Time-varying β-model for dynamic directed networks

Yuqing Du[1] | **Lianqiang Qu[1]** | **Ting Yan[1]** | **Yuan Zhang[2]**

[1]School of Mathematics and Statistics, Central China Normal University, Wuhan, Hubei, 430079, China

[2]Department of Statistics, The Ohio State University, Columbus, Ohio, 43210, USA

Correspondence
Lianqiang Qu, School of Mathematics and Statistics , Central China Normal University, Wuhan, China.
Email: qulianq@ccnu.edu.cn

Funding information
National Natural Science Foundation of China, Grant/Award Numbers: 11771171, 12001219.

We extend the well-known β-model for directed graphs to dynamic network setting, where we observe snapshots of adjacency matrices at different time points. We propose a kernel-smoothed likelihood approach for estimating $2n$ time-varying parameters in a network with n nodes, from N snapshots. We establish consistency and asymptotic normality properties of our kernel-smoothed estimators as either n or N diverges. Our results contrast their counterparts in single-network analyses, where $n \to \infty$ is invariantly required in asymptotic studies. We conduct comprehensive simulation studies that confirm our theory's prediction and illustrate the performance of our method from various angles. We apply our method to an email data set and obtain meaningful results.

KEYWORDS
β-model, directed networks, dynamic networks, kernel smoothing

1 | INTRODUCTION

In statistical network analysis, node degrees play a fundamentally important role, both for its meaningfulness in modeling and high computational efficiency in parametric and nonparametric inference (Zhang and Xia, 2022). Among existing literature, a family of degree-driven models, namely, the β-model and its variants, received extensive attention and research interest (Chatterjee et al., 2011; Hillar and Wibisono, 2013; Olhede and Wolfe, 2012). Compared to other popular network models such as stochastic block model, the β-model provides a simple yet elegant way to characterize networks with potentially high heterogeneity. A particularly attractive advantage of the β-model is its

convenient and highly efficient parameter estimation (Chen et al., 2021), see also Hillar and Wibisono (2013); Shao et al. (2021). An incomplete list of notable results also includes: MLE existence (Rinaldo et al., 2013), central limit theorems (Yan and Xu, 2013), directed and bipartite β-model (Yan et al., 2016a; Fan et al., 2022), weighted edges (Hillar and Wibisono, 2013), incorporating covariates (Graham, 2017), regularization (Chen et al., 2021; Stein and Leng, 2021; Shao et al., 2021) and so on. But existing literature on the β-model almost exclusively focused on static networks; whereas time-varying data appear in many applications, such as dynamic email networks and online social networks (Cui and Chen, 2022+). Therefore, it is certainly of interest to build a time-varying β-model for dynamic networks.

In this paper, we fill in the significant blank in the modeling, methodology and theory of the β-model for addressing dynamic networks. Our contributions are three-folds. First, we propose a time-varying directed β-model for characterizing the evolution of dynamic bi-degrees in directed dynamic networks by extending the static β-model. The model contains $2n$ time-varying parameters, for which each node is attached to an out-degree varying parameter and an in-degree varying parameter. Second, we propose a kernel-smoothed likelihood approach to estimate $2n$ unknown parameter functions by borrowing information from nearby time points. This is inspired by the classical β-model estimation methods for static networks and kernel methods in classical nonparametric statistics. Third, we establish consistency and asymptotic normality of our proposed estimator. In contrast to the results for static directed networks (Yan et al., 2016a), our analysis reveals that borrowing information from nearby time points can significantly improve estimation accuracy and relax the assumptions for consistency and asymptotic normality. The results hold even when n stays fixed, as long as the number of observed times goes to infinity. For a sparse static network with zero in-degrees or zero out-degrees, the MLE in the directed β-model does not exist. Our kernel smoothed estimator eliminates this case as long as those nodes are connected to other nodes in the full observation period. Numerical studies and a real data application demonstrate our theoretical findings.

The rest of this paper is organized as follows. We first set up some notation; Section 2 describes our proposed model, devises estimation methods and inference procedures, and establishes theoretical justifications; in Section 3, we conduct comprehensive simulation studies to address the question of selecting the tuning parameter, assess the performance of our method from different aspects, and validate our theoretical prediction; in Section 4, we apply our method to a real-world data set and interpret the results; we conclude our paper with some discussion in Section 5.

1.1 | Notation

We define the notation used in this paper. For a vector $\mathbf{x} = (x_1, x_2, \ldots x_n)^\top$, define $\|\mathbf{x}\|_\infty = \max_{1 \le i \le n} |x_i|$. For a matrix $V := (v_{i,j})$, define $\|V\|_\infty$ to be $\|V\|_\infty = \sup_{\mathbf{x} \neq 0} \|V\mathbf{x}\|_\infty / \|\mathbf{x}\|_\infty = \max_{1 \le i \le n} \sum_{j=1}^n |V_{i,j}|$. Define $\|V\|_{\max} := \max_{1 \le \{i,j\} \le n} |V_{i,j}|$. We will need to approximate the inverse of some V matrix of certain structures. Specifically, following Yan et al. (2016a), we call a $(2n-1) \times (2n-1)$ matrix $V = (v_{i,j})$ to belong to the matrix class $\mathcal{L}_n(m, M)$, if V satisfies: $m \le v_{i,j} - \sum_{j=n+1}^{2n-1} v_{i,j} \le M$, $i = 1, \ldots, n-1$; $v_{n,n} = \sum_{j=n+1}^{2n-1} v_{nj}$; $v_{i,j} = 0$, $i, j = 1, \ldots, n, i \neq j$; $v_{i,j} = 0$, $i, j = n+1, \ldots, 2n-1, i \neq j$; $m \le v_{i,j} = v_{ji} \le M$, $i = 1, \ldots, n, j = n+1, \ldots, 2n-1, j \neq n+i$; $v_{i,n+i} = v_{n+i,i} = 0$, $i = 1, \cdots, n-1$; and $v_{i,i} = \sum_{k=1}^n v_{ki} = \sum_{k=1}^n v_{ik}$, $i = n+1, \cdots, 2n-1$. If $V \in \mathcal{L}_n(m, M)$, then V is symmetric, element-wise non-negative and diagonally dominant, thus invertible. Moreover, by Simons and

Yao (1999) and Yan et al. (2016a), V^{-1} can be well-approximated by $S = \mathcal{S}(V)$, as follows

$$
s_{i,j} = \begin{cases}
\dfrac{\delta_{i,j}}{v_{i,i}} + \dfrac{1}{v_{2n,2n}}, & i,j = 1,\ldots,n, \\[2mm]
-\dfrac{1}{v_{2n,2n}}, & i = 1,\ldots,n,\ j = n+1,\ldots,2n-1, \\[2mm]
-\dfrac{1}{v_{2n,2n}}, & i = n+1,\ldots,2n-1,\ j = 1,\ldots,n, \\[2mm]
\dfrac{\delta_{i,j}}{v_{i,i}} + \dfrac{1}{v_{2n,2n}}, & i,j = n+1,\ldots,2n-1,
\end{cases}
\tag{1}
$$

where $\delta_{i,j} := \mathbb{1}_{[i=j]}$ and $v_{2n,i} = v_{i,2n} := v_{ii} - \sum_{j=1,j\neq i}^{2n-1} v_{ij}$ for $i = 1,\ldots,2n-1$ and $v_{2n,2n} = \sum_{i=1}^{2n-1} v_{2n,i}$.

2 | OUR METHOD

2.1 | Data structure and our model

We start with presenting the data structure. We observe network snapshots at N randomly selected time points $T_1,\ldots,T_N$ selected by a probability density function $f_T(t)$ with sample space $[a,b]$ for some constants $a < b$. All snapshots share a common node set $[n] = \{1,\ldots,n\}$ across different time points. At each time point $t \in \mathcal{T} := \{T_1,\ldots,T_N\}$, we observe a directed binary network, represented by its adjacency matrix $A(t) := \{A_{i,j}(t)\}_{1\leq\{i,j\}\leq n} \in \{0,1\}^{n\times n}$, where $A_{i,j}(t) = 1$ if there is a directed edge from i to j at time t, and $A_{i,j}(t) = 0$ otherwise. For simplicity, we assume all edges are generated independently from each other. We assume no self-loops: $A_{i,i}(t) \equiv 0$.

To model this data structure, we propose a time-varying β-model for directed networks. At any time $t \in \mathcal{T}$, each node i is associated with two parameters $\alpha_i(t)$ and $\beta_i(t)$, encoding the strengths of the sender's effect and receiver's effect, respectively. The edge probability from i to j is

$$
P(A_{i,j}(t) = 1) = \frac{e^{\alpha_i(t)+\beta_j(t)}}{1 + e^{\alpha_i(t)+\beta_j(t)}} \quad \text{and} \quad P(A_{i,j}(t) = 0) = 1 - P(A_{i,j}(t) = 1).
\tag{2}
$$

Similar to directed and bipartite β-models for static networks (Yan et al., 2016a; Fan et al., 2022), here, we need to enforce an additional regularity conditions to ensure parameter identifiability. For simplicity, we set $\beta_n(t) = 0$ (Yan et al., 2016a). This will also guarantee the uniqueness of the solution to our estimation equation set (6) and (7), which we shall present in Section 2.2. Let

$$
\theta(t) := \Big(\alpha_1(t),\ldots,\alpha_n(t),\beta_1,\ldots,\beta_{n-1}(t)\Big)^\top
$$

to denote all free parameters at t.

Next, we describe our assumption on the relationship between the model parameters at different time points. We naturally anticipate that $\theta(t)$ and $\theta(t')$ should be similar for close-by time points $t \approx t'$, unless there exists a change point in between; while those parameters at distant time points might possibly be very different (or they might not). For simplicity, we stick to the continuous $\theta(t)$ setting in our method development and analysis. We will discuss how our method can be slightly tweaked to handle change points at the end of Section 5.

2.2 | Parameter estimation

Our estimation method is semi-parametric. The parametric flavor of our approach is reflected by (2); whereas the nonparametric flavor lies in that we utilize the smoothness of $\theta(t)$ over t to enhance estimation accuracy. Given the observed time points $\mathcal{T}$, the log-likelihood function is

$$
\begin{aligned}
\mathcal{L}_0(\{(t, \theta(t))\}_{t \in \mathcal{T}}) &= \log \left[\prod_{t \in \mathcal{T}} \prod_{(i,j):1 \leq \{i \neq j\} \leq n} \frac{e^{A_{i,j}(t)\{\alpha_i(t) + \beta_j(t)\}}}{1 + e^{\alpha_i(t) + \beta_j(t)}} \right] \\
&= \sum_{t \in \mathcal{T}} \left\{ \sum_{i=1}^n \left(d_i(t)\alpha_i(t) + b_i(t)\beta_i(t) \right) - \sum_{(i,j):1 \leq \{i \neq j\} \leq n} \log \left(1 + e^{\alpha_i(t) + \beta_j(t)} \right) \right\},
\end{aligned} \tag{3}
$$

where $d_i(t) = \sum_{j:1 \leq j \leq n, j \neq i} A_{i,j}(t)$ and $b_j(t) = \sum_{i:1 \leq i \leq n, i \neq j} A_{i,j}(t)$ denote the out-degree of node i and the in-degree of node j, respectively. Straightly applying a maximum-likelihood method on (3), separately for each time point, would produce an estimation of $\theta(t)$ using only the $A(t)$ at time t. That is, for every $i \in [n]$ and $j \in [n-1]$, define

$$
F_{i;0}(t, \theta(t)) := d_i(t) - \sum_{\substack{j:1 \leq j \leq n \\ j \neq i}} \frac{e^{\alpha_i(t) + \beta_j(t)}}{1 + e^{\alpha_i(t) + \beta_j(t)}}, \quad t \in \{T_1, \dots, T_N\} \tag{4}
$$

$$
F_{n+j;0}(t, \theta(t)) := b_j(t) - \sum_{\substack{i:1 \leq i \leq n \\ i \neq j}} \frac{e^{\alpha_i(t) + \beta_j(t)}}{1 + e^{\alpha_i(t) + \beta_j(t)}}, \quad t \in \{T_1, \dots, T_N\} \tag{5}
$$

$$
F_0(t, \theta(t)) := (F_1(t, \theta(t)), \dots, F_{2n-1}(t, \theta(t)))^\top.
$$

This approach estimates $\theta(t)$ by the solution to $F_0 = 0$. Then one can directly apply the algorithm and theory of Yan et al. (2016a). We call this method "point-wise estimation" throughout this paper and will use it as a benchmark for comparison in our simulation studies.

Our goal is to estimate $\theta(t)$ for all $t \in [a, b]$, not just at those observed time points $t \in \mathcal{T}$. This could not be achieved by the point-wise estimator (4) and (5). Our main idea to estimate $\theta(t)$ for all $t \in [a, b]$ is to borrow information from T_ℓ's close to t – in fact, this will also improve the estimation at time points $t \in \mathcal{T}$. Bearing this in mind, we generalize (4) and (5), replacing the observed out- and in-degrees $d_i(t)$ and $b_i(t)$ at time t by their kernel-smoothed versions incorporating information from nearby time points. More precisely speaking, define

$$
F_i(t, \theta(t)) := \sum_{\ell \in [N]} K_h(t - T_\ell) \left\{ d_i(T_\ell) - \sum_{\substack{j:1 \leq j \leq n \\ j \neq i}} \frac{e^{\alpha_i(t) + \beta_j(t)}}{1 + e^{\alpha_i(t) + \beta_j(t)}} \right\}, \quad t \in [a, b], 1 \leq i \leq n \tag{6}
$$

$$
F_{n+j}(t, \theta(t)) := \sum_{\ell \in [N]} K_h(t - T_\ell) \left\{ b_j(T_\ell) - \sum_{\substack{i:1 \leq i \leq n \\ i \neq j}} \frac{e^{\alpha_i(t) + \beta_j(t)}}{1 + e^{\alpha_i(t) + \beta_j(t)}} \right\}, \quad t \in [a, b], 1 \leq j \leq n - 1 \tag{7}
$$

where the kernel function $K_h(u) := (1/h) \cdot K(u/h)$ is even, supported on $[-1, 1]$ and satisfies $\int_{\mathbb{R}} K(u) \, du = 1$.

Our estimator, denoted by $\widehat{\theta}(t)$, is the solution to the estimation equation set $F(t, \theta(t)) = 0$. For a fluent narration, we relegate the choice of the bandwidth h to Section 3.1 and focus on solving (6) and (7), for which we shall employ

Newton's method. The Jacobian matrix of $F(t, \theta(t))/(Nn)$, denoted by $V(t, \theta(t))$, is

$$V(t, \theta(t)) := \frac{1}{N} \sum_{\ell \in [N]} K_h(t - T_\ell) V_0(t, \theta(t)),$$

where $V_0(t, \theta(t)) = V_0(t, (\alpha(t), \beta(t)))$ is further defined as follows

$$\{V_0(t, \theta(t))\}_{i,n+j} = \{V_0(t, \theta(t))\}_{n+i,j} := \frac{1}{n} \frac{e^{\alpha_i(t) + \beta_j(t)}}{\{1 + e^{\alpha_i(t) + \beta_j(t)}\}^2}, \qquad i \in [n], \ j \in [n-1], \ i \neq j,$$

$$\{V_0(t, \theta(t))\}_{i,i} := \sum_{\substack{i:1 \leq i \leq n \\ j \neq i}} \{V_0(t, \theta(t))\}_{i,n+j}, \qquad\qquad i \in [n],$$

$$\{V_0(t, \theta(t))\}_{n+j,n+j} := \sum_{\substack{j:1 \leq j \leq n-1 \\ i \neq j}} \{V_0(t, \theta(t))\}_{i,n+j}, \qquad\qquad j \in [n-1],$$

and $\{V_0(t, \theta(t))\}_{i,j} = 0$ otherwise. Since $V(t, \theta(t))$ is element-wise non-negative and strictly diagonally dominant, it is positive definite. Therefore, when the solution to the estimation equations (6) and (7) exists, it must be unique and can be found by a gradient descent or Newton's method (Bubeck, 2015). We will address the existence of $\widehat{\theta}(t)$ and formally characterize its accuracy in Theorem 1.

2.3 | Theoretical properties of the estimator

Recall that our estimator $\widehat{\theta}(t)$ is obtained by solving (6) and (7). In this section, we establish three aspects of theoretical guarantee for $\widehat{\theta}(t)$: existence, error rate and asymptotic normality. Let $\theta^*(t) := \left(\alpha_1^*(t), \ldots, \alpha_n^*(t), \beta_1^*(t), \ldots, \beta_{n-1}^*(t)\right)^\top$ be the true parameters. Define

$$Q_{Nnh} = \sup_{t \in [a,b]} \max_{i,j:1 \leq \{i \neq j\} \leq n} \frac{(1 + e^{\alpha_i^*(t) + \beta_j^*(t)})^2}{e^{\alpha_i^*(t) + \beta_j^*(t)}}.$$

Readers familiar with the β-model literature may immediately notice the analogy between our Q_{Nnh} and its counterpart, usually denoted by b_n (Fan et al., 2022), in the static network setting. Now we describe some mild regularity conditions that we would need.

Condition 1 *The parameters $\theta^*(t)$ is element-wise twice continuously differentiable. Moreover, there exists a constant or diverging deterministic series C_{Nn} such that*

$$\max_{i \in [2n-1]} \left| \frac{d\theta_i^*(t)}{dt} \right| \leq C_{Nn} \quad and \quad \max_{i \in [2n-1]} \left| \frac{d^2\theta_i^*(t)}{dt^2} \right| \leq C_{Nn},$$

where C_{Nn} satisfies

$$C_{Nn} h \to 0 \quad as \quad Nn \to \infty.$$

Condition 2 *Assume $f(t) > 0$ for all $t \in [a, b]$, and is twice continuously differentiable.*

Remark Condition 1 allows $\theta^*(t)$ to change increasingly rapidly as $Nn \to \infty$. This is not surprising since a growing Nn

brings an increasing amount of information. Condition 2 is a very mild condition that we choose observation times from $[a, b]$ in a balanced fashion, so no part of the entire time period $[a, b]$ is ignored. Let $k_{21} := \int v^2 K(v) dv < \infty$. Now we are ready to present the main theorem on the existence and uniform consistency of our estimator.

Theorem 1 *Suppose Conditions 1 and 2 hold. Assume $h \to 0$, $Nh \to \infty$ and either n remains constant or $n \to \infty$. If*

$$Q_{Nnh} = o\left\{ \frac{1}{\left(\sqrt{C_1 \log(Nnh)/(Nnh)} + C_2 C_{Nn} h^2 \right)^{1/6}} \right\}, \tag{8}$$

then the estimator $\widehat{\theta}(t)$ exists and satisfies

$$\sup_{t \in [a,b]} \|\widehat{\theta}(t) - \theta^*(t)\|_\infty = O_p\left\{ Q_{Nnh}^3 \left(\sqrt{C_1 \frac{\log(Nnh)}{Nnh}} + C_2 C_{Nn} h^2 \right) \right\}, \tag{9}$$

where $C_1 > 6$ is a constant and $C_2 > k_{21} \sup_{t \in [a,b]} (f'(t) + f(t)/2)$.

The condition (8) in Theorem 1 appears stronger than what is needed to send the right hand side of (9) to zero. One may naturally wonder if it can be substantively improved. The answer is "not easily". The precise reason is technical, but a quick explanation is that (8) not only contributes to uniform consistency, but also the existence of the solution $\widehat{\theta}(t)$. It turns out that the existence of solution demands stronger assumptions on Q_{Nnh} in the analysis, but this point is not explicitly reflected in (9). The phenomenon that the assumption on Q_{Nnh} is stronger than the apparent need (for the error bound to diminish) is frequently reported in many related works on the β-model, including (Chatterjee et al., 2011; Yan et al., 2016a; Chen et al., 2021; Stein and Leng, 2021; Shao et al., 2021).

The uniform convergence rate in Theorem 1 has the familiar bias-variance trade-off form, thus the bandwidth should be carefully selected to properly balance bias and variance. Theoretically, if we know C_{Nn}, then we can optimize the choice of h analytically. For example, if C_{Nn} is constant, then we can choose $h \asymp (Nn)^{-1/5}$, which ensures that the right hand side of (9) is $o_p(1)$. But in practice, it might not always be easy to know C_{Nn}. Therefore, in Section 3, we develop a leave-one-out cross-validation procedure to select h. We find it to perform well in our numerical studies.

Remark The classical literature on the β-model typically assumes very dense networks, see the overviews of this issue in Chen et al. (2021); Shao et al. (2021) and the references therein. In our work, to guarantee consistency, we need $Q_{Nnh} = o\left\{ 1/(\sqrt{C_1 \log(Nnh)/(Nnh)} + C_2 C_{Nn} h^2)^{1/6} \right\}$. This yields the following lower bound on network sparsity

$$\min_{1 \le i < j \le n} P(A_{i,j}(t) = 1) \ge \frac{1}{Q_{Nnh}} = \omega\left(\left\{ \sqrt{C_1 \frac{\log(Nnh)}{Nnh}} + C_2 C_{Nn} h^2 \right\}^{1/6} \right). \tag{10}$$

To better decipher (10), let $N = O(n^\psi)$ and notice that when $\psi > 1/4$, we can set $h = O\{(NnC_{Nn})^{-1/5}\}$ and further simplify (10) into

$$\left\{ \sqrt{C_1 \frac{\log(Nnh)}{Nnh}} + C_2 C_{Nn} h^2 \right\}^{1/6} = \frac{\log(Nn)^{1/12} C_{Nn}^{1/30}}{(Nn)^{1/15}}.$$

Consider a simple case where $C_{Nn} = O(1)$. Compared to the requirements of $\rho_n \gtrsim \log^{-C} n$ for some C in Yan and Xu (2013) and $\rho_n \gtrsim n^{-1/12}$ in Yan et al. (2016a), we accommodate sparser networks (recall $N \gg n^{1/4}$). Notice that in this simple illustration, our choice of h is still tailored to minimize the right hand side of (9). If the network becomes

sparser, we would need to choose larger h values – this is understandable, since if we observe too few edges from network snapshots near one time point, the natural thing to do is to expand h to incorporate data from a wider time window; but this of course may inflate bias. An interesting future work is to handle very sparse networks, for which purpose, some regularization might be necessary (Chen et al., 2021; Stein and Leng, 2021; Shao et al., 2021).

Next, we establish the asymptotic normality of our estimator under mild conditions. As a preparation, let us set up some shorthand. Define $k_{02} := \int K^2(v) dv$, $u_{ij}(t) := \left\{ e^{\alpha_i^*(t) + \beta_j^*(t)} \right\} / \left\{ 1 + e^{\alpha_i^*(t) + \beta_j^*(t)} \right\}$ and $\mu_{ij}(t) := u_{ij}'(t) f'(t) + \frac{1}{2} u_{ij}''(t) f(t)$. To explicitly express the asymptotic distribution formula, define $\bar{S}(t, \theta(t)) = \mathcal{S}(\bar{V}(t, \theta(t)))$, where recall $\mathcal{S}(\cdot)$ from (1) and define $\bar{V}(t, \theta(t)) = f(t) V_0(t, \theta(t))$. Now we are ready to state the result.

Theorem 2 *Suppose Conditions 1 and 2 hold. Assume $Nnh^5 C_{Nn}^2 Q_{Nnh}^2 = o(1)$, $h \to 0$, $Nh \to \infty$, n either stays constant or diverges, and*

$$Q_{Nnh} = o\left\{ \frac{1}{(Nnh)^{1/18} \left(\sqrt{C_1 \log(Nnh)/(Nnh)} + C_2 C_{Nn} h^2 \right)^2} \right\},$$

where C_1 and C_2 are defined in Theorem 1. Then for any fixed $1 \le p \le 2n-1$ and $t \in [a, b]$, $\sqrt{Nnh} \left[\{ \widehat{\theta}_1(t) - \theta_1^*(t), \dots, \widehat{\theta}_p(t) - \theta_p^*(t) \}^\top - k_{21} h^2 \{ \mu_1(t), \dots, \mu_p(t) \}^\top \right]$ converges in distribution to a p-dimensional multivariate normal, with mean zero and covariance matrix given by the upper left $p \times p$ block of $k_{02} \bar{S}(t, \theta^*(t)) \bar{V}(t, \theta^*(t)) \bar{S}(t, \theta^*(t))^\top$, where

$$\mu_q(t) = \left\{ \frac{\sum_{j \in [n], j \ne q} \mu_{qj}(t)/n}{f(t) \{ V_0(t, \theta^*(t)) \}_{q,q}} \right\} \{ 1 - \mathbb{1}_{n+1 \le q \le 2n-1} \} + \left\{ \frac{\sum_{j \in [n], j \ne q-n} \mu_{j,q-n}(t)/n}{f(t) \{ V_0(t, \theta^*(t)) \}_{q,q}} \right\} \mathbb{1}_{n+1 \le q \le 2n-1}$$

$$+ (-1)^{n+1 \le q \le 2n-1} \frac{\sum_{j \in [n-1]} \mu_{jn}(t)/n}{f(t) \{ V_0(t, \theta^*(t)) \}_{2n,2n}},$$

for all $q = 1, \dots, p$, and $\{ V_0(t, \theta^*(t)) \}_{2n,2n} = 2 \sum_{i \in [n]} \{ V_0(t, \theta^*(t)) \}_{i,i} - \sum_{i \in [n]} \sum_{j \in [2n-1]} \{ V_0(t, \theta^*(t)) \}_{i,j}$.

Remark We present two quick understandings of the result of Theorem 2. First, this theorem allows n to be fixed or divergent. If $n \to \infty$, then $p \le n$ always holds after a certain point. Second, the bias term $\{ \sqrt{Nnh} k_{21} h^2 \mu_p(t) \}$ in Theorem 2 is uniformly bounded by $O(\sqrt{Nnh} C_{Nn} Q_{Nnh} h^2)$, thus asymptotically diminishing.

To put the result of Theorem 2 to practice, we estimate the asymptotic variance. Note that k_{02} is known and

$$\Sigma(t, \theta(t)) := \bar{S}(t, \theta(t)) \bar{V}(t, \theta(t)) \bar{S}(t, \theta(t))^\top$$

is indeed determined by $\bar{V}(t, \theta(t))$ where recall the definition of $\mathcal{S}(\cdot)$ from Section 1.1. Setting

$$\widehat{V}(t, \widehat{\theta}(t)) = \frac{1}{N} \sum_{\ell \in [N]} K_h(t - T_\ell) V_0(t, \widehat{\theta}(t)) \tag{11}$$

and $S(t, \widehat{\theta}(t)) = \mathcal{S}(\widehat{V}(t, \widehat{\theta}(t)))$, the estimated asymptotic variance is then $k_{02} \cdot \widehat{\Sigma}(t, \widehat{\theta}(t))$, where

$$\widehat{\Sigma}(t, \widehat{\theta}(t)) = S(t, \widehat{\theta}(t)) \widehat{V}(t, \widehat{\theta}(t)) S(t, \widehat{\theta}(t))^\top.$$

Now we confirm that the variance estimator is indeed consistent.

Proposition 3 *Suppose that Conditions 1 and 2 hold. If $Nh \to \infty$ as $N \to \infty$, then we have*

$$\|\widehat{\Sigma}(t, \widehat{\theta}(t)) - \Sigma(t, \theta^*(t))\|_{\max} = o_p(1).$$

In practice, we first estimate $\widehat{V}(t, \widehat{\theta}(t))$ using (11); then plug it into the approximate inversion formula $\mathcal{S}(\cdot)$ in Section 1.1 to obtain $\mathcal{S}(t, \widehat{\theta}(t))$ and $\widehat{\Sigma}(t, \widehat{\theta}(t))$.

3 | SIMULATIONS

In this section, we conduct numerical experiments to gain some empirical understandings in several aspects of our method, including: the impact of different h choices, estimation accuracy, computation speed and the match between numerical result and our theory's prediction. Throughout all experiments, we set $[a, b] = [0.1, 0.9]$ and $f(t) \sim \text{Uniform}(a, b)$. To evaluate our method's effectiveness in estimating different parameter evolution dynamics, we set up $\alpha_i^*(t)$'s and $\beta_j^*(t)$'s each of four different types, as is shown in Table 1.

TABLE 1 Simulation parameter set up.

Index	$1 \le i \le n/4$	$n/4 + 1 \le i \le 2n/4$	$2n/4 + 1 \le i \le 3n/4$	$3n/4 + 1 \le i \le n - 1$
$\alpha_i^*(t)$	$-0.8(3t - 0.6)$	$-3.6(t - 0.3)^2/(t - 3)$	-0.4^t	$-3.2(t - 0.5)^2/(1 + t^2)$
$\beta_i^*(t)$	$2(t - 0.5)^2/(t - 2)$	$-1.6t(t - 0.3)^3 + (t - 0.2)^2 + 0.2t$	$-1.8(t + 0.6)\sin(0.2\pi t)$	$-1.6(t - 0.2)^2 \sin(\pi t)$

Set $\alpha_n^*(t) = \alpha_{n-1}^*(t)$ and $\beta_n^*(t) \equiv 0$. For narration simplicity, we stick to the kernel function $K(x) = 0.75(1 - x^2)\mathbb{1}_{[|x|<1]}$ in all simulations.

3.1 | Simulation 1: Select the tuning parameter h

In this simulation, we consider two settings: $(n, N) = (40, 100)$ and $(160, 200)$. The purpose of this simulation is to compare the optimal h's between these two settings and check if their ratio matches the theoretical suggestion of our Theorem 1. In each setting, we select the optimal h from the range $[0.05, 0.3]$ by a leave-one-out cross-validation with loss function

$$L_{\text{CV}}(h) := \sum_{\ell=1}^{N} \|A(T_\ell) - \widehat{W}_h(T_\ell)\|_F^2,$$

where $A(T_\ell)$ is the observed adjacency matrix at time $t = T_\ell$, and $\widehat{W}_h(T_\ell)$ is the estimated edge probability matrix based on data at times $\{T_{\ell'} : \ell' = 1, \ldots, N, \ell' \neq \ell\}$.

We repeated experiments 50 times for both settings. The average optimal $h_{\text{opt}}^{(n,N)}$ selected by cross-validation in the two settings are $\bar{h}_{\text{opt}}^{(40,100)} = 0.225$ and $\bar{h}_{\text{opt}}^{(160,200)} = 0.140$. Their ratio is 0.622, which well-matches our Theorem 1's prediction of $\{(40 \times 100)/(160 \times 200)\}^{(1/5)} = 0.659$.
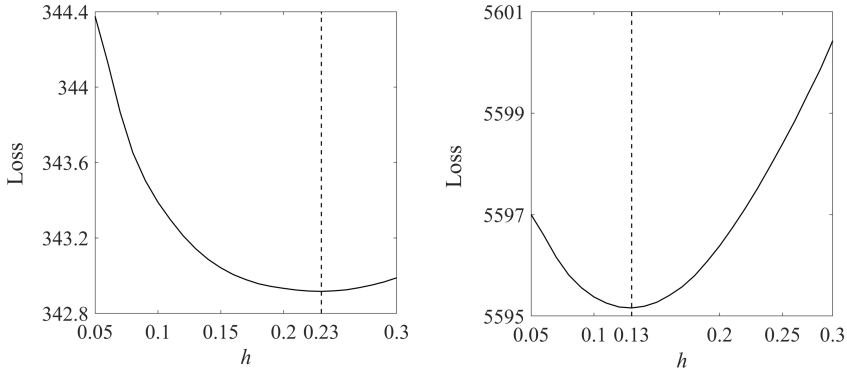
FIGURE 1 Left: $n = 40$, $N = 100$; Right: $n = 160$, $N = 200$. Both plots show only one experiment.

3.2 | Simulation 2: RMSE diminishing rate and time cost

In this simulation, we focus on evaluating estimation errors as the network size n and number of time points N vary. Following the result of Simulation 1 and our theory, we set $h = 0.23\{(40 \times 100)/(Nn)\}^{1/5}$ from this simulation onward. We vary $n \in \{40, 80, 160, 320\}$ and $N \in \{25, 50, 100, 200\}$. The estimation error is measured by the following RMSE:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{\ell=1}^{N} (\widehat{\theta}(T_\ell) - \theta^*(T_\ell))^2}$$

We repeat all experiments 100 times and record the average value of RMSE. The estimation errors and running times are reported in Figure 2. From the plots, we see that $\log(\text{RMSE})$ diminishes at the rate roughly $n^{-1/5}$. Our theory
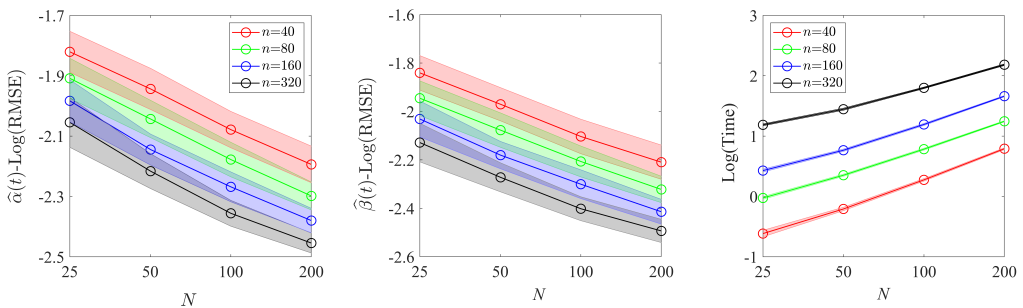

FIGURE 2 Left and Middle: log estimation errors for $\alpha(t)$ and $\beta(t)$ parameters, respectively. Right: log time cost. Shades indicate Monte Carlo standard deviations across repeated experiments.

predicts an error rate of $O(Q_{Nnh}^3 \cdot n^{-2/5})$ that agrees with the observed result of this experiment. Specifically, the time cost seems linear in N, which also matches our theoretical prediction; also, the vertical increment in log time cost as n

doubles is reasonably close to $\log 2 = 0.69$, matching our theoretical understanding that the computation complexity scales quadratically in n.

3.3 | Simulation 3: Fitted curves with comparison to point-wise method

To assess the performance of our method, we plot the true curve $\theta(t)$ and compare it to $\widehat{\theta}(t)$. We compare our method with two alternative approaches: (1) the point-wise method as described in Section 2.2; and (2) the smoothed point-wise method which smooths the point-wise estimates using a kernel smoothing procedure. Here, we set $n = 160$, $N = 200$ and $h = 0.13$ as suggested by the result of Simulation 1. We repeat the experiment 100 times. See Table 1 for the settings of $\alpha_i^*(t)$ and $\beta_j^*(t)$.

Figures 3 and 4 illustrate the result; and the numerical outputs are summarized in Tables 2 and 3. Our method shows smaller variance and significantly smaller bias compared to the smoothed point-wise method and a much bigger advantage over the point-wise method. As aforementioned, our method can borrow data from neighboring time points to improve estimation; whereas the point-wise method's curve fluctuates across different time points due to a limited n – this aligns with our theoretical understanding by comparing our Theorem 1 with classical error rate results at one time point, c.f. Yan et al. (2016a).

Next, we experiment in a sparse network setting. The settings for parameters are given in Table 4. Here we set $n = 40$, $N = 100$ and $h = 0.23$ as suggested by the result of Simulation 1. The results are shown in Figure 5. We see that the estimators of the point-wise method do not exist at some time points, and this also impacts the smoothed point-wise method because it smooths the result of the point-wise estimation method. In sharp contrast, our method works stably by incorporating data from neighboring time points at the earliest stage in the estimation.

TABLE 2 Simulation results of the Bias and SD for $\widehat{\alpha}_i(t)$ when $n = 160$, $N = 200$, $h = 0.13$.

|		Our method		Point-wise method		Smoothed point-wise method	
t	$\alpha^*(t)$	Bias	SD	Bias	SD	Bias	SD
0.2	$i = 1$	-0.004	0.041	0.019	0.233	-0.019	0.050
	$i = 41$	-0.008	0.029	-0.026	0.181	-0.017	0.049
	$i = 81$	0.013	0.037	0.016	0.204	0.033	0.057
	$i = 121$	-0.008	0.033	-0.022	0.205	-0.011	0.048
0.4	$i = 1$	-0.002	0.040	-0.037	0.227	0.012	0.054
	$i = 41$	-0.011	0.034	-0.047	0.216	-0.023	0.052
	$i = 81$	0.008	0.034	-0.016	0.214	0.018	0.055
	$i = 121$	-0.026	0.032	-0.056	0.209	-0.041	0.047
0.6	$i = 1$	0.004	0.042	-0.025	0.218	0.024	0.056
	$i = 41$	-0.019	0.038	-0.026	0.242	-0.019	0.050
	$i = 81$	-0.008	0.037	-0.018	0.238	0.000	0.049
	$i = 121$	-0.030	0.036	-0.045	0.208	-0.029	0.053
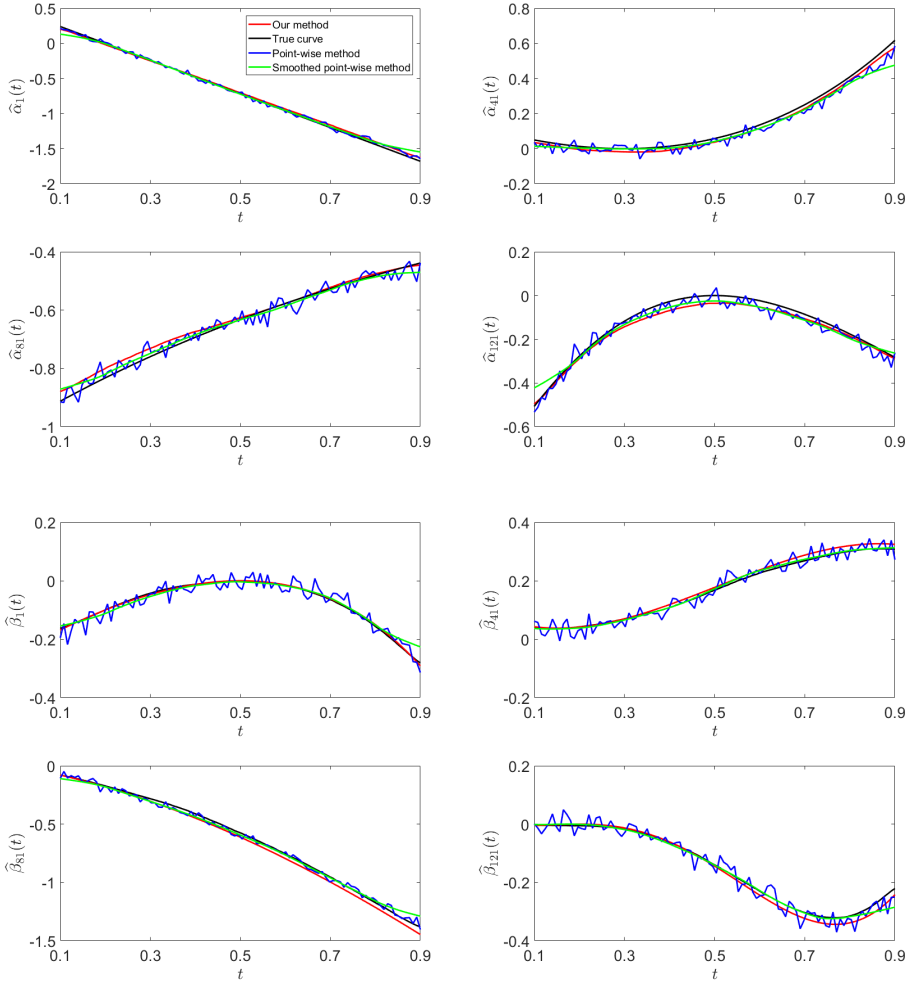
FIGURE 3 The estimate results of $\alpha_i^*(t)$ and $\beta_i^*(t)$ ($i = 1, 41, 81, 121$) with $n = 160$, $N = 200$ and $h = 0.13$.

3.4 | Simulation 4: Asymptotic normality

This simulation focuses on validating our asymptotic normality theory, namely, Theorem 2. For simplicity, we consider one particular setting $(n, N) = (160, 200)$. According to the result of Simulation 1, we set $h = 0.13$. We simulate 1000 Monte Carlo repetitions. Figure 6 and 7 illustrate the results. The plots suggest that the simulation result very well matches the prediction of our Theorem 2.
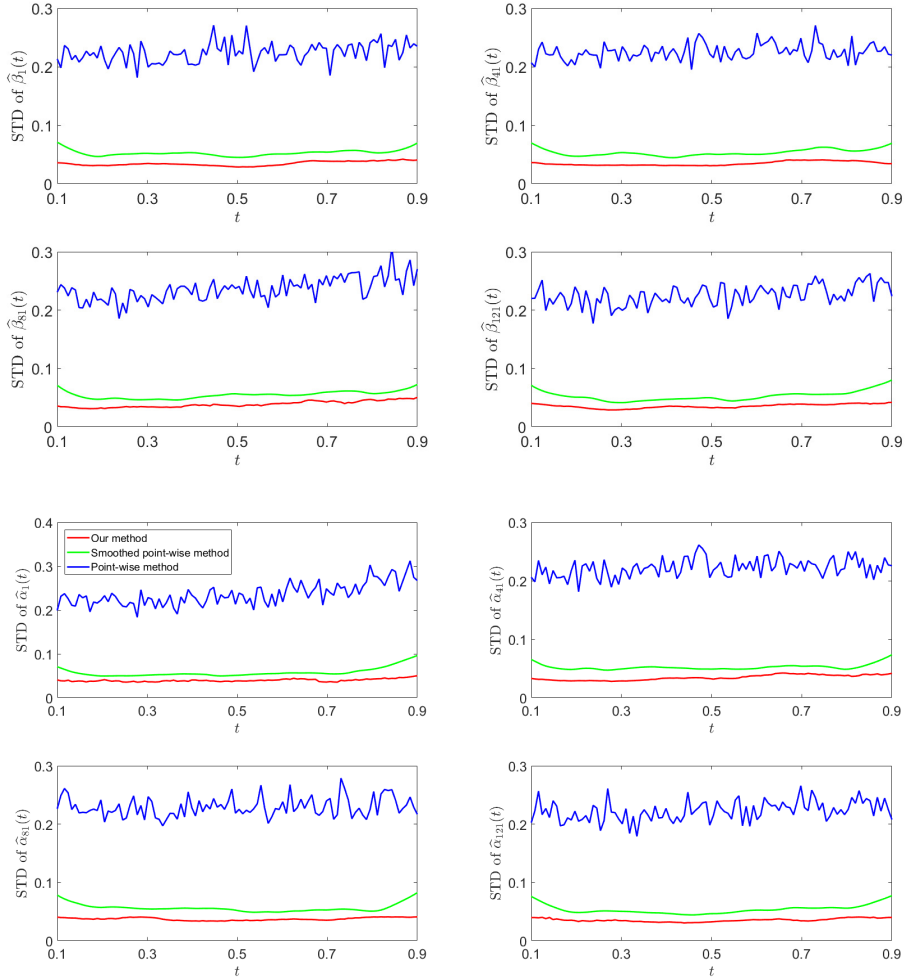
FIGURE 4 The sample standard deviations (SD) of estimators $\widehat{\alpha}_i(t)$ and $\widehat{\beta}_i(t)$ ($i = 1$, 41, 81, 121) with $n = 160$, $N = 200$ and $h = 0.13$.

4 | DATA EXAMPLE

We analyze an email data set Leskovec et al. (2007); Paranjape et al. (2017), available at `http://snap.stanford.edu/data/email-Eu-core-temporal.html`. It records email exchanges between members of a large European institution. Each entry (i, j, t) signifies that i emailed j at time t, where t is accurate to "day". We transcribe the individual entries into network snapshots by aggregating all communications by month – using a too small time unit will result too sparse network snapshots. This produces $N = 19$ snapshots over $(n_1, n_2, n_3, n_4) = (298, 154, 89, 140)$ nodes in four selected departments in the institute, respectively. For convenience, we rescale the entire time span to $[0, 1]$.

We use the cross-validation method in Section 3.1 to tune the bandwidth h and select $(h_1, h_2, h_3, h_4) =$
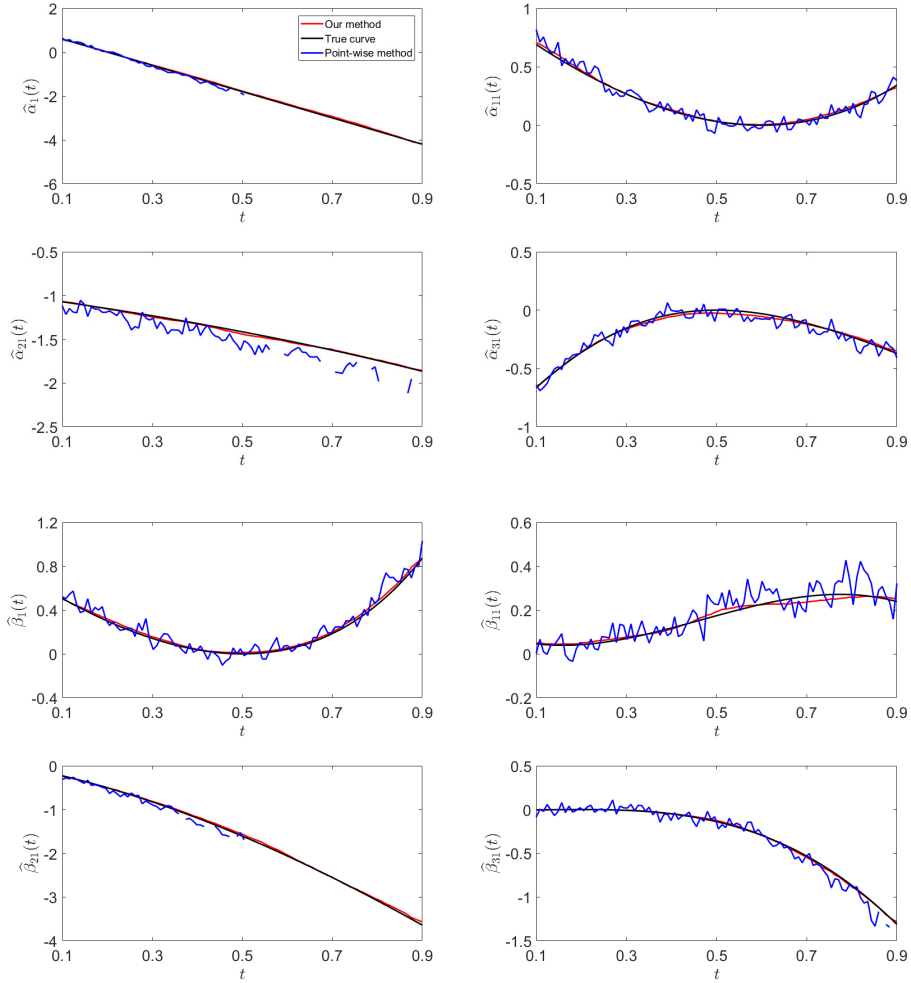
FIGURE 5 The estimate results of $\alpha_i^*(t)$ and $\beta_i^*(t)$ ($i = 1, 11, 21, 31$) with $n = 40$, $N = 100$ and $h = 0.23$.

$(0.21, 0.15, 0.46, 0.18)$ for the four departments, respectively. Using the selected h values, we estimate $\{\widehat{\alpha}_i(t)\}_{i=1}^n$ and $\{\widehat{\beta}_j(t)\}_{j=1}^{n-1}$ (recall $\widehat{\beta}_n(t) := \beta_n(t) \equiv 0$).

However, in real data analysis, we need to be very cautious with the interpretation of estimated $\alpha_i(t)$'s and $\beta_j(t)$'s. Let us consider a toy example, where $n = 2$; over time, node 1 becomes more active as both sender and receiver, whereas node 2 grows more inactive. One might think this leads to increasing $\alpha_1(t)$ and $\beta_1(t)$ and decreasing $\alpha_2(t)$ and $\beta_2(t)$. However, due to identifiability concerns, we follow the tradition in directed β-model analysis and set $\beta_2(t) \equiv 0$. The increasingly inactive behavior of node 2 as a receiver will "bounce" to all other parameters $\alpha_1(t), \alpha_2(t)$ and $\beta_1(t)$. Consequently, we cannot even be sure that $\alpha_1(t)$ and $\alpha_2(t)$ will be increasing.

Therefore, we should not merely look at the trends in each individual $\alpha_i(t)$ and $\beta_j(t)$ marginally. Fortunately, the above-mentioned identifiability issue will not affect $\alpha_i(t) - \alpha_j(t)$ or $\beta_i(t) - \beta_j(t)$ for any $i \neq j$. In view of this, after

TABLE 3 Simulation results of the Bias and SD for $\widehat{\beta}_i(t)$ when $n = 160$, $N = 200$, $h = 0.13$.

t	$\beta^*(t)$	Our method		Point-wise method		Smoothed point-wise method	
		Bias	SD	Bias	SD	Bias	SD
0.2	$i = 1$	0.001	0.031	0.018	0.212	-0.012	0.048
	$i = 41$	0.001	0.032	-0.007	0.223	-0.003	0.047
	$i = 81$	0.008	0.031	0.001	0.214	0.000	0.048
	$i = 121$	-0.005	0.034	-0.015	0.228	-0.001	0.051
0.4	$i = 1$	0.000	0.032	0.031	0.203	-0.004	0.053
	$i = 41$	-0.011	0.031	0.018	0.224	-0.011	0.045
	$i = 81$	0.037	0.038	0.056	0.247	0.010	0.050
	$i = 121$	-0.001	0.035	-0.003	0.212	-0.007	0.047
0.6	$i = 1$	-0.003	0.032	-0.018	0.220	-0.003	0.051
	$i = 41$	-0.010	0.035	-0.011	0.258	-0.002	0.050
	$i = 81$	0.040	0.040	0.059	0.260	0.031	0.054
	$i = 121$	0.014	0.035	-0.001	0.233	0.016	0.049

TABLE 4 The second set up of true parameters in Simulation 3.

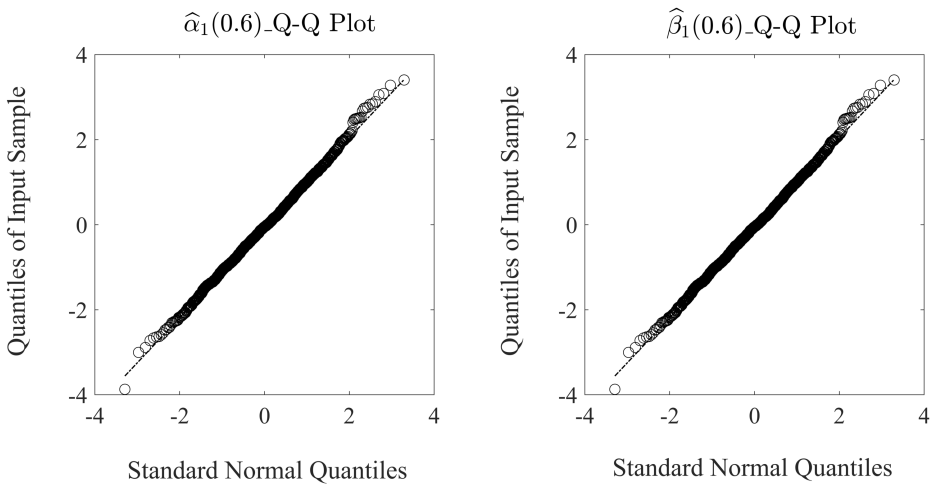| Index	$1 \le i \le n/4$	$n/4 + 1 \le i \le 2n/4$	$2n/4 + 1 \le i \le 3n/4$	$3n/4 + 1 \le i \le n - 1$
$\alpha_i^*(t)$	$-2(3t - 0.6)$	$-8(t - 0.6)^2/(t - 3)$	-2^t	$-4.2(t - 0.5)^2/(1 + t^2)$
$\beta_i^*(t)$	$-6(t - 0.5)^2/(t - 2)$	$-2(t - 0.3)^3 + (t - 0.2)^2 + 0.2t$	$-4(t + 0.8)\sin(0.2\pi t)$	$-5(t - 0.2)^2 \sin(0.2\pi t)$

FIGURE 6 One-dimensional marginal normality.

obtaining the estimation $\widehat{\alpha}_i(t)$'s, we compute a pairwise distance matrix $\mathcal{D}_\alpha$ as follows

$$\left(\mathcal{D}_\alpha\right)^2_{i,j} := \int_a^b |\widehat{\alpha}_i(t) - \widehat{\alpha}_j(t)|^2 dt \tag{12}$$
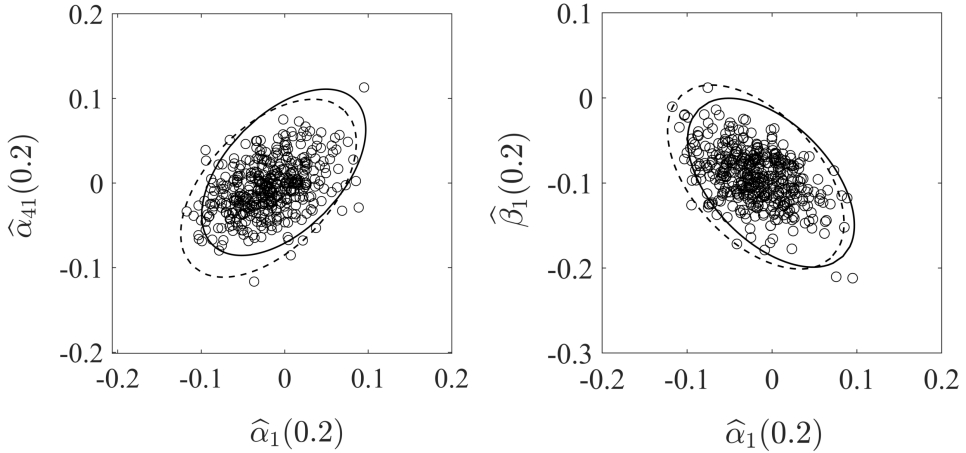
FIGURE 7 Two-dimensional marginal normality. Solid ellipse is the contour of $(x - \mu)^\top \Sigma^{-1}(x - \mu) = 9$, where (μ, Σ) are the asymptotic center and covariance matrix predicted by Theorem 2, respectively. Dashed ellipse is a similar contour but with (μ, Σ) being the sample mean and covariance matrix, respectively.

Then we perform a K-means clustering using $\mathcal{D}_\alpha$ and obtain an α-cluster label for each node. We clarify that the clustering here refers to the classification of nodes based on how their $\alpha_i(t)$ curve changes over time and has a very different meaning than the concept of "varying community memberships" in dynamic stochastic block models. Also, performing a multi-dimensional scaling (MDS) on $\mathcal{D}_\alpha$, projecting into $\mathbb{R}$, produces a one-dimensional latent space position for each node. Do the same using $\widehat{\beta}_j(t)$'s, we obtain a β-cluster label and β-MDS position coordinate on each node. A natural question is how to select K, the number of clusters, for each department's two clustering structures (α- and β-clustering). For this, we compute the ratio between the average between-cluster distance over the average within-cluster distance, for different choices of K. This suggests that we select $K = 4$ for α- and β-clustering for department 1; and $K = 5$ for clustering for departments 2–4 when $N = 19$. And we select $K = 4$ for α- and β-clustering for department 1 and department 4; and $K = 5$ for clustering for departments 2–3 when $N = 38$. And select $K = 5$ for α- and β-clustering for department 2; and $K = 4$ for clustering for other departments when $N = 57$. All details of this step are reported in Figure 8.

Next, in Figure 9, we show the clustering result in a colored and marked scatter plot of all nodes when $N = 19$ and the clustering results when $N = 38$ and $N = 57$ are in the Supplementary Material. Each node's X- and Y-coordinates are its α- and β-MDS coordinates, respectively. Each node's color indicates its α-clustering membership and marker shows its β-clustering membership. Now we present some interpretations of the results in Figure 9. First, we notice that the α- and β-clustering results mostly reflect nodes' α- and β-MDS cooridnates, but they are not entirely aligned – recall that α- and β-clustering use their corresponding distance matrices, whereas their MDS coordinates can be viewed as the distance matrices nonlinearly projected into one-dimensional subspace. All four plots in Figure 9 suggest that the trends in which nodes' sender and receiver effects evolve over time are overall concordant with each other, since most points align along diagonal straight lines; but they do not fully align with each other, which is reflected by the difference between α- and β- clustering results.
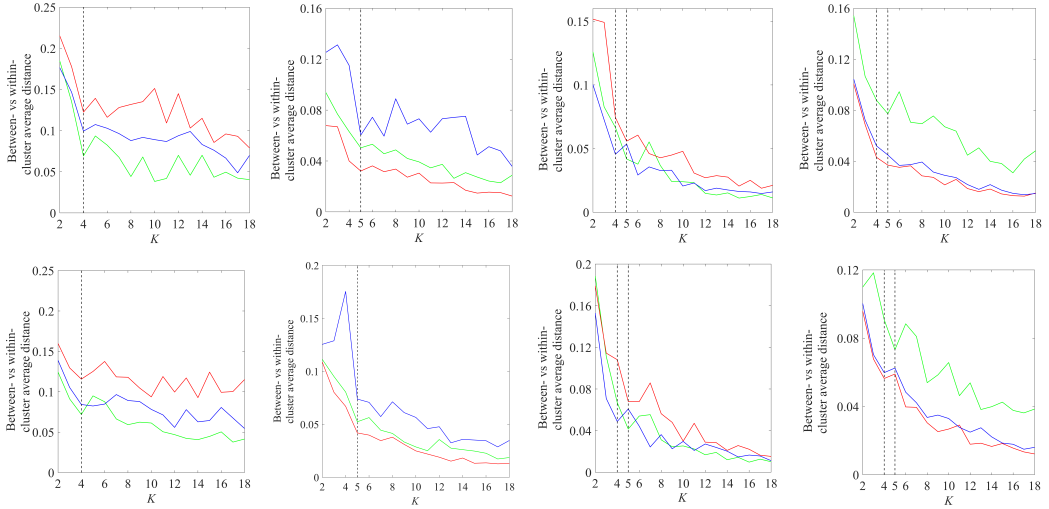
FIGURE 8 Green lines: $N = 19$; Red lines: $N = 38$; Blue lines: $N = 57$. Selecting the numbers of clusters for α- and β-clustering of each department in the email data set. Our selection balances two considerations: (i). selecting a small K; (ii) and minimizing the between- vs within- cluster average distance.

5 | DISCUSSION

In this paper, we extend the β-model for directed binary networks to the dynamic setting by letting the parameters evolve over time. This work can be viewed as a first step in the exploration of this interesting direction. Many intriguing questions remain. Here, we select a few future directions for discussion.

First, our study concentrates on binary edges. Retrospecting our approach, we see that the key ingredient is to kernel-smooth the partial differentiation formulae of the likelihood function at each time point, using the counterparts at adjacent time points. For instance, we can extend the edge generation scheme from Bernoulli to Poisson with $P(a_{ij} = x) = e^{x(\alpha_i + \beta_j)}/x! \cdot e^{-e^{\alpha_i + \beta_j}}$, $x = 0, 1, 2, \ldots$. The idea is very similar to how Yan et al. (2016b) extended the ordinary β-model from Bernoulli to Poisson and many other general edge distributions. We leave the detailed formulation and theoretical study to future work.

The second discussion topic is to relax our global smoothness assumption on all $\alpha_i(t)$'s and $\beta_j(t)$'s to accommodate change points (Wang et al., 2021; Bhamidi et al., 2018; Yu et al., 2021). To tweak our approach to this end, we consider a single change point setting: $\alpha_i(t)$'s and $\beta_j(t)$'s contain a common jump point at $t = t_0$, and are otherwise smooth conformal to Condition 1. Suppose $t_0 \in [a_1, b_1] \subset [a + C_a, b - C_b]$ for some global constants C_a and $C_b > 0$ (we assume a_1 and b_1 are known, but they can be set conservatively, close to a and b, respectively). A simple
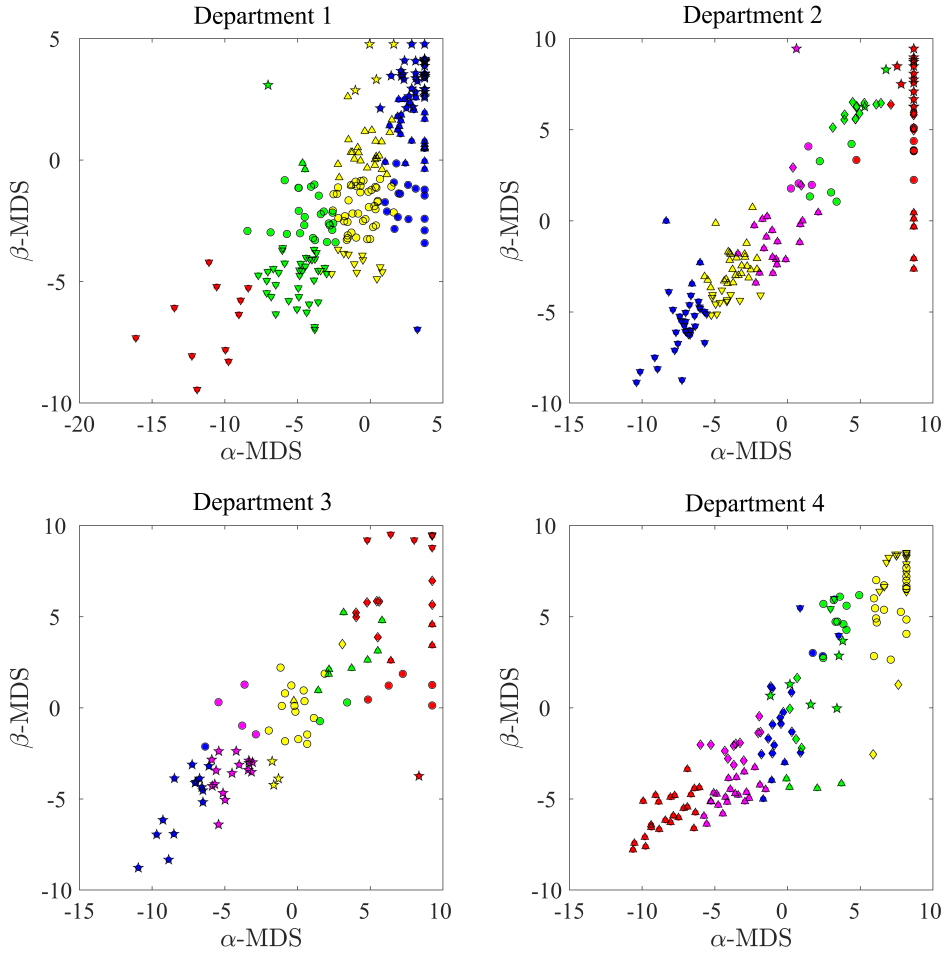
FIGURE 9 Illustration of $\widehat{\alpha}_i(t)$'s and $\widehat{\beta}_j(t)$'s. Each point's X-coordinate comes from a multi-dimensional scaling of the ℓ_2 distance matrix between $\widehat{\alpha}_i(t)$'s; similarly obtain its Y-coordinate. Color represents α-clustering result, and marker represents β-clustering result.

adaptation of our method for change point detection in this scenario is

$$\widehat{t}_0 := \arg \max_{t:t \in [a_1, b_1]} \left\| \left(\widehat{\alpha}_+(t) - \widehat{\alpha}_-(t), \widehat{\beta}_+(t) - \widehat{\beta}_-(t) \right) \right\|_2^2$$

where $\alpha_i(t) = \left(\widehat{\alpha}_+(t) \right)_i$ solves: $\sum_{\ell \in [N]:T_\ell \geq t} K_h(t - T_\ell) \left\{ d_i(T_\ell) - e^{\alpha_i(t)} \left(\sum_{j:j \neq i} e^{\beta_j(t)} \right) \right\} = 0,$

and $\alpha_i(t) = \left(\widehat{\alpha}_-(t) \right)_i$ solves: $\sum_{\ell \in [N]:T_\ell < t} K_h(t - T_\ell) \left\{ d_i(T_\ell) - e^{\alpha_i(t)} \left(\sum_{j:j \neq i} e^{\beta_j(t)} \right) \right\} = 0,$ (13)

and similarly define $\widehat{\beta}_+(t)$ and $\widehat{\beta}_-(t)$. Aside from (13), another way to extend the model is to generalize the weight

function, taking into account not only the closeness in time (reflected by $|t - T_\ell|$ for each t and T_ℓ), but also the closeness between their estimated parameters (Li et al., 2011). Both interesting extensions require further study.

Acknowledgments

References

Bhamidi, S., Jin, J. and Nobel, A. (2018) Change point detection in network models: Preferential attachment and long range dependence. *The Annals of Applied Probability*, **28**, 35–78.

Bubeck, S. (2015) Convex optimization: Algorithms and complexity. *Foundations and Trends in Machine Learning*, **8**, 231–357.

Chatterjee, S., Diaconis, P. and Sly, A. (2011) Random graphs with a given degree sequence. *The Annals of Applied Probability*, **21**, 1400–1435.

Chen, M., Kato, K. and Leng, C. (2021) Analysis of networks via the sparse β-model. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **83**, 887–910.

Cui, X. and Chen, Y. (2022+) Inferring social influence in dynamic networks. *Statistical Sinica (In press)*.

Fan, Y., Jiang, B., Yan, T. and Zhang, Y. (2022) Asymptotic theory in bipartite graph models with a growing number of parameters. *Canadian Journal of Statistics (In press)*.

Graham, B. S. (2017) An econometric model of network formation with degree heterogeneity. *Econometrica*, **85**, 1033–1063.

Hillar, C. and Wibisono, A. (2013) Maximum entropy distributions on graphs. *arXiv preprint arXiv:1301.3321*.

Leskovec, J., Kleinberg, J. and Faloutsos, C. (2007) Graph evolution: Densification and shrinking diameters. *ACM transactions on Knowledge Discovery from Data (TKDD)*, **1**, 2–es.

Li, Y., Zhu, H., Shen, D., Lin, W., Gilmore, J. H. and Ibrahim, J. G. (2011) Multiscale adaptive regression models for neuroimaging data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **73**, 559–578.

Masry, E. (1996) Multivariate local polynomial regression for time series: uniform strong consistency and rates. *Journal of Time Series Analysis*, **17**, 571–599.

Olhede, S. C. and Wolfe, P. J. (2012) Degree-based network models. *arXiv preprint arXiv:1211.6537*.

Paranjape, A., Benson, A. R. and Leskovec, J. (2017) Motifs in temporal networks. In *Proceedings of the tenth ACM international conference on web search and data mining*, 601–610.

Rinaldo, A., Petrovic, S. and Fienberg, S. E. (2013) Maximum likelihood estimation in the β-model, supplementary materials. *Ann. Stat.*, **41**, 1085–1110.

Shao, M., Zhang, Y., Wang, Q., Zhang, Y., Luo, J. and Yan, T. (2021) L-2 regularized maximum likelihood for β-model in large and sparse networks. *arXiv preprint arXiv:2110.11856*.

Simons, G. and Yao, Y.-C. (1999) Asymptotics when the number of parameters tends to infinity in the bradley-terry model for paired comparisons. *The Annals of Statistics*, **27**, 1041–1060.

Stein, S. and Leng, C. (2021) A sparse random graph model for sparse directed networks. *arXiv preprint arXiv:2108.09504.*

Wang, D., Yu, Y. and Rinaldo, A. (2021) Optimal change point detection and localization in sparse dynamic networks. *The Annals of Statistics*, **49**, 203–232.

Yan, T., Leng, C. and Zhu, J. (2016a) Asymptotics in directed exponential random graph models with an increasing bi-degree sequence. *The Annals of Statistics*, **44**, 31–57.

Yan, T., Qin, H. and Wang, H. (2016b) Asymptotics in undirected random graph models parameterized by the strengths of vertices. *Statistica Sinica*, 273–293.

Yan, T. and Xu, J. (2013) A central limit theorem in the β-model for undirected random graphs with a diverging number of vertices. *Biometrika*, **100**, 519–524.

Yu, Y., Padilla, O. H. M., Wang, D. and Rinaldo, A. (2021) Optimal network online change point localisation. *arXiv preprint arXiv:2101.05477.*

Zhang, Y. and Xia, D. (2022) Edgeworth expansions for network moments. *The Annals of Statistics*, **50**, 726–753.

Appendix: Proofs of Theorem 1, Theorem 2 and Proposition 3

Proof of Theorem 1. We show the consistency of the estimators by checking the conditions of Lemma 1 in the online Supplementary Material. We first calculate

$$r = \|(F'(t, \theta^{(0)}(t))/(Nn))^{-1} F(t, \theta^{(0)}(t)/(Nn))\|_\infty.$$

Note that $-F'(t, \theta(t))/(Nn) \in \mathcal{L}_n(m, M)$. As shown in Lemma 6, we have

$$\frac{1}{Nn} F'_i(t, \theta(t)) \le \epsilon + \frac{1}{Nn} \mathbb{E} F'_i(t, \theta(t)), \ i \in [2n-1],$$

where $\epsilon > 0$ is some constant. In addition, by some arguments similar to the proof of Lemma 6, we have

$$m = \frac{1}{NnQ_{Nnh}} \sum_{\ell \in [N]} K_h(t - T_\ell) \ge (f(t) + \frac{1}{2}k_{21} f''(t)h^2 + o(h^2))/(2nQ_{Nnh}),$$

$$M = \frac{1}{4Nn} \sum_{\ell \in [N]} K_h(t - T_\ell) \le (f(t) + \frac{1}{2}k_{21} f''(t)h^2 + o(h^2))/(2n).$$

By Lemmas 2 and 3, we have

$$
\begin{aligned}
r =& \| (F'(t, \theta^{(0)}(t))/(Nn))^{-1} F(t, \theta^{(0)}(t)/(Nn)) \|_{\infty} \\
\leq & \left(\frac{2c_4(2n-1)M^2 \|F(t,\theta^{(0)}(t)/(Nn))\|_{\infty}}{m^3(n-1)^2} + \max_i \frac{|F_i(t,\theta^{(0)}(t))/(Nn)|}{\{V(t,\theta^{(0)}(t))\}_{i,i}} + \frac{|F_{2n}(t,\theta^{(0)}(t))/(Nn)|}{\{V(t,\theta^{(0)}(t))\}_{2n,2n}} \right) \\
\leq & \left(\frac{2c_4(2n-1)M^2}{m^3(n-1)^2} + \frac{c_5}{(2n-1)mN} \right) \|F(t,\theta^{(0)}(t))/(Nn)\|_{\infty} \\
\leq & O\left(\frac{Q_{Nnh}^3 + Q_{Nnh}}{f(t) + k_{21}f''(t)h^2/2 + o(h^2)} \right) \left(\sqrt{C_1 \frac{\log(Nnh)}{Nnh}} + C_2 C_{Nn} h^2 \right) \\
\leq & O\left(Q_{Nnh}^3 \left(\sqrt{C_1 \frac{\log(Nnh)}{Nnh}} + C_2 C_{Nn} h^2 \right) \right),
\end{aligned}
$$

where c_4 and c_5 are some positive constants, and

$$
\{V(t,\theta(t))\}_{2n,2n} = 2 \sum_{i \in [n]} \{V(t,\theta(t))\}_{i,i} - \sum_{i \in [n]} \sum_{j \in [2n-1]} \{V(t,\theta(t))\}_{i,j}.
$$

The last equation holds under Condition 2. By Lemma 1, note that

$$
\rho = \frac{c_1(2n-1)M^2 H_1}{2m^3 n^2} + \frac{H_2}{(n-1)m} \leq O\left(Q_{Nnh}^3 + Q_{Nnh} \right)
$$

with probability tending to one, where c_1 is some positive constant. Therefore,

$$
\begin{aligned}
\rho r =& O\left(Q_{Nnh}^3 + Q_{Nnh} \right) \times O\left(Q_{Nnh}^3 \left(\sqrt{C_1 \frac{\log(Nnh)}{Nnh}} + C_2 C_{Nn} h^2 \right) \right) \\
=& O\left(Q_{Nnh}^6 \sqrt{C_1 \frac{\log(Nnh)}{Nnh}} + Q_{Nnh}^4 \sqrt{C_1 \frac{\log(Nnh)}{Nnh}} + C_2 C_{Nn} Q_{Nnh}^6 h^2 + C_2 C_{Nn} Q_{Nnh}^4 h^2 \right) \\
=& o(1).
\end{aligned}
$$

If $Q_{Nnh} = o\left(1/(\sqrt{C_1 \log(Nnh)/(Nnh)} + C_2 C_{Nn} h^2)^{1/6} \right)$, which is less than $1/2$ as $Nnh \rightarrow \infty$. Consequently,

$$
\sup_{t \in [a,b]} \|\widehat{\theta}(t) - \theta^*(t)\|_{\infty} = O_p\left(Q_{Nnh}^3 \left(\sqrt{C_1 \frac{\log(Nnh)}{Nnh}} + C_2 C_{Nn} h^2 \right) \right) = o_p(1).
$$

It completes the proof.

Proof of Theorem 2. By Theorem 1, we have

$$
\widehat{b}_{Nnh} := \sup_{t \in [a,b]} \max_{i \in [2n-1]} \left| \{\widehat{\theta}(t) - \theta^*(t)\}_i \right| = O_p\left\{ Q_{Nnh}^3 \left(\sqrt{C_1 \frac{\log(Nnh)}{Nnh}} + C_2 C_{Nn} h^2 \right) \right\}.
$$

Let $\widehat{\omega}_{ij}(t) = \widehat{\alpha}_i(t) + \widehat{\beta}_j(t) - \alpha_i^*(t) - \beta_j^*(t)$. By Taylor's expansion for $F(t, \widehat{\theta}(t))$, we have

$$
\begin{aligned}
F(t, \widehat{\theta}(t)) &= F(t, \theta^*(t)) + F'(t, \theta^*(t))(\widehat{\theta}(t) - \theta^*(t)) + \widehat{\gamma}(t, \theta^*(t)) \\
&= F(t, \theta^*(t)) + \frac{1}{Nn} F'(t, \theta^*(t))(Nn)(\widehat{\theta}(t) - \theta^*(t)) + \widehat{\gamma}(t, \theta^*(t)) \\
&= F(t, \theta^*(t)) - V(t, \theta^*(t))(Nn)(\widehat{\theta}(t) - \theta^*(t)) + \widehat{\gamma}(t, \theta^*(t)),
\end{aligned}
$$

which implies that

$$
\begin{aligned}
&\widehat{\theta}(t) - \theta^*(t) - \frac{1}{Nn} V^{-1}(t, \theta^*(t)) \mathbb{E} F(t, \theta^*(t)) \\
&= \frac{1}{Nn} V^{-1}(t, \theta^*(t)) \{F(t, \theta^*(t)) - \mathbb{E} F(t, \theta^*(t))\} + \frac{1}{Nn} V^{-1}(t, \theta^*(t)) \widehat{\gamma}(t, \theta^*(t)).
\end{aligned}
$$

Here $\widehat{\gamma}(t, \theta^*(t)) = \left(\widehat{\gamma}_1(t, \theta^*(t)), \widehat{\gamma}_2(t, \theta^*(t)), \ldots, \widehat{\gamma}_{2n-1}(t, \theta^*(t))\right)^\top$, where

$$
\widehat{\gamma}_i(t, \theta^*(t)) = \sum_{\substack{j:1 \le j \le n \\ j \ne i}} \sum_{\ell \in [N]} K_h(t - T_\ell) \widehat{\gamma}_{ij}(t, \theta^*(t)), \quad i \in [n],
$$

$$
\widehat{\gamma}_{n+j}(t, \theta^*(t)) = \sum_{\substack{i:1 \le i \le n \\ i \ne j}} \sum_{\ell \in [N]} K_h(t - T_\ell) \widehat{\gamma}_{ij}(t, \theta^*(t)), \quad j \in [n-1],
$$

with

$$
\widehat{\gamma}_{ij}(t, \theta^*(t)) = -\frac{e^{\alpha_i^*(t) + \beta_j^*(t) + \widehat{\omega}_{ij}(t)\phi_{ij}} \left(1 - e^{\alpha_i^*(t) + \beta_j^*(t) + \phi_{ij}\widehat{\omega}_{ij}(t)}\right)}{2(1 + e^{\alpha_i^*(t) + \beta_j^*(t) + \phi_{ij}\widehat{\omega}_{ij}(t)})^3} \widehat{\omega}_{ij}^2(t),
$$

and $0 \le \phi_{ij} \le 1$. Since $\left|e^x(1 - e^x)/(1 + e^x)^3\right| \le 1$, we have

$$
\begin{aligned}
&|\widehat{\gamma}_{ij}(t, \theta^*(t))| \le |\widehat{\omega}_{ij}^2(t)/2| \le 2\widehat{b}_{Nnh}^2, \\
&|\widehat{\gamma}_i(t, \theta^*(t))| = \sum_{\substack{j:1 \le j \le n \\ j \ne i}} \sum_{\ell \in [N]} K_h(t - T_\ell)|\widehat{\gamma}_{ij}(t, \theta^*(t))| \le 2N(n-1)\widehat{b}_{Nnh}^2/h, \quad i \in [n], \\
&|\widehat{\gamma}_{n+j}(t, \theta^*(t))| = \sum_{\substack{i:1 \le i \le n \\ i \ne j}} \sum_{\ell \in [N]} K_h(t - T_\ell)|\widehat{\gamma}_{ij}(t, \theta^*(t))| \le 2N(n-1)\widehat{b}_{Nnh}^2/h, \quad j \in [n-1].
\end{aligned}
$$

By Lemmas 8-10 given below, we have

$$
\left\| \frac{1}{Nn} V^{-1}(t, \theta^*(t)) \widehat{\gamma}(t, \theta^*(t)) \right\|_{\max} = o_p\left((Nnh)^{-1/2}\right),
$$

$$
\left\| \frac{1}{Nn} V^{-1}(t, \theta^*(t)) \mathbb{E} F(t, \theta^*(t)) - k_{21} h^2 \mu(t) \right\|_{\max} = o_p\left((Nnh)^{-1/2}\right)
$$

and

$$\frac{1}{Nn} V^{-1}(t, \theta^*(t)) \{ F(t, \theta^*(t)) - \mathbb{E}F(t, \theta^*(t)) \}$$
$$= \frac{1}{Nn} \bar{S}(t, \theta^*(t)) \{ F(t, \theta^*(t)) - \mathbb{E}F(t, \theta^*(t)) \} + o_p((Nnh)^{-1/2}),$$

where $\mu(t)$ is defined in Lemma 9. That is,

$$\sqrt{Nnh} \Big[\{ \widehat{\theta}(t) - \theta^*(t) \}_i - k_{21} h^2 \mu_i(t) \Big]$$
$$= \sqrt{Nnh} \Big[\{ \frac{1}{Nn} \bar{S}(t, \theta^*(t)) (F(t, \theta^*(t)) - \mathbb{E}F(t, \theta^*(t))) \}_i \Big] + o_p(1), \quad i \in [2n-1].$$

For any $i \in [n]$, it can be shown

$$\sum_{\substack{j:1 \le j \le n \\ j \ne i}} \sum_{\ell \in [N]} \mathbb{E} \Big| K_h(t - T_\ell)(a_{ij}(T_\ell) - u_{ij}(t)) - \mathbb{E}\{ K_h(t - T_\ell)(a_{ij}(T_\ell) - u_{ij}(t)) \} \Big|^3$$
$$\le \sum_{\substack{j:1 \le j \le n \\ j \ne i}} \sum_{\ell \in [N]} \mathbb{E} \Big| K_h(t - T_\ell)(a_{ij}(T_\ell) - u_{ij}(t)) - \mathbb{E}\{ K_h(t - T_\ell)(a_{ij}(T_\ell) - u_{ij}(t)) \} \Big|^2 / h$$
$$= O(Nn/h^2),$$

and by Lemma 5, we have

$$\text{Var} \Big\{ \sum_{\substack{j:1 \le j \le n \\ j \ne i}} \sum_{\ell \in [N]} K_h(t - T_\ell)(a_{ij}(T_\ell) - u_{ij}(t)) \Big\} = O(Nn/h).$$

Therefore,

$$\frac{\sum_{j \in [n], j \ne i} \sum_{\ell \in [N]} \mathbb{E} \Big| K_h(t - T_\ell)(a_{ij}(T_\ell) - u_{ij}(t)) - \mathbb{E}\{ K_h(t - T_\ell)(a_{ij}(T_\ell) - u_{ij}(t)) \} \Big|^3}{\Big(\text{Var}\{ \sum_{j \in [n], j \ne i} \sum_{\ell \in [N]} K_h(t - T_\ell)(a_{ij}(T_\ell) - u_{ij}(t)) \} \Big)^{3/2}} \le O\Big(\frac{1}{\sqrt{Nnh}} \Big).$$

Similarly, for any $j \in [n-1]$,

$$\frac{\sum_{i \in [n], i \ne j} \sum_{\ell \in [N]} \mathbb{E} \Big| K_h(t - T_\ell)(a_{ij}(T_\ell) - u_{ij}(t)) - \mathbb{E}\{ K_h(t - T_\ell)(a_{ij}(T_\ell) - u_{ij}(t)) \} \Big|^3}{\Big(\text{Var}\{ \sum_{i \in [n], i \ne j} \sum_{\ell \in [N]} K_h(t - T_\ell)(a_{ij}(T_\ell) - u_{ij}(t)) \} \Big)^{3/2}} \le O\Big(\frac{1}{\sqrt{Nnh}} \Big).$$

This implies that the condition for the Lyapunov's central limit theorem holds under Conditions 1-2, $C_{Nn}^2 Nnh^5 Q_{Nnh}^2 = o(1)$, $Nh \to \infty$ and

$$Q_{Nnh} = o\Big[1 / \{ (Nnh)^{1/18} (\sqrt{C_1 \log(Nnh)/(Nnh)} + C_2 C_{Nn} h^2)^2 \} \Big].$$

Thus for any fixed $1 \le p \le 2n-1$ and $t \in [a, b]$, $\sqrt{Nnh} \Big[\{ \widehat{\theta}_1(t) - \theta_1^*(t), \ldots, \widehat{\theta}_p(t) - \theta_p^*(t) \}^\top - k_{21} h^2 \{ \mu_1(t), \ldots, \mu_p(t) \}^\top \Big]$ converges in distribution to a p-dimensional multivariate normal random vector with mean zero and covariance $k_{02} \Sigma(t, \theta^*(t))$.

A direct calculation yields

$$\mathrm{Cov}\left(\frac{\sqrt{Nnh}}{Nn}\bar{S}(t,\theta^*(t))\left\{F(t,\theta^*(t)) - \mathbb{E}F(t,\theta^*(t))\right\}\right) = \bar{S}(t,\theta^*(t))\Omega(t,\theta^*(t))\bar{S}(t,\theta^*(t))^\top,$$

where $\Omega(t,\theta^*(t)) = \mathrm{Cov}\{\sqrt{h/(Nn)}F(t,\theta^*(t))\}$.

We next calculate each element of $\Omega(t,\theta^*(t))$. For $i \in [n]$, by Lemma 5, the ith diagonal element of $\Omega(t,\theta^*(t))$ is

$$\begin{aligned}
\{\Omega(t,\theta^*(t))\}_{i,i} &= \frac{h}{Nn}\mathrm{Var}\{F_i(t,\theta^*(t))\} \\
&= \frac{1}{n}\sum_{\substack{j:1\le j\le n \\ j\ne i}}\left[k_{02}f(t)u_{ij}(t)(1-u_{ij}(t)) + k_{12}h\widetilde{\psi}_{ij}(t) + o(C_{Nn}h)\right] \\
&= \frac{1}{n}\sum_{\substack{j:1\le j\le n \\ j\ne i}}k_{02}f(t)u_{ij}(t)(1-u_{ij}(t)) + \frac{1}{n}\sum_{\substack{j:1\le j\le n \\ j\ne i}}k_{12}h\widetilde{\psi}_{ij}(t) + o(C_{Nn}h) \\
&= k_{02}f(t)\{V_0(t,\theta^*(t))\}_{i,i} + k_{12}h\{\widetilde{V}(t,\theta^*(t))\}_{i,i} + o(C_{Nn}h) \\
&= k_{02}\{\bar{V}(t,\theta^*(t))\}_{i,i} + k_{12}h\{\widetilde{V}(t,\theta^*(t))\}_{i,i} + o(C_{Nn}h) + O(Nnh^5C_{Nn}^2),
\end{aligned}$$

where $k_{12} := \int vK^2(v)dv$, $\widetilde{\psi}_{ij}(t) := u_{ij}(t)(1-u_{ij}(t))f'(t) + f(t)u'_{ij}(t)(1-2u_{ij}(t))$, and

$$\{\widetilde{V}(t,\theta^*(t))\}_{i,i} = \frac{1}{n}\sum_{\substack{j:1\le j\le n \\ j\ne i}}\widetilde{\psi}_{ij}(t).$$

For any $j \in [n-1]$,

$$\begin{aligned}
\{\Omega(t,\theta^*(t))\}_{n+j,n+j} &= \frac{h}{Nn}\mathrm{Var}\{F_{n+j}(t,\theta^*(t))\} \\
&= k_{02}\{\bar{V}(t,\theta^*(t))\}_{n+j,n+j} + k_{12}h\{\widetilde{V}(t,\theta^*(t))\}_{n+j,n+j} \\
&\quad + o(C_{Nn}h) + O(Nnh^5C_{Nn}^2),
\end{aligned}$$

where $\{\widetilde{V}(t,\theta^*(t))\}_{n+j,n+j} = \frac{1}{n}\sum_{\substack{i:1\le i\le n \\ i\ne j}}\widetilde{\psi}_{ij}(t)$. In addition,

$$\begin{aligned}
\{\Omega(t,\theta^*(t))\}_{i,j} &= \frac{h}{Nn}\mathrm{Cov}\{F_i(t,\theta^*(t)), F_j(t,\theta^*(t))\} = 0,\ i \in [n],\ j \in [n],\ i \ne j, \\
\{\Omega(t,\theta^*(t))\}_{n+i,n+j} &= \frac{h}{Nn}\mathrm{Cov}\{F_{n+i}(t,\theta^*(t)), F_{n+j}(t,\theta^*(t))\} = 0,\ i \in [n-1],\ j \in [n-1],\ i \ne j, \\
\{\Omega(t,\theta^*(t))\}_{i,n+i} &= \frac{h}{Nn}\mathrm{Cov}\{F_i(t,\theta^*(t)), F_{n+i}(t,\theta^*(t))\} = 0,\ i \in [n-1], \\
\{\Omega(t,\theta^*(t))\}_{n+i,i} &= \frac{h}{Nn}\mathrm{Cov}\{F_{n+i}(t,\theta^*(t)), F_i(t,\theta^*(t))\} = 0,\ i \in [n-1].
\end{aligned}$$

For any $i \in [n], j \in [n-1], i \neq j$,

$$
\begin{aligned}
\{\Omega(t, \theta^*(t))\}_{i,n+j} &= \frac{h}{Nn} \mathrm{Cov}\{F_i(t, \theta^*(t)), F_{n+j}(t, \theta^*(t))\} \\
&= k_{02}\{\bar{V}(t, \theta^*(t))\}_{i,n+j} + k_{12}h\{\widetilde{V}(t, \theta^*(t))\}_{i,n+j} + o(C_{Nn}h/n) + O(Nh^5 C_{Nn}^2),
\end{aligned}
$$

$$
\begin{aligned}
\{\Omega(t, \theta^*(t))\}_{n+j,i} &= \frac{h}{Nn} \mathrm{Cov}\{F_{n+j}(t, \theta^*(t)), F_i(t, \theta^*(t))\} \\
&= k_{02}\{\bar{V}(t, \theta^*(t))\}_{n+j,i} + k_{12}h\{\widetilde{V}(t, \theta^*(t))\}_{n+j,i} + o(C_{Nn}h/n) + O(Nh^5 C_{Nn}^2),
\end{aligned}
$$

where $\{\widetilde{V}(t, \theta^*(t))\}_{n+j,i} = \{\widetilde{V}(t, \theta^*(t))\}_{i,n+j} = \widetilde{\psi}_{ij}(t)$. Now we establish

$$
\left\| \bar{S}(t, \theta^*(t))\Omega(t, \theta^*(t))\bar{S}(t, \theta^*(t)) - k_{02}\bar{S}(t, \theta^*(t))\bar{V}(t, \theta^*(t))\bar{S}(t, \theta^*(t)) \right\|_{\max} = o(1).
$$

For this, define $\widetilde{V}(t, \theta^*(t)) \in \mathbb{R}^{(2n-1) \times (2n-1)}$ and its (i,j)th element is $\{\widetilde{V}(t, \theta^*(t))\}_{i,j}$. Then, we have

$$
\Omega(t, \theta^*(t)) = k_{02}\bar{V}(t, \theta^*(t)) + k_{12}h\widetilde{V}(t, \theta^*(t)) + \left[o(C_{Nn}h) + O(Nnh^5 C_{Nn}^2)\right]\mathbb{D}.
$$

Here $\mathbb{D} = (d_{ij}) \in \mathbb{R}^{(2n-1) \times (2n-1)}$, and $d_{ii} = 1$ ($i \in [n]$), $d_{n+j,n+j} = 1$ ($j \in [n-1]$), $d_{i,n+j} = d_{n+i,j} = 1/n$ ($i \in [n], j \in [n-1], i \neq j$), and $d_{ij} = 0$ otherwise. We first show

$$
\left\| \bar{S}(t, \theta^*(t))\widetilde{V}(t, \theta^*(t))\bar{S}(t, \theta^*(t)) \right\|_{\max} = O(C_{Nn}Q_{Nnh}^2).
$$

Write

$$
\bar{S}(t, \theta^*(t))\widetilde{V}(t, \theta^*(t)) = \left(\sum_{k=1}^{2n-1} \{\bar{S}(t, \theta^*(t))\}_{ik} \widetilde{\psi}_{kj}(t, \theta^*(t)) \right).
$$

Specifically, for $i = j \in [n]$, we have

$$
\begin{aligned}
\left| \sum_{k=1}^{2n-1} \{\bar{S}(t, \theta^*(t))\}_{i,k} \widetilde{\psi}_{kj}(t, \theta^*(t)) \right| &\leq \frac{|\widetilde{\psi}_{ii}(t, \theta^*(t))|}{\{\bar{V}(t, \theta^*(t))\}_{i,i}} + \frac{|\widetilde{\psi}_{ii}(t, \theta^*(t)) - \sum_{k=1, k \neq j}^{2n-1} \widetilde{\psi}_{kj}(t, \theta^*(t))|}{\{\bar{V}(t, \theta^*(t))\}_{2n,2n}} \\
&= \frac{|\widetilde{\psi}_{ii}(t, \theta^*(t))|}{\{\bar{V}(t, \theta^*(t))\}_{i,i}} + \frac{|\widetilde{\psi}_{2n,j}(t, \theta^*(t))|}{\{\bar{V}(t, \theta^*(t))\}_{2n,2n}} \\
&\leq \frac{C_2 C_{Nn}Q_{Nnh}}{f(t)}\left(1 + \frac{1}{n}\right),
\end{aligned}
$$

where $\{\bar{V}(t, \theta^*(t))\}_{2n,2n} = 2\sum_{i=1}^n \{\bar{V}(t, \theta^*(t))\}_{i,i} - \sum_{i=1}^n \sum_{j=1}^{2n-1} \{\bar{V}(t, \theta^*(t))\}_{i,j}$ and

$$
\widetilde{\psi}_{2n,i}(t) = \widetilde{\psi}_{i,2n}(t) := \widetilde{\psi}_{ii}(t) - \sum_{j=1, j \neq i}^{2n-1} \widetilde{\psi}_{ij}(t), \quad i \in [2n-1].
$$

For $i = j = n+1, \ldots, 2n-1$, we have

$$\left| \sum_{k=1}^{2n-1} \left\{ \bar{S}(t, \theta^*(t)) \right\}_{i,k} \widetilde{\psi}_{kj}(t, \theta^*(t)) \right| \leq \frac{|\widetilde{\psi}_{ii}(t, \theta^*(t))|}{\{\bar{V}(t, \theta^*(t))\}_{i,i}} + \frac{|\widetilde{\psi}_{ii}(t, \theta^*(t)) - \sum_{k=1, k \neq j}^{2n-1} \widetilde{\psi}_{kj}(t, \theta^*(t))|}{\{\bar{V}(t, \theta^*(t))\}_{2n, 2n}}$$

$$= \frac{|\widetilde{\psi}_{ii}(t, \theta^*(t))|}{\{\bar{V}(t, \theta^*(t))\}_{i,i}} \leq C_2 C_{Nn} Q_{Nnh}/f(t).$$

For $i \in [n]$, $j \in [n-1]$, $i \neq j$, we have

$$\left| \sum_{k=1}^{2n-1} \left\{ \bar{S}(t, \theta^*(t)) \right\}_{i,k} \widetilde{\psi}_{kj}(t, \theta^*(t)) \right| = \frac{|\widetilde{\psi}_{jj}(t, \theta^*(t)) - \sum_{k=1, k \neq j}^{2n-1} \widetilde{\psi}_{kj}(t, \theta^*(t))|}{\{\bar{V}(t, \theta^*(t))\}_{2n, 2n}}$$

$$\leq C_2 C_{Nn} Q_{Nnh}/(nf(t)).$$

For $i \in [n]$, $j = n+1, \ldots, 2n-1$, $j \neq n+i$, we have

$$\left| \sum_{k=1}^{2n-1} \left\{ \bar{S}(t, \theta^*(t)) \right\}_{i,k} \widetilde{\psi}_{kj}(t, \theta^*(t)) \right| = \frac{|\widetilde{\psi}_{ij}(t, \theta^*(t))|}{\{\bar{V}(t, \theta^*(t))\}_{i,i}} \leq C_2 C_{Nn} Q_{Nnh}/(nf(t)).$$

For $i = n+1, \ldots, 2n-1, j \in [n]$, we have

$$\left| \sum_{k=1}^{2n-1} \left\{ \bar{S}(t, \theta^*(t)) \right\}_{i,k} \widetilde{\psi}_{kj}(t, \theta^*(t)) \right| \leq C_2 C_{Nn} Q_{Nnh}/(nf(t)).$$

Otherwise, we have

$$\left| \sum_{k=1}^{2n-1} \left\{ \bar{S}(t, \theta^*(t)) \right\}_{i,k} \widetilde{\psi}_{kj}(t, \theta^*(t)) \right| = 0.$$

This implies that

$$\left\| \bar{S}(t, \theta^*(t)) \widetilde{V}(t, \theta^*(t)) - \frac{C_2 C_{Nn} Q_{Nnh}}{f(t)} \mathbb{1} \right\|_{\max} \leq 2 C_2 C_{Nn} Q_{Nnh}/(nf(t)),$$

where $\mathbb{1} \in \mathbb{R}^{(2n-1) \times (2n-1)}$ is the identity matrix. Thus, we have

$$\left\| \bar{S}(t, \theta^*(t)) \widetilde{V}(t, \theta^*(t)) \bar{S}(t, \theta^*(t)) \right\|_{\max}$$

$$= \left\| \left(\bar{S}(t, \theta^*(t)) \widetilde{V}(t, \theta^*(t)) - \frac{C_2 C_{Nn} Q_{Nnh}}{f(t)} \mathbb{1} \right) \bar{S}(t, \theta^*(t)) + \frac{C_2 C_{Nn} Q_{Nnh}}{f(t)} \bar{S}(t, \theta^*(t)) \right\|_{\max}$$

$$\leq \left\| \left(\bar{S}(t, \theta^*(t)) \widetilde{V}(t, \theta^*(t)) - \frac{C_2 C_{Nn} Q_{Nnh}}{f(t)} \mathbb{1} \right) \bar{S}(t, \theta^*(t)) \right\|_{\max} + \frac{C_2 C_{Nn} Q_{Nnh}}{f(t)} \left\| \bar{S}(t, \theta^*(t)) \right\|_{\max}$$

$$\leq \left\| \bar{S}(t, \theta^*(t)) \widetilde{V}(t, \theta^*(t)) - \frac{C_2 C_{Nn} Q_{Nnh}}{f(t)} \mathbb{1} \right\|_{\max} \times \max_{1 \leq j \leq 2n-1} \sum_{k=1}^{2n-1} \left| \{\bar{S}(t, \theta^*(t))\}_{k,j} \right| + \frac{C_2 C_{Nn} Q_{Nnh}}{f(t)} \left\| \bar{S}(t, \theta^*(t)) \right\|_{\max}$$

$$= O(C_{Nn} Q_{Nnh}^2).$$

That is,

$$\left\|\bar{S}(t,\theta^*(t))\widetilde{V}(t,\theta^*(t))\bar{S}(t,\theta^*(t))\right\|_{\max} = O(C_{Nn}Q_{Nnh}^2).$$

Similarly, we can show that

$$\left[o(C_{Nn}h) + O(Nnh^5C_{Nn}^2)\right]\left\|\bar{S}(t,\theta^*(t))\mathbb{D}\bar{S}(t,\theta^*(t))\right\|_{\max} = O(hC_{Nn}Q_{Nnh}^2).$$

These facts imply that

$$\left\|\bar{S}(t,\theta^*(t))\Omega(t,\theta^*(t))\bar{S}(t,\theta^*(t)) - k_{02}\bar{S}(t,\theta^*(t))\bar{V}(t,\theta^*(t))\bar{S}(t,\theta^*(t))\right\|_{\max}$$
$$\leq\left\|\bar{S}(t,\theta^*(t))\widetilde{V}(t,\theta^*(t))\bar{S}(t,\theta^*(t))\right\|_{\max} + \left[o(C_{Nn}h) + O(Nnh^5C_{Nn}^2)\right]\left\|\bar{S}(t,\theta^*(t))\mathbb{D}\bar{S}(t,\theta^*(t))\right\|_{\max}$$
$$=O(hC_{Nn}Q_{Nnh}^2).$$

Thus, if $hC_{Nn}Q_{Nnh}^2 = o(1)$, then we have

$$\left\|\bar{S}(t,\theta^*(t))\Omega(t,\theta^*(t))\bar{S}(t,\theta^*(t)) - k_{02}\bar{S}(t,\theta^*(t))\bar{V}(t,\theta^*(t))\bar{S}(t,\theta^*(t))\right\|_{\max} = o(1).$$

This completes the proof.

Proof of Proposition 3 As in the proof of Lemma 10, we have

$$S(t,\theta(t)) = \frac{1}{\widehat{f}(t)}S_0(t,\theta(t)) \text{ and } \bar{S}(t,\theta(t)) = \frac{1}{f(t)}S_0(t,\theta(t)).$$

Then, a direct calculation yields

$$\left\|\widehat{\Sigma}(t,\widehat{\theta}(t)) - \Sigma(t,\theta^*(t))\right\|_{\max}$$
$$=\left\|S(t,\widehat{\theta}(t))\widehat{V}(t,\widehat{\theta}(t))S(t,\widehat{\theta}(t))^\top - k_{02}\bar{S}(t,\theta^*(t))\bar{V}(t,\theta^*(t))\bar{S}(t,\theta^*(t))^\top\right\|_{\max}$$
$$=\left\|\frac{1}{\widehat{f}^2(t)}S_0(t,\widehat{\theta}(t))\widehat{V}(t,\widehat{\theta}(t))S_0(t,\widehat{\theta}(t))^\top - \frac{1}{f^2(t)}k_{02}S_0(t,\widehat{\theta}(t))\bar{V}(t,\theta^*(t))S_0(t,\widehat{\theta}(t))^\top\right\|_{\max}$$
$$\leq\left\|\left(\frac{1}{\widehat{f}^2(t)} - \frac{1}{f^2(t)}\right)S_0(t,\widehat{\theta}(t))\widehat{V}(t,\widehat{\theta}(t))S_0(t,\widehat{\theta}(t))^\top\right\|_{\max}$$
$$+\left\|\frac{1}{f^2(t)}S_0(t,\widehat{\theta}(t))\left(\widehat{V}(t,\widehat{\theta}(t)) - k_{02}\bar{V}(t,\theta^*(t))\right)S_0(t,\widehat{\theta}(t))^\top\right\|_{\max}.$$

We next show that

$$\left\|\left(\frac{1}{\widehat{f}^2(t)} - \frac{1}{f^2(t)}\right)S_0(t,\widehat{\theta}(t))\widehat{V}(t,\widehat{\theta}(t))S_0(t,\widehat{\theta}(t))^\top\right\|_{\max} =O_p\left(Q_{Nnh}\left\{\sqrt{\frac{\log(1/h)}{Nh}} + h^2\right\}\right),$$

$$f^{-2}(t)\left\|S_0(t,\widehat{\theta}(t))\left(\widehat{V}(t,\widehat{\theta}(t)) - k_{02}\bar{V}(t,\theta^*(t))\right)S_0(t,\widehat{\theta}(t))^\top\right\|_{\max} =O_p\left(Q_{Nnh}\left\{\sqrt{\frac{\log(1/h)}{Nh}} + h^2\right\}\right).$$

Write

$$S_0(t, \widehat{\theta}(t)) \widehat{V}(t, \widehat{\theta}(t)) = \Big(\sum_{k \in [2n-1]} \big\{ S_0(t, \widehat{\theta}(t)) \big\}_{i,k} \big\{ \widehat{V}(t, \widehat{\theta}(t)) \big\}_{k,j} \Big).$$

Then, if $i = j \in [n]$, we have

$$\Big| \sum_{k \in [2n-1]} \big\{ S_0(t, \widehat{\theta}(t)) \big\}_{i,k} \big\{ \widehat{V}(t, \widehat{\theta}(t)) \big\}_{k,j} \Big|$$

$$\leq \frac{\big| \{ \widehat{V}(t, \widehat{\theta}(t)) \}_{i,i} \big|}{\{ V_0(t, \widehat{\theta}(t)) \}_{i,i}} + \frac{\big| \{ \widehat{V}(t, \widehat{\theta}(t)) \}_{i,i} - \sum_{k=1, k \neq j}^{2n-1} \{ \widehat{V}(t, \widehat{\theta}(t)) \}_{k,j} \big|}{\{ V_0(t, \widehat{\theta}(t)) \}_{2n,2n}}$$

$$= \frac{\big| \{ \widehat{V}(t, \widehat{\theta}(t)) \}_{i,i} \big|}{\{ V_0(t, \widehat{\theta}(t)) \}_{i,i}} + \frac{\big| \{ \widehat{V}(t, \widehat{\theta}(t)) \}_{2n,j} \big|}{\{ V_0(t, \widehat{\theta}(t)) \}_{2n,2n}}.$$

It can be shown that with probability tending to 1,

$$\big| \{ \widehat{V}(t, \widehat{\theta}(t)) \}_{i,i} \big| \leq 2 \big| \mathbb{E} \{ \widehat{V}(t, \theta^*(t)) \}_{i,i} \big|.$$

Then, we have

$$\Big| \sum_{k=1}^{2n-1} \big\{ S_0(t, \widehat{\theta}(t)) \big\}_{i,k} \big\{ \widehat{V}(t, \widehat{\theta}(t)) \big\}_{k,j} \Big| = O_p(1 + 1/n).$$

Similarly, we can show

$$\Big| \sum_{k=1}^{2n-1} \big\{ S_0(t, \widehat{\theta}(t)) \big\}_{i,k} \big\{ \widehat{V}(t, \widehat{\theta}(t)) \big\}_{k,j} \Big| = O_p(1), \quad i = j = n+1, \dots, 2n-1,$$

$$\Big| \sum_{k=1}^{2n-1} \big\{ S_0(t, \widehat{\theta}(t)) \big\}_{i,k} \big\{ \widehat{V}(t, \widehat{\theta}(t)) \big\}_{k,j} \Big| = O_p(1/n), \quad i \in [n], j \in [n-1], i \neq j,$$

$$\Big| \sum_{k=1}^{2n-1} \big\{ S_0(t, \widehat{\theta}(t)) \big\}_{i,k} \big\{ \widehat{V}(t, \widehat{\theta}(t)) \big\}_{k,j} \Big| = O_p(1/n) \quad i \in [n], j = n+1, \dots, 2n-1, j \neq n+i,$$

$$\Big| \sum_{k=1}^{2n-1} \big\{ S_0(t, \widehat{\theta}(t)) \big\}_{i,k} \big\{ \widehat{V}(t, \widehat{\theta}(t)) \big\}_{k,j} \Big| = O(1/n), \quad i = n+1, \dots, 2n-1, j \in [n],$$

$$\Big| \sum_{k=1}^{2n-1} \big\{ S_0(t, \widehat{\theta}(t)) \big\}_{i,k} \big\{ \widehat{V}(t, \widehat{\theta}(t)) \big\}_{k,j} \Big| = 0, \quad \text{otherwise}.$$

These facts imply that

$$\big\| S_0(t, \widehat{\theta}(t)) \widehat{V}(t, \widehat{\theta}(t)) - C_4 \mathbb{1} \big\|_{\max} = O(1/n).$$

with probability tending to 1 for some constant C_4. By Theorem 6 in Masry (1996), we have

$$\sup_{t \in [a,b]} |\widehat{f}(t) - f(t)| = O\Big(\Big\{ \frac{\log(1/h)}{Nh} \Big\}^{1/2} + h^2 \Big).$$

This implies that

$$
\begin{aligned}
&\left\|\frac{1}{\widehat{f}^2(t)} - \frac{1}{f^2(t)}\right| \times \left\|S_0(t,\widehat{\theta}(t))\widehat{V}(t,\widehat{\theta}(t))S_0(t,\widehat{\theta}(t))^\top\right\|_{\max} \\
&\leq \left|\frac{1}{\widehat{f}^2(t)} - \frac{1}{f^2(t)}\right| \times \left\|(S_0(t,\widehat{\theta}(t))\widehat{V}(t,\widehat{\theta}(t)) - C_4\mathbb{1})S_0(t,\widehat{\theta}(t)) + C_4\mathbb{1}S_0(t,\widehat{\theta}(t))\right\|_{\max} \\
&\leq \left|\frac{1}{\widehat{f}^2(t)} - \frac{1}{f^2(t)}\right| \times \left\|(S_0(t,\widehat{\theta}(t))\widehat{V}(t,\widehat{\theta}(t)) - C_4\mathbb{1})S_0(t,\widehat{\theta}(t))\right\|_{\max} \\
&\quad + \left|\frac{1}{\widehat{f}^2(t)} - \frac{1}{f^2(t)}\right| \times C_4\left\|S_0(t,\widehat{\theta}(t))\right\|_{\max} \\
&\leq \left|\frac{1}{\widehat{f}^2(t)} - \frac{1}{f^2(t)}\right| \times \left\|S_0(t,\widehat{\theta}(t))\widehat{V}(t,\widehat{\theta}(t)) - C_4\mathbb{1}\right\|_{\max} \times \max_{j\in[2n-1]} \sum_{k=1}^{2n-1}\left|\{S_0(t,\widehat{\theta}(t))\}_{k,j}\right| \\
&\quad + \left|\frac{1}{\widehat{f}^2(t)} - \frac{1}{f^2(t)}\right| \times C_4\|S_0(t,\widehat{\theta}(t))\|_{\max} \\
&= O_p\left(Q_{Nnh}\left\{\sqrt{\frac{\log(1/h)}{Nh}} + h^2\right\}\right).
\end{aligned}
$$

By some arguments similar to the above, we can show

$$
\left\|\frac{1}{f^2(t)}S_0(t,\widehat{\theta}(t))(\widehat{V}(t,\widehat{\theta}(t)) - k_{02}\bar{V}(t,\theta^*(t)))S_0(t,\widehat{\theta}(t))^\top\right\|_{\max} = O_p\left(Q_{Nnh}\left\{\sqrt{\frac{\log(1/h)}{Nh}} + h^2\right\}\right).
$$

Thus, if $Nh \to \infty$, then

$$
\|\widehat{\Sigma}(t,\widehat{\theta}(t)) - \Sigma(t,\theta^*(t))\|_{\max} = o_p(1)
$$

under the conditions of Theorems 1 and 2. It completes the proof.

Supplement to:

"Time-varying β-model for dynamic directed networks"

Yuqing Du[†], Lianqiang Qu[†], Ting Yan[†] and Yuan Zhang[*]

[†]Central China Normal University and [*]The Ohio State University

April 13, 2023

S1. Newton iterative theorem

Let $\mathbb{G}$ be an open convex subset of $\mathbb{R}^{2n-1}$, $\Omega(\mathbf{x}, r)$ denote the open ball $\{\mathbf{y} \in \mathbb{R}^{2n-1} : \|\mathbf{x} - \mathbf{y}\|_\infty < r\}$ and $\bar{\Omega}(\mathbf{x}, r)$ be its closure, where $\mathbf{x} \in \mathbb{R}^{2n-1}$. We will use Newton's iterative sequence to prove the consistency of the MLE.

Lemma 1. *(Theorem 7, Yan et al. (2016)) Define a system of equations*

$$F_i(\theta) = d_i - \sum_{k=1, k\neq i}^{n} f(\alpha_i + \beta_k), i = 1, \ldots, n,$$

$$F_{n+j}(\theta) = b_j - \sum_{k=1, k\neq j}^{n} f(\alpha_k + \beta_j), j = 1, \ldots, n - 1,$$

$$F(\theta) = (F_1(\theta), \ldots, F_n(\theta), F_{n+1}(\theta), \ldots, F_{2n-1}(\theta))^\top.$$

where $f(\cdot)$ is third continuously differentiable. Let $\mathbb{G} \subset \mathbb{R}^{2n-1}$ be a convex set and assume for any $\mathbf{x}, \mathbf{y}, \mathbf{v} \in \mathbb{G}$, we have

$$\|[F'(\mathbf{x}) - F'(\mathbf{y})]\mathbf{v}\|_\infty \le H_1 \|\mathbf{x} - \mathbf{y}\|_\infty \|\mathbf{v}\|_\infty,$$

$$\max_{i \in [2n-1]} \|F_i'(\mathbf{x}) - F_i'(\mathbf{y})\|_\infty \le H_2 \|\mathbf{x} - \mathbf{y}\|_\infty.$$

---

[*]Correspondence: `qulianq@ccnu.edu.cn`

where $F'(\theta)$ is the Jacobian matrix of F on θ and $F_i'(\theta)$ is the gradient function of F_i on θ. Consider $\theta^{(0)} \in \mathbb{G}$ with $\Omega(\theta^{(0)}, 2r) \subset \mathbb{G}$, where $r = \|[F'(\theta^{(0)})^{-1}]F(\theta^{(0)})\|_\infty$. For any $\theta^{(0)} \in \Omega(\theta^{(0)}, 2r)$, we assume

$$F'(\theta) \in \mathcal{L}_n(m, M) \quad or \quad -F'(\theta) \in \mathcal{L}_n(m, M).$$

For $k = 1, 2, \ldots$, define the Newton iterates $\theta^{(k+1)} = \theta^{(k)} - [F'(\theta^{(k)})]^{-1}F(\theta^{(k)})$. Let

$$\rho = \frac{c_1(2n-1)M^2 H_1}{2m^3 n^2} + \frac{H_2}{(n-1)m}.$$

where c_1 is a positive constant. If $\rho r < 1/2$, then $\theta^{(k)} \in \Omega(\theta^{(0)}, 2r), k = 1, 2, \ldots$, are well defined and satisfy

$$\|\theta^{(k+1)} - \theta^{(0)}\|_\infty \le r/(1 - \rho r).$$

Further, $\lim_{k \to \infty} \theta^{(k)}$ exists and the limiting point is precisely the solution of $F(\theta) = 0$ in the range of $\theta \in \Omega(\theta^{(0)}, 2r)$.

Lemma 2. *(Proposition 1, Yan et al. (2016)) If $V \in \mathcal{L}_n(m, M)$ with $M/m = o(n)$, then for large enough n,*

$$\|V^{-1} - S\|_{\max} \le \frac{C_3 M^2}{m^3(n-1)^2},$$

where C_3 is a positive constant.

Lemma 3. *(Lemma 1, Yan et al. (2016)) If $V \in \mathcal{L}_n(m, M)$ with $M/m = o(n)$, then for a vector $\mathbf{x} \in \mathbb{R}^{2n-1}$,*

$$\|V^{-1}\mathbf{x}\|_\infty \le \|(V^{-1} - S)\mathbf{x}\|_\infty + \|S\mathbf{x}\|_\infty \le \frac{2c_2(2n-1)M^2\|\mathbf{x}\|_\infty}{m^3(n-1)^2} + \frac{x_{2n}}{v_{2n,2n}} + \max_{1=1,\ldots,2n-1} \frac{|x_i|}{v_{ii}}.$$

where c_2 is a positive constant and $x_{2n} := \sum_{i=1}^n x_i - \sum_{i=n+1}^{2n-1} x_i$.

S2. Proofs of Lemmas

Lemma 4. *Suppose X is a random variables with mean zero and variance σ^2, bounded by c_3, which is a positive constant, then the following holds:*

$$P(X > \epsilon) \le e^{-\epsilon^2/4\sigma^2}$$

where $\epsilon > 0$ is a small constant.

Remark. Lemma 4 is very similar to well-known concentration inequalities such as Hoeffding and Bernstein, but they speak for centered random variables. Here we set up and prove Lemma 4 just for convenience.

Proof. Note that

$$P(X > \epsilon) \leq e^{-\lambda\epsilon} Ee^{\lambda X}.$$

By the Taylor series for $Ee^{\lambda X}$, we have

$$
\begin{aligned}
Ee^{\lambda X} &= 1 + \lambda EX + \frac{1}{2}\lambda^2 E(X^2) + \sum_{j=3}^{\infty} \frac{\lambda^j}{j!} E(X^j) \\
&= 1 + \frac{1}{2}\lambda^2\sigma^2 + \sum_{j=3}^{\infty} \frac{\lambda^2\lambda^{j-2}}{j!} E(X^{j-2}X^2) \quad \text{when } j \geq 3, j! \geq 2 \times 3^{j-2} \\
&\leq 1 + \frac{1}{2}\lambda^2\sigma^2 + \sum_{j=3}^{\infty} (\frac{\lambda c_3}{3})^{j-2}\frac{\lambda^2\sigma^2}{2} \\
&\leq 1 + \frac{1}{2}\lambda^2\sigma^2 + \frac{1}{2}\lambda^2\sigma^2 \leq e^{\lambda^2\sigma^2},
\end{aligned}
$$

where λ is a positive constant. If $\lambda = \epsilon/(2\sigma^2)$ and $\lambda \leq 3/(2c_3)$, then we obtain

$$P(X > \epsilon) \leq e^{-\epsilon^2/4\sigma^2}.$$

It completes the proof. $\qquad\square$

Lemma 5. *Suppose that Conditions 1 and 2 hold. Then,*

$$
P\left(\sup_{t\in[a,b]} \max_{i\in[2n-1]} |F_i(t, \theta^*(t))| \geq \sqrt{C_1 \frac{Nn\log(Nnh)}{h} + C_2 C_{Nn} Nnh^2} \right) = o(1),
$$

where $C_1 > 6$ is a constant and $C_2 > k_{21}\sup_{t\in[a,b]}(f'(t) + f(t)/2)$.

Proof. We begin with the following decomposition

$$F_i(t, \theta^*(t)) = F_i(t, \theta^*(t)) - \mathbb{E}F_i(t, \theta^*(t)) + \mathbb{E}F_i(t, \theta^*(t)), \ i \in [2n-1].$$

The proof consists of four steps:

    (1) We establish that the bias term $\mathbb{E}F_i(t, \theta^*(t))$ is of order $O(C_{Nn}h^2)$;

(2) We show the variance $\mathrm{Var}\big(F_i(t, \theta^*(t)) - \mathbb{E}F_i(t, \theta^*(t))\big) = O(Nn/h)$;

(3) For any fixed $t \in [a, b]$, we show

$$P\Big(\max_{i \in [2n-1]} |F_i(t, \theta^*(t))| > \sqrt{C_1 Nn \log(Nnh)/h} + C_2 C_{Nn} Nnh^2\Big) = o(1);$$

(4) Combining (1)-(3), we show that Lemma 5 holds.

We first show (1). For $i \in [n]$,

$$\mathbb{E}F_i(t, \theta^*(t))$$

$$= \sum_{\substack{j:1 \leq j \leq n \\ j \neq i}} \sum_{\ell \in [N]} \mathbb{E}\Big\{K_h(t - T_\ell)\big[a_{ij}(T_\ell) - u_{ij}(t)\big]\Big\}$$

$$= \sum_{\substack{j:1 \leq j \leq n \\ j \neq i}} \sum_{\ell \in [N]} \mathbb{E}\Big\{K_h(t - T_\ell)\big[u_{ij}(T_\ell) - u_{ij}(t)\big]\Big\}$$

$$= \frac{N}{h} \sum_{\substack{j:1 \leq j \leq n \\ j \neq i}} \int K\Big(\frac{s-t}{h}\Big)\big(u_{ij}(s) - u_{ij}(t)\big)f(s)ds$$

$$= N \sum_{\substack{j:1 \leq j \leq n \\ j \neq i}} \int K(w)\big(u_{ij}(wh + t) - u_{ij}(t)\big)f(wh + t)dw$$

$$= N \sum_{\substack{j:1 \leq j \leq n \\ j \neq i}} \int K(w)\Big(u'_{ij}(t)wh + \frac{h^2}{2}u''_{ij}(t)w^2 + o(h^2)\Big)\Big(f(t) + f'(t)wh + \frac{h^2}{2}f''(t)w^2 + o(h^2)\Big)dw$$

$$= N \sum_{\substack{j:1 \leq j \leq n \\ j \neq i}} \Big(hf(t)u'_{ij}(t)\int wK(w)dw + h^2\Big(u'_{ij}(t)f'(t) + \frac{1}{2}u''_{ij}(t)f(t)\Big)k_{21} + o(h^2)\Big)$$

$$\leq C_2 C_{Nn} Nnh^2(1 + o(1)).$$

The last inequality holds due to Condition 1 and $\int wK(w)dw = 0$. Then we have

$$\mathbb{E}F_i(t, \theta^*(t)) = O(C_{Nn} Nnh^2), \quad i \in [n].$$

Similarly, we can show

$$\mathbb{E}F_{n+j}(t, \theta^*(t)) = O(C_{Nn} Nnh^2), \quad j \in [n-1].$$

We next show (2). Note that

$$\mathrm{Var}\Big(F_i(t, \theta^*(t)) - \mathbb{E}F_i(t, \theta^*(t))\Big) = N \sum_{\substack{j:1 \leq j \leq n \\ j \neq i}} \mathrm{Var}\Big\{K_h(t - T_1)\big[a_{ij}(T_1) - u_{ij}(t)\big]\Big\}.$$

It can be shown that

$$\mathbb{E}\left\{K_h^2(t-T_1)\big[a_{ij}(T_1)-u_{ij}(t)\big]^2\right\}$$

$$=\frac{1}{h^2}\int K^2\left(\frac{t-s}{h}\right)\left(u_{ij}(s)(1-2u_{ij}(t))+u_{ij}^2(t)\right)f(s)ds$$

$$=\frac{1}{h}\int K^2(w)\left(u_{ij}(wh+t)(1-2u_{ij}(t))+u_{ij}^2(t)\right)f(wh+t)dw$$

$$=\frac{1}{h}\int K^2(w)\left[\left((1-2u_{ij}(t))(u_{ij}(t)+u_{ij}'(t)wh+\frac{1}{2}u_{ij}''(t)w^2h^2+o(h^2))\right)+u_{ij}^2(t)\right]$$

$$\times\left(f(t)+f'(t)wh+\frac{1}{2}f''(t)w^2h^2+o(h^2)\right)dw$$

$$=\frac{1}{h}\Big[u_{ij}(t)(1-u_{ij}(t))f(t)\int K^2(w)dw$$

$$+h\int wK^2(w)dw\left(u_{ij}(t)(1-u_{ij}(t))f'(t)+f(t)(u_{ij}'(t)-2u_{ij}(t)u_{ij}'(t))\right)+O(C_{Nn}h^2)\Big]$$

$$\leq O(1/h).$$

The last inequality holds due to Condition 1, $\int wK(w)dw = 0$ and the fact $u_{ij}(t)(1-u_{ij}(t)) \leq 1/4$. Thus, we obtain

$$\mathrm{Var}\Big(F_i(t,\theta^*(t))\Big) = O\big(Nn/h\big),\ i\in[n].$$

Similarly, it can be shown

$$\mathrm{Var}\Big(F_{n+j}(t,\theta^*(t))\Big) = O\big(Nn/h\big),\ j\in[n-1].$$

For (3), applying the Lemma 4 yields

$$P\Big(\max_{i\in[2n-1]}\Big|F_i(t,\theta^*(t))-\mathbb{E}F_i(t,\theta^*(t))\Big| > \epsilon\Big) \leq Nnh\exp\left\{\frac{-\epsilon^2}{\mathrm{Var}(F_i(t,\theta^*(t)))}\right\}.$$

By setting

$$\epsilon = \sqrt{C_1Nn\log(Nnh)/h} > \sqrt{4Nn\log(Nnh)/h},$$

we have

$$Nnh\exp\left\{\frac{-\epsilon^2}{4\mathrm{Var}(F_i(t,\theta^*(t)))}\right\} = (Nnh)^{1-C_1/4},$$

which converges to 0 as $Nnh \to \infty$ if $C_1 > 6$.

To show (4), let $\{t_k, k\in[J_n-1]\}$ be the equal grid points on $\mathcal{T} = [a,b]$, and let t_k^0 be a interior point of $[t_{k-1},t_k)$. Let $\delta = \sqrt{C_1Nn\log(Nnh)/h}+C_2C_{Nn}Nnh^2$. Then, it can be

shown

$$P\Big(\sup_{t\in[a,b]} \max_{i\in[2n-1]} |F_i(t,\theta^*(t))| > \delta\Big)$$

$$=P\Big(\bigcup_{k\in[J_n]} \sup_{t\in[t_{k-1},t_k)} \max_{i\in[2n-1]} |F_i(t,\theta^*(t))| > \delta\Big)$$

$$\leq \sum_{k\in[J_n]} P\Big(\sup_{t\in[t_{k-1},t_k)} \max_{i\in[2n-1]} |F_i(t,\theta^*(t))| > \delta\Big)$$

$$\leq \sum_{k\in[J_n]} P\Big(\sup_{t\in[t_{k-1},t_k)} \max_{i\in[2n-1]} |F_i(t_k^0,\theta^*(t_k^0)) - F_i(t,\theta^*(t)| + \sup_{t\in[t_{k-1},t_k)} \max_{i\in[2n-1]} |F_i(t_k^0,\theta^*(t_k^0))| > \delta\Big)$$

$$\leq \sum_{k\in[J_n]} P\Big(\sup_{t\in[t_{k-1},t_k)} \max_{i\in[2n-1]} |F_i(t_k^0,\theta^*(t_k^0)) - F_i(t,\theta^*(t))| > \frac{\delta}{2}\Big)$$

$$+ \sum_{k\in[J_n]} P\Big(\max_{i\in[2n-1]} |F_i(t_k^0,\theta^*(t_k^0))| > \frac{\delta}{2}\Big)$$

$$=B_1 + B_2$$

Since $F_i(t,\theta^*(t))$ is the Lipschitz continuous with respect to t, we have

$$|F_i(t,\theta^*(t)) - F_i(t_k^0,\theta^*(t))| \leq C_4 Nn|t - t_k^0| < \delta/2$$

by choosing $J_n > 2(b-a)C_4/\delta$. Therefore, $B_1 = 0$. By steps (1)-(3), $B_2 = o(1)$. That is,

$$P\Big(\sup_{t\in[a,b]} \max_{i\in[2n-1]} |F_i(t,\theta^*(t))| > \sqrt{C_1 \frac{Nn\log(Nnh)}{h} + C_2 C_{Nn} Nnh^2}\Big) = o(1).$$

It completes the proof. $\square$

Lemma 6. *Suppose that Conditions 1-2 hold. Then the Jacobian matrix of $F(t,\theta(t))/(Nn)$ at $\theta(t)$ is Lipschitz continuous on $\mathbb{R}^{2n-1}$.*

Proof. Let

$$\frac{\partial F_i(t,\theta(t))/(Nn)}{\partial\theta(t)} = \frac{1}{Nn}\Big\{\frac{\partial F_i(t,\theta(t))}{\partial\alpha_1(t)}, \ldots, \frac{\partial F_i(t,\theta(t))}{\partial\alpha_n(t)}, \frac{\partial F_i(t,\theta(t))}{\partial\beta_1(t)}, \ldots, \frac{\partial F_i(t,\theta(t))}{\partial\beta_{n-1}(t)}\Big\}$$

denote the ith column of $F'(t, \theta(t))/(Nn)$. Under model (2), for $i \in [n]$, we have

$$\frac{\partial F_i(t, \theta(t))/(Nn)}{\partial \alpha_i(t)} = -\frac{1}{Nn} \sum_{\ell \in [N]} \sum_{\substack{j:1 \leq j \leq n \\ j \neq i}} K_h(t - T_\ell) \frac{e^{\alpha_i(t) + \beta_j(t)}}{(1 + e^{\alpha_i(t) + \beta_j(t)})^2},$$

$$\frac{\partial F_i(t, \theta(t))/(Nn)}{\partial \alpha_s(t)} = 0, \quad s \in [n], \ s \neq i,$$

$$\frac{\partial F_i(t, \theta(t))/(Nn)}{\partial \beta_j(t)} = -\frac{1}{Nn} \sum_{\ell \in [N]} K_h(t - T_\ell) \frac{e^{\alpha_i(t) + \beta_j(t)}}{(1 + e^{\alpha_i(t) + \beta_j(t)})^2}, \quad j \in [n-1], j \neq i,$$

$$\frac{\partial F_i(t, \theta(t))/(Nn)}{\partial \beta_i(t)} = 0, \quad \text{otherwise.}$$

Moreover, for $i \in [n]$, we have

$$\frac{\partial^2 F_i(t, \theta(t))/(Nn)}{\partial \alpha_i^2(t)} = -\frac{1}{Nn} \sum_{\ell \in [N]} \sum_{\substack{j:1 \leq j \leq n \\ j \neq i}} K_h(t - T_\ell) \frac{e^{\alpha_i(t) + \beta_j(t)} (1 - e^{\alpha_i(t) + \beta_j(t)})}{(1 + e^{\alpha_i(t) + \beta_j(t)})^3},$$

$$\frac{\partial^2 F_i(t, \theta(t))/(Nn)}{\partial \alpha_i(t)\beta_j(t)} = -\frac{1}{Nn} \sum_{\ell \in [N]} K_h(t - T_\ell) \frac{e^{\alpha_i(t) + \beta_j(t)} (1 - e^{\alpha_i(t) + \beta_j(t)})}{(1 + e^{\alpha_i(t) + \beta_j(t)})^3}, \quad j \in [n-1], j \neq i,$$

$$\frac{\partial^2 F_i(t, \theta(t))/(Nn)}{\partial \beta_j^2(t)} = -\frac{1}{Nn} \sum_{\ell \in [N]} K_h(t - T_\ell) \frac{e^{\alpha_i(t) + \beta_j(t)} (1 - e^{\alpha_i(t) + \beta_j(t)})}{(1 + e^{\alpha_i(t) + \beta_j(t)})^3} \quad j \in [n-1].$$

$$\frac{\partial^2 F_i(t, \theta(t))}{\partial \alpha_i(t)\partial \beta_i(t)} = 0, \quad \frac{\partial^2 F_i(t, \theta(t))}{\partial \alpha_s(t)\partial \alpha_k(t)} = 0, \quad s \neq k, \quad \frac{\partial^2 F_i(t, \theta(t))}{\partial \beta_s(t)\partial \beta_k(t)} = 0, \quad s \neq k.$$

We next calculate H_1 and H_2 defined in Lemma 1. By some arguments similar to the proof of (2) in Lemma 5, we have $\text{Var}(F_i'(t, \theta(t))/(Nn)) = O(1/(Nnh))$. Then by Chebyshev's inequality, for any $\epsilon > 0$, we have

$$P\Big(\Big| \frac{1}{Nn} F_i'(t, \theta(t)) - \frac{1}{Nn} \mathbb{E} F_i'(t, \theta(t)) \Big| > \epsilon \Big) = O\Big(\frac{1}{Nnh\epsilon^2} \Big).$$

Thus, it holds

$$\frac{1}{Nn} F_i'(t, \theta(t)) \leq \epsilon + \frac{1}{Nn} \mathbb{E} F_i'(t, \theta(t))$$

with probability tending to one. Similarly, it can be shown that

$$\frac{1}{Nn} F_i''(t, \theta(t)) \leq \epsilon + \frac{1}{Nn} \mathbb{E} F_i''(t, \theta(t))$$

with probability tending to one.

Note that

$$\left| \frac{e^x(1 - e^x)}{(1 + e^x)^3} \right| \leq 1/4.$$

7

For $i \in [n]$, we have

$$\left|\frac{\partial^2 F_i(t, \theta(t))/(Nn)}{\partial \alpha_i^2(t)}\right| \leq \frac{1}{Nn} \sum_{\ell \in [N]} \sum_{\substack{j:1 \leq j \leq n \\ j \neq i}} K_h(t - T_\ell) \left|\frac{e^{\alpha_i(t) + \beta_j(t)}(1 - e^{\alpha_i(t) + \beta_j(t)})}{(1 + e^{\alpha_i(t) + \beta_j(t)})^3}\right|$$

$$\leq \frac{1}{4Nn} \sum_{\ell \in [N]} \sum_{\substack{j:1 \leq j \leq n \\ j \neq i}} K_h(t - T_\ell)$$

$$\left|\frac{\partial^2 F_i(t, \theta(t))/(Nn)}{\partial \alpha_i(t)\beta_j(t)}\right| \leq \frac{1}{4Nn} \sum_{\ell \in [N]} K_h(t - T_\ell), \quad j \in [n-1], j \neq i,$$

$$\left|\frac{\partial^2 F_i(t, \theta(t))/(Nn)}{\partial \beta_j^2(t)}\right| \leq \frac{1}{4Nn} \sum_{\ell \in [N]} K_h(t - T_\ell), \quad j \in [n-1],$$

$$\sum_{j \in [n-1]} \left|\frac{\partial^2 F_i(t, \theta(t))/(Nn)}{\partial \alpha_i(t)\beta_j(t)}\right| \leq \frac{1}{4N} \sum_{\ell \in [N]} K_h(t - T_\ell), \quad j \in [n-1], j \neq i,$$

$$\sum_{k \in [n-1]} \left|\frac{\partial^2 F_i(t, \theta(t))/(Nn)}{\partial \alpha_j(t)\beta_k(t)}\right| \leq \frac{1}{4N} \sum_{\ell \in [N]} K_h(t - T_\ell), \quad j \neq i, k \neq i, j.$$

By the mean value theorem for vector-valued functions (Lang, 1993), for any $\mathbf{x}, \mathbf{y} \in \mathbb{G}$, we have

$$\frac{1}{Nn} F_i'(\mathbf{x}, t) - \frac{1}{Nn} F_i'(\mathbf{y}, t) = J_i(\mathbf{x} - \mathbf{y}, t),$$

where

$$J_{i,j,k}(\mathbf{x} - \mathbf{y}, t) = \int_0^1 \frac{\partial F_{i,j}'((s\mathbf{x} + (1-s)\mathbf{y}), t)/(Nn)}{\partial \theta_k(t)} ds, \quad j, k \in [2n-1].$$

Note that for $i \in [n]$,

$$\left|\mathbb{E}\frac{\partial^2 F_i(t, \theta(t))/(Nn)}{\partial \alpha_i^2(t)}\right| \leq (f(t) + \frac{1}{2}k_{21}f''(t)h^2 + o(h^2)) \left|\frac{e^{\alpha_i(t) + \beta_j(t)}(1 - e^{\alpha_i(t) + \beta_j(t)})}{(1 + e^{\alpha_i(t) + \beta_j(t)})^3}\right|$$

$$\leq (f(t) + \frac{1}{2}k_{21}f''(t)h^2 + o(h^2))/4,$$

$$\left|\mathbb{E}\frac{\partial^2 F_i(t, \theta(t))/(Nn)}{\partial \alpha_i(t)\beta_j(t)}\right| \leq (f(t) + \frac{1}{2}k_{21}f''(t)h^2 + o(h^2))/(4n), \quad j \in [n-1], j \neq i,$$

$$\left|\mathbb{E}\frac{\partial^2 F_i(t, \theta(t))/(Nn)}{\partial \beta_j^2(t)}\right| \leq (f(t) + \frac{1}{2}k_{21}f''(t)h^2 + o(h^2))/(4n), \quad j \in [n-1],$$

$$\sum_{j \in [n-1]} \left|\mathbb{E}\frac{\partial^2 F_i(t, \theta(t))/(Nn)}{\partial \alpha_i(t)\beta_j(t)}\right| \leq (f(t) + \frac{1}{2}k_{21}f''(t)h^2 + o(h^2))/4, \quad j \in [n-1], j \neq i,$$

$$\sum_{k \in [n-1]} \left|\mathbb{E}\frac{\partial^2 F_i(t, \theta(t))/(Nn)}{\partial \alpha_j(t)\beta_k(t)}\right| \leq (f(t) + \frac{1}{2}k_{21}f''(t)h^2 + o(h^2))/4, \quad j \neq i, k \neq i, j.$$

Therefore,

$$\max_j \sum_{k \in [2n-1]} |J_{i,j,k}(\mathbf{x} - \mathbf{y}, t)| \le \frac{1}{4N} \sum_{\ell \in [N]} K_h(t - T_\ell) \le (f(t) + \frac{1}{2}k_{21}f''(t)h^2 + o(h^2))/2,$$

$$\sum_{k,j} |J_{i,j,k}(\mathbf{x} - \mathbf{y}, t)| \le f(t) + \frac{1}{2}k_{21}f''(t)h^2 + o(h^2).$$

Similarly, for $i = n + 1, \ldots, 2n - 1$, we have

$$\frac{1}{Nn}F_i'(\mathbf{x}, t) - \frac{1}{Nn}F_i'(\mathbf{y}, t) = J_i(\mathbf{x} - \mathbf{y}, t)$$

and

$$\sum_{k,j} |J_{i,j,k}(\mathbf{x} - \mathbf{y}, t)| \le f(t) + \frac{1}{2}k_{21}f''(t)h^2 + o(h^2).$$

Thus, for $i \in [2n - 1]$ and any $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{G}$ we have

$$\left\| [\frac{1}{Nn}F_i'(\mathbf{x}, t) - \frac{1}{Nn}F_i'(\mathbf{y}, t)]\mathbf{z} \right\|_\infty$$
$$\le \max_i \sum_{j \in [2n-1]} \left| [\frac{1}{Nn}F_{ij}'(\mathbf{x}, t) - \frac{1}{Nn}F_{ij}'(\mathbf{y}, t)]z_j \right|$$
$$\le \max_i \sum_{j \in [2n-1]} \int_0^1 \left| \frac{\partial F_{i,j}'((s\mathbf{x} + (1-s)\mathbf{y}), t)/(Nn)}{\partial \theta_k(t)} \right| ds \times \|\mathbf{z}\|_\infty \times \|\mathbf{x} - \mathbf{y}\|_\infty$$
$$\le (f(t) + \frac{1}{2}k_{21}f''(t)h^2 + o(h^2))/2\|\mathbf{z}\|_\infty \|\mathbf{x} - \mathbf{y}\|_\infty$$

with probability tending to one. This implies that we can take $H_1 = f(t)$ as $h \to 0$. In addition, we can show

$$\|[\frac{1}{Nn}F_i'(\mathbf{x}, t) - \frac{1}{Nn}F_i'(\mathbf{y}, t)]\|_\infty \le \|J_i(\mathbf{x} - \mathbf{y}, t)\|_\infty \times \|\mathbf{x} - \mathbf{y}\|_\infty$$
$$\le \max_j \sum_{k \in [2n-1]} \left| J_{i,j,k}(\mathbf{x} - \mathbf{y}, t) \right| \times \|\mathbf{x} - \mathbf{y}\|_\infty$$
$$\le (f(t) + \frac{1}{2}k_{21}f''(t)h^2 + o(h^2))\|\mathbf{x} - \mathbf{y}\|_\infty$$

with probability tending to one, which implies that we can set $H_2 = 2f(t)$ as $h \to 0$. It completes the proof. $\square$

Lemma 7. *Let* $U(t, \theta^*(t)) = Cov\{R(t, \theta^*(t))\{F(t, \theta^*(t)) - \mathbb{E}F(t, \theta^*(t))\}/(Nn)\}$, *where*

9

$R(t, \theta^*(t)) = V^{-1}(t, \theta^*(t)) - S(t, \theta^*(t))$. *Then,*

$$\|U(t, \theta^*(t))\|_{\max} = O\Big(\frac{Q_{Nnh}^6 h}{Nn}\Big).$$

Proof. Let

$$C(t, \theta^*(t)) = \big\{R(t, \theta^*(t))/(Nn)\big\}\mathrm{Var}\big\{F(t, \theta^*(t)) - \mathbb{E}F(t, \theta^*(t))\big\},$$

where $\mathrm{Var}\big\{F(t, \theta^*(t)) - \mathbb{E}F(t, \theta^*(t))\big\} = (\sigma_{ij}(t, \theta^*(t)))$. By Lemma 2, we have

$$\begin{aligned}
\|C(t, \theta^*(t))\|_{\max} &= \big\|\{R(t, \theta^*(t))/(Nn)\}\mathrm{Var}\{F(t, \theta^*(t)) - \mathbb{E}F(t, \theta^*(t))\}\big\|_{\max} \\
&\leq \max_k \big|R_{ik}(t, \theta^*(t))/(Nn)\big| \times \sum_{k \in [2n-1]} \Big|\sigma_{kj}(t, \theta^*(t))\Big| \\
&\leq O\Big(\frac{Q_{Nnh}^3 h}{Nn^2}\Big) \times O\Big(\frac{Nn^2}{h}\Big) \\
&= O\Big(Q_{Nnh}^3\Big).
\end{aligned}$$

Then we have

$$\begin{aligned}
\|U(t, \theta^*(t))\|_{\max} &= \big\|C(t, \theta^*(t))\{R(t, \theta^*(t))/(Nn)\}^\top\big\|_{\max} \\
&\leq \sum_{k \in [2n-1]} c_{ik}(t, \theta^*(t)) \max_j |R_{kj}(t, \theta^*(t))/(Nn)| \\
&\leq O\Big(nQ_{Nnh}^3\Big) \times O\Big(\frac{Q_{Nnh}^3 h}{Nn^2}\Big) \\
&= O\Big(\frac{Q_{Nnh}^6 h}{Nn}\Big).
\end{aligned}$$

It completes the proof. $\qquad\square$

Lemma 8. *Under the conditions of Theorem 2, we have*

$$\Big\|\frac{1}{Nn}V^{-1}(t, \theta^*(t))\widehat{\gamma}(t, \theta^*(t)\Big\|_{\max} = o_p\big((Nnh)^{-1/2}\big).$$

Proof. Let $V(t, \theta(t)) = \widehat{f}(t)V_0(t, \theta(t))$ and $S(t, \theta(t)) = \mathcal{S}(V(t, \theta(t)))$, where

$$\widehat{f}(t) = N^{-1}\sum_{\ell \in [N]} K_h(t - T_\ell).$$

Then, we can write

$$\frac{1}{Nn}V^{-1}(t,\theta^*(t))\widehat{\gamma}(t,\theta^*(t)) = \frac{1}{Nn}S(t,\theta^*(t))\widehat{\gamma}(t,\theta^*(t)) + \frac{1}{Nn}R(t,\theta^*(t))\widehat{\gamma}(t,\theta^*(t)),$$

where $R(t,\theta^*(t)) = V^{-1}(t,\theta^*(t)) - S(t,\theta^*(t))$. For $i \in [2n-1]$, note that

$$\left\{S(t,\theta^*(t))\widehat{\gamma}(t,\theta^*(t))\right\}_i = \widehat{\gamma}_i(t,\theta^*(t))/\{V(t,\theta^*(t))\}_{i,i} + (-1)^{\mathbb{1}_{i>n}}\widehat{\gamma}_{2n}(t,\theta^*(t))/\{V(t,\theta^*(t))\}_{2n,2n},$$

where $\{V(t,\theta^*(t))\}_{2n,2n} = 2\sum_{i=1}^n\{V(t,\theta^*(t))\}_{i,i} - \sum_{i=1}^n\sum_{j=1}^{2n-1}\{V(t,\theta^*(t))\}_{i,j}$. A direct calculation yields

$$\left|\frac{1}{Nn}\left\{S(t,\theta^*(t))\widehat{\gamma}(t,\theta^*(t))\right\}_i\right| \le \frac{1}{Nn}\left\{\frac{|\widehat{\gamma}_i(t,\theta^*(t))|}{\{V(t,\theta^*(t))\}_{i,i}} + \frac{|\widehat{\gamma}_{2n}(t,\theta^*(t))|}{\{V(t,\theta^*(t))\}_{2n,2n}}\right\}$$
$$\le O\left(Q_{Nnh}^7\left(\sqrt{C_1\frac{\log(Nnh)}{Nnh}} + C_2C_{Nn}h^2\right)^2\right),$$

where

$$\widehat{\gamma}_{2n}(t,\theta^*(t)) = \sum_{i=1}^n\widehat{\gamma}_i(t,\theta^*(t)) - \sum_{j=1}^{n-1}\widehat{\gamma}_{n+j}(t,\theta^*(t)) = \sum_{k=1}^{n-1}\widehat{\gamma}_{kn}(t,\theta^*(t)).$$

By Lemma 3, we have

$$\left|\frac{1}{Nn}\left\{R(t,\theta^*(t))\widehat{\gamma}(t,\theta^*(t))\right\}_i\right| \le \|\frac{1}{Nn}R(t,\theta^*(t))\|_\infty \times \left[2(n-1)\max_i|\widehat{\gamma}_i(t,\theta^*(t))|\right]$$
$$\le O\left(Q_{Nnh}^9\left(\sqrt{C_1\frac{\log(Nnh)}{Nnh}} + C_2C_{Nn}h^2\right)^2\right).$$

This together with $Q_{Nnh} = o\left[1/\{(Nnh)^{1/18}(\sqrt{C_1\log(Nnh)/(Nnh)} + C_{Nn}h^2)^2\}\right]$, implies that

$$\left|\frac{1}{Nn}\left\{S(t,\theta^*(t))\widehat{\gamma}(t,\theta^*(t))\right\}_i\right| = o_p\left((Nnh)^{-1/2}\right) \quad \text{and}$$
$$\left|\frac{1}{Nn}\left(R(t,\theta^*(t))\widehat{\gamma}(t,\theta^*(t))\right)_i\right| = o_p\left((Nnh)^{-1/2}\right).$$

Thus,

$$V^{-1}(t,\theta^*(t))\widehat{\gamma}(t,\theta^*(t))/(Nn) = o_p\left((Nnh)^{-1/2}\right).$$

This completes the proof. $\square$

Lemma 9. *Under the conditions of Theorem 2, we have*

$$\left\| \frac{1}{Nn} V^{-1}(t, \theta^*(t)) \mathbb{E} F(t, \theta^*(t)) - k_{21} h^2 \mu(t) \right\|_{\max} = o_p\left((Nnh)^{-1/2} \right),$$

where $\mu(t) = (\mu_1(t), \ldots, \mu_{2n-1}(t))^\top$ with

$$\mu_i(t) = \left\{ \frac{\sum_{j \in [n], j \neq i} \mu_{ij}(t)/n}{f(t) \{V_0(t, \theta^*(t))\}_{i,i}} \right\} \{1 - \mathbb{1}_{n+1 \leq i \leq 2n-1}\} + \left\{ \frac{\sum_{j \in [n], j \neq i-n} \mu_{j,i-n}(t)/n}{f(t) \{V_0(t, \theta^*(t))\}_{i,i}} \right\} \mathbb{1}_{n+1 \leq i \leq 2n-1}$$

$$+ (-1)^{n+1 \leq i \leq 2n-1} \frac{\sum_{j=1}^{n-1} \mu_{jn}(t)/n}{f(t) \{V_0(t, \theta^*(t))\}_{2n,2n}}.$$

Proof. We begin with the following decomposition

$$V^{-1}(t, \theta^*(t)) \mathbb{E} F(t, \theta^*(t)) = \bar{S}(t, \theta^*(t)) \mathbb{E} F(t, \theta^*(t)) + R(t, \theta^*(t)) \mathbb{E} F(t, \theta^*(t))$$
$$+ \left\{ S(t, \theta^*(t)) - \bar{S}(t, \theta^*(t)) \right\} \mathbb{E} F(t, \theta^*(t)).$$

Write

$$\frac{1}{Nn} \mathbb{E} F(t, \theta^*(t)) = \left(l_1(t, \theta^*(t)), l_2(t, \theta^*(t)), \ldots, l_{2n-1}(t, \theta^*(t)) \right)^\top,$$

where

$$l_i(t, \theta^*(t)) = \sum_{\substack{j: 1 \leq j \leq n \\ j \neq i}} l_{ij}(t, \theta^*(t)) = \frac{1}{n} \sum_{\substack{j: 1 \leq j \leq n \\ j \neq i}} \left(h^2 k_{21} \mu_{ij}(t) + o(h^2) \right), \quad i, j \in [n],$$

$$l_{n+j}(t, \theta^*(t)) = \sum_{\substack{i: 1 \leq i \leq n \\ i \neq j}} l_{ij}(t, \theta^*(t)) = \frac{1}{n} \sum_{\substack{i: 1 \leq i \leq n \\ i \neq j}} \left(h^2 k_{21} \mu_{ij}(t) + o(h^2) \right), \quad i \in [n], j \in [n-1].$$

Then, we have

$$\left| \left[\frac{1}{Nn} \{S(t, \theta^*(t)) - \bar{S}(t, \theta^*(t))\} \mathbb{E} F(t, \theta^*(t)) \right]_i \right|$$

$$= \left| \frac{1}{\widehat{f}(t)} - \frac{1}{f(t)} \right| \times \left| \left[S_0(t, \theta^*(t)) \mathbb{E} F(t, \theta^*(t))/(Nn) \right]_i \right|$$

$$= \left| \frac{1}{\widehat{f}(t)} - \frac{1}{f(t)} \right| \times \left| \frac{l_i(t, \theta^*(t))}{\{V_0(t, \theta^*(t))\}_{i,i}} + \frac{l_{2n}(t, \theta^*(t))}{\{V_0(t, \theta^*(t))\}_{2n,2n}} \right|$$

$$\leq O\left(Q_{Nnh} \left(\left\{ \frac{\log(1/h)}{Nh} \right\}^{1/2} + h^2 \right) \right) O(C_{Nn} h^2)$$

$$= O\left(C_{Nn} Q_{Nnh} \sqrt{\frac{h^3 \log(1/h)}{N}} \right),$$

12

where

$$l_{2n}(t, \theta^*(t)) = \sum_{i=1}^{n} l_i(t, \theta^*(t)) - \sum_{j=1}^{n-1} l_{n+j}(t, \theta^*(t)) = \sum_{j=1}^{n-1} l_{jn}(t, \theta^*(t)).$$

In addition, direct calculations yield

$$\left\| \frac{1}{Nn} R(t, \theta^*(t)) \mathbb{E} F(t, \theta^*(t)) \right\| = \max_{k} \left\{ R(t, \theta^*(t)) \right\}_{ik} \times \frac{1}{Nn} \sum_{k=1}^{2n-1} \left\{ \mathbb{E} F(t, \theta^*(t))(t, \theta^*(t)) \right\}_{k}$$

$$\leq \frac{h Q_{Nnh}^3}{n} O\big(n C_{Nn} h^2 \big) = O\big(C_{Nn} h^3 Q_{Nnh}^3 \big).$$

Furthermore, for $i \in [2n - 1]$, we have

$$\left\{ \frac{1}{Nn} \bar{S}(t, \theta^*(t)) \mathbb{E} F(t, \theta^*(t)) \right\}_{i} = \left\{ \frac{1}{f(t)} S_0(t, \theta^*(t)) \mathbb{E} \frac{1}{Nn} F(t, \theta^*(t)) \right\}_{i}$$

$$= k_{21} h^2 \mu_i(t) + O(h^2 C_{Nn} Q_{Nnh}),$$

These facts imply that

$$\left\{ \frac{1}{Nn} V^{-1}(t, \theta^*(t)) \mathbb{E} F(t, \theta^*(t)) \right\}_{i}$$
$$= k_{21} h^2 \mu_i(t) + O\Big(h^2 C_{Nn} Q_{Nnh} + C_{Nn} Q_{Nnh} \sqrt{h^3 \log(1/h)/N} + C_{Nn} h^3 Q_{Nnh}^3 \Big).$$

If $Q_{Nnh} = o\big[1 / \{ (Nnh)^{1/18} (\sqrt{C_1 \log(Nnh)/(Nnh)} + C_{Nn} h^2)^2 \} \big]$ and $C_{Nn}^2 Nnh^5 Q_{Nnh}^2 = o(1)$, then we have

$$\sqrt{Nnh} \Big(h^2 C_{Nn} Q_{Nnh} + C_{Nn} Q_{Nnh} \sqrt{h^3 \log(1/h)/N} + C_{Nn} h^3 Q_{Nnh}^3 \Big) = o(1),$$

which complets the proof. □

Lemma 10. *Under the conditions of Theorem 2, we have*

$$\frac{1}{Nn} V^{-1}(t, \theta^*(t)) \big\{ F(t, \theta^*(t)) - \mathbb{E} F(t, \theta^*(t)) \big\}$$
$$= \frac{1}{Nn} \bar{S}(t, \theta^*(t)) \big\{ F(t, \theta^*(t)) - \mathbb{E} F(t, \theta^*(t)) \big\} + o_p\big((Nnh)^{-1/2} \big).$$

Proof. Note that

$$\frac{1}{Nn}V^{-1}(t,\theta^*(t))\big\{F(t,\theta^*(t)) - \mathbb{E}F(t,\theta^*(t))\big\}$$

$$= \frac{1}{Nn}\bar{S}(t,\theta^*(t))\big\{F(t,\theta^*(t)) - \mathbb{E}F(t,\theta^*(t))\big\}$$

$$+ \frac{1}{Nn}\big\{S(t,\theta^*(t)) - \bar{S}(t,\theta^*(t))\big\}\big\{F(t,\theta^*(t)) - \mathbb{E}F(t,\theta^*(t))\big\}$$

$$+ \frac{1}{Nn}R(t,\theta^*(t))\big\{F(t,\theta^*(t)) - \mathbb{E}F(t,\theta^*(t))\big\}.$$

Next we show

$$\Big\{\frac{1}{Nn}\big[S(t,\theta^*(t)) - \bar{S}(t,\theta^*(t))\big]\big[F(t,\theta^*(t)) - \mathbb{E}F(t,\theta^*(t))\big]\Big\}_i = o_p\big((Nnh)^{-1/2}\big),$$

$$\Big\{\frac{1}{Nn}R(t,\theta^*(t))\big[F(t,\theta^*(t)) - \mathbb{E}F(t,\theta^*(t))\big]\Big\}_i = o_p\big((Nnh)^{-1/2}\big).$$

Recall that

$$V(t,\theta(t)) = \widehat{f}(t)V_0(t,\theta(t)) \quad \text{and} \quad \bar{V}(t,\theta(t)) = f(t)V_0(t,\theta(t)),$$

which implies

$$S(t,\theta(t)) = \frac{1}{\widehat{f}(t)}S_0(t,\theta(t)) \quad \text{and} \quad \bar{S}(t,\theta(t)) = \frac{1}{f(t)}S_0(t,\theta(t))$$

where $S_0(t,\theta(t)) = \mathcal{S}(V_0(t,\theta(t)))$. By Theorem 6 in [Masry (1996)](#), we have

$$\sup_{t\in[a,b]}|\widehat{f}(t) - f(t)| = O\Big(\Big\{\frac{\log(1/h)}{Nh}\Big\}^{1/2} + h^2\Big).$$

Let

$$G(t,\theta^*(t)) = \frac{1}{Nn}\big\{F(t,\theta^*(t)) - \mathbb{E}F(t,\theta^*(t))\big\}$$

and its ith element $g_i(t,\theta^*(t))$ is defined as

$$g_i(t,\theta^*(t)) = \sum_{\substack{j:1\leq j\leq n\\ j\neq i}} g_{ij}(t,\theta^*(t)) = \frac{1}{Nn}\big\{F_i(t,\theta^*(t)) - \mathbb{E}F_i(t,\theta^*(t))\big\}, \quad i,j \in [n],$$

$$g_{n+j}(t,\theta^*(t)) = \sum_{\substack{i:1\leq i\leq n\\ i\neq j}} g_{ij}(t,\theta^*(t)) = \frac{1}{Nn}\big\{F_{n+j}(t,\theta^*(t)) - \mathbb{E}F_{n+j}(t,\theta^*(t))\big\}, \quad i \in [n], j \in [n-1].$$

14

Then, we have

$$
\left| \left(\frac{1}{Nn} \{ S(t, \theta^*(t)) - \bar{S}(t, \theta^*(t)) \} \{ F(t, \theta^*(t)) - \mathbb{E} F(t, \theta^*(t)) \} \right)_i \right|
$$

$$
= \left| \frac{1}{\widehat{f}(t)} - \frac{1}{f(t)} \right| \times \left| \left(S_0(t, \theta^*(t)) G(t, \theta^*(t)) \right)_i \right|
$$

$$
= \left| \frac{1}{\widehat{f}(t)} - \frac{1}{f(t)} \right| \times \left| \frac{g_i(t, \theta^*(t))}{\{ V_0(t, \theta^*(t)) \}_{i,i}} + \frac{g_{2n}(t, \theta^*(t))}{\{ V_0(t, \theta^*(t)) \}_{2n,2n}} \right|
$$

$$
\leq O_p\left(\left\{ \frac{\log(1/h)}{Nh} \right\}^{1/2} + h^2 \right) O_p\left(Q_{Nnh} \sqrt{\frac{C_1 \log(Nnh)}{Nnh}} \right)
$$

$$
= O_p\left(Q_{Nnh} \sqrt{\frac{C_1 \log(Nnh) \log(1/h)}{N^2 nh^2}} + h^2 Q_{Nnh} \sqrt{\frac{C_1 \log(Nnh)}{Nnh}} \right)
$$

$$
= O_p\left(Q_{Nnh} \sqrt{\frac{C_1 \log(Nnh) \log(1/h)}{N^2 nh^2}} \right),
$$

where $g_{2n}(t, \theta^*(t)) = \sum_{k=1}^{n-1} g_{kn}(t, \theta^*(t))$ and

$$
\{ V(t, \theta^*(t)) \}_{2n,2n} = 2 \sum_{i=1}^{n} \{ V(t, \theta^*(t)) \}_{i,i} - \sum_{i=1}^{n} \sum_{j=1}^{2n-1} \{ V(t, \theta^*(t)) \}_{i,j}.
$$

It is of order $o_p((Nnh)^{-1/2})$ under $Q_{Nnh} = o\left[1 / \{ (Nnh)^{1/18} (\sqrt{C_1 \log(Nnh)/(Nnh)} + C_2 C_{Nn} h^2)^2 \} \right]$. In addition, by Lemma 7, we have

$$
\left\{ \frac{1}{Nn} R(t, \theta^*(t)) \left[F(t, \theta^*(t)) - \mathbb{E} F(t, \theta^*(t)) \right] \right\}_i = o_p((Nnh)^{-1/2}).
$$

These facts imply

$$
\frac{1}{Nn} V^{-1}(t, \theta^*(t)) \{ F(t, \theta^*(t)) - \mathbb{E} F(t, \theta^*(t)) \}
$$

$$
= \frac{1}{Nn} \bar{S}(t, \theta^*(t)) \{ F(t, \theta^*(t)) - \mathbb{E} F(t, \theta^*(t)) \} + o_p\left((Nnh)^{-1/2} \right).
$$

This completes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad\square$

References

S. Lang. *Real and functional analysis.* Real and Functional Analysis, 1993.

E. Masry. Multivariate local polynomial regression for time series: uniform strong consistency and rates. *Journal of Time Series Analysis*, 17(6):571–599, 1996.

T. Yan, C. Leng, and J. Zhu. Asymptotics in directed exponential random graph models with an increasing bi-degree sequence. *The Annals of Statistics*, 44(1):31–57, 2016.
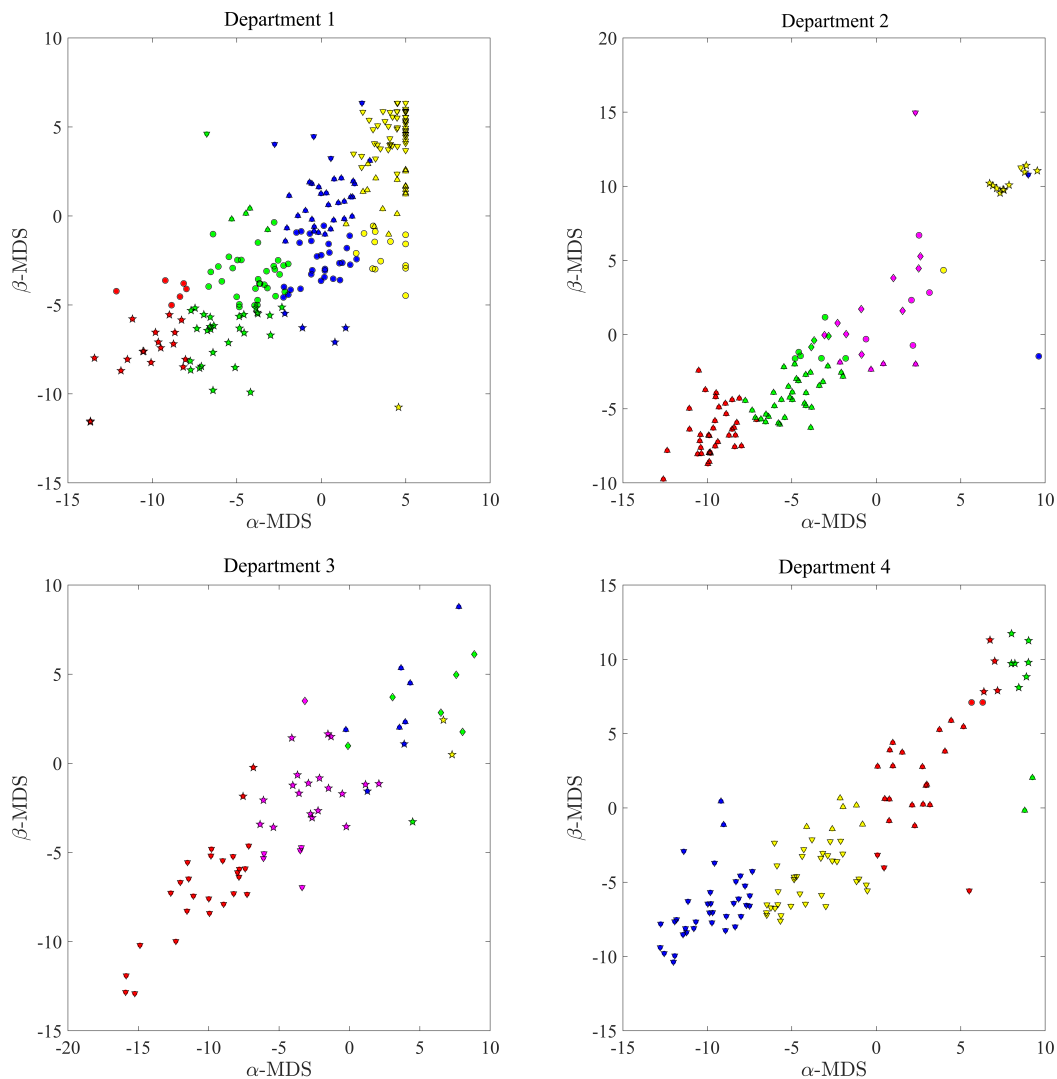
S3. More results of real data analysis

Figure S1: Illustration of $\widehat{\alpha}_i(t)$'s and $\widehat{\beta}_j(t)$'s when $N = 38$. Each point's X-coordinate comes from a multi-dimensional scaling of the ℓ_2 distance matrix between $\widehat{\alpha}_i(t)$'s; similarly obtain its Y-coordinate. Color represents α-clustering result, and marker represents β-clustering result.
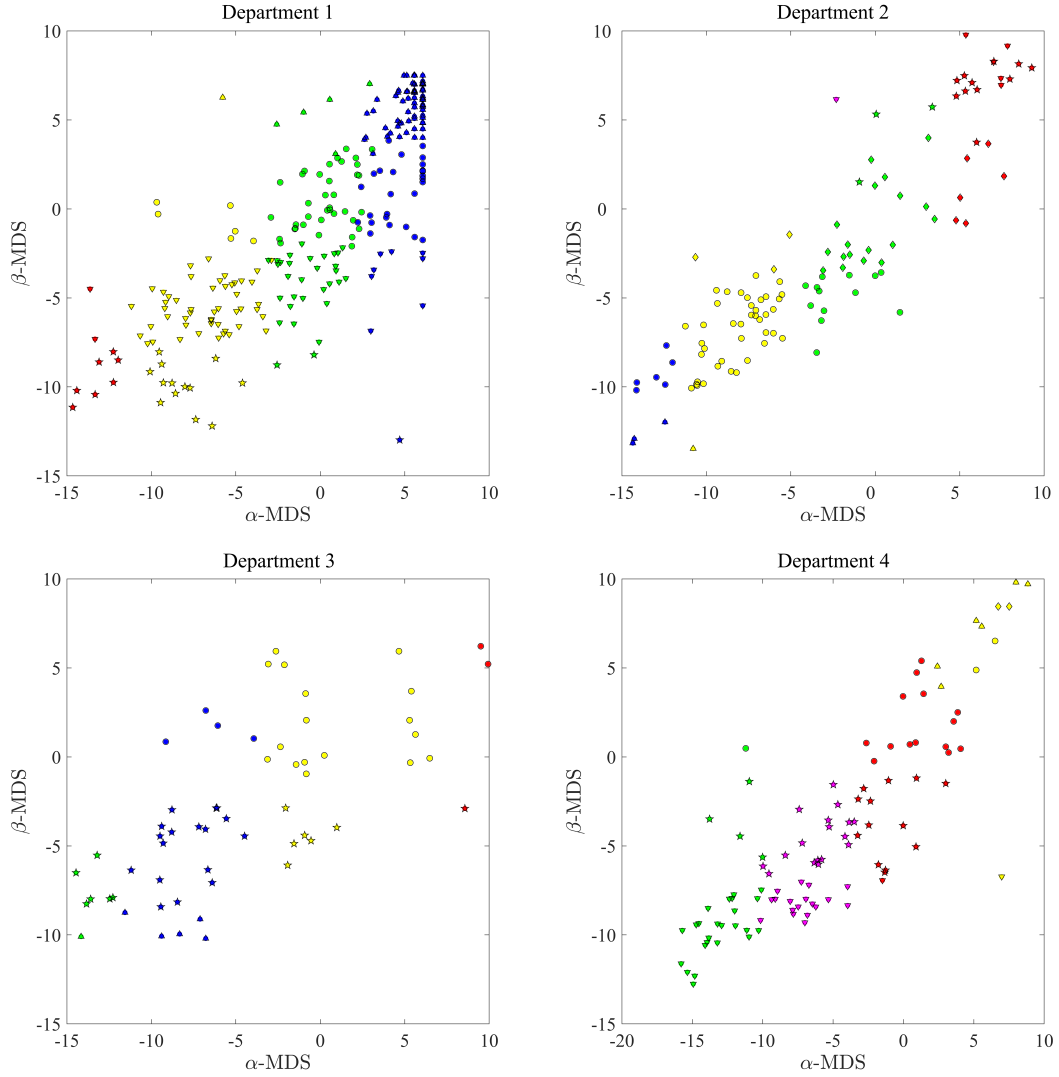
Figure S2: Illustration of $\widehat{\alpha}_i(t)$'s and $\widehat{\beta}_j(t)$'s when $N = 57$. Each point's X-coordinate comes from a multi-dimensional scaling of the ℓ_2 distance matrix between $\widehat{\alpha}_i(t)$'s; similarly obtain its Y-coordinate. Color represents α-clustering result, and marker represents β-clustering result.