, we assume that T is finite when applying the following Lemma.

Lemma 28 *Let $\mathbf{h}_i \in \mathbb{R}^{j_i}$ for $j_1 = K$, $j_2 = p$ be real vectors such that $\|\mathbf{h}_i\| \leq 1$. Then under the conditions of Theorem 4, there exist $N_0 > 0$ and $\upsilon_0(N_0) > 0$ such that the statements 0–13 hold for all $\upsilon \in (0, \upsilon_0)$ and $N > N_0$:*

0. $\mathbb{P}\left(\frac{1}{N\rho_N}\left|\mathbf{h}_1{}^\intercal\left(|\hat{\mathbf{S}}| - \mathbf{S}\right)\mathbf{h}_1\right| > \upsilon\right) < 1/N$
1. $\mathbb{P}\left(\frac{1}{NT\sqrt{\rho_N}}\left|\mathbf{h}_1{}^\intercal \sum_t (\hat{\mathbf{U}}_e^\intercal \mathbf{y}_t - \tilde{\mathbf{U}}_e^\intercal \varphi_e)\right| > \upsilon\right) \lesssim e^{-d_0\upsilon^2} + 1/N$
2. $\mathbb{P}\left[\frac{1}{NT\sqrt{\rho_N}}\left|\mathbf{h}_1{}^\intercal \sum_t \left(\hat{\mathbf{U}}_e^\intercal \mathcal{L}_{\mathbf{A}}\mathbf{y}_t - \tilde{\mathbf{U}}_e^\intercal \mathcal{L}_{\mathbf{A}}\varphi_e\right)\right| > \upsilon\right] \lesssim e^{-d_0\upsilon^2} + 1/N$
3. $\mathbb{P}\left(\frac{1}{NT\sqrt{\rho_N}}\left|\sum_t \mathbf{h}_1{}^\intercal \hat{\mathbf{U}}_e^\intercal \mathbf{Z}_t \mathbf{h}_2\right| > \upsilon\right) \lesssim e^{-d_0\upsilon^2} + 1/N$
4. $\mathbb{P}\left[\frac{1}{NT}\left|\sum_t \left\{\mathbf{y}_t{}^\intercal \mathbf{y}_t - \mathbb{E}\,\mathrm{tr}\,(\Gamma) - \varphi_e{}^\intercal \varphi_e\right\}\right| > \upsilon\right] \lesssim e^{-d_0\upsilon^2} + e^{-d_1\upsilon^2}$
5. $\mathbb{P}\left[\frac{1}{NT}\left|\sum_t \left\{\mathbf{y}_t{}^\intercal \mathcal{L}_{\mathbf{A}}\mathbf{y}_t - \mathbb{E}\,\mathrm{tr}\,(\mathcal{L}_{\mathbf{A}}\Gamma) - \varphi_e{}^\intercal \mathcal{L}_{\mathbf{A}}\varphi_e\right\}\right| > \upsilon\right] \lesssim e^{-d_0\upsilon^2} + e^{-d_1\upsilon^2}$

6. $\mathbb{P}\left(\frac{1}{NT}\left|\sum_t \mathbf{y}_t{}^\intercal \mathbf{Z}_t \mathbf{h}_2\right| > \upsilon\right) \lesssim e^{-d_2\upsilon^2} + e^{-d_1\upsilon^2} + 1/N$

7. $\mathbb{P}\left[\frac{1}{NT}\left|\sum_t \left\{\mathbf{y}_t{}^\intercal \mathcal{L}_{\mathbf{A}}^2 \mathbf{y}_t - \mathbb{E}\,\mathrm{tr}\left(\mathcal{L}_{\mathbf{A}}^2 \Gamma\right) - \varphi_e{}^\intercal \mathcal{L}_{\mathbf{A}}^2 \varphi_e\right\}\right| > \upsilon\right] \lesssim e^{-d_0\upsilon^2} + e^{-d_1\upsilon^2}$

8. $\mathbb{P}\left(\frac{1}{NT}\left|\sum_t \mathbf{y}_t{}^\intercal \mathcal{L}_{\mathbf{A}} \mathbf{Z}_t \mathbf{h}_2\right| > \upsilon\right) \lesssim e^{-d_2\upsilon^2} + e^{-d_1\upsilon^2} + 1/N$

9. $\mathbb{P}\left[\frac{1}{NT}\left|\mathbf{h}_2{}^\intercal \sum_t \left(\mathbf{Z}_t{}^\intercal \mathbf{Z}_t - \Sigma_z\right) \mathbf{h}_2\right| > \upsilon\right] \lesssim e^{-d_0\upsilon^2}$

10. $\frac{\sqrt{T}}{\sqrt{N\rho_N}}\left|\mathbf{h}_1{}^\intercal \hat{\mathbf{U}}_e^\intercal \left(\hat{\mathbf{U}}_e - \tilde{\mathbf{U}}_e\right) \mathbf{H}_e^\intercal \beta_e\right| = O(\sqrt{T\pi_e})$ *whp.*

11. $\frac{1}{\sqrt{NT}}\left|\sum_t \mathbf{y}_t{}^\intercal \left(\hat{\mathbf{U}}_e - \tilde{\mathbf{U}}_e\right) \mathbf{H}_e^\intercal \beta_e\right| = O\left(\sqrt{\pi_e T \rho_N}\right)$ *whp.*

12. $\frac{1}{\sqrt{NT}}\left|\sum_t \mathbf{y}_t{}^\intercal \mathcal{L}_{\mathbf{A}} \left(\hat{\mathbf{U}}_e - \tilde{\mathbf{U}}_e\right) \mathbf{H}_e^\intercal \beta_e\right| = O\left(\sqrt{\pi_e T \rho_N}\right)$ *whp.*

13. $\mathbb{P}\left[\frac{1}{\sqrt{NT}}\left|\mathbf{h}_2{}^\intercal \sum_t \mathbf{Z}_t{}^\intercal \left(\hat{\mathbf{U}}_e - \tilde{\mathbf{U}}_e\right) \mathbf{H}_e^\intercal \beta_e\right| > \upsilon\right] \lesssim e^{-d_3\upsilon^2} + 1/N$

where $d_0 \triangleq c_0 NT$, $d_1 \triangleq \frac{c_1 NT}{\rho_N}$, $d_2 \triangleq \frac{c_2 NT}{p}$, *and* $d_3 = \frac{c_3 N}{K}$.

Proof [of 0] First, $\left\||\hat{\mathbf{S}}| - \mathbf{S}\right\| = O(\log N)$ *whp.* by inhomogeneous Erdős–Rényi model singular value perturbation bounds Cape et al. (2017) and $N\rho_N = \omega\left(\log^4 N\right)$. Therefore,

$$\mathbb{P}\left(\frac{1}{N\rho_N}\left|\mathbf{h}_1{}^\intercal \left(|\hat{\mathbf{S}}| - \mathbf{S}\right) \mathbf{h}_1\right| > \upsilon\right) \leq \mathbb{P}\left(\left|\mathbf{h}_1{}^\intercal \left(|\hat{\mathbf{S}}| - \mathbf{S}\right) \mathbf{h}_1\right| > \frac{C \log N}{N\rho_N}\right) < 1/N$$

for some constant $C > 0$ and any $\upsilon \geq \frac{C \log N}{N\rho_N}$. So we have the desired bound for large enough N. ■

Proof [of 1] Note that

$$\begin{aligned}\sum_{t=1}^{T} \mathbf{h}_1{}^\intercal \left(\hat{\mathbf{U}}_e^\intercal \mathbf{y}_t - \tilde{\mathbf{U}}_e^\intercal \varphi_e\right) &= \left\{\left(\mathbf{1}_T \otimes \hat{\mathbf{U}}_e\right) \mathbf{h}_1\right\}^\intercal {\Gamma_{\mathbf{y}}}^{1/2} \tilde{\mathbf{y}} + T\left(\Delta_e \mathbf{h}_1\right)^\intercal \varphi_e \\ &= \left(\hat{\mathbf{S}} \mathbf{h}_1\right)^\intercal \tilde{\mathbf{y}} + T\left(\Delta_e \mathbf{h}_1\right)^\intercal \varphi_e\end{aligned}$$

where $\hat{\mathbf{S}} \triangleq {\Gamma_{\mathbf{y}}}^{1/2}\left(\mathbf{1}_T \otimes \hat{\mathbf{U}}_e\right)$. Since $\left\|\hat{\mathbf{S}}\mathbf{h}_1\right\| = O\left(\sqrt{TN\rho_N}\right)$ *whp.*, by Lemma 20 we have

$$\mathbb{P}^*\left(\left|\left(\hat{\mathbf{S}}\mathbf{h}_1\right)^\intercal \tilde{\mathbf{y}}\right| > NT\sqrt{\rho_N}\upsilon\right) < 2\exp\left(\frac{-cN^2T^2\rho_N\upsilon^2}{TN\rho_N}\right) + 1/N.$$

Since the RHS does not depend on $\mathbf{A}$, we can replace $\mathbb{P}^*$ with $\mathbb{P}$. Next, since $\|\varphi_e\| = O(\sqrt{N\rho_N})$ $\mathbb{P}^* - as.$ (hence is $\mathbb{P} - as.$), $|(\Delta_e \mathbf{h}_1)^\intercal \varphi_e| = O(\sqrt{\pi_e N \rho_N})$ *whp.*, i.e., we have $\frac{T}{NT\sqrt{\rho_N}}\left|(\Delta_e \mathbf{h}_1)^\intercal \varphi_e\right| = O\left(\sqrt{\pi_e/N}\right)$ *whp.* We can obtain the desired bound using the same logic as 0, provided that $N = \omega(\pi_e)$. ■

Proof [of 2]

$$\mathbf{h}_1{}^\intercal \sum_t \left(\hat{\mathbf{U}}_e^\intercal \mathcal{L}_{\mathbf{A}} \mathbf{y}_t - \tilde{\mathbf{U}}_e^\intercal \mathcal{L}_{\mathbf{A}} \varphi_e\right) = \mathbf{h}_1{}^\intercal \left\{\mathbf{1}_T \otimes \left(\mathcal{L}_{\mathbf{A}} \hat{\mathbf{U}}_e\right)\right\}^\intercal \Gamma_{\mathbf{y}}{}^{1/2} \tilde{\mathbf{y}} + T \left(\Delta_e \mathbf{h}_1\right)^\intercal \mathcal{L}_{\mathbf{A}} \varphi_e$$
$$= \mathbf{h}_1{}^\intercal \hat{\mathbf{S}}^\intercal \tilde{\mathbf{y}} + T \left(\Delta_e \mathbf{h}_1\right)^\intercal \mathcal{L}_{\mathbf{A}} \varphi_e.$$

For $\hat{\mathbf{S}} \triangleq \Gamma_{\mathbf{y}}{}^{1/2} \left\{\mathbf{1}_T \otimes \left(\mathcal{L}_{\mathbf{A}} \hat{\mathbf{U}}_e\right)\right\}$, we have $\left\|\hat{\mathbf{S}} \mathbf{h}_1\right\| = O(\sqrt{TN\rho_N})$ *whp.* because $\rho\left(\mathcal{L}_{\mathbf{A}}\right) = O(1)$ *as.* The tail probability of $\left|T \left(\Delta_e \mathbf{h}_1\right)^\intercal \mathcal{L}_{\mathbf{A}} \varphi_e\right|$ can be bounded similarly to 1. ■

Proof [of 3] Let $\mathbf{Z} \triangleq [\mathbf{Z}_0{}^\intercal, ..., \mathbf{Z}_{T-1}{}^\intercal]^\intercal$ and $\hat{\mathbf{S}} \triangleq \mathbf{1}_T \otimes \hat{\mathbf{U}}_e$. Then, we can express $\sum_t \mathbf{h}_1{}^\intercal \hat{\mathbf{U}}_e^\intercal \mathbf{Z}_t \mathbf{h}_2$ as $\operatorname{tr}\left(\mathbf{Z}\mathbf{h}_2\mathbf{h}_1{}^\intercal \hat{\mathbf{S}}^\intercal\right)$. This is equal to $\operatorname{vec}\left(\mathbf{h}_2\mathbf{h}_1{}^\intercal\right)^\intercal \left(\mathbf{I}_p \otimes \hat{\mathbf{S}}^\intercal\right) \operatorname{vec}\left(\mathbf{Z}\right)$ and note that $\left\|\mathbf{I}_p \otimes \hat{\mathbf{S}}^\intercal\right\| = O\left(\sqrt{TN\rho_N}\right)$ *whp.* Therefore, by Lemma 20, we have

$$\mathbb{P}\left[\left|\sum_t \mathbf{h}_1{}^\intercal \hat{\mathbf{U}}_e^\intercal \mathbf{Z}_t \mathbf{h}_2\right| > NT\sqrt{\rho_N}\upsilon\right] \leq 2\exp\left(-\frac{cN^2T^2\rho_N\upsilon^2}{TN\rho_N}\right) + 1/N.$$

■

Proof [of 4] Now we observe that

$$\sum_t \left(\mathbf{y}_t{}^\intercal \mathbf{y}_t - \varphi_e^\intercal \varphi_e\right) = \tilde{\mathbf{y}}^\intercal \Gamma_{\mathbf{y}} \tilde{\mathbf{y}} + 2\Phi_e^\intercal \Gamma_{\mathbf{y}}{}^{1/2} \tilde{\mathbf{y}}.$$

Note that $\|\Gamma_{\mathbf{y}}\|_F^2 = O(NT)$ *as.* By Lemma 22, we have

$$\mathbb{P}\left(\left|\tilde{\mathbf{y}}^\intercal \Gamma_{\mathbf{y}} \tilde{\mathbf{y}} - T \operatorname{tr} \mathbb{E}\Gamma\right| > \upsilon\right) < 2\exp\left(-c\min\left\{\frac{\upsilon^2}{NT}, \upsilon\right\}\right).$$

By Lemma 20,

$$\mathbb{P}\left(2|\Phi_e^\intercal \Gamma_{\mathbf{y}}{}^{1/2} \tilde{\mathbf{y}}| > \upsilon\right) < 2\exp\left(-\frac{c'\upsilon^2}{TN\rho_N}\right)$$

since $\left\|\Gamma_{\mathbf{y}}{}^{1/2}\Phi_e\right\|^2 \leq \|\Gamma_{\mathbf{y}}\| T\varphi_e^\intercal \varphi_e = O(TN\rho_N)$ *as.* ■

Proof [of 6] We have

$$\sum_t \mathbf{y}_t{}^\intercal \mathbf{Z}_t \mathbf{h}_2 = \left(\Gamma_{\mathbf{y}}{}^{1/2} \tilde{\mathbf{y}} + \Phi_e\right)^\intercal \mathbf{Z}\mathbf{h}_2 = \operatorname{tr}\left(\mathbf{h}_2 \tilde{\mathbf{y}}^\intercal \Gamma_{\mathbf{y}}{}^{1/2} \mathbf{Z}\right) + \operatorname{tr}\left(\mathbf{h}_2 \Phi_e{}^\intercal \mathbf{Z}\right)$$
$$= \operatorname{vec}\left(\tilde{\mathbf{y}}\mathbf{h}_2{}^\intercal\right)^\intercal \left(\mathbf{I}_p \otimes \Gamma_{\mathbf{y}}{}^{1/2}\right) \operatorname{vec}\left(\mathbf{Z}\right) + \operatorname{vec}\left(\Phi_e \mathbf{h}_2{}^\intercal\right)^\intercal \left(\mathbf{I}_p \otimes \mathbf{I}_{NT}\right) \operatorname{vec}\left(\mathbf{Z}\right).$$

Therefore, since $\|\Gamma_{\mathbf{y}}\| = O(1)$ *as.* , $\left\|\left(\mathbf{I}_p \otimes \Gamma_{\mathbf{y}}{}^{1/2}\right) \operatorname{vec}\left(\tilde{\mathbf{y}}\mathbf{h}_2{}^\intercal\right)\right\| \lesssim \left\|\operatorname{vec}\left(\tilde{\mathbf{y}}\mathbf{h}_2{}^\intercal\right)\right\|$ which is concentrated around $\sqrt{NTp}$ (Vershynin, 2018) hence is $O(\sqrt{NTp})$ *whp.* Also, we have

$$\left\|\left(\mathbf{I}_p \otimes \mathbf{I}_{NT}\right) \operatorname{vec}\left(\Phi_e \mathbf{h}_2{}^\intercal\right)\right\| = O(\sqrt{TN\rho_N}) \text{ as.}$$

By Lemma 20,

$$\mathbb{P}\left(\left|\sum_t \mathbf{y}_t{}^\intercal \mathbf{Z}_t \mathbf{h}_2\right| > \upsilon\right) \leq 2\exp\left(-\frac{c\upsilon^2}{NTp}\right) + 2\exp\left(-\frac{c\upsilon^2}{TN\rho_N}\right) + \frac{1}{N}.$$

■

Proof [of 10] Noting that $\|\mathbf{H}_e^\intercal \beta_e\| = O(1)$, we have $\left|\mathbf{h}_1{}^\intercal \hat{\mathbf{U}}_e^\intercal \Delta_e \mathbf{H}_e^\intercal \beta_e\right| = O\left(\sqrt{\pi_e N \rho_N}\right)$ *whp.* which implies $\sqrt{T/(N\rho_N)}\left|\mathbf{h}_1{}^\intercal \hat{\mathbf{U}}_e^\intercal \Delta_e \mathbf{H}_e^\intercal \beta_e\right| = O\left(\sqrt{T\pi_e}\right)$ *whp.* ■

Proof [of 11] Similarly to the proof of 1., taking $\hat{\mathbf{S}} \triangleq \Gamma_{\mathbf{y}}{}^{1/2}\left(\mathbf{1}_T \otimes \Delta_e\right)$ which has the spectral norm of $O\left(\sqrt{KT}\right)$ *whp.*, we have

$$\sum_t \mathbf{y}_t{}^\intercal \Delta_e \beta_e = \left(\Gamma_{\mathbf{y}}{}^{1/2}\tilde{\mathbf{y}} + \Phi_e\right)^\intercal \left(\mathbf{1}_T \otimes \Delta_e\right) \mathbf{H}_e^\intercal \beta_e = \tilde{\mathbf{y}}^\intercal \hat{\mathbf{S}} \mathbf{H}_e^\intercal \beta_e + \Phi_e{}^\intercal \Gamma_{\mathbf{y}}{}^{-\frac{1}{2}} \hat{\mathbf{S}} \mathbf{H}_e^\intercal \beta_e.$$

It is straightforward by noting that $\|\beta_e\| = O(1)$ as $K \to \infty$,

$$\mathbb{P}\left(\frac{1}{\sqrt{NT}}\left|\tilde{\mathbf{y}}^\intercal \hat{\mathbf{S}} \mathbf{H}_e^\intercal \beta_e\right| \geq \upsilon\right) \lesssim \exp\left(-cN\upsilon^2/K\right) + 1/N$$

and $\Phi_e{}^\intercal \Gamma_{\mathbf{y}}{}^{-\frac{1}{2}} \hat{\mathbf{S}} \mathbf{H}_e^\intercal \beta_e = O\left(T\sqrt{\pi_e N \rho_N}\right)$ *whp.* ■

Proof [of 13] Note that $\mathbf{h}_2{}^\intercal \sum_t \mathbf{Z}_t{}^\intercal \Delta_e \mathbf{H}_e^\intercal \beta_e = \mathbf{h}_2{}^\intercal \mathbf{Z}^\intercal \left(\mathbf{1}_T \otimes \Delta_e\right) \mathbf{H}_e^\intercal \beta_e$ which equals

$$\operatorname{tr}\left\{\mathbf{Z}\mathbf{h}_2 \beta_e^\intercal \mathbf{H}_e \left(\mathbf{1}_T \otimes \Delta_e\right)^\intercal\right\} = \operatorname{vec}\left(\mathbf{h}_2 \beta_e^\intercal \mathbf{H}_e\right)^\intercal \left\{\mathbf{I}_p \otimes \left(\mathbf{1}_T \otimes \Delta_e\right)^\intercal\right\} \operatorname{vec}\left(\mathbf{Z}\right).$$

Since the norm of $\frac{1}{\sqrt{NT}} \operatorname{vec}\left(\mathbf{h}_2 \beta_e^\intercal \mathbf{H}_e\right)^\intercal \left\{\mathbf{I}_p \otimes \left(\mathbf{1}_T \otimes \Delta_e\right)^\intercal\right\}$ is of $O\left(\sqrt{\pi_e/N}\right)$ *whp.*, we obtain the bound provided that $\upsilon = \omega\left(\sqrt{\pi_e/N}\right)$ and $N = \omega(\pi_e)$. ■

Lemma 29 *There exist $N_0, T_0 > 0$ and $\upsilon_0(N_0, T_0) \in (0,1)$ such that for all $\upsilon \in (0, \upsilon_0)$ and $N > N_0, T > T_0$,*

$$\mathbb{P}\left[\sup_{\mathbf{h} \in B^{K+p+1}} \left|\mathbf{h}^\intercal \left(\hat{\Sigma}_e - \Sigma_e\right) \mathbf{h}\right| > \upsilon\right] \lesssim K^2 \left\{\frac{1}{N} + \exp\left(-\frac{cNT\upsilon^2}{K^2 p}\right)\right\}.$$

Proof Let $\mathbf{h} \in B^{K+p+1}$ and $\mathbf{e}_i \in S^{K+p+1}$ be the i^{th} canonical basis of $\mathbb{R}^{K+p+2}$, i.e., $(\mathbf{e}_i)_j \triangleq \mathbb{I}_{\{i\}}(j)$ for $j = 1, \ldots, K+p+2$. By 0–9 of Lemma 28, we have

$$\max_{i,j} \mathbb{P}\left(\left|\mathbf{e}_i{}^\intercal \left(\hat{\Sigma}_e - \Sigma_e\right) \mathbf{e}_j\right| > \upsilon\right) \lesssim \frac{1}{N} + \exp\left(-c_1 \upsilon^2 NT/p\right).$$

Since

$$\sup_{\mathbf{h} \in B^{K+p+1}} \left|\mathbf{h}^\intercal \left(\hat{\Sigma}_e - \Sigma_e\right) \mathbf{h}\right| \lesssim K \max_{i,j} \left|\hat{\sigma}_{e,ij} - \sigma_{e,ij}\right|,$$

we have the conclusion. ■

Appendix D. Over-selection of K by $\hat{K}$

This is the main theorem related to not-under-selection guarantee with the general information criteria.

Theorem 30 *Assume that $K \leq J_N$ and $T = \omega(\log K)$. Let p_{NT} be a penalty such that $p_{NT}(J_N) = o(1)$ as $N, T \to \infty$. Then, we have*

$$\mathbb{P}\left[\liminf_{N,K,T\to\infty} \left\{\hat{K} \geq K\right\}\right] = 1.$$

Remark 31 *(Practical guidance on model selection) It is well known that $N\rho_N = \omega\left(\log^2 N\right)$ guarantees $\|\mathbf{A}\|_\infty \to \|\mathbf{P}\|_\infty$ as. (see Proposition A.2 of Tang et al., 2017 for example) where $\|\mathbf{P}\|_\infty = O(N\rho_N)$. Then, with an additional condition on K as $K = o(N\rho_N)$, we can choose J_N such that $J_N = O\left(\|\mathbf{A}\|_\infty\right)$ and $p_{NT}(k) = \frac{k}{N+T}$ for the search of K, for example. A second choice of penalty that may be more appropriate for larger T is $p_{NT}(k) = \frac{k^2}{N+T}$. With this penalty, when $T = \omega(N)$, we may choose $J_N = \sqrt{N}$, while for $T = o(N)$ we may choose any $J_N = o(\sqrt{N})$.*

Proof [of Theorem 30] Let $p_{l,NT}$ be a penalty function that can vary over $l \in \{v, e\}$. First, we see that $\hat{K} < K$ implies $Cr_l(k) \leq Cr_l(k')$ for some $k < K$ and all $k' \in \{K, \ldots, J_N\}$, i.e.,

$$\left\{\hat{K} < K\right\} \subseteq \bigcup_{k<K} \bigcap_{k'=K}^{J_N} \left\{Cr_l(k) \leq Cr_l(k')\right\}$$

so

$$\mathbb{P}\left(\hat{K} < k\right) \leq \sum_{k<K} \mathbb{P}\left[\cap_{k'\geq K} \left\{Cr_l(k) \leq Cr_l(k')\right\}\right] \leq K \max_{k<K} \max_{k'\geq K} \mathbb{P}\left[Cr_l(k) \leq Cr_l(k')\right].$$

Note that for any k such that $0 \leq k < K$ we have

$$\begin{aligned}
\max_{k'\geq K} \mathbb{P}\left[\left\{Cr_l(k) \leq Cr_l(k')\right\}\right] &= \max_{k'\geq K} \mathbb{P}\left[\frac{1}{T}\mathbf{y}^\intercal\left(\hat{\Pi}_l^{(k')} - \hat{\Pi}_l^{(k)}\right)\mathbf{y} \leq p_{l,NT}(k') - p_{l,NT}(k)\right] \\
&\leq \mathbb{P}\left[\frac{1}{T}\mathbf{y}^\intercal\left(\hat{\Pi}_l^{(K)} - \hat{\Pi}_l^{(k)}\right)\mathbf{y} \leq p_{l,NT}(J_N) - p_{l,NT}(k)\right], \\
&\max_{k<K} \mathbb{P}\left[\frac{1}{T}\mathbf{y}^\intercal\left(\hat{\Pi}_l^{(K)} - \hat{\Pi}_l^{(k)}\right)\mathbf{y} \leq p_{l,NT}(J_N) - p_{l,NT}(k)\right] \\
&\leq \mathbb{P}\left[\frac{1}{T}\mathbf{y}^\intercal\left(\hat{\Pi}_l^{(K)} - \hat{\Pi}_l^{(K-1)}\right)\mathbf{y} \leq p_{l,NT}(J_N) - p_{l,NT}(0)\right]
\end{aligned}$$

by the fact that $\hat{\Pi}_l^{(i)} \succcurlyeq \hat{\Pi}_l^{(j)}$ for any $i \geq j$. Let us denote $\hat{\Pi}_l^{(K)} - \hat{\Pi}_l^{(K-1)}$ by $\hat{\Pi}_{\Delta,l}$. $\mathbf{y}^\intercal\left(\hat{\Pi}_l^{(K)} - \hat{\Pi}_l^{(K-1)}\right)\mathbf{y}$ can be decomposed as

$$\begin{aligned}
\mathbf{y}^\intercal\hat{\Pi}_{\Delta,l}\mathbf{y} &= \tilde{\mathbf{y}}^\intercal{\Gamma_{\mathbf{y}}}^{1/2}\hat{\Pi}_{\Delta,l}{\Gamma_{\mathbf{y}}}^{1/2}\tilde{\mathbf{y}} + 2\tilde{\mathbf{y}}^\intercal{\Gamma_{\mathbf{y}}}^{1/2}\hat{\Pi}_{\Delta,l}{\Gamma_{\mathbf{y}}}^{1/2}\Phi_l + \Phi_l^\intercal{\Gamma_{\mathbf{y}}}^{1/2}\hat{\Pi}_{\Delta,l}{\Gamma_{\mathbf{y}}}^{1/2}\Phi_l \\
&\geq \Phi_l^\intercal{\Gamma_{\mathbf{y}}}^{1/2}\hat{\Pi}_{\Delta,l}{\Gamma_{\mathbf{y}}}^{1/2}\Phi_l - 2\left|\tilde{\mathbf{y}}^\intercal{\Gamma_{\mathbf{y}}}^{1/2}\hat{\Pi}_{\Delta,l}{\Gamma_{\mathbf{y}}}^{1/2}\Phi_l\right|.
\end{aligned}$$

Now we realize that for $\hat{\mathbf{U}}_{i:j,l}$ with $l = e, v$, $\hat{\mathbf{U}}_{i:j,l}{}^{\intercal}\hat{\mathbf{U}}_{i:j,l} = \hat{\Lambda}_{i:j,l}$ where $\hat{\Lambda}_{i:j,v} = \mathbf{I}_{j-i}$ and $\hat{\Lambda}_{i:j,e} = |\hat{\mathbf{S}}_{i:j}|$. So,

$$\begin{aligned}\hat{\Pi}_{\Delta,l} &= \frac{1}{T}\mathbf{J}_T \otimes \left(\hat{\mathbf{U}}_{:K,l}\hat{\Lambda}_{:K,l}\hat{\mathbf{U}}^{\intercal}_{:K,l} - \hat{\mathbf{U}}_{:(K-1),l}\hat{\Lambda}_{:(K-1),l}\hat{\mathbf{U}}^{\intercal}_{:(K-1),l}\right) \\ &= \frac{\hat{\lambda}_{K,l}}{T}\mathbf{J}_T \otimes \left(\hat{\mathbf{U}}_{K:K,l}\hat{\mathbf{U}}^{\intercal}_{K:K,l}\right), \quad \hat{\lambda}_{K,l} = \begin{cases} 1, & l = v \\ s_K(\mathbf{A}), & l = e \end{cases}\end{aligned}$$

where $\mathbf{J}_T \triangleq \mathbf{1}_T\mathbf{1}_T{}^{\intercal}$. This is a rank-one matrix. Let $\mathbf{v}_l \triangleq \mathbf{1}_T \otimes \hat{\mathbf{U}}_{K:K,l}$ so that $\hat{\Pi}_{\Delta,l} = \frac{\hat{\lambda}_{k,l}}{T}\mathbf{v}_l\mathbf{v}_l^{\intercal}$. Note that $\hat{\lambda}_{K,e} = \Theta(N\rho_N)$ *whp.* by Weyl's inequality and the magnitude of $\lambda_K(\mathbf{P})$. So, on $\left\{\hat{\lambda}_{K,l} \gtrsim 1\right\} \cap \left\{\Phi_l^{\intercal}\Gamma_{\mathbf{y}}^{1/2}\mathbf{v}_l \neq 0\right\}$, we have

$$\left(\Gamma_{\mathbf{y}}{}^{1/2}\Phi_l\right)^{\intercal}\hat{\Pi}_{\Delta,l}\Gamma_{\mathbf{y}}{}^{1/2}\Phi_l = \frac{\hat{\lambda}_{K,l}}{T}\left(\Phi_l{}^{\intercal}\Gamma_{\mathbf{y}}^{1/2}\mathbf{v}_l\right)^2 \gtrsim T\left\|\varphi_l\right\|^2 s_{NT}\left(\Gamma_{\mathbf{y}}\right)$$

by variational characterization of singular values. The event $\left\{\Phi_l^{\intercal}\Gamma_{\mathbf{y}}^{1/2}\mathbf{v}_l \neq 0\right\}$ happens almost surely as the distribution of the LHS is continuous. By Proposition 2.2 and 2.3 of Basu and Michailidis (2015) again, we have

$$s_{NT}(\Gamma_{\mathbf{y}}) \geq \frac{\sigma^2}{m_{\max}(\mathbf{G})} \geq \frac{\sigma^2}{\left(1 + \|\mathbf{G}\|\right)^2} \; as.$$

where $m_{\max}(\mathbf{G}) \triangleq \max_{\{z\in\mathcal{C};|z|=1\}}\left(\mathbf{I}_N - \mathbf{G}z\right)^*\left(\mathbf{I}_N - \mathbf{G}z\right)$. For the last lower bound, we refer to (i) of the proof of Proposition 2.2 in Basu and Michailidis (2015). Therefore, $s_{NT}\left(\Gamma_{\mathbf{y}}\right) > \frac{\sigma^2}{4}$ *as.* Next, noting that $\left\|\Gamma_{\mathbf{y}}{}^{1/2}\hat{\Pi}_{\Delta,l}\Gamma_{\mathbf{y}}{}^{1/2}\Phi_l\right\| \leq \|\Phi_l\|\,\|\Gamma_{\mathbf{y}}\| = O\left(\sqrt{T}\left\|\varphi_l\right\|\right)$ *as.*, we have

$$\mathbb{P}\left[2\left|\tilde{\mathbf{y}}^{\intercal}\Gamma_{\mathbf{y}}{}^{1/2}\hat{\Pi}_{\Delta,l}\Gamma_{\mathbf{y}}{}^{1/2}\Phi\right| > \upsilon\right] < 2\exp\left(-\frac{c\upsilon^2}{T\left\|\varphi_l\right\|^2}\right)$$

for any $\upsilon > 0$. So, denoting $p_{l,NT}(J_N) - p_{l,NT}(0)$ by $\Delta p_{l,NT}$,

$$\begin{aligned}&K\mathbb{P}\left[\mathbf{y}^{\intercal}\left(\hat{\Pi}_l^{(K)} - \hat{\Pi}_l^{(K-1)}\right)\mathbf{y} \leq T\Delta p_{l,NT}\right] \\ &\quad\leq K\mathbb{P}\left[\Phi_l^{\intercal}\Gamma_{\mathbf{y}}{}^{1/2}\hat{\Pi}_{\Delta,l}\Gamma_{\mathbf{y}}{}^{1/2}\Phi_l - T\Delta p_{l,NT} \leq 2\left|\tilde{\mathbf{y}}^{\intercal}\Gamma_{\mathbf{y}}{}^{1/2}\hat{\Pi}_{\Delta,l}\Gamma_{\mathbf{y}}{}^{1/2}\Phi\right|\right] \\ &\quad\leq K\mathbb{P}\left[T\upsilon_* \leq 2\left|\tilde{\mathbf{y}}^{\intercal}\Gamma_{\mathbf{y}}{}^{1/2}\hat{\Pi}_{\Delta,l}\Gamma_{\mathbf{y}}{}^{1/2}\Phi\right|\right] \leq 2K\exp\left(\frac{-cT\upsilon_*^2}{\left\|\varphi_l\right\|^2}\right)\end{aligned}$$

for $\upsilon_* \triangleq \frac{\sigma^2}{4}\left\|\varphi_l\right\|^2 - \Delta p_{l,NT}$ provided that there exist large enough N, T such that $\frac{\sigma^2}{4}\left\|\varphi_l\right\|^2 > \Delta p_{l,NT}$. Note that $\left\|\varphi_l\right\|^2 \gtrsim s_K(\mathbf{V}^{\intercal}\mathbf{V}) \geq 1$ and $\left\|\varphi_l\right\|^2 \leq \left\|\mathbf{U}_e\right\|^2 \lesssim N\rho_N$ hence

$$\frac{\upsilon_*}{\left\|\varphi_l\right\|} = \frac{\sigma^2}{4}\left\|\varphi_l\right\| - \frac{\Delta p_{l,NT}}{\left\|\varphi_l\right\|} \gtrsim 1 - \frac{p_{l,NT}(J_N)}{\sqrt{N\rho_N}}.$$

So if $p_{l,NT}(J_N) = o(1)$, then the conclusion follows by Borel-Cantelli Lemma provided that $T = \omega\left(\log K\right)$. ■

Appendix E. Additional Simulation and Real Data Results

E.1 More simulations

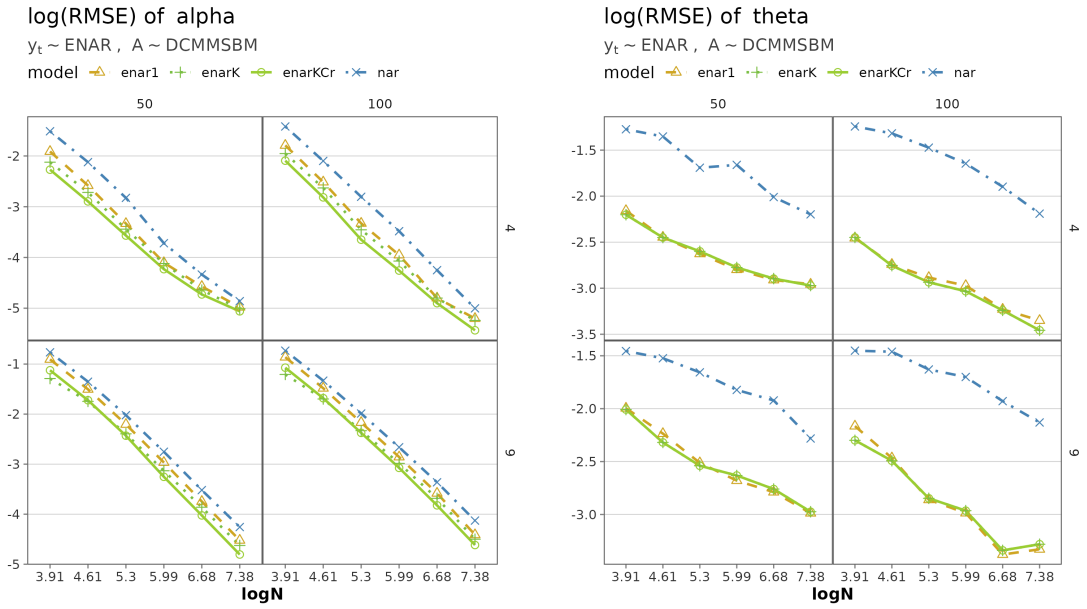


Figure C1: Plot of estimates of θ and α from ENAR with $\mathbf{U} = \mathbf{V}$ and increasing N and T, for $K_o \in \left\{1, K, \hat{K}\right\}$, and NAR model when data is generated from ENAR model with DCMMSBM.

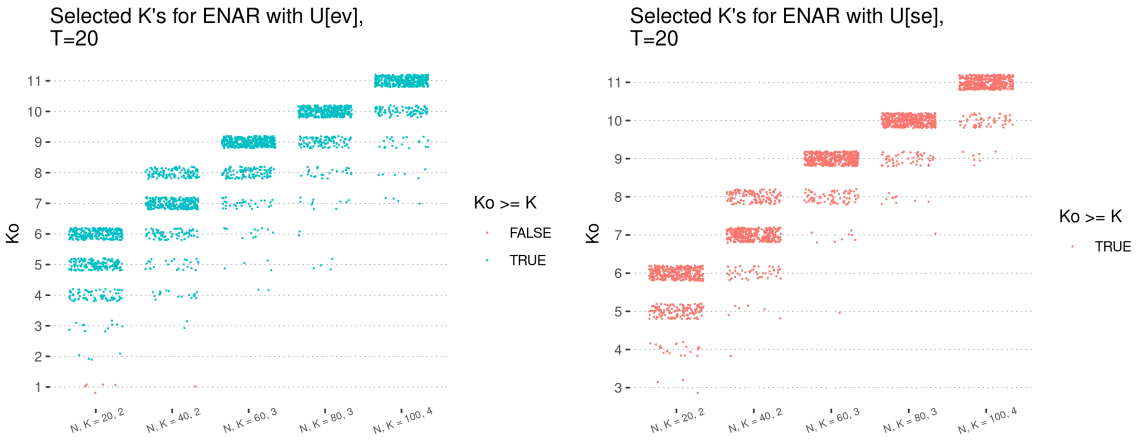


Figure C2: Jitter plots of selected $\hat{K}$ when $T = 20$ and N increases when $\mathbf{U} = \mathbf{V}$ (left) and $\mathbf{U} = \mathbf{U}_e$ (right).

E.1.1 Details of simulation setup for ENAR model

First, we used the matrix $2q\mathbf{I}_K + q\mathbf{1}_K\mathbf{1}_K^\intercal$ where $q = \frac{9}{40}$ to generate the $K \times K$ block matrix of connection probabilities. As a result, the ratio of inter-community and between-community connectivity is 3. The maximum expected degree for each graph was set to be $N\rho_N$ where $\rho_N \triangleq N^{-1/2}$, ensuring that the graphs are sparse. The degree heterogeneity parameters associated with both DCSBM and DCMMSBM were sampled from a standard log-normal distribution. For DCMMSBM, the (mixed) block memberships were generated from a Dirichlet distribution with parameter vector $(1, \ldots, 1)$. For DCSBM, the block memberships were sampled from a categorical distribution with equal probabilities. For model parametrization, we set the parameters β^{enar} associated with the latent effects as $\left(1, -1/2, ..., (-1)^{K-1}/K\right)^\intercal \in \mathbb{R}^K$. For the latent peer effect measured by the community structure of CNAR, we used $\mathcal{B}_1 \triangleq \frac{1}{10}\operatorname{diag}(\beta^{\text{enar}})$ where the definition of $\mathcal{B}_1$ comes from Chen et al. (2023). For ENAR, and NAR we set $\alpha = \theta = \frac{1}{5}$, and set the covariate effects as $\gamma = \left(\frac{1}{3}, -\frac{1}{6}, 0\right)^\intercal$ for all of considered models. Throughout simulations we generate the covariates $\mathbf{z}_{it}$ from $\mathcal{N}(\mathbf{o}_3, \operatorname{diag}(3,2,1))$ and $\mathcal{E}_t$ from $\mathcal{N}(\mathbf{o}_N, 0.25\,\mathbf{I}_N)$.

E.1.2 Misspecification of K

In Figure C1, we plot the RMSE plots for growing N and T with $\mathbf{U} = \mathbf{V}$ for $K_o \in \left\{1, K, \hat{K}\right\}$, and the NAR model when data is generated from the ENAR model with DCMMSBM.

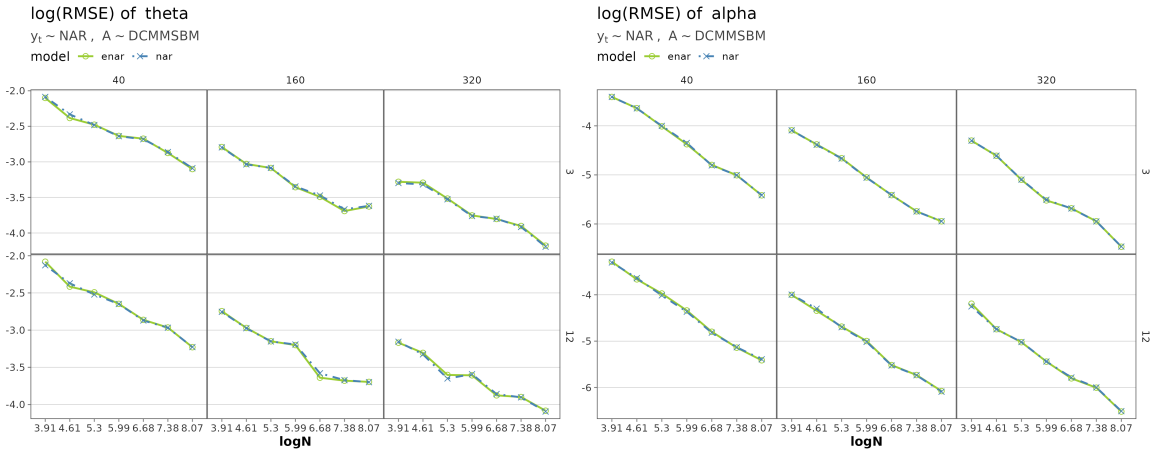


Figure C3: Plot of estimates of θ and α from ENAR, and NAR model when data is generated from NAR model with DCMMSBM.

E.1.3 Over-selection of K by IC

Here we present the simulation results for the over-selection procedure of latent dimension K introduced in section 5.1. We pose that the network density is $N\rho_N = \log^{2.5} N$ and choose the penalty function as $p_{NT}(k) = \frac{k}{N+T}$. We use almost the same simulation setting we used in section 6, but we only consider the case when $\mathbf{A}$ is generated from DCMMSBM. For the grid of $N, T \in \{20, 40, 60, 80, 100\}$, we grow K to be 2, 2, 3, 3, and 4 along each

of N. We plot the $\frac{1}{M}\sum_{i=1}^{M}\mathbb{I}_{\left\{\hat{K}\geq K\right\}}$ for $M = 500$ replications of Monte Carlo simulations. When applying the criteria, we consider a grid of $k = 0, 1, 2, \ldots, J_N = \lceil \|\mathbf{A}\|_\infty^{1/2} \rceil$ to search for $\hat{K}$. The lowest empirical probability of over-selection was 0.99 for all choices of N, T and both $l \in \{v, se\}$, suggesting an almost sure over-selection. The jitter plots of selected $\hat{K}$ in figure C2 clearly show their concentration around values greater than K.

E.1.4 Generating data from NAR with growing N, T

The figure C3 displays the plots of estimates of θ and α parameters with growing N and T when the data is generated from the NAR model. The simulation setup and the figure have been described in the main text.