# Estimation of Grouped Time-Varying Network Vector Autoregression Models

Degui Li[*], Bin Peng[†], Songqiao Tang[‡], Weibiao Wu[§]

*University of Macau, Monash University, Zhejiang University, University of Chicago*

Version: March 24, 2025

## Abstract

This paper introduces a flexible time-varying network vector autoregressive model framework for large-scale time series. A latent group structure is imposed on the heterogeneous and node-specific time-varying momentum and network spillover effects so that the number of unknown time-varying coefficients to be estimated can be reduced considerably. A classic agglomerative clustering algorithm with nonparametrically estimated distance matrix is combined with a ratio criterion to consistently estimate the latent group number and membership. A post-grouping local linear smoothing method is proposed to estimate the group-specific time-varying momentum and network effects, substantially improving the convergence rates of the preliminary estimates which ignore the latent structure. We further modify the methodology and theory to allow for structural breaks in either the group membership, group number or group-specific coefficient functions. Numerical studies including Monte-Carlo simulation and an empirical application are presented to examine the finite-sample performance of the developed model and methodology.

*Keywords*: cluster analysis, network VAR, latent groups, local linear estimator, time-varying coefficients

---

[*]Faculty of Business Administration, University of Macau, China.
[†]Department of Econometrics and Business Statistics, Monash University, Australia.
[‡]School of Mathematical Sciences, Zhejiang University, China.
[§]Department of Statistics, University of Chicago, US.

# 1 Introduction

Modeling large-scale time series has been the main frontier of recent advances of time series analysis and is of fundamental importance in various fields of applications such as climatology, economics, finance and social networks. Since Sims (1980)'s seminal work, the vector autoregressive (VAR) model has become a commonly-used statistical tool to tackle multivariate time series, see Lütkepohl (2006) and Kilian and Lütkepohl (2017) for a comprehensive review of classic estimation and forecasting techniques. However, the VAR-based estimation and forecasting are challenging when the number of time series sequences $N$ diverges to infinity. In this case, the number of unknown parameters in VAR transition matrices is of order $O(N^2)$, which may be substantially larger than the time series length $T$. In order to construct sensible estimate and forecast, two dimension-reduction approaches are often employed: VAR with sparse transition matrices and regularised estimation (e.g., Song & Bickel 2011, Bonhomme & Manresa 2015, Kock & Callot 2015, Davis, Zang & Zheng 2016, Miao, Phillips & Su 2023) and factor-augmented VAR (e.g., Bernanke, Boivin & Eliasz 2005, Bai & Ng 2006, Bai, Li & Lu 2016). Although some sound asymptotic properties have been developed for the sparse or factor-augmented VAR estimates, they often neglect possible network structures in large-scale time series and cannot directly capture dynamic network effects.

Consider time series observation vectors $X_t = (x_{1,t}, \cdots, x_{N,t})^\top$ with $N$ being the number of nodes in the large-scale network, and denote an adjacency matrix by $\mathbf{W} = (w_{ij})_{N \times N}$, where $w_{ii} = 0$, $w_{ij} = 1$ for $i \neq j$ if there exists a directed edge from $i$ to $j$ and $w_{ij} = 0$ otherwise. The matrix $\mathbf{W}$ is assumed to be observable and can be either directed ($\mathbf{W}^\top \neq \mathbf{W}$) or undirected ($\mathbf{W}^\top = \mathbf{W}$). The classic network VAR model is defined by

$$x_{i,t} = \beta_1 \sum_{j \neq i} \widetilde{w}_{ij} x_{j,t-1} + \beta_2 x_{i,t-1} + \varepsilon_{i,t}, \ \ i = 1, \cdots, N, \tag{1.1}$$

where $\widetilde{w}_{ij} = w_{ij}/n_i$ with $n_i = \sum_{j \neq i} w_{ij}$, $\beta_1$ and $\beta_2$ are unknown parameters, and $\varepsilon_t = (\varepsilon_{1,t}, \cdots, \varepsilon_{N,t})^\top$ is a sequence of independent and identically distributed (i.i.d.) random vectors. The above network VAR model formulation contains two regression components: $\beta_1 \sum_{j \neq i} \widetilde{w}_{ij} x_{j,t-1}$ and $\beta_2 x_{i,t-1}$, corresponding to the network (cross-lag) and momentum (own-lag) effects, respectively. Zhu *et al* (2017) discuss stationarity conditions for an extended network VAR model with extra nodal effects, propose the least squares estimation method and derive the relevant asymptotic theory. Although the classic linear network VAR model is easy to interpret and implement, it may be invalid in empirical applications. In particular, there exist two practical issues: (i) the stable network VAR model cannot capture smooth structural changes in the underlying data generating process over a long time span; and (ii) it is often too restrictive to impose the homogeneity assumption on the autoregressive coefficients over $N$ nodes. Consequently, the homogenous linear network VAR (1.1) may suffer from the model misspecification problem, resulting in biased estimates, inaccurate forecast and misleading inference. There have been some attempts in recent

years to address one of the aforementioned two issues. To incorporate structural changes in the autoregressive structures, Sun (2016), Sun & Malikov (2018), Wu (2019), Chen *et al* (2023) and Yin, Safikhani & Michailidis (2024) extend the linear network VAR models, allowing the coefficients to vary smoothly over time or with a stationary index variable. On the other hand, to relax the homogeneity restriction in linear VAR models, Yin, Safikhani & Michailidis (2023, 2024) introduce fully heterogenous network VAR models with the momentum and network spillover effects varying over nodes. However, the number of unknown coefficients in the latter models grow with the number of nodes, resulting in slow estimation convergence when the time span is not sufficiently long.

This paper aims to jointly tackle the aforementioned two issues by introducing a general time-varying network VAR model framework satisfying a latent group structure, i.e., the time-varying network autoregressive relationships are invariant within a group of nodes, but change over different groups. The grouped time-varying network VAR model achieves a good balance between model flexibility and parsimony. It not only covers the homogenous network VAR models (Zhu *et al* 2017, Wu 2019) as a special case, but also provides a more parsimonious model formulation than the fully heterogenous network VAR models (Yin, Safikhani & Michailidis 2023, 2024), achieving dimension reduction in estimation and improving the subsequent out-of-sample forecasting performance. The main methodological and theoretical contributions of our paper with connection to the existing literature are summarized as follows.

- *General network VAR model framework with a latent group structure on time-varying momentum and network spillover effects*. There has been increasing interest in recent years to explore a group structure under the classic stable VAR or network VAR model framework. For example, Zhu & Pan (2020) introduce a grouped linear network VAR model via a mixture Gaussian distribution and use an EM estimation algorithm; Guðmundsson & Brownlees (2021) propose a stochastic block VAR model and detect a latent group structure on the network spillover effects; Chen, Fan & Zhu (2023) study a community network VAR model and allow network effects to vary over different communities; and Zhu, Xu & Fan (2023) introduce a least squares algorithm to simultaneously estimate the parameters and identify the group structure for heterogeneous network VAR models. In this paper, we relax the somehow restricted stable model assumption in the aforementioned literature, allowing for structural changes in the underlying data generating process, a typical dynamic feature for large-scale network time series collected over a long time span. With the latent group structure, we substantially reduce the number of unknown coefficient functions for momentum and network spillover effects, which is appealing when the model is applied to the out-of-sample prediction. As in Zhu, Xu & Fan (2023), we allow the time-varying network effects to depend on both the sender and receiver's group information, resulting in a further homogeneity structure over the network effects.

- *Easy-to-implement clustering algorithm and post-grouping nonparametric estimation*. Since neither the group number nor membership is known a priori, we combine a classic agglom-

erative clustering algorithm and a simple ratio criterion to consistently estimate the group structure. With the nonparametrically estimated distance matrix, we may directly adopt a standard package in the computing software to implement the cluster analysis. The three-stage procedure introduced in Section 3.1 does not require iterative computation (e.g., Bonhomme & Manresa 2015, Zhu, Xu & Fan 2023) to obtain consistent group membership estimates. The developed clustering methodology complements recent developments on latent group estimation in the context of large panel data (e.g., Bonhomme & Manresa 2015, Ke, Li & Zhang 2016, Su, Shi & Phillips 2016, Ando & Bai 2017, Vogt & Linton 2017, 2020, Chen 2019). However, since the underlying high-dimensional time series process is locally stationary, it is technically more challenging to derive the asymptotic property of the developed methodology. To improve the convergence rates of the fully heterogenous time-varying coefficient estimation which only uses the sample information from one node and its direct neighbors, we propose a post-grouping local linear smoothing method in Section 4 to estimate the group-specific time-varying momentum and network effects by making use of the consistently estimated group structure. The asymptotic normal distribution theory is derived with the convergence rates comparable to those for homogenous time-varying coefficient estimation.

- *Structural breaks in the group structure.* The existing literature on grouped network VAR models requires the assumption that the latent group structure is time-invariant. This assumption may be restrictive for some empirical case studies. For example, a macroeconomic shock such as the global financial crisis may not only lead to a structural break in the group structure among a large number of countries, but also result in abrupt changes in the vector autoregressive structure of macroeconomic time series. This paper extends the model, methodology and theory, allowing for structural breaks in either the group membership, group number or group-specific coefficient functions. With the two-stage estimation procedure introduced in Section 5, we consistently estimate the scaled break location and the group structures over the two time periods separated by the break point. The online supplement (Li *et al* 2024) further introduces a refined break point estimation using the consistently estimated group structure to improve the break point estimation accuracy. Our model and methodology can be seen as an extension of the linear panel model framework (with break in the group structure) considered by Lumsdaine, Okui & Wang (2023) and Wang, Phillips & Su (2023), taking into account the network structure and allowing for structural changes over time.

The finite-sample Monte-Carlo simulation study shows that the proposed clustering algorithm and ratio criterion can consistently estimate the latent group membership and number as long as the time series length T is moderately large; the post-grouping local linear estimates perform significantly better than the naive heterogeneous local linear estimates; and the developed two-stage estimation procedure can precisely locate the break point and estimate the group structure before and after the break. The developed model and methodology are further applied to analyze

the monthly temperature data collected in 37 UK weather stations over the time period between January 1950 and February 2023. The empirical result reveals that there exist two groups over the node-specific time-varying coefficients: weather stations in Northern Ireland and Wales tend to form one group and those in England and Scotland form the other. In addition, our proposed method outperforms other competing methods in terms of out-of-sample prediction.

The rest of the paper is organized as follows. Section 2 introduces the time-varying network VAR model together with the latent group structure and provides some fundamental model assumptions. Section 3 describes the estimation methodology for the group structure and establishes the consistency property for the resulting estimation. Section 4 proposes the post-grouping nonparametric estimation and derives the relevant asymptotic theory. Section 5 extends the model, method and theory to the case with breaks in either the group structure or group-specific coefficient functions. Section 6 reports both the simulation and empirical studies. Section 7 concludes the paper. Appendix A introduces the clustering algorithm to identify the homogeneity structure on the network effects. The online supplement (Li *et al* 2024) contains proofs of the main asymptotic theorems, some technical lemmas with their proofs, discussion on the refined break point estimation, and extra numerical results. Throughout the paper, we let $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ be the floor and ceiling functions, respectively. For a vector $x = (x_1, \cdots, x_p)^\top \in \mathscr{R}^p$, we write $|x|_q = (\sum_{i=1}^p |x_i|^q)^{1/q}$ for $q \geqslant 1$ and $|x|_\infty = \max_{1 \leqslant i \leqslant p} |x_i|$; for a matrix $\Sigma = (\sigma_{ij})_{p \times p}$, write $|\Sigma|_F = (\sum_{i=1}^p \sum_{j=1}^p \sigma_{ij}^2)^{1/2}$, $|\Sigma|_\infty = \max_{1 \leqslant i,j \leqslant p} |\sigma_{ij}|$ and $|\Sigma|_O = \max_{x \in \mathscr{R}^p : |x|_2 = 1} |\Sigma x|$; and for a p-dimensional random vector $Z$, write $Z \in \mathscr{L}^\kappa$ if $\|Z\|_\kappa := [\mathsf{E}(|Z|_2^\kappa)]^{1/\kappa} < \infty, \kappa \geqslant 1$. Let $\mathbf{I}_k$ be a $k \times k$ identity matrix and $\mathbf{O}_{k \times l}$ a $k \times l$ zero matrix. For a square matrix, $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ denote the minimum and maximum eigenvalues and $\det(\cdot)$ denotes its determinant. Let $a_n = o(b_n)$, $a_n = o_P(b_n)$ and $a_n \propto b_n$ denote that $a_n / b_n \to 0$ as $n \to \infty$, $a_n / b_n \to 0$ with probability approaching one (*w.p.a.1*), and $0 < \underline{c} \leqslant a_n / b_n \leqslant \overline{c} < \infty$, respectively.

# 2 Time-varying network VAR and latent groups

In this section, we introduce the main model framework, i.e., the time-varying network VAR model with a latent group structure, and impose some fundamental assumptions, ensuring the network time series are locally stable.

## 2.1 Grouped time-varying network VAR

Suppose that there exists a partition of the node index set $\{1, 2, \cdots, N\}$, denoted by $\mathscr{G} = \{\mathscr{G}_1, \mathscr{G}_2, \cdots, \mathscr{G}_{K_0}\}$, such that $\mathscr{G}_i \cap \mathscr{G}_j = \emptyset$ for $1 \leqslant i \neq j \leqslant K_0$. Let $g_i \in \{1, \cdots, K_0\}$ be the group membership of the $i$-th node, i.e., $g_i = k$ is equivalent to $i \in \mathscr{G}_k$. Neither the group membership nor the group number is

known a priori. Consider the following grouped time-varying network VAR model:

$$x_{i,t} = \sum_{j \neq i} \alpha_{g_i g_j}(\tau_t) \widetilde{w}_{ij} x_{j,t-1} + \alpha_{g_i}(\tau_t) x_{i,t-1} + \varepsilon_{i,t}, \ \ i = 1, \cdots, N, \ \ t = 1, \cdots, T, \qquad (2.1)$$

where $\tau_t = t/T$ denotes the scaled time point, $\alpha_{g_i g_j}(\cdot)$ and $\alpha_{g_i}(\cdot)$ are smooth coefficient functions, and the remaining elements are defined as those in (1.1). In contrast to the linear network VAR model (1.1), the grouped time-varying network VAR model (2.1) provides a much more flexible framework, allowing the network and momentum effects to change over time and nodes. In particular, the interaction between nodes from two groups share the same time-varying network effects which is appealing for modeling social networks with smooth structural changes. Although we only consider one lag in model (2.1) for simplicity of exposition, the method and theory developed in Sections 3 and 4 below can be easily extended to the model setting with finite lags. It is worth pointing out that our network structure is deterministic (Zhu *et al* 2017) and the group membership is determined by node-specific time-varying momentum and network spillover effects. This is substantially different from the community network VAR model studied by Chen, Fan & Zhu (2023), where the network structure is random and the community structure is used for the network generating mechanism.

Let $\widetilde{W}$ be the row-normalised adjacency matrix with the $(i, j)$-entry being $\widetilde{w}_{ij}$, $B_1(\cdot)$ be an $N \times N$ matrix being the diagonal entries being zeros and the off-diagonal $(i, j)$-entry being $\alpha_{g_i g_j}(\cdot)$ and $B_2(\cdot) = \mathrm{diag}\{\alpha_{g_1}(\cdot), \cdots, \alpha_{g_N}(\cdot)\}$. Then, we may rewrite model (2.1) as

$$X_t = B(\tau_t) X_{t-1} + \varepsilon_t, \ \ B(\tau_t) = B_1(\tau_t) \odot \widetilde{W} + B_2(\tau_t), \ \ t = 1, \cdots, T, \qquad (2.2)$$

where $\odot$ denotes the Hadamard product between matrices. Model (2.2) thus falls within the high-dimensional time-varying VAR model framework which has received increasing attention in recent years. For instance, Ding, Qiu & Chen (2017) propose a kernel-weighted $\ell_1$-regularised estimation for a time-varying VAR model; Xu, Chen & Wu (2020) study a high-dimensional VAR model with multiple breaks and estimate smooth time-varying covariance and precision matrices between the break points; Chen *et al* (2023) estimate dual network structures via directed Granger causality and undirected partial correlation linkages within the high-dimensional time-varying VAR framework. However, the aforementioned literature often assumes a sparsity condition on the time-varying VAR transition matrices to facilitate the use of the regularised estimation techniques and cannot directly capture possible time-varying network effects. In this paper, we decompose the time-varying transition matrix into two components: $B_2(\tau_t)$ capturing the momentum effects and $B_1(\tau_t) \odot \widetilde{W}$ capturing the dynamic network spillover effects.

The proposed model (2.1) contains the homogenous time-varying network VAR model (e.g., Wu 2019) as a special case, i.e.,

$$x_{i,t} = \alpha_1^\dagger(\tau_t) \sum_{j \neq i} \widetilde{w}_{ij} x_{j,t-1} + \alpha_2^\dagger(\tau_t) x_{i,t-1} + \varepsilon_{i,t}.$$

In contrast to the fully heterogenous network VAR model (e.g., Yin, Safikhani & Michailidis 2024), our model achieves substantial dimension reduction, reducing the number of unknown coefficient functions to $K_0^2 + K_0$ which is finite (as $K_0$ is assumed to be fixed). In the homogenous or grouped linear network VAR (e.g., Zhu *et al* 2017, Chen, Fan & Zhu 2023, Zhu, Xu & Fan 2023), the so-called nodal effect is often incorporated in the model formulation. Hence, we may further extend model (2.1) to

$$x_{i,t} = \sum_{j \neq i} \alpha_{g_i g_j}(\tau_t) \widetilde{w}_{ij} x_{j,t-1} + \alpha_{g_i}(\tau_t) x_{i,t-1} + Z_i^\intercal \gamma_{g_i}(\tau_t) + \varepsilon_{i,t}, \tag{2.3}$$

where $Z_i$ is a p-dimensional vector of node-specific exogenous covariates and $\gamma_{g_i}(\cdot)$ is a vector of smooth coefficient functions. Letting $\{Z_i\}$ be independent of $\{\varepsilon_{i,t}\}$, the model framework and methodology developed in the sequel can be extended to tackle (2.3) with slight modification. However, for notational simplicity, we mainly focus on model (2.1) without the nodal effect throughout the paper.

## 2.2 Fundamental assumptions and functional dependence measure

Let $f'(\cdot)$ and $f''(\cdot)$ be the first- and second-order derivatives of $f(\cdot)$. We impose the following assumption on model (2.1).

**Assumption 1.** *(i) For* $1 \leqslant g, g^* \leqslant K_0$, $\alpha_{gg^*}(\cdot)$ *and* $\alpha_g(\cdot)$ *are second-order continuously differentiable functions with*

$$\max_{1 \leqslant g, g^* \leqslant K_0} \sup_{0 \leqslant \tau \leqslant 1} \left\{ \left| \alpha'_{gg^*}(\tau) \right| + \left| \alpha''_{gg^*}(\tau) \right| \right\} + \max_{1 \leqslant g \leqslant K_0} \sup_{0 \leqslant \tau \leqslant 1} \left\{ \left| \alpha'_g(\tau) \right| + \left| \alpha''_g(\tau) \right| \right\} \leqslant c_\alpha,$$

*where* $c_\alpha$ *is a positive constant. In addition,*

$$\max_{1 \leqslant g, g^* \leqslant K_0} \sup_{0 \leqslant \tau \leqslant 1} |\alpha_{gg^*}(\tau)| + \max_{1 \leqslant g \leqslant K_0} \sup_{0 \leqslant \tau \leqslant 1} |\alpha_g(\tau)| < 1. \tag{2.4}$$

*(ii) Let* $\{\varepsilon_t\}$ *be a sequence of i.i.d. random vectors with zero mean, positive definite covariance matrix denoted by* $\Sigma_\varepsilon$, *and*

$$\max_{1 \leqslant i \leqslant N} \mathsf{E}\left[ |\varepsilon_{i,t}|^q \right] \leqslant c_\varepsilon,$$

*where* $q > 8$ *and* $c_\varepsilon$ *is a positive constant.*

**Remark 2.1.** *(i) The smoothness condition on* $\alpha_{gg^*}(\cdot)$ *and* $\alpha_g(\cdot)$ *in Assumption 1(i) is common for the local linear estimation method and theory (e.g., Fan & Gijbels 1996). We may replace it by the condition that* $\alpha_{gg^*}(\cdot)$ *and* $\alpha_g(\cdot)$ *belong to the Hölder class (e.g., Definition 1.2 in Tsybakov 2008) when adopting a general local polynomial estimation. The condition (2.4) in Assumption 1(i) is a natural extension of the stability assumption for grouped linear network VAR (Zhu, Xu & Fan 2023), ensuring that the underlying time*

*series process is locally stable[1]. We require a relatively strong moment condition ($q > 8$) in Assumption 1(ii) to derive the uniform convergence of the kernel-weighted quantities, see Lemmas D.1 and D.2 in Appendix D of the supplement (Li et al 2024).*

*(ii) For $X_t$ defined in (2.2), its stationary approximation is given by*

$$X_t^\circ(\tau) = \mathbf{B}(\tau)X_{t-1}^\circ(\tau) + \varepsilon_t, \ \ 0 \leqslant \tau \leqslant 1. \tag{2.5}$$

*It follows from (2.4) in Assumption 1(i) that $|\mathbf{B}(\tau)|_O < 1$ uniformly over $0 \leqslant \tau \leqslant 1$. Hence, we obtain the following Wold representation:*

$$X_t^\circ(\tau) = G(\tau, \mathscr{F}_t) := \sum_{j=0}^{\infty} \mathbf{B}^j(\tau)\varepsilon_{t-j}, \tag{2.6}$$

*where $\mathscr{F}_t = (\cdots, \varepsilon_{t-1}, \varepsilon_t)$. Assume that*

$$\sup_{0 \leqslant \tau \leqslant 1} \||X_t^\circ(\tau)|_\infty\|_q \leqslant \theta_{N,q}, \tag{2.7}$$

*which is weaker than the condition in Example 2.1 of Zhang & Wu (2021) as we allow $\theta_{N,q}$ to diverge as $N$ increases. When $N$ is fixed, the above condition can be simplified to $\sup_{0 \leqslant \tau \leqslant 1} \mathsf{E}[|X_t^\circ(\tau)|_q^q] \leqslant \theta_q$ with $\theta_q$ be a positive constant. Letting $X_t^\circ = X_t^\circ(\tau_t)$, under Assumption 1(i) and (2.7), we may show that*

$$\max_{1 \leqslant t \leqslant T} \||X_t - X_t^\circ|_\infty\|_q = O\left(\theta_{N,q}/T\right), \tag{2.8}$$

*indicating that $X_t$ may be replaced by $X_t^\circ$ in the asymptotic derivation by restricting the divergence rate of $\theta_{N,q}$. The proof of (2.8) is provided in Appendix B of the supplement (Li et al 2024), which also discusses the connection of the proposed model to the nonlinear functional dependence measure introduced by Wu (2005).*

# 3 Estimation of the latent group structure

In this section, we introduce the methodology for estimating the latent group membership and number, and present the relevant asymptotic properties.

## 3.1 Group membership estimation when $K_0$ is pre-specified

We next introduce the nonparametric estimation and clustering methods and obtain the consistent estimation of the group membership $\mathscr{G}$ when $K_0$ is known a priori. The methodology can be split

---

[1]The condition in (2.4) may be replaced by the following assumption: uniformly over $\tau \in [0, 1]$, $\det(\mathbf{I}_N - z\mathbf{B}(\tau)) \neq 0$ for all $|z| \leqslant 1$. In fact, (2.4) is a sufficient condition to guarantee the latter assumption.

into the following three stages.

**Stage 1**: As there is no prior information on the latent group structure, we start with the fully heterogenous time-varying network VAR model:

$$x_{i,t} = \sum_{j \in \mathscr{N}_i} \beta_{ij}(\tau_t)\widetilde{w}_{ij}x_{j,t-1} + \beta_i(\tau_t)x_{i,t-1} + \varepsilon_{i,t}, \tag{3.1}$$

where $\mathscr{N}_i = \{j \neq i : w_{ij} = 1\}$ is the index set of nodes which the $i$-th node follows, $\beta_{ij}(\cdot) = \alpha_{g_i g_j}(\cdot)$ and $\beta_i(\cdot) = \alpha_{g_i}(\cdot)$. Model (3.1) is similar to that in [Yin, Safikhani & Michailidis (2024)](#). Note that $\beta_{ij}(\cdot), j \notin \mathscr{N}_i$, are unidentifiable and thus not estimable. With the smoothness condition in Assumption 1(i), we adopt the local linear smoothing (e.g., [Fan & Gijbels 1996](#)) to estimate the heterogenous time-varying coefficient functions, only using the sample information from the $i$-th node and its direct neighbors, i.e., $\widetilde{w}_{ij} \neq 0$.

Define

$$\widetilde{X}_{i,t-1} = \left[(\widetilde{w}_{ij}x_{j,t-1} : j \in \mathscr{N}_i)^\intercal, x_{i,t-1}\right]^\intercal,$$

which is a random vector with dimension $n_i + 1$, where $n_i = \mathrm{card}(\mathscr{N}_i)$ is allowed to diverge slowly to infinity and $\mathrm{card}(\cdot)$ denotes the cardinality of a set. Letting

$$\beta_{i\bullet}(\tau) = \left[(\beta_{ij}(\tau) : j \in \mathscr{N}_i)^\intercal, \beta_i(\tau)\right]^\intercal,$$

with Assumption 1(i), we have the following Taylor expansion:

$$\beta_{i\bullet}(\tau_t) \approx \beta_{i\bullet}(\tau) + \beta'_{i\bullet}(\tau)(\tau_t - \tau)$$

when $\tau_t$ falls in a small neighborhood of $\tau$. Define the node-specific local linear weighted objective function:

$$\mathcal{L}_i(a, b) = \sum_{t=1}^{T} \left[x_{i,t} - a^\intercal \widetilde{X}_{i,t-1} - b^\intercal \widetilde{X}_{i,t-1}(\tau_t - \tau)\right]^2 K_h(\tau_t - \tau), \tag{3.2}$$

where $a$ and $b$ are $(n_i + 1)$-dimensional vectors, $K_h(\cdot) = \frac{1}{h}K(\cdot/h)$, $K(\cdot)$ is a kernel function and $h$ is a bandwidth. Minimizing $\mathcal{L}_i(a, b)$ with respect to the vectors $a$ and $b$, we obtain the solution denoted as $\widehat{a}$ and $\widehat{b}$, and then the local linear estimate of $\beta_{i\bullet}(\tau)$ as

$$\widehat{\beta}_{i\bullet}(\tau) = \left[\left(\widehat{\beta}_{ij}(\tau) : j \in \mathscr{N}_i\right)^\intercal, \widehat{\beta}_i(\tau)\right]^\intercal = \widehat{a}. \tag{3.3}$$

In practice, we obtain the local linear estimates at $\tau_l^*$, $l = 1, \cdots, L$, a sequence of user-specified equidistant grid points between 0 and 1 satisfying $L \to \infty$ and $L = O(T)$.

**Stage 2**: Let $\bar{\mathscr{N}}_N = \{(i, j) : 1 \leqslant i \leqslant N, j \in \mathscr{N}_i\}$. It follows from (2.1) that there exists a latent homogeneity structure for $\beta_{ij}(\cdot)$, $(i, j) \in \bar{\mathscr{N}}_N$, and the number of distinct time-varying coefficient

9

functions is at most $K_0^2$. Let $\beta_m^\circ(\cdot)$, $m = 1, \cdots, M_0$, denote the true distinct time-varying coefficient functions for network spillover effects, $M_0 \leqslant K_0^2$, and $g_{ij} \in \{1, \cdots, M_0\}$ be the group membership for the index pair $(i, j) \in \bar{\mathcal{N}}_N$. With $\{\widehat{\beta}_{ij}(\tau_l^*) : 1 \leqslant l \leqslant L, (i, j) \in \bar{\mathcal{N}}_N\}$ obtained in Stage 1, combining the clustering algorithm and the ratio criterion with details provided in Appendix A (see also Stage 3 and Section 3.2), we may obtain a consistent estimate of $M_0$ denoted by $\widehat{M}$, and the estimated membership $\widehat{g}_{ij}$. For the $i$-th node, we construct

$$\widehat{\beta}_{i\bullet}^\circ(\tau) = \left[\widehat{\beta}_{i1}^\circ(\tau), \cdots, \widehat{\beta}_{i\widehat{M}}^\circ(\tau)\right]^{\mathsf{T}} \quad \text{with} \quad \widehat{\beta}_{im}^\circ(\tau) = \sum_{j \in \mathcal{N}_i} \widehat{\beta}_{ij}(\tau)\widehat{\omega}_{ij,m}, \tag{3.4}$$

where $\tau$ is chosen as the grid points $\tau_l^*$ defined in Stage 1,

$$\widehat{\omega}_{ij,m} = \begin{cases} I(\widehat{g}_{ij} = m) / \sum_{j \in \mathcal{N}_i} I(\widehat{g}_{ij} = m), & \text{if } \sum_{j \in \mathcal{N}_i} I(\widehat{g}_{ij} = m) > 0, \\ 0, & \text{if } \sum_{j \in \mathcal{N}_i} I(\widehat{g}_{ij} = m) = 0, \end{cases}$$

and $I(\cdot)$ denotes the indicator function.

**Stage 3**: With the estimates $\widehat{\beta}_i(\cdot)$ and $\widehat{\beta}_{i\bullet}^\circ(\cdot)$ defined in Stages 1 and 2, respectively, we may compute the point-wise distance between nodes $i$ and $j$:

$$\widehat{d}_{ij}(\tau) = \left|\widehat{\beta}_i(\tau) - \widehat{\beta}_j(\tau)\right| + \left|\widehat{\beta}_{i\bullet}^\circ(\tau) - \widehat{\beta}_{j\bullet}^\circ(\tau)\right|_2, \tag{3.5}$$

and subsequently define the distance matrix:

$$\widehat{\mathbf{D}} = \left\{\widehat{D}_{ij}\right\}_{N \times N}, \quad \widehat{D}_{ij} = \frac{1}{L}\sum_{l=1}^{L} \widehat{d}_{ij}(\tau_l^*).$$

It is clear that the diagonal elements of $\widehat{\mathbf{D}}$ are zeros. With the distance matrix $\widehat{\mathbf{D}}$, we may adopt the agglomerative hierarchical clustering algorithm which is commonly used in unsupervised cluster analysis (e.g., Hastie, Tibshirani & Friedman 2009, Everitt *et al* 2011). This clustering algorithm has been recently combined with the kernel-based estimation technique to identify the homogeneity/group structure in nonparametric panel regression models. For instance, Chen (2019) constructs a similar distance matrix and further estimates the latent group structure in time-varying coefficient panel data models; and Vogt & Linton (2020) introduce a bandwidth-free normalized distance measure in the clustering algorithm but assume the panel observations are independent over subjects. The latter assumption may be too restrictive for large-scale network time series data and is thus removed in this paper. Another relevant paper is Zhang (2013) which clusters nonlinear trend functions based on parallelism and allows the number of time series to grow at a slow polynomial rate of T. In contrast, the number of nodes N can be much larger than T in this paper, see (3.4) in Assumption 2(iii).

Assuming the true group number $K_0$ is known a priori, we start with N clusters each of which

corresponds to one node, search for the smallest off-diagonal entry in $\widehat{\mathbf{D}}$ (which is the smallest estimated distance between nodes), and merge the two corresponding nodes. Consequently the cluster number reduces from N to $N - 1$. Use a linkage technique (such as the single or complete linkage) to calculate the distance between the merged cluster and the remaining ones and update the estimated distance matrix with size $(N - 1) \times (N - 1)$. Repeat the previous steps with the updated distance matrix, and stop the algorithm when the number of clusters reaches $K_0$. We denote the estimated clusters by $\widehat{\mathscr{G}}_k$, $k = 1, \cdots, K_0$.

Let $\boldsymbol{\Delta}_{i,t} = \mathsf{E}[\widetilde{X}_{i,t} \widetilde{X}_{i,t}^\mathsf{T}]$ with $\widetilde{X}_{i,t}$ defined in Stage 1, and write

$$\beta_{i\bullet}^\circ(\tau) = \left[ \beta_{i1}^\circ(\tau), \cdots, \beta_{iM_0}^\circ(\tau) \right]^\mathsf{T} \quad \text{with} \quad \beta_{im}^\circ(\tau) = \sum_{j \in \mathscr{N}_i} \beta_{ij}(\tau) \omega_{ij,m} \tag{3.6}$$

and

$$\omega_{ij,m} = \begin{cases} I(g_{ij} = m) / \sum_{j \in \mathscr{N}_i} I(g_{ij} = m), & \text{if } \sum_{j \in \mathscr{N}_i} I(g_{ij} = m) > 0, \\ 0, & \text{if } \sum_{j \in \mathscr{N}_i} I(g_{ij} = m) = 0. \end{cases}$$

The latter is estimated by $\widehat{\beta}_{i\bullet}^\circ(\tau)$ defined in (3.4) (up to permutation). The following conditions are required to derive the consistency property of $\widehat{\mathscr{G}}_k$, $k = 1, \cdots, K_0$.

**Assumption 2.** *(i) The kernel function $K(\cdot)$ is a symmetric probability density function that is Lipschitz-continuous and has a compact support $[-1, 1]$.*

*(ii) There exist two finite positive constants: $\underline{\lambda}$ and $\overline{\lambda}$, such that*

$$0 < \underline{\lambda} \leqslant \min_{1 \leqslant i \leqslant N} \min_{0 \leqslant t \leqslant T-1} \lambda_{\min}(\boldsymbol{\Delta}_{i,t}) \leqslant \max_{1 \leqslant i \leqslant N} \sup_{0 \leqslant t \leqslant T-1} \lambda_{\max}(\boldsymbol{\Delta}_{i,t}) \leqslant \overline{\lambda} < \infty.$$

*(iii) Let T, N, and h satisfy $h \to 0$, $Th \to \infty$ and*

$$\frac{N \theta_{N,q}^q}{T^{\frac{q^2 - 6q - 8}{4(q+2)}} [h \log(N \vee T)]^{q/4}} \to 0 \tag{3.7}$$

*with q defined in Assumption 1(ii) and $\theta_{N,q}$ is defined in (2.7).*

*(iv) Letting $\mathscr{N}_i(j) = \{k \in \mathscr{N}_i : g_k = j\}$ and $\overline{n} = \max_{1 \leqslant i \leqslant N} n_i$,*

$$\min_{1 \leqslant i,j \leqslant K_0} \mathrm{card}(\mathscr{N}_i(j)) \geqslant 1, \quad \overline{n} = o\left( \sqrt{Th/\log(N \vee T)} \right). \tag{3.8}$$

**Assumption 3.** *Let*

$$\sqrt{\overline{n}} \left( h^2 + \sqrt{\frac{\log(N \vee T)}{Th}} \right) + L^{-1} = o\left( \zeta_{NT}^\dagger \wedge \zeta_{NT}^\ddagger \right), \tag{3.9}$$

11

*where*

$$\zeta_{NT}^{\dagger} = \min_{1 \leqslant g_i \neq g_j \leqslant K_0} \int_0^1 \left[ \left| \alpha_{g_i}(\tau) - \alpha_{g_j}(\tau) \right| + \left| \beta_{i\bullet}^{\circ}(\tau) - \beta_{j\bullet}^{\circ}(\tau) \right|_2 \right] d\tau$$

*and*

$$\zeta_{NT}^{\ddagger} = \min_{1 \leqslant m \neq m^* \leqslant M_0} \int_0^1 \left| \beta_m^{\circ}(\tau) - \beta_{m^*}^{\circ}(\tau) \right| d\tau.$$

**Remark 3.1.** *(i) Assumption 2(i) contains some commonly-used conditions imposed on the kernel function (e.g., Fan & Gijbels 1996). Assumption 2(ii) is crucial to ensure that the kernel-weighted random denominator in local linear estimation is non-singular. In fact, the consistency property in Theorem 3.1 below continues to hold by allowing $\underline{\lambda}$ to slowly converge to zero and strengthening other relevant conditions, e.g., the second restriction in (3.8) and (3.9) would be strengthened to*

$$\bar{n} = o\left( \underline{\lambda} \sqrt{Th/\log(N \vee T)} \right) \quad \text{and} \quad \sqrt{\bar{n}} \left( h^2 + \sqrt{\frac{\log(N \vee T)}{Th}} \right) / \underline{\lambda} + L^{-1} = o\left( \zeta_{NT}^{\dagger} \right),$$

*respectively. Note that $h \to 0$ and $Th \to \infty$ in Assumption 2(iii) are regular conditions for kernel-based smoothing, whereas the condition (3.7) indicates that there is a trade-off between the network size and the required moment condition (i.e., when $q$ increases, $N$ may diverge at a faster rate). The first condition in (3.8) indicates that node $i$ follows nodes in each of the $K_0$ groups and $\alpha_{g_i g_j}(\cdot)$ is estimable via the local linear method in Stage 1, whereas the second condition in (3.8) restricts the divergence rate of $n_i$ so that the first-stage local linear estimation of the heterogenous time-varying coefficient functions is uniformly consistent, see Lemma D.3 in the supplement (Li et al 2024).*

*(ii) Assumption 3 indicates that the minimum distance between groups may converge to zero at a rate slower than a typical nonparametric uniform convergence rate if the grid number $L$ is of order $T$ and $\bar{n}$ is bounded. The restriction (3.9) is automatically satisfied if $\zeta_{NT}^{\dagger}$ and $\zeta_{NT}^{\ddagger}$ are strictly larger than a positive constant (e.g., Zhu, Xu & Fan 2023).*

Theorem 3.1 below establishes the consistency property of the group membership estimation when $K_0$ is pre-specified.

**Theorem 3.1.** *Suppose that Assumptions 1–3 hold and $K_0$ is known a priori. Then, as $N, T \to \infty$ jointly, we have*

$$P\left( \{ \widehat{\mathscr{G}}_k, \ 1 \leqslant k \leqslant K_0 \} = \{ \mathscr{G}_k, \ 1 \leqslant k \leqslant K_0 \} \right) \to 1. \tag{3.10}$$

**Remark 3.2.** *The consistency property (3.10) is similar to the consistency results of group membership estimation in nonparametric panel/longitudinal data models, see Theorem 3.1 in Vogt & Linton (2017) and Theorem 4.1(a) in Vogt & Linton (2020). The key step of proving Theorem 3.1 is to show that*

$$\max_{1 \leqslant k \leqslant K_0} \max_{i,j \in \mathscr{G}_k} \widehat{D}_{ij} < \min_{1 \leqslant k \neq l \leqslant K_0} \min_{i \in \mathscr{G}_k, j \in \mathscr{G}_l} \widehat{D}_{ij}, \quad w.p.a.1.$$

*This can be proved by using the uniform convergence property of $\widehat{d}_{ij}(\cdot)$.*

## 3.2 Estimation of the group number

In practice, the true number of groups is unknown and a data-driven criterion is thus required to obtain its consistent estimation. We next introduce an easy-to-implement ratio criterion to estimate $K_0$ consistently. Assuming the group number to be $K$, we may terminate the clustering algorithm in Stage 3 when the cluster number reaches $K$ and obtain the estimated clusters denoted by $\widehat{\mathscr{G}}_{k|K}$, $k = 1, \cdots, K$. With these estimated clusters, we pool the heterogenous time-varying coefficient estimates $\widehat{\beta}_i(\cdot)$ and $\widehat{\beta}^\circ_{i\bullet}(\cdot)$ over $i \in \widehat{\mathscr{G}}_{k|K}$, and obtain

$$\widehat{\beta}_{k|K}(\tau) = \frac{1}{\operatorname{card}\left(\widehat{\mathscr{G}}_{k|K}\right)} \sum_{i \in \widehat{\mathscr{G}}_{k|K}} \widehat{\beta}_i(\tau), \quad \widehat{\beta}^\circ_{k|K}(\tau) = \frac{1}{\operatorname{card}\left(\widehat{\mathscr{G}}_{k|K}\right)} \sum_{i \in \widehat{\mathscr{G}}_{k|K}} \widehat{\beta}^\circ_{i\bullet}(\tau). \tag{3.11}$$

Then, we define the average deviation:

$$\widehat{R}(K) = \frac{1}{KL} \sum_{k=1}^{K} \sum_{l=1}^{L} \frac{1}{\operatorname{card}\left(\widehat{\mathscr{G}}_{k|K}\right)} \sum_{i \in \widehat{\mathscr{G}}_{k|K}} \left[ \left| \widehat{\beta}_i(\tau^*_l) - \widehat{\beta}_{k|K}(\tau^*_l) \right| + \left| \widehat{\beta}^\circ_{i\bullet}(\tau^*_l) - \widehat{\beta}^\circ_{k|K}(\tau^*_l) \right|_2 \right]. \tag{3.12}$$

The grouped time-varying network VAR model is either correctly- or over-fitted when $K \geqslant K_0$, indicating that $\widehat{R}(K)$ converges to zero, and is under-fitted when $K < K_0$. For the latter scenario, at least two groups would be falsely merged, leading to biased estimation of some group-specific time-varying coefficient functions and a relatively large value of $\widehat{R}(K)$. Hence, it is sensible to estimate $K_0$ by

$$\widehat{K} = \operatorname*{argmin}_{1 \leqslant K \leqslant \overline{K}} \frac{\widehat{R}(K)}{\widehat{R}(K-1)}, \tag{3.13}$$

where $\overline{K}$ is a pre-specified positive integer larger than $K_0$. In practical implementation, we set $\widehat{R}(1)/\widehat{R}(0) \equiv 1$, $\widehat{R}(K) = 0$ if it is smaller than $\rho_{NT}$, a user-specified tuning parameter, and define $0/0 \equiv 1$. A similar ratio criterion is adopted by Yang *et al* (2023) to consistently estimate the group number in nonparametric grouped panel quantile regression models. Other applications of the ratio criterion can be found in Lam & Yao (2012) and Li, Robinson and Shang (2020).

We require some further conditions to derive the consistency property of $\widehat{K}$.

**Assumption 4.** *(i) There exists a positive constant $c_{\mathscr{G}}$ such that*

$$\min_{1 \leqslant k \leqslant K_0} \operatorname{card}(\mathscr{G}_k) \geqslant c_{\mathscr{G}} \cdot N.$$

*(ii) The tuning parameter $\rho_{NT}$ satisfies that*

$$\rho_{NT} = o\left(\zeta^\dagger_{NT} \wedge \zeta^\ddagger_{NT}\right), \quad \sqrt{\bar{n}}\left(h^2 + \sqrt{\frac{\log(N \vee T)}{Th}}\right) = o\left(\rho_{NT}\right). \tag{3.14}$$

**Remark 3.3.** *Assumption 4(i) indicates that the cardinality of $\mathscr{G}_k$ is of the same order over $k$. A similar restriction is also adopted by Chen (2019) and Zhu, Xu & Fan (2023). Assumption 4(ii) indicates that the order of $\rho_{NT}$ lies between $\sqrt{\bar{n}}(h^2 + \sqrt{\log(N \vee T)/(Th)})$ and $\xi_{NT}^\dagger \wedge \xi_{NT}^\ddagger$, which is not unreasonable due to Assumption 3. In particular, when $\xi_{NT}^\dagger \wedge \xi_{NT}^\ddagger$ is bounded away from zero and $\bar{n}$ is upper bounded by a positive constant, the conditions in (3.14) can be simplified to*

$$\rho_{NT} \to 0, \quad h^2 + \sqrt{\frac{\log(N \vee T)}{Th}} = o\left(\rho_{NT}\right).$$

We establish the consistency property of the ratio criterion in the following theorem.

**Theorem 3.2.** *Suppose that Assumptions 1–4 are satisfied. Then, as $N, T \to \infty$ jointly,*

$$P\left(\widehat{K} = K_0\right) \to 1. \tag{3.15}$$

Finally, we use the estimated group number $\widehat{K}$ and terminate the clustering algorithm in Section 3.1 when the cluster number reaches $\widehat{K}$. In order to avoid unnecessary notational burden, we still denote the estimated groups as $\widehat{\mathscr{G}}_k$, $k = 1, \cdots, \widehat{K}$. Combining Theorems 3.1 and 3.2, we readily have the following result.

**Corollary 3.3.** *Suppose that Assumptions 1–4 are satisfied. Then, as $N, T \to \infty$ jointly,*

$$P\left(\{\widehat{\mathscr{G}}_k, 1 \leqslant k \leqslant \widehat{K}\} = \{\mathscr{G}_k, 1 \leqslant k \leqslant K_0\}\right) \to 1. \tag{3.16}$$

# 4 Post-grouping local linear estimation

The heterogenous local linear estimation defined in (3.3) only makes use of the sample information from node $i$ and its direct neighbors, resulting in rather slow convergence rates (see Lemma D.3 in the supplement) and unstable numerical performance if $T$ is not sufficiently large in finite samples. We next aim to address this issue by pooling the sample information over nodes in the same cluster and proposing a post-grouping local linear estimation.

It follows from Corollary 3.3 that, for any $k = 1, \cdots, K_0$, there exists $1 \leqslant k^\dagger \leqslant \widehat{K}$ such that $\mathscr{G}_k = \widehat{\mathscr{G}}_{k^\dagger}$ *w.p.a.1*. Without loss of generality, we may consider $\mathscr{G}_k = \widehat{\mathscr{G}}_k$ (conditioning on $\widehat{K} = K_0$) throughout this section. For $i \in \widehat{\mathscr{G}}_k$, define

$$\check{X}_{i,t-1} = \left[\sum_{j \in \widehat{\mathscr{G}}_1} \widetilde{w}_{ij} x_{j,t-1}, \cdots, \sum_{j \in \widehat{\mathscr{G}}_{\widehat{K}}} \widetilde{w}_{ij} x_{j,t-1}, x_{i,t-1}\right]^\top,$$

which is a random vector with dimension $\widehat{K} + 1$. For the k-th group, let

$$\alpha_{k\bullet}(\tau) = [\alpha_{k1}(\tau), \cdots, \alpha_{kK_0}(\tau), \alpha_k(\tau)]^{\mathsf{T}}$$

be a vector of true group-specific time-varying coefficient functions to be estimated. For each k and given $\tau \in (0,1)$, we define the following post-grouping local linear objective function:

$$\sum_{i \in \widehat{\mathscr{G}}_k} \sum_{t=1}^{T} \left[ x_{i,t} - a^{\mathsf{T}} \check{X}_{i,t-1} - b^{\mathsf{T}} \check{X}_{i,t-1}(\tau_t - \tau) \right]^2 K_{h_\dagger}(\tau_t - \tau), \tag{4.1}$$

where $h_\dagger$ is a bandwidth which may be different from h used in the heterogenous local linear estimation (3.2) and (3.3). Minimizing the post-grouping objective function with respect to the vectors a and b, we obtain the solutions denoted by $\check{a}$ and $\check{b}$, and construct the post-grouping local linear estimation as

$$\check{\alpha}_{k\bullet}(\tau) = [\check{\alpha}_{k1}(\tau), \cdots, \check{\alpha}_{kK_0}(\tau), \check{\alpha}_k(\tau)]^{\mathsf{T}} = \check{a}. \tag{4.2}$$

Let $\sigma_{ij} = \mathsf{E}(\varepsilon_{i,t}\varepsilon_{j,t})$ and

$$\Delta_{ij}^{\diamond}(\tau) = \mathsf{E}\left[ X_{i,t}^{\diamond}(\tau) X_{j,t}^{\diamond\mathsf{T}}(\tau) \right], \quad 1 \leqslant i,j \leqslant N,$$

where

$$X_{i,t}^{\diamond} = \left[ \sum_{j \in \mathscr{G}_1} \widetilde{w}_{ij} x_{j,t}^{\circ}(\tau), \cdots, \sum_{j \in \mathscr{G}_{K_0}} \widetilde{w}_{ij} x_{j,t}^{\circ}(\tau), x_{i,t}^{\circ}(\tau) \right]^{\mathsf{T}}$$

with $x_{i,t}^{\circ}(\tau)$ being the i-th element of $X_t^{\circ}(\tau)$ defined in (2.6).

**Assumption 5.** *(i) The bandwidth $h_\dagger$ satisfies that $h_\dagger \to 0$ and $Th_\dagger / \log(N \vee T) \to \infty$. In addition, (3.7) holds when h is replaced by $h_\dagger$.*

*(ii) There exists a positive definite matrix $\Upsilon_{\mathscr{G}_k}(\tau)$ such that*

$$\frac{1}{\mathrm{card}(\mathscr{G}_k)} \sum_{i,j \in \mathscr{G}_k} \sigma_{ij} \Delta_{ij}^{\diamond}(\tau) \to \Upsilon_{\mathscr{G}_k}(\tau) \tag{4.3}$$

*as $\mathrm{card}(\mathscr{G}_k) \to \infty$, and in addition,*

$$\frac{1}{\mathrm{card}(\mathscr{G}_k)} \sum_{i \in \mathscr{G}_k} \Delta_i^{\diamond}(\tau) \to \Delta_{\mathscr{G}_k}(\tau),$$

*which is positive definite, where $\Delta_i^{\diamond}(\tau) = \Delta_{ii}^{\diamond}(\tau)$.*

**Remark 4.1.** *The bandwidth restriction in Assumption 5(i) is comparable to that in Assumption 2(iii). Assumption 5(ii) allows weak correlation between nodes and can be substantially simplified when $\varepsilon_{i,t}$ are*

*independent over* $i$ *(e.g.,* [Zhu](#) *et al 2017). For example, if* $\sigma_{ij} = 0$ *when* $i \neq j$ *and* $\sigma_{ii} \equiv \sigma^2$, *(4.3) would be simplified to*

$$\frac{1}{\text{card}(\mathscr{G}_k)} \sum_{i,j \in \mathscr{G}_k} \sigma_{ij} \mathbf{\Delta}_{ij}^\diamond(\tau) = \frac{\sigma^2}{\text{card}(\mathscr{G}_k)} \sum_{i \in \mathscr{G}_k} \mathbf{\Delta}_i^\diamond(\tau) \to \sigma^2 \mathbf{\Delta}_{\mathscr{G}_k}(\tau).$$

**Theorem 4.1.** *Suppose that Assumptions [1–5](#) are satisfied. For any* $\tau \in (0,1)$,

$$\sqrt{\text{card}(\mathscr{G}_k)\text{Th}_\dagger} \left[ \check{\alpha}_{k\bullet}(\tau) - \alpha_{k\bullet}(\tau) - \frac{1}{2}h_\dagger^2 \mu_2 \alpha_{k\bullet}''(\tau) \right] \xrightarrow{d} N\left(\mathbf{0}, \mathbf{\Omega}_{\mathscr{G}_k}(\tau)\right), \tag{4.4}$$

*as* $N, T \to \infty$ *jointly, where* $\mathbf{\Omega}_{\mathscr{G}_k}(\tau) = \nu_0 \mathbf{\Delta}_{\mathscr{G}_k}^{-1}(\tau) \mathbf{\Upsilon}_{\mathscr{G}_k}(\tau) \mathbf{\Delta}_{\mathscr{G}_k}^{-1}(\tau)$, $\nu_\kappa = \int u^\kappa K^2(u) du$ *and* $\mu_\kappa = \int u^\kappa K(u) du$ *for* $k = 0, 1, \cdots$.

**Remark 4.2.** *Since* $\text{card}(\mathscr{G}_k)$ *is of order* $N$ *by Assumption [4](#)(i), Theorem [4.1](#) shows that the post-grouping local linear estimation* $\check{\alpha}_{k\bullet}(\tau)$ *has the point-wise convergence rate* $1/\sqrt{NTh_\dagger} + h_\dagger^2$, *which is substantially faster than that for the heterogenous local linear estimation (ignoring the group structure). This is unsurprising since more sample information is used in the post-grouping estimation procedure. Note that [Yin, Safikhani & Michailidis (2024)](#) use the spline-based estimation for heterogenous functional coefficients, which only achieve the root-$T$ convergence rate. If, in addition,* $\varepsilon_{i,t}$ *are independent over* $i$, *as discussed in Remark [4.1](#), we may simplify the asymptotic covariance matrix, i.e.,* $\mathbf{\Omega}_{\mathscr{G}_k}(\tau) = \nu_0 \sigma^2 \mathbf{\Delta}_{\mathscr{G}_k}^{-1}(\tau)$.

# 5 Breaks in the group structue

The model, methodology and theory developed in Sections [2–4](#) rely on the assumption that the latent group structure is time invariant and the group-specific coefficient functions are smooth over the entire time span. As discussed in the introductory section, this assumption may be restrictive for some empirical applications. Hence, we next make a further extension of the model, methodology and theory, allowing structural breaks in either the group membership, group number or group-specific coefficient functions. Our main interest lies in locating the break point and estimating the group structure before and after the break. We mainly consider the case of a single break for notational brevity and will briefly discuss its extension to the case of multiple breaks later in Remark [5.2](#)(ii).

Assume that the break occurs at an unknown time point $t_0$. Let $\mathscr{G}^1 = \{\mathscr{G}_1^1, \cdots, \mathscr{G}_{K_1}^1\}$ and $g_i^1 \in \{1, \cdots, K_1\}$ be the group structure and membership label (for node $i$) before the break, whereas let $\mathscr{G}^2 = \{\mathscr{G}_1^2, \cdots, \mathscr{G}_{K_2}^2\}$ and $g_i^2 \in \{1, \cdots, K_2\}$ be defined similarly for those after the break. Consider the time-varying network VAR model with break in the group structure:

$$x_{i,t} = \begin{cases} \sum_{j \neq i} \alpha_{g_i^1 g_j^1}^1(\tau_t) \widetilde{w}_{ij} x_{j,t-1} + \alpha_{g_i^1}^1(\tau_t) x_{i,t-1} + \varepsilon_{i,t}, & 1 \leqslant t \leqslant t_0, \\ \sum_{j \neq i} \alpha_{g_i^2 g_j^2}^2(\tau_t) \widetilde{w}_{ij} x_{j,t-1} + \alpha_{g_i^2}^2(\tau_t) x_{i,t-1} + \varepsilon_{i,t}, & t_0 + 1 \leqslant t \leqslant T, \end{cases} \tag{5.1}$$

where $\alpha^1_{g^1_i g^1_j}(\cdot)$ (or $\alpha^2_{g^2_i g^2_j}(\cdot)$) and $\alpha^1_{g^1_i}(\cdot)$ (or $\alpha^2_{g^2_i}(\cdot)$) are the smooth time-varying network spillover and momentum effects before (or after) the break. Model (5.1) can be seen as an extension of the linear panel model framework (with break in the group structure) in Lumsdaine, Okui & Wang (2023) and Wang, Phillips & Su (2023), taking into account the smooth time-varying feature and network structure.

Throughout this section, assume that $\bar{n}$ is bounded. We next introduce a two-stage estimation procedure with break location estimation in Stage 1 and then group estimation in Stage 2.

**Stage 1**: As the time-varying group structure is latent, similar to (3.1), we first consider the fully heterogenous time-varying network VAR model with break at $t_0$:

$$x_{i,t} = \sum_{j \in \mathcal{N}_i} \beta^{\ddagger}_{ij}(\tau_t) \widetilde{w}_{ij} x_{j,t-1} + \beta^{\ddagger}_i(\tau_t) x_{i,t-1} + \varepsilon_{i,t}, \tag{5.2}$$

where

$$\beta^{\ddagger}_{ij}(\tau_t) = \begin{cases} \beta^1_{ij}(\tau_t), & 1 \leqslant t \leqslant t_0, \\ \beta^2_{ij}(\tau_t), & t_0 + 1 \leqslant t \leqslant T, \end{cases} \qquad \beta^{\ddagger}_i(\tau_t) = \begin{cases} \beta^1_i(\tau_t), & 1 \leqslant t \leqslant t_0, \\ \beta^2_i(\tau_t), & t_0 + 1 \leqslant t \leqslant T. \end{cases} \tag{5.3}$$

Write

$$\beta^{\ddagger}_{i\bullet}(\tau_t) = \left[ \left( \beta^{\ddagger}_{ij}(\tau_t) : j \in \mathcal{N}_i \right)^{\mathsf{T}}, \beta^{\ddagger}_i(\tau_t) \right]^{\mathsf{T}}$$

as in Section 3.1, and let

$$\beta^{\ddagger,l}_{i\bullet}(\tau) = \lim_{x \uparrow \tau} \beta^{\ddagger}_{i\bullet}(x) \quad \text{and} \quad \beta^{\ddagger,r}_{i\bullet}(\tau) = \lim_{x \downarrow \tau} \beta^{\ddagger}_{i\bullet}(x)$$

denote the left and right limits of $\beta^{\ddagger}_{i\bullet}(\tau)$, respectively. Define

$$\delta_\beta(t) = \max_{1 \leqslant i \leqslant N} \left| \beta^{\ddagger,r}_{i\bullet}(\tau_t) - \beta^{\ddagger,l}_{i\bullet}(\tau_t) \right|_2. \tag{5.4}$$

It follows from (5.1)–(5.3) that $\delta_\beta(t)$, $t = 1, \cdots, T$, achieve the maximum at $t = t_0$, which motivates the subsequent estimation procedure. Specifically, we estimate $\delta_\beta(t)$ by a one-sided kernel smoothing method[2] and then locate the break point by maximizing the estimate of $\delta_\beta(t)$ over $t$.

Let $K^{\ddagger}(\cdot)$ be a one-sided kernel function with a compact support $[0, 1]$, say, the one-sided version of the Epanechnikov kernel. Define

$$\widetilde{\Gamma}^l_{it} = \frac{1}{Th_{\ddagger}} \sum_{s=1}^{T} \widetilde{X}_{i,s-1} \widetilde{X}^{\mathsf{T}}_{i,s-1} K^{\ddagger} \left( \frac{\tau_t - \tau_s}{h_{\ddagger}} \right), \quad \widetilde{\Gamma}^r_{it} = \frac{1}{Th_{\ddagger}} \sum_{s=1}^{T} \widetilde{X}_{i,s-1} \widetilde{X}^{\mathsf{T}}_{i,s-1} K^{\ddagger} \left( \frac{\tau_s - \tau_t}{h_{\ddagger}} \right),$$

---

[2]The extension to the one-sided local polynomial smoothing is straightforward, see Chen, Wang & Wu (2022).

$$\overline{\Gamma}^l_{it} = \frac{1}{Th_{\ddagger}} \sum_{s=1}^{T} \widetilde{X}_{i,s-1} x_{i,s} K^{\ddagger}\left(\frac{\tau_t - \tau_s}{h_{\ddagger}}\right), \quad \overline{\Gamma}^r_{it} = \frac{1}{Th_{\ddagger}} \sum_{s=1}^{T} \widetilde{X}_{i,s-1} x_{i,s} K^{\ddagger}\left(\frac{\tau_s - \tau_t}{h_{\ddagger}}\right),$$

where $h_{\ddagger}$ is the bandwidth and $\widetilde{X}_{i,s-1}$ is defined as in Section 3.1. We estimate $\beta^{\ddagger,l}_{i\bullet}(\tau_t)$ and $\beta^{\ddagger,r}_{i\bullet}(\tau_t)$ by

$$\widehat{\beta}^{\ddagger,l}_{i\bullet}(\tau_t) = \left(\widetilde{\Gamma}^l_{it}\right)^{-1} \overline{\Gamma}^l_{it} \quad \text{and} \quad \widehat{\beta}^{\ddagger,r}_{i\bullet}(\tau_t) = \left(\widetilde{\Gamma}^r_{it}\right)^{-1} \overline{\Gamma}^r_{it}, \tag{5.5}$$

respectively, and then construct

$$\widehat{\delta}_{\beta}(t) = \max_{1 \leqslant i \leqslant N} \left|\widehat{\beta}^{\ddagger,r}_{i\bullet}(\tau_t) - \widehat{\beta}^{\ddagger,l}_{i\bullet}(\tau_t)\right|_2. \tag{5.6}$$

The estimation of $t_0$ is defined as

$$\widehat{t} = \operatorname*{argmax}_{t} \widehat{\delta}_{\beta}(t). \tag{5.7}$$

**Stage 2**: Let

$$\mathcal{T}_1 = \{1, 2, \cdots, \widehat{t} - \lfloor \epsilon_0 T \rfloor\} \quad \text{and} \quad \mathcal{T}_1 = \{\widehat{t} + \lfloor \epsilon_0 T \rfloor, \cdots, T-1, T\},$$

where $\epsilon_0$ is an arbitrarily small positive number (say 0.01). From Theorem 5.1(i), the group membership and number are time invariant *w.p.a.1* over the two time periods $\mathcal{T}_1$ and $\mathcal{T}_2$. Hence, we may adopt the clustering algorithm and ratio criterion developed in Section 3 to estimate $\mathcal{G}^1_k$ and $K_1$ (using the network time series sample over $\mathcal{T}_1$) as well as $\mathcal{G}^2_k$ and $K_2$ (using the sample over $\mathcal{T}_2$). We denote the resulting estimates as $\widehat{\mathcal{G}}^1_k$, $\widehat{K}_1$, $\widehat{\mathcal{G}}^2_k$ and $\widehat{K}_2$, whose consistency property is derived in Theorem 5.1(ii) below.

The following conditions are required to derive the asymptotic property of the above two-stage estimation method.

**Assumption 6.** *(i) For $1 \leqslant g, g^* \leqslant K_1$, $\alpha^1_{gg^*}(\cdot)$ and $\alpha^1_g(\cdot)$ have bounded first-order derivatives and satisfy*

$$\max_{1 \leqslant g, g^* \leqslant K_1} \sup_{0 \leqslant \tau \leqslant 1} \left|\alpha^1_{gg^*}(\tau)\right| + \max_{1 \leqslant g \leqslant K_1} \sup_{0 \leqslant \tau \leqslant 1} \left|\alpha^1_g(\tau)\right| < 1.$$

*The same conditions hold for $\alpha^2_{gg^*}(\cdot)$ and $\alpha^2_g(\cdot)$, $1 \leqslant g, g^* \leqslant K_2$.*

*(ii) $K^{\ddagger}(\cdot)$ is positive and Lipschitz-continuous with a compact support $[0, 1]$.*

*(iii) There exist positive constants $c_1 \in (0, 1)$ and $c_2$ such that $t_0 = c_1 T$ and $\delta_{\beta}(t_0) > c_2$.*

*(iv) There exist positive constant $c_3$ and $c_4$ such that*

$$\min_{1 \leqslant k \leqslant K_1} \operatorname{card}(\mathcal{G}^1_k) \geqslant c_3 \cdot N \quad \text{and} \quad \min_{1 \leqslant k \leqslant K_2} \operatorname{card}(\mathcal{G}^2_k) \geqslant c_4 \cdot N.$$

**Remark 5.1.** *Assumption 6(i)(ii) contains some typical conditions on the time-varying coefficient functions and kernel function which are often required when the one-sided kernel smoothing is adopted. In*

*particular, Assumption 6(i) ensures that the grouped time-varying network VAR process is locally stable over the two time periods separated by the break point. Assumption 6(iii) indicates that the break location is well separated from the endpoints and the break size is bounded away from zero. The structural break may be due to abrupt changes in the group-specific time-varying coefficients functions or breaks in the group membership or number. More discussion and examples are available in Appendix E of the online supplement (Li et al 2024). In fact, by slightly modifying the proof and theory, we may allow $\delta_\beta(t_0)$ to slowly approach zero as in Chen, Wang & Wu (2022). Assumption 6(iv) is a natural extension of Assumption 4(i) to the time-varying group structure.*

Let $\beta_{i\bullet}^{\circ 1}(\cdot)$ be defined similarly to $\beta_{i\bullet}^{\circ}(\cdot)$ in (3.6) but with $\beta_{ij}(\cdot)$ replaced by $\beta_{ij}^1(\cdot)$. As $\beta_m^{\circ}(\cdot)$ defined in Stage 2 of Section 3.1, we let $\beta_m^{\circ 1}(\cdot)$, $m = 1, \cdots, M_1$, denote the distinct coefficient functions for time-varying network effects before the break, where $M_1 \leqslant K_1^2$. The definitions of $\beta_{i\bullet}^{\circ 2}(\cdot)$ and $\beta_m^{\circ 2}(\cdot)$, $m = 1, \cdots, M_2$, are analogous. Define

$$\zeta_{NT}^{\dagger 1} = \min_{1 \leqslant g_i^1 \neq g_j^1 \leqslant K_1} \int_0^{c_1} \left[ \left| \alpha_{g_i^1}^1(\tau) - \alpha_{g_j^1}^1(\tau) \right| + \left| \beta_{i\bullet}^{\circ 1}(\tau) - \beta_{j\bullet}^{\circ 1}(\tau) \right|_2 \right] d\tau,$$

$$\zeta_{NT}^{\dagger 2} = \min_{1 \leqslant g_i^2 \neq g_j^2 \leqslant K_2} \int_{c_1}^1 \left[ \left| \alpha_{g_i^2}^2(\tau) - \alpha_{g_j^2}^2(\tau) \right| + \left| \beta_{i\bullet}^{\circ 2}(\tau) - \beta_{j\bullet}^{\circ 2}(\tau) \right|_2 \right] d\tau,$$

$$\zeta_{NT}^{\ddagger 1} = \min_{1 \leqslant m \neq m^* \leqslant M_1} \int_0^{c_1} \left| \beta_m^{\circ 1}(\tau) - \beta_{m^*}^{\circ 1}(\tau) \right| d\tau,$$

$$\zeta_{NT}^{\ddagger 2} = \min_{1 \leqslant m \neq m^* \leqslant M_2} \int_{c_1}^1 \left| \beta_m^{\circ 2}(\tau) - \beta_{m^*}^{\circ 2}(\tau) \right| d\tau.$$

Let $\mathscr{N}_i^1(j) = \{k \in \mathscr{N}_i : g_k^1 = j\}$ and $\mathscr{N}_i^2(j) = \{k \in \mathscr{N}_i : g_k^2 = j\}$.

**Theorem 5.1.** *Suppose that Assumptions 1(ii), 2(ii) and 6(i)–(iii) are satisfied.*

*(i) The break location estimate $\widehat{t}$ has the following approximation order:*

$$\left| \frac{\widehat{t} - t_0}{T} \right| = O_P \left( \sqrt{\frac{\bar{n} h_\ddagger \log(N \vee T)}{T}} + h_\ddagger^2 \right). \tag{5.8}$$

*(ii) If, in addition, Assumption 6(iv) is satisfied and (3.8), (3.9) and (3.14) continue to hold when $h$, $\mathrm{card}(\mathscr{N}_i(j))$ and $\zeta_{NT}^\dagger \wedge \zeta_{NT}^\ddagger$ and replaced by $h_\ddagger$, $\mathrm{card}(\mathscr{N}_i^1(j)) \wedge \mathrm{card}(\mathscr{N}_i^2(j))$ and $\min\{\zeta_{NT}^{\dagger 1}, \zeta_{NT}^{\dagger 2}, \zeta_{NT}^{\ddagger 1}, \zeta_{NT}^{\ddagger 2}\}$, respectively, we have*

$$\mathsf{P}\left( \{\widehat{\mathscr{G}}_k^1, 1 \leqslant k \leqslant \widehat{K}_1\} = \{\mathscr{G}_k^1, 1 \leqslant k \leqslant K_1\} \right) \to 1,$$

$$\mathsf{P}\left( \{\widehat{\mathscr{G}}_k^2, 1 \leqslant k \leqslant \widehat{K}_2\} = \{\mathscr{G}_k^2, 1 \leqslant k \leqslant K_2\} \right) \to 1,$$

$$\mathsf{P}\left( \widehat{K}_1 = K_1 \right) \to 1, \quad \mathsf{P}\left( \widehat{K}_2 = K_2 \right) \to 1,$$

*as $N, T \to \infty$ jointly.*

**Remark 5.2.** *(i) Theorem [5.1](i) shows that the scaled break point estimation is consistent. Although the convergence rate in (5.8) is conservative, it is sufficient to consistently estimate the time-varying group membership and number in Stage 2. The approximation rate may be improved if we replace the one-sided kernel by one-sided local linear smoothing (e.g., Chen, Wang & Wu 2022) in Stage 1.*

*(ii) In practice, there are often multiple breaks in the latent group structure, i.e., breaks occur at some unknown but well separated break points $\mathsf{t}_1, \mathsf{t}_2, \cdots, \mathsf{t}_p$. In this setting, the methodology and theory continue to hold with minor amendments. For example, we may use the recursive algorithm in Xu, Chen & Wu (2020) to locate the $\mathsf{p}$ break points, an idea similar to the binary segmentation commonly used to estimate multiple breaks in parametric models (e.g., Cho & Fryzlewicz 2012, 2015).*

*(iii) In Appendix E of the online supplement (Li et al 2024), we discuss a refined break point estimation, making use of the consistently estimated group structures. Under some high-level conditions, we show that the refined estimation of the break location is consistent. Furthermore, we provide a few examples to verify the high-level conditions.*

# 6  Numerical studies

In this section, we conduct both the simulation and empirical studies. Sections 6.1 and 6.2 assess the finite-sample performance of the developed methodology and verify the main convergence properties via simulation, whereas Section 6.3 reports the empirical application to a network time series data set for UK temperature.

## 6.1  Simulation study without break

We use the grouped network time-varying VAR model (2.1) for data generation. The entries of the adjacency matrix $\mathbf{W}$ are defined by $w_{ij} = I(u_{ij} \leqslant \bar{w})$, where $u_{ij} \sim \mathsf{U}(0,1)$ and $0 < \bar{w} < 1$, controlling the sparsity level of the network structure. The innovation vectors are generated by $\varepsilon_t \sim \mathsf{N}(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon)$ independently over $t$, where $\boldsymbol{\Sigma}_\varepsilon = \{\sigma_{ij}\}_{N \times N}$ with $\sigma_{ij} = 0.1^{|i-j|}$, allowing cross-sectional dependence over components. Consider $K_0 = 2$ and define the group-specific coefficients as

$$\alpha_{g_i}(\tau) = \begin{cases} 0.49 \cos(\pi\tau), & i \in \mathscr{G}_1, \\ -0.2, & i \in \mathscr{G}_2, \end{cases} \qquad \alpha_{g_i g_j}(\tau) = \begin{cases} 0.5 - \tau, & \text{if } i, j \in \mathscr{G}_1, \\ \tau^3 - 0.5, & \text{otherwise.} \end{cases}$$

The group membership is generated as follows: assign each node $i \in \{1, \ldots, N\}$ to $\mathscr{G}_1$ and $\mathscr{G}_2$ with respective probabilities 0.65 and 0.35. Note that there exists a further group structure on the time-varying spillover effects $\alpha_{g_i g_j}(\cdot)$ with $M_0 = 2$, as described in Stage 2 of Section 3.1. In the simulation, we consider two scenarios for generating the group membership: (i) fixed group, i.e., the group membership is only generated once and remains the same over replications;

and (ii) random group, i.e., the group membership is randomly generated for each replication. We conduct the simulation over $R = 1000$ replications and set $N = 100, 200$, $T = 300, 600$, and $\bar{w} = 0.025, 0.075$.

We use the Epanechnikov kernel in the local linear smoothing, where the bandwidth is determined by the rule of thumb (Su & Wang 2017): $h = (2.35/\sqrt{12})T^{-1/5}$ for the fully heterogenous local linear estimation (3.2) and $h_\dagger = (2.35/\sqrt{12})[\text{card}(\widehat{\mathscr{G}}_j)T]^{-1/5}$ for the post-grouping estimation (4.1), where $\widehat{\mathscr{G}}_j$ denotes the estimated group. We notice that the clustering results are insensitive to the bandwidth choice in our simulation. For each simulated data set, we first estimate the group membership and number as in Section 3, and then conduct the post-grouping estimation as in Section 4. To evaluate the group structure estimation accuracy, we adopt the following two measurements:

$$\mathsf{AC}(K_0) = \frac{1}{R} \sum_{r=1}^{R} I(\widehat{K}_r = K_0) \quad \text{and} \quad \mathsf{Purity}(\mathscr{G}) = \frac{1}{RN} \sum_{r=1}^{R} \sum_{k=1}^{\widehat{K}_r} \max_{1 \leqslant j \leqslant K_0} \left| \widehat{\mathscr{G}}_{k,r} \cap \mathscr{G}_j \right|,$$

where $\widehat{K}_r$ and $\widehat{\mathscr{G}}_{k,r}$ denote the estimates of the group number and membership in the $r$-th replication. The purity quantity $\mathsf{Purity}(\mathscr{G})$ is a simple and transparent evaluation measure with value close to one when the clustering method is precise. To compare the estimation performance between the pre-grouping local linear estimation and the post-grouping one, we compute the root mean squared errors for the estimated momentum and spillover effects:

$$\mathsf{RMSE}_{M,r} = \left\{ \frac{1}{NS} \sum_{i=1}^{N} \sum_{s=1}^{S} \left| \widetilde{\beta}_{r,i}(\tau_s) - \beta_i(\tau_s) \right|^2 \right\}^{1/2},$$

$$\mathsf{RMSE}_{S,r} = \left\{ \frac{1}{NS} \sum_{i=1}^{N} \sum_{s=1}^{S} \left| \widetilde{\beta}^*_{r,i\bullet}(\tau_s) - \beta^*_{i\bullet}(\tau_s) \right|_2^2 \right\}^{1/2},$$

where $\beta^*_{i\bullet}(\cdot) = (\beta_{ij}(\cdot) : j \in \mathscr{N}_i)^{\mathsf{T}}$, $\widetilde{\beta}_{r,i}(\cdot)$ and $\widetilde{\beta}^*_{r,i\bullet}(\cdot)$ stand for the estimates in the $r$-th replication, $\tau_s = 0.05, 0.1, \cdots, 0.95$ and $S = 19$.

The simulation results are summarized in Tables 1 and 2. The numbers in parentheses of Table 2 are standard deviations of $\mathsf{RMSE}_{M,r}$ and $\mathsf{RMSE}_{S,r}$ over 1000 replications. It follows from Table 1 that both $\mathsf{AC}(K_0)$ and $\mathsf{Purity}(\mathscr{G})$ converge to one as the time series length $T$ increases from 300 to 600, and the results remain stable when the sparsity level $\bar{w}$ changes from 0.025 to 0.075. Table 2 shows that the post-grouping time-varying coefficient estimation substantially outperforms the pre-grouping one, confirming that the estimation accuracy is significantly improved by making use of the estimated group structure. The standard deviations are generally small, indicating that the nonparametric estimation performance is stable over replications. In addition, both the pre-grouping and post-grouping local linear estimates deteriorate when $\bar{w}$ increases from 0.025 to 0.075.

21

Table 1: Estimation performance of the group number and membership

|  |  |  | Fixed group | | Random group | |
| --- | --- | --- | --- | --- | --- | --- |
| Sparsity | Measurement | T \ N | 100 | 200 | 100 | 200 |
| $\bar{w} = 0.025$ | AC($K_0$) | 300 | 1.000 | 0.995 | 0.999 | 0.997 |
|  |  | 600 | 1.000 | 1.000 | 1.000 | 1.000 |
|  | Purity($\mathscr{G}$) | 300 | 0.997 | 0.996 | 0.996 | 0.995 |
|  |  | 600 | 1.000 | 1.000 | 1.000 | 1.000 |
|  |  |  |  |  |  |  |
| $\bar{w} = 0.075$ | AC($K_0$) | 300 | 0.999 | 0.989 | 0.989 | 0.993 |
|  |  | 600 | 1.000 | 1.000 | 1.000 | 1.000 |
|  | Purity($\mathscr{G}$) | 300 | 0.988 | 0.958 | 0.985 | 0.955 |
|  |  | 600 | 1.000 | 1.000 | 1.000 | 0.999 |

Table 2: Estimation performance of the time-varying momentum and network spillover effects

|  |  |  | Fixed group | | | | Random group | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  | RMSE$_M$ | | RMSE$_S$ | | RMSE$_M$ | | RMSE$_S$ | |
| Sparsity | Estimation | T \ N | 100 | 200 | 100 | 200 | 100 | 200 | 100 | 200 |
| $\bar{w} = 0.025$ | Pre-grouping | 300 | 0.148 | 0.316 | 0.330 | 0.768 | 0.189 | 0.312 | 0.419 | 0.745 |
|  |  |  | (0.023) | (0.015) | (0.011) | (0.012) | (0.027) | (0.019) | (0.038) | (0.037) |
|  |  | 600 | 0.078 | 0.211 | 0.234 | 0.493 | 0.116 | 0.223 | 0.295 | 0.515 |
|  |  |  | (0.023) | (0.014) | (0.007) | (0.006) | (0.026) | (0.016) | (0.025) | (0.025) |
|  | Post-grouping | 300 | 0.035 | 0.048 | 0.107 | 0.118 | 0.036 | 0.047 | 0.108 | 0.117 |
|  |  |  | (0.010) | (0.007) | (0.004) | (0.003) | (0.011) | (0.008) | (0.004) | (0.003) |
|  |  | 600 | 0.021 | 0.024 | 0.077 | 0.081 | 0.021 | 0.026 | 0.078 | 0.082 |
|  |  |  | (0.004) | (0.003) | (0.002) | (0.002) | (0.004) | (0.003) | (0.002) | (0.002) |
| $\bar{w} = 0.075$ | Pre-grouping | 300 | 0.399 | 0.751 | 0.967 | 2.268 | 0.424 | 0.768 | 1.075 | 2.333 |
|  |  |  | (0.023) | (0.129) | (0.015) | (0.020) | (0.034) | (0.146) | (0.063) | (0.083) |
|  |  | 600 | 0.289 | 0.505 | 0.660 | 1.450 | 0.306 | 0.509 | 0.729 | 1.457 |
|  |  |  | (0.014) | (0.048) | (0.009) | (0.010) | (0.019) | (0.047) | (0.038) | (0.042) |
|  | Post-grouping | 300 | 0.062 | 0.102 | 0.125 | 0.156 | 0.063 | 0.104 | 0.126 | 0.157 |
|  |  |  | (0.013) | (0.019) | (0.004) | (0.003) | (0.015) | (0.020) | (0.005) | (0.004) |
|  |  | 600 | 0.032 | 0.049 | 0.086 | 0.099 | 0.033 | 0.050 | 0.086 | 0.100 |
|  |  |  | (0.003) | (0.004) | (0.003) | (0.002) | (0.004) | (0.005) | (0.003) | (0.003) |

Table 3: Measurements of the (scaled) break point estimation

| Sparsity | T \ N | Fixed group | | Random group | |
|---|---|---|---|---|---|
| | | 100 | 200 | 100 | 200 |
| $\bar{w} = 0.025$ | 400 | -0.002 (0.031) | -0.003 (0.033) | -0.001 (0.032) | -0.003 (0.033) |
| | 800 | -0.001 (0.016) | -0.002 (0.017) | -0.002 (0.016) | 0.000 (0.017) |
| $\bar{w} = 0.075$ | 400 | -0.003 (0.033) | -0.002 (0.034) | -0.003 (0.034) | -0.002 (0.034) |
| | 800 | -0.001 (0.017) | 0.000 (0.017) | 0.000 (0.017) | -0.001 (0.017) |

## 6.2   Simulation study with a break in the group structure

We next examine the numerical performance of the estimation method with a break in the group structure introduced in Section 5. The break point is set at $t_0 = \lfloor T/2 \rfloor + 1$ when we re-assign each node $i$ to $\mathscr{G}_1$ and $\mathscr{G}_2$ with respective probabilities 0.65 and 0.35. This results in a break in the group membership. Before the break time, the data generating process is the same as that in Section 6.1, whereas, after the break, the group-specific time-varying coefficients are defined as

$$\alpha_{g_i}(\tau) = \begin{cases} -0.49\sin(\pi\tau/2), & i \in \mathscr{G}_1, \\ 0.49\sin(\pi\tau/2), & i \in \mathscr{G}_2, \end{cases} \qquad \alpha_{g_i g_j}(\tau) = \begin{cases} 0.49\sin(\pi\tau/2), & \text{if } i,j \in \mathscr{G}_1, \\ -0.49\sin(\pi\tau/2), & \text{otherwise.} \end{cases}$$

As in Section 6.1, we consider both the fixed and random groups when generating the group membership (with a break) over $R = 1000$ replications. In order to obtain stable finite-sample performance, we slightly increase $T$ from $(300, 600)$ to $(400, 800)$. The number of nodes remains as $N = 100$ and 200.

The one-sided version of the Epanechnikov kernel function is adopted in our nonparametric method. We first estimate the break point via (5.7), compute the (scaled) measurement $\Delta_{t_0} = (\hat{t} - t_0)/T$, and then report the means and standard deviations (in parentheses) of $\Delta_{t_0}$ in Table 3. It is clear that the scaled break point $t_0/T$ can be accurately detected, and its estimation accuracy is not sensitive to the sparsity level of the adjacency matrix $\mathbf{W}$. With the estimated break point, we may split the entire time period into the "pre-break" and "post-break" periods and compute their respective $AC(K_0)$ and $Purity(\mathscr{G})$. We further take a simple average of those values over the two periods and report them in Table 4. We note that both the group number and membership are estimated very accurately even when there exists a break in the group structure. We finally compare the estimation performance between the pre-grouping and post-grouping local linear estimation after the break point is detected. When computing $RMSE_M$ and $RMSE_S$, we choose $\tau_s = 0.05, 0.1, \ldots, 0.4$ for the "pre-break" period and $\tau_s = 0.6, 0.65, \ldots, 0.95$ for the "post-break" period, avoiding possible boundary effect in the estimation. The general pattern in Table 5 is very similar to that in Table 2, again confirming the significant advantage of the post-grouping estimation.

Appendix F in the online supplement (Li *et al* 2024) contains extra simulation results: the

Table 4: Estimation performance of the group number and membership (with a break)

| Sparsity | Measurement | T \ N | Fixed group | | Random group | |
|---|---|---|---|---|---|---|
| | | | 100 | 200 | 100 | 200 |
| $\bar{w} = 0.025$ | $AC(K_0)$ | 400 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | 800 | 1.000 | 1.000 | 1.000 | 1.000 |
| | $Purity(\mathscr{G})$ | 400 | 0.999 | 0.999 | 0.999 | 0.998 |
| | | 800 | 1.000 | 1.000 | 1.000 | 1.000 |
| $\bar{w} = 0.075$ | $AC(K_0)$ | 400 | 0.994 | 0.966 | 0.993 | 0.968 |
| | | 800 | 1.000 | 1.000 | 1.000 | 1.000 |
| | $Purity(\mathscr{G})$ | 400 | 0.996 | 0.982 | 0.995 | 0.983 |
| | | 800 | 1.000 | 1.000 | 1.000 | 1.000 |

Table 5: Estimation performance of the time-varying coefficients with a break in the group structure

| | | | Fixed group | | | | Random group | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | $RMSE_M$ | | $RMSE_S$ | | $RMSE_M$ | | $RMSE_S$ | |
| Sparsity | Estimation | T \ N | 100 | 200 | 100 | 200 | 100 | 200 | 100 | 200 |
| $\bar{w} = 0.025$ | Pre-grouping | 400 | 0.490 | 0.562 | 0.553 | 0.785 | 0.477 | 0.545 | 0.569 | 0.765 |
| | | | (0.006) | (0.010) | (0.007) | (0.008) | (0.020) | (0.016) | (0.027) | (0.028) |
| | | 800 | 0.485 | 0.490 | 0.521 | 0.622 | 0.461 | 0.494 | 0.515 | 0.627 |
| | | | (0.004) | (0.005) | (0.005) | (0.006) | (0.020) | (0.016) | (0.021) | (0.020) |
| | Post-grouping | 400 | 0.032 | 0.040 | 0.093 | 0.101 | 0.034 | 0.040 | 0.095 | 0.101 |
| | | | (0.006) | (0.006) | (0.003) | (0.002) | (0.009) | (0.007) | (0.003) | (0.002) |
| | | 800 | 0.020 | 0.023 | 0.067 | 0.071 | 0.021 | 0.024 | 0.068 | 0.072 |
| | | | (0.003) | (0.002) | (0.002) | (0.002) | (0.004) | (0.003) | (0.002) | (0.002) |
| $\bar{w} = 0.075$ | Pre-grouping | 400 | 0.613 | 1.023 | 0.930 | 1.890 | 0.636 | 1.035 | 0.998 | 1.910 |
| | | | (0.026) | (0.070) | (0.012) | (0.015) | (0.027) | (0.066) | (0.046) | (0.058) |
| | | 800 | 0.535 | 0.730 | 0.731 | 1.260 | 0.538 | 0.743 | 0.759 | 1.291 |
| | | | (0.008) | (0.021) | (0.008) | (0.008) | (0.017) | (0.027) | (0.029) | (0.033) |
| | Post-grouping | 400 | 0.046 | 0.075 | 0.106 | 0.126 | 0.050 | 0.075 | 0.108 | 0.126 |
| | | | (0.010) | (0.016) | (0.004) | (0.003) | (0.011) | (0.016) | (0.004) | (0.004) |
| | | 800 | 0.026 | 0.038 | 0.074 | 0.083 | 0.028 | 0.038 | 0.074 | 0.083 |
| | | | (0.003) | (0.003) | (0.002) | (0.002) | (0.004) | (0.005) | (0.002) | (0.002) |

finite-sample performance of the clustering algorithm introduced in Appendix A on estimating the homogeneity structure for the network spillover effects, and the clustering result by using Zhu, Xu & Fan (2023)'s grouped network VAR model with constant coefficients.

## 6.3 An Empirical Study

There has been increasing interest in investigating the spatial pattern of climate data, see, for example, Portmann, Solomon & Hegerl (2009), Kendon *et al* (2020), Hanlon *et al* (2021) and the references therein. We next apply the proposed model and methodology to analyze a set of UK climate data, exploring the network and latent group structures and allowing for smooth structural changes to account for possible climate changes in the past few decades. The data that we use are collected from the UK Meteorological Office,[3] containing temperature recordings (in Celsius) of 37 weather stations with their geographical locations presented in Figure 1. The original dataset collects minimum and maximum temperatures per month over the period from January 1950 to February 2023. Hence, the time series length is $T = 878$. We consider the following two scenarios when building the network VAR model: (i) both the minimum and maximum temperatures are used as elements of $X_t$; (ii) the averaged temperature per month is used as elements of $X_t$. In model (ii), each weather station is treated as a node and $N = 37$; whereas in model (i), the minimum and maximum recordings in each weather station are treated as two nodes and $N = 74$. The time series observations are standardized to have zero mean and unit standard deviation. The adjacency matrix **W** is constructed by following the UK climate region map[4]. Specifically, it is classified into the following five regions: Southern England, Northern England, Wales, Scotland, and Northern Ireland. When stations $i$ and $j$ are in the same region, we set $w_{ij} = 1$, otherwise, $w_{ij} = 0$. Consequently, the percentage of non-zero elements of the adjacency matrix is 0.0559 and 0.2267 for the two models.

The primary interest of this empirical study lies in identifying potential group structure over the weather stations to achieve dimension reduction in the subsequent network VAR model estimation. With the clustering algorithm in Section 3, we obtain two estimated groups, i.e., $\widehat{K} = 2$, for both models. Table 6 reports the estimated group membership which is very similar between the two models. For model (i), all the stations of Group 1 are in Northern Ireland and Wales, while those of Group 2 are in England and Scotland. This indicates that Northern Ireland and Wales often have common patterns in terms of temperature change whereas the weather stations in the island of Great Britain usually have similar temperature recordings. It is noteworthy that, although each weather station in model (i) contains the minimum and maximum temperatures (as two nodes), both nodes are classified into the same group (thus we only report the station name in Table 6). For model (ii), three weather stations in the coastal towns (Eastbourne, Tiree and Whitby) move from Group 2 to Group 1. The estimated time-varying coefficient functions

---

[3]https://www.metoffice.gov.uk/research/climate/maps-and-data/historic-station-data
[4]https://www.metoffice.gov.uk/research/climate/maps-and-data/about/regions-map
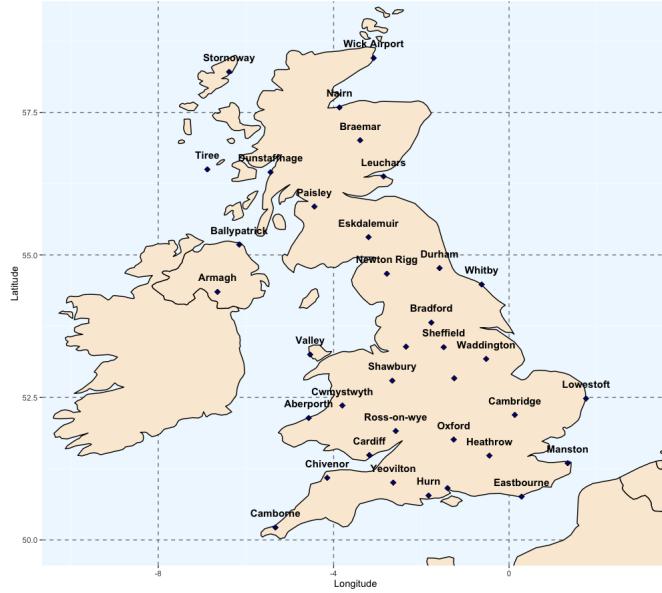
Figure 1: Geographical locations of the UK weather stations

are plotted in the online supplement (Li *et al* 2024).

To further show the necessity of accounting for the group structure and smooth structural changes, we compare the out-of-sample prediction performance among the following three methods: fully heterogeneous time-varying network VAR model (Yin, Safikhani & Michailidis 2024); the proposed grouped time-varying network VAR model; and the grouped network VAR model with constant coefficients (Zhu, Xu & Fan 2023), denoted by "fully heterogeneous", "grouped+TV" and "grouped+linear", respectively, in Table 7. We leave the last $T_{pre}$ observations out for prediction, where $T_{pre} = 12, 24$ and 36, corresponding to one, two and three years, respectively. For any time point $t_{\bullet}$ in the prediction (or test) period, we use the observations over $t = 1, \cdots, t_{\bullet} - 1$ to get the estimates of time-varying or constant coefficients, which are subsequently used to forecast the value of $X_{t_{\bullet}}$. We conduct this expanding-window one-step ahead forecasting exercise for all the three models and report their out-of-sample RMSE in Table 7. It follows from Table 7 that our proposed model produces the most accurate out-of-sample forecasting results with much smaller RMSE than the other two competing methods.

# 7 Conclusion

In this paper we have introduced a general nonlinear network VAR model for high-dimensional time series, where the momentum and network spillover effects are allowed to change over time and nodes. To achieve dimension reduction and obtain satisfactory estimation convergence rates,

Table 6: The estimated group membership using the UK temperature time series

| Model | Group 1 | Group 2 | | | |
|---|---|---|---|---|---|
| Model (i) | Aberporth | Bradford | Eskdalemuir | Newton Rigg | Stornoway |
| | Armagh | Braemar | Heathrow | Oxford | Sutton Bonington |
| | Ballypatrick | Camborne | Hurn | Paisley | Tiree |
| | Cardiff | Cambridge | Lerwick | Ringway | Waddington |
| | Cwmystwyth | Chivenor | Leuchars | Ross-on-wye | Whitby |
| | Valley | Dunstaffnage | Lowestoft | Shawbury | Wick Airport |
| | | Durham | Manston | Sheffield | Yeovilton |
| | | Eastbourne | Nairn | Southampton | |
| | | | | | |
| Model (ii) | Aberporth | Bradford | Eskdalemuir | Nairn | Sheffield |
| | Armagh | Braemar | Heathrow | Newton Rigg | Southampton |
| | Ballypatrick | Camborne | Hurn | Oxford | Stornoway |
| | Cardiff | Cambridge | Lerwick | Paisley | Sutton Bonington |
| | Cwmystwyth | Chivenor | Leuchars | Ringway | Waddington |
| | Valley | Dunstaffnage | Lowestoft | Ross-on-wye | Wick Airport |
| | Eastbourne | Durham | Manston | Shawbury | Yeovilton |
| | Tiree | | | | |
| | Whitby | | | | |

Table 7: Comparison of out-of-sample RMSE

| Model | $T_{pre} = 12$ | $T_{pre} = 24$ | $T_{pre} = 36$ |
|---|---|---|---|
| Model (i): fully heterogeneous | 0.939 | 1.048 | 1.030 |
| Model (i): grouped+TV | 0.609 | 0.618 | 0.624 |
| Model (i): grouped+linear | 1.113 | 1.102 | 1.086 |
| | | | |
| Model (ii): fully heterogeneous | 0.711 | 0.703 | 0.773 |
| Model (ii): grouped+TV | 0.684 | 0.686 | 0.709 |
| Model (ii): grouped+linear | 1.124 | 1.110 | 1.083 |

we impose a latent group structure on time-varying coefficients in the heterogenous network VAR model. The unknown group number is determined by an easy-to-implement criterion whereas the group membership is estimated by the agglomerative clustering algorithm with the nonparametrically estimated distance matrix. Theorems 3.1 and 3.2 show that the developed methodology consistently estimates the latent group structure. To further improve the convergence rates of the time-varying coefficient estimation, we have proposed a post-grouping local linear smoothing to estimate the group-specific time-varying momentum and network effects. In addition, we further extend the model, methodology and theory to allow for structural breaks in either the group structure or group-specific coefficient functions. The simulation study demonstrates that (i) the developed method can accurately estimate the latent group structure in finite samples; (ii) the post-grouping local linear estimation significantly outperforms the naive heterogenous estimation which ignores the latent structure; and (iii) the developed two-stage method can accurately locate the break point and estimate the group structure before and after the break. The empirical study of the UK temperature time series shows that there exist two groups over the 37 UK weather stations and our proposed method has better out-of-sample prediction performance than the other two competing methods which ignore either the grouped or time-varying feature in network VAR model building.

## Acknowledgements

## A    Cluster analysis of heterogenous network effects

Let $\beta_{ij}(\cdot) = \alpha_{g_i g_j}(\cdot)$ be the heterogenous time-varying network spillover effects defined as in Section 3.1. Due to the homogeneity structure for $\beta_{ij}(\cdot)$, $(i,j) \in \bar{\mathcal{N}}_N$, there exists a partition of the index pair set $\bar{\mathcal{N}}_N$, denoted by $\mathscr{G}^\circ = \{\mathscr{G}_1^\circ, \cdots, \mathscr{G}_{M_0}^\circ\}$, such that

$$\mathscr{G}_i^\circ \cap \mathscr{G}_j^\circ = \emptyset, \quad 1 \leqslant i \neq j \leqslant M_0, \quad \beta_{ij}(\cdot) = \beta_m^\circ(\cdot) \text{ for } (i,j) \in \mathscr{G}_m^\circ, \tag{A.1}$$

where $M_0$ is a finite positive integer upper bounded by $G_0^2$. Neither the group membership $\mathscr{G}^\circ$ nor the group number $M_0$ is known a priori. Define the distance between index pairs $(i_1, j_1)$ and

$(i_2, j_2)$:

$$\widetilde{D}_{(i_1 j_1)(i_2 j_2)} = \frac{1}{L} \sum_{l=1}^{L} \left| \widehat{\beta}_{i_1 j_1}(\tau_l^*) - \widehat{\beta}_{i_2 j_2}(\tau_l^*) \right|,$$

where $\widehat{\beta}_{ij}(\cdot)$ and $\tau_l^*$ are defined in Section 3.1. With $\widetilde{D}_{(i_1 j_1)(i_2 j_2)}$ as the entries, we may further construct an $\bar{N} \times \bar{N}$ distance matrix denoted by $\widehat{\mathbf{D}}$, where $\bar{N} = |\bar{\mathcal{N}}_N|$. Assuming the group number as $M$, we adopt the agglomerative hierarchical clustering algorithm described in Section 3.1 with the distance matrix $\widehat{\mathbf{D}}$, terminate it when the group number reaches $M$, and denote the resulting group estimates as $\widetilde{\mathscr{G}}_{1|M}^{\circ}, \cdots, \widetilde{\mathscr{G}}_{M|M}^{\circ}$.

As in Section 3.2, we pool the estimated heterogenous time-varying network spillover effects over $(i, j) \in \widetilde{\mathscr{G}}_{m|M}^{\circ}$:

$$\widetilde{\beta}_{m|M}(\tau) = \frac{1}{\left| \widetilde{\mathscr{G}}_{m|M}^{\circ} \right|} \sum_{(i,j) \in \widetilde{\mathscr{G}}_{m|M}^{\circ}} \widehat{\beta}_{ij}(\tau),$$

and define the average deviation:

$$\widetilde{R}(M) = \frac{1}{ML} \sum_{m=1}^{M} \frac{1}{\left| \widetilde{\mathscr{G}}_{m|M}^{\circ} \right|} \sum_{(i,j) \in \widetilde{\mathscr{G}}_{m|M}^{\circ}} \sum_{l=1}^{L} \left| \widehat{\beta}_{ij}(\tau_l^*) - \widetilde{\beta}_{m|M}(\tau_l^*) \right|.$$

The group number $M_0$ is estimated by the following ratio criterion:

$$\widehat{M} = \operatorname*{argmin}_{1 \leqslant M \leqslant \overline{M}} \frac{\widetilde{R}(M)}{\widetilde{R}(M-1)}, \tag{A.2}$$

where $\overline{M}$ is a pre-specified positive integer. With the above consistent group number number estimate, we again run the agglomerative clustering algorithm, terminate it when the group number reaches $\widehat{M}$, and denote the final estimates of group membership by $\widehat{\mathscr{G}}_m^{\circ}$, $m = 1, \cdots, \widehat{M}$, from which we obtain the group label estimates $\widehat{g}_{ij} = m$ if $(i, j) \in \widehat{\mathscr{G}}_m^{\circ} \cap \bar{\mathcal{N}}_N$.

**Proposition A.1.** *Suppose that the conditions of Theorems 3.1 and 3.2 hold. Then*

$$\mathsf{P}\left(\widehat{M} = M_0\right) \to 1, \quad \mathsf{P}\left(\{\widehat{\mathscr{G}}_m^{\circ}, 1 \leqslant m \leqslant \widehat{M}\} = \{\mathscr{G}_m^{\circ}, 1 \leqslant m \leqslant M_0\}\right) \to 1. \tag{A.3}$$

The above proposition establishes the consistency properties of the group number estimate $\widehat{M}$ and the membership estimate $\widehat{\mathscr{G}}_m^{\circ}$, $m = 1, \cdots, \widehat{M}$. The finite-sample performance of these estimators is reported in Appendix F of the online supplement (Li *et al* 2024).

# References

Ando, T. and Bai, J. (2017). Clustering huge number of time series: A panel data approach with high-dimensional predictors and factor structures. *Journal of the American Statistical Association* **112**, 1182–1198.

Bai, J., Li, K. and Lu, L. (2016). Estimation and inference of FAVAR models. *Journal of Business and Economic Statistics* **34**, 620–641.

Bai, J. and Ng, S. (2006). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica* **74**, 1135–1150.

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

Bernanke, B., Boivin, J. and Eliasz, P. S. (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach. *Quarterly Journal of Economics* **120**, 387–422.

Bonhomme, S. and Manresa, E. (2015). Grouped patterns of heterogeneity in panel data. *Econometrica* **83**, 1147–1184.

Cai, Z. (2007). Trending time-varying coefficient time series models with serially correlated errors. *Journal of Econometrics* **136**, 163–188.

Chen, E., Fan, J. and Zhu, X. (2023). Community network autoregression for high-dimensional time series. *Journal of Econometrics* **235**, 1239–1256.

Chen, J. (2019). Estimating latent group structure in time-varying coefficient panel data models. *Econometrics Journal* **22**, 223–240.

Chen, J., Li, D., Li, Y. and Linton, O. (2023). Estimating time-varying networks for high-dimensional time series. Working paper available at https://arxiv.org/abs/2302.02476.

Chen, L. and Wu. W. (2019). Testing for trends in high-dimensional time series. *Journal of the American Statistical Association* **114**, 869–881.

Chen, L., Wang, W. and Wu. W. (2022). Inference of breakpoints in high-dimensional time series. *Journal of the American Statistical Association* **117**,1951–1963.

Cho, H. and Fryzlewicz P. (2012). Multiscale and multilevel technique for consistent segmentation of nonstationary time series. *Statistica Sinica* **22**, 207–229.

Cho, H. and Fryzlewicz P. (2015). Multiple change-point detection for high-dimensional time series via Sparsified Binary Segmentation. *Journal of the Royal Statistical Society Series B* **77**, 475–507.

Davis, R., Zang, P. and Zheng, T. (2016). Sparse vector autoregressive modeling. *Journal of Computational and Graphical Statistics* **25**, 1077–1096.

Ding, X., Qiu, Z. and Chen, X. (2017). Sparse transition matrix estimation for high-dimensional and locally stationary vector autoregressive models. *Electronic Journal of Statistics* **11**, 3871–3902.

Everitt, B. S., Landau, S., Leese, M. and Stahl, D. (2011). *Cluster Analysis* (5th Edition). Wiley Series in Probability and Statistics, Wiley.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman and Hall/CRC.

Guðmundsson, G. and Brownlees C. (2021). Detecting groups in large vector autoregressions. *Journal of Econometrics* **225**, 2–26.

Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference and Prediction* (2nd Edition). Springer-Verlag, New York.

Hanlon, H.M., Bernie, D., Carigi, G. and Lowe, J.A. (2021). Future changes to high impact weather in the UK. *Climatic Change* **166**, 50.

Kilian, K. and Lütkepohl, H. (2017). *Structural Vector Autoregressive Analysis*. Cambridge University Press.

Ke, Y., Li, J. and Zhang, W. (2016). Structure identification in panel data analysis. *The Annals of Statistics* **44**, 1193–1233.

Kendon, M., McCarthy, M., Jevrejeva, S., Matthews, A., Sparks, T. and Garforth, J. (2020). State of the UK Climate 2019. *International Journal of Climatology* **40**, 1–69.

Kock, A.B. and Callot, L. (2015). Oracle inequalities for high dimensional vector autoregressions. *Journal of Econometrics* **186**, 325–344.

Lam, C. and Yao, Q. (2012). Factor modelling for high-dimensional time series: Inference for the number of factor. *The Annals of Statistics*, **40**, 694–726.

LI, D., PENG, B., TANG, S. AND WU, W. (2024). Supplement to "Inference of grouped time-varying network vector autoregression models".

Li, D., Robinson, P. and Shang, H. (2020). Long-range dependent curve time series. *Journal of the American Statistical Association*, **115**, 957–971.

Lumsdaine, R., Okui, R. and Wang W. (2023). Estimation of panel group structure models with structural breaks in group memberships and coefficients. *Journal of Econometrics*, **233**, 45–65.

Lütkepohl, H. (2006). *New Introduction to Multiple Time Series Analysis*. Springer.

Miao, K., Phillips, P.C.B. & Su, L. (2023). High-dimensional VARs with common factors. *Journal of Econometrics*, **233**, 155–183.

Portmann, R. W., Solomon, S. & Hegerl, G. C. (2009). Spatial and seasonal patterns in climate change, temperatures, and precipitation across the United States. *Proceedings of the National Academy of Sciences* **106**, 7324-7329.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica* **48**, 1–48.

Song, S. and Bickel, P. (2011). Large vector auto regressions. Working paper available at https://arxiv.org/abs/1106.3915.

Su, L., Shi, Z. and Phillips, P. C. B. (2016). Identifying latent structures in panel data. *Econometrica* **84**, 2215–2264.

SU, L. & WANG, X. (2017). On time-varying factor models: estimation and testing. *Journal of Econometrics* **198**, 84–101.

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

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

Tsybakov, A. (2008). *Introduction to Nonparametric Estimation*. Springer Series in Statistics, Springer.

Vogt, M. (2012). Nonparametric regression for locally stationary time series. *The Annals of Statistics* **40**, 2601–2633.

Vogt, M. and Linton, O. (2017). Classification of nonparametric regression functions in longitudinal data models. *Journal of the Royal Statistical Society, Series B* **79**, 5–27.

Vogt, M. and Linton, O. (2020). Multiscale clustering of nonparametric regression curves. *Journal of Econometrics* **216**, 305–325.

Wang, Y., Phillips, P.C.B. and Su, L. (2023). Panel data models with time-varying latent group structures. Working paper available at https://arxiv.org/abs/2307.15863.

Wu, B. (2019). Time-varying network vector autoregression model. Working paper available at http://dx.doi.org/10.2139/ssrn.3491608.

Wu, W. (2005). Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences* **102**, 14150–14154.

Xu, M., Chen, X. and Wu, W. (2020). Estimation of dynamic networks for high-dimensional nonstationary time series. *Entropy* **22**, 55.

Yang, X., Chen, J., Li, D. and Li, R. (2023). Functional-coefficient quantile regression for panel data with latent group structure. Forthcoming in *Journal of Business and Economic Statistics*.

Yin, H., Safikhani, A. and Michailidis, G. (2023). A general modeling framework for network autoregressive processes. *Technometrics* **65**, 579–589.

Yin, H., Safikhani, A. and Michailidis, G. (2024). A functional coefficients network autoregressive model. Forthcoming in *Statistica Sinica*.

Zhang, D. and Wu, W. (2021). Convergence of covariance and spectral density estimates for high-dimensional locally stationary processes. *The Annals of Statistics* **49**, 233–254.

Zhang, T. (2013). Clustering high-dimensional time series based on parallelism. *Journal of the American Statistical Association* **108**, 577–588.

Zhu, X. and Pan, R. (2020). Grouped network vector autoregression. *Statistica Sinica* **30**, 1437–1462.

Zhu, X., Pan, R., Li, G., Liu, Y. and Wang, H. (2017). Network vector autoregression. *The Annals of Statistics* **45**, 1096–1123.

Zhu, X., Xu, G. and Fan, J. (2023). Simultaneous estimation and group identification for network vector autoregressive model with heterogeneous nodes. Forthcoming in *Journal of Econometrics*.

# Online Supplement to "Estimation of Grouped Time-Varying Network Vector Autoregression Models"

This supplemental material contains five appendices: Appendix B gives the proof of (2.8) in Remark 1(ii) and discusses connection of the proposed model to the nonlinear functional dependence measure; Appendix C provides the proofs of main theorems; Appendix D includes the proofs of some technical lemmas; Appendix E discusses a refined estimation of the break location; and Appendix F reports some extra numerical results. Unless explicitly stated differently, all the assumptions, equations, propositions, remarks, theorems and sections mentioned in this supplemental material refer to those presented in the main text. Throughout the supplement, we let $C, C_1, C_2, \cdots$ denote some generic positive constants.

## B  Theoretical justification of Remark 1(ii)

In this appendix, we first provide the proof of (2.8) in Remark 1(ii) and then discuss the nonlinear functional dependence measure introduced by Wu (2005).

**Proof of (2.8) in Remark 1(ii).** The proof is similar to the proof of Lemma A.1 in Zhang & Wu (2021). It follows from (2.4) in Assumption 1(i) that there exists $0 < \chi_0 < 1$ such that

$$
\begin{aligned}
\left\| \left| X_t - X_t^\circ \right|_\infty \right\|_q &= \left\| \left| \mathbf{B}(\tau_t) \left[ X_{t-1} - X_{t-1}^\circ(\tau_t) \right] \right|_\infty \right\|_q \\
&\leqslant \chi_0 \left\| \left| X_{t-1} - X_{t-1}^\circ(\tau_t) \right|_\infty \right\|_q .
\end{aligned}
\tag{B.1}
$$

Letting

$$
\Xi_{t-1,1} = \left\| \left| X_{t-1} - X_{t-1}^\circ(\tau_{t-1}) \right|_\infty \right\|_q \quad \text{and} \quad \Xi_{t-1,2} = \left\| \left| X_{t-1}^\circ(\tau_{t-1}) - X_{t-1}^\circ(\tau_t) \right|_\infty \right\|_q ,
$$

and using (B.1) and the triangle inequality, we obtain

$$
\left\| \left| X_t - X_t^\circ \right|_\infty \right\|_q \leqslant \chi_0 \left( \Xi_{t-1,1} + \Xi_{t-1,2} \right) .
\tag{B.2}
$$

For any $0 \leqslant \tau, \tau^* \leqslant 1$, by the smoothness condition in Assumption 1(i) and (2.7), we may show that

$$
\begin{aligned}
\left\| \left| X_t^\circ(\tau) - X_t^\circ(\tau^*) \right|_\infty \right\|_q &= \left\| \left| \mathbf{B}(\tau) X_{t-1}^\circ(\tau) - \mathbf{B}(\tau^*) X_{t-1}^\circ(\tau^*) \right|_\infty \right\|_q \\
&\leqslant \left\| \left| \mathbf{B}(\tau) X_{t-1}^\circ(\tau) - \mathbf{B}(\tau^*) X_{t-1}^\circ(\tau) \right|_\infty \right\|_q + \\
&\quad \left\| \left| \mathbf{B}(\tau^*) X_{t-1}^\circ(\tau) - \mathbf{B}(\tau^*) X_{t-1}^\circ(\tau^*) \right|_\infty \right\|_q \\
&\leqslant C_1 \left\| \left| X_{t-1}^\circ(\tau) \right|_\infty \right\|_q |\tau - \tau^*| + \chi_0 \left\| \left| X_{t-1}^\circ(\tau) - X_{t-1}^\circ(\tau^*) \right|_\infty \right\|_q \\
&\leqslant C_1 \theta_{N,q} |\tau - \tau^*| + \chi_0 \left\| \left| X_{t-1}^\circ(\tau) - X_{t-1}^\circ(\tau^*) \right|_\infty \right\|_q ,
\end{aligned}
$$

indicating that

$$
\left\| \left| X_t^\circ(\tau) - X_t^\circ(\tau^*) \right|_\infty \right\|_q \leqslant \frac{C_1 \theta_{N,q} |\tau - \tau^*|}{1 - \chi_0}
$$

1

and thus

$$\Xi_{t-1,2} \leqslant \frac{C_1}{1-\chi_0} \cdot \frac{\theta_{N,q}}{T}. \tag{B.3}$$

By virtue of (B.2) and (B.3), we readily have that

$$\left\| |X_t - X_t^\circ|_\infty \right\|_q \leqslant \chi_0 \left\| |X_{t-1} - X_{t-1}^\circ|_\infty \right\|_q + \frac{C_1 \chi_0}{1-\chi_0} \cdot \frac{\theta_{N,q}}{T}. \tag{B.4}$$

With (B.4), using the argument similar to the proof of Lemma 4.5 in Dahlhaus, Richter & Wu (2019), we have

$$\max_{1 \leqslant t \leqslant T} \left\| |X_t - X_t^\circ|_\infty \right\|_q \leqslant \frac{C_1 \chi_0}{(1-\chi_0)^2} \cdot \frac{\theta_{N,q}}{T},$$

completing the proof of (2.8). $\qquad\qquad\square$

We next connect the grouped time-varying network VAR model to the nonlinear functional dependence measure introduced by Wu (2005), facilitating the development of our main asymptotic theory. Let $\{\varepsilon_t^*\}$ be an independent copy of $\{\varepsilon_t\}$ and $\mathscr{F}_t^{\{l\}} = (\cdots, \varepsilon_{l-1}, \varepsilon_l^*, \varepsilon_{l+1}, \cdots, \varepsilon_t)$ be a coupled version of $\mathscr{F}_t$ replacing $\varepsilon_l$ by $\varepsilon_l^*$. Letting $X_t^{\circ\{l\}}(\tau) = G(\tau, \mathscr{F}_t^{\{l\}})$, as in Zhang & Wu (2021), we define the node-wise functional dependence measure:

$$\delta_{i,t,q} = \sup_{\tau \in [0,1]} \left\| x_{i,t}^\circ(\tau) - x_{i,t}^{\circ\{0\}}(\tau) \right\|_q,$$

where $x_{i,t}^\circ(\tau)$ and $x_{i,t}^{\circ\{0\}}(\tau)$ are the $i$-th element of $X_t^\circ(\tau)$ and $X_t^{\circ\{0\}}(\tau)$, respectively. Furthermore, we construct the node-wise dependence adjusted norm:

$$\|x_{i\bullet}\|_{q,\iota} = \sup_{m \geqslant 0} (m+1)^\iota \Delta_{i,m,q}, \quad \Delta_{i,m,q} = \sum_{t=m}^{\infty} \delta_{i,t,q},$$

where $\iota \geqslant 0$ depicts the decay rate of the cumulative dependence measure $\Delta_{i,m,q}$. Letting $\mathbf{B}_{i\bullet}^j(\tau)$ be the $i$-th row vector of $\mathbf{B}^j(\tau)$, by (2.4) in Assumption 1(i), $\sup_{\tau \in [0,1]} |\mathbf{B}_{i\bullet}^j(\tau)|_2$ decays at a geometric rate of $j$ and $\delta_{i,t,q}$ decays at a geometric rate of $t$ for all $i$. For simplicity, we may set $\iota = 1$ and write $\|x_{i\bullet}\|_q = \|x_{i\bullet}\|_{q,1}$. Then we have $\max_{1 \leqslant i \leqslant N} \|x_{i\bullet}\|_q \leqslant C_2$, where $C_2$ is a positive constant.

# C  Proofs of the main asymptotic theorems

We next provide the detailed proofs of the main asymptotic theorems in the main text.

## C.1  Proof of Theorem 3.1

To prove (3.10), we only need to show that

$$P\left( \max_{1 \leqslant k \leqslant K_0} \max_{i,j \in \mathscr{G}_k} \widehat{D}_{ij} < \min_{1 \leqslant k \neq l \leqslant K_0} \min_{i \in \mathscr{G}_k, j \in \mathscr{G}_l} \widehat{D}_{ij} \right) \to 1 \tag{C.1}$$

as $T$ tends to infinity. Define

$$D_{ij} = \int_0^1 d_{ij}(\tau) d\tau \text{ with } d_{ij}(\tau) = \left| \beta_i(\tau) - \beta_j(\tau) \right| + \left| \beta_{i\bullet}^\circ(\tau) - \beta_{j\bullet}^\circ(\tau) \right|_2.$$

Notice that $D_{ij} = 0$ if $i, j \in \mathscr{G}_k$, and

$$\zeta_{NT}^\dagger = \min_{1 \leqslant k \neq l \leqslant K_0} \min_{i \in \mathscr{G}_k, j \in \mathscr{G}_l} D_{ij} > 0.$$

Hence, to prove (C.1), it is sufficient to show that

$$\max_{1 \leqslant i \neq j \leqslant N} \left| \widehat{D}_{ij} - D_{ij} \right| = o_P \left( \zeta_{NT}^\dagger \right). \tag{C.2}$$

Letting $\overline{D}_{ij} = \frac{1}{L} \sum_{l=1}^L d_{ij}(\tau_l^*)$, we have

$$\left| \widehat{D}_{ij} - D_{ij} \right| \leqslant \left| \widehat{D}_{ij} - \overline{D}_{ij} \right| + \left| \overline{D}_{ij} - D_{ij} \right|.$$

By the definition of the Riemann integral and Assumption 1(i), we readily have that

$$\max_{1 \leqslant i \neq j \leqslant N} \left| \overline{D}_{ij} - D_{ij} \right| = O\left(1/L\right) = o\left( \zeta_{NT}^\dagger \right), \tag{C.3}$$

where $1/L = o(\zeta_{NT}^\dagger)$ in Assumption 3 has been used.

By Proposition A.1, without loss of generality, we may assume that $\widehat{M} = M_0$ and $\widehat{\mathscr{G}}_m^\circ = \mathscr{G}_m^\circ$, $m = 1, \cdots, M_0$, hold *w.p.a.1* in the remaining proof. By the triangle inequality, we have

$$\begin{aligned}
\left| \widehat{D}_{ij} - \overline{D}_{ij} \right| &= \left| \frac{1}{L} \sum_{l=1}^L \left[ \widehat{d}_{ij}(\tau_l^*) - d_{ij}(\tau_l^*) \right] \right| \\
&\leqslant \frac{1}{L} \sum_{l=1}^L \left| \left| \widehat{\beta}_i(\tau_l^*) - \widehat{\beta}_j(\tau_l^*) \right| - \left| \beta_i(\tau_l^*) - \beta_j(\tau_l^*) \right| \right| + \\
&\quad \frac{1}{L} \sum_{l=1}^L \left| \left| \widehat{\beta}_{i\bullet}^\circ(\tau_l^*) - \widehat{\beta}_{j\bullet}^\circ(\tau_l^*) \right|_2 - \left| \beta_{i\bullet}^\circ(\tau_l^*) - \beta_{j\bullet}^\circ(\tau_l^*) \right|_2 \right| \\
&\leqslant \frac{1}{L} \sum_{l=1}^L \left| \widehat{\beta}_i(\tau_l^*) - \beta_i(\tau_l^*) \right| + \frac{1}{L} \sum_{l=1}^L \left| \widehat{\beta}_j(\tau_l^*) - \beta_j(\tau_l^*) \right| + \\
&\quad \frac{1}{L} \sum_{l=1}^L \left| \widehat{\beta}_{i\bullet}^\circ(\tau_l^*) - \beta_{i\bullet}^\circ(\tau_l^*) \right|_2 + \frac{1}{L} \sum_{l=1}^L \left| \widehat{\beta}_{j\bullet}^\circ(\tau_l^*) - \beta_{j\bullet}^\circ(\tau_l^*) \right|_2,
\end{aligned}$$

which, together with Lemmas D.3 and D.4 and Assumption 3, leads to

$$\max_{1 \leqslant i \neq j \leqslant N} \left| \widehat{D}_{ij} - \overline{D}_{ij} \right| = O_P \left( \sqrt{\frac{\bar{n} \log(N \vee T)}{Th}} + \sqrt{\bar{n}} h^2 \right) = o_P \left( \zeta_{NT}^\dagger \right). \tag{C.4}$$

3

With (C.3) and (C.4), we complete the proof of (C.2). □

## C.2 Proof of Theorem 3.2

Let $\mathscr{E}_{\mathscr{G}}$ denote the event that $\{\widehat{\mathscr{G}}_1, \cdots, \widehat{\mathscr{G}}_{K_0}\} = \{\mathscr{G}_1, \cdots, \mathscr{G}_{K_0}\}$. It follows from Theorem 3.1 that $\mathsf{P}(\mathscr{E}_{\mathscr{G}}) \to 1$. Hence, to prove Theorem 3.2, it is sufficient to show that

$$\mathsf{P}\left(\widehat{K} = K_0 \mid \mathscr{E}_{\mathscr{G}}\right) \to 1. \tag{C.5}$$

By the definition of $\widehat{K}$, we only need to show that

$$\mathsf{P}\left(\frac{\widehat{R}(K_0)}{\widehat{R}(K_0 - 1)} = \min_{1 \leqslant K \leqslant \overline{K}} \frac{\widehat{R}(K)}{\widehat{R}(K - 1)} \mid \mathscr{E}_{\mathscr{G}}\right) \to 1. \tag{C.6}$$

To prove (C.6), we next consider the following two scenarios: (i) $1 \leqslant K \leqslant K_0 - 1$ and (ii) $K_0 + 1 \leqslant K \leqslant \overline{K}$, corresponding to the under-fitted and over-fitted grouped time-varying network VAR models, respectively. For the case $K_0 + 1 \leqslant K \leqslant \overline{K}$, conditional on the event $\mathscr{E}_{\mathscr{G}}$, by Lemma D.5(i), we have $\widehat{R}(K) = o_\mathsf{P}(\rho_{NT})$ for $K = K_0 + 1, \cdots, \overline{K}$. Consequently, we may set $\widehat{R}(K) \equiv 0$ *w.p.a.1*, and

$$\mathsf{P}\left(\frac{\widehat{R}(K)}{\widehat{R}(K - 1)} = \frac{0}{0} \equiv 1, \ \ K = K_0 + 1, \cdots, \overline{K} \mid \mathscr{E}_{\mathscr{G}}\right) \to 1. \tag{C.7}$$

For the case $1 \leqslant K \leqslant K_0 - 1$, conditional on $\mathscr{E}_{\mathscr{G}}$, by Lemma D.5(ii), we have $\widehat{R}(K) \geqslant \underline{c}\zeta_{NT}^\dagger$ *w.p.a.1* with $\underline{c}$ being a positive constant strictly larger than zero. Hence, there exists a positive constant $\underline{c}_*$ such that

$$\mathsf{P}\left(\frac{\widehat{R}(K)}{\widehat{R}(K - 1)} \geqslant \underline{c}_*, \ \ K = 1, \cdots, K_0 - 1 \mid \mathscr{E}_{\mathscr{G}}\right) \to 1, \tag{C.8}$$

setting $\frac{\widehat{R}(1)}{\widehat{R}(0)} = 1$, and

$$\mathsf{P}\left(\frac{\widehat{R}(K_0)}{\widehat{R}(K_0 - 1)} = 0 \mid \mathscr{E}_{\mathscr{G}}\right) \to 1. \tag{C.9}$$

With (C.7)–(C.9), we prove (C.6), completing the proof of Theorem 3.2. □

## C.3 Proof of Theorem 4.1

By the consistency properties in Theorem 3.2 and Corollary 3.3, we may prove the asymptotic distribution theory conditional on $\widehat{K} = K_0$ and $\widehat{\mathscr{G}}_k = \mathscr{G}_k$, $k = 1, \cdots, K_0$. For $k = 1, \cdots, K_0$, let

$$\Xi_{kX}(\tau) = \frac{1}{\operatorname{card}(\mathscr{G}_k)Th_\dagger} \sum_{t=1}^T \sum_{i \in \mathscr{G}_k} \left[ \begin{array}{cc} X_{i,t-1}^\diamond X_{i,t-1}^{\diamond\intercal} K_{t0}^\dagger(\tau) & X_{i,t-1}^\diamond X_{i,t-1}^{\diamond\intercal} K_{t1}^\dagger(\tau) \\ X_{i,t-1}^\diamond X_{i,t-1}^{\diamond\intercal} K_{t1}^\dagger(\tau) & X_{i,t-1}^\diamond X_{i,t-1}^{\diamond\intercal} K_{t2}^\dagger(\tau) \end{array} \right],$$

4

$$\Xi_{k\alpha}(\tau) = \frac{1}{\text{card}(\mathscr{G}_k)Th_\dagger} \sum_{t=1}^{T} \sum_{i\in\mathscr{G}_k} \left[ \begin{array}{c} X_{i,t-1}^\diamond X_{i,t-1}^{\diamond\mathsf{T}} \left[ \alpha_{k\bullet}(\tau_t) - \alpha_{k\bullet}(\tau) - \alpha'_{k\bullet}(\tau)(\tau_t - \tau) \right] K_{t0}^\dagger(\tau) \\ X_{i,t-1}^\diamond X_{i,t-1}^{\diamond\mathsf{T}} \left[ \alpha_{k\bullet}(\tau_t) - \alpha_{k\bullet}(\tau) - \alpha'_{k\bullet}(\tau)(\tau_t - \tau) \right] K_{t1}^\dagger(\tau) \end{array} \right],$$

$$\Xi_{k\varepsilon}(\tau) = \frac{1}{\text{card}(\mathscr{G}_k)Th_\dagger} \sum_{t=1}^{T} \sum_{i\in\mathscr{G}_k} \left[ \begin{array}{c} X_{i,t-1}^\diamond \varepsilon_{i,t} K_{t0}^\dagger(\tau) \\ X_{i,t-1}^\diamond \varepsilon_{i,t} K_{t1}^\dagger(\tau) \end{array} \right], \quad K_{t\kappa}^\dagger(\tau) = \left( \frac{\tau_t - \tau}{h_\dagger} \right)^\kappa K \left( \frac{\tau_t - \tau}{h_\dagger} \right).$$

By the definition of the post-grouping local linear estimation defined in (4.2), we have

$$\check{\alpha}_{k\bullet}(\tau) - \alpha_{k\bullet}(\tau) = \mathbf{E}_\bullet \cdot \Xi_{kX}^{-1}(\tau) \left[ \Xi_{k\alpha}(\tau) + \Xi_{k\varepsilon}(\tau) \right]$$

conditional on $\widehat{K} = K_0$ and $\widehat{\mathscr{G}}_k = \mathscr{G}_k$, $k = 1, \cdots, K_0$, where $\mathbf{E}_\bullet = \left( \mathbf{I}_{K_0+1}, \mathbf{O}_{(K_0+1)\times(K_0+1)} \right)$. In order to prove (4.4), we only need to show that

$$\Xi_{kX}(\tau) \overset{P}{\longrightarrow} \text{diag}\{1, \mu_2\} \otimes \mathbf{\Delta}_{\mathscr{G}_k}(\tau), \tag{C.10}$$

$$\Xi_{k\alpha}(\tau) = \left[ \begin{array}{c} \frac{1}{2}h_\dagger^2 \mu_2 \mathbf{\Delta}_{\mathscr{G}_k}(\tau)\alpha''_{k\bullet}(\tau) \\ \mathbf{0} \end{array} \right] + o_P(h_\dagger^2), \tag{C.11}$$

$$\left[ \text{card}(\mathscr{G}_k)Th_\dagger \right]^{1/2} \Xi_{k\varepsilon}(\tau) \overset{d}{\longrightarrow} \mathsf{N} \left( \mathbf{0}, \text{diag}\{\nu_0, \nu_2\} \otimes \mathbf{\Upsilon}_{\mathscr{G}_k}(\tau) \right), \tag{C.12}$$

where $\otimes$ denotes the Kronecker product between matrices.

By Lemma D.6 and the approximation result (2.8), we can prove (C.10). By the smoothness condition in Assumption 1(i), Taylor's expansion of $\alpha_{k\bullet}(\cdot)$ and (C.10), we can prove (C.11). We next turn to the proof of (C.12). Let

$$W_{\mathscr{G}_k,t}(\tau) = \left[ W_{\mathscr{G}_k,t}^\mathsf{T}(\tau, 0), W_{\mathscr{G}_k,t}^\mathsf{T}(\tau, 1) \right]^\mathsf{T}$$

with

$$W_{\mathscr{G}_k,t}(\tau, \kappa) = \frac{1}{\sqrt{\text{card}(\mathscr{G}_k)}} \sum_{i\in\mathscr{G}_k} \varepsilon_{i,t} X_{i,t-1}^\diamond K_{t\kappa}^\dagger(\tau),$$

and write

$$\sqrt{\text{card}(\mathscr{G}_k)Th_\dagger} \Xi_{k\varepsilon}(\tau) = \frac{1}{\sqrt{Th_\dagger}} \sum_{t=1}^{T} W_{\mathscr{G}_k,t}(\tau).$$

Using Lemma D.6 and the approximation (2.8), we have

$$\frac{1}{Th_\dagger} \sum_{t=1}^{T} \mathsf{E} \left[ W_{\mathscr{G}_k,t}(\tau, 0) W_{\mathscr{G}_k,t}^\mathsf{T}(\tau, 0) | \mathscr{F}_{t-1} \right]$$

$$= \frac{1}{\text{card}(\mathscr{G}_k)} \sum_{i,j\in\mathscr{G}_k} \sigma_{ij} \left[ \frac{1}{Th_\dagger} \sum_{t=1}^{T} X_{i,t-1}^\diamond X_{i,t-1}^{\diamond\mathsf{T}} K_{t,0}^\ddagger(\tau) \right]$$

$$\overset{P}{\longrightarrow} \frac{\nu_0}{\text{card}(\mathscr{G}_k)} \sum_{i,j\in\mathscr{G}_k} \sigma_{ij} \mathbf{\Delta}_{ij}^\diamond(\tau).$$

Similarly, we can also prove that

$$\frac{1}{Th_\dagger} \sum_{t=1}^{T} E\left[W_{\mathscr{G}_k,t}(\tau,1)W_{\mathscr{G}_k,t}^\mathsf{T}(\tau,1)|\mathcal{F}_{t-1}\right] \xrightarrow{P} \frac{\nu_2}{\mathrm{card}(\mathscr{G}_k)} \sum_{i,j\in\mathscr{G}_k} \sigma_{ij}\Delta_{ij}^\diamond(\tau),$$

and

$$\frac{1}{Th_\dagger} \sum_{t=1}^{T} E\left[W_{\mathscr{G}_k,t}(\tau,0)W_{\mathscr{G}_k,t}^\mathsf{T}(\tau,1)|\mathcal{F}_{t-1}\right] = o_P(1).$$

Hence, we have

$$\frac{1}{Th_\dagger} \sum_{t=1}^{T} E\left[W_{\mathscr{G}_k,t}(\tau)W_{\mathscr{G}_k,t}^\mathsf{T}(\tau)|\mathcal{F}_{t-1}\right] \xrightarrow{P} \mathrm{diag}\{\nu_0,\nu_2\} \otimes \frac{1}{\mathrm{card}(\mathscr{G}_k)} \sum_{i,j\in\mathscr{G}_k} \sigma_{ij}\Delta_{ij}^\diamond(\tau),$$

which, together with (4.3) and the martingale central limit theorem (e.g., Hall & Heyde 1980), leads to (C.12). The proof of Theorem 4.1 is completed. $\qquad\square$

## C.4 Proof of Theorem 5.1

Let

$$\overline{\Gamma}_{it}^{l,\varepsilon} = \frac{1}{Th_\ddagger} \sum_{s=1}^{T} \widetilde{X}_{i,s-1}\varepsilon_{i,s}K^\ddagger\left(\frac{\tau_t - \tau_s}{h_\ddagger}\right), \quad \overline{\Gamma}_{it}^{r,\varepsilon} = \frac{1}{Th_\ddagger} \sum_{s=1}^{T} \widetilde{X}_{i,s-1}\varepsilon_{i,s}K^\ddagger\left(\frac{\tau_s - \tau_t}{h_\ddagger}\right),$$

$$\overline{\beta}_{i\bullet}^{\ddagger,l}(\tau_t) = \left(\widetilde{\Gamma}_{it}^l\right)^{-1}\left(\overline{\Gamma}_{it}^l - \overline{\Gamma}_{it}^{l,\varepsilon}\right), \quad \overline{\beta}_{i\bullet}^{\ddagger,r}(\tau_t) = \left(\widetilde{\Gamma}_{it}^r\right)^{-1}\left(\overline{\Gamma}_{it}^r - \overline{\Gamma}_{it}^{r,\varepsilon}\right).$$

Define

$$\overline{\delta}_{\beta,i}(t) = \overline{\beta}_{i\bullet}^{\ddagger,r}(\tau_t) - \overline{\beta}_{i\bullet}^{\ddagger,l}(\tau_t) \quad \text{and} \quad \delta_{\beta,i}(t) = \beta_{i\bullet}^{\ddagger,r}(\tau_t) - \beta_{i\bullet}^{\ddagger,l}(\tau_t).$$

By Assumption 2(ii) and 6(i)(ii), we may show that

$$\overline{\delta}_{\beta,i}(t) = \widetilde{\delta}_{\beta,i}(t) + O_P(h_\ddagger) \tag{C.13}$$

uniformly over $i = 1, \cdots, N$, where

$$\widetilde{\delta}_{\beta,i}(t) = \begin{cases} 0, & t > t_0 + Th_\ddagger \text{ or } t < t_0 - Th_\ddagger, \\ \left[1 - g_i^l\left(\frac{t_0-t}{Th_\ddagger}\right)\right]\delta_{\beta,i}(t_0), & t_0 - Th_\ddagger \leqslant t < t_0, \\ \left[1 - g_i^r\left(\frac{t-t_0}{Th_\ddagger}\right)\right]\delta_{\beta,i}(t_0), & t_0 < t \leqslant t_0 + Th_\ddagger, \\ \delta_{\beta,i}(t_0), & t = t_0, \end{cases}$$

and $g_i^l(\cdot)$ and $g_i^r(\cdot)$ are positive functions satisfying $g_i^l(0) = g_i^r(0) = 0$, $g_i^l(1) = g_i^r(1) = 1$ and

$$\min_{1\leqslant i\leqslant N}\{g_i^l(x), g_i^r(x)\} \geqslant \underline{c}_0 x, \quad 0 < x < 1,$$

6

with $\underline{c}_0$ being a positive constant.

Writing
$$\overline{\delta}_\beta(t) = \left[\overline{\delta}_{\beta,1}(t), \cdots, \overline{\delta}_{\beta,N}(t)\right]^\intercal \quad \text{and} \quad \widetilde{\delta}_\beta(t) = \left[\widetilde{\delta}_{\beta,1}(t), \cdots, \widetilde{\delta}_{\beta,N}(t)\right]^\intercal,$$
by (C.13) and the triangle inequality, we have
$$\begin{aligned}
\left|\overline{\delta}_\beta(t_0)\right|_\infty - \left|\overline{\delta}_\beta(t)\right|_\infty &\geqslant \left|\widetilde{\delta}_\beta(t_0)\right|_\infty - \left|\widetilde{\delta}_\beta(t)\right|_\infty - O_P(h_\ddagger) \\
&\geqslant \underline{c}_0 \left|\frac{t - t_0}{Th_\ddagger}\right| \cdot \delta_\beta(t_0) - O_P(h_\ddagger)
\end{aligned} \tag{C.14}$$
when $t_0 - Th_\ddagger \leqslant t \leqslant t_0 + Th_\ddagger$, and otherwise
$$\left|\overline{\delta}_\beta(t_0)\right|_\infty - \left|\overline{\delta}_\beta(t)\right|_\infty \geqslant \left|\delta_\beta(t_0)\right|_\infty - O_P(h_\ddagger).$$

Following the proofs of Lemma D.1 and (D.28) in Appendix D, we may show that
$$\max_{1 \leqslant i \leqslant N} \max_{1 \leqslant t \leqslant T} \left|\left(\widetilde{\Gamma}_{it}^\mathsf{l}\right)^{-1} \overline{\Gamma}_{it}^{\mathsf{l},\varepsilon}\right|_2 = O_P\left(\sqrt{\frac{\bar{n}\log(N \vee T)}{Th_\ddagger}}\right), \tag{C.15}$$
$$\max_{1 \leqslant i \leqslant N} \max_{1 \leqslant t \leqslant T} \left|\left(\widetilde{\Gamma}_{it}^\mathsf{r}\right)^{-1} \overline{\Gamma}_{it}^{\mathsf{r},\varepsilon}\right|_2 = O_P\left(\sqrt{\frac{\bar{n}\log(N \vee T)}{Th_\ddagger}}\right). \tag{C.16}$$

By the definition of $\widehat{t}$, we have $\widehat{\delta}_\beta(\widehat{t}) \geqslant \widehat{\delta}_\beta(t_0)$, which, together with (C.15), (C.16) and the triangle inequality, leads to
$$\begin{aligned}
\left|\overline{\delta}_\beta(t_0)\right|_\infty - \left|\overline{\delta}_\beta(\widehat{t})\right|_\infty &\leqslant \widehat{\delta}_\beta(t_0) - \widehat{\delta}_\beta(\widehat{t}) + O_P\left(\sqrt{\frac{\bar{n}\log(N \vee T)}{Th_\ddagger}}\right) \\
&\leqslant O_P\left(\sqrt{\frac{\bar{n}\log(N \vee T)}{Th_\ddagger}}\right).
\end{aligned} \tag{C.17}$$

By virtue of (C.14) and (C.17), we readily have that
$$\underline{c}_0 \left|\frac{\widehat{t} - t_0}{Th_\ddagger}\right| \cdot \delta_\beta(t_0) = O_P\left(\sqrt{\frac{\bar{n}\log(N \vee T)}{Th_\ddagger}} + h_\ddagger\right),$$
which completes the proof of (5.8) as $\delta_\beta(t_0)$ is bounded away from zero in Assumption 6(iii).

We next turn to the proof of Theorem 5.1(ii). Let
$$\mathscr{T}_1^\circ = \{1, 2, \cdots, t_0\} \quad \text{and} \quad \mathscr{T}_2^\circ = \{t_0 + 1, t_0 + 2, \cdots, T\}.$$
It follows from (5.8) that
$$\mathsf{P}(\mathscr{T}_1 \subset \mathscr{T}_1^\circ) \to 1 \quad \text{and} \quad \mathsf{P}(\mathscr{T}_2 \subset \mathscr{T}_2^\circ) \to 1,$$
indicating that the group structure is time invariant *w.p.a.1* over the two separate time periods $\mathscr{T}_1$ and $\mathscr{T}_2$.

Following the proofs of Theorems 3.1, 3.2 and Corollary 3.3, we can prove Theorem 5.1(ii). Details are omitted to save the space. □

## C.5 Proof of Proposition A.1

Since the proof is analogous to the arguments used in the proofs of Theorems 3.1 and 3.2, we next only sketch the proof.

In the first step, we need to prove that

$$
\mathsf{P}\left(\{\widehat{\mathscr{G}}_m^\circ,\ 1 \leqslant m \leqslant M_0\} = \{\mathscr{G}_m^\circ,\ 1 \leqslant m \leqslant M_0\}\right) \to 1. \tag{C.18}
$$

For the index pairs $(i_1, j_1)$ and $(i_2, j_2)$ taken from $\mathscr{N}_N$, recall

$$
\widetilde{D}_{(i_1 j_1)(i_2 j_2)} = \frac{1}{L} \sum_{l=1}^{L} \left| \widehat{\beta}_{i_1 j_1}(\tau_l^*) - \widehat{\beta}_{i_2 j_2}(\tau_l^*) \right|,
$$

and define

$$
D_{(i_1 j_1)(i_2 j_2)} = \frac{1}{L} \sum_{l=1}^{L} \left| \beta_{i_1 j_1}(\tau_l^*) - \beta_{i_2 j_2}(\tau_l^*) \right|.
$$

By (3.9) in Assumption 3, in order to prove (C.18), it is sufficient to show that

$$
\max_{(i_1, j_1),(i_2, j_2) \in \mathscr{N}_N} \left| \widetilde{D}_{(i_1 j_1)(i_2 j_2)} - D_{(i_1 j_1)(i_2 j_2)} \right| = o_\mathsf{P}\left( \zeta_{NT}^\ddagger \right). \tag{C.19}
$$

With Lemma D.3 and (3.9), following the proof of (C.2), we can easily prove (C.19).

In the second step, we need to prove that

$$
\mathsf{P}\left( \widehat{M} = M_0 \mid \mathscr{E}_{\mathscr{G}}^\circ \right) \to 1, \tag{C.20}
$$

where $\mathscr{E}_{\mathscr{G}}^\circ$ denotes the event that $\{\widehat{\mathscr{G}}_m^\circ,\ 1 \leqslant m \leqslant M_0\} = \{\mathscr{G}_m^\circ,\ 1 \leqslant m \leqslant M_0\}$. The proof of (C.20) is similar to the proof of (C.5).

In the final step, combining (C.18) and (C.20), we readily have that

$$
\mathsf{P}\left(\{\widehat{\mathscr{G}}_m^\circ,\ 1 \leqslant m \leqslant \widehat{M}\} = \{\mathscr{G}_m^\circ,\ 1 \leqslant m \leqslant M_0\}\right) \to 1,
$$

completing the proof of Proposition A.1. □

# D Technical lemmas with proofs

We next prove some technical lemmas which have been used to prove the asymptotic theorems in Appendix C. For any $1 \leqslant i \leqslant N$, we define

$$
\boldsymbol{\Gamma}_{iX}(\tau) = \left[ \begin{array}{cc} \frac{1}{Th} \sum_{t=1}^{T} \widetilde{X}_{i,t-1} \widetilde{X}_{i,t-1}^{\intercal} K_{t0}(\tau) & \frac{1}{Th} \sum_{t=1}^{T} \widetilde{X}_{i,t-1} \widetilde{X}_{i,t-1}^{\intercal} K_{t1}(\tau) \\ \frac{1}{Th} \sum_{t=1}^{T} \widetilde{X}_{i,t-1} \widetilde{X}_{i,t-1}^{\intercal} K_{t1}(\tau) & \frac{1}{Th} \sum_{t=1}^{T} \widetilde{X}_{i,t-1} \widetilde{X}_{i,t-1}^{\intercal} K_{t2}(\tau) \end{array} \right]
$$

$$
=: \left[ \begin{array}{cc} \boldsymbol{\Gamma}_{iX}(\tau, 0) & \boldsymbol{\Gamma}_{iX}(\tau, 1) \\ \boldsymbol{\Gamma}_{iX}(\tau, 1) & \boldsymbol{\Gamma}_{iX}(\tau, 2) \end{array} \right],
$$

$$
\boldsymbol{\Gamma}_{i\beta}(\tau) = \left[ \begin{array}{c} \frac{1}{Th} \sum_{t=1}^{T} \widetilde{X}_{i,t-1} \widetilde{X}_{i,t-1}^{\intercal} \left[ \boldsymbol{\beta}_{i\bullet}(\tau_t) - \boldsymbol{\beta}_{i\bullet}(\tau) - \boldsymbol{\beta}_{i\bullet}'(\tau)(\tau_t - \tau) \right] K_{t0}(\tau) \\ \frac{1}{Th} \sum_{t=1}^{T} \widetilde{X}_{i,t-1} \widetilde{X}_{i,t-1}^{\intercal} \left[ \boldsymbol{\beta}_{i\bullet}(\tau_t) - \boldsymbol{\beta}_{i\bullet}(\tau) - \boldsymbol{\beta}_{i\bullet}'(\tau)(\tau_t - \tau) \right] K_{t1}(\tau) \end{array} \right]
$$

$$
=: \left[ \begin{array}{c} \boldsymbol{\Gamma}_{i\beta}(\tau, 0) \\ \boldsymbol{\Gamma}_{i\beta}(\tau, 1) \end{array} \right],
$$

$$
\boldsymbol{\Gamma}_{i\varepsilon}(\tau) = \left[ \begin{array}{c} \frac{1}{Th} \sum_{t=1}^{T} \widetilde{X}_{i,t-1} \varepsilon_{i,t} K_{t0}(\tau) \\ \frac{1}{Th} \sum_{t=1}^{T} \widetilde{X}_{i,t-1} \varepsilon_{i,t} K_{t1}(\tau) \end{array} \right] =: \left[ \begin{array}{c} \boldsymbol{\Gamma}_{i\varepsilon}(\tau, 0) \\ \boldsymbol{\Gamma}_{i\varepsilon}(\tau, 1) \end{array} \right],
$$

where $K_{tq}(\tau) = \left( \frac{\tau_t - \tau}{h} \right)^q K \left( \frac{\tau_t - \tau}{h} \right)$. From the definition of $\widehat{\boldsymbol{\beta}}_{i\bullet}(\cdot)$ in (3.3), we write

$$
\widehat{\boldsymbol{\beta}}_{i\bullet}(\tau) - \boldsymbol{\beta}_{i\bullet}(\tau) = \mathbf{E}_* \cdot \boldsymbol{\Gamma}_{iX}^{-1}(\tau) \left[ \boldsymbol{\Gamma}_{i\beta}(\tau) + \boldsymbol{\Gamma}_{i\varepsilon}(\tau) \right], \tag{D.1}
$$

where $\mathbf{E}_* = \left( \mathbf{I}_{n_i+1}, \mathbf{O}_{(n_i+1) \times (n_i+1)} \right)$. As in the previous proofs, we let C denote a generic positive constant whose value may change from one place to another.

**Lemma D.1.** *Suppose that Assumptions 1 and 2(i)–(iv) and (2.7) in the main text are satisfied. Then we have*

$$
\max_{1 \leqslant i \leqslant N} \sup_{0 \leqslant \tau \leqslant 1} \left| \boldsymbol{\Gamma}_{iX}(\tau) - \boldsymbol{\Gamma}_{iX}^{\dagger}(\tau) \right|_F = O_P \left( \sqrt{\frac{\bar{n}^2 \log(N \vee T)}{Th}} \right), \tag{D.2}
$$

*and*

$$
\boldsymbol{\Gamma}_{iX}^{\dagger}(\tau) = \left[ \begin{array}{cc} \frac{1}{T} \sum_{t=1}^{T} \mathsf{E} \left( \widetilde{X}_{i,t-1} \widetilde{X}_{i,t-1}^{\intercal} \right) K_{t0}(\tau) & \frac{1}{T} \sum_{t=1}^{T} \mathsf{E} \left( \widetilde{X}_{i,t-1} \widetilde{X}_{i,t-1}^{\intercal} \right) K_{t1}(\tau) \\ \frac{1}{T} \sum_{t=1}^{T} \mathsf{E} \left( \widetilde{X}_{i,t-1} \widetilde{X}_{i,t-1}^{\intercal} \right) K_{t1}(\tau) & \frac{1}{T} \sum_{t=1}^{T} \mathsf{E} \left( \widetilde{X}_{i,t-1} \widetilde{X}_{i,t-1}^{\intercal} \right) K_{t2}(\tau) \end{array} \right]
$$

*is positive definite uniformly over $\tau \in [0, 1]$.*

**Proof of Lemma D.1.** We start with the proof of (D.2). We only prove the uniform convergence for $\boldsymbol{\Gamma}_{iX}(\cdot, 0)$ since the proofs for the other block-matrices $\boldsymbol{\Gamma}_{iX}(\cdot, 1)$ and $\boldsymbol{\Gamma}_{iX}(\cdot, 2)$ are exactly the same. Let

$$
\widehat{\boldsymbol{\Delta}}_{iT}(\tau) = \boldsymbol{\Gamma}_{iX}(\tau, 0), \quad \boldsymbol{\Delta}_i^{\dagger}(\tau) = \frac{1}{T} \sum_{t=1}^{T} \mathsf{E} \left[ \widetilde{X}_{i,t-1} \widetilde{X}_{i,t-1}^{\intercal} \right] K_{t0}(\tau) = \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\Delta}_{i,t-1} K_{t0}(\tau), \tag{D.3}
$$

where $\Delta_{i,t}$ is defined as in Assumption 2(ii). Hence, we only show that

$$\max_{1\leqslant i\leqslant N}\sup_{0\leqslant\tau\leqslant1}\left|\widehat{\Delta}_{iT}(\tau)-\Delta_i^\dagger(\tau)\right|_F = O_P\left(\sqrt{\frac{\bar{n}^2\log(N\vee T)}{Th}}\right). \tag{D.4}$$

As

$$\left|\widehat{\Delta}_{iT}(\tau)-\Delta_i^\dagger(\tau)\right|_F \leqslant n_i\left|\widehat{\Delta}_{iT}(\tau)-\Delta_i^\dagger(\tau)\right|_\infty \leqslant \bar{n}\left|\widehat{\Delta}_{iT}(\tau)-\Delta_i^\dagger(\tau)\right|_\infty,$$

it is sufficient to show that

$$\max_{1\leqslant i\leqslant N}\sup_{0\leqslant\tau\leqslant1}\left|\widehat{\Delta}_{iT}(\tau)-\Delta_i^\dagger(\tau)\right|_\infty = O_P\left(\sqrt{\frac{\log(N\vee T)}{Th}}\right). \tag{D.5}$$

Let

$$\widehat{\Delta}_{iT}^\circ(\tau) = \frac{1}{T}\sum_{t=1}^T \widetilde{X}_{i,t-1}^\circ\widetilde{X}_{i,t-1}^{\circ\intercal}K_{t0}(\tau), \quad \Delta_i^\circ(\tau) = \frac{1}{T}\sum_{t=1}^T \mathsf{E}\left[\widetilde{X}_{i,t-1}^\circ\widetilde{X}_{i,t-1}^{\circ\intercal}\right]K_{t0}(\tau),$$

where $\widetilde{X}_{i,t}^\circ$ is defined similarly to $\widetilde{X}_{i,t}$ but with elements in $X_t$ replaced by those in $X_t^\circ$. By Lemma D.2 below, to prove (D.5), we only need to show that

$$\max_{1\leqslant i\leqslant N}\sup_{0\leqslant\tau\leqslant1}\left|\widehat{\Delta}_{iT}^\circ(\tau)-\Delta_i^\circ(\tau)\right|_\infty = O_P\left(\sqrt{\frac{\log(N\vee T)}{Th}}\right). \tag{D.6}$$

Let $\widetilde{V}_{i,t} = \widetilde{X}_{i,t-1}^\circ\widetilde{X}_{i,t-1}^{\circ\intercal}$. We consider covering the closed interval $[0,1]$ by some disjoint sub-intervals $\mathcal{I}_b$ with centres $\tau_b$ and length $\gamma_{NT} = h^2\sqrt{\log(N\vee T)/Th}/\xi_{NT}, 1\leqslant b\leqslant B$, where $\xi_{NT} = T^{(q-2)/[2(q+2)]}[h\log(N\vee T)]^1$ The number of sub-intervals, B, is upper bounded by $\xi_{NT}/(h^{3/2}\sqrt{\log(N\vee T)/T})$. Observe that

$$\max_{1\leqslant i\leqslant N}\sup_{0\leqslant\tau\leqslant1}\left|\widehat{\Delta}_{iT}^\circ(\tau)-\Delta_i^\circ(\tau)\right|_\infty \leqslant \max_{1\leqslant i\leqslant N}\max_{1\leqslant b\leqslant B}\left|\frac{1}{Th}\sum_{t=1}^T\left\{\widetilde{V}_{i,t}-\mathsf{E}[\widetilde{V}_{i,t}]\right\}K_{t0}(\tau_b)\right|_\infty +$$

$$\max_{1\leqslant i\leqslant N}\max_{1\leqslant b\leqslant B}\sup_{\tau\in\mathcal{I}_b}\left|\frac{1}{Th}\sum_{t=1}^T\left\{\widetilde{V}_{i,t}-\mathsf{E}[\widetilde{V}_{i,t}]\right\}[K_{t0}(\tau)-K_{t0}(\tau_b)]\right|_\infty,$$

We first prove that

$$\max_{1\leqslant i\leqslant N}\max_{1\leqslant b\leqslant B}\sup_{\tau\in\mathcal{I}_b}\left|\frac{1}{Th}\sum_{t=1}^T\left\{\widetilde{V}_{i,t}-\mathsf{E}[\widetilde{V}_{i,t}]\right\}[K_{t0}(\tau)-K_{t0}(\tau_b)]\right|_\infty$$

$$= O_P\left(\sqrt{\frac{\log(N\vee T)}{Th}}\right). \tag{D.7}$$

Let

$$\widetilde{V}_{i,t}^\dagger = \widetilde{V}_{i,t}I\left\{|\widetilde{V}_{i,t}|_\infty\leqslant\xi_{NT}\right\} \quad\text{and}\quad \widetilde{V}_{i,t}^\ddagger = \widetilde{V}_{i,t}-\widetilde{V}_{i,t}^\dagger.$$

By Assumption 2(i), we readily have that

$$\max_{1\leqslant b\leqslant B}\sup_{\tau\in\mathcal{J}_b}|K_{t0}(\tau)-K_{t0}(\tau_b)|=O_P\left(\gamma_{NT}h^{-2}\right).$$

Hence, we may show that

$$\max_{1\leqslant i\leqslant N}\max_{1\leqslant b\leqslant B}\sup_{\tau\in\mathcal{J}_b}\left|\frac{1}{Th}\sum_{t=1}^{T}\left\{\widetilde{V}_{i,t}-E[\widetilde{V}_{i,t}]\right\}[K_{t0}(\tau)-K_{t0}(\tau_b)]\right|_{\infty}$$

$$\leqslant\gamma_{NT}h^{-2}\left\{\max_{1\leqslant i\leqslant N}\frac{1}{T}\sum_{t=1}^{T}\left|\widetilde{V}_{i,t}^{\dagger}-E[\widetilde{V}_{i,t}^{\dagger}]\right|_{\infty}+\max_{1\leqslant i\leqslant N}\frac{1}{T}\sum_{t=1}^{T}\left|\widetilde{V}_{i,t}^{\ddagger}-E[\widetilde{V}_{i,t}^{\ddagger}]\right|_{\infty}\right\}$$

$$=O_P\left(\sqrt{\frac{\log(N\vee T)}{Th}}\right)+\gamma_{NT}h^{-2}\max_{1\leqslant i\leqslant N}\frac{1}{T}\sum_{t=1}^{T}\left|\widetilde{V}_{i,t}^{\ddagger}-E[\widetilde{V}_{i,t}^{\ddagger}]\right|_{\infty}. \tag{D.8}$$

It follows from (2.7) in the main text that

$$\max_{1\leqslant i\leqslant N}\max_{1\leqslant t\leqslant T}\left\|\,|\widetilde{V}_{i,t}|_{\infty}\right\|_{q/2}^{q/2}\leqslant\theta_{N,q}^{q}, \tag{D.9}$$

which, together with the Markov inequality, indicates that, for any $\eta>0$,

$$P\left(\max_{1\leqslant i\leqslant N}\frac{1}{T}\sum_{t=1}^{T}\left|\widetilde{V}_{i,t}^{\ddagger}-E[\widetilde{V}_{i,t}^{\ddagger}]\right|_{\infty}>\eta\xi_{NT}\right)\leqslant\sum_{i=1}^{N}\sum_{t=1}^{T}P\left(\left|\widetilde{V}_{i,t}\right|_{\infty}>\xi_{NT}\right)$$

$$=O\left(NT\theta_{N,q}^{q}\xi_{NT}^{-q/2}\right)=o(1), \tag{D.10}$$

where the last equality is due to (3.7) in Assumption 2(iii). With (D.8) and (D.10), we complete the proof of (D.6).

It remains to show that

$$\max_{1\leqslant i\leqslant N}\max_{1\leqslant b\leqslant B}\left|\frac{1}{Th}\sum_{t=1}^{T}\left\{\widetilde{V}_{i,t}-E[\widetilde{V}_{i,t}]\right\}K_{t0}(\tau_b)\right|_{\infty}=O_p\left(\sqrt{\frac{\log(N\vee T)}{Th}}\right). \tag{D.11}$$

Letting $\widetilde{V}_{i,t}(\tau)=\widetilde{V}_{i,t}K_{t0}(\tau)$ for notational simplicity, as $K(\cdot)$ has the support $[-1,1]$,

$$\sum_{t=1}^{T}\widetilde{V}_{i,t}(\tau)=\sum_{t=T_1(\tau)}^{T_2(\tau)}\widetilde{V}_{i,t}(\tau)=\sum_{t=T_1(\tau)}^{T_2(\tau)}\widetilde{X}_{i,t-1}^{\circ}\widetilde{X}_{i,t-1}^{\circ\intercal}K_{t0}(\tau),$$

where $T_1(\tau)=\lfloor T\tau\rfloor-\lfloor Th\rfloor+1$ and $T_2(\tau)=\lfloor T\tau\rfloor+\lfloor Th\rfloor$. We next adopt some standard techniques in the literature on high-dimensional locally stationary processes (e.g., Zhang & Wu 2021) to prove (D.11). Let $M=2\lfloor Th\rfloor, M^{\dagger}=\lfloor\log M/\log 2\rfloor, u_l=2^l$ for $1\leqslant l\leqslant M^{\dagger}-1$ and $u_{M^{\dagger}}=M$. Define $\widetilde{V}_{i,t,u}(\tau)=$

$\mathsf{E}[\widetilde{V}_{i,t}(\tau)|\mathscr{F}_{t-u}^{t-1}]$ with $\mathscr{F}_s^t = (\varepsilon_s, \cdots, \varepsilon_t)$, and

$$S_{i,l,T}(\tau) = \sum_{t=T_1(\tau)}^{T_2(\tau)} \left[ \widetilde{V}_{i,t,u_l}(\tau) - \widetilde{V}_{i,t,u_{l-1}}(\tau) \right], \quad l = 1, \cdots, M^\dagger.$$

Then, for any $1 \leqslant i \leqslant N$ and $1 \leqslant b \leqslant B$, we may decompose

$$\frac{1}{Th} \sum_{t=1}^{T} \left\{ \widetilde{V}_{i,t} - \mathsf{E}\left[ \widetilde{V}_{i,t} \right] \right\} K_{t0}(\tau_b) = \sum_{t=T_1(\tau_b)}^{T_2(\tau_b)} \left\{ \widetilde{V}_{i,t}(\tau_b) - \mathsf{E}\left[ \widetilde{V}_{i,t}(\tau_b) \right] \right\}$$

$$= \frac{1}{Th} \left( \sum_{t=T_1(\tau_b)}^{T_2(\tau_b)} \left[ \widetilde{V}_{i,t}(\tau_b) - \widetilde{V}_{i,t,M}(\tau_b) \right] + \sum_{l=2}^{M^\dagger} S_{i,l,T}(\tau_b) + \right.$$

$$\left. \sum_{t=T_1(\tau_b)}^{T_2(\tau_b)} \left\{ \widetilde{V}_{i,t,2}(\tau_b) - \mathsf{E}\left[ \widetilde{V}_{i,t,2}(\tau_b) \right] \right\} \right). \tag{D.12}$$

We next separately tackle the three terms on the right side of (D.12). Since

$$\widetilde{V}_{i,t}(\tau) - \widetilde{V}_{i,t,M}(\tau) = \sum_{k=M+1}^{\infty} \left[ \widetilde{V}_{i,t,k}(\tau) - \widetilde{V}_{i,t,k-1}(\tau) \right],$$

we have

$$\left\| \left\| \sum_{t=T_1(\tau_b)}^{T_2(\tau_b)} \left[ \widetilde{V}_{i,t}(\tau_b) - \widetilde{V}_{i,t,M}(\tau_b) \right] \right\|_\infty \right\|_{q/2}$$

$$= \left\| \left\| \sum_{t=T_1(\tau_b)}^{T_2(\tau_b)} \sum_{k=M+1}^{\infty} \left[ \widetilde{V}_{i,t,k}(\tau_b) - \widetilde{V}_{i,t,k-1}(\tau_b) \right] \right\|_\infty \right\|_{q/2}$$

$$\leqslant C \sum_{k=M+1}^{\infty} \left\| \left\| \sum_{t=T_1(\tau_b)}^{T_2(\tau_b)} \left[ \widetilde{V}_{i,t,k}(\tau_b) - \widetilde{V}_{i,t,k-1}(\tau_b) \right] \right\|_\infty \right\|_{q/2}.$$

Note that, for any $1 \leqslant i \leqslant N$, $1 \leqslant b \leqslant B$ and $k \geqslant M+1$, $\mathsf{vec}(\widetilde{V}_{i,t,k}(\tau_b) - \widetilde{V}_{i,t,k-1}(\tau_b))$, $T_1(\tau) \leqslant t \leqslant T_2(\tau)$, is a sequence of martingale difference vectors. By Lemma D.3 of Zhang & Wu (2021), we have

$$\sum_{k=M+1}^{\infty} \left\| \left\| \sum_{t=T_1(\tau_b)}^{T_2(\tau_b)} \left( \widetilde{V}_{i,t,k}(\tau_b) - \widetilde{V}_{i,t,k-1}(\tau_b) \right) \right\|_\infty \right\|_{q/2}$$

$$= \sum_{k=M+1}^{\infty} \left\| \left\| \sum_{t=T_1(\tau_b)}^{T_2(\tau_b)} \mathsf{vec}\left( \widetilde{V}_{i,t,k}(\tau_b) - \widetilde{V}_{i,t,k-1}(\tau_b) \right) \right\|_\infty \right\|_{q/2}$$

$$\leqslant C \sum_{k=M+1}^{\infty} \sqrt{\sum_{t=T_1(\tau_b)}^{T_2(\tau_b)} \left\| \left| \mathsf{vec}\left( \widetilde{V}_{i,t,k}(\tau_b) - \widetilde{V}_{i,t,k-1}(\tau_b) \right) \right|_{\infty} \right\|_{q/2}^2}$$

$$= C \sum_{k=M+1}^{\infty} \sqrt{\sum_{t=T_1(\tau_b)}^{T_2(\tau_b)} \left\| \left| \widetilde{V}_{i,t,k}(\tau_b) - \widetilde{V}_{i,t,k-1}(\tau_b) \right|_{\infty} \right\|_{q/2}^2}.$$

With the triangle inequality,

$$\left\| \left| \widetilde{V}_{i,t,k}(\tau_b) - \widetilde{V}_{i,t,k-1}(\tau_b) \right|_{\infty} \right\|_{q/2}$$

$$= \left\| \left| \mathsf{E}\left( \widetilde{V}_{i,t}(\tau_b) - \widetilde{V}_{i,t}^{\{t-k\}}(\tau_b) \big| \mathscr{F}_{t-k}^{t-1} \right) \right|_{\infty} \right\|_{q/2}$$

$$\leqslant \left\| \left| \left( \widetilde{X}_{i,t-1}^{\circ} - \widetilde{X}_{i,t-1}^{\circ\{t-k\}} \right) \widetilde{X}_{i,t-1}^{\circ\intercal} K_{t0}(\tau_b) \right|_{\infty} \right\|_{q/2} +$$

$$\left\| \left| \widetilde{X}_{i,t-1}^{\circ\{t-k\}} \left( \widetilde{X}_{i,t-1}^{\circ} - \widetilde{X}_{i,t-1}^{\circ\{t-k\}} \right)^{\intercal} K_{t0}(\tau_b) \right|_{\infty} \right\|_{q/2},$$

where $\widetilde{V}_{i,t}^{\{l\}}(\tau) = \widetilde{X}_{i,t-1}^{\circ\{l\}} \widetilde{X}_{i,t-1}^{\circ\{l\}\intercal} K_{t0}(\tau)$ and

$$\widetilde{X}_{i,t}^{\circ\{l\}} = \left[ \left( \widetilde{w}_{ij} x_{j,t}^{\circ\{l\}} : j \in \mathscr{N}_i \right)^{\intercal}, x_{i,t}^{\circ\{l\}} \right]^{\intercal}$$

with $x_{i,t}^{\circ\{l\}}$ being the $i$-th element of $X_t^{\circ\{l\}} = X_t^{\circ\{l\}}(\tau_t)$ defined in Appendix B. Following the discussion in Appendix B and noting that $\sum_{j \in \mathscr{N}_i} \widetilde{w}_{ij} = 1$, we have

$$\left\| \left| \widetilde{X}_{i,t-1}^{\circ} - \widetilde{X}_{i,t-1}^{\circ\{t-k\}} \right|_{\infty} \right\|_q \leqslant \delta_{i,k-1,q} + \sum_{j \in \mathscr{N}_i} \widetilde{w}_{ij} \delta_{j,k-1,q},$$

and

$$\left\| \left| \widetilde{X}_{i,t-1}^{\circ} K_{t0}(\tau_b) \right|_{\infty} \right\|_q \leqslant \sum_{k=0}^{\infty} \left\| \left| \left[ \mathsf{E}(\widetilde{X}_{i,t-1}^{\circ} | \mathscr{F}_{t-k-1}) - \mathsf{E}(\widetilde{X}_{i,t-1}^{\circ} | \mathscr{F}_{t-k-2}) \right] K_{t0}(\tau_b) \right|_{\infty} \right\|_q$$

$$\leqslant C \sum_{k=0}^{\infty} \left( \delta_{i,k,q} + \sum_{j \in \mathscr{N}_i} \widetilde{w}_{ij} \delta_{j,k,q} \right)$$

$$\leqslant C \max_{1 \leqslant i \leqslant N} \|x_{i\bullet}\|_{q,\iota}$$

which is bounded, where $\mathscr{F}_t = (\cdots, \varepsilon_{t-1}, \varepsilon_t)$ as in Section 2 of the main text. As discussed in Appendix B, we set $\iota = 1$ in the rest of the proof. By the Jensen and Cauchy-Schwarz inequalities, we have

$$\left\| \left| \left( \widetilde{X}_{i,t-1}^{\circ} - \widetilde{X}_{i,t-1}^{\circ\{t-k\}} \right) \widetilde{X}_{i,t-1}^{\circ\intercal} K_{t0}(\tau_b) \right|_{\infty} \right\|_{q/2}$$

$$\leqslant \left\| \left| \widetilde{X}^{\circ}_{i,t-1} - \widetilde{X}^{\circ\{t-k\}}_{i,t-1} \right|_{\infty} \right\|_q \cdot \left\| \left| \widetilde{X}^{\circ}_{i,t-1} K_{t0}(\tau_b) \right|_{\infty} \right\|_q$$

$$\leqslant C \left( \delta_{i,k-1,q} + \sum_{j \in \mathscr{N}_i} \widetilde{w}_{ij} \delta_{j,k-1,q} \right),$$

and similarly,

$$\left\| \left| \widetilde{X}^{\circ\{t-k\}}_{i,t-1} \left( \widetilde{X}^{\circ}_{i,t-1} - \widetilde{X}^{\circ\{t-k\}}_{i,t-1} \right)^{\mathsf{T}} K_{t0}(\tau_b) \right|_{\infty} \right\|_{q/2} \leqslant C \left( \delta_{i,k-1,q} + \sum_{j \in \mathscr{N}_i} \widetilde{w}_{ij} \delta_{j,k-1,q} \right).$$

Since $M = 2\lfloor Th \rfloor$, we have

$$\sum_{k=M+1}^{\infty} \sqrt{\sum_{t=T_1(\tau_b)}^{T_2(\tau_b)} \left\| \left| \widetilde{V}_{i,t,k}(\tau_b) - \widetilde{V}_{i,t,k-1}(\tau_b) \right|_{\infty} \right\|_{q/2}^2}$$

$$\leqslant C(Th)^{1/2} \sum_{k=M+1}^{\infty} \left( \delta_{i,k-1,q} + \sum_{j \in \mathscr{N}_i} \widetilde{w}_{ij} \delta_{j,k-1,q} \right)$$

$$= C(Th)^{1/2} \left( \Delta_{i,M,q} + \sum_{j \in \mathscr{N}_i} \widetilde{w}_{ij} \Delta_{j,M,q} \right)$$

$$\leqslant C(Th)^{-1/2} \max_{1 \leqslant i \leqslant N} \|x_{i\bullet}\|_{q,1}, \tag{D.13}$$

where the last inequality is due to the fact that

$$\max_{1 \leqslant i \leqslant N} \Delta_{i,M,q} \leqslant M^{-1} \max_{1 \leqslant i \leqslant N} \|x_{i\bullet}\|_{q,1}$$

using the definition of $\|x_{i\bullet}\|_{q,\iota}$ and setting $\iota = 1$. Then, by (D.13) and the Markov inequality, we have for any $z > 0$,

$$\mathsf{P} \left( \left| \sum_{t=T_1(\tau_b)}^{T_2(\tau_b)} \left[ \widetilde{V}_{i,t}(\tau_b) - \widetilde{V}_{i,t,M}(\tau_b) \right] \right|_{\infty} > z \right) \leqslant \frac{C}{z^{q/2}(Th)^{q/4}} \max_{1 \leqslant i \leqslant N} \|x_{i\bullet}\|_{q,1}^{q/2}, \tag{D.14}$$

where $C$ is a positive constant independent of $1 \leqslant i \leqslant N$ and $1 \leqslant b \leqslant B$.

We next turn to the second term on the right side of (D.12). For each $2 \leqslant l \leqslant M^{\dagger}$, define

$$U_{i,k,l}(\tau) = \sum_{t=(k-1)u_l+T_1(\tau)}^{(ku_l+T_1(\tau)) \wedge T_2(\tau)} \left[ \widetilde{V}_{i,t,u_l}(\tau) - \widetilde{V}_{i,t,u_{l-1}}(\tau) \right], \ \ 1 \leqslant k \leqslant \lceil M/u_l \rceil,$$

and

$$S^e_{i,l,T}(\tau) = \sum_{k \text{ is even}} U_{i,k,l}(\tau), \ \ S^o_{i,l,T}(\tau) = \sum_{k \text{ is odd}} U_{i,k,l}(\tau).$$

Let $\lambda_l = (l-1)^{-2}/(\pi^2/3)$ if $2 \leqslant l \leqslant M^{\dagger}/2$ and $\lambda_l = (M^{\dagger}+1-l)^{-2}/(\pi^2/3)$ if $M^{\dagger}/2 < l < M^{\dagger}$. It is easy to

14

verify that $\sum_{l=2}^{M^\dagger} \lambda_l \leqslant 1$. Since $U_{i,k_1,l}(\tau)$ and $U_{i,k_2,l}(\tau)$ are independent for $|k_1 - k_2| > 1$, by Lemma D.4 of Zhang & Wu (2021), we have, for any $z > 0$,

$$P\left(\left|S_{i,l,T}^e(\tau_b)\right|_\infty \geqslant \lambda_l z\right) \leqslant C\left((\lambda_l z)^{-q/2} \sum_{k \text{ is even}} E\left[|U_{i,k,l}(\tau_b)|_\infty^{q/2}\right] + \right.$$
$$\left. \exp\left\{-\frac{(\lambda_l z)^2}{C \sum_{k \text{ is even}} E\left[|U_{i,k,l}(\tau_b)|_\infty^2\right]}\right\}\right),$$

where $2 \leqslant l \leqslant M^\dagger$. Similarly to the proof of (D.13), we may show that

$$\||U_{i,j,l}(\tau)|_\infty\|_{q/2} \leqslant C \cdot u_l^{-1/2} \max_{1 \leqslant i \leqslant N} \|x_{i\bullet}\|_{q,1}$$

and

$$\||U_{i,j,l}(\tau)|_\infty\|_2 \leqslant C \cdot u_l^{-1/2} \max_{1 \leqslant i \leqslant N} \|x_{i\bullet}\|_{4,1}.$$

Similar results also hold for $|S_{i,l,T}^o(\tau)|_\infty$ with details omitted to save space. A combination of the above arguments yields that

$$P\left(\left|\sum_{l=2}^{M^\dagger} S_{i,l,T}(\tau_b)\right|_\infty \geqslant 2z\right) \leqslant \sum_{l=2}^{M^\dagger} P\left(|S_{i,l,T}(\tau_b)|_\infty \geqslant 2\lambda_l z\right)$$
$$\leqslant \sum_{l=2}^{M^\dagger} P\left(|S_{i,l,T}^e(\tau_b)|_\infty \geqslant \lambda_l z\right) + \sum_{l=2}^{M^\dagger} P\left(|S_{i,l,T}^o(\tau_b)|_\infty \geqslant \lambda_l z\right)$$
$$\leqslant C\left\{\frac{Th}{z^{q/2}} \cdot \sum_{l=2}^{M^\dagger} u_l^{-(q/4+1)} \lambda_l^{-q/2} \left(\max_{1 \leqslant i \leqslant N} \|x_{i\bullet}\|_{q,1}^{q/2}\right) + \right.$$
$$\left. \sum_{l=2}^{M^\dagger} \exp\left(-\frac{\lambda_l^2 u_l^2 z^2}{C(Th) \max_{1 \leqslant i \leqslant N} \|x_{i\bullet}\|_{4,1}^2}\right)\right\}.$$

Making use of the fact that $\sum_{l=2}^{M^\dagger} \lambda_l^{-q/2} u_l^{-(q/4+1)}$ is bounded and $\min_{l \geqslant 1} \lambda_l^2 u_l^2 > 0$, we have

$$P\left(\left|\sum_{l=2}^{M^\dagger} S_{i,l,T}(\tau_b)\right|_\infty \geqslant 2z\right) \leqslant C\left(\frac{Th}{z^{q/2}} \left(\max_{1 \leqslant i \leqslant N} \|x_{i\bullet}\|_{q,1}^{q/2}\right) + \right.$$
$$\left. \exp\left\{-\frac{z^2}{C(Th)} \left(\max_{1 \leqslant i \leqslant N} \|x_{i\bullet}\|_{4,1}^2\right)^{-1}\right\}\right), \tag{D.15}$$

where $C$ is a positive constant independent of $1 \leqslant i \leqslant N$ and $1 \leqslant b \leqslant B$.

We next tackle the third term on the right side of (D.12). Note that

$$\widetilde{V}_{i,t,2}(\tau) - \mathsf{E}\left[\widetilde{V}_{i,t,2}(\tau)\right] = \sum_{s=-\infty}^{t} \left\{ \mathsf{E}\left[\widetilde{V}_{i,t,2}(\tau)\big|\mathscr{F}_{s-1}\right] - \mathsf{E}\left[\widetilde{V}_{i,t,2}(\tau)\big|\mathscr{F}_{s-2}\right] \right\}.$$

For any s, by the triangle and Jensen inequalities, we have

$$\left\| \left| \mathsf{E}\left[\widetilde{V}_{i,t,2}(\tau_b)\big|\mathscr{F}_{s-1}\right] - \mathsf{E}\left[\widetilde{V}_{i,t,2}(\tau_b)\big|\mathscr{F}_{s-2}\right]\right|_{\infty} \right\|_{q/2}$$

$$\leqslant \left\| \left|\left(\widetilde{X}_{i,t-1}^{\circ} - \widetilde{X}_{i,t-1}^{\circ\{s-1\}}\right)\widetilde{X}_{i,t-1}^{\circ\mathsf{T}}\mathsf{K}_{t0}(\tau_b)\right|_{\infty}\right\|_{q/2} + \left\|\left|\widetilde{X}_{i,t-1}^{\circ\{s-1\}}\left(\widetilde{X}_{i,t-1}^{\circ} - \widetilde{X}_{i,t-1}^{\circ\{s-1\}}\right)^{\mathsf{T}}\mathsf{K}_{t0}(\tau_b)\right|_{\infty}\right\|_{q/2}$$

$$\leqslant C\left(\delta_{i,t-s,q} + \sum_{j\in\mathscr{N}_i}\widetilde{w}_{ij}\delta_{j,t-s,q}\right).$$

Consequently,

$$\left\|\left|\widetilde{V}_{i,t,2}(\tau_b) - \mathsf{E}\left[\widetilde{V}_{i,t,2}(\tau_b)\right]\right|_{\infty}\right\|_{q/2} \leqslant C\sum_{s=-\infty}^{t}\left(\delta_{i,t-s,q} + \sum_{j\in\mathscr{N}_i}\widetilde{w}_{ij}\delta_{j,t-s,q}\right) \leqslant C\max_{1\leqslant i\leqslant N}\Delta_{i,0,q},$$

and

$$\sum_{t=T_1(\tau_b)}^{T_2(\tau_b)}\left\|\left|\widetilde{V}_{i,t,2}(\tau_b) - \mathsf{E}\left[\widetilde{V}_{i,t,2}(\tau_b)\right]\right|_{\infty}\right\|_{q/2}^{q/2} \leqslant C(Th)\max_{1\leqslant i\leqslant N}\|x_{i\bullet}\|_{q,\alpha}^{q/2}. \tag{D.16}$$

Note that $\widetilde{V}_{i,t_1,2}(\tau)$ and $\widetilde{V}_{i,t_2,2}(\tau)$ are independent when $|t_1 - t_2| > 2$. Following the proof of (D.15) and using (D.16), we have

$$\mathsf{P}\left(\left|\sum_{t=T_1(\tau_b)}^{T_2(\tau_b)}\left\{\widetilde{V}_{i,t,2}(\tau_b) - \mathsf{E}\left[\widetilde{V}_{i,t,2}(\tau_b)\right]\right\}\right|_{\infty} \geqslant z\right)$$

$$\leqslant C\left(\frac{Th}{z^{q/2}}\left(\max_{1\leqslant i\leqslant N}\|x_{i\bullet}\|_{q,1}^{q/2}\right) + \exp\left\{-\frac{z^2}{C(Th)}\left(\max_{1\leqslant i\leqslant N}\|x_{i\bullet}\|_{4,1}^2\right)^{-1}\right\}\right), \tag{D.17}$$

where C is a positive constant independent of $1 \leqslant i \leqslant N$ and $1 \leqslant b \leqslant B$.

By virtue of (D.12), (D.14), (D.15) and (D.17), we readily have that

$$\mathsf{P}\left(\left|\sum_{t=1}^{T}\left\{\widetilde{V}_{i,t} - \mathsf{E}[\widetilde{V}_{i,t}]\right\}\mathsf{K}_{t0}(\tau_b)\right|_{\infty} > z\right)$$

$$\leqslant C\left(\frac{Th}{z^{q/2}}\left(\max_{1\leqslant i\leqslant N}\|x_{i\bullet}\|_{q,1}^{q/2}\right) + \exp\left\{-\frac{z^2}{C(Th)}\left(\max_{1\leqslant i\leqslant N}\|x_{i\bullet}\|_{4,1}^2\right)^{-1}\right\}\right). \tag{D.18}$$

As discussed in Appendix B, $\|x_{i\bullet}\|_{q,1} + \|x_{i\bullet}\|_{4,1}$ is bounded uniformly over i. Setting $z = \eta\sqrt{Th\log(N\vee T)}$

in (D.18), we may show that, for any $\eta > 0$,

$$
\begin{aligned}
&\mathsf{P}\left( \max_{1 \leqslant i \leqslant N} \max_{1 \leqslant b \leqslant B} \left| \frac{1}{Th} \sum_{t=1}^{T} \left\{ \widetilde{V}_{i,t} - \mathsf{E}[\widetilde{V}_{i,t}] \right\} K_{t0}(\tau_b) \right|_{\infty} > \eta \sqrt{\frac{\log(N \vee T)}{Th}} \right) \\
&\leqslant \sum_{i=1}^{N} \sum_{b=1}^{B} \mathsf{P}\left( \left| \sum_{t=1}^{T} \left\{ \widetilde{V}_{i,t} - \mathsf{E}[\widetilde{V}_{i,t}] \right\} K_{t0}(\tau_b) \right|_{\infty} > \eta \sqrt{Th \log(N \vee T)} \right) \\
&= O\left( NB \exp\left\{ -C\eta^2 \log(N \vee T) \right\} \right) + O\left( \frac{NB}{\eta^{q/2}(Th)^{q/4-1}(\log(N \vee T))^{q/4}} \right).
\end{aligned}
\tag{D.19}
$$

Noting that B diverge at a certain polynomial rate of T and letting $\eta$ be sufficiently large, the first order on the right side of (D.19) converges to zero. By (3.7) in Assumption 2(iii), we may show that the second order also converges to zero. The proof of (D.11) is completed.

By Assumption 2(i)(ii), it is easy to show that $\Gamma_{iX}^{\dagger}(\tau)$ is positive definite uniformly over $0 \leqslant \tau \leqslant 1$. The proof of Lemma D.1 is completed. $\qquad \square$

**Lemma D.2.** *Suppose that Assumptions 1 and 2(i)(iii) and (2.7) in the main text are satisfied. Let*

$$
\widehat{\boldsymbol{\Delta}}_{iT}^{\circ}(\tau) = \frac{1}{T} \sum_{t=1}^{T} \widetilde{X}_{i,t-1}^{\circ} \widetilde{X}_{i,t-1}^{\circ\mathsf{T}} K_{t0}(\tau), \quad \boldsymbol{\Delta}_{i}^{\circ}(\tau) = \frac{1}{T} \sum_{t=1}^{T} \mathsf{E}\left[ \widetilde{X}_{i,t-1}^{\circ} \widetilde{X}_{i,t-1}^{\circ\mathsf{T}} \right] K_{t0}(\tau),
$$

*where $\widetilde{X}_{i,t}^{\circ}$ is defined similarly to $\widetilde{X}_{i,t}$ in Section 3.1 but with elements in $X_t$ replaced by those in $X_t^{\circ}$. Then we have*

$$
\max_{1 \leqslant i \leqslant N} \sup_{0 \leqslant \tau \leqslant 1} \left| \widehat{\boldsymbol{\Delta}}_{iT}(\tau) - \widehat{\boldsymbol{\Delta}}_{iT}^{\circ}(\tau) \right|_{\infty} = o_{\mathsf{P}}\left( \sqrt{\frac{\log(N \vee T)}{Th}} \right)
\tag{D.20}
$$

*and*

$$
\max_{1 \leqslant i \leqslant N} \sup_{0 \leqslant \tau \leqslant 1} \left| \boldsymbol{\Delta}_{i}^{\dagger}(\tau) - \boldsymbol{\Delta}_{i}^{\circ}(\tau) \right|_{\infty} = o\left( \sqrt{\frac{\log(N \vee T)}{Th}} \right),
\tag{D.21}
$$

*where $\widehat{\boldsymbol{\Delta}}_{iT}(\tau)$ and $\boldsymbol{\Delta}_{i}^{\dagger}(\tau)$ are defined in (D.3).*

**Proof of Lemma D.2.** With (2.7), by slightly modifying the proof of (2.8) in Appendix B, we may show that

$$
\max_{0 \leqslant t \leqslant T-1} \||X_t - X_t^{\circ}|_{\infty}\|_q = O\left( \theta_{N,q}/T \right).
\tag{D.22}
$$

By (2.7), (D.22), the classic $c_r$- and Cauchy-Schwarz inequalities, we have

$$
\begin{aligned}
&\max_{1 \leqslant i \leqslant N} \max_{1 \leqslant t \leqslant T} \left\| \left| \widetilde{X}_{i,t-1} \widetilde{X}_{i,t-1}^{\mathsf{T}} - \widetilde{X}_{i,t-1}^{\circ} \widetilde{X}_{i,t-1}^{\circ\mathsf{T}} \right|_{\infty} \right\|_{q/2}^{q/2} \\
&\leqslant C \max_{1 \leqslant i \leqslant N} \max_{1 \leqslant t \leqslant T} \left\| \left| \widetilde{X}_{i,t-1} - \widetilde{X}_{i,t-1}^{\circ} \right|_{\infty} \right\|_{q}^{q/2} \left[ \left\| \left| \widetilde{X}_{i,t-1} \right|_{\infty} \right\|_{q}^{q/2} + \left\| \left| \widetilde{X}_{i,t-1}^{\circ} \right|_{\infty} \right\|_{q}^{q/2} \right] \\
&\leqslant C \max_{0 \leqslant t \leqslant T-1} \||X_t - X_t^{\circ}|_{\infty}\|_q^{q/2} \left[ \||X_t|_{\infty}\|_q^{q/2} + \||X_t^{\circ}|_{\infty}\|_q^{q/2} \right]
\end{aligned}
$$

17

$$= O\left(\theta_{N,q}^q T^{-q/2}\right). \tag{D.23}$$

With (D.23), the Bonferroni and Markov inequalities and (3.7) in Assumption 2(iii), we may show that, for any $\eta > 0$,

$$P\left(\max_{1\leqslant i\leqslant N}\max_{1\leqslant t\leqslant T}\left|\widetilde{X}_{i,t-1}\widetilde{X}_{i,t-1}^{\intercal} - \widetilde{X}_{i,t-1}^{\circ}\widetilde{X}_{i,t-1}^{\circ\intercal}\right|_{\infty} > \eta\sqrt{\frac{\log(N\vee T)}{Th}}\right)$$

$$\leqslant \sum_{i=1}^{N}\sum_{t=1}^{T} P\left(\left|\widetilde{X}_{i,t-1}\widetilde{X}_{i,t-1}^{\intercal} - \widetilde{X}_{i,t-1}^{\circ}\widetilde{X}_{i,t-1}^{\circ\intercal}\right|_{\infty} > \eta\sqrt{\frac{\log(N\vee T)}{Th}}\right)$$

$$\leqslant \left(\eta^2\frac{\log(N\vee T)}{Th}\right)^{-q/4}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\left|\widetilde{X}_{i,t-1}\widetilde{X}_{i,t-1}^{\intercal} - \widetilde{X}_{i,t-1}^{\circ}\widetilde{X}_{i,t-1}^{\circ\intercal}\right|_{\infty}\right\|_{q/2}^{q/2}$$

$$\leqslant C\eta^{-q/2}\cdot\frac{N\theta_{N,q}^q h^{q/4}}{T^{q/4-1}[\log(N\vee T)]^{q/4}} = o(1),$$

indicating that

$$\max_{1\leqslant i\leqslant N}\max_{1\leqslant t\leqslant T}\left|\widetilde{X}_{i,t-1}\widetilde{X}_{i,t-1}^{\intercal} - \widetilde{X}_{i,t-1}^{\circ}\widetilde{X}_{i,t-1}^{\circ\intercal}\right|_{\infty} = o_P\left(\sqrt{\frac{\log(N\vee T)}{Th}}\right). \tag{D.24}$$

By (D.24) and Assumption 2(i), we may show that

$$\max_{1\leqslant i\leqslant N}\sup_{0\leqslant\tau\leqslant 1}\left|\widehat{\Delta}_{iT}(\tau) - \widehat{\Delta}_{iT}^{\circ}(\tau)\right|_{\infty}$$

$$\leqslant \max_{1\leqslant i\leqslant N}\sup_{0\leqslant\tau\leqslant 1}\frac{1}{T}\sum_{t=1}^{T}\left|\widetilde{X}_{i,t-1}\widetilde{X}_{i,t-1}^{\intercal} - \widetilde{X}_{i,t-1}^{\circ}\widetilde{X}_{i,t-1}^{\circ\intercal}\right|_{\infty}K_{t0}(\tau)$$

$$\leqslant \max_{1\leqslant i\leqslant N}\max_{1\leqslant s\leqslant T}\left|\widetilde{X}_{i,s-1}\widetilde{X}_{i,s-1}^{\intercal} - \widetilde{X}_{i,s-1}^{\circ}\widetilde{X}_{i,s-1}^{\circ\intercal}\right|_{\infty}\sup_{0\leqslant\tau\leqslant 1}\frac{1}{T}\sum_{t=1}^{T}K_{t0}(\tau)$$

$$= o_P\left(\sqrt{\frac{\log(N\vee T)}{Th}}\right),$$

completing the proof of (D.20). With (D.24) and Assumption 2(i), we can also prove (D.21). $\qquad\square$

In the following lemma, we derive the uniform consistency of the local linear estimation of the heterogenous time-varying coefficient functions defined in (3.3).

**Lemma D.3.** *Suppose that Assumptions 1 and 2(i)–(iv) and (2.7) in the main text are satisfied. Letting $\widehat{\beta}_{i\bullet}(\cdot)$ be defined in (3.3), we have*

$$\max_{1\leqslant i\leqslant N}\max_{1\leqslant l\leqslant L}\left|\widehat{\beta}_{i\bullet}(\tau_l^*) - \beta_{i\bullet}(\tau_l^*)\right|_2 = O_P\left(\sqrt{\bar{n}}\left(\sqrt{\frac{\log(N\vee T)}{Th}} + h^2\right)\right), \tag{D.25}$$

*where the number of grid points $L$ satisfies $L = O(T)$.*

**Proof of Lemma D.3.** By Lemma D.1 and Assumptions 1(i) and 2(ii), we may show that

$$\max_{1\leqslant i\leqslant N}\max_{1\leqslant l\leqslant L}\lambda_{\max}\left(\Gamma_{iX}^{-1}(\tau_l^*)\right)=O_P(1) \tag{D.26}$$

and

$$\max_{1\leqslant i\leqslant N}\max_{1\leqslant l\leqslant L}\left|\Gamma_{iX}^{-1}(\tau_l^*)\Gamma_{i\beta}(\tau_l^*)\right|_2=O_P\left(\sqrt{\bar{n}}h^2\right). \tag{D.27}$$

In view of (D.26) and (D.27), in order to prove (D.25), we only need to show that

$$\max_{1\leqslant i\leqslant N}\max_{1\leqslant l\leqslant L}|\Gamma_{i\varepsilon}(\tau_l^*)|_2=O_P\left(\sqrt{\frac{\bar{n}\log(N\vee T)}{Th}}\right).$$

As in the proof of Lemma D.1, to save the space, we only provide the proof of

$$\max_{1\leqslant i\leqslant N}\max_{1\leqslant l\leqslant L}\left|\frac{1}{Th}\sum_{t=1}^{T}\widetilde{X}_{i,t-1}\varepsilon_{i,t}K_{t0}(\tau_l^*)\right|_2=O_P\left(\sqrt{\frac{\bar{n}\log(N\vee T)}{Th}}\right). \tag{D.28}$$

We next use the truncation technique and the concentration inequality for martingale to prove (D.28). Let $\xi_{NT}=T^{(q-2)/[2(q+2)]}[h\log(N\vee T)]^{1/2}$ be defined as in the proof of Lemma D.1, $\widetilde{W}_{i,t}=\widetilde{X}_{i,t-1}\varepsilon_{i,t}$,

$$\widetilde{W}_{i,t}^{\dagger}=\widetilde{X}_{i,t-1}\varepsilon_{i,t}I\left\{|\widetilde{X}_{i,t-1}|_{\infty}\leqslant\xi_{NT}^{1/2},\ |\varepsilon_{i,t}|\leqslant\xi_{NT}^{1/2}\right\}\ \text{ and }\ \widetilde{W}_{i,t}^{\ddagger}=\widetilde{W}_{i,t}-\widetilde{W}_{i,t}^{\dagger}.$$

Note that

$$\max_{1\leqslant i\leqslant N}\max_{1\leqslant l\leqslant L}\left|\frac{1}{Th}\sum_{t=1}^{T}\widetilde{X}_{i,t-1}\varepsilon_{i,t}K_{t0}(\tau_l^*)\right|_2$$

$$\leqslant\max_{1\leqslant i\leqslant N}\max_{1\leqslant l\leqslant L}\left|\frac{1}{Th}\sum_{t=1}^{T}\left\{\widetilde{W}_{i,t}^{\dagger}-\mathsf{E}\left[\widetilde{W}_{i,t}^{\dagger}\right]\right\}K_{t0}(\tau_l^*)\right|_2+$$

$$\max_{1\leqslant i\leqslant N}\max_{1\leqslant l\leqslant L}\left|\frac{1}{Th}\sum_{t=1}^{T}\left\{\widetilde{W}_{i,t}^{\ddagger}-\mathsf{E}\left[\widetilde{W}_{i,t}^{\ddagger}\right]\right\}K_{t0}(\tau_l^*)\right|_2.$$

For any $\eta>0$, by the Markov inequality, Assumptions 1(ii) and 2(iii), and noting that

$$\max_{1\leqslant i\leqslant N}\max_{0\leqslant t\leqslant T-1}\left\||\widetilde{X}_{i,t}|_{\infty}\right\|_q^q\leqslant\theta_{N,q}^q,$$

we have

$$\mathsf{P}\left(\max_{1\leqslant i\leqslant N}\max_{1\leqslant l\leqslant L}\left|\frac{1}{Th}\sum_{t=1}^{T}\left\{\widetilde{W}_{i,t}^{\ddagger}-\mathsf{E}\left[\widetilde{W}_{i,t}^{\ddagger}\right]\right\}K_{t0}(\tau_l^*)\right|_2>\eta\sqrt{\frac{\bar{n}\log(N\vee T)}{Th}}\right) \tag{D.29}$$

$$\leqslant\mathsf{P}\left(\max_{1\leqslant i\leqslant N}\max_{0\leqslant t\leqslant T-1}|\widetilde{X}_{i,t}|_{\infty}>\xi_{NT}^{1/2}\right)+\mathsf{P}\left(\max_{1\leqslant i\leqslant N}\max_{1\leqslant t\leqslant T}|\varepsilon_{i,t}|>\xi_{NT}^{1/2}\right)$$

$$\leqslant \sum_{i=1}^{N} \sum_{t=0}^{T-1} \mathsf{P}\left(|\widetilde{X}_{i,t}|_{\infty} > \xi_{NT}^{1/2}\right) + \sum_{i=1}^{N} \sum_{t=1}^{T} \mathsf{P}\left(|\varepsilon_{i,t}| > \xi_{NT}^{1/2}\right)$$

$$= O\left(NT\theta_{N,q}^{q}\xi_{NT}^{-q/2} + NT\xi_{NT}^{-q/2}\right) = o(1).$$

By the concentration inequality for martingales (e.g., Freedman 1975), we have

$$\mathsf{P}\left(\max_{1\leqslant i\leqslant N}\max_{1\leqslant l\leqslant L}\left|\frac{1}{Th}\sum_{t=1}^{T}\left\{\widetilde{W}_{i,t}^{\dagger} - \mathsf{E}\left[\widetilde{W}_{i,t}^{\dagger}\right]\right\}K_{t0}(\tau_{l}^{*})\right|_{2} > \eta\sqrt{\frac{\bar{n}\log(N\vee T)}{Th}}\right)$$

$$\leqslant \sum_{i=1}^{N}\sum_{l=1}^{L}\mathsf{P}\left(\left|\frac{1}{Th}\sum_{t=1}^{T}\left\{\widetilde{W}_{i,t}^{\dagger} - \mathsf{E}\left[\widetilde{W}_{i,t}^{\dagger}\right]\right\}K_{t0}(\tau_{l}^{*})\right|_{2} > \eta\sqrt{\frac{\bar{n}\log(N\vee T)}{Th}}\right)$$

$$= O\left(NL\exp\left\{-\frac{\eta^{2}(Th)\log(N\vee T)}{CTh}\right\}\right) = o(1), \tag{D.30}$$

by letting $\eta$ be sufficiently large. With (D.29) and (D.30), we complete the proof of (D.28). $\qquad\square$

**Lemma D.4.** *Suppose that the conditions of Lemma D.3 are satisfied and*

$$\mathsf{P}\left(\widehat{M} = M_{0} \quad \text{and} \quad \widehat{\mathscr{G}}_{m}^{\circ} = \mathscr{G}_{m}^{\circ}, \, 1 \leqslant m \leqslant M_{0}\right) \to 1. \tag{D.31}$$

*Letting $\widehat{\beta}_{i\bullet}^{\circ}(\cdot)$ be defined in (3.4) of the main text, we have*

$$\max_{1\leqslant i\leqslant N}\max_{1\leqslant l\leqslant L}\left|\widehat{\beta}_{i\bullet}^{\circ}(\tau_{l}^{*}) - \beta_{i\bullet}^{\circ}(\tau_{l}^{*})\right|_{2} = O_{\mathsf{P}}\left(\sqrt{\bar{n}}\left(\sqrt{\frac{\log(N\vee T)}{Th}} + h^{2}\right)\right). \tag{D.32}$$

**Proof of Lemma D.4.** Define

$$\mathscr{E}_{g} = \left\{\widehat{g}_{ij} = g_{ij}, (i,j) \in \bar{\mathscr{N}}_{N}\right\} \cap \left\{\widehat{M} = M_{0}\right\},$$

where $\widehat{g}_{ij}$ and $\bar{\mathscr{N}}_{N}$ are defined in Appendix A, and $g_{ij}$ is defined in Section 3.1. By (D.31), we readily have that $\mathsf{P}(\mathscr{E}_{g}) \to 1$. Hence, it is sufficient to prove (D.32) conditional on the event $\mathscr{E}_{g}$. Let $\widehat{\omega}_{ij,m}$ and $\omega_{ij,m}$ be defined as in Section 3.1. Note that

$$\widehat{\beta}_{im}^{\circ}(\tau) = \sum_{j\in\mathscr{N}_{i}}\widehat{\beta}_{ij}(\tau)\widehat{\omega}_{ij,m} = \sum_{j\in\mathscr{N}_{i}}\widehat{\beta}_{ij}(\tau)\omega_{ij,m} =: \overline{\beta}_{im}^{\circ}(\tau) \tag{D.33}$$

and

$$\overline{\beta}_{i\bullet}^{\circ}(\tau) - \beta_{i\bullet}^{\circ}(\tau) = \left\{\sum_{j\in\mathscr{N}_{i}}\left[\widehat{\beta}_{ij}(\tau) - \beta_{ij}(\tau)\right]\omega_{ij,1}, \cdots, \sum_{j\in\mathscr{N}_{i}}\left[\widehat{\beta}_{ij}(\tau) - \beta_{ij}(\tau)\right]\omega_{ij,M_{0}}\right\}^{\top} \tag{D.34}$$

conditional on $\mathscr{E}_{g}$, where $\overline{\beta}_{i\bullet}^{\circ}(\tau) = \left[\overline{\beta}_{i1}^{\circ}(\tau), \cdots, \overline{\beta}_{iM_{0}}^{\circ}(\tau)\right]^{\top}$. Combining (D.33), (D.34) and Lemma D.3, we can complete the proof of (D.32) conditional on $\mathscr{E}_{g}$. $\qquad\square$

The following lemma is crucial to prove the consistency property of the ratio criterion.

**Lemma D.5.** *Suppose that the conditions of Theorem 3.2 are satisfied. Define $\mathcal{E}_{\mathcal{G}}$ as the event that $\{\widehat{\mathcal{G}}_1, \cdots, \widehat{\mathcal{G}}_{K_0}\} = \{\mathcal{G}_1, \cdots, \mathcal{G}_{K_0}\}$. Then, conditional on $\mathcal{E}_{\mathcal{G}}$, we have (i) for $K = K_0, K_0 + 1, \cdots, \overline{K}$, $\widehat{R}(K) = o_P(\rho_{NT})$, where $\rho_{NT}$ is defined in Section 3.2; and (ii) for $K = 1, \cdots, K_0 - 1$, $\widehat{R}(K) \geqslant \underline{c}\zeta_{NT}^{\dagger}$ w.p.a.1, where $\zeta_{NT}^{\dagger}$ is defined in Assumption 3 and $\underline{c}$ is a positive constant strictly larger than zero.*

**Proof of Lemma D.5.** For the case $K_0 \leqslant K \leqslant \overline{K}$, conditional on the event $\mathcal{E}_{\mathcal{G}}$, the grouped time-varying network VAR model is either correctly- or over-fitted. Some of $\mathcal{G}_k$, $k = 1, \cdots, K_0$, are further split into smaller groups when $K > K_0$. Without loss of generality, we only consider the case of $K = K_0 + 1$ (conditional on $\mathcal{E}_{\mathcal{G}}$) and assume that $\mathcal{G}_{K_0}$ is split into $\mathcal{G}_{K_0}^{\dagger}$ and $\mathcal{G}_{K_0}^{\ddagger}$. It follows from the latent group assumption in Section 2.1 that there exist $\alpha_{k\bullet}^{\circ}(\cdot)$, $k = 1, \cdots, K_0$, such that $\beta_{i\bullet}^{\circ}(\cdot) = \alpha_{k\bullet}^{\circ}(\cdot)$ when $g_i = k$. By Assumption 4(i) as well as Lemmas D.3 and D.4, we may show that

$$\widehat{R}(K_0 + 1) = \frac{1}{L(K_0 + 1)} \sum_{k=1}^{K_0-1} \frac{1}{\mathrm{card}(\mathcal{G}_k)} \sum_{i \in \mathcal{G}_k} \sum_{l=1}^{L} [|\beta_i(\tau_l^*) - \alpha_k(\tau_l^*)| + |\beta_{i\bullet}^{\circ}(\tau_l^*) - \alpha_{k\bullet}^{\circ}(\tau_l^*)|_2] +$$

$$\frac{1}{L(K_0 + 1)\mathrm{card}(\mathcal{G}_{K_0}^{\dagger})} \sum_{i \in \mathcal{G}_{K_0}^{\dagger}} \sum_{l=1}^{L} \left[|\beta_i(\tau_l^*) - \alpha_{K_0}(\tau_l^*)| + \left|\beta_{i\bullet}^{\circ}(\tau_l^*) - \alpha_{K_0\bullet}^{\circ}(\tau_l^*)\right|_2\right] +$$

$$\frac{1}{L(K_0 + 1)\mathrm{card}(\mathcal{G}_{K_0}^{\ddagger})} \sum_{i \in \mathcal{G}_{K_0}^{\ddagger}} \sum_{l=1}^{L} \left[|\beta_i(\tau_l^*) - \alpha_{K_0}(\tau_l^*)| + \left|\beta_{i\bullet}^{\circ}(\tau_l^*) - \alpha_{K_0\bullet}^{\circ}(\tau_l^*)\right|_2\right] +$$

$$O_P\left(\sqrt{\overline{n}}\left(\sqrt{\frac{\log(N \vee T)}{Th}} + h^2\right)\right)$$

$$= O_P\left(\sqrt{\overline{n}}\left(\sqrt{\frac{\log(N \vee T)}{Th}} + h^2\right)\right) = o_P(\rho_{NT})$$

conditional on $\mathcal{E}_{\mathcal{G}}$. The proof is similar for $\widehat{R}(K)$, $K = K_0, K_0 + 2, \cdots, \overline{K}$.

For the case $1 \leqslant K \leqslant K_0 - 1$, conditional on the event $\mathcal{E}_{\mathcal{G}}$, the model is under-fitted and at least two groups are falsely merged. Without loss of generality, we only consider the case of $K = K_0 - 1$ (conditional on $\mathcal{E}_{\mathcal{G}}$) and assume that $\mathcal{G}_{K_0-1}$ and $\mathcal{G}_{K_0}$ are merged. Define

$$\alpha_*(\tau) = \frac{1}{\mathrm{card}(\mathcal{G}_{K_0-1} \cup \mathcal{G}_{K_0})} [|\mathcal{G}_{K_0-1}| \alpha_{K_0-1}(\tau) + |\mathcal{G}_{K_0}| \alpha_{K_0}(\tau)]$$

and

$$\alpha_*^{\circ}(\tau) = \frac{1}{\mathrm{card}(\mathcal{G}_{K_0-1} \cup \mathcal{G}_{K_0})} \left[|\mathcal{G}_{K_0-1}| \alpha_{K_0-1\bullet}^{\circ}(\tau) + |\mathcal{G}_{K_0}| \alpha_{K_0\bullet}^{\circ}(\tau)\right].$$

By Assumptions 3 and 4, and Lemmas D.3 and D.4, we can prove that

$$\widehat{R}(K_0 - 1) = \frac{1}{L(K_0 - 1)} \sum_{k=1}^{K_0-2} \frac{1}{\mathrm{card}(\mathcal{G}_k)} \sum_{i \in \mathcal{G}_k} \sum_{l=1}^{L} [|\beta_i(\tau_l^*) - \alpha_k(\tau_l^*)| + |\beta_{i\bullet}^{\circ}(\tau_l^*) - \alpha_{k\bullet}^{\circ}(\tau_l^*)|_2] +$$

$$\frac{1}{L(K_0-1)\mathrm{card}(\mathscr{G}_{K_0-1}\cup\mathscr{G}_{K_0})}\sum_{i\in\mathscr{G}_{K_0-1}\cup\mathscr{G}_{K_0}}\sum_{l=1}^{L}[|\beta_i(\tau_l^*)-\alpha_*(\tau_l^*)|+|\beta_{i\bullet}^\circ(\tau_l^*)-\alpha_*^\circ(\tau_l^*)|_2]$$

$$+O_P\left(\sqrt{\bar{n}}\left(\sqrt{\frac{\log N\vee T}{Th}}+h^2\right)\right)$$

$$=2\frac{\mathrm{card}(\mathscr{G}_{K_0-1})\mathrm{card}(\mathscr{G}_{K_0})}{L(K_0-1)\mathrm{card}(\mathscr{G}_{K_0-1}\cup\mathscr{G}_{K_0})}\sum_{l=1}^{L}\left[|\alpha_{K_0}(\tau_l^*)-\alpha_{K_0-1}(\tau_l^*)|+\left|\alpha_{K_0\bullet}^\circ(\tau_l^*)-\alpha_{K_0-1\bullet}^\circ(\tau_l^*)\right|_2\right]$$

$$+O_P\left(\sqrt{\bar{n}}\left(\sqrt{\frac{\log N\vee T}{Th}}+h^2\right)\right)$$

$$\geqslant\underline{c}\xi_{NT}^\dagger\quad w.p.a.1,$$

conditional on $\mathscr{E}_\mathscr{G}$, where $\underline{c}$ is a positive constant. The same result also holds for $\widehat{R}(K)$, $K=1,\cdots,K_0-2$. The proof of Lemma D.5 is completed. $\qquad\square$

The following lemma is useful to prove the limit distribution theory of the post-grouping local linear estimation.

**Lemma D.6.** *Suppose that Assumptions 1, 2(i) and 5(i) in the main text are satisfied. Let*

$$\check{\Delta}_{ij,\kappa}(\tau)=\frac{1}{Th_\dagger}\sum_{t=1}^{T}X_{i,t-1}^\diamond X_{i,t-1}^{\diamond\intercal}K_{t\kappa}^\dagger(\tau),\quad K_{t\kappa}^\dagger(\tau)=\left(\frac{\tau_t-\tau}{h_\dagger}\right)^\kappa K\left(\frac{\tau_t-\tau}{h_\dagger}\right),$$

*where $X_{i,t}^\diamond$ is defined in Section 4.1. Then we have*

$$\max_{1\leqslant i,j\leqslant N}\left|\check{\Delta}_{ij,\kappa}(\tau)-E\left[\check{\Delta}_{ij,\kappa}(\tau)\right]\right|_F=O_P\left(\sqrt{\frac{\log(N\vee T)}{Th}}\right) \tag{D.35}$$

*for $\kappa=0,1,2$ and $\tau\in[0,1]$.*

**Proof of Lemma D.6.** The proof is similar to the proof of Lemma D.1. Details are omitted here to save space. $\qquad\square$

# E   Refined estimation of the break point

## E.1   Refined estimation methodology

In Section 5 of the main text, we construct an estimation of the break location $t_0$, which is shown to be consistent with scaling, i.e.,

$$\left|\frac{\widehat{t}-t_0}{T}\right|=o_P(1),$$

see Theorem 5.1(i). Although this consistency property is sufficient to consistently estimate the group membership and number in the subsequent stage, there is a natural question on whether the precision of the

break location estimation can be improved. We next aim to tackle this issue by introducing a refined break point estimation which achieves improvement of the estimation accuracy by making use of the estimated group structure.

As in Section 4, we define

$$
\check{X}^1_{i,t} = \left[ \sum_{j \in \widehat{\mathscr{G}}^1_1} \widetilde{w}_{ij} x_{j,t}, \cdots, \sum_{j \in \widehat{\mathscr{G}}^1_{\check{K}_1}} \widetilde{w}_{ij} x_{j,t}, x_{i,t} \right]^\mathsf{T},
$$

$$
\check{X}^2_{i,t} = \left[ \sum_{j \in \widehat{\mathscr{G}}^2_1} \widetilde{w}_{ij} x_{j,t}, \cdots, \sum_{j \in \widehat{\mathscr{G}}^2_{\check{K}_2}} \widetilde{w}_{ij} x_{j,t}, x_{i,t} \right]^\mathsf{T}.
$$

For any t, we use the cross-node regression to estimate the group-specific coefficient functions:

$$
\check{\alpha}^1_{k\bullet}(\tau_t) = \left[ \sum_{i \in \widehat{\mathscr{G}}^1_k} \check{X}^1_{i,t-1} \left( \check{X}^1_{i,t-1} \right)^\mathsf{T} \right]^{-1} \sum_{i \in \widehat{\mathscr{G}}^1_k} \check{X}^1_{i,t-1} x_{i,t}, \quad k = 1, \cdots, \widehat{K}_1,
$$

$$
\check{\alpha}^2_{k\bullet}(\tau_t) = \left[ \sum_{i \in \widehat{\mathscr{G}}^2_k} \check{X}^2_{i,t-1} \left( \check{X}^2_{i,t-1} \right)^\mathsf{T} \right]^{-1} \sum_{i \in \widehat{\mathscr{G}}^2_k} \check{X}^2_{i,t-1} x_{i,t}. \quad k = 1, \cdots, \widehat{K}_2.
$$

As in Section 4, without loss of generality, we assume

$$
\widehat{\mathscr{G}}^1_k = \mathscr{G}^1_k, \quad 1 \leqslant k \leqslant K_1, \quad \text{and} \quad \widehat{K}_1 = K_1,
$$
$$
\widehat{\mathscr{G}}^2_k = \mathscr{G}^2_k, \quad 1 \leqslant k \leqslant K_2, \quad \text{and} \quad \widehat{K}_2 = K_2.
$$

We expect $\check{\alpha}^1_{k\bullet}(\tau_t)$ to be a consistent estimate of $\alpha^1_{k\bullet}(\tau_t) := \left[ \alpha^1_{k1}(\tau_t), \cdots, \alpha^1_{kK_1}(\tau_t), \alpha^1_k(\tau_t) \right]^\mathsf{T}$ when $t \leqslant t_0$ but this consistency becomes invalid for at least one k when $t > t_0$. Similarly, we expect $\check{\alpha}^2_{k\bullet}(\tau_t)$ to be a consistent estimate of $\alpha^2_{k\bullet}(\tau_t) := \left[ \alpha^2_{k1}(\tau_t), \cdots, \alpha^2_{kK_2}(\tau_t), \alpha^2_k(\tau_t) \right]^\mathsf{T}$ when $t > t_0$ but the consistency becomes invalid for at least one k when $t \leqslant t_0$. Define

$$
\check{\delta}_\alpha(t) = \max_{1 \leqslant k \leqslant \widehat{K}_1} \left| \check{\alpha}^1_{k\bullet}(\tau_{t+1}) - \check{\alpha}^1_{k\bullet}(\tau_t) \right|_2 + \max_{1 \leqslant k \leqslant \widehat{K}_2} \left| \check{\alpha}^2_{k\bullet}(\tau_{t+1}) - \check{\alpha}^2_{k\bullet}(\tau_t) \right|_2. \tag{E.1}
$$

The refined break location estimation is then obtained by

$$
\check{t} = \operatorname*{argmax}_t \check{\delta}_\alpha(t). \tag{E.2}
$$

## E.2 Assumptions and consistency

We next provide some high-level conditions and then derive the consistency property of $\check{t}$. We start with the introduction of some notation. Let

$$
X_{i,t}^{1,\diamond} = \left[ \sum_{g_j^1=1} \widetilde{w}_{ij} x_{j,t}, \cdots, \sum_{g_j^1=K_1} \widetilde{w}_{ij} x_{j,t}, x_{i,t} \right]^{\mathsf{T}},
$$

$$
X_{i,t}^{2,\diamond} = \left[ \sum_{g_j^2=1} \widetilde{w}_{ij} x_{j,t}, \cdots, \sum_{g_j^2=K_2} \widetilde{w}_{ij} x_{j,t}, x_{i,t} \right]^{\mathsf{T}},
$$

$$
\Delta_{ij,t}^{1,\diamond} = \mathsf{E}\left[ X_{i,t}^{1,\diamond}\left(X_{j,t}^{1,\diamond}\right)^{\mathsf{T}} \right], \quad \Delta_{i,t}^{1,\diamond} = \Delta_{ii,t}^{1,\diamond},
$$

$$
\Delta_{ij,t}^{2,\diamond} = \mathsf{E}\left[ X_{i,t}^{2,\diamond}\left(X_{j,t}^{2,\diamond}\right)^{\mathsf{T}} \right], \quad \Delta_{i,t}^{2,\diamond} = \Delta_{ii,t}^{2,\diamond}.
$$

**Assumption E.1.** *(i) Let*

$$
\Delta_{\mathscr{G}_k^1,t} = \frac{1}{\mathrm{card}(\mathscr{G}_k^1)} \sum_{g_i^1=k} \Delta_{i,t}^{1,\diamond} \qquad \Delta_{\mathscr{G}_k^2,t} = \frac{1}{\mathrm{card}(\mathscr{G}_k^2)} \sum_{g_i^2=k} \Delta_{i,t}^{2,\diamond}
$$

*be positive definite uniformly over* $k$ *and* $t$.

*(ii) There exist* $\Delta_{\mathscr{G}_k^1,t}^{\alpha}$, $k = 1, \cdots, K_1$, *and* $\Delta_{\mathscr{G}_k^2,t}^{\alpha}$, $k = 1, \cdots, K_2$, *such that*

$$
\max_{1 \leqslant k \leqslant K_1} \max_{t_0 < t \leqslant T} \left| \frac{1}{\mathrm{card}(\mathscr{G}_k^1)} \sum_{g_i^1=k} X_{i,t}^{1,\diamond} \left[ \sum_{j \neq i} \alpha_{g_i^2 g_j^2}^2(\tau_t) \widetilde{w}_{ij} x_{j,t-1} + \alpha_{g_i^2}^2(\tau_t) x_{i,t-1} \right] - \Delta_{\mathscr{G}_k^1,t}^{\alpha} \right|_2 = o_{\mathsf{P}}(1),
$$

$$
\max_{1 \leqslant k \leqslant K_2} \max_{1 < t \leqslant t_0} \left| \frac{1}{\mathrm{card}(\mathscr{G}_k^2)} \sum_{g_i^2=k} X_{i,t}^{2,\diamond} \left[ \sum_{j \neq i} \alpha_{g_i^1 g_j^1}^1(\tau_t) \widetilde{w}_{ij} x_{j,t-1} + \alpha_{g_i^1}^1(\tau_t) x_{i,t-1} \right] - \Delta_{\mathscr{G}_k^2,t}^{\alpha} \right|_2 = o_{\mathsf{P}}(1).
$$

**Assumption E.2.** *For any group* $\mathscr{G}$ *with cardinality sufficiently large,*

$$
\max_{1 \leqslant t \leqslant T} \mathsf{E}\left( \left| \sum_{i \in \mathscr{G}} \left[ X_{i,t}^{1,\diamond}\left(X_{i,t}^{1,\diamond}\right)^{\mathsf{T}} - \Delta_{i,t}^{1,\diamond} \right] \right|_{\mathsf{F}}^{\iota} + \left| \sum_{i \in \mathscr{G}} \left[ X_{i,t}^{2,\diamond}\left(X_{i,t}^{2,\diamond}\right)^{\mathsf{T}} - \Delta_{i,t}^{2,\diamond} \right] \right|_{\mathsf{F}}^{\iota} \right) = O\left( [\mathrm{card}(\mathscr{G})]^{\iota/2} \right),
$$

$$
\max_{1 \leqslant t \leqslant T} \mathsf{E}\left( \left| \sum_{i \in \mathscr{G}} X_{i,t-1}^{1,\diamond} \varepsilon_{i,t} \right|_2^{\iota} + \left| \sum_{i \in \mathscr{G}} X_{i,t-1}^{2,\diamond} \varepsilon_{i,t} \right|_2^{\iota} \right) = O\left( [\mathrm{card}(\mathscr{G})]^{\iota/2} \right),
$$

*where* $\iota$ *is a positive constant larger than 2. In addition,* $T/N^{\iota/2} \to 0$.

**Assumption E.3.** *Let*

$$
\delta_{\alpha} := \max_{1 \leqslant k \leqslant K_1} \left| \alpha_{k\bullet}^{\dagger}(\tau_{t_0+1}) - \alpha_{k\bullet}^1(\tau_{t_0}) \right|_2 + \max_{1 \leqslant k \leqslant K_2} \left| \alpha_{k\bullet}^2(\tau_{t_0+1}) - \alpha_{k\bullet}^{\ddagger}(\tau_{t_0}) \right|_2 \geqslant c_5 > 0
$$

24

*where*

$$\alpha_{k\bullet}^{\dagger}(\tau_t) = \Delta_{\mathscr{G}_k^1,t}^{-1} \Delta_{\mathscr{G}_k^1,t}^{\alpha}, \quad t_0 < t \leqslant T,$$

$$\alpha_{k\bullet}^{\ddagger}(\tau_t) = \Delta_{\mathscr{G}_k^2,t}^{-1} \Delta_{\mathscr{G}_k^2,t}^{\alpha}, \quad 1 < t \leqslant t_0.$$

Assumption E.1(i) is similar to Assumption 5(ii) in the main text. Meanwhile, a combination of the conditions in Assumption E.1(i)(ii) ensures that the limits of $\check{\alpha}_{k\bullet}^1(\cdot)$ and $\check{\alpha}_{k\bullet}^2(\cdot)$ are well defined even when groups are misclassified, i.e., $\alpha_{k\bullet}^{\dagger}(\cdot)$ and $\alpha_{k\bullet}^{\ddagger}(\cdot)$ in Assumption E.3 are well defined. Assumption E.2 imposes some high-level moment conditions to restrict weak correlation between nodes and indicates that $T$ can be either smaller or larger than $N$ (depending on the value of $\iota$). Assumption E.3 indicates that the break size at $t_0$ needs to be bounded away from zero. This condition can be satisfied when there are breaks in the group-specific coefficient functions, group membership or number. Some examples are provided in Appendix E.3 to verify this condition.

**Proposition E.1.** *Suppose that the conditions of Theorem 5.1 and Assumptions E.1–E.3 are satisfied. Then we have* $P(\check{t} = t_0) \to 1$.

**Proof of Proposition E.1.** Let $\mathscr{E}_{\mathscr{G}}^1$ denote the event that $\widehat{\mathscr{G}}_k^1 = \mathscr{G}_k^1$, $1 \leqslant k \leqslant K_1$, and $\widehat{K}_1 = K_1$, and let $\mathscr{E}_{\mathscr{G}}^2$ denote the event that $\widehat{\mathscr{G}}_k^2 = \mathscr{G}_k^2$, $1 \leqslant k \leqslant K_2$, and $\widehat{K}_2 = K_2$. For $1 \leqslant t \leqslant t_0$, conditional on $\mathscr{E}_{\mathscr{G}}^1 \cap \mathscr{E}_{\mathscr{G}}^2$, the grouped time-varying network VAR model is correctly fitted. Hence, we have

$$
\begin{aligned}
\check{\alpha}_{k\bullet}^1(\tau_t) &= \left[ \sum_{i\in\widehat{\mathscr{G}}_k^1} \check{X}_{i,t-1}^1 \left(\check{X}_{i,t-1}^1\right)^{\mathsf{T}} \right]^{-1} \sum_{i\in\widehat{\mathscr{G}}_k^1} \check{X}_{i,t-1}^1 x_{i,t} \\
&= \left[ \sum_{i\in\mathscr{G}_k^1} X_{i,t-1}^{1,\diamond} \left(X_{i,t-1}^{1,\diamond}\right)^{\mathsf{T}} \right]^{-1} \sum_{i\in\mathscr{G}_k^1} X_{i,t-1}^{1,\diamond} x_{i,t} \\
&= \alpha_{k\bullet}^1(\tau_t) + \left[ \sum_{i\in\mathscr{G}_k^1} X_{i,t-1}^{1,\diamond} \left(X_{i,t-1}^{1,\diamond}\right)^{\mathsf{T}} \right]^{-1} \sum_{i\in\mathscr{G}_k^1} X_{i,t-1}^{1,\diamond} \varepsilon_{i,t}
\end{aligned}
\tag{E.3}
$$

conditional on $\mathscr{E}_{\mathscr{G}}^1 \cap \mathscr{E}_{\mathscr{G}}^2$. By Assumptions E.2 and 6(iv) and the Bonferroni and Markov inequalities, we may show that, for any $\eta > 0$

$$
\begin{aligned}
&P\left( \max_{1\leqslant t\leqslant t_0} \left| \frac{1}{\mathrm{card}(\mathscr{G}_k^1)} \sum_{i\in\mathscr{G}_k^1} X_{i,t-1}^{1,\diamond} \varepsilon_{i,t} \right|_2 > \eta \right) \\
&\leqslant \sum_{t=1}^{t_0} P\left( \left| \frac{1}{\mathrm{card}(\mathscr{G}_k^1)} \sum_{i\in\mathscr{G}_k^1} X_{i,t-1}^{1,\diamond} \varepsilon_{i,t} \right|_2 > \eta \right) \\
&\leqslant \sum_{t=1}^{t_0} [\mathrm{card}(\mathscr{G}_k^1)]^{-\iota} \mathsf{E} \left| \sum_{i\in\mathscr{G}_k^1} X_{i,t-1}^{1,\diamond} \varepsilon_{i,t} \right|_2^{\iota}
\end{aligned}
$$

25

$$= O\left(TN^{-\iota/2}\right) = o(1),$$

which leads to

$$\max_{1\leqslant t\leqslant t_0}\left|\frac{1}{\operatorname{card}(\mathscr{G}_k^1)}\sum_{i\in\mathscr{G}_k^1}X_{i,t-1}^{1,\diamond}\varepsilon_{i,t}\right|_2 = o_P(1). \tag{E.4}$$

Similarly, we can also prove that

$$\max_{1\leqslant t\leqslant t_0}\left|\frac{1}{\operatorname{card}(\mathscr{G}_k^1)}\sum_{i\in\mathscr{G}}\left[X_{i,t}^{1,\diamond}\left(X_{i,t}^{1,\diamond}\right)^{\intercal}-\Delta_{i,t}^{1,\diamond}\right]\right|_F = o_P(1). \tag{E.5}$$

With (E.3)–(E.5), as $K_1$ is fixed, we readily have that

$$\max_{1\leqslant k\leqslant K_1}\max_{1\leqslant t\leqslant t_0}\left|\breve{\alpha}_{k\bullet}^1(\tau_t)-\alpha_{k\bullet}^1(\tau_t)\right|_2 = o_P(1). \tag{E.6}$$

Similarly, by Assumptions E.1 and E.2, we also have

$$\max_{1\leqslant k\leqslant K_1}\max_{t_0+1\leqslant t\leqslant T}\left|\breve{\alpha}_{k\bullet}^1(\tau_t)-\alpha_{k\bullet}^\dagger(\tau_t)\right|_2 = o_P(1), \tag{E.7}$$

$$\max_{1\leqslant k\leqslant K_1}\max_{t_0+1\leqslant t\leqslant T}\left|\breve{\alpha}_{k\bullet}^2(\tau_t)-\alpha_{k\bullet}^2(\tau_t)\right|_2 = o_P(1), \tag{E.8}$$

$$\max_{1\leqslant k\leqslant K_1}\max_{1\leqslant t\leqslant t_0}\left|\breve{\alpha}_{k\bullet}^2(\tau_t)-\alpha_{k\bullet}^\ddagger(\tau_t)\right|_2 = o_P(1). \tag{E.9}$$

Note that $\alpha_{k\bullet}^1(\cdot)$ and $\alpha_{k\bullet}^\ddagger(\cdot)$ are continuous over $[0,\tau_{t_0}]$ whereas $\alpha_{k\bullet}^2(\cdot)$ and $\alpha_{k\bullet}^\dagger(\cdot)$ are continuous over $[\tau_{t_0},1]$. By virtue of (E.6)–(E.9), we have $\breve{\delta}_\alpha(t) = o_P(1)$ when $t\neq t_0$ and $\breve{\delta}_\alpha(t_0)-\delta_\alpha(t_0) = o_P(1)$. Finally, by Assumption E.3, we may prove that $P(\breve{t}=t_0)\to 1$. $\qquad\square$

## E.3   Verification of Assumption E.3

We next provide a few examples to verify the condition $\delta_\alpha(t_0) > c_5$ in Assumption E.3.

**Example E.1.** Suppose that the group structure is time invariant, i.e., $\mathscr{G}^1 = \mathscr{G}^2$ and $K_1 = K_2 = K_0$, but there exists a $k_0 \in \{1, 2, \cdots, K_0\}$ such that $\alpha_{k_0\bullet}(\cdot)$ defined in Section 4 is discontinuous at $\tau_{t_0}$ satisfying

$$|\alpha_{k_0\bullet}(\tau_{t_0+1})-\alpha_{k_0\bullet}(\tau_{t_0})|_2 \geqslant c_6 > 0,$$

where $\alpha_{k_0\bullet}^r(\cdot)$ and $\alpha_{k_0\bullet}^l(\cdot)$ denote the right and left limits of $\alpha_{k_0\bullet}(\cdot)$, respectively. In this case, we may verify that

$$\alpha_{k\bullet}^\dagger(\tau_t) = \alpha_{k\bullet}(\tau_t), \quad \alpha_{k\bullet}^\ddagger(\tau_t) = \alpha_{k\bullet}(\tau_t).$$

Hence, it is easy to show that

$$\delta_\alpha \geqslant |\alpha_{k_0\bullet}(\tau_{t_0+1})-\alpha_{k_0\bullet}(\tau_{t_0})|_2 \geqslant c_6,$$

verifying Assumption E.3. Similarly, the condition of $\delta_\beta(t_0) > c_2$ in Assumption 6(iii) is also satisfied.

In the following examples, for notational simplicity, we consider a special time-varying network VAR model with break in either the group membership or number:

$$x_{i,t} = \begin{cases} \gamma_{g_i^1}(\tau_t) \sum_{j \neq i} \widetilde{w}_{ij} x_{j,t-1} + \alpha_{g_i^1}(\tau_t) x_{i,t-1} + \varepsilon_{i,t}, & 1 \leqslant t \leqslant t_0, \\ \gamma_{g_i^2}(\tau_t) \sum_{j \neq i} \widetilde{w}_{ij} x_{j,t-1} + \alpha_{g_i^2}(\tau_t) x_{i,t-1} + \varepsilon_{i,t}, & t_0 + 1 \leqslant t \leqslant T, \end{cases} \tag{E.10}$$

where $\gamma_{g_i^1}(\cdot)$ and $\gamma_{g_i^2}(\cdot)$ denote the time-varying network effects invariant over nodes which the $i$-th node follows. This model is considered in an earlier version of our paper with no break in the group structure.

**Example E.2.** Suppose that model (E.10) holds, the group number is time invariant, $K_1 = K_2 = 2$, but there is a break in the group membership at the time point $t_0$. Before the break point, $g_i^1 = 1$ for $1 \leqslant i \leqslant \lfloor N/2 \rfloor$ and $g_i^1 = 2$ for $\lfloor N/2 \rfloor + 1 \leqslant i \leqslant N$; after the break point, $g_i^2 = 1$ for $1 \leqslant i \leqslant \lfloor N/4 \rfloor$ and $g_i^2 = 2$ for $\lfloor N/4 \rfloor + 1 \leqslant i \leqslant N$. The group-specific coefficient functions $\gamma_1(\cdot)$, $\gamma_2(\cdot)$, $\alpha_1(\cdot)$ and $\alpha_2(\cdot)$ are continuous over $[0, 1]$, and

$$|\gamma_2(\tau_{t_0}) - \gamma_1(\tau_{t_0})| + |\alpha_2(\tau_{t_0}) - \alpha_1(\tau_{t_0})| \geqslant c_7 > 0. \tag{E.11}$$

Note that

$$\alpha_{1\bullet}(\tau_t) = [\gamma_1(\tau_t), \alpha_1(\tau_t)]^\intercal, \quad \alpha_{2\bullet}(\tau_t) = [\gamma_2(\tau_t), \alpha_2(\tau_t)]^\intercal.$$

With some elementary calculations, we may show that

$$\alpha_{1\bullet}^\dagger(\tau_t) = \mathbf{W}_1^\dagger \begin{bmatrix} \gamma_1(\tau_t) \\ \alpha_1(\tau_t) \end{bmatrix} + \mathbf{W}_2^\dagger \begin{bmatrix} \gamma_2(\tau_t) \\ \alpha_2(\tau_t) \end{bmatrix}, \quad \alpha_{2\bullet}^\dagger(\tau_t) = [\gamma_2(\tau_t), \alpha_2(\tau_t)]^\intercal,$$

$$\alpha_{1\bullet}^\ddagger(\tau_t) = [\gamma_1(\tau_t), \alpha_1(\tau_t)]^\intercal, \quad \alpha_{2\bullet}^\ddagger(\tau_t) = \mathbf{W}_1^\ddagger \begin{bmatrix} \gamma_1(\tau_t) \\ \alpha_1(\tau_t) \end{bmatrix} + \mathbf{W}_2^\ddagger \begin{bmatrix} \gamma_2(\tau_t) \\ \alpha_2(\tau_t) \end{bmatrix},$$

where $\mathbf{W}_1^\dagger$, $\mathbf{W}_2^\dagger$, $\mathbf{W}_1^\ddagger$ and $\mathbf{W}_2^\ddagger$ are all positive definite satisfying $\mathbf{W}_1^\dagger + \mathbf{W}_2^\dagger = \mathbf{I}_2$ and $\mathbf{W}_1^\ddagger + \mathbf{W}_2^\ddagger = \mathbf{I}_2$. Under the condition (E.11), it is easy to verify Assumption E.3. The verification of $\delta_\beta(t_0) > c_2$ in Assumption 6(iii) is similar.

It is worth pointing out that Assumption E.3 may be invalid when the break in the group membership is sparse. In Example E.2, consider $g_i^1 = 1$ for $1 \leqslant i \leqslant \lfloor N/2 \rfloor$ and $g_i^1 = 2$ for $\lfloor N/2 \rfloor + 1 \leqslant i \leqslant N$ before the break point; $g_i^2 = 1$ for $1 \leqslant i \leqslant \lfloor N/2 \rfloor - s$ and $g_i^2 = 2$ for $\lfloor N/2 \rfloor - s + 1 \leqslant i \leqslant N$ after the break point, where $s$ is a small and fixed positive integer. In this case, $\delta_\alpha$ would converge to zero, indicating that Assumption E.3 is violated.

**Example E.3.** Suppose that model (E.10) holds. Consider a break in the group number: $K_1 = 2$ and $K_2 = 1$. Before the break time $t_0$, $g_i^1 = 1$ for $1 \leqslant i \leqslant \lfloor N/2 \rfloor$ and $g_i^1 = 2$ for $\lfloor N/2 \rfloor + 1 \leqslant i \leqslant N$; after the break point, the two groups merge, i.e., $g_i^2 \equiv 1$. Using the arguments in Example E.2, we can similarly verify Assumption E.3 as well as $\delta_\beta(t_0) > c_2$ in Assumption 6(iii).

Table F.8: Estimation performance of the group number $M_0$ and membership $\mathcal{G}^\circ$

|  |  |  | Fixed group | | Random group | |
| --- | --- | --- | --- | --- | --- | --- |
| Sparsity | Measurement | $T \setminus N$ | 100 | 200 | 100 | 200 |
| $\bar{w} = 0.025$ | $\mathsf{AC}(M_0)$ | 300 | 0.890 | 0.972 | 0.916 | 0.978 |
|  |  | 600 | 1.000 | 0.999 | 0.974 | 0.998 |
|  | $\mathsf{Purity}(\mathcal{G}^\circ)$ | 300 | 0.939 | 0.761 | 0.901 | 0.759 |
|  |  | 600 | 0.982 | 0.874 | 0.961 | 0.860 |
|  |  |  |  |  |  |  |
| $\bar{w} = 0.075$ | $\mathsf{AC}(M_0)$ | 300 | 0.904 | 0.702 | 0.856 | 0.697 |
|  |  | 600 | 0.999 | 0.829 | 0.988 | 0.857 |
|  | $\mathsf{Purity}(\mathcal{G}^\circ)$ | 300 | 0.649 | 0.471 | 0.622 | 0.467 |
|  |  | 600 | 0.782 | 0.547 | 0.762 | 0.556 |

# F   Extra numerical results

## F.1   Results of the clustering algorithm in Appendix A

We next report the finite-sample performance of the clustering algorithm introduced in Appendix A to determine the homogeneity structure on the time-varying network spillover effects. We start with the data generating process provided in Section 6.1. Note that $M_0 = 2$ and

$$\mathcal{G}_1^\circ = \left\{ (i,j) \in \bar{\mathcal{N}}_N : i \in \mathcal{G}_1, j \in \mathcal{G}_1 \right\}, \quad \mathcal{G}_2^\circ = \bar{\mathcal{N}}_N / \mathcal{G}_1^\circ.$$

To evaluate the homogeneity structure estimation accuracy, as in Section 6.1, we compute

$$\mathsf{AC}(M_0) = \frac{1}{R} \sum_{r=1}^{R} \mathrm{I}(\widehat{M}_r = M_0) \quad \text{and} \quad \mathsf{Purity}(\mathcal{G}^\circ) = \frac{1}{RN} \sum_{r=1}^{R} \sum_{m=1}^{\widehat{M}_r} \max_{1 \leqslant j \leqslant M_0} \left| \widehat{\mathcal{G}}_{m,r}^\circ \cap \mathcal{G}_j^\circ \right|,$$

where $\widehat{M}_r$ and $\widehat{\mathcal{G}}_{m,r}^\circ$ are the estimates of the group number and membership in the $r$-th replication. The relevant results are summarized in Table F.8. In general, the performance of the group number and membership estimates improve as $T$ increases from 300 to 600 whereas the membership estimation is sensitive to the network sparsity level with $\mathsf{Purity}(\mathcal{G})$ decreasing significantly as $\bar{w}$ increases from 0.025 to 0.075.

In addition, we report the clustering results when there exists a break in the group membership. The data generating process is described in Section 6.2. Table F.9 summarizes the estimation performance for the group number $M_0$ and membership $\mathcal{G}^\circ$. The findings drawn from Table F.9 are similar to those from Table F.8.

Table F.9: Estimation performance of the group number $M_0$ and membership $\mathcal{G}^\circ$ (with a break)

| Sparsity | Measurement | T \ N | Fixed group 100 | Fixed group 200 | Random group 100 | Random group 200 |
|---|---|---|---|---|---|---|
| $\bar{w} = 0.025$ | AC($M_0$) | 400 | 1.000 | 0.995 | 1.000 | 0.996 |
| | | 800 | 1.000 | 0.998 | 1.000 | 0.999 |
| | Purity($\mathcal{G}^\circ$) | 400 | 0.961 | 0.811 | 0.930 | 0.808 |
| | | 800 | 0.989 | 0.903 | 0.975 | 0.895 |
| $\bar{w} = 0.075$ | AC($M_0$) | 400 | 0.989 | 0.933 | 0.975 | 0.938 |
| | | 800 | 1.000 | 0.980 | 0.995 | 0.979 |
| | Purity($\mathcal{G}^\circ$) | 400 | 0.729 | 0.578 | 0.712 | 0.577 |
| | | 800 | 0.827 | 0.658 | 0.812 | 0.654 |

## F.2 Clustering results for grouped network VAR with constant coefficients

We next report the additional simulation result by treating the data generating process in Section 6.1 as Zhu, Xu & Fan (2023)'s grouped network VAR with constant coefficients and adopting their clustering algorithm. The estimation results for the group structure are shown in Table F.10. By comparing the results with those in Table 1, we note that failure to account for smooth structural changes in network VAR would substantially affect the estimation accuracy of the latent group structure for network time series, resulting in lower values of AC($K_0$) and Purity($\mathcal{G}$).

Table F.10: Estimation performance of the group structure using Zhu, Xu & Fan (2023)'s model and algorithm

| Sparsity | Measurement | T \ N | Fixed group 100 | Fixed group 200 | Random group 100 | Random group 200 |
|---|---|---|---|---|---|---|
| $\bar{w} = 0.025$ | AC($K_0$) | 300 | 0.844 | 0.851 | 0.827 | 0.830 |
| | | 600 | 0.670 | 0.714 | 0.557 | 0.743 |
| | Purity($\mathcal{G}$) | 300 | 0.555 | 0.577 | 0.561 | 0.582 |
| | | 600 | 0.538 | 0.681 | 0.551 | 0.675 |
| $\bar{w} = 0.075$ | AC($K_0$) | 300 | 0.802 | 0.773 | 0.796 | 0.823 |
| | | 600 | 0.763 | 0.705 | 0.719 | 0.733 |
| | Purity($\mathcal{G}$) | 300 | 0.616 | 0.644 | 0.604 | 0.663 |
| | | 600 | 0.711 | 0.750 | 0.685 | 0.757 |

## F.3 Extra empirical result

In addition, we plot the estimated time-varying momentum effects from both models in Figures F.2 and F.3, respectively. In each figure, the left plot is for Group 1, while the right one is for Group 2. Although
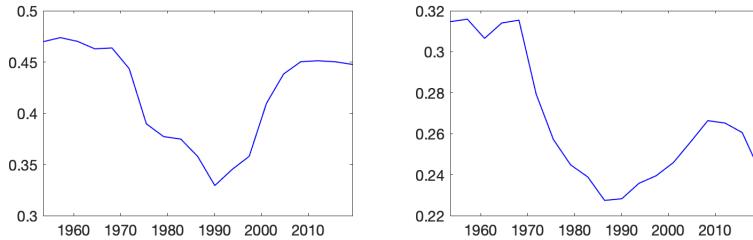
Figure F.2: Estimated time-varying momentum effects in model (i)
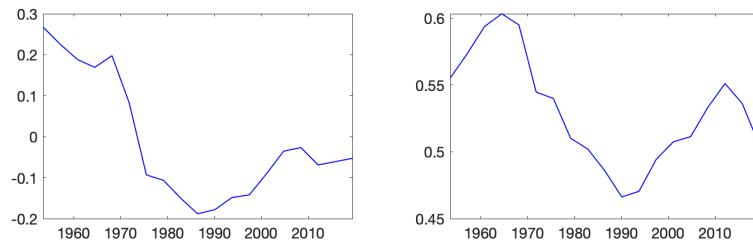


Figure F.3: Estimated time-varying momentum effects in model (ii)

both figures present similar patterns of movement across groups, the Y-axis covers different ranges.

# References

Billingsley, P. (1995). *Probability and Measure* (3rd Edition). John Wiley and Sons.

Burkholder, D. L. (1973). Distribution function inequalities for martingales. *The Annals of Probability* **1**, 19–42.

Dahlhaus, R., Richter, S. and Wu, W. (2009). Towards a general theory for nonlinear locally stationary processes. *Bernoulli* **25**, 1013–1044.

Freedman D.A. (1975). On tail probabilities for martingales. *The Annals of Probability* **3**, 100–118.

Hall, P. and Heyde, C.C. (1980). *Martingale limit theory and its application*. Academic press.

Wu, W. (2005). Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences* **102**, 14150–14154.

Zhang, D. and Wu, W. (2021). Convergence of covariance and spectral density estimates for high-dimensional locally stationary processes. *The Annals of Statistics* **49**, 233–254.

Zhu, X., Xu, G. and Fan, J. (2023). Simultaneous estimation and group identification for network vector autoregressive model with heterogeneous nodes. Forthcoming in *Journal of Econometrics*.