entropy

MDPI

Article

Estimation of Dynamic Networks for High-Dimensional Nonstationary Time Series

Mengyu Xu [1], Xiaohui Chen [2] and Wei Biao Wu [3,*]

[1] Department of Statistics and Data Science, University of Central Florida, 4000 Central Florida Blvd, Orlando, FL 32816, USA; Mengyu.Xu@ucf.edu
[2] Department of Statistics, University of Illinois at Urbana-Champaign, S. Wright Street, Champaign, IL 61820, USA; xhchen@illinois.edu
[3] Department of Statistics, University of Chicago, 5747 S. Ellis Avenue, Jones 311, Chicago, IL 60637, USA
* Correspondence: wbwu@galton.uchicago.edu

Received: 14 November 2019; Accepted: 26 December 2019; Published: 31 December 2019

Abstract: This paper is concerned with the estimation of time-varying networks for high-dimensional nonstationary time series. Two types of dynamic behaviors are considered: structural breaks (i.e., abrupt change points) and smooth changes. To simultaneously handle these two types of time-varying features, a two-step approach is proposed: multiple change point locations are first identified on the basis of comparing the difference between the localized averages on sample covariance matrices, and then graph supports are recovered on the basis of a kernelized time-varying constrained L_1-minimization for inverse matrix estimation (CLIME) estimator on each segment. We derive the rates of convergence for estimating the change points and precision matrices under mild moment and dependence conditions. In particular, we show that this two-step approach is consistent in estimating the change points and the piecewise smooth precision matrix function, under a certain high-dimensional scaling limit. The method is applied to the analysis of network structure of the S&P 500 index between 2003 and 2008.

Keywords: high-dimensional time series; nonstationarity; network estimation; change points; kernel estimation

1. Introduction

Networks are useful tools to visualize the relational information among a large number of variables. An undirected graphical model belongs to a rich class of statistical network models that encodes conditional independence [1]. Canonically, Gaussian graphical models (or their normalized version partial correlations [2]) can be represented by the inverse covariance matrix (i.e., the precision matrix), where a zero entry is associated with a missing edge between two vertices in the graph. Specifically, two vertices are not connected if and only if they are conditionally independent, given the value of all other variables.

On one hand, there is a large volume of literature on estimating the (static) precision matrix for graphical models in the high-dimensional setting, where the sample size and the dimension are both large [3–16]. Most of the earlier work along this line assumes that the underlying network is time-invariant. This assumption is quite restrictive in practice and hardly plausible for many real-world applications, such as gene regulatory networks, social networks, and stocking market, where the underlying data generating mechanisms are often dynamic. On the other hand, dynamic random networks have been extensively studied from the perspective of large random graphs, such as community detection and edge probability estimation for dynamic stochastic block models (DSBMs) [17–30]. Such approaches do not model the sampling distributions of the error (or noise),

since the "true" networks are connected with random edges sampled from certain probability models, such as the Erdős–Rényi graphs [31] and random geometric graphs [32].

In this paper, we view the (time-varying) networks of interests as non-random graphs. We adopt the graph signal processing approach for denoising the nonstationary time series and target on estimating the *true unknown* underlying graphs. Despite the recent attempts towards more flexible time-varying models [33–40], there are still a number of major limitations in the current high-dimensional literature. First, theoretical analysis was derived under the fundamental assumption that the observations are either temporally *independent*, or the temporal dependence has very specific forms, such as Gaussian processes or (linear) vector autoregression (VAR) [14,33,34,37,41–43]. Such dynamic structures are unduly demanding in view that many time series encountered in real applications have very complex nonlinear spatial-temporal dependency [44,45]. Second, most existing work assumes the data have time-varying distributions with sufficiently light tails, such as Gaussian graphical models and Ising models [33,34,36,41,42]. Third, in change point estimation problems for high-dimensional time series, piecewise constancy is widely used [41,42,46,47], which can be fragile in practice. For instance, financial data often appears to have time-dependent cross-volatility with structural breaks [48]. For resting-state fMRI signals, correlation analysis reveals both slowly varying and abruptly changing characteristics corresponding to modularities in brain functional networks [49,50].

Advances in analyzing high-dimensional (stationary) time series have been made recently to address the aforementioned nonlinear spatial-temporal dependency issue [14,37,43,51–57]. In [53,56,57], the authors considered the theoretical properties of regularized estimation of covariance and precision matrices, based on various dependence measures of high-dimensional time series. Reference [38] considered the non-paranormal graphs that evolve with a random variable. Reference [37] discussed the joint estimation of Gaussian graphical models based on a stationary VAR(1) model with special coefficient matrices, which may also depend on certain covariates. The authors applied a constrained L_1-minimization for inverse matrix estimation (CLIME) estimator with a kernel estimator of covariance matrix and developed consistency in the graph recovery at a given time point. Reference [14] studied the recovery of the Granger causality across time and nodes assuming a stationary Gaussian VAR model with unknown order.

In this paper, we focus on the recovery of time-varying undirected graphs on the basis of the regularized estimation of the precision matrices for a general class of nonstationary time series. We simultaneously model two types of dynamics: abrupt changes with an unknown number of change points and the smooth evolution between the change points. In particular, we study a class of high-dimensional *piecewise locally stationary processes* in a general nonlinear temporal dependency framework, where the observations are allowed to have a finite polynomial moment.

More specifically, there are two main goals of this paper: first, to estimate the change point locations, as well as the number of change points, and second, to estimate the smooth precision matrix functions between the change points. Accordingly, our proposed method contains two steps. In the first step, the maximum norm of the local difference matrix is computed at each time point and the jumps in the covariance matrices are detected at the location where the maximum norms are above a certain threshold. In the second step, the precision matrices before and after the jump are estimated by a regularized kernel smoothing estimator. These two steps are recursively performed until a stopping criterion is met. Moreover, a boundary correction procedure based on data reflection is considered to reduce the bias near the change point.

We provide an asymptotic theory to justify the proposed method in high dimensions: point-wise and uniform rates of convergence are derived for the change point estimation and graph recovery under mild and interpretable conditions. The convergence rates are determined via subtle interplay among the sample size, dimensionality, temporal dependence, moment condition, and the choice of bandwidth in the kernel estimator. Our results are significantly more involved than problems for sub-Gaussian tails and independent samples. We highlight that uniform consistency in terms

of time-varying network structure recovery is much more challenging and difficult than pointwise consistency. For the multiple change point detection problem, we also characterize the threshold of the difference statistic that gives a consistent selection of the number of change points.

We fix some notations: Positive, finite, and non-random constants, independent of the sample size n and dimension p, are denoted by $C, C_1, C_2, \ldots$, whose values may differ from line to line. For the sequence of real numbers, a_n and b_n, we write $a_n = O(b_n)$ or $a_n \lesssim b_n$ if $\limsup_{n\to\infty}(a_n/b_n) \leq C$ for some constant $C < \infty$ and $a_n = o(b_n)$ if $\lim_{n\to\infty}(a_n/b_n) = 0$. We say $a_n \asymp b_n$ if $a_n = O(b_n)$ and $b_n = O(a_n)$. For a sequence of random variables Y_n and a corresponding set of constants a_n, denote $Y_n = O_{\mathbb{P}}(a_n)$ if for any $\varepsilon > 0$ there is a constant $C > 0$ such that $\mathbb{P}(|Y_n|/a_n > C) < \varepsilon$ for all n. For a vector $\mathbf{x} \in \mathbb{R}^p$, we write $|\mathbf{x}| = (\sum_{j=1}^p x_j^2)^{1/2}$. For a matrix Σ, $|\Sigma|_1 = \sum_{j,k}|\sigma_{jk}|$, $|\Sigma|_\infty = \max_{j,k}|\sigma_{jk}|$, $|\Sigma|_{L_1} = \max_k \sum_j |\sigma_{jk}|$, $|\Sigma|_F = (\sum_{j,k}\sigma_{jk}^2)^{1/2}$ and $\rho(\Sigma) = \max\{|\Sigma\mathbf{x}| : |\mathbf{x}| = 1\}$. For a random vector $\mathbf{z} \in \mathbb{R}^p$, write $\mathbf{z} \in \mathcal{L}^a$, $a > 0$, if $\|\mathbf{z}\|_a =: [\mathbb{E}(|\mathbf{z}|^a)]^{1/a} < \infty$. Let $\|\mathbf{z}\| = \|\mathbf{z}\|_2$. Denote $a \wedge b = \min(a, b)$ and $a \vee b = \max(a, b)$.

The rest of the paper is organized as follows: Section 2 presents the time series model, as well as the main assumptions, which can simultaneously capture the smooth and abrupt changes. In Section 3, we introduce the two-step method that first segments the time series based on the difference between the localized averages on sample covariance matrices and then recovers the graph support based on a kernelized CLIME estimator. In Section 4, we state the main theoretical results for the change point estimation and support recovery. Simulation examples are presented in Section 5 and a real data application is given in Section 6. Proof of main results can be found in Section 7.

2. Time Series Model

We first introduce a class of causal vector stochastic processes. Next, we state the assumptions to derive an asymptotic theory in Section 4 and explain their implications. Let $\boldsymbol{\varepsilon}_i \in \mathbb{R}^p, i \in \mathbb{Z}$ be independent and identically distributed (i.i.d.) random vectors and $\mathcal{F}_i = (\ldots, \boldsymbol{\varepsilon}_{i-1}, \boldsymbol{\varepsilon}_i)$ be a shift process. Let $\mathbf{X}_i^\circ(t) = (X_{i1}^\circ(t), \ldots, X_{ip}^\circ(t))$ be a p-dimensional nonstationary time series generated by

$$\mathbf{X}_i^\circ(t) = \mathbf{H}(\mathcal{F}_i; t), \tag{1}$$

where $\mathbf{H}(\cdot;\cdot) = \big(H_1(\cdot;\cdot), \ldots, H_p(\cdot;\cdot)\big)$ is an $\mathbb{R}^p$-valued jointly measurable function. Suppose we observe the data points $\mathbf{X}_i = \mathbf{X}_{i,n} = \mathbf{X}_i^\circ(t_i)$ at the evenly spaced time intervals $t_i = i/n, i = 1, 2, \ldots, n$,

$$\mathbf{X}_{i,n} = \mathbf{H}(\mathcal{F}_i; i/n). \tag{2}$$

We drop the subscription n in $\mathbf{X}_{i,n}$ in the rest of this section. Since our focus is to study the second-order properties, the data is assumed to have a mean of zero.

Model (1) is first introduced in [58]. The stochastic process $\big(X_i^\circ(t)\big)_{i\in\mathbb{Z}, t\in[0,1)}$ can be thought of as a triangular array system, double indexed by i and t, while the observations $(X_i)_{i=1}^n$ are sampled from the diagonal of the array. On one hand, when fixing the time index t, the (vertical) process $\big(X_i^\circ(t)\big)_{i\in\mathbb{Z}}$ is stationary. On the other hand, since $\mathbf{H}(\mathcal{F}_i; t_i)$ is allowed to vary with t_i, the diagonal process (2) is able to capture nonstationarity.

The process $(\mathbf{X}_i)_{i\in\mathbb{Z}}$ is causal or non-anticipative as $\mathbf{X}_i$ is an output of the past innovations $(\boldsymbol{\varepsilon}_j)_{j\leq i}$ and does not depend on future innovations. In fact, it covers a broad range of linear and nonlinear, stationary and non-stationary processes, such as vector auto-regressive moving average processes, locally stationary processes, Markov chains, and nonlinear functional processes [53,58–61].

Motivated by real applications where nonstationary time series data can involve both abrupt breaks and smooth varies between the breaks, we model the underlying processes as piecewise locally stationary with a finite number of structural breaks.

Definition 1 (Piecewise locally stationary time series model)**.** *Define* $\mathrm{PLS}_\iota([0,1],L)$ *as the collection of mean-zero piecewise locally stationary processes on* $[0,1]$*, if for each* $(X(t))_{0\le t\le 1}\in \mathrm{PLS}_\iota([0,1],L)$*, there is a nonnegative integer* ι *such that* $X(t)$ *is piecewise stochastic Lipschitz continuous in* t *with Lipschitz constant* L *on the interval* $[t^{(l)},t^{(l+1)}), l=0,\cdots,\iota$*, where* $0=t^{(0)}<t^{(1)}\cdots<t^{(\iota)}<t^{(\iota+1)}=1$*. A vector stochastic process* $(\mathbf{X}(t))_{0\le t\le 1}\in \mathrm{PLS}_\iota([0,1],L)$ *if all coordinates belong to* $\mathrm{PLS}_\iota([0,1],L)$*. For the process* $(X_0^\circ(t))_{0\le t\le 1}$ *defined in (1), this means that there exists a non-negative integer* ι *and a constant* $L>0$*, such that*

$$\max_{1\le j\le p}\left\|H_j(\mathcal{F}_0;t)-H_j(\mathcal{F}_0;t')\right\|\le L|t-t'| \text{ for all } t^{(l)}\le t,t'<t^{(l+1)}, 0\le l\le \iota.$$

Remark 1. *If we assume* $(\mathbf{X}_i^\circ(t))_{0\le t\le 1}\in \mathrm{PLS}_\iota([0,1],L), i\in\mathbb{Z}$*, then it follows that for each* $i'=i-k,\ldots,i+k$*, where* $k/n\to 0$*, and that* $t^{(l)}\le i,i'<t^{(l+1)}$ *for some* $0\le l\le\iota$*, we have*

$$\max_{1\le j\le p}\|H_j(\mathcal{F}_{i'};i/n)-H_j(\mathcal{F}_{i'};i'/n)\|\le Lk/n=o(1).$$

In other words, within a locally stationary time period, in a local window of i*,* $(X_{i'j})_{i-k\le i'\le i+k}$ *can be approximated by the stationary process* $(X_{i'j}^\circ(i/n))_{i-k\le i'\le i+k}$ *for each* $j=1,\ldots,p$*. This justifies the terminology of local stationarity.*

The covariance matrix function of the underlying process is $\Sigma(t)=\left(\sigma_{jk}(t)\right)_{1\le j,k\le p}$, $t\in[0,1]$, where $\sigma_{jk}(t)=\mathbb{E}\big(H_j(\mathcal{F}_0;t)H_k(\mathcal{F}_0;t)\big)$, and the precision matrix function is $\Omega(t)=\Sigma(t)^{-1}=\left(\omega_{jk}(t)\right)_{1\le j,k\le p}$. The graph at time t is denoted by $G(t)=(\mathcal{V},\mathcal{E}(t))$, where $\mathcal{V}$ is the vertex set and $\mathcal{E}(t)=\{(j,k):\omega_{jk}(t)\neq 0\}$. Note that $(\mathbf{X}_i^\circ(t))_t\in\mathrm{PLS}_\iota([0,1],L), i\in\mathbb{Z}$ implies piecewise Lipschitz continuity in $\Sigma(t)$ except at the breaks $t^{(1)},\ldots,t^{(\iota)}$. In particular, if $\sup_{0\le t\le 1}\max_{1\le j\le p}\left\|H_j(\mathcal{F}_0;t)\right\|\le C$ for some constant $C>0$, then

$$|\Sigma(s)-\Sigma(t)|_\infty \quad\le\quad 2CL|s-t|, \qquad \forall s,t\in[t^{(l)},t^{(l+1)}), l=0,\ldots,\iota. \tag{3}$$

The reverse direction is not necessarily true, i.e., (3) does not indicate $(\mathbf{X}_i^\circ(t))_t\in\mathrm{PLS}_\iota([0,1],L)$, $i\in\mathbb{Z}$ in general. As a trivial example, let $\varepsilon_{ij}=2^{-1/2}$ with probability $2/3$ and $\sqrt{2}$ with probability $1/3$ i.i.d for all i,j. At time $t_k=k/n$, let $X_{ij}^\circ(t_k)=(-1)^k\sqrt{t_k}\varepsilon_{ij}$. Then for any k and k' such that $k+k'$ is odd, $|\Sigma(t_k)-\Sigma(t_{k'})|_\infty=|t_k-t_{k'}|$, while $\|X_{01}^\circ(t_k)-X_{01}^\circ(t_{k'})\|_2=\sqrt{t_k}+\sqrt{t_{k'}}$.

Assumption 1 (Piecewise smoothness)**.** *(i) Assume* $(\mathbf{X}_i^\circ(t))_{0\le t\le 1}\in\mathrm{PLS}_\iota([0,1],L)$ *for each* $i\in\mathbb{Z}$*, where* $L>0$ *and* $\iota\ge 0$ *are constants independent of* n *and* p*. (ii) For each* $l=0,\ldots,\iota$*, and* $1\le j,k\le p$*, we have* $\sigma_{jk}(t)\in\mathcal{C}^2[t^{(l)},t^{(l+1)})$*.*

Now we introduce the temporal dependence measure. We quantify the dependence of $\left(\mathbf{X}_i^\circ(t)\right)_{i\in\mathbb{Z}}$ by the dependence adjusted norm (DAN) (cf. [62]). Let $\boldsymbol{\varepsilon}_i'$ be an independent copy of $\boldsymbol{\varepsilon}_i$ and $\mathcal{F}_{i,\{m\}}=(\ldots,\boldsymbol{\varepsilon}_{i-m-1},\boldsymbol{\varepsilon}_{i-m}',\boldsymbol{\varepsilon}_{i-m+1},\ldots,\boldsymbol{\varepsilon}_i)$. Denote $\mathbf{X}_{i,\{m\}}^\circ(t)=\left(X_{i1,\{m\}}^\circ(t),\ldots,X_{ip,\{m\}}^\circ(t)\right)$, where $X_{ij,\{m\}}^\circ(t)=H_j(\mathcal{F}_{i,\{m\}};t)$, $1\le j\le p$. Here $\mathbf{X}_{i,\{m\}}^\circ(t)$ is a coupled version of $\mathbf{X}_i^\circ(t)$, with the same generating mechanism and input, except that $\boldsymbol{\varepsilon}_{i-m}$ is replaced by an independent copy $\boldsymbol{\varepsilon}_{i-m}'$.

Definition 2 (Dependence adjusted norm (DAN))**.** *Let constants* $a\ge 1, A>0$*. Assume* $\sup_{0\le t\le 1}\|X_{1j}^\circ(t)\|_a<\infty, j=1,\ldots,p$*. Define the uniform functional dependence measure for the sequences* $(X_{ij}^\circ(t))_{i\in\mathbb{Z},t\in[0,1]}$ *of form (1) as*

$$\theta_{m,a,j}=\sup_{0\le t\le 1}\|X_{ij}^\circ(t)-X_{ij,\{m\}}^\circ(t)\|_a, \quad j=1,\ldots,p,$$

and $\Theta_{m,a,j} = \sum_{i=m}^{\infty} \theta_{i,a,j}$. *The dependence adjusted norm of* $(X_{ij}^{\circ}(t))_{i\in\mathbb{Z},t\in[0,1]}$ *is defined as*

$$\|X_{\cdot,j}\|_{a,A} = \sup_{m\geq 0}(m+1)^{A}\Theta_{m,a,j},$$

whenever $\|X_{\cdot,j}\|_{a,A} < \infty$.

Intuitively, the physical dependence measure quantifies the adjusted stochastic difference between the random variable and its coupled version by replacing past innovations. Indeed, $\theta_{m,a,j}$ measures the impact on $X_{ij}^{\circ}(t)$ uniform over t by replacing $\boldsymbol{\varepsilon}_{i-m}$ while freezing all the other inputs, while $\Theta_{m,a,j}$ quantifies the cumulative influence of replacing $\boldsymbol{\varepsilon}_{-m}$ on $(X_{ij}^{\circ}(t))_{i\geq 0}$ uniform over t. Then $\|X_{\cdot,j}\|_{a,A}$ controls the uniform polynomial decay in the lag of the cumulative physical dependence, where a depends on the the tail of marginal distributions of $X_{1,j}^{\circ}(t)$ and A quantifies the polynomial decay power and thus the temporal dependence strength. It is clear that $\|X_{\cdot,j}\|_{a,A}$ is a semi-norm, i.e., it is subaddative and absolutely homogeneous.

Assumption 2 (Dependence and moment conditions)**.** *Let* $\mathbf{X}_i^{\circ}(t)$ *be defined in (1) and* $\mathbf{X}_i$ *in (2). There exist* $q > 2$ *and* $A > 0$ *such that*

$$\nu_{2q} := \sup_{t\in[0,1]} \max_{1\leq j\leq p} \mathbb{E}|X_j^{\circ}(t)|^{2q} < \infty \qquad \textit{and} \qquad N_{X,2q} := \max_{1\leq j\leq p} \|X_{\cdot,j}\|_{2q,A} < \infty. \tag{4}$$

We let $M_{X,q} := \left(\sum_{1\leq j\leq p} \|X_{\cdot,j}\|_{2q,A}^{q}\right)^{1/q}$ and write $N_X = N_{X,4}$, $M_X = M_{X,2}$. The quantities $M_{X,q}$ and $N_{X,2q}$ measure the L^q-norm aggregated effect and the largest effect of the element-wise DANs respectively. Both quantities play a role in the convergence rates of our estimator.

Obviously, we have $\|X_{ij} - X_{ij,\{m\}}\|_a \leq \theta_{m,a,j}$ and $\max_{1\leq j\leq p} \mathbb{E}|X_{ij}|^{2q} \leq \nu_{2q}$ for all $1 \leq i \leq n$. In contrast to other works in a high-dimensional covariance matrix and network estimation, where sub-Gaussian tails and independence are the keys to ensure consistent estimation. Assumption 2 only requires that the time series have a finite polynomial moment, and it allows linear and nonlinear processes with short memory in the time domain.

Example 1 (Vector linear process)**.** *Consider the following vector linear process model*

$$\mathbf{H}(\mathcal{F}_i; t) = \sum_{m=0}^{\infty} A_m(t)\boldsymbol{\varepsilon}_{i-m},$$

where $\boldsymbol{\varepsilon}_i = (\varepsilon_1, \ldots, \varepsilon_p)$ *and* ε_{ij} *are i.i.d. with mean* 0 *and variance* 1*, and* $\|\varepsilon_{ij}\|_q \leq C_q$ *for each* $i \in \mathbb{Z}$ *and* $1 \leq j \leq p$ *with some constants* $q > 2$ *and* $C_q > 0$*. The vector linear process is commonly seen in literature and application [63]. It includes the time-varying VAR model where* $A_m(t) = A(t)^m$ *as a special example.*

Suppose that the coefficient matrices $A_m(t) = (a_{m,jk}(t))_{1\leq j,k\leq p}, m = 0, 1, \ldots$ *satisfy the following condition.*

(A1) For each $1 \leq j, k \leq p$, $a_{m,jk}(t) \in \mathcal{C}^2[0,1]$.

(A2) For each $1 \leq j \leq p$*, there is a constant* $C_{A,j} > 0$ *such that for each* $t \in [0,1]$, $\sum_{k=1}^{p} a_{m,jk}(t)^2 \leq C_{A,j}(m+1)^{-2(A+1)}$ *for all* $m \geq 0$.

(A3) For any $t, t' \in [0,1]$, $\sum_{m=0}^{\infty}\sum_{k=1}^{p}[a_{m,jk}(t) - a_{m,jk}(t')]^2 \leq L^2|t-t'|^2$ *for each* $j = 1, \ldots, p$.

Note that

$$\sigma_{jk}(t) = \sum_{m\geq 0} A_{m,j\cdot}^{\top}(t) A_{m,k\cdot}(t),$$

$$\Theta_{m,q,j} \leq 2C_q\sqrt{q-1}\sum_{m=0}^{\infty}(A_{m,j\cdot}^{\top}A_{m,j\cdot})^{1/2},$$

$$\|X_{ij}^{\circ}(t) - X_{ij}^{\circ}(t')\|^2 = \sum_{m=0}^{\infty} A_{m,j\cdot}\sum_{k=1}^{p}[a_{m,jk}(t) - a_{m,jk}(t')]^2,$$

where $A_{m,j\cdot}(t)$ is the jth row of $A_m(t)$. Under conditions (A1)–(A3), one can easily verify that for each $1 \leq j,k \leq p$, the process satisfies: (1) $\sigma_{jk}(t) \in \mathcal{C}^2[0,1]$; (2) $\|X_{\cdot,j}\|_{q,A} \leq C_q\sqrt{(q-1)C_{A,j}}$ (due to Burkholder's inequality, cf. [64]); (3) $\|H_j(\mathcal{F}_0;t) - H_j(\mathcal{F}_0;t')\| \leq L|t-t'|$.

Conditions (A1)–(A3) implicitly impose smoothness in each entry of the coefficient matrices, sparseness in each column of the entry and evolution, and polynomial decay rate in the lag m of each entry and its derivative.

For $1 \leq l \leq \iota$, let $\delta_{jk}(t^{(l)}) := \sigma_{jk}(t^{(l)}) - \sigma_{jk}(t^{(l)}-)$ and $\Delta(t^{(l)}) = \left(\delta_{jk}(t^{(l)})\right)_{1\leq j,k\leq p}$, where $\sigma_{jk}(t^{(l)}-) = \lim_{t\to t^{(l)}-}\sigma_{jk}(t)$ is well-defined in view of (3). We assume that the change points are separated and sizeable.

Assumption 3 (Separability and sizeability of change points)**.** *There exist positive constants $c_1 \in (0,1)$ and $c_2 > 0$ independent of n and p such that $\max_{0\leq l\leq\iota}(t^{(l+1)} - t^{(l)}) \geq c_1$ and $\delta(t_l) := |\Delta(t_l)|_\infty \geq c_2$.*

In the high-dimensional context, we assume that the inverse covariance matrices are sparse in the sense of their L_1 norms.

Assumption 4 (Sparsity of precision matrices)**.** *The precision matrix $|\Omega(t)|_{L^1} \leq \kappa_p$ for each $t \in [0,1]$, where κ_p is allowed to grow with p.*

If we further assume that the eigenvalues of the covariance matrices are bounded from below and above, i.e., there exists a constant $0 < c < 1$, such that $c \leq \inf_{t\in[0,1]}|\Sigma(t)|_2 \leq \sup_{t\in[0,1]}|\Sigma(t)|_2 \leq c^{-1}$, then the covariance matrices and precision matrices are well-conditioned. In particular, as $|\Omega(t) - \Omega(t')| \leq c^{-2}|\Sigma(t) - \Sigma(t')|$, a small perturbation in the covariance matrix would guarantee a small change of the same order in the precision matrix under the spectral norm.

3. Method: Change Point Estimation and Support Recovery

In graphical models (such as the Gaussian graphical model or partial correlation graph), network structures relevant to correlations or partial correlations are second-order characteristics of the data distributions. Specifically, the existence of edges coincides with non-zero entries of the inverse covariance matrix. We consider the dynamics of time series with both structural breaks and smooth changes. The piecewise stochastic Lipschitz continuity in Definition 1 allows the time series to have discontinuity in the covariance matrix function at time points $t^{(l)}, l = 1,\ldots,\iota$ (i.e., change points), while only smooth changes (i.e., twice continuous differentiability of the covariance matrix function in Assumptions 1) can occur between the change points.

In the presence of change points, we must first remove the change points before applying any smoothing procedures since $|\Omega(t) - \Omega(t-)|_\infty \geq |\Sigma(t)|_{L^1}^{-1}|\Sigma(t-)|_{L^1}^{-1}|\Delta(t)|_\infty$, i.e., a non-negligible abrupt change in the covariance matrix will result in a substantial change of the graph structure for sparse and smooth covariance matrices. Thus our proposed graph recovery method consists of two steps: change point detection and support recovery.

Let $h \equiv h_n > 0$ be a bandwidth parameter such that $h = o(1)$ and $n^{-1} = o(h)$, and $\mathcal{D}_h(0) = \{h, h+1/n, \ldots, 1-h\}$ be a search grid in $(0,1)$. Define

$$D(s) = n^{-1}\left(\sum_{i=0}^{hn-1} \mathbf{X}_{ns-i}\mathbf{X}_{ns-i}^\top - \sum_{i=1}^{hn} \mathbf{X}_{ns+i}\mathbf{X}_{ns+i}^\top\right), \qquad s \in \mathcal{D}_h(0). \tag{5}$$

To estimate the change points, compute

$$\hat{s}_1 = \text{argmax}_{s\in\mathcal{D}_h(0)}|D(s)|_\infty. \tag{6}$$

The following steps are performed recursively. For $l = 1, 2, \ldots$, let

$$\mathcal{D}_h(l) = \mathcal{D}_h(l-1) \cap \{\hat{s}_l - 2h, \cdots, \hat{s}_l + 2h\}^c, \tag{7}$$

$$\hat{s}_{l+1} = \arg\max_{s\in\mathcal{D}_h(l)} |D(s)|_\infty, \tag{8}$$

until the following criterion is attained:

$$\max_{s\in\mathcal{D}_h(l)} |D(s)|_\infty < \nu, \tag{9}$$

where ν is an early stopping threshold. The value of ν is determined in Section 4, which depends on the dimension and sample size, as well as the serial dependence level, tail condition, and local smoothness. Since our method only utilizes data in the localized neighborhood, multiple change points can be estimated and ranked in a single pass, which offers some computational advantage than the binary segmentation algorithm [41,46].

Once the change points are claimed, in the second step, we consider recovering the networks from the locally stationary time series before and after the structural breaks. In [11], where $X_i, i = 1, \ldots, n$ are assumed with an identical covariance matrix, the precision matrix $\hat{\Omega}$ is estimated as,

$$\hat{\Omega}_\lambda = \arg\min_{\Omega\in\mathbb{R}^{p\times p}} |\Omega|_1 \quad \text{s.t. } |\hat{\Sigma}\Omega - \text{Id}_p|_\infty \le \lambda, \tag{10}$$

where $\hat{\Sigma}$ is the sample covariance matrix. Inspired by (10), we apply a kernelized time-varying (tv-) CLIME estimator for the covariance matrix functions of the multiple pieces of locally stationary processes before and after the structural breaks. Let

$$\hat{\Sigma}(t) = \sum_{i=1}^{n} w(t, t_i)\mathbf{X}_i\mathbf{X}_i^\top, \tag{11}$$

where

$$w(t,i) = \frac{K_b(t_i, t)}{\sum_{i=1}^{n} K_b(t_i, t)} \tag{12}$$

and $K_b(u,v) = K(|u-v|/b)/b$. The bandwidth parameter b satisfies that $b = o(1)$ and $n^{-1} = o(b)$. Denote $B_n = nb$. The kernel function $K(\cdot)$ is chosen to have properties as follows.

Assumption 5 (Regularity of kernel function)**.** *The kernel function $K(\cdot)$ is non-negative, symmetric, and Lipschitz continuous with bounded support in $[-1,1]$, and that $\int_{-1}^{1} K(u)du = 1$.*

Assumption 5 is a common requirement on the kernel functions and can be fulfilled by a range of kernel functions, such as the uniform kernel, triangular kernel, and the Epanechnikov kernel.

Now the tv-CLIME estimator of the precision matrix $\Omega(t)$ is defined by $\tilde{\Omega}(t) = \left(\tilde{\omega}_{jk}(t)\right)_{1\le j,k\le p}$, where $\tilde{\omega}_{jk}(t) = \min(\hat{\omega}_{jk}(t), \hat{\omega}_{kj}(t))$, and $\hat{\Omega}(t) \equiv \hat{\Omega}_\lambda(t) = (\hat{\omega}_{jk}(t))_{1\le j,k\le p}$,

$$\hat{\Omega}_\lambda(t) = \arg\min_{\Omega\in\mathbb{R}^{p\times p}} |\Omega|_1 \quad \text{s.t. } |\hat{\Sigma}(t)\Omega - \mathrm{Id}_p|_\infty \le \lambda. \tag{13}$$

Similar hybridized kernel smoothing and the CLIME method for estimating the sparse and smooth transition matrices in high-dimensional VAR model has been considered in [65], where change point is not considered. Thus in the current setting we need to carefully control effect of (consistently) removing the change points before smoothing.

Then, the network is estimated by the "effective support" defined as follows.

$$\hat{G}(t;u) = (\hat{g}_{jk}(t;u))_{1\le j,k\le p}, \quad \text{where } \hat{g}_{jk}(t;u) = \mathbb{I}\left\{|\tilde{\omega}_{jk}(t)| \ge u\right\}. \tag{14}$$

It should be noted that the (vanilla) kernel smoothing estimator (11) of the covariance matrix does not adjust for the boundary effect due to the change points in the covariance matrice function. Thus, in the neighborhood of the change points, a larger bias can be induced in estimating $\Sigma(t)$ by $\hat{\Sigma}(t)$. As a remedy, we apply the following reflection procedure for boundary correction. Suppose $t \in \hat{\mathcal{T}}_{b+h^2}(j)$ for $1 \le j \le \iota$, Denote $\hat{\mathcal{T}}_d(j) := [\hat{s}_j - d, \hat{s}_j + d)$ for $d \in (0,1)$. We replace (11) by

$$\hat{\Sigma}(t) = \sum_{i=1}^{n} w(t,t_i)\check{\mathbf{x}}_i\check{\mathbf{x}}_i^\top,$$

and then apply the rest of the tv-CLIME approach. Here

$$\check{\mathbf{x}}_i = \begin{cases} \mathbf{x}_i & \text{if } (i-\hat{s}_j n)(t-\hat{s}_j n) \ge 0; \\ \mathbf{x}_{2\hat{s}_j n - i} & \text{otherwise.} \end{cases} \tag{15}$$

4. Theoretical Results

In this section, we derive the theoretical guarantees for the change point estimation and graph support recovery. Roughly speaking, Proposition 1 and 2 below show that under appropriate conditions, if each element of the covariance matrix varies smoothly in time, one can obtain an accurate snapshot estimation of the precision matrices as well as the time-varying graphs with high probability via the proposed kernel smoothed constrained l_1 minimization approach.

Define $J_{q,A}(n,p) = M_{X,q}(p\varpi_{q,A}(n))^{1/q}$, where $\varpi_{q,A}(n) = n, n(\log n)^{1+2q}, n^{q/2-Aq}$ if $A > 1/2 - 1/q$, $A = 1/2 - 1/q$, and $0 < A < 1/2 - 1/q$, respectively.

Proposition 1 (Rate of convergence for estimating precision matrices: pointwise and uniform)**.** *Suppose Assumptions 2, 4, and 5 hold with $\iota = 0$. Let $B_n = bn$ for $n^{-1} = o(b)$ and $b = o(1)$.*

(i) **Pointwise.** *Choose the parameter $\lambda^\circ \ge C\kappa_p(b^2 + B_n^{-1}J_{q,A}(B_n,p) + N_X(\log p/B_n)^{1/2})$ in the tv-CLIME estimator $\hat{\Omega}_{\lambda^\circ}(t)$ in (13), where C is a sufficiently large constant independent of n and p. Then for any $t \in [b, 1-b]$, we have*

$$|\hat{\Omega}_{\lambda^\circ}(t) - \Omega(t)|_\infty = O_{\mathbb{P}}(\kappa_p\lambda^\circ). \tag{16}$$

(ii) **Uniform.** *Choose $\lambda^\diamond \ge C\kappa_p\left(b^2 + B_n^{-1}J_{q,A}(n,p) + N_X B_n^{-1}(n\log(p))^{1/2}\right)$ in the tv-CLIME estimator $\hat{\Omega}_{\lambda^\circ}(t)$ in (13), where C is a sufficiently large constant independent of n and p. Then we have*

$$\sup_{t\in[b,1-b]} |\hat{\Omega}_{\lambda^\diamond}(t) - \Omega(t)|_\infty = O_{\mathbb{P}}(\kappa_p\lambda^\diamond). \tag{17}$$

The optimal order of the bandwidth parameter $b = b_\sharp$ in (17) is the solution to the following equation:

$$b^2 = B_n^{-1} \max(J_{q,A}(n,p),\, N_X(n\log(p^2))^{1/2}),$$

which implies that the closed-form expression for $b_\sharp$ is given by

$$b_\sharp = C_1\big(n^{-1}J_{q,A}(n,p)\big)^{1/3} + C_2 N_X^{1/3} n^{-1/6}\log(p)^{1/6}$$

for some constants C_1 and C_2 that are independent of n and p.

Given a finite sample, to distinguish the small entries in the precision matrix from the noise is challenging. Since a smaller magnitude of a certain element of the precision matrix implies a weaker connection of the edge in the graphical model, we instead consider the estimation of *significant* edges in the graph. Define the set of *significant* edges at level u as $\mathcal{E}^*(t;u) = \Big\{(j,k) : g^*_{jk}(t;u) \neq 0\Big\}$, where

$$g^*_{jk}(t;u) = \mathbb{I}\Big\{|\omega_{jk}(t)| > u\Big\}.$$

Then, as a consequence of (17), we have the following support recovery consistency result.

Proposition 2 (Consistency of support recovery: significant edges)**.** *Choose u as $u_\sharp = C_0\kappa_p^2 b_\sharp^2$, where C_0 is taken as a sufficiently large constant independent of n and p. Suppose that $u_\sharp = o(1)$ as $n, p \to \infty$. Then under conditions of Proposition 1, we have that as $n, p \to \infty$,*

$$\mathbb{P}\Big(\sup_{t\in[b,1-b]} \sum_{(j,k)\in\mathcal{E}^c(t)} \mathbb{I}\Big\{\hat{g}_{jk}(t;u_\sharp) \neq 0\Big\} \neq 0\Big) \to 0, \tag{18}$$

$$\mathbb{P}\Big(\sup_{t\in[b,1-b]} \sum_{(j,k)\in\mathcal{E}^*(t;2u_\sharp)} \mathbb{I}\Big\{\hat{g}_{jk}(t;u_\sharp) = 0\Big\} \neq 0\Big) \to 0. \tag{19}$$

Proposition 2 shows that the pattern of significant edges in the time-varying true graphs $G(t), t \in [b, 1-b]$, can be correctly recovered with high probability. However, it is still an open question to what extent the edges with magnitude below u can be consistently estimated, which can be naturally studied in the multiple hypothesis testing framework. Nonetheless, hypothesis testing for graphical models on the nonstationary high-dimensional time series is rather challenging. We leave it as a future problem.

Propositions 1 and 2 together yield that the consistent estimation of the precision matrices and the graphs can be achieved before and after the change points. Now, we provide the theoretical result of the change point estimation. Theorem 1 below shows that if the change points are separated and sizable, then we can consistently identify them via the single pass segmentation approach under suitable conditions. Denote

$$h_\diamond = C_1\big(n^{-1}J_{q,A}(n,p)\big)^{1/3} + C_2 N_X^{1/3} n^{-1/6}\log(p)^{1/6},$$

where C_1 and C_2 are constants independent of n and p.

Theorem 1 (Consistency of change point estimation)**.** *Assume $\mathbf{X}_i \in \mathbb{R}^p$ admits the form (2). Suppose that Assumptions 2 to 3 are satisfied. Choose the bandwidth $h = h_\diamond$, and $\nu = (1+L)h_\diamond^2$ in (5) and (9) respectively. Assume that $h_\diamond = o(1)$ as $n, p \to \infty$. We find that there exist constants C_1, C_2, C_3 independent of n and p, such that*

$$\mathbb{P}(|\hat{\imath} - \imath| > 0) \leq C_1\Big(\frac{p\varpi_{q,A}(n)M_{X,q}^q \nu_{2q}^q}{n^q c_2^q}\Big)^{1/3} + C_2 p^2 \exp\Big\{-C_3\big(\frac{n\log^2(p)}{N_X^2}\big)^{1/3}\Big\}. \tag{20}$$

Furthermore, in the event $\{\iota = \hat{\iota}\}$*, the ordered change-point estimator* $(\hat{s}_{(1)} < \hat{s}_{(2)} < \cdots < \hat{s}_{(\hat{\iota})})$ *defined in (7) satisfies*

$$\max_{1\le j\le \iota} |\hat{s}_{(j)} - t^{(j)}| = O_{\mathbb{P}}(h_\diamond^2). \tag{21}$$

Proposition 2 and Theorem 1 together indicate the consistency in the snapshot estimation of the time-varying graphs before and after the change points. In a close neighborhood of the change points, we have the following result for the recovery of the time-varying network. Denote $\mathcal{S} := [b_\sharp, 1 - b_\sharp] \cap (\cup_{1\le j\le \hat{\iota}} \hat{\mathcal{T}}^c_{h_\diamond^2 + b_\sharp}(j))$ as the time intervals between the estimated change points, and $\mathcal{N} := [0, b_\sharp) \cup \big(\cup_{1\le j\le \hat{\iota}} (\hat{\mathcal{T}}_{h_\diamond^2+b_\sharp} \cap \hat{\mathcal{T}}^c_{h_\diamond^2})\big) \cup (1 - b_\sharp, 1]$ as the recoverable neighborhood of the jump.

Theorem 2. *Let Assumptions 2 to 5 be satisfied. We have the following results as* $n, p \to \infty$*.*

(i) **Between change points.** *For* $t \in \mathcal{S}$*, take* $b = b_\sharp$ *and* $u = u_\sharp$*, where* $b_\sharp$ *and* $u_\sharp$ *are defined in Proposition 2. Suppose* $u_\sharp = o(1)$*. We have*

$$\sup_{t\in\mathcal{S}} \max_{j,k} |\hat{\sigma}_{j,k}(t) - \sigma_{j,k}(t)| = O_{\mathbb{P}}(b_\sharp^2). \tag{22}$$

Choose the penalty parameter as $\lambda_\sharp := C_1 \kappa_p b_\sharp^2$*, where* C_1 *is a constant independent of* n *and* p*. Then*

$$\sup_{t\in\mathcal{S}} |\hat{\Omega}_{\lambda_\sharp}(t) - \Omega(t)|_\infty = O_{\mathbb{P}}(\kappa_p^2 b_\sharp^2).$$

Moreover,

$$\mathbb{P}\Big(\sup_{t\in\mathcal{S}} \sum_{(j,k)\in\mathcal{E}^c(t)} \mathbb{I}\Big\{\hat{g}_{j,k}(t; u_\sharp) \neq 0\Big\} = 0\Big) \to 1, \tag{23}$$

$$\mathbb{P}\Big(\sup_{t\in\mathcal{S}} \sum_{(j,k)\in\mathcal{E}^*(t;2u_\sharp)} \mathbb{I}\Big\{\hat{g}_{jk}(t; u_\sharp) = 0\Big\} = 0\Big) \to 1. \tag{24}$$

(ii) **Around change points.** *For* $s \in \mathcal{N}$*, take* $b = b_\star := C_1\big(n^{-1} J_{q,A}(n,p)\big)^{1/2} + C_2 N_X^{1/2} n^{-1/4} \log(p)^{1/4}$*, and* $u = u_\star := C_0 \kappa_p^2 b_\star$*, where* C_0*,* C_1 *and* C_2 *are constants independent of* n *and* p*. Suppose* $u_\star = o(1)$*. We have*

$$\sup_{t\in\mathcal{N}} \max_{j,k} |\hat{\sigma}_{j,k}(t) - \sigma_{j,k}(t)| = O_{\mathbb{P}}(b_\star).$$

Choose the penalty parameter as $\lambda_\star := C_1 \kappa_p b_\star$*, where* C_1 *is a constant independent of* n *and* p*. Then*

$$\sup_{t\in\mathcal{N}} |\hat{\Omega}_{\lambda_\star}(t) - \Omega(t)|_\infty = O_{\mathbb{P}}(\kappa_p^2 b_\star). \tag{25}$$

Moreover,

$$\mathbb{P}\Big(\sup_{t\in\mathcal{N}} \sum_{(j,k)\in\mathcal{E}^c(t)} \mathbb{I}\Big\{\hat{g}_{j,k}(t; u_\star) \neq 0\Big\} = 0\Big) \to 1, \tag{26}$$

$$\mathbb{P}\Big(\sup_{t\in\mathcal{N}} \sum_{(j,k)\in\mathcal{E}^*(t;2u_\star)} \mathbb{I}\Big\{\hat{g}_{j,k}(t; u_\star) = 0\Big\} = 0\Big) \to 1. \tag{27}$$

Note that the convergence rates for the covariance matrix entries and precision matrix entries in case (ii) around the jump locations are slower than those for points well separated from the jump locations in case (i). This is because on the boundary due to the reflection, the smooth condition may no longer hold true. Indeed, we only take advantage of the Lipschitz continuous property of the

covariance matrix function. Thus, we lose one degree of regularity in the covariance matrix function, and the bias term b^2 in the convergence rate of the between-jump area becomes b around the jumps. We also note that around the smaller neighborhood of the jump $\mathcal{J} := \cup_{1 \le j \le \hat{\iota}} \hat{\mathcal{T}}_{h_\diamond^2}$, due to the larger error in the change point estimation, consistent recovery of the graphs is not achievable.

5. A Simulation Study

We simulate data from the following multivariate time series model:

$$X_i = \sum_{m=0}^{100} A_m(i) \boldsymbol{\epsilon}_{i-m}, i = 1, \ldots, n,$$

where $A_m(i) \in \mathbb{R}^{p \times p}, 1 \le m \le 100, 1 \le i \le n$, and $\boldsymbol{\epsilon}_{i-m} = (\epsilon_{i-m,1}, \ldots, \epsilon_{i-m,p})^\top$, with $\epsilon_{m,k}$, $m \in \mathbb{Z}$, $j = 1, \ldots, p$ generated as i.i.d. standardized $T(8)$ random variables. In the simulation, we fix $n = 1000$ and vary $p = 50$ and $p = 100$. For each $m = 1, \ldots, 100$, the coefficient matrices $A_m(i) = (1+m)^{-\beta} B_m(i)$, where $\beta = 1$, and $B_m(1)$ is an $R^{p \times p}$ block diagonal matrix. The 5×5 diagonal blocks in $B_m(i)$ are fixed with i.i.d. $N(0,1)$ entries and all the other entries are 0.

We consider the number of abrupt changes is $\iota = 2$ and $(nt^{(1)}, nt^{(2)}) = (300, 650)$. The matrix $A_0(i)$ is set to be a zero matrix for $i = 1, 2, \ldots, 299$, while $A_0(i) = A_0(299) + \boldsymbol{\alpha}\boldsymbol{\alpha}^\top$, $i = 300, 301, \ldots, 649$, and $A_0(i) = A_0(649) - \boldsymbol{\alpha}\boldsymbol{\alpha}^\top$, $i = 650, 651, \ldots, 1000$, where the first 20 entries in $\boldsymbol{\alpha}$ are taken to be a constant δ_0 and the others are 0.

We let the coefficient matrices $A_1(i) = \{a_{m,jk}(i)\}_{1 \le j,k \le p}$ evolve at each time point, such that two entries are soft-thresholded and another two elements increase. Specifically, at time i, we randomly select two elements from the support of $A_1(i)$, which are denoted as $\{a_{1,j_l^\star k_l^\star}(i)\}, l = 1, 2$ and that $a_{1,j^\star k^\star}(i) \neq 0$, and set them to $a^\star_{1,j_l^\star k_l^\star}(i) = \text{sign}(a_{1,j_l^\star k_l^\star}(i))(|a_{1,j_l^\star k_l^\star}(i) - 0.05|)$. We also randomly select two elements from $A_1^\star(i)$ and increase their values by 0.03.

Figures 1 and 2 show the support of the true covariance matrices at $i = 100, 200, \ldots, 900$.

In detecting the change points, the cutoff value ν of detection is chosen as follows. After removing the neighborhood of detected change points, we obtain $\mathcal{D}_h^{(l)}$ by ordering $\mathcal{D}_h^{(l)}, \ldots \mathcal{D}_h^{(\mathfrak{l})}$, where $\mathfrak{l}$ is obtained from (9) with $\nu = 0$. For $l = 1, 2, \ldots, \mathfrak{l} - 1$, compute

$$\mathcal{R}_h^{(l)} = \frac{\mathcal{D}_h^{(l)}}{\mathcal{D}_h^{(l+1)}}.$$

We let $\hat{\iota} = \arg\max_{0 \le l \le \mathfrak{l}-1} \mathcal{R}_h^{(l)}$ and set $\nu = \mathcal{D}_h^{(\hat{\iota})}$.

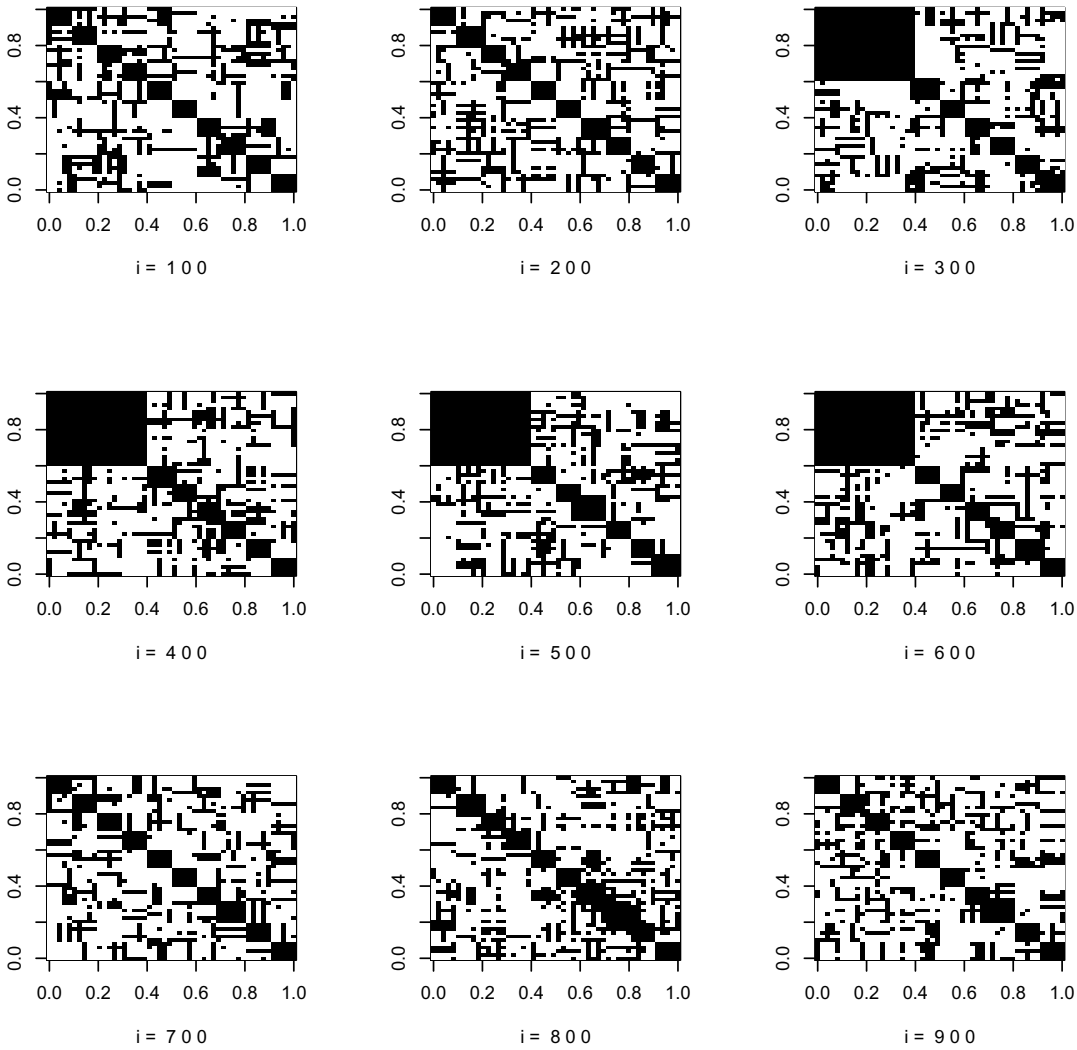


Figure 1. Support of the true covariance matrices, $p = 50$.

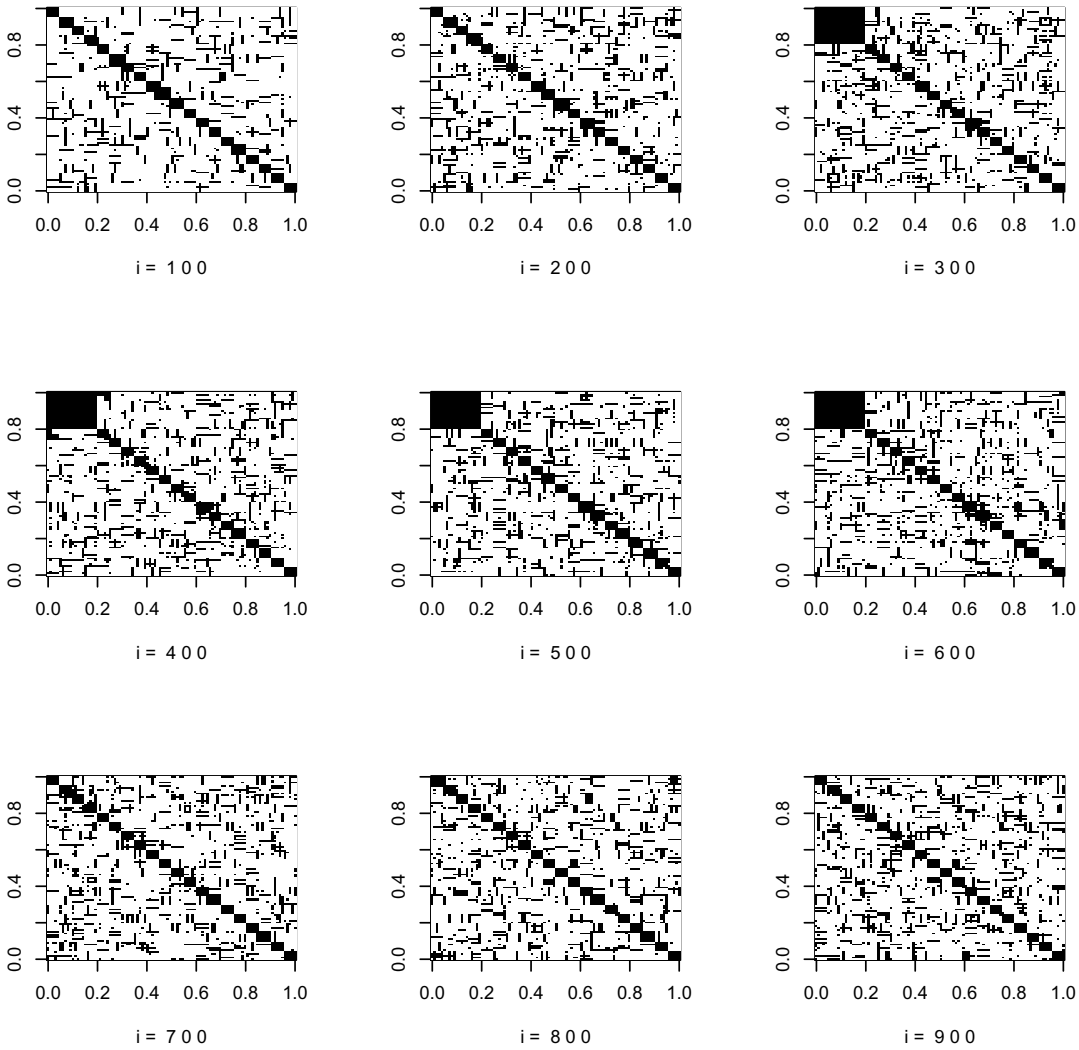


Figure 2. Support of the true covariance matrices, $p = 100$.

We report the number of estimated jumps and the average absolute estimation error, where the average absolute estimation error is the mean of the distance between the estimated change points and the true change points. As is shown in Tables 1 and 2, there is an apparent improvement in the estimation accuracy as the jump magnitude increases and dimension decreases. The detection is relatively robust to the choice of bandwidth.

Table 1. Average distance.

	Bandwidth	**0.14**	**0.16**	**0.18**	**0.2**	**0.22**	**0.24**
$p = 50$	$\delta_0 = 1$	23.4	21.0	17.47	16.6	14.7	16.5
	$\delta_0 = 2$	7.4	6.9	8.3	8.1	7.2	6.3
$p = 100$	$\delta_0 = 1$	37.2	30.1	26.4	25.5	21.2	21.3
	$\delta_0 = 2$	7.8	8.2	9.9	6.9	8.9	7.6

Table 2. Number of estimated change points.

	Bandwidth	**0.14**	**0.16**	**0.18**	**0.2**	**0.22**	**0.24**
$p = 50$	$\delta_0 = 1$	2.38	2.16	1.99	2.00	2.00	2.00
	$\delta_0 = 2$	2.46	2.31	2.00	2.00	2.00	2.00
$p = 100$	$\delta_0 = 1$	2.25	2.09	1.99	1.99	2.00	2.00
	$\delta_0 = 2$	2.38	2.19	2.00	2.00	2.00	2.00

We evaluate the support recovery performance of the time-varying CLIME at the lattice $100, 200, \ldots, 900$ with $\lambda = 0.02, 0.06, 0.1$. We take the uniform kernel function and the bandwidth is fixed as 0.2. At each time point t_0, two quantities are computed: sensitivity and specificity, which are defined as:

$$\text{sensitivity} = \frac{\sum_{1 \le j,k \le p} \mathbb{I}\{\hat{g}_{jk}(t_0; u) \neq 0, g_{jk}(t_0; u) \neq 0\}}{\sum_{1 \le j,k \le p} \mathbb{I}\{g_{jk}(t_0; u) \neq 0\}},$$
$$\text{specificity} = \frac{\sum_{1 \le j,k \le p} \mathbb{I}\{\hat{g}_{jk}(t_0; u) = 0, g_{jk}(t_0; u) = 0\}}{\sum_{1 \le j,k \le p} \mathbb{I}\{g_{jk}(t_0; u) = 0\}}.$$

We plot the Receiver Operating Characteristic (ROC) curve, that is, sensitivity against 1-specificity. From Figures 3 and 4 we observe that, due to a screening step, the support recovery is robust to the choice of λ, except at the change points, where a non-negligible estimation error of the covariance matrix is induced and the overall estimation is less accurate. As the effective dimension of the network remains the same at $p = 50$ and $p = 100$ by the construction of the coefficient matrix $A_m(i)$, there is no significant difference in the ROC curves at different dimensions.

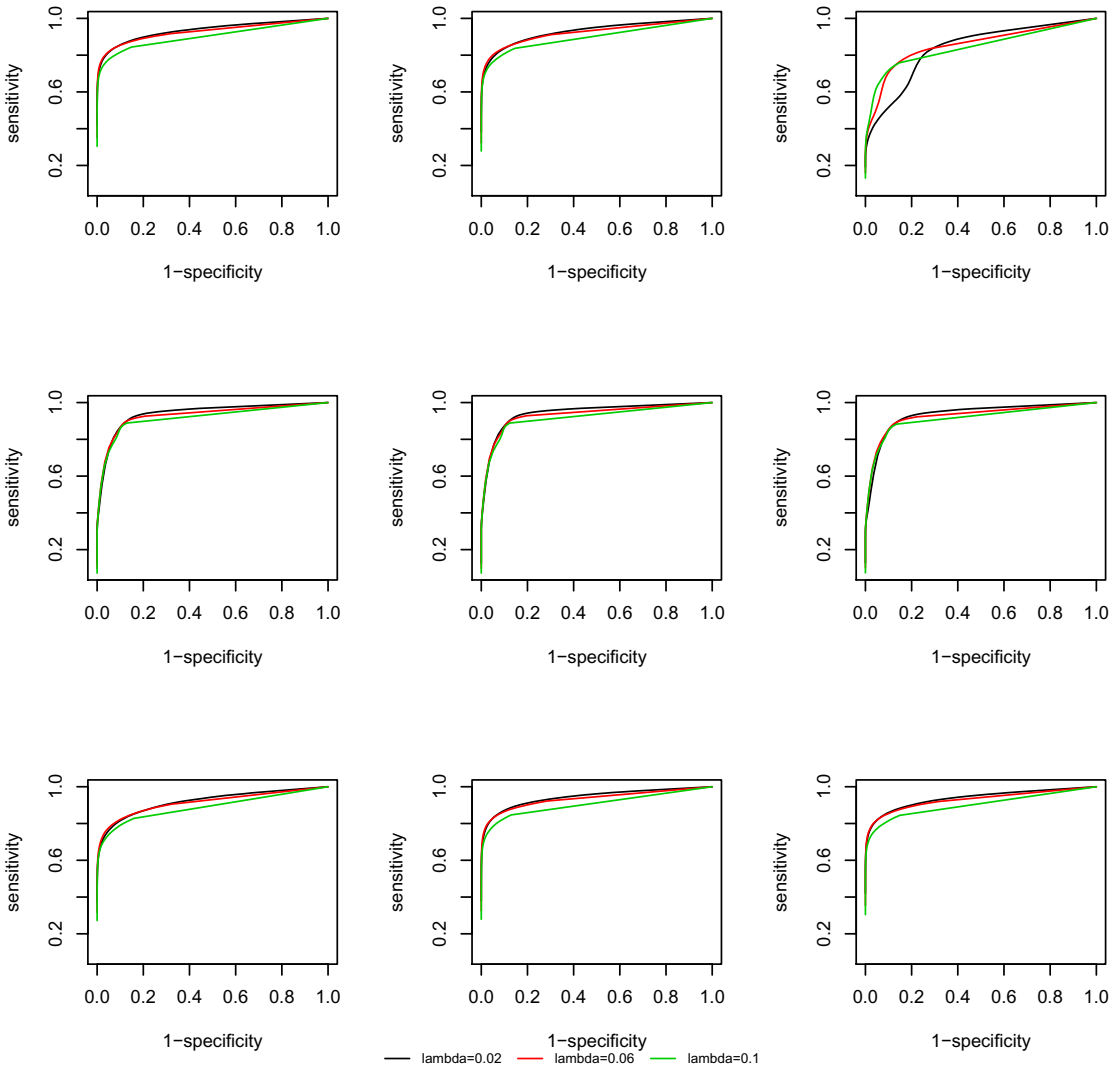


Figure 3. ROC curve of the time-varying CLIME, $p = 50$.

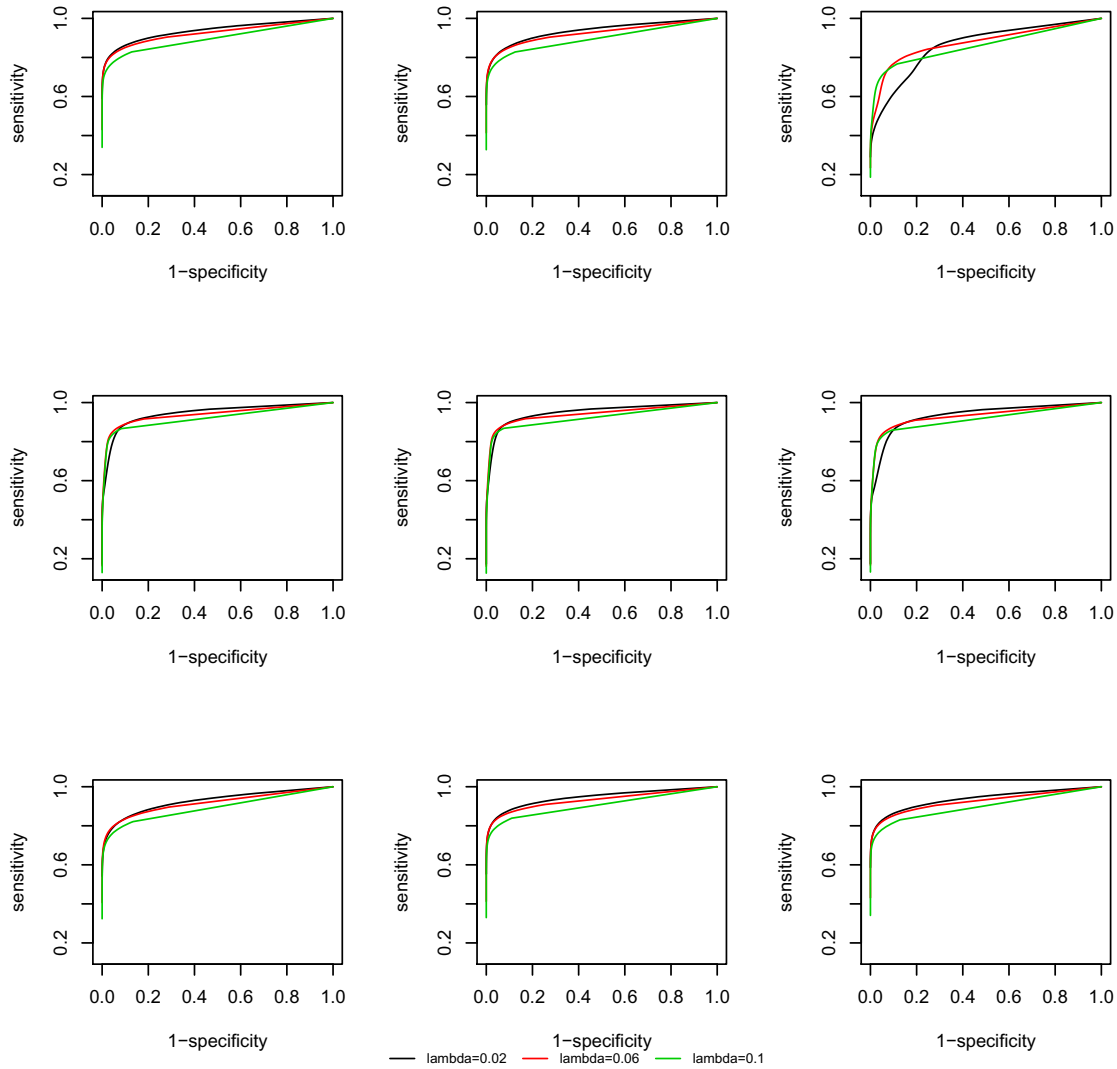


Figure 4. ROC curve of the time-varying CLIME, $p = 100$.

6. A Real Data Application

Understanding the interconnection among financial entities and how they vary over time provides investors and policy makers with insights into risk control and decision making. Reference [66] presents a comprehensive study of the applications of network theory in financial systems. In this section, we apply our method to a real financial dataset from Yahoo! Finance (finance.yahoo.com). The data matrix contains daily closing prices of 420 stocks that are always in the S&P 500 index between 2 January 2002 through 30 December 2011. In total, there are $n = 2519$ time points. We select 100 stocks with the largest volatility and consider their log-returns; that is, for $j = 1, \ldots, 100$,

$$X_{ij} = \log\left(p_{i+1,j}/p_{ij}\right),$$

where p_{ij} is the daily closing price of the stock j at time point i. We first compute the statistic (5) and (6) for the change point detection. We look at the top three statistics for different bandwidths. For bandwidth $k = n^{-1/5} = 0.21$, we rank the test statistic and find that the location for the top change point is: 7 February 2008 ($n_{\hat{s}_1} = 1536$), which is shown in Figure 5. The detected change point is quite robust to a variety of choices of bandwidth. Our result is partially consistent with the change point

detection method in [48]. In particular, the two breaks in 2006 and 2007 were also found in [48] and it is conjectured that the 2007 break may be associated to the U.S. house market collapse. Meanwhile, it is interesting to observe the increased volatility before the 2008 financial crisis.

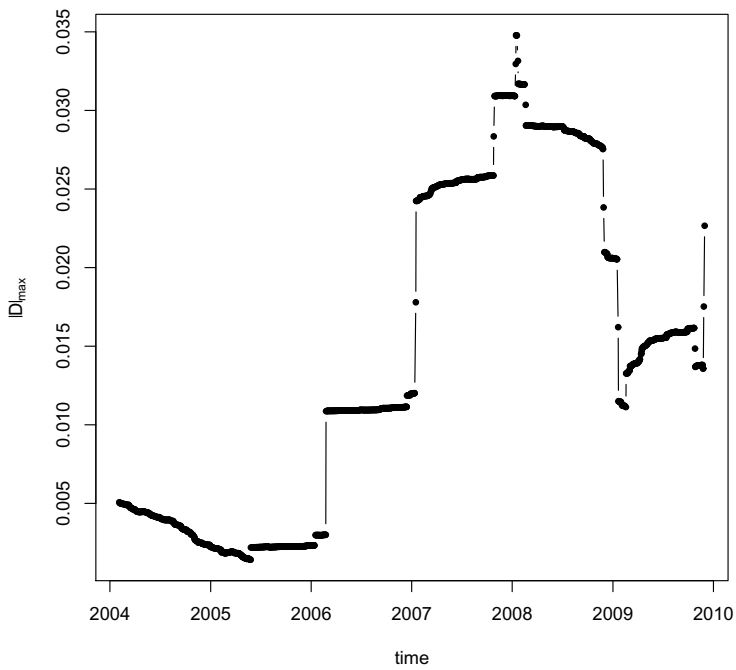
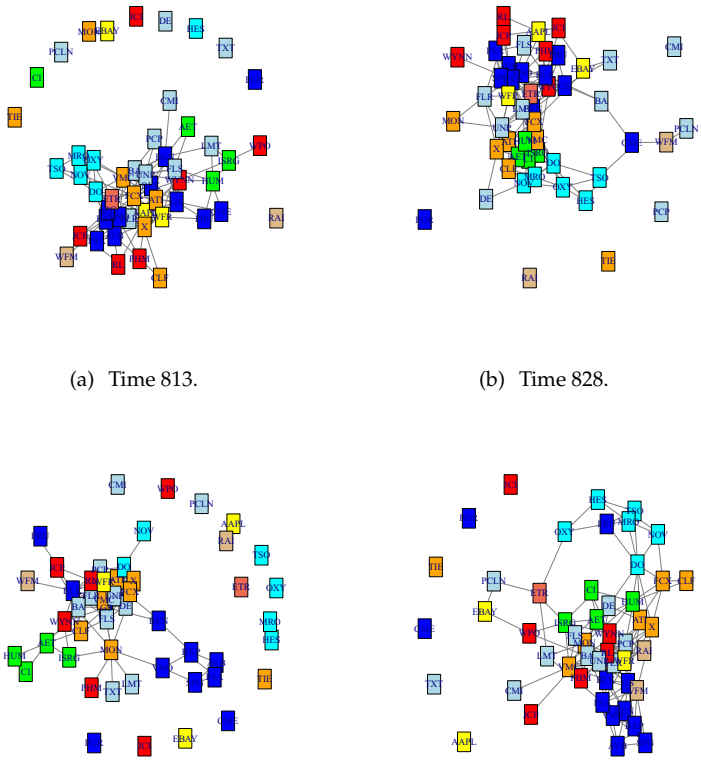


Figure 5. Break size $|D_s|_\infty$. From 4 February 2004, to 30 November 2009.

Next, we estimate the time-varying networks before and after the change point at 26 May 2006 with the largest jump size. Specifically, we look at four time points at: 813, 828, 888, and 903, corresponding to 23 March 2006, 13 April 2006, 11 July 2006, and 1 August 2006. We use tv-CLIME (13) with the Epanechnikov kernel with the same bandwidth as in the change point detection to estimate the networks at the four points. Optimal tuning parameter λ is automatically selected according to the stability approach [67]. The following matrix shows the number of different edges at those four time points. It is observed that the time of the first two time points (813 and 828) and the last two (888 and 903) has a higher similarity than across the change point at time 858. The estimated networks are shown in Figure 6. Networks in the first and second row are estimated before and after the estimated change point at time 858, respectively. It is observed that at each time point the companies in the same section tend to be clustered together such as companies in the `Energy` section: OXY, NOV, TSO, MRO, and DO (highlighted in cyan). In addition, the distance matrix of estimated networks is estimated as

$$\begin{pmatrix} 0 & 332 & 350 & 396 \\ 332 & 0 & 394 & 428 \\ 350 & 394 & 0 & 234 \\ 396 & 428 & 234 & 0 \end{pmatrix}.$$



(a) Time 813. (b) Time 828.

(c) Time 888. (d) Time 903.

Figure 6. Estimated networks at time points 813, 828, 888, and 903, corresponding to 23 March 2006, 13 April 2006, 11 July 2006, and 1 August 2006. Colors correspond to the nine sections in the S&P dataset.

7. Proof of Main Results

7.1. Preliminary Lemmas

Lemma 1. *Let $(Y_i)_{i\in Z}$ be a sequence that admits (2). Assume $Y_i \in \mathcal{L}^q$ for $i = 1, 2, \ldots$, and the dependence adjusted norm (DAN) of the corresponding underlying array $(Y_i^\circ(t))$ satisfies $\|Y.\|_{q,A} < \infty$ for $q > 2$ and $A > 0$. Let $(\omega(t, t_i))_{i=1}^n$ be defined in (12) and suppose that the kernel function $K(\cdot)$ satisfies Assumption 5. Denote $\varpi_{q,A}(n) = n, n(\log n)^{1+2q}, n^{q/2-Aq}$ if $A > 1/2 - 1/q$, $A = 1/2 - 1/q$, and $0 < A < 1/2 - 1/q$, respectively. Then there exist constants C_1, C_2 and C_3 independent of n, such that for all $x > 0$,*

$$\sup_{t\in(0,1)} \mathbb{P}\left(\left|\sum_{i=1}^n w(t,t_i)\big(Y_i - \mathbb{E}(Y_i)\big)\right| > x\right) \le C_1 \frac{\varpi_{q,A}(B_n)\,\|Y.\|_{q,A}^q}{B_n^q x^q} + C_2 \exp\left(\frac{-C_3 B_n x^2}{\|Y.\|_{2,A}^2}\right). \tag{28}$$

$$\mathbb{P}\left(\sup_{t\in(0,1)} \left|\sum_{i=1}^n w(t,t_i)\big(Y_i - \mathbb{E}(Y_i)\big)\right| > x\right) \le C_1 \frac{\varpi_{q,A}(n)\,\|Y.\|_{q,A}^q}{B_n^q x^q} + C_2 \exp\left(\frac{-C_3 B_n^2 x^2}{n\,\|Y.\|_{2,A}^2}\right). \tag{29}$$

Proof. Let $S_i = \sum_{j=1}^{i} \big(Y_i - \mathbb{E}(Y_i)\big)$. Note that

$$\begin{aligned}\sup_{t\in(0,1)} \left|\sum_{i=1}^{n} w(t,t_i)Y_i\right| &= \sup_{t\in(0,1)} \left|\sum_{i=1}^{n} w(t,t_i)(S_i - S_{i-1})\right| \\ &\le \sup_{t} \left|\sum_{i=1}^{n-1} \big[\big(w(t,t_i) - w(t,t_{i+1})\big)S_i\big]\right| + \sup_{t} |w(t,1)S_n| \\ &\lesssim B_n^{-1} \max_{1\le i\le n} |S_i|,\end{aligned}$$

where the last inequality follows from the fact that $\sup_t \sum_{i=1}^{n} |w(t,t_i) - w(t-t_{i+1})| \asymp B_n^{-1}$, due to Assumption 5.

To see (29), it suffices to show

$$\mathbb{P}\left(\max_{1\le i\le n} |S_i| > x\right) \le C_1 \frac{\varpi_{q,A}(n)\, \|Y.\|_{q,A}^{q}}{x^q} + C_2 \exp\left(\frac{-C_3 x^2}{n\, \|Y.\|_{2,A}^{2}}\right). \tag{30}$$

Now, we develop a probability deviation inequality for $\max_{1\le i\le n} |\sum_{j=1}^{i} \alpha_j Y_j|$, where $\alpha_j \ge 0$, $1 \le j \le n$ are constants such that $\sum_{1\le j\le n} \alpha_j = 1$. Denote $\mathcal{P}_0(Y_i) = \mathbb{E}(Y_i|\varepsilon_i) - \mathbb{E}(Y_i)$ and

$$\mathcal{P}_k(Y_i) = \mathbb{E}(Y_i|\varepsilon_{i-k},\ldots,\varepsilon_i) - \mathbb{E}(Y_i|\varepsilon_{i-k+1},\ldots,\varepsilon_i).$$

Then we can write

$$\begin{aligned}\max_{1\le i\le n} |\textstyle\sum_{j=1}^{i} \alpha_j Y_j| \quad &\le \max_{1\le i\le n} |\textstyle\sum_{j=1}^{i} \alpha_j \mathcal{P}_0(Y_j)| + \max_{1\le i\le n} |\textstyle\sum_{k=1}^{n}\sum_{j=1}^{i} \alpha_j \mathcal{P}_k(Y_j)| \\ &\quad + \max_{1\le i\le n} |\textstyle\sum_{k=n+1}^{\infty}\sum_{j=1}^{i} \alpha_j \mathcal{P}_k(Y_j)|.\end{aligned} \tag{31}$$

Note that $(\mathcal{P}_0(Y_j))_{j\in\mathbb{Z}}$ is an independent sequence. By Nagaev's inequality and Ottaviani's inequality, we have that

$$\begin{aligned}\mathbb{P}(\max_{1\le i\le n} |\textstyle\sum_{j=1}^{i} \alpha_j \mathcal{P}_0(Y_j)| \ge x) \quad &\lesssim \frac{\sum_{j=1}^{n} \alpha_j^q \|\mathcal{P}_0(Y_j)\|_q^q}{x^q} + \exp\big(-\frac{C_3 x^2}{\sum_{j=1}^{n} \alpha_j^2 \|\mathcal{P}_0(Y_j)\|_2^2}\big) \\ &\lesssim \frac{\sum_{j=1}^{n} \alpha_j^q}{x^q \|Y_j\|_q} + \exp\big(-C_3 \frac{x^2}{\sum_{j=1}^{n} \alpha_j^2}\big),\end{aligned} \tag{32}$$

where the last inequality holds because $\|\mathcal{P}_0(Y_j)\|_q \le 2\|Y_j\|_q$ by Jensen's inequality. Since $\sum_{j=i+1}^{\infty} \alpha_j \mathcal{P}_k(Y_j)$ is a martingale difference sequence with respect to $\sigma(\varepsilon_{i+1-k}, \varepsilon_{i+2-k}, \ldots)$, we have that $|\sum_{k=1+n}^{\infty}\sum_{j=i+1}^{n} \alpha_j \mathcal{P}_k(Y_j)|$ is a non-negative sub-martingale. Then by Doob's inequality and Burkholder's inequality, we have

$$\begin{aligned}&\mathbb{P}\big(\max_{1\le i\le n} |\textstyle\sum_{k=n+1}^{\infty}\sum_{j=1}^{i} \alpha_j \mathcal{P}_k(Y_j)| \ge x\big) \\ &\le \mathbb{P}\big(|\textstyle\sum_{k=n+1}^{\infty}\sum_{j=1}^{n} \alpha_j \mathcal{P}_k(Y_j)| \ge \frac{x}{2}\big) + \mathbb{P}\big(\max_{1\le i\le n} |\textstyle\sum_{k=n+1}^{\infty}\sum_{j=1+i}^{n} \alpha_j \mathcal{P}_k(Y_j)| \ge \frac{x}{2}\big) \\ &\lesssim \frac{\big\|\sum_{k=1+n}^{\infty}\sum_{j=1}^{n} \alpha_j \mathcal{P}_k(Y_j)\big\|_q^q}{x^q} \\ &\lesssim \frac{(\sum_{j=1}^{n} \alpha_j^2)^{q/2} \Theta_{n,q}^q}{x^q} \le \frac{\Theta_{n,q}^q n^{q/2-1} \sum_{j=1}^{n} \alpha_j^q}{x^q}.\end{aligned} \tag{33}$$

Now, we deal with the term $\max_{1\le i\le n} |\sum_{k=1}^{n}\sum_{j=1}^{i} \alpha_j \mathcal{P}_k(Y_j)|$. Define $a_m = \min(2^m, n)$ and $M_n = \lceil \log n / \log 2 \rceil$. Then

$$\max_{1\le i\le n} \Big|\sum_{k=1}^{n}\sum_{j=1}^{i} \alpha_j \mathcal{P}_k(Y_j)\Big| \le \sum_{m=1}^{M_n} \max_{1\le i\le n} \Big|\sum_{l=1}^{\lceil i/a_m \rceil} \sum_{j=1+(l-1)a_m}^{\min(la_m,i)} \sum_{k=1+a_{m-1}}^{a_m} \alpha_j \mathcal{P}_k(Y_j)\Big|. \tag{34}$$

Let $\mathcal{A}_{odd} = \{1 \le l \le \lceil i/a_m \rceil, l \text{ is odd}\}$ and $\mathcal{A}_{even} = \{1 \le l \le \lceil i/a_m \rceil, l \text{ is even}\}$. We have

$$\mathbb{P}\big(\max_{1\le i\le n} \big|\sum_{l=1}^{\lceil i/a_m\rceil} Z_{l,m,i}\big| \ge x\big) \le \mathbb{P}\big(\max_{1\le i\le n} \big|\sum_{\mathcal{A}_{odd}} Z_{l,m,i}\big| \ge x/2\big) + \mathbb{P}\big(\max_{1\le i\le n} \big|\sum_{\mathcal{A}_{even}} Z_{l,m,i}\big| \ge x/2\big),$$

where we have that $Z_{l,m,i} := \sum_{j=1+(l-1)a_m}^{\min(la_m,i)} \alpha_j \mathcal{P}_{a_{m-1}}^{a_m}(Y_j)$ is independent of $Z_{l+2,m,i}$ for $1 \le l \le \lceil i/a_m \rceil, 1 \le m \le M_n, 1 \le i \le n$, as $\mathcal{P}_{a_{m-1}}^{a_m}(Y_j) := \sum_{k=1+a_{m-1}}^{a_m} \mathcal{P}_k(Y_j)$ is a_m-dependent. Therefore, we can apply Ottaviani's inequality and Nagaev's inequality for independent variables. As a consequence,

$$\mathbb{P}\big(\max_{1\le i\le n} \big|\sum_{l=1}^{\lceil i/a_m\rceil} Z_{l,m,i}\big| \ge x\big) \lesssim \frac{\sum_{1\le l\le \lceil n/a_m\rceil} \|Z_{l,m,n}\|_q^q}{x^q} + \exp\big(-\frac{C_3x^2}{\sum_{1\le l\le \lceil n/a_m\rceil} \|Z_{l,m,n}\|_2^2}\big).$$

Again, by Burkholder's inequality, we have that for $q \ge 2$,

$$\begin{aligned}\|Z_{l,m,n}\|_q &\le \sum_{k=1+a_{m-1}}^{a_m} \Big\|\sum_{j=1+(l-1)a_m}^{\min(la_m,n)} \alpha_j \mathcal{P}_k(Y_j)\Big\|_q \\ &\lesssim \Big(\sum_{j=1+(l-1)a_m}^{\min(la_m,n)} \alpha_j^2\Big)^{1/2} (\Theta_{a_{m-1}} - \Theta_{a_m}).\end{aligned}$$

Note $\sum_{j=1+(l-1)a_m}^{\min(la_m,n)} \alpha_j^2 \le a_m^{(q-2)/q} (\sum_{j=1+(l-1)a_m}^{\min(la_m,n)} \alpha_j^q)^{2/q}$. Let $\tau_m = m^{-2}/\sum_{m=1}^{M_n} m^{-2}$, and we have $\tau_m \asymp m^{-2}$ as $1 \le \sum_{m=1}^{M_n} m^{-2} \le \pi^2/6$. In respect to (34), we have that

$$\begin{aligned}\mathbb{P}\big(\max_{1\le i\le n} \big|\textstyle\sum_{k=1}^n \sum_{j=1}^i \mathcal{P}_k(Y_j)\big| \ge x\big) &\le \textstyle\sum_{m=1}^{M_n} \mathbb{P}\big(\max_{1\le i\le n} \big|\sum_{l=1}^{\lceil i/a_m\rceil} Z_{l,m,i}\big| \ge \tau_m x\big) \\ &\lesssim \frac{\sum_{i=1}^n \alpha_j^q}{x^q} \|Y.\|_{q,A}^q \textstyle\sum_{m=1}^{M_n} \tau_m^{-q} a_m^{(1/2-A)q-1} + \sum_{m=1}^{M_n} \exp\big(-\frac{C_3x^2\tau_m^2 a_m^{2A}}{\sum_{j=1}^n \alpha_j^2 \|Y.\|_{2,A}^2}\big).\end{aligned} \tag{35}$$

Note $\sum_{m=1}^{M_n} \tau_m^{-q} a_m^{(1/2-A)q-1} \asymp n^{-1}\varpi_{q,A}(n)$, and

$$\sum_{m=1}^{M_n} \exp\big(-\frac{C_3x^2\tau_m^2 a_m^{2A}}{\sum_{j=1}^n \alpha_j^2 \|Y.\|_{2,A}^2}\big) \lesssim \exp\big(-\frac{C_3x^2}{\sum_{j=1}^n \alpha_j^2 \|Y.\|_{2,A}^2}\big).$$

Combining (31), (32), (33), and (35), we obtain

$$\begin{aligned}&\mathbb{P}\big(\max_{1\le i\le n} \big|\textstyle\sum_{j=1}^i \alpha_j\big(Y_j - \mathbb{E}(Y_j)\big)\big| > x\big) \\ &\quad\le C_1 \frac{\varpi_{q,A}(n) \sum_{j=1}^n \alpha_j^q \|Y.\|_{q,A}^q}{nx^q} + C_2 \exp\big(\frac{-C_3x^2}{\sum_{j=1}^n \alpha_j^2 \|Y\|_{2,A}^2}\big).\end{aligned} \tag{36}$$

Now, we have (30) by taking $\alpha_j = n^{-1}$ for $j = 1, \ldots, n$. Note that since $K(\cdot)$ has bounded support, for any given $t \in [b, 1-b]$, we have

$$\begin{aligned}&\mathbb{P}\big(\big|\sum_{i=1}^n w(t,t_i)(Y_i - \mathbb{E}Y_i)\big| > x\big) \le \mathbb{P}\big(\big|\sum_{i=-B_n}^{B_n} w(t,t_{tn+i})(Y_{tn+i} - \mathbb{E}Y_{tn+i})\big| > x\big) \\ &\le C_1 \frac{\varpi_{q,A}(B_n) \sum_{i=-B_n}^{B_n} w(t,t_{tn+i})^q \|Y.\|_{q,A}^q}{B_n x^q} + C_2 \exp\big(\frac{-C_3x^2}{\sum_{i=-B_n}^{B_n} w(t,t_{tn+i})^2 \|Y.\|_{2,A}^2}\big).\end{aligned}$$

Therefore (28) follows from (36) by taking $\alpha_j = w(t, tn+j)$, and note that for any $t \in [b, 1-b]$, $\sum_{i=-B_n}^{B_n} w(t,t_{tn+i})^\beta \asymp B_n^{1-\beta}$ for a constant $\beta \ge 2$. □

Lemma 2. *Suppose $(X_{ij})_{i\in\mathbb{Z},1\le j\le p}$ satisfys Assumption 2. Furthermore, let Assumption 5 hold. Let $\varpi_{q,A}(n)$ be defined as in Lemma 1. Then there exist constants C_1, C_2, and C_3 independent of n and p, such that for all $x > 0$, we have*

$$\begin{aligned}&\sup_{t\in(0,1)} \mathbb{P}\big(\big|\textstyle\sum_{i=1}^n \omega(t,t_i)\big(\mathbf{X}_i\mathbf{X}_i^\top - \mathbb{E}(\mathbf{X}_i\mathbf{X}_i^\top)\big)\big|_\infty \ge x\big)\\ &\qquad \le C_1\nu_{2q}^q \frac{p\varpi_{q,A}(B_n)M_{X,q}^q}{B_n^q x^q} + C_2 p^2 \exp\left(-C_3\frac{B_n x^2}{\nu_4^2 N_X^2}\right),\end{aligned} \tag{37}$$

and

$$\begin{aligned}&\mathbb{P}\big(\sup_{t\in(0,1)} \big|\textstyle\sum_{i=1}^n w(t,t_i)\big(\mathbf{X}_i\mathbf{X}_i^\top - \mathbb{E}(\mathbf{X}_i\mathbf{X}_i^\top)\big)\big|_\infty \ge x\big)\\ &\qquad \le C_1\nu_{2q}^q \frac{p\varpi_{q,A}(n)M_{X,q}^q}{B_n^q x^q} + C_2 p^2 \exp\left(-C_3\frac{B_n^2 x^2}{n\nu_4^2 N_X^2}\right).\end{aligned} \tag{38}$$

Proof. For $1 \le j,k \le p$, let $Y_{i,jk} = X_{ij}X_{ik}$. We now check the conditions in Lemma 1 for $(Y_{i,jk})_{1\le i\le n}$. Denote $Y_{i,jk,\{m\}} = X_{ij,\{m\}}X_{ik,\{m\}}$. Then the uniform functional dependence measure of $(Y_{i,jk})_i$ is

$$\begin{aligned}\theta_{m,q,jk}^Y &= \sup_i \|Y_{i,jk} - Y_{i,jk,\{m\}}\|_q\\ &= \sup_i \|X_{ij}X_{ik} - X_{ij,\{m\}}X_{ik,\{m\}}\|_q\\ &\le \sup_i \|X_{ij}(X_{ik} - X_{ik,\{m\}})\|_q + \sup_i \|X_{ik,\{m\}}(X_{ij} - X_{ij,\{m\}})\|_q.\end{aligned}$$

Thus the DAN of the process $Y_{\cdot,jk}$ satisfies that

$$\|Y_{\cdot,jk}\|_{q,A} \le \sup_i \|X_{ij}\|_{2q}\ \|X_{\cdot,k}\|_{2q,A} + \sup_i \|X_{ik}\|_{2q}\ \|X_{\cdot,j}\|_{2q,A} \le \nu_q(\|X_{\cdot,k}\|_{2q,A} + \|X_{\cdot,j}\|_{2q,A}).$$

The result follows immediately from Lemma 1 and the Bonferroni inequality. □

Lemma 3. *We adopt the notation in Lemma 2. Suppose Assumptions 2, 1, and 5 hold with $\iota = 0$. Recall $B_n = nb$, where $b \to 0$ and $B_n/\sqrt{n} \to \infty$ as $n \to \infty$. Then there exists a constant C independent of n and p such that $\hat{\Sigma}(t)$ in (11) satisfies that for any $t \in [c, 1-c]$,*

$$|\hat{\Sigma}(t) - \Sigma(t)|_\infty = O_{\mathbb{P}}\left(b^2 + M_{X,q}\nu_{2q}B_n^{-1}(p\varpi_{q,A}(B_n))^{1/q} + \nu_4 N_X(\log p/B_n)^{1/2}\right). \tag{39}$$

Furthermore,

$$\sup_{t\in[c,1-c]} |\hat{\Sigma}(t) - \Sigma(t)|_\infty = O_{\mathbb{P}}\left(b^2 + M_{X,q}\nu_{2q}B_n^{-1}(p\varpi_{q,A}(n))^{1/q} + \nu_4 N_X B_n^{-1}[n\log p]^{1/2}\right). \tag{40}$$

Proof. First, we have

$$\mathbb{E}\hat{\sigma}_{jk}(t) - \sigma_{jk}(t) = \sum_{i=1}^n w(t,t_i)[\sigma_{jk}(t_i) - \sigma_{jk}(t)].$$

Approximating the discrete summation with integral, we obtain for all $1 \le j,k \le p$,

$$\sup_{t\in[b,1-b]} \left|\mathbb{E}\hat{\sigma}_{jk}(t) - \sigma_{jk}(t) - \int_{-1}^1 K(u)[\sigma_{jk}(ub+t) - \sigma_{jk}(t)]du\right| = O\left(B_n^{-1}\right).$$

By Assumption 1, we have

$$\sigma_{jk}(ub+t) - \sigma_{jk}(t) = ub\sigma_{jk}'(t) + \frac{1}{2}u^2b^2\sigma_{jk}''(t) + o(b^2u^2).$$

Thus we have $\sup_{t\in[c,1-c]} |\mathbb{E}\hat{\sigma}(t) - \sigma(t)|_\infty = O\left(B_n^{-1} + b^2\right)$, in view of Assumption 5. By Lemma 2, we have

$$\sup_{t\in(0,1)} \mathbb{P}\left(\left|\hat{\Sigma}(t) - \mathbb{E}\hat{\Sigma}(t)\right|_\infty \geq x\right) \leq C_1 p \nu_q^q \frac{M_{X,q}^q \varpi_{q,A}(B_n)}{B_n^q x^q} + C_2 p^2 \exp\left(-C_3 \frac{B_n x^2}{N_X^2}\right).$$

Denote $u = C_4\left(M_{X,q}\nu_{2q}B_n^{-1}(p\varpi_{q,A}(B_n))^{1/q} + \nu_4 N_X(\log p/B_n)^{1/2}\right)$ for a large enough constant C_4, then for any $t \in (0,1)$,

$$\left|\hat{\Sigma}(t) - \mathbb{E}\hat{\Sigma}(t)\right|_\infty = O_{\mathbb{P}}(u).$$

Thus (39) is proved. The result (40) can be obtained similarly. □

7.2. Proof of Main Results

Proof of Proposition 1. Given (39) and (40), the proof of (16) is standard. (See, e.g., Theorem 6 of [11]). For λ° and λ^* given in Proposition 1, by Lemma 3, we have that, respectively,

$$\lambda^\circ \geq \sup_t \mathbb{E}\left(\kappa_p|\hat{\Sigma}(t) - \Sigma(t)|_\infty\right), \tag{41}$$

$$\lambda^\diamond \geq \mathbb{E}\left(\kappa_p \sup_t |\hat{\Sigma}(t) - \Sigma(t)|_\infty\right). \tag{42}$$

Then note that for any $t \in [0,1]$, for any $\lambda > 0$,

$$\begin{aligned}|\hat{\Omega}_\lambda(t) - \Omega(t)|_\infty &\leq |\Omega(t)|_{L_1}|\Sigma(t)\hat{\Omega}_\lambda(t) - \mathrm{Id}_p|_\infty \\ &\leq |\Omega(t)|_{L_1}\left[|\hat{\Sigma}(t)\hat{\Omega}_\lambda(t) - \mathrm{Id}_p|_\infty + |(\Sigma(t) - \hat{\Sigma}(t))\Omega(t)|_\infty + |\hat{\Omega}_\lambda(t) - \Omega(t)|_{L^1}|\hat{\Sigma}(t) - \Sigma(t)|_\infty\right]\end{aligned}$$

where by construction, we have $|\hat{\Sigma}(t)\hat{\Omega}_\lambda(t) - \mathrm{Id}_p|_\infty \leq \lambda$ and $|\hat{\Omega}_\lambda(t) - \Omega(t)|_{L^1} \leq 2\kappa_p$. Consequently,

$$|\hat{\Omega}_\lambda(t) - \Omega(t)|_\infty \leq \kappa_p\left(\lambda + 3\kappa_p|\hat{\Sigma}(t) - \Sigma(t)|_\infty\right). \tag{43}$$

Then (16) and (17) follow from (41) to (43). □

Proof of Proposition 2. Theorem 2 is an immediate result of (17). □

Proof of Theorem 1. Denote $r_j, 1 \leq j \leq \iota$ as the time point(s) of the time of jump ordered decreasingly in the sense of the infinite norm of covariance matrices, i.e., $|\Delta(r_1)|_\infty \geq |\Delta(r_2)|_\infty \geq \ldots \geq |\Delta(r_\iota)|_\infty \geq |\Delta(s)|_\infty$ for $s \in (0,1) \cap \{r_1, \ldots, r_\iota\}^c$. (Temporal order is applied if there is a tie.) Let $\mathcal{T}_h(j) = [r_j - h, r_j + h)$. For $h = o(1)$, as a result of Assumption 3, $\mathcal{T}_h(j) \cap \mathcal{T}_h(i) = \emptyset$ if $i \neq j$ for n sufficiently large. That is to say, each time point $s \in (0,1)$ is in the neighborhood of, at most, one change point.

For any $s \in [t^{(j)}, t^{(j+1)})$, $j = 0, 1, \ldots, \iota$, denote $\mathbb{D}(s) = \mathbb{E}[D(s)]$ and

$$\mathbb{D}^\diamond(s) = \begin{cases} (h - s + t^{(j)})\Delta(t^{(j)}), & t^{(j)} \leq s < t^{(j)} + h \\ 0, & t^{(j)} + h \leq s < t^{(j+1)} - h \\ (h + s - r)\Delta(t^{(j+1)}), & t^{(j+1)} - h \leq s \leq t^{(j+1)}. \end{cases} \tag{44}$$

Then, for $s \in \cup_{1\leq j\leq \iota}[t^{(j)} + h, < t^{(j+1)} - h)$, by (3), we have

$$|\Sigma(s+t) - \Sigma(s)|_\infty \leq Lt, \qquad \forall |t| \leq h,$$

we can easily verify that

$$\sup_{s\in[0,1]} |\mathbb{D}(s) - \mathbb{D}^\diamond(s)|_\infty \leq Lh^2. \tag{45}$$

Note that $|\mathbb{D}^\diamond(s)|_\infty$ is maximized at $s = r_1$ and $|\mathbb{D}^\diamond(r_1)|_\infty = h|\Delta(r_1)|_\infty$. By the triangle inequalities, we have that for some positive constant C, for any $s \in [0,1]$,

$$\begin{aligned} |\mathbb{D}(r_1)|_\infty - |\mathbb{D}(s)|_\infty &\geq hc_2 - |\mathbb{D}(r_1) - \mathbb{D}^\diamond(r_1)|_\infty - |\mathbb{D}^\diamond(s)|_\infty - |\mathbb{D}(s) - \mathbb{D}^\diamond(s)|_\infty \\ &\geq hc_2 - |\mathbb{D}^\diamond(s)|_\infty - 2Lh^2 \\ &\geq c_2(|s - r_1| \wedge h) - 2Lh^2. \end{aligned} \tag{46}$$

On the other hand, since $|D(r_1)|_\infty \leq |D(\hat{s}_1)|_\infty$, we have

$$\begin{aligned} |\mathbb{D}(r_1)|_\infty - |\mathbb{D}(\hat{s}_1)|_\infty &\leq |D(r_1)|_\infty - |D(\hat{s}_1)|_\infty + |\mathbb{D}(r_1) - D(r_1)|_\infty + |\mathbb{D}(\hat{s}_1) - D(\hat{s}_1)|_\infty \\ &\leq |\mathbb{D}(r_1) - D(r_1)|_\infty + |\mathbb{D}(\hat{s}_1) - D(\hat{s}_1)|_\infty. \end{aligned} \tag{47}$$

Denote the event $\mathcal{A} := \{\sup_{s\in[h,1-h]} |D(s) - \mathbb{D}(s)|_\infty \leq h_\diamond^2\}$ and let $\mathbf{Y_i} = (Y_{i,jk})_{1\leq j,k\leq p}$, $Y_{i,jk} = X_{ij}X_{ik} - \sigma_{i,jk}$. Note that

$$|D_{jk}(s) - \mathbb{D}_{jk}(s)| = \frac{1}{n}\left|\sum_{i=1}^{hn} Y_{n_s+1-i,jk} - \sum_{i=1}^{hn} Y_{n_s+i,jk}\right|. \tag{48}$$

By Lemma 2, we have for any $x > 0$,

$$\mathbb{P}\left(\sup_{s\in[h,1-h]} |D(s) - \mathbb{D}(s)|_\infty \geq x\right) \leq C_1 \frac{p\varpi_{q,A}(n) M_{X,q}^q \nu_{2q}^q}{n^q x^q} + C_2 p^2 \exp\left(-C_3 \frac{nx^2}{N_X^2}\right). \tag{49}$$

It follows that

$$|\mathbb{D}(r_1)|_\infty - |\mathbb{D}(\hat{s}_1)|_\infty = O_\mathbb{P}\big(h^{-1} J_{q,A}(n,p) + N_X h^{-1}(n^{-1}\log(p))^{1/2}\big).$$

Taking $h = h_\diamond$, we have

$$|\hat{s}_1 - r_1| = O_\mathbb{P}(h_\diamond^2).$$

Furthermore, we have

$$\mathbb{P}(\mathcal{A}) \geq 1 - C_1\big(\frac{p\varpi_{q,A}(n) M_{X,q}^q \nu_{2q}^q}{n^q c_2^q}\big)^{1/3} - C_2 p^2 \exp\big(-C_3\big(\frac{n\log^2(p)}{N_X^2}\big)^{1/3}\big).$$

Let $\mathcal{A}_k := \{\max_{1\leq j\leq k} |\hat{s}_j - r_j| \leq c_2^{-1} 2(L+1)h_\diamond^2\}$ for some $1 \leq k \leq \iota$. Assume $\mathcal{A}_k \subset \mathcal{A}$. Under $\mathcal{A}_k$ we have that $[r_j - h_\diamond, r_j + h_\diamond) \subset \hat{\mathcal{T}}_{2h_\diamond}(j) =: [\hat{s}_j - 2h_\diamond, \hat{s}_j + 2h_\diamond)$ for $1 \leq j \leq k$ and $r_{k+1} \notin \cup_{1\leq j\leq k} \hat{\mathcal{T}}_{2h_\diamond}(j)$ as a consequence of Assumption 3. According to (46) and (47), we have if $\mathcal{A}$ is true, $|\hat{s}_{k+1} - r_{k+1}| \leq c_2^{-1} 2(L+1)h_\diamond^2$, which implies $\mathcal{A}_{k+1} \subset \mathcal{A}$. The result (21) follows from deduction.

Suppose $\mathcal{A}$ holds. By the choice of ν, as a consequence of (45) and (49), and that $\nu \ll h_\diamond$, we have that

$$\sup_{s\in[0,1]} |D(s) - \mathbb{D}^\diamond(s)|_\infty \leq \nu.$$

As a result,

$$\min_{1\leq j\leq \iota} |D(r_j)|_\infty \geq c_2 h_\diamond - \nu \geq \nu,$$

i.e., $\hat{\iota} \geq \iota$. On the other hand, since $\cup_{1\leq j\leq \iota} \hat{\mathcal{T}}_{2h_\diamond}(j)$ is excluded from the searching region for $s_{\iota+1}$, we have

$$\sup_{s\in\left(\cup_{1\leq j\leq \iota} \hat{\mathcal{T}}_{2h_\diamond}(j)\right)^c} |D(s)|_\infty \leq \nu.$$

In other words, $\{\hat{\iota} = \iota\} \subset \mathcal{A}$. Thus (20) is proved. □

Proof of Theorem 2. We adopt the notations in the proof of Theorem 1 and assume that $\mathcal{E}$ holds. Similar to Lemma 3, we have that by Lemma 2, for any $t \in (0,1)$,

$$\left|\hat{\Sigma}(t) - \mathbb{E}\hat{\Sigma}(t)\right|_\infty = O_{\mathbb{P}}(u),$$

where $u = C_4\big(M_{X,q}\nu_{2q}B_n^{-1}(p\varpi_{q,A}(B_n))^{1/q} + \nu_4 N_X(\log p/B_n)^{1/2}\big)$ for a large enough constant C_4.

Since under $\mathcal{E}$, $\mathcal{T}_b(j) \subset \hat{\mathcal{T}}_{b+h_\diamond^2}(j)$. For $t \in \big(\cup_{1\le j\le \iota}\hat{\mathcal{T}}_{b+h_\diamond^2}(j)\big)^c \cap [b, 1-b]$, we have that for all $1 \le j,k \le p$,

$$\begin{aligned}\left|\mathbb{E}\hat{\sigma}_{jk}(t) - \sigma_{jk}(t)\right| &= \int_{-1}^{1} K(u)[\sigma_{jk}(ub+t) - \sigma_{jk}(t)]du + O\left(B_n^{-1}\right)\\ &= b\sigma'_{jk}(t)\int_{-1}^{1} uK(u)du + \big(\frac{1}{2}b^2\sigma''_{jk}(t) + o(b^2)\big)\int_{-1}^{1} u^2K(u)du + O\left(B_n^{-1}\right)\\ &= O(b^2 + B_n^{-1}).\end{aligned}$$

On the other hand, for $t \in \cup_{1\le j\le \iota}\big(\hat{\mathcal{T}}_{b+h_\diamond^2}(j) \cap \mathcal{T}^c_{h_\diamond^2}(j)\big) \cup [0,b] \cup [1-b,1]$, due to reflection, we no longer have that differentiability. As a result of the Lipschitz continuity, we get

$$\left|\mathbb{E}\hat{\sigma}_{jk}(t) - \sigma_{jk}(t)\right| = \int_{-1}^{1} K(u)[\sigma_{jk}(ub+t) - \sigma_{jk}(t)]du + O\left(B_n^{-1}\right) = O(b + B_n^{-1}).$$

The result (22) follows by the choices of b. The rest of the proof are similar to that of Proposition 1 and Theorem 2. □

Author Contributions: Methodology, M.X., X.C., W.B.W.; writing–original draft preparation, M.X., X.C., W.B.W.; writing–review and editing, M.X., X.C., W.B.W., software, M.X. All authors have read and agreed to the published version of the manuscript.

Funding: X.C.'s research is supported in part by NSF CAREER Award DMS-1752614 and UIUC Research Board Award RB18099. W.B.W.'s research is supported in part by NSF DMS-1405410.

Acknowledgments: X.C. acknowledges that part of this work was carried out at the MIT Institute for Data, System, and Society (IDSS).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lauritzen, S. *Graphical Models*; Clarendon Press: Oxford, UK, 1996.
2. Peng, J.; Wang, P.; Zhou, N.; Zhu, J. Partial Correlation Estimation by Joint Sparse Regression Models. *J. Am. Stat. Assoc.* **2009**, *104*, 735–746. [CrossRef]
3. Meinshausen, N.; Bühlmann, P. High-dimensional graphs and variable selection with the lasso. *Ann. Stat.* **2006**, *34*, 1436–1462. [CrossRef]
4. Friedman, J.; Hastie, T.; Tibshirani, R. Sparse Inverse Covariance Estimation with the Graphical Lasso. *Biostatistics* **2008**, *9*, 432–441. [CrossRef]
5. Banerjee, O.; El Ghaoui, L.; d'Aspremont, A. Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data. *J. Mach. Learn. Res.* **2008**, *9*, 485–516.
6. Rothman, A.J.; Bickel, P.J.; Levina, E.; Zhu, J. Sparse Permutation Invariant Covariance Estimation. *Electron. J. Stat.* **2008**, *2*, 494–515. [CrossRef]
7. Yuan, M. High Dimensional Inverse Covariance Matrix Estimation via Linear Programming. *J. Mach. Learn. Res.* **2010**, *11*, 2261–2286.
8. Yuan, M.; Lin, Y. Model selection and estimation in the Gaussian graphical model. *Biometrika* **2007**, *94*, 19–35. [CrossRef]
9. Ravikumar, P.; Wainwright, M.J.; Raskutti, G.; Yu, B. High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence. *Electron. J. Stat.* **2011**, *5*, 935–980. [CrossRef]
10. Candès, E.; Tao, T. Rejoinder: "The Dantzig selector: Statistical estimation when p is much larger than n" *Ann. Stat.* **2007**, *35*, 2392–2404. [CrossRef]

11. Cai, T.; Liu, W.; Luo, X. A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *J. Am. Stat. Assoc.* **2011**, *106*, 594–607. [CrossRef]
12. Cai, T.; Liu, W. Adaptive thresholding for sparse covariance matrix estimation. *J. Am. Stat. Assoc.* **2011**, *106*, 672–684. [CrossRef]
13. Fan, J.; Feng, Y.; Wu, Y. Network Exploration via the Adaptive Lasso and SCAD penalties. *Ann. Appl. Stat.* **2009**, *3*, 521–541. [CrossRef]
14. Basu, S.; Shojaie, A.; Michailidis, G. Network Granger causality with inherent grouping structure. *J. Mach. Learn. Res.* **2015**, *16*, 417–453.
15. Loh, P.L.; Bühlmann, P. High-dimensional learning of linear causal networks via inverse covariance estimation. *J. Mach. Learn. Res.* **2014**, *15*, 3065–3105.
16. Loh, P.L.; Wainwright, M.J. Structure estimation for discrete graphical models: Generalized covariance matrices and their inverses. *Ann. Stat.* **2013**, *41*, 3022–3049. [CrossRef]
17. Lèbre, S.; Becq, J.; Devaux, F.; Stumpf, M.P.; Lelandais, G. Statistical inference of the time-varying structure of gene-regulation networks. *BMC Syst. Biol.* **2010**, *4*, 1–16. [CrossRef]
18. Przytycka, T.M.; Singh, M.; Slonim, D.K. Toward the dynamic interactome: It's Toward the dynamic interactome: it's about time. *Brief. Bioinform.* **2010**, *11*, 15–29. [CrossRef]
19. Khandani, A.E.; Lo, A.W. What happened to the quants in August 2007? Evidence from factors and transactions data. *J. Financ. Mark.* **2011**, *14*, 1–46. [CrossRef]
20. Chi, K.T.; Liu, J.; Lau, F.C. A network perspective of the stock market. *J. Empir. Financ.* **2010**, *17*, 659–667.
21. Durante, D.; Dunson, D.B.; Vogelstein, J.T. Nonparametric Bayes modeling of populations of networks. *J. Am. Stat. Assoc.* **2017**, *112*, 1516–1530. [CrossRef]
22. Durante, D.; Dunson, D.B. Locally adaptive dynamic networks. *Ann. Appl. Stat.* **2016**, *10*, 2203–2232. [CrossRef]
23. Han, Q.; Xu, K.; Airoldi, E. Consistent estimation of dynamic and multi-layer block models. In Proceedings of the 32nd International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 1511–1520.
24. Danaher, P.; Wang, P.; Witten, D.M. The joint graphical lasso for inverse covariance estimation across multiple classes. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2014**, *76*, 373–397. [CrossRef]
25. Dondelinger, F.; Lèbre, S.; Husmeier, D. Non-homogeneous dynamic Bayesian networks with Bayesian regularization for inferring gene regulatory networks with gradually time-varying structure. *Mach. Learn.* **2013**, *90*, 191–230. [CrossRef]
26. Pensky, M. Dynamic network models and graphon estimation. *Ann. Stat.* **2019**, *47*, 2378–2403. [CrossRef]
27. Pensky, M.; Zhang, T. Spectral clustering in the dynamic stochastic block model. *Electron. J. Stat.* **2019**, *13*, 678–709. [CrossRef]
28. Bhattacharjee, M.; Banerjee, M.; Michailidis, G. Change Point Estimation in a Dynamic Stochastic Block Model. *arXiv* **2018**, arXiv:1812.03090.
29. Bartlett, T.E.; Kosmidis, I.; Silva, R. Two-way sparsity for time-varying networks, with applications in genomics. *arXiv* **2018**, arXiv:1802.08114.
30. Gaucher, S.; Klopp, O. Maximum likelihood estimation of sparse networks with missing observations. *arXiv* **2019**, arXiv:1902.10605.
31. Erdös, P.; Rényi, A. On Random Graphs I. *Publ. Math. Debr.* **1959**, *6*, 290–297.
32. Penrose, M. *Random Geometric Graphs*; Oxford University Press: Oxford, UK, 2003.
33. Zhou, S.; Lafferty, J.; Wasserman, L. Time Varying Undirected Graphs. *Mach. Learn.* **2010**, *80*, 295–319. [CrossRef]
34. Kolar, M.; Xing, E. On time varying undirected graphs. In Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS 2011), Ft. Lauderdale, FL, USA, 11–13 April 2011.
35. Kolar, M.; Song, L.; Xing, E. Estimating time-varying networks. *Ann. Appl. Stat.* **2010**, *4*, 94–123. [CrossRef]
36. Kolar, M.; Xing, E.P. Sparsistent Estimation Of Time-Varying Markov Sparsistent Estimation Of Time-Varying Markov Random Fields. *arXiv* **2009**, arXiv:0907.2337.
37. Qiu, H.; Han, F.; Liu, H.; Caffo, B. Joint estimation of multiple graphical models from high dimensional time series. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2015**, *78*, 487–504. [CrossRef]
38. Lu, J.; Kolar, M.; Liu, H. Post-regularization Inference for Dynamic Nonparanormal Graphical Models. *arXiv* **2015**, arXiv:1512.08298.

39. Ahmed, A.; Xing, E.P. Recovering time-varying networks of dependencies Recovering time-varying networks of dependencies in social and biological studies. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 11878–11883. [CrossRef]
40. Tibshirani, R.; Saunders, M.; Rosset, S.; Zhu, J.; Knight, K. Sparsity and smoothness via the fused lasso. *J. R. Stat. Soc. Ser. B* **2005**, *67*, 91–108. [CrossRef]
41. Cho, H.; Fryzlewicz, P. Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2015**, *77*, 475–507. [CrossRef]
42. Roy, S.; Atchadè, Y.; Michailidis, G. Change-point estimation in high-dimensional Markov random field models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2017**, *79*, 1187–1206. [CrossRef]
43. Zhou, S. Gemini: Graph estimation with matrix variate normal instances. *Ann. Stat.* **2014**, *42*, 532–562. [CrossRef]
44. Tong, H. *Non-Linear Time Series: A Dynamical System Approach*; Oxford University Press: Oxford, UK, 1993.
45. Fan, J.; Yao, Q. *Nonlinear Time Series: Nonparmatric and Parametric Methods*; Springer: Berlin/Heidelberg, Germany, 2003.
46. Fryzlewicz, P. Wild Binary Segmentation for multiple change-point detection. *Ann. Stat.* **2014**, *42*, 2243–2281. [CrossRef]
47. Kokoszka, P.; Leipus, R. Change-point estimation in ARCH models. *Bernoulli* **2000**, *6*, 513–539. [CrossRef]
48. Aue, A.; Hörmann, S.; Horváth, L.; Reimherr, M. Break detection in the covariance structure of multivariate time series models. *Ann. Stat.* **2009**, *37*, 4046–4087. [CrossRef]
49. Chang, C.; Glover, G.H. Time-frequency dynamics of resting-state brain connectivity measured with fMRI. *NeuroImage* **2010**, *50*, 81–98. [CrossRef] [PubMed]
50. Hutchison, M.; Womelsdorf, T.; Gati, J.; Everling, S.; Menon, R. Resting-state networks show dynamic functional connectivity in awake humans and anesthetized macaques. *Hum. Brain Mapp.* **2013**, *34*, 2154–2177. [CrossRef]
51. Wiesel, A.; Bibi, O.; Globerson, A. Time varying autoregressive moving average models for covariance estimation. *IEEE Trans. Signal Process.* **2013**, *61*, 2791–2801. [CrossRef]
52. Qiu, H.; Han, F.; Liu, H.; Caffo, B. *Robust Portfolio Optimization under High Dimensional Heavy-Tailed Time Series*; Technical Report; Johns Hopkins University: Baltimore, MD, USA, 2014.
53. Chen, X.; Xu, M.; Wu, W.B. Covariance and precision matrix estimation for high-dimensional time series. *Ann. Stat.* **2013**, *41*, 2994–3021. [CrossRef]
54. Chen, X.; Xu, M.; Wu, W.B. Regularized Estimation of Linear Functionals of Precision Matrices for High-Dimensional Time Series. *IEEE Trans. Signal Process.* **2016**, *64*, 6459–6470. [CrossRef]
55. Basu, S.; Michailidis, G. Regularized estimation in sparse high-dimensional time series models. *Ann. Stat.* **2015**, *43*, 1535–1567. [CrossRef]
56. Bhattacharjee, M.; Bose, A. Consistency of large dimensional sample covariance matrix under weak dependence. *Stat. Methodol.* **2014**, *20*, 11–26. [CrossRef]
57. Shu, H.; Nan, B. Estimation of Large Covariance and Precision Matrices from Temporally Dependent Observations. *arXiv* **2014**, arXiv:1412.5059.
58. Draghicescu, D.; Guillas, S.; Wu, W.B. Quantile curve estimation and visualization for nonstationary time series. *J. Comput. Graph. Stat.* **2009**, *18*, 1–20. [CrossRef]
59. Wu, W.B. Nonlinear system theory: Another look at dependence. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 14150–14154. [CrossRef] [PubMed]
60. Zhou, Z.; Wu, W.B. Local linear quantile estimation for nonstationary time series. *Ann. Stat.* **2009**, *37*, 2696–2729. [CrossRef]
61. Zhou, Z.; Wu, W.B. Simultaneous inference of linear models with time varying coefficients. *J. R. Stat. Soc.* **2010**, *72*, 513–531. [CrossRef]
62. Wu, W.B.; Wu, Y.N. Performance bounds for parameter estimates of high-dimensional linear models with correlated errors. *Electron. J. Stat.* **2016**, *10*, 352–379. [CrossRef]
63. Ltkepohl, H. *New Introduction to Multiple Time Series Analysis*; Springer: Berlin/Heidelberg, Germany, 2007.
64. Chow, Y.; Teicher, H. *Probability Theory: Independence, Interchangeability, Martingales*; Springer: New York, NY, USA, 1997; p. 414.
65. Ding, X.; Qiu, Z.; Chen, X. Sparse transition matrix estimation for high-dimensional and locally stationary vector autoregressive models. *Electron. J. Stat.* **2017**, *11*, 3871–3902. [CrossRef]

66. Allen, F.; Babus, A. Networks in Finance. In *The Network Challenge: Strategy, Profit, and Risk in an Interlinked World*; FT Press: Hoboken, NJ, USA, 2009.
67. Liu, H.; Roeder, K.; Wasserman, L. Stability Approach to Regularization Selection (StARS) for High-Dim Graphical Models. In Proceedings of the 23rd International Conference on Neural Information Processing Systems (NIPS'10), Vancouver, BC, Canada, 6–9 December 2010.