Nonparametric Network Vector Autoregression

Zixin Yang[a], Xiaojun Song[b], and Jihai Yu[b]

[a]School of Statistics and Data Science, Shanghai University of Finance and Economics, China

[b]Guanghua School of Management, Peking University, China

Abstract

This paper examines the nonlinear dynamics of a high-dimensional time series collected at the nodes of a large-scale network. We propose a first-order vector autoregressive model that allows the network and momentum effects to be nonlinear and nonparametric. For the model estimation, we propose a sieve least squares estimator and establish its consistency and asymptotic normality. The large sample properties are shown to be valid when either the number of nodes or the number of time periods tends to infinity. We also put forward two specification tests for the linearity of the nonparametric components. The test statistics have an asymptotically standard normal distribution under the null hypothesis and a sequence of Pitman local alternatives. The usefulness of the proposed nonparametric model is illustrated through a user activity analysis using Sina Weibo, China's largest Twitter-like social media platform.

Keywords: *Multivariate time series, Vector autoregression, Sieve estimation, Social network.*

JEL Classification: C14, C32, C36

1 Introduction

Network models are pivotal tools for studying interconnected systems and are widely used in various fields. Classic applications include the diffusion of spillover effects across large networks (He and Song, 2023), financial contagion in stock markets (Chen et al., 2023; Härdle et al., 2016), and the spread of epidemic diseases (Keeling and Eames, 2005). From an empirical standpoint, analyzing networks requires developing novel statistical techniques that can account for complex interactions and handle large datasets, which may be recorded over time. In this work, we study nonlinear autoregressive processes with network dependence. We focus on estimation and inference in the framework of network vector autoregression (Zhu et al., 2017), where the network and momentum effects are allowed to be nonlinear and nonparametric.

Consider a network of N nodes indexed by $i \in \{1, 2, \ldots, N\}$. The network structure is represented by a nonstochastic adjacency matrix $A = (a_{ij})$, where $a_{ij} = 1$ if a direct link exists from i to j and $a_{ij} = 0$ otherwise. Let $Y_{it} \in \mathbb{R}$ be the outcome of interest obtained from node i at time t, and $\mathbb{Y}_t = (Y_{1t}, \ldots, Y_{Nt})^\top \in \mathbb{R}^N$. Modeling the temporal dynamics of $\{\mathbb{Y}_t\}$ is a crucial task that arises frequently in practice. A popular approach is vector autoregression (VAR), where the dynamics of $\{\mathbb{Y}_t\}$ are modeled through a linear operation on its past values (Lütkepohl, 1991). Despite its simple interpretation, inferences based on VAR require estimation of an $N \times N$ transition matrix, which can be challenging when N is much larger than the number of total time periods. Moreover, standard VARs do not exploit known network structure and therefore cannot directly represent dynamic transmission across links. To address this issue, several approaches have been proposed to integrate network structures into the VAR framework, such as the global VAR by Pesaran et al. (2004) and the network autoregression (NAR) by Zhu et al. (2017). Such methods reduce dimensionality by imposing structure, but mostly retain linear dependence. For example, in the NAR model, each Y_{it} is a linear function of its own lag and an average of neighbors' lags, implying constant marginal effects. Such linearity improves tractability but precludes state-dependent propagation mechanisms that are often empirically relevant.

A motivating example is volatility spillovers. When nodes are institutions and Y_{it} denotes return volatility, policymakers are interested in how shocks propagate across firms. Linear NAR has documented significant volatility spillovers (Chen et al., 2023; Zhu et al., 2025), but empirical work indicates transmission is state-dependent: spillovers intensify in periods of extreme stress (Hoque et al., 2024) and only sufficiently large shocks raise conditional volatility in some markets (Jin, 2015). Similar nonlinearities arise in other domains: regional economic shocks may only spill over to neighboring areas beyond certain thresholds, and diffusion of behaviors or innovations on social networks often exhibits threshold, saturation, or asymmetric responses. These patterns imply that spillover magnitudes depend on the state of neighbors rather than being constant. To account for nonlinear network effects, parametric nonlinear NAR models have been proposed (e.g., Armillotta and Fokianos, 2023). Though less restrictive than the linear NAR, these models are still of a parametric form, and thus remain vulnerable to functional misspecification. To fully capture the complex network dependencies, time series models that allow response functions to be estimated in a fully data-driven way are desirable. This motivates the study of nonparametric NAR theory.

Nonparametric dynamic models for multivariate time series have been the subject of a wide range of literature. The seminal work by Härdle et al. (1998) introduced nonlinear VARs in which both the conditional mean and the conditional variance matrix are unknown functions of the past. Jeliazkov (2013) modeled the conditional mean using a Bayesian hierarchical representation of generalized additive models, while Kalli and Griffin (2018) modeled the stationary and transition densities using Bayesian nonparametric methods. Such approaches capture rich nonlinear dependence but suffer from the curse of dimensionality as the cross-sectional dimension grows. When network information is available, exploiting network structure reduces complexity and aids interpretation. Recent NAR extensions follow this strategy – for example, time-varying NAR that allows network and momentum effects to vary over time and nodes (Li et al., 2024; Ding et al., 2025), and functional-coefficient NAR (Yin et al., 2024). Despite these advances, there remains no general methodology with a fully nonparametric treatment of network and autoregressive effects for high-dimensional time series. The present paper aims to fill in this gap.

We propose a first-order VAR model that embeds network structure to characterize dependencies in high-dimensional time series. Our framework extends the linear NAR model by allowing both network and momentum effects to be represented as nonparametric functions of neighbors' lagged responses and a node's historical states, respectively. This nonparametric formulation enables flexible modeling of nonlinear network interactions across large-scale temporal systems, thus ensuring broad applicability in various fields. For model estimation, we employ series approximation techniques and develop a least squares (LS) estimator for the sieve coefficients. While the sieve LS estimation approach is standard, the asymptotic analysis presents technical challenges due to nonlinear interdependencies across time and network links. To handle this, we establish the near-epoch dependence (NED) property of the data-generating process relative to an independent and identically distributed (i.i.d.) innovation sequence, extending the spatial NED methods (Jenish and Prucha, 2012) to a space-time setting. Under standard regularity conditions, the sieve LS estimator is shown to be consistent and asymptotically normal when $\max\{N, T\} \to \infty$, and it also attains the optimal nonparametric rate of Stone (1982). Furthermore, we provide two specification tests for the linearity of the network and momentum effects. The tests are based on the Wald and Lagrangian Multiplier (LM) principles and have an asymptotically standard normal distribution under the null of linearity. The power properties under a sequence of local alternatives are also investigated. Monte Carlo experiments demonstrate the satisfactory finite-sample performance of the proposed estimators and tests. An empirical application to Sina Weibo data illustrates how nonparametric autoregressive effects reveal state-dependent diffusion patterns omitted by linear models.

The rest of the paper is organized as follows. In Section 2, we discuss the stability conditions of the proposed nonparametric NAR model and introduce the sieve LS estimator. The asymptotic properties of the estimator are investigated in Section 3. In Section 4, we propose two specification tests for the null hypothesis of linearity and joint linearity and study their asymptotic distributions. Section 5 presents Monte Carlo experiments to evaluate the finite-sample performance of the proposed estimator and tests. Section 6 presents an empirical analysis of the Sina Weibo data. The last section concludes. The proofs of the theorems, some additional simulation results, and a robustness check of the empirical analysis are relegated to the Online Supplementary Appendix.

Notation. For natural numbers m and n, I_n denotes an $n \times n$ identity matrix, and $\mathbf{0}_{m\times n}$

($\mathbf{1}_{m\times n}$) denotes an $m \times n$ matrix of zeros (ones). For a random variable X, let $\|X\|_p = (\mathbb{E}|X|^p)^{1/p}$. For a matrix A, let $\|A\| = \sqrt{\mathrm{tr}(AA^\top)}$ where $\mathrm{tr}(\cdot)$ is a trace of a matrix. When A is square, $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote the minimum and maximum eigenvalues of A, respectively. Moreover, A^- denotes the Moore–Penrose generalized inverse of A. For two sequences of nonnegative numbers $\{a_N : N \geq 1\}$ and $\{b_N : N \geq 1\}$, $a_N \lesssim b_N$ means there exists a finite positive C such that $a_N < Cb_N$ for all N sufficiently large, and $a_N \asymp b_N$ means $a_N \lesssim b_N$ and $b_N \lesssim a_N$.

2 Model and estimation procedure

2.1 Model and stability conditions

As mentioned in the introduction, we consider a network with N nodes indexed by $i = 1, \ldots, N$, where the neighborhood structure is described by a binary adjacency matrix $A = (a_{ij})$. Self-relationships are excluded, i.e., $a_{ii} = 0$ for any $i = 1, \ldots, N$. For the ith node, a time series of continuous variable Y, denoted by $\{Y_{it}\}_{t=0}^T$, is collected together with a set of node specific covariates $Z_i \in \mathcal{R}_Z$, where $\mathcal{R}_Z$ is a compact subset of $\mathbb{R}^{d_Z}$. To model the dynamics of $\{Y_{it}\}$, we consider the model

$$Y_{it} = n_i^{-1} \sum_{j=1}^{N} a_{ij} f_1(Y_{j,t-1}) + f_2(Y_{i,t-1}) + Z_i^\top \gamma + \epsilon_{it}, \quad i = 1, \ldots, N, \ \ t = 1, \ldots, T, \tag{1}$$

where $n_i = \sum_{j=1}^N a_{ij}$ is the out-degree of node i, f_1 and f_2 are unknown nonparametric functions on $\mathbb{R}$, and ϵ_{it}'s are i.i.d. random noises. Model (1) allows the response of node i to depend on the past value of itself as well as its connections in a nonlinear way. The term $n_i^{-1} \sum_{j=1}^N a_{ij} f_1(Y_{j,t-1})$ captures the average impact of i's neighbors on the response of i, while the term $f_2(Y_{i,t-1})$ captures the impact of i's own lagged response. Hereafter, the functions f_1 and f_2 will be referred to as the network function and the momentum function, respectively. The term $Z_i^\top \gamma$ captures the nodal impact of the ith node, where $\gamma \in \mathbb{R}^{d_Z}$ is the associated coefficient (nodal effect) vector. For identification purpose, we assume that $f_1(0) = f_2(0) = 0$, and the first element of Z_i is 1. The linear NAR model in Zhu et al. (2017)

is a special case of (1) with f_1 and f_2 both being linear. We can also consider the NAR model where the covariate Z_i also enters nonparametrically as $m(Z_i)$. If the dimension of Z_i is not high, standard nonparametric estimation techniques can be used to estimate $m(\cdot)$. However, if the dimension of Z_i is high, a fully nonparametric specification of $m(\cdot)$ could suffer from the curse of dimensionality. It would be desirable to consider a partially linear form or a single-index form of $m(\cdot)$, which will lead to more difficult estimation procedures. Since we focus more on the inferences of the network and momentum effects, we restrict our attention to model (1) for analytical simplicity.

Let $\mathbb{Y}_t = (Y_{1t}, \ldots, Y_{Nt})^\top$, $\mathbb{Z} = (Z_1, \ldots, Z_N)^\top$, and $\mathcal{E}_t = (\epsilon_{1t}, ..., \epsilon_{Nt})^\top$. The matrix form of (1) is

$$\mathbb{Y}_t = WF_1(\mathbb{Y}_{t-1}) + F_2(\mathbb{Y}_{t-1}) + \mathbb{Z}\gamma + \mathcal{E}_t, \quad t = 1, \ldots, T, \tag{2}$$

where $W = \text{diag}\{n_1^{-1}, \ldots, n_N^{-1}\}A$ is the row-normalized adjacency matrix, and $F_\ell(\mathbb{Y}_{t-1}) = (f_\ell(Y_{1,t-1}), \ldots, f_\ell(Y_{N,t-1}))^\top$ for $\ell = 1, 2$. Since $\{\mathbb{Y}_t\}$ is a nonlinear vector autoregressive process, it is of interest to study sufficient conditions that guarantee its asymptotic stability. According to whether the network size N is fixed or $N \to \infty$, Zhu et al. (2017) defined two types of stationarities. They are referred to as, respectively, Type I (N is fixed) and Type II ($N \to \infty$). The following conditions ensure that a stationary solution of (2) exists and is unique for both cases.

Assumption 1. *The errors $\{\epsilon_{it} : 1 \le i \le N, 1 \le t \le T\}$ are i.i.d. with mean zero, variance σ^2, and $\mathbb{E}|\epsilon_{it}|^4 < \infty$; they are also independent of $\{\mathbb{Y}_0\}$. The distribution of ϵ_{it} is absolutely continuous w.r.t. Lebesgue measure on $\mathbb{R}$, and has positive continuous density.*

Assumption 2. *For any $y, y' \in \mathbb{R}$, $|f_\ell(y) - f_\ell(y')| \le \kappa_\ell |y - y'|$, $\ell = 1, 2$, and $\kappa_1 + \kappa_2 < 1$.*

If one adopts the linear NAR specification $Y_{it} = \beta_1 n_i^{-1} \sum_{j=1}^N a_{ij} Y_{j,t-1} + \beta_2 Y_{i,t-1} + Z_i^\top \gamma + \epsilon_{it}$, Assumption 2 reduces to a familiar condition on the space of the network and momentum effect parameter (β_1, β_2), namely, $|\beta_1| + |\beta_2| < 1$.

Proposition 1. *Let Assumptions 1-2 hold. The network size N is either fixed or diverging to infinity. Then there exists a unique strictly stationary solution $\{\mathbb{Y}_t \in \mathbb{R}^N : t \in \mathbb{Z}\}$ to model (2). In addition, if $\mathbb{E}|\epsilon_{it}|^a \le C_{\epsilon,a} < \infty$ for some $a \ge 1$, we have $\max_{i \ge 1} \mathbb{E}|Y_{it}|^a \le C_a < \infty$.*[1]

[1]When $N \to \infty$, $\{\mathbb{Y}_t\}$ is a vector time series of increasing dimension. In this case, the stationarity of

Proposition 1 extends the fixed and increasing network-type results in Zhu et al. (2017), Theorems 1-2 therein, to the nonlinear NAR model (2). When Assumptions 1-2 are not met, model (2) may allow multiple (or no) stationary solutions; however, we do not consider such cases in this paper. Not only for the stability of $\{\mathbb{Y}_t\}$, Assumption 2 also plays a key role in ensuring that the data follow an NED process (see Definition 1 in the Online Supplementary Appendix).

2.2 Sieve estimation

For the unknown functions f_1 and f_2, suppose that $f_1, f_2 \in \mathcal{F}$, where $\mathcal{F}$ is a certain subset of square-integrable continuous functions on $\mathbb{R}$. Then we can approximate them in a finite-dimensional sieve space (e.g., power series, splines, Fourier series, wavelets). Let $\{p_j(\cdot) : j = 1, 2, \ldots\}$ be a sequence of basis functions on $\mathbb{R}$, excluding a constant function, and $p^J(y) = (p_1(y), \ldots, p_J(y))^\top$ for some $J \equiv J_{NT}$. The constraint that $f_1(0) = f_2(0) = 0$ is enforced by translating the basis functions vertically so that $p_j(0) = 0$ for all $j = 1, \ldots, J$.

Denote $\mathcal{F}_J = \{f_J(\cdot) = p^J(\cdot)^\top \beta : \beta \in \mathbb{R}^J\}$ as the sieve space of dimension J. For sufficiently large J_1 and J_2, we can find vectors $\beta_1 = (\beta_{11}, \ldots, \beta_{1J_1})^\top$ and $\beta_2 = (\beta_{21}, \ldots, \beta_{2J_2})^\top$ such that $f_1(\cdot)$ and $f_2(\cdot)$ can be well approximated by $\beta_1^\top p^{J_1}(\cdot) \in \mathcal{F}_{J_1}$ and $\beta_2^\top p^{J_2}(\cdot) \in \mathcal{F}_{J_2}$, respectively. Then, model (1) can be rewritten as $Y_{it} = \beta_1^\top \bar{P}^{J_1}_{i,t-1} + \beta_2^\top P^{J_2}_{i,t-1} + Z_i^\top \gamma + u_{it}$, where $\bar{P}^{J_1}_{i,t-1} = n_i^{-1} \sum_{j=1}^N a_{ij} p^{J_1}(Y_{j,t-1})$, $P^{J_2}_{i,t-1} = p^{J_2}(Y_{i,t-1})$, and u_{it} is the new error term. In matrix form, we have

$$\mathbb{Y}_t = W\mathbb{P}^{J_1}_{t-1}\beta_1 + \mathbb{P}^{J_2}_{t-1}\beta_2 + \mathbb{Z}\gamma + U_t,$$

where $\mathbb{P}^{J_\ell}_{t-1} = (p^{J_\ell}(Y_{1,t-1}), \ldots, p^{J_\ell}(Y_{N,t-1}))^\top$ for $\ell = 1, 2$, and $U_t = (u_{1t}, \ldots, u_{Nt})^\top$. Denote $\mathbb{X}_{t-1} = (W\mathbb{P}^{J_1}_{t-1}, \mathbb{P}^{J_2}_{t-1}, \mathbb{Z})$, and $\theta = (\beta_1^\top, \beta_2^\top, \gamma^\top)^\top$. The sieve LS estimator of the coefficient θ is given by

$$\widehat{\theta} = \left(\sum_{t=1}^T \mathbb{X}_{t-1}^\top \mathbb{X}_{t-1} \right)^- \sum_{t=1}^T \mathbb{X}_{t-1}^\top \mathbb{Y}_t. \tag{3}$$

Denote $\mathbb{Q}_{t-1} = (W\mathbb{P}^{J_1}_{t-1}, \mathbb{P}^{J_2}_{t-1})$ and $\beta = (\beta_1^\top, \beta_2^\top)^\top$. For any $m \times n$ matrix A, let $\mathbf{M}_A = I_m - A(A^\top A)^- A^\top$. By the formula for partitioned regression, we can obtain the estimators

$\{\mathbb{Y}_t\}$ is defined according to Definition 1 in Zhu et al. (2017). Moreover, the notation $\max_{i \geq 1} x_i$ means $\max_{1 \leq i < \infty} x_i$.

of γ and β, respectively, as follows

$$\widehat{\gamma} = \left(\mathbf{Z}^\top \mathbf{M}_{\mathbf{Q}} \mathbf{Z}\right)^{-} \mathbf{Z}^\top \mathbf{M}_{\mathbf{Q}} \mathbf{Y}, \quad \widehat{\beta} = \left(\mathbf{Q}^\top \mathbf{M}_{\mathbf{Z}} \mathbf{Q}\right)^{-} \mathbf{Q}^\top \mathbf{M}_{\mathbf{Z}} \mathbf{Y}, \tag{4}$$

where $\mathbf{Z} = \mathbf{1}_T \otimes \mathbb{Z}$, $\mathbf{Q} = (\mathbb{Q}_0^\top, \ldots, \mathbb{Q}_{T-1}^\top)^\top$, and $\mathbf{Y} = (\mathbb{Y}_1^\top, \ldots, \mathbb{Y}_T^\top)^\top$. Denote $\widehat{\beta} = (\widehat{\beta}_1^\top, \widehat{\beta}_2^\top)^\top$, where $\widehat{\beta}_1$ and $\widehat{\beta}_2$ are of length J_1 and J_2, respectively. The sieve LS estimator of $f_\ell(\cdot)$ is given by $\widehat{f}_\ell(\cdot) = \widehat{\beta}_\ell^\top p^{J_\ell}(\cdot)$ for $\ell = 1, 2$.

Remark 1. *It may be desirable to include some fixed or random effect α_i in model (1) to better characterize the nodal heteroskedasticity. However, since α_i is correlated with $Y_{i,t-1}$, the sieve LS estimator may be inconsistent. Following established practices in the dynamic panel data literature, we can first eliminate the fixed effect through data transformation and then use the instrumental variable method to estimate f_1 and f_2. Since the derivation of the asymptotic theory is much more involved, we leave this for further study.*

3 Asymptotic properties

We consider three asymptotic regimes: (i) $N \to \infty$ and T is fixed, (ii) $T \to \infty$ and N is fixed, and (iii) $(N, T) \to \infty$. In the following, we focus on cases (i) and (iii) to establish the asymptotic properties of the sieve LS estimator while relegating the discussion of case (ii) to Remark 4 and Section C.2 of the Online Supplementary Appendix.

Assumption 3. *(i)* $\max_{i \geq 1} \mathbb{E}|Y_{it}|^{2\omega} < \infty$ *for some* $\omega \geq 1$. *(ii)* $\{Z_i : 1 \leq i \leq N\}$ *are fixed covariates with* $\max_{1 \leq i \leq N} |Z_i|_\infty < \infty$. *The matrix* $\Psi_Z = \lim_{N \to \infty} \mathbb{Z}^\top \mathbb{Z}/N$ *exists and is nonsingular.*

The moment condition in Assumption 3(i) is introduced to deal with responses with unbounded support. If f_1 and f_2 are uniformly bounded on $\mathbb{R}$ or satisfy Assumption 2, then $\max_{i \geq 1} \mathbb{E}|\epsilon_{it}|^{2\omega} < \infty$ and uniform boundedness of Z_i's together imply Assumption 3(i). The nonrandomness of Z_i's facilitates the establishment of NED properties of $\{Y_{it}\}$ along the cross-sectional dimension. It is also essential for the asymptotic analysis under the regime $T \to \infty$ and N is fixed. If Z_i's are stochastic and have bounded support, the theorems in the present paper can be understood as conditional on $\mathbb{Z}$.

Assumption 4. *Let* $J = \max\{J_1, J_2\}$. *(i)* $\max_{i\geq 1}\max_{1\leq j\leq J}\mathbb{E}|p_j(Y_{it})|^2 < \infty$. *(ii) Let* $\Sigma_{N,J_1J_2} = \mathbb{E}(\mathbb{X}_{t-1}^\top\mathbb{X}_{t-1}/N)$, $\bar{\nu}_{J_1J_2} = \limsup_{N\to\infty}\lambda_{max}(\Sigma_{N,J_1J_2})$, *and* $\underline{\nu}_{J_1J_2} = \liminf_{N\to\infty}\lambda_{min}(\Sigma_{N,J_1J_2})$. *Uniformly in* J_1 *and* J_2, *there exists a constant* $0 < \bar{c} < \infty$ *such that* $\bar{\nu}_{J_1J_2} \leq \bar{c}$, *and* $\underline{\nu}_{J_1J_2} > 0$.

Assumption 4(i) is standard in the series estimation literature and can be trivially satisfied by uniformly bounded bases such as B-spline bases and Fourier series. Regrading Assumption 4(ii), the quantities $\bar{\nu}_{J_1J_2}$ and $\underline{\nu}_{J_1J_2}$ are contingent on the distribution of the data, the network structure, and the choice of basis. We allow that $\underline{\nu}_{J_1J_2} \to 0$ as $(J_1, J_2) \to \infty$. Such a setting accommodates the situation where the degree of collinearity between $W\mathbb{P}_{t-1}^{J_1}$ and $\mathbb{P}_{t-1}^{J_2}$ slowly increases as $(J_1, J_2) \to \infty$. In general, it is not easy to show how $\underline{\nu}_{J_1J_2}$ relates to (J_1, J_2) and other model parameters. Under some specific model settings, we conduct Monte Carlo experiments to examine the decay rate of $\underline{\nu}_{J_1J_2}$, the results of which demonstrate a nearly polynomial relationship between $\underline{\nu}_{J_1J_2}^{-1}$ and (J_1, J_2) for commonly used networks. See Section D.1 in the Online Supplementary Appendix for details.

To introduce the following assumption, we define some quantities relevant for the law of large numbers (LLN) under network dependence. Let $D_N = \{1, 2, \ldots, N\}$ be the set of node indices, and let $G_N = (D_N, E_N)$ represent an undirected network on D_N, where E_N is a set of links such that $\{i, j\} \in E_N$ if and only if $a_{ij} = 1$ or $a_{ji} = 1$.

For nodes $i, j \in D_N$, we define $d_N(i, j)$ as the length of the shortest path between them in G_N. Let $\mathcal{N}_N^\partial(i; s)$ denote the set of nodes exactly at the distance s from node i, i.e., $\mathcal{N}_N^\partial(i; s) = \{j \in D_N : d_N(i, j) = s\}$, and let $\delta_N^\partial(s) = N^{-1}\sum_{i=1}^N |\mathcal{N}_N^\partial(i; s)|$ denote the average s-neighborhood size of G_N. For the LLN to hold, the number of neighbors at distance s should not grow too fast as s increases (Kojevnikov et al., 2021). In our context, this condition is formalized by the quantity $\Xi_{N,d} = N^{-1}\sum_{s=1}^\infty s^d \delta_N^\partial(s)(\kappa_1 + \kappa_2)^{\lfloor s/3\rfloor}$, where $d \in \{0, 1\}$, κ_1 and κ_2 are defined in Assumption 2, and $\lfloor\cdot\rfloor$ denotes the floor function. Clearly, $\Xi_{N,d}$ quantifies the trade-off between the decay rate of network dependence (governed by $(\kappa_1 + \kappa_2)^{\lfloor s/3\rfloor}$) and the network denseness (captured by $\delta_N^\partial(s)$). For arbitrary network structure and $d \in \{0, 1\}$, we have $\Xi_{N,d} \leq \sum_{s=1}^\infty s^d(\kappa_1 + \kappa_2)^{\lfloor s/3\rfloor} < \infty$ under Assumption 2. Moreover, $\Xi_{N,d} \to 0$ as $N \to \infty$ if the average s-neighborhood size grows sufficiently slowly with s. For instance, for spatial networks[2] such as latent space (Hoff et al., 2002) and RGG models (Penrose, 2003), we have $\sup_N \sup_{i\in D_N} |\mathcal{N}_N^{\partial}(i;s)| = Cs^k$ for some $C > 0$ and $k \geq 1$ (Leung, 2022). This implies that $\delta_N^{\partial}(s) \lesssim s^k$, and thus $\Xi_{N,d} = O(N^{-1})$.

Let $\zeta_{0,J} = \sup_{y\in\mathcal{R}_\mathcal{Y}} \|p^J(y)\|$ and $\zeta_{1,J} = \sup_{y,y'\in\mathcal{R}_\mathcal{Y}} \|p^J(y) - p^J(y')\|/|y-y'|$.

Assumption 5. *As $N \to \infty$, T is either fixed or tends to ∞, and $\underline{\nu}_{J_1J_2}^{-4} r_{NT,J} J \to 0$, where $r_{NT,J} = \zeta_{0,J}^2 J/(NT) + \zeta_{0,J}^2 \zeta_{1,J} J^{1/2} \min\{\Xi_{N,0}, T^{-1}\Xi_{N,1}\}$.*

Assumption 5 is mainly used to control the convergence in probability of the matrix $(NT)^{-1}\sum_{t=1}^T \mathbb{X}_{t-1}^\top \mathbb{X}_{t-1}$ to its expectation in the Euclidean norm. The term $\min\{\Xi_{N,0}, T^{-1}\Xi_{N,1}\}$ accounts for the dominant regime of dependence decay: $\Xi_{N,0}$ dominates when T is fixed, while $T^{-1}\Xi_{N,1}$ becomes relevant as $T \to \infty$. Notably, when T is fixed, the assumption reduces to (**a**) $\zeta_{0,J}^2 J^2/(\underline{\nu}_{J_1J_2}^4 N) \to 0$ and (**b**) $\zeta_{0,J}^2 \zeta_{1,J} J^{3/2} \Xi_{N,0}/\underline{\nu}_{J_1J_2}^4 \to 0$. For condition (**b**) to hold, the component $\Xi_{N,0}$ must shrink to zero as N goes to infinity, implying the necessity of "sparse" networks to build limit theorems under a fixed T setting. Furthermore, the "sparser" the networks (the faster $\Xi_{N,0}$ converges to zero), the larger J is allowed, which enables fitting f_1 and f_2 with less smoothness. Finally, we can find primitive conditions on the order of J for commonly used basis functions. For example, consider spatial networks where $\Xi_{N,1} = O(N^{-1})$. Then, if the degree of collinearity is not severe so that $\underline{\nu}_{J_1J_2}^{-1} = O(J^r)$ for some $r > 0$, Assumption 5 reduces to $J = o((NT)^{1/(4r+4)})$ when polynomial or trigonometric splines are used, and $J = o((NT)^{2/(8r+11)})$ when power series or orthogonal polynomial bases are used.[3]

To state the next assumption, we introduce a weighted sup-norm defined as

$$|f|_{\infty,\omega} = \sup_{y\in\mathcal{R}_\mathcal{Y}} \left[|f(y)|(1+|y|^2)^{-\omega/2}\right],$$

for some $\omega \geq 0$. As noted by Chen et al. (2005), one may regard the weight function $(1+|y|^2)^{-\omega/2}$ as an alternative to the trimming function used in kernel estimation when the

[2]Spatial networks refer to models in which nodes are embedded in a geometrically meaningful space (e.g., Euclidean space) and the probability of forming a link between two nodes is governed by their spatial proximity.

[3]Here we use the fact that if $\mathcal{R}_\mathcal{Y}$ is compact, $\zeta_{0,J} \lesssim \sqrt{J}$ and $\zeta_{1,J} \lesssim J^{3/2}$ for univariate polynomial or trigonometric splines; and $\zeta_{0,J} \lesssim J$ and $\zeta_{1,J} \lesssim J^2$ for power series or orthogonal polynomial bases; see, e.g, Newey (1997) and Huang (1998).

support is unbounded. When $\omega = 0$, this leads to the standard sup-norm.

Assumption 6. *There exist sequences of vectors $\beta_1 \in \mathbb{R}^{J_1}$, $\beta_2 \in \mathbb{R}^{J_2}$, and a constant $\mu > 0$ such that $|f_\ell(\cdot) - \beta_\ell^\top p^{J_\ell}(\cdot)|_{\infty,\omega} = O(J_\ell^{-\mu})$ for $\ell = 1, 2$.*

Assumption 6 quantifies the approximation error of f_ℓ by the sieve space in terms of the weighted sup-norm. The number μ relies on the smoothness of f_ℓ and the choice of sieve basis. For instance, if f_1 and f_2 belong to a weighted Hölder ball with some finite radius and smoothness μ, this assumption is satisfied by commonly used basis functions such as splines, wavelets, and polynomials.[4]

Denote $\Psi_{N,Z} = \mathbb{Z}^\top \mathbb{Z}/N$, $\Psi_{N,J} = \mathbb{E}(\mathbb{Q}_t^\top \mathbb{Q}_t/N)$ and $C_{N,ZJ} = \mathbb{E}(\mathbb{Z}^\top \mathbb{Q}_t/N)$. The following assumption guarantees the $\sqrt{NT}$-consistency of the sieve estimator $\widehat{\gamma}$.

Assumption 7. *The matrix $\Sigma_{N,Z} = \Psi_{N,Z} - C_{N,ZJ}\Psi_{N,J}^{-1}C_{N,ZJ}^\top$ is asymptotically positive definite.*

Theorem 1. *Suppose that Assumptions 1-7 hold. Then,*

(i) $\|\widehat{\gamma} - \gamma\| = O_p((NT)^{-1/2} + J_1^{-\mu} + J_2^{-\mu})$;

(ii) If $(NT)^{1/2}(J_1^{-\mu} + J_2^{-\mu}) = o(1)$, $\sqrt{NT}\Sigma_{N,Z}^{1/2}(\widehat{\gamma} - \gamma)/\sigma \xrightarrow{d} N(0, I_{d_Z})$.

Theorem 1(i) implies that if the number of basis terms is sufficiently large so that $J_1^{-\mu} \asymp (NT)^{-1/2}$ and $J_2^{-\mu} \asymp (NT)^{-1/2}$, the estimator $\widehat{\gamma}$ becomes $\sqrt{NT}$-consistent. To derive the limiting distribution of $\widehat{\gamma}$, further undersmoothing is required so that the sieve approximation error does not affect the limiting distribution; that is, J_1 and J_2 increase to infinity at a rate faster than $(NT)^{1/(2\mu)}$. In conjunction with Assumption 5, when $J_1 \asymp J_2 \asymp (NT)^\kappa$ for some constant $0 < \kappa < \infty$, we need to choose κ satisfying $\kappa > 1/(2\mu)$ and $[(NT)^{4\kappa}\min\{\Xi_{N,0}, T^{-1}\Xi_{N,1}\} + (NT)^{3\kappa-1}]/\underline{\nu}_{J_1J_2}^4 = o(1)$ simultaneously, for polynomial or trigonometric splines. In particular, if $\min\{\Xi_{N,0}, T^{-1}\Xi_{N,1}\} = O((NT)^{-1})$ and $\underline{\nu}_{J_1J_2}^{-1} = O(J^r)$ for some $r > 0$, these conditions can be reduced to $1/(2\mu) < \kappa < 1/(4r+4)$. Note that this automatically implies that μ must be larger than $2r + 2$.

[4]Let $\Lambda^\mu(\mathcal{R}_\mathcal{Y})$ denote a set of functions: $f : \mathcal{R}_\mathcal{Y} \to \mathbb{R}$ which is $\lfloor \mu \rfloor$-times continuously differentiable, and the $\lfloor \mu \rfloor$th derivative is Hölder continuous with exponent $\mu - \lfloor \mu \rfloor$. Further, let $\Lambda^\mu(\mathcal{R}_\mathcal{Y}, \omega_1) = \{f(\cdot) : f(\cdot)[1 + |\cdot|^2]^{-\omega_1/2} \in \Lambda^\mu(\mathcal{R}_\mathcal{Y})\}$. The set of functions $f \in \Lambda^\mu(\mathcal{R}_\mathcal{Y}, \omega_1)$ such that the Hölder norm of $f(\cdot)[1 + |\cdot|^2]^{-\omega_1/2}$ is bounded by c is called a weighted Hölder ball with radius c and smoothness μ, and is denoted as $\Lambda_c^\mu(\mathcal{R}_\mathcal{Y}, \omega_1)$. Then, if $f_\ell \in \Lambda_c^\mu(\mathcal{R}_\mathcal{Y}, \omega_1)$, Assumption 6 is met for $\omega > \mu + \omega_1$. For more details, see Ai and Chen (2003) and Chen et al. (2005).

Now, we examine the asymptotic properties of $\widehat{f}_1(\cdot)$ and $\widehat{f}_2(\cdot)$. Let

$$\Sigma_{N,f} = \Psi_{N,J} - C_{N,ZJ}^{\top} \Psi_{N,Z}^{-1} C_{N,ZJ}. \tag{5}$$

Also, denote $\Sigma_{N,f_1} = (\mathcal{S}_1 \Sigma_{N,f}^{-1} \mathcal{S}_1^{\top})^{-1}$ and $\Sigma_{N,f_2} = (\mathcal{S}_2 \Sigma_{N,f}^{-1} \mathcal{S}_2^{\top})^{-1}$, where $\mathcal{S}_1 = (I_{J_1}, \mathbf{0}_{J_1 \times J_2})$ and $\mathcal{S}_2 = (\mathbf{0}_{J_2 \times J_1}, I_{J_2})$ are two selection matrices.

Theorem 2. *Suppose that Assumptions 1-6 hold. Let $\ell = 1, 2$.*

(i) If $\max_{i \geq 1} \lambda_{max}(\mathbb{E}[p^{J_\ell}(Y_{it}) p^{J_\ell}(Y_{it})^{\top}]) < \infty$, then $\max_{i \geq 1} \int_{\mathcal{R}_\mathcal{Y}} [\widehat{f}_\ell(y) - f_\ell(y)]^2 dF_i(y) = O_p(\underline{\nu}_{J_1 J_2}^{-1}(J_\ell/(NT) + J_1^{-2\mu} + J_2^{-2\mu}))$, where $F_i(y)$ is the cumulative distribution function of Y_{it}.

(ii) If $\zeta_{1,J} \lesssim J^{\xi}$ for some finite constant $\xi > 0$ and $\zeta_{0,J}^4 \ln J_\ell / (\underline{\nu}_{J_1 J_2}^2 NT) = o(1)$, then $\left| \widehat{f}_\ell(\cdot) - f_\ell(\cdot) \right|_{\infty,\omega} = O_p\big(\zeta_{0,J_\ell} \underline{\nu}_{J_1 J_2}^{-1/2} \big(\sqrt{\ln J_\ell/(NT)} + J_1^{-\mu} + J_2^{-\mu}\big)\big)$.

(iii) If $\sqrt{NT} \underline{\nu}_{J_1 J_2}^{-1/2}(J_1^{-\mu} + J_2^{-\mu}) = o(1)$, then for a given finite $y \in \mathcal{R}_\mathcal{Y}$ such that $\|p^{J_\ell}(y)\| > 0$, we have $\frac{\sqrt{NT}}{v_{\ell N}(y)} \big(\widehat{f}_\ell(y) - f_\ell(y)\big) \xrightarrow{d} N(0,1)$ with $v_{\ell N}^2(y) = \sigma^2 p^{J_\ell}(y)^{\top} \Sigma_{N,f_\ell}^{-1} p^{J_\ell}(y)$.

Results (i) and (ii) of Theorem 2 provide the convergence rate of the sieve LS estimator under the L^2 and sup-norm, respectively. Suppose that $J_1 \asymp J_2 \asymp J$ and $\nu_J^{-1} \equiv \nu_{J_1 J_2}^{-1} \lesssim 1$. Regarding Theorem 2(i), the optimal choice of J would balance the standard deviation part $O_p(\sqrt{J/(NT)})$ and the bias part $O_p(J^{-\mu})$ by choosing $J \asymp (NT)^{1/(1+2\mu)}$, which yields $\max_{i \geq 1} \mathbb{E}|\widehat{f}_\ell(Y_{it}) - f_\ell(Y_{it})|^2 = O_p((NT)^{-2\mu/(1+2\mu)})$. This coincides with the optimal convergence rate of Stone (1982) for nonparametric LS regression under independent data. Regarding Theorem 2(ii), the uniform convergence rate will not be optimal. Chen and Christensen (2015) derived optimal sup-norm convergence rates for nonparametric regression without cross-sectional dependence; that is, in our context, $|\widehat{f}_\ell(\cdot) - f_\ell(\cdot)|_{\infty,0} = O_p(\zeta_{0,J}\sqrt{\ln J/(\underline{\nu}_J NT)}) + O_p(J^{-\mu})$. While the variance term of our estimator attains the optimal rate, the bias term appears to be $O(\zeta_{0,J} \underline{\nu}_J^{-1/2})$ times bigger. Theorem 2(iii) states a pointwise asymptotic normality property of the sieve LS estimator. Notice that $[\lambda_{\max}(\Sigma_{N,f})]^{-1} \|p^{J_\ell}(y)\|^2 \leq p^{J_\ell}(y)^{\top} \mathcal{S}_\ell \Sigma_{N,f}^{-1} \mathcal{S}_\ell^{\top} p^{J_\ell}(y) \leq [\lambda_{\min}(\Sigma_{N,f})]^{-1} \|p^{J_\ell}(y)\|^2$, which together with Assumption 4 implies that the convergence rate of $\widehat{f}_\ell(y)$ is $O((NT/(J_\ell \underline{\nu}_{J_1 J_2}))^{1/2})$ as $\|p^{J_\ell}(y)\| = O(J_\ell^{1/2})$ and $\lambda_{\min}^{-1}(\Sigma_{N,f}) = O(\underline{\nu}_{J_1 J_2}^{-1})$.

Remark 2. *For inferences based on Theorems 1 and 2, it is necessary to consistently estimate* $\sigma^2\Sigma_{N,Z}^{-1}$*,* $v_{1N}^2(y)$*, and* $v_{2N}^2(y)$*. The matrices* $\Sigma_{N,Z}$*,* Σ_{N,f_1} *and* Σ_{N,f_2} *can be easily estimated by their sample analogues. For example, an estimator for* Σ_{N,f_1}^{-1} *is given by* $\widehat{\Sigma}_{NT,f_1}^{-1} = \mathcal{S}_1\widehat{\Sigma}_{NT,f}^{-1}\mathcal{S}_1^\top$*, where* $\widehat{\Sigma}_{NT,f} = \widehat{\Psi}_{NT,J} - \widehat{C}_{NT,ZJ}^\top\Psi_{N,Z}^{-1}\widehat{C}_{NT,ZJ}$ *with* $\widehat{\Psi}_{NT,J} = \sum_{t=1}^T \mathbb{Q}_{t-1}^\top\mathbb{Q}_{t-1}/(NT)$ *and* $\widehat{C}_{NT,ZJ} = \sum_{t=1}^T \mathbb{Z}^\top\mathbb{Q}_{t-1}/(NT)$*. The convergence in probability of* $\widehat{\Sigma}_{NT,f}^{-1}$ *to* $\Sigma_{N,f}^{-1}$ *in the Euclidean norm is shown in Lemma 3 in the Online Supplementary Appendix. For* σ^2*, a straightforward estimator is* $\widehat{\sigma}^2 = (NT)^{-1}\sum_{i=1}^N\sum_{t=1}^T \widehat{\epsilon}_{it}^2$ *with* $\widehat{\epsilon}_{it} = Y_{it} - \sum_{j=1}^N w_{ij}\widehat{f}_1(Y_{j,t-1}) - \widehat{f}_2(Y_{i,t-1}) - Z_i^\top\widehat{\gamma}$*. We show in Appendix C.1 that* $\widehat{\sigma}^2$ *is a consistent estimator of* σ^2 *under Assumptions 2-6. As such, Theorem 2(iii) still holds when* $v_{1N}^2(y)$ *is replaced by* $\widehat{v}_{1N}^2(y) = \widehat{\sigma}^2 p^{J_1}(y)^\top\widehat{\Sigma}_{NT,f_1}^{-}p^{J_1}(y)$*.*

Remark 3. *The above results can be extended to the estimation of various functionals of* $f_\ell(\cdot)$*. Let* $\phi(f_\ell)$ *denote the functional of interest, where* ϕ *is a function from* $\mathcal{F}$ *to* $\mathbb{R}$*. We focus on two cases of* ϕ *: (i)* $\phi(f_\ell) = \partial f_\ell(y)/\partial y$ *for some* $y \in \mathcal{R}_\mathcal{Y}$*, and (ii)* $\phi(f_\ell) = \int_{\mathcal{R}_\mathcal{Y}} \partial f_\ell(y)/\partial y w(y)dy$ *for some weight function* w*. One could allow the weight function in (ii) to depend on* (N,T)*. For each case, we can estimate* $\phi(f_\ell)$ *by* $\phi(\widehat{f}_\ell)$*. As the functional* ϕ *is linear here,* $\phi(\widehat{f}_\ell) = \phi(p^{J_\ell}(\cdot)^\top\widehat{\beta}_\ell) = \varphi^{J_\ell\top}\widehat{\beta}_\ell$*, where* $\varphi^{J_\ell} = (\varphi_1,\ldots,\varphi_{J_\ell})^\top \in \mathbb{R}^{J_\ell}$ *and* $\varphi_j = \phi(p_j(\cdot)) \in \mathbb{R}$ *for* $j = 1,\ldots,J_\ell$*. For cases (i) and (ii), the* j*th element of* φ^{J_ℓ} *is* $\partial p_j(y)/\partial y$ *and* $\int_{\mathcal{R}_\mathcal{Y}} \partial p_j(y)/\partial y w(y)dy$*, respectively.*

Assume that ***(A)*** *there exists some norm* $|\cdot|_s$ *such that* $|\phi(f_\ell)| \leq C|f_\ell|_s$ *for some* $C < \infty$*, and* ***(B)*** $\sqrt{NT}|f_\ell(\cdot) - p^{J_\ell}(\cdot)^\top\beta_\ell|_s = o(1)$ *as* $N \to \infty$*. If* $\mathcal{R}_\mathcal{Y}$ *is compact, it is reasonable to define* $|f_\ell|_s$ *as the Sobolev norm of derivative order 1, i.e.,* $|f_\ell|_s = \sup_{y\in\mathcal{R}_\mathcal{Y}}|f_\ell(y)| + \sup_{y\in\mathcal{R}_\mathcal{Y}}|\partial f_\ell(y)/\partial y|$*, and condition* ***(A)*** *is trivially satisfied for both cases we investigate. Assumption* ***(B)*** *plays the role of undersmoothing. Under assumptions* ***(A)****,* ***(B)*** *and the conditions of Theorem 2(iii),* $\sqrt{NT}\tilde{v}_{\ell N}^{-1}[\phi(\widehat{f}_\ell) - \phi(f_\ell)] \xrightarrow{d} N(0,1)$*, where* $\tilde{v}_{\ell N}^2 = \sigma^2\varphi^{J_\ell\top}\Sigma_{N,f_\ell}^{-1}\varphi^{J_\ell}$*. A consistent estimator of* $\tilde{v}_{\ell N}$ *can be constructed by replacing* Σ_{N,f_ℓ} *by* $\widehat{\Sigma}_{NT,f_\ell}$ *given in Remark 2.*

Remark 4. *The large sample properties of the sieve estimator can also be established under the setting where* $T \to \infty$ *and* N *is fixed. The derivation of the asymptotic theory in this case can be easier than the large* N *case because the degree of the network dependence is* *always at a controllable level with fixed N. To establish the related limit theorems, we only need to explore the NED properties of $\{\mathbb{Y}_t\}$ along the time dimension. Specifically, we show in the Section C.2 of the Online Supplementary Appendix that under Assumptions 1-7 (with 4, 5, and 7 replaced by $4^\ddagger$, $5^\ddagger$, and $7^\ddagger$ therein, respectively), Theorems 1 and 2 still hold when $T \to \infty$ and N is fixed. It is worth noting that the requirement of the growth order of (J_1, J_2) in this case reduces to $\underline{\nu}_{J_1 J_2}^{-4} \zeta_{0,J}^2 J^{3/2}(J^{1/2} + \zeta_{1,J})/T \to 0$, which is irrelevant with the network structure.*

Remark 5. *(General NAR(p) model) The methodology above can be extended to an NAR model which includes the past p time lags, that is,*

$$Y_{it} = \sum_{j=1}^{N} w_{ij} f_1(Y_{j,t-1}, Y_{j,t-2}, \dots, Y_{j,t-p}) + f_2(Y_{i,t-1}, Y_{i,t-2}, \dots, Y_{i,t-p}) + Z_i^\top \gamma + \epsilon_{it}, \tag{6}$$

where f_1 and f_2 are unknown nonparametric functions on $\mathbb{R}^p$. We can show that there exists a unique stationary solution to the NAR model (6) as long as Assumption 1 holds and there exist non-negative constants $\{a_{1,k}\}_{k=1}^p$ and $\{a_{2,k}\}_{k=1}^p$ such that for any $\boldsymbol{y}, \boldsymbol{y}' \in \mathcal{R}_{\mathcal{Y}}^p$, $\left|f_\ell(\boldsymbol{y}) - f_\ell(\boldsymbol{y}')\right| \le \sum_{k=1}^p a_{\ell,k}|y_k - y_k'|$, and $\sum_{k=1}^p (a_{1,k} + a_{2,k}) < 1$. To estimate f_1 and f_2, we can approximate them in sieve space as $f_\ell(y) \approx \sum_{j=1}^{J_\ell} \beta_{\ell j} p_j(y)$ for $\ell = 1, 2$, where $\{p_j(\cdot) : j = 1, 2, \dots\}$ is a sequence of basis functions on $\mathbb{R}^p$, and then construct a LS estimator for $\beta_{\ell j}$'s in a similar way as Section 2. The asymptotic theory of the corresponding estimator can be similarly derived as in the NAR(1) case, see Section C.3 of the Online Supplementary Appendix for more details.

4 Specification tests

In this section, we consider testing the functional forms of the network and momentum effects given by f_1 and f_2, respectively, which will determine whether the more complicated nonparametric NAR model is indeed necessary. In particular, we are interested in testing whether the widely used linear specification is appropriate for f_1, f_2, or both.

4.1 Testing the linearity of f_1 or f_2

Our first aim is to test the following two null hypotheses separately:

$$\begin{aligned} H_0^n &: f_1(Y_{i,t-1}) = \rho_1 Y_{i,t-1}, \ \text{ a.s. for some } \rho_1 \in \mathcal{B} \subset \mathbb{R}, \\ H_0^m &: f_2(Y_{i,t-1}) = \rho_2 Y_{i,t-1}, \ \text{ a.s. for some } \rho_2 \in \mathcal{B} \subset \mathbb{R}, \end{aligned} \tag{7}$$

where $i = 1, \ldots, N$, $t = 1, \ldots, T$, and $\mathcal{B}$ is a compact subset of $\mathbb{R}$. The alternative hypotheses H_1^n and H_1^m are the negation of H_0^n and H_0^m, respectively. Since both f_1 and f_2 pass the origin, we do not include an intercept term in the expressions of f_1 and f_2 in (7). In the following, we focus on the test of H_0^n, and the test of H_0^m can be derived similarly.

A natural idea to test H_0^n is to calculate the distance between the restricted and unrestricted LS estimates of f_1 in the spirit of Härdle and Mammen (1993). Note that the null model can be written as

$$Y_{it} = \rho_1 \sum_{j=1}^{N} w_{ij} Y_{j,t-1} + f_2(Y_{i,t-1}) + Z_i^\top \gamma + \epsilon_{it}. \tag{8}$$

Let $\widehat{\rho}_1$ be a $\sqrt{NT}$-consistent estimator of ρ_1, e.g., the restricted sieve LS estimator

$$\widehat{\rho}_{1,LS} = e_1^\top \left(\sum_{t=1}^{T} \mathbb{X}_{t-1,1}^\top \mathbb{X}_{t-1,1} \right)^{-1} \sum_{t=1}^{T} \mathbb{X}_{t-1,1}^\top \mathbb{Y}_t, \tag{9}$$

where $\mathbb{X}_{t-1,1} = (W\mathbb{Y}_{t-1}, \mathbb{P}_{t-1}^{J_2}, \mathbb{Z})$, and e_1 is a vector of length $(J_2 + d_Z + 1)$ with first element being one and others being zero. If H_0^n is true, both the restricted and unrestricted estimators of f_1 are consistent, and they would yield close estimates; otherwise, at least for some of the Y_{it}'s, the estimates should be different. This motivates us to consider the test statistic

$$\mathbf{T}_{NT,1} = \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\widehat{f}_1(Y_{i,t-1}) - \widehat{\rho}_1 Y_{i,t-1} \right)^2. \tag{10}$$

To test the linearity of nonparametric functions, such a type of test statistic has been considered in various contexts in the literature, see, e.g., Su and Lu (2013) and Hoshino (2022). We will show that after being properly centered and scaled, $\mathbf{T}_{NT,1}$ in (10) is asymptotically normal under H_0^n.

The test statistic $\mathbf{T}_{NT,1}$ requires estimation of f_1 under the null and alternative. Since both models are semiparametric, the computation of $\mathbf{T}_{NT,1}$ could be cumbersome. Therefore, we also propose an LM test for H_0^n that utilizes only the estimates under the null. To this end, we first nest the null models in the alternative by taking $p_1(y) = y$. Then the variable matrix can be split as $W\mathbb{P}_{t-1}^{J_1} = (W\mathbb{Y}_{t-1}, W\tilde{\mathbb{P}}_{t-1}^{J_1})$ and the sieve coefficients as $\beta_1 = (\beta_{1,1}, \beta_{1,2}^\top)^\top$, where $\tilde{\mathbb{P}}_{t-1}^{J_1}$ is an $N \times (J_1 - 1)$ matrix of high-order series terms and $\beta_{1,2}$ is the corresponding sieve coefficients of length $J_1 - 1$. The alternative model (2) could thus be written as $\mathbb{Y}_t = \beta_{1,1}W\mathbb{Y}_{t-1} + W\tilde{\mathbb{P}}_{t-1}^{J_1}\beta_{1,2} + \mathbb{P}_{t-1}^{J_2}\beta_2 + \mathbb{Z}\gamma + U_t$. If H_0^n is correct, the series terms $W\tilde{\mathbb{P}}_{t-1}^{J_1}$ should not enter the model. This implies that we can test H_0^n by way of the approximate null $H_{0,\text{app}}^n : \beta_{1,2} = 0$, or equivalently, by testing the joint significance of high-order sieve terms in the network component. Denote $\mathbb{X}_{t-1,1} = (W\mathbb{Y}_{t-1}, \mathbb{P}_{t-1}^{J_2}, \mathbb{Z})$, $\mathbb{X}_{t-1,2} = W\tilde{\mathbb{P}}_{t-1}^{J_1}$, and $\theta_1 = (\beta_{1,1}, \beta_2^\top, \gamma^\top)^\top$. The LM test measures how close to zero the gradient of the sieve LS criterion is when evaluated at the constrained estimator, which is given by $\bar{\theta}_1 = (\sum_{t=1}^T \mathbb{X}_{t-1,1}^\top \mathbb{X}_{t-1,1})^{-1} \sum_{t=1}^T \mathbb{X}_{t-1,1}^\top \mathbb{Y}_t$ in our context. The test statistic is

$$\mathbf{LM}_{NT,1} = (\mathbf{Y} - \mathbf{X}_1\bar{\theta}_1)^\top \mathbf{X}_2 \big(\mathbf{X}_2^\top \mathbf{M}_{\mathbf{X}_1} \mathbf{X}_2\big)^{-1} \mathbf{X}_2^\top (\mathbf{Y} - \mathbf{X}_1\bar{\theta}_1)/\widehat{\sigma}^2, \tag{11}$$

where $\mathbf{X}_1 = (\mathbb{X}_{0,1}^\top, \ldots, \mathbb{X}_{T-1,1}^\top)^\top$, $\mathbf{X}_2 = (\mathbb{X}_{0,2}^\top, \ldots, \mathbb{X}_{T-1,2}^\top)^\top$, and $\widehat{\sigma}^2$ is a consistent estimator for σ^2. The number of restrictions in the approximate null is $J_1 - 1$, where $J_1 \to \infty$ as $N \to \infty$. For a fixed J_1, the statistic $\mathbf{LM}_{NT,1}$ has an asymptotic $\chi^2_{J_1-1}$ distribution. Below, we show that the standardized LM statistic converges to the standard normal distribution $N(0,1)$ under H_0^n.

To formally establish the desired result, we modify the assumptions mentioned above and introduce some additional assumptions as follows.

Assumption 1′. *Assumption 1 holds. In addition,* $\mathbb{E}(\epsilon_{it}^6) < \infty$.

Assumption 4′. *Assumption 4 holds. In addition, (i)* $\max_{i\geq 1}\max_{1\leq j\leq J}\mathbb{E}|p_j(Y_{it})|^6 < \infty$*; and (ii) there exists a constant* $\underline{c}_p > 0$ *such that* $\lambda_{min}(\mathbb{E}(\mathbb{P}_t^{J_1\top}\mathbb{P}_t^{J_1}/N)) > \underline{c}_p$ *and* $\|\mathbb{E}(\mathbb{Y}_t^\top \mathbb{P}_t^{J_1}/N)\| = O(J_1^{1/4})$.

Assumption 5′. *Assumption 5 holds. In addition, (i)* $J^{3/2} r_{NT,J}/\underline{\nu}_{J_1 J_2}^5 = o(1)$*, (ii)*

$\big(J^{3/2} r_{NT,J}^{1/4} + J^{1/2}\zeta_{0,J}^{5/3} r_{NT,J}^{1/3}\big)/\underline{\nu}_{J_1J_2}^4 = o(J_1)$, *(iii)* $\sqrt{NT}(\zeta_{0,J_1} J_1^{-\mu} + J_2^{-\mu})/\underline{\nu}_{J_1J_2} = o(1)$, *and (iv)* $\underline{\nu}_{J_1J_2} J_1^{1/2} \to \infty$ *as* $J_1 \to \infty$.

Assumption 8. *Under* H_0^n, $\widehat{\rho}_1$ *is a* $\sqrt{NT}$*-consistent estimator of* ρ_1.

Since both test statistics have a quadratic form, it is necessary to assume the existence of higher moments of the errors as in Assumption 1′. Assumption 4′ is only required for deriving the limiting distribution of $\mathbf{T}_{NT,1}$. Assumption 5′(i)-(iii) are slightly stronger than those required for Theorem 2(ii). Assumption 5′(iv) requires that the degree of collinearity is not severe. Assumption 8 can be easily satisfied by sieve LS estimator of ρ_1 given in (9) provided that f_2 is smooth enough and the growth order of J_2 is chosen properly.

Denote $\mathbf{B}_{N,1} = \sigma^2 \mathrm{tr}\{\Phi_{N,J_1}\Sigma_{N,f_1}^{-1}\}$ and $\mathbf{s}_{N,1}^2 = 2\sigma^4 \mathrm{tr}\big\{(\Phi_{N,J_1}\Sigma_{N,f_1}^{-1})^2\big\}$, where $\Phi_{N,J_1} = \mathbb{E}(\mathbb{P}_t^{J_1\top}\mathbb{P}_t^{J_1}/N)$. We consider the standardized test statistics $\overline{\mathbf{T}}_{NT,1} = (\mathbf{T}_{NT,1} - \mathbf{B}_{N,1})/\mathbf{s}_{N,1}$ and $\overline{\mathbf{LM}}_{NT,1} = (\mathbf{LM}_{NT,1} - (J_1 - 1))/\sqrt{2(J_1 - 1)}$.

Theorem 3. *Suppose that Assumptions 1-6 and 8 hold (where 1, 4 and 5 are replaced by 1′, 4′ and 5′, respectively). Under* H_0^n, $\overline{\mathbf{T}}_{NT,1} \xrightarrow{d} N(0,1)$ *and* $\overline{\mathbf{LM}}_{NT,1} \xrightarrow{d} N(0,1)$.

If H_0^n is not true, the test statistics $\overline{\mathbf{T}}_{NT,1}$ and $\overline{\mathbf{LM}}_{NT,1}$ tend to deviate to a positive value. Theorem 3 implies that we can implement a one-sided test by comparing the value of $\overline{\mathbf{T}}_{NT,1}$ or $\overline{\mathbf{LM}}_{NT,1}$ with z_α, the upper α-percentile of $N(0,1)$. To implement the test, we need to consistently estimate $\mathbf{B}_{N,1}$ and $\mathbf{s}_{N,1}$, which can be carried out by the sample analogs: $\widehat{\mathbf{B}}_{NT,1} = \widehat{\sigma}^2 \mathrm{tr}\{\widehat{\Phi}_{NT,J_1}\widehat{\Sigma}_{NT,f_1}^{-1}\}$ and $\widehat{\mathbf{s}}_{NT,1}^2 = 2\widehat{\sigma}^4 \mathrm{tr}\{(\widehat{\Phi}_{NT,J_1}\widehat{\Sigma}_{NT,f_1}^{-1})^2\}$, where $\widehat{\Phi}_{NT,J_1} = \sum_{t=1}^T \mathbb{P}_{t-1}^{J_1\top}\mathbb{P}_{t-1}^{J_1}/(NT)$ and $\widehat{\Sigma}_{NT,f_1}$ is given in Remark 2. In view of Lemmas 2 and 3 in Appendix A, it is straightforward to see that $\widehat{\mathbf{B}}_{NT,1} - \mathbf{B}_{N,1} = o_p(1)$ and $\widehat{\mathbf{s}}_{NT,1}^2/\mathbf{s}_{N,1}^2 \xrightarrow{p} 1$ under Assumptions 2-4 and 5′.

Next, we study the local power properties of $\overline{\mathbf{T}}_{NT,1}$ and $\overline{\mathbf{LM}}_{NT,1}$. Consider the following sequence of Pitman-type local alternatives

$$H_1^n(\alpha_{NT}) : f_1(Y_{i,t-1}) = \rho_1 Y_{i,t-1} + \alpha_{NT} r(Y_{i,t-1}) \;\text{ a.s. for some } \rho_1 \in \mathcal{B} \subset \mathbb{R}, \tag{12}$$

where $\alpha_{NT} \to 0$ as $N \to \infty$, and $r(\cdot)$ is a measurable nonlinear function on $\mathcal{R}_{\mathcal{Y}}$ assumed to satisfy the following conditions.

Assumption 9. *The function $r(\cdot)$ is Lipschitz continuous on $\mathcal{R}_{\mathcal{Y}}$ with Lipschitz constant $\bar{c}_r$, and $\max_{i\geq 1}\mathbb{E}|r(Y_{it})|^4 < \infty$.*

To study the local power of $\overline{\mathbf{T}}_{NT,1}$, we need an asymptotic representation of the restricted estimator $\widehat{\rho}_1$ in (10). In the following, we consider $\widehat{\rho}_1$ as the sieve LS estimator given in (9). Denote $\widetilde{\mathbb{X}}_t = (\mathbb{X}_{t,1}, \mathbb{X}_{t,2})$, $\Psi_N = \mathbb{E}(\widetilde{\mathbb{X}}_t^\top \widetilde{\mathbb{X}}_t/N)$, $\Gamma_N = \mathbb{E}(\widetilde{\mathbb{X}}_t^\top W R(\mathbb{Y}_t)/N)$, and $R(\mathbb{Y}_t) = (r(Y_{1t}), \ldots, r(Y_{Nt}))^\top$. The "non-centrality parameters" of the $\overline{\mathbf{T}}_{NT,1}$ and $\overline{\mathbf{LM}}_{NT,1}$ tests under the local alternatives (12) take the form

$$\Delta_{N,T} = \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\left(r(Y_{i,t-1}) - e_1^\top \Psi_{N,11}^{-1}\Gamma_{N,1}Y_{i,t-1}\right)^2\right],$$

and

$$\Delta_{N,LM} = \Gamma_N^\top \Psi_N^{-1}\mathcal{S}_3^\top\left(\mathcal{S}_3\Psi_N^{-1}\mathcal{S}_3^\top\right)^{-1}\mathcal{S}_3\Psi_N^{-1}\Gamma_N/(\sqrt{2}\sigma^2),$$

respectively, where $\Psi_{N,11} = \mathbb{E}(\mathbb{X}_{t,1}^\top \mathbb{X}_{t,1}/N)$, $\Gamma_{N,1} = \mathbb{E}(\mathbb{X}_{t,1}^\top W R(\mathbb{Y}_t)/N)$, and $\mathcal{S}_3 = (\mathbf{0}_{(J_1-1)\times(J_2+d_Z+1)}, I_{J_1-1})$ is a selection matrix.

Theorem 4. *Suppose that Assumptions 1-6 and 9 hold (where 1, 4 and 5 are replaced by 1′, 4′ and 5′, respectively). Also, $\Delta_T = \lim_{N\to\infty}\Delta_{N,T}$ and $\Delta_{LM} = \lim_{N\to\infty}\Delta_{N,LM}$ exist and are finite.*

(i) Under $H_1^n(\alpha_{NT})$ with $\alpha_{NT} = \mathbf{s}_{N,1}^{1/2}(NT)^{-1/2}$, $\overline{\mathbf{T}}_{NT,1} \xrightarrow{d} N(\Delta_T, 1)$.

(ii) Under $H_1^n(\alpha_{NT})$ with $\alpha_{NT} = J_1^{1/4}(NT)^{-1/2}$, $\overline{\mathbf{LM}}_{NT,1} \xrightarrow{d} N(\Delta_{LM}, 1)$.

Theorem 4 states that $\mathbf{T}_{NT,1}$ and $\mathbf{LM}_{NT,1}$ have nontrivial power to detect local alternatives that converge to the null hypothesis at the rate $O(\mathbf{s}_{N,1}^{1/2}(NT)^{-1/2})$ and $O(J_1^{1/4}(NT)^{-1/2})$, respectively. In view that $\mathbf{s}_{N,1}^2 = O(J_1/\underline{\nu}_{J_1J_2}^2)$ and $J_1 \to \infty$ as $N \to \infty$, both rates are slower than the parametric rate $(NT)^{-1/2}$. Such rates have also been found by de Jong and Bierens (1994) and Gupta (2018).

4.2 Testing the joint linearity of f_1 and f_2

Tests considered in Section 4.1 deal with semiparametric null models versus semiparametric (or nonparametric) alternatives. In some applications, it is also interesting to directly test the validity of the linear NAR model. The corresponding null hypothesis is given by

$$H_0^\dagger : f_1(Y_{i,t-1}) = \rho_1 Y_{i,t-1} \text{ and } f_2(Y_{i,t-1}) = \rho_2 Y_{i,t-1} \text{ a.s. for some } (\rho_1, \rho_2) \in \mathcal{B} \subset \mathbb{R}^2.$$

The alternative hypothesis is the negation of $H_0^\dagger$. A naive test for $H_0^\dagger$ is based on the Bonferroni correction. That is, one considers separately testing the hypotheses H_0^n and H_0^m with the adjusted significance level $\alpha/2$. Then, $H_0^\dagger$ is rejected at significance level α if either H_0^n or H_0^m is rejected. Though easy to implement, such a procedure may suffer from inaccurate size in finite samples. To better control the Type-I error, we consider distance-based and LM test statistics as above. The null model now is of the parametric form:

$$Y_{it} = \rho_1 \sum_{j=1}^{N} w_{ij} Y_{j,t-1} + \rho_2 Y_{i,t-1} + Z_i^\top \gamma + \epsilon_{it}. \tag{13}$$

Let $\widehat{\rho} = (\widehat{\rho}_1, \widehat{\rho}_2)$ be a $\sqrt{NT}$-consistent estimator of $\rho = (\rho_1, \rho_2)$, for example, the OLS estimator in Zhu et al. (2017). The L_2-distance-based test statistic for $H_0^\dagger$ is defined as

$$\mathbf{T}_{NT} = \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\left(\widehat{f}_1(Y_{i,t-1}) - \widehat{\rho}_1 Y_{i,t-1} \right)^2 + \left(\widehat{f}_2(Y_{i,t-1}) - \widehat{\rho}_2 Y_{i,t-1} \right)^2 \right]. \tag{14}$$

We reject $H_0^\dagger$ for the realized value of $\mathbf{T}_{NT}$ appearing in the right tail of its asymptotic null distribution, which is given in Theorem 5.

The LM test statistic can be constructed in a similar way to (11). In this case, we partition the network and momentum variable matrices as $W\mathbb{P}_{t-1}^{J_1} = (W\mathbb{Y}_{t-1}, W\tilde{\mathbb{P}}_{t-1}^{J_1})$ and $\mathbb{P}_{t-1}^{J_2} = (\mathbb{Y}_{t-1}, \tilde{\mathbb{P}}_{t-1}^{J_2})$, and the corresponding sieve coefficients as $\beta_1 = (\beta_{1,1}, \beta_{1,2}^\top)^\top$ and $\beta_2 = (\beta_{2,1}, \beta_{2,2}^\top)^\top$. If $H_0^\dagger$ is correct, the series terms used to capture the nonlinear network and momentum effects, i.e., $W\tilde{\mathbb{P}}_{t-1}^{J_1}$ and $\tilde{\mathbb{P}}_{t-1}^{J_2}$, should not enter the model. Thus we can test $H_0^\dagger$ by testing the approximate null $H_{0,\text{app}}^\dagger : \beta_{1,2} = 0$ and $\beta_{2,2} = 0$. Denote $\mathcal{X}_{t-1,1} = (W\mathbb{Y}_{t-1}, \mathbb{Y}_{t-1}, \mathbb{Z})$, $\mathcal{X}_{t-1,2} = (W\tilde{\mathbb{P}}_{t-1}^{J_1}, \tilde{\mathbb{P}}_{t-1}^{J_2})$, $\vartheta_1 = (\beta_{1,1}, \beta_{2,1}, \gamma^\top)^\top$ and $\bar{\vartheta}_1 = (\sum_{t=1}^T \mathcal{X}_{t-1,1}^\top \mathcal{X}_{t-1,1})^{-1} \sum_{t=1}^T \mathcal{X}_{t-1,1}^\top \mathbb{Y}_t$. The LM test statistic to test $H_0^\dagger$ is given by

$$\mathbf{LM}_{NT} = (\mathbf{Y} - \boldsymbol{\mathcal{X}}_1 \bar{\vartheta}_1)^\top \boldsymbol{\mathcal{X}}_2 \left(\boldsymbol{\mathcal{X}}_2^\top \mathbf{M}_{\boldsymbol{\mathcal{X}}_1} \boldsymbol{\mathcal{X}}_2 \right)^{-} \boldsymbol{\mathcal{X}}_2^\top (\mathbf{Y} - \boldsymbol{\mathcal{X}}_1 \bar{\vartheta}_1)/\widehat{\sigma}^2, \tag{15}$$

where $\boldsymbol{\mathcal{X}}_1 = (\mathcal{X}_{0,1}^\top, \ldots, \mathcal{X}_{T-1,1}^\top)^\top$ and $\boldsymbol{\mathcal{X}}_2 = (\mathcal{X}_{0,2}^\top, \ldots, \mathcal{X}_{T-1,2}^\top)^\top$. Note that the number of restrictions in $H_{0,\text{app}}^\dagger$ is $J_1 + J_2 - 2$. For fixed J_1 and J_2, $\mathbf{LM}_{NT}$ converges to $\chi^2(J_1 + J_2 - 2)$ distribution. In the following, we will show that the standardized LM statistic $\overline{\mathbf{LM}}_{NT} = (\mathbf{LM}_{NT} - (J_1 + J_2 - 2))/\sqrt{2(J_1 + J_2 - 2)}$ is asymptotically $N(0,1)$ distributed under $H_0^\dagger$.

To establish the desired result, we modify the assumptions mentioned above as follows.

Assumption 4†. *Assumptions 4 and 4′(i) hold. In addition, there exists a constant* $0 < \underline{C}_p < \infty$ *such that* $\lambda_{min}\big(\mathbb{E}(\mathbb{P}_t^{J_\ell \top}\mathbb{P}_t^{J_\ell}/N)\big) > \underline{C}_p$ *and* $\|\mathbb{E}(\mathbb{Y}_t^\top \mathbb{P}_t^{J_\ell}/N)\| = O(J_\ell^{1/4})$ *for* $\ell = 1, 2$.

Assumption 5†. *Assumption 5 holds. In addition, (i)* $J(r_{NT,J} + N^{-1})/\underline{\nu}_{J_1J_2}^5 = o(1)$*, (ii)* $J^{-1/2}\zeta_{0,J}^{5/3}(N^{-1/3} + r_{NT,J}^{1/3})/\underline{\nu}_{J_1J_2}^4 + J^{1/2}(N^{-1/4} + r_{NT,J}^{1/4})/\underline{\nu}_{J_1J_2}^4 = o(1)$*, (iii)* $(NT)^{1/2}\zeta_{0,J}(J_1^{-\mu} + J_2^{-\mu})/\underline{\nu}_{J_1J_2} = o(1)$*, and (iv)* $\underline{\nu}_{J_1J_2}J^{1/2} \to \infty$ *as* $J \to \infty$.

Assumption 10. *Under* $H_0^\dagger$*,* $(\widehat{\rho}_1, \widehat{\rho}_2)$ *is a* $\sqrt{NT}$*-consistent estimator of* (ρ_1, ρ_2).

Denote $\mathbf{B}_N = \sigma^2 \text{tr}\{\Phi_{N,J}\Sigma_{N,f}^{-1}\}$ and $\mathbf{s}_N^2 = 2\sigma^4\text{tr}\{(\Phi_{N,J}\Sigma_{N,f}^{-1})^2\}$, where $\Phi_{N,J} = \text{Diag}(\mathbb{E}(\mathbb{P}_t^{J_1\top}\mathbb{P}_t^{J_1}/N), \mathbb{E}(\mathbb{P}_t^{J_2\top}\mathbb{P}_t^{J_2}/N))$ is a $(J_1 + J_2) \times (J_1 + J_2)$ block-diagonal matrix and $\Sigma_{N,f}$ is defined in (5). The following theorem gives the limiting null distribution of the standardized test statistics $\overline{\mathbf{T}}_{NT} = (\mathbf{T}_{NT} - \mathbf{B}_N)/\mathbf{s}_N$ and $\overline{\mathbf{LM}}_{NT}$.

Theorem 5. *Suppose that Assumptions 1-6 and 10 hold (where 1, 4 and 5 are replaced by 1′, 4† and 5†, respectively). Under* $H_0^\dagger$*,* $\overline{\mathbf{T}}_{NT} \xrightarrow{d} N(0,1)$ *and* $\overline{\mathbf{LM}}_{NT} \xrightarrow{d} N(0,1)$.

Theorem 5 implies that we can implement a one-sided test by comparing the value of $\overline{\mathbf{T}}_{NT}$ or $\overline{\mathbf{LM}}_{NT}$ with z_α. Consistent estimators of $\mathbf{B}_N$ and $\mathbf{s}_N$ can be obtained from their sample analogs: $\widehat{\mathbf{B}}_{NT} = \widehat{\sigma}^2\text{tr}\{\widehat{\Phi}_{NT,J}\widehat{\Sigma}_{N,f}^{-1}\}$ and $\widehat{\mathbf{s}}_{NT}^2 = 2\widehat{\sigma}^4\text{tr}\big\{(\widehat{\Phi}_{NT,J}\widehat{\Sigma}_{N,f}^{-1})^2\big\}$, where $\widehat{\sigma}^2$ and $\widehat{\Sigma}_{NT,f}$ are defined in Remark 2 and $\Phi_{N,J} = \text{Diag}(\widehat{\Phi}_{NT,J_1}, \widehat{\Phi}_{NT,J_2})$. The consistency of these estimators is from Lemmas 2-3 in the Online Supplementary Appendix.

5 Simulation

5.1 Sieve LS estimation

In this subsection, we examine the finite-sample performance of the sieve estimators through Monte Carlo experiments. Samples are generated from the following data-generating processes (DGPs):

$$Y_{it} = n_i^{-1} \sum_{j=1}^{N} a_{ij} f_1(Y_{j,t-1}) + f_2(Y_{i,t-1}) + Z_{i1}\gamma_1 + Z_{i2}\gamma_2 + \epsilon_{it}, \tag{16}$$

where

$$\text{DGP 1: } f_1(y) = 0.5y, \;\; f_2(y) = 0.4y,$$
$$\text{DGP 2: } f_1(y) = \cos(0.4y), \;\; f_2(y) = 0.5y,$$
$$\text{DGP 3: } f_1(y) = 0.5y, \;\; f_2(y) = \cos(0.4y),$$
$$\text{DGP 4: } f_1(y) = \cos(0.5y), \;\; f_2(y) = \cos(0.4y),$$

with $\gamma_1 = \gamma_2 = 0.5$, $Z_i = (Z_{i1}, Z_{i2})$ from a bivariate normal distribution with zero mean, unit variance, and correlation coefficient 0.5, and ϵ_{it}'s $\overset{\text{i.i.d.}}{\sim} N(0,1)$. We generate the vector $\mathbb{Y}_0$ by contraction mapping iterations. Specifically, let $\mathbb{Y}_0^{(0)}$ be an initial candidate value of $\mathbb{Y}_0$ and we update it by $\mathbb{Y}_0^{(l+1)} = M(\mathbb{Y}_0^{(l)})$, where $M(\cdot)$ is a mapping from $\mathbb{R}^N$ to $\mathbb{R}^N$ defined as $M(\boldsymbol{y}) = WF_1(\boldsymbol{y}) + F_2(\boldsymbol{y}) + \mathbb{Z}\gamma + \mathcal{E}_t$ for $\boldsymbol{y} \in \mathbb{R}^N$. The iteration stops when $\|\mathbb{Y}_0^{(l+1)} - \mathbb{Y}_0^{(l)}\| < 10^{-4}$. Given $\mathbb{Y}_0$, the observations $\mathbb{Y}_t$'s for $t = 1, \ldots, T$ are generated according to model (16). Apart from DGPs 1–4, we also consider DGPs that involve mildly nonlinear functions f_1 and f_2, see Section D.7 in the Online Supplementary Appendix for more details.

We consider the sample sizes $N \in \{100, 200\}$ and $T \in \{10, 20\}$. The adjacency matrix $A_N = (a_{ij})$ is generated according to the following network models.

(1) Erdős–Rényi (ER) Model. This model assumes that the network is constructed by randomly connecting N nodes. Let $D_{ij} = (a_{ij}, a_{ji})$ for $1 \le i < j \le N$. Following Zhu et al. (2017), we set $P(D_{ij} = (1,1)) = 20N^{-1}$, $P(D_{ij} = (0,1)) = P(D_{ij} = (1,0)) = 0.5N^{-0.8}$, and $P(D_{ij} = (0,0)) = 1 - 20N^{-1} - N^{-0.8}$. Different D_{ij}'s are independent.

(2) Stochastic Block Model (SBM). In this model, a block label $(k = 1, \ldots, K)$ is assigned for each node with equal probability, where K is the total number of blocks. Then, let $P(a_{ij} = 1) = N^{-1/4}$ if i and j belong to the same block, and $P(a_{ij} = 1) = N^{-1}$ otherwise. Accordingly, the nodes in the same block are more likely to be connected than nodes from

different blocks. We take $K = 2$ in the simulation.

We also consider a power-law distribution model where the majority of nodes have few followers, but a small percentage of them have a large number of followers. Due to space limitations, the simulation results under such a model are deferred to Section D.5 of the Online Supplementary Appendix. For the above networks, we examine their neighborhood shell size, represented by $\delta_N^{\partial}(s)$, using Monte Carlo experiments; see Section D.2 for details.

For the basis function p, we use the cubic B-splines. For simplicity, we take $J_1 = J_2 = J$. To evaluate how our sieve estimator is sensitive to the choice of J, we consider two values for each sample size: $J = 4$ and $J = 5$. The locations of the knots are determined based on the empirical quantiles. The number of Monte Carlo repetitions for each setup is $M = 1000$.

The performance of $\widehat{f}_1$ and $\widehat{f}_2$ is evaluated by the integrated squared bias (ISB), the integrated mean squared error (IMSE), and the coverage rate of the 95% confidence interval (CR95) of $f_\ell(y)$ at some fixed points y. The ISB and IMSE are defined as

$$\text{ISB: } \int_{y_{2.5}}^{y_{97.5}} \left[\frac{1}{M}\sum_{r=1}^{M} \widehat{f}_\ell^{(r)}(y) - f_\ell^o(y)\right]^2 dy, \quad \text{IMSE: } \int_{y_{2.5}}^{y_{97.5}} \left[\frac{1}{M}\sum_{r=1}^{M} \left(\widehat{f}_\ell^{(r)}(y) - f_\ell^o(y)\right)^2\right] dy,$$

where $\widehat{f}_\ell^{(r)}(y)$ is the estimate of f_ℓ obtained from the rth replicated data, and $f_\ell^o(y) = f_\ell(y) - f_\ell(0)$. The coverage rate is considered for $f_\ell(y)$ with $y \in \{y_{25}, y_{50}, y_{75}\}$, where for $\alpha \in (0, 100)$, y_α represents the α% empirical quantiles of $\{Y_{i,t-1}\}$ averaged over all repetitions.

Table 1 presents the performance of the sieve estimator of γ. We see that the bias and RMSE (root mean squared error) of both γ_1 and γ_2 are satisfactorily small, and the CR95 is close to the nominal level in all setups. The performance is not sensitive to the choice of J.

Table 2 summarizes the simulation results for f_1 and f_2. Our findings are as follows. In DGPs 2 and 4, where the function f_1 is highly nonlinear, we can significantly reduce the ISB value by increasing J, as expected in theory. However, a larger J also leads to a larger variance. For example, when $(N, T) = (200, 10)$ and W is from the ER model, using $J = 5$ seems too flexible for estimating f_1, and this results in an IMSE that is 1.5 times as large as the case with $J = 4$. These results illustrate the bias-variance trade-off of our estimator with respect to J. On the other hand, when f_1 is a linear function, as in DGPs 1 and 3, using a relatively small number of J always outperforms the cases with larger J. A similar

Table 1: Estimation results of (γ_1, γ_2)

				DGP 1			DGP 2			DGP 3			DGP 4		
Model	N	T	J	Bias γ_1	RMSE γ_1	CR95 γ_1	Bias γ_1	RMSE γ_1	CR95 γ_1	Bias γ_1	RMSE γ_1	CR95 γ_1	Bias γ_1	RMSE γ_1	CR95 γ_1
ER	100	10	4	0.0057	0.0455	0.945	0.0047	0.0449	0.952	0.0018	0.0406	0.938	0.0003	0.0395	0.952
	100	10	5	0.0057	0.0455	0.945	0.0049	0.0450	0.952	0.0018	0.0406	0.938	0.0002	0.0395	0.950
	200	10	4	0.0007	0.0306	0.953	0.0024	0.0317	0.947	-0.0003	0.0280	0.946	0.0007	0.0279	0.950
	200	10	5	0.0008	0.0307	0.953	0.0024	0.0318	0.947	-0.0003	0.0280	0.945	0.0007	0.0279	0.949
	100	20	4	0.0005	0.0305	0.959	0.0033	0.0340	0.942	-0.0011	0.0278	0.952	0.0008	0.0280	0.951
	100	20	5	0.0005	0.0306	0.957	0.0033	0.0341	0.940	-0.0011	0.0278	0.953	0.0007	0.0280	0.952
	200	20	4	-0.0001	0.0216	0.946	0.0023	0.0227	0.943	0.0008	0.0193	0.958	-0.0006	0.0197	0.951
	200	20	5	-0.0001	0.0216	0.946	0.0023	0.0227	0.946	0.0008	0.0193	0.957	-0.0006	0.0198	0.950
SBM	100	10	4	0.0052	0.0455	0.943	0.0084	0.0460	0.944	-0.0011	0.0382	0.956	0.0005	0.0411	0.938
	100	10	5	0.0052	0.0458	0.943	0.0085	0.0460	0.945	-0.0010	0.0383	0.955	0.0006	0.0410	0.938
	200	10	4	0.0024	0.0306	0.949	0.0036	0.0319	0.944	0.0017	0.0269	0.965	-0.0003	0.0277	0.940
	200	10	5	0.0024	0.0306	0.952	0.0036	0.0319	0.947	0.0017	0.0269	0.964	-0.0002	0.0278	0.941
	100	20	4	0.0019	0.0318	0.936	0.0015	0.0329	0.951	0.0004	0.0271	0.960	0.0010	0.0281	0.948
	100	20	5	0.0018	0.0317	0.935	0.0015	0.0329	0.951	0.0005	0.0271	0.960	0.0010	0.0281	0.949
	200	20	4	0.0013	0.0217	0.954	0.0024	0.0223	0.959	0.0000	0.0190	0.950	0.0010	0.0196	0.949
	200	20	5	0.0013	0.0217	0.953	0.0025	0.0223	0.962	0.0000	0.0190	0.950	0.0010	0.0196	0.948

				DGP 1			DGP 2			DGP 3			DGP 4		
Model	N	T	J	Bias γ_2	RMSE γ_2	CR95 γ_2	Bias γ_2	RMSE γ_2	CR95 γ_2	Bias γ_2	RMSE γ_2	CR95 γ_2	Bias γ_2	RMSE γ_2	CR95 γ_2
ER	100	10	4	0.0063	0.0450	0.940	0.0058	0.0452	0.951	-0.0011	0.0408	0.937	0.0005	0.0398	0.938
	100	10	5	0.0063	0.0451	0.936	0.0059	0.0452	0.955	-0.0011	0.0408	0.938	0.0005	0.0398	0.942
	200	10	4	0.0034	0.0313	0.947	0.0029	0.0305	0.958	0.0016	0.0280	0.941	-0.0005	0.0278	0.948
	200	10	5	0.0033	0.0313	0.949	0.0029	0.0306	0.956	0.0016	0.0280	0.942	-0.0005	0.0279	0.949
	100	20	4	0.0014	0.0316	0.950	0.0022	0.0336	0.937	0.0011	0.0271	0.959	0.0004	0.0294	0.938
	100	20	5	0.0014	0.0316	0.948	0.0023	0.0337	0.935	0.0012	0.0270	0.959	0.0004	0.0294	0.936
	200	20	4	0.0009	0.0217	0.952	0.0011	0.0218	0.951	0.0001	0.0200	0.949	0.0001	0.0203	0.941
	200	20	5	0.0009	0.0217	0.953	0.0011	0.0219	0.952	0.0001	0.0200	0.950	0.0001	0.0203	0.941
SBM	100	10	4	0.0020	0.0450	0.947	0.0064	0.0465	0.941	0.0009	0.0398	0.957	0.0010	0.0406	0.940
	100	10	5	0.0021	0.0449	0.951	0.0064	0.0467	0.935	0.0008	0.0399	0.955	0.0009	0.0407	0.945
	200	10	4	0.0012	0.0316	0.940	0.0027	0.0319	0.952	0.0001	0.0268	0.957	0.0005	0.0266	0.957
	200	10	5	0.0011	0.0316	0.939	0.0028	0.0318	0.950	0.0001	0.0268	0.958	0.0005	0.0266	0.958
	100	20	4	-0.0001	0.0310	0.952	0.0057	0.0327	0.955	0.0005	0.0273	0.958	-0.0005	0.0279	0.953
	100	20	5	-0.0001	0.0310	0.954	0.0058	0.0327	0.958	0.0005	0.0273	0.956	-0.0004	0.0279	0.952
	200	20	4	0.0011	0.0223	0.949	0.0014	0.0216	0.964	0.0001	0.0198	0.956	-0.0001	0.0192	0.951
	200	20	5	0.0012	0.0223	0.948	0.0014	0.0216	0.964	0.0001	0.0198	0.960	0.0000	0.0192	0.949

Table 2: Estimation results of f_1 and f_2

				DGP 1		DGP 2		DGP 3		DGP 4	
Model	N	T	J	ISB f_1	IMSE f_1	ISB f_1	IMSE f_1	ISB f_1	IMSE f_1	ISB f_1	IMSE f_1
ER	100	10	4	0.0174	1.5507	0.0077	0.9298	0.0094	0.5094	0.0001	0.5088
	100	10	5	0.0177	2.0531	0.0066	1.2957	0.0107	0.6076	0.0002	0.6034
	200	10	4	0.0006	0.5186	0.0029	0.4238	0.0022	0.2550	0.0004	0.2432
	200	10	5	0.0009	0.6564	0.0009	0.6165	0.0023	0.2995	0.0001	0.2921
	100	20	4	0.0074	0.6860	0.0035	0.4216	0.0033	0.2625	0.0002	0.2525
	100	20	5	0.0089	0.8372	0.0049	0.5915	0.0034	0.3022	0.0000	0.2967
	200	20	4	0.0022	0.2185	0.0017	0.2018	0.0002	0.1222	0.0004	0.1233
	200	20	5	0.0019	0.2882	0.0012	0.2766	0.0003	0.1425	0.0002	0.1450
SBM	100	10	4	0.0021	1.1805	0.0061	0.6624	0.0008	0.3629	0.0005	0.3659
	100	10	5	0.0050	1.6077	0.0055	0.9431	0.0006	0.4351	0.0002	0.4418
	200	10	4	0.0029	0.5998	0.0078	0.5336	0.0027	0.3160	0.0014	0.3060
	200	10	5	0.0032	0.8007	0.0069	0.7491	0.0024	0.3761	0.0013	0.3653
	100	20	4	0.0026	0.4893	0.0027	0.3052	0.0010	0.1866	0.0002	0.1749
	100	20	5	0.0024	0.7085	0.0025	0.4189	0.0010	0.2259	0.0001	0.2107
	200	20	4	0.0021	0.2637	0.0022	0.2863	0.0017	0.1488	0.0002	0.1577
	200	20	5	0.0024	0.3621	0.0004	0.4001	0.0016	0.1797	0.0002	0.1862
				DGP 1		DGP 2		DGP 3		DGP 4	
Model	N	T	J	ISB f_2	IMSE f_2	ISB f_2	IMSE f_2	ISB f_2	IMSE f_2	ISB f_2	IMSE f_2
ER	100	10	4	0.0018	0.0809	0.0025	0.0646	0.0001	0.0313	0.0001	0.0303
	100	10	5	0.0019	0.1166	0.0026	0.0825	0.0001	0.0359	0.0000	0.0350
	200	10	4	0.0003	0.0318	0.0003	0.0323	0.0001	0.0152	0.0001	0.0152
	200	10	5	0.0002	0.0381	0.0003	0.0417	0.0001	0.0174	0.0000	0.0175
	100	20	4	0.0004	0.0397	0.0005	0.0330	0.0000	0.0150	0.0000	0.0144
	100	20	5	0.0003	0.0506	0.0005	0.0410	0.0000	0.0172	0.0000	0.0169
	200	20	4	0.0001	0.0142	0.0002	0.0157	0.0000	0.0076	0.0001	0.0079
	200	20	5	0.0000	0.0168	0.0002	0.0199	0.0000	0.0087	0.0001	0.0089
SBM	100	10	4	0.0017	0.1148	0.0022	0.0677	0.0003	0.0319	0.0001	0.0317
	100	10	5	0.0020	0.1639	0.0022	0.0867	0.0002	0.0368	0.0001	0.0368
	200	10	4	0.0002	0.0309	0.0006	0.0329	0.0000	0.0157	0.0000	0.0147
	200	10	5	0.0002	0.0372	0.0006	0.0417	0.0000	0.0179	0.0000	0.0170
	100	20	4	0.0005	0.0408	0.0009	0.0330	0.0000	0.0146	0.0000	0.0148
	100	20	5	0.0005	0.0524	0.0009	0.0412	0.0000	0.0168	0.0000	0.0173
	200	20	4	0.0001	0.0152	0.0003	0.0160	0.0000	0.0086	0.0000	0.0073
	200	20	5	0.0001	0.0179	0.0003	0.0203	0.0000	0.0096	0.0000	0.0084

pattern can be found for the estimation of f_2. Finally, for all the DGPs we consider, the ISB and IMSE of f_1 are larger than those of f_2, given the same set of (N, T, J). This implies that recovering the nonlinearity in the network effect is generally more difficult.

The coverage rates of $f_1(y)$ and $f_2(y)$ are reported in Table D.1 in the Online Supplementary Appendix. It is observed that for all the quantiles considered, the coverage rates are close to the nominal level. They are also insensitive to the choice of J and the network structure.

5.2 Testing linearity

We now examine the finite sample performance of the LM and **T** tests. The DGPs are the same as above. The hypothesis H_0^n is true in DGPs 1 and 3, and H_0^m is true in DGPs 1 and 2.

Table 3 reports the rejection frequencies of the LM test at 10%, 5%, and 1% significance levels. We see that the size of the LM test is reasonably well controlled. When the function f_1 (f_2) is nonlinear, as in DGPs 2 and 4 (3 and 4), the LM test for H_0^n (H_0^m) has a good power property for all values of J and moderate sample sizes. Besides, the power of the LM test shows little sensitivity to the network structure. This may be due to the setting of the network parameters, which result in a comparable density of ER network and SBM (given the same sample size N).[5] Finally, the choice of J has some influence on the performance of our test. Specifically, if too many basis terms are used to approximate f_1 and f_2, the increase in the estimation variance diminishes the power of the test.

The rejection frequencies of the **T** test are reported in Table 4 below and Table D.2 in the Online Supplementary Appendix. It is seen that the **T** test is undersized when either H_0^n or H_0^m is considered, although it can be improved by increasing the sample size. The empirical power of the **T** test is reasonably high under moderate sample sizes. In practice, the LM test is recommended since it has better finite sample performance and is easier to

[5]To better examine the impact of the network typology on the test, we conduct additional simulations in which richer network structures and densities are considered. The results imply that the power of the tests diminishes with an increase in network density, given a constant network structure. However, it will not undergo significant alterations if only the network structure is modified, while the density remains largely unchanged; see Section D.6 for more details.

Table 3: Rejection probabilities of the LM test statistic

(1) Null: H_0^n				DGP 1			DGP 2			DGP 3			DGP 4		
Model	N	T	J	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
ER	100	10	4	0.098	0.053	0.014	0.642	0.531	0.313	0.098	0.050	0.010	0.273	0.178	0.067
	100	10	5	0.103	0.048	0.012	0.601	0.488	0.273	0.098	0.052	0.010	0.258	0.158	0.055
	200	10	4	0.110	0.058	0.011	0.921	0.850	0.696	0.104	0.050	0.012	0.431	0.309	0.134
	200	10	5	0.108	0.055	0.011	0.893	0.827	0.661	0.100	0.050	0.013	0.389	0.276	0.105
	100	20	4	0.119	0.061	0.014	0.898	0.827	0.635	0.098	0.044	0.006	0.455	0.330	0.144
	100	20	5	0.117	0.063	0.016	0.863	0.780	0.579	0.087	0.039	0.010	0.405	0.296	0.124
	200	20	4	0.082	0.043	0.008	0.991	0.987	0.949	0.110	0.058	0.009	0.681	0.579	0.338
	200	20	5	0.101	0.054	0.012	0.992	0.979	0.931	0.109	0.053	0.013	0.636	0.518	0.282
SBM	100	10	4	0.092	0.039	0.009	0.752	0.659	0.432	0.105	0.048	0.012	0.330	0.220	0.091
	100	10	5	0.100	0.052	0.011	0.716	0.609	0.393	0.096	0.055	0.014	0.302	0.209	0.076
	200	10	4	0.110	0.060	0.012	0.873	0.794	0.595	0.093	0.034	0.005	0.339	0.237	0.097
	200	10	5	0.110	0.058	0.010	0.830	0.764	0.536	0.090	0.038	0.007	0.306	0.213	0.087
	100	20	4	0.104	0.058	0.015	0.946	0.901	0.773	0.117	0.051	0.017	0.550	0.432	0.227
	100	20	5	0.103	0.052	0.017	0.919	0.876	0.720	0.108	0.064	0.012	0.507	0.391	0.196
	200	20	4	0.112	0.051	0.006	0.967	0.948	0.856	0.102	0.045	0.010	0.615	0.493	0.262
	200	20	5	0.111	0.052	0.008	0.963	0.933	0.825	0.099	0.052	0.009	0.568	0.432	0.223

(2) Null: H_0^m				DGP 1			DGP 2			DGP 3			DGP 4		
Model	N	T	J	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
ER	100	10	4	0.079	0.032	0.006	0.100	0.051	0.007	0.970	0.945	0.857	0.985	0.963	0.878
	100	10	5	0.063	0.033	0.006	0.102	0.045	0.011	0.955	0.931	0.834	0.974	0.951	0.842
	200	10	4	0.105	0.046	0.011	0.091	0.050	0.013	0.999	0.999	0.998	1	1	0.999
	200	10	5	0.094	0.050	0.012	0.095	0.059	0.012	0.999	0.999	0.996	1	1	0.999
	100	20	4	0.101	0.046	0.008	0.101	0.040	0.004	0.999	0.999	0.998	1	1	0.996
	100	20	5	0.088	0.043	0.008	0.095	0.048	0.004	0.999	0.999	0.996	1	1	0.996
	200	20	4	0.095	0.048	0.007	0.089	0.041	0.007	1.000	1	1	1	1	1
	200	20	5	0.076	0.032	0.002	0.101	0.043	0.011	1.000	1	1	1	1	1
SBM	100	10	4	0.106	0.052	0.012	0.119	0.064	0.011	0.980	0.952	0.866	0.980	0.965	0.872
	100	10	5	0.105	0.054	0.009	0.120	0.064	0.015	0.963	0.934	0.838	0.975	0.955	0.837
	200	10	4	0.093	0.044	0.006	0.104	0.054	0.008	1.000	0.999	0.995	1	0.999	0.997
	200	10	5	0.084	0.043	0.008	0.110	0.057	0.012	0.999	0.999	0.993	1	0.999	0.996
	100	20	4	0.098	0.044	0.010	0.117	0.061	0.012	1.000	1	0.999	1	1	0.999
	100	20	5	0.086	0.046	0.011	0.108	0.062	0.015	1.000	1	0.998	1	1	0.997
	200	20	4	0.100	0.059	0.014	0.099	0.052	0.009	1.000	1	1	1	1	1
	200	20	5	0.099	0.056	0.012	0.104	0.059	0.008	1.000	1	1	1	1	1

Table 4: Rejection probabilities of the distance-based test statistic

(1) Null: H_0^n				DGP 1			DGP 2			DGP 3			DGP 4		
Model	N	T	J	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
ER	100	10	4	0.053	0.031	0.016	0.495	0.444	0.322	0.033	0.015	0.007	0.077	0.048	0.015
	100	10	5	0.061	0.038	0.024	0.395	0.334	0.225	0.034	0.017	0.009	0.072	0.042	0.020
	200	10	4	0.062	0.044	0.021	0.835	0.792	0.699	0.027	0.018	0.005	0.125	0.082	0.032
	200	10	5	0.068	0.050	0.028	0.754	0.697	0.569	0.036	0.019	0.009	0.127	0.079	0.029
	100	20	4	0.076	0.052	0.025	0.811	0.751	0.635	0.023	0.015	0.009	0.152	0.109	0.049
	100	20	5	0.080	0.059	0.029	0.704	0.644	0.507	0.027	0.019	0.009	0.154	0.113	0.042
	200	20	4	0.052	0.034	0.016	0.985	0.979	0.951	0.033	0.013	0.006	0.304	0.233	0.119
	200	20	5	0.069	0.046	0.027	0.959	0.941	0.894	0.028	0.017	0.005	0.295	0.215	0.123
SBM	100	10	4	0.058	0.041	0.020	0.658	0.577	0.455	0.028	0.016	0.007	0.107	0.070	0.031
	100	10	5	0.072	0.051	0.030	0.532	0.457	0.357	0.029	0.017	0.008	0.112	0.071	0.027
	200	10	4	0.077	0.058	0.029	0.761	0.702	0.571	0.024	0.010	0.007	0.101	0.067	0.030
	200	10	5	0.077	0.053	0.028	0.650	0.563	0.441	0.024	0.017	0.006	0.100	0.058	0.026
	100	20	4	0.085	0.054	0.033	0.903	0.858	0.774	0.037	0.021	0.010	0.218	0.155	0.075
	100	20	5	0.084	0.058	0.034	0.811	0.763	0.644	0.038	0.028	0.011	0.211	0.154	0.075
	200	20	4	0.080	0.053	0.021	0.941	0.924	0.867	0.028	0.013	0.005	0.254	0.186	0.099
	200	20	5	0.084	0.056	0.027	0.893	0.852	0.753	0.028	0.019	0.009	0.247	0.173	0.091
(2) Null: H_0^m				DGP 1			DGP 2			DGP 3			DGP 4		
Model	N	T	J	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
ER	100	10	4	0.019	0.007	0.002	0.007	0.004	0.000	0.697	0.613	0.436	0.734	0.635	0.456
	100	10	5	0.025	0.012	0.005	0.021	0.013	0.006	0.689	0.595	0.415	0.718	0.621	0.446
	200	10	4	0.017	0.008	0.002	0.013	0.004	0.001	0.973	0.952	0.896	0.988	0.977	0.915
	200	10	5	0.027	0.020	0.005	0.023	0.017	0.008	0.974	0.947	0.892	0.990	0.975	0.920
	100	20	4	0.011	0.003	0.001	0.005	0.001	0.000	0.985	0.969	0.905	0.988	0.975	0.932
	100	20	5	0.028	0.013	0.004	0.014	0.010	0.002	0.982	0.955	0.886	0.986	0.968	0.920
	200	20	4	0.008	0.004	0.001	0.006	0.003	0.001	1	1	0.999	1	1	1
	200	20	5	0.009	0.004	0.003	0.017	0.012	0.006	1	1	0.999	1	1	1
SBM	100	10	4	0.022	0.011	0.003	0.010	0.005	0.002	0.701	0.590	0.404	0.735	0.638	0.474
	100	10	5	0.022	0.017	0.005	0.034	0.023	0.010	0.676	0.559	0.389	0.731	0.626	0.456
	200	10	4	0.011	0.005	0.002	0.011	0.005	0.000	0.977	0.958	0.902	0.984	0.972	0.921
	200	10	5	0.021	0.011	0.005	0.026	0.020	0.008	0.976	0.954	0.894	0.983	0.968	0.919
	100	20	4	0.014	0.008	0.003	0.013	0.007	0.001	0.980	0.959	0.896	0.989	0.980	0.936
	100	20	5	0.028	0.013	0.004	0.022	0.014	0.006	0.970	0.952	0.886	0.989	0.977	0.929
	200	20	4	0.021	0.014	0.007	0.004	0.001	0.000	1	1	1	1	1	1
	200	20	5	0.022	0.014	0.007	0.021	0.012	0.002	1	1	1	1	1	1

compute.[6]

In the above, we present the simulation results for large N and small T. For the case where N is small and T is relatively large, the performance of the estimator and the tests is also satisfactory. Due to space limitations, the related results are given in Section D.4 of the Online Supplementary Appendix.

6 Empirical application

In this section, we analyze user posting activities on Sina Weibo, a Twitter-like online social networking platform in China, using the same dataset as Zhu et al. (2017). The dataset records the posting activities of $N = 2982$ active users for $T = 4$ consecutive weeks. Zhu et al. (2017) studied the Weibo user activities based on the linear NAR model and provided strong empirical evidence that the activeness of a user is positively related to his/her neighbors. Zhu and Pan (2020) and Zhu et al. (2025) generalized the linear NAR model to the grouped NAR model, allowing the network effect of posting activities to be heterogeneous among groups. The above analysis is restricted to a linear model.

To account for possibly nonlinear network and momentum effects, we apply the proposed method to investigate the posting activities of Weibo users. The model is given by (1), where the response Y_{it} is $\ln(1 + P_{it})$ with P_{it} being the length of posts made by the ith user in the tth week. Two node-specific covariates are collected: the number of personal labels and the gender of each node (male = 1, female = 0). The adjacency matrix A is defined by the followee–follower relationship. The resulting network density is around 2.2%. For comparison, we also fit (i) the linear NAR model (13) using the OLS procedure; (ii) the semiparametric model (1) assuming $f_1(y) = \rho_1 y$, and (iii) the semiparametric model (1) assuming $f_2(y) = \rho_2 y$ using the sieve LS method. We use third-order polynomial splines with one internal knot to approximate f_1 and f_2. This yields $J_1 = J_2 = 4$.[7]

[6] A possible reason why the LM test outperforms the distance-based test is that the asymptotic mean and variance of the former are fixed numbers, while they need to be estimated in the latter case. Errors in estimation mean that the approximation to the normal distribution is not very good in finite samples. Besides, the LM test only requires estimation under the null, while the **T** test requires estimation under both the null and alternative. This renders the selection of J a more substantial issue for the **T** test.

[7] We check the robustness of the estimation and testing results in this section to the number of basis functions, the details of which are presented in Section E in the Online Supplementary Appendix.

The standardized $\mathbf{T}$ and LM tests for joint testing of H_0^n and H_0^m are 4.50 and 418.48, respectively, thus the linear model (13) is rejected at the 1% significance level. To detect the source of rejection, we further conduct tests for the hypotheses H_0^n and H_0^m in (7). The standardized $\mathbf{T}$ and LM tests for testing H_0^n are 0.051 and 2.072, respectively, implying that the linearity of f_1 cannot be rejected at the 1% level for both tests. In contrast, the standardized $\mathbf{T}$ and LM tests for testing H_0^m are 3.165 and 307.46, respectively; thus, the linearity of f_2 is rejected at the 1% level. The estimation results for f_2 are shown in Figure 1. The estimated f_2 indicates a threshold-driven autoregressive behavior in user activeness. The steep rise in f_2 at higher $\ln(1 + P_{it})$ suggests that highly active users tend to become more active in the future. This aligns with habit formation or content momentum – once users are highly engaged, they self-reinforce.

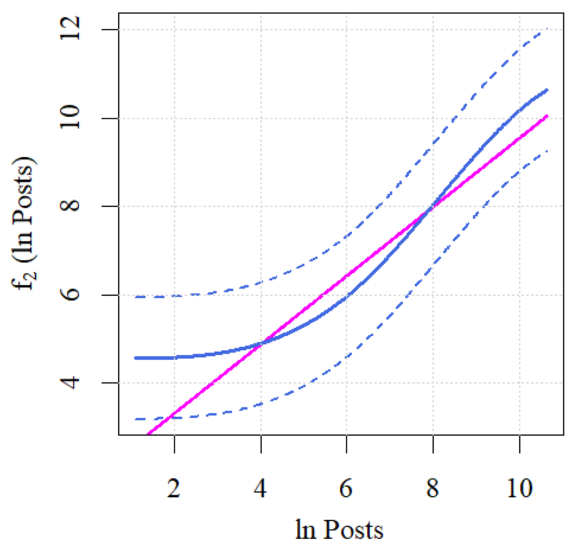

Figure 1: Estimation of f_2 based on (1) for the Sina Weibo dataset. Solid red and blue lines represent the linear and sieve estimates for f_2, respectively, and dashed blue lines represent the 95% pointwise confidence interval.

Table 5 presents the estimation results for ρ_1, ρ_2 and γ. From the first column, the network effect ρ_1 is significantly positive based on the linear NAR model (13). However, when we allow the momentum effect to be nonlinear and adopt the semiparametric specification as in (8), the network effect ρ_1 becomes smaller and insignificant at the 10% level. Moreover, the estimation results of the coefficient γ suggest that male users with more self-created

Table 5: Parameter estimates for the Sina Weibo dataset.

	Linear NAR		Semi-NAR (1)		Semi-NAR (2)		Semi-NAR (3)	
	Coef.	t-value	Coef.	t-value	Coef.	t-value	Coef.	t-value
intercept	0.5342	4.0697						
ρ_1	0.0873	4.8160	0.0138	0.8137				
ρ_2	0.7828	115.7912			0.7804	4.6592		
Number of Labels	0.0177	5.7483	0.0139	4.8422	0.0182	6.3638	0.0145	5.0668
Gender	0.0970	4.0040	0.0698	3.1013	0.0962	4.2726	0.0692	3.0785

Note. The column "Linear NAR" displays the parametric LS estimates for model (13), "Semi-NAR (1)" and "Semi-NAR (2)" the sieve LS estimates for model (1) with $f_1(y) = \rho_1 y$ and $f_2(y) = \rho_2 y$, respectively, and "Semi-NAR (3)" the sieve LS estimates for model (1).

labels tend to be more active, which is in line with the findings of Zhu et al. (2017). The estimates differ slightly among the four specifications in terms of magnitude and significance. In view that we reject the linearity of f_2, it is expected that the estimates obtained from the semiparametric model (1) or (8) are more reliable.

Finally, to evaluate the out-of-sample prediction performance, we use the data from the first 3 weeks for estimation, and observations in the last week to evaluate its prediction accuracy. The forecasting performance of the semiparametric NAR model (1) is compared to the linear NAR model (13), as well as a baseline AR(1) model fitted separately to each node. The mean absolute prediction error (MAPE) and mean squared prediction error (MSPE) for the linear and semiparametric NAR models are shown in Table 6. It is seen that the MAPE for the semiparametric NAR model (8) (0.710) is considerably smaller than the MAPE obtained by the linear NAR model (0.785). For the AR(1) model fitted to each node, the MAPE is 3.304, which is substantially larger than that of the NAR models. A similar pattern can be found for MSPE. We conclude that the NAR model (8) provides significant accuracy improvements in prediction while also achieving parsimony.

Table 6: Forecasting performance of the NAR models.

	Linear NAR	Semi-NAR (1)	Semi-NAR (2)	Semi-NAR (3)	AR
MAPE	0.785	0.710	0.784	0.709	3.304
MSPE	1.077	0.988	1.077	0.987	10.637

7 Conclusion

This paper considers the estimation and inference of NAR models with nonparametric network and momentum effects. A sieve LS estimation method is developed, whose large sample properties are established under various asymptotic settings. We also propose nonparametric specification tests for testing the null hypothesis that the network (momentum) effect is linear in the past values of others' (or one's own) outcomes. An empirical illustration using Sina Weibo data indicates the usefulness of the proposed model.

Some issues remain unaddressed in this study. These include choosing an optimal order for the basis expansion of J_1 and J_2, accounting for unobserved individual heterogeneity, and investigating sieve estimation with shape restrictions. Moreover, as suggested by many authors (e.g., Li et al., 2024; Zhu and Pan, 2020), allowing the functions f_1 and f_2 to be heterogeneous across units could be important in some empirical situations. We leave these promising topics for future research.

References

Ai, C. and Chen, X. (2003), 'Efficient estimation of models with conditional moment restrictions containing unknown functions', *Econometrica* **71**(6), 1795–1843.

Armillotta, M. and Fokianos, K. (2023), 'Nonlinear network autoregression', *The Annals of Statistics* **51**(6), 2526 – 2552.

Chen, E. Y., Fan, J. and Zhu, X. (2023), 'Community network auto-regression for high-dimensional time series', *Journal of Econometrics* **235**(2), 1239–1256.

Chen, X. (2007), Chapter 76 Large Sample Sieve Estimation of Semi-Nonparametric Models, Vol. 6 of *Handbook of Econometrics*, Elsevier, pp. 5549–5632.

Chen, X. and Christensen, T. M. (2015), 'Optimal uniform convergence rates and asymptotic normality for series estimators under weak dependence and weak conditions', *Journal of Econometrics* **188**(2), 447–465.

Chen, X., Hong, H. and Tamer, E. (2005), 'Measurement error models with auxiliary data', *The Review of Economic Studies* **72**(2), 343–366.

de Jong, R. M. and Bierens, H. J. (1994), 'On the limit behavior of a Chi-square type test if the number of conditional moments tested approaches infinity', *Econometric Theory* **10**(1), 70–90.

Ding, Y., Zhu, X., Pan, R. and Zhang, B. (2025), 'Network vector autoregression with time-varying nodal influence', *Computational Economics* .

Gupta, A. (2018), 'Nonparametric specification testing via the trinity of tests', *Journal of Econometrics* **203**(1), 169–185.

He, X. and Song, K. (2023), 'Measuring diffusion over a large network', *The Review of Economic Studies* p. rdad115.

Hoff, P. D., Raftery, A. E. and Handcock, M. S. (2002), 'Latent space approaches to social network analysis', *Journal of the American Statistical Association* **97**(460), 1090–1098.

Hoque, M. E., Billah, M., Kapar, B. and Naeem, M. A. (2024), 'Quantifying the volatility spillover dynamics between financial stress and US financial sectors: Evidence from QVAR connectedness', *International Review of Financial Analysis* **95**, 103434.

Hoshino, T. (2022), 'Sieve IV estimation of cross-sectional interaction models with nonparametric endogenous effect', *Journal of Econometrics* **229**(2), 263–275.

Huang, J. Z. (1998), 'Projection estimation in multiple regression with application to functional ANOVA models', *The Annals of Statistics* **26**(1), 242 – 272.

Huang, J. Z. and Shen, H. (2004), 'Functional coefficient regression models for non-linear time series: a polynomial spline approach', *Scandinavian Journal of Statistics* **31**(4), 515–534.

Härdle, W. K., Wang, W. and Yu, L. (2016), 'TENET: Tail-Event driven NETwork risk', *Journal of Econometrics* **192**(2), 499–513.

Härdle, W. and Mammen, E. (1993), 'Comparing nonparametric versus parametric regression fits', *The Annals of Statistics* **21**(4), 1926–1947.

Härdle, W., Tsybakov, A. and Yang, L. (1998), 'Nonparametric vector autoregression', *Journal of Statistical Planning and Inference* **68**(2), 221–245.

Jeliazkov, I. (2013), Nonparametric vector autoregressions: Specification, estimation, and inference, *in* 'VAR Models in Macroeconomics – New Developments and Applications: Essays in Honor of Christopher A. Sims', Emerald Group Publishing Limited.

Jenish, N. (2012), 'Nonparametric spatial regression under near-epoch dependence', *Journal of Econometrics* **167**(1), 224–239.

Jenish, N. and Prucha, I. R. (2012), 'On spatial processes and asymptotic inference under near-epoch dependence', *Journal of Econometrics* **170**(1), 178–190.

Jin, X. (2015), 'Volatility transmission and volatility impulse response functions among the greater china stock markets', *Journal of Asian Economics* **39**, 43–58.

Kalli, M. and Griffin, J. E. (2018), 'Bayesian nonparametric vector autoregressive models', *Journal of Econometrics* **203**(2), 267–282.

Keeling, M. J. and Eames, K. T. (2005), 'Networks and epidemic models', *Journal of The Royal Society Interface* **2**(4), 295–307.

Kojevnikov, D., Marmer, V. and Song, K. (2021), 'Limit theorems for network dependent random variables', *Journal of Econometrics* **222**(2), 882–908.

Leung, M. P. (2022), 'Causal inference under approximate neighborhood interference', *Econometrica* **90**(1), 267–293.

Li, D., Peng, B., Tang, S. and Wu, W. (2024), 'Estimation of grouped time-varying network vector autoregression models'. *arXiv*: 2303.10117.

Lütkepohl, H. (1991), *Introduction to multiple time series analysis*, Springer Berlin, Heidelberg.

Newey, W. K. (1997), 'Convergence rates and asymptotic normality for series estimators', *Journal of Econometrics* **79**(1), 147–168.

Penrose, M. (2003), *Random Geometric Graphs*, Oxford University Press.

Pesaran, M. H., Schuermann, T. and Weiner, S. M. (2004), 'Modeling regional interdependencies using a global error-correcting macroeconometric model', *Journal of Business & Economic Statistics* **22**(2), 129–162.

Stone, C. J. (1982), 'Optimal global rates of convergence for nonparametric regression', *The Annals of Statistics* **10**(4), 1040 – 1053.

Su, L. and Lu, X. (2013), 'Nonparametric dynamic panel data models: Kernel estimation and specification testing', *Journal of Econometrics* **176**(2), 112–133.

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

Zhu, X. and Pan, R. (2020), 'Grouped network vector autoregression', *Statistica Sinica* **30**(3), 1437–1462.

Zhu, X., Pan, R., Li, G., Liu, Y. and Wang, H. (2017), 'Network vector autoregression', *The Annals of Statistics* **45**(3).

Zhu, X., Xu, G. and Fan, J. (2025), 'Simultaneous estimation and group identification for network vector autoregressive model with heterogeneous nodes', *Journal of Econometrics* **249**, 105564.

Online Supplementary Appendix to

"Nonparametric Network Vector Autoregression"

This Online Supplementary Appendix contains some useful auxiliary results and proofs for the theorems in the main text. It also includes discussions of assumptions and remarks in Section 3, some additional simulation results, and a robustness check of the empirical analysis in Section 6.

A Some Lemmas

We first introduce some lemmas that are useful in dealing with the convergence in probability of the sample mean of nonlinear transformations of Y_{it}. We consider the sequence $\{Y_{it}\}$ as a random field laid on a network over time for analysis. Let $\mathcal{D}_{NT_N} = \{(i,t) : 1 \leq i \leq N, 1 \leq t \leq T_N, i,t \in \mathbb{Z}\}$ be the set of cross-sectional and time indices. Also, recall that $D_N = \{1, 2, ..., N\}$ and G_N is an undirected network on D_N given by $G_N = (D_N, E_N)$, where E_N denotes a set of links such that $\{i_1, i_2\} \in E_N$ if either $a_{i_1 i_2} = 1$ or $a_{i_2 i_1} = 1$. Modeling spatial-time dependence usually requires a metric on $\mathcal{D}_{NT_N}$ (Kojevnikov et al., 2021; Qu et al., 2017). Similar to Qu et al. (2017), we define the metric as

$$\Delta(l_1, l_2) = \max\{|t_1 - t_2|, d_N(i_1, i_2)\},$$

where $l_1 = (i_1, t_1)$ and $l_2 = (i_2, t_2)$ are two generic indices in $\mathcal{D}_{NT_N}$, and $d_N(i_1, i_2)$ is the distance between nodes i_1 and i_2 in G_N, i.e., the length of the shortest path between nodes i_1 and i_2 given G_N. We extend the notion of spatial near-epoch-dependent (NED) process in Jenish and Prucha (2012) to incorporate network typology and time dimension as follows.

Definition 1. *Let $\boldsymbol{\xi} = \{\xi_{it} : (i,t) \in \mathcal{D}_{NT_N}, N \geq 1\}$ and $\boldsymbol{e} = \{e_{it} : (i,t) \in \mathcal{D}_{NT_N}, N \geq 1\}$ be two random fields. Then, the random field $\boldsymbol{\xi}$ is said to be L^p-NED on $\boldsymbol{e}$ if*

$$\left\| \xi_{it} - \mathbb{E}\left[\xi_{it} \mid \mathcal{F}_{it}(s)\right] \right\|_p \leq d_{it} \upsilon_s$$

for an array of finite positive constants $\{d_{it} : (i,t) \in \mathcal{D}_{NT_N}; N \geq 1\}$ and some sequence

$\upsilon_s \geq 0$ *with* $\upsilon_s \to 0$ *as* $s \to \infty$, *where* $\mathcal{F}_{it}(s)$ *is the* σ*-field generated by the random variables* $\{e_{j\tau} : \Delta((i,t),(j,\tau)) \leq s\}$. *The* d_{it}*'s and* υ_s *are called the NED scaling factors and NED coefficient, respectively.* $\boldsymbol{\xi}$ *is said to be uniformly* L^p*-NED on* $\boldsymbol{e}$ *if* d_{it} *is uniformly bounded. If* $\upsilon_s = O(\varrho^s)$ *for some* $0 < \varrho < 1$, *then it is called geometrically* L^p*-NED.*

The following lemma establishes the L^2-NED property of $\{Y_{it}\}$ generated by model (1).

Lemma 1. *Suppose that Assumptions 1-3 hold. Then,* $\{Y_{it}\}$ *is uniformly and geometrically* L^2*-NED on* $\{\epsilon_{it}\}$.

Proof. We denote $|x|_r = (\sum_{i=1}^p |x_i|^r)^{1/r}$ the l^r-norm of a $p \times 1$ vector x. If $r = \infty$, $|x|_\infty = \max_{1\leq i\leq p} |x_i|$. Let $|M|_{vec} = (|m_{ij}|) \in \mathbb{R}^{n\times p}$ for any arbitrary matrix $M = (m_{ij}) \in \mathbb{R}^{n\times p}$, and $\|M\|_\infty = \max_{1\leq i\leq n} \sum_{j=1}^p |m_{ij}|$. For matrices $M_1 = (m_{ij}^{(1)}) \in \mathbb{R}^{n\times p}$ and $M_2 = (m_{ij}^{(2)}) \in \mathbb{R}^{n\times p}$, define $M_1 \preceq M_2$ as $m_{ij}^{(1)} \leq m_{ij}^{(2)}$ for $1 \leq i \leq n$ and $1 \leq j \leq p$. We complete the proof using arguments similar to Lemma S-2 in Armillotta and Fokianos (2023). For $K > 0$, define

$$\bar{\mathbb{Y}}_t = \begin{cases} \bar{F}(\bar{\mathbb{Y}}_{t-1}), & t > 0 \\ \mathbb{Y}_0, & t \leq 0 \end{cases} \quad \text{and} \quad \hat{\mathbb{Y}}_{t-K}^s = \begin{cases} \bar{F}(\hat{\mathbb{Y}}_{t-K}^{s-1}) + \mathbb{Z}\gamma + \mathcal{E}_s, & \max\{t-K, 0\} < s \leq t \\ \bar{\mathbb{Y}}_s, & s \leq \max\{t-K, 0\} \end{cases} \tag{1}$$

where $\bar{F}(A_t) \equiv WF_1(A_t) + F_2(A_t)$ for any $A_t \in \mathbb{R}^N$. We first aim to show that under Assumption 2, it holds that

$$\left|\mathbb{Y}_t - \hat{\mathbb{Y}}_{t-K}^t\right|_{vec} \preceq G^K \sum_{j=0}^{t-K-1} G^j \left|\mathbb{Z}\gamma + \mathcal{E}_{t-K-j}\right|_{vec}, \tag{2}$$

where $G \equiv \kappa_1 W + \kappa_2 I_N$. Indeed, for $t - K > 0$,

$$\begin{aligned} \left|\mathbb{Y}_t - \hat{\mathbb{Y}}_{t-K}^t\right|_{vec} &= |\bar{F}(\mathbb{Y}_{t-1}) + \mathbb{Z}\gamma + \mathcal{E}_t - \bar{F}(\hat{\mathbb{Y}}_{t-K}^{t-1}) - \mathbb{Z}\gamma - \mathcal{E}_t|_{vec} \\ &= |\bar{F}(\mathbb{Y}_{t-1}) - \bar{F}(\hat{\mathbb{Y}}_{t-K}^{t-1})|_{vec} \preceq G|\mathbb{Y}_{t-1} - \hat{\mathbb{Y}}_{t-K}^{t-1}|_{vec} \preceq G^K|\mathbb{Y}_{t-K} - \bar{\mathbb{Y}}_{t-K}|. \end{aligned} \tag{3}$$

By construction, we have for $t > 0$, $|\mathbb{Y}_t - \bar{\mathbb{Y}}_t|_{vec} = |\bar{F}(\mathbb{Y}_{t-1}) + \mathbb{Z}\gamma + \mathcal{E}_t - \bar{F}(\bar{\mathbb{Y}}_{t-1})|_{vec} \preceq G|\mathbb{Y}_{t-1} - \bar{\mathbb{Y}}_{t-1}|_{vec} + |\mathbb{Z}\gamma + \mathcal{E}_t|_{vec} \preceq G\left(G|\mathbb{Y}_{t-2} - \bar{\mathbb{Y}}_{t-2}|_{vec} + |\mathbb{Z}\gamma + \mathcal{E}_{t-1}|_{vec}\right) + |\mathbb{Z}\gamma + \mathcal{E}_t|_{vec} \preceq \sum_{l=0}^{t-1} G^l |\mathbb{Z}\gamma + \mathcal{E}_{t-l}|_{vec}$. This, together with (3), implies (2) with $t - K > 0$. For $t - K \leq 0$, we can easily show that $|\mathbb{Y}_t - \hat{\mathbb{Y}}^t_{t-K}|_\infty \leq (\kappa_1+\kappa_2)^t |\mathbb{Y}_0 - \bar{\mathbb{Y}}_0|_\infty = 0$. (2) follows by combining the above results.

Denote the ith element of $\hat{\mathbb{Y}}^t_{t-K}$ by $\hat{Y}^t_{i,t-K}$. Note that $\|G^j\|_\infty \leq \|G\|^j_\infty = (\kappa_1+\kappa_2)^j$, for $j \in \mathbb{N}$. Then, by (2) and Minkowski's inequality, we have for each $1 \leq i \leq N$,

$$\begin{aligned}\|Y_{it} - \hat{Y}^t_{i,t-K}\|_2 &\leq \sum_{j=0}^{t-K-1} \|G^{j+K}\|_\infty \sup_{i\geq 1} \left\| Z_i^\top \gamma + \epsilon_{i,t-K-j} \right\|_2 \\ &\leq \sum_{j=0}^{t-K-1} \|G\|^{j+K}_\infty C_v \leq C_v d^K/(1-d), \end{aligned} \tag{4}$$

where $d \equiv \kappa_1 + \kappa_2 < 1$, and $C_v = \sup_{1\leq i\leq N} \|Z_i^\top \gamma + \epsilon_{it}\|_2 < \infty$ under the maintained assumptions. Note that $\hat{Y}^t_{i,t-K}$ is a $\mathcal{F}_{it}(K)$-measurable approximation to Y_{it}. It follows from Theorem 10.12 of Davidson (1994) that $\|Y_{it} - \mathbb{E}[Y_{it} \mid \mathcal{F}_{it}(K)]\|_2 \leq \|Y_{it} - \hat{Y}^t_{i,t-K}\|_2 \leq C_v d^K/(1-d)$. The lemma immediately follows. □

Let $\widehat{C}_{NT,ZJ} = (NT)^{-1}\sum_{t=1}^T \mathbb{Z}^\top \mathbb{Q}_{t-1}$ and $\widehat{\Psi}_{NT,J} = (NT)^{-1}\sum_{t=1}^T \mathbb{Q}_{t-1}^\top \mathbb{Q}_{t-1}$ with $\mathbb{Q}_{t-1} = (W\mathbb{P}^{J_1}_{t-1}, \mathbb{P}^{J_2}_{t-1})$. Further, recall that $C_{N,ZJ} = \mathbb{E}(\widehat{C}_{NT,ZJ})$ and $\Psi_{N,J} = \mathbb{E}(\widehat{\Psi}_{NT,J})$. Lemma 2 gives the convergence rate of $\widehat{C}_{NT,ZJ}$ and $\widehat{\Psi}_{NT,J}$ in the Euclidean norm.

Lemma 2. *Under Assumptions 1-5, (i)* $\|\widehat{C}_{NT,ZJ} - C_{N,ZJ}\| = O_p\big(\zeta^{-1}_{0,J}\sqrt{r_{NT,J}}\big) = o_p(1)$, *and (ii)* $\|\widehat{\Psi}_{NT,J} - \Psi_{N,J}\| = O_p\big(\sqrt{r_{NT,J}}\big) = o_p(\underline{\nu}_{J_1 J_2})$, *where $r_{NT,J}$ is defined in Assumption 5.*

Proof. In the following, we suppress the dependence of $\widehat{C}_{NT,ZJ}$, $\widehat{\Psi}_{NT,J}$, $C_{N,ZJ}$ and $\Psi_{N,J}$ on (N,T) and write $\widehat{C}_{ZJ}$, $\widehat{\Psi}_J$, C_{ZJ} and Ψ_J for notational simplicity.

(i) Recall that $\bar{P}^{J_1}_{it} = \sum_{j=1}^N w_{ij} p^{J_1}(Y_{jt})$ and $P^{J_2}_{it} = p^{J_2}(Y_{it})$. Partition $\widehat{C}_{ZJ}$ as

$$\widehat{C}_{ZJ} = \big(\widehat{C}_{ZJ,1}, \widehat{C}_{ZJ,2}\big) \equiv \left(\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T Z_i \bar{P}^{J_1\top}_{i,t-1}, \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T Z_i P^{J_2\top}_{i,t-1}\right),$$

where $\widehat{C}_{ZJ,1}$ and $\widehat{C}_{ZJ,2}$ are $d_Z \times J_1$ and $d_Z \times J_2$ submatrices of $\widehat{C}_{ZJ}$, respectively.

Recall that $\mathcal{F}_{it}(s)$ is the σ-field generated by the random variables $\{\epsilon_{j\tau} : (j,\tau) \in \mathcal{D}_{NT_N}, \Delta\big((i,t),(j,\tau)\big) \leq s\}$. Let $\hat{\bar{P}}^t_{i,t-s} \equiv \sum_{j=1}^N w_{ij} p^{J_1}(\hat{Y}^t_{j,t-s})$ and $\hat{P}^t_{i,t-s} \equiv p^{J_2}(\hat{Y}^t_{i,t-s})$, where

$\hat{Y}_{i,t-s}^{t}$ is a $\mathcal{F}_{it}(s)$-measurable approximation to Y_{it} defined in Lemma 1. Note that

$$
\begin{aligned}
\mathbb{E}\big\|Z_i\bar{P}_{i,t-1}^{J_1\top}-Z_i\hat{\bar{P}}_{i,t-s}^{(t-1)\top}\big\|^2 &= \mathbb{E}\|Z_i\|^2\left\|\sum_{j=1}^N w_{ij}\big(p^{J_1}(Y_{j,t-1})-p^{J_1}(\hat{Y}_{j,t-s}^{t-1})\big)\right\|^2\\
&\le C_z^2\sum_{j=1}^N w_{ij}\mathbb{E}\left\|p^{J_1}(Y_{j,t-1})-p^{J_1}(\hat{Y}_{j,t-s}^{t-1})\right\|^2\\
&\le C_z^2\sup_{1\le i\le N}\mathbb{E}\left\|p^{J_1}(Y_{i,t-1})-p^{J_1}(\hat{Y}_{i,t-s}^{t-1})\right\|^2\\
&\le C_z^2\zeta_{1,J_1}^2\sup_{1\le i\le N}\mathbb{E}\big\|Y_{i,t-1}-\hat{Y}_{i,t-s}^{t-1}\big\|^2,
\end{aligned}
\tag{5}
$$

where we have used the facts that $\max_{1\le i\le N}\|Z_i\|\le C_z<\infty$ for some constant C_z by Assumption 3, $\sum_{j=1}^N w_{ij}=1$, and the Cauchy–Schwarz (CS) inequality.

Let $C_{ZJ,\ell}=\mathbb{E}(\widehat{C}_{ZJ,\ell})$ for $\ell=1,2$. In the following, we denote $\mathrm{Cov}(A,B)=\mathbb{E}\mathrm{tr}\{(A-\mathbb{E}(A))(B-\mathbb{E}(B))^\top\}=\sum_{i=1}^p\sum_{j=1}^q\mathrm{Cov}(a_{ij},b_{ij})$ for any two $p\times q$ random matrices $A=(a_{ij})$ and $B=(b_{ij})$. Also, let $\mathbb{E}[A|\mathcal{F}]$ denote the $p\times q$ matrix with (i,j)th element $\mathbb{E}[a_{ij}|\mathcal{F}]$, $1\le i\le p, 1\le j\le q$. Then,

$$
\begin{aligned}
\mathbb{E}\big\|\widehat{C}_{ZJ,1}-C_{ZJ,1}\big\|^2 &= \mathbb{E}\left\|\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\big(Z_i\bar{P}_{i,t-1}^{J_1\top}-\mathbb{E}\big(Z_i\bar{P}_{i,t-1}^{J_1\top}\big)\big)\right\|^2\\
&=\frac{1}{(NT)^2}\sum_{(i,t)\in\mathcal{D}_{NT}}\mathbb{E}\big\|Z_i\bar{P}_{i,t-1}^{J_1\top}-\mathbb{E}(Z_i\bar{P}_{i,t-1}^{J_1\top})\big\|^2\\
&\quad+\frac{1}{(NT)^2}\sum_{(i,t)\in\mathcal{D}_{NT}}\sum_{(j,\tau)\in\mathcal{D}_{NT}\setminus(i,t)}\mathrm{Cov}\big(Z_i\bar{P}_{i,t-1}^{J_1\top},Z_j\bar{P}_{j,\tau-1}^{J_1\top}\big),
\end{aligned}
\tag{6}
$$

where $\mathcal{D}_{NT}$ is a shorthand for $\mathcal{D}_{NT_N}$. By Assumptions 3(ii) and 4(i), we have

$$
\frac{1}{NT}\sum_{(i,t)\in\mathcal{D}_{NT}}\mathbb{E}\big\|Z_i\bar{P}_{i,t-1}^{J_1\top}-\mathbb{E}(Z_i\bar{P}_{i,t-1}^{J_1\top})\big\|^2\le\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\mathbb{E}\|Z_i\bar{P}_{i,t-1}^{J_1\top}\|^2=O(J_1). \tag{7}
$$

We control the second term on the r.h.s. of (6) using similar arguments to those used in Lemma A.3 of Jenish and Prucha (2012). For any $(i,t)\in\mathcal{D}_{NT}$ and $m\in\mathbb{N}$, let

$$
\xi_{i,t-1}^m\equiv\mathbb{E}\big[Z_i\bar{P}_{i,t-1}^{J_1\top}\,|\,\mathcal{F}_{i,t-1}(m)\big],\quad \eta_{i,t-1}^m\equiv Z_i\bar{P}_{i,t-1}^{J_1\top}-\mathbb{E}\big[Z_i\bar{P}_{i,t-1}^{J_1\top}\,|\,\mathcal{F}_{i,t-1}(m)\big].
$$

By Jensen's inequality and Assumption 4(i), we have $\mathbb{E}\|\xi_{i,t-1}^m\|^2 = \mathbb{E}\big\|\mathbb{E}\big[Z_i\bar{P}_{i,t-1}^{J_1\top}|\mathcal{F}_{i,t-1}(m)\big]\big\|^2 \leq \mathbb{E}\big\|Z_i\bar{P}_{i,t-1}^{J_1\top}\big\|^2 \lesssim J_1$, and similarly $\mathbb{E}\|\eta_{i,t-1}^m\|^2 \leq 4\mathbb{E}\|Z_i\bar{P}_{i,t-1}^{J_1\top}\|^2 \lesssim J_1$. Also, (4) and (5) imply that for $m \geq 1$, $\mathbb{E}\|\eta_{i,t-1}^m\|^2 \leq \mathbb{E}\big\|Z_i\bar{P}_{i,t-1}^{J_1\top} - Z_i\hat{\bar{P}}_{i,t-m-1}^{(t-1)\top}\big\|^2 \leq C_z\zeta_{1,J_1}^2 \sup_{i\geq 1}\mathbb{E}\big\|Y_{i,t-1} - \hat{Y}_{i,t-m-1}^{t-1}\big\|^2 \lesssim \zeta_{1,J_1}^2 d^{2m}$.

Decompose $Z_i\bar{P}_{i,t-1}^{J_1\top}$ and $Z_j\bar{P}_{j,\tau-1}^{J_1\top}$ as

$$Z_i\bar{P}_{i,t-1}^{J_1\top} = \xi_{i,t-1}^{\lfloor h/3\rfloor} + \eta_{i,t-1}^{\lfloor h/3\rfloor}, \quad Z_j\bar{P}_{j,\tau-1}^{J_1\top} = \xi_{j,\tau-1}^{\lfloor h/3\rfloor} + \eta_{j,\tau-1}^{\lfloor h/3\rfloor}, \tag{8}$$

with $h = \Delta((i,t),(j,\tau))$ and $\lfloor a\rfloor$ denotes the largest integer not larger than a. Note that

$$\begin{aligned}\Big|\mathrm{Cov}\Big(Z_i\bar{P}_{i,t-1}^{\top}, Z_j\bar{P}_{j,\tau-1}^{\top}\Big)\Big| &= \Big|\mathrm{Cov}\Big(\xi_{i,t-1}^{\lfloor h/3\rfloor} + \eta_{i,t-1}^{\lfloor h/3\rfloor}, \xi_{j,\tau-1}^{\lfloor h/3\rfloor} + \eta_{j,\tau-1}^{\lfloor h/3\rfloor}\Big)\Big| \\ &\leq \Big|\mathrm{Cov}\Big(\xi_{i,t-1}^{\lfloor h/3\rfloor}, \xi_{j,\tau-1}^{\lfloor h/3\rfloor}\Big)\Big| + \Big|\mathrm{Cov}\Big(\eta_{i,t-1}^{\lfloor h/3\rfloor}, \xi_{j,\tau-1}^{\lfloor h/3\rfloor}\Big)\Big| \\ &\quad + \Big|\mathrm{Cov}\Big(\xi_{i,t-1}^{\lfloor h/3\rfloor}, \eta_{j,\tau-1}^{\lfloor h/3\rfloor}\Big)\Big| + \Big|\mathrm{Cov}\Big(\eta_{i,t-1}^{\lfloor h/3\rfloor}, \eta_{j,\tau-1}^{\lfloor h/3\rfloor}\Big)\Big|.\end{aligned}$$

We will now bound separately each term on the r.h.s. of the last inequality. First, by Assumption 1, the σ-fields $\mathcal{F}_{i,t-1}(\lfloor h/3\rfloor)$ and $\mathcal{F}_{j,\tau-1}(\lfloor h/3\rfloor)$ are conditionally independent given $\mathbb{Y}_0$, which implies that $\mathrm{Cov}\big(\xi_{i,t-1}^{\lfloor h/3\rfloor}, \xi_{j,\tau-1}^{\lfloor h/3\rfloor}\big) = 0$ conditional on $\mathbb{Y}_0$. Second, the CS and Hölder inequalities give the following bound on the second and third terms:

$$\begin{aligned}\big|\mathrm{Cov}\big(\eta_{i,t-1}^{\lfloor h/3\rfloor}, \xi_{j,\tau-1}^{\lfloor h/3\rfloor}\big)\big| &= \Big|\mathbb{E}\mathrm{tr}\left\{\big(\eta_{i,t-1}^{\lfloor h/3\rfloor} - \mathbb{E}\big(\eta_{i,t-1}^{\lfloor h/3\rfloor}\big)\big)\big(\xi_{j,\tau-1}^{\lfloor h/3\rfloor} - \mathbb{E}\big(\xi_{j,\tau-1}^{\lfloor h/3\rfloor}\big)\big)^{\top}\right\}\Big| \\ &\leq 2\left\{\mathbb{E}\left\|\eta_{i,t-1}^{\lfloor h/3\rfloor}\right\|^2 \mathbb{E}\left\|\xi_{j,\tau-1}^{\lfloor h/3\rfloor}\right\|^2\right\}^{1/2} \leq C\zeta_{1,J_1}J_1^{1/2}d^{\lfloor h/3\rfloor},\end{aligned}$$

where C is a constant that does not depend on i, t, N, T and J_1. Lastly, $\big|\mathrm{Cov}\big(\eta_{i,t-1}^{\lfloor h/3\rfloor}, \eta_{j,\tau-1}^{\lfloor h/3\rfloor}\big)\big| \leq 2\Big\{\mathbb{E}\big\|\eta_{i,t-1}^{\lfloor h/3\rfloor}\big\|^2\mathbb{E}\big\|\eta_{j,\tau-1}^{\lfloor h/3\rfloor}\big\|^2\Big\}^{1/2} \leq C\zeta_{1,J_1}J_1^{1/2}d^{\lfloor h/3\rfloor}$. A combination of the above results yields $\big|\mathrm{Cov}\big(Z_i\bar{P}_{i,t-1}^{\top}, Z_j\bar{P}_{j,\tau-1}^{\top}\big)\big| \leq C\zeta_{1,J_1}J_1^{1/2}d^{\lfloor h/3\rfloor}$.

We now derive the bound for the second term on the r.h.s. of (6). Using the above result, we have for any $(i,t) \in \mathcal{D}_{NT}$,

$$\sum_{j\neq i}^{N}\sum_{\tau\neq t}^{T}\mathrm{Cov}\big(Z_i\bar{P}_{i,t-1}^{\top}, Z_j\bar{P}_{j,\tau-1}^{\top}\big)$$

$$
\begin{aligned}
&= \sum_{s=1}^{N} \sum_{j\in\mathcal{N}_N^{\partial}(i;s)} \sum_{l=t-T,\, l\neq 0}^{t-1} \mathrm{Cov}\big(Z_i \bar{P}_{i,t-1}^{\top}, Z_j \bar{P}_{j,t-l-1}^{\top}\big) \\
&\leq C\zeta_{1,J_1} J_1^{1/2} \sum_{s=1}^{N} \sum_{j\in\mathcal{N}_N^{\partial}(i;s)} \sum_{l=t-T,\, l\neq 0}^{t-1} d^{\lfloor \max\{s,|l|\}/3\rfloor} \\
&\leq C\zeta_{1,J_1} J_1^{1/2} \sum_{s=1}^{N} \sum_{j\in\mathcal{N}_N^{\partial}(i;s)} \sum_{l=1}^{T-1} d^{\lfloor \max\{s,l\}/3\rfloor} \\
&\leq C\zeta_{1,J_1} J_1^{1/2} \sum_{s=1}^{N} \big|\mathcal{N}_N^{\partial}(i;s)\big| \left(\sum_{l=1}^{\min\{s,T-1\}} d^{\lfloor s/3\rfloor} + \sum_{l=\min\{s+1,T-1\}}^{T-1} d^{\lfloor l/3\rfloor} \right) \\
&\leq C\zeta_{1,J_1} J_1^{1/2} \sum_{s=1}^{N} \min\{s,T\} \big|\mathcal{N}_N^{\partial}(i;s)\big| d^{\lfloor s/3\rfloor}, \qquad (9)
\end{aligned}
$$

where the last inequality follows from the observation that $\sum_{s=1}^{N} \sum_{l=\min\{s+1,T-1\}}^{T-1} |\mathcal{N}_N^{\partial}(i;s)| d^{\lfloor l/3\rfloor} \leq \sum_{s=1}^{N} \big(|\mathcal{N}_N^{\partial}(i;s)| d^{\lfloor s/3\rfloor} \cdot \sum_{q=1}^{T-s} d^{\lfloor q/3\rfloor} \big) \leq \big(\sum_{s=1}^{N} |\mathcal{N}_N^{\partial}(i;s)| d^{\lfloor s/3\rfloor} \big) \cdot \big(\sum_{q=1}^{\infty} d^{\lfloor q/3\rfloor} \big) \leq C \sum_{s=1}^{N} |\mathcal{N}_N^{\partial}(i;s)| d^{\lfloor s/3\rfloor}$ for some finite constant C. Similarly, we can show that

$$
\sum_{\tau\neq t}^{T} \mathrm{Cov}\big(Z_i \bar{P}_{i,t-1}^{\top}, Z_i \bar{P}_{i,\tau-1}^{\top}\big) \leq C\zeta_{1,J_1} J_1^{1/2} \sum_{s=1}^{T-1} d^{\lfloor s/3\rfloor} \leq C\zeta_{1,J_1} J_1^{1/2}, \qquad (10)
$$

and

$$
\sum_{j\neq i}^{N} \mathrm{Cov}\big(Z_i \bar{P}_{i,t-1}^{\top}, Z_j \bar{P}_{j,t-1}^{\top}\big) \leq C\zeta_{1,J_1} J_1^{1/2} \sum_{s=1}^{N} \big|\mathcal{N}_N^{\partial}(i;s)\big| d^{\lfloor s/3\rfloor}. \qquad (11)
$$

Combining the above results and recalling that $\delta_N^{\partial}(s) = N^{-1} \sum_{i=1}^{N} |\mathcal{N}_N^{\partial}(i;s)|$, we conclude that the second term on the r.h.s. of (6) is bounded by a multiple of

$$
\begin{aligned}
\frac{\zeta_{1,J_1} J_1^{1/2}}{(NT)^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{N} \min\{s,T\} |\mathcal{N}_N^{\partial}(i;s)| d^{\lfloor s/3\rfloor} &= \frac{\zeta_{1,J_1} J_1^{1/2}}{NT} \sum_{s=1}^{N} \min\{s,T\} \delta_N^{\partial}(s) d^{\lfloor s/3\rfloor} \\
&\leq \zeta_{1,J_1} J_1^{1/2} \min\left\{T^{-1}\Xi_{N,1}, \Xi_{N,0}\right\}.
\end{aligned}
$$

This result, together with (6) and (7) leads to $\mathbb{E}\|\widehat{C}_{ZJ,1} - C_{ZJ,1}\|^2 = O\big(J_1/(NT) + \zeta_{1,J_1} J_1^{1/2} \min\left\{T^{-1}\Xi_{N,1}, \Xi_{N,0}\right\}\big)$. Similarly, we can show that the same order holds for $\mathbb{E}\|\widehat{C}_{ZJ,2} - C_{ZJ,2}\|^2$ with J_1 replaced by J_2. Therefore, $\mathbb{E}\|\widehat{C}_{ZJ} - C_{ZJ}\|^2 = O(J/(NT) +$

$\zeta_{1,J} J^{1/2} \min\{T^{-1}\Xi_{N,1}, \Xi_{N,0}\})$ as $N \to \infty$. The desired result follows from Markov's inequality.

(ii) Next, for $\widehat{\Psi}_J$, it can be written as

$$\widehat{\Psi}_J = \begin{pmatrix} \widehat{\Psi}_{J,11} & \widehat{\Psi}_{J,12} \\ \widehat{\Psi}_{J,21} & \widehat{\Psi}_{J,22} \end{pmatrix} = \begin{pmatrix} \frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T \bar{P}_{i,t-1}\bar{P}_{i,t-1}^\top & \frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T \bar{P}_{i,t-1}P_{i,t-1}^\top \\ \frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T P_{i,t-1}\bar{P}_{i,t-1}^\top & \frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T P_{i,t-1}P_{i,t-1}^\top \end{pmatrix},$$

where $\widehat{\Psi}_{J,21} = \widehat{\Psi}_{J,12}^\top$, $\widehat{\Psi}_{J,11}$, $\widehat{\Psi}_{J,12}$ and $\widehat{\Psi}_{J,22}$ are $J_1 \times J_1$, $J_1 \times J_2$ and $J_2 \times J_2$ submatrices of $\widehat{\Psi}_J$, respectively. Let $\Psi_{J,11} = \mathbb{E}(\widehat{\Psi}_{J,11})$, $\Psi_{J,12} = \mathbb{E}(\widehat{\Psi}_{J,12})$ and $\Psi_{J,22} = \mathbb{E}(\widehat{\Psi}_{J,22})$.

First, from the definition of $\zeta_{0,J}$, $\|\bar{P}_{i,t-1}\|^2 = \|\sum_{j=1}^N w_{ij} p^{J_1}(Y_{j,t-1})\|^2 \leq \sum_{j=1}^N w_{ij}\|p^{J_1}(Y_{j,t-1})\|^2 \leq \sup_{1\leq i\leq N}\|p^{J_1}(Y_{i,t-1})\|^2 \leq C\zeta_{0,J_1}^2$, and similarly $\|\hat{\bar{P}}_{i,t-1}\|^2 \leq C\zeta_{0,J_1}^2$ a.s.. By the triangular and CS inequalities,

$$\begin{aligned}
&\mathbb{E}\left\|\bar{P}_{i,t-1}\bar{P}_{i,t-1}^\top - \hat{\bar{P}}_{i,t-s}^{t-1}\hat{\bar{P}}_{i,t-s}^{(t-1)\top}\right\|^2 \\
&\leq 2\mathbb{E}\left\|\bar{P}_{i,t-1}\big(\bar{P}_{i,t-1} - \hat{\bar{P}}_{i,t-s}^{t-1}\big)^\top\right\|^2 + 2\mathbb{E}\left\|\big(\bar{P}_{i,t-1} - \hat{\bar{P}}_{i,t-s}^{t-1}\big)\hat{\bar{P}}_{i,t-s}^{(t-1)\top}\right\|^2 \\
&= 2\mathbb{E}\big(\|\bar{P}_{i,t-1}\|^2 + \|\hat{\bar{P}}_{i,t-s}^{t-1}\|^2\big)\left\|\sum_{j=1}^N w_{ij}\big(p^{J_1}(Y_{j,t-1}) - p^{J_1}(\hat{Y}_{j,t-s}^{t-1})\big)\right\|^2 \\
&\leq C\zeta_{0,J_1}^2 \sum_{j=1}^N w_{ij}\mathbb{E}\left\|p^{J_1}(Y_{j,t-1}) - p^{J_1}(\hat{Y}_{j,t-s}^{t-1})\right\|^2 \\
&\leq C\zeta_{0,J_1}^2 \sup_{1\leq i\leq N}\mathbb{E}\left\|p^{J_1}(Y_{i,t-1}) - p^{J_1}(\hat{Y}_{i,t-s}^{t-1})\right\|^2 \\
&\leq C\zeta_{0,J_1}^2\zeta_{1,J_1}^2 \sup_{1\leq i\leq N}\mathbb{E}\big\|Y_{i,t-1} - \hat{Y}_{i,t-s}^{t-1}\big\|^2. \qquad (12)
\end{aligned}$$

We focus on the block $\widehat{\Psi}_{J,11}$, and the rest of the blocks can be treated similarly. With a decomposition,

$$\begin{aligned}
\mathbb{E}\left\|\widehat{\Psi}_{J,11} - \Psi_{J,11}\right\|^2 &= \mathbb{E}\left\|\frac{1}{NT}\sum_{(i,t)\in\mathcal{D}_{NT}}\big(\bar{P}_{i,t-1}\bar{P}_{i,t-1}^\top - \mathbb{E}\big(\bar{P}_{i,t-1}\bar{P}_{i,t-1}^\top\big)\big)\right\|^2 \qquad (13) \\
&= \frac{1}{(NT)^2}\sum_{(i,t)\in\mathcal{D}_{NT}}\mathbb{E}\left\|\bar{P}_{i,t-1}\bar{P}_{i,t-1}^\top - \mathbb{E}(\bar{P}_{i,t-1}\bar{P}_{i,t-1}^\top)\right\|^2 \\
&\quad + \frac{1}{(NT)^2}\sum_{(i,t)\in\mathcal{D}_{NT}}\sum_{(j,\tau)\in\mathcal{D}_{NT}\setminus(i,t)}\mathrm{Cov}\big(\bar{P}_{i,t-1}\bar{P}_{i,t-1}^\top, \bar{P}_{j,\tau-1}\bar{P}_{j,\tau-1}^\top\big).
\end{aligned}$$

By Assumption 4, we have the first term on the r.h.s. of (13) is bounded by

$$\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}\|\bar{P}_{i,t-1}\bar{P}_{i,t-1}^{\top}\|^2 \leq \frac{\zeta_{0,J_1}^2}{NT}\mathrm{tr}\big\{\mathbb{E}\big(\mathbb{P}_{t-1}^{J_1\top}W^{\top}W\mathbb{P}_{t-1}^{J_1}/N\big)\big\} = O\left(\frac{\zeta_{0,J_1}^2 J_1}{NT}\right). \tag{14}$$

The rest of the proof proceeds analogously to the proof of (9)-(11). For ease of comparability, we use the same notation as that used there. For all $(i,t) \in \mathcal{D}_{NT}$ and $m \in \mathbb{N}$, denote

$$\xi_{i,t-1}^m = \mathbb{E}\big[\bar{P}_{i,t-1}\bar{P}_{i,t-1}^{\top}|\mathcal{F}_{i,t-1}(m)\big], \quad \eta_{i,t-1}^m = \bar{P}_{i,t-1}\bar{P}_{i,t-1}^{\top} - \mathbb{E}\big[\bar{P}_{i,t-1}\bar{P}_{i,t-1}^{\top}|\mathcal{F}_{i,t-1}(m)\big].$$

By Jensen's inequality and Assumption 4, we have $\mathbb{E}\|\xi_{i,t-1}^m\|^2 \leq \mathbb{E}\big\|\bar{P}_{i,t-1}\bar{P}_{i,t-1}^{\top}\big\|^2 \lesssim J_1\zeta_{0,J_1}^2$, and $\mathbb{E}\|\eta_{i,t-1}^m\|^2 \leq 4\mathbb{E}\|\bar{P}_{i,t-1}\bar{P}_{i,t-1}^{\top}\|^2 \lesssim J_1\zeta_{0,J_1}^2$. Also, (4) and (12) imply that $\mathbb{E}\|\eta_{i,t-1}^m\|^2 \leq C\zeta_{0,J_1}^2\zeta_{1,J_1}^2\sup_{1\leq i\leq N}\mathbb{E}\big\|Y_{i,t-1} - \hat{Y}_{i,t-m-1}^{t-1}\big\|^2 \lesssim \zeta_{0,J_1}^2\zeta_{1,J_1}^2 d^{2m}$. Next, by decomposing $\bar{P}_{i,t-1}\bar{P}_{i,t-1}^{\top}$ and $\bar{P}_{j,\tau-1}\bar{P}_{j,\tau-1}^{\top}$, $(i,t) \neq (j,\tau)$ in the same way as (8), we have from above results that

$$\big|\mathrm{Cov}\big(\bar{P}_{i,t-1}\bar{P}_{i,t-1}^{\top}, \bar{P}_{j,\tau-1}\bar{P}_{j,\tau-1}^{\top}\big)\big| \leq CJ_1^{1/2}\zeta_{0,J_1}^2\zeta_{1,J_1}d^{\lfloor h/3\rfloor},$$

where $h = \Delta((i,t),(j,\tau))$. Therefore, using similar arguments as those used in part (i), we can easily show that the second term on the r.h.s. of (13) is $O\Big(J_1^{1/2}\zeta_{0,J_1}^2\zeta_{1,J_1}\min\{\Xi_{N,0}, \Xi_{N,1}/T\}\Big)$. Combining the above results, we conclude that $\mathbb{E}\big\|\widehat{\Psi}_{J,11} - \Psi_{J,11}\big\|^2 = O(\zeta_{0,J_1}^2(J_1/(NT) + \zeta_{1,J_1}J_1^{1/2}\min\{\Xi_{N,0}, T^{-1}\Xi_{N,1}\}))$.

The convergence rate of other blocks of the matrix $\widehat{\Psi}_J$ can be derived analogously. We finally have $\mathbb{E}\big\|\widehat{\Psi}_J - \Psi_J\big\|^2 = O(\zeta_{0,J}^2(J/(NT) + J^{1/2}\zeta_{1,J}\min\{\Xi_{N,0}, T^{-1}\Xi_{N,1}\}))$. By Markov's inequality, $\big\|\widehat{\Psi}_J - \Psi_J\big\| = O_p(\sqrt{r_{NT,J}})$. □

The following lemma establishes the convergence rates of matrices related to the variances of the sieve estimator. For proof of this, we introduce two facts concerning the norm of the subtraction of two matrices.

Fact 1. *(Fact A.2 in Hoshino (2022)) Let A, A', B, and B' be matrices of dimensions $k \times l$,*

$k \times l$, $k \times k$ and $k \times k$, respectively. Then,

$$\left\|A^\top BA - A'^\top B'A'\right\| \le \sigma^2_{max}(A)\|B - B'\| + \lambda_{max}(B')\|A - A'\|\left[\|A - A'\| + 2\sigma_{max}(A')\right],$$

where $\sigma_{\max}(A) = \sqrt{\lambda_{\max}(A^\top A)}$ denotes the largest singular value of the matrix A.

Fact 2. *(Fact A.3 in Hoshino (2022)) Let A and B be symmetric matrices with full rank and A^- and B^- be their generalized inverse matrices, respectively. Then, $\|A^- - B^-\| = [\lambda_{min}(A)\lambda_{min}(B)]^{-1}\|A - B\|$.*

Lemma 3. *Let $\widehat{\Sigma}_{NT,Z} = \Psi_{N,Z} - \widehat{C}_{NT,ZJ}\widehat{\Psi}^-_{NT,J}\widehat{C}^\top_{NT,ZJ}$ and $\widehat{\Sigma}_{NT,f} = \widehat{\Psi}_{NT,J} - \widehat{C}^\top_{NT,ZJ}\Psi^{-1}_{N,Z}\widehat{C}_{NT,ZJ}$. Under Assumptions 1-5 and 7, (i) $\|\widehat{\Sigma}^{-1}_{NT,Z} - \Sigma^{-1}_{N,Z}\| = O_p\big(\underline{\nu}^{-2}_{J_1J_2}r^{1/2}_{NT,J}\big)$; and (ii) $\|\widehat{\Sigma}^-_{NT,f} - \Sigma^{-1}_{N,f}\| = O_p\big(\underline{\nu}^{-2}_{J_1J_2}r^{1/2}_{NT,J}\big)$, where $\Sigma_{N,Z}$ and $\Sigma_{N,f}$ are defined in Assumption 7 and (5), respectively.*

Proof. (i) Note that $\|\widehat{\Sigma}_{NT,Z} - \Sigma_{N,Z}\| = \left\|\widehat{C}_{NT,ZJ}\widehat{\Psi}^-_{NT,J}\widehat{C}^\top_{NT,ZJ} - C_{N,ZJ}\Psi^{-1}_{N,J}C^\top_{N,ZJ}\right\|$. First, by Assumption 4, we have $\lambda_{\min}(\Psi_{N,J}) \ge \underline{\nu}_{J_1J_2} > 0$ and $\sigma_{\max}(C_{N,ZJ}) \le C < \infty$ for sufficiently large N. This, Lemma 2 and Assumption 5 together imply that $\sigma_{\max}(\widehat{C}_{NT,ZJ}) = O_p(1)$ and $\lambda_{\min}(\widehat{\Psi}_{NT,J}) \ge c\underline{\nu}_{J_1J_2} > 0$ for some $0 < c < \infty$ with probability approaching one (wpa1). Combining the latter result, Lemma 2(ii) and Fact 2, we obtain $\|\widehat{\Psi}^-_{NT,J} - \Psi^{-1}_{N,J}\| = O(\underline{\nu}^{-2}_{J_1J_2}r^{1/2}_{NT,J})$. An application of Lemma 2(i) and Fact 1 yields

$$\left\|\widehat{C}_{NT,ZJ}\widehat{\Psi}^-_{NT,J}\widehat{C}^\top_{NT,ZJ} - C_{N,ZJ}\Psi^{-1}_{N,J}C^\top_{N,ZJ}\right\| = O_p\big(\underline{\nu}^{-2}_{J_1J_2}r^{1/2}_{NT,J}\big) = o_p(1).$$

Assumption 7 shows that $\lambda_{\min}(\Sigma_{N,Z}) \ge C$ for some positive constant C for sufficiently large N. Hence, the above result implies that $\lambda_{\min}(\widehat{\Sigma}_{NT,Z}) \ge C'$ for some $0 < C' < \infty$ wpa1. Then, Lemma 3(i) directly follows from Fact 2.

(ii) Using Fact 1, Assumptions 3-4, Lemma 2 and triangular inequality, we have

$$\left\|\widehat{\Sigma}_{NT,f} - \Sigma_{N,f}\right\| \le \underbrace{\left\|\widehat{\Psi}_{NT,J} - \Psi_{N,J}\right\|}_{=O_p(\sqrt{r_{NT,J}})} + \underbrace{\left\|\widehat{C}^\top_{NT,ZJ}\Psi^{-1}_{N,Z}\widehat{C}_{NT,ZJ} - C^\top_{N,ZJ}\Psi^{-1}_{N,Z}C_{N,ZJ}\right\|}_{=O_p(\zeta^{-1}_{0,J}\sqrt{r_{NT,J}})} = o_p(1),$$

This and Assumption 4 imply that $\lambda_{\min}(\widehat{\Sigma}_{NT,f}) \ge C\underline{\nu}_{J_1J_2}$ for some $0 < C < \infty$ wpa1. Then, result (ii) follows from Fact 2. □

B Proof of Main Theorems

B.1 Proof of Proposition 1

For a random variable X, denote $\|X\|_r = [\mathbb{E}|X|^r]^{1/r}$, and for a random vector $Z = (Z_1, \ldots, Z_k)^\top$, denote $\|Z\|_{r,vec} = (\|Z_1\|_r, \ldots, \|Z_k\|_r)^\top$. We also adopt the same notation as in Lemma 1. For the Type I stationarity (fixed N), Proposition 1 follows from Theorem 1 of Debaly and Truquet (2021), provided that Assumptions A1-A3 therein are satisfied. For $\mathbf{y} = (y_1, \ldots, y_N)^\top \in \mathbb{R}^N$, let $\mathbf{F}(\mathbf{y}; \mathcal{E}_t) = WF_1(\mathbf{y}) + F_2(\mathbf{y}) + \mathbb{Z}\gamma + \mathcal{E}_t$. Recall that $f_1(0) = f_2(0) = 0$. By Assumption 2,

$$\begin{aligned}\mathbb{E}|\mathbf{F}(\mathbf{y}; \mathcal{E}_t)|_1 &\le |WF_1(\mathbf{y})|_1 + |F_2(\mathbf{y})|_1 + \mathbb{E}|\mathbb{Z}\gamma + \mathcal{E}_t|_1 \\ &\le \mathbf{1}_N^\top \big[W(|F_1(\mathbf{0})|_{vec} + \kappa_1|\mathbf{y}|_{vec}) + (|F_2(\mathbf{0})|_{vec} + \kappa_2|\mathbf{y}|_{vec})\big] + C_{\mathcal{E},1} < \infty\end{aligned}$$

where $C_{\mathcal{E},1} = \mathbb{E}|\mathbb{Z}\gamma + \mathcal{E}_t|_1 < \infty$. Moreover, for $\mathbf{y}, \mathbf{y}^* \in \mathbb{R}^N$,

$$\mathbb{E}|\mathbf{F}(\mathbf{y}; \mathcal{E}_t) - \mathbf{F}(\mathbf{y}^*; \mathcal{E}_t)|_{vec} = |W(F_1(\mathbf{y}) - F_1(\mathbf{y}^*)) + F_2(\mathbf{y}) - F_2(\mathbf{y}^*)|_{vec} \preceq G|\mathbf{y} - \mathbf{y}^*|_{vec},$$

where $G = \kappa_1 W + \kappa_2 I_N$, and $\rho(G) < 1$ with $\rho(\cdot)$ being the spectral radius. Therefore, by Theorem 1 of Debaly and Truquet (2021), $\{\mathbb{Y}_t : t \in \mathbb{Z}\}$ is a stationary and ergodic process with $\mathbb{E}|\mathbb{Y}_t|_1 < \infty$. Furthermore, $\|\mathbf{F}(\mathbf{y}; \mathcal{E}_t)\|_{a,vec} = \|W(F_1(\mathbf{y}) - F_1(\mathbf{0})) + F_2(\mathbf{y}) - F_2(\mathbf{0}) + \mathbb{Z}\gamma + \mathcal{E}_t\|_{a,vec} \preceq C_{0,a} + G|\mathbf{y}|_{vec}$, where $C_{0,a} = \|\mathbb{Z}\gamma + \mathcal{E}_t\|_{a,vec}$. By Theorem 1(ii) of Debaly and Truquet (2021), if $\mathbb{E}|\epsilon_{it}|^a < \infty$, we have $\mathbb{E}|\mathbb{Y}_t|_1^a < \infty$, which implies that $\max_{1 \le i \le N} \mathbb{E}|Y_{it}|^a < \infty$.

Next, consider the Type II stationarity ($N \to \infty$). Recall from Zhu et al. (2017) that $\mathcal{W} = \{\omega \in \mathbb{R}^\infty : \sum |\omega_i| < \infty\}$, where $\omega = (\omega_i \in \mathbb{R}^1 : 1 \le i \le \infty)^\top \in \mathbb{R}^\infty$. For each $\omega \in \mathcal{W}$, let $\mathbf{w}_N = (\omega_1, ..., \omega_N)^\top \in \mathbb{R}^N$ be the its truncated N-dimensional version. Note that

$$\mathbb{Y}_t = W(F_1(\mathbb{Y}_{t-1}) - F_1(\mathbf{0})) + F_2(\mathbb{Y}_{t-1}) - F_2(\mathbf{0}) + \mathbb{Z}\gamma + \mathcal{E}_t$$

since $f_1(0) = f_2(0) = 0$. By Assumption 2, $|\mathbb{Y}_t|_{vec} \preceq G|\mathbb{Y}_{t-1}|_{vec} + |\mathbb{Z}\gamma|_{vec} + |\mathcal{E}_t|_{vec}$. Then, by infinite backward substitution, we have

$$|\mathbb{Y}_t|_{vec} \preceq G|\mathbb{Y}_{t-1}|_{vec} + |\mathbb{Z}\gamma|_{vec} + |\mathcal{E}_t|_{vec} \preceq \sum_{j=0}^{\infty} G^j(|\mathbb{Z}\gamma|_{vec} + |\mathcal{E}_{t-j}|_{vec}). \tag{15}$$

Let $C_0 = \max_{i\geq 1} |Z_i^\top\gamma| + \mathbb{E}|\epsilon_{it}|$. Since $\mathbb{E}\big[|\mathbb{Z}\gamma|_{vec} + |\mathcal{E}_{t-j}|_{vec}\big] \preceq C_0\mathbf{1}_N$, and $|G|_{vec}^j\mathbf{1}_N = (\kappa_1 W + \kappa_2 I_N)^j\mathbf{1}_N = (\kappa_1+\kappa_2)^j\mathbf{1}_N$, we have $\mathbb{E}|\mathbf{w}_N^\top\mathbb{Y}_t| \leq \mathbb{E}[|\mathbf{w}_N|_{vec}^\top\sum_{j=0}^\infty G^j(|\mathbb{Z}\gamma|_{vec} + |\mathcal{E}_{t-j}|_{vec})] \leq |\mathbf{w}_N|_1 C_0\sum_{j=0}^\infty(\kappa_1+\kappa_2)^j < \infty$. This implies that $\lim_{N\to\infty}\mathbf{w}_N^\top\mathbb{Y}_t$ exists and is finite with probability 1. Denote $Y_t^\omega = \lim_{N\to\infty}\mathbf{w}_N^\top\mathbb{Y}_t$. Obviously, Y_t^ω is strictly stationary and therefore $\{\mathbb{Y}_t\}$ is strictly stationary according to Definition 1 of Zhu et al. (2017). Assume that $\{\tilde{\mathbb{Y}}_t\}$ is another strictly stationary solution to the NAR model (1) with finite first order moment. Using similar arguments as above, $\mathbb{E}|\mathbf{w}_N^\top(\mathbb{Y}_t - \tilde{\mathbb{Y}}_t)| \leq |\mathbf{w}_N|_{vec}^\top G\mathbb{E}|\mathbb{Y}_{t-1} - \tilde{\mathbb{Y}}_{t-1}|_{vec} = 0$ by infinite backward substitution, for any N and weight ω. Consequently, $Y_t^\omega = \tilde{Y}_t^\omega$ with probability one. Finally, observe that for $a \geq 1$, $\|Y_{it}\|_a \leq \|e_i^\top\sum_{j=0}^\infty G^j(|\mathbb{Z}\gamma|_{vec} + |\mathcal{E}_{t-j}|_{vec})\|_a \leq \sum_{j=0}^\infty(\kappa_1+\kappa_2)^j\max_{i\geq 1}\|Z_i^\top\gamma + \epsilon_{it}\|_a$, where e_i is an $N\times 1$ vector with the ith element being 1 and others being zero. It follows that the first ath moments of $\mathbb{Y}_t$ are uniformly bounded if $\mathbb{E}|\epsilon_{it}|^a < \infty$ and $\max_{i\geq 1}|Z_i|_1^a < \infty$. The second condition is implied by the compactness of $\mathcal{R}_Z$. □

B.2 Proof of Theorem 1

(i) Recall that $\boldsymbol{\mathcal{E}} = (\mathcal{E}_1^\top,\ldots,\mathcal{E}_T^\top)^\top$, $\mathbf{Z} = \mathbf{1}_T\otimes\mathbb{Z}$, $\mathbf{Q} = (\mathbb{Q}_0^\top,\ldots,\mathbb{Q}_{T-1}^\top)^\top$, $\mathbf{Y} = (\mathbb{Y}_1^\top,\ldots,\mathbb{Y}_T^\top)^\top$, and $\widehat{\Sigma}_{NT,Z} = \mathbf{Z}^\top\mathbf{M_Q}\mathbf{Z}/(NT)$. Decompose

$$\begin{aligned}
\widehat{\gamma} - \gamma &= (\mathbf{Z}^\top\mathbf{M_Q}\mathbf{Z})^{-1}\mathbf{Z}^\top\mathbf{M_Q}\mathbf{Y} - \gamma \\
&= \widehat{\Sigma}_{NT,Z}^{-1}\mathbf{Z}^\top\mathbf{M_Q}[\mathbf{G}(\mathbf{Y}_{-1}) - \mathbf{Q}\beta]/(NT) + \widehat{\Sigma}_{NT,Z}^{-1}\mathbf{Z}^\top\mathbf{M_Q}\boldsymbol{\mathcal{E}}/(NT) \\
&\equiv D_1 + D_2,
\end{aligned}$$

where $\beta = (\beta_1^\top,\beta_2^\top)^\top$ and $\mathbf{G}(\mathbf{Y}_{-1}) = (G(\mathbb{Y}_0)^\top,\ldots,G(\mathbb{Y}_{T-1})^\top)^\top$ with $G(\mathbb{Y}_t) = WF_1(\mathbb{Y}_t) + F_2(\mathbb{Y}_t)$.

As shown in the proof of Lemma 3(i), $\lambda_{\min}(\widehat{\Sigma}_{NT,Z}) \geq C$ for some $0 < C < \infty$ wpa1.

Since the eigenvalue of an idempotent matrix is either 1 or 0, we have

$$\begin{aligned}\|D_1\|^2 &= [\mathbf{G}(\mathbf{Y}_{-1}) - \mathbf{Q}\beta]^\top \mathbf{M}_\mathbf{Q}\mathbf{Z}\widehat{\Sigma}_{NT,Z}^{-2}\mathbf{Z}^\top \mathbf{M}_\mathbf{Q}[\mathbf{G}(\mathbf{Y}_{-1}) - \mathbf{Q}\beta]/(NT)^2 \\ &\le O_p(1)\cdot [\mathbf{G}(\mathbf{Y}_{-1}) - \mathbf{Q}\beta]^\top[\mathbf{G}(\mathbf{Y}_{-1}) - \mathbf{Q}\beta]/(NT) \\ &= O_p(1)\cdot \frac{1}{NT}\sum_{t=1}^T \Big\{\big\|W(F_1(\mathbb{Y}_{t-1}) - \mathbb{P}_{t-1}^{J_1}\beta_1)\big\|^2 + \big\|F_2(\mathbb{Y}_{t-1}) - \mathbb{P}_{t-1}^{J_2}\beta_2\big\|^2\Big\}.\end{aligned}$$

Let $u_1(y) = f_1(y) - p^{J_1}(y)^\top\beta_1$ and $u_{1,\omega}(y) = u_1(y)(1+|y|^2)^{-\omega/2}$. Recall that $|u_1|_{\infty,\omega} = \sup_y |u_{1,\omega}(y)|$. By Theorem 1, Assumptions 3 and 6, we have

$$\begin{aligned}\frac{1}{NT}\sum_{t=1}^T \big\|W(F_1(\mathbb{Y}_{t-1}) - \mathbb{P}_{t-1}^{J_1}\beta_1)\big\|^2 &= \frac{1}{NT}\sum_{t=1}^T\sum_{i=1}^N\left(\sum_{l=1}^N w_{il}u_1(Y_{l,t-1})\right)^2 \\ &= \frac{1}{NT}\sum_{t=1}^T\sum_{i=1}^N\left(\sum_{l=1}^N w_{il}u_{1,\omega}(Y_{l,t-1})(1+|Y_{l,t-1}|^2)^{\omega/2}\right)^2 \\ &\le \frac{|u_1|_{\infty,\omega}^2}{NT}\sum_{i=1}^N\sum_{t=1}^T\left(\sum_{l=1}^N w_{il}(1+|Y_{l,t-1}|^2)^{\omega}\right) = O_p(J_1^{-2\mu}),\end{aligned}$$

and similarly $(NT)^{-1}\sum_{t=1}^T \big\|F_2(\mathbb{Y}_{t-1}) - \mathbb{P}_{t-1}^{J_2}\beta_2\big\|^2 = O_p(J_2^{-2\mu})$. Hence, $\|D_1\| = O_p(J_1^{-\mu} + J_2^{-\mu})$.

For D_2, write $D_2 = D_{21} + D_{22}$, where $D_{21} = \Sigma_{N,Z}^{-1}\boldsymbol{\xi}^\top\boldsymbol{\mathcal{E}}/(NT)$, $D_{22} = \big(\widehat{\Sigma}_{NT,Z}^{-1}\mathbf{Z}^\top\mathbf{M}_\mathbf{Q} - \Sigma_{N,Z}^{-1}\boldsymbol{\xi}^\top\big)\boldsymbol{\mathcal{E}}/(NT)$, and $\boldsymbol{\xi} = \mathbf{Z} - \mathbf{Q}\Psi_{N,J}^{-1}C_{N,ZJ}^\top$. Noting that $\mathbb{Z}$ is nonrandom and $\mathbb{Q}_{t-1}$ only depends on $\{\mathcal{E}_s : s < t\}$, we have $\{\big(\mathbb{Z}^\top - C_{N,ZJ}\Psi_{N,J}^{-1}\mathbb{Q}_{t-1}^\top\big)\mathcal{E}_t, \mathcal{B}_t, t\ge 1\}$ forms a martingale difference sequence, where $\mathcal{B}_t$ is the σ-field generated by $\{\mathcal{E}_s : s \le t\}$. This implies that

$$\begin{aligned}\mathrm{Var}\left(\frac{1}{NT}\boldsymbol{\xi}^\top\boldsymbol{\mathcal{E}}\right) &= \mathrm{Var}\left(\frac{1}{NT}\sum_{t=1}^T\big(\mathbb{Z}^\top - C_{N,ZJ}\Psi_{N,J}^{-1}\mathbb{Q}_{t-1}^\top\big)\mathcal{E}_t\right) \\ &= \frac{1}{(NT)^2}\sum_{t=1}^T \mathrm{Var}\big(\big(\mathbb{Z}^\top - C_{N,ZJ}\Psi_{N,J}^{-1}\mathbb{Q}_{t-1}^\top\big)\mathcal{E}_t\big) = \sigma^2(NT)^{-1}\Sigma_{N,Z}.\end{aligned}$$

Hence, we have $\mathbb{E}\|D_{21}\|^2 = \sigma^2\mathrm{tr}(\Sigma_{N,Z}^{-1}\Sigma_{N,Z}\Sigma_{N,Z}^{-1}/(NT)) = O((NT)^{-1})$ by Assumption 7. It follows from Markov's inequality that $\|D_{21}\| = O_p((NT)^{-1/2})$.

For D_{22}, we decompose it into two terms as $D_{22} = \widehat{\Sigma}_{NT,Z}^{-1}(\mathbf{Z}^\top\mathbf{M}_\mathbf{Q}\boldsymbol{\mathcal{E}} - \boldsymbol{\xi}^\top\boldsymbol{\mathcal{E}})/(NT) - (\Sigma_{N,Z}^{-1} - \widehat{\Sigma}_{NT,Z}^{-1})\boldsymbol{\xi}^\top\boldsymbol{\mathcal{E}}/(NT)$. Using Lemmas 2, 3 and Assumption 4, we can easily show that $\|\widehat{C}_{NT,ZJ}\widehat{\Psi}_{NT,J}^{-1} - C_{N,ZJ}\Psi_{N,J}^{-1}\| = \|C_{N,ZJ}(\widehat{\Psi}_{NT,J}^{-1} - \Psi_{N,J}^{-1}) + (\widehat{C}_{NT,ZJ} - C_{N,ZJ})\widehat{\Psi}_{NT,J}^{-1}\| = O_p\big(\underline{\nu}_{J_1J_2}^{-2} r_{NT,J}^{1/2}\big) = o_p(1)$. In addition, it follows from Markov's inequality and Assumption 4 that $\|\mathbf{Q}^\top \boldsymbol{\mathcal{E}}/(NT)\| = O_p(\sqrt{J/(NT)})$. Thus,

$$\begin{aligned}\big\|\big(\mathbf{Z}^\top \mathbf{M}_{\mathbf{Q}}\boldsymbol{\mathcal{E}} - \boldsymbol{\xi}^\top \boldsymbol{\mathcal{E}}\big)/(NT)\big\| &= \big\|\big(\widehat{C}_{NT,ZJ}\widehat{\Psi}_{NT,J}^{-1} - C_{N,ZJ}\Psi_{N,J}^{-1}\big)\mathbf{Q}^\top \boldsymbol{\mathcal{E}}/(NT)\big\| \\ &= O_p\left(\underline{\nu}_{J_1J_2}^{-2}\sqrt{Jr_{NT,J}/(NT)}\right).\end{aligned}$$

This, Assumption 4 and Lemma 3(i) imply that

$$\begin{aligned}\|D_{22}\| &= \|\widehat{\Sigma}_{NT,Z}^{-1}(\mathbf{Z}^\top \mathbf{M}_{\mathbf{Q}}\boldsymbol{\mathcal{E}} - \boldsymbol{\xi}^\top \boldsymbol{\mathcal{E}})/(NT) - (\Sigma_{N,Z}^{-1} - \widehat{\Sigma}_{NT,Z}^{-1})\boldsymbol{\xi}^\top \boldsymbol{\mathcal{E}}/(NT)\| \\ &\leq \lambda_{\max}(\widehat{\Sigma}_{NT,Z}^{-1})\|(\mathbf{Z}^\top \mathbf{M}_{\mathbf{Q}}\boldsymbol{\mathcal{E}} - \boldsymbol{\xi}^\top \boldsymbol{\mathcal{E}})/(NT)\| + \|\widehat{\Sigma}_{NT,Z}^{-1} - \Sigma_{N,Z}^{-1}\|\|\boldsymbol{\xi}^\top \boldsymbol{\mathcal{E}}/(NT)\| \\ &= O_p\left(\underline{\nu}_{J_1J_2}^{-2}\sqrt{Jr_{NT,J}/(NT)}\right).\end{aligned}$$

Consequently, we have $D_2 = \Sigma_{N,Z}^{-1}\boldsymbol{\xi}^\top \boldsymbol{\mathcal{E}}/(NT) + o_p((NT)^{-1/2}) = O_p((NT)^{-1/2})$ by Assumption 5. This completes the proof of Theorem 1(i).

(ii) From the undersmoothing condition $(NT)^{1/2}(J_1^{-\mu} + J_2^{-\mu}) = o(1)$ and the fact that $\|D_{22}\| = o_p((NT)^{-1/2})$, we have $\sqrt{NT}(\widehat{\gamma} - \gamma) = \Sigma_{N,Z}^{-1}\boldsymbol{\xi}^\top \boldsymbol{\mathcal{E}}/\sqrt{NT} + o_p(1)$. Let $\boldsymbol{c}$ be an arbitrary $d_Z \times 1$ vector such that $\|\boldsymbol{c}\| = 1$. Denote $\Omega_{N,Z} = \sigma^2\Sigma_{N,Z}$ and

$$a_{NT} = \boldsymbol{c}^\top \Omega_{N,Z}^{-1/2}\boldsymbol{\xi}^\top \boldsymbol{\mathcal{E}}/\sqrt{NT}.$$

By construction, $\mathbb{E}(a_{NT}) = 0$ and $\mathrm{Var}(a_{NT}) = 1$. We aim to show that $a_{NT} \xrightarrow{d} N(0,1)$.

Define $k_N = NT$ and $a_{NT} = \sum_{v=1}^{k_N} X_{N,v}$, where for $t = 1, \ldots, T$, $i = 1, \ldots, N$,

$$X_{N,(t-1)N+i} = (NT)^{-1/2}\boldsymbol{c}^\top \Omega_{N,Z}^{-1/2}(Z_i - C_{N,ZJ}\Psi_{N,J}^{-1}Q_{i,t-1})\epsilon_{it}.$$

Define $\epsilon_{jt}^o = (\epsilon_{j1}, \ldots, \epsilon_{jt})^\top$ and the sub σ-fields $(i = 1, \ldots, N)$ of $\mathcal{F}$ as

$$
\begin{aligned}
\mathcal{F}_{N,i} &= \sigma(\{Y_{j0}\}_{j=1}^N, \{\epsilon_{j1}\}_{j=1}^i), \\
\mathcal{F}_{N,N+i} &= \sigma(\{Y_{j0}\}_{j=1}^N, \{\epsilon_{j1}^o\}_{j=1}^N, \{\epsilon_{j2}\}_{j=1}^i), \\
&\vdots \\
\mathcal{F}_{N,(T-1)N+i} &= \sigma(\{Y_{j0}\}_{j=1}^N, \{\epsilon_{j,T-1}^o\}_{j=1}^N, \{\epsilon_{jT}\}_{j=1}^i),
\end{aligned} \tag{16}
$$

with $\mathcal{F}_{N,0} = \sigma(\{Y_{j0}\}_{j=1}^N)$. Given the construction of the random variables $X_{N,v}$ and information sets $\mathcal{F}_{N,v}$ with $v = (t-1)N + i$, we have that $\mathcal{F}_{N,v} \subseteq \mathcal{F}_{N,v+1}$, that $X_{N,v}$ is $\mathcal{F}_{N,v}$-measurable, and that $\mathbb{E}[X_{N,v} \mid \mathcal{F}_{N,v-1}] = 0$ in light of Assumption 3. This establishes that $\{X_{N,v}, \mathcal{F}_{N,v}, 1 \leq v \leq NT, N \geq 1\}$ is a martingale difference array. Then, it suffices to check the following two conditions for the central limit theorem (CLT) of Kuersteiner and Prucha (2013, Theorem 1):

$$
\sum_{v=1}^{k_N} \mathbb{E}[|X_{N,v}|^{2+\delta}] \to 0 \tag{17}
$$

for some $\delta > 0$ and

$$
\sum_{v=1}^{k_N} \mathbb{E}[X_{N,v}^2 \mid \mathcal{F}_{N,v-1}] \xrightarrow{p} 1. \tag{18}
$$

For verification of condition (17), let $\delta = 2$ and $v = (t-1)N + i$. Decompose $X_{N,v} = X_{N1,v} - X_{N2,v}$, where $X_{N1,v} = \boldsymbol{c}^\top \Omega_{N,Z}^{-1/2} Z_i \epsilon_{it}/\sqrt{NT}$, and $X_{N2,v} = \boldsymbol{c}^\top \Omega_{N,Z}^{-1/2} C_{N,ZJ} \Psi_{N,J}^{-1} Q_{i,t-1} \epsilon_{it}/\sqrt{NT}$. By Assumptions 3 and 4, we have

$$
\begin{aligned}
\mathbb{E}\big(X_{N2,v}^4 \mid \mathcal{F}_{N,v-1}\big) &= \frac{1}{(NT)^2} \mathbb{E}\Big[\big(\boldsymbol{c}^\top \Omega_{N,Z}^{-1/2} C_{N,ZJ} \Psi_{N,J}^{-1} Q_{i,t-1} \epsilon_{it}\big)^4 \,\Big|\, \mathcal{F}_{N,v-1}\Big] \\
&= \frac{1}{(NT)^2} \big(\boldsymbol{c}^\top \Omega_{N,Z}^{-1/2} C_{N,ZJ} \Psi_{N,J}^{-1} Q_{i,t-1}\big)^4 \cdot \mathbb{E}(\epsilon_{it}^4) \\
&\leq \frac{C}{(NT)^2 \underline{\nu}_{J_1 J_2}^4} \|Q_{i,t-1}\|^4.
\end{aligned}
$$

Assumption 4(i) implies that $\mathbb{E}\|Q_{i,t-1}\|^2 = \mathbb{E}\|P_{i,t-1}\|^2 + \mathbb{E}\|\sum_{j=1}^N w_{ij} P_{j,t-1}\|^2 = O(J)$ and

$\max_{1\le i\le N}\|Q_{i,t-1}\|^2 = O(\zeta_{0,J}^2)$. This and the last display yield that

$$\sum_{v=1}^{k_N}\mathbb{E}\left[\mathbb{E}\big(X_{N2,v}^4 \mid \mathcal{F}_{N,v-1}\big)\right] \le \frac{C}{(NT)^2\underline{\nu}_{J_1J_2}^4}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}\|Q_{i,t-1}\|^4 = O\bigg(\frac{\zeta_{0,J}^2 J}{\underline{\nu}_{J_1J_2}^4 NT}\bigg).$$

Similarly, it is straightforward to show that $\sum_{v=1}^{k_N}\mathbb{E}\big(X_{N1,v}^4\big) = O(1/(NT)) = o(1)$. The desired result is obtained by the c_r-inequality: $\sum_{v=1}^{k_N}\mathbb{E}\big(X_{N,v}^4\big) = \sum_{v=1}^{k_N}\mathbb{E}|X_{N1,v} - X_{N2,v}|^4 \le 8\sum_{v=1}^{k_N}\mathbb{E}|X_{N1,v}|^4 + 8\sum_{v=1}^{k_N}\mathbb{E}|X_{N2,v}|^4 = o(1)$. This verifies condition (17).

Next, observe that

$$\sum_{v=1}^{k_N}\mathbb{E}\big[X_{N,v}^2|\mathcal{F}_{N,v-1}\big] = \frac{1}{NT}\sum_{t=1}^{T}\boldsymbol{c}^\top\Sigma_{N,Z}^{-1/2}(\mathbb{Z}^\top - C_{N,ZJ}\Psi_{N,J}^{-1}\mathbb{Q}_{t-1}^\top)(\mathbb{Z} - \mathbb{Q}_{t-1}\Psi_{N,J}^{-1}C_{N,ZJ}^\top)\Sigma_{N,Z}^{-1/2}\boldsymbol{c}.$$

Using Lemma 2, we can easily see that $\big\|1/(NT)\sum_{t=1}^{T}(\mathbb{Z}^\top - C_{N,ZJ}\Psi_{N,J}^{-1}\mathbb{Q}_{t-1}^\top)(\mathbb{Z} - \mathbb{Q}_{t-1}\Psi_{N,J}^{-1}C_{N,ZJ}^\top) - \Sigma_{N,Z}\big\| = o_p(1)$. This yields that $\sum_{v=1}^{k_N}\mathbb{E}\big[X_{N,v}^2|\mathcal{F}_{N,v-1}\big] \xrightarrow{p} \boldsymbol{c}^\top\Sigma_{N,Z}^{-1/2}\Sigma_{N,Z}\Sigma_{N,Z}^{-1/2}\boldsymbol{c} = 1$, which verifies condition (18). $\square$

B.3 Proof of Theorem 2

Lemma 4. *Suppose that Assumptions 1-5 hold. If, additionally, the conditions in Theorem 2(ii) hold, then for $\ell = 1, 2$,*

$$\Big|p^{J_\ell}(\cdot)^\top\mathcal{S}_\ell(\mathbf{Q}^\top\mathbf{M_Z}\mathbf{Q})^-\mathbf{Q}^\top\mathbf{M_Z}\boldsymbol{\mathcal{E}}\Big|_{\infty,\omega} = O_p\left(\zeta_{0,J_\ell}\sqrt{\frac{\ln J_\ell}{\underline{\nu}_{J_1J_2}NT}}\right), \tag{19}$$

where $\mathbf{Q} = (\mathbb{Q}_0^\top,\ldots,\mathbb{Q}_{T-1}^\top)^\top$ with $\mathbb{Q}_t = (W\mathbb{P}_t^{J_1},\mathbb{P}_t^{J_2})$, $\mathcal{S}_1 = (I_{J_1},\mathbf{0}_{J_1\times J_2})$ and $\mathcal{S}_2 = (\mathbf{0}_{J_2\times J_1}, I_{J_2})$.

Proof. Note that

$$\big(\mathbf{Q}^\top\mathbf{M_Z}\mathbf{Q}\big)^-\mathbf{Q}^\top\mathbf{M_Z}\boldsymbol{\mathcal{E}} = U_1 + U_2,$$

where

$$U_1 = \Sigma_{N,f}^{-1}\big(\mathbf{Q}^\top\boldsymbol{\mathcal{E}} - C_{N,ZJ}^\top\Psi_{N,Z}^{-1}\mathbf{Z}^\top\boldsymbol{\mathcal{E}}\big)/(NT), \tag{20}$$

$$U_2 = \big[(\widehat{\Sigma}_{NT,f}^- - \Sigma_{N,f}^{-1})\mathbf{Q}^\top\boldsymbol{\mathcal{E}} + \big(\widehat{\Sigma}_{NT,f}^-\widehat{C}_{NT,ZJ}^\top\Psi_{N,Z}^{-1} - \Sigma_{N,f}^{-1}C_{N,ZJ}^\top\Psi_{N,Z}^{-1}\big)\mathbf{Z}^\top\boldsymbol{\mathcal{E}}\big]/(NT), \tag{21}$$

with $\Sigma_{N,f} = \Psi_{N,J} - C_{N,ZJ}^\top \Psi_{N,Z}^{-1} C_{N,ZJ}$, and $\widehat{\Sigma}_{NT,f} = \mathbf{Q}^\top \mathbf{M}_\mathbf{Z} \mathbf{Q}/(NT)$.

We focus on the case where $\ell = 1$, and the other case can be proved similarly. Denote $\widetilde{U}_1 = \mathcal{S}_1 U_1$ and $\widetilde{U}_2 = \mathcal{S}_1 U_2$. We first examine $|p^{J_1}(\cdot)^\top \widetilde{U}_2|_{\infty,\omega}$. Notice that

$$\widetilde{U}_2 = \mathcal{S}_1(\widehat{\Sigma}_{NT,f}^{-} - \Sigma_{N,f}^{-1})\mathbf{H}^\top \boldsymbol{\mathcal{E}}/(NT) - \mathcal{S}_1 \widehat{\Sigma}_{NT,f}^{-}\big(\widehat{C}_{NT,ZJ}^\top \Psi_{N,Z}^{-1} - C_{N,ZJ}^\top \Psi_{N,Z}^{-1}\big)\mathbf{Z}^\top \boldsymbol{\mathcal{E}}/(NT),$$

where $\mathbf{H} \equiv (\mathbb{H}_0^\top, \ldots, \mathbb{H}_{T-1}^\top)^\top$ with $\mathbb{H}_{t-1} \equiv \mathbb{Q}_{t-1} - \mathbb{Z}\Psi_{N,Z}^{-1} C_{N,ZJ}$. First, it can be easily seen that $\mathbb{E}\|\mathbf{H}^\top \boldsymbol{\mathcal{E}}/(NT)\|^2 = (NT)^{-1}\mathrm{tr}(\mathbb{E}[\mathbb{H}_{t-1}^\top \mathcal{E}_t \mathcal{E}_t^\top \mathbb{H}_{t-1}/N]) = \sigma^2 (NT)^{-1}\mathrm{tr}(\Sigma_{N,f}) = O(J/(NT))$. It follows from Markov's inequality that $\|\mathbf{H}^\top \boldsymbol{\mathcal{E}}/(NT)\| = O_p(\sqrt{J/(NT)})$. According to Lemma 3(ii),

$$\big\|\mathcal{S}_1(\widehat{\Sigma}_{NT,f}^{-} - \Sigma_{N,f}^{-1})\mathbf{H}^\top \boldsymbol{\mathcal{E}}/(NT)\| \le \big\|\widehat{\Sigma}_{NT,f}^{-} - \Sigma_{N,f}^{-1}\big\|\big\|\mathbf{H}^\top \boldsymbol{\mathcal{E}}/(NT)\big\| = O_p\big(\underline{\nu}_{J_1 J_2}^{-2} r_{NT,J}^{1/2}\sqrt{J/(NT)}\big).$$

Besides, from Lemma 2(i) and the fact that $\|\mathbf{Z}^\top \boldsymbol{\mathcal{E}}/(NT)\| = O_p((NT)^{-1/2})$, we have

$$\left\|\mathcal{S}_1 \widehat{\Sigma}_{NT,f}^{-}(\widehat{C}_{NT,ZJ}\Psi_{N,Z}^{-1} - C_{N,ZJ}\Psi_{N,Z}^{-1})\mathbf{Z}^\top \boldsymbol{\mathcal{E}}/(NT)\right\| = O_p\left(\underline{\nu}_{J_1 J_2}^{-1}\zeta_{0,J}^{-1}\sqrt{r_{NT,J}/(NT)}\right).$$

A combination of above results yields $\|\widetilde{U}_2\| = O_p(\underline{\nu}_{J_1 J_2}^{-2} r_{NT,J}^{1/2}\sqrt{J/(NT)})$. As $\sup_{y \in \mathcal{R}_\mathcal{Y}} \|p^{J_1}(y)\| = O(\zeta_{0,J_1})$, we have

$$\big|p^{J_1}(\cdot)^\top \widetilde{U}_2\big|_{\infty,\omega} = O_p\left(\underline{\nu}_{J_1 J_2}^{-2}\zeta_{0,J_1} r_{NT,J}^{1/2}\sqrt{J/(NT)}\right) = O_p\left(\zeta_{0,J_1}\sqrt{\frac{\ln J_1}{\underline{\nu}_{J_1 J_2} NT}}\right). \qquad (22)$$

The second equality is due to the fact that $r_{NT,J} J/(\underline{\nu}_{J_1 J_2}^3 \ln J_1) = O(1)$ which is implied by Assumption 5.

Next, $\mathbb{E}\|\widetilde{U}_1\|^2 = \sigma^2 \mathrm{tr}(\mathcal{S}_1 \Sigma_{N,f}^{-1}\mathbb{E}[\mathbf{H}^\top \mathbf{H}/(NT)]\Sigma_{N,f}^{-1}\mathcal{S}_1^\top/(NT)) = \sigma^2 \mathrm{tr}(\mathcal{S}_1 \Sigma_{N,f}^{-1}\Sigma_{N,f}\Sigma_{N,f}^{-1}\mathcal{S}_1^\top)/(NT) = \sigma^2 \mathrm{tr}(\Sigma_{N,f_1}^{-1})/(NT) = O(J_1/(\underline{\nu}_{J_1 J_2} NT))$ by Assumption 4, whence $\|\widetilde{U}_1\| = O_p(\sqrt{J_1/(\underline{\nu}_{J_1 J_2} NT)})$. Let c_{NT} be a sequence of positive constants tending to 1 at the rate $c_{NT} = O(J_1^{1/(2\omega)})$, where ω is given in Assumption 3(i). Using similar arguments as Hoshino (2022, Lemma A.9), we have

$$\big|p^{J_1}(\cdot)^\top \widetilde{U}_1\big|_{\infty,\omega} \le \sup_{y \in \mathcal{R}_\mathcal{Y}} \big|p^{J_1}(y)^\top \widetilde{U}_1 \mathbf{1}(|y| \le c_{NT}) \cdot (1 + |y|^2)^{-\omega/2}\big|$$

$$
\begin{aligned}
&+ \sup_{y \in \mathcal{R}_{\mathcal{Y}}} \left| p^{J_1}(y)^\top \widetilde{U}_1 \mathbf{1}(|y| > c_{NT}) \cdot (1 + |y|^2)^{-\omega/2} \right| \\
&\leq \sup_{y \in \mathcal{R}_{\mathcal{Y}} : |y| \leq c_{NT}} \left| p^{J_1}(y)^\top \widetilde{U}_1 \right| + \sup_{y \in \mathcal{R}_{\mathcal{Y}}} \| p^{J_1}(y) \| \| \widetilde{U}_1 \| \cdot \sup_{y : |y| > c_{NT}} (1 + |y|^2)^{-\omega/2} \\
&= \sup_{y \in \mathcal{R}_{\mathcal{Y}} : |y| \leq c_{NT}} \left| p^{J_1}(y)^\top \widetilde{U}_1 \right| + \underbrace{O_p\left(\frac{\zeta_{0,J_1} \sqrt{J_1}}{c_{NT}^{\omega} \sqrt{\underline{\nu}_{J_1 J_2} NT}} \right)}_{= O_p(\zeta_{0,J_1} / \sqrt{\underline{\nu}_{J_1 J_2} NT})}. \qquad (23)
\end{aligned}
$$

Since the set $\{y \in \mathcal{R}_{\mathcal{Y}} : |y| \leq c_{NT}\}$ is compact, we can partition it into countably many sub-intervals. Let the set of the partitioning points (including the boundary points) be $\mathcal{T}_{NT}$. We can construct the partition that, for any $y \in \mathcal{R}_{\mathcal{Y}}$ satisfying $|y| \leq c_{NT}$, there exists a point $a_y \in \mathcal{T}_{NT}$ satisfying $|y - a_y| = O(J_1^{-\xi})$, where ξ is as given in Theorem 2(ii). Then, similar to Lemma A.9 of Hoshino (2022), we have

$$
\begin{aligned}
\sup_{y \in \mathcal{R}_{\mathcal{Y}} : |y| \leq c_{NT}} \left| p^{J_1}(y)^\top \widetilde{U}_1 \right| &\leq \max_{a \in \mathcal{T}_{NT}} |p^{J_1}(a)^\top \widetilde{U}_1| + \max_{y \in \mathcal{R}_{\mathcal{Y}} : |y| \leq c_{NT}} |(p^{J_1}(y) - p^{J_1}(a_y))^\top \widetilde{U}_1| \\
&\leq \max_{a \in \mathcal{T}_{NT}} |p^{J_1}(a)^\top \widetilde{U}_1| + O_p(1) \cdot \|\widetilde{U}_1\| \\
&= \max_{a \in \mathcal{T}_{NT}} |p^{J_1}(a)^\top \widetilde{U}_1| + O_p\left(\sqrt{\frac{J_1}{\underline{\nu}_{J_1 J_2} NT}} \right). \qquad (24)
\end{aligned}
$$

To derive the bound of the first term of (24), we decompose $\epsilon_{it} = \epsilon_{1,it} + \epsilon_{2,it}$, where

$$
\begin{aligned}
\epsilon_{1,it} &= \epsilon_{it} \mathbf{1}\{|\epsilon_{it}| \leq M_{NT}\} - \mathbb{E}[\epsilon_{it} \mathbf{1}\{|\epsilon_{it}| \leq M_{NT}\}], \\
\epsilon_{2,it} &= \epsilon_{it} \mathbf{1}\{|\epsilon_{it}| > M_{NT}\} - \mathbb{E}[\epsilon_{it} \mathbf{1}\{|\epsilon_{it}| > M_{NT}\}],
\end{aligned}
$$

and M_{NT} is a sequence of positive constants tending to ∞. Let $\mathcal{E}_{1,t} = (\epsilon_{1,1t}, \ldots, \epsilon_{1,Nt})^\top$, $\mathcal{E}_{2,t} = (\epsilon_{2,1t}, \ldots, \epsilon_{2,Nt})^\top$, $\boldsymbol{\mathcal{E}}_1 = (\mathcal{E}_{1,1}^\top, \ldots, \mathcal{E}_{1,T}^\top)^\top$, and $\boldsymbol{\mathcal{E}}_2 = (\mathcal{E}_{2,1}^\top, \ldots, \mathcal{E}_{2,T}^\top)^\top$, so that

$$
\widetilde{U}_1 = \mathcal{S}_1 \Sigma_{N,f}^{-1} \mathbf{H}^\top (\boldsymbol{\mathcal{E}}_1 + \boldsymbol{\mathcal{E}}_2)/(NT).
$$

Let $q_{N,i,t-1}(a) \equiv p^{J_1}(a)^\top \mathcal{S}_1 \Sigma_{N,f}^{-1} (Q_{i,t-1} - C_{N,ZJ}^\top \Psi_{N,Z}^{-1} Z_i)/(NT)$, and $\xi_{N,v}(a) \equiv q_{N,i,t-1}(a) \epsilon_{1,it}$, where $v = (t-1)N + i$. Then,

$$\sum_{v=1}^{k_N} \xi_{N,v}(a) = \sum_{i=1}^{N}\sum_{t=1}^{T} q_{N,i,t-1}(a)\epsilon_{1,it} = p^{J_1}(a)^\top \mathcal{S}_1 \Sigma_{N,f}^{-1}\mathbf{H}^\top \boldsymbol{\mathcal{E}}_1/(NT) \tag{25}$$

for $k_N \equiv NT$. Note that $\mathbb{E}[\xi_{N,v}(a) \mid \mathcal{F}_{N,v-1}] = 0$, where $\{\mathcal{F}_{N,v}, v \geq 1\}$ is a sequence of σ-fields defined in (16). We now apply the concentration inequality for martingale sequences to obtain the bound for $\max_{a\in\mathcal{T}_{NT}} |\sum_{v=1}^{k_N} \xi_{N,v}(a)|$. First, it is straightforward to see that there exists a positive constant c_1 such that

$$|\xi_{N,v}(a)| = |q_{N,i,t-1}(a)\epsilon_{1,it}| \leq c_1\zeta_{0,J_1}\zeta_{0,J} M_{NT}/(\underline{\nu}_{J_1J_2}NT).$$

Besides, denoting $\sigma_1^2 \equiv \mathrm{Var}(\epsilon_{1,it})$, we have $\sum_{v=1}^{k_N} \mathbb{E}[|\xi_{N,v}(a)|^2|\mathcal{F}_{N,v-1}] = \sigma_1^2 \sum_{i=1}^{N}\sum_{t=1}^{T} |q_{N,i,t-1}(a)|^2 = \sigma_1^2 p^{J_1}(a)^\top \mathcal{S}_1\Sigma_{N,f}^{-1}\mathbf{H}^\top\mathbf{H}\Sigma_{N,f}^{-1}\mathcal{S}_1^\top p^{J_1}(a)/(NT)^2$. According to Lemma 2, $\|(NT)^{-1}\mathbf{H}^\top\mathbf{H} - \Sigma_{N,f}\| = O_p(r_{NT,J}^{1/2}) = o_p(1)$. It follows that $|\sum_{v=1}^{k_N}\mathbb{E}[|\xi_{N,v}(a)|^2|\mathcal{F}_{N,v-1}] - \sigma_1^2 p^{J_1}(a)^\top\Sigma_{N,f_1}^{-1}p^{J_1}(a)/(NT)| = (NT)^{-1}\sigma_1^2\big|p^{J_1}(a)^\top\mathcal{S}_1\Sigma_{N,f}^{-1}[(NT)^{-1}\mathbf{H}^\top\mathbf{H} - \Sigma_{N,f}]\Sigma_{N,f}^{-1}\mathcal{S}_1^\top p^{J_1}(a)\big| \leq C\underline{\nu}_{J_1J_2}^{-2}\|p^{J_1}(a)\|^2\|(NT)^{-1}\mathbf{H}^\top\mathbf{H} - \Sigma_{N,f}\|/(NT) = O_p(\underline{\nu}_{J_1J_2}^{-2}\zeta_{0,J_1}^2 r_{NT,J}^{1/2}/(NT))$. By the triangle inequality, we have

$$\begin{aligned}
\max_{a\in\mathcal{T}_{NT}} \sum_{v=1}^{k_N}\mathbb{E}[|\xi_{N,v}(a)|^2|\mathcal{F}_{N,v-1}] &\leq \max_{a\in\mathcal{T}_{NT}}\left|\sum_{v=1}^{k_N}\mathbb{E}[|\xi_{N,v}(a)|^2|\mathcal{F}_{N,v-1}] - \sigma_1^2 p^{J_1}(a)^\top\Sigma_{N,f_1}^{-1}p^{J_1}(a)/(NT)\right| \\
&\quad + \sigma_1^2 \max_{a\in\mathcal{T}_{NT}} p^{J_1}(a)^\top\Sigma_{N,f_1}^{-1}p^{J_1}(a)/(NT) \\
&= O_p(\zeta_{0,J_1}^2 r_{NT,J}^{1/2}/(\underline{\nu}_{J_1J_2}^2 NT)) + O_p(\zeta_{0,J_1}^2/(\underline{\nu}_{J_1J_2}NT)) \\
&= O_p(\zeta_{0,J_1}^2/(\underline{\nu}_{J_1J_2}NT))
\end{aligned}$$

since $r_{NT,J}^{1/2}/\underline{\nu}_{J_1J_2} \to 0$ as $N \to \infty$ by Assumption 5. This implies that for any $\eta > 0$, there exists some $L_1 \equiv L_{1,\eta} > 0$ such that $\mathbb{P}\big\{\max_{a\in\mathcal{T}_{NT}}\sum_{v=1}^{k_N}\mathbb{E}[|\xi_{N,v}(a)|^2|\mathcal{F}_{N,v-1}] > L_1\zeta_{0,J_1}^2/(NT\underline{\nu}_{J_1J_2})\big\} < \eta$ for sufficiently large N. Using Bernstein's inequality for martingale difference sequence, see, e.g., Freedman (1975) and Bercu et al. (2015, Theorem 3.14), we have

$$
\begin{aligned}
&\mathbb{P}\bigg(\bigg|\sum_{v=1}^{k_N}\xi_{N,v}(a)\bigg| > e_{NT},\ \sum_{v=1}^{k_N}\mathbb{E}\big[\xi_{N,v}^2(a)|\mathcal{F}_{N,v-1}\big] \le L_1\zeta_{0,J_1}^2/(\underline{\nu}_{J_1J_2}NT)\bigg)\\
&\le 2\exp\left(-\frac{e_{NT}^2}{2\big(L_1\zeta_{0,J_1}^2/(\underline{\nu}_{J_1J_2}NT)+c_1\zeta_{0,J}\zeta_{0,J_1}M_{NT}e_{NT}/(\underline{\nu}_{J_1J_2}NT)\big)}\right)\\
&\le 2\exp\left(-\frac{e_{NT}^2}{\tilde{L}_1(\zeta_{0,J_1}^2/(\underline{\nu}_{J_1J_2}NT))(1+e_{NT}M_{NT}\zeta_{0,J}/\zeta_{0,J_1})}\right)
\end{aligned}
$$

for some positive constant $\tilde{L}_1$ (depending on η). Let $\mathscr{V}_{NT}$ denote the event $\big\{\max_{a\in\mathcal{T}_{NT}}\sum_{v=1}^{k_N}\mathbb{E}[|\xi_{N,v}(a)|^2|\mathcal{F}_{N,v-1}] > L_1\zeta_{0,J_1}^2/(NT\underline{\nu}_{J_1J_2})\big\}$. Then,

$$
\begin{aligned}
&\mathbb{P}\bigg(\max_{a\in\mathcal{T}_{NT}}\bigg|\sum_{v=1}^{k_N}\xi_{N,v}(a)\bigg| > e_{NT}\bigg)\\
&\le \mathbb{P}\bigg(\max_{a\in\mathcal{T}_{NT}}\bigg|\sum_{v=1}^{k_N}\xi_{N,v}(a)\bigg| > e_{NT},\ \max_{a\in\mathcal{T}_{NT}}\sum_{v=1}^{k_N}\mathbb{E}\big[|\xi_{N,v}(a)|^2|\mathcal{F}_{N,v-1}\big] \le \frac{L_1\zeta_{0,J_1}^2}{\underline{\nu}_{J_1J_2}NT}\bigg)+\mathbb{P}(\mathscr{V}_{NT})\\
&\le \sum_{a\in\mathcal{T}_{NT}}\mathbb{P}\bigg(\bigg|\sum_{v=1}^{k_N}\xi_{N,v}(a)\bigg| > e_{NT},\ \max_{a\in\mathcal{T}_{NT}}\sum_{v=1}^{k_N}\mathbb{E}\big[|\xi_{N,v}(a)|^2|\mathcal{F}_{N,v-1}\big] \le \frac{L_1\zeta_{0,J_1}^2}{\underline{\nu}_{J_1J_2}NT}\bigg)+\mathbb{P}(\mathscr{V}_{NT})\\
&\le \sum_{a\in\mathcal{T}_{NT}}\mathbb{P}\left(\bigg|\sum_{v=1}^{k_N}\xi_{N,v}(a)\bigg| > e_{NT},\ \sum_{v=1}^{k_N}\mathbb{E}\big[|\xi_{N,v}(a)|^2|\mathcal{F}_{N,v-1}\big] \le \frac{L_1\zeta_{0,J_1}^2}{\underline{\nu}_{J_1J_2}NT}\right)+\mathbb{P}(\mathscr{V}_{NT})\\
&\le 2|\mathcal{T}_{NT}|\exp\left(-\frac{e_{NT}^2}{\tilde{L}_1(\zeta_{0,J_1}^2/(\underline{\nu}_{J_1J_2}NT))(1+e_{NT}M_{NT}\zeta_{0,J}/\zeta_{0,J_1})}\right)+\mathbb{P}(\mathscr{V}_{NT}),
\end{aligned}
$$

where $|\mathcal{T}_{NT}|$ denotes the cardinality of the set $\mathcal{T}_{NT}$. Then, setting $e_{NT} = L_2\zeta_{0,J_1}\sqrt{\ln J_1/(NT\underline{\nu}_{J_1J_2})}$ for a large constant $L_2 > 0$, provided that M_{NT} grows sufficiently slowly so that $e_{NT}M_{NT}\zeta_{0,J}/\zeta_{0,J_1} = o(1)$, we have $\ln|\mathcal{T}_{NT}| - e_{NT}^2/[\tilde{L}_1(\zeta_{0,J_1}^2/(\underline{\nu}_{J_1J_2}NT))(1+e_{NT}M_{NT}\zeta_{0,J}/\zeta_{0,J_1})] \asymp \ln\big(|\mathcal{T}_{NT}|/J_1^{L_2^2}\big)$. Recall that since the length of $\mathcal{T}_{NT}$ is of order $O(J_1^{1/(2\omega)})$ and that of each sub-interval is $O(J_1^{-\xi})$, the cardinality of $\mathcal{T}_{NT}$ grows at the rate $J_1^{1/(2\omega)+\xi}$. Thus, for sufficiently large L_2, we have $|\mathcal{T}_{NT}|/J_1^{L_2^2} \to 0$. Also, for any $\eta > 0$, $\mathbb{P}(\mathscr{V}_{NT}) < \eta$ for sufficiently large N. Combining the above results, we have

$$
\max_{a\in\mathcal{T}_{NT}}\bigg|\sum_{v=1}^{k_N}\xi_{N,v}(a)\bigg| = \max_{a\in\mathcal{T}_{NT}}\Big|p^{J_1}(a)^\top\mathcal{S}_1\Sigma_{N,f}^{-1}\mathbf{H}^\top\boldsymbol{\mathcal{E}}_1/(NT)\Big| = O_p\left(\frac{\zeta_{0,J_1}\sqrt{\ln J_1}}{\sqrt{\underline{\nu}_{J_1J_2}NT}}\right). \tag{26}
$$

Next, by Markov's inequality and Assumption 1, it holds that

$$
\begin{aligned}
\mathbb{P}\left(\max_{a\in\mathcal{T}_{NT}}\left|p^{J_1}(a)^\top\mathcal{S}_1\Sigma_{N,f}^{-1}\mathbf{H}^\top\boldsymbol{\mathcal{E}}_2/(NT)\right|>e_{NT}\right) &= \mathbb{P}\left(\max_{a\in\mathcal{T}_{NT}}\left|\sum_{i=1}^N\sum_{t=1}^T q_{N,i,t-1}(a)\epsilon_{2,it}\right|>e_{NT}\right)\\
&\le \mathbb{P}\left(c_1\zeta_{0,J_1}\zeta_{0,J}/(\underline{\nu}_{J_1J_2}NT)\sum_{i=1}^N\sum_{t=1}^T|\epsilon_{2,it}|>e_{NT}\right)\\
&\le \frac{2c_1\zeta_{0,J_1}\zeta_{0,J}}{\underline{\nu}_{J_1J_2}NTe_{NT}}\sum_{i=1}^N\sum_{t=1}^T\mathbb{E}[|\epsilon_{it}|\mathbf{1}\{|\epsilon_{it}|>M_{NT}\}]\\
&\le \frac{2c_1\zeta_{0,J_1}\zeta_{0,J}}{\underline{\nu}_{J_1J_2}NTe_{NT}M_{NT}^3}\sum_{i=1}^N\sum_{t=1}^T\mathbb{E}[|\epsilon_{it}|^4\mathbf{1}\{|\epsilon_{it}|>M_{NT}\}]\\
&= O\left(\frac{\zeta_{0,J_1}\zeta_{0,J}}{\underline{\nu}_{J_1J_2}e_{NT}M_{NT}^3}\right).
\end{aligned}
$$

Recall that $e_{NT}=L_2\zeta_{0,J_1}\sqrt{\ln J_1/(NT\underline{\nu}_{J_1J_2})}$. If $\zeta_{0,J}\sqrt{NT}/\sqrt{\underline{\nu}_{J_1J_2}\ln J_1}=O(M_{NT}^3)$, then $\zeta_{0,J_1}\zeta_{0,J}/(\underline{\nu}_{J_1J_2}e_{NT}M_{NT}^3)=\zeta_{0,J}\sqrt{NT}/[L_2\sqrt{\underline{\nu}_{J_1J_2}\ln J_1}M_{NT}^3]\to 0$ as $L_2\to\infty$, which implies that

$$
\max_{a\in\mathcal{T}_{NT}}\left|p^{J_1}(a)^\top\mathcal{S}_1\Sigma_{N,f}^{-1}\mathbf{H}^\top\boldsymbol{\mathcal{E}}_2/(NT)\right|=O_p\left(\frac{\zeta_{0,J_1}\sqrt{\ln J_1}}{\sqrt{\underline{\nu}_{J_1J_2}NT}}\right). \tag{27}
$$

It should be noted that the two requirements on M_{NT}, namely, $e_{NT}M_{NT}\zeta_{0,J}/\zeta_{0,J_1}=o(1)$ and $\zeta_{0,J}\sqrt{NT}/\sqrt{\underline{\nu}_{J_1J_2}\ln J_1}=O(M_{NT}^3)$, can be satisfied simultaneously under the second assumption of Theorem 2(ii). A combination of (23)-(27) yields

$$
\left|p^{J_1}(\cdot)^\top\widetilde{U}_1\right|_{\infty,\omega}=O_p\left(\frac{\zeta_{0,J_1}\sqrt{\ln J_1}}{\sqrt{\underline{\nu}_{J_1J_2}NT}}\right)+O_p\left(\frac{\zeta_{0,J_1}+J_1^{1/2}}{\sqrt{\underline{\nu}_{J_1J_2}NT}}\right). \tag{28}
$$

Lemma 4 with $\ell=1$ immediately follows from (22) and (28). The case where $\ell=2$ can be similarly derived, and we omit the details. □

Proof of Theorem 2.

(i) Recall that $\mathbf{Q}=(\mathbb{Q}_0^\top,\ldots,\mathbb{Q}_{T-1}^\top)^\top$, $\widehat{\Sigma}_{NT,f}=\mathbf{Q}^\top\mathbf{M_Z}\mathbf{Q}/(NT)$ and $\mathbf{G}(\mathbf{Y}_{-1})\equiv(G(\mathbb{Y}_0)^\top,\ldots,G(\mathbb{Y}_{T-1})^\top)^\top$ with $G(\mathbb{Y}_t)\equiv WF_1(\mathbb{Y}_t)+F_2(\mathbb{Y}_t)$. By the formula for partitioned regression,

$$
\widehat{\beta}-\beta=\left[\mathbf{Q}^\top\mathbf{M_Z}\mathbf{Q}\right]^-\mathbf{Q}^\top\mathbf{M_Z}\mathbf{Y}-\beta=U_1+U_2+U_3, \tag{29}
$$

where U_1 and U_2 are defined in (20) and (21), and $U_3 \equiv \widehat{\Sigma}_{NT,f}^{-}\mathbf{Q}^\top \mathbf{M}_\mathbf{Z}[\mathbf{G}(\mathbf{Y}_{-1}) - \mathbf{Q}\beta]/(NT)$. It is shown in the proof of Lemma 4 that $\|\mathcal{S}_1 U_1\| = O_p(\sqrt{J_1/(\underline{\nu}_{J_1J_2}NT)})$ and $\|\mathcal{S}_1 U_2\| = O_p(\underline{\nu}_{J_1J_2}^{-2} r_{NT,J}^{1/2}\sqrt{J/(NT)}) = O_p(\sqrt{J_1/(\underline{\nu}_{J_1J_2}NT)})$ under the stated conditions. We also have that the bias term $\|U_3\| = O_p(\underline{\nu}_{J_1J_2}^{-1/2}(J_1^{-\mu} + J_2^{-\mu}))$. To see this, note that according to Lemma 3(ii), $\|\widehat{\Sigma}_{NT,f} - \Sigma_{N,f}\| = O_p(r_{NT,J}^{1/2}) = o_p(1)$, which implies that $\lambda_{\min}(\widehat{\Sigma}_{NT,f}) \geq c\underline{\nu}_{J_1J_2} > 0$ for some $0 < c < \infty$ wpa1. Thus,

$$
\begin{aligned}
\|U_3\|^2 &\leq O_p(\underline{\nu}_{J_1J_2}^{-1}) \cdot [\mathbf{G}(\mathbf{Y}_{-1}) - \mathbf{Q}\beta]^\top \mathbf{M}_\mathbf{Z}\mathbf{Q}\widehat{\Sigma}_{NT,f}^{-}\mathbf{Q}^\top \mathbf{M}_\mathbf{Z}[\mathbf{G}(\mathbf{Y}_{-1}) - \mathbf{Q}\beta]/(NT)^2 \\
&\leq O_p(\underline{\nu}_{J_1J_2}^{-1}) \cdot \left\|\mathbf{G}(\mathbf{Y}_{-1}) - \mathbf{Q}\beta\right\|^2/(NT) = O_p\big(\underline{\nu}_{J_1J_2}^{-1}(J_1^{-2\mu} + J_2^{-2\mu})\big). \qquad (30)
\end{aligned}
$$

We conclude that $\|\widehat{\beta}_1 - \beta_1\| = \|\mathcal{S}_1(\widehat{\beta} - \beta)\| = O_p((J_1/(\underline{\nu}_{J_1J_2}NT))^{1/2} + \underline{\nu}_{J_1J_2}^{-1/2}(J_1^{-\mu} + J_2^{-\mu}))$.

Let $u_1(y) = f_1(y) - p^{J_1}(y)^\top\beta_1$ and $F_i(\cdot)$ be the c.d.f. of $\{Y_{it}\}$, $i = 1, \ldots, N$. Then,

$$
\begin{aligned}
&\int \left|\widehat{f}_1(y) - f_1(y)\right|^2 dF_i(y) \\
&= \int \left|p^{J_1}(y)^\top(\widehat{\beta}_1 - \beta_1) - \big(f_1(y) - p^{J_1}(y)^\top\beta_1\big)\right|^2 dF_i(y) \\
&\leq 2\int \left|p^{J_1}(y)^\top(\widehat{\beta}_1 - \beta_1)\right|^2 dF_i(y) + 2\int \left|f_1(y) - p^{J_1}(y)^\top\beta_1\right|^2 dF_i(y) \\
&\leq 2\mathbb{E}\Big[p^{J_1}(Y_{it})p^{J_1}(Y_{it})^\top\Big]\|\widehat{\beta}_1 - \beta_1\|^2 + 2|u_1|_{\infty,\omega}^2 \mathbb{E}[(1 + |Y_{it}|^2)^\omega] \\
&= O_p(J_1/(\underline{\nu}_{J_1J_2}NT) + \underline{\nu}_{J_1J_2}^{-1}(J_1^{-2\mu} + J_2^{-2\mu})),
\end{aligned}
$$

where the last equality follows from Assumptions 3(i), 6 and the condition that $\max_{1\leq i\leq N}\lambda_{\max}(\mathbb{E}[p^{J_1}(Y_{it})p^{J_1}(Y_{it})^\top]) < \infty$. The mean squared error of $\widehat{f}_2$ can be obtained similarly, and the details are omitted.

(ii) From (29), Lemma 4 and Assumption 6, we have

$$
\begin{aligned}
\left|\widehat{f}_1(\cdot) - f_1(\cdot)\right|_{\infty,\omega} &\leq \left|p^{J_1}(\cdot)^\top(\widehat{\beta}_1 - \beta_1)\right|_{\infty,\omega} + \left|p^{J_1}(\cdot)^\top\beta_1 - f_1(\cdot)\right|_{\infty,\omega} \\
&\leq \left|p^{J_1}(\cdot)^\top\mathcal{S}_1(U_1 + U_2)\right|_{\infty,\omega} + \left|p^{J_1}(\cdot)^\top\mathcal{S}_1U_3\right|_{\infty,\omega} + O(J_1^{-\mu}) \\
&= O_p\left(\zeta_{0,J_1}\sqrt{\ln J_1/(\underline{\nu}_{J_1J_2}NT)}\right) + \left|p^{J_1}(\cdot)^\top\mathcal{S}_1U_3\right|_{\infty,\omega} + O(J_1^{-\mu}).
\end{aligned}
$$

Besides, $\left|p^{J_1}(\cdot)^\top\mathcal{S}_1U_3\right|_{\infty,\omega} \leq \left|p^{J_1}(\cdot)\right|_\infty\|U_3\| = O_p(\zeta_{0,J_1}\underline{\nu}_{J_1J_2}^{-1/2}(J_1^{-\mu} + J_2^{-\mu}))$. Then the result for

$\ell = 1$ follows. The uniform convergence rate of $\widehat{f}_2$ can be shown similarly, and the details are omitted.

(iii) Note that $\Psi_{N,J} - \Sigma_{N,f}$ is non-negative definite and $\lambda_{\max}(\Psi_{N,J}) \leq \lambda_{\max}(\Sigma_{N,J_1J_2}) \leq \bar{c} < \infty$. By Assumption 4, we have $\lambda_{\max}(\Sigma_{N,f}) \leq \bar{c} < \infty$, and thus $v_{1N}^2(y) = \sigma^2 p^{J_1}(y)^\top \mathcal{S}_1 \Sigma_{N,f}^{-1} \mathcal{S}_1^\top p^{J_1}(y) \geq \sigma^2 \|p^{J_1}(y)\|^2 \lambda_{\min}(\mathcal{S}_1 \Sigma_{N,f}^{-1} \mathcal{S}_1^\top) \geq \sigma^2 \|p^{J_1}(y)\|^2 \lambda_{\min}(\Sigma_{N,f}^{-1}) \geq \sigma^2 \bar{c}^{-1} \|p^{J_1}(y)\|^2 > 0$. From the proof of part (i), we have

$$\begin{aligned}
&\frac{\sqrt{NT}}{v_{1N}(y)} \big(\widehat{f}_1(y) - f_1(y)\big) \\
&= \frac{\sqrt{NT}}{v_{1N}(y)} p^{J_1}(y)^\top \mathcal{S}_1 (U_1 + U_2 + U_3) + \frac{\sqrt{NT}}{v_{1N}(y)} \big(p^{J_1}(y)^\top \beta_1 - f_1(y)\big) \\
&= \frac{\sqrt{NT}}{v_{1N}(y)} \left[p^{J_1}(y)^\top \mathcal{S}_1 U_1 + \|p^{J_1}(y)\| \cdot O_p\left(\frac{1}{\underline{\nu}_{J_1J_2}^2} \sqrt{\frac{J r_{NT,J}}{NT}} + (J_1^{-\mu} + J_2^{-\mu})/\sqrt{\underline{\nu}_{J_1J_2}} \right) \right] \\
&= \frac{\sqrt{NT}}{v_{1N}(y)} p^{J_1}(y)^\top \mathcal{S}_1 U_1 + o_p(1),
\end{aligned}$$

under the conditions that $\underline{\nu}_{J_1J_2}^{-2} \sqrt{J r_{NT,J}} = o(1)$ and $\sqrt{NT}(J_1^{-\mu} + J_2^{-\mu})/\sqrt{\underline{\nu}_{J_1J_2}} = o(1)$. The rest of the proof proceeds in a similar way to the proof of Theorem 1(ii). We use the same notation as that used there. Let

$$a_{NT} \equiv \sqrt{NT} v_{1N}^{-1}(y) p^{J_1}(y)^\top \widetilde{U}_1 = v_{1N}^{-1}(y) p^{J_1}(y)^\top \mathcal{S}_1 \Sigma_{N,f}^{-1} \sum_{t=1}^{T} \mathbb{H}_{t-1}^\top \mathcal{E}_t / \sqrt{NT},$$

where $\mathbb{H}_{t-1} = \mathbb{Q}_{t-1} - \mathbb{Z} \Psi_{N,Z}^{-1} C_{N,ZJ}$. To show $a_{NT} \xrightarrow{d} N(0,1)$, we write $a_{NT} = \sum_{v=1}^{NT} \tilde{X}_{N,v}$, where for $t = 1, \ldots, T$ and $i = 1, \ldots, N$,

$$\tilde{X}_{N,(t-1)N+i} = (NT)^{-1/2} v_{1N}^{-1}(y) p^{J_1}(y)^\top \mathcal{S}_1 \Sigma_{N,f}^{-1} (Q_{i,t-1} - C_{N,ZJ}^\top \Psi_{N,Z}^{-1} Z_i) \epsilon_{it}.$$

In a similar manner to Theorem 1(ii), we can show that $\{\tilde{X}_{N,v}, \mathcal{F}_{N,v}, 1 \leq v \leq NT, N \geq 1\}$ forms a martingale difference array, where the σ-field $\mathcal{F}_{N,v}$ is defined in (16). It remains to check the conditions (17) and (18). To verify condition (17), note that

$$\mathbb{E}\big(\tilde{X}_{N,v}^4 | \mathcal{F}_{N,v-1}\big) = \frac{1}{(NT)^2} \frac{1}{v_{1N}^4(y)} \mathbb{E}\Big[\big(p^{J_1}(y)^\top \mathcal{S}_1 \Sigma_{N,f}^{-1} (Q_{i,t-1} - C_{N,ZJ}^\top \Psi_{N,Z}^{-1} Z_i) \epsilon_{it}\big)^4 \Big| \mathcal{F}_{N,v-1} \Big]$$

$$
\begin{aligned}
&= \frac{1}{(NT)^2}\frac{1}{v_{1N}^4(y)}\left[p^{J_1}(y)^\top \mathcal{S}_1\Sigma_{N,f}^{-1}(Q_{i,t-1}-C_{N,ZJ}^\top\Psi_{N,Z}^{-1}Z_i)\right]^4\cdot\mathbb{E}(\epsilon_{it}^4)\\
&\le \frac{C}{(NT)^2\underline{\nu}_{J_1J_2}^4}\|Q_{i,t-1}-C_{N,ZJ}^\top\Psi_{N,Z}^{-1}Z_i\|^4.
\end{aligned}
$$

Further, Assumption 4 implies that $\mathbb{E}\|Q_{i,t-1}-C_{N,ZJ}^\top\Psi_{N,Z}^{-1}Z_i\|^2 \le 2\mathbb{E}\|Q_{i,t-1}\|^2 + 2\|C_{N,ZJ}^\top\Psi_{N,Z}^{-1}Z_i\|^2 = O(J)$ and $\max_{1\le i\le N}\|Q_{i,t-1}-C_{N,ZJ}^\top\Psi_{N,Z}^{-1}Z_i\|^2 = O(\zeta_{0,J}^2)$. This and the last display yield that

$$
\mathbb{E}\big(X_{N,v}^4\big) = \mathbb{E}\left[\mathbb{E}\big(X_{N,v}^4\mid\mathcal{F}_{N,v-1}\big)\right] \le \frac{C}{\underline{\nu}_{J_1J_2}^4(NT)^2}\mathbb{E}\|Q_{i,t-1}-C_{N,ZJ}^\top\Psi_{N,Z}^{-1}Z_i\|^4 = O\bigg(\frac{\zeta_{0,J}^2J}{\underline{\nu}_{J_1J_2}^4(NT)^2}\bigg).
$$

Hence, $\sum_{v=1}^{NT}\mathbb{E}\big(X_{N,v}^4\big) = O(\zeta_{0,J}^2J/(\underline{\nu}_{J_1J_2}^4NT)) = o(1)$. This verifies condition (17).

To verify condition (18), observe that

$$
\sum_{v=1}^{NT}\mathbb{E}\big[\tilde{X}_{N,v}^2|\mathcal{F}_{N,v-1}\big] = \frac{\sigma^2}{NT}v_{1N}^{-2}(y)p^{J_1}(y)^\top\mathcal{S}_1\Sigma_{N,f}^{-1}\bigg(\sum_{t=1}^{T}\mathbb{H}_{t-1}^\top\mathbb{H}_{t-1}\bigg)\Sigma_{N,f}^{-1}\mathcal{S}_1^\top p^{J_1}(y).
$$

By Lemma 2 and Assumptions 4-5, we have $\|(NT)^{-1}\sum_{t=1}^{T}\mathbb{H}_{t-1}^\top\mathbb{H}_{t-1}-\Sigma_{N,f}\| = \|(NT)^{-1}\sum_{t=1}^{T}\big(\mathbb{H}_{t-1}^\top\mathbb{H}_{t-1}-\mathbb{E}(\mathbb{H}_{t-1}^\top\mathbb{H}_{t-1})\big)\| = O_p(\sqrt{r_{NT,J}}) = o_p(1)$, which implies that $\sum_{v=1}^{k_N}\mathbb{E}\big[\tilde{X}_{N,v}^2|\mathcal{F}_{N,v-1}\big]-1 = \sigma^2v_{1N}^{-2}(y)p^{J_1}(y)^\top\mathcal{S}_1\Sigma_{N,f}^{-1}[(NT)^{-1}\sum_{t=1}^{T}\mathbb{H}_{t-1}^\top\mathbb{H}_{t-1}-\Sigma_{N,f}]\Sigma_{N,f}^{-1}\mathcal{S}_1^\top p^{J_1}(y) = O_p(\underline{\nu}_{J_1J_2}^{-2}r_{NT,J}^{1/2}) = o_p(1)$. Thus, condition (18) is met. □

B.4 Proof of Theorem 3

(i) By Assumption 4′, we have

$$
\mathbf{s}_{N,1}^2 \ge c\cdot\|\Sigma_{f_1}^{-1}\|^2 \ge c'J_1, \tag{31}
$$

where c and c' are some finite positive constants. Decompose $\mathbf{T}_{NT,1}$ into three parts as follows:

$$
\mathbf{T}_{NT,1} = \sum_{i=1}^{N}\sum_{t=1}^{T}\big(\widehat{f}_1(Y_{i,t-1})-f_1(Y_{i,t-1})+f_1(Y_{i,t-1})-\widehat{\rho}_1Y_{i,t-1}\big)^2
$$

$$
\begin{aligned}
&= \sum_{i=1}^{N}\sum_{t=1}^{T}\Big[\big(\widehat{f}_1(Y_{i,t-1}) - f_1(Y_{i,t-1})\big)^2 + \big(f_1(Y_{i,t-1}) - \widehat{\rho}_1 Y_{i,t-1}\big)^2 \\
&\qquad + 2\big(\widehat{f}_1(Y_{i,t-1}) - f_1(Y_{i,t-1})\big)\big(f_1(Y_{i,t-1}) - \widehat{\rho}_1 Y_{i,t-1}\big)\Big] \\
&\equiv \mathbf{T}_{NT,1a} + \mathbf{T}_{NT,1b} + 2\mathbf{T}_{NT,1c}. \qquad (32)
\end{aligned}
$$

Under H_0^n, we have $f_1(Y_{i,t-1}) = \rho_1 Y_{i,t-1}$ for some $\rho_1 \in \mathbb{R}$, and $\mathbf{T}_{NT,1b} = (\rho_1 - \widehat{\rho}_1)^2 \sum_{i=1}^N \sum_{t=1}^T Y_{i,t-1}^2$. By Assumption 3 and the $\sqrt{NT}$-consistency of $\widehat{\rho}_1$, we have $\mathbf{T}_{NT,1b} = O_p(1)$, and thus $\mathbf{T}_{NT,1b}/\mathbf{s}_{N,1} = o_p(1)$ by (31). In Lemma 5 below, we will show that $\mathbf{T}_{NT,1c} = o_p(J_1^{1/2})$. For $\mathbf{T}_{NT,1a}$, we write

$$
\widehat{f}_1(Y_{i,t-1}) - f_1(Y_{i,t-1}) = p^{J_1}(Y_{i,t-1})^\top \widetilde{U}_1 + p^{J_1}(Y_{i,t-1})^\top \widetilde{U}_2 + p^{J_1}(Y_{i,t-1})^\top \widetilde{U}_3 - u_1(Y_{i,t-1}),
$$

where $\widetilde{U}_j = \mathcal{S}_1 U_j$ for $j = 1,2,3$. Here, U_1 and U_2 are given in (20) and (21), respectively, U_3 is defined below (29), and $u_1(y) \equiv f_1(y) - p^{J_1}(y)^\top \beta_1$. In addition, $\|\widetilde{U}_1\| = O_p(\sqrt{J_1/(\underline{\nu}_{J_1J_2}NT)})$, $\|\widetilde{U}_2\| = O_p(\underline{\nu}_{J_1J_2}^{-2} r_{NT,J}^{1/2}\sqrt{J/(NT)})$, and $\|\widetilde{U}_3\| = O_p(\underline{\nu}_{J_1J_2}^{-1/2}(J_1^{-\mu} + J_2^{-\mu}))$, as is shown in the proof of Lemma 4 and Theorem 2(i). Besides, Assumption 6 implies that $|u_1(Y_{i,t-1})| = O_p(J_1^{-\mu})$ for all i. Denote $\mathbf{P} \equiv (\mathbb{P}_0^{J_1\top}, \ldots, \mathbb{P}_{T-1}^{J_1\top})^\top$ as an $(NT) \times J_1$ matrix. By Lemma 2(ii) and Assumption 4, we have $\lambda_{\max}(\mathbf{P}^\top\mathbf{P}) = \lambda_{\max}(\sum_{t=1}^T \mathbb{P}_{t-1}^{J_1\top}\mathbb{P}_{t-1}^{J_1}) = O_p(NT)$. Combining these, we have

$$
\begin{aligned}
\mathbf{T}_{NT,1a} = {} & \widetilde{U}_1^\top(\mathbf{P}^\top\mathbf{P})\widetilde{U}_1 + \underbrace{\widetilde{U}_2^\top(\mathbf{P}^\top\mathbf{P})\widetilde{U}_2}_{=O_p(\underline{\nu}_{J_1J_2}^{-4} r_{NT,J} J)} + \underbrace{\widetilde{U}_3^\top(\mathbf{P}^\top\mathbf{P})\widetilde{U}_3}_{=O_p(NT\underline{\nu}_{J_1J_2}^{-1}(J_1^{-2\mu}+J_2^{-2\mu}))} \\
& + \underbrace{2\widetilde{U}_1^\top(\mathbf{P}^\top\mathbf{P})\widetilde{U}_2}_{=O_p(r_{NT,J}^{1/2}\sqrt{J_1 J/\underline{\nu}_{J_1J_2}^5})} + \underbrace{2\widetilde{U}_1^\top(\mathbf{P}^\top\mathbf{P})\widetilde{U}_3}_{=O_p(\sqrt{J_1NT/\underline{\nu}_{J_1J_2}^2}(J_1^{-\mu}+J_2^{-\mu}))} \\
& \underbrace{- 2\sum_{i=1}^{N}\sum_{t=1}^{T} p^{J_1}(Y_{i,t-1})^\top \widetilde{U}_1 \cdot u_1(Y_{i,t-1})}_{=O_p(\sqrt{NT}\underline{\nu}_{J_1J_2}^{-1/2}\zeta_{0,J_1}J_1^{1/2-\mu})} + \underbrace{2\widetilde{U}_2^\top(\mathbf{P}^\top\mathbf{P})\widetilde{U}_3}_{=O_p(\sqrt{JNT/\underline{\nu}_{J_1J_2}^5}r_{NT,J}^{1/2}(J_1^{-\mu}+J_2^{-\mu}))} \\
& \underbrace{- 2\sum_{i=1}^{N}\sum_{t=1}^{T} p^{J_1}(Y_{i,t-1})^\top \widetilde{U}_2 \cdot u_1(Y_{i,t-1})}_{=O_p(\sqrt{NT}\underline{\nu}_{J_1J_2}^{-2}\zeta_{0,J_1}r_{NT,J}^{1/2}J^{1/2}J_1^{-\mu})} + O_p\big(NT\zeta_{0,J_1}J_1^{-\mu}(J_1^{-\mu} + J_2^{-\mu})/\sqrt{\underline{\nu}_{J_1J_2}}\big).
\end{aligned}
$$

By Assumption 5′, we have $\mathbf{T}_{NT,1a} = \widetilde{U}_1^\top(\mathbf{P}^\top\mathbf{P})\widetilde{U}_1 + o_p(J_1^{1/2})$. Consequently, by (31),

$$\frac{\mathbf{T}_{NT,1} - \mathbf{B}_{N,1}}{\mathbf{s}_{N,1}} = \frac{\widetilde{U}_1^\top(\mathbf{P}^\top\mathbf{P})\widetilde{U}_1 - \mathbf{B}_{N,1}}{\mathbf{s}_{N,1}} + o_p(1),$$

and Theorem 3(i) follows from Lemma 6.

(ii) Note that $\mathbf{Y} - \mathbf{X}_1\bar{\theta}_1 = \mathbf{M}_{\mathbf{X}_1}\mathbf{Y} = \mathbf{M}_{\mathbf{X}_1}(\mathbf{X}_1\theta_1 + \mathbf{X}_2\beta_{1,2} + \mathbf{R} + \boldsymbol{\mathcal{E}})$, where $\mathbf{R} \equiv (\mathbb{R}_0^\top, \mathbb{R}_1^\top, \ldots, \mathbb{R}_{T-1}^\top)^\top$ with $\mathbb{R}_{t-1} = W(F_1(\mathbb{Y}_{t-1}) - \mathbb{P}_{t-1}^{J_1}\beta_1) + F_2(\mathbb{Y}_{t-1}) - \mathbb{P}_{t-1}^{J_2}\beta_2$. Under H_0^n and using the choice $p_1(y) = y$, we have $\beta_{1,2} = 0$ and $F_1(\mathbb{Y}_{t-1}) - \mathbb{P}_{t-1}^{J_1}\beta_1 = F_1(\mathbb{Y}_{t-1}) - \beta_{1,1}\mathbb{Y}_{t-1} = 0$, and thus $\mathbf{Y} - \mathbf{X}_1\bar{\theta}_1 = \mathbf{M}_{\mathbf{X}_1}(\mathbf{R}_2 + \boldsymbol{\mathcal{E}})$, where $\mathbf{R}_2 \equiv (\mathbb{R}_{0,2}^\top, \mathbb{R}_{1,2}^\top, \ldots, \mathbb{R}_{T-1,2}^\top)^\top$ with $\mathbb{R}_{t-1,2} = F_2(\mathbb{Y}_{t-1}) - \mathbb{P}_{t-1}^{J_2}\beta_2$. Hence, the test statistic in (11) can be written as

$$\mathbf{LM}_{NT,1} = (\mathbf{R}_2 + \boldsymbol{\mathcal{E}})^\top\mathbf{M}_{\mathbf{X}_1}\mathbf{X}_2\big(\mathbf{X}_2^\top\mathbf{M}_{\mathbf{X}_1}\mathbf{X}_2\big)^{-}\mathbf{X}_2^\top\mathbf{M}_{\mathbf{X}_1}(\mathbf{R}_2 + \boldsymbol{\mathcal{E}})/\widehat{\sigma}^2.$$

Decompose $\mathbf{LM}_{NT,1} = \mathbf{LM}_{NT,a} + \mathbf{LM}_{NT,b} + \mathbf{LM}_{NT,c}$, where

$$\begin{aligned}
\mathbf{LM}_{NT,a} &\equiv \boldsymbol{\mathcal{E}}^\top\mathbf{M}_{\mathbf{X}_1}\mathbf{X}_2\big(\mathbf{X}_2^\top\mathbf{M}_{\mathbf{X}_1}\mathbf{X}_2\big)^{-}\mathbf{X}_2^\top\mathbf{M}_{\mathbf{X}_1}\boldsymbol{\mathcal{E}}/\widehat{\sigma}^2,\\
\mathbf{LM}_{NT,b} &\equiv \mathbf{R}_2^\top\mathbf{M}_{\mathbf{X}_1}\mathbf{X}_2\big(\mathbf{X}_2^\top\mathbf{M}_{\mathbf{X}_1}\mathbf{X}_2\big)^{-}\mathbf{X}_2^\top\mathbf{M}_{\mathbf{X}_1}\mathbf{R}_2/\widehat{\sigma}^2,\\
\mathbf{LM}_{NT,c} &\equiv 2\boldsymbol{\mathcal{E}}^\top\mathbf{M}_{\mathbf{X}_1}\mathbf{X}_2\big(\mathbf{X}_2^\top\mathbf{M}_{\mathbf{X}_1}\mathbf{X}_2\big)^{-}\mathbf{X}_2^\top\mathbf{M}_{\mathbf{X}_1}\mathbf{R}_2/\widehat{\sigma}^2.
\end{aligned}$$

We first show that

$$\frac{\mathbf{LM}_{NT,a} - (J_1 - 1)}{\sqrt{2(J_1 - 1)}} \xrightarrow{d} N(0,1). \tag{33}$$

Let $\Psi_{N,11} \equiv \mathbb{E}(\mathbb{X}_{t,1}^\top\mathbb{X}_{t,1}/N)$, $\Psi_{N,12} \equiv \mathbb{E}(\mathbb{X}_{t,1}^\top\mathbb{X}_{t,2}/N)$, $\Psi_{N,22} \equiv \mathbb{E}(\mathbb{X}_{t,2}^\top\mathbb{X}_{t,2}/N)$, and $\Theta_N \equiv \Psi_{N,22} - \Psi_{N,12}^\top\Psi_{N,11}^{-1}\Psi_{N,12}$. We can further decompose $\mathbf{LM}_{NT,a}$ as $\sum_{l=1}^3 \mathbf{LM}_{NT,a}^{(l)}$, where

$$\begin{aligned}
&\mathbf{LM}_{NT,a}^{(1)} \equiv (NT)^{-1}\boldsymbol{\mathcal{E}}^\top\boldsymbol{\nu}\Theta_N^{-1}\boldsymbol{\nu}^\top\boldsymbol{\mathcal{E}}/\widehat{\sigma}^2,\\
&\mathbf{LM}_{NT,a}^{(2)} \equiv (NT)^{-1}\boldsymbol{\mathcal{E}}^\top\boldsymbol{\nu}\big[\big(\mathbf{X}_2^\top\mathbf{M}_{\mathbf{X}_1}\mathbf{X}_2/(NT)\big)^{-} - \Theta_N^{-1}\big]\boldsymbol{\nu}^\top\boldsymbol{\mathcal{E}}/\widehat{\sigma}^2,\\
&\mathbf{LM}_{NT,a}^{(3)} \equiv \big[\boldsymbol{\mathcal{E}}^\top\mathbf{M}_{\mathbf{X}_1}\mathbf{X}_2\big(\mathbf{X}_2^\top\mathbf{M}_{\mathbf{X}_1}\mathbf{X}_2\big)^{-}\mathbf{X}_2^\top\mathbf{M}_{\mathbf{X}_1}\boldsymbol{\mathcal{E}} - \boldsymbol{\mathcal{E}}^\top\boldsymbol{\nu}\big(\mathbf{X}_2^\top\mathbf{M}_{\mathbf{X}_1}\mathbf{X}_2\big)^{-}\boldsymbol{\nu}^\top\boldsymbol{\mathcal{E}}\big]/\widehat{\sigma}^2,
\end{aligned}$$

and $\boldsymbol{\nu} \equiv \mathbf{X}_2 - \mathbf{X}_1\Psi_{N,11}^{-1}\Psi_{N,12}$. First, under the assumptions of Theorem 3(ii), $\mathbf{LM}_{NT,a}^{(1)} - (J_1 - 1)/\sqrt{2(J_1 - 1)} \xrightarrow{d} N(0,1)$, which can be shown in a similar manner to Lemma 6. Second, it holds that $|\mathbf{LM}^{(2)}_{NT,a}| = o_p(J_1^{1/2})$. To see this, note that $\mathbb{E}\|\boldsymbol{\nu}^\top\boldsymbol{\mathcal{E}}/(NT)\|^2 = \sigma^2\mathrm{tr}(\Theta_N)/(NT) = O(J_1/(NT))$. Thus $\|\boldsymbol{\nu}^\top\boldsymbol{\mathcal{E}}/(NT)\| = O_p(\sqrt{J_1/(NT)})$. By Lemma 2, we have $\|\widehat{\Sigma}_{NT} - \Sigma_N\| = O_p(r_{NT,J}^{1/2}) = o_p(1)$, where $\widehat{\Sigma}_{NT} = (NT)^{-1}\sum_{t=1}^T \mathbb{X}_{t-1}^\top\mathbb{X}_{t-1}$ and $\Sigma_N = \mathbb{E}(\mathbb{X}_{t-1}^\top\mathbb{X}_{t-1}/N)$. This together with Assumption 4 implies that $\lambda_{\min}(\widehat{\Sigma}_{NT}) \geq c\underline{\nu}_{J_1J_2} > 0$ for some $0 < c < \infty$ wpa1. Applying Fact 2 gives $\|\widehat{\Sigma}_{NT}^{-1} - \Sigma_N^{-1}\| = O_p(\underline{\nu}_{J_1J_2}^{-2} r_{NT,J}^{1/2})$. Since $(\mathbf{X}_2\mathbf{M}_{\mathbf{X}_1}\mathbf{X}_2/(NT))^{-1}$ and Θ_N^{-1} are submatrices of $\widehat{\Sigma}_{NT}^{-1}$ and Σ_N^{-1}, respectively, we have $\|\big(\mathbf{X}_2^\top\mathbf{M}_{\mathbf{X}_1}\mathbf{X}_2/(NT)\big)^{-1} - \Theta_N^{-1}\| = O_p(\underline{\nu}_{J_1J_2}^{-2} r_{NT,J}^{1/2})$. Therefore, $|\mathbf{LM}^{(2)}_{NT,a}| \leq NT\|\big(\mathbf{X}_2^\top\mathbf{M}_{\mathbf{X}_1}\mathbf{X}_2/(NT)\big)^{-1} - \Theta_N^{-1}\|\|\boldsymbol{\nu}^\top\boldsymbol{\mathcal{E}}/(NT)\|^2 = O_p(\underline{\nu}_{J_1J_2}^{-2} r_{NT,J}^{1/2} J_1) = o_p(J_1^{1/2})$. Further,

$$
\begin{aligned}
|\mathbf{LM}^{(3)}_{NT,a}| &= \big|2\boldsymbol{\mathcal{E}}^\top(\mathbf{M}_{\mathbf{X}_1}\mathbf{X}_2 - \boldsymbol{\nu})\big(\mathbf{X}_2^\top\mathbf{M}_{\mathbf{X}_1}\mathbf{X}_2\big)^{-1}\boldsymbol{\nu}^\top\boldsymbol{\mathcal{E}}/\widehat{\sigma}^2 \\
&\qquad + \boldsymbol{\mathcal{E}}^\top(\mathbf{M}_{\mathbf{X}_1}\mathbf{X}_2 - \boldsymbol{\nu})\big(\mathbf{X}_2^\top\mathbf{M}_{\mathbf{X}_1}\mathbf{X}_2\big)^{-1}(\mathbf{X}_2^\top\mathbf{M}_{\mathbf{X}_1} - \boldsymbol{\nu}^\top)\boldsymbol{\mathcal{E}}/\widehat{\sigma}^2\big| \\
&\leq O_p(1)\cdot\big\|[\mathbf{X}_2^\top\mathbf{X}_1(\mathbf{X}_1^\top\mathbf{X}_1)^{-1} - \Psi_{N,12}^\top\Psi_{N,11}^{-1}]\mathbf{X}_1^\top\boldsymbol{\mathcal{E}}\big\|\cdot\|(\mathbf{X}_2^\top\mathbf{M}_{\mathbf{X}_1}\mathbf{X}_2)^{-1}\boldsymbol{\nu}^\top\boldsymbol{\mathcal{E}}\| \\
&\qquad + O_p((\underline{\nu}_{J_1J_2}NT)^{-1})\big\|[\mathbf{X}_2^\top\mathbf{X}_1(\mathbf{X}_1^\top\mathbf{X}_1)^{-1} - \Psi_{N,12}^\top\Psi_{N,11}^{-1}]\mathbf{X}_1^\top\boldsymbol{\mathcal{E}}\big\|^2.
\end{aligned}
$$

Following the lines of Lemmas 2-3, $\|[\mathbf{X}_2^\top\mathbf{X}_1(\mathbf{X}_1^\top\mathbf{X}_1)^{-1} - \Psi_{N,12}^\top\Psi_{N,11}^{-1}]\mathbf{X}_1^\top\boldsymbol{\mathcal{E}}\| \leq \|\mathbf{X}_2^\top\mathbf{X}_1(\mathbf{X}_1^\top\mathbf{X}_1)^{-1} - \Psi_{N,12}^\top\Psi_{N,11}^{-1}\|\cdot\|\mathbf{X}_1^\top\boldsymbol{\mathcal{E}}\| = O_p(\underline{\nu}_{J_1J_2}^{-2} r_{NT,J}^{1/2}\sqrt{NTJ_2})$. Also,

$$
\begin{aligned}
\|(\mathbf{X}_2^\top\mathbf{M}_{\mathbf{X}_1}\mathbf{X}_2)^{-1}\boldsymbol{\nu}^\top\boldsymbol{\mathcal{E}}\| &\leq \|(\mathbf{X}_2^\top\mathbf{M}_{\mathbf{X}_1}\mathbf{X}_2/(NT))^{-1} - \Theta_N^{-1}\|\|\boldsymbol{\nu}^\top\boldsymbol{\mathcal{E}}/(NT)\| + \|\Theta_N^{-1}\boldsymbol{\nu}^\top\boldsymbol{\mathcal{E}}/(NT)\| \\
&= O_p\left(\underline{\nu}_{J_1J_2}^{-2} r_{NT,J}^{1/2}\sqrt{J_1/(NT)} + \sqrt{J_1/(\underline{\nu}_{J_1J_2}NT)}\right) \\
&= O_p\big(\sqrt{J_1/(\underline{\nu}_{J_1J_2}NT)}\big)
\end{aligned}
$$

by Assumption 5. Above results yield $\|\mathbf{LM}^{(3)}_{NT,a}\| = O_p(\underline{\nu}_{J_1J_2}^{-5/2} r_{NT,J}^{1/2}\sqrt{J_1J_2} + \underline{\nu}_{J_1J_2}^{-5} r_{NT,J}J) = o_p(J_1^{1/2})$ by Assumption 5′(i). Thus, the convergence in (33) holds.

Next, since $\mathbf{M}_{\mathbf{X}_1}$ is idempotent,

$$
\begin{aligned}
\mathbf{LM}_{NT,b} &\equiv \mathbf{R}_2^\top\mathbf{M}_{\mathbf{X}_1}\mathbf{X}_2\big(\mathbf{X}_2^\top\mathbf{M}_{\mathbf{X}_1}\mathbf{X}_2\big)^{-1}\mathbf{X}_2^\top\mathbf{M}_{\mathbf{X}_1}\mathbf{R}_2/\widehat{\sigma}^2 \\
&\leq \|\mathbf{R}_2\|^2/\widehat{\sigma}^2 = O_p(NTJ_2^{-2\mu})
\end{aligned}
$$

by Assumption 6 and

$$
\begin{aligned}
|\mathbf{LM}_{NT,c}| \equiv & \left|\boldsymbol{\mathcal{E}}^\top \mathbf{M}_{\mathbf{X}_1}\mathbf{X}_2\big(\mathbf{X}_2^\top\mathbf{M}_{\mathbf{X}_1}\mathbf{X}_2\big)^{-1}\mathbf{X}_2^\top\mathbf{M}_{\mathbf{X}_1}\mathbf{R}_2/\widehat{\sigma}^2\right| \\
\le & \sqrt{\mathbf{LM}_{NT,a}\cdot\mathbf{LM}_{NT,b}} = O_p(\sqrt{J_1NT}J_2^{-\mu}),
\end{aligned}
$$

which are both $o_p(J_1^{1/2})$ by Assumption 5′(iii). A combination of the above results yields that $(\mathbf{LM}_{NT,1}-(J_1-1))/\sqrt{2(J_1-1)} = (\mathbf{LM}_{NT,a}+\mathbf{LM}_{NT,b}+2\mathbf{LM}_{NT,c}-(J_1-1))/\sqrt{2(J_1-1)} = \big(\mathbf{LM}_{NT,a}-(J_1-1)\big)/\sqrt{2(J_1-1)}+o_p(1)$ and Theorem 3(ii) immediately follows from (33). □

Lemma 5. *Suppose the assumptions in Theorem 3(i) hold. Then* $\mathbf{T}_{NT,1c} = o_p(J_1^{1/2})$ *under* H_0^n.

Proof. Using the decomposition (29), we can write $\mathbf{T}_{NT,1c}$ as

$$
\begin{aligned}
&\sum_{i=1}^N\sum_{t=1}^T\big(\widehat{f}_1(Y_{i,t-1})-f_1(Y_{i,t-1})\big)\big(f_1(Y_{i,t-1})-\widehat{\rho}_1Y_{i,t-1}\big) \\
&=(\rho_1-\widehat{\rho}_1)\sum_{i=1}^N\sum_{t=1}^T\big(p^{J_1}(Y_{i,t-1})^\top\widetilde{U}_1+p^{J_1}(Y_{i,t-1})^\top\widetilde{U}_2+p^{J_1}(Y_{i,t-1})^\top\widetilde{U}_3-u_1(Y_{i,t-1})\big)Y_{i,t-1} \\
&=(\rho_1-\widehat{\rho}_1)\bigg(\sum_{t=1}^T\mathbb{Y}_{t-1}^\top\mathbb{P}_{t-1}^{J_1}\bigg)\widetilde{U}_1+O_p\big(\underline{\nu}_{J_1J_2}^{-2}r_{NT,J}^{1/2}\sqrt{J_1J}\big) \\
&\quad+O_p\left(\sqrt{NTJ_1/\underline{\nu}_{J_1J_2}}(J_1^{-\mu}+J_2^{-\mu})\right) \\
&=(\rho_1-\widehat{\rho}_1)\bigg(\sum_{t=1}^T\mathbb{Y}_{t-1}^\top\mathbb{P}_{t-1}^{J_1}\bigg)\widetilde{U}_1+o_p(J_1^{1/2}),
\end{aligned}
$$

where the second equality uses the condition that $\sqrt{NT}(\widehat{\rho}_1-\rho_1) = O_p(1)$ and $\|\sum_{t=1}^T\mathbb{P}_{t-1}^{J_1\top}\mathbb{Y}_{t-1}\| = O_p(NT\sqrt{J_1})$, and the last equality is from Assumption 5′. Now, decompose $\big(\sum_{t=1}^T\mathbb{Y}_{t-1}^\top\mathbb{P}_{t-1}^{J_1}\big)\widetilde{U}_1$ into the following two terms:

$$
\begin{aligned}
\tilde{U}_1^{(1)} \equiv & \mathbb{E}(\mathbb{Y}_{t-1}^\top\mathbb{P}_{t-1}^{J_1})\mathcal{S}_1\Sigma_{N,f}^{-1}\mathbf{H}^\top\boldsymbol{\mathcal{E}}/N, \\
\tilde{U}_1^{(2)} \equiv & \sum_{t=1}^T\big[\mathbb{Y}_{t-1}^\top\mathbb{P}_{t-1}^{J_1}-\mathbb{E}(\mathbb{Y}_{t-1}^\top\mathbb{P}_{t-1}^{J_1})\big]\underbrace{\mathcal{S}_1\Sigma_{N,f}^{-1}\mathbf{H}^\top\boldsymbol{\mathcal{E}}/(NT)}_{=\widetilde{U}_1},
\end{aligned}
$$

where $\mathbf{H}$ is defined in the proof of Lemma 4. From Assumption 4′ and Markov's inequality,

$$\mathbb{E}\big[|\tilde{U}_1^{(1)}|^2\big] = \sigma^2 NT \cdot \mathbb{E}\big(\mathbb{Y}_{t-1}^\top \mathbb{P}_{t-1}^{J_1}/N\big)\Sigma_{N,f_1}^{-1}\mathbb{E}\big(\mathbb{P}_{t-1}^{J_1\top}\mathbb{Y}_{t-1}/N\big) = O(J_1^{1/2}NT/\underline{\nu}_{J_1J_2}),$$

implying that $|\tilde{U}_1^{(1)}| = O_p(J_1^{1/4}\sqrt{NT/\underline{\nu}_{J_1J_2}})$. Therefore, $(\widehat{\rho}_1 - \rho_1)\tilde{U}_1^{(1)} = O_p(J_1^{1/4}\underline{\nu}_{J_1J_2}^{-1/2}) = o_p(J_1^{1/2})$ under Assumption 5′(iv).

Recall from the proof of Lemma 1 that $\hat{Y}_{i,t-s}^{t-1}$ is a $\mathcal{F}_{i,t-1}(s)$-measurable approximation to $Y_{i,t-1}$, where $\mathcal{F}_{i,t}(s)$ is the σ-field generated by $\{\epsilon_{j\tau} : \Delta((i,t),(j,\tau)) \leq s\}$. In view of Assumption 1′ and (2), we have $\sup_{1\leq i\leq N}\|Y_{i,t-1} - \hat{Y}_{i,t-s}^{t-1}\|_4 \leq C_1 d^{s-1}$ for some constant C_1. Denote $P_{i,t-1} \equiv p^{J_1}(Y_{i,t-1})$ and $\hat{P}_{i,t-s}^{t-1} \equiv p^{J_1}(\hat{Y}_{i,t-s}^{t-1})$. Thus, $\mathbb{E}\|Y_{i,t-1}P_{i,t-1}\|^2 \leq C_2\zeta_{0,J_1}^2$ and

$$\begin{aligned}
\mathbb{E}\big\|Y_{i,t-1}P_{i,t-1} - \hat{Y}_{i,t-s}^{t-1}\hat{P}_{i,t-s}^{t-1}\big\|^2 \leq\ & 2\mathbb{E}\big\|(Y_{i,t-1} - \hat{Y}_{i,t-s}^{t-1})\hat{P}_{i,t-s}^{t-1}\big\|^2 + 2\mathbb{E}\big\|Y_{i,t-1}(\hat{P}_{i,t-s}^{t-1} - P_{i,t-1})\big\|^2 \\
\leq\ & 2\zeta_{0,J_1}^2\mathbb{E}\big|Y_{i,t-1} - \hat{Y}_{i,t-s}^{t-1}\big|^2 + 2\zeta_{1,J_1}^2\mathbb{E}\big|Y_{i,t-1}(\hat{Y}_{i,t-s}^{t-1} - Y_{i,t-1})\big|^2 \\
\leq\ & 2\zeta_{0,J_1}^2\mathbb{E}\big|Y_{i,t-1} - \hat{Y}_{i,t-s}^{t-1}\big|^2 + 2\zeta_{1,J_1}^2\big\|Y_{i,t-1}\big\|_4^2\big\|\hat{Y}_{i,t-s}^{t-1} - Y_{i,t-1}\big\|_4^2 \\
\leq\ & C_3(\zeta_{0,J_1}^2 + \zeta_{1,J_1}^2)d^{2(s-1)}
\end{aligned}$$

by the Hölder's inequality. As such, we can easily show that $\big\|(NT)^{-1}\sum_{t=1}^T\big[\mathbb{P}_{t-1}^{J_1\top}\mathbb{Y}_{t-1} - \mathbb{E}(\mathbb{P}_{t-1}^{J_1\top}\mathbb{Y}_{t-1})\big]\big\|^2 = O_p(\zeta_{0,J_1}^2/(NT) + \zeta_{0,J_1}(\zeta_{0,J_1} + \zeta_{1,J_1})\min\{\Xi_{N,0}, \Xi_{N,1}/T\})$ in the same way as in Lemma 2. Also, $\big\|\mathcal{S}_1\Sigma_{N,f}^{-1}\big(\mathbf{Q} - \mathbf{Z}\Psi_{N,Z}^{-1}C_{N,ZJ}\big)^\top\boldsymbol{\mathcal{E}}\big\| = O_p(\sqrt{J_1NT/\underline{\nu}_{J_1J_2}})$ by Markov's inequality. Thus

$$\begin{aligned}
|\tilde{U}_1^{(2)}| &\leq \left\|\frac{1}{NT}\sum_{t=1}^T\big[\mathbb{Y}_{t-1}^\top\mathbb{P}_{t-1}^{J_1} - \mathbb{E}\big(\mathbb{Y}_{t-1}^\top\mathbb{P}_{t-1}^{J_1}\big)\big]\right\| \cdot \left\|\mathcal{S}_1\Sigma_{N,f}^{-1}\big(\mathbf{Q} - \mathbf{Z}\Psi_{N,Z}^{-1}C_{N,ZJ}\big)^\top\boldsymbol{\mathcal{E}}\right\| \\
&= O_p\Big(\sqrt{J_1\zeta_{0,J_1}^2/\underline{\nu}_{J_1J_2}} + \sqrt{J_1\zeta_{0,J_1}(\zeta_{0,J_1} + \zeta_{1,J_1})\min\{\Xi_{N,0}, \Xi_{N,1}/T\}NT/\underline{\nu}_{J_1J_2}}\Big). \quad (34)
\end{aligned}$$

This result, Assumptions 5 and 8 together imply that $(\widehat{\rho}_1 - \rho_1)\tilde{U}_1^{(2)} = O_p\big(\zeta_{0,J_1}\sqrt{J_1/(\underline{\nu}_{J_1J_2}NT)}\big) + O_p\big(\sqrt{J_1\zeta_{0,J_1}(\zeta_{0,J_1} + \zeta_{1,J_1})\min\{\Xi_{N,0}, T^{-1}\Xi_{N,1}\}/\underline{\nu}_{J_1J_2}}\big) = o_p(J_1^{1/2})$. A combination of the above results yields that $\mathbf{T}_{NT,1c} = o_p(J_1^{1/2})$. □

Lemma 6. *Suppose the assumptions in Theorem 3(i) hold. Then,*

$$\frac{\widetilde{U}_1^\top \big(\mathbf{P}^\top \mathbf{P}\big) \widetilde{U}_1 - \mathbf{B}_{N,1}}{\mathbf{s}_{N,1}} \xrightarrow{d} N(0,1).$$

Proof. Recall that $\widetilde{U}_1 = \mathcal{S}_1 \Sigma_{N,f}^{-1} \sum_{i=1}^N \sum_{t=1}^T H_{i,t-1}\epsilon_{it}/(NT)$ with $H_{i,t-1} \equiv Q_{i,t-1} - C_{N,ZJ}^\top \Psi_{N,Z}^{-1} Z_i$, and $Q_{i,t-1} = (\bar{P}_{i,t-1}^{J_1\top}, P_{i,t-1}^{J_2\top})^\top$. For $i = 1, \ldots, N$ and $t = 1, \ldots, T$, denote

$$\pi_{N,(t-1)N+i} \equiv H_{i,t-1}\epsilon_{it} \tag{35}$$

so that $\widetilde{U}_1 = \mathcal{S}_1 \Sigma_{N,f}^{-1} \sum_{v=1}^{NT} \pi_{N,v}/(NT)$. Then,

$$\begin{aligned}
\widetilde{U}_1^\top (\mathbf{P}^\top \mathbf{P}) \widetilde{U}_1 &= \widetilde{U}_1^\top \mathbb{E}(\mathbf{P}^\top \mathbf{P}) \widetilde{U}_1 + \widetilde{U}_1^\top \big[\mathbf{P}^\top \mathbf{P} - \mathbb{E}(\mathbf{P}^\top \mathbf{P})\big] \widetilde{U}_1 \\
&= \frac{2}{NT} \sum_{1 \le u < v \le k_N} \pi_{N,u}^\top \Omega_{f_1} \pi_{N,v} + \frac{1}{NT} \sum_{v=1}^{k_N} \pi_{N,v}^\top \Omega_{f_1} \pi_{N,v} + \underbrace{O_p\big(\zeta_{0,J} r_{NT,J} J_1 / \underline{\nu}_{J_1 J_2}\big)}_{=o_p(J_1^{1/2})},
\end{aligned}$$

where $k_N = NT$ and $\Omega_{f_1} \equiv \Sigma_{N,f}^{-1} \mathcal{S}_1^\top \Phi_{N,J_1} \mathcal{S}_1 \Sigma_{N,f}^{-1}$ with $\Phi_{N,J_1} = \mathbb{E}(\mathbb{P}_t^{J_1\top} \mathbb{P}_t^{J_1}/N)$, and the second equality follows from Lemma 2(ii). Further,

$$\begin{aligned}
\left| \frac{1}{NT} \sum_{v=1}^{k_N} \pi_{N,v}^\top \Omega_{f_1} \pi_{N,v} - \mathbf{B}_{N,1} \right| &= \left| \mathrm{tr}\left(\Omega_{f_1} \left(\frac{1}{NT} \sum_{v=1}^{k_N} \pi_{N,v} \pi_{N,v}^\top - \sigma^2 \Sigma_{N,f} \right) \right) \right| \\
&\le \underbrace{\|\Omega_{f_1}\|}_{=O(\underline{\nu}_{J_1 J_2}^{-2} J_1^{1/2})} \cdot \left\| \frac{1}{NT} \sum_{v=1}^{k_N} \pi_{N,v} \pi_{N,v}^\top - \sigma^2 \Sigma_{N,f} \right\|,
\end{aligned}$$

and it follows from Assumption 4, Lemmas 2 and 7 that $\|(NT)^{-1} \sum_{v=1}^{k_N} \pi_{N,v} \pi_{N,v}^\top - \sigma^2 \Sigma_{N,f}\| = \|(NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T H_{i,t-1} H_{i,t-1}^\top \epsilon_{it}^2 - \sigma^2 \Sigma_{N,f}\| \le \|(NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T H_{i,t-1} H_{i,t-1}^\top (\epsilon_{it}^2 - \sigma^2)\| + \|\sigma^2 (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T [H_{i,t-1} H_{i,t-1}^\top - \mathbb{E}(H_{i,t-1} H_{i,t-1}^\top)]\| = O_p(\zeta_{0,J}\sqrt{J/NT} + r_{NT,J}^{1/2})$. From this and (31), we have

$$\frac{(NT)^{-1} \sum_{v=1}^{k_N} \pi_{N,v}^\top \Omega_{f_1} \pi_{N,v} - \mathbf{B}_{N,1}}{\mathbf{s}_{N,1}} = O_p\left(\underline{\nu}_{J_1 J_2}^{-2} r_{NT,J}^{1/2} \right) = o_p(1) \tag{36}$$

by Assumption 5. Therefore,

$$\frac{\widetilde{U}_1^\top(\mathbf{P}^\top\mathbf{P})\widetilde{U}_1 - \mathbf{B}_{N,1}}{\mathbf{s}_{N,1}} = \sum_{v=1}^{k_N} \zeta_{N,v} + o_p(1),$$

where $\zeta_{N,1} = 0$ and $\zeta_{N,v} \equiv (NT\mathbf{s}_{N,1})^{-1} 2\pi_{N,v}^\top \Omega_{f_1}(\sum_{l=1}^{v-1} \pi_{N,l})$ with $v = (t-1)N + i \geq 2$. By Assumption 6, we have $\zeta_{N,v}$ is $\mathcal{F}_{N,v}$-measurable and $\mathbb{E}[\zeta_{N,v}|\mathcal{F}_{N,v-1}] = 0$. Then, to derive the limiting distribution of $\sum_{v=1}^{NT} \zeta_{N,v}$, we can use the central limit theorem for martingale difference sequence, see e.g., Kuersteiner and Prucha (2013). In the proof of Theorem 2 therein, it is shown that the sufficient conditions for the CLT to hold are

$$\sum_{v=1}^{k_N} \mathbb{E}\big[|\zeta_{N,v}|^{2+\eta}\big] \to 0 \tag{37}$$

for some $\eta > 0$, and

$$\sum_{v=1}^{k_N} \mathbb{E}\big[\zeta_{N,v}^2|\mathcal{F}_{N,v-1}\big] \xrightarrow{p} 1. \tag{38}$$

We first verify condition (37). Let $\eta = 2$. By the definition (35), Assumptions 1 and 4,

$$\begin{aligned}
\mathbb{E}\big[|\zeta_{N,v}|^4|\mathcal{F}_{N,v-1}\big] &= \frac{16\mu_4}{(NT\mathbf{s}_{N,1})^4}\left|H_{i,t-1}^\top \Omega_{f_1}\left(\sum_{j=1}^{i-1} H_{j,t-1}\epsilon_{jt} + \sum_{j=1}^{N}\sum_{s=1}^{t-1} H_{j,s-1}\epsilon_{js}\right)\right|^4 \\
&\leq \frac{16\mu_4}{(NT\underline{\nu}_{J_1J_2}^2\mathbf{s}_{N,1})^4}\,\|H_{i,t-1}\|^4\left\|\sum_{j=1}^{i-1} H_{j,t-1}\epsilon_{jt} + \sum_{j=1}^{N}\sum_{s=1}^{t-1} H_{j,s-1}\epsilon_{js}\right\|^4 \\
&\leq \frac{C\zeta_{0,J}^4}{(NT\underline{\nu}_{J_1J_2}^2\mathbf{s}_{N,1})^4}\left\|\sum_{j=1}^{i-1} H_{j,t-1}\epsilon_{jt} + \sum_{j=1}^{N}\sum_{s=1}^{t-1} H_{j,s-1}\epsilon_{js}\right\|^4.
\end{aligned}$$

From Lemma 7(ii), $\mathbb{E}\big\|\sum_{j=1}^{i-1} H_{j,t-1}\epsilon_{jt} + \sum_{j=1}^{N}\sum_{s=1}^{t-1} H_{j,s-1}\epsilon_{js}\big\|^4 \leq C(NT)^2J^2$. Combining these and (31), the $\sum_{v=1}^{k_N} \mathbb{E}\{\mathbb{E}[\zeta_{N,v}^4|\mathcal{F}_{N,v-1}]\}$ is bounded by

$$\frac{C\zeta_{0,J}^4}{(NT\underline{\nu}_{J_1J_2}^2\mathbf{s}_{N,1})^4}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}\left\|\sum_{j=1}^{i-1} H_{j,t-1}\epsilon_{jt} + \sum_{j=1}^{N}\sum_{s=1}^{t-1} H_{j,s-1}\epsilon_{js}\right\|^4 = O\left(\frac{\zeta_{0,J}^4 J^2}{\underline{\nu}_{J_1J_2}^8 J_1^2 NT}\right) = o(1),$$

which completes the verification of condition (37).

Next, observe that

$$\sum_{v=1}^{k_N}\mathbb{E}\big[\zeta_{N,v}^2|\mathcal{F}_{N,v-1}\big]=\frac{4}{(NT\mathbf{s}_{N,1})^2}\sum_{v=1}^{NT}\mathrm{tr}\bigg\{\Omega_{f_1}\bigg(\sum_{u=1}^{v-1}\pi_{N,u}\bigg)\bigg(\sum_{u=1}^{v-1}\pi_{N,u}\bigg)^{\top}\Omega_{f_1}\mathbb{E}[\pi_{N,v}\pi_{N,v}^{\top}|\mathcal{F}_{N,v-1}]\bigg\}.$$

To show (38), it suffices to show that

$$\frac{4}{(NT)^2}\sum_{v=1}^{NT}\mathrm{tr}\bigg\{\Omega_{f_1}\bigg(\sum_{u=1}^{v-1}\pi_{N,u}\bigg)\bigg(\sum_{u=1}^{v-1}\pi_{N,u}\bigg)^{\top}\Omega_{f_1}\mathbb{E}[\pi_{N,v}\pi_{N,v}^{\top}|\mathcal{F}_{N,v-1}]\bigg\}-\mathbf{s}_{N,1}^2=o_p(\mathbf{s}_{N,1}^2). \quad (39)$$

Noting that $\sigma^2\Sigma_{N,f}=\sigma^2\mathbb{E}(\mathbb{H}_t^{\top}\mathbb{H}_t/N)=\sum_{v=1}^{k_N}\mathbb{E}(\pi_{N,v}\pi_{N,v}^{\top})/(NT)$, and $\mathbf{s}_{N,1}^2=2\sigma^4\mathrm{tr}(\mathbb{E}(\mathbb{P}_t^{\top}\mathbb{P}_t/N)\Sigma_{N,f_1}^{-1}\mathbb{E}(\mathbb{P}_t^{\top}\mathbb{P}_t/N)\Sigma_{N,f_1}^{-1})=2\sigma^4\mathrm{tr}(\Omega_{f_1}\Sigma_{N,f}\Omega_{f_1}\Sigma_{N,f})$, we can decompose the l.h.s. of the last equation as

$$\begin{aligned}&\frac{4}{(NT)^2}\sum_{v=1}^{k_N}\mathrm{tr}\bigg\{\Omega_{f_1}\bigg(\sum_{u=1}^{v-1}\pi_{N,u}\bigg)\bigg(\sum_{u=1}^{v-1}\pi_{N,u}\bigg)^{\top}\Omega_{f_1}\mathbb{E}[\pi_{N,v}\pi_{N,v}^{\top}|\mathcal{F}_{N,v-1}]\bigg\}-2\sigma^4\mathrm{tr}(\Omega_{f_1}\Sigma_{N,f}\Omega_{f_1}\Sigma_{N,f})\\ &=\frac{4}{(NT)^2}\sum_{v=1}^{k_N}\bigg[\mathrm{tr}\bigg\{\Omega_{f_1}\bigg(\sum_{u=1}^{v-1}\pi_{N,u}\bigg)\bigg(\sum_{u=1}^{v-1}\pi_{N,u}\bigg)^{\top}\Omega_{f_1}\mathbb{E}[\pi_{N,v}\pi_{N,v}^{\top}|\mathcal{F}_{N,v-1}]\bigg\}\\ &\quad-\mathrm{tr}\bigg\{\Omega_{f_1}\sum_{u=1}^{v-1}\mathbb{E}\big(\pi_{N,u}\pi_{N,u}^{\top}\big)\Omega_{f_1}\mathbb{E}(\pi_{N,v}\pi_{N,v}^{\top})\bigg\}\bigg]-\frac{2}{(NT)^2}\sum_{v=1}^{k_N}\mathrm{tr}\big(\Omega_{f_1}\mathbb{E}\big(\pi_{N,v}\pi_{N,v}^{\top}\big)\Omega_{f_1}\mathbb{E}\big(\pi_{N,v}\pi_{N,v}^{\top}\big)\big).\end{aligned}$$

It can be seen that $(NT)^{-2}\sum_{v=1}^{k_N}\mathrm{tr}\big(\Omega_{f_1}\mathbb{E}(\pi_{N,v}\pi_{N,v}^{\top})\Omega_{f_1}\mathbb{E}(\pi_{N,v}\pi_{N,v}^{\top})\big)=O(J^2/(\underline{\nu}_{J_1J_2}^4NT))=o(1)=o(\mathbf{s}_{N,1}^2)$. The first term on the r.h.s. of the last equation can be decomposed as $Q_{NT,1}+Q_{NT,2}$, where

$$\begin{aligned}Q_{NT,1}&\equiv\frac{4}{(NT)^2}\sum_{v=1}^{NT}\mathrm{tr}\left\{\Omega_{f_1}\bigg(\sum_{u=1}^{v-1}\pi_{N,u}\bigg)\bigg(\sum_{u=1}^{v-1}\pi_{N,u}\bigg)^{\top}\Omega_{f_1}\left(\mathbb{E}[\pi_{N,v}\pi_{N,v}^{\top}|\mathcal{F}_{N,v-1}]-\mathbb{E}(\pi_{N,v}\pi_{N,v}^{\top})\right)\right\},\\ Q_{NT,2}&\equiv\frac{4}{(NT)^2}\sum_{v=1}^{NT}\mathrm{tr}\bigg\{\Omega_{f_1}\bigg[\bigg(\sum_{u=1}^{v-1}\pi_{N,u}\bigg)\bigg(\sum_{u=1}^{v-1}\pi_{N,u}\bigg)^{\top}-\sum_{u=1}^{v-1}\mathbb{E}\big(\pi_{N,u}\pi_{N,u}^{\top}\big)\bigg]\Omega_{f_1}\mathbb{E}\big(\pi_{N,v}\pi_{N,v}^{\top}\big)\bigg\}.\end{aligned}$$

First we show that $Q_{NT,1} = o_p(J_1)$. Note that $Q_{NT,1} = Q_{NT,1a} + Q_{NT,1b}$, where

$$Q_{NT,1a} \equiv \frac{4}{(NT)^2} \sum_{1\le u<v\le NT} \mathrm{tr}\left\{\Omega_{f_1}\pi_{N,u}\pi_{N,u}^\top\Omega_{f_1}\left[\mathbb{E}[\pi_{N,v}\pi_{N,v}^\top|\mathcal{F}_{N,v-1}] - \mathbb{E}(\pi_{N,v}\pi_{N,v}^\top)\right]\right\},$$

$$Q_{NT,1b} \equiv \frac{8}{(NT)^2} \sum_{1\le q<u<v\le NT} \mathrm{tr}\left\{\Omega_{f_1}\pi_{N,u}\pi_{N,q}^\top\Omega_{f_1}\left[\mathbb{E}[\pi_{N,v}\pi_{N,v}^\top|\mathcal{F}_{N,v-1}] - \mathbb{E}(\pi_{N,v}\pi_{N,v}^\top)\right]\right\}.$$

For $Q_{NT,1a}$, we have

$$Q_{NT,1a} = \frac{4}{(NT)^2}\mathrm{tr}\left\{\Omega_{f_1}\sum_{u=1}^{NT}\left(\pi_{N,u}\pi_{N,u}^\top\right)\Omega_{f_1}\sum_{v=1}^{NT}\left[\mathbb{E}[\pi_{N,v}\pi_{N,v}^\top|\mathcal{F}_{N,v-1}] - \mathbb{E}(\pi_{N,v}\pi_{N,v}^\top)\right]\right\} \tag{40}$$

$$- \frac{4}{(NT)^2} \sum_{1\le u\le v\le NT} \mathrm{tr}\left\{\Omega_{f_1}\pi_{N,v}\pi_{N,v}^\top\Omega_{f_1}\left[\mathbb{E}[\pi_{N,u}\pi_{N,u}^\top|\mathcal{F}_{N,u-1}] - \mathbb{E}(\pi_{N,u}\pi_{N,u}^\top)\right]\right\}. \tag{41}$$

By the triangular inequality, Assumption 4, Lemmas 2 and 7, we have $\left\|(NT)^{-1}\sum_{u=1}^{NT}\pi_{N,u}\pi_{N,u}^\top\right\| \le \|(NT)^{-1}\sum_{u=1}^{NT}\left(\pi_{N,u}\pi_{N,u}^\top - \mathbb{E}[\pi_{N,u}\pi_{N,u}^\top|\mathcal{F}_{N,u-1}]\right)\| + \|(NT)^{-1}\sum_{u=1}^{NT}\left[\mathbb{E}[\pi_{N,u}\pi_{N,u}^\top|\mathcal{F}_{N,u-1}] - \mathbb{E}(\pi_{N,u}\pi_{N,u}^\top)\right]\| + \sigma^2\|\Sigma_{N,f}\| = O_p(J^{1/2}\zeta_{0,J}(NT)^{-1/2}) + O_p(r_{NT,J}^{1/2}) + O(J) = O_p(J)$. Then,

$$\left|\frac{1}{(NT)^2}\sum_{u=1}^{NT}\sum_{v=1}^{NT}\mathrm{tr}\left\{\Omega_{f_1}\pi_{N,u}\pi_{N,u}^\top\Omega_{f_1}\left[\mathbb{E}[\pi_{N,v}\pi_{N,v}^\top|\mathcal{F}_{N,v-1}] - \mathbb{E}(\pi_{N,v}\pi_{N,v}^\top)\right]\right\}\right|$$
$$\le \lambda^2_{\max}(\Omega_{f_1})\underbrace{\left\|\frac{1}{NT}\sum_{u=1}^{NT}\pi_{N,u}\pi_{N,u}^\top\right\|}_{=O_p(J)}\cdot\underbrace{\left\|\frac{1}{NT}\sum_{v=1}^{NT}\left[\mathbb{E}[\pi_{N,v}\pi_{N,v}^\top|\mathcal{F}_{N,v-1}] - \mathbb{E}(\pi_{N,v}\pi_{N,v}^\top)\right]\right\|}_{=O_p(r_{NT,J}^{1/2})}.$$

Therefore, the term (40) is $O_p(Jr_{NT,J}^{1/2}/\underline{\nu}^4_{J_1J_2}) = o_p(J_1)$ by Assumption 5.

For $1 \le v \le k_N$, let $\psi_{v,1} \equiv \mathrm{tr}\{\Omega_{f_1}(\pi_{N,v}\pi_{N,v}^\top - \mathbb{E}[\pi_{N,v}\pi_{N,v}^\top|\mathcal{F}_{N,v-1}])\Omega_{f_1}\sum_{u=1}^{v}\left[\mathbb{E}[\pi_{N,u}\pi_{N,u}^\top|\mathcal{F}_{N,u-1}] - \mathbb{E}(\pi_{N,u}\pi_{N,u}^\top)\right]\}$, and $\psi_{v,2} \equiv \sum_{u=1}^{v}\mathrm{tr}\{\Omega_{f_1}\mathbb{E}[\pi_{N,v}\pi_{N,v}^\top|\mathcal{F}_{N,v-1}]\Omega_{f_1}[\mathbb{E}[\pi_{N,u}\pi_{N,u}^\top|\mathcal{F}_{N,u-1}] - \mathbb{E}(\pi_{N,u}\pi_{N,u}^\top)]\}$. Then the term (41) can be written as $4\sum_{v=1}^{k_N}\left(\psi_{v,1}+\psi_{v,2}\right)/(NT)^2$. Noting that $\{\psi_{v,1}, \mathcal{F}_{N,v} : 1 \le v \le NT, N \ge 1\}$ forms a martingale difference sequence, we have

$$\mathbb{E}\left[\frac{1}{(NT)^2}\sum_{v=1}^{k_N}\psi_{v,1}\right]^2 = \frac{1}{(NT)^4}\sum_{v=1}^{k_N}\mathbb{E}(\psi_{v,1}^2)$$

$$
\begin{aligned}
&\leq \frac{\lambda_{\max}^4(\Omega_{f_1})}{(NT)^4} \sum_{v=1}^{k_N} \mathbb{E} \left\| \pi_{N,v}\pi_{N,v}^\top - \mathbb{E}[\pi_{N,v}\pi_{N,v}^\top | \mathcal{F}_{N,v-1}] \right\|^2 \\
&\quad \times \left\| \sum_{u=1}^{v} \left[\mathbb{E}[\pi_{N,u}\pi_{N,u}^\top | \mathcal{F}_{N,u-1}] - \mathbb{E}(\pi_{N,u}\pi_{N,u}^\top) \right] \right\|^2 \\
&= \frac{\lambda_{\max}^4(\Omega_{f_1})}{(NT)^4} \sum_{i=1}^{N}\sum_{t=1}^{T} \mathbb{E} \left\| H_{i,t-1}H_{i,t-1}^\top (\epsilon_{it}^2 - \sigma^2) \right\|^2 \left\| \sigma^2 \Upsilon_{it} \right\|^2 \\
&\leq \frac{C}{(NT\underline{\nu}_{J_1J_2}^2)^4} \sum_{i=1}^{N}\sum_{t=1}^{T} \mathbb{E} \left\| H_{i,t-1}H_{i,t-1}^\top \right\|^2 \left\| \Upsilon_{it} \right\|^2 \\
&\leq \frac{C\zeta_{0,J}^4}{(NT\underline{\nu}_{J_1J_2}^2)^4} \sum_{i=1}^{N}\sum_{t=1}^{T} \mathbb{E} \left\| \Upsilon_{it} \right\|^2,
\end{aligned}
$$

where $\Upsilon_{it} \equiv \sum_{j=1}^{i} \left[H_{j,t-1}H_{j,t-1}^\top - \mathbb{E}\big(H_{j,t-1}H_{j,t-1}^\top\big) \right] + \sum_{j=1}^{N}\sum_{s=1}^{t-1} \left[H_{j,s-1}H_{j,s-1}^\top - \mathbb{E}\big(H_{j,s-1}H_{j,s-1}^\top\big) \right]$. The first inequality uses the independence of ϵ_{it} and H_{js} for all $1 \leq j \leq N$ and $1 \leq s < t$ and that $\mathbb{E}|\epsilon_{it}^2 - \sigma^2|^2 < \infty$. Further, using similar arguments as Lemma 2, we have

$$
\mathbb{E} \left\| \Upsilon_{it} \right\|^2 \leq C\zeta_{0,J}^2 \big(J/(NT) + \zeta_{1,J} J^{1/2} \min\{\Xi_{N,0}, T^{-1}\Xi_{N,1}\} \big) (NT)^2. \tag{42}
$$

A combination of above results gives that $\mathbb{E}\left[(NT)^{-2}\sum_{v=1}^{k_N} \psi_{v,1}\right]^2 = O(\underline{\nu}_{J_1J_2}^{-8}\zeta_{0,J}^6(J/(NT)^2 + \zeta_{1,J}J^{1/2}\min\{\Xi_{N,0}, T^{-1}\Xi_{N,1}\}/(NT))) = o(J_1^2)$. It follows from Markov's inequality that $|(NT)^{-2}\sum_{v=1}^{k_N} \psi_{v,1}| = o_p(J_1)$. Next, by the Hölder's inequality,

$$
\begin{aligned}
&\mathbb{E} \left| \frac{1}{(NT)^2} \sum_{v=1}^{k_N} \psi_{v,2} \right| \\
&= \mathbb{E} \left| \frac{1}{(NT)^2} \sum_{1 \leq u \leq v \leq k_N} \operatorname{tr}\left\{ \Omega_{f_1} \mathbb{E}\left[\pi_{N,v}\pi_{N,v}^\top | \mathcal{F}_{N,v-1}\right] \Omega_{f_1} \left[\mathbb{E}[\pi_{N,u}\pi_{N,u}^\top | \mathcal{F}_{N,u-1}] - \mathbb{E}(\pi_{N,u}\pi_{N,u}^\top) \right] \right\} \right| \\
&\leq \frac{\lambda_{\max}^2(\Omega_{f_1})}{(NT)^2} \sum_{v=1}^{k_N} \mathbb{E} \left\| \mathbb{E}\left[\pi_{N,v}\pi_{N,v}^\top | \mathcal{F}_{N,v-1}\right] \right\| \cdot \left\| \sum_{u=1}^{v} \left[\mathbb{E}[\pi_{N,u}\pi_{N,u}^\top | \mathcal{F}_{N,u-1}] - \mathbb{E}(\pi_{N,u}\pi_{N,u}^\top) \right] \right\| \\
&\leq \frac{C}{(\underline{\nu}_{J_1J_2}^2 NT)^2} \sum_{v=1}^{k_N} \left\| \mathbb{E}\left[\pi_{N,v}\pi_{N,v}^\top | \mathcal{F}_{N,v-1}\right] \right\|_2 \left\| \sum_{u=1}^{v} \left[\mathbb{E}[\pi_{N,u}\pi_{N,u}^\top | \mathcal{F}_{N,u-1}] - \mathbb{E}(\pi_{N,u}\pi_{N,u}^\top) \right] \right\|_2,
\end{aligned}
$$

where $\|A\|_2 = \sqrt{\mathbb{E}\|A\|^2}$. By Assumption 4 and (35), we have $\max_v \left\| \mathbb{E}\left[\pi_{N,v}\pi_{N,v}^\top | \mathcal{F}_{N,v-1}\right] \right\|_2 =$

$\sigma^2 \max_{i,t} \|H_{i,t-1}H_{i,t-1}^\top\|_2 = \sigma^2 \max_i \{\mathbb{E}(H_{i,t-1}^\top H_{i,t-1})^2\}^{1/2} = O(J^{1/2}\zeta_{0,J})$. Also, (42) implies that $\big\| \sum_{u=1}^{v} \big[\mathbb{E}[\pi_{N,u}\pi_{N,u}^\top|\mathcal{F}_{N,u-1}] - \mathbb{E}(\pi_{N,u}\pi_{N,u}^\top)\big] \big\|_2 = O(\zeta_{0,J}\sqrt{JNT + \zeta_{1,J}J^{1/2}\min\{\Xi_{N,0}, T^{-1}\Xi_{N,1}\}(NT)^2})$. Combining the above results, we conclude that $\mathbb{E}|(NT)^{-2}\sum_{v=1}^{NT}\psi_{v,2}| = O(\zeta_{0,J}^2 J/(\underline{\nu}_{J_1J_2}^4\sqrt{NT}) + \zeta_{0,J}^2 J^{3/4}\sqrt{\zeta_{1,J}\min\{\Xi_{N,0}, T^{-1}\Xi_{N,1}\}}/\underline{\nu}_{J_1J_2}^4) = o(J_1)$. Hence, the term (41) is also $o_p(J_1)$. Consequently, $Q_{NT,1a} = o_p(J_1)$.

Next, we decompose $Q_{NT,1b}$ as $Q_{NT,1b} = 8(Q_{NT,1b}^{(1)} - Q_{NT,1b}^{(2)})$ with

$$Q_{NT,1b}^{(1)} = \frac{1}{(NT)^2}\mathrm{tr}\Big\{\Omega_{f_1}\sum_{u=1}^{k_N}\sum_{s=1}^{u-1}\big(\pi_{N,u}\pi_{N,s}^\top\big)\Omega_{f_1}\sum_{v=1}^{k_N}\big[\mathbb{E}[\pi_{N,v}\pi_{N,v}^\top|\mathcal{F}_{N,v-1}] - \mathbb{E}(\pi_{N,v}\pi_{N,v}^\top)\big]\Big\},$$

$$Q_{NT,1b}^{(2)} = \frac{1}{(NT)^2}\mathrm{tr}\Big\{\Omega_{f_1}\sum_{u=1}^{k_N}\sum_{s=1}^{u-1}\big(\pi_{N,u}\pi_{N,s}^\top\big)\Omega_{f_1}\sum_{v=1}^{u}\big[\mathbb{E}[\pi_{N,v}\pi_{N,v}^\top|\mathcal{F}_{N,v-1}] - \mathbb{E}(\pi_{N,v}\pi_{N,v}^\top)\big]\Big\}.$$

First we show $Q_{NT,1b}^{(1)} = o_p(J_1)$. Note that $\mathbb{E}(\sum_{u=1}^{k_N}\sum_{s=1}^{u-1}\pi_{N,u}\pi_{N,s}^\top) = \mathbf{0}$ and

$$\begin{aligned}\mathbb{E}\Big\|\frac{1}{NT}\sum_{u=1}^{k_N}\sum_{s=1}^{u-1}\pi_{N,u}\pi_{N,s}^\top\Big\|^2 &= \frac{1}{(NT)^2}\sum_{u=1}^{k_N}\mathbb{E}\left[\mathbb{E}\big[\pi_{N,u}^\top\pi_{N,u}|\mathcal{F}_{N,u-1}\big]\Big(\sum_{v=1}^{u-1}\pi_{N,v}\Big)^\top\Big(\sum_{v=1}^{u-1}\pi_{N,v}\Big)\right]\\ &\le \frac{C\zeta_{0,J}^2}{(NT)^2}\sum_{u=1}^{k_N}\mathbb{E}\Big(\sum_{v=1}^{u-1}\pi_{N,v}\Big)^\top\Big(\sum_{v=1}^{u-1}\pi_{N,v}\Big)\\ &\le \frac{C\zeta_{0,J}^2}{(NT)^2}\sum_{u=1}^{k_N}\sum_{v=1}^{k_N}\mathbb{E}(\pi_{N,v}^\top\pi_{N,v}) = O(J\zeta_{0,J}^2),\end{aligned}$$

where we have used $\mathbb{E}(\pi_{N,v}^\top\pi_{N,v}) = O(J)$ according to Assumption 4. This implies that $\|(NT)^{-1}\sum_{u=1}^{k_N}\sum_{s=1}^{u-1}\pi_{N,u}\pi_{N,s}^\top\| = O_p(J^{1/2}\zeta_{0,J})$. This and Lemma 2(ii) yield

$$|Q_{NT,1b}^{(1)}| \le \frac{C}{\underline{\nu}_{J_1J_2}^4}\underbrace{\Big\|\frac{1}{NT}\sum_{u=1}^{k_N}\sum_{s=1}^{u-1}\pi_{N,u}\pi_{N,s}^\top\Big\|}_{=O_p(J^{1/2}\zeta_{0,J})}\cdot\underbrace{\Big\|\frac{1}{NT}\sum_{v=1}^{k_N}\big[\mathbb{E}[\pi_{N,v}\pi_{N,v}^\top|\mathcal{F}_{N,v-1}] - \mathbb{E}(\pi_{N,v}\pi_{N,v}^\top)\big]\Big\|}_{=O_p(r_{NT,J}^{1/2})} = o_p(J_1).$$

Second, we also have $Q_{NT,1b}^{(2)} = o_p(J_1)$. To see this, we write $Q_{NT,1b}^{(2)} = (NT)^{-2}\sum_{u=1}^{k_N}\phi_{N,u}$, where $\phi_{N,u} \equiv \sum_{s=1}^{u-1}\pi_{N,s}^\top\Omega_{f_1}\sum_{v=1}^{u}\big[\mathbb{E}[\pi_{N,v}\pi_{N,v}^\top|\mathcal{F}_{N,v-1}] - \mathbb{E}(\pi_{N,v}\pi_{N,v}^\top)\big]\Omega_{f_1}\pi_{N,u}$. By construction, $\phi_{N,u}$ is $\mathcal{F}_{N,u}$-measurable and $\mathbb{E}[\phi_{N,u}|\mathcal{F}_{N,u-1}] = 0$. Therefore,

$$
\begin{aligned}
\mathbb{E}|Q^{(2)}_{NT,1b}|^2 &= \frac{1}{(NT)^4}\sum_{u=1}^{NT}\mathbb{E}|\phi_{N,u}|^2 \\
&= \frac{1}{(NT)^4}\sum_{u=1}^{NT}\mathbb{E}\Big[\sum_{s=1}^{u-1}\pi_{N,s}^\top\Omega_{f_1}\sum_{v=1}^{u}\big[\mathbb{E}[\pi_{N,v}\pi_{N,v}^\top|\mathcal{F}_{N,v-1}] - \mathbb{E}(\pi_{N,v}\pi_{N,v}^\top)\big]\Omega_{f_1}\pi_{N,u}\Big]^2 \\
&\le \frac{1}{(NT)^4}\sum_{u=1}^{NT}\mathbb{E}\left\|\Omega_{f_1}\sum_{v=1}^{u}\big[\mathbb{E}[\pi_{N,v}\pi_{N,v}^\top|\mathcal{F}_{N,v-1}] - \mathbb{E}(\pi_{N,v}\pi_{N,v}^\top)\big]\Omega_{f_1}\sum_{s=1}^{u-1}\pi_{N,s}\right\|^2\|\pi_{N,u}\|^2 \\
&\le \frac{C\zeta_{0,J}^2}{(NT)^4}\sum_{u=1}^{NT}\mathbb{E}\left\|\Omega_{f_1}\sum_{v=1}^{u}\big[\mathbb{E}[\pi_{N,v}\pi_{N,v}^\top|\mathcal{F}_{N,v-1}] - \mathbb{E}(\pi_{N,v}\pi_{N,v}^\top)\big]\Omega_{f_1}\sum_{s=1}^{u-1}\pi_{N,s}\right\|^2 \\
&\le \frac{C\zeta_{0,J}^2}{(NT\underline{\nu}_{J_1J_2}^2)^4}\sum_{u=1}^{NT}\mathbb{E}\left\|\sum_{v=1}^{u}\big[\mathbb{E}[\pi_{N,v}\pi_{N,v}^\top|\mathcal{F}_{N,v-1}] - \mathbb{E}(\pi_{N,v}\pi_{N,v}^\top)\big]\right\|^2\left\|\sum_{s=1}^{u-1}\pi_{N,s}\right\|^2 \\
&\le \frac{C\zeta_{0,J}^2}{(NT\underline{\nu}_{J_1J_2}^2)^4}\sum_{u=1}^{NT}\left\|\sum_{v=1}^{u}\big[\mathbb{E}[\pi_{N,v}\pi_{N,v}^\top|\mathcal{F}_{N,v-1}] - \mathbb{E}(\pi_{N,v}\pi_{N,v}^\top)\big]\right\|_3^2\left\|\sum_{s=1}^{u-1}\pi_{N,s}\right\|_6^2,
\end{aligned}
$$

where the second inequality follows from the fact that $\mathbb{E}[\|\pi_{N,u}\|^2|\mathcal{F}_{N,u-1}] \le C\zeta_{0,J}^2$ and the law of iterated expectation, and the last follows from the Hölder's inequality. By Lemma 7, we have $\big\|\sum_{s=1}^{u-1}\pi_{N,s}\big\|_6^2 \le CNTJ$. Also, $\mathbb{E}\big\|\sum_{v=1}^{u}\big[\mathbb{E}[\pi_{N,v}\pi_{N,v}^\top|\mathcal{F}_{N,v-1}] - \mathbb{E}(\pi_{N,v}\pi_{N,v}^\top)\big]\big\|^3 \le (CNT\zeta_{0,J}^2)\mathbb{E}\big\|\sum_{v=1}^{u}\big[\mathbb{E}[\pi_{N,v}\pi_{N,v}^\top|\mathcal{F}_{N,v-1}] - \mathbb{E}(\pi_{N,v}\pi_{N,v}^\top)\big]\big\|^2 \le C\zeta_{0,J}^2N^2T^3 + C\zeta_{0,J}^2 r_{NT,J}(NT)^3$. Thus the $\mathbb{E}|Q^{(2)}_{NT,1b}|^2$ is bounded by $CJ\zeta_{0,J}^{10/3}(N^{-2/3} + r_{NT,J}^{2/3})\underline{\nu}_{J_1J_2}^{-8}$, which is $o(J_1^2)$ by Assumption $5'$. The above results show that $Q^{(2)}_{NT,1b} = o_p(J_1)$.

It remains to show that $Q_{NT,2} = o_p(\mathbf{s}_{N,1}^2)$. Note that $Q_{NT,2} = 4(Q_{NT,2a} + 2Q_{NT,2b})$, where

$$
\begin{aligned}
Q_{NT,2a} &\equiv \frac{1}{(NT)^2}\sum_{v=1}^{k_N}\mathrm{tr}\left\{\Omega_{f_1}\sum_{s=1}^{v-1}\left[\pi_{N,s}\pi_{N,s}^\top - \mathbb{E}\big(\pi_{N,s}\pi_{N,s}^\top\big)\right]\Omega_{f_1}\mathbb{E}\big(\pi_{N,v}\pi_{N,v}^\top\big)\right\}, \\
Q_{NT,2b} &\equiv \frac{1}{(NT)^2}\sum_{v=1}^{k_N}\sum_{u=1}^{v-1}\sum_{s=1}^{u-1}\mathrm{tr}\left\{\Omega_{f_1}\pi_{N,s}\pi_{N,u}^\top\Omega_{f_1}\mathbb{E}\big(\pi_{N,v}\pi_{N,v}^\top\big)\right\}.
\end{aligned}
$$

First, we observe that

$$
\mathbb{E}\big|Q_{NT,2a}\big| = \frac{1}{(NT)^2}\mathbb{E}\bigg|\sum_{v=1}^{k_N}\mathrm{tr}\bigg\{\Omega_{f_1}\sum_{s=1}^{v-1}\left[\pi_{N,s}\pi_{N,s}^\top - \mathbb{E}\big(\pi_{N,s}\pi_{N,s}^\top\big)\right]\Omega_{f_1}\mathbb{E}\big(\pi_{N,v}\pi_{N,v}^\top\big)\bigg\}\bigg|
$$

$$\leq \frac{1}{(NT)^2} \sum_{v=1}^{k_N} \mathbb{E} \left\| \sum_{s=1}^{v-1} \left[\pi_{N,s} \pi_{N,s}^\top - \mathbb{E}\big(\pi_{N,s} \pi_{N,s}^\top\big) \right] \right\| \left\| \Omega_{f_1} \mathbb{E}\big(\pi_{N,v} \pi_{N,v}^\top\big) \Omega_{f_1} \right\|.$$

From the proof of Lemmas 2 and 7(i), we have $\mathbb{E}\big\| \sum_{s=1}^{v} [\pi_{N,s}\pi_{N,s}^\top - \mathbb{E}\big(\pi_{N,s}\pi_{N,s}^\top\big)] \big\| \leq \big\| \sum_{s=1}^{v} [\pi_{N,s}\pi_{N,s}^\top - \mathbb{E}\big(\pi_{N,s}\pi_{N,s}^\top\big)] \big\|_2 \leq \big\| \sum_{s=1}^{v} [\pi_{N,s}\pi_{N,s}^\top - \mathbb{E}[\pi_{N,s}\pi_{N,s}^\top | \mathcal{F}_{N,s-1}]] \big\|_2 + \big\| \sum_{s=1}^{v} \big[\mathbb{E}\big[\pi_{N,s}\pi_{N,s}^\top | \mathcal{F}_{N,s-1}\big] - \mathbb{E}(\pi_{N,s}\pi_{N,s}^\top)\big] \big\|_2 = O(NTr_{NT,J}^{1/2})$, where we have used the CS and Minkowski's inequalities. Also, $\big\|\Omega_{f_1}\mathbb{E}(\pi_{N,v}\pi_{N,v}^\top)\Omega_{f_1}\big\| = O(\underline{\nu}_{J_1J_2}^{-4} J)$. Combining the above results, we have $\mathbb{E}\big|Q_{NT,2a}\big| = O(\underline{\nu}_{J_1J_2}^{-4} J r_{NT,J}^{1/2}) = o(J_1)$. It follows from Markov's inequality and (31) that $|Q_{NT,2a}| = o_p(J_1) = o_p(\mathbf{s}_N^2)$.

Lastly, to show $|Q_{NT,2b}| = o_p(\mathbf{s}_{N,1}^2)$, we note that $\mathbb{E}(Q_{NT,2b}) = 0$ and $\mathrm{Var}(Q_{NT,2b})$ is given by

$$\begin{aligned}
&\frac{1}{(NT)^4} \sum_{u=2}^{k_N-1} \mathbb{E} \left[\left(\sum_{s=1}^{u-1} \pi_{N,s}^\top \right) \Omega_{f_1} \sum_{v=u+1}^{k_N} \mathbb{E}\big(\pi_{N,v}\pi_{N,v}^\top\big) \Omega_{f_1} \pi_{N,u} \right]^2 \\
&= \frac{1}{(NT)^4} \sum_{u=2}^{k_N-1} \mathbb{E}\Bigg[\left(\sum_{s=1}^{u-1} \pi_{N,s}^\top \right) \Omega_{f_1} \sum_{v=u+1}^{k_N} \mathbb{E}\big(\pi_{N,v}\pi_{N,v}^\top\big) \Omega_{f_1} \mathbb{E}\big[\pi_{N,u}\pi_{N,u}^\top \mid \mathcal{F}_{N,u-1}\big] \\
&\quad \times \Omega_{f_1} \sum_{v=u+1}^{k_N} \mathbb{E}\big(\pi_{N,v}\pi_{N,v}^\top\big) \Omega_{f_1} \left(\sum_{s=1}^{u-1} \pi_{N,s} \right) \Bigg] \\
&\leq \frac{1}{(NT)^4} \sum_{u=2}^{k_N-1} \mathbb{E}\Bigg[\left(\sum_{s=1}^{NT} \pi_{N,s}^\top \right) \Omega_{f_1} \sum_{v=u+1}^{k_N} \mathbb{E}\big(\pi_{N,v}\pi_{N,v}^\top\big) \Omega_{f_1} \mathbb{E}\big[\pi_{N,u}\pi_{N,u}^\top \mid \mathcal{F}_{N,u-1}\big] \\
&\quad \times \Omega_{f_1} \sum_{v=u+1}^{k_N} \mathbb{E}\big(\pi_{N,v}\pi_{N,v}^\top\big) \Omega_{f_1} \left(\sum_{s=1}^{NT} \pi_{N,s} \right) \Bigg] \\
&= \frac{1}{(NT)^4} \sum_{u=1}^{k_N-1} \mathbb{E}\Bigg[\left(\sum_{s=1}^{NT} \pi_{N,s}^\top \right) \Omega_{f_1} \sum_{v=u+1}^{k_N} \mathbb{E}\big(\pi_{N,v}\pi_{N,v}^\top\big) \Omega_{f_1} \\
&\quad \times \Big[\mathbb{E}\big[\pi_{N,u}\pi_{N,u}^\top \mid \mathcal{F}_{N,u-1}\big] - \mathbb{E}\big(\pi_{N,u}\pi_{N,u}^\top\big) \Big] \Omega_{f_1} \sum_{v=u+1}^{k_N} \mathbb{E}\big(\pi_{N,v}\pi_{N,v}^\top\big) \Omega_{f_1} \left(\sum_{s=1}^{NT} \pi_{N,s} \right) \Bigg] \\
&\quad + \frac{1}{(NT)^4} \sum_{u=1}^{k_N-1} \sum_{s=1}^{k_N} \mathbb{E}\left[\pi_{N,s}^\top \Omega_{f_1} \sum_{v=u+1}^{k_N} \mathbb{E}\big(\pi_{N,v}\pi_{N,v}^\top\big) \Omega_{f_1} \mathbb{E}\big(\pi_{N,u}\pi_{N,u}^\top\big) \Omega_{f_1} \sum_{v=u+1}^{k_N} \mathbb{E}\big(\pi_{N,v}\pi_{N,v}^\top\big) \Omega_{f_1} \pi_{N,s} \right] \\
&= \mathcal{A}_{NT,1} + \mathcal{A}_{NT,2},
\end{aligned}$$

where the first inequality follows from the fact that $\{\pi_{N,v}\}$ is a martingale difference sequence, and thus $\mathbb{E}[\pi_{N,u}^\top \Omega_{f_1}\mathbb{E}(\pi_{N,v_1}\pi_{N,v_1}^\top)\Omega_{f_1}\mathbb{E}[\pi_{N,s}\pi_{N,s}^\top|\mathcal{F}_{N,s-1}]\Omega_{f_1}\mathbb{E}(\pi_{N,v_2}\pi_{N,v_2}^\top)\Omega_{f_1}\pi_{N,l}] = 0$ for all $1 \le u < s \le l \le k_N$ and $1 \le v_1, v_2 \le k_N$. Besides, it can be easily seen that $\mathbb{E}[(\sum_{s=u}^{k_N}\pi_{N,s})^\top \Omega_{f_1}\sum_{v=u+1}^{k_N}\mathbb{E}(\pi_{N,v}\pi_{N,v}^\top)\Omega_{f_1}\mathbb{E}[\pi_{N,u}\pi_{N,u}^\top|\mathcal{F}_{N,u-1}]\Omega_{f_1}\sum_{v=u+1}^{k_N}\mathbb{E}(\pi_{N,v}\pi_{N,v}^\top)\times \Omega_{f_1}(\sum_{s=u}^{k_N}\pi_{N,s})] \ge 0$ for any $1 \le u \le k_N$.

To show $Q_{NT,2b} = o_p(\mathbf{s}_{N,1}^2)$, it is sufficient to prove that $\mathcal{A}_{NT,1} = o(\mathbf{s}_{N,1}^4)$ and $\mathcal{A}_{NT,2} = o(\mathbf{s}_{N,1}^4)$. Letting $\overline{\Omega}_{f_1,v} \equiv \Omega_{f_1}\mathbb{E}(\pi_{N,v}\pi_{N,v}^\top)\Omega_{f_1}$, we can rewrite $|\mathcal{A}_{NT,1}|$ as

$$
\begin{aligned}
&\frac{1}{(NT)^4}\Bigg|\sum_{v=2}^{NT}\mathbb{E}\Bigg[\left(\sum_{s=1}^{NT}\pi_{N,s}\right)^\top \overline{\Omega}_{f_1,v}\sum_{u=1}^{v-1}\Big(\mathbb{E}[\pi_{N,u}\pi_{N,u}^\top|\mathcal{F}_{N,u-1}] - \mathbb{E}\big(\pi_{N,u}\pi_{N,u}^\top\big)\Big)\overline{\Omega}_{f_1,v}\left(\sum_{s=1}^{NT}\pi_{N,s}\right)\Bigg] \\
&+ 2\sum_{1<v_2<v_1\le NT}\mathbb{E}\Bigg[\left(\sum_{s=1}^{NT}\pi_{N,s}\right)^\top \overline{\Omega}_{f_1,v_1}\sum_{u=1}^{v_2-1}\Big(\mathbb{E}[\pi_{N,u}\pi_{N,u}^\top|\mathcal{F}_{N,u-1}] - \mathbb{E}\big(\pi_{N,u}\pi_{N,u}^\top\big)\Big)\overline{\Omega}_{f_1,v_2}\left(\sum_{s=1}^{NT}\pi_{N,s}\right)\Bigg]\Bigg| \\
&\le \frac{1}{(NT)^4}\sum_{v_1=2}^{NT}\sum_{v_2\ne v_1}^{NT}\|\overline{\Omega}_{f_1,v_1}\|\|\overline{\Omega}_{f_1,v_2}\|\mathbb{E}\left\|\sum_{s=1}^{NT}\pi_{N,s}\right\|^2\left\|\sum_{u=1}^{(v_1\wedge v_2)-1}\Big[\mathbb{E}[\pi_{N,u}\pi_{N,u}^\top|\mathcal{F}_{N,u-1}] - \mathbb{E}\big(\pi_{N,u}\pi_{N,u}^\top\big)\Big]\right\| \\
&\le \frac{CJ^2}{\underline{\nu}_{J_1J_2}^8(NT)^4}\sum_{v_1=1}^{NT}\sum_{v_2\ne v_1}^{NT}\left\|\sum_{s=1}^{NT}\pi_{N,s}\right\|_4^2\left\|\sum_{u=1}^{(v_1\wedge v_2)-1}\Big[\mathbb{E}[\pi_{N,u}\pi_{N,u}^\top|\mathcal{F}_{N,u-1}] - \mathbb{E}\big(\pi_{N,u}\pi_{N,u}^\top\big)\Big]\right\|_2 \\
&= O\big(\underline{\nu}_{J_1J_2}^{-8}J^3 r_{NT,J}^{1/2}\big) = o(\mathbf{s}_{N,1}^4),
\end{aligned}
$$

where we make use of triangular and CS inequalities, the fact that $\max_{1\le v\le k_N}\|\overline{\Omega}_{f_1,v}\| \le \lambda_{\max}^2(\Omega_{f_1})\cdot\max_{1\le v\le k_N}\|\mathbb{E}(\pi_{N,v}\pi_{N,v}^\top)\| \le C\underline{\nu}_{J_1J_2}^{-4}J$, (42) and Lemma 7. Next, $\mathcal{A}_{NT,2}$ can be written as

$$
\begin{aligned}
&\frac{1}{(NT)^3}\sum_{u=1}^{NT}\sum_{v_1=u+1}^{NT}\sum_{v_2=u+1}^{NT}\mathrm{tr}\big\{\Sigma_{N,f}\Omega_{f_1}\mathbb{E}\big(\pi_{N,v_1}\pi_{N,v_1}^\top\big)\Omega_{f_1}\mathbb{E}\big(\pi_{N,u}\pi_{N,u}^\top\big)\Omega_{f_1}\mathbb{E}\big(\pi_{N,v_2}\pi_{N,v_2}^\top\big)\Omega_{f_1}\big\} \\
&\le \frac{C}{\underline{\nu}_{J_1J_2}(NT)^3}\sum_{u=1}^{NT}\sum_{v_1=u+1}^{NT}\sum_{v_2=u+1}^{NT}\mathrm{tr}\big\{\mathbb{E}\big(\pi_{N,v_1}\pi_{N,v_1}^\top\big)\Omega_{f_1}\mathbb{E}\big(\pi_{N,u}\pi_{N,u}^\top\big)\Omega_{f_1}\mathbb{E}\big(\pi_{N,v_2}\pi_{N,v_2}^\top\big)\Omega_{f_1}\big\} \\
&= \frac{2C}{\underline{\nu}_{J_1J_2}(NT)^3}\sum_{1\le u<v_1<v_2\le NT}\mathrm{tr}\big\{\mathbb{E}\big(\pi_{N,v_1}\pi_{N,v_1}^\top\big)\Omega_{f_1}\mathbb{E}\big(\pi_{N,u}\pi_{N,u}^\top\big)\Omega_{f_1}\mathbb{E}\big(\pi_{N,v_2}\pi_{N,v_2}^\top\big)\Omega_{f_1}\big\} \\
&\quad \underbrace{+\frac{C}{\underline{\nu}_{J_1J_2}(NT)^3}\sum_{1\le u<v\le NT}\mathrm{tr}\big\{\mathbb{E}\big(\pi_{N,v}\pi_{N,v}^\top\big)\Omega_{f_1}\mathbb{E}\big(\pi_{N,u}\pi_{N,u}^\top\big)\Omega_{f_1}\mathbb{E}\big(\pi_{N,v}\pi_{N,v}^\top\big)\Omega_{f_1}\big\}}_{=O(J^3/(\underline{\nu}_{J_1J_2}^7 NT))},
\end{aligned} \tag{43}
$$

where we have used the fact that $\lambda_{\max}(\Omega_{f_1}\Sigma_f) = \lambda_{\max}(\mathbb{E}[\mathbb{P}_t^\top\mathbb{P}_t/N]\mathcal{S}_1\Sigma_{N,f}^{-1}\mathcal{S}_1^\top) \le C\underline{\nu}_{J_1J_2}^{-1}$ according to Assumption 5′. Further, we note that

$$
\begin{aligned}
&\operatorname{tr}((\Sigma_{N,f}\Omega_{f_1})^3) \\
&= \frac{1}{(NT)^3}\sum_{u=1}^{NT}\sum_{v_1=1}^{NT}\sum_{v_2=1}^{NT}\operatorname{tr}\big\{\mathbb{E}\big(\pi_{N,v_1}\pi_{N,v_1}^\top\big)\Omega_{f_1}\mathbb{E}\big(\pi_{N,u}\pi_{N,u}^\top\big)\Omega_{f_1}\mathbb{E}\big(\pi_{N,v_2}\pi_{N,v_2}^\top\big)\Omega_{f_1}\big\} \\
&= \frac{6}{(NT)^3}\sum_{1\le u<v_1<v_2\le NT}\operatorname{tr}\big\{\mathbb{E}\big(\pi_{N,v_1}\pi_{N,v_1}^\top\big)\Omega_{f_1}\mathbb{E}\big(\pi_{N,u}\pi_{N,u}^\top\big)\Omega_{f_1}\mathbb{E}\big(\pi_{N,v_2}\pi_{N,v_2}^\top\big)\Omega_{f_1}\big\} \\
&\quad + \frac{3}{(NT)^3}\sum_{v_1=1}^{NT}\sum_{v_2\neq v_1}^{NT}\operatorname{tr}\big\{\mathbb{E}\big(\pi_{N,v_1}\pi_{N,v_1}^\top\big)\Omega_{f_1}\mathbb{E}\big(\pi_{N,v_1}\pi_{N,v_1}^\top\big)\Omega_{f_1}\mathbb{E}\big(\pi_{N,v_2}\pi_{N,v_2}^\top\big)\Omega_{f_1}\big\} \\
&\quad + \frac{1}{(NT)^3}\sum_{v=1}^{NT}\operatorname{tr}\big\{\mathbb{E}\big(\pi_{N,v}\pi_{N,v}^\top\big)\Omega_{f_1}\mathbb{E}\big(\pi_{N,v}\pi_{N,v}^\top\big)\Omega_{f_1}\mathbb{E}\big(\pi_{N,v}\pi_{N,v}^\top\big)\Omega_{f_1}\big\} \\
&= \frac{6}{(NT)^3}\sum_{1\le u<v_1<v_2\le NT}\operatorname{tr}\big\{\mathbb{E}\big(\pi_{N,v_1}\pi_{N,v_1}^\top\big)\Omega_{f_1}\mathbb{E}\big(\pi_{N,u}\pi_{N,u}^\top\big)\Omega_{f_1}\mathbb{E}\big(\pi_{N,v_2}\pi_{N,v_2}^\top\big)\Omega_{f_1}\big\} + O\bigg(\frac{J^3}{\underline{\nu}_{J_1J_2}^6 NT}\bigg).
\end{aligned}
$$

By (31), Assumptions 4′ and 5′, $\operatorname{tr}((\Sigma_{N,f}\Omega_{f_1})^3) \le C\underline{\nu}_{J_1J_2}^{-1}\operatorname{tr}((\Sigma_{N,f}\Omega_{f_1})^2) \le C'\underline{\nu}_{J_1J_2}^{-1}\mathbf{s}_{N,1}^2$. Combining the above results, we have

$$
\begin{aligned}
&\frac{6}{\underline{\nu}_{J_1J_2}(NT)^3}\sum_{1\le u<v_1<v_2\le NT}\operatorname{tr}\big\{\mathbb{E}\big(\pi_{N,v_1}\pi_{N,v_1}^\top\big)\Omega_{f_1}\mathbb{E}\big(\pi_{N,u}\pi_{N,u}^\top\big)\Omega_{f_1}\mathbb{E}\big(\pi_{N,v_2}\pi_{N,v_2}^\top\big)\Omega_{f_1}\big\} \\
&= \underline{\nu}_{J_1J_2}^{-1}\operatorname{tr}((\Sigma_{N,f}\Omega_{f_1})^3) + O(J^3/(\underline{\nu}_{J_1J_2}^7 NT)) = O(\underline{\nu}_{J_1J_2}^{-2}\mathbf{s}_{N,1}^2) + O(J^3/(\underline{\nu}_{J_1J_2}^7 NT)) = o(\mathbf{s}_{N,1}^4)
\end{aligned}
$$

by (31) and Assumption 5′. Hence, $\mathcal{A}_{NT,2} = o(\mathbf{s}_{N,1}^4)$. This shows that $\mathbb{E}|Q_{NT,2b}|^2 = o(\mathbf{s}_{N,1}^4)$, implying that $|Q_{NT,2b}| = o_p(\mathbf{s}_{N,1}^2)$. We have now verified condition (38). By Theorem 1 in Kuersteiner and Prucha (2013), we have $\sum_{v=1}^{k_N}\zeta_{N,v} \xrightarrow{d} N(0,1)$, which implies the desired convergence in distribution. □

Lemma 7. *Suppose that Assumptions 1′, 2, 3 and 4′ hold. Then, (i)* $\big\|(NT)^{-1}\sum_{v=1}^{NT}\big(\pi_{N,v}\pi_{N,v}^\top - \mathbb{E}[\pi_{N,v}\pi_{N,v}^\top|\mathcal{F}_{N,v-1}]\big)\big\| = O_p(J^{1/2}\zeta_{0,J}/\sqrt{NT})$; *(ii)* $\max_{1\le s\le NT}\mathbb{E}\|\sum_{v=1}^{s}\pi_{N,v}\|^4 \le C(NT)^2J^2$, *and* $\max_{1\le s\le NT}\mathbb{E}\|\sum_{v=1}^{s}\pi_{N,v}\|^6 \le C(NT)^3J^3$ *for some finite constant C.*

Proof. Denote the $J\times J$ matrix $\mathscr{G}_{NT} \equiv (NT)^{-1}\sum_{v=1}^{NT}\big(\pi_{N,v}\pi_{N,v}^\top - \mathbb{E}[\pi_{N,v}\pi_{N,v}^\top|\mathcal{F}_{N,v-1}]\big)$. It is easy to see that each element of $\mathscr{G}_{NT}$ is a sum of the martingale difference array with respect to the filtration defined in (16). By Assumption 1′,

$$
\begin{aligned}
\mathbb{E}\|\mathscr{G}_{NT}\|^2 &= \frac{1}{(NT)^2}\left\|\sum_{i=1}^{N}\sum_{t=1}^{T} H_{i,t-1}H_{i,t-1}^\top(\epsilon_{it}^2-\sigma^2)\right\|^2 \\
&= \frac{1}{(NT)^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}(\epsilon_{it}^2-\sigma^2)^2\cdot \mathrm{tr}\big(H_{i,t-1}H_{i,t-1}^\top H_{i,t-1}H_{i,t-1}^\top\big) \\
&= \frac{1}{(NT)^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}\mathrm{tr}\big(H_{i,t-1}H_{i,t-1}^\top H_{i,t-1}H_{i,t-1}^\top\big)\mathbb{E}(\epsilon_{it}^2-\sigma^2)^2 \\
&\le \frac{C\zeta_{0,J}^2}{NT}\mathbb{E}\mathrm{tr}\left(\frac{1}{NT}\sum_{t=1}^{T}\mathbb{H}_{t-1}^\top\mathbb{H}_{t-1}\right) \\
&= O_p(J\zeta_{0,J}^2/(NT)),
\end{aligned}
$$

where the second equality is due to the martingale property, the third holds because $H_{i,t-1}$ is independent of ϵ_{it} due to Assumption 1, the first inequality uses the fact that $\max_i |H_{i,t-1}|^2 = O(\zeta_{0,J}^2)$ and the last equality is from Assumption 4. Lemma 7(i) follows immediately from Markov's inequality.

To show result (ii), denote the jth element of $\pi_{N,v}$ as $\pi_{N,v}^{(j)}$, for $j = 1, \ldots, (J_1+J_2)$. For random variable X, denote $\|X\|_p \equiv \{\mathbb{E}|X|^p\}^{1/p}$ for $p \ge 1$. Then, by Minkowski's inequality, for $1 \le s \le NT$,

$$
\begin{aligned}
\left[\mathbb{E}\left\|\sum_{v=1}^{s}\pi_{N,v}\right\|^4\right]^{1/2} &= \left\|\sum_{j=1}^{J_1+J_2}\left(\sum_{v=1}^{s}\pi_{N,v}^{(j)}\right)^2\right\|_2 \le \sum_{j=1}^{J_1+J_2}\left\|\sum_{v=1}^{s}\pi_{N,v}^{(j)}\right\|_4^2 \\
&\le C_1\sum_{j=1}^{J_1+J_2}\sum_{v=1}^{s}\left\|\pi_{N,v}^{(j)}\right\|_4^2 \le C_2(J_1+J_2)s,
\end{aligned}
$$

where the second inequality follows from Lemma 1 of Wu and Shao (2007), and the last inequality follows from Assumptions 1′ and 4′. This proves the first part of Lemma 7(ii). The second part can be proved similarly. The details are omitted for brevity. □

B.5 Proof of Theorem 4

(i) As in (32), decompose $\mathbf{T}_{NT,1}$ into three parts $\mathbf{T}_{NT,1a}+\mathbf{T}_{NT,1b}+\mathbf{T}_{NT,1c}$. Note that $\mathbf{s}_{N,1}^{-1}(\mathbf{T}_{NT,1}-\mathbf{B}_{N,1})=\mathbf{s}_{N,1}^{-1}(\mathbf{T}_{NT,1a}-\mathbf{B}_{N,1})+\mathbf{s}_{N,1}^{-1}\mathbf{T}_{NT,1b}+\mathbf{s}_{N,1}^{-1}\mathbf{T}_{NT,1c}$. Theorem 4(i) follows if we can prove that (i) $\mathbf{s}_{N,1}^{-1}(\mathbf{T}_{NT,1a}-\mathbf{B}_{N,1})\xrightarrow{d}N(0,1)$, (ii) $\mathbf{s}_{N,1}^{-1}\mathbf{T}_{NT,1b}=\Delta_1+o_p(1)$, and (iii) $\mathbf{s}_{N,1}^{-1}\mathbf{T}_{NT,1c}=o_p(1)$ under $H_1^n(\alpha_{NT})$. The assertion (i) has been shown in Lemma 6. To show (ii), note that under $H_1^n(\alpha_{NT})$, $f_1(Y_{i,t-1})=\rho_1 Y_{i,t-1}+\alpha_{NT}r(Y_{i,t-1})$. Thus,

$$\mathbf{T}_{NT,1b}=\sum_{i=1}^{N}\sum_{t=1}^{T}\left(f_1(Y_{i,t-1})-\widehat{\rho}_1 Y_{i,t-1}\right)^2=\sum_{i=1}^{N}\sum_{t=1}^{T}\left[(\rho_1-\widehat{\rho}_1)Y_{i,t-1}+\alpha_{NT}r(Y_{i,t-1})\right]^2.$$

To derive the asymptotic expansion of $\widehat{\rho}_1-\rho_1$, we denote $\mathbb{U}_{t-1}\equiv(\mathbb{P}_{t-1}^{J_2},\mathbb{Z})$, $\mathbf{U}\equiv(\mathbb{U}_0^\top,\ldots,\mathbb{U}_{T-1}^\top)^\top$, $\mathbf{Y}_{-1}\equiv(\mathbb{Y}_0^\top,\ldots,\mathbb{Y}_{T-1}^\top)^\top$, $\mathbf{P}^J\equiv(\mathbb{P}_0^{J\top},\ldots,\mathbb{P}_{T-1}^{J\top})^\top$, and $\mathbf{W}\equiv I_T\otimes W$. Moreover, define $\mathbf{R}(\mathbf{Y}_{-1})\equiv(R(\mathbb{Y}_0)^\top,\ldots,R(\mathbb{Y}_{T-1})^\top)^\top$, and $\mathbf{F}_2(\mathbf{Y}_{-1})\equiv(F_2(\mathbb{Y}_0)^\top,\ldots,F_2(\mathbb{Y}_{T-1})^\top)^\top$. Then,

$$\begin{aligned}\widehat{\rho}_1-\rho_1&=\left(\mathbf{Y}_{-1}^\top\mathbf{W}^\top\mathbf{M}_{\mathbf{U}}\mathbf{W}\mathbf{Y}_{-1}\right)^{-1}\mathbf{Y}_{-1}^\top\mathbf{W}^\top\mathbf{M}_{\mathbf{U}}\mathbf{Y}-\rho_1\\&=\widehat{\Sigma}_{NT,1}^{-1}\mathbf{Y}_{-1}^\top\mathbf{W}^\top\mathbf{M}_{\mathbf{U}}\left(\alpha_{NT}\mathbf{W}\mathbf{R}(\mathbf{Y}_{-1})+\boldsymbol{\mathcal{E}}+\mathbf{F}_2(\mathbf{Y}_{-1})-\mathbf{P}^{J_2}\beta_2\right)/(NT)\\&\equiv\mathcal{Q}_{NT,1}+\mathcal{Q}_{NT,2}+\mathcal{Q}_{NT,3},\end{aligned}$$

where $\widehat{\Sigma}_{NT,1}\equiv\mathbf{Y}_{-1}^\top\mathbf{W}^\top\mathbf{M}_{\mathbf{U}}\mathbf{W}\mathbf{Y}_{-1}/(NT)$.

We will show that (a) $\mathcal{Q}_{NT,1}=O_p(\alpha_{NT})$, (b) $\mathcal{Q}_{NT,2}=O_p((NT)^{-1/2})$, and (c) $\mathcal{Q}_{NT,3}=O_p(J_2^{-\mu})$. Denote $\Psi_{J_2,11}\equiv\mathbb{E}((W\mathbb{Y}_t)^\top W\mathbb{Y}_t/N)$, $\Psi_{J_2,12}\equiv\mathbb{E}((W\mathbb{Y}_t)^\top\mathbb{U}_t/N)$, $\Psi_{J_2,22}\equiv\mathbb{E}(\mathbb{U}_t^\top\mathbb{U}_t/N)$, and $\Sigma_{N,1}\equiv\Psi_{J_2,11}-\Psi_{J_2,12}\Psi_{J_2,22}^{-1}\Psi_{J_2,12}^\top$. Then,

$$|\widehat{\Sigma}_{NT,1}^{-1}-\Sigma_{N,1}^{-1}|=\left|e_1^\top\left(\widehat{\Psi}_{NT,11}^{-1}-\Psi_{N,11}^{-1}\right)e_1\right|\le\|\widehat{\Psi}_{NT,11}^{-1}-\Psi_{N,11}^{-1}\|=O_p(\underline{\nu}_{J_1J_2}^{-2}r_{NT,J}^{1/2}),$$

which can be shown in a similar way to Lemmas 2 and 3. By Assumption 4′(iii), $|e_1^\top\Psi_{N,11}^{-1}e_1|<\bar{c}<\infty$ for sufficiently large N. Hence, the last display implies that $|\widehat{\Sigma}_{NT,1}^{-1}|<c<\infty$ for some $0 < c < \infty$ wpa1. Then,

$$\begin{aligned}|\mathcal{Q}_{NT,1}|^2 =& \alpha_{NT}^2 (NT)^{-2} \mathbf{R}^\top(\mathbf{Y}_{-1})\mathbf{W}^\top \mathbf{M}_\mathbf{U} \mathbf{W}\mathbf{Y}_{-1}\widehat{\Sigma}_{NT,1}^{-2}\mathbf{Y}_{-1}^\top \mathbf{W}^\top \mathbf{M}_\mathbf{U}\mathbf{W}\mathbf{R}(\mathbf{Y}_{-1})\\ \le& O_p(\alpha_{NT}^2)\cdot \mathbf{R}^\top(\mathbf{Y}_{-1})\mathbf{W}^\top\mathbf{W}\mathbf{R}(\mathbf{Y}_{-1})/(NT) = O_p(\alpha_{NT}^2).\end{aligned}$$

Next, for $\mathcal{Q}_{NT,2}$, we write $\mathcal{Q}_{NT,2} = \mathcal{Q}_{NT,2a} + \mathcal{Q}_{NT,2b}$, where

$$\begin{aligned}\mathcal{Q}_{NT,2a} &\equiv \Sigma_{N,1}^{-1}\boldsymbol{\eta}^\top \boldsymbol{\mathcal{E}}/(NT),\\ \mathcal{Q}_{NT,2b} &\equiv (\widehat{\Sigma}_{NT,1}^{-1}\mathbf{Y}_{-1}^\top \mathbf{W}^\top \mathbf{M}_\mathbf{U} - \Sigma_{N,1}^{-1}\boldsymbol{\eta}^\top)\boldsymbol{\mathcal{E}}/(NT),\end{aligned}$$

and $\boldsymbol{\eta} \equiv \mathbf{W}\mathbf{Y}_{-1} - \mathbf{U}\Psi_{J_2,22}^{-1}\Psi_{J_2,12}^\top$. By Assumption 4′, $\mathbb{E}|\mathcal{Q}_{NT,2a}|^2 = \sigma^2\mathrm{tr}(\Sigma_{N,1}^{-1}\Sigma_{N,1}\Sigma_{N,1}^{-1})/(NT) = O((NT)^{-1})$. It follows from Markov's inequality that $|\mathcal{Q}_{NT,2a}| = O_p((NT)^{-1/2})$. Let $\widehat{\Psi}_{J_2,12} = (NT)^{-1}\sum_{t=1}^T (W\mathbb{Y}_{t-1})^\top \mathbb{U}_{t-1}$ and $\widehat{\Psi}_{J_2,22} = (NT)^{-1}\sum_{t=1}^T \mathbb{U}_{t-1}^\top\mathbb{U}_{t-1}$. By triangular inequality,

$$\begin{aligned}\left|\mathcal{Q}_{NT,2b}\right| =& \left|\widehat{\Sigma}_{NT,1}^{-1}(\Psi_{J_2,12}\Psi_{J_2,22}^{-1} - \widehat{\Psi}_{J_2,12}\widehat{\Psi}_{J_2,22}^{-1})\mathbf{U}^\top\boldsymbol{\mathcal{E}}/(NT) + (\widehat{\Sigma}_{NT,1}^{-1} - \Sigma_{N,1}^{-1})\boldsymbol{\eta}^\top\boldsymbol{\mathcal{E}}/(NT)\right|\\ \le& O_p(1)\cdot\left\|\Psi_{J_2,12}\Psi_{J_2,22}^{-1} - \widehat{\Psi}_{J_2,12}\widehat{\Psi}_{J_2,22}^{-1}\right\|\|\mathbf{U}^\top\boldsymbol{\mathcal{E}}/(NT)\| + |\widehat{\Sigma}_{NT,1}^{-1} - \Sigma_{N,1}^{-1}|\|\boldsymbol{\eta}^\top\boldsymbol{\mathcal{E}}/(NT)\|\\ =& O_p(\underline{\nu}_{J_1J_2}^{-2} r_{NT,J}^{1/2}\sqrt{J_2/(NT)}) = o_p((NT)^{-1/2}),\end{aligned}$$

where we have used the fact that $\|\widehat{\Psi}_{J_2} - \Psi_{J_2}\| = O_p(r_{NT,J}^{1/2})$, $\sigma_{\max}^2(\widehat{\Psi}_{J_2,12}) = O_p(1)$, $\lambda_{\max}(\widehat{\Psi}_{J_2,22}^{-1}) = O_p(\underline{\nu}_{J_1J_2}^{-1})$, $\|\mathbf{U}^\top\boldsymbol{\mathcal{E}}/(NT)\| = O_p(\sqrt{J_2/(NT)})$ and Assumption 5′. Hence, $\mathcal{Q}_{NT,2} = \Sigma_{N,1}^{-1}\boldsymbol{\eta}^\top\boldsymbol{\mathcal{E}}/(NT) + o_p((NT)^{-1/2}) = O_p((NT)^{-1/2})$. Finally, using the same arguments as (30), we can show that $\mathcal{Q}_{NT,3} = O_p(J_2^{-\mu})$. Combining the results (a)-(c), we have

$$(\widehat{\rho}_1 - \rho_1)^2 = (\mathcal{Q}_{NT,1} + \mathcal{Q}_{NT,2} + \mathcal{Q}_{NT,3})^2 = \mathcal{Q}_{NT,1}^2 + o_p(\alpha_{NT}^2)$$

observing that $\mathcal{Q}_{NT,2} = o_p(\alpha_{NT})$ and $\mathcal{Q}_{NT,3} = o_p(\alpha_{NT})$ under the maintained assumptions and the choice that $\alpha_{NT} = \mathbf{s}_{N,1}^{1/2}(NT)^{-1/2}$.

Further, $\mathcal{Q}_{NT,1} = \alpha_{NT}e_1^\top\widehat{\Psi}_{NT,11}^{-1}\widehat{\Gamma}_{NT,1}$, where $\widehat{\Psi}_{NT,11} = (NT)^{-1}\sum_{t=1}^T \mathbb{X}_{t-1,1}^\top\mathbb{X}_{t-1,1}$ and

$\widehat{\Gamma}_{NT,1} = (NT)^{-1}\sum_{t=1}^{T}\mathbb{X}_{t-1,1}^{\top}WR(\mathbb{Y}_{t-1})$. Then,

$$\begin{aligned}
\mathcal{Q}_{NT,1} &= \alpha_{NT}e_1^{\top}\Psi_{N,11}^{-1}\Gamma_{N,1} + \alpha_{NT}e_1^{\top}(\widehat{\Psi}_{NT,11}^{-1}\widehat{\Gamma}_{NT,1} - \Psi_{N,11}^{-1}\Gamma_{N,1}) \\
&= \alpha_{NT}e_1^{\top}\Psi_{N,11}^{-1}\Gamma_{N,1} + \alpha_{NT}e_1^{\top}\left[(\widehat{\Psi}_{NT,11}^{-1} - \Psi_{N,11}^{-1})\Gamma_{N,1} + \widehat{\Psi}_{NT,11}^{-1}(\widehat{\Gamma}_{NT,1} - \Gamma_{N,1})\right] \\
&= \alpha_{NT}e_1^{\top}\Psi_{N,11}^{-1}\Gamma_{N,1} + O_p(\alpha_{NT}r_{NT,J}^{1/2}J_2^{1/2}) + O_p(\alpha_{NT}\underline{\nu}_{J_1J_2}^{-1}r_{NT,J}^{1/2}),
\end{aligned}$$

where we have used the fact that $\|\Gamma_{N,1}\| = O(J_2^{1/2})$ and $\|\widehat{\Gamma}_{NT,1} - \Gamma_{N,1}\| = O_p(\sqrt{\zeta_{0,J}^2/(NT) + \zeta_{0,J}(\zeta_{0,J}+\zeta_{1,J})\min\{\Xi_{N,0}, T^{-1}\Xi_{N,1}\}}) = O_p(r_{NT,J}^{1/2})$. It follows that $\mathcal{Q}_{NT,1}^2 = \alpha_{NT}^2(e_1^{\top}\Psi_{N,11}^{-1}\Gamma_{N,1})^2 + o_p(\alpha_{NT}^2)$ under Assumption 5′. Combining the above results, we have

$$(\widehat{\rho}_1 - \rho_1)^2 = \mathcal{Q}_{NT,1}^2 + o_p(\alpha_{NT}^2) = \alpha_{NT}^2(e_1^{\top}\Psi_{N,11}^{-1}\Gamma_{N,1})^2 + o_p(\alpha_{NT}^2).$$

Using similar arguments, we can easily show that $(\widehat{\rho}_1-\rho_1)\alpha_{NT} = \alpha_{NT}^2 e_1^{\top}\Psi_{N,11}^{-1}\Gamma_{N,1} + o_p(\alpha_{NT}^2)$. Recalling that $\alpha_{NT} = \mathbf{s}_{N,1}^{1/2}(NT)^{-1/2}$, we have

$$\begin{aligned}
&\mathbf{s}_{N,1}^{-1}\mathbf{T}_{NT,1b} \\
&= \mathbf{s}_{N,1}^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{(\widehat{\rho}_1 - \rho_1)^2Y_{i,t-1}^2 - 2(\widehat{\rho}_1-\rho_1)\alpha_{NT}Y_{i,t-1}r(Y_{i,t-1}) + \alpha_{NT}^2r^2(Y_{i,t-1})\right\} \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{(e_1^{\top}\Psi_{N,11}^{-1}\Gamma_{N,1})^2Y_{i,t-1}^2 - 2(e_1^{\top}\Psi_{N,11}^{-1}\Gamma_{N,1})Y_{i,t-1}r(Y_{i,t-1}) + r^2(Y_{i,t-1})\right\} + o_p(1) \\
&= \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[r(Y_{i,t-1}) - (e_1^{\top}\Psi_{N,11}^{-1}\Gamma_{N,1})Y_{i,t-1}\right]^2 + o_p(1) = \Delta_1 + o_p(1),
\end{aligned}$$

where we have used the fact that $(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\xi_{i,t-1} = N^{-1}\sum_{i=1}^{N}\mathbb{E}\xi_{i,t-1} + o_p(1)$ with $\xi_{i,t-1}$ being one of $Y_{i,t-1}^2$, $r^2(Y_{i,t-1})$ and $Y_{i,t-1}r(Y_{i,t-1})$. This proves assertion (b).

Next, we show (iii) $\mathbf{s}_{N,1}^{-1}\mathbf{T}_{NT,1c} = o_p(1)$. In the same manner as Lemma 5, we have the following decomposition of $\mathbf{T}_{NT,1c}$ under $H_1^n(\alpha_{NT})$:

$$\begin{aligned}
&\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\widehat{f}_1(Y_{i,t-1}) - f_1(Y_{i,t-1})\right)\left(f_1(Y_{i,t-1}) - \widehat{\rho}_1Y_{i,t-1}\right) \\
&= (\rho_1 - \widehat{\rho}_1)\sum_{i=1}^{N}\sum_{t=1}^{T}\left(p^{J_1}(Y_{i,t-1})^{\top}\widetilde{U}_1 + p^{J_1}(Y_{i,t-1})^{\top}\widetilde{U}_2 + p^{J_1}(Y_{i,t-1})^{\top}\widetilde{U}_3 - u_1(Y_{i,t-1})\right)Y_{i,t-1}
\end{aligned}$$

$$
\begin{aligned}
&+\alpha_{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\big(p^{J_1}(Y_{i,t-1})^\top\widetilde{U}_1+p^{J_1}(Y_{i,t-1})^\top\widetilde{U}_2+p^{J_1}(Y_{i,t-1})^\top\widetilde{U}_3-u_1(Y_{i,t-1})\big)r(Y_{i,t-1})\\
=&\sum_{t=1}^{T}\big((\rho_1-\widehat{\rho}_1)\mathbb{Y}_{t-1}+\alpha_{NT}R(\mathbb{Y}_{t-1})\big)^\top\mathbb{P}_{t-1}^{J_1}\widetilde{U}_1\\
&+\underbrace{\sum_{t=1}^{T}\big((\rho_1-\widehat{\rho}_1)\mathbb{Y}_{t-1}+\alpha_{NT}R(\mathbb{Y}_{t-1})\big)^\top\mathbb{P}_{t-1}^{J_1}\widetilde{U}_2}_{=O_p\left(\alpha_{NT}\underline{\nu}_{J_1J_2}^{-2}r_{NT,J}^{1/2}\sqrt{J_1JNT}\right)}+\underbrace{\sum_{t=1}^{T}\big((\rho_1-\widehat{\rho}_1)\mathbb{Y}_{t-1}+\alpha_{NT}R(\mathbb{Y}_{t-1})\big)^\top\mathbb{P}_{t-1}^{J_1}\widetilde{U}_3}_{=O_p\left(\alpha_{NT}NT\sqrt{J_1/\underline{\nu}_{J_1J_2}}(J_1^{-\mu}+J_2^{-\mu})\right)}\\
&-\underbrace{\sum_{i=1}^{N}\sum_{t=1}^{T}\big((\rho_1-\widehat{\rho}_1)Y_{i,t-1}+\alpha_{NT}r(Y_{i,t-1})\big)u_1(Y_{i,t-1})}_{=O_p(\alpha_{NT}NTJ_1^{-\mu})}\\
=&(\rho_1-\widehat{\rho}_1)\left(\sum_{t=1}^{T}\mathbb{Y}_{t-1}^\top\mathbb{P}_{t-1}^{J_1}\right)\widetilde{U}_1+\alpha_{NT}\left(\sum_{t=1}^{T}R(\mathbb{Y}_{t-1})^\top\mathbb{P}_{t-1}^{J_1}\right)\widetilde{U}_1+o_p(J_1^{1/2}), \qquad (44)
\end{aligned}
$$

where the second equality uses the fact that $\widehat{\rho}_1-\rho_1=O_p(\alpha_{NT})$, $\|\sum_{t=1}^{T}\mathbb{P}_{t-1}^{J_1\top}R(\mathbb{Y}_{t-1})\|=O_p(NT\sqrt{J_1})$, and $\|\sum_{t=1}^{T}\mathbb{P}_{t-1}^{J_1\top}\mathbb{Y}_{t-1}\|=O_p(NT\sqrt{J_1})$, and the last equality is from Assumption 5′. Recall from the proof of Lemma 5 that $\big(\sum_{t=1}^{T}\mathbb{Y}_{t-1}^\top\mathbb{P}_{t-1}^{J_1}\big)\widetilde{U}_1=\tilde{U}_1^{(1)}+\tilde{U}_1^{(2)}$, and $|\tilde{U}_1^{(1)}|=O_p(\sqrt{NT/\underline{\nu}_{J_1J_2}})$. Therefore, $(\widehat{\rho}_1-\rho_1)\tilde{U}_1^{(1)}=O_p(\alpha_{NT}\sqrt{NT/\underline{\nu}_{J_1J_2}})=O_p(\mathbf{s}_{N,1}^{1/2}\underline{\nu}_{J_1J_2}^{-1/2})=o_p(\mathbf{s}_{N,1})$ under Assumption 5′. Moreover, by (34), we have $(\widehat{\rho}_1-\rho_1)\tilde{U}_1^{(2)}=O_p(\alpha_{NT}\sqrt{J_1\zeta_{0,J_1}^2/\underline{\nu}_{J_1J_2}}+\alpha_{NT}\sqrt{J_1\zeta_{0,J_1}(\zeta_{0,J_1}+\zeta_{1,J_1})\min\{\Xi_{N,0},T^{-1}\Xi_{N,1}\}NT/\underline{\nu}_{J_1J_2}})=o_p(\mathbf{s}_{N,1})$. Consequently, the first term on the r.h.s. of (44) is $o_p(\mathbf{s}_{N,1})$. Similarly, we can show that the second term on the r.h.s. of (44) is also $o_p(\mathbf{s}_{N,1})$. This completes the proof of assertion (iii).

(ii) Theorem 4(ii) can be shown by slightly modifying the arguments used in Theorem 3(ii). Indeed, under $H_1^n(\alpha_{NT})$, we have $F_1(\mathbb{Y}_{t-1})-\mathbb{P}_{t-1}^{J_1}\beta_1=F_1(\mathbb{Y}_{t-1})-\beta_{1,1}\mathbb{Y}_{t-1}=\alpha_{NT}R(\mathbb{Y}_{t-1})$, and thus $\mathbf{Y}-\mathbf{X}_1\bar{\theta}_1=\mathbf{M}_{\mathbf{X}_1}(\alpha_{NT}\mathbf{W}\mathbf{R}(\mathbf{Y}_{-1})+\mathbf{R}_2+\boldsymbol{\mathcal{E}})$, where $\mathbf{R}_2\equiv(\mathbb{R}_{0,2}^\top,\mathbb{R}_{1,2}^\top,\ldots,\mathbb{R}_{T-1,2}^\top)^\top$ with $\mathbb{R}_{t-1,2}=F_2(\mathbb{Y}_{t-1})-\mathbb{P}_{t-1}^{J_2}\beta_2$. Hence, the test statistic in (11) can be decomposed as $\mathbf{LM}_{NT,1}=\mathbf{LM}_{NT,1}+2\mathbf{LM}_{NT,2}+\mathbf{LM}_{NT,3}$, where

$$
\begin{aligned}
\mathbf{LM}_{NT,1}&\equiv(\mathbf{R}_2+\boldsymbol{\mathcal{E}})^\top\mathbf{M}_{\mathbf{X}_1}\mathbf{X}_2\big(\mathbf{X}_2^\top\mathbf{M}_{\mathbf{X}_1}\mathbf{X}_2\big)^-\mathbf{X}_2^\top\mathbf{M}_{\mathbf{X}_1}(\mathbf{R}_2+\boldsymbol{\mathcal{E}})/\widehat{\sigma}^2,\\
\mathbf{LM}_{NT,2}&\equiv\alpha_{NT}(\mathbf{W}\mathbf{R}(\mathbf{Y}_{-1}))^\top\mathbf{M}_{\mathbf{X}_1}\mathbf{X}_2\big(\mathbf{X}_2^\top\mathbf{M}_{\mathbf{X}_1}\mathbf{X}_2\big)^-\mathbf{X}_2^\top\mathbf{M}_{\mathbf{X}_1}(\mathbf{R}_2+\boldsymbol{\mathcal{E}})/\widehat{\sigma}^2,\\
\mathbf{LM}_{NT,3}&\equiv\alpha_{NT}^2(\mathbf{W}\mathbf{R}(\mathbf{Y}_{-1}))^\top\mathbf{M}_{\mathbf{X}_1}\mathbf{X}_2\big(\mathbf{X}_2^\top\mathbf{M}_{\mathbf{X}_1}\mathbf{X}_2\big)^-\mathbf{X}_2^\top\mathbf{M}_{\mathbf{X}_1}\mathbf{W}\mathbf{R}(\mathbf{Y}_{-1})/\widehat{\sigma}^2.
\end{aligned}
$$

We have shown in the proof of Theorem 3 that $(\mathbf{LM}_{NT,1}-(J_1-1))/\sqrt{2(J_1-1)} \xrightarrow{d} N(0,1)$. The desired result follows if we can show that (d) $\mathbf{LM}_{NT,2}=o_p(J_1^{1/2})$, and (e) $\mathbf{LM}_{NT,3}/\sqrt{2(J_1-1)} = \Delta_2+o_p(1)$. We decompose $\mathbf{LM}_{NT,2}$ into two parts $\mathbf{LM}_{NT,2a}+\mathbf{LM}_{NT,2b}$. First, noting that the eigenvalue of an idempotent matrix is 1 or 0, we have

$$\begin{aligned}\left|\mathbf{LM}_{NT,2a}\right| \equiv& \left|\alpha_{NT}(\mathbf{WR}(\mathbf{Y}_{-1}))^\top\mathbf{M}_{\mathbf{X}_1}\mathbf{X}_2\big(\mathbf{X}_2^\top\mathbf{M}_{\mathbf{X}_1}\mathbf{X}_2\big)^{-}\mathbf{X}_2^\top\mathbf{M}_{\mathbf{X}_1}\mathbf{R}_2/\widehat{\sigma}^2\right|\\ &\le O_p(1)\|\alpha_{NT}\mathbf{WR}(\mathbf{Y}_{-1})\|\|\mathbf{R}_2\| = O_p(J_1^{1/4}\sqrt{NT}J_2^{-\mu}) = o_p(J_1^{1/2}).\end{aligned}$$

Moreover, it also holds that

$$\left|\mathbf{LM}_{NT,2b}\right| \equiv \left|\alpha_{NT}(\mathbf{WR}(\mathbf{Y}_{-1}))^\top\mathbf{M}_{\mathbf{X}_1}\mathbf{X}_2\big(\mathbf{X}_2^\top\mathbf{M}_{\mathbf{X}_1}\mathbf{X}_2\big)^{-}\mathbf{X}_2^\top\mathbf{M}_{\mathbf{X}_1}\boldsymbol{\mathcal{E}}/\widehat{\sigma}^2\right| = o_p(J_1^{1/2}).$$

To show this, we write $\mathbf{LM}_{NT,2b}=\mathbf{LM}^{(1)}_{NT,2b}+\mathbf{LM}^{(2)}_{NT,2b}+\mathbf{LM}^{(3)}_{NT,2b}$, where $\mathbf{LM}^{(1)}_{NT,2b}\equiv \alpha_{NT}\Lambda_N^\top\Theta_N^{-1}\boldsymbol{\nu}^\top\boldsymbol{\mathcal{E}}/\widehat{\sigma}^2$, $\mathbf{LM}^{(2)}_{NT,2b} \equiv \alpha_{NT}(\widehat{\Lambda}_{NT}^\top-\Lambda_N^\top)\Theta_N^{-1}\boldsymbol{\nu}^\top\boldsymbol{\mathcal{E}}/\widehat{\sigma}^2$, and $\mathbf{LM}^{(3)}_{NT,2b}\equiv \alpha_{NT}\widehat{\Lambda}_{NT}^\top\big(\widehat{\Theta}_{NT}^{-1}\mathbf{X}_2^\top\mathbf{M}_{\mathbf{X}_1}-\Theta_N^{-1}\boldsymbol{\nu}^\top\big)\boldsymbol{\mathcal{E}}/\widehat{\sigma}^2$, with $\boldsymbol{\nu}=\mathbf{X}_2-\mathbf{X}_1\Psi_{N,11}^{-1}\Psi_{N,12}$, $\Lambda_N\equiv\Gamma_{N,2}-\Psi_{N,12}^\top\Psi_{N,11}^{-1}\Gamma_{N,1}$, $\widehat{\Lambda}_{NT}\equiv(NT)^{-1}\mathbf{X}_2^\top\mathbf{M}_{\mathbf{X}_1}\mathbf{WR}(\mathbf{Y}_{-1})$, $\Theta_N\equiv\Psi_{N,22}-\Psi_{N,12}^\top\Psi_{N,11}^{-1}\Psi_{N,12}$, and $\widehat{\Theta}_{NT}\equiv(NT)^{-1}\mathbf{X}_2^\top\mathbf{M}_{\mathbf{X}_1}\mathbf{X}_2$. First, note that $\mathbf{E}|\Lambda_N^\top\Theta_N^{-1}\boldsymbol{\nu}^\top\boldsymbol{\mathcal{E}}|^2=\sigma^2(NT)\Lambda_N^\top\Theta_N^{-1}\Theta_N\Theta_N^{-1}\Lambda_N = \sigma^2(NT)\Lambda_N^\top\Theta_N^{-1}\Lambda_N = O(NT)$. It follows from Markov's inequality that $|\mathbf{LM}^{(1)}_{NT,2b}| = O_p(\alpha_{NT}\sqrt{NT})=O_p(J_1^{1/4})$. Next, for $\mathbf{LM}^{(2)}_{NT,2b}$, note that

$$\begin{aligned}\|\widehat{\Lambda}_{NT}-\Lambda_N\| =& \|\widehat{\Gamma}_{NT,2}-\Gamma_{N,2}-(\widehat{\Psi}_{NT,12}^\top\widehat{\Psi}_{NT,11}^{-1}\widehat{\Gamma}_{NT,1}-\Psi_{N,12}^\top\Psi_{N,11}^{-1}\Gamma_{N,1})\|\\ &\le \|\widehat{\Gamma}_{NT,2}-\Gamma_{N,2}\|+\|(\widehat{\Psi}_{NT,12}^\top\widehat{\Psi}_{NT,11}^{-1}-\Psi_{N,12}^\top\Psi_{N,11}^{-1})\Gamma_{N,1}\|\\ &\quad+\|\widehat{\Psi}_{NT,12}^\top\widehat{\Psi}_{NT,11}^{-1}(\widehat{\Gamma}_{NT,1}-\Gamma_{N,1})\|\\ &= O_p(\underline{\nu}_{J_1J_2}^{-2}r_{NT,J}^{1/2}J^{1/2})\end{aligned}$$

in light of Lemma 2. Thus, we have $|\mathbf{LM}^{(2)}_{NT,2b}| \le O_p(\alpha_{NT})\big\|\widehat{\Lambda}_{NT}-\Lambda_N\big\|\|\Theta_N^{-1}\boldsymbol{\nu}^\top\boldsymbol{\mathcal{E}}\| = O_p(\alpha_{NT}\sqrt{JJ_1}\underline{\nu}_{J_1J_2}^{-5/2}\sqrt{NT}r_{NT,J}^{1/2}) = O_p(J_1^{3/4}J^{1/2}r_{NT,J}^{1/2}\underline{\nu}_{J_1J_2}^{-5/2}) = o_p(J_1^{1/2})$. Finally, it has been shown in the proof of Theorem 2(ii) that $\|\widehat{\Theta}_{NT}^{-1}-\Theta_N^{-1}\|=O_p(\underline{\nu}_{J_1J_2}^{-2}r_{NT,J}^{1/2})$ and $\|\mathbf{X}_2^\top\mathscr{M}(\mathbf{X}_1)\boldsymbol{\mathcal{E}}-\boldsymbol{\nu}^\top\boldsymbol{\mathcal{E}}\| = O_p(\underline{\nu}_{J_1J_2}^{-2}r_{NT,J}^{1/2}\sqrt{NTJ_2})$. Also, noting that $\lambda_{\min}(\widehat{\Theta}_{NT}^{-1})\|\widehat{\Lambda}_{NT}\|^2\le\widehat{\Lambda}_{NT}^\top\widehat{\Theta}_{NT}^{-1}\widehat{\Lambda}_{NT}\le (NT)^{-1}\|\mathbf{WR}(\mathbf{Y}_{-1})\|^2 = O_p(1)$, we have $\|\widehat{\Lambda}_{NT}\| = O_p(\underline{\nu}_{J_1J_2}^{-1/2})$ by Assumption 4. Combining these results, we have $|\mathbf{LM}^{(3)}_{NT,2b}| = O_p(\alpha_{NT}\underline{\nu}^{-7/2}_{J_1J_2}r^{1/2}_{NT,J}\sqrt{NTJ}) = o_p(J_1^{1/2})$. This proves assertion (d).

To show (e), note that

$$
\begin{aligned}
&\frac{\mathbf{LM}_{NT,3}}{\sqrt{2(J_1-1)}} - \Delta_{N,2} \\
&= \frac{1}{\sqrt{2}\widehat{\sigma}^2}\widehat{\Lambda}^\top_{NT}\widehat{\Theta}^{-1}_{NT}\widehat{\Lambda}_{NT} - \frac{1}{\sqrt{2}\sigma^2}\Lambda^\top_N\Theta^{-1}_N\Lambda_N \\
&= \frac{1}{\sqrt{2}\widehat{\sigma}^2}\left[\|\widehat{\Lambda}_{NT}\|^2\|\widehat{\Theta}^{-1}_{NT} - \Theta^{-1}_N\| + \lambda_{\max}(\Theta^{-1}_N)\|\widehat{\Lambda}_{NT} - \Lambda_N\|\|\widehat{\Lambda}_{NT} + \Lambda_N\|\right] \\
&\quad + \frac{\sigma^2 - \widehat{\sigma}^2}{\sqrt{2}\sigma^2\widehat{\sigma}^2}\Lambda^\top_N\Theta^{-1}_N\Lambda_N \\
&= O_p\big(\underline{\nu}^{-3}_{J_1J_2}r^{1/2}_{NT,J}\big) + O_p\big(\underline{\nu}^{-7/2}_{J_1J_2}r^{1/2}_{NT,J}J^{1/2}\big) + o_p(1),
\end{aligned}
$$

which is $o_p(1)$ according to Assumption 5′. Consequently, $\mathbf{LM}_{NT,3}/\sqrt{2(J_1-1)} = \Delta_2 + o_p(1)$. This completes the proof of Theorem 4(ii). □

C Additional Discussion

C.1 Proof of Remark 2

Note that

$$
\begin{aligned}
\widehat{\sigma}^2 - \sigma^2 =& \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\Big(Y_{it} - \sum_{j=1}^N w_{ij}\widehat{f}_1(Y_{j,t-1}) - \widehat{f}_2(Y_{i,t-1}) - Z_i^\top\widehat{\gamma}\Big)^2 - \sigma^2 \\
=& \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\Bigg(\epsilon_{it} + \sum_{j=1}^N w_{ij}(f_1(Y_{j,t-1}) - \widehat{f}_1(Y_{j,t-1})) \\
&\quad + f_2(Y_{i,t-1}) - \widehat{f}_2(Y_{i,t-1}) + Z_i^\top(\gamma - \widehat{\gamma})\Bigg)^2 - \sigma^2 \\
=& \phi_1 + \phi_2 + \phi_3 + 2\phi_4 + 2\phi_5 + 2\phi_6,
\end{aligned}
$$

where $\phi_1 \equiv (NT)^{-1}\sum_{v=1}^{NT}\epsilon^2_{it} - \sigma^2$, $\phi_2 \equiv \sum_{i=1}^N(Z_i^\top(\widehat{\gamma} - \gamma))^2/N$, $\phi_3 \equiv (NT)^{-1}\sum_{v=1}^{NT}(\sum_{j=1}^N w_{ij}(\widehat{f}_1(Y_{j,t-1}) - f_1(Y_{j,t-1})) + \widehat{f}_2(Y_{i,t-1}) - f_2(Y_{i,t-1}))^2$, $\phi_4 \equiv (NT)^{-1}\sum_{v=1}^{NT}\epsilon_{it}Z_i^\top(\widehat{\gamma} - \gamma)$, $\phi_5 \equiv (NT)^{-1}\sum_{v=1}^{NT}\epsilon_{it}(\sum_{j=1}^N w_{ij}(\widehat{f}_1(Y_{j,t-1}) - f_1(Y_{j,t-1})) +$

$\widehat{f}_2(Y_{i,t-1}) - f_2(Y_{i,t-1}))$ and $\phi_6 \equiv (NT)^{-1}\sum_{v=1}^{NT}(\sum_{j=1}^{N} w_{ij}(\widehat{f}_1(Y_{j,t-1}) - f_1(Y_{j,t-1})) + \widehat{f}_2(Y_{i,t-1}) - f_2(Y_{i,t-1}))Z_i^\top(\widehat{\gamma} - \gamma)$.

First, by the law of large numbers, $|\phi_1| = o_p(1)$ under Assumption 1. Using the boundedness of Z_i and Theorem 1(i), we have $|\phi_2| = O_p((NT)^{-1} + J_1^{-2\mu} + J_2^{-2\mu})$. Also, ϕ_3 can be written as

$$\frac{1}{NT}\sum_{t=1}^{T}\|W\mathbb{P}_{t-1}^{J_1}(\beta_1 - \widehat{\beta}_1) + \mathbb{P}_{t-1}^{J_2}(\beta_2 - \widehat{\beta}_2) + W(F_1(\mathbb{Y}_{t-1}) - \mathbb{P}_{t-1}^{J_1}\beta_1) + F_2(\mathbb{Y}_{t-1}) - \mathbb{P}_{t-1}^{J_2}\beta_2\|^2$$
$$\leq 2(\widehat{\beta} - \beta)^\top\widehat{\Psi}_{NT,J}(\widehat{\beta} - \beta) + 2(NT)^{-1}\sum_{t=1}^{T}\left\|W(F_1(\mathbb{Y}_{t-1}) - \mathbb{P}_{t-1}^{J_1}\beta_1) + F_2(\mathbb{Y}_{t-1}) - \mathbb{P}_{t-1}^{J_2}\beta_2\right\|^2$$

where $\widehat{\Psi}_{NT,J} = \sum_{t=1}^{T}\mathbb{Q}_{t-1}^\top\mathbb{Q}_{t-1}/(NT)$ with $\mathbb{Q}_{t-1} = (W\mathbb{P}_{t-1}^{J_1}, \mathbb{P}_{t-1}^{J_2})$. By Lemma 2 and Assumption 4, $\lambda_{\max}(\widehat{\Psi}_{N,J}) = O_p(1)$. Hence, it follows from the proof of Theorem 2(i) that $(\widehat{\beta} - \beta)^\top\widehat{\Psi}_{NT,J}(\widehat{\beta} - \beta) \leq O_p(1)\cdot\|\widehat{\beta} - \beta\|^2 = O_p(J/(NT) + J_1^{-2\mu} + J_2^{-2\mu})$. Moreover, the second term on the r.h.s. of the last display is $O_p(J_1^{-2\mu} + J_2^{-2\mu})$. Thus $|\phi_3| = O_p(J/(NT) + J_1^{-2\mu} + J_2^{-2\mu}) = o_p(1)$.

Denote $\tilde{\phi}_1 \equiv (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\epsilon_{it}^2$. It can be easily seen that $|\phi_4| \leq \sqrt{\tilde{\phi}_1\phi_2}$, $|\phi_5| \leq \sqrt{\tilde{\phi}_1\phi_3}$ and $|\phi_6| \leq \sqrt{\phi_2\phi_3}$. It follows from the above arguments and the fact $\tilde{\phi}_1 = O_p(1)$ that $|\phi_j| = o_p(1)$ for $j = 4, 5, 6$. Combining the above results, we have $\widehat{\sigma}^2 - \sigma^2 \xrightarrow{p} 0$. $\square$

C.2 Asymptotics with fixed N

The asymptotic results presented in Section 3 are established under the regime $N \to \infty$ and T can be fixed or tends to ∞. In this subsection, we discuss another regime, that is, $T \to \infty$ and N is fixed. We first modify our assumptions as follows.

Assumption 4‡. *Assumption 4(i) holds. In addition, let* $\bar{\nu}_{J_1J_2} = \lambda_{\max}(\Sigma_{N,J_1J_2})$ *and* $\underline{\nu}_{J_1J_2} = \lambda_{\min}(\Sigma_{N,J_1J_2})$. *Uniformly in J_1 and J_2, there exists a constant* $0 < \bar{c} < \infty$ *such that* $\bar{\nu}_{J_1J_2} \leq \bar{c}$ *and* $\underline{\nu}_{J_1J_2} > 0$.

Assumption 5†. *As* $T \to \infty$, *N is fixed, and* $\underline{\nu}_{J_1J_2}^{-4}\zeta_{0,J}^2 J^{3/2}(J^{1/2} + \zeta_{1,J})/T \to 0$.

Assumption 7†. *The matrix* $\Sigma_{N,Z} = \Psi_{N,Z} - C_{N,ZJ}\Psi_{N,J}^{-1}C_{N,ZJ}^\top$ *is positive definite.*

The major difference between Assumption $4^\ddagger$ ($7^\ddagger$) and 4 (7) is that the former imposes conditions on the matrix Σ_{N,J_1J_2} ($\Sigma_{N,Z}$) rather than its limit version. The requirement of (J_1, J_2) in Assumption $5^\ddagger$ reduces to $\underline{\nu}_{J_1J_2}^{-4}J^4/T = o(1)$ for univariate polynomial splines or trigonometric polynomials, and $\underline{\nu}_{J_1J_2}^{-4}J^{11/2}/T = o(1)$ for power series or orthogonal polynomial bases.

Proposition 2. *Suppose that Assumptions 1-7 hold (where 4, 5 and 7 are replaced by $4^\ddagger$, $5^\ddagger$ and $7^\ddagger$, respectively). Then, (i) $\|\hat{\gamma}-\gamma\| = O_p(T^{-1/2}+J_1^{-\mu}+J_2^{-\mu})$; (ii) If $T^{1/2}(J_1^{-\mu}+J_2^{-\mu}) = o(1)$, $\sqrt{T}(\widehat{\gamma}-\gamma) \xrightarrow{d} N(0, \sigma^2\Sigma_{N,Z}^{-1}/N)$.*

Proof. When N is fixed, the NED property of $\{\mathbb{Y}_t : t \in \mathbb{Z}\}$ can be considered only along the time dimension. Following the lines of Lemma 1, we can easily see that

$$\max_{1\le i\le N} \|Y_{it} - \hat{Y}_{i,t-K}^t\|_2 \le d^K \sum_{j=0}^{t-K-1} d^j \max_{1\le i\le N} \|Z_i^\top\gamma + \epsilon_{it}\|_2 \le Cd^K/(1-d),$$

where $\hat{Y}_{i,t-K}^t$ is the ith element of $\hat{\mathbb{Y}}_{t-K}^t$ (a $\sigma(\mathcal{E}_{t-K},\ldots,\mathcal{E}_t)$-measurable approximation to $\mathbb{Y}_t$), $d = \kappa_1+\kappa_2 < 1$, and C is a finite constant. Using the above result, we can verify that the results in Lemma 2 still hold with $r_{NT,J}$ replaced by $\tilde{r}_{T,J} \equiv \zeta_{0,J}^2(J+\zeta_{1,J}J^{1/2})/T$ under the fixed N setting. Lemma 3 also holds with $r_{NT,J}$ replaced by $\tilde{r}_{T,J}$ under the modified assumptions. Proposition 2(i) can then be shown using the same arguments as the large N case.

To show Theorem 1(ii), note that $\sqrt{T}(\widehat{\gamma}-\gamma) = (N\Sigma_{N,Z})^{-1}\boldsymbol{\xi}^\top\boldsymbol{\mathcal{E}}/\sqrt{T}+o_p(1)$. Let $\boldsymbol{c}$ be an arbitrary $d_Z\times 1$ vector such that $\|\boldsymbol{c}\| = 1$, and

$$a_T = \boldsymbol{c}^\top\widetilde{\Omega}_{N,Z}^{-1/2}\boldsymbol{\xi}^\top\boldsymbol{\mathcal{E}}/\sqrt{T},$$

where $\widetilde{\Omega}_{N,Z} \equiv \sigma^2N\Sigma_{N,Z}$. Then, it suffices to show that a_T weakly converges to $N(0,1)$ as $T\to\infty$. For this purpose, write $a_T = \sum_{t=1}^T X_t$, where $X_t \equiv T^{-1/2}\boldsymbol{c}^\top\widetilde{\Omega}_{N,Z}^{-1/2}(\mathbb{Z}^\top - C_{N,ZJ}\Psi_{N,J}^{-1}\mathbb{Q}_{t-1}^\top)\mathcal{E}_t$. Denote $\mathcal{B}_t$ the σ-field generated by the random variables $\{\mathcal{E}_s : s\le t\}$. We can easily check that X_t is $\mathcal{B}_t$-measurable and $\mathbb{E}[X_t|\mathcal{B}_{t-1}] = 0$. Hence, $\{X_t, \mathcal{B}_t, \infty < t \le T\}$ forms a martingale difference array. Applying Corollary 3.1 in Hall and Heyde (1980), we have $a_T \xrightarrow{d} N(0,1)$ provided the following conditions hold: (a) for some $\delta > 0$,

$\sum_{t=1}^{T}\mathbb{E}|X_t|^{2+\delta}\to 0$, and (b) $\sum_{t=1}^{T}\mathbb{E}[|X_t|^2|\mathcal{B}_{t-1}]\xrightarrow{p}1$. To show (a), write $X_t = X_{1,t}-X_{2,t}$ where $X_{1,t}=T^{-1/2}\boldsymbol{c}^\top\widetilde{\Omega}_{N,Z}^{-1/2}\mathbb{Z}^\top\mathcal{E}_t$ and $X_{2,t}=T^{-1/2}\boldsymbol{c}^\top\widetilde{\Omega}_{N,Z}^{-1/2}C_{N,ZJ}\Psi_{N,J}^{-1}\mathbb{Q}_{t-1}^\top\mathcal{E}_t$. By Assumptions 3 and $4^{\ddagger}$,

$$\begin{aligned}\mathbb{E}\big(|X_{2,t}|^4\,|\,\mathcal{B}_{t-1}\big) &= \frac{1}{T^2}\mathbb{E}\left[\big(\boldsymbol{c}^\top\widetilde{\Omega}_{N,Z}^{-1/2}C_{N,ZJ}\Psi_{N,J}^{-1}\mathbb{Q}_{t-1}^\top\mathcal{E}_t\big)^4|\mathcal{B}_{t-1}\right]\\ &\le \frac{C}{T^2}\big(\boldsymbol{c}^\top\widetilde{\Omega}_{N,Z}^{-1/2}C_{N,ZJ}\Psi_{N,J}^{-1}\mathbb{Q}_{t-1}^\top\mathbb{Q}_{t-1}\Psi_{N,J}^{-1}C_{N,ZJ}^\top\widetilde{\Omega}_{N,Z}^{-1/2}\boldsymbol{c}\big)^2\le\frac{C}{T^2\underline{\nu}_{J_1J_2}^4}\|\mathbb{Q}_{t-1}\|^4,\end{aligned}$$

where $C=2\sigma^4+|\mathbb{E}(\epsilon_{it}^4)-3\sigma^4|$. Also, Assumption 4(i) implies that $\max_{t\ge1}\mathbb{E}\|\mathbb{Q}_{t-1}\|^4 = O(J^2\zeta_{0,J}^2)$. Hence,

$$\sum_{t=1}^{T}\mathbb{E}\big[\mathbb{E}\big(|X_{2,t}|^4\,|\,\mathcal{B}_{t-1}\big)\big]\le\frac{C}{T^2\underline{\nu}_{J_1J_2}^4}\sum_{t=1}^{T}\mathbb{E}\|\mathbb{Q}_{t-1}\|^4 = O\left(\frac{J^2\zeta_{0,J}^2}{T\underline{\nu}_{J_1J_2}^4}\right)=o(1).$$

Analogously, we can show that $\sum_{t=1}^{T}\mathbb{E}(X_{1,t}^4)=O(1/T)$. The desired result follows directly from the c_r inequality. Condition (b) can be verified in a similar way to Theorem 1(ii). This completes the proof of the second part of the proposition. □

Proposition 3. *Suppose that Assumptions 1-3, $4^{\ddagger}$, $5^{\ddagger}$, and 6 hold. Let $\ell=1,2$.*

(i) If $\max_{1\le i\le N}\lambda_{max}(\mathbb{E}[p^{J_\ell}(Y_{it})p^{J_\ell}(Y_{it})^\top])<\infty$, then $\max_{1\le i\le N}\int_{\mathcal{R}_\mathcal{Y}}[\widehat{f}_\ell(y)-f_\ell(y)]^2dF_i(y) = O_p(\underline{\nu}_{J_1J_2}^{-1}(J_\ell/T+J_1^{-2\mu}+J_2^{-2\mu}))$, where $F_i(y)$ is the cumulative distribution function of Y_{it}.

(ii) If $\sqrt{T}\underline{\nu}_{J_1J_2}^{-1/2}(J_1^{-\mu}+J_2^{-\mu})=o(1)$, then for a given finite $y\in\mathcal{R}_\mathcal{Y}$ such that $\|p^{J_\ell}(y)\|>0$, we have $\sqrt{T}\tilde{v}_\ell^{-1}(y)\big(\widehat{f}_\ell(y)-f_\ell(y)\big)\xrightarrow{d}N(0,1)$ with $\tilde{v}_\ell^2(y)=\sigma^2N^{-1}p^{J_\ell}(y)^\top\Sigma_{N,f_\ell}^{-1}p^{J_\ell}(y)$.

Proposition 3 can be proved in a similar way to Theorem 2 and Proposition 2. The details are omitted for brevity.

C.3 General NAR(p) model

In this subsection, we study the estimation of the semiparametric NAR(p) model given by

$$Y_{it}=\sum_{j=1}^{N}w_{ij}f_1(Y_{j,t-1},Y_{j,t-2},\dots,Y_{j,t-p})+f_2(Y_{i,t-1},Y_{i,t-2},\dots,Y_{i,t-p})+Z_i^\top\gamma+\epsilon_{it},\qquad(45)$$

where $i = 1, \ldots, N$, $t = p, \ldots, T$, and f_1 and f_2 are unknown functions on $\mathbb{R}^p$. For identification, assume that $f_1(\mathbf{0}) = 0$, $f_2(\mathbf{0}) = 0$ and the intercept term is included in Z_i. Similar to Section 2, we consider estimating f_1 and f_2 by sieve methods. Let $\{p_j(y) : j = 1, 2, \ldots\}$ be a sequence of known basis functions on $\mathbb{R}^p$, and $p^J(y) = (p_1(y), \ldots, p_J(y))^\top$ for some $J \equiv J_{NT}$. Assuming that the unknown functions f_1 and f_2 are smooth enough, we can find two vectors $\beta_1 = (\beta_{11}, \ldots, \beta_{1J_1})^\top$ and $\beta_2 = (\beta_{21}, \ldots, \beta_{2J_2})^\top$ such that $f_1(\cdot)$ and $f_2(\cdot)$ can be well approximated by $\beta_1^\top p^{J_1}(\cdot)$ and $\beta_2^\top p^{J_2}(\cdot)$, respectively, for sufficiently large J_1 and J_2. Denote $\underline{Y}_{i,t-1}^{(p)} = (Y_{i,t-1}, Y_{i,t-2}, \ldots, Y_{i,t-p})$. Then, model (45) can be rewritten as $Y_{it} = \beta_1^\top \bar{P}_{i,t-1}^{J_1} + \beta_2^\top P_{i,t-1}^{J_2} + Z_i^\top \gamma + u_{it}$ where $\bar{P}_{i,t-1}^{J_1} = \sum_{j=1}^N w_{ij} p^{J_1}(\underline{Y}_{j,t-1}^{(p)})$, $P_{i,t-1}^{J_2} = p^{J_2}(\underline{Y}_{i,t-1}^{(p)})$, and u_{it} is the new error term. In matrix form,

$$\mathbb{Y}_t = W\mathbb{P}_{t-1}^{J_1}\beta_1 + \mathbb{P}_{t-1}^{J_2}\beta_2 + \mathbb{Z}\gamma + U_t,$$

where $\mathbb{P}_{t-1}^{J_\ell} = (p^{J_\ell}(\underline{Y}_{1,t-1}^{(p)}), \ldots, p^{J_\ell}(\underline{Y}_{N,t-1}^{(p)}))^\top$ for $\ell = 1, 2$, and $U_t = (u_{1t}, \ldots, u_{Nt})^\top$. Denote $\mathbb{X}_{t-1} = (W\mathbb{P}_{t-1}^{J_1}, \mathbb{P}_{t-1}^{J_2}, \mathbb{Z})$, and $\theta = (\beta_1^\top, \beta_2^\top, \gamma^\top)^\top$. The sieve LS estimator of the coefficient θ is given by

$$\widehat{\theta} = (\widehat{\beta}_1^\top, \widehat{\beta}_2^\top, \widehat{\gamma}^\top)^\top = \left(\sum_{t=1}^T \mathbb{X}_{t-1}^\top \mathbb{X}_{t-1}\right)^- \sum_{t=1}^T \mathbb{X}_{t-1}^\top \mathbb{Y}_t. \tag{46}$$

For $\ell = 1, 2$, the sieve estimator of f_ℓ is given by $\widehat{f}_\ell(\cdot) = p^{J_\ell}(\cdot)^\top \widehat{\beta}_\ell$.

To study the stationarity of $\{\mathbb{Y}_t\}$ in (45), we modify Assumption 2 in the main text as follows.

Assumption 2′. *There exist non-negative constants $\{a_{1,k}\}_{k=1}^p$ and $\{a_{2,k}\}_{k=1}^p$ such that for any $\boldsymbol{y}$, $\boldsymbol{y}' \in \mathbb{R}^p$, $|f_\ell(\boldsymbol{y}) - f_\ell(\boldsymbol{y}')| \leq \sum_{k=1}^p a_{\ell,k}|y_k - y_k'|$, $\ell = 1, 2$, and $\sum_{k=1}^p (a_{1,k} + a_{2,k}) < 1$.*

Using similar arguments to Proposition 1, we can show that under Assumptions 1 and 2′, there exists a unique stationary solution to the NAR model (6) regardless of whether N is fixed or tends to infinity. In addition, if Z_i's are uniformly bounded and the ath moment of ϵ_{it} exists, we have $\max_{i\geq 1} \mathbb{E}|Y_{it}|^a \leq C_a < \infty$.

Redefine $\zeta_{0,J} = \sup_{\boldsymbol{y}\in\mathbb{R}^p} \|p^J(\boldsymbol{y})\|$ and $\zeta_{1,J} = \sup_{\boldsymbol{y},\boldsymbol{y}'\in\mathbb{R}^p} \|p^J(\boldsymbol{y}) - p^J(\boldsymbol{y}')\|/\|\boldsymbol{y} - \boldsymbol{y}'\|$. Also, under the multivariate case, the weighted sup-norm is adapted as $|f|_{\infty,\omega} = \sup_{\boldsymbol{y}\in\mathcal{R}_{\mathcal{Y}}^p} |f(\boldsymbol{y})|(1+\|\boldsymbol{y}\|^2)^{-\omega/2}$.

Proposition 4. *Consider model (45) and sieve estimator (46). Under Assumptions 1, 2', 3-7, the conclusions in Theorems 1 and 2 still hold.*

Proof. First, in a similar way to Lemma 1, we can show that $\{Y_{it}\}$ generated by model (6) satisfies $\|Y_{it} - \mathbb{E}[Y_{it} \mid \mathcal{F}_{it,p}(s)]\|_2 \leq Cd^s$, where $d = \sum_{k=1}^{p}(a_{1,k} + a_{2,k}) < 1$, and $\mathcal{F}_{it,p}(s)$ denotes the σ-field generated by the random variables $\{\epsilon_{j\tau} : d_N(i,j) \leq s, |t-\tau| \leq sp\}$. Using this property, we can adapt the arguments of Lemma 2 to see that the convergence rates of $\widehat{C}_{NT,ZJ}$ and $\widehat{\Psi}_{NT,J}$ therein still apply to the general NAR(p) model (45). We take the proof of (9) for illustration. First, under the GNAR(p) model, we have $\left|\mathrm{Cov}\big(Z_i \bar{P}_{i,t-1}^\top, Z_j \bar{P}_{j,\tau-1}^\top\big)\right| \leq C\zeta_{1,J_1} J_1^{1/2} d^{\lfloor \tilde{h}/3 \rfloor}$, where $\tilde{h} = \max\{d_N(i,j), \lfloor |t-\tau|/p \rfloor\}$. Then,

$$
\begin{aligned}
&\sum_{j\neq i}^{N}\sum_{\tau\neq t}^{T}\mathrm{Cov}\big(Z_i \bar{P}_{i,t-1}^\top, Z_j \bar{P}_{j,\tau-1}^\top\big)\\
&= \sum_{s=1}^{N}\sum_{j\in\mathcal{N}_N^\partial(i;s)}\sum_{l=t-T,\, l\neq 0}^{t-1}\mathrm{Cov}\big(Z_i \bar{P}_{i,t-1}^\top, Z_j \bar{P}_{j,t-l-1}^\top\big)\\
&\leq\ C\zeta_{1,J_1} J_1^{1/2}\sum_{s=1}^{N}\sum_{j\in\mathcal{N}_N^\partial(i;s)}\sum_{l=t-T,\, l\neq 0}^{t-1} d^{\lfloor \max\{s,\lfloor l/p \rfloor\}/3 \rfloor}\\
&\leq\ C\zeta_{1,J_1} J_1^{1/2}\sum_{s=1}^{N}\sum_{j\in\mathcal{N}_N^\partial(i;s)}\sum_{l=0}^{\lfloor T/p \rfloor} d^{\lfloor \max\{s,l\}/3 \rfloor}\\
&\leq C\zeta_{1,J_1} J_1^{1/2}\sum_{s=1}^{N}\left|\mathcal{N}_N^\partial(i;s)\right|\left(\sum_{l=0}^{\min\{s,\lfloor T/p \rfloor\}} d^{\lfloor s/3 \rfloor} + \sum_{l=\min\{s+1,\lfloor T/p \rfloor\}}^{\lfloor T/p \rfloor} d^{\lfloor l/3 \rfloor}\right)\\
&\leq\ C\zeta_{1,J_1} J_1^{1/2}\sum_{s=1}^{N}\min\{s,\lfloor T/p \rfloor\}\left|\mathcal{N}_N^\partial(i;s)\right| d^{\lfloor s/3 \rfloor},
\end{aligned}
$$

where C denotes finite constants that vary from line to line. As such, Lemma 2 can be proved under the GNAR(p) model (45). Then, Proposition 4 can be shown in a similar way to Theorems 1 and 2. The details are omitted for brevity. □

D Additional Simulation Results

D.1 Estimating the degree of collinearity

In this subsection, we present some numerical results on how the measure of collinearity $\underline{\nu}_{J_1 J_2}^{-1/2}$ depends on (J_1, J_2) and the structure of the adjacency matrix A. We consider the following model:

$$Y_{it} = n_i^{-1} \sum_{j=1}^{N} a_{ij} f_1(Y_{j,t-1}) + f_2(Y_{i,t-1}) + \epsilon_{it},$$

where $\epsilon_{it} \overset{\text{i.i.d.}}{\sim} N(0,1)$. For the functional form of f_1 and f_2, we consider the same DGPs as in Section 5. The sample size is set to $N = 1000$ and $T = 10$. For the adjacency matrix A, we consider three different network structures.

(1) Erdos–Rényi (ER) Model. This model assumes that the network is constructed by randomly connecting N nodes. Let $D_{ij} = (a_{ij}, a_{ji})$ for $1 \leq i < j \leq N$. Following Zhu et al. (2017), we set $P(D_{ij} = (1,1)) = cN^{-1}$, $P(D_{ij} = (0,1)) = P(D_{ij} = (1,0)) = 0.5N^{-0.8}$ and $P(D_{ij} = (0,0)) = 1 - cN^{-1} - N^{-0.8}$, where c is fixed to be 20 in this simulation. Different D_{ij}'s are independent.

(2) Stochastic Block Model (SBM). Randomly assign for each node a block label ($k = 1, \ldots, K$) with equal probability. Here, K is the total number of blocks, which is fixed at 5 in this simulation. Next, set $P(a_{ij} = 1) = 0.3N^{-0.3}$ if i and j belong to the same block, and $P(a_{ij} = 1) = 0.3N^{-1}$ otherwise. Accordingly, the nodes within the same block are more likely to be connected than nodes from different blocks.

(3) Power-law distribution Model (PLD). The power-law distribution captures a network phenomenon in which the majority of nodes have few edges, while a small number have a huge number of edges. To mimic this phenomenon, we generate for each node its in-degree $d_i = \sum_{j=1}^{N} a_{ji}$ according to the discrete power-law distribution, that is, $P(d_i = k) = ck^{-\alpha}$ for a normalizing constant c and the exponent parameter $\alpha \in \{1.5, 3.0\}$.

For sieve estimation, we use the cubic B-splines for p, with the number of basis functions $J_1 = J_2 = J$. We choose the value of J from $J \in \{3, 4, 5, 6, 7\}$.

In our context, the measure of collinearity refers to the quantity $(\lambda_{\min}(\Sigma_{N,J}))^{-1/2}$, see Assumption 4 and related discussion. However, since $\Sigma_{N,J}$ is an unknown population object, we replace it with its sample counterparts $\widehat{\Sigma}_{NT,J} = (NT)^{-1}\sum_{t=1}^{T}\mathbb{X}_{t-1}^{\top}\mathbb{X}_{t-1}$, and compute $(\lambda_{\min}(\widehat{\Sigma}_{NT,J}))^{-1/2}$ for each set-up. The results are summarized in Figure D.1. The values reported in these figures are those averaged over 100 Monte Carlo repetitions.

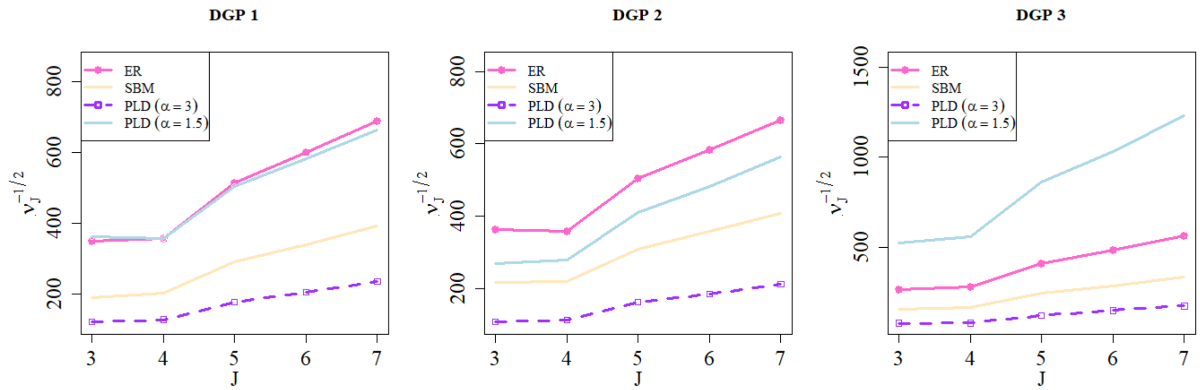

Figure D.1: Degree of collinearity: $(\lambda_{\min}(\widehat{\Sigma}_{NT,J_1J_2}))^{-1/2}$

In Figure D.1, we observe that the degree of collinearity increases monotonically as J increases. More precisely, the (estimated) quantity $\underline{\nu}_J^{-1/2}$ goes to infinity at a nearly polynomial rate of J for all the DGPs we consider. Moreover, the structure of the adjacency matrix also has a significant impact on the degree of collinearity.[8] It can be clearly observed that the collinearity is mild if the network density is small. A possible explanation is that the larger the number of connected nodes, the smaller the variance of $\bar{P}_{it}^{J}$. The smaller variance further makes the relevance between $\bar{P}_{it}^{J}$ and P_{it}^{J} less sharp.

D.2 Neighborhood shell size

In this subsection, we examine the neighborhood shell properties of the commonly used networks using Monte Carlo simulations. Our focus is on the growth rate of $\delta_N^{\partial}(s)$ with s, where $\delta_N^{\partial}(s)$ is the average number of the neighbors at distance s in the network G_N (see Assumption 5 and its related discussion for detailed definition). We consider networks generated by ER and stochastic block models. The network generation mechanism is the same as Section D.1, where the parameter c for the ER model is set to be $c \in \{5, 20\}$, and

[8]In the simulation, the network density for the ER, stochastic block, PLD ($\alpha = 3$), and PLD ($\alpha = 1.5$) models is around 2.21%, 0.78%, 0.15%, and 2.51%, respectively.

the number of blocks K for the SBM is set to be $K \in \{5, 20\}$. For the ER model, a larger c implies a denser network, whereas for SBM, a larger K implies a sparser network.

Figures D.2 and D.3 present the average s-neighborhood shell size for the networks generated by ER and stochastic block models under the sample size $N = 1000$ and 2000, respectively. We see that for both models and sample sizes, $\delta_N^\partial(s)$ increases with s and then declines to zero. The peak of $\delta_N^\partial(s)$ depends on both the network structure and density. For a fixed network structure, the growth and decline rates of $\delta_N^\partial(s)$ are faster for the denser case. From the pattern implied by Figures D.2 and D.3, the average s-neighborhood shell size for ER and SBM networks can be bounded as $\delta_N^\partial(s) \leq C_N e^{\lambda s}$, where C_N is a positive constant that might depend on N and $\lambda \geq 1$.[9] If this relationship holds, the quantity $\Xi_{N,d}$ defined above Assumption 5 satisfies that $\Xi_{N,d} = N^{-1} \sum_{s=1}^{\infty} s^d \delta_N^\partial(s)(\kappa_1 + \kappa_2)^{\lfloor s/3 \rfloor} \to 0$ if $e^\lambda(\kappa_1 + \kappa_2)^{1/3} < 1$ and $C_N/N \to 0$.

D.3 Additional simulation results

This subsection contains the simulation tables that were initially intended for Section 5, but are presented here due to space limitations. The content includes: (1) the estimation results of γ, (2) the coverage rate of $f_1(y)$ and $f_2(y)$ at some fixed points y, (3) the rejection probabilities of the $\mathbf{T}$ test for testing H_0^n and H_0^m, and (4) the rejection probabilities of the $\mathbf{T}$ and LM tests for testing $H_0^\dagger$ in Section 4.2.

The performance of the sieve LS estimator of γ is shown in Table 1. We see that the bias and RMSE (root mean squared error) of both γ_1 and γ_2 are satisfactorily small, and the coverage rate of the 95% confidence interval (CR95) is close to the nominal level in all set-ups. The performance is not sensitive to the choice of the number of basis functions J.

The coverage rates of 95% confidence interval for $f_\ell(y_{25})$, $f_\ell(y_{50})$, and $f_\ell(y_{75})$ are presented in Table D.1, where $\ell = 1, 2$, and y_α represents the α% empirical quantiles of $\{Y_{i,t-1}\}$ averaged over the replications. The computation is based on Theorem 2(iii). The results show that the coverage rates are close to the nominal levels in all cases. It is also insensitive to the choice of J and the network structure.

[9] We refer the readers to Section 14 of Bollobás et al. (2007) for theoretical analysis of the neighborhood shell properties of ER and stochastic block models.

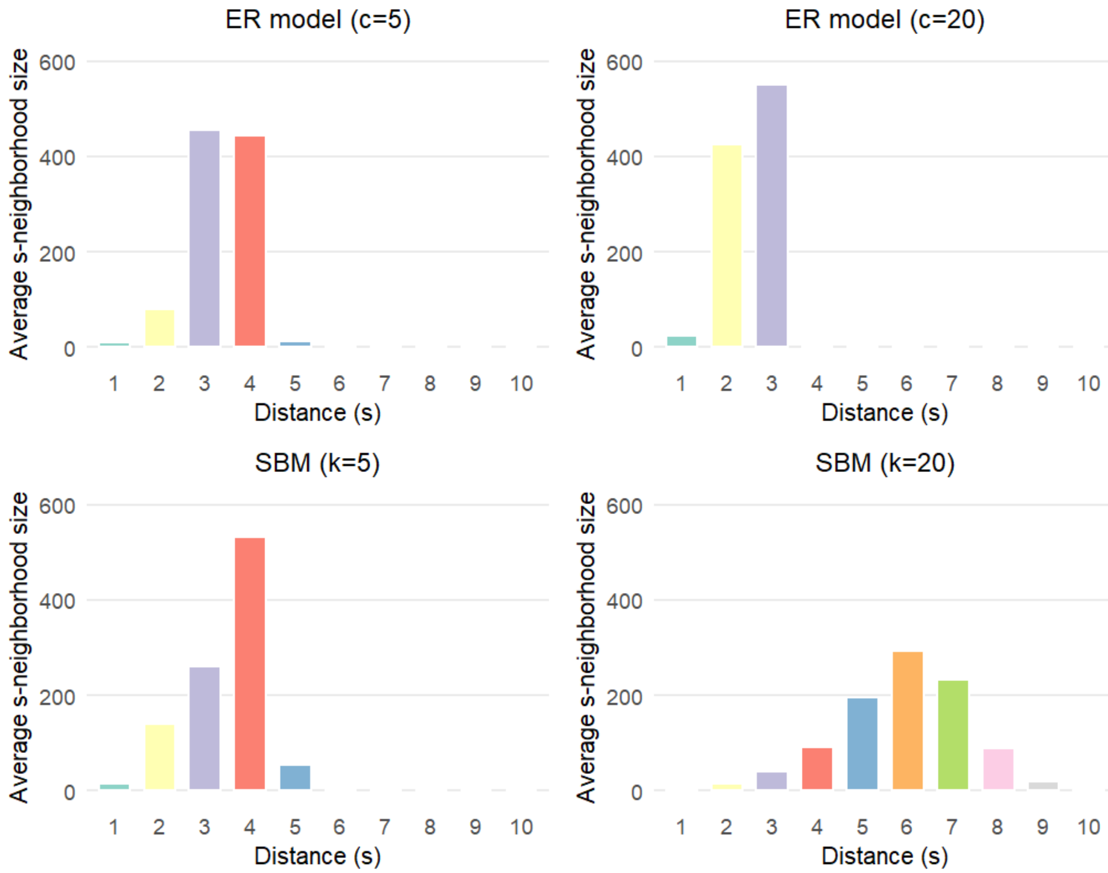

Figure D.2: Average s-neighborhood shell size $\delta_N^{\partial}(s)$ for $N = 1000$

Tables 4 and D.2 show the rejection frequencies of the $\mathbf{T}$ test at 10%, 5%, and 1% significance levels. The sample sizes we consider are $N \in \{100, 200, 500, 1000\}$ and $T \in \{10, 20\}$. Under a moderate sample size, the $\mathbf{T}$ test is severely undersized, particularly when the hypothesis H_0^m is considered. However, this distortion becomes milder as the sample size N increases. The empirical power of the $\mathbf{T}$ test is reasonably high under moderate sample sizes. Finally, the performance of the distance-based test is more sensitive to the choice of J than the LM test. This may be because the $\mathbf{T}$ test requires both estimation of both the null and alternative models, the precision of which depends on a more delicate selection of J.

Table D.3 reports the rejection frequencies of the $\mathbf{T}$ and LM tests for testing the joint linearity of f_1 and f_2. We see from the results for DGP 1 that the size of the LM test is close to the nominal levels, while the $\mathbf{T}$ test is undersized at the 10% significance level. When either f_1 or f_2 is highly nonlinear, as in DGPs 2–4, the LM test has high empirical power for

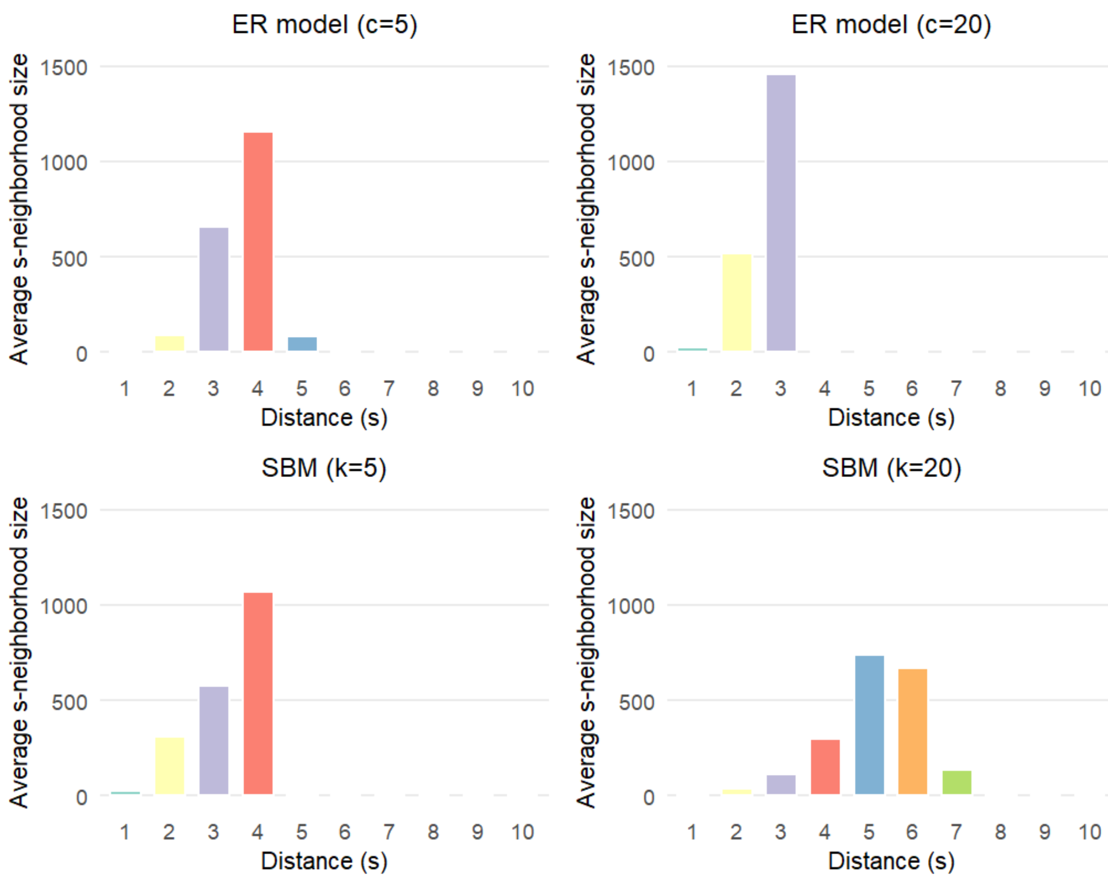

Figure D.3: Average s-neighborhood shell size $\delta_N^\partial(s)$ for $N = 2000$

all choices of J and all sample sizes. It also outperforms the $\mathbf{T}$ test in all cases we consider. Finally, the empirical power of both tests is not very sensitive to the choice of J and the network structure.

D.4 Simulations with large T

In Section 5 of the main text, we consider simulations under the asymptotic setting with large N and small T. In what follows, we report the estimation and testing results under the DGPs in Section 5 with small N and large T. Specifically, we set $N \in \{10, 50\}$ and $T \in \{100, 200\}$. For the number of basis functions, we consider $J \in \{3, 4\}$ for $(N, T) \in \{(10, 100), (10, 200), (50, 100)\}$, and $J \in \{4, 5\}$ for $(N, T) = (50, 200)$.

Tables D.4 and D.5 below report the estimation results of γ and (f_1, f_2), respectively. It is seen from Table D.4 that the bias and RMSE of γ_1 and γ_2 are quite small, and the

Table D.1: Coverage rate of the 95% confidence interval of $f_1(y)$ and $f_2(y)$

				DGP 1			DGP 2			DGP 3			DGP 4	
Model	N	T	J	$f_1(y_{25})$	$f_1(y_{50})$	$f_1(y_{75})$	$f_1(y_{25})$	$f_1(y_{50})$	$f_1(y_{75})$	$f_1(y_{25})$	$f_1(y_{50})$	$f_1(y_{75})$	$f_1(y_{25})$	$f_1(y_{50})$	$f_1(y_{75})$
ER	100	10	4	0.951	0.946	0.945	0.942	0.942	0.944	0.940	0.943	0.963	0.953	0.950	0.935
	100	10	5	0.946	0.952	0.945	0.950	0.952	0.950	0.950	0.944	0.958	0.959	0.949	0.942
	200	10	4	0.943	0.940	0.940	0.945	0.940	0.946	0.943	0.953	0.942	0.957	0.954	0.955
	200	10	5	0.945	0.940	0.951	0.941	0.945	0.945	0.944	0.944	0.934	0.960	0.955	0.956
	100	20	4	0.943	0.949	0.940	0.950	0.960	0.957	0.952	0.950	0.942	0.937	0.945	0.942
	100	20	5	0.938	0.937	0.940	0.961	0.950	0.946	0.962	0.955	0.941	0.944	0.943	0.955
	200	20	4	0.955	0.949	0.943	0.930	0.925	0.941	0.952	0.950	0.957	0.937	0.946	0.947
	200	20	5	0.946	0.946	0.949	0.927	0.953	0.960	0.958	0.951	0.957	0.941	0.938	0.944
SBM	100	10	4	0.942	0.947	0.951	0.946	0.949	0.964	0.951	0.953	0.948	0.953	0.957	0.954
	100	10	5	0.942	0.933	0.945	0.942	0.952	0.951	0.955	0.956	0.945	0.946	0.956	0.938
	200	10	4	0.942	0.943	0.947	0.956	0.955	0.943	0.938	0.942	0.962	0.951	0.951	0.959
	200	10	5	0.951	0.946	0.946	0.957	0.955	0.956	0.946	0.940	0.965	0.948	0.948	0.954
	100	20	4	0.945	0.946	0.951	0.966	0.949	0.955	0.954	0.952	0.956	0.956	0.959	0.960
	100	20	5	0.946	0.956	0.952	0.959	0.962	0.960	0.944	0.948	0.953	0.951	0.954	0.953
	200	20	4	0.936	0.944	0.953	0.934	0.923	0.934	0.958	0.954	0.966	0.960	0.956	0.952
	200	20	5	0.945	0.945	0.945	0.932	0.936	0.939	0.953	0.957	0.965	0.956	0.958	0.944

				DGP 1			DGP 2			DGP 3			DGP 4	
Model	N	T	J	$f_2(y_{25})$	$f_2(y_{50})$	$f_2(y_{75})$	$f_2(y_{25})$	$f_2(y_{50})$	$f_2(y_{75})$	$f_2(y_{25})$	$f_2(y_{50})$	$f_2(y_{75})$	$f_2(y_{25})$	$f_2(y_{50})$	$f_2(y_{75})$
ER	100	10	4	0.946	0.955	0.955	0.944	0.947	0.951	0.945	0.951	0.956	0.954	0.950	0.953
	100	10	5	0.951	0.960	0.947	0.948	0.948	0.944	0.954	0.958	0.954	0.945	0.945	0.951
	200	10	4	0.947	0.941	0.943	0.952	0.949	0.952	0.947	0.943	0.956	0.946	0.946	0.952
	200	10	5	0.947	0.946	0.951	0.951	0.948	0.941	0.942	0.947	0.956	0.947	0.943	0.953
	100	20	4	0.938	0.936	0.944	0.943	0.941	0.946	0.952	0.948	0.954	0.943	0.951	0.944
	100	20	5	0.931	0.934	0.952	0.944	0.938	0.952	0.955	0.955	0.951	0.946	0.945	0.945
	200	20	4	0.951	0.949	0.949	0.955	0.960	0.956	0.946	0.952	0.947	0.955	0.946	0.941
	200	20	5	0.951	0.955	0.956	0.956	0.946	0.956	0.943	0.950	0.947	0.959	0.949	0.942
SBM	100	10	4	0.949	0.944	0.946	0.952	0.950	0.962	0.939	0.941	0.939	0.935	0.934	0.931
	100	10	5	0.948	0.940	0.946	0.949	0.943	0.947	0.942	0.941	0.945	0.946	0.927	0.931
	200	10	4	0.959	0.965	0.945	0.943	0.943	0.950	0.948	0.948	0.947	0.954	0.953	0.950
	200	10	5	0.961	0.954	0.953	0.944	0.946	0.948	0.946	0.947	0.940	0.957	0.951	0.956
	100	20	4	0.943	0.943	0.949	0.943	0.950	0.953	0.958	0.953	0.956	0.943	0.938	0.950
	100	20	5	0.947	0.958	0.956	0.939	0.948	0.948	0.953	0.956	0.953	0.954	0.938	0.950
	200	20	4	0.955	0.958	0.967	0.961	0.954	0.944	0.949	0.944	0.943	0.956	0.952	0.955
	200	20	5	0.953	0.956	0.960	0.955	0.956	0.950	0.948	0.948	0.946	0.955	0.952	0.949

coverage rate of the 95% confidence interval is close to the nominal level in all set-ups. The magnitude of the bias and RMSE of γ therein is similar to that in Table 1, for fixed total sample size NT. The performance of the sieve estimators of f_1 and f_2 exhibits a similar pattern as discussed in Section 5. Comparing Table D.5 with Table 2, we see that a large N plays an important role in guaranteeing the estimation accuracy of $\widehat{f}_1$.

The rejection frequencies of the distance-based and LM tests under the large T setting are reported in Tables D.6 and D.7, respectively. The distance-based test statistic $\mathbf{T}$ still suffers from size distortion, especially when testing H_0^m. However, the $\mathbf{T}$ test is very powerful in detecting deviations from the null for all cases we consider. The performance of the LM test is satisfactory. The empirical sizes are quite close to nominal ones, and the power all approach one. Compared with Table 3, the empirical power of the LM test significantly

Table D.2: Rejection probabilities of the distance-based test statistic with $N \in \{500, 1000\}$

(1) Null : H_0^n				DGP 1			DGP 2			DGP 3			DGP 4		
Model	N	T	J	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
ER	500	10	6	0.061	0.032	0.017	0.996	0.995	0.981	0.057	0.030	0.006	0.638	0.558	0.413
	500	10	7	0.053	0.039	0.015	0.995	0.989	0.976	0.056	0.027	0.010	0.610	0.525	0.384
	1000	10	7	0.077	0.039	0.012	1	1	1	0.058	0.038	0.013	0.903	0.869	0.774
	1000	10	8	0.071	0.047	0.012	1	1	1	0.056	0.039	0.011	0.896	0.850	0.753
	500	20	7	0.072	0.045	0.015	1	1	1	0.059	0.028	0.012	0.920	0.880	0.789
	500	20	8	0.081	0.040	0.019	1	1	1	0.057	0.038	0.010	0.909	0.858	0.772
	1000	20	8	0.061	0.038	0.015	1	1	1	0.055	0.031	0.012	0.998	0.998	0.987
	1000	20	9	0.056	0.037	0.012	1	1	1	0.049	0.031	0.010	0.998	0.996	0.985
SBM	500	10	6	0.084	0.051	0.025	0.817	0.761	0.625	0.071	0.041	0.011	0.265	0.201	0.110
	500	10	7	0.089	0.068	0.025	0.793	0.715	0.574	0.063	0.038	0.017	0.261	0.191	0.095
	1000	10	7	0.076	0.047	0.018	0.926	0.884	0.758	0.062	0.042	0.013	0.318	0.249	0.145
	1000	10	8	0.076	0.051	0.014	0.910	0.861	0.717	0.065	0.039	0.014	0.304	0.231	0.134
	500	20	7	0.076	0.049	0.016	0.977	0.954	0.913	0.063	0.034	0.012	0.488	0.412	0.269
	500	20	8	0.084	0.052	0.018	0.968	0.945	0.896	0.060	0.043	0.014	0.468	0.393	0.252
	1000	20	8	0.072	0.039	0.014	0.994	0.989	0.977	0.080	0.045	0.014	0.563	0.473	0.332
	1000	20	9	0.085	0.046	0.013	0.993	0.987	0.967	0.089	0.051	0.019	0.534	0.459	0.309
(2) Null:H_0^m				DGP 1			DGP 2			DGP 3			DGP 4		
Model	N	T	J	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
ER	500	10	6	0.032	0.024	0.008	0.016	0.010	0.002	1	1	1	1	1	1
	500	10	7	0.046	0.030	0.021	0.028	0.015	0.007	1	1	1	1	1	1
	1000	10	7	0.045	0.029	0.012	0.037	0.025	0.011	1	1	1	1	1	1
	1000	10	8	0.059	0.038	0.013	0.049	0.036	0.012	1	1	1	1	1	1
	500	20	7	0.059	0.036	0.018	0.019	0.013	0.007	1	1	1	1	1	1
	500	20	8	0.060	0.043	0.024	0.042	0.027	0.009	1	1	1	1	1	1
	1000	20	8	0.049	0.033	0.013	0.047	0.035	0.018	1	1	1	1	1	1
	1000	20	9	0.060	0.032	0.019	0.051	0.040	0.021	1	1	1	1	1	1
SBM	500	10	6	0.033	0.017	0.007	0.018	0.009	0.004	1	1	1	1	1	1
	500	10	7	0.039	0.023	0.010	0.020	0.011	0.005	1	1	1	1	1	1
	1000	10	7	0.049	0.032	0.015	0.023	0.018	0.004	1	1	1	1	1	1
	1000	10	8	0.047	0.035	0.014	0.046	0.033	0.015	1	1	1	1	1	1
	500	20	7	0.037	0.023	0.007	0.023	0.014	0.005	1	1	1	1	1	1
	500	20	8	0.049	0.034	0.019	0.040	0.029	0.016	1	1	1	1	1	1
	1000	20	8	0.057	0.040	0.022	0.032	0.021	0.008	1	1	1	1	1	1
	1000	20	9	0.063	0.043	0.020	0.044	0.026	0.007	1	1	1	1	1	1

improves, especially for DGP 4.

D.5 Simulation under the power-law distribution model

In this subsection, we present the simulation results when the network is generated by a power-law distribution model. The power-law distribution captures the network phenomenon where the majority of nodes have few edges but a small amount have a huge number of edges. To mimic this phenomenon, we follow Zhu et al. (2017) and simulate A as follows. First, generate for each node its in-degree $d_i = \sum_{j=1}^{N} a_{ji}$ according to the discrete power-

Table D.3: Rejection probabilities of the LM and distance-based tests for testing of $H_0^\dagger$

(1) Test statistic: **T**				DGP 1			DGP 2			DGP 3			DGP 4		
Model	N	T	J	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
ER	100	10	4	0.048	0.031	0.014	0.493	0.430	0.316	0.037	0.019	0.008	0.102	0.065	0.019
	100	10	5	0.059	0.038	0.022	0.392	0.330	0.220	0.042	0.022	0.009	0.094	0.058	0.024
	200	10	4	0.059	0.039	0.023	0.832	0.784	0.688	0.048	0.027	0.011	0.192	0.121	0.057
	200	10	5	0.068	0.047	0.026	0.752	0.695	0.565	0.053	0.033	0.013	0.184	0.118	0.042
	100	20	4	0.072	0.049	0.024	0.803	0.742	0.627	0.041	0.024	0.011	0.222	0.153	0.075
	100	20	5	0.080	0.058	0.028	0.703	0.636	0.506	0.041	0.031	0.014	0.225	0.157	0.077
	200	20	4	0.049	0.035	0.016	0.984	0.978	0.950	0.100	0.054	0.018	0.531	0.410	0.240
	200	20	5	0.069	0.044	0.027	0.957	0.940	0.891	0.090	0.050	0.017	0.508	0.387	0.220
SBM	100	10	4	0.052	0.038	0.019	0.649	0.564	0.446	0.036	0.024	0.006	0.140	0.092	0.036
	100	10	5	0.070	0.050	0.028	0.528	0.451	0.354	0.040	0.024	0.009	0.147	0.095	0.038
	200	10	4	0.077	0.055	0.026	0.756	0.698	0.571	0.042	0.021	0.009	0.148	0.095	0.039
	200	10	5	0.074	0.051	0.028	0.646	0.551	0.438	0.041	0.024	0.009	0.149	0.090	0.031
	100	20	4	0.076	0.050	0.031	0.895	0.850	0.767	0.066	0.042	0.018	0.336	0.246	0.125
	100	20	5	0.079	0.054	0.030	0.809	0.758	0.645	0.071	0.047	0.019	0.318	0.230	0.114
	200	20	4	0.079	0.052	0.019	0.938	0.922	0.863	0.052	0.031	0.011	0.402	0.304	0.163
	200	20	5	0.084	0.053	0.027	0.889	0.849	0.748	0.056	0.034	0.013	0.380	0.292	0.155
(2) Test statistic: **LM**				DGP 1			DGP 2			DGP 3			DGP 4		
Model	N	T	J	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
ER	100	10	4	0.089	0.043	0.009	0.558	0.432	0.212	0.936	0.900	0.779	0.966	0.944	0.842
	100	10	5	0.088	0.039	0.006	0.519	0.371	0.181	0.920	0.873	0.722	0.956	0.913	0.788
	200	10	4	0.098	0.055	0.009	0.837	0.761	0.567	0.999	0.997	0.987	1	1	0.999
	200	10	5	0.095	0.045	0.014	0.812	0.721	0.525	0.998	0.997	0.975	1	1	0.997
	100	20	4	0.104	0.061	0.014	0.810	0.722	0.495	0.999	0.998	0.989	1	0.998	0.992
	100	20	5	0.111	0.064	0.011	0.777	0.660	0.446	0.998	0.997	0.983	0.998	0.997	0.991
	200	20	4	0.088	0.048	0.006	0.988	0.970	0.900	1	1	1	1	1	1
	200	20	5	0.087	0.047	0.007	0.977	0.950	0.868	1	1	1	1	1	1
SBM	100	10	4	0.089	0.048	0.007	0.667	0.541	0.345	0.956	0.912	0.789	0.971	0.943	0.837
	100	10	5	0.108	0.037	0.005	0.634	0.499	0.302	0.933	0.887	0.735	0.953	0.924	0.798
	200	10	4	0.097	0.042	0.004	0.801	0.691	0.454	0.998	0.996	0.990	1	1	0.992
	200	10	5	0.091	0.048	0.010	0.753	0.635	0.394	0.998	0.994	0.984	1	1	0.989
	100	20	4	0.108	0.059	0.014	0.889	0.831	0.665	1	0.999	0.991	1	1	0.997
	100	20	5	0.105	0.052	0.015	0.866	0.794	0.578	0.999	0.998	0.982	1	0.999	0.995
	200	20	4	0.116	0.056	0.011	0.947	0.902	0.752	1	1	1	1	1	1
	200	20	5	0.101	0.056	0.015	0.915	0.870	0.718	1	1	1	1	1	1

law distribution, that is, $P(d_i = k) = ck^{-\alpha}$ for a normalizing constant c and the exponent parameter $\alpha \in \{1.5, 3.0\}$. Next, for the ith node, we randomly select d_i nodes to be its followers.

Table D.8 reports the estimation results of (f_1, f_2) under the power-law distribution model. The DGPs we consider and other simulation settings are the same as in Section 5. We see a similar bias-variance trade-off pattern as in Table 2. Moreover, compared with the case $\alpha = 3$, the ISB, IV, and IMSE of f_1 are much larger under the case where $\alpha = 1.5$. This implies that a heavier distribution tail (thus a denser network) poses difficulty in recovering nonlinearity in f_1. The performance of the estimator of γ under the PLD model is also satisfactory. The detailed results are omitted for brevity.

Tables D.9 and D.10 present the rejection frequencies of the LM and $\mathbf{T}$ tests, respectively, under the power-law distribution model. Our findings are summarized as follows. First, the size of the LM test is reasonably well controlled at all nominal levels, while the $\mathbf{T}$ test is undersized. When the function f is highly nonlinear, the LM test has a good power property for all choices of J and all sample sizes, and also outperforms the $\mathbf{T}$ test in all cases. Second, when the tail of the distribution gets heavier (smaller α), the power to detect nonlinearity in f decreases. This is expected because a denser network implies a less precise estimate of f_1. The increase in the estimation bias then diminishes the power of the test considerably. Finally, the choice of J has a minor influence on the performance of the tests.

D.6 Impact of the network properties on the tests

As shown in Section D.5, the network density is a key factor influencing the empirical power of the tests. Apart from density, the network structure (topology) could also affect the performance of the tests. In this subsection, we examine the impact of the network structure on the tests using a larger set of simulations.

We consider two network models, each with two different densities: (1) stochastic block model with $K \in \{2, 5\}$, and (2) power-law distribution model with $\alpha \in \{1.1, 1.5\}$. The configurations of the network parameters are the same as Sections 5 and D.5, respectively. The sample sizes $N \in \{100, 200\}$ and $T = 20$. The parameters are chosen so that the SBM with $K = 2$ ($K = 5$) and PLD model with $\alpha = 1.1$ ($\alpha = 1.5$) have very similar densities. Specifically, when $N = 100$, the network density of the SBM is 16.32% for $K = 2$ and 7.13% for $K = 5$, and the network density of the PLD model is 16.35% for $\alpha = 1.1$ and 7.84% for $\alpha = 1.5$. When $N = 200$, the network density of the SBM is 13.56% for $K = 2$ and 5.73% for $K = 5$, and the network density of the PLD model is 14.07% for $\alpha = 1.1$ and 5.59% for $\alpha = 1.5$. From each network configuration, we compute the rejection frequencies of the $\mathbf{T}$ and LM tests for testing the hypotheses H_0^n and H_0^m. The results are recorded in Tables D.11 and D.12 below.

Table D.4: Estimation results of (γ_1, γ_2) when $N \in \{10, 50\}$ and $T \in \{100, 200\}$.

				DGP 1			DGP 2			DGP 3			DGP 4		
Model	N	T	J	Bias γ_1	RMSE γ_1	CR95 γ_1	Bias γ_1	RMSE γ_1	CR95 γ_1	Bias γ_1	RMSE γ_1	CR95 γ_1	Bias γ_1	RMSE γ_1	CR95 γ_1
ER	10	100	3	0.0036	0.0539	0.955	0.0081	0.0585	0.946	-0.0004	0.0482	0.955	-0.0026	0.0495	0.945
	10	100	4	0.0034	0.0538	0.954	0.0081	0.0588	0.947	-0.0004	0.0484	0.954	-0.0027	0.0496	0.949
	50	100	3	0.0020	0.0201	0.954	0.0022	0.0211	0.953	0.0013	0.0189	0.938	0.0004	0.0187	0.938
	50	100	4	0.0020	0.0201	0.957	0.0024	0.0211	0.954	0.0013	0.0189	0.939	0.0004	0.0187	0.936
	10	200	3	0.0018	0.0399	0.931	0.0012	0.0400	0.953	-0.0001	0.0335	0.949	0.0009	0.0347	0.951
	10	200	4	0.0019	0.0401	0.931	0.0012	0.0404	0.953	-0.0001	0.0335	0.949	0.0008	0.0347	0.952
	50	200	4	0.0012	0.0145	0.945	0.0016	0.0146	0.955	-0.0005	0.0130	0.940	-0.0004	0.0132	0.940
	50	200	5	0.0012	0.0145	0.945	0.0016	0.0147	0.955	-0.0005	0.0130	0.939	-0.0004	0.0132	0.940
SBM	10	100	3	0.0056	0.0593	0.949	0.0076	0.0624	0.940	0.0009	0.0575	0.946	0.0039	0.0545	0.953
	10	100	4	0.0060	0.0598	0.949	0.0081	0.0630	0.938	0.0008	0.0573	0.949	0.0040	0.0547	0.954
	50	100	3	0.0015	0.0205	0.943	0.0006	0.0210	0.946	0.0005	0.0180	0.945	0.0003	0.0177	0.955
	50	100	4	0.0016	0.0205	0.943	0.0009	0.0209	0.941	0.0005	0.0180	0.945	0.0002	0.0177	0.956
	10	200	3	0.0033	0.0416	0.952	0.0046	0.0448	0.956	0.0000	0.0391	0.932	0.0014	0.0397	0.940
	10	200	4	0.0032	0.0421	0.952	0.0048	0.0450	0.951	-0.0001	0.0391	0.933	0.0013	0.0397	0.940
	50	200	4	0.0008	0.0140	0.958	0.0003	0.0158	0.938	0.0001	0.0131	0.940	0.0003	0.0132	0.942
	50	200	5	0.0008	0.0140	0.955	0.0003	0.0158	0.941	0.0001	0.0131	0.939	0.0003	0.0132	0.945

				DGP 1			DGP 2			DGP 3			DGP 4		
Model	N	T	J	Bias γ_2	RMSE γ_2	CR95 γ_2	Bias γ_2	RMSE γ_2	CR95 γ_2	Bias γ_2	RMSE γ_2	CR95 γ_2	Bias γ_2	RMSE γ_2	CR95 γ_2
ER	10	100	3	0.0043	0.0523	0.956	0.0065	0.0581	0.949	0.0009	0.0487	0.948	0.0024	0.0493	0.940
	10	100	4	0.0042	0.0525	0.953	0.0064	0.0585	0.946	0.0008	0.0486	0.947	0.0024	0.0496	0.937
	50	100	3	0.0009	0.0196	0.962	0.0019	0.0222	0.943	-0.0003	0.0176	0.960	0.0001	0.0180	0.953
	50	100	4	0.0010	0.0196	0.962	0.0021	0.0222	0.939	-0.0004	0.0176	0.960	0.0001	0.0180	0.955
	10	200	3	0.0023	0.0370	0.959	0.0043	0.0409	0.952	-0.0002	0.0353	0.946	0.0005	0.0337	0.952
	10	200	4	0.0024	0.0372	0.958	0.0045	0.0411	0.956	-0.0003	0.0353	0.944	0.0005	0.0337	0.951
	50	200	4	0.0010	0.0141	0.949	0.0011	0.0152	0.947	0.0015	0.0129	0.942	0.0005	0.0128	0.950
	50	200	5	0.0010	0.0141	0.948	0.0011	0.0152	0.946	0.0015	0.0129	0.941	0.0005	0.0128	0.950
SBM	10	100	3	0.0051	0.0594	0.947	0.0089	0.0624	0.940	-0.0004	0.0572	0.944	-0.0003	0.0582	0.947
	10	100	4	0.0053	0.0597	0.952	0.0096	0.0630	0.938	-0.0003	0.0572	0.943	-0.0005	0.0582	0.946
	50	100	3	0.0023	0.0206	0.951	0.0019	0.0214	0.950	-0.0004	0.0183	0.947	0.0007	0.0177	0.967
	50	100	4	0.0024	0.0206	0.949	0.0023	0.0213	0.950	-0.0005	0.0183	0.945	0.0007	0.0177	0.967
	10	200	3	0.0033	0.0428	0.952	0.0029	0.0457	0.954	0.0013	0.0391	0.946	0.0005	0.0400	0.938
	10	200	4	0.0035	0.0432	0.952	0.0039	0.0464	0.951	0.0013	0.0393	0.948	0.0005	0.0400	0.940
	50	200	4	0.0003	0.0144	0.954	0.0006	0.0156	0.931	-0.0002	0.0127	0.959	0.0003	0.0129	0.946
	50	200	5	0.0002	0.0144	0.955	0.0006	0.0156	0.930	-0.0002	0.0127	0.960	0.0003	0.0129	0.948

Table D.5: Estimation results of f_1 and f_2 when $N \in \{10, 50\}$ and $T \in \{100, 200\}$.

				DGP 1		DGP 2		DGP 3		DGP 4	
Model	N	T	J	ISB f_1	IMSE f_1	ISB f_1	IMSE f_1	ISB f_1	IMSE f_1	ISB f_1	IMSE f_1
ER	10	100	3	0.0326	7.1772	0.0166	0.7369	0.0019	0.2623	0.0011	0.2256
	10	100	4	0.0505	22.8713	0.0291	0.8844	0.0017	0.3182	0.0003	0.2731
	50	100	3	0.0027	0.2261	0.0225	0.1910	0.0005	0.0911	0.0006	0.0854
	50	100	4	0.0024	0.4048	0.0010	0.1976	0.0005	0.1076	0.0004	0.0978
	10	200	3	0.0018	2.4315	0.0122	0.3993	0.0008	0.1385	0.0006	0.1131
	10	200	4	0.0035	7.9021	0.0094	0.4467	0.0006	0.1667	0.0004	0.1351
	50	200	4	0.0007	0.1763	0.0010	0.0983	0.0001	0.0538	0.0001	0.0520
	50	200	5	0.0006	0.2715	0.0002	0.1287	0.0001	0.0624	0.0000	0.0602
SBM	10	100	3	0.0006	0.6981	0.0098	0.1744	0.0001	0.0839	0.0006	0.0703
	10	100	4	0.0025	1.9120	0.0003	0.1952	0.0002	0.1043	0.0002	0.0844
	50	100	3	0.0010	0.1138	0.0195	0.0917	0.0000	0.0351	0.0011	0.0339
	50	100	4	0.0013	0.1846	0.0010	0.0793	0.0000	0.0427	0.0001	0.0385
	10	200	3	0.0001	0.3097	0.0080	0.1064	0.0002	0.0390	0.0009	0.0353
	10	200	4	0.0004	0.9183	0.0002	0.1133	0.0002	0.0478	0.0003	0.0400
	50	200	4	0.0001	0.1179	0.0011	0.0395	0.0000	0.0212	0.0002	0.0202
	50	200	5	0.0002	0.1501	0.0002	0.0517	0.0000	0.0251	0.0001	0.0241

				DGP 1		DGP 2		DGP 3		DGP 4	
Model	N	T	J	ISB f_2	IMSE f_2	ISB f_2	IMSE f_2	ISB f_2	IMSE f_2	ISB f_2	IMSE f_2
ER	10	100	3	0.0105	0.7713	0.0033	0.0715	0.0003	0.0303	0.0002	0.0282
	10	100	4	0.0030	3.9542	0.0037	0.0840	0.0001	0.0358	0.0001	0.0334
	50	100	3	0.0002	0.0154	0.0001	0.0130	0.0003	0.0060	0.0003	0.0057
	50	100	4	0.0003	0.0246	0.0002	0.0146	0.0000	0.0065	0.0000	0.0062
	10	200	3	0.0012	0.2745	0.0006	0.0365	0.0003	0.0157	0.0002	0.0135
	10	200	4	0.0086	1.1279	0.0007	0.0418	0.0001	0.0180	0.0000	0.0155
	50	200	4	0.0001	0.0122	0.0001	0.0069	0.0000	0.0032	0.0000	0.0031
	50	200	5	0.0001	0.0171	0.0001	0.0085	0.0000	0.0036	0.0000	0.0035
SBM	10	100	3	0.0061	0.2487	0.0033	0.0743	0.0006	0.0312	0.0005	0.0273
	10	100	4	0.0037	0.5339	0.0041	0.0884	0.0003	0.0356	0.0002	0.0312
	50	100	3	0.0002	0.0187	0.0001	0.0120	0.0004	0.0061	0.0002	0.0057
	50	100	4	0.0002	0.0283	0.0002	0.0136	0.0000	0.0066	0.0000	0.0061
	10	200	3	0.0017	0.1093	0.0009	0.0379	0.0002	0.0158	0.0003	0.0141
	10	200	4	0.0019	0.2464	0.0014	0.0477	0.0000	0.0180	0.0000	0.0160
	50	200	4	0.0001	0.0145	0.0000	0.0075	0.0000	0.0033	0.0000	0.0031
	50	200	5	0.0001	0.0179	0.0000	0.0092	0.0000	0.0037	0.0000	0.0035

Table D.6: Rejection probabilities of the distance-based test statistic when $N \in \{10, 50\}$ and $T \in \{100, 200\}$.

(1) Null: H_0^n				DGP 1			DGP 2			DGP 3			DGP 4		
Model	N	T	J	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
ER	10	100	3	0.049	0.035	0.020	0.427	0.340	0.246	0.011	0.006	0.003	0.173	0.122	0.070
	10	100	4	0.073	0.056	0.024	0.433	0.344	0.266	0.020	0.013	0.005	0.162	0.114	0.068
	50	100	3	0.065	0.040	0.020	0.970	0.957	0.916	0.022	0.010	0.002	0.509	0.422	0.267
	50	100	4	0.077	0.047	0.020	0.968	0.949	0.908	0.027	0.021	0.010	0.442	0.348	0.213
	10	200	3	0.062	0.040	0.019	0.674	0.626	0.496	0.007	0.005	0.000	0.363	0.300	0.190
	10	200	4	0.056	0.038	0.017	0.677	0.615	0.510	0.023	0.016	0.010	0.339	0.266	0.159
	50	200	4	0.060	0.043	0.021	1	0.998	0.996	0.028	0.012	0.006	0.813	0.731	0.563
	50	200	5	0.059	0.047	0.023	0.999	0.998	0.990	0.017	0.012	0.005	0.799	0.718	0.549
SBM	10	100	3	0.035	0.026	0.012	0.943	0.925	0.894	0.036	0.024	0.012	0.687	0.621	0.492
	10	100	4	0.050	0.035	0.015	0.928	0.908	0.857	0.044	0.033	0.016	0.624	0.557	0.402
	50	100	3	0.051	0.033	0.013	1	1	1	0.019	0.009	0.002	0.915	0.887	0.812
	50	100	4	0.062	0.043	0.019	1	1	1	0.033	0.019	0.009	0.890	0.837	0.720
	10	200	3	0.034	0.023	0.013	0.992	0.987	0.981	0.032	0.016	0.011	0.943	0.915	0.841
	10	200	4	0.038	0.021	0.010	0.987	0.980	0.966	0.044	0.028	0.014	0.906	0.871	0.756
	50	200	4	0.052	0.039	0.015	1	1	1	0.035	0.020	0.012	0.997	0.994	0.987
	50	200	5	0.063	0.042	0.017	1	1	1	0.039	0.022	0.010	0.995	0.993	0.987
(2) Null: H_0^m				DGP 1			DGP 2			DGP 3			DGP 4		
Model	N	T	J	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
ER	10	100	3	0.016	0.010	0.004	0.010	0.005	0.003	0.728	0.653	0.516	0.765	0.683	0.527
	10	100	4	0.021	0.015	0.008	0.010	0.006	0.003	0.682	0.592	0.445	0.709	0.623	0.458
	50	100	3	0.014	0.007	0.001	0.008	0.004	0.001	1	1	1	1	1	1
	50	100	4	0.014	0.007	0.001	0.010	0.006	0.001	1	1	1	1	1	0.999
	10	200	3	0.014	0.011	0.005	0.007	0.005	0.004	0.977	0.963	0.918	0.990	0.974	0.921
	10	200	4	0.023	0.017	0.007	0.010	0.006	0.003	0.962	0.934	0.859	0.964	0.941	0.872
	50	200	4	0.013	0.007	0.002	0.006	0.002	0.000	1	1	1	1	1	1
	50	200	5	0.016	0.010	0.005	0.022	0.011	0.006	1	1	1	1	1	1
SBM	10	100	3	0.007	0.006	0.002	0.015	0.008	0.002	0.840	0.783	0.654	0.809	0.734	0.591
	10	100	4	0.014	0.008	0.004	0.020	0.010	0.003	0.795	0.722	0.610	0.761	0.660	0.535
	50	100	3	0.007	0.002	0.002	0.004	0.003	0.002	1	1	1	1	1	1
	50	100	4	0.013	0.006	0.001	0.007	0.003	0.001	1	1	1	1	1	1
	10	200	3	0.012	0.008	0.002	0.010	0.006	0.001	0.995	0.990	0.967	0.985	0.976	0.942
	10	200	4	0.014	0.009	0.002	0.013	0.009	0.003	0.986	0.976	0.940	0.970	0.955	0.910
	50	200	4	0.012	0.007	0.000	0.007	0.003	0.000	1	1	1	1	1	1
	50	200	5	0.010	0.007	0.001	0.017	0.010	0.003	1	1	1	1	1	1

Table D.7: Rejection probabilities of the LM test statistic when $N \in \{10, 50\}$ and $T \in \{100, 200\}$.

(1) Null:H_0^n				DGP 1			DGP 2			DGP 3			DGP 4		
Model	N	T	J	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
ER	10	100	3	0.096	0.047	0.007	0.720	0.588	0.371	0.112	0.050	0.007	0.454	0.358	0.154
	10	100	4	0.122	0.052	0.013	0.659	0.543	0.315	0.104	0.053	0.009	0.432	0.323	0.134
	50	100	3	0.110	0.059	0.007	0.991	0.983	0.944	0.110	0.057	0.011	0.838	0.756	0.564
	50	100	4	0.113	0.060	0.008	0.986	0.979	0.925	0.120	0.048	0.011	0.791	0.707	0.489
	10	200	3	0.113	0.059	0.010	0.907	0.842	0.684	0.096	0.046	0.005	0.724	0.623	0.408
	10	200	4	0.106	0.051	0.008	0.867	0.803	0.646	0.097	0.046	0.009	0.686	0.576	0.357
	50	200	4	0.092	0.047	0.012	1	1	0.999	0.102	0.052	0.006	0.974	0.948	0.856
	50	200	5	0.097	0.046	0.012	1	1	0.998	0.091	0.049	0.006	0.965	0.933	0.823
SBM	10	100	3	0.107	0.058	0.007	0.983	0.968	0.917	0.107	0.056	0.010	0.917	0.864	0.709
	10	100	4	0.121	0.063	0.009	0.975	0.956	0.903	0.104	0.059	0.012	0.889	0.818	0.641
	50	100	3	0.093	0.052	0.008	1	1	1	0.085	0.050	0.009	0.991	0.982	0.932
	50	100	4	0.101	0.056	0.011	1	1	1	0.088	0.052	0.007	0.985	0.971	0.905
	10	200	3	0.111	0.065	0.007	1	0.998	0.991	0.106	0.046	0.009	0.988	0.980	0.953
	10	200	4	0.116	0.057	0.009	1	0.996	0.990	0.100	0.059	0.016	0.987	0.973	0.933
	50	200	4	0.096	0.047	0.008	1	1	1	0.115	0.064	0.009	1	0.999	0.998
	50	200	5	0.096	0.040	0.008	1	1	1	0.115	0.053	0.011	1	1	0.998
(2) Null:H_0^m				DGP 1			DGP 2			DGP 3			DGP 4		
Model	N	T	J	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
ER	10	100	3	0.094	0.042	0.007	0.113	0.057	0.010	0.946	0.917	0.818	0.971	0.943	0.854
	10	100	4	0.095	0.048	0.010	0.101	0.052	0.006	0.931	0.891	0.766	0.953	0.919	0.823
	50	100	3	0.123	0.069	0.015	0.100	0.062	0.011	1	1	1	1	1	1
	50	100	4	0.106	0.056	0.013	0.105	0.051	0.011	1	1	1	1	1	1
	10	200	3	0.099	0.041	0.010	0.114	0.062	0.010	1	0.998	0.988	0.999	0.999	0.994
	10	200	4	0.106	0.045	0.010	0.109	0.055	0.009	0.999	0.993	0.98	0.999	0.997	0.990
	50	200	4	0.100	0.056	0.009	0.106	0.054	0.010	1	1	1	1	1	1
	50	200	5	0.098	0.041	0.007	0.091	0.050	0.012	1	1	1	1	1	1
SBM	10	100	3	0.108	0.049	0.005	0.085	0.043	0.012	0.986	0.970	0.899	0.984	0.968	0.888
	10	100	4	0.096	0.039	0.004	0.079	0.040	0.010	0.977	0.953	0.865	0.973	0.949	0.857
	50	100	3	0.109	0.054	0.012	0.099	0.051	0.013	1	1	1	1	1	1
	50	100	4	0.113	0.062	0.013	0.095	0.056	0.016	1	1	1	1	1	1
	10	200	3	0.094	0.048	0.008	0.087	0.046	0.013	1	1	0.996	1	0.999	0.991
	10	200	4	0.088	0.041	0.008	0.092	0.038	0.009	1	1	0.997	0.998	0.996	0.990
	50	200	4	0.100	0.052	0.006	0.101	0.053	0.011	1	1	1	1	1	1
	50	200	5	0.100	0.045	0.007	0.088	0.043	0.010	1	1	1	1	1	1

Table D.8: Estimation results of f_1 and f_2 under the power-law distributional model

				DGP 1		DGP 2		DGP 3		DGP 4	
α	N	T	J	ISB f_1	IMSE f_1	ISB f_1	IMSE f_1	ISB f_1	IMSE f_1	ISB f_1	IMSE f_1
1.5	100	10	4	0.0072	7.5084	0.0037	0.4135	0.0003	0.2272	0.0011	0.2104
	100	20	5	0.0048	17.8325	0.0022	0.5544	0.0002	0.2782	0.0009	0.2468
	200	10	4	0.0007	1.7182	0.0017	0.2261	0.0007	0.1348	0.0002	0.1200
	200	20	5	0.0031	3.8193	0.0005	0.3120	0.0010	0.1606	0.0001	0.1426
	100	10	4	0.0036	2.9190	0.0013	0.1968	0.0001	0.1123	0.0003	0.0942
	100	20	5	0.0007	5.6990	0.0002	0.2610	0.0001	0.1323	0.0003	0.1124
	200	10	4	0.0004	0.6698	0.0018	0.1140	0.0001	0.0633	0.0003	0.0602
	200	20	5	0.0010	1.1138	0.0002	0.1543	0.0001	0.0755	0.0002	0.0719
3.0	100	10	4	0.0005	0.1505	0.0012	0.0911	0.0001	0.0479	0.0002	0.0459
	100	20	5	0.0004	0.1794	0.0014	0.1206	0.0002	0.0609	0.0001	0.0569
	200	10	4	0.0002	0.0878	0.0014	0.0453	0.0000	0.0207	0.0004	0.0212
	200	20	5	0.0003	0.1103	0.0007	0.0586	0.0000	0.0277	0.0001	0.0267
	100	10	4	0.0001	0.0646	0.0008	0.0425	0.0000	0.0236	0.0002	0.0221
	100	20	5	0.0001	0.0794	0.0009	0.0588	0.0001	0.0297	0.0000	0.0272
	200	10	4	0.0002	0.0382	0.0011	0.0226	0.0000	0.0112	0.0004	0.0108
	200	20	5	0.0003	0.0470	0.0007	0.0299	0.0000	0.0147	0.0001	0.0137

				DGP 1		DGP 2		DGP 3		DGP 4	
α	N	T	J	ISB f_2	IMSE f_2	ISB f_2	IMSE f_2	ISB f_2	IMSE f_2	ISB f_2	IMSE f_2
1.5	100	10	4	0.0119	0.6354	0.0011	0.0687	0.0001	0.0336	0.0002	0.0307
	100	20	5	0.0166	1.0486	0.0011	0.0866	0.0001	0.0385	0.0002	0.0357
	200	10	4	0.0006	0.1037	0.0002	0.0319	0.0001	0.0160	0.0001	0.0148
	200	20	5	0.0009	0.1505	0.0002	0.0395	0.0000	0.0183	0.0000	0.0171
	100	10	4	0.0009	0.1992	0.0009	0.0343	0.0000	0.0157	0.0001	0.0153
	100	20	5	0.0008	0.2778	0.0009	0.0433	0.0000	0.0178	0.0001	0.0174
	200	10	4	0.0003	0.0452	0.0002	0.0154	0.0000	0.0080	0.0000	0.0076
	200	20	5	0.0004	0.0572	0.0002	0.0195	0.0000	0.0091	0.0000	0.0088
3.0	100	10	4	0.0018	0.0819	0.0027	0.0698	0.0001	0.0326	0.0003	0.0282
	100	20	5	0.0014	0.0966	0.0028	0.0875	0.0000	0.0396	0.0003	0.0337
	200	10	4	0.0010	0.0566	0.0011	0.0341	0.0001	0.0145	0.0001	0.0152
	200	20	5	0.0008	0.0660	0.0011	0.0429	0.0000	0.0181	0.0001	0.0182
	100	10	4	0.0005	0.0377	0.0009	0.0360	0.0001	0.0151	0.0001	0.0148
	100	20	5	0.0003	0.0439	0.0009	0.0439	0.0000	0.0184	0.0000	0.0180
	200	10	4	0.0005	0.0262	0.0003	0.0176	0.0001	0.0076	0.0001	0.0075
	200	20	5	0.0003	0.0299	0.0003	0.0213	0.0000	0.0093	0.0000	0.0090

Table D.9: Rejection probabilities of the LM test under the power-law distributional model

(1) Null: H_0^n				DGP 1			DGP 2			DGP 3			DGP 4		
α	N	T	J	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
1.5	100	10	4	0.087	0.034	0.008	0.886	0.815	0.670	0.103	0.058	0.011	0.536	0.398	0.202
	100	10	5	0.090	0.039	0.007	0.855	0.787	0.629	0.111	0.062	0.013	0.488	0.363	0.165
	200	10	4	0.110	0.055	0.010	0.992	0.983	0.941	0.099	0.056	0.013	0.726	0.588	0.382
	200	10	5	0.105	0.054	0.007	0.985	0.975	0.924	0.097	0.045	0.011	0.671	0.545	0.349
	100	20	4	0.098	0.045	0.005	0.992	0.981	0.937	0.104	0.053	0.009	0.782	0.698	0.486
	100	20	5	0.096	0.046	0.010	0.986	0.976	0.920	0.103	0.048	0.005	0.751	0.648	0.423
	200	20	4	0.093	0.055	0.014	0.999	0.999	0.996	0.108	0.044	0.010	0.933	0.895	0.774
	200	20	5	0.104	0.056	0.008	0.999	0.998	0.994	0.099	0.049	0.008	0.925	0.878	0.731
3.0	100	10	4	0.107	0.051	0.006	1	1	1	0.109	0.053	0.012	0.993	0.983	0.933
	100	10	5	0.115	0.046	0.009	1	1	1	0.094	0.044	0.015	0.988	0.970	0.925
	200	10	4	0.097	0.045	0.012	1	1	1	0.093	0.049	0.007	1	1	1
	200	10	5	0.101	0.047	0.010	1	1	1	0.103	0.052	0.012	1	1	0.999
	100	20	4	0.105	0.061	0.014	1	1	1	0.101	0.063	0.009	1	1	0.999
	100	20	5	0.112	0.066	0.012	1	1	1	0.104	0.052	0.012	1	1	0.999
	200	20	4	0.126	0.062	0.014	1	1	1	0.099	0.042	0.010	1	1	1
	200	20	5	0.120	0.063	0.011	1	1	1	0.092	0.052	0.009	1	1	1
(2) Null: H_0^m				DGP 1			DGP 2			DGP 3			DGP 4		
α	N	T	J	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
1.5	100	10	4	0.106	0.050	0.014	0.089	0.043	0.009	0.975	0.948	0.853	0.978	0.955	0.861
	100	10	5	0.099	0.047	0.012	0.086	0.047	0.009	0.966	0.933	0.821	0.962	0.939	0.829
	200	10	4	0.104	0.048	0.012	0.092	0.048	0.004	1	1	0.997	1	1	0.998
	200	10	5	0.097	0.047	0.015	0.097	0.041	0.006	1	1	0.996	1	0.998	0.996
	100	20	4	0.102	0.053	0.013	0.101	0.058	0.012	0.998	0.998	0.993	1	0.999	0.997
	100	20	5	0.097	0.057	0.012	0.107	0.063	0.015	0.998	0.996	0.987	1	0.998	0.997
	200	20	4	0.112	0.046	0.010	0.101	0.052	0.008	1	1	1	1	1	1
	200	20	5	0.108	0.054	0.009	0.097	0.044	0.008	1	1	1	1	1	1
3.0	100	10	4	0.109	0.059	0.012	0.099	0.059	0.015	0.996	0.992	0.972	0.989	0.983	0.941
	100	10	5	0.109	0.052	0.008	0.102	0.057	0.014	0.997	0.990	0.957	0.984	0.972	0.919
	200	10	4	0.090	0.046	0.009	0.105	0.060	0.009	1	1	1	1	1	1
	200	10	5	0.096	0.043	0.010	0.103	0.058	0.019	1	1	1	1	1	0.999
	100	20	4	0.100	0.047	0.007	0.117	0.059	0.011	1	1	1	1	1	1
	100	20	5	0.093	0.050	0.007	0.113	0.046	0.011	1	1	1	1	1	0.999
	200	20	4	0.101	0.061	0.015	0.098	0.048	0.010	1	1	1	1	1	1
	200	20	5	0.099	0.057	0.017	0.085	0.048	0.009	1	1	1	1	1	1

Table D.10: Rejection probabilities of the distance-based test under the power-law distributional model

(1) Null: H_0^n				DGP 1			DGP 2			DGP 3			DGP 4		
α	N	T	J	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
1.5	100	10	4	0.056	0.038	0.015	0.787	0.730	0.642	0.042	0.031	0.012	0.195	0.131	0.070
	100	10	5	0.068	0.046	0.019	0.708	0.648	0.532	0.055	0.039	0.014	0.177	0.122	0.068
	200	10	4	0.070	0.041	0.016	0.961	0.939	0.891	0.040	0.029	0.017	0.317	0.240	0.138
	200	10	5	0.069	0.043	0.017	0.928	0.896	0.825	0.044	0.032	0.015	0.287	0.216	0.118
	100	20	4	0.048	0.031	0.014	0.963	0.950	0.913	0.043	0.025	0.009	0.432	0.337	0.203
	100	20	5	0.059	0.041	0.023	0.938	0.913	0.864	0.048	0.026	0.014	0.383	0.298	0.177
	200	20	4	0.066	0.048	0.026	0.998	0.998	0.993	0.029	0.019	0.008	0.690	0.594	0.449
	200	20	5	0.076	0.055	0.028	0.999	0.996	0.986	0.046	0.028	0.010	0.657	0.567	0.413
3.0	100	10	4	0.054	0.033	0.014	1	1	1	0.050	0.028	0.008	0.930	0.892	0.803
	100	10	5	0.063	0.039	0.016	1	1	1	0.049	0.027	0.013	0.913	0.864	0.750
	200	10	4	0.037	0.021	0.012	1	1	1	0.056	0.031	0.014	0.999	0.998	0.996
	200	10	5	0.043	0.025	0.010	1	1	1	0.056	0.038	0.017	0.998	0.998	0.992
	100	20	4	0.067	0.037	0.015	1	1	1	0.054	0.032	0.010	0.998	0.997	0.988
	100	20	5	0.076	0.045	0.021	1	1	1	0.057	0.034	0.012	0.999	0.993	0.984
	200	20	4	0.051	0.035	0.015	1	1	1	0.044	0.020	0.006	1	1	1
	200	20	5	0.052	0.037	0.017	1	1	1	0.046	0.021	0.006	1	1	1

(2) Null: H_0^m				DGP 1			DGP 2			DGP 3			DGP 4		
α	N	T	J	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
1.5	100	10	4	0.023	0.014	0.006	0.010	0.008	0.001	0.698	0.618	0.458	0.725	0.610	0.432
	100	10	5	0.025	0.015	0.010	0.024	0.014	0.004	0.693	0.616	0.433	0.719	0.599	0.425
	200	10	4	0.015	0.006	0.004	0.006	0.005	0.001	0.979	0.968	0.902	0.989	0.975	0.928
	200	10	5	0.015	0.009	0.003	0.013	0.005	0.001	0.980	0.966	0.892	0.987	0.971	0.924
	100	20	4	0.013	0.007	0.001	0.010	0.007	0.002	0.972	0.956	0.904	0.986	0.976	0.924
	100	20	5	0.020	0.012	0.004	0.025	0.016	0.010	0.967	0.951	0.893	0.985	0.965	0.923
	200	20	4	0.010	0.006	0.004	0.008	0.004	0.001	1	1	0.998	1	1	1
	200	20	5	0.022	0.015	0.011	0.010	0.004	0.002	1	0.999	0.997	1	1	1
3.0	100	10	4	0.023	0.014	0.004	0.015	0.012	0.008	0.947	0.919	0.836	0.866	0.818	0.664
	100	10	5	0.025	0.013	0.004	0.028	0.020	0.011	0.942	0.898	0.798	0.853	0.793	0.629
	200	10	4	0.021	0.010	0.005	0.019	0.011	0.002	1	1	0.998	0.999	0.998	0.986
	200	10	5	0.021	0.016	0.009	0.034	0.023	0.012	1	1	0.997	0.999	0.997	0.988
	100	20	4	0.018	0.008	0.001	0.018	0.009	0.001	1	1	0.996	0.997	0.993	0.983
	100	20	5	0.025	0.009	0.002	0.022	0.012	0.006	1	1	0.996	0.995	0.993	0.980
	200	20	4	0.026	0.019	0.010	0.008	0.005	0.000	1	1	1	1	1	1
	200	20	5	0.027	0.018	0.008	0.024	0.012	0.004	1	1	1	1	1	1

Table D.11: Rejection probabilities of the distance-based test under SBM and PLD model with different densities

(1) Null : H_0^n			DGP 1			DGP 2			DGP 3			DGP 4		
Model	N	J	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
SBM ($K = 2$)	100	4	0.070	0.052	0.029	0.889	0.843	0.740	0.045	0.025	0.014	0.384	0.311	0.196
	100	5	0.082	0.054	0.024	0.865	0.801	0.702	0.058	0.040	0.015	0.352	0.272	0.165
	200	4	0.085	0.049	0.025	0.931	0.908	0.854	0.050	0.030	0.013	0.480	0.392	0.275
	200	5	0.079	0.059	0.026	0.918	0.882	0.809	0.048	0.032	0.010	0.446	0.363	0.233
SBM ($K = 5$)	100	4	0.043	0.032	0.013	1	1	0.995	0.039	0.024	0.010	0.788	0.735	0.623
	100	5	0.046	0.025	0.006	1	0.999	0.992	0.047	0.030	0.008	0.768	0.704	0.574
	200	4	0.057	0.035	0.013	1	1	1	0.047	0.029	0.010	0.896	0.864	0.783
	200	5	0.060	0.034	0.018	1	1	1	0.053	0.033	0.015	0.879	0.845	0.733
PLD ($\alpha = 1.1$)	100	4	0.083	0.053	0.024	0.754	0.699	0.569	0.060	0.040	0.018	0.334	0.272	0.171
	100	5	0.093	0.056	0.030	0.697	0.622	0.500	0.074	0.045	0.018	0.323	0.242	0.148
	200	4	0.072	0.057	0.020	0.888	0.854	0.774	0.051	0.031	0.018	0.450	0.392	0.267
	200	5	0.089	0.051	0.029	0.864	0.818	0.691	0.062	0.036	0.023	0.441	0.355	0.231
PLD ($\alpha = 1.5$)	100	4	0.054	0.028	0.013	0.973	0.960	0.927	0.050	0.025	0.012	0.666	0.596	0.478
	100	5	0.054	0.035	0.016	0.953	0.938	0.894	0.049	0.035	0.011	0.611	0.536	0.407
	200	4	0.067	0.042	0.019	0.996	0.994	0.988	0.059	0.039	0.022	0.790	0.734	0.605
	200	5	0.071	0.046	0.018	0.991	0.989	0.968	0.069	0.048	0.019	0.736	0.663	0.534
(2) Null: H_0^m			DGP 1			DGP 2			DGP 3			DGP 4		
Model	N	J	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
SBM ($K = 2$)	100	4	0.017	0.008	0.002	0.008	0.003	0.001	0.975	0.954	0.893	0.987	0.976	0.926
	100	5	0.022	0.009	0.004	0.016	0.010	0.006	0.969	0.950	0.876	0.987	0.973	0.921
	200	4	0.018	0.009	0.003	0.007	0.003	0.000	1	1	1	1	1	1
	200	5	0.022	0.010	0.001	0.019	0.011	0.003	1	1	1	1	1	1
SBM ($K = 5$)	100	4	0.010	0.005	0.001	0.008	0.005	0.002	0.987	0.970	0.918	0.991	0.982	0.936
	100	5	0.016	0.007	0.003	0.014	0.008	0.004	0.984	0.963	0.911	0.989	0.977	0.935
	200	4	0.011	0.006	0.004	0.011	0.005	0.002	1	1	1	1	1	1
	200	5	0.017	0.008	0.003	0.018	0.010	0.006	1	1	1	1	1	1
PLD ($\alpha = 1.1$)	100	4	0.009	0.006	0.002	0.007	0.003	0.001	0.979	0.952	0.880	0.980	0.975	0.927
	100	5	0.014	0.011	0.004	0.019	0.009	0.003	0.970	0.937	0.870	0.979	0.973	0.925
	200	4	0.012	0.008	0.003	0.007	0.002	0.000	1	1	1	1	1	1
	200	5	0.017	0.009	0.004	0.012	0.005	0.002	1	1	0.998	1	1	1
PLD ($\alpha = 1.5$)	100	4	0.009	0.004	0.000	0.005	0.003	0.001	0.983	0.963	0.904	0.990	0.971	0.935
	100	5	0.010	0.004	0.000	0.023	0.010	0.002	0.977	0.955	0.893	0.986	0.971	0.935
	200	4	0.011	0.006	0.000	0.008	0.004	0.001	1	1	0.999	1	1	1
	200	5	0.017	0.007	0.004	0.018	0.011	0.003	1	0.999	0.999	1	1	1

Table D.12: Rejection probabilities of the LM test under SBM and PLD model with different densities

(1) Null: H_0^n			DGP 1			DGP 2			DGP 3			DGP 4		
Model	N	J	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
SBM ($K=2$)	100	4	0.106	0.058	0.015	0.955	0.908	0.765	0.095	0.039	0.013	0.514	0.407	0.181
	100	5	0.102	0.057	0.015	0.935	0.879	0.724	0.089	0.055	0.012	0.488	0.346	0.148
	200	4	0.107	0.056	0.013	0.971	0.951	0.868	0.102	0.051	0.008	0.614	0.472	0.247
	200	5	0.104	0.054	0.017	0.970	0.928	0.830	0.095	0.047	0.010	0.558	0.419	0.206
SBM ($K=5$)	100	4	0.107	0.057	0.010	1	1	0.996	0.093	0.047	0.011	0.882	0.827	0.640
	100	5	0.100	0.051	0.006	1	1	0.995	0.106	0.053	0.010	0.864	0.795	0.588
	200	4	0.091	0.046	0.006	1	1	1	0.105	0.053	0.009	0.948	0.904	0.768
	200	5	0.097	0.042	0.009	1	1	1	0.111	0.052	0.016	0.927	0.879	0.714
PLD model ($\alpha=1.1$)	100	4	0.115	0.064	0.011	0.881	0.822	0.652	0.110	0.056	0.014	0.506	0.374	0.192
	100	5	0.122	0.067	0.016	0.844	0.785	0.603	0.109	0.051	0.009	0.472	0.351	0.171
	200	4	0.105	0.058	0.016	0.972	0.944	0.838	0.105	0.049	0.010	0.635	0.510	0.272
	200	5	0.109	0.058	0.019	0.962	0.927	0.803	0.104	0.053	0.013	0.596	0.482	0.253
PLD model ($\alpha=1.5$)	100	4	0.111	0.057	0.012	0.991	0.985	0.962	0.104	0.048	0.010	0.845	0.768	0.542
	100	5	0.105	0.050	0.013	0.989	0.980	0.953	0.098	0.048	0.009	0.833	0.714	0.510
	200	4	0.114	0.062	0.013	1	1	0.995	0.107	0.058	0.009	0.912	0.840	0.648
	200	5	0.106	0.061	0.013	1	1	0.995	0.115	0.057	0.008	0.884	0.803	0.602
(2) Null: H_0^m			DGP 1			DGP 2			DGP 3			DGP 4		
Model	N	J	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
SBM ($K=2$)	100	4	0.100	0.050	0.005	0.101	0.049	0.008	0.999	0.999	0.993	1	1	0.996
	100	5	0.099	0.045	0.010	0.096	0.058	0.009	0.999	0.999	0.992	1	1	0.995
	200	4	0.106	0.046	0.009	0.100	0.053	0.012	1	1	1	1	1	1
	200	5	0.091	0.048	0.005	0.103	0.049	0.008	1	1	1	1	1	1
SBM ($K=5$)	100	4	0.084	0.047	0.011	0.098	0.051	0.012	1	1	1	1	1	0.998
	100	5	0.097	0.048	0.011	0.084	0.045	0.010	1	1	0.999	1	0.999	0.994
	200	4	0.102	0.047	0.008	0.088	0.050	0.011	1	1	1	1	1	1
	200	5	0.103	0.043	0.008	0.094	0.052	0.009	1	1	1	1	1	1
PLD model ($\alpha=1.1$)	100	4	0.096	0.044	0.006	0.112	0.058	0.008	1	1	0.998	1	1	0.995
	100	5	0.094	0.043	0.006	0.116	0.046	0.009	1	1	0.995	1	0.999	0.995
	200	4	0.099	0.049	0.010	0.092	0.051	0.013	1	1	1	1	1	1
	200	5	0.110	0.055	0.009	0.090	0.050	0.011	1	1	1	1	1	1
PLD model ($\alpha=1.5$)	100	4	0.101	0.052	0.010	0.093	0.046	0.005	1	1	0.997	1	1	0.998
	100	5	0.100	0.046	0.010	0.091	0.039	0.006	1	1	0.995	1	0.998	0.995
	200	4	0.095	0.052	0.010	0.096	0.044	0.009	1	1	1	1	1	1
	200	5	0.090	0.046	0.014	0.089	0.033	0.009	1	1	1	1	1	1

Our findings from Tables D.11 and D.12 are as follows. First, given identical network structures, the empirical power for testing H_0^n shows an inverse relationship with network density, whereas the power for testing H_0^m remains relatively stable across varying density levels. Second, structural changes that preserve network density have a negligible impact on test power. Finally, given the sample sizes N and T, the size of the LM test is robust to both network density and structure.

D.7 Simulations with mildly nonlinear f_1 and f_2

The DGPs considered in Section 5 involve trigonometric functions f_1 and f_2, which may be rarely seen in practice. In this subsection, we examine the finite sample performance of the estimators and tests under more empirically relevant DGPs. Specifically, we generate the data according to model in Section 5 with

$$\begin{aligned}
&\text{DGP 5: } f_1(y) = 0.4\ln(|y-1|+1) * \text{sgn}(y-1), \;\; f_2(y) = 0.5y, \\
&\text{DGP 6: } f_1(y) = 0.5y, \;\; f_2(y) = 0.4\ln(|y-1|+1) * \text{sgn}(y-1), \\
&\text{DGP 7: } f_1(y) = 0.5\ln(|y-1|+1) * \text{sgn}(y-1), \;\; f_2(y) = 0.4\ln(|y-1|+1) * \text{sgn}(y-1), \\
&\text{DGP 8: } f_1(y) = 0.5\ln(|y-1|+1) * \text{sgn}(y-1), \;\; f_2(y) = \cos(0.4y),
\end{aligned}$$

where $\text{sgn}(\cdot)$ denotes the sign function. Note that f_1 in DGPs 5, 7, and 8 is less smooth compared to that in DGPs 1–4. The derivative of f_1 has a kink at $y = 1$ (this design is borrowed from Newey and Powell (2003)).

Table D.13 presents the results of estimating (γ_1, γ_2) under DGPs 5–8. They show that our estimator performs quite well under the new DGPs. The magnitudes of the bias and RMSE are similar to those under DGPs 1–4 when the sample size is held fixed. The results of estimating f_1 and f_2 under DGPs 5–8 are presented in Table D.14. Compared with Table 2, the IMSE of $\widehat{f}_1$ and $\widehat{f}_2$ is relatively smaller, while the ISB is slightly larger for the same set of (N, T, J). This is probably because f is less smooth than DGPs 1–4. Lastly, the choice of J is a crucial factor that determines the performance of the sieve estimator.

Tables D.15 and D.16 present the rejection frequencies of the **T** and LM tests, respectively.

Table D.13: Estimation results of (γ_1, γ_2) under DGPs 5-8

				DGP 5			DGP 6			DGP 7			DGP 8		
Model	N	T	J	Bias γ_1	RMSE γ_1	CR95 γ_1	Bias γ_1	RMSE γ_1	CR95 γ_1	Bias γ_1	RMSE γ_1	CR95 γ_1	Bias γ_1	RMSE γ_1	CR95 γ_1
ER	100	10	4	0.0057	0.0447	0.957	-0.0002	0.0423	0.948	0.0010	0.0428	0.940	-0.0002	0.0392	0.957
	100	10	5	0.0058	0.0448	0.955	-0.0004	0.0424	0.949	0.0010	0.0428	0.940	-0.0001	0.0392	0.953
	300	10	4	0.0026	0.0249	0.963	0.0009	0.0242	0.939	0.0008	0.0237	0.950	0.0003	0.0232	0.930
	300	10	5	0.0026	0.0250	0.963	0.0008	0.0242	0.939	0.0007	0.0237	0.949	0.0002	0.0231	0.931
	100	30	4	0.0026	0.0276	0.946	0.0002	0.0242	0.957	0.0016	0.0242	0.951	0.0011	0.0220	0.961
	100	30	5	0.0026	0.0276	0.945	0.0001	0.0242	0.959	0.0015	0.0242	0.951	0.0011	0.0220	0.961
	300	30	4	0.0018	0.0157	0.938	0.0004	0.0137	0.953	0.0004	0.0139	0.945	0.0006	0.0130	0.948
	300	30	5	0.0018	0.0157	0.938	0.0003	0.0137	0.954	0.0003	0.0139	0.946	0.0006	0.0130	0.947
SBM	100	10	4	0.0079	0.0456	0.948	0.0010	0.0407	0.954	-0.0006	0.0425	0.952	-0.0003	0.0418	0.947
	100	10	5	0.0079	0.0457	0.949	0.0009	0.0408	0.954	-0.0008	0.0425	0.952	-0.0004	0.0419	0.946
	300	10	4	0.0027	0.0267	0.943	0.0005	0.0230	0.956	0.0000	0.0242	0.939	0.0005	0.0228	0.955
	300	10	5	0.0027	0.0268	0.941	0.0005	0.0231	0.955	0.0000	0.0242	0.939	0.0005	0.0228	0.957
	100	30	4	0.0023	0.0261	0.956	0.0017	0.0230	0.959	0.0005	0.0243	0.950	0.0011	0.0241	0.934
	100	30	5	0.0022	0.0261	0.957	0.0016	0.0230	0.960	0.0004	0.0243	0.951	0.0010	0.0241	0.934
	300	30	4	0.0000	0.0152	0.947	0.0007	0.0140	0.947	0.0005	0.0139	0.949	0.0010	0.0135	0.942
	300	30	5	0.0000	0.0151	0.948	0.0007	0.0140	0.948	0.0004	0.0139	0.949	0.0010	0.0135	0.943
				DGP 5			DGP 6			DGP 7			DGP 8		
Model	N	T	J	Bias γ_2	RMSE γ_2	CR95 γ_2	Bias γ_2	RMSE γ_2	CR95 γ_2	Bias γ_2	RMSE γ_2	CR95 γ_2	Bias γ_2	RMSE γ_2	CR95 γ_2
ER	100	10	4	0.0073	0.0471	0.942	0.0019	0.0423	0.949	0.0014	0.0420	0.952	0.0029	0.0403	0.945
	100	10	5	0.0073	0.0474	0.942	0.0017	0.0424	0.949	0.0013	0.0420	0.953	0.0029	0.0402	0.947
	300	10	4	0.0015	0.0260	0.955	0.0015	0.0246	0.944	0.0009	0.0237	0.946	0.0005	0.0231	0.947
	300	10	5	0.0015	0.0260	0.955	0.0014	0.0246	0.945	0.0009	0.0237	0.945	0.0005	0.0231	0.949
	100	30	4	0.0029	0.0266	0.949	0.0019	0.0243	0.950	-0.0004	0.0249	0.940	-0.0004	0.0231	0.951
	100	30	5	0.0030	0.0267	0.950	0.0018	0.0243	0.948	-0.0006	0.0250	0.939	-0.0004	0.0231	0.950
	300	30	4	0.0011	0.0150	0.956	0.0007	0.0139	0.941	0.0001	0.0137	0.952	0.0007	0.0135	0.948
	300	30	5	0.0011	0.0150	0.957	0.0007	0.0139	0.942	0.0000	0.0137	0.949	0.0007	0.0135	0.945
SBM	100	10	4	0.0065	0.0458	0.947	0.0039	0.0428	0.944	0.0049	0.0423	0.942	0.0014	0.0389	0.958
	100	10	5	0.0064	0.0458	0.943	0.0039	0.0429	0.942	0.0046	0.0424	0.938	0.0014	0.0388	0.959
	300	10	4	0.0025	0.0249	0.958	0.0004	0.0233	0.954	0.0012	0.0236	0.946	0.0001	0.0230	0.953
	300	10	5	0.0025	0.0249	0.960	0.0003	0.0233	0.952	0.0011	0.0236	0.949	0.0001	0.0230	0.952
	100	30	4	0.0025	0.0278	0.936	0.0013	0.0243	0.951	0.0021	0.0248	0.939	-0.0008	0.0236	0.945
	100	30	5	0.0025	0.0278	0.937	0.0012	0.0243	0.952	0.0020	0.0248	0.941	-0.0008	0.0236	0.945
	300	30	4	0.0002	0.0154	0.945	0.0001	0.0142	0.939	0.0006	0.0141	0.946	0.0003	0.0128	0.956
	300	30	5	0.0002	0.0153	0.945	0.0000	0.0142	0.937	0.0005	0.0141	0.947	0.0002	0.0128	0.957

Table D.14: Estimation results of f_1 and f_2 under DGPs 5-8

				DGP 5		DGP 6		DGP 7		DGP 8	
Model	N	T	J	ISB f_1	IMSE f_1	ISB f_1	IMSE f_1	ISB f_1	IMSE f_1	ISB f_1	IMSE f_1
ER	100	10	4	0.0131	1.0000	0.0069	0.6105	0.0138	0.5592	0.0077	0.4504
	100	10	5	0.0156	1.2864	0.0069	0.8507	0.0100	0.8067	0.0028	0.6594
	300	10	4	0.0105	0.2888	0.0010	0.1842	0.0072	0.1740	0.0065	0.1529
	300	10	5	0.0125	0.3690	0.0009	0.2574	0.0022	0.2410	0.0005	0.2131
	100	30	4	0.0092	0.3112	0.0007	0.2041	0.0077	0.1877	0.0063	0.1585
	100	30	5	0.0088	0.3907	0.0009	0.2760	0.0015	0.2560	0.0004	0.2204
	300	30	4	0.0104	0.1026	0.0001	0.0601	0.0070	0.0641	0.0076	0.0567
	300	30	5	0.0112	0.1246	0.0002	0.0840	0.0014	0.0818	0.0006	0.0701
SBM	100	10	4	0.0172	0.8156	0.0049	0.4881	0.0106	0.4747	0.0080	0.3646
	100	10	5	0.0121	1.0474	0.0045	0.6588	0.0066	0.6620	0.0031	0.5253
	300	10	4	0.0108	0.5383	0.0048	0.3035	0.0095	0.3236	0.0071	0.2648
	300	10	5	0.0122	0.6765	0.0052	0.4275	0.0037	0.4515	0.0015	0.3739
	100	30	4	0.0115	0.2462	0.0014	0.1512	0.0083	0.1561	0.0049	0.1223
	100	30	5	0.0117	0.3069	0.0014	0.2088	0.0038	0.2210	0.0011	0.1686
	300	30	4	0.0097	0.1710	0.0005	0.1063	0.0063	0.1162	0.0071	0.0954
	300	30	5	0.0102	0.2107	0.0005	0.1477	0.0012	0.1562	0.0011	0.1272

				DGP 5		DGP 6		DGP 7		DGP 8	
Model	N	T	J	ISB f_2	IMSE f_2	ISB f_2	IMSE f_2	ISB f_2	IMSE f_2	ISB f_2	IMSE f_2
ER	100	10	4	0.0026	0.0708	0.0037	0.0416	0.0054	0.0422	0.0001	0.0271
	100	10	5	0.0024	0.0843	0.0007	0.0477	0.0009	0.0492	0.0001	0.0368
	300	10	4	0.0002	0.0213	0.0037	0.0153	0.0046	0.0156	0.0000	0.0090
	300	10	5	0.0002	0.0249	0.0009	0.0161	0.0008	0.0154	0.0000	0.0123
	100	30	4	0.0004	0.0229	0.0038	0.0166	0.0042	0.0157	0.0000	0.0098
	100	30	5	0.0003	0.0270	0.0008	0.0171	0.0006	0.0159	0.0000	0.0128
	300	30	4	0.0001	0.0076	0.0041	0.0080	0.0043	0.0082	0.0000	0.0029
	300	30	5	0.0001	0.0086	0.0010	0.0063	0.0007	0.0058	0.0000	0.0038
SBM	100	10	4	0.0035	0.0686	0.0041	0.0409	0.0049	0.0411	0.0001	0.0273
	100	10	5	0.0032	0.0802	0.0008	0.0487	0.0007	0.0484	0.0001	0.0375
	300	10	4	0.0004	0.0226	0.0042	0.0155	0.0045	0.0161	0.0001	0.0090
	300	10	5	0.0003	0.0260	0.0010	0.0155	0.0007	0.0160	0.0000	0.0120
	100	30	4	0.0003	0.0215	0.0039	0.0163	0.0046	0.0164	0.0001	0.0090
	100	30	5	0.0002	0.0258	0.0010	0.0168	0.0008	0.0164	0.0000	0.0119
	300	30	4	0.0000	0.0069	0.0039	0.0076	0.0044	0.0084	0.0000	0.0030
	300	30	5	0.0000	0.0081	0.0009	0.0057	0.0007	0.0059	0.0000	0.0041

Table D.15: Rejection frequencies of the distance-based test statistic under DGPs 5-8

(1) Null: H_0^n				DGP 5			DGP 6			DGP 7			DGP 8		
Model	N	T	J	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
ER	100	10	4	0.098	0.081	0.040	0.052	0.037	0.015	0.082	0.049	0.021	0.053	0.042	0.018
	100	10	5	0.106	0.079	0.050	0.069	0.038	0.020	0.088	0.060	0.023	0.081	0.050	0.031
	300	10	4	0.175	0.138	0.075	0.060	0.036	0.018	0.117	0.078	0.038	0.089	0.065	0.039
	300	10	5	0.158	0.113	0.059	0.069	0.043	0.023	0.104	0.083	0.033	0.110	0.082	0.053
	100	30	4	0.182	0.129	0.074	0.067	0.047	0.020	0.130	0.082	0.037	0.093	0.066	0.032
	100	30	5	0.153	0.115	0.068	0.064	0.045	0.022	0.115	0.081	0.046	0.122	0.096	0.051
	300	30	4	0.414	0.333	0.233	0.052	0.032	0.010	0.294	0.229	0.147	0.168	0.123	0.080
	300	30	5	0.371	0.277	0.185	0.062	0.036	0.017	0.278	0.219	0.129	0.246	0.200	0.125
SBM	100	10	4	0.108	0.084	0.050	0.074	0.051	0.027	0.090	0.057	0.033	0.072	0.049	0.022
	100	10	5	0.118	0.088	0.043	0.083	0.056	0.023	0.100	0.072	0.033	0.091	0.071	0.042
	300	10	4	0.155	0.122	0.071	0.060	0.043	0.016	0.094	0.061	0.037	0.073	0.054	0.023
	300	10	5	0.159	0.117	0.066	0.068	0.044	0.022	0.097	0.068	0.043	0.088	0.065	0.031
	100	30	4	0.192	0.144	0.080	0.068	0.039	0.015	0.137	0.090	0.049	0.096	0.073	0.040
	100	30	5	0.164	0.113	0.065	0.072	0.047	0.023	0.140	0.105	0.057	0.130	0.096	0.059
	300	30	4	0.246	0.192	0.116	0.064	0.047	0.022	0.216	0.167	0.093	0.119	0.080	0.050
	300	30	5	0.210	0.168	0.098	0.064	0.047	0.027	0.215	0.162	0.094	0.169	0.131	0.070
(2) Null: H_0^m				DGP 5			DGP 6			DGP 7			DGP 8		
Model	N	T	J	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
ER	100	10	4	0.021	0.009	0.004	0.300	0.238	0.148	0.252	0.205	0.116	0.970	0.957	0.913
	100	10	5	0.028	0.013	0.005	0.329	0.265	0.163	0.305	0.219	0.138	0.942	0.916	0.845
	300	10	4	0.017	0.009	0.004	0.861	0.802	0.703	0.809	0.737	0.621	1	1	1
	300	10	5	0.022	0.016	0.007	0.852	0.792	0.692	0.811	0.747	0.613	1	1	1
	100	30	4	0.015	0.007	0.002	0.834	0.771	0.653	0.803	0.741	0.606	1	1	1
	100	30	5	0.022	0.010	0.002	0.833	0.782	0.658	0.793	0.724	0.616	1	1	1
	300	30	4	0.025	0.012	0.004	0.999	0.999	0.999	1	1	0.999	1	1	1
	300	30	5	0.028	0.016	0.009	0.999	0.999	0.999	1	1	0.997	1	1	1
SBM	100	10	4	0.023	0.011	0.007	0.315	0.254	0.147	0.268	0.215	0.136	0.972	0.956	0.913
	100	10	5	0.026	0.014	0.005	0.330	0.265	0.180	0.313	0.233	0.155	0.943	0.919	0.859
	300	10	4	0.030	0.024	0.008	0.846	0.790	0.677	0.821	0.762	0.615	1	1	1
	300	10	5	0.037	0.023	0.006	0.831	0.780	0.638	0.807	0.754	0.610	1	1	1
	100	30	4	0.012	0.007	0.003	0.812	0.749	0.644	0.788	0.709	0.580	1	1	1
	100	30	5	0.024	0.013	0.004	0.806	0.745	0.630	0.782	0.711	0.563	1	1	1
	300	30	4	0.015	0.008	0.003	1	1	0.999	1	1	0.999	1	1	1
	300	30	5	0.021	0.012	0.008	1	1	0.998	1	1	0.998	1	1	1

Table D.16: Rejection frequencies of the LM test statistic under DGPs 5-8

(1) Null: H_0^n				DGP 5			DGP 6			DGP 7			DGP 8		
Model	N	T	J	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
ER	100	10	4	0.156	0.090	0.025	0.103	0.055	0.012	0.128	0.076	0.014	0.111	0.055	0.012
	100	10	5	0.166	0.082	0.027	0.108	0.058	0.016	0.125	0.071	0.016	0.103	0.063	0.018
	300	10	4	0.256	0.146	0.045	0.101	0.054	0.011	0.202	0.110	0.037	0.152	0.094	0.024
	300	10	5	0.232	0.148	0.035	0.098	0.049	0.009	0.178	0.101	0.025	0.162	0.099	0.027
	100	30	4	0.232	0.144	0.035	0.116	0.052	0.006	0.195	0.124	0.038	0.143	0.089	0.020
	100	30	5	0.213	0.129	0.039	0.104	0.053	0.009	0.187	0.123	0.029	0.167	0.079	0.018
	300	30	4	0.495	0.356	0.171	0.102	0.043	0.008	0.405	0.310	0.127	0.228	0.154	0.044
	300	30	5	0.452	0.324	0.147	0.104	0.046	0.009	0.386	0.267	0.113	0.251	0.158	0.055
SBM	100	10	4	0.153	0.085	0.029	0.106	0.052	0.010	0.158	0.089	0.018	0.109	0.057	0.015
	100	10	5	0.148	0.087	0.020	0.099	0.043	0.008	0.155	0.085	0.020	0.127	0.068	0.015
	300	10	4	0.216	0.132	0.043	0.097	0.049	0.010	0.160	0.096	0.029	0.130	0.074	0.012
	300	10	5	0.198	0.132	0.032	0.096	0.051	0.008	0.163	0.096	0.022	0.136	0.067	0.017
	100	30	4	0.275	0.160	0.055	0.103	0.044	0.007	0.228	0.129	0.034	0.146	0.090	0.022
	100	30	5	0.237	0.143	0.043	0.099	0.050	0.014	0.228	0.129	0.033	0.151	0.084	0.018
	300	30	4	0.319	0.219	0.089	0.119	0.058	0.012	0.302	0.207	0.078	0.175	0.102	0.030
	300	30	5	0.291	0.197	0.065	0.117	0.048	0.012	0.290	0.197	0.067	0.185	0.110	0.030
(2) Null: H_0^m				DGP 5			DGP 6			DGP 7			DGP 8		
Model	N	T	J	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
ER	100	10	4	0.104	0.048	0.009	0.539	0.407	0.189	0.517	0.403	0.194	0.992	0.986	0.956
	100	10	5	0.099	0.054	0.008	0.498	0.374	0.181	0.501	0.382	0.194	0.989	0.981	0.939
	300	10	4	0.087	0.044	0.008	0.957	0.921	0.792	0.948	0.905	0.749	1	1	1
	300	10	5	0.084	0.044	0.006	0.943	0.900	0.760	0.936	0.875	0.736	1	1	1
	100	30	4	0.090	0.043	0.007	0.941	0.894	0.759	0.946	0.902	0.764	1	1	1
	100	30	5	0.082	0.038	0.008	0.925	0.881	0.737	0.931	0.886	0.739	1	1	1
	300	30	4	0.095	0.055	0.013	1	1	0.999	1	1	0.999	1	1	1
	300	30	5	0.099	0.056	0.015	1	0.999	0.999	1	1	0.999	1	1	1
SBM	100	10	4	0.111	0.057	0.014	0.548	0.418	0.212	0.538	0.419	0.214	0.996	0.992	0.947
	100	10	5	0.105	0.061	0.011	0.532	0.402	0.196	0.525	0.409	0.205	0.995	0.980	0.933
	300	10	4	0.113	0.066	0.015	0.943	0.903	0.766	0.951	0.913	0.766	1	1	1
	300	10	5	0.121	0.065	0.011	0.938	0.890	0.724	0.947	0.899	0.739	1	1	1
	100	30	4	0.099	0.048	0.011	0.936	0.893	0.755	0.943	0.901	0.725	1	1	1
	100	30	5	0.094	0.051	0.011	0.927	0.872	0.712	0.931	0.880	0.708	1	1	1
	300	30	4	0.083	0.045	0.010	1	1	1	1	1	1	1	1	1
	300	30	5	0.079	0.041	0.008	1	1	1	1	1	1	1	1	1

We summarize our findings as follows. First, the size of the LM test is reasonably well controlled at all nominal levels, while the $\mathbf{T}$ test is undersized, especially when testing for H_0^m. When the function f is nonlinear, both tests have a good power property for all choices of J and all sample sizes. Second, compared with Table 3, the power of the LM test under DGPs 5–8 is significantly lower, although it can be improved by increasing the sample size. This is expected because DGPs 5–7 involve mildly nonlinear functions f_1 and f_2.

E Robustness Check of the Empirical Analysis

In this section, we check the robustness of the empirical results in Section 6 to the number of basis functions. Specifically, we consider using the third-order polynomial splines with zero or two internal knots to approximate f_1 and f_2. For the former case, we have $J_1 = J_2 = 3$. For the latter case, we have $J_1 = J_2 = 5$.

The values of the standardized LM and $\mathbf{T}$ test statistics for testing $H_0^\dagger$, H_0^n and H_0^m are presented in Table E.17. The results show that the linear model (13) is rejected at the 1% significance level for the LM test, while it cannot be rejected by the $\mathbf{T}$ test when J is chosen as 5. The semiparametric model (8) is rejected by the LM test at the 1% level, but cannot be rejected by the $\mathbf{T}$ test. In contrast, the linearity of the momentum function f_2 is rejected by both tests at any reasonable significance level. These results support the nonlinearity of the momentum effect and are essentially identical to the findings in Section 6.

Table E.17: Standardized $\mathbf{T}$ and LM test statistics for testing the hypotheses $H_0^\dagger$, H_0^n and H_0^m for the Sina Weibo dataset.

	$H_0^\dagger$		H_0^n		H_0^m	
	$\overline{\mathbf{T}}$	$\overline{\mathbf{LM}}$	$\overline{\mathbf{T}}$	$\overline{\mathbf{LM}}$	$\overline{\mathbf{T}}$	$\overline{\mathbf{LM}}$
$J = 3$	14.0255	491.2760	0.3216	6.0479	386.9114	695.1488
$J = 5$	-0.1365	367.4518	-0.6867	6.8323	20.8691	515.4243

The estimation results for f_2 are given in Figure E.4. We observe that the estimated f_2 has a similar shape to that in Figure 1, which is nonlinear and convex. These results imply the same conclusion as in Section 6, that is, a node with a higher activeness level in the past tends to be more active in the future.

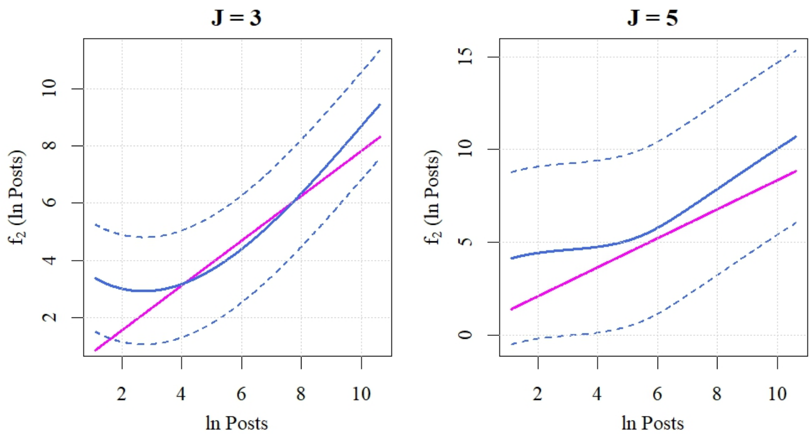

Figure E.4: Estimation of f_2 based on (1) for the Sina Weibo dataset when the number of basis functions $J \in \{3, 5\}$. Solid red and blue lines represent the linear and sieve estimates for f_2, respectively, and dashed blue lines represent the 95% pointwise confidence interval.

Table E.18 shows the estimation results for ρ_1, ρ_2 and γ when $J \in \{3, 5\}$. It is seen that the magnitude and significance of the parameter γ are very similar to those presented in Table 5. All the specifications we consider support the conclusion that male users with more self-created labels tend to be more active. Since we reject the linearity of f_2, the estimates obtained from the semiparametric model (1) or (8) could be more reliable. Compared with Table 5, we see that for the semiparametric model (1) with linear f_1, the network effect ρ_1 still appears to be insignificant when fewer or more basis functions are used.

Table E.18: Parameter estimates for the Sina Weibo dataset with $J \in \{3, 5\}$.

(1) $J = 3$	Semi-NAR (1)		Semi-NAR (2)		Semi-NAR (3)	
	Coef.	t-value	Coef.	t-value	Coef.	t-value
ρ_1	0.0119	0.6984				
ρ_2			0.7823	115.6374		
Number of Labels	0.0145	5.0480	0.0178	5.7822	0.0148	5.1386
Gender	0.0677	2.9996	0.0975	4.0224	0.0684	3.0342
(2) $J = 5$	Semi-NAR (1)		Semi-NAR (2)		Semi-NAR (3)	
	Coef.	t-value	Coef.	t-value	Coef.	t-value
ρ_1	0.0138	0.8137				
ρ_2			0.7804	114.9351		
Number of Labels	0.0139	4.8422	0.0182	5.9144	0.0145	5.0668
Gender	0.0698	3.1013	0.0962	3.9704	0.0692	3.0785

References

Armillotta, M. and Fokianos, K. (2023), 'Count network autoregression', *Journal of Time Series Analysis* .

Bercu, B., Delyon, B. and Rio, E. (2015), *Concentration inequalities for martingales*, Springer International Publishing, Cham, pp. 61–98.

Bollobás, B., Janson, S. and Riordan, O. (2007), 'The phase transition in inhomogeneous random graphs', *Random Structures & Algorithms* **31**(1), 3–122.

Davidson, J. (1994), *Stochastic Limit Theory: An Introduction for Econometricians*, Oxford University Press, New York.

Debaly, Z. M. and Truquet, L. (2021), 'A note on the stability of multivariate non-linear time series with an application to time series of counts', *Statistics & Probability Letters* **179**, 109196.

Freedman, D. A. (1975), 'On tail probabilities for martingales', *The Annals of Probability* **3**(1), 100 – 118.

Hall, P. and Heyde, C. (1980), 3 - the central limit theorem, *in* P. Hall and C. Heyde, eds, 'Martingale Limit Theory and its Application', Probability and Mathematical Statistics: A Series of Monographs and Textbooks, Academic Press, pp. 51–96.

Hoshino, T. (2022), 'Sieve IV estimation of cross-sectional interaction models with nonparametric endogenous effect', *Journal of Econometrics* **229**(2), 263–275.

Jenish, N. and Prucha, I. R. (2012), 'On spatial processes and asymptotic inference under near-epoch dependence', *Journal of Econometrics* **170**(1), 178–190.

Kojevnikov, D., Marmer, V. and Song, K. (2021), 'Limit theorems for network dependent random variables', *Journal of Econometrics* **222**(2), 882–908.

Kuersteiner, G. M. and Prucha, I. R. (2013), 'Limit theory for panel data models with cross sectional dependence and sequential exogeneity', *Journal of Econometrics* **174**(2), 107–126.

Newey, W. K. and Powell, J. L. (2003), 'Instrumental variable estimation of nonparametric models', *Econometrica* **71**(5), 1565–1578.

Qu, X., Lee, L.-F. and Yu, J. (2017), 'QML estimation of spatial dynamic panel data models with endogenous time varying spatial weights matrices', *Journal of Econometrics* **197**(2), 173–201.

Wu, W. B. and Shao, X. (2007), 'A limit theorem for quadratic forms and its applications', *Econometric Theory* **23**(5), 930–951.

Zhu, X., Pan, R., Li, G., Liu, Y. and Wang, H. (2017), 'Network vector autoregression', *The Annals of Statistics* **45**(3).