

DEPARTMENT OF ECONOMETRICS
AND BUSINESS STATISTICS

ISSN 1440-771X

WORKING PAPER SERIES

Inference of Grouped Time-Varying Network Vector Autoregression Models

Degui Li*, Bin Peng†, Songqiao Tang‡, Weibiao Wu§
University of York, Monash University, Zhejiang University, University of
Chicago

Inference of Grouped Time-Varying Network Vector Autoregression Models

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

University of York, Monash University, Zhejiang University, University of Chicago

Version: April 4, 2023

Abstract

This paper considers statistical inference of time-varying network vector autoregression models 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 normalized distance matrix estimates is combined with a generalized information 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. In addition, a post-grouping specification test is conducted to verify the validity of the parametric model assumption for group-specific time-varying coefficient functions, and the asymptotic theory is derived for the test statistic constructed via a kernel weighted quadratic form under the null and alternative hypotheses. Numerical studies including Monte-Carlo simulation and an empirical application to the global trade flow data are presented to examine the finite-sample performance of the developed model and methodology.

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

^{*}Department of Mathematics, University of York, UK. Email: degui.li@york.ac.uk

[†]Department of Econometrics and Business Statistics, Monash University, Australia. Email: Bin.Peng@monash.edu

[‡]School of Mathematical Sciences, Zhejiang University, China. Email: statsq@zju.edu.cn

[§]Department of Statistics, University of Chicago, US. Email: wbwu@uchicago.edu

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 autoregression (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 much 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 and Bickel, 2011](#); [Bonhomme and Manresa, 2015](#); [Kock and Callot, 2015](#); [Davis, Zang and Zheng, 2016](#); [Miao, Phillips and Su, 2022](#)) and factor-augmented VAR (e.g., [Bernanke, Boivin and Elias, 2005](#); [Bai and Ng, 2006](#); [Bai, Li and 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 $X_t = (x_{1,t}, \dots, x_{N,t})^\top$ 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 network VAR model is defined by

$$x_{i,t} = \beta_1 \sum_{j \neq i} \tilde{w}_{ij} x_{j,t-1} + \beta_2 x_{i,t-1} + \varepsilon_{i,t}, \quad i = 1, \dots, N, \quad (1.1)$$

where $\tilde{w}_{ij} = w_{ij}/n_i$ with $n_i = \sum_{j \neq i} w_{ij}$, β_1 and β_2 are unknown parameters, and $\varepsilon_t = (\varepsilon_{1,t}, \dots, \varepsilon_{N,t})^\top$ is a sequence of independent and identically distributed (i.i.d.) error vectors. The above network VAR model formulation contains two regression components: $\beta_1 \sum_{j \neq i} \tilde{w}_{ij} x_{j,t-1}$ 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 linear network VAR model is easy to interpret and implement, it may be invalid in empirical applications. In particular, there often 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 autoregression 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 and Malikov \(2018\)](#), [Wu \(2019\)](#) and [Chen *et al* \(2023\)](#) extend the linear network VAR models, allowing the coefficients to vary smoothly over time or with a stationary index variable. To relax the homogeneity restriction in linear VAR models, [Zhu and Pan \(2020\)](#) introduce a grouped linear network VAR model via a mixture Gaussian distribution and use an EM estimation algorithm; [Guðmundsson and Brownlees \(2021\)](#) propose a stochastic block VAR model and detect a latent group structure on the network spillover effects; [Chen, Fan and Zhu \(2022\)](#) study a community network VAR model and allow network effects to vary over different communities; and more recently [Zhu, Xu and Fan \(2022\)](#) introduce a least squares algorithm to simultaneously estimate the parameters and identify the group structure for heterogeneous network VAR models. As far as we know, however, there is virtually no work which tackles the two issues jointly. The present paper aims to fill in this gap by introducing a general time-varying network VAR model framework satisfying a latent group structure, i.e., the time-varying network autoregression relationships are invariant within a group of nodes, but change over different groups.

With the latent group structure, we substantially reduce the number of unknown coefficient functions in the network autoregression for time-varying momentum and network spillover effects. Since neither the group number nor membership is known a priori, we combine a classic agglomerative clustering algorithm (with the normalized distance matrix estimator) and a generalized information criterion to consistently estimate the group structure. The developed estimation methodology complements recent developments on latent group estimation in the context of large panel data (e.g., [Bonhomme and Manresa, 2015](#); [Ke, Li and Zhang, 2016](#); [Su, Shi and Phillips, 2016](#); [Ando and Bai, 2017](#); [Vogt and Linton, 2017, 2020](#); [Chen, 2019](#); [Su, Wang and Jin, 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. In addition, the normalized distance estimates used in our clustering algorithm have the approximate unit variance, making their sizes comparable over nodes and resulting in stable membership estimation.

To improve the convergence rates of preliminary time-varying coefficient estimation which only uses the sample information from one node and its direct neighbors (see [Section 3.1](#)), we propose a post-grouping local linear smoothing method to estimate the group-specific time-varying momentum, network and nodal 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. We also conduct the post-grouping statistical inference by testing the hypothesis that some group-specific time-varying coefficient functions belong to a given parametric function family. For example, it is often

interesting to test whether the time-varying coefficients can be approximated by some constant values or linear functions. The test statistic is constructed via the kernel-weighted quadratic form and the asymptotic theory is established for the developed test statistic under both the null and alternative hypotheses.

The finite-sample Monte-Carlo simulation study shows that the proposed clustering algorithm and information criterion can accurately estimate the latent group membership and number as long as the time series length T is moderately large; and the post-grouping local linear estimates perform significantly better than the naive heterogeneous local linear estimates. The test statistic with the kernel-weighted quadratic form also has stable finite-sample performance in both the test size and power. The developed model and methodology are further applied to analyze the export flow data collected for 90 countries over the time period between January 1995 and December 2021. The empirical result reveals that there exist three groups over the node-specific time-varying coefficients, and the momentum effects tend to have negative and time-varying patterns which are sensible from the demand and supply perspective.

The rest of the paper is organized as follows. Section 2 introduces the heterogeneous 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 membership and number and establishes the consistency property. Section 4 proposes the post-grouping nonparametric estimation and inference and derives the relevant asymptotic theory. Section 5 reports both the simulation and empirical studies. Section 6 concludes the paper. Appendix A contains proofs of the main asymptotic theorems and Appendix B contains some technical lemmas with their proofs. Throughout the paper, we let $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ be the floor and ceiling functions, respectively. For a vector $\mathbf{x} = (x_1, \dots, x_p)^\top \in \mathcal{R}^p$, we write $\|\mathbf{x}\| = (\sum_{i=1}^p x_i^2)^{1/2}$ as the Euclidean norm; for a matrix $\mathbf{\Sigma} = (\sigma_{ij})_{p \times p}$, write $\|\mathbf{\Sigma}\| = (\sum_{i=1}^p \sum_{j=1}^p \sigma_{ij}^2)^{1/2}$, $\|\mathbf{\Sigma}\|_\infty = \max_{1 \leq i, j \leq p} |\sigma_{ij}|$ and $\|\mathbf{\Sigma}\|_2 = \max_{\mathbf{x} \in \mathcal{R}^p: \|\mathbf{x}\|=1} |\mathbf{\Sigma}\mathbf{x}|$; and for a p -dimensional random vector \mathbf{Z} , write $\mathbf{Z} \in \mathcal{L}^q$ if $\|\mathbf{Z}\|_q := [\mathbf{E}(|\mathbf{Z}|^q)]^{1/q} < \infty$, $q \geq 1$. 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 $a_n/b_n \rightarrow 0$ as $n \rightarrow \infty$, $a_n/b_n \rightarrow 0$ with probability approaching one (*w.p.a.1*), and $0 < \underline{c} \leq a_n/b_n \leq \bar{c} < \infty$, 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 Heterogeneous time-varying network VAR

Consider the following time-varying network VAR model:

$$x_{i,t} = \beta_{i,1}(\tau_t) \sum_{j \neq i} \tilde{w}_{ij} x_{j,t-1} + \beta_{i,2}(\tau_t) x_{i,t-1} + \varepsilon_{i,t}, \quad i = 1, \dots, N, \quad t = 1, \dots, T, \quad (2.1)$$

where $\tau_t = t/T$ denotes the scaled time point, $\beta_{i,1}(\cdot)$ and $\beta_{i,2}(\cdot)$ are node-specific 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 heterogeneous 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. 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. Let $\tilde{\mathbf{W}}$ be the row-normalized adjacency matrix with the (i, j) -entry being \tilde{w}_{ij} ,

$$\mathbf{B}_1(\tau_t) = \text{diag}\{\beta_{1,1}(\tau_t), \dots, \beta_{N,1}(\tau_t)\} \quad \text{and} \quad \mathbf{B}_2(\tau_t) = \text{diag}\{\beta_{1,2}(\tau_t), \dots, \beta_{N,2}(\tau_t)\}.$$

Then, we may rewrite model (2.1) as

$$\mathbf{X}_t = \mathbf{B}(\tau_t) \mathbf{X}_{t-1} + \boldsymbol{\varepsilon}_t, \quad \mathbf{B}(\cdot) = \mathbf{B}_1(\cdot) \tilde{\mathbf{W}} + \mathbf{B}_2(\cdot), \quad t = 1, \dots, T. \quad (2.2)$$

Model (2.2) falls within the high-dimensional time-varying VAR model framework which has received increasing attention in recent years. For instance, [Ding, Qiu and Chen \(2017\)](#) propose a kernel-weighted ℓ_1 -regularised estimation for a time-varying VAR model when the number of time series variables diverges at a sub-exponential rate of T ; [Xu, Chen and 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 time-varying network effects. In this paper, we do not require the sparsity restriction by decomposing the time-varying transition matrix $\mathbf{B}(\tau_t)$ into two components: $\mathbf{B}_2(\tau_t)$ capturing the momentum effects; and $\mathbf{B}_1(\tau_t) \tilde{\mathbf{W}}$ capturing the dynamic network spillover effects.

In the homogenous or grouped linear network VAR (e.g., [Zhu *et al*, 2017](#); [Chen, Fan and Zhu, 2022](#)), the so-called nodal effect is often incorporated in the model formulation. Hence, we may

extend model (2.1) to

$$x_{i,t} = \beta_{i,1}(\tau_t) \sum_{j \neq i} \tilde{w}_{ij} x_{j,t-1} + \beta_{i,2}(\tau_t) x_{i,t-1} + Z_i^\top \gamma(\tau_t) + \varepsilon_{i,t}, \quad (2.3)$$

where Z_i is a p -dimensional vector of node-specific exogenous covariates and $\gamma(\cdot)$ is a vector of smooth coefficient functions. Since Z_i are node-specific covariates, we do not need to further assume heterogeneity on $\gamma(\cdot)$. 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) or (2.2) throughout the paper.

2.2 Fundamental assumptions and functional dependence measure

We impose the following assumption on model (2.1) or (2.2).

Assumption 1. (i) Uniformly over $\tau \in [0, 1]$, $\det(\mathbf{I}_N - z\mathbf{B}(\tau)) \neq 0$ for all $|z| \leq 1$.

(ii) For $i = 1, \dots, N$, $\beta_{i,1}(\cdot)$ and $\beta_{i,2}(\cdot)$ are second-order continuously differentiable functions with

$$\max_{1 \leq i \leq N} \sup_{0 \leq \tau \leq 1} \{|\beta'_{i,1}(\tau)| + |\beta'_{i,2}(\tau)| + |\beta''_{i,1}(\tau)| + |\beta''_{i,2}(\tau)|\} \leq c_\beta,$$

where $\beta'_{i,j}(\cdot)$ and $\beta''_{i,j}(\cdot)$ are the first- and second-order derivatives of $\beta_{i,j}(\cdot)$, $j = 1, 2$, and c_β is a positive constant.

(iii) Let $\{\varepsilon_t\}$ be a sequence of i.i.d. random vectors with zero mean, covariance matrix Σ_ε and

$$\max_{1 \leq i \leq N} \mathbf{E} [|\varepsilon_{i,t}|^q] \leq c_\varepsilon,$$

where $q > 8$ and c_ε is a positive constant.

Assumption 1(i) is a natural extension of the stability assumption on linear network VAR, see Theorems 1 and 2 in [Zhu et al \(2017\)](#). This condition ensures that the underlying time series process is locally stable and implies the following local Wold representation:

$$X_t(\tau) = G(\tau, \mathcal{F}_t) := \sum_{m=0}^{\infty} \mathbf{B}_m(\tau) \varepsilon_{t-m} \quad \text{for } 0 < \tau < 1, \quad (2.4)$$

where $\mathbf{B}_0(\tau) = \mathbf{I}_N$, $\mathbf{B}_m(\tau) = \prod_{k=\lfloor T\tau \rfloor - m + 1}^{\lfloor T\tau \rfloor} \mathbf{B}(\tau_k)$ and $\mathcal{F}_t = (\dots, \varepsilon_{t-1}, \varepsilon_t)$. By (2.2) and (2.4), we may write $X_t = X_t(\tau_t)$. The smoothness condition on $\beta_{i,1}(\cdot)$ and $\beta_{i,2}(\cdot)$ in Assumption 1(ii) is common for the local linear estimation method and theory (e.g., [Fan and Gijbels, 1996](#)). We require a relatively strong moment condition ($q > 8$) in Assumption 1(iii) to derive the uniform

convergence of the kernel-weighted quantities, see Lemmas B.1 and B.2 in Appendix B.

We next connect the proposed time-varying network VAR model to the nonlinear functional dependence measure introduced by Wu (2005), which will facilitate the development of our main asymptotic theory. Let $\{\varepsilon_t^*\}$ be an independent copy of $\{\varepsilon_t\}$ and $\mathcal{F}_t^{\{1\}} = (\cdots, \varepsilon_{l-1}, \varepsilon_l^*, \varepsilon_{l+1}, \cdots, \varepsilon_t)$ be a coupled version of \mathcal{F}_t replacing ε_l by ε_l^* . Define $X_t^{\{1\}}(\tau) = G(\tau, \mathcal{F}_t^{\{1\}})$ and $X_t^{\{0\}} = X_t^{\{1\}}(\tau_t)$. As in Xu, Chen and Wu (2020) and Zhang and Wu (2021), we define the node-wise functional dependence measure:

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

where q is defined in Assumption 1(iii), $x_{i,t}(\tau)$ is the i -th element of $X_t(\tau)$, and $x_{i,t}^{\{0\}}(\tau)$ is the i -th element of $X_t^{\{0\}}(\tau)$, and further construct the node-wise dependence adjusted norm:

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

where $\alpha \geq 0$ depicts the decay rate of the cumulative dependence measure $\Delta_{i,m,q}$. Letting $\mathbf{B}_{m,i\bullet}(\tau)$ be the i -th row vector of $\mathbf{B}_m(\tau)$, assume that there exist $\alpha > 1/2 - 2/q$ and $0 < c < \infty$, such that for all $t \geq 0$ and $1 \leq i \leq N$, $\sup_{\tau \in [0,1]} |\mathbf{B}_{t,i\bullet}(\tau)| \leq c(t+1)^{-(\alpha+1)}$. Then, using (2.4) and following the argument of Example 2.2 of Zhang and Wu (2021), we have $\max_{1 \leq i \leq N} \|x_{i\bullet}\|_{q,\alpha} \leq c_{q,\alpha}$ where $c_{q,\alpha}$ is a positive constant depending on q and α .

2.3 The latent group structure

Note that the number of unknown coefficient functions in (2.1) for describing time-varying network and momentum effects is $2N$, which diverges to infinity as $N \rightarrow \infty$ and may be larger than T . Without further restriction on heterogeneity, we can only make use of the sample information from the i -th node and the average impact from its neighbors to nonparametrically estimate the node-specific coefficient functions $\beta_i(\cdot) = [\beta_{i,1}(\cdot), \beta_{i,2}(\cdot)]^\top$ (see Section 3.1), leading to slow estimation convergence rates and unstable numerical performance in particular when T is not sufficiently large. To address this problem, we impose a latent group structure on $\beta_i(\cdot)$, i.e., suppose there exists a partition of the index set $\{1, 2, \cdots, N\}$, denoted by $\mathcal{G} = \{\mathcal{G}_1, \mathcal{G}_2, \cdots, \mathcal{G}_{K_0}\}$, such that

$$\mathcal{G}_i \cap \mathcal{G}_j = \emptyset \text{ for } 1 \leq i \neq j \leq K_0, \text{ and } \beta_i(\cdot) = \alpha_k(\cdot) \text{ for } i \in \mathcal{G}_k, \quad (2.5)$$

where $\alpha_k(\cdot) = [\alpha_{k,1}(\cdot), \alpha_{k,2}(\cdot)]^\top$ denotes a two-dimensional vector of group-specific coefficient functions. The time-varying network and momentum effects have the same functional patterns over nodes within the same group, but vary over different groups. In this paper, neither the group

membership \mathcal{G} nor the number K_0 is known a priori.

With (2.1) and (2.5), we may re-write the main model as

$$x_{i,t} = \alpha_{k,1}(\tau_t) \sum_{j \neq i} \tilde{w}_{ij} x_{j,t-1} + \alpha_{k,2}(\tau_t) x_{i,t-1} + \varepsilon_{i,t}, \quad i \in \mathcal{G}_k, \quad t = 1, \dots, T, \quad (2.6)$$

which is called as the grouped time-varying network VAR model. Let \mathbf{M} be an $N \times K_0$ membership matrix such that it has exactly one element taking the value of 1 in each row and at least one element taking the value of 1 in each column. Let $m_{i,k} \in \{0, 1\}$ be the (i, k) -entry of \mathbf{M} with $m_{i,k} = 1$ if the i -th node belongs to the k -th group, and $M_i = (m_{i,1}, \dots, m_{i,K_0})^\top$. Then, we may further write model (2.6) in the vector/matrix form as in (2.2), letting

$$\mathbf{B}_1(\tau_t) = \text{diag} \{M_1^\top \boldsymbol{\alpha}_{\bullet 1}(\tau_t), \dots, M_N^\top \boldsymbol{\alpha}_{\bullet 1}(\tau_t)\} \quad \text{and} \quad \mathbf{B}_2(\tau_t) = \text{diag} \{M_1^\top \boldsymbol{\alpha}_{\bullet 2}(\tau_t), \dots, M_N^\top \boldsymbol{\alpha}_{\bullet 2}(\tau_t)\},$$

where $\boldsymbol{\alpha}_{\bullet 1}(\cdot) = [\alpha_{1,1}(\cdot), \dots, \alpha_{K_0,1}(\cdot)]^\top$ and $\boldsymbol{\alpha}_{\bullet 2}(\cdot) = [\alpha_{1,2}(\cdot), \dots, \alpha_{K_0,2}(\cdot)]^\top$. It follows from (2.6) that the number of unknown coefficient functions is reduced from $2N$ to $2K_0$. Since K_0 is assumed to be fixed, we achieve dimension reduction and can improve the estimation convergence in the subsequent nonparametric estimation procedure (see Section 4.1) if the latent group structure is consistently estimated.

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 Estimation of \mathcal{G} when K_0 is pre-specified

3.1.1. Point-wise distance estimation between nodes

As there is no prior information on the latent group structure, we start with the fully heterogeneous time-varying network VAR model (2.1). Letting

$$\tilde{\mathbf{X}}_{i,t-1} = \left(\sum_{j \neq i} \tilde{w}_{ij} x_{j,t-1}, x_{i,t-1} \right)^\top,$$

we re-write (2.1) as

$$x_{i,t} = \beta_i^\top(\tau_t) \tilde{\mathbf{X}}_{i,t-1} + \varepsilon_{i,t}, \quad i = 1, \dots, N, \quad t = 1, \dots, T. \quad (3.1)$$

With the smoothness condition in Assumption 1(ii), the node-specific coefficient functions can be estimated by the local linear smoothing (e.g., Fan and Gijbels, 1996), only using the time series sample information from the i -th node and the average impact from its direct neighbors, i.e., $\sum_{j \neq i} \tilde{w}_{ij} x_{j,t-1}$. Letting $\beta'_i(\cdot) = [\beta'_{i,1}(\cdot), \beta'_{i,2}(\cdot)]^\top$, for any given $\tau \in [0, 1]$, we have the following Taylor expansion:

$$\beta_i(\tau_t) \approx \beta_i(\tau) + \beta'_i(\tau)(\tau_t - \tau),$$

when τ_t is in a small neighborhood of τ . Define the node-specific local linear weighted objective function:

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

where $\mathbf{a} = (a_1, a_2)^\top$, $\mathbf{b} = (b_1, b_2)^\top$, $K_h(\cdot) = \frac{1}{h} K(\cdot/h)$, $K(\cdot)$ is a kernel function and h is a bandwidth. Minimizing $\mathcal{L}_i(\mathbf{a}, \mathbf{b})$ with respect to the vectors \mathbf{a} and \mathbf{b} , we obtain the solutions: $\hat{\mathbf{a}} = (\hat{a}_1, \hat{a}_2)^\top$ and $\hat{\mathbf{b}} = (\hat{b}_1, \hat{b}_2)^\top$, and then the local linear estimate of $\beta_i(\tau)$ as

$$\hat{\beta}_i(\tau) = \left[\hat{\beta}_{i,1}(\tau), \hat{\beta}_{i,2}(\tau) \right]^\top = (\hat{a}_1, \hat{a}_2)^\top. \quad (3.3)$$

With the preliminary local linear estimates defined in (3.3), we may compute the point-wise distance between nodes which is defined by

$$\hat{\mathbf{d}}_{ij}(\tau) = \hat{\beta}_i(\tau) - \hat{\beta}_j(\tau) = \begin{bmatrix} \hat{\beta}_{i,1}(\tau) - \hat{\beta}_{j,1}(\tau) \\ \hat{\beta}_{i,2}(\tau) - \hat{\beta}_{j,2}(\tau) \end{bmatrix}.$$

This is a natural estimate of $\mathbf{d}_{ij}(\tau) = \beta_i(\tau) - \beta_j(\tau)$, $1 \leq i \neq j \leq N$. Let $\Delta_i(\tau) = \mathbf{E}[\tilde{\mathbf{X}}_{i,t}(\tau) \tilde{\mathbf{X}}_{i,t}^\top(\tau)]$, where $\tilde{\mathbf{X}}_{i,t}(\tau)$ is defined similarly to $\tilde{\mathbf{X}}_{i,t}$ but with $x_{i,t}$ replaced by $x_{i,t}(\tau)$, the i -th element of $\mathbf{X}_t(\tau)$ defined in (2.4). The following conditions are required to establish the asymptotic property of $\hat{\mathbf{d}}_{ij}(\cdot)$.

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) For each $1 \leq i \leq N$, the matrix $\Delta_i(\tau)$ is positive-definite uniformly over $\tau \in [0, 1]$. In addition, there exist two finite positive constants: $\underline{\lambda}$ and $\bar{\lambda}$, such that

$$0 < \underline{\lambda} \leq \min_{1 \leq i \leq N} \inf_{0 \leq \tau \leq 1} \lambda_{\min}(\Delta_i(\tau)) \leq \max_{1 \leq i \leq N} \sup_{0 \leq \tau \leq 1} \lambda_{\max}(\Delta_i(\tau)) \leq \bar{\lambda} < \infty.$$

(iii) The time series length T , the number of nodes N , and the bandwidth h satisfy $h \rightarrow 0$, $Th \rightarrow \infty$ and

$$\frac{N^2}{T^{\frac{q^2-6q-8}{4(q+2)}} [h \log(N \vee T)]^{q/4}} \rightarrow 0. \quad (3.4)$$

Assumption 2(i) contains some commonly-used conditions imposed on the kernel function (e.g., Fan and Gijbels, 1996). Assumption 2(ii) is crucial to ensure that the kernel-weighted random denominator in local linear estimation is non-singular (e.g., the proof of Proposition 3.1 in Appendix A). Note that $h \rightarrow 0$ and $Th \rightarrow \infty$ in Assumption 2(iii) are regular conditions for kernel-based smoothing, whereas the condition (3.4) 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).

Write $\sigma_{ij} = \mathbf{E}(\varepsilon_{i,t}\varepsilon_{j,t})$, $\sigma_i^2 = \sigma_{ii}$ and

$$\Delta_{ij}(\tau) = \mathbf{E} \left[\tilde{X}_{i,t}(\tau) \tilde{X}_{j,t}^T(\tau) \right], \quad 1 \leq i, j \leq N.$$

We next give the point-wise asymptotic distribution theory for $\hat{d}_{ij}(\tau)$, which is similar to the conventional kernel-based estimation theory of the time-varying coefficient functions (e.g., Cai, 2007; Zhou and Wu, 2010; Vogt, 2012; Zhang and Wu, 2012).

Proposition 3.1. *Suppose that Assumptions 1 and 2 hold. For any given $\tau \in (0, 1)$ and $1 \leq i \neq j \leq N$,*

$$\sqrt{Th} \left[\hat{d}_{ij}(\tau) - d_{ij}(\tau) - \frac{1}{2} h^2 \mu_2 d''_{ij}(\tau) \right] \xrightarrow{d} \mathbf{N}(\mathbf{0}, \Sigma_{ij}(\tau)), \quad (3.5)$$

where $d''_{ij}(\tau) = \beta''_i(\tau) - \beta''_j(\tau)$ with $\beta''_i(\tau) = [\beta''_{i,1}(\tau), \beta''_{i,2}(\tau)]^T$, and

$$\Sigma_{ij}(\tau) = \nu_0 \left[\sigma_i^2 \Delta_i^{-1}(\tau) + \sigma_j^2 \Delta_j^{-1}(\tau) - \Sigma_{ij}^*(\tau) \right]$$

with

$$\Sigma_{ij}^*(\tau) = \sigma_{ij} \left[\Delta_i^{-1}(\tau) \Delta_{ij}(\tau) \Delta_j^{-1}(\tau) + \Delta_j^{-1}(\tau) \Delta_{ji}(\tau) \Delta_i^{-1}(\tau) \right],$$

$\mu_k = \int_{-1}^1 u^k K(u) du$ and $\nu_k = \int_{-1}^1 u^k K^2(u) du$.

We next construct the consistent estimation of the asymptotic variance matrix $\Sigma_{ij}(\tau)$ to facilitate further inference. Let

$$\Sigma_{ij,T}(\tau) = \nu_0 \left[\sigma_i^2 \Delta_{i,T}^{-1}(\tau) + \sigma_j^2 \Delta_{j,T}^{-1}(\tau) - \Sigma_{ij,T}^*(\tau) \right],$$

where $\Sigma_{ij,\Gamma}^*(\tau)$ is defined similarly to $\Sigma_{ij}^*(\tau)$ but with $\Delta_{ij}(\tau)$ and $\Delta_i(\tau)$ replaced by

$$\Delta_{ij,\Gamma}(\tau) = \frac{1}{T} \sum_{t=1}^T \mathbb{E} \left[\tilde{X}_{i,t-1} \tilde{X}_{j,t-1}^\top \right] K_h(\tau_t - \tau) \quad \text{and} \quad \Delta_{i,\Gamma}(\tau) = \Delta_{ii,\Gamma}(\tau),$$

respectively. Following the proof of Proposition 3.1, we may show that $\Sigma_{ij,\Gamma}(\tau) - \Sigma_{ij}(\tau) = o(1)$. Hence, we next only need to consistently estimate $\Sigma_{ij,\Gamma}(\tau)$.

Letting $\hat{\varepsilon}_{i,t} = x_{i,t} - \hat{\beta}_i^\top(\tau_t) \tilde{X}_{i,t-1}$, we may consistently estimate σ_{ij} and σ_i^2 by

$$\hat{\sigma}_{ij} = \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_{i,t} \hat{\varepsilon}_{j,t}, \quad 1 \leq i, j \leq N, \quad \text{and} \quad \hat{\sigma}_i^2 = \hat{\sigma}_{ii},$$

respectively. Define

$$\hat{\Delta}_{ij}(\tau) = \frac{1}{T} \sum_{t=1}^T \tilde{X}_{i,t-1} \tilde{X}_{j,t-1}^\top K_h(\tau_t - \tau), \quad 1 \leq i, j \leq N, \quad \hat{\Delta}_i(\tau) = \hat{\Delta}_{ii}(\tau),$$

and

$$\hat{\Sigma}_{ij}(\tau) = \nu_0 \left[\hat{\sigma}_i^2 \hat{\Delta}_i^{-1}(\tau) + \hat{\sigma}_j^2 \hat{\Delta}_j^{-1}(\tau) - \hat{\Sigma}_{ij}^*(\tau) \right],$$

where

$$\hat{\Sigma}_{ij}^*(\tau) = \hat{\sigma}_{ij} \left[\hat{\Delta}_i^{-1}(\tau) \hat{\Delta}_{ij}(\tau) \hat{\Delta}_j^{-1}(\tau) + \hat{\Delta}_j^{-1}(\tau) \hat{\Delta}_{ji}(\tau) \hat{\Delta}_i^{-1}(\tau) \right].$$

In Appendix A, we prove that $\hat{\Sigma}_{ij}(\tau)$ is a uniformly consistent estimate of $\Sigma_{ij,\Gamma}(\tau)$, see (A.11).

3.1.2. The clustering algorithm with normalized distance matrix

Let τ_l^* , $l = 1, \dots, L$, be a sequence of user-specified equidistant grid points between 0 and 1, where $L \rightarrow \infty$ and $L = O(T)$. Define the following normalized time-varying difference estimate between the i -th and j -th nodes at each of the grid points:

$$\hat{D}_{ij}(\tau_l^*) = \hat{\Sigma}_{ij}^{-1/2}(\tau_l^*) \hat{d}_{ij}(\tau_l^*), \quad l = 1, \dots, L.$$

Then we construct

$$\hat{D}_{ij} = \frac{1}{L} \sum_{l=1}^L \hat{D}_{ij}^\top(\tau_l^*) \hat{D}_{ij}(\tau_l^*) = \frac{1}{L} \sum_{l=1}^L \hat{d}_{ij}^\top(\tau_l^*) \hat{\Sigma}_{ij}^{-1}(\tau_l^*) \hat{d}_{ij}(\tau_l^*). \quad (3.6)$$

Let \hat{D} be an $N \times N$ estimated distance matrix between nodes with the (i, j) -entry being \hat{D}_{ij} . It is clear that the diagonal elements of \hat{D} are zeros. With the distance matrix estimate, we may adopt

the agglomerative clustering algorithm which is commonly used in unsupervised cluster analysis (e.g., [Hastie, Tibshirani and 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 the distance matrix using the unnormalized distance measure, i.e., $\widehat{D}_{ij} = \frac{1}{L} \sum_{l=1}^L \widehat{d}_{ij}^{\top}(\tau_l^*) \widehat{d}_{ij}(\tau_l^*)$, and further estimate the latent group structure in time-varying coefficient panel data models; and [Vogt and 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).

We next describe the clustering algorithm when the true group number K_0 is known a priori. Start with N clusters each of which corresponds to one node, search for the smallest off-diagonal entry in \widehat{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 $\widehat{\mathcal{G}}_k$, $k = 1, \dots, K_0$.

The following theorem states the consistency property.

Theorem 3.2. *Suppose that Assumptions 1 and 2 hold and the group number K_0 is known a priori. In addition,*

$$\underline{D} := \min_{1 \leq k \neq l \leq K_0} \min_{i \in \mathcal{G}_k, j \in \mathcal{G}_l} D_{ij} > 0, \quad h^2 + \sqrt{\frac{\log(N \vee T)}{Th}} + L^{-1} = o(\underline{D}), \quad (3.7)$$

where $D_{ij} = \int_0^1 D_{ij}^{\top}(\tau) D_{ij}(\tau) d\tau$ with $D_{ij}(\tau) = \Sigma_{ij, \tau}^{-1/2}(\tau) d_{ij}(\tau)$, and $\Sigma_{ij}(\tau)$ is positive definite uniformly over i, j and τ . Then we have

$$P \left(\{\widehat{\mathcal{G}}_k, 1 \leq k \leq K_0\} = \{\mathcal{G}_k, 1 \leq k \leq K_0\} \right) \rightarrow 1. \quad (3.8)$$

The consistency property (3.8) is similar to the consistency results of group membership estimation in nonparametric panel/longitudinal data models, see Theorem 3.1 in [Vogt and Linton \(2017\)](#) and Theorem 4.1(a) in [Vogt and Linton \(2020\)](#). The condition (3.7) indicates that the minimum scaled distance between groups may converge to zero at a rate slower than $h^2 + \sqrt{\frac{\log(N \vee T)}{Th}}$ (a typical nonparametric uniform convergence rate) if the grid number L is of order T . In fact, to

prove Theorem 3.2, it suffices to show that

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

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

To further reduce the local linear estimation bias, we may adopt a bias-corrected normalized distance measure between nodes. Let $\widehat{\beta}_i(\tau | h)$ be the local linear estimate of $\beta_i(\tau)$, where we make the dependence on the bandwidth h explicitly. As recommended by Xia (1998) and Zhou and Wu (2010), we may use the jackknife bias-corrected local linear estimates:

$$\widehat{\beta}_i^{\text{bc}}(\tau) = 2\widehat{\beta}_i(\tau | 2^{-1/2}h) - \widehat{\beta}_i(\tau | h)$$

and

$$\widehat{d}_{ij}^{\text{bc}}(\tau) = \widehat{\beta}_i^{\text{bc}}(\tau) - \widehat{\beta}_j^{\text{bc}}(\tau).$$

It is easy to show that the bias-corrected estimation has a smaller bias order h^4 , and implementing the bias correction in local linear estimation is equivalent to the conventional local linear estimation using the fourth-order kernel $K^*(x) = 2\sqrt{2}K(\sqrt{2}x) - K(x)$. The bias-corrected version of the normalized time-varying coefficient difference estimate can be constructed similarly, and Theorem 3.2 continues to hold under the following condition weaker than (3.7):

$$h^4 + \sqrt{\frac{\log(N \vee T)}{Th}} + L^{-1} = o(\underline{D}).$$

3.2 Estimation of the group number K_0

In practice, the true group number K_0 is often unknown and a data-driven criterion is thus required to estimate K_0 . Assuming the group number to be K , we may terminate the clustering algorithm in Section 3.1 when the cluster number reaches K and obtain the estimated clusters denoted by $\widehat{\mathcal{G}}_{k|K}$, $k = 1, \dots, K$. Note that the grouped time-varying network VAR model is either correctly- or over-fitted if $K \geq K_0$, and is under-fitted if $K < K_0$. For the latter scenario, at least two groups would be falsely merged, leading to biased estimation of some group-specific time-varying coefficient functions. When the group number is set as K , we may estimate the group-specific coefficient functions using the estimated groups $\widehat{\mathcal{G}}_{k|K}$, $k = 1, \dots, K$. For each k and given $\tau \in [0, 1]$, we define the local linear objective function by pooling the sample information over $i \in \widehat{\mathcal{G}}_{k|K}$, i.e.,

$$\frac{1}{T|\widehat{\mathcal{G}}_{k|K}|} \sum_{i \in \widehat{\mathcal{G}}_{k|K}} \sum_{t=1}^T \left[x_{i,t} - \mathbf{a}^\top \widetilde{X}_{i,t-1} - \mathbf{b}^\top \widetilde{X}_{i,t-1}(\tau_t - \tau) \right]^2 K_{h_1}(\tau_t - \tau), \quad (3.9)$$

where h_1 is a bandwidth which may be different from h used in the heterogeneous local linear estimation (3.2) and (3.3). Minimizing the pooled objective function in (3.9) with respect to the vectors \mathbf{a} and \mathbf{b} , we obtain the solutions denoted by $\tilde{\mathbf{a}}$ and $\tilde{\mathbf{b}}$, and define the pooled local linear estimates of group-specific coefficient functions (when the group number is assumed to be K):

$$\hat{\alpha}_{k|K}(\tau) = [\hat{\alpha}_{k|K,1}(\tau), \hat{\alpha}_{k|K,2}(\tau)]^\top = \tilde{\mathbf{a}}. \quad (3.10)$$

When $K \geq K_0$, we may show that the local linear estimates defined in (3.10) are uniformly consistent estimates of the true group-specific coefficient functions in (2.6); and when $K < K_0$, as the model is under-fitted, the local linear estimates in (3.10) are biased for at least one k .

We next propose a generalized information criterion to estimate K_0 . Define

$$\text{IC}(K) = \log \hat{\sigma}_{NT}^2(K) + K\rho_{NT}, \quad (3.11)$$

where ρ_{NT} is a user-specified tuning parameter and

$$\hat{\sigma}_{NT}^2(K) = \frac{1}{NT} \sum_{k=1}^K \sum_{i \in \hat{\mathcal{G}}_{k|K}} \sum_{t=1}^T [x_{i,t} - \hat{\alpha}_{k|K}^\top(\tau_t) \tilde{X}_{i,t-1}]^2. \quad (3.12)$$

We estimate the group number by

$$\hat{K} = \arg \min_{1 \leq K \leq \bar{K}} \text{IC}(K), \quad (3.13)$$

where \bar{K} is a pre-specified positive integer larger than K_0 . Similar information criteria are used by [Su, Shi and Phillips \(2016\)](#), [Chen \(2019\)](#) and [Su, Wang and Jin \(2019\)](#) to determine the group number in linear and time-varying coefficient longitudinal data models, see also [Zhu, Xu and Fan \(2022\)](#). We require some further conditions to derive the consistency property of \hat{K} .

Assumption 3. (i) There exist two positive constants $\kappa_1 \in (0, 1)$ and κ_2 such that

$$\min_{1 \leq k \leq K_0} |\mathcal{G}_k| \geq \kappa_1 \cdot N, \quad \min_{1 \leq k_1 \neq k_2 \leq K_0} \int_0^1 |\alpha_{k_1}(\tau) - \alpha_{k_2}(\tau)|^2 d\tau > \kappa_2.$$

(ii) The bandwidth h_1 satisfies $h_1 \propto h$ and $\log(N \vee T) = O(T h_1^3)$, and the tuning parameter ρ_{NT} satisfies $\rho_{NT} \rightarrow 0$ and $\frac{1}{NT h_1} + h_1^4 = o(\rho_{NT})$.

(iii) For any index set $\mathcal{G} \subset \mathcal{G}_k, k = 1, \dots, K_0$,

$$\max_{i \in \mathcal{G}} \sum_{j \in \mathcal{G}} |\sigma_{ij}| \leq \bar{c}, \quad \max_{1 \leq t \leq T} \mathbb{E} \left[\left| \sum_{i \in \mathcal{G}} \varepsilon_{i,t} \tilde{X}_{i,t-1} \right|^4 \right] \leq \bar{c} |\mathcal{G}|^2, \quad (3.14)$$

where $\sigma_{ij} = \mathbf{E}(\varepsilon_{i,t}\varepsilon_{j,t})$ and \bar{c} is a positive constant independent of the choice of \mathcal{G} .

Assumption 3(i) is crucial to derive a lower bound of $\hat{\sigma}_{NT}^2(K)$ when $K < K_0$. The condition $\min_{1 \leq k \leq K_0} |\mathcal{G}_k| \geq \kappa_1 \cdot N$ indicates that the cardinality of groups \mathcal{G}_k is of the same order. By Assumption 3(ii), we may set $\rho_{NT} = \frac{\log(N \vee T)}{NTh_1}$, in which case the proposed information criterion is similar to Schwarz (1978)'s Bayesian information criterion by treating NTh_1 as the "effective" sample size in the kernel-based estimation. Assumption 3(iii) allows weak correlation between nodes and is sufficient to verify the high-level conditions such as Assumption 3.6 in Chen (2019). In fact, the first condition in (3.14) is slightly weaker than $\max_{1 \leq i \leq N} \sum_{j=1}^N |\sigma_{i,j}| \leq \bar{c}$, a commonly-used sparsity restriction for large covariance matrix estimation. The second condition in (3.14) can be easily verified when $\varepsilon_{i,t}$ is independent over i (e.g., Zhu et al, 2017).

Theorem 3.3. *Suppose that Assumptions 1–3 are satisfied. Then we have*

$$\mathbf{P}(\hat{K} = K_0) \rightarrow 1. \quad (3.15)$$

Finally, we use the estimated group number \hat{K} and terminate the clustering algorithm in Section 3.1 when the cluster number reaches \hat{K} . In order to avoid unnecessary notational burden, we still denote the estimated groups as $\hat{\mathcal{G}}_k, k = 1, \dots, \hat{K}$.

4 Post-grouping estimation and inference

In this section we make use of the consistently estimated group structure to more efficiently estimate the group-specific time-varying network and momentum effects in (2.6) and then conduct statistical inference.

4.1 Post-grouping local linear estimation

Note that the heterogeneous local linear estimates defined in (3.3) only use the sample information from one node and its direct neighbors, resulting in rather slow convergence rates (see Proposition 3.1 and Lemma B.3) and unstable numerical performance if T is not sufficiently large in finite samples. It follows from Theorems 3.2 and 3.3 that, for any $k = 1, \dots, K_0$, there exists $1 \leq k_\circ \leq \hat{K}$ such that $\mathcal{G}_k = \hat{\mathcal{G}}_{k_\circ}$ w.p.a.1. Hence, without loss of generality, we may consider $\mathcal{G}_k = \hat{\mathcal{G}}_k$ (conditioning on $\hat{K} = K_0$) throughout this section. For each k and given $\tau \in (0, 1)$, similarly to (3.9),

we define the post-grouping local linear objective function:

$$\frac{1}{T|\mathcal{G}_k|} \sum_{i \in \mathcal{G}_k} \sum_{t=1}^T \left[x_{i,t} - \mathbf{a}^\top \tilde{\mathbf{X}}_{i,t-1} - \mathbf{b}^\top \tilde{\mathbf{X}}_{i,t-1}(\tau_t - \tau) \right]^2 K_{h_2}(\tau_t - \tau), \quad (4.1)$$

where h_2 is a bandwidth which may be different from h used in the heterogeneous local linear estimation (3.2) and (3.3). Minimizing the post-grouping objective function with respect to the vectors \mathbf{a} and \mathbf{b} , we obtain the solutions denoted by $\check{\mathbf{a}}$ and $\check{\mathbf{b}}$, and define the post-grouping local linear estimates as

$$\hat{\alpha}_k(\tau) = [\hat{\alpha}_{k,1}(\tau), \hat{\alpha}_{k,2}(\tau)]^\top = \check{\mathbf{a}}. \quad (4.2)$$

Theorem 4.1. *Suppose that Assumptions 1–3 are satisfied, $h_2 \rightarrow 0$, $NTh_2 \rightarrow \infty$ and there exists a positive definite matrix $\Upsilon_{\mathcal{G}_k}(\tau)$ such that*

$$\frac{1}{|\mathcal{G}_k|} \sum_{i,j \in \mathcal{G}_k} \sigma_{ij} \Delta_{ij}(\tau) \rightarrow \Upsilon_{\mathcal{G}_k}(\tau) \quad (4.3)$$

as $|\mathcal{G}_k| \rightarrow \infty$, where σ_{ij} and $\Delta_{ij}(\tau)$ are defined in Section 3.1.1. For any $\tau \in (0, 1)$,

$$\sqrt{|\mathcal{G}_k|Th_2} \left[\hat{\alpha}_k(\tau) - \alpha_k(\tau) - \frac{1}{2} h_2^2 \mu_2 \alpha_k''(\tau) \right] \xrightarrow{d} \mathbf{N}(\mathbf{0}, \mathbf{\Omega}_{\mathcal{G}_k}(\tau)) \quad (4.4)$$

where

$$\mathbf{\Omega}_{\mathcal{G}_k}(\tau) = \nu_0 \mathbf{\Delta}_{\mathcal{G}_k}^{-1}(\tau) \Upsilon_{\mathcal{G}_k}(\tau) \mathbf{\Delta}_{\mathcal{G}_k}^{-1}(\tau)$$

with

$$\mathbf{\Delta}_{\mathcal{G}_k}(\tau) = \frac{1}{|\mathcal{G}_k|} \sum_{i \in \mathcal{G}_k} \mathbf{\Delta}_i(\tau) = \frac{1}{|\mathcal{G}_k|} \sum_{i \in \mathcal{G}_k} \mathbf{E} \left[\tilde{\mathbf{X}}_{i,t}(\tau) \tilde{\mathbf{X}}_{i,t}^\top(\tau) \right].$$

Since $|\mathcal{G}_k|$ is of order N , Theorem 4.1 shows that the post-grouping local linear estimation $\hat{\alpha}_k(\tau)$ has the point-wise convergence rate of $O_p(1/\sqrt{NTh_2} + h_2^2)$, which is substantially faster than that for the preliminary local linear estimates (see Lemma B.3). This is unsurprising since more sample information is used in the post-grouping estimation procedure.

4.2 Post-grouping specification testing

In practice, one is often interested in testing the hypothesis that some group-specific time-varying coefficient functions belong to a given parametric function family, i.e., for $k = 1, \dots, K_0$,

$$H_0 : \alpha_k(\tau) = g_k(\tau; \theta_k), \quad \tau \in [0, 1], \quad (4.5)$$

where θ_k is a d_k -dimensional vector of unknown parameters which lies in the interior of a compact set Θ , and $g_k(\cdot; \cdot)$ is a pre-specified function. A typical example of (4.5) is to test whether the time-varying coefficient functions are constants, i.e., for $k = 1, \dots, K_0$,

$$H_0 : \alpha_k(\tau) = \alpha_k, \quad \tau \in [0, 1], \quad (4.6)$$

where α_k is a constant vector. Under the null hypothesis in (4.6), model (2.6) reduces to the grouped parametric network VAR model (e.g., [Zhu and Pan, 2020](#); [Chen, Fan and Zhu, 2022](#); [Zhu, Xu and Fan, 2022](#)). The statistical inference of the time-varying coefficient functions or trending functions have been extensively studied for either univariate or fixed dimensional time series (e.g., [Wu and Zhao, 2007](#); [Li, Chen and Lin, 2011](#); [Zhang and Wu, 2011, 2012](#); [Degras et al, 2012](#)). [Chen and Wu \(2019\)](#) consider testing the trend functions in high-dimensional time series. We next make a further extension by developing a testing tool under a more general network VAR model framework.

Without loss of generality, we assume that $\mathcal{G}_k = \widehat{\mathcal{G}}_k$ (conditioning on $\widehat{K} = K_0$) as in the post-grouping local linear estimation. For $i \in \widehat{\mathcal{G}}_k$, we let $\tilde{\varepsilon}_{i,t} = x_{i,t} - g_k^\top(\tau_t; \tilde{\theta}_k) \tilde{X}_{i,t-1}$, where $\tilde{\theta}_k$ is the parametric estimate of θ_k . Partly motivated by [Chen and Wu \(2019\)](#), we construct the test statistic via the following kernel weighted quadratic form:

$$\tilde{Q}_T(k) = \sum_{t=1}^T \sum_{s \neq t} \mathbb{K} \left(\frac{\tau_t - \tau_s}{h_3} \right) \tilde{\varepsilon}_t^\top(k) \tilde{\varepsilon}_s(k), \quad (4.7)$$

where $\tilde{\varepsilon}_t(k) = (\tilde{\varepsilon}_{i,t}, i \in \widehat{\mathcal{G}}_k)$ and h_3 is a bandwidth. Here we simply use the kernel weight rather than a more complex local linear weight (e.g., [Chen and Wu, 2019](#)), avoiding possible random denominators in the test statistic. Similar test statistics have been proposed in the model specification testing of univariate or multiple time series (e.g., [Li and Wang, 1998](#); [Li, 1999](#); [Gao and Gijbels, 2008](#); [Gao et al, 2009](#)) and high-dimensional time series (e.g., [Bai and Saranadasa, 1996](#); [Chen and Wu, 2019](#)).

Assumption 4. (i) *There exist two positive constants c_g^\dagger and c_g^\ddagger such that*

$$\max_{1 \leq k \leq K_0} \sup_{\tau \in [0,1]} |\partial_\theta g_k(\tau; \theta_k)|_2 < c_g^\dagger,$$

and

$$\max_{1 \leq k \leq K_0} \sup_{\tau \in [0,1]} |g_k(\tau; \theta) - g_k(\tau; \theta_k) - \partial_\theta g_k(\tau; \theta_k)(\theta - \theta_k)|_\infty \leq c_g^\ddagger |\theta - \theta_k|^2.$$

(ii) Letting $\delta_\theta = |\tilde{\theta}_k - \theta_k|$ and $\Sigma_\varepsilon(k) = \mathbf{E} [\varepsilon_t(k)\varepsilon_t^\top(k)]$ with $\varepsilon_t(k) = (\varepsilon_{i,t}, i \in \mathcal{G}_k)$, assume

$$\frac{\delta_\theta^2 \mathbf{N}\mathbf{T}}{\|\Sigma_\varepsilon(k)\|} \left(h_3^{1/2} + \|\Sigma_\varepsilon(k)\|^{-1} \right) \rightarrow 0.$$

(iii) Let $h_3 \rightarrow 0$ and $\|\Sigma_\varepsilon(k)\|\mathbf{T}h_3^{1/2} \rightarrow \infty$.

The smoothness conditions on $g_k(\tau; \cdot)$ in Assumption 4(i) are not uncommon in the literature (e.g., Gao and Gijbels, 2008; Chen and Wu, 2019). If δ_θ has the root-NT rate, $\|\Sigma_\varepsilon(k)\|$ is of the order $\mathbf{N}^{1/2}$ (which is satisfied by imposing a sparsity restriction), $h_3 \rightarrow 0$ and $\mathbf{T}h_3 \rightarrow \infty$, it is easy to verify Assumption 4(ii)(iii).

Theorem 4.2. *Suppose that Assumptions 1–4 hold. Under H_0 ,*

$$\frac{1}{(2\nu_0)^{1/2}\|\Sigma_\varepsilon(k)\|\mathbf{T}h_3^{1/2}} \tilde{Q}_\mathbf{T}(k) \xrightarrow{d} \mathbf{N}(0, 1). \quad (4.8)$$

The normalization rate in Theorem 4.2 depends on \mathbf{T} , h_3 and \mathbf{N} (implicitly via $\|\Sigma_\varepsilon(k)\|$). In particular, if $\Sigma_\varepsilon(k)$ is sparse (see Assumption 3(iii)), we may show that $\|\Sigma_\varepsilon(k)\| \propto \mathbf{N}^{1/2}$ and the normalization rate in (4.8) is $\mathbf{N}^{1/2}\mathbf{T}h_3^{1/2}$.

We next study the asymptotic power of the test under the alternative hypothesis. Denote

$$\theta_k^\diamond = \arg \min_{\theta_k \in \Theta} \frac{1}{|\mathcal{G}_k|} \sum_{i \in \mathcal{G}_k} \int_0^1 \left\{ [\alpha_k(\tau) - g_k(\tau; \theta_k)]^\top \tilde{X}_{i,t}(\tau) \right\}^2 d\tau,$$

where Θ is a compact set. Consider the alternative hypothesis as

$$H_A : \alpha_k(\tau) = g_k(\tau; \theta_k^\diamond) + \varphi g_k(\tau), \quad (4.9)$$

where $\varphi > 0$ but may tend to zero at an appropriate rate, and $g_k(\cdot)$ is continuous over $[0, 1]$. The following conditions are required to derive the asymptotic property of the test under H_A .

Assumption 5. (i) *There exist two positive constants $c_{g,\diamond}^\dagger$ and $c_{g,\diamond}^\ddagger$ such that*

$$\max_{1 \leq k \leq K_0} \sup_{\tau \in [0,1]} |\partial_\theta g_k(\tau; \theta_k^\diamond)|_2 \leq c_{g,\diamond}^\dagger$$

and

$$\max_{1 \leq k \leq K_0} \sup_{\tau \in [0,1]} |g_k(\tau; \theta) - g_k(\tau; \theta_k^\diamond) - \partial_\theta g_k(\tau; \theta_k^\diamond)(\theta - \theta_k^\diamond)|_\infty \leq c_{g,\diamond}^\ddagger |\theta - \theta_k^\diamond|^2.$$

(ii) Assumption 4(ii) when δ_θ is replaced by $\delta_\theta^\diamond = |\tilde{\theta}_k - \theta_k^\diamond|$. There exists ϕ_k such that

$$\frac{1}{|\mathcal{G}_k|T^2h_3} \sum_{t=1}^T \sum_{s \neq t} \mathbb{K} \left(\frac{\tau_t - \tau_s}{h_3} \right) \omega_{\mathcal{G}_k,t}^\top(g) \omega_{\mathcal{G}_k,s}(g) \xrightarrow{P} \phi_k, \quad (4.10)$$

where $\omega_{\mathcal{G}_k,t}(g) = \left\{ g_k^\top(\tau_t) \tilde{\chi}_{i,t-1}, i \in \mathcal{G}_k \right\}$.

(iii) Let $\delta_\theta^\diamond = o(\varphi)$ and $\varphi^2 N T h_3 \rightarrow \infty$.

The following theorem establishes the asymptotic distribution of $\tilde{Q}_T(k)$ under the local alternative hypothesis H_A with φ possibly converging to zero.

Theorem 4.3. Suppose that Assumptions 1–3, 4(iii) and 5 hold. Under H_A , (i) if $\frac{\varphi^2 \phi_k |\mathcal{G}_k| T h_3^{1/2}}{(2\nu_0)^{1/2} \|\Sigma_\varepsilon(k)\|} \rightarrow \kappa$,

$$\frac{1}{(2\nu_0)^{1/2} \|\Sigma_\varepsilon(k)\| T h_3^{1/2}} \tilde{Q}_T(k) \xrightarrow{d} \mathbf{N}(\kappa, 1); \quad (4.11)$$

and (ii) if $\frac{\varphi^2 \phi_k |\mathcal{G}_k| T h_3^{1/2}}{\|\Sigma_\varepsilon(k)\|} \rightarrow \infty$,

$$\frac{1}{(2\nu_0)^{1/2} \|\Sigma_\varepsilon(k)\| T h_3^{1/2}} \tilde{Q}_T(k) \xrightarrow{P} \infty. \quad (4.12)$$

The above asymptotic distribution theory under the local alternative is comparable to Proposition 2 in Chen and Wu (2019). Let

$$Q_T(k) = \sum_{t=1}^T \sum_{s \neq t} \mathbb{K} \left(\frac{\tau_t - \tau_s}{h_3} \right) \varepsilon_t^\top(k) \varepsilon_s(k).$$

In fact, following the proofs of Theorems 4.2 and 4.3 in Appendix A, we may show that

$$\frac{1}{(2\nu_0)^{1/2} \|\Sigma_\varepsilon(k)\| T h_3^{1/2}} \tilde{Q}_T(k) = \frac{1}{(2\nu_0)^{1/2} \|\Sigma_\varepsilon(k)\| T h_3^{1/2}} [Q_T(k) + \varphi^2 \phi_k \cdot |\mathcal{G}_k| T^2 h_3 (1 + o_P(1))]$$

under H_A , where

$$\frac{1}{(2\nu_0)^{1/2} \|\Sigma_\varepsilon(k)\| T h_3^{1/2}} Q_T(k) \xrightarrow{d} \mathbf{N}(0, 1).$$

Hence, Theorem 4.2 can be viewed as a special case of Theorem 4.3(i). Theorem 4.3(ii) indicates that the null hypothesis is rejected *w.p.a.1* when $\frac{\varphi^2 \phi_k |\mathcal{G}_k| T h_3^{1/2}}{\|\Sigma_\varepsilon(k)\|} \rightarrow \infty$.

5 Numerical studies

In this section, we conduct both the simulation and empirical studies. Section 5.1 assesses the finite-sample performance of the developed methodology and verifies the main theoretical properties via the Monte-Carlo simulation, whereas Section 5.2 applies the model and methodology to the export flow data collected in 90 counties. Throughout this section, we use the Epanechnikov kernel as $K(\cdot)$ in the estimation and testing so that Assumption 2(i) is satisfied.

5.1 Simulation

We consider the following data generating process (DGP):

$$x_{i,t} = \beta_{i,1}(\tau_t) \sum_{j \neq i} \tilde{w}_{ij} x_{j,t-1} + \beta_{i,2}(\tau_t) x_{i,t-1} + \varepsilon_{i,t}, \quad i = 1, \dots, N, \quad t = 1, \dots, T, \quad (5.1)$$

in which $\beta_i(\cdot) = [\beta_{i,1}(\cdot), \beta_{i,2}(\cdot)]^\top$ satisfies the following latent group structure:

$$\begin{aligned} \beta_i(\tau) = \alpha_1(\tau) &= [\alpha_{1,1}(\tau), \alpha_{1,2}(\tau)]^\top = [-0.95 \sin(\pi\tau), -0.6 \cos(\pi\tau) - 0.3]^\top, \quad i \in \mathcal{G}_1, \\ \beta_i(\tau) = \alpha_2(\tau) &= [\alpha_{2,1}(\tau), \alpha_{2,2}(\tau)]^\top = [-0.7 \sin(\pi\tau) + 0.8, -0.5 \cos(\pi\tau) + 0.45]^\top, \quad i \in \mathcal{G}_2, \\ \beta_i(\tau) = \alpha_3(\tau) &= [\alpha_{3,1}(\tau), \alpha_{3,2}(\tau)]^\top = [\sin(\pi\tau) - 0.2, 0.9 \cos(\pi\tau)]^\top, \quad i \in \mathcal{G}_3. \end{aligned}$$

Clearly the number of groups is $K_0 = 3$. The group membership $\mathcal{G} = \{\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3\}$ is generated by either of the following two partitions.

- Random group structure: Randomly assign node i to one of the three groups (with equal probability) in each simulation replication.
- Fixed group structure: Assign node i to the three groups such that $\mathbf{P}(i \in \mathcal{G}_1) = 0.25$, $\mathbf{P}(i \in \mathcal{G}_2) = 0.25$ and $\mathbf{P}(i \in \mathcal{G}_3) = 0.5$, which is only done once and then fixed over replications.

Let $\tilde{w}_{ij} = w_{ij}/n_i$ with $n_i = \sum_{j \neq i} w_{ij}$, where $\mathbf{P}(w_{ij} = 1) = 0.1$ and $\mathbf{P}(w_{ij} = 0) = 0.9$. The errors $\varepsilon_t = (\varepsilon_{1,t}, \dots, \varepsilon_{N,t})^\top$ are independently generated from $\mathbf{N}(\mathbf{0}, \Sigma_\varepsilon)$, where $\Sigma_\varepsilon = \{\sigma_{ij}\}_{N \times N}$ with $\sigma_{ij} = 0.3^{|i-j|}$.

5.1.1. Latent group estimation

We first examine the numerical performance of the clustering algorithm and information criterion in finite samples. Consider $N = 100, 200$, $T = 200, 300, 400$, and the replication number $R = 400$. In the local linear estimation procedure, there are three bandwidths involved: h in the preliminary

estimation, and h_1 and h_2 in the pooled and post-grouping estimation. To reduce the computational burden, we adopt the rule of thumb, setting the bandwidths as $h = 2.34\sigma T^{-1/5}$ and $h_1 = h_2 = 2.34\sigma(NT)^{-1/5}$, where the constant 2.34 is due to use of the Epanechnikov kernel in local linear smoothing and $\sigma = 1/\sqrt{12}$ is due to the fact that the scaled time point sequence $\{\tau_t\}$ behaves like the observations following a uniform distribution on $[0, 1]$. The tuning parameter in the information criterion is set as $\rho_{NT} = \frac{\log(N\sqrt{T})}{NT h_1}$ as discussed in Section 3.2.

To assess the performance of the information of the proposed information criterion, we compute the empirical probabilities of correct-, over- and under-estimation of the group number, i.e.,

$$\text{EP}_{\hat{K}=K_0} = \frac{1}{R} \sum_{r=1}^R I\{\hat{K}_r = K_0\}, \quad \text{EP}_{\hat{K}>K_0} = \frac{1}{R} \sum_{r=1}^R I\{\hat{K}_r > K_0\}, \quad \text{EP}_{\hat{K}<K_0} = \frac{1}{R} \sum_{r=1}^R I\{\hat{K}_r < K_0\},$$

where \hat{K}_r is the group number estimate in the r -th replication, and $I\{\cdot\}$ denotes the indicator function. We use the so-called purity measurement to examine the accuracy of the group membership estimate, i.e.,

$$\text{Purity} = \frac{1}{R} \sum_{r=1}^R P_r, \quad P_r = \frac{1}{N} \sum_{k=1}^{\hat{K}_r} \max_{1 \leq j \leq K_0} |\hat{\mathcal{G}}_{k,r} \cap \mathcal{G}_j|,$$

where $\hat{\mathcal{G}}_{k,r}$ denotes the group member estimate in the r -th replication.

Table 1: Estimation of the group number and membership

		Random groups			Fixed groups		
N \ T		200	300	400	200	300	400
EP $_{\hat{K}=K_0}$	100	1.000	0.985	0.950	1.000	0.995	1.000
	200	0.725	0.995	1.000	0.815	0.995	1.000
EP $_{\hat{K}>K_0}$	100	0.000	0.015	0.050	0.000	0.005	0.000
	200	0.260	0.005	0.000	0.160	0.005	0.000
EP $_{\hat{K}<K_0}$	100	0.000	0.000	0.000	0.000	0.000	0.000
	200	0.015	0.000	0.000	0.025	0.000	0.000
Purity	100	1.000	1.000	0.996	1.000	1.000	1.000
	200	0.986	1.000	1.000	0.979	1.000	1.000

Table 1 presents the empirical probabilities and purity values over 400 replications. In general, the information criterion can accurately estimate the group number as long as the time series length is moderately large. The performance of the group number estimation improves, i.e., $\text{EP}_{\hat{K}=K_0}$ increases whereas $\text{EP}_{\hat{K}>K_0}$ and $\text{EP}_{\hat{K}<K_0}$ generally decreases, as T increases, but slightly deteriorates

as N increases from 100 to 200. When $N = 200$ and $T = 200$ or 300, if $\widehat{K} \neq K_0$, the information criterion tends to over-estimate the group number, which is not as harmful as under-estimation of the group number. Note that the closer the values of purity are to 1, the more accurate the estimated groups are to the true ones. The obtained purity values are all larger than 0.975, indicating that only a small proportion of nodes are misclassified and verifying the consistency property of the clustering algorithm in Theorem 3.2. The estimation performance for the random group structure is similar to that for the fixed group structure.

5.1.2. Time-varying coefficient estimation

We next compare the numerical performance between the preliminary heterogeneous local linear estimation and post-grouping estimation and show the advantage of making use of the group structure to achieve dimension reduction and improve estimation efficiency. Define

$$\text{RMSE}_r = \sqrt{\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \left\| \widehat{\beta}_i^r(\tau_t) - \beta_i^r(\tau_t) \right\|^2},$$

for each replication, where $\widehat{\beta}_i^r(\cdot)$ denotes the pre-grouping estimate of $\beta_i(\cdot)$. We further define RMSE_{pre} as the average of RMSE_r over 400 replications. Let $\text{RMSE}_{\text{post}}$ and RMSE_{ora} be defined similarly using the post-grouping local linear estimation, and using the oracle local linear estimation with the true group structure, respectively. The computed RMSE values are reported in Table 2, where numbers in the parentheses are the standard errors. As we can see from Table 2, the $\text{RMSE}_{\text{post}}$ values are much smaller than the RMSE_{pre} values, confirming the convergence improvement by the post-grouping estimation and verifying Theorem 4.1 and the subsequent discussion. In particular, the improvement becomes more significant when N increases from 100 to 200. Meanwhile, the performance of the post-grouping estimation is similar to that of the oracle estimation which is infeasible in practice, and the difference between $\text{RMSE}_{\text{post}}$ and RMSE_{ora} is nearly invisible when $T = 300$ or 400.

5.1.3. Post-grouping inference

We next assess the size and power performance of the proposed kernel-weighted test statistic in finite samples. The DGP is similar to that used for estimation assessment, i.e., consider (5.1) but with the group-specific coefficients defined as

$$\begin{aligned} \alpha_1(\tau) &= [-0.7, -0.6]^\top, \\ \alpha_2(\tau) &= [\alpha_{1,1}(\tau), \alpha_{1,2}(\tau)]^\top = [-0.7 \sin(\pi\tau) + 0.8, -0.5 \cos(\pi\tau) + 0.45]^\top, \\ \alpha_3(\tau) &= [\alpha_{3,1}(\tau), \alpha_{3,2}(\tau)]^\top = [\sin(\pi\tau) - 0.2, 0.9 \cos(\pi\tau)]^\top. \end{aligned}$$

Table 2: Estimation performance of the time-varying coefficients

		Random groups			Fixed groups		
N \ T		200	300	400	200	300	400
RMSE _{pre}	100	0.061 (0.058)	0.050 (0.031)	0.048 (0.052)	0.066 (0.006)	0.052 (0.004)	0.030 (0.002)
	200	0.116 (0.009)	0.074 (0.005)	0.055 (0.003)	0.123 (0.010)	0.076 (0.005)	0.055 (0.003)
RMSE _{post}	100	0.048 (0.007)	0.041 (0.011)	0.043 (0.044)	0.049 (0.007)	0.040 (0.006)	0.036 (0.005)
	200	0.068 (0.058)	0.043 (0.007)	0.039 (0.006)	0.064 (0.005)	0.043 (0.006)	0.038 (0.005)
RMSE _{ora}	100	0.048 (0.007)	0.040 (0.006)	0.038 (0.007)	0.049 (0.007)	0.040 (0.006)	0.036 (0.005)
	200	0.052 (0.008)	0.043 (0.006)	0.039 (0.006)	0.051 (0.007)	0.043 (0.006)	0.038 (0.005)

We aim to test the null hypothesis $H_0 : \alpha_k(\tau) = \alpha_k, \tau \in [0, 1]$ for $k = 1, 2, 3$, where α_k is a bivariate vector of constants. It is clear that H_0 is only valid for $k = 1$. Hence, the network time series sample in \mathcal{G}_1 is used to assess the finite-sample size performance, whereas those in \mathcal{G}_2 and \mathcal{G}_3 are used to assess the power performance.

Since the optimal bandwidth is of order $T^{-2/5}$ under the null (e.g., Casas *et al*, 2021), we adopt the rule of thumb for each group \mathcal{G}_k by setting $h_3 = 2.34\tilde{\sigma}_{\varepsilon,k}T^{-2/5}$, where $\tilde{\sigma}_{\varepsilon,k}^2 = \frac{1}{|\mathcal{G}_k|(T-1)} \sum_{t=2}^T \|\tilde{\varepsilon}_t(k)\|^2$. We use the asymptotic null distribution in Theorem 4.2 to determine the test critical value. Conditioning on $\hat{K} = K_0$, we compute the test rejection rates (at the significance level α), which is defined as

$$\text{TRR}_{\hat{K}=K_0} = \frac{1}{R_*} \sum_{r=1}^{R_*} \mathbb{I} \left(|\hat{Q}_T^r(k)| > z_{1-\alpha/2} \right), \quad \hat{Q}_T^r(k) = \frac{1}{\|\tilde{\Sigma}_\varepsilon^r(k)\| (2\nu_0 T^2 h_3)^{1/2}} \tilde{Q}_T^r(k)$$

where R_* denotes the number of replications when $\hat{K} = K_0$, z_α is the α -quantile of $N(0, 1)$ and $\tilde{\Sigma}_\varepsilon^r(k)$ is the sample version of $\Sigma_\varepsilon(k)$ in the r -th replication (e.g., Chen and Wu, 2019). We also consider the oracle kernel-weighted test directly using the true group structure and compute

$$\text{TRR}_{\text{ora}} = \frac{1}{R} \sum_{r=1}^R \mathbb{I} \left(|\hat{Q}_{\text{ora}}^r(k)| > z_{1-\alpha/2} \right),$$

where $\hat{Q}_{\text{ora}}^r(k)$ is defined similarly to $\hat{Q}_T^r(k)$, using the latent group structure when constructing $\tilde{Q}_T^r(k)$ and $\tilde{\Sigma}_\varepsilon^r(k)$. The finite-sample test results are summarized in Table 3, where we consider $\alpha = 1\%, 5\%$ and 10% . For \mathcal{G}_1 , both $\text{TRR}_{\hat{K}=K_0}$ and TRR_{ora} are generally close to the nominal rates, indicating that it is reasonable to choose the test critical value via the asymptotic null distribution.

The test results for \mathcal{G}_2 and \mathcal{G}_3 show that the proposed test statistic has reliable power performance. In general, the $\text{TRR}_{\hat{\kappa}=\kappa_0}$ values are very close to the TRR_{ora} values.

Table 3: Rejection rates of the post-grouping tests under the levels 1%, 5% and 10%

		$\alpha = 1\%$			$\alpha = 5\%$			$\alpha = 10\%$		
N/T		200	300	400	200	300	400	200	300	400
$\text{TRR}_{\hat{\kappa}=\kappa_0}$ — Random groups										
Group1	100	0.010	0.013	0.014	0.040	0.043	0.060	0.105	0.083	0.090
	200	0.016	0.008	0.008	0.056	0.045	0.053	0.107	0.088	0.105
Group2	100	0.341	0.806	0.932	0.466	0.856	0.943	0.554	0.881	0.957
	200	0.273	0.499	0.835	0.365	0.579	0.868	0.429	0.642	0.893
Group3	100	1.000	0.843	0.546	1.000	0.874	0.709	1.000	0.907	0.785
	200	1.000	0.995	0.895	1.000	0.995	0.933	1.000	0.997	0.950
TRR_{ora} — Random groups										
Group1	100	0.010	0.013	0.013	0.040	0.043	0.060	0.105	0.083	0.093
	200	0.018	0.007	0.008	0.058	0.045	0.053	0.105	0.090	0.105
Group2	100	0.343	0.808	0.938	0.468	0.858	0.948	0.555	0.883	0.960
	200	0.283	0.498	0.835	0.373	0.578	0.868	0.435	0.640	0.893
Group3	100	1.000	0.840	0.528	1.000	0.873	0.695	1.000	0.905	0.778
	200	1.000	0.995	0.895	1.000	0.995	0.933	1.000	0.998	0.950
$\text{TRR}_{\hat{\kappa}=\kappa_0}$ — Fixed groups										
Group1	100	0.025	0.020	0.008	0.068	0.048	0.063	0.111	0.083	0.106
	200	0.011	0.008	0.010	0.059	0.055	0.040	0.111	0.113	0.088
Group2	100	0.158	0.256	0.808	0.296	0.421	0.856	0.392	0.524	0.881
	200	0.376	0.649	0.735	0.469	0.711	0.803	0.539	0.743	0.830
Group3	100	1.000	1.000	0.924	1.000	1.000	0.952	1.000	1.000	0.957
	200	1.000	1.000	0.998	1.000	1.000	0.998	1.000	1.000	0.998
TRR_{ora} — Fixed groups										
Group1	100	0.025	0.020	0.008	0.068	0.048	0.055	0.110	0.083	0.108
	200	0.010	0.008	0.010	0.060	0.055	0.040	0.103	0.113	0.088
Group2	100	0.160	0.255	0.808	0.298	0.420	0.855	0.395	0.523	0.880
	200	0.375	0.665	0.735	0.468	0.725	0.803	0.538	0.753	0.830
Group3	100	1.000	1.000	0.925	1.000	1.000	0.953	1.000	1.000	0.958
	200	1.000	1.000	1.000	1.000	1.000	0.998	1.000	1.000	0.998

5.2 An Empirical Study

Most countries are now connected to each other through varieties of bilateral trade. In order to provide a detailed insight into the most recent trading patterns, it is crucial to obtain reliable and up-to-date statistics on trade flows between different countries and explore possible dynamic network structures. There has been increasing interest in understanding such a complex network and developing different network models over the past decade (e.g., Moench, Ng and Potter, 2013; Beck, Hubrich and Marcellino, 2016). It is noteworthy that trade flows between a dyad represented by an origin and a destination country do not solely depend on these two countries directly involved, but also on how this relates to the opportunities of moving to other destinations. Hence, it is sensible to apply the proposed network time series model (2.1) to the trade flow data and further explore smooth structural changes and the latent group structure on time-varying coefficients.

Let $x_{i,t}$ denote the growth rates of bilateral trade flows divided by GDP per capita (e.g., Rodrik, 2013) of 90 countries, i.e., $N = 90$. The export flow data are collected from the Direction of Trade Statistics (DOTS) of International Monetary Fund (IMF) available at <https://www.imf.org/external/index.htm>. We use the FOB (free on board) value of exports of goods (denominated in U.S. dollars) and restrict the sample over a 324-month period between January 1995 and December 2021, i.e., $T = 324$. The data of GDP per capita are available at <https://www.worldbank.org/en/home>. In the preliminary analysis, we standardize each time series observation to ensure zero mean and unit variance. For the adjacency matrix $\mathbf{W} = (w_{ij})_{N \times N}$, we set $w_{ij} = 1$ if there is a trade flow between i and j , and 0 otherwise. The row-normalized adjacency matrix $\widetilde{\mathbf{W}}$ is then used in the developed model.

With the estimation methodology developed in Section 3, we first estimate the heterogeneous node-specific time-varying coefficients (see Figure 1), and then determine the latent group structure. The tuning parameters involved are selected as in the simulation study. By the proposed information criterion, we identify 3 groups, i.e., $\widehat{K} = 3$. The group membership details are summarized in Table 4. Group 1 mainly consists of some Asia-Pacific countries such as CHN, KOR, MYS, THA, USA and VNM; whereas Group 3 includes some European countries such as DEU, ESP, GBR, POL and SWE as well as some members of Commonwealth countries (BGD, CAN, DOM, IND, NZL, TTO) which traditionally have close ties to GBR. It seems difficult to interpret Group 2 which contains 63 countries, and may be possible to further split it into some smaller groups via a different clustering technique.

We further conduct post-grouping local linear estimation and kernel-weighted inference. The estimated group-specific time-varying coefficients and their 95% point-wise confidence intervals are plotted in Figures 2. It follows from Figure 2 that the estimated momentum effects $\widehat{\alpha}_{k,2}(\cdot)$ have

Table 4: Group members for the empirical analysis of the trade flow data.

Group 1	Group 2					Group 3	
CHL	DZA	CYP	ITA	PAK	URY	BGD	SVK
CHN	AGO	DNK	JPN	PAN	UZB	CAN	SWE
COL	ARG	ECU	KAZ	PNG	ZMB	CIV	TTO
KOR	AUS	EGY	KWT	PRY		CZE	
MYS	AUT	EST	LAO	PER		DOM	
THA	AZE	FIN	LVA	PRT		DEU	
ARE	BLR	FRA	LBY	QAT		ESP	
USA	BOL	GAB	LTU	RUS		GBR	
VNM	BRA	CHA	MEX	SAU		HUN	
	BRN	GRC	MNG	SVN		IND	
	BGR	GTM	MAR	LKA		JOR	
	KHM	IDN	MMR	CHE		NZL	
	COG	IRN	NLD	TUN		OMN	
	CRI	IRL	NGA	TUR		PHL	
	HRV	ISR	NOR	UKR		POL	

Note: We use the ISO 3166-1 alpha-3 codes to represent countries.

more significant time-varying patterns than the estimated network effects $\hat{\alpha}_{k,1}(\cdot)$. The proposed kernel-weighted test reject the null hypothesis of constant coefficients at the 5% significance level for all the groups with respective p-values 0.047, 0.033, 0.029. Note that the estimated time-varying momentum effect are generally negative, indicating that a fast trade growth rate at time t often leads to a slow rate at $t + 1$. This may be sensible from the demand and supply perspective, as a fast growth rate is often related to a strong demand which results in over-supply in the following time period.

6 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, we impose a latent group structure on time-varying coefficients in the heterogeneous network VAR model. The unknown group number is determined by the extended information criterion whereas the group membership is estimated by the agglomerative clustering algorithm with normalized distance matrix estimates. Theorems 3.2 and 3.3 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

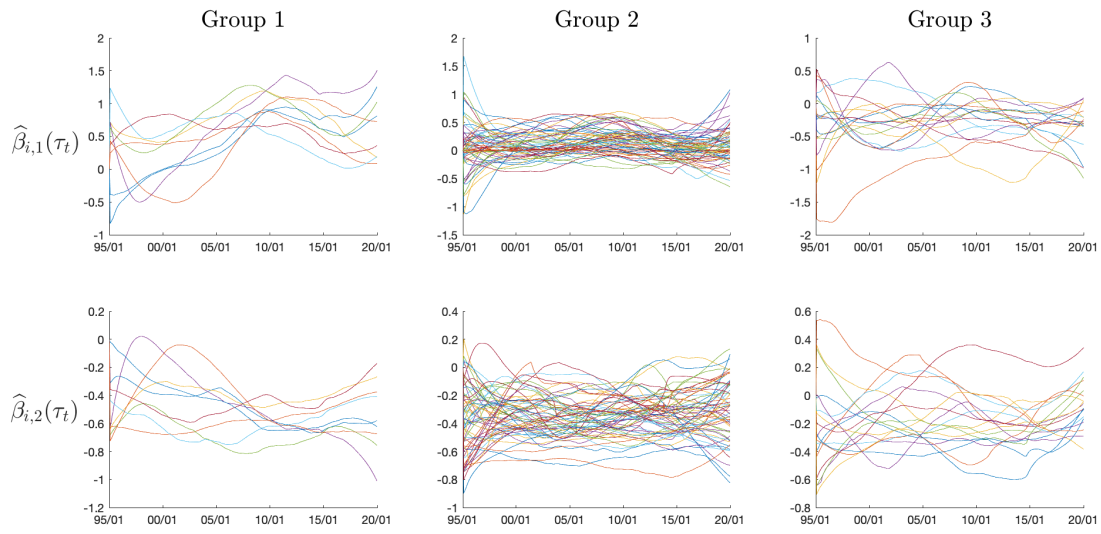


Figure 1: Estimates of the node-specific time-varying coefficients

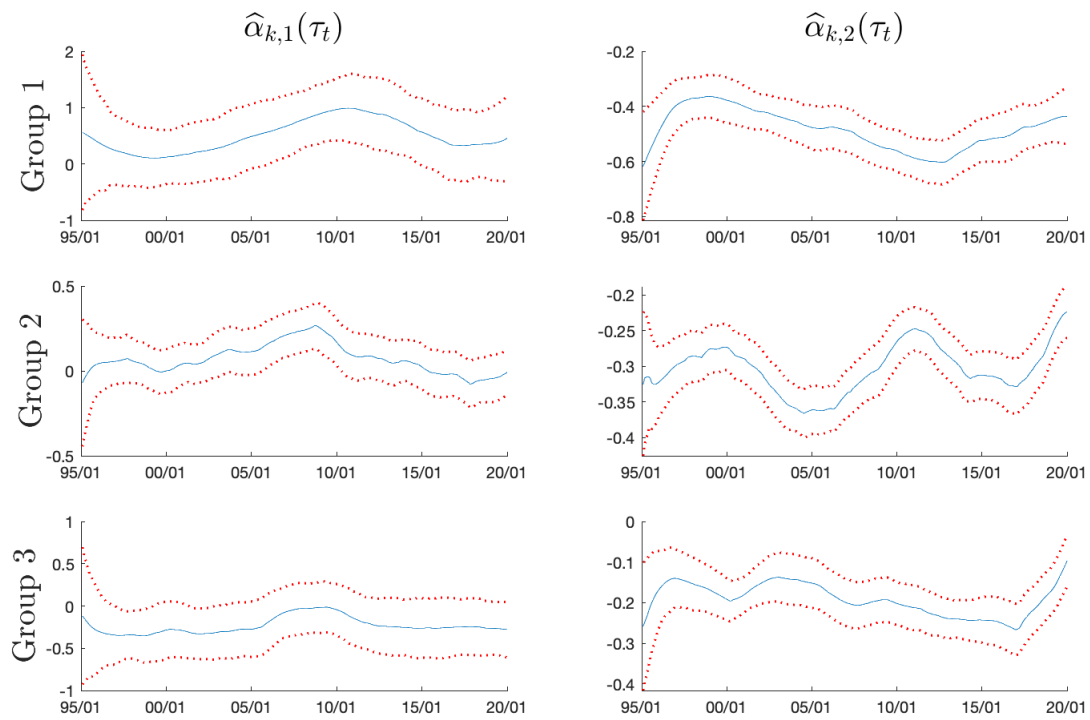


Figure 2: Estimates of the group-specific time-varying coefficients

to estimate the group-specific time-varying momentum and network effects. Meanwhile a post-grouping kernel-weighted test is developed to examine the parametric model specification for group-specific time-varying coefficients. 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 heterogeneous estimation which ignores the latent structure; and (iii) the kernel-weighted test statistic has stable finite-sample performance under either the null or alternative hypothesis. The empirical study shows that there exist three groups over the node-specific time-varying coefficients for the export flow data collected over 90 countries and the momentum effects tend to evolve smoothly over time.

Appendix A: Proofs of the main asymptotic theorems

We next provide the detailed proofs of the main asymptotic theorems stated in Sections 3 and 4. Let C denote a generic positive constant throughout the proofs.

Proof of Proposition 3.1. For any $1 \leq i \leq N$, we define

$$\begin{aligned}
\Gamma_{i,\mathcal{X}}(\tau) &= \begin{bmatrix} \frac{1}{T_h} \sum_{t=1}^T \tilde{X}_{i,t-1} \tilde{X}_{i,t-1}^\top K_{t,0}(\tau) & \frac{1}{T_h} \sum_{t=1}^T \tilde{X}_{i,t-1} \tilde{X}_{i,t-1}^\top K_{t,1}(\tau) \\ \frac{1}{T_h} \sum_{t=1}^T \tilde{X}_{i,t-1} \tilde{X}_{i,t-1}^\top K_{t,1}(\tau) & \frac{1}{T_h} \sum_{t=1}^T \tilde{X}_{i,t-1} \tilde{X}_{i,t-1}^\top K_{t,2}(\tau) \end{bmatrix} \\
&=: \begin{bmatrix} \Gamma_{i,\mathcal{X}}(\tau,0) & \Gamma_{i,\mathcal{X}}(\tau,1) \\ \Gamma_{i,\mathcal{X}}(\tau,1) & \Gamma_{i,\mathcal{X}}(\tau,2) \end{bmatrix}, \\
\Gamma_{i,\beta}(\tau) &= \begin{bmatrix} \frac{1}{T_h} \sum_{t=1}^T \tilde{X}_{i,t-1} \tilde{X}_{i,t-1}^\top [\beta_i(\tau_t) - \beta_i(\tau) - \beta'_i(\tau)(\tau_t - \tau)] K_{t,0}(\tau) \\ \frac{1}{T_h} \sum_{t=1}^T \tilde{X}_{i,t-1} \tilde{X}_{i,t-1}^\top [\beta_i(\tau_t) - \beta_i(\tau) - \beta'_i(\tau)(\tau_t - \tau)] K_{t,1}(\tau) \end{bmatrix} \\
&=: \begin{bmatrix} \Gamma_{i,\beta}(\tau,0) \\ \Gamma_{i,\beta}(\tau,1) \end{bmatrix}, \\
\Gamma_{i,\varepsilon}(\tau) &= \begin{bmatrix} \frac{1}{T_h} \sum_{t=1}^T \tilde{X}_{i,t-1} \varepsilon_{i,t} K_{t,0}(\tau) \\ \frac{1}{T_h} \sum_{t=1}^T \tilde{X}_{i,t-1} \varepsilon_{i,t} K_{t,1}(\tau) \end{bmatrix} =: \begin{bmatrix} \Gamma_{i,\varepsilon}(\tau,0) \\ \Gamma_{i,\varepsilon}(\tau,1) \end{bmatrix},
\end{aligned}$$

where $K_{t,j}(\tau) = \left(\frac{\tau_t - \tau}{h}\right)^j K\left(\frac{\tau_t - \tau}{h}\right)$. From the definition of $\hat{\beta}_i(\cdot)$ in (3.3), we write

$$\hat{\beta}_i(\tau) - \beta_i(\tau) = \mathbf{E}_* \cdot \Gamma_{i,\mathcal{X}}^{-1}(\tau) [\Gamma_{i,\beta}(\tau) + \Gamma_{i,\varepsilon}(\tau)], \quad (\text{A.1})$$

where $\mathbf{E}_* = (\mathbf{I}_2, \mathbf{O}_{2 \times 2})$.

Recall that $\Delta_{ij}(\tau) = \mathbf{E} \left[\tilde{X}_{i,t}(\tau) \tilde{X}_{j,t}^\top(\tau) \right]$ and $\Delta_i(\tau) = \Delta_{ii}(\tau)$. Define $\Lambda_i(\tau) = \text{diag}\{1, \mu_2\} \otimes \Delta_i(\tau)$. By Lemma B.1(i) and Assumption 1, we have

$$\max_{1 \leq i \leq N} \|\Gamma_{i,\mathcal{X}}(\tau) - \Lambda_i(\tau)\| = o_P(1). \quad (\text{A.2})$$

By Taylor's expansion and following the proof of Lemma B.1(i), we have for any $\tau \in (0, 1)$,

$$\Gamma_{i,\beta}(\tau) = \begin{bmatrix} \frac{1}{2}h^2\mu_2\Delta_i(\tau)\beta_i''(\tau) \\ \mathbf{0} \end{bmatrix} + o_P(h^2). \quad (\text{A.3})$$

By (A.1)–(A.3), we may show that, for $i \neq j$,

$$\begin{aligned} \widehat{d}_{ij}(\tau) - d_{ij}(\tau) &= \mathbf{E}_* \cdot \left\{ \left[\Gamma_{i,\mathcal{X}}^{-1}(\tau)\Gamma_{i,\beta}(\tau) - \Gamma_{j,\mathcal{X}}^{-1}(\tau)\Gamma_{j,\beta}(\tau) \right] + \left[\Gamma_{i,\mathcal{X}}^{-1}(\tau)\Gamma_{i,\varepsilon}(\tau) - \Gamma_{j,\mathcal{X}}^{-1}(\tau)\Gamma_{j,\varepsilon}(\tau) \right] \right\} \\ &= \mathbf{E}_* \cdot \left\{ \left[\Lambda_i^{-1}(\tau)\Gamma_{i,\varepsilon}(\tau) - \Lambda_j^{-1}(\tau)\Gamma_{j,\varepsilon}(\tau) \right] + \left[\frac{1}{2}h^2\mu_2 d_{ij}''(\tau), \mathbf{0}^\top \right]^\top \right. \\ &\quad \left. + o_P\left((Th)^{-1/2} + h^2\right) \right\}. \end{aligned} \quad (\text{A.4})$$

We next only need to show that

$$\sqrt{Th} \cdot \left[\Lambda_i^{-1}(\tau)\Gamma_{i,\varepsilon}(\tau) - \Lambda_j^{-1}(\tau)\Gamma_{j,\varepsilon}(\tau) \right] \xrightarrow{d} \mathbf{N}(\mathbf{0}, \Sigma_{ij}^*(\tau)), \quad (\text{A.5})$$

where $\Sigma_{ij}^*(\tau) = \text{diag}\{v_0, v_2\mu_2^{-2}\} \otimes \Sigma_{ij}(\tau)$ with $\Sigma_{ij}(\tau)$ defined in Proposition 3.1. By the Cramér-Wold device (e.g., Billingsley, 1995), we only need to prove that for any four-dimensional vector \mathbf{u} of numbers,

$$\sqrt{Th} \cdot \mathbf{u}^\top \left[\Lambda_i^{-1}(\tau)\Gamma_{i,\varepsilon}(\tau) - \Lambda_j^{-1}(\tau)\Gamma_{j,\varepsilon}(\tau) \right] \xrightarrow{d} \mathbf{N}(0, \mathbf{u}^\top \Sigma_{ij}^*(\tau) \mathbf{u}). \quad (\text{A.6})$$

Note that $\mathbf{u}^\top [\Lambda_i^{-1}(\tau)\Gamma_{i,\varepsilon}(\tau) - \Lambda_j^{-1}(\tau)\Gamma_{j,\varepsilon}(\tau)]$ is martingale with respect to \mathcal{F}_t . Write $\mathbf{u} = (\mathbf{u}_1^\top, \mathbf{u}_2^\top)^\top$ with \mathbf{u}_1 and \mathbf{u}_2 being two-dimensional vectors of numbers. Let

$$\begin{aligned} \sigma_{ij,T}^2 &= \frac{1}{Th} \sum_{t=1}^T \mathbf{E} \left(\left\{ \left[\mathbf{u}_1^\top \Delta_i^{-1}(\tau) \widetilde{X}_{i,t-1} K_{t,0}(\tau) + \mathbf{u}_2^\top \mu_2^{-1} \Delta_i^{-1} \widetilde{X}_{i,t-1} K_{t,1}(\tau) \right] \varepsilon_{i,t} \right. \right. \\ &\quad \left. \left. - \left[\mathbf{u}_1^\top \Delta_j^{-1}(\tau) \widetilde{X}_{j,t-1} K_{t,0}(\tau) + \mathbf{u}_2^\top \mu_2^{-1} \Delta_j^{-1}(\tau) \widetilde{X}_{j,t-1} K_{t,1}(\tau) \right] \varepsilon_{j,t} \right\}^2 \middle| \mathcal{F}_{t-1} \right), \end{aligned}$$

Using Assumption 1(iii) and Lemmas B.1(i) and B.2(i), we may show that $\sigma_{ij,T}^2 \xrightarrow{P} \mathbf{u}^\top \Sigma_{ij}^*(\tau) \mathbf{u}$ as $T \rightarrow \infty$. Then, by the martingale central limit theorem (e.g., Corollary 3.1 in Hall and Heyde, 1980), we prove (A.6) and then (A.5). By virtue of (A.4) and (A.5), we complete the proof of (3.5). \square

Proof of Theorem 3.2. To prove (3.8), we only need to show that

$$\mathbf{P} \left(\max_{1 \leq k \leq K_0} \max_{i,j \in \mathcal{G}_k} \widehat{D}_{ij} < \min_{1 \leq k \neq l \leq K_0} \min_{i \in \mathcal{G}_k, j \in \mathcal{G}_l} \widehat{D}_{ij} \right) \rightarrow 1 \quad (\text{A.7})$$

as T tends to infinity. Notice that we have $D_{ij} = 0$ if $i, j \in \mathcal{G}_k$, and $\min_{1 \leq k \neq l \leq K_0} \min_{i \in \mathcal{G}_k, j \in \mathcal{G}_l} D_{ij} > 0$. Hence,

to prove (A.7), it is sufficient to show that

$$\max_{1 \leq i, j \leq N} |\widehat{D}_{ij} - D_{ij}| = o_P(\underline{D}), \quad (\text{A.8})$$

where \underline{D} is defined in (3.7).

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

$$|\widehat{D}_{ij} - D_{ij}| \leq |\widehat{D}_{ij} - \widetilde{D}_{ij}| + |\widetilde{D}_{ij} - D_{ij}|.$$

By the definition of the Riemann integral and $1/L = o(\underline{D})$ in (3.7), we readily have

$$\max_{1 \leq i, j \leq N} |\widetilde{D}_{ij} - D_{ij}| = O_P(1/L) = o_P(\underline{D}). \quad (\text{A.9})$$

On the other hand, note that $1 \leq i, j \leq N$,

$$\begin{aligned} |\widehat{D}_{ij} - \widetilde{D}_{ij}| &= \left| \frac{1}{L} \sum_{l=1}^L \left[\widehat{\mathbf{d}}_{ij}^T(\tau_l^*) \widehat{\boldsymbol{\Sigma}}_{ij}^{-1}(\tau_l^*) \widehat{\mathbf{d}}_{ij}(\tau_l^*) - \mathbf{d}_{ij}^T(\tau_l^*) \boldsymbol{\Sigma}_{ij, T}^{-1}(\tau_l^*) \mathbf{d}_{ij}(\tau_l^*) \right] \right| \\ &\leq \frac{1}{L} \sum_{l=1}^L \left| \widehat{\mathbf{d}}_{ij}^T(\tau_l^*) \left[\widehat{\boldsymbol{\Sigma}}_{ij}^{-1}(\tau_l^*) - \boldsymbol{\Sigma}_{ij, T}^{-1}(\tau_l^*) \right] \widehat{\mathbf{d}}_{ij}(\tau_l^*) \right| + \\ &\quad \frac{1}{L} \sum_{l=1}^L \left| \left[\widehat{\mathbf{d}}_{ij}(\tau_l^*) + \mathbf{d}_{ij}(\tau_l^*) \right]^T \boldsymbol{\Sigma}_{ij, T}^{-1}(\tau_l^*) \left[\widehat{\mathbf{d}}_{ij}(\tau_l^*) - \mathbf{d}_{ij}(\tau_l^*) \right] \right| \\ &=: I_{ij}(1) + I_{ij}(2). \end{aligned}$$

Note that $\left\| \boldsymbol{\Sigma}_{ij, T}^{-1}(\tau) \right\|$ is uniformly bounded since $\boldsymbol{\Sigma}_{ij}(\tau)$ is positive definite uniformly over i, j and τ and $\boldsymbol{\Sigma}_{ij, T}(\tau)$ uniformly converges to $\boldsymbol{\Sigma}_{ij}(\tau)$. Then, for $I_{ij}(2)$, using Lemma B.3, we readily have that

$$\max_{1 \leq i, j \leq N} I_{ij}(2) = O_P \left(\sqrt{\frac{\log(N \vee T)}{Th}} + h^2 \right). \quad (\text{A.10})$$

We next consider $I_{ij}(1)$. Note that

$$\begin{aligned} \widehat{\sigma}_{ij} - \sigma_{ij} &= \frac{1}{T} \sum_{t=1}^T \widehat{\varepsilon}_{i,t} \widehat{\varepsilon}_{j,t} - \frac{1}{T} \sum_{t=1}^T \varepsilon_{i,t} \varepsilon_{j,t} + \frac{1}{T} \sum_{t=1}^T \varepsilon_{i,t} \varepsilon_{j,t} - \mathbb{E}(\varepsilon_{i,t} \varepsilon_{j,t}) \\ &= \frac{1}{T} \sum_{t=1}^T (\widehat{\varepsilon}_{i,t} \widehat{\varepsilon}_{j,t} - \varepsilon_{i,t} \varepsilon_{j,t}) + O_P \left(\sqrt{\frac{\log N}{T}} \right), \end{aligned}$$

and

$$|\widehat{\varepsilon}_{i,t} - \varepsilon_{i,t}| \leq \left| \left[\widehat{\boldsymbol{\beta}}_i(\tau_t) - \boldsymbol{\beta}_i(\tau_t) \right]^T \widetilde{X}_{i,t-1} \right|.$$

By Lemma B.3, we then have

$$\max_{1 \leq i, j \leq N} |\hat{\sigma}_{ij} - \sigma_{ij}| = O_p \left(\sqrt{\frac{\log(N \vee T)}{Th}} + h^2 \right),$$

which, together with Lemmas B.1(ii) and B.2(ii), leads to

$$\max_{1 \leq i, j \leq N} \sup_{\tau \in [0,1]} \left\| \hat{\Sigma}_{ij}(\tau) - \Sigma_{ij,T}(\tau) \right\| = O_p \left(\sqrt{\frac{\log(N \vee T)}{Th}} + h^2 \right). \quad (\text{A.11})$$

This indicates that

$$\max_{1 \leq i, j \leq N} I_{ij}(1) \leq C \cdot \frac{1}{L} \sum_{l=1}^L \left\| \hat{\Sigma}_{ij}^{-1}(\tau_l^*) - \Sigma_{ij,T}^{-1}(\tau_l^*) \right\| = O_p \left(\sqrt{\frac{\log(N \vee T)}{Th}} + h^2 \right). \quad (\text{A.12})$$

By (3.7), (A.10) and (A.12), we have

$$\max_{1 \leq i, j \leq N} |\hat{D}_{ij} - \tilde{D}_{ij}| = O_p \left(\sqrt{\frac{\log(N \vee T)}{Th}} + h^2 \right) = o_p(\underline{D}). \quad (\text{A.13})$$

With (A.9) and (A.13), we complete the proof of (A.8). \square

Proof of Theorem 3.3. By (3.13), we only need to show that

$$\mathbb{P} \left(\text{IC}(K_0) = \min_{1 \leq K \leq \bar{K}} \text{IC}(K) \right) \rightarrow 1. \quad (\text{A.14})$$

To prove (A.14), we next consider the following two scenarios: (i) $1 \leq K \leq K_0 - 1$ and (ii) $K_0 + 1 \leq K \leq \bar{K}$, corresponding to the under-fitted and over-fitted models, respectively. Let $\mathcal{E}(\mathcal{G})$ denote the event that $\{\hat{\mathcal{G}}_1, \dots, \hat{\mathcal{G}}_{K_0}\} = \{\mathcal{G}_1, \dots, \mathcal{G}_{K_0}\}$. For scenario (i), by Theorem 3.2 and Lemma B.4 in Appendix B, we have

$$\begin{aligned} & \mathbb{P}(\text{IC}(K_0) < \text{IC}(K), 1 \leq K \leq K_0 - 1) \\ &= \mathbb{P}(\text{IC}(K_0) < \text{IC}(K), 1 \leq K \leq K_0 - 1, \mathcal{E}(\mathcal{G})) + o(1) \\ &= 1 + o(1). \end{aligned} \quad (\text{A.15})$$

On the other hand, for scenario (ii), by Theorem 3.2 and Lemma B.5, we have

$$\begin{aligned} & \mathbb{P}(\text{IC}(K_0) < \text{IC}(K), K_0 + 1 \leq K \leq \bar{K}) \\ &= \mathbb{P}(\text{IC}(K_0) < \text{IC}(K), K_0 + 1 \leq K \leq \bar{K}, \mathcal{E}(\mathcal{G})) + o(1) \\ &= 1 + o(1). \end{aligned} \quad (\text{A.16})$$

Combine (A.15) and (A.16), we complete the proof of Theorem 3.3. \square

Proof of Theorem 4.1. By the consistency properties in Theorems 3.2 and 3.3, we may prove the asymptotic

distribution theory conditional on $\widehat{\mathcal{G}}_k = \mathcal{G}_k$ and $\widehat{K} = K_0$. For $k = 1, \dots, K_0$, let

$$\begin{aligned}\Xi_{k,X}(\tau) &= \begin{bmatrix} \frac{1}{|\mathcal{G}_k|Th_2} \sum_{t=1}^T \sum_{i \in \mathcal{G}_k} \widetilde{X}_{i,t-1} \widetilde{X}_{i,t-1}^\top K_{t,0}^\dagger(\tau) & \frac{1}{|\mathcal{G}_k|Th_2} \sum_{t=1}^T \sum_{i \in \mathcal{G}_k} \widetilde{X}_{i,t-1} \widetilde{X}_{i,t-1}^\top K_{t,1}^\dagger(\tau) \\ \frac{1}{|\mathcal{G}_k|Th_2} \sum_{t=1}^T \sum_{i \in \mathcal{G}_k} \widetilde{X}_{i,t-1} \widetilde{X}_{i,t-1}^\top K_{t,1}^\dagger(\tau) & \frac{1}{|\mathcal{G}_k|Th_2} \sum_{t=1}^T \sum_{i \in \mathcal{G}_k} \widetilde{X}_{i,t-1} \widetilde{X}_{i,t-1}^\top K_{t,2}^\dagger(\tau) \end{bmatrix}, \\ \Xi_{k,\alpha}(\tau) &= \begin{bmatrix} \frac{1}{|\mathcal{G}_k|Th_2} \sum_{t=1}^T \sum_{i \in \mathcal{G}_k} \widetilde{X}_{i,t-1} \widetilde{X}_{i,t-1}^\top [\alpha_k(\tau_t) - \alpha_k(\tau) - \alpha'_k(\tau)(\tau_t - \tau)] K_{t,0}^\dagger(\tau) \\ \frac{1}{|\mathcal{G}_k|Th_2} \sum_{t=1}^T \sum_{i \in \mathcal{G}_k} \widetilde{X}_{i,t-1} \widetilde{X}_{i,t-1}^\top [\alpha_k(\tau_t) - \alpha_k(\tau) - \alpha'_k(\tau)(\tau_t - \tau)] K_{t,1}^\dagger(\tau) \end{bmatrix}, \\ \Xi_{k,\varepsilon}(\tau) &= \begin{bmatrix} \frac{1}{|\mathcal{G}_k|Th_2} \sum_{t=1}^T \sum_{i \in \mathcal{G}_k} \widetilde{X}_{i,t-1} \varepsilon_{i,t} K_{t,0}^\dagger(\tau) \\ \frac{1}{|\mathcal{G}_k|Th_2} \sum_{t=1}^T \sum_{i \in \mathcal{G}_k} \widetilde{X}_{i,t-1} \varepsilon_{i,t} K_{t,1}^\dagger(\tau) \end{bmatrix},\end{aligned}$$

where $K_{t,j}^\dagger(\tau) = \left(\frac{\tau_t - \tau}{h_2}\right)^j K\left(\frac{\tau_t - \tau}{h_2}\right)$. By the definition of the post-grouping local linear estimation defined in (4.2), we have

$$\widehat{\alpha}_k(\tau) - \alpha_k(\tau) = \mathbf{E}_* \cdot \Xi_{k,X}^{-1}(\tau) [\Xi_{k,\alpha}(\tau) + \Xi_{k,\varepsilon}(\tau)].$$

In order to prove (4.4) in Theorem 4.1, we only need to show

$$\Xi_{k,X}(\tau) \xrightarrow{P} \text{diag}\{1, \mu_2\} \otimes \Delta_{\mathcal{G}_k}(\tau), \quad (\text{A.17})$$

$$\Xi_{k,\alpha}(\tau) = \begin{bmatrix} \frac{1}{2} h_2^2 \mu_2 \Delta_{\mathcal{G}_k}(\tau) \alpha_k''(\tau) \\ \mathbf{0} \end{bmatrix} + o_P(h_2^2), \quad (\text{A.18})$$

$$\sqrt{|\mathcal{G}_k|Th_2} \Xi_{k,\varepsilon}(\tau) \xrightarrow{d} N(\mathbf{0}, \text{diag}\{\nu_0, \nu_2\} \otimes \Upsilon_{\mathcal{G}_k}(\tau)). \quad (\text{A.19})$$

As the proof of (A.17) is similar to the proof of (A.2), details are omitted here. By the smoothness condition in Assumption 1(ii), Taylor's expansion of $\alpha_k(\cdot)$ and (A.17), we can prove (A.18). We next turn to the proof of (A.19). Let

$$W_{\mathcal{G}_k,t}(\tau) = [W_{\mathcal{G}_k,t}^\top(\tau, 0), W_{\mathcal{G}_k,t}^\top(\tau, 1)]^\top \quad \text{with} \quad W_{\mathcal{G}_k,t}(\tau, j) = \frac{1}{\sqrt{|\mathcal{G}_k|}} \sum_{i \in \mathcal{G}_k} \varepsilon_{i,t} \widetilde{X}_{i,t-1} K_{t,j}^\dagger(\tau),$$

and write

$$\sqrt{|\mathcal{G}_k|Th_2} \Xi_{k,\varepsilon}(\tau) = \frac{1}{\sqrt{Th_2}} \sum_{t=1}^T W_{\mathcal{G}_k,t}(\tau).$$

Using Lemma B.2(i), we may show that

$$\begin{aligned}\frac{1}{Th_2} \sum_{t=1}^T \mathbf{E} [W_{\mathcal{G}_k,t}(\tau, 0) W_{\mathcal{G}_k,t}^\top(\tau, 0) | \mathcal{F}_{t-1}] &= \frac{1}{|\mathcal{G}_k|} \sum_{i,j \in \mathcal{G}_k} \sigma_{ij} \left[\frac{1}{Th_2} \sum_{t=1}^T \widetilde{X}_{i,t-1} \widetilde{X}_{j,t-1}^\top K_{t,0}^\dagger(\tau) \right] \\ &\xrightarrow{P} \frac{\nu_0}{|\mathcal{G}_k|} \sum_{i,j \in \mathcal{G}_k} \sigma_{ij} \Delta_{ij}(\tau).\end{aligned}$$

Similarly, we can also prove that

$$\frac{1}{T h_2} \sum_{t=1}^T \mathbb{E} [W_{\mathcal{G}_k,t}(\tau, 1) W_{\mathcal{G}_k,t}^\top(\tau, 1) | \mathcal{F}_{t-1}] \xrightarrow{P} \frac{\nu_2}{|\mathcal{G}_k|} \sum_{i,j \in \mathcal{G}_k} \sigma_{ij} \Delta_{ij}(\tau),$$

and

$$\frac{1}{T h_2} \sum_{t=1}^T \mathbb{E} [W_{\mathcal{G}_k,t}(\tau, 0) W_{\mathcal{G}_k,t}^\top(\tau, 1) | \mathcal{F}_{t-1}] = o_P(1).$$

Hence, we have

$$\frac{1}{T h_2} \sum_{t=1}^T \mathbb{E} [W_{\mathcal{G}_k,t}(\tau) W_{\mathcal{G}_k,t}^\top(\tau) | \mathcal{F}_{t-1}] \xrightarrow{P} \text{diag}\{\nu_0, \nu_2\} \otimes \frac{1}{|\mathcal{G}_k|} \sum_{i,j \in \mathcal{G}_k} \sigma_{ij} \Delta_{ij}(\tau),$$

which, together with (4.3) and the martingale central limit theorem (e.g., Hall and Heyde, 1980), leads to (A.19). The proof of Theorem 4.1 is completed. \square

Proof of Theorem 4.2. Denote

$$Q_T(\mathbf{k}) = \sum_{t=1}^T \sum_{s \neq t} K \left(\frac{\tau_t - \tau_s}{h_3} \right) \varepsilon_t^\top(\mathbf{k}) \varepsilon_s(\mathbf{k}),$$

where $\varepsilon_t(\mathbf{k})$ is defined in Assumption 4(ii). It suffices to prove that

$$(2\nu_0 T^2 h_3)^{-1/2} \|\boldsymbol{\Sigma}_\varepsilon(\mathbf{k})\|^{-1} Q_T(\mathbf{k}) \xrightarrow{d} \mathbf{N}(0, 1), \quad (\text{A.20})$$

$$(T^2 h_3)^{-1/2} \|\boldsymbol{\Sigma}_\varepsilon(\mathbf{k})\|^{-1} [\tilde{Q}_T(\mathbf{k}) - Q_T(\mathbf{k})] = o_P(1). \quad (\text{A.21})$$

Letting $Y_t(\mathbf{k}) = 2\varepsilon_t^\top(\mathbf{k}) \sum_{s < t} K \left(\frac{\tau_t - \tau_s}{h_3} \right) \varepsilon_s(\mathbf{k})$, we readily have that

$$Q_T(\mathbf{k}) = 2 \sum_{t=1}^T \sum_{s < t} K \left(\frac{\tau_t - \tau_s}{h_3} \right) \varepsilon_t^\top(\mathbf{k}) \varepsilon_s(\mathbf{k}) = \sum_{t=1}^T Y_t(\mathbf{k}),$$

which is a martingale with respect to \mathcal{F}_t . Note that

$$\begin{aligned} \frac{1}{T^2 h_3} \sum_{t=1}^T \mathbb{E} [Y_t^2(\mathbf{k}) | \mathcal{F}_{t-1}] &= \frac{4}{T^2 h_3} \sum_{t=1}^T \sum_{s_1, s_2 < t} K \left(\frac{\tau_t - \tau_{s_1}}{h_3} \right) K \left(\frac{\tau_t - \tau_{s_2}}{h_3} \right) \varepsilon_{s_1}^\top(\mathbf{k}) \mathbb{E} [\varepsilon_t(\mathbf{k}) \varepsilon_t^\top(\mathbf{k}) | \mathcal{F}_{t-1}] \varepsilon_{s_2}(\mathbf{k}) \\ &= \frac{4}{T^2 h_3} \sum_{t=1}^T \sum_{s < t} \varepsilon_s^\top(\mathbf{k}) \boldsymbol{\Sigma}_\varepsilon(\mathbf{k}) \varepsilon_s(\mathbf{k}) K^2 \left(\frac{\tau_t - \tau_s}{h_3} \right) + O_P(T^{-1/2}) \\ &= 2 \cdot \text{trace}\{\boldsymbol{\Sigma}_\varepsilon(\mathbf{k}) \boldsymbol{\Sigma}_\varepsilon(\mathbf{k})\} \frac{2}{T^2 h_3} \sum_{t=1}^T \sum_{s < t} K^2 \left(\frac{\tau_t - \tau_s}{h_3} \right) + o_P(1) \end{aligned}$$

$$\begin{aligned}
&= 2\|\boldsymbol{\Sigma}_\varepsilon(\mathbf{k})\|^2 \frac{1}{T^2 h_3} \sum_{t=1}^T \sum_{s \neq t} K^2 \left(\frac{\tau_t - \tau_s}{h_3} \right) + o_P(1) \\
&= 2\|\boldsymbol{\Sigma}_\varepsilon(\mathbf{k})\|^2 \nu_0 (1 + o_P(1)).
\end{aligned} \tag{A.22}$$

Then, by (A.22) and the martingale central limit theorem, we prove (A.20).

We next turn to the proof of (A.21) conditioning on $\widehat{\mathcal{G}}_k = \mathcal{G}_k$. Define

$$h_t(\mathbf{k}) = \left(\left[g_k(\tau_t; \tilde{\theta}_k) - g_k(\tau_t; \theta_k) \right]^\top \tilde{X}_{i,t-1}, i \in \mathcal{G}_k \right).$$

Consider the following decomposition:

$$\tilde{Q}_T(\mathbf{k}) - Q_T(\mathbf{k}) = \sum_{t=1}^T \sum_{s \neq t} K \left(\frac{\tau_t - \tau_s}{h_3} \right) h_t^\top(\mathbf{k}) h_s(\mathbf{k}) - 2 \sum_{t=1}^T \sum_{s \neq t} K \left(\frac{\tau_t - \tau_s}{h_3} \right) \varepsilon_t^\top(\mathbf{k}) h_s(\mathbf{k}).$$

Let

$$\omega_t(\mathbf{k}) = \left(\left[\partial_{\theta} g_k(\tau_t; \theta_k) (\tilde{\theta}_k - \theta_k) \right]^\top \tilde{X}_{i,t-1}, i \in \mathcal{G}_k \right)$$

and $r_t(\mathbf{k}) = h_t(\mathbf{k}) - \omega_t(\mathbf{k})$. By Assumption 4(i), we have

$$|\omega_t(\mathbf{k})|^2 = O(\delta_\theta^2) \sum_{i \in \mathcal{G}_k} |\tilde{X}_{i,t-1}|^2, \quad |r_t(\mathbf{k})|^2 = O(\delta_\theta^4) \sum_{i \in \mathcal{G}_k} |\tilde{X}_{i,t-1}|^2. \tag{A.23}$$

Hence, by (A.23) and Assumption 4(ii) we may show that

$$\begin{aligned}
&\frac{1}{(T^2 h_3)^{1/2} \|\boldsymbol{\Sigma}_\varepsilon(\mathbf{k})\|} \sum_{t=1}^T \sum_{s \neq t} K \left(\frac{\tau_t - \tau_s}{h_3} \right) h_t^\top(\mathbf{k}) h_s(\mathbf{k}) \\
&= \frac{1}{(T^2 h_3)^{1/2} \|\boldsymbol{\Sigma}_\varepsilon(\mathbf{k})\|} \sum_{t=1}^T \sum_{s \neq t} K \left(\frac{\tau_t - \tau_s}{h_3} \right) [\omega_t^\top(\mathbf{k}) \omega_s(\mathbf{k}) + \omega_t^\top(\mathbf{k}) r_s(\mathbf{k}) + r_t^\top(\mathbf{k}) \omega_s(\mathbf{k}) + r_t^\top(\mathbf{k}) r_s(\mathbf{k})] \\
&= \frac{O(\delta_\theta^2 + \delta_\theta^4)}{(T^2 h_3)^{1/2} \|\boldsymbol{\Sigma}_\varepsilon(\mathbf{k})\|} \sum_{t=1}^T \sum_{s \neq t} K \left(\frac{\tau_t - \tau_s}{h_3} \right) \sum_{i \in \mathcal{G}_k} |\tilde{X}_{i,t-1}|^2 \\
&= O(\delta_\theta^2) \frac{O_P(|\mathcal{G}_k| T^2 h_3)}{(T^2 h_3)^{1/2} \|\boldsymbol{\Sigma}_\varepsilon(\mathbf{k})\|} \\
&= O_P \left(\frac{\delta_\theta^2 N T h_3^{1/2}}{\|\boldsymbol{\Sigma}_\varepsilon(\mathbf{k})\|} \right) = o_P(1).
\end{aligned} \tag{A.24}$$

On the other hand, using (A.23) and noting that

$$E \left[\left| \sum_{i \in \mathcal{G}_k} \sum_{t=1}^T \sum_{t \neq s} K \left(\frac{\tau_t - \tau_s}{h_3} \right) \partial_{\theta} g_k(\tau_s; \theta_k) \tilde{X}_{i,s-1} \varepsilon_{i,t} \right|^2 \right] = O(|\mathcal{G}_k| T^3 h_3)$$

which can be proved using the first condition in (3.14), we may show that

$$\begin{aligned}
& \frac{1}{(T^2 h_3)^{1/2} \|\boldsymbol{\Sigma}_\varepsilon(\mathbf{k})\|} \sum_{t=1}^T \sum_{s \neq t} K\left(\frac{\tau_t - \tau_s}{h_3}\right) \varepsilon_t^\top(\mathbf{k}) h_s(\mathbf{k}) \\
& \leq \frac{|\tilde{\theta}_k - \theta_k|}{(T^2 h_3)^{1/2} \|\boldsymbol{\Sigma}_\varepsilon(\mathbf{k})\|} \left| \sum_{i \in \mathcal{G}_k} \sum_{t=1}^T \sum_{t \neq s} K\left(\frac{\tau_t - \tau_s}{h_3}\right) \partial_{\theta} g_k(\tau_s; \theta_k) \tilde{X}_{i,s-1} \varepsilon_{i,t} \right| \\
& = \frac{O(\delta_\theta)}{(T^2 h_3)^{1/2} \|\boldsymbol{\Sigma}_\varepsilon(\mathbf{k})\|} \cdot O_P\left(|\mathcal{G}_k|^{1/2} T^{3/2} h_3^{1/2}\right) \\
& = O_P\left(\frac{\delta_\theta (NT)^{1/2}}{\|\boldsymbol{\Sigma}_\varepsilon(\mathbf{k})\|}\right) = o_P(1). \tag{A.25}
\end{aligned}$$

By (A.24) and (A.25), we complete the proof of (A.21). \square

Proof of Theorem 4.3. Let $\mathcal{Q}_T^\diamond(\mathbf{k})$ be defined similarly to $\tilde{\mathcal{Q}}_T(\mathbf{k})$ but with $\tilde{\theta}_k$ replaced by θ_k^\diamond . Denote

$$\begin{aligned}
\omega_t(\mathbf{k}) &= \left(\alpha_k^\top(\tau_t) \tilde{X}_{i,t-1}, i \in \mathcal{G}_k \right), \\
\omega_t^\diamond(\mathbf{k}) &= \left(g_k^\top(\tau_t; \theta_k^\diamond) \tilde{X}_{i,t-1}, i \in \mathcal{G}_k \right), \\
\omega_t^\dagger(\mathbf{k}) &= \left(g_k^\top(\tau_t; \tilde{\theta}_k) \tilde{X}_{i,t-1}, i \in \mathcal{G}_k \right),
\end{aligned}$$

and $\omega_t^\ddagger(\mathbf{k}) = \omega_t^\dagger(\mathbf{k}) - \omega_t^\diamond(\mathbf{k})$. Write

$$\begin{aligned}
\tilde{\mathcal{Q}}_T(\mathbf{k}) - \mathcal{Q}_T^\diamond(\mathbf{k}) &= -2 \sum_{t=1}^T \sum_{s \neq t} K\left(\frac{\tau_t - \tau_s}{h_3}\right) [\omega_t^\ddagger(\mathbf{k})]^\top \varepsilon_s(\mathbf{k}) + \sum_{t=1}^T \sum_{s \neq t} K\left(\frac{\tau_t - \tau_s}{h_3}\right) [\omega_t^\ddagger(\mathbf{k})]^\top \omega_s^\ddagger(\mathbf{k}) - \\
& \quad 2 \sum_{t=1}^T \sum_{s \neq t} K\left(\frac{\tau_t - \tau_s}{h_3}\right) [\omega_t^\ddagger(\mathbf{k})]^\top [\omega_s(\mathbf{k}) - \omega_s^\diamond(\mathbf{k})].
\end{aligned}$$

Following the proofs of (A.24) and (A.25), we can similarly prove that

$$\frac{1}{(T^2 h_3)^{1/2} \|\boldsymbol{\Sigma}_\varepsilon(\mathbf{k})\|} \sum_{t=1}^T \sum_{s \neq t} K\left(\frac{\tau_t - \tau_s}{h_3}\right) [\omega_t^\ddagger(\mathbf{k})]^\top \omega_s^\ddagger(\mathbf{k}) = o_P(1), \tag{A.26}$$

$$\frac{1}{(T^2 h_3)^{1/2} \|\boldsymbol{\Sigma}_\varepsilon(\mathbf{k})\|} \sum_{t=1}^T \sum_{s \neq t} K\left(\frac{\tau_t - \tau_s}{h_3}\right) [\omega_t^\ddagger(\mathbf{k})]^\top \varepsilon_s(\mathbf{k}) = o_P(1). \tag{A.27}$$

By the Cauchy-Schwarz inequality and Assumption 5(i), we may show that

$$\frac{1}{(T^2 h_3)^{1/2} \|\boldsymbol{\Sigma}_\varepsilon(\mathbf{k})\|} \sum_{t=1}^T \sum_{s \neq t} K\left(\frac{\tau_t - \tau_s}{h_3}\right) \left| [\omega_t^\ddagger(\mathbf{k})]^\top [\omega_s(\mathbf{k}) - \omega_s^\diamond(\mathbf{k})] \right|$$

$$\begin{aligned}
&\leq \frac{1}{(\mathbb{T}^2 h_3)^{1/2} \|\boldsymbol{\Sigma}_\varepsilon(\mathbf{k})\|} \sum_{t=1}^{\mathbb{T}} \left| \boldsymbol{\omega}_t^\dagger(\mathbf{k}) \right| \sum_{s \neq t} \mathbb{K} \left(\frac{\tau_t - \tau_s}{h_3} \right) |\boldsymbol{\omega}_s(\mathbf{k}) - \boldsymbol{\omega}_s^\diamond(\mathbf{k})| \\
&= O_P \left(\frac{|\mathcal{G}_k| \mathbb{T}^2 h_3 \delta_\theta^\diamond \varphi}{(\mathbb{T}^2 h_3)^{1/2} \|\boldsymbol{\Sigma}_\varepsilon(\mathbf{k})\|} \right) = O_P \left(\frac{\mathbb{N} \mathbb{T} h_3^{1/2} \delta_\theta^\diamond \varphi}{\|\boldsymbol{\Sigma}_\varepsilon(\mathbf{k})\|} \right). \tag{A.28}
\end{aligned}$$

By (A.26)–(A.28) and Assumption 5(iii), we readily have that

$$\frac{1}{(\mathbb{T}^2 h_3)^{1/2} \|\boldsymbol{\Sigma}_\varepsilon(\mathbf{k})\|} \left| \tilde{\mathcal{Q}}_T(\mathbf{k}) - \mathcal{Q}_T^\diamond(\mathbf{k}) \right| = o_P \left(\frac{\mathbb{N} \mathbb{T} h_3^{1/2} \varphi^2}{\|\boldsymbol{\Sigma}_\varepsilon(\mathbf{k})\|} \vee 1 \right). \tag{A.29}$$

On the other hand, we note that

$$\begin{aligned}
\mathcal{Q}_T^\diamond(\mathbf{k}) - \mathcal{Q}_T(\mathbf{k}) &= 2 \sum_{t=1}^{\mathbb{T}} \sum_{s \neq t} \mathbb{K} \left(\frac{\tau_t - \tau_s}{h_3} \right) \boldsymbol{\varepsilon}_t^\top(\mathbf{k}) [\boldsymbol{\omega}_s(\mathbf{k}) - \boldsymbol{\omega}_s^\diamond(\mathbf{k})] + \\
&\quad \sum_{t=1}^{\mathbb{T}} \sum_{s \neq t} \mathbb{K} \left(\frac{\tau_t - \tau_s}{h_3} \right) [\boldsymbol{\omega}_t(\mathbf{k}) - \boldsymbol{\omega}_t^\diamond(\mathbf{k})]^\top [\boldsymbol{\omega}_s(\mathbf{k}) - \boldsymbol{\omega}_s^\diamond(\mathbf{k})].
\end{aligned}$$

Following the proof of (A.25) and using the continuity condition on $g_k(\cdot)$, we can show that

$$\frac{1}{(\mathbb{T}^2 h_3)^{1/2} \|\boldsymbol{\Sigma}_\varepsilon(\mathbf{k})\|} \sum_{t=1}^{\mathbb{T}} \sum_{s \neq t} \mathbb{K} \left(\frac{\tau_t - \tau_s}{h_3} \right) \boldsymbol{\varepsilon}_t^\top(\mathbf{k}) [\boldsymbol{\omega}_s(\mathbf{k}) - \boldsymbol{\omega}_s^\diamond(\mathbf{k})] = O_P \left(\frac{\varphi(\mathbb{N} \mathbb{T})^{1/2}}{\|\boldsymbol{\Sigma}_\varepsilon(\mathbf{k})\|} \right). \tag{A.30}$$

By Assumption 5(ii), we can prove that

$$\begin{aligned}
&\frac{1}{(\mathbb{T}^2 h_3)^{1/2} \|\boldsymbol{\Sigma}_\varepsilon(\mathbf{k})\|} \sum_{t=1}^{\mathbb{T}} \sum_{s \neq t} \mathbb{K} \left(\frac{\tau_t - \tau_s}{h_3} \right) [\boldsymbol{\omega}_t(\mathbf{k}) - \boldsymbol{\omega}_t^\diamond(\mathbf{k})]^\top [\boldsymbol{\omega}_s(\mathbf{k}) - \boldsymbol{\omega}_s^\diamond(\mathbf{k})] \\
&= \frac{\varphi^2 \phi_k \cdot |\mathcal{G}_k| \mathbb{T}^2 h_3}{(\mathbb{T}^2 h_3)^{1/2} \|\boldsymbol{\Sigma}_\varepsilon(\mathbf{k})\|} (1 + o_P(1)) = \frac{\varphi^2 \phi_k \cdot |\mathcal{G}_k| \mathbb{T} h_3^{1/2}}{\|\boldsymbol{\Sigma}_\varepsilon(\mathbf{k})\|} (1 + o_P(1)). \tag{A.31}
\end{aligned}$$

By (A.20) and (A.29)–(A.31), we complete the proof of Theorem 4.3. \square

Appendix B: Technical lemmas with proofs

We next prove some technical lemmas which have been used to prove the main asymptotic theorems in Appendix A. As in Appendix A, we let C denote a generic positive constant whose value may change from one place to another.

Lemma B.1. *Suppose that the conditions of Proposition 3.1 are satisfied. Then we have*

(i) For any fixed $\tau \in (0, 1)$,

$$\max_{1 \leq i \leq N} \|\Gamma_{i,X}(\tau) - \Delta_{i,T}^*(\tau)\| = O_P \left(\sqrt{\frac{\log N}{Th}} \right), \quad (\text{B.1})$$

and (ii)

$$\max_{1 \leq i \leq N} \sup_{\tau \in [0,1]} \|\Gamma_{i,X}(\tau) - \Delta_{i,T}^*(\tau)\| = O_P \left(\sqrt{\frac{\log(N \vee T)}{Th}} \right), \quad (\text{B.2})$$

where

$$\Delta_{i,T}^*(\tau) = \begin{bmatrix} \frac{1}{T} \sum_{t=1}^T \mathbb{E} \left(\tilde{X}_{i,t-1} \tilde{X}_{i,t-1}^\top \right) K_{t,0}(\tau) & \frac{1}{T} \sum_{t=1}^T \mathbb{E} \left(\tilde{X}_{i,t-1} \tilde{X}_{i,t-1}^\top \right) K_{t,1}(\tau) \\ \frac{1}{T} \sum_{t=1}^T \mathbb{E} \left(\tilde{X}_{i,t-1} \tilde{X}_{i,t-1}^\top \right) K_{t,1}(\tau) & \frac{1}{T} \sum_{t=1}^T \mathbb{E} \left(\tilde{X}_{i,t-1} \tilde{X}_{i,t-1}^\top \right) K_{t,2}(\tau) \end{bmatrix}.$$

Proof of Lemma B.1. We only prove the uniform convergence for $\Gamma_{i,X}(\cdot, 0)$ since the proofs for the other block-matrices $\Gamma_{i,X}(\cdot, 1)$ and $\Gamma_{i,X}(\cdot, 2)$ are very similar. Let

$$\Delta_{i,T}(\tau) = \frac{1}{T} \sum_{t=1}^T \mathbb{E} \left[\tilde{X}_{i,t-1} \tilde{X}_{i,t-1}^\top \right] K_h(\tau_t - \tau),$$

and note that $\hat{\Delta}_i(\tau) = \Gamma_{i,X}(\tau, 0)$. Hence, we only show that for any fixed $\tau \in (0, 1)$,

$$\max_{1 \leq i \leq N} \|\hat{\Delta}_i(\tau) - \Delta_{i,T}(\tau)\| = O_P \left(\sqrt{\frac{\log N}{Th}} \right) \quad (\text{B.3})$$

Noting that

$$\hat{\Delta}_i(\tau) - \Delta_{i,T}(\tau) = \frac{1}{Th} \sum_{t=1}^T \left\{ \tilde{X}_{i,t-1} \tilde{X}_{i,t-1}^\top - \mathbb{E} \left[\tilde{X}_{i,t-1} \tilde{X}_{i,t-1}^\top \right] \right\} K_{t,0}(\tau),$$

to prove (B.3), it is sufficient to show that

$$\max_{1 \leq i \leq N} \left| \hat{\Delta}_i(\tau) - \Delta_{i,T}(\tau) \right|_\infty = O_P \left(\sqrt{\frac{\log N}{Th}} \right). \quad (\text{B.4})$$

Letting $V_{i,t}(\tau) = \tilde{X}_{i,t-1} \tilde{X}_{i,t-1}^\top K_{t,0}(\tau)$ and noting that $K(\cdot)$ has the support $[-1, 1]$,

$$\sum_{t=1}^T V_{i,t}(\tau) = \sum_{t=1}^T \tilde{X}_{i,t-1} \tilde{X}_{i,t-1}^\top K_{t,0}(\tau) = \sum_{t=T_1(\tau)}^{T_2(\tau)} \tilde{X}_{i,t-1} \tilde{X}_{i,t-1}^\top K_{t,0}(\tau),$$

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., [Xu, Chen and Wu, 2020](#); [Zhang and Wu, 2021](#)) to prove (B.4).

Let $M = 2\lfloor Th \rfloor$, $M_\dagger = \lfloor \log M / \log 2 \rfloor$, $u_l = 2^l$ for $1 \leq l \leq M_\dagger - 1$ and $u_{M_\dagger} = M$. Define $V_{i,t,u}(\tau) =$

$\mathbb{E} [V_{i,t}(\tau) | \mathcal{F}_{t-u}^{t-1}]$ with $\mathcal{F}_s^t = (\varepsilon_s, \dots, \varepsilon_t)$, and

$$S_{i,l,T}(\tau) = \sum_{t=T_1(\tau)}^{T_2(\tau)} [V_{i,t,u_l}(\tau) - V_{i,t,u_l-1}(\tau)], \quad l = 1, \dots, M_+.$$

Then we may write

$$\widehat{\Delta}_i(\tau) - \Delta_{i,T}(\tau) = \frac{1}{Th} \left(\sum_{t=T_1(\tau)}^{T_2(\tau)} [V_{i,t}(\tau) - V_{i,t,M}(\tau)] + \sum_{l=2}^{M_+} S_{i,l,T}(\tau) + \sum_{t \ni \phi = T_1(\tau)}^{T_2(\tau)} \{V_{i,t,2}(\tau) - \mathbb{E} [V_{i,t,2}(\tau)]\} \right). \quad (\text{B.5})$$

We next tackle the three terms on the right side of (B.5) separately. Since

$$V_{i,t}(\tau) - V_{i,t,M}(\tau) = \sum_{k=M+1}^{\infty} [V_{i,t,k}(\tau) - V_{i,t,k-1}(\tau)],$$

we have

$$\begin{aligned} \left\| \left\| \sum_{t=T_1(\tau)}^{T_2(\tau)} [V_{i,t}(\tau) - V_{i,t,M}(\tau)] \right\|_{\infty} \right\|_{q/2} &= \left\| \left\| \sum_{t=T_1(\tau)}^{T_2(\tau)} \sum_{k=M+1}^{\infty} [V_{i,t,k}(\tau) - V_{i,t,k-1}(\tau)] \right\|_{\infty} \right\|_{q/2} \\ &\leq C \sum_{k=M+1}^{\infty} \left\| \left\| \sum_{t=T_1(\tau)}^{T_2(\tau)} [V_{i,t,k}(\tau) - V_{i,t,k-1}(\tau)] \right\|_{\infty} \right\|_{q/2}. \end{aligned}$$

Note that, for any $1 \leq i \leq N$ and $k \geq M+1$, $\text{vec}(V_{i,t,k}(\tau) - V_{i,t,k-1}(\tau))$, $T_1(\tau) \leq t \leq T_2(\tau)$, is a sequence of martingale difference vectors. By Lemma D.3 of [Zhang and Wu \(2021\)](#), we have

$$\begin{aligned} &\sum_{k=M+1}^{\infty} \left\| \left\| \sum_{t=T_1(\tau)}^{T_2(\tau)} (V_{i,t,k}(\tau) - V_{i,t,k-1}(\tau)) \right\|_{\infty} \right\|_{q/2} \\ &= \sum_{k=M+1}^{\infty} \left\| \left\| \sum_{t=T_1(\tau)}^{T_2(\tau)} \text{vec}(V_{i,t,k}(\tau) - V_{i,t,k-1}(\tau)) \right\|_{\infty} \right\|_{q/2} \\ &\leq C \sum_{k=M+1}^{\infty} \sqrt{\left\| \left\| \sum_{t=T_1(\tau)}^{T_2(\tau)} \text{vec}(V_{i,t,k}(\tau) - V_{i,t,k-1}(\tau)) \right\|_{\infty} \right\|_{q/2}^2} \\ &= C \sum_{k=M+1}^{\infty} \sqrt{\left\| \left\| \sum_{t=T_1(\tau)}^{T_2(\tau)} V_{i,t,k}(\tau) - V_{i,t,k-1}(\tau) \right\|_{\infty} \right\|_{q/2}^2}. \end{aligned}$$

With the triangle inequality,

$$\begin{aligned} & \left\| \left\| V_{i,t,k}(\tau) - V_{i,t,k-1}(\tau) \right\|_{\infty} \right\|_{q/2} = \left\| \left\| \mathbf{E} \left(V_{i,t}(\tau) - V_{i,t}^{\{t-k\}}(\tau) \middle| \mathcal{F}_{t-k}^{\tau-1} \right) \right\|_{\infty} \right\|_{q/2} \\ & \leq \left\| \left\| \left(\tilde{X}_{i,t-1} - \tilde{X}_{i,t-1}^{\{t-k\}} \right) \tilde{X}_{i,t-1}^{\top} \mathbf{K}_{t,0}(\tau) \right\|_{\infty} \right\|_{q/2} + \left\| \left\| \tilde{X}_{i,t-1}^{\{t-k\}} \left(\tilde{X}_{i,t-1} - \tilde{X}_{i,t-1}^{\{t-k\}} \right)^{\top} \mathbf{K}_{t,0}(\tau) \right\|_{\infty} \right\|_{q/2}, \end{aligned}$$

where $V_{i,t}^{\{l\}}(\tau) = \tilde{X}_{i,t-1}^{\{l\}} \tilde{X}_{i,t-1}^{\{l\}\top} \mathbf{K}_{t,0}(\tau)$ and $\tilde{X}_{i,t}^{\{l\}} = \left(\sum_{j \neq i} \tilde{w}_{ij} x_{j,t}^{\{l\}}, x_{i,t}^{\{l\}} \right)^{\top}$ with $x_{i,t}^{\{l\}}$ being the i -th element of $X_t^{\{l\}}$.

Following the discussion in Section 2.2 and noting that $\sum_{j=1, \neq i}^N \tilde{w}_{ij} = 1$, we have

$$\left\| \left\| \tilde{X}_{i,t-1} - \tilde{X}_{i,t-1}^{\{t-k\}} \right\|_{\infty} \right\|_q \leq \delta_{i,k-1,q} + \sum_{j=1, \neq i}^N \tilde{w}_{ij} \delta_{j,k-1,q},$$

and

$$\begin{aligned} \left\| \left\| \tilde{X}_{i,t-1} \mathbf{K}_{t,0}(\tau) \right\|_{\infty} \right\|_q & \leq \sum_{k=0}^{\infty} \left\| \left\| \left[\mathbf{E}(\tilde{X}_{i,t-1} | \mathcal{F}_{t-k-1}) - \mathbf{E}(\tilde{X}_{i,t-1} | \mathcal{F}_{t-k-2}) \right] \mathbf{K}_{t,0}(\tau) \right\|_{\infty} \right\|_q \\ & \leq C \sum_{k=0}^{\infty} \left(\delta_{i,k,q} + \sum_{j=1, \neq i}^N \tilde{w}_{ij} \delta_{j,k,q} \right) \\ & \leq C \max_{1 \leq i \leq N} \|x_{i \bullet}\|_{q,\alpha} \end{aligned}$$

which is bounded. Then, by Jensen's inequality, applying the Cauchy-Schwarz inequality, we have

$$\begin{aligned} \left\| \left\| \left(\tilde{X}_{i,t-1} - \tilde{X}_{i,t-1}^{\{t-k\}} \right) \tilde{X}_{i,t-1}^{\top} \mathbf{K}_{t,0}(\tau) \right\|_{\infty} \right\|_{q/2} & \leq \left\| \left\| \tilde{X}_{i,t-1} - \tilde{X}_{i,t-1}^{\{t-k\}} \right\|_{\infty} \right\|_q \cdot \left\| \left\| \tilde{X}_{i,t-1} \mathbf{K}_{t,0}(\tau) \right\|_{\infty} \right\|_q \\ & \leq C \left(\delta_{i,k-1,q} + \sum_{j=1, \neq i}^N \tilde{w}_{ij} \delta_{j,k-1,q} \right). \end{aligned}$$

Similarly we can also prove that

$$\left\| \left\| \left(\tilde{X}_{i,t-1} - \tilde{X}_{i,t-1}^{\{t-k\}} \right) \left(\tilde{X}_{i,t-1}^{\{t-k\}} \right)^{\top} \mathbf{K}_{t,0}(\tau) \right\|_{\infty} \right\|_{q/2} \leq C \left(\delta_{i,k-1,q} + \sum_{j=1, \neq i}^N \tilde{w}_{ij} \delta_{j,k-1,q} \right).$$

Since $M = 2\lceil Th \rceil$, we have

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

$$\begin{aligned}
&\leq C(\text{Th})^{1/2} \sum_{k=M+1}^{\infty} \left(\delta_{i,k-1,q} + \sum_{j=1, \neq i}^N \tilde{w}_{ij} \delta_{j,k-1,q} \right) \\
&= C(\text{Th})^{1/2} \left(\Delta_{i,M,q} + \sum_{j=1, \neq i}^N \tilde{w}_{ij} \Delta_{j,M,q} \right) \\
&\leq C(\text{Th})^{1/2-\alpha} \max_{1 \leq i \leq N} \|x_{i\bullet}\|_{q,\alpha}, \tag{B.6}
\end{aligned}$$

where the last inequality is due to the fact that

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

By (B.6) and Markov's inequality, we have for any $z > 0$,

$$\mathbb{P} \left(\left| \sum_{t=T_1(\tau)}^{T_2(\tau)} (V_{i,t}(\tau) - V_{i,t,M}(\tau)) \right|_{\infty} > z \right) \leq \frac{C(\text{Th})^{(1-2\alpha)q/4}}{z^{q/2}} \max_{1 \leq i \leq N} \|x_{i\bullet}\|_{q,\alpha}^{q/2}. \tag{B.7}$$

For each $2 \leq l \leq M_{\dagger}$, define

$$U_{i,k,l}(\tau) = \sum_{t=(k-1)u_l+T_1(\tau)}^{(ku_l) \wedge T_2(\tau)} [V_{i,t,u_l}(\tau) - V_{i,t,u_l-1}(\tau)], \quad 1 \leq k \leq \lceil M/u_l \rceil,$$

and

$$S_{i,l,T}^e(\tau) = \sum_{k \text{ is even}} U_{i,k,l}(\tau), \quad S_{i,l,T}^o(\tau) = \sum_{k \text{ is odd}} U_{i,k,l}(\tau).$$

Let $\lambda_l = (l-1)^{-2}/(\pi^2/3)$ if $2 \leq l \leq M_{\dagger}/2$ and $\lambda_l = (M_{\dagger}+1-l)^{-2}/(\pi^2/3)$ if $M_{\dagger}/2 < l < M_{\dagger}$. It is easy to verify that $\sum_{l=2}^{M_{\dagger}} \lambda_l \leq 1$. Since $U_{i,k,l}(\tau)$ and $U_{i,k',l}(\tau)$ are independent for $|k-k'| > 1$, by Lemma D.4 of [Zhang and Wu \(2021\)](#), we have for any $z > 0$ and each $2 \leq l \leq M_{\dagger}$,

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

Similarly to (B.6), we may show that

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

Similar results also hold for $|S_{i,l,T}^o(\tau)|_{\infty}$. A combination of these arguments yields that

$$\mathbb{P} \left(\left| \sum_{l=2}^{M_{\dagger}} S_{i,l,T}(\tau) \right|_{\infty} \geq 2z \right)$$

$$\begin{aligned}
&\leq \sum_{l=2}^{M_{\dagger}} \mathbf{P}(|S_{i,l,T}(\tau)|_{\infty} \geq 2\lambda_l z) \\
&\leq \sum_{l=2}^{M_{\dagger}} \mathbf{P}(|S_{i,l,T}^e(\tau)|_{\infty} \geq \lambda_l z) + \sum_{l=2}^{M_{\dagger}} \mathbf{P}(|S_{i,l,T}^o(\tau)|_{\infty} \geq \lambda_l z) \\
&\leq C \left\{ \frac{\text{Th}}{z^{q/2}} \cdot \sum_{l=2}^{M_{\dagger}} \frac{u_l^{(1-2\alpha)q/4-1}}{\lambda_l^{q/2}} \left(\max_{1 \leq i \leq N} \|x_{i\bullet}\|_{q,\alpha}^{q/2} \right) + \sum_{l=2}^{M_{\dagger}} \exp \left(-\frac{\lambda_l^2 u_l^{2\alpha} z^2}{C(\text{Th}) \max_{1 \leq i \leq N} \|x_{i\bullet}\|_{4,\alpha}^2} \right) \right\}.
\end{aligned}$$

Noting that $\sum_{l=2}^{M_{\dagger}} \lambda_l^{-q/2} u_l^{(1-2\alpha)q/4-1}$ is bounded (since $\alpha < 1/2 - 2/q$) and $\min_{l \geq 1} \lambda_l^2 u_l^{2\alpha} > 0$, we have

$$\mathbf{P} \left(\left| \sum_{l=2}^{M_{\dagger}} S_{i,l,T}(\tau) \right|_{\infty} \geq 2z \right) \leq C \left(\frac{\text{Th}}{z^{q/2}} \left(\max_{1 \leq i \leq N} \|x_{i\bullet}\|_{q,\alpha}^{q/2} \right) + \exp \left\{ -\frac{z^2}{C(\text{Th})} \left(\max_{1 \leq i \leq N} \|x_{i\bullet}\|_{4,\alpha}^2 \right)^{-1} \right\} \right). \quad (\text{B.8})$$

We next consider $\sum_{t=T_1(\tau)}^{T_2(\tau)} \{V_{i,t,2}(\tau) - \mathbf{E}[V_{i,t,2}(\tau)]\}$. Note that

$$V_{i,t,2}(\tau) - \mathbf{E}[V_{i,t,2}(\tau)] = \sum_{s=-\infty}^t \{ \mathbf{E}[V_{i,t,2}(\tau) | \mathcal{F}_{s-1}] - \mathbf{E}[V_{i,t,2}(\tau) | \mathcal{F}_{s-2}] \}.$$

For any s , by Jensen's inequality and the triangle inequality, we have

$$\begin{aligned}
&\left\| \mathbf{E}[V_{i,t,2}(\tau) | \mathcal{F}_{s-1}] - \mathbf{E}[V_{i,t,2}(\tau) | \mathcal{F}_{s-2}] \right\|_{\infty, q/2} \\
&\leq \left\| \left(\tilde{X}_{i,t-1} - \tilde{X}_{i,t-1}^{[s-1]} \right) \tilde{X}_{i,t-1}^{\top} \mathbf{K}_{t,0}(\tau) \right\|_{\infty, q/2} + \left\| \tilde{X}_{i,t-1, \{r-1\}}^{[s-1]} \left(\tilde{X}_{i,t-1} - \tilde{X}_{i,t-1}^{[s-1]} \right)^{\top} \mathbf{K}_{t,0}(\tau) \right\|_{\infty, q/2} \\
&\leq C \left(\delta_{i,t-s,q} + \sum_{j=1, \neq i}^N \tilde{w}_{ij} \delta_{j,t-s,q} \right).
\end{aligned}$$

Consequently,

$$\left\| V_{i,t,2}(\tau) - \mathbf{E}[V_{i,t,2}(\tau)] \right\|_{\infty, q/2} \leq C \sum_{s=-\infty}^t \left(\delta_{i,t-s,q} + \sum_{j=1, \neq i}^N \tilde{w}_{ij} \delta_{j,t-s,q} \right) \leq C \max_{1 \leq i \leq N} \Delta_{i,0,q},$$

and

$$\sum_{t=T_1(\tau)}^{T_2(\tau)} \mathbf{E} \left| V_{i,t,2}(\tau) - \mathbf{E}[V_{i,t,2}(\tau)] \right|_{\infty}^{q/2} \leq C(\text{Th}) \max_{1 \leq i \leq N} \|x_{i\bullet}\|_{q,\alpha}^{q/2}. \quad (\text{B.9})$$

By the independence of $\varepsilon_{i,t}$ over t , $V_{i,t,2}(\tau)$ and $V_{i,t',2}(\tau)$ are independent for $|t - t'| > 2$. Following the

proof of (B.8) and using (B.9), we have

$$\begin{aligned} & \mathbb{P} \left(\left| \sum_{t=T_1(\tau)}^{T_2(\tau)} \{V_{i,t,2}(\tau) - \mathbb{E}[V_{i,t,2}(\tau)]\} \right|_{\infty} \geq z \right) \\ & \leq C \left(\frac{\text{Th}}{z^{q/2}} \left(\max_{1 \leq i \leq N} \|x_{i\bullet}\|_{q,\alpha}^{q/2} \right) + \exp \left\{ -\frac{z^2}{C(\text{Th})} \left(\max_{1 \leq i \leq N} \|x_{i\bullet}\|_{4,\alpha}^2 \right)^{-1} \right\} \right). \end{aligned} \quad (\text{B.10})$$

A combination of (B.5), (B.7), (B.8) and (B.10) yields that

$$\mathbb{P} \left(\left| \widehat{\Delta}_i(\tau) - \Delta_{i,T}(\tau) \right|_{\infty} > \frac{z}{\text{Th}} \right) \leq C \left(\frac{\text{Th}}{z^{q/2}} \left(\max_{1 \leq i \leq N} \|x_{i\bullet}\|_{q,\alpha}^{q/2} \right) + \exp \left\{ -\frac{z^2}{C(\text{Th})} \left(\max_{1 \leq i \leq N} \|x_{i\bullet}\|_{4,\alpha}^2 \right)^{-1} \right\} \right). \quad (\text{B.11})$$

By the discussion in Section 2.2, $\max_{1 \leq i \leq N} (\|x_{i\bullet}\|_{q,\alpha} + \|x_{i\bullet}\|_{4,\alpha})$ is bounded. Setting $z = \eta \sqrt{\text{Th} \log N}$ in (B.11), we may show that

$$\begin{aligned} & \mathbb{P} \left(\max_{1 \leq i \leq N} \left| \widehat{\Delta}_i(\tau) - \Delta_{i,T}(\tau) \right|_{\infty} > \eta \sqrt{\frac{\log N}{\text{Th}}} \right) \\ & \leq \sum_{i=1}^N \mathbb{P} \left(\left| \widehat{\Delta}_i(\tau) - \Delta_{i,T}(\tau) \right|_{\infty} > \eta \sqrt{\frac{\log N}{\text{Th}}} \right) \\ & \leq CN \left\{ \exp \left(-\frac{\eta^2 \text{Th} \log N}{C\text{Th}} \right) + \frac{C\text{Th}}{(\eta \sqrt{\text{Th} \log N})^{q/2}} \right\} \\ & = O(N \exp \{ -(\eta^2/C) \log N \}) + O \left(\frac{N}{\eta^{q/2} (\text{Th})^{q/4-1} (\log N)^{q/4}} \right). \end{aligned} \quad (\text{B.12})$$

Choosing η to be sufficiently large and using Assumption 2(iii), we have

$$\mathbb{P} \left(\max_{1 \leq i \leq N} \left| \widehat{\Delta}_i(\tau) - \Delta_{i,T}(\tau) \right|_{\infty} > \eta \sqrt{\frac{\log N}{\text{Th}}} \right) \rightarrow 0.$$

This completes the proof of (B.4).

We next turn to the proof of Lemma B.1(ii). As in the proof of Lemma B.1(i), we only prove that

$$\max_{1 \leq i \leq N} \sup_{\tau \in [0,1]} \left| \widehat{\Delta}_i(\tau) - \Delta_{i,T}(\tau) \right|_{\infty} = O_{\mathbb{P}} \left(\sqrt{\frac{\log(N \vee T)}{\text{Th}}} \right). \quad (\text{B.13})$$

Let $\xi_{NT} = T^{(q-2)/[2(q+2)]} [h \log(N \vee T)]^{1/2}$. We consider covering the closed interval $[0, 1]$ by some disjoint sub-intervals \mathcal{J}_b , $1 \leq b \leq B$, with centres τ_b and length $\gamma_{NT} = h^2 \sqrt{\log(N \vee T)/\text{Th}}/\xi_{NT}$. The number B is upper bounded by $\xi_{NT}/(h^2 \sqrt{\log(N \vee T)/\text{Th}})$. Then we have

$$\max_{1 \leq i \leq N} \sup_{\tau \in [0,1]} \left| \widehat{\Delta}_i(\tau) - \Delta_{i,T}(\tau) \right|_{\infty}$$

$$\begin{aligned} &\leq \max_{1 \leq i \leq N} \max_{1 \leq b \leq B} \left| \frac{1}{Th} \sum_{t=1}^T \left\{ \tilde{X}_{i,t-1} \tilde{X}_{i,t-1}^\top - \mathbf{E} \left[\tilde{X}_{i,t-1} \tilde{X}_{i,t-1}^\top \right] \right\} \mathbf{K}_{t,0}(\tau_b) \right|_\infty + \\ &\quad \max_{1 \leq i \leq N} \max_{1 \leq b \leq B} \sup_{\tau \in \mathcal{J}_b} \left| \frac{1}{Th} \sum_{t=1}^T \left\{ \tilde{X}_{i,t-1} \tilde{X}_{i,t-1}^\top - \mathbf{E} \left[\tilde{X}_{i,t-1} \tilde{X}_{i,t-1}^\top \right] \right\} [\mathbf{K}_{t,0}(\tau) - \mathbf{K}_{t,0}(\tau_b)] \right|_\infty. \end{aligned}$$

We first prove that

$$\max_{1 \leq i \leq N} \max_{1 \leq b \leq B} \sup_{\tau \in \mathcal{J}_b} \left| \frac{1}{Th} \sum_{t=1}^T \left\{ \tilde{X}_{i,t-1} \tilde{X}_{i,t-1}^\top - \mathbf{E} \left[\tilde{X}_{i,t-1} \tilde{X}_{i,t-1}^\top \right] \right\} [\mathbf{K}_{t,0}(\tau) - \mathbf{K}_{t,0}(\tau_b)] \right|_\infty = O_P \left(\sqrt{\frac{\log(N \vee T)}{Th}} \right). \quad (\text{B.14})$$

By Assumption 2(i), we readily have that

$$\max_{1 \leq b \leq B} \sup_{\tau \in \mathcal{J}_b} |\mathbf{K}_{t,0}(\tau) - \mathbf{K}_{t,0}(\tau_b)| = O_P(\gamma_{NT} h^{-2}).$$

Define $V_{i,t}^X = \tilde{X}_{i,t-1} \tilde{X}_{i,t-1}^\top$,

$$\bar{V}_{i,t}^X = \tilde{X}_{i,t-1} \tilde{X}_{i,t-1}^\top \mathbf{I} \left\{ \|\tilde{X}_{i,t-1} \tilde{X}_{i,t-1}^\top\| \leq \xi_{NT} \right\} \quad \text{and} \quad \tilde{V}_{i,t}^X = V_{i,t}^X - \bar{V}_{i,t}^X.$$

Then, we may show that

$$\begin{aligned} &\max_{1 \leq i \leq N} \max_{1 \leq b \leq B} \sup_{\tau \in \mathcal{J}_b} \left| \frac{1}{Th} \sum_{t=1}^T \left\{ \tilde{X}_{i,t-1} \tilde{X}_{i,t-1}^\top - \mathbf{E} \left[\tilde{X}_{i,t-1} \tilde{X}_{i,t-1}^\top \right] \right\} [\mathbf{K}_{t,0}(\tau) - \mathbf{K}_{t,0}(\tau_b)] \right|_\infty \\ &\leq \gamma_{NT} h^{-2} \left\{ \max_{1 \leq i \leq N} \frac{1}{T} \sum_{t=1}^T \left| \bar{V}_{i,t}^X - \mathbf{E}[\bar{V}_{i,t}^X] \right|_\infty + \max_{1 \leq i \leq N} \frac{1}{T} \sum_{t=1}^T \left| \tilde{V}_{i,t}^X - \mathbf{E}[\tilde{V}_{i,t}^X] \right|_\infty \right\} \\ &= O_P \left(\sqrt{\frac{\log(N \vee T)}{Th}} \right) + \gamma_{NT} h^{-2} \max_{1 \leq i \leq N} \frac{1}{T} \sum_{t=1}^T \left| \tilde{V}_{i,t}^X - \mathbf{E}[\tilde{V}_{i,t}^X] \right|_\infty, \end{aligned} \quad (\text{B.15})$$

and, for any $\eta > 0$,

$$\begin{aligned} &\mathbf{P} \left(\max_{1 \leq i \leq N} \frac{1}{T} \sum_{t=1}^T \left| \tilde{V}_{i,t}^X - \mathbf{E}[\tilde{V}_{i,t}^X] \right|_\infty > \eta \xi_{NT} \right) \\ &\leq \sum_{i=1}^N \sum_{t=1}^T \mathbf{P} \left(\left\| \tilde{X}_{i,t-1} \tilde{X}_{i,t-1}^\top \right\| > \xi_{NT} \right) \\ &= O \left(NT \xi_{NT}^{-q/2} \right) = o(1) \end{aligned} \quad (\text{B.16})$$

by Assumptions 1(iii) and 2(iii). With (B.15) and (B.16), we prove (B.14).

It remains to show that

$$\max_{1 \leq i \leq N} \max_{1 \leq b \leq B} \left| \widehat{\Delta}_i(\tau_b) - \Delta_{i,T}(\tau_b) \right|_{\infty} = O_P \left(\sqrt{\frac{\log(N \vee T)}{Th}} \right). \quad (\text{B.17})$$

Following the same arguments as Lemma B.1(i), we have for any $\eta > 0$,

$$\begin{aligned} & \mathbb{P} \left(\max_{1 \leq i \leq N} \max_{1 \leq b \leq B} \left| \widehat{\Delta}_i(\tau_b) - \Delta_{i,T}(\tau_b) \right|_{\infty} > \eta \sqrt{\frac{\log(N \vee T)}{Th}} \right) \\ & \leq \sum_{i=1}^N \sum_{b=1}^B \mathbb{P} \left(\left| \widehat{\Delta}_i(\tau_b) - \Delta_{i,T}(\tau_b) \right|_{\infty} > \eta \sqrt{\frac{\log(N \vee T)}{Th}} \right) \\ & = O \left(NB \exp \{ -C\eta^2 \log(N \vee T) \} \right) + O \left(\frac{NB}{\eta^{q/2} (Th)^{q/4-1} (\log(N \vee T))^{q/4}} \right). \end{aligned} \quad (\text{B.18})$$

Noting that B diverge at a certain polynomial rate of T and letting η be sufficiently large, the first order on the right side of (B.18) converges to zero. By Assumption 2(iii), we may also show that the second order also converges to zero. Hence, we complete the proof of (B.17). By virtue of (B.14) and (B.17), we prove (B.13). The proof of Lemma B.1 is completed. \square

Lemma B.2. *Suppose that the conditions of Proposition 3.1 are satisfied. Then we have*

(i) *For any fixed $\tau \in (0, 1)$,*

$$\max_{1 \leq i, j \leq N} \left\| \widehat{\Delta}_{ij}(\tau) - \Delta_{ij,T}(\tau) \right\| = O_P \left(\sqrt{\frac{\log N}{Th}} \right), \quad (\text{B.19})$$

and (ii)

$$\max_{1 \leq i, j \leq N} \sup_{\tau \in [0, 1]} \left\| \widehat{\Delta}_{ij}(\tau) - \Delta_{ij,T}(\tau) \right\| = O_P \left(\sqrt{\frac{\log(N \vee T)}{Th}} \right), \quad (\text{B.20})$$

where $\widehat{\Delta}_{ij}(\tau)$ and $\Delta_{ij,T}(\tau)$ are defined in Section 3.1.1.

Proof of Lemma B.2. The proofs of (B.19) and (B.20) are similar to the proofs of (B.4) and (B.13), respectively. Details are omitted here. \square

Lemma B.3. *Suppose that the conditions of Theorem 3.2 are satisfied. Then we have*

$$\max_{1 \leq i \leq N} \max_{1 \leq k \leq L} \left| \widehat{\beta}_i(\tau_k^*) - \beta_i(\tau_k^*) \right| = O_P \left(\sqrt{\frac{\log N}{Th}} + h^2 \right), \quad (\text{B.21})$$

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

Proof of Lemma B.3. Letting $\Gamma_{i,X}(\cdot)$, $\Gamma_{i,\beta}(\cdot)$ and $\Gamma_{i,\varepsilon}(\cdot)$ be defined as in the proof of Proposition 3.1,

$$\widehat{\beta}_i(\tau_k^*) - \beta_i(\tau_k^*) = \mathbf{E}_* \cdot \Gamma_{i,X}^{-1}(\tau_k^*) \left[\Gamma_{i,\beta}(\tau_k^*) + \Gamma_{i,\varepsilon}(\tau_k^*) \right] \quad (\text{B.22})$$

as in (A.1). Following the proof of Lemma B.1(ii), we may show that

$$\max_{1 \leq i \leq N} \max_{1 \leq k \leq L} \left\| \Gamma_{i,X}^{-1}(\tau_k^*) \right\| = O_P(1), \quad \max_{1 \leq i \leq N} \max_{1 \leq k \leq L} |\Gamma_{i,\beta}(\tau_k^*)| = O_P(h^2). \quad (\text{B.23})$$

In view of (B.22) and (B.23), in order to prove (B.21), we only need to show that

$$\max_{1 \leq i \leq N} \max_{1 \leq k \leq L} |\Gamma_{i,\varepsilon}(\tau_k^*)| = O_P\left(\sqrt{\frac{\log(N \vee T)}{Th}}\right).$$

To save the space, we only provide the proof of

$$\max_{1 \leq i \leq N} \max_{1 \leq k \leq L} \left| \frac{1}{Th} \sum_{t=1}^T \tilde{X}_{i,t-1} \varepsilon_{i,t} K_{t,0}(\tau_k^*) \right| = O_P\left(\sqrt{\frac{\log(N \vee T)}{Th}}\right). \quad (\text{B.24})$$

We next use the truncation technique and Freedman (1975)'s concentration inequality for martingale to prove (B.24). Let $\xi_{NT} = T^{(q-2)/[2(q+2)]} [h \log(N \vee T)]^{1/2}$ be defined as in the proof of Lemma B.1(ii), $W_{i,t}^{X,\varepsilon} = \tilde{X}_{i,t-1} \varepsilon_{i,t}$,

$$\bar{W}_{i,t}^{X,\varepsilon} = \tilde{X}_{i,t-1} \varepsilon_{i,t} \mathbf{I} \left\{ |\tilde{X}_{i,t-1}| \leq \xi_{NT}^{1/2}, |\varepsilon_{i,t}| \leq \xi_{NT}^{1/2} \right\} \quad \text{and} \quad \widetilde{W}_{i,t}^{X,\varepsilon} = W_{i,t}^{X,\varepsilon} - \bar{W}_{i,t}^{X,\varepsilon}.$$

Note that

$$\begin{aligned} & \max_{1 \leq i \leq N} \max_{1 \leq k \leq L} \left| \frac{1}{Th} \sum_{t=1}^T \tilde{X}_{i,t-1} \varepsilon_{i,t} K_{t,0}(\tau_k^*) \right| \\ &= \max_{1 \leq i \leq N} \max_{1 \leq k \leq L} \left| \frac{1}{Th} \sum_{t=1}^T \left[\bar{W}_{i,t}^{X,\varepsilon} - \mathbb{E} \left(\bar{W}_{i,t}^{X,\varepsilon} \right) \right] K_{t,0}(\tau_k^*) \right| + \\ & \quad \max_{1 \leq i \leq N} \max_{1 \leq k \leq L} \left| \frac{1}{Th} \sum_{t=1}^T \left[\widetilde{W}_{i,t}^{X,\varepsilon} - \mathbb{E} \left(\widetilde{W}_{i,t}^{X,\varepsilon} \right) \right] K_{t,0}(\tau_k^*) \right|. \end{aligned}$$

For any $\eta > 0$, by Markov's inequality, Assumption 2(iii) and $L = O(T)$,

$$\begin{aligned} & \mathbb{P} \left(\max_{1 \leq i \leq N} \max_{1 \leq k \leq L} \left| \frac{1}{Th} \sum_{t=1}^T \left[\widetilde{W}_{i,t}^{X,\varepsilon} - \mathbb{E} \left(\widetilde{W}_{i,t}^{X,\varepsilon} \right) \right] K_{t,0}(\tau_k^*) \right| > \eta \sqrt{\frac{\log(N \vee T)}{Th}} \right) \\ & \leq \mathbb{P} \left(\max_{1 \leq i \leq N} \max_{1 \leq k \leq L} |\tilde{X}_{i,t-1}| > \xi_{NT}^{1/2} \right) + \mathbb{P} \left(\max_{1 \leq i \leq N} \max_{1 \leq k \leq L} |\varepsilon_{i,t}| > \xi_{NT}^{1/2} \right) \\ & \leq \sum_{i=1}^N \sum_{k=1}^L \mathbb{P} \left(|\tilde{X}_{i,t-1}| > \xi_{NT}^{1/2} \right) + \sum_{i=1}^N \sum_{k=1}^L \mathbb{P} \left(|\varepsilon_{i,t}| > \xi_{NT}^{1/2} \right) \\ & = O \left(NL \xi_{NT}^{-q/2} \right) = o(1). \end{aligned} \quad (\text{B.25})$$

On the other hand, by [Freedman \(1975\)](#)'s concentration inequality for martingale, we have

$$\begin{aligned}
& \mathbb{P} \left(\max_{1 \leq i \leq N} \max_{1 \leq k \leq L} \left| \frac{1}{T\mathfrak{h}} \sum_{t=1}^T [\overline{W}_{i,t}^{X,\varepsilon} - \mathbb{E}(\overline{W}_{i,t}^{X,\varepsilon})] \mathbb{K}_{t,0}(\tau_k^*) \right| > \eta \sqrt{\frac{\log(N \vee T)}{T\mathfrak{h}}} \right) \\
& \leq \sum_{i=1}^N \sum_{k=1}^L \mathbb{P} \left(\left| \frac{1}{T\mathfrak{h}} \sum_{t=1}^T [\overline{W}_{i,t}^{X,\varepsilon} - \mathbb{E}(\overline{W}_{i,t}^{X,\varepsilon})] \mathbb{K}_{t,0}(\tau_k^*) \right| > \eta \sqrt{\frac{\log(N \vee T)}{T\mathfrak{h}}} \right) \\
& = O \left(N L \exp \left\{ -\frac{\eta^2 (T\mathfrak{h}) \log(N \vee T)}{C T \mathfrak{h}} \right\} \right) = o(1), \tag{B.26}
\end{aligned}$$

by letting η be sufficiently large. With [\(B.25\)](#) and [\(B.26\)](#), we complete the proof of [\(B.24\)](#). \square

Lemma B.4. *Suppose that the conditions of [Theorem 3.3](#) are satisfied. Then we have*

$$\mathbb{P}(\mathbb{IC}(K_0) < \mathbb{IC}(K), 1 \leq K \leq K_0 - 1) \rightarrow 1 \tag{B.27}$$

conditional on the event $\mathcal{E}(\mathcal{G})$.

Proof of Lemma B.4. Without loss of generality, we only consider $K = K_0 - 1$ because the other scenarios can be dealt with in the same manner. Conditional on $\mathcal{E}(\mathcal{G})$, two clusters from $\mathcal{G}_1, \dots, \mathcal{G}_{K_0}$ are falsely merged when the clustering algorithm terminates at $K_0 - 1$. Without loss of generality we assume that \mathcal{G}_{K_0-1} and \mathcal{G}_{K_0} are falsely merged, i.e., $\widehat{\mathcal{G}}_k = \mathcal{G}_k, k = 1, \dots, K_0 - 2$ and $\widehat{\mathcal{G}}_{K_0-1} = \mathcal{G}_{K_0-1}^* := \mathcal{G}_{K_0-1} \cup \mathcal{G}_{K_0}$. Let $\alpha_{k|K_0-1,1}(\cdot) = \alpha_{k,1}(\cdot), \alpha_{k|K_0-1,2}(\cdot) = \alpha_{k,2}(\cdot), k = 1, \dots, K_0 - 2$, and write $\alpha_{k|K_0-1}(\cdot) = [\alpha_{k|K_0-1,1}(\cdot), \alpha_{k|K_0-1,2}(\cdot)]^\top$. Let $\alpha_{K_0-1}^*(\cdot) = [\alpha_{K_0-1,1}^*(\cdot), \alpha_{K_0-1,2}^*(\cdot)]^\top$ be the vector of ‘‘pseudo’’ coefficient functions associated with $\mathcal{G}_{K_0-1}^*$.

For $i \in \mathcal{G}_k, k = 1, \dots, K_0 - 2$, following the proof of [Lemma B.3](#), we may show that

$$\max_{1 \leq k \leq K_0-2} \max_{1 \leq t \leq T} |\widehat{\alpha}_{k|K_0-1}(\tau_t) - \alpha_{k|K_0-1}(\tau_t)| = O_{\mathbb{P}} \left(\mathfrak{h}_1^2 + \sqrt{\frac{\log(N \vee T)}{T\mathfrak{h}_1}} \right),$$

indicating that

$$\frac{1}{NT} \sum_{k=1}^{K_0-2} \sum_{i \in \mathcal{G}_k} \sum_{t=1}^T \left[x_{i,t} - \widehat{\alpha}_{k|K_0-1}^\top(\tau_t) \widetilde{X}_{i,t-1} \right]^2 = \frac{1}{NT} \sum_{k=1}^{K_0-2} \sum_{i \in \mathcal{G}_k} \sum_{t=1}^T \varepsilon_{i,t}^2 + o_{\mathbb{P}}(1). \tag{B.28}$$

Conditional on $\mathcal{E}(\mathcal{G})$, for $i \in \mathcal{G}_{K_0-1}$, we write

$$x_{i,t} - \widehat{\alpha}_{K_0-1|K_0-1}^\top(\tau_t) \widetilde{X}_{i,t-1} = \varepsilon_{i,t} - [\widehat{\alpha}_{K_0-1|K_0-1}(\tau_t) - \alpha_{K_0-1}^*(\tau_t)]^\top \widetilde{X}_{i,t-1} \tag{B.29}$$

and

$$\widehat{\alpha}_{K_0-1|K_0-1}(\tau) - \alpha_{K_0-1}^*(\tau) = \mathbf{E}_* \cdot \Gamma_{*,X}^{-1}(\tau) [\Gamma_{*,\beta}(\tau) + \Gamma_{*,\varepsilon}(\tau)],$$

where \mathbf{E}_* is defined as in (A.1),

$$\begin{aligned}\Gamma_{*,X}(\tau) &= \begin{bmatrix} \frac{1}{\mathbb{T}h_1} \sum_{i \in \mathcal{G}_{K_0-1}^*} \sum_{t=1}^{\mathbb{T}} \tilde{X}_{i,t-1} \tilde{X}_{i,t-1}^\top \mathbf{K}_{t,0}^\dagger(\tau) & \frac{1}{\mathbb{T}h_1} \sum_{i \in \mathcal{G}_{K_0-1}^*} \sum_{t=1}^{\mathbb{T}} \tilde{X}_{i,t-1} \tilde{X}_{i,t-1}^\top \mathbf{K}_{t,1}^\dagger(\tau) \\ \frac{1}{\mathbb{T}h_1} \sum_{i \in \mathcal{G}_{K_0-1}^*} \sum_{t=1}^{\mathbb{T}} \tilde{X}_{i,t-1} \tilde{X}_{i,t-1}^\top \mathbf{K}_{t,1}^\dagger(\tau) & \frac{1}{\mathbb{T}h_1} \sum_{i \in \mathcal{G}_{K_0-1}^*} \sum_{t=1}^{\mathbb{T}} \tilde{X}_{i,t-1} \tilde{X}_{i,t-1}^\top \mathbf{K}_{t,2}^\dagger(\tau) \end{bmatrix}, \\ \Gamma_{*,\beta}(\tau) &= \begin{bmatrix} \frac{1}{\mathbb{T}h_1} \sum_{i \in \mathcal{G}_{K_0-1}^*} \sum_{t=1}^{\mathbb{T}} \tilde{X}_{i,t-1} \tilde{X}_{i,t-1}^\top [\beta_i(\tau_t) - \beta_i(\tau) - \beta'_i(\tau)(\tau_t - \tau)] \mathbf{K}_{t,0}^\dagger(\tau) \\ \frac{1}{\mathbb{T}h_1} \sum_{i \in \mathcal{G}_{K_0-1}^*} \sum_{t=1}^{\mathbb{T}} \tilde{X}_{i,t-1} \tilde{X}_{i,t-1}^\top [\beta_i(\tau_t) - \beta_i(\tau) - \beta'_i(\tau)(\tau_t - \tau)] \mathbf{K}_{t,1}^\dagger(\tau) \end{bmatrix}, \\ \Gamma_{*,\varepsilon}(\tau) &= \begin{bmatrix} \frac{1}{\mathbb{T}h_1} \sum_{i \in \mathcal{G}_{K_0-1}^*} \sum_{t=1}^{\mathbb{T}} \tilde{X}_{i,t-1} \varepsilon_{i,t} \mathbf{K}_{t,0}^\dagger(\tau) \\ \frac{1}{\mathbb{T}h_1} \sum_{i \in \mathcal{G}_{K_0-1}^*} \sum_{t=1}^{\mathbb{T}} \tilde{X}_{i,t-1} \varepsilon_{i,t} \mathbf{K}_{t,1}^\dagger(\tau) \end{bmatrix},\end{aligned}$$

$\mathbf{K}_{t,j}^\dagger(\tau) = \left(\frac{\tau_t - \tau}{h_1}\right)^j \mathbf{K}\left(\frac{\tau_t - \tau}{h_1}\right)$. Letting $\Delta_{\mathcal{G}_k}(\tau) = \frac{1}{|\mathcal{G}_k|} \sum_{i \in \mathcal{G}_k} \Delta_i(\tau)$, $1 \leq k \leq K_0$, we may explicitly write the "pseudo" time-varying coefficient vector as

$$\alpha_{K_0-1}^*(\tau) = \left[|\mathcal{G}_{K_0-1}| \Delta_{\mathcal{G}_{K_0-1}}(\tau) + |\mathcal{G}_{K_0}| \Delta_{\mathcal{G}_{K_0}}(\tau) \right]^{-1} \left[|\mathcal{G}_{K_0-1}| \Delta_{\mathcal{G}_{K_0-1}}(\tau) \alpha_{K_0-1}(\tau) + |\mathcal{G}_{K_0}| \Delta_{\mathcal{G}_{K_0}}(\tau) \alpha_{K_0}(\tau) \right],$$

a weighted average of $\alpha_{K_0-1}(\tau)$ and $\alpha_{K_0}(\tau)$. Following the proof of Lemma B.3, we may show that

$$\max_{1 \leq t \leq \mathbb{T}} |\hat{\alpha}_{K_0-1|K_0-1}(\tau_t) - \alpha_{K_0-1}^*(\tau_t)| = O_P \left(h_1^2 + \sqrt{\frac{\log(N \vee \mathbb{T})}{\mathbb{T}h_1}} \right). \quad (\text{B.30})$$

By (B.29) and (B.30), letting $d_{\alpha,1}(\tau) = \alpha_{K_0-1}(\tau) - \alpha_{K_0-1}^*(\tau)$, we have

$$\begin{aligned}& \frac{1}{N\mathbb{T}} \sum_{i \in \mathcal{G}_{K_0-1}} \sum_{t=1}^{\mathbb{T}} \left[x_{i,t}^\dagger - \hat{\alpha}_{K_0-1|K_0-1}^\top(\tau_t) \tilde{X}_{i,t-1} \right]^2 \\ &= \frac{1}{N\mathbb{T}} \sum_{i \in \mathcal{G}_{K_0-1}} \sum_{t=1}^{\mathbb{T}} \varepsilon_{i,t}^2 + \frac{1}{N\mathbb{T}} \sum_{i \in \mathcal{G}_{K_0-1}} \sum_{t=1}^{\mathbb{T}} d_{\alpha,1}^\top(\tau_t) \tilde{X}_{i,t-1} \tilde{X}_{i,t-1}^\top d_{\alpha,1}(\tau_t) + \\ & \quad \frac{2}{N\mathbb{T}} \sum_{i \in \mathcal{G}_{K_0-1}} \sum_{t=1}^{\mathbb{T}} \varepsilon_{i,t} d_{\alpha,1}^\top(\tau_t) \tilde{X}_{i,t-1} + o_P(1) \\ &\geq \frac{1}{N\mathbb{T}} \sum_{i \in \mathcal{G}_{K_0-1}} \sum_{t=1}^{\mathbb{T}} \varepsilon_{i,t}^2 + (\underline{\lambda} \cdot \kappa_1) \int_0^1 |d_{\alpha,1}(\tau)|^2 d\tau + o_P(1),\end{aligned} \quad (\text{B.31})$$

where $\underline{\lambda}$ and κ_1 are defined in Assumptions 2(ii) and 3(i), respectively. Analogously, we can also prove that

$$\frac{1}{N\mathbb{T}} \sum_{i \in \mathcal{G}_{K_0}} \sum_{t=1}^{\mathbb{T}} \left[x_{i,t}^\dagger - \hat{\alpha}_{K_0-1|K_0-1}^\top(\tau_t) \tilde{X}_{i,t-1} \right]^2 \geq \frac{1}{N\mathbb{T}} \sum_{i \in \mathcal{G}_{K_0}} \sum_{t=1}^{\mathbb{T}} \varepsilon_{i,t}^2 + (\underline{\lambda} \cdot \kappa_1) \int_0^1 |d_{\alpha,2}(\tau)|^2 d\tau + o_P(1), \quad (\text{B.32})$$

where $d_{\alpha,2}(\tau) = \alpha_{K_0}(\tau) - \alpha_{K_0-1}^*(\tau)$. By (B.28), (B.31), (B.32) and Assumption 3(i)(ii), we have

$$\begin{aligned}
\text{IC}(K_0 - 1) &= \log \widehat{\sigma}_{NT}^2(K_0 - 1) + (K_0 - 1) \cdot \rho_{NT} \\
&\geq \log \left\{ \frac{1}{NT} \sum_{k=1}^{K_0} \sum_{i \in \mathcal{G}_k} \sum_{t=1}^T \varepsilon_{i,t}^2 + (\underline{\lambda}_{K_1}) \int_0^1 [|d_{\alpha,1}(\tau)|^2 + |d_{\alpha,2}(\tau)|^2] d\tau \right\} + \\
&\quad (K_0 - 1) \cdot \rho_{NT} + o_P(1) \\
&\geq \log \left\{ \frac{1}{NT} \sum_{k=1}^{K_0} \sum_{i \in \mathcal{G}_k} \sum_{t=1}^T \varepsilon_{i,t}^2 + \frac{1}{2} (\underline{\lambda}_{K_1} \underline{\lambda}_{K_2}) \right\} + (K_0 - 1) \cdot \rho_{NT} + o_P(1) \\
&> \log \left\{ \frac{1}{NT} \sum_{k=1}^{K_0} \sum_{i \in \mathcal{G}_k} \sum_{t=1}^T \varepsilon_{i,t}^2 \right\} + K_0 \cdot \rho_{NT} + o_P(1) \\
&= \log \widehat{\sigma}_{NT}^2(K_0) + K_0 \cdot \rho_{NT} + o_P(1) = \text{IC}(K_0) + o_P(1),
\end{aligned}$$

completing the proof of Lemma B.4. \square

Lemma B.5. *Suppose that the conditions of Theorem 3.3 are satisfied. Then we have*

$$P(\text{IC}(K_0) < \text{IC}(K), K_0 + 1 \leq K \leq \bar{K}) \rightarrow 1 \quad (\text{B.33})$$

conditional on the event $\mathcal{E}(\mathcal{G})$.

Proof of Lemma B.5. Without loss of generality, we only consider $K = K_0 + 1$ and prove that

$$P(\text{IC}(K_0) < \text{IC}(K_0 + 1), \mathcal{E}(\mathcal{G})) \rightarrow 1.$$

Conditional on $\mathcal{E}(\mathcal{G})$, one of $\mathcal{G}_1, \dots, \mathcal{G}_{K_0}$ is split into two sub-clusters when the clustering algorithm stops at $K_0 + 1$. Without loss of generality, we assume that \mathcal{G}_{K_0} is divided into two sub-clusters and denote the resulting $K_0 + 1$ clusters as $\mathcal{G}_1^\diamond, \dots, \mathcal{G}_{K_0}^\diamond, \mathcal{G}_{K_0+1}^\diamond$ with $\mathcal{G}_k^\diamond = \mathcal{G}_k$ for $k = 1, \dots, K_0 - 1$ and $\mathcal{G}_{K_0}^\diamond \cup \mathcal{G}_{K_0+1}^\diamond = \mathcal{G}_{K_0}$. Accordingly, we let $\alpha_k^\diamond(\cdot) = \alpha_k(\cdot)$ when $k = 1, \dots, K_0 - 1$ and $\alpha_k^\diamond(\cdot) = \alpha_{K_0}(\cdot)$ when $k = K_0, K_0 + 1$.

Observe that

$$\widehat{\sigma}_{NT}^2(K_0 + 1) = \frac{1}{NT} \sum_{k=1}^{K_0+1} \sum_{i \in \mathcal{G}_k^\diamond} \sum_{t=1}^T [x_{i,t} - \widehat{\alpha}_{k|K_0+1}^\top(\tau_t) \widetilde{X}_{i,t-1}]^2.$$

Let $\Gamma_{k|K_0+1,X}(\tau), \Gamma_{k|K_0+1,\beta}(\tau)$ and $\Gamma_{k|K_0+1,\varepsilon}(\tau)$ be defined similarly to $\Gamma_{i,X}(\tau), \Gamma_{i,\beta}(\tau)$ and $\Gamma_{i,\varepsilon}(\tau)$, respectively, but with $\sum_{t=1}^T$ and h replaced by $\sum_{i \in \mathcal{G}_k^\diamond} \sum_{t=1}^T$ and h_1 , respectively. Write $\Delta_{\mathcal{G}_k^\diamond}(\tau) = \frac{1}{|\mathcal{G}_k^\diamond|} \sum_{i \in \mathcal{G}_k^\diamond} \Delta_i(\tau)$. For any $i \in \mathcal{G}_k^\diamond$, we write

$$x_{i,t} - \widehat{\alpha}_{k|K_0+1}^\top(\tau_t) \widetilde{X}_{i,t-1} = \varepsilon_{i,t} - [\widehat{\alpha}_{k|K_0+1}(\tau_t) - \alpha_k^\diamond(\tau_t)]^\top \widetilde{X}_{i,t-1}. \quad (\text{B.34})$$

By Lemma B.1(ii), (B.24) and the bandwidth condition $\log(N \vee T) = O(\text{Th}_1^3)$ in Assumption 3(ii), we may

show that

$$\begin{aligned}
\widehat{\alpha}_{k|K_0+1}(\tau_t) - \alpha_k^\diamond(\tau_t) &= \mathbf{E}_* \cdot \Gamma_{k|K_0+1, \mathcal{X}}^{-1}(\tau_t) [\Gamma_{k|K_0+1, \beta}(\tau_t) + \Gamma_{k|K_0+1, \varepsilon}(\tau_t)] \\
&= \Delta_{\mathcal{G}_k^\diamond}^{-1}(\tau) \left[\frac{1}{|\mathcal{G}_k^\diamond| \text{Th}_1} \sum_{i \in \mathcal{G}_k^\diamond} \sum_{s=1}^T \widetilde{X}_{i,s-1} \varepsilon_{i,s} \mathbf{K} \left(\frac{\tau_s - \tau_t}{h_1} \right) \right] + \text{O}_P \left(h_1^2 + \frac{\log(N \vee T)}{\text{Th}_1} \right) \\
&=: \mathbf{R}_k(\tau_t) + \text{O}_P(h_1^2)
\end{aligned}$$

uniformly over $\tau_t, t = 1, \dots, T$. By Assumption 2(ii), we may show that

$$\begin{aligned}
&\frac{1}{NT} \sum_{i \in \mathcal{G}_k^\diamond} \sum_{t=1}^T \left[\widetilde{X}_{i,t-1}^\top \mathbf{R}_k(\tau_t) \right]^2 \\
&\leq \frac{1}{NT} \sum_{i \in \mathcal{G}_k^\diamond} \sum_{t=1}^T \left\{ \widetilde{X}_{i,t-1}^\top \Delta_{\mathcal{G}_k^\diamond}^{-1}(\tau_t) \left[\frac{1}{|\mathcal{G}_k^\diamond| \text{Th}_1} \sum_{j \in \mathcal{G}_k^\diamond} \sum_{s=1}^T \widetilde{X}_{j,s-1} \varepsilon_{j,s} \mathbf{K} \left(\frac{\tau_s - \tau_t}{h_1} \right) \right] \right\}^2 \\
&\leq \Delta^{-2} \cdot \frac{1}{NT} \sum_{i \in \mathcal{G}_k^\diamond} \sum_{t=1}^T \left| \widetilde{X}_{i,t-1} \right|^2 \left| \frac{1}{|\mathcal{G}_k^\diamond| \text{Th}_1} \sum_{j \in \mathcal{G}_k^\diamond} \sum_{s=1}^T \widetilde{X}_{j,s-1} \varepsilon_{j,s} \mathbf{K} \left(\frac{\tau_s - \tau_t}{h_1} \right) \right|^2. \tag{B.35}
\end{aligned}$$

By the Cauchy-Schwarz inequality,

$$\begin{aligned}
&\sum_{i \in \mathcal{G}_k^\diamond} \sum_{t=1}^T \mathbf{E} \left[\left| \widetilde{X}_{i,t-1} \right|^2 \left| \sum_{j \in \mathcal{G}_k^\diamond} \sum_{s=1}^T \widetilde{X}_{j,s-1} \varepsilon_{j,s} \mathbf{K} \left(\frac{\tau_s - \tau_t}{h_1} \right) \right|^2 \right] \\
&\leq \sum_{i \in \mathcal{G}_k^\diamond} \sum_{t=1}^T \left\{ \mathbf{E} \left[\left| \widetilde{X}_{i,t-1} \right|^4 \right] \right\}^{1/2} \left\{ \mathbf{E} \left[\left| \sum_{j \in \mathcal{G}_k^\diamond} \sum_{s=1}^T \widetilde{X}_{j,s-1} \varepsilon_{j,s} \mathbf{K} \left(\frac{\tau_s - \tau_t}{h_1} \right) \right|^4 \right] \right\}^{1/2} \\
&\leq C^{|\mathcal{G}_k^\diamond|} T \left(\mathbf{E} \left[\left| \sum_{j \in \mathcal{G}_k^\diamond} \sum_{s=1}^T \widetilde{X}_{j,s-1} \varepsilon_{j,s} \mathbf{K} \left(\frac{\tau_s - \tau_t}{h_1} \right) \right|^4 \right] \right)^{1/2}. \tag{B.36}
\end{aligned}$$

By Burkholder's inequality for martingale differences (e.g., [Burkholder, 1973](#)), the conditional Hölder's inequality and Assumption 3(iii), we have

$$\begin{aligned}
&\mathbf{E} \left[\left| \sum_{s=1}^T \sum_{j \in \mathcal{G}_k^\diamond} \widetilde{X}_{j,s-1} \varepsilon_{j,s} \mathbf{K} \left(\frac{\tau_s - \tau_t}{h_1} \right) \right|^4 \right] \\
&\leq C \left(\mathbf{E} \left\{ \left[\sum_{s=1}^T \mathbf{K}^2 \left(\frac{\tau_s - \tau_t}{h_1} \right) \mathbf{E} \left(\left| \sum_{j \in \mathcal{G}_k^\diamond} \widetilde{X}_{j,s-1} \varepsilon_{j,s} \right|^2 \middle| \mathcal{F}_{s-1} \right) \right]^2 \right\} \right) +
\end{aligned}$$

$$\begin{aligned}
& \sum_{s=1}^T K^4 \left(\frac{\tau_s - \tau_t}{h_1} \right) \mathbb{E} \left[\left| \sum_{j \in \mathcal{G}_k^\circ} \tilde{X}_{j,s-1} \varepsilon_{j,s} \right|^4 \right] \\
& \leq C \sum_{s_1=1}^T \sum_{s_2=1}^T K^2 \left(\frac{\tau_{s_1} - \tau_t}{h_1} \right) K^2 \left(\frac{\tau_{s_2} - \tau_t}{h_1} \right) \mathbb{E} \left[\mathbb{E} \left(\left| \sum_{j \in \mathcal{G}_k^\circ} \tilde{X}_{j,s_1-1} \varepsilon_{j,s_1} \right|^2 \middle| \mathcal{F}_{s_1-1} \right) \right. \\
& \quad \left. \mathbb{E} \left(\left| \sum_{j \in \mathcal{G}_k^\circ} \tilde{X}_{j,s_2-1} \varepsilon_{j,s_2} \right|^2 \middle| \mathcal{F}_{s_2-1} \right) \right] + O(|\mathcal{G}_k^\circ|^2 T h_1) \\
& \leq C \left[\sum_{s=1}^T K^2 \left(\frac{\tau_s - \tau_t}{h_1} \right) \left\{ \mathbb{E} \left[\mathbb{E} \left(\left| \sum_{j \in \mathcal{G}_k^\circ} \tilde{X}_{j,s-1} \varepsilon_{j,s} \right|^4 \middle| \mathcal{F}_{s-1} \right) \right] \right\}^{1/2} \right]^2 + O(|\mathcal{G}_k^\circ|^2 T h_1) \\
& = C \left\{ \sum_{s=1}^T K^2 \left(\frac{\tau_s - \tau_t}{h_1} \right) \left[\mathbb{E} \left(\left| \sum_{j \in \mathcal{G}_k^\circ} \tilde{X}_{j,s-1} \varepsilon_{j,s} \right|^4 \right) \right]^{1/2} \right\}^2 + O(|\mathcal{G}_k^\circ|^2 T h_1) \\
& = O(|\mathcal{G}_k^\circ|^2 T^2 h_1^2 + |\mathcal{G}_k^\circ|^2 T h_1) = O(|\mathcal{G}_k^\circ|^2 T^2 h_1^2),
\end{aligned}$$

which, together with (B.35) and (B.36), leads to

$$\frac{1}{NT} \sum_{i \in \mathcal{G}_k^\circ} \sum_{t=1}^T \left[\tilde{X}_{i,t-1}^\top R_k(\tau_t) \right]^2 = O_P \left(\frac{|\mathcal{G}_k^\circ|^2 T^2 h_1}{N |\mathcal{G}_k^\circ|^2 T^3 h_1^2} \right) = O_P \left(\frac{1}{NT h_1} \right). \quad (\text{B.37})$$

On the other hand, note that

$$\begin{aligned}
& \frac{1}{NT} \sum_{i \in \mathcal{G}_k^\circ} \sum_{t=1}^T \varepsilon_{i,t} \tilde{X}_{i,t-1}^\top R_k(\tau_t) \\
& = \frac{1}{NT} \sum_{t=2}^T \sum_{i \in \mathcal{G}_k^\circ} \varepsilon_{i,t} \tilde{X}_{i,t-1}^\top \Delta_{\mathcal{G}_k^\circ}^{-1}(\tau_t) \left[\frac{1}{|\mathcal{G}_k^\circ| T h_1} \sum_{j \in \mathcal{G}_k^\circ} \sum_{s < t} \tilde{X}_{j,s-1} \varepsilon_{j,s} K \left(\frac{\tau_s - \tau_t}{h_1} \right) \right] + \\
& \quad \frac{1}{NT} \sum_{t=1}^T \sum_{i \in \mathcal{G}_k^\circ} \varepsilon_{i,t} \tilde{X}_{i,t-1}^\top \Delta_{\mathcal{G}_k^\circ}^{-1}(\tau_t) \left[\frac{1}{|\mathcal{G}_k^\circ| T h_1} \sum_{j \in \mathcal{G}_k^\circ} \sum_{s > t} \tilde{X}_{j,s-1} \varepsilon_{j,s} K \left(\frac{\tau_s - \tau_t}{h_1} \right) \right] + \\
& \quad \frac{1}{N |\mathcal{G}_k^\circ| T^2 h_1} \sum_{t=1}^T \sum_{i \in \mathcal{G}_k^\circ} \sum_{j \in \mathcal{G}_k^\circ} \varepsilon_{i,t} \varepsilon_{j,t} \tilde{X}_{i,t-1}^\top \Delta_{\mathcal{G}_k^\circ}^{-1}(\tau_t) \tilde{X}_{j,t-1} K(0).
\end{aligned}$$

As $\varepsilon_{i,t}$ is independent of $\tilde{X}_{i,t-1}^\top \sum_{j \in \mathcal{G}_k^\circ} \sum_{s < t} \tilde{X}_{j,s-1} \varepsilon_{j,s}$, following the arguments in the proof of (B.37) and

using the first condition in (3.14), we may show that

$$\begin{aligned}
& \mathbb{E} \left\{ \frac{1}{N\bar{T}} \sum_{t=2}^T \sum_{i \in \mathcal{G}_k^\circ} \varepsilon_{i,t} \tilde{X}_{i,t-1}^\top \Delta_{\mathcal{G}_k^\circ}^{-1}(\tau_t) \left[\frac{1}{|\mathcal{G}_k^\circ| T h_1} \sum_{j \in \mathcal{G}_k^\circ} \sum_{s < t} \tilde{X}_{j,s-1} \varepsilon_{j,s} \mathbb{K} \left(\frac{\tau_s - \tau_t}{h_1} \right) \right] \right\}^2 \\
&= \frac{1}{N^2 |\mathcal{G}_k^\circ|^2 T^4 h_1^2} \sum_{t=2}^T \mathbb{E} \left\{ \sum_{i \in \mathcal{G}_k^\circ} \varepsilon_{i,t} \tilde{X}_{i,t-1}^\top \Delta_{\mathcal{G}_k^\circ}^{-1}(\tau_t) \left[\sum_{j \in \mathcal{G}_k^\circ} \sum_{s < t} \tilde{X}_{j,s-1} \varepsilon_{j,s} \mathbb{K} \left(\frac{\tau_s - \tau_t}{h_1} \right) \right] \right\}^2 \\
&= \frac{1}{N^2 |\mathcal{G}_k^\circ|^2 T^4 h_1^2} \sum_{i,j \in \mathcal{G}} |\sigma_{i,j}| \cdot O(|\mathcal{G}_k^\circ| T^2 h_1) = O\left(\frac{1}{N^2 T^2 h_1}\right),
\end{aligned}$$

indicating that

$$\frac{1}{N\bar{T}} \sum_{t=2}^T \sum_{i \in \mathcal{G}_k^\circ} \varepsilon_{i,t} \tilde{X}_{i,t-1}^\top \Delta_{\mathcal{G}_k^\circ}^{-1}(\tau_t) \left[\frac{1}{|\mathcal{G}_k^\circ| T h_1} \sum_{j \in \mathcal{G}_k^\circ} \sum_{s < t} \tilde{X}_{j,s-1} \varepsilon_{j,s} \mathbb{K} \left(\frac{\tau_s - \tau_t}{h_1} \right) \right] = O_P \left(\frac{1}{N T h_1^{1/2}} \right).$$

Similarly, we can also prove that

$$\frac{1}{N\bar{T}} \sum_{t=1}^T \sum_{i \in \mathcal{G}_k^\circ} \varepsilon_{i,t} \tilde{X}_{i,t-1}^\top \Delta_{\mathcal{G}_k^\circ}^{-1}(\tau_t) \left[\frac{1}{|\mathcal{G}_k^\circ| T h_1} \sum_{j \in \mathcal{G}_k^\circ} \sum_{s > t} \tilde{X}_{j,s-1} \varepsilon_{j,s} \mathbb{K} \left(\frac{\tau_s - \tau_t}{h_1} \right) \right] = O_P \left(\frac{1}{N T h_1^{1/2}} \right).$$

By (3.14), we have

$$\frac{1}{N |\mathcal{G}_k^\circ| T^2 h_1} \sum_{t=1}^T \sum_{i \in \mathcal{G}_k^\circ} \sum_{j \in \mathcal{G}_k^\circ} \mathbb{E} [|\varepsilon_{i,t} \varepsilon_{j,t}|] \mathbb{E} \left[\left| \tilde{X}_{i,t-1}^\top \Delta_{\mathcal{G}_k^\circ}^{-1}(\tau_t) \tilde{X}_{j,t-1} \right| \right] = O \left(\frac{1}{N T h_1} \right).$$

Combining the above arguments, we prove that

$$\frac{1}{N\bar{T}} \sum_{i \in \mathcal{G}_k^\circ} \sum_{t=1}^T \varepsilon_{i,t} \tilde{X}_{i,t-1}^\top R_k(\tau_t) = O_P \left(\frac{1}{N T h_1} \right). \tag{B.38}$$

By (B.34), (B.37), (B.38) and Assumption 3(ii), we can prove that

$$\begin{aligned}
\text{IC}(K_0 + 1) &= \log \hat{\sigma}_{N\bar{T}}^2(K_0 + 1) + (K_0 + 1) \cdot \rho_{N\bar{T}} \\
&= \log \left\{ \frac{1}{N\bar{T}} \sum_{k=1}^{K_0+1} \sum_{i \in \mathcal{G}_k^\circ} \sum_{t=1}^T \varepsilon_{i,t}^2 \right\} + O_P \left(\frac{1}{N T h_1} + h_1^4 \right) + (K_0 + 1) \cdot \rho_{N\bar{T}} \\
&= \log \left\{ \frac{1}{N\bar{T}} \sum_{k=1}^{K_0+1} \sum_{i \in \mathcal{G}_k^\circ} \sum_{t=1}^T \varepsilon_{i,t}^2 \right\} + (K_0 + 1) \cdot \rho_{N\bar{T}} + o_P(\rho_{N\bar{T}})
\end{aligned}$$

$$\begin{aligned}
&> \log \left\{ \frac{1}{NT} \sum_{k=1}^{K_0} \sum_{i \in \mathcal{I}_k} \sum_{t=1}^T \varepsilon_{i,t}^2 \right\} + K_0 \cdot \rho_{NT} + o_P(\rho_{NT}) \\
&= IC(K_0) + o_P(\rho_{NT}),
\end{aligned}$$

completing the proof of Lemma B.5. □

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.
- Bai, Z. and Saranadasa, H. (1996). Effect of high dimension: By an example of a two sample problem. *Statistica Sinica* **6**, 311–329.
- Basu, S. and Michailidis, G. (2015). Regularized estimation in sparse high-dimensional time series models. *The Annals of Statistics* **43**, 1535–1567.
- Beck, G. W., Hubrich, K. and Marcellino, M. (2016). On the importance of sectoral and regional shocks for price-setting. *Journal of Applied Econometrics* **31**, 1234–1253.
- Bernanke, B., Boivin, J. and Elias, P. S. (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach. *Quarterly Journal of Economics* **120**, 387–422.
- Billingsley, P. (1995). *Probability and Measure* (3rd Edition). John Wiley and Sons.
- Bonhomme, S. and Manresa, E. (2015). Grouped patterns of heterogeneity in panel data. *Econometrica* **83**, 1147–1184.
- Burkholder, D. L. (1973). Distribution function inequalities for martingales. *The Annals of Probability* **1**, 19–42.
- Cai, Z. (2007). Trending time-varying coefficient time series models with serially correlated errors. *Journal of Econometrics* **136**, 163–188.
- Casas, I., Gao, J., Peng, B. and Xie, S. (2021). Trending time-varying coefficient time series models with serially correlated errors. *Journal of Applied Econometrics* **36**, 328–345.
- Chen, E., Fan, J. and Zhu, X. (2022). Community network autoregression for high-dimensional time series. Forthcoming in *Journal of Econometrics*.
- 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.
- Davis, R., Zang, P. and Zheng, T. (2016). Sparse vector autoregressive modeling. *Journal of Computational and Graphical Statistics* **25**, 1077–1096.
- Degras, D., Xu, Z., Zhang, T. and Wu, W. (2012). Testing for parallelism among trends in multiple time series. *IEEE Transactions on Signal Processing* **60**, 1087–1097.
- 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.
- Feng, G., Gao, J. and Peng, B. (2022). An integrated panel data approach to modelling economic growth. *Journal of Econometrics* **228**, 379–397.
- Freedman D.A. (1975). On tail probabilities for martingales. *The Annals of Probability* **3**, 100–118.
- Gao, J. and Gijbels I. (2008). Bandwidth selection in nonparametric kernel testing. *Journal of the American Statistical Association* **103**, 1584–1594.
- Gao, J., King, M., Lu, Z. and Tjøstheim, D. (2009). Nonparametric specification testing for nonlinear time series with nonstationarity. *Econometric Theory* **25**, 1869–1892.
- Guðmundsson, G. and Brownlees C. (2021). Detecting groups in large vector autoregressions. *Journal of Econometrics* **225**, 2–26.
- Hall, P. and Heyde, C.C. (1980). *Martingale limit theory and its application*. Academic press.
- Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference and Prediction* (2nd Edition). Springer-Verlag, New York.
- 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.
- Kock, A.B. and Callot, L. (2015). Oracle inequalities for high dimensional vector autoregressions. *Journal of Econometrics* **186**, 325–344.
- Li, D., Chen, J. and Lin, Z. (2011). Statistical inference in partially time-varying coefficient models. *Journal of Statistical Planning and Inference* **141**, 995–1013.
- Li, Q. (1999). Consistent model specification tests for time series econometric models. *Journal of Econometrics* **92**, 101–147.

- Li, Q. and Wang, S. (1998). A simple consistent bootstrap test for a parametric regression function. *Journal of Econometrics* **87**, 145–165.
- Lütkepohl, H. (2006). *New Introduction to Multiple Time Series Analysis*. Springer.
- Miao, K., Phillips, P.C.B. and Su, L. (2022). High-dimensional VARs with common factors. Forthcoming in *Journal of Econometrics*.
- Moench, E., Ng, S. and Potter, S. (2013). Dynamic hierarchical factor models. *The Review of Economics and Statistics* **95**, 1811–1817.
- Rodrik, D. (2013). Unconditional convergence in manufacturing. *Quarterly Journal of Economics* **128**, 165–204.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics* **6**, 461–464.
- 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. and Jin, S. (2019). Sieve estimation of time-varying panel data models with latent structures. *Journal of Business and Economic Statistics* **37**, 334–349.
- 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.
- 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.
- 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.
- Wu, W. and Zhao, Z. (2007). Inference of trends in time series. *Journal of the Royal Statistical Society, Series B* **69**, 391–410.
- Xia, Y. (1998). Bias-corrected confidence bands in nonparametric regression. *Journal of the Royal Statistical Society Series B* **60**, 797–811.

- Xu, M., Chen, X. and Wu, W. (2020). Estimation of dynamic networks for high-dimensional nonstationary time series. *Entropy* **22**, 55.
- 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.
- Zhang, T. and Wu, W. (2011). Testing parametric assumptions of trends of a nonstationary time series. *Biometrika* **98**, 599–614.
- Zhang, T. and Wu, W. (2012). Inference of time varying regression models. *The Annals of Statistics* **40**, 1376–1402.
- Zhou, Z. and Wu, W. (2010). Simultaneous inference of linear models with time varying coefficients. *Journal of the Royal Statistical Society, Series B* **72**, 513–531.
- 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. (2022). Simultaneous estimation and group identification for network vector autoregressive model with heterogeneous nodes. Working paper available at <https://arxiv.org/abs/2209.12229>.