Nonparametric Network Autoregression

Zixin Yang
Peking University

Xiaojun Song
Peking University

Jihai Yu
Peking University

December 3, 2024

Abstract

This paper is concerned with the nonlinear dynamics of a high-dimensional time series measured on 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 with Sina Weibo, the largest Twitter-type social media in China.

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

JEL Classification: C14, C32, C36

1 Introduction

Network models are key tools for analyzing interconnected systems and are increasingly used in many research areas. Classic applications involve diffusion of spillover effects across large networks (He and Song, 2023), the financial contagion in stock markets (Chen et al., 2019; Härdle et al., 2016; Mitchener and Richardson, 2019) as well as the spread of epidemic diseases (Keeling and Eames, 2005). From an empirical perspective, the analysis of networks calls for developing novel statistical techniques that can account for complex interactions and handle large datasets, possibly 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 (NAR, Zhu et al., 2017), where the network and momentum effects are allowed to be nonlinear and nonparametric.

1.1 Motivation

Consider a large network with N nodes indexed by $i \in \{1, 2, \ldots, N\}$. The network structure is described by a nonstochastic adjacency matrix $A = (a_{ij})$ such that $a_{ij} = 1$ if a direct link exists from i to j and $a_{ij} = 0$ otherwise. Self-relationships are excluded, that is, $a_{ii} = 0$ for $i = 1, \ldots, N$. Let $Y_{it} \in \mathbb{R}$ be the outcome of interest obtained from node i at time point t, and $\mathbb{Y}_t = (Y_{1t}, \ldots, Y_{Nt})^\top \in \mathbb{R}^N$. Modeling the dynamics of $\{\mathbb{Y}_t\}$ over time is an important task that arises commonly in practice. A popular approach is vector autoregression (VAR), where the dynamics of $\{\mathbb{Y}_t\}$ are modeled via a linear operation on its past values, see 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, the VAR-based estimates often neglect the network structures in large-scale time series and cannot directly capture dynamic network effects. To address this issue, several approaches have been proposed to integrate the network structures into the VAR framework, e.g., the global vector autoregressive model by Pesaran et al. (2004), and the NAR model by Zhu et al. (2017). While abundant literature is available, most of the existing work limits attention to time series $\{\mathbb{Y}_t\}$ with linear dependencies among its components. For example, in the NAR models, the outcome of node i, say Y_{it}, is assumed to depend linearly on the past value of itself, $Y_{i,t-1}$, as well as an average of its neighbors collected by $\{j : a_{ij} = 1\}$. Such specification directly implies that the impact of one-point increase in $Y_{i,t-1}$ on Y_{it} is always a constant, and so is the impact of one-point increase in $Y_{j,t-1}$ on Y_{it} for a neighboring node j. In the words of the impulse response analysis, this property means that the dynamic behavior of the nodes can be well represented by random impulses being propagated over time (and across the network) by an invariant linear structure. Such representation simplifies the theoretical analysis but can be restrictive in many cases.

For example, consider a network where a node represents a financial institution, and the outcome of interest is the stock return volatility. Evaluating the risk spillover effect – the impact of shocks to one financial institution on other financial institutions – is important from the policy perspective. Based on the linear NAR model discussed above, many empirical studies have provided evidence of significant volatility spillover effect in stock markets (Chen et al., 2023; Zhu et al., 2023). However, it would be worth reexamining the results as growing studies have suggested that the transmission of volatility across markets exhibits complex nonlinear patterns. For instance, Hoque et al. (2024) examined volatility connectedness among a global financial stress index and US financial sectors and found that spillover effects among the series are higher during extreme volatility than during low and moderate volatility periods. Jin (2015) examined volatility spillover across the Greater China stock markets and showed that only "large" shocks compared to the current level of volatility would increase the expected conditional volatility. These results imply that the spillover effect is not a constant but depends on the level of one's neighbors' outcome. In this situation, using the linear NAR model in Zhu et al. (2017) is inadequate and may lead to inaccurate inferences. To account for nonlinear network and momentum effects, Armillotta and Fokianos (2023) proposed general nonlinear NAR models. Though less restrictive than the linear NAR model, their specifications are still of parametric form, and if they are incorrect, the resulting parameter estimators are not consistent. 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.

1.2 Related works

Nonparametric dynamic models for multivariate time series have been the subject of a wide range of literature. A line of research targets nonparametric modeling of the conditional mean and variance matrix of multivariate time series. The seminal work by Härdle et al. (1998) considered nonlinear vector autoregressive models in which both the conditional mean and the conditional variance matrix are unknown functions of the past. Jeliazkov (2013) modeled conditional mean of a vector time series using a Bayesian hierarchical representation of generalized additive models. Such models fully capture the nonlinear dependence among the series but may suffer from the "curse of dimensionality", especially when the dimension of the vector time series is large. To alleviate this problem, many approaches have been proposed to reduce the model complexity while maintaining easy interpretations. In particular, when the network information about the time series is known, a feasible approach is to integrate the network structure into the model specification. For example, the NAR model in Zhu et al. (2017) has recently been extended to allow for various kinds of nonlinearity in empirical networks, e.g., Li et al. (2024) who considered time-varying NAR models where the network and momentum effects are allowed to change over time and nodes, and Yin et al. (2024) who studied functional-coefficient NAR models. Although some works are available, no methods have been proposed to directly deal with time series with nonparametric network dependence.

1.3 Our contribution

In this paper, we consider a first-order VAR model that utilizes the network structure to characterize the dependence of a high-dimensional time series. The model greatly increases the flexibility and generality of the NAR model proposed by Zhu et al. (2017) by allowing the network and momentum functions to be nonlinear and nonparametric. Following the seminal works such as Newey (1997), Chen and Shen (1998), Huang and Shen (2004), and Chen and Christensen (2015), we estimate the model parameters by a sieve least squares (LS) method. The asymptotic properties of the estimator are investigated under the setting $\max\{N, T\} \to \infty$, that is, either N or T tends to infinity. In particular, we show that the convergence rates of the sieve estimators attain the optimal rate of Stone (1982) under some regularity conditions. 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) methods and have an asymptotically standard normal distribution after proper standardization. We conduct a set of Monte Carlo simulations, which show that our estimator and test perform well in finite samples under various network structures. As an empirical illustration, we apply the proposed method to analyze the posting behavior of Sina Weibo users as in Zhu et al. (2017). The results show that the momentum effect is highly nonlinear with a near piecewise-linear shape, whereas the network effect is insignificant in the semiparametric models.

The rest of the paper is organized as follows. In Section 2, we discuss the stability conditions of the nonparametric NAR model and propose our sieve LS estimator. The asymptotic properties of the proposed estimator are investigated in Section 3. In Section 4, we propose two test statistics for the null hypothesis of linearity and study their asymptotic distributions. Section 5 presents a set of 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 are relegated to the 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 $\|X\|_p$ denotes its L^p-norm: $\|X\|_p = (\mathbb{E}|X|^p)^{1/p}$. For any $m \times n$ matrix A, let $\|A\| = \sqrt{\operatorname{tr}(AA^\top)}$ be the Frobenius norm of A, where $\operatorname{tr}(\cdot)$ is a trace of a matrix. When A is square, we use $\lambda_{\min}(A)$

and $\lambda_{\max}(A)$ to denote the minimum and maximum eigenvalues of A, respectively. Moreover, A^{-} denotes a symmetric generalized inverse of A.

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 an adjacency matrix $A = (a_{ij})$ with $a_{ij} = 1$ if the ith node follows the jth node and 0 otherwise. Also, we assume that A has zero diagonals. 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 nonparametric NAR 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 independent and identically distributed (i.i.d.) random noises. Model (1) allows the outcome of node i to depend on the past value of itself as well as its connections in a nonparametric way. Specifically, 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 outcome of i, while the term $f_2(Y_{i,t-1})$ captures the impact of i's own lagged outcome. The functions f_1 and f_2 will be, respectively, referred to as network effect and momentum effect hereafter. The term $Z_i^{\top}\gamma$ characterizes the nodal impact of the ith node, where $\gamma \in \mathbb{R}^{d_Z}$ is the associated coefficient (nodal effect). For identification purpose, we assume that $f_1(0) = 0$, $f_2(0) = 0$, and the first element of the vector 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.

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 the model is

$$\mathbb{Y}_t = W F_1(\mathbb{Y}_{t-1}) + F_2(\mathbb{Y}_{t-1}) + \mathbb{Z}\gamma + \mathcal{E}_t, \quad t = 1, \ldots, T, \tag{2}$$

where $W = \mathrm{diag}\{n_1^{-1}, \ldots, n_N^{-1}\}A$ 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. *f_1 and f_2 are Lipschitz continuous on $\mathbb{R}$ with Lipschitz constants κ_1 and κ_2, respectively: 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. *Suppose that Assumptions 1-2 hold. The network size N is either fixed or diverges to infinity. Then there exists a unique strictly stationary solution $\{\mathbb{Y}_t \in \mathbb{R}^N : t \in \mathbb{Z}\}$ to the NAR 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]

Proposition 1 extends the fixed and increasing network-type results in Zhu et al. (2017), Theorems 1-2, 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 a near-epoch dependent (NED) process (see Definition 1 in the Appendix A).

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 $\{p_j(y) : j = 1, 2, \ldots\}$ be a sequence of basis functions on $\mathbb{R}$, and $p^J(y) = (p_1(y), \ldots, p_J(y))^\top$ for some $J = J_{NT}$. Further, 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

[1] When $N \to \infty$, $\{\mathbb{Y}_t\}$ is a vector time series of increasing dimension. In this case, the stationarity of $\{\mathbb{Y}_t\}$ is defined according to Definition 1 in Zhu et al. (2017). Moreover, the notation $\max_{i \ge 1} x_i$ means $\max_{1 \le i < \infty} x_i$.

$\beta_2^\top p^{J_2}(\cdot) \in \mathcal{F}_{J_2}$, respectively. (A more precise statement is given later.) Then, model (1) can be rewritten as

$$Y_{it} = \beta_1^\top \bar{P}_{i,t-1}^{J_1} + \beta_2^\top P_{i,t-1}^{J_2} + Z_i^\top \gamma + u_{it}, \tag{3}$$

where $\bar{P}_{i,t-1}^{J_1} = n_i^{-1} \sum_{j=1}^N a_{ij} p^{J_1}(Y_{j,t-1})$, $P_{i,t-1}^{J_2} = p^{J_2}(Y_{i,t-1})$, 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 $\mathbb{P}_{t-1}^{J_\ell} = (p^{J_\ell}(Y_{1,t-1}), \ldots, p^{J_\ell}(Y_{N,t-1}))^\top$ is an $N \times J_\ell$ matrix of transformed lagged responses 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

$$\widehat{\theta} = \left(\sum_{t=1}^T \mathbb{X}_{t-1}^\top \mathbb{X}_{t-1} \right)^{-1} \sum_{t=1}^T \mathbb{X}_{t-1}^\top \mathbb{Y}_t. \tag{4}$$

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

$$\begin{aligned} \widehat{\gamma} &= \left(\mathbf{Z}^\top \mathscr{M}(\mathbf{Q})\mathbf{Z}\right)^{-} \mathbf{Z}^\top \mathscr{M}(\mathbf{Q})\mathbf{Y}, \\ \widehat{\beta} &= \left(\mathbf{Q}^\top \mathscr{M}(\mathbf{Z})\mathbf{Q}\right)^{-} \mathbf{Q}^\top \mathscr{M}(\mathbf{Z})\mathbf{Y}, \end{aligned} \tag{5}$$

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 simply given by $\widehat{f}_\ell(\cdot) = \widehat{\beta}_\ell^\top p^{J_\ell}(\cdot)$ for $\ell = 1, 2$.

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 3 and Section C of the Supplementary Appendix.

Assumption 3. *(i)* $\max_{i \geq 1} \mathbb{E}|Y_{it}|^{2\omega} < \infty$ *for some* $\omega \geq 1$. *(ii) The covariates* $\{Z_i : 1 \leq i \leq N\}$ *are independent random variables, and* $\mathcal{R}_Z$ *is a compact subset of* $\mathbb{R}^{d_Z}$. *Moreover,* $\{Z_i\}$ *and* $\{\epsilon_{it}\}$ *are mutually independent. Let* $\Psi_{N,Z} = \mathbb{E}(\mathbb{Z}^\top \mathbb{Z}/N)$. *The matrix* $\Psi_Z = \lim_{N\to\infty} \Psi_{N,Z}$ *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 bounded functions 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). We assume independence among $\{Z_i\}$ for analytical simplicity. This assumption can be moderately relaxed so that different Z_i's are weakly dependent, as long as the law of large numbers and central limit theorem hold.

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 $\Sigma_{N,J_1J_2}=\mathbb{E}(\mathbb{X}_{t-1}^\top\mathbb{X}_{t-1}/N)$, $\bar{\nu}_{N,J_1J_2}=\lambda_{max}(\Sigma_{N,J_1J_2})$, and $\underline{\nu}_{N,J_1J_2}=\lambda_{min}(\Sigma_{N,J_1J_2})$. For each J_1, J_2 and N, there exists a constant $0<\bar{c}<\infty$ such that $\bar{\nu}_{N,J_1J_2}\leq\bar{c}$ and $\underline{\nu}_{N,J_1J_2}>0$.*

Assumption 4(i) imposes a moment condition on the variables $p_j(Y_{it})$'s and can be trivially satisfied by uniformly bounded basis such as B-spline basis and Fourier series. For Assumption 4(ii), the quantities $\bar{\nu}_{N,J_1J_2}$ and $\underline{\nu}_{N,J_1J_2}$ generally depend on the distribution of the data, network structure and choice of basis. We allow that $\underline{\nu}_{N,J_1J_2}\to 0$ as $N\to\infty$. Such a setting incorporates the case where the degree of collinearity between $W\mathbb{P}_{t-1}^{J_1}$ and $\mathbb{P}_{t-1}^{J_2}$ slowly increases as N becomes larger.

To state the next assumption, we introduce some relevant quantities for the limit theorems under network-time dependence. Let $D_N=\{1,2,\ldots,N\}$ be the set of cross-sectional unit indices. Define an undirected network G_N on D_N, where $G_N=(D_N,E_N)$, and E_N denotes a set of links such that $\{i,j\}\in E_N$ if and only if $a_{ij}=1$ or $a_{ji}=1$. For $i,j\in D_N$, define $d_N(i,j)$ to be the distance between i and j in G_N, i.e., the length of the shortest path between nodes i and j given G_N. Let $\mathcal{N}_N^\partial(i;s)$ denote the set of nodes exactly at the distance s from node i, that is,

$$\mathcal{N}_N^\partial(i;s)=\{j\in D_N:d_N(i,j)=s\}$$

and $\delta_N^\partial(s)=N^{-1}\sum_{i=1}^N|\mathcal{N}_N^\partial(i;s)|$ denotes the average s-neighborhood size of G_N.

The quantity related to the limit theorems under network-time dependence is given by $\Xi_{N,d}=N^{-1}\sum_{s=1}^\infty s^d\delta_N^\partial(s)(\kappa_1+\kappa_2)^{\lfloor s/3\rfloor}$, where $d\in\mathbb{N}$, κ_1 and κ_2 are defined in Assumption 2 and $\lfloor a\rfloor$ denotes the greatest integer less than or equal to a. Clearly, $\Xi_{N,d}$ measures the relative strength between the decay rate of dependence across a network and the network's denseness (Kojevnikov et al., 2021). Note that $\delta_N^\partial(s)\leq N$. Then, for arbitrary network structure and $d\in\mathbb{N}$, we have $\Xi_{N,d}\leq\sum_{s=1}^\infty s^d(\kappa_1+\kappa_2)^{\lfloor s/3\rfloor}<\infty$. Particularly, $\Xi_{N,d}$ will tend to zero as $N\to\infty$ when the s-neighborhood size grows at a sufficiently slow rate with s. For example, for spatial networks such as latent space (Hoff et al., 2002) and RGG models (Penrose, 2003), $\delta_N(s)$ grows polynomially with s, implying that $\Xi_{N,d}=O(N^{-1})$ for these network structures (Leung, 2022).

As in Li and Liao (2020), we introduce the sequence $\zeta_{0,J}=\sup_{y\in\mathcal{R}_\mathcal{Y}}\|p^J(y)\|$ and $\zeta_{1,J}=\sup_{y,y'\in\mathcal{R}_\mathcal{Y}}\|p^J(y)-p^J(y')\|/|y-y'|$. It will be assumed throughout that $\zeta_{0,J}\geq 1$ and $\zeta_{1,J}\geq 1$ for large enough J. If $\mathcal{R}_\mathcal{Y}$ is compact, we have $\zeta_{0,J}\lesssim\sqrt{J}$, $\zeta_{1,J}\lesssim J^{3/2}$ for univariate polynomial spline and trigonometric polynomial; and $\zeta_{0,J}\lesssim J$, $\zeta_{1,J}\lesssim J^2$ for power series or orthogonal polynomial bases (Newey, 1997).

Assumption 5. *As $N \to \infty$, T is either fixed or tends to ∞, and $\underline{\nu}_{N,J_1J_2}^{-4} r_{NT,J} \to 0$, 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 specifies the growth order of the smoothing parameter J, the s-neighborhood sizes $\delta_N^\partial(s)$, and the degree of singularity $\underline{\nu}_{N,J_1J_2}^{-1}$. This condition is mainly used in controlling the convergence in probability of the matrix $\widehat{\Sigma}_{NT,J_1J_2} = (NT)^{-1}\sum_{t=1}^T \mathbb{X}_{t-1}^\top \mathbb{X}_{t-1}$ to its expectation Σ_{N,J_1J_2} in the Euclidean norm. When T is fixed, the condition reduces to (**a**) $\zeta_{0,J}^2 J/(\underline{\nu}_{N,J_1J_2}^4 N) \to 0$ and (**b**) $\zeta_{0,J}^2\zeta_{1,J}J^{1/2}\Xi_{N,0}/\underline{\nu}_{N,J_1J_2}^4 \to 0$. For condition (**b**) to hold, the component $\Xi_{N,0}$ must shrink to zero as N goes to infinity, implying that "sparse" networks are necessary to build limit theorems under fixed T setting. Further, the "sparser" the networks (the faster $\Xi_{N,0}$ converges to zero), the larger J is allowed. When T also diverges, the restriction on the sparsity of the network can be relaxed. Indeed, condition (**b**) is now replaced by $\zeta_{0,J}^2\zeta_{1,J}J^{1/2}\min\{\Xi_{N,0}, T^{-1}\Xi_{N,1}\}/\underline{\nu}_{N,J_1J_2}^4 \to 0$. A sufficient condition is $\zeta_{0,J}^2\zeta_{1,J}J^{1/2}/(\underline{\nu}_{N,J_1J_2}^4 T) \to 0$, a condition unrelated to network G_N. Lastly, the asymptotic setting with $T \to \infty$ and fixed N is also allowed. Since the conditions needed differ slightly from the setting, we defer the related discussion to Remark 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 $\omega \geq 0$, see, e.g., Chen et al. (2005), Su and Jin (2012) and Hoshino (2022). As stated 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 support is unbounded.

Assumption 6. *There exist sequences of vectors $\beta_1 = \beta_{1J_1}$, $\beta_2 = \beta_{2J_2}$ 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 the functions f_1 and f_2 by the sieve space in terms of the weighted sup-norm. 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 Fourier series, see Chen (2007).

Denote $\Psi_N = \mathbb{E}(\mathbb{Q}_t^\top\mathbb{Q}_t/N)$, $C_N = \mathbb{E}(\mathbb{Z}^\top\mathbb{Q}_t/N)$ and $\Sigma_{N,Z} = \Psi_{N,Z} - C_{N,ZJ}\Psi_{N,J}^{-1}C_{N,ZJ}^\top$.

Assumption 7. *The matrix $\Sigma_Z = \lim_{N\to\infty}\Sigma_{N,Z}$ exists and is positive definite.*

The following theorem provides the convergence rate and asymptotic distribution of $\widehat{\gamma}$.

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

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

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

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 combination 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 $[(NT)^{3\kappa} \min\{\Xi_{N,0}, T^{-1}\Xi_{N,1}\} + (NT)^{2\kappa-1}]/\underline{\nu}_{N,J_1J_2}^4 = o(1)$ simultaneously, for polynomial spline and trigonometric polynomial. In particular, if $\min\{\Xi_{N,0}, T^{-1}\Xi_{N,1}\} = O((NT)^{-1})$ and the degree of singularity is not severe so that $1/\sqrt{\underline{\nu}_{N,J_1J_2}} = J^r$ for some $r > 0$, these conditions can be reduced to $1/(2\mu) < \kappa < 1/(8r+3)$.

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

$$\Sigma_{N,f} = \Psi_{N,J} - C_{N,ZJ}^{\top} \Psi_{N,Z}^{-1} C_{N,ZJ}. \tag{6}$$

Further, 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 $(N^{-1} + r_{NT,J})J/(J_\ell \underline{\nu}_{N,J_1J_2}^3) = O(1)$ and $\max_{i\geq 1} \lambda_{max}(\mathbb{E}[p^{J_\ell}(Y_{it}) p^{J_\ell}(Y_{it})^{\top}]) < \infty$, then

$$\max_{i\geq 1} \int_{\mathcal{R}_{\mathcal{Y}}} [\widehat{f}_\ell(y) - f_\ell(y)]^2 dF_i(y) = O_p(\underline{\nu}_{N,J_1J_2}^{-1}(J_\ell/(NT) + J_1^{-2\mu} + J_2^{-2\mu})),$$

where $F_i(y)$ is the cumulative distribution function of Y_{it}.

(ii) If $\zeta_{1,J} \lesssim J^{\xi}$ for some $\xi > 0$, $\zeta_{0,J}^4 \ln J_\ell/(\underline{\nu}_{N,J_1J_2}^2 NT) = o(1)$, and $(N^{-1} + r_{NT,J})J/(\underline{\nu}_{N,J_1J_2}^3 \ln J_\ell) = O(1)$, then

$$\left|\widehat{f}_\ell(\cdot) - f_\ell(\cdot)\right|_{\infty,\omega} = O_p\left(\zeta_{0,J_\ell}\sqrt{\ln J_\ell/(\underline{\nu}_{N,J_1J_2} NT)} + \zeta_{0,J_\ell} \underline{\nu}_{N,J_1J_2}^{-1/2}(J_1^{-\mu} + J_2^{-\mu})\right).$$

(iii) If $\underline{\nu}_{N,J_1J_2}^{-4} J(N^{-1} + r_{NT,J}) = o(1)$ and $\sqrt{NT}\underline{\nu}_{N,J_1J_2}^{-1/2}(J_1^{-\mu} + J_2^{-\mu}) = o(1)$, then for a given finite $y \in \mathcal{R}_{\mathcal{Y}}$ such that $\|p^{J_\ell}(y)\| > 0$, we have $\frac{\sqrt{NT}}{v_{\ell N}(y)}(\widehat{f}_\ell(y) - f_\ell(y)) \xrightarrow{d} N(0,1)$ 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 $\nu_{N,J_1J_2}^{-1} \lesssim 1$. 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 that $\max_{i\geq 1} \mathbb{E}|\widehat{f}_\ell(Y_{it}) - f_\ell(Y_{it})|^2 = O_p((NT)^{-2\mu/(1+2\mu)})$. 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, $|\widehat{f}_\ell(\cdot) - f_\ell(\cdot)|_{\infty,0} = O_p(\zeta_{0,J}\sqrt{\ln J/(\underline{\nu}_{N,J_1J_2}NT)}) + O_p(J^{-\mu})$, see Lemma 2.4 in Chen and Christensen (2015). While the variance term of our estimator attains the optimal rate, the bias term appears to be $O(\zeta_{0,J}\underline{\nu}_{N,J_1J_2}^{-1/2})$ times bigger. Theorem 2(iii) states a pointwise asymptotic normality property of the sieve LS estimator. Notice that $[\lambda_{\max}(\Sigma_{N,f})]^{-1}\|p^{J_\ell}(y)\|^2 \le p^{J_\ell}(y)^\top \mathcal{S}_\ell \Sigma_{N,f}^{-1}\mathcal{S}_\ell^\top p^{J_\ell}(y) \le [\lambda_{\min}(\Sigma_{N,f})]^{-1}\|p^{J_\ell}(y)\|^2$. This and Assumption 4 imply that the convergence rate of $\widehat{f}_\ell(y)$ is $O(\sqrt{NT/(J_\ell \underline{\nu}_{N,J_1J_2})})$ as $\|p^{J_\ell}(y)\| = O(J_\ell^{1/2})$ and $\lambda_{\min}^{-1}(\Sigma_{N,f}) = O(\underline{\nu}_{N,J_1J_2}^{-1})$.

Remark 1. *(Consistent Estimation of Covariance Matrices). For statistical 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 estimate 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 $\widehat{\Sigma}_{NT,f} = \widehat{\Psi}_{NT,J} - \widehat{C}_{NT,ZJ}^\top \widehat{\Psi}_{N,Z}^{-1}\widehat{C}_{NT,ZJ}$ with $\widehat{\Psi}_{N,Z} = \mathbb{Z}^\top\mathbb{Z}/N$, $\widehat{\Psi}_{NT,J} = \sum_{t=1}^T \mathbb{Q}_{t-1}^\top\mathbb{Q}_{t-1}/(NT)$ and $\widehat{C}_{NT,ZJ} = \sum_{t=1}^T \mathbb{Z}^\top\mathbb{Q}_{t-1}/(NT)$. The convergence in probability of such sample second moment matrices to their population counterparts is shown in Lemma 3 in the Appendix A. 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 Section C of the Supplementary Appendix that $\widehat{\sigma}^2$ is a consistent estimator of σ^2 under Assumptions 2-6. As such, Theorem 2 still holds with $v_{1N}^2(y)$ replaced by $\widehat{v}_{1N}^2(y) = \widehat{\sigma}^2 p^{J_\ell}(y)^\top\widehat{\Sigma}_{f_1}^{-1}p^{J_\ell}(y)$.*

Remark 2. *The above results can be extended to the estimation of various functionals of $f_\ell(\cdot)$, such as the derivatives, average derivatives, or weighted derivatives 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 $w(y)$. 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, we have*

$$\phi(\widehat{f}_\ell) = \phi(p^{J_\ell}(\cdot)^\top\widehat{\beta}_\ell) = \varphi^{J_\ell\top}\widehat{\beta}_\ell,$$

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 jth 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 $|\phi(f_\ell)| \le C|f_\ell|_s$ 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 assumption (**A**) is trivially satisfied for both cases we investigate. Assumption (**B**) plays the role of undersmoothing. Then, under assumptions (**A**), (**B**) and* *the conditions of Theorem 2(iii), we have*

$$\sqrt{NT}\tilde{v}_{\ell N}^{-1}[\phi(\widehat{f}_\ell) - \phi(f_\ell)] \xrightarrow{d} N(0,1),$$

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} *in this definition by* $\widehat{\Sigma}_{NT,f_\ell}$ *given in Remark 1.*

Remark 3. *(Asymptotics with Large* T *and Fixed* N*) The asymptotic properties in Theorems 1-2 can also be established under the setting where* $T \to \infty$ *and* N *is fixed. In this case, we modify Assumption 5 as follows.*

Assumption 8. *As* $T \to \infty$*,* N *is fixed, and* $\underline{\nu}_{N,J_1J_2}^{-4}\zeta_{0,J}^2(J + \zeta_{1,J}J^{1/2})/T \to 0$*.*

This condition does not impose any restrictions on the network structure. Then, under Assumptions 1-4, 6-8, we have Theorems 1-2 hold conditional on $\{Z_i : 1 \le i \le N\}$*, see Section C.2 in the Supplementary Appendix for details.*

4 Specification tests

In this section, we consider testing the functional form of the network and momentum effects given by f_1 and f_2, respectively. In particular, we focus on testing whether the widely used linear specification is appropriate.

4.1 Testing the linearity of f_1 or f_2

We first consider individually testing the following null hypotheses over the domain of the variable $Y_{i,t-1}$:

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

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. Note that $f_\ell(0) = 0$ for $\ell = 1, 2$, so that we do not include an intercept term in the expression of f_ℓ in (7). Since the tests for the two hypotheses are similar, we focus on the test of H_0^n.

We first propose a L_2-distance-based test for H_0^n. The basic idea is to compare the restricted estimate under H_0^n and the unrestricted estimate under H_1^n in the same 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 given by

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

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 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, (10) is asymptotically normally distributed 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. To avoid this problem, we propose an LM test for H_0^n, which only utilizes the estimates under the null. 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 used to approximate the nonlinear network effect, and $\beta_{1,2}$ is the corresponding sieve coefficients of length $J_1 - 1$. Then the alternative model (2) could be written as

$$\mathbb{Y}_t = \beta_{1,1} W\mathbb{Y}_{t-1} + W\tilde{\mathbb{P}}_{t-1}^{J_1}\beta_{1,2} + \mathbb{P}_{t-1}^{J_2}\beta_2 + \mathbb{Z}\gamma + U_t, \ \ t = 1, \ldots, T.$$

If H_0^n is correct, the additional 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. In this connection, we propose a new test for H_0^n based on LM principles as in Gupta (2018). 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 constrained sieve LS estimator of θ_1 is given by $\bar{\theta}_1 = (\sum_{t=1}^T \mathbb{X}_{t-1,1}^\top \mathbb{X}_{t-1,1})^{-1} \sum_{t=1}^T \mathbb{X}_{t-1,1}^\top \mathbb{Y}_t$. The LM test measures how close to zero the gradient of the sieve LS criterion is when evaluated at the constrained estimator. The test statistic is defined as

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

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 $\mathcal{N}(0,1)$ under H_0^n.

To formally establish the asymptotic null distribution of the test statistics, 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* $0 < \underline{c}_p < \infty$ *such that* $\lambda_{min}\big(\mathbb{E}(\mathbb{P}_t^{J_1\top}\mathbb{P}_t^{J_1}/N)\big) > \underline{c}_p$ *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)* $J^{3/2}(r_{NT,J}+N^{-1})/\underline{\nu}^5_{N,J_1J_2} = o(1)$*, (ii)* $J^{1/2}\zeta_{0,J}^{5/3}(N^{-1/3}+r_{NT,J}^{1/3})/\underline{\nu}^4_{N,J_1J_2}+J^{3/2}(N^{-1/4}+r_{NT,J}^{1/4})/\underline{\nu}^4_{N,J_1J_2} = o(J_1)$*, (iii)* $(NT)^{1/2}(\zeta_{0,J_1}J_1^{-\mu}+J_2^{-\mu})/\underline{\nu}_{N,J_1J_2} = o(1)$*, and (iv)* $\underline{\nu}_{N,J_1J_2}J_1^{1/2} \to \infty$ *as* $J_1 \to \infty$.

Assumption 9. *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. 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 9 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 $\Phi_{N,J_1} = \mathbb{E}(\mathbb{P}_t^{J_1\top}\mathbb{P}_t^{J_1}/N)$. As shown below, $\mathbf{B}_{N,1}$ and $\mathbf{s}_{N,1}^2$ serve as the mean and variance of the test statistic (10), respectively. The following theorem establishes the limiting null distribution of 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 9 hold (where 1, 4 and 5 are replaced by 1′, 4′ and 5′, respectively). Under* H_0^n*,* $\overline{\mathbf{T}}_{NT,1} \xrightarrow{d} \mathcal{N}(0,1)$ *and* $\overline{\mathbf{LM}}_{NT,1} \xrightarrow{d} \mathcal{N}(0,1)$.

If H_0^n is not true, the standardized 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 $\mathcal{N}(0,1)$. To implement the test in practice, 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 1. 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 $\widehat{\mathbf{s}}^2_{NT,1}/\mathbf{s}^2_{N,1} \xrightarrow{p} 1$ under Assumptions 2-4 and 5′.

Next, we study the local power properties of $\overline{\mathbf{T}}_{NT,1}$ and $\overline{\mathbf{L}}\mathbf{M}_{NT,1}$. 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 $r(\cdot)$ is a measurable nonlinear function on $\mathcal{R}_{\mathcal{Y}}$, and $\alpha_{NT} \to 0$ as $N \to \infty$ is a scalar that specifies the speed at which the local alternatives converge to the null. We impose the following conditions on $r(\cdot)$.

Assumption 10. *The function $r(\cdot)$ is Lipschitz continuous on $\mathcal{R}_{\mathcal{Y}}$ with Lipschitz constant $\bar{c}_r$, and $\max_i \mathbb{E}|r(Y_{it})|^4 < \infty$.*

To study the asymptotic 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 as given in (9). Denote $\widetilde{\mathbb{X}}_t = (\mathbb{X}_{t,1}, \mathbb{X}_{t,2})$, $\Psi_N = \mathbb{E}(\widetilde{\mathbb{X}}_t^\top \widetilde{\mathbb{X}}_t / N)$ and $\Gamma_N = \mathbb{E}(\widetilde{\mathbb{X}}_t^\top W R(\mathbb{Y}_t)/N)$, where $R(\mathbb{Y}_t) = (r(Y_{1t}), \ldots, r(Y_{Nt}))^\top$. Further, let $\Psi_{N,11} = \mathbb{E}(\mathbb{X}_{t,1}^\top \mathbb{X}_{t,1}/N)$ and $\Gamma_{N,1} = \mathbb{E}(\mathbb{X}_{t,1}^\top W R(\mathbb{Y}_t)/N)$. The "non-centrality parameters" of the $\overline{\mathbf{T}}_{NT,1}$ and $\overline{\mathbf{L}}\mathbf{M}_{NT,1}$ tests under the sequence of local alternatives given by (12) take the form

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

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 $\mathcal{S}_3 = (\mathbf{0}_{(J_1-1)\times(J_2+d_Z+1)}, I_{J_1-1})$ is a selection matrix.

Theorem 4. *Suppose that Assumptions 1-6 and 10 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. Then,*

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

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

Theorem 4 states that our test statistics $\mathbf{T}_{NT,1}$ 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 of the facts that $\mathbf{s}^2_{N,1} = O(J_1/\underline{\nu}^2_{N,J_1J_2})$ and $J_1 \to \infty$ as $N \to \infty$, both rates are slower than the parametric rate $(NT)^{-1/2}$. Such rates have been found also by de Jong and Bierens (1994) and Gupta (2018).

Remark 4. *The above results can be modified to test the existence of network or momentum effects. In this case, the null hypothesis is* $\tilde{H}_0^n : f_1(Y_{i,t-1}) = 0$ *a.s., and the alternative is* $\tilde{H}_1^n : \mathbb{P}(f_1(Y_{i,t-1}) \neq 0) > 0$*. To construct a Wald-type test for* $\tilde{H}_0^n$*, one can directly compare the estimated function* $\widehat{f}_1$ *with zero. A feasible test statistic is given by* $\tilde{\mathbf{T}}_{NT,1} = \sum_{i=1}^N \sum_{t=1}^T (\widehat{f}_1(Y_{i,t-1}))^2$*. For the LM test, note that the problem can be transformed to testing the approximate null* $\tilde{H}_0^n : \beta_1 = 0$*. Denote* $\tilde{\mathbb{X}}_{t-1,1} = (\mathbb{P}_{t-1}^{J_2}, \mathbb{Z})$ *and* $\tilde{\mathbb{X}}_{t-1,2} = W\mathbb{P}_{t-1}^{J_1}$*, and* $\tilde{\theta}_1 = (\sum_{t=1}^T \tilde{\mathbb{X}}_{t-1,1}^\top \tilde{\mathbb{X}}_{t-1,1})^{-1} \sum_{t=1}^T \tilde{\mathbb{X}}_{t-1,1}^\top \mathbb{Y}_t$*. The corresponding LM test statistic is defined as in (11) with* $\mathbb{X}_{t-1,1}$*,* $\mathbb{X}_{t-1,2}$ *and* $\bar{\theta}_1$ *replaced by* $\tilde{\mathbb{X}}_{t-1,1}$*,* $\tilde{\mathbb{X}}_{t-1,2}$ *and* $\tilde{\theta}_1$*, respectively. The asymptotic results of the tests can be derived similarly to those above. Compared with the test of* $\rho_1 = 0$ *under the parametric model (13), such tests do not assume the correct specification of linear network effect and thus can detect more general forms of network dependence.*

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. It is also interesting to directly test the validity of the linear NAR model in Zhu et al. (2017). 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 Bonferroni correction. That is, one considers individually 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 as (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

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

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

Assumption 11. *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 $\mathbf{s}_N^2 = 2\sigma^4 \text{tr}\big\{\Phi_{N,J}\Sigma_{N,f}^{-1}\Phi_{N,J}\Sigma_{N,f}^{-1}\big\}$, 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 $(J_1 + J_2) \times (J_1 + J_2)$ matrix and $\Sigma_{N,f}$ is defined in (6). 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′, 2, 3, 4†, 5†, 6 and 11 hold. Under* $H_0^\dagger$, $\overline{\mathbf{T}}_{NT} \xrightarrow{d} \mathcal{N}(0,1)$ *and* $\overline{\mathbf{LM}}_{NT} \xrightarrow{d} \mathcal{N}(0,1)$.

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 $\mathbf{s}_N$ and $\mathbf{B}_N$ 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 1 and $\Phi_{N,J} = \text{Diag}(\widehat{\Phi}_{NT,J_1}, \widehat{\Phi}_{NT,J_2})$. The consistency of these estimators can be easily seen from Lemmas 2-3 in Appendix A.

5 Simulation

5.1 Sieve LS estimation

We now examine the finite sample performance of the sieve LS estimator by 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

$$\begin{aligned}
&\text{DGP 1: } f_1(y) = 0.5y, \;\; f_2(y) = 0.4y,\\
&\text{DGP 2: } f_1(y) = \cos(0.3y), \;\; f_2(y) = 0.5y,\\
&\text{DGP 3: } f_1(y) = 0.5y, \;\; f_2(y) = \cos(0.3y),\\
&\text{DGP 4: } f_1(y) = \cos(0.4y), \;\; f_2(y) = \cos(0.3y)
\end{aligned}$$

with $\gamma_1 = \gamma_2 = 0.5$, $Z_i = (Z_{i1}, Z_{i2})^\top$ 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 dependent variable $\mathbb{Y}_0$ by contraction mapping iterations. Specifically, let $\mathbb{Y}_0^{(0)}$ be an initial candidate value of $\mathbb{Y}_0$ and 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 $\|\mathbb{Y}_0^{(l+1)} - \mathbb{Y}_0^{(l)}\| < 10^{-4}$ is met. Given $\mathbb{Y}_0$, the observations $\mathbb{Y}_t$'s for $t = 1, \ldots, T$ are generated according to model (1).

We consider the following combinations of sample sizes: $N \in \{100, 200\}$ and $T \in \{10, 20, 100, 200\}$. Following Zhu et al. (2017) and Armillotta and Fokianos (2023), we generate the adjacency matrix $A_N = (a_{ij})$ using the following network models.

(1) Erdos–Rényi (ER) Model. This model assumes that the network is constructed by connecting N nodes randomly. 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). Randomly assign for each node a block label ($k = 1, \ldots, K$) with equal probability, where $K = 5$ is the total number of blocks. 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.

In the simulations, for the ER model, the observed network density (i.e., $\sum_{i,j} a_{ij}/\{N(N-1)\}$) is 21.75% and 10.70% for $N = 100$ and 200, respectively. For the SBM with $K = 5$, the observed density is 1.86% and 1.31% for $N = 100$ and 200, respectively.

As for the choice of the basis function p, we use the third-order polynomial splines.[2] 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 settings of J for each T: $J \in \{4, 5\}$ for $T \in \{10, 20\}$ and $J \in \{6, 7\}$ for $T \in \{100, 200\}$. The intercept term is excluded from the basis. The number of Monte Carlo repetitions for each set-up is set to $M = 1000$.

As in Hoshino (2022), the performance of the sieve estimator $\widehat{f}_1$ and $\widehat{f}_2$ is evaluated by the integrated squared bias (ISB) and integrated mean squared error (IMSE):

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

for $\ell = 1, 2$, where $\widehat{f}_\ell^{(r)}(y)$ is the estimate of f_ℓ obtained from the rth replicated data, $\bar{f}_\ell(y) = f_\ell(y) - f_\ell(0)$, and $y_{2.5}$ and $y_{97.5}$ represent the 2.5% and 97.5% empirical quantiles of $\{Y_{i,t-1}\}$ averaged over all repetitions, respectively.

Table 1 presents the performance of the sieve LS 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 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 J. The above pattern is robust to the choice of network structure.

Table 2 summarizes the simulation results for f_1 and f_2. The ISB and IMSE of $\widehat{f}_1$ and $\widehat{f}_2$ are quite small for all DGPs and decrease significantly as N or T gets larger. 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 cases with $J = 4$. These results illustrate the bias-variance trade-off of our estimator concerning 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 in terms of IMSE. A similar pattern can be found for the estimation of f_2. Finally, for all DGPs we consider, the IMSE of f_1 is larger than that of f_2, especially for the ER model where the network density is relatively large. This implies that recovering the nonlinearity in the network effect is generally difficult.

[2] The polynomial splines of order q is given by $\text{Pol}(J) = \{\sum_{j=0}^{q} \pi_j y^j + \sum_{l=1}^{r} \pi_{q+l}(y - \xi_l)_+^q : \pi_j \in \mathbb{R}, j = 0, \ldots, q+r\}$ for $J = q+r+1$, where $(y-\xi)_+ = \max\{y-\xi, 0\}$ and $\xi_1, \ldots, \xi_r$ are knots. In the simulation, given the value of knots number r, the location of the knots $\xi_1, \ldots, \xi_r$ are determined by the empirical quantiles, this is, $\xi_l =$ the $l/(r+1)$th quantile of $\{Y_{i,t-1} : 1 \leq i \leq N, 1 \leq t \leq T\}$. See Chen (2007) for more details.

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	0.0025	0.0020	0.940	0.0030	0.0021	0.944	-0.0008	0.0016	0.948	0.0012	0.0016	0.954
	100	10	5	0.0025	0.0020	0.945	0.0030	0.0021	0.945	-0.0008	0.0016	0.948	0.0011	0.0016	0.956
	200	10	4	0.0014	0.0010	0.942	0.0003	0.0010	0.946	0.0010	0.0008	0.954	0.0009	0.0007	0.950
	200	10	5	0.0014	0.0010	0.942	0.0004	0.0010	0.945	0.0010	0.0008	0.955	0.0009	0.0007	0.949
	100	20	4	0.0020	0.0010	0.954	0.0047	0.0012	0.934	0.0005	0.0008	0.957	0.0010	0.0008	0.957
	100	20	5	0.0020	0.0010	0.953	0.0047	0.0012	0.932	0.0005	0.0008	0.956	0.0010	0.0008	0.956
	200	20	4	0.0025	0.0005	0.944	0.0012	0.0005	0.964	0.0003	0.0004	0.955	0.0001	0.0004	0.953
	200	20	5	0.0025	0.0005	0.943	0.0012	0.0005	0.964	0.0003	0.0004	0.955	0.0001	0.0004	0.953
SBM	100	10	4	0.0058	0.0018	0.948	0.0074	0.0021	0.933	0.0002	0.0015	0.957	0.0028	0.0016	0.942
	100	10	5	0.0060	0.0019	0.948	0.0076	0.0022	0.933	0.0001	0.0015	0.955	0.0027	0.0016	0.942
	200	10	4	0.0017	0.0009	0.956	0.0050	0.0010	0.943	-0.0008	0.0008	0.953	0.0004	0.0009	0.932
	200	10	5	0.0019	0.0009	0.956	0.0051	0.0011	0.944	-0.0008	0.0008	0.953	0.0003	0.0009	0.934
	100	20	4	0.0024	0.0009	0.948	0.0033	0.0010	0.954	0.0013	0.0008	0.935	0.0006	0.0008	0.958
	100	20	5	0.0025	0.0009	0.947	0.0034	0.0010	0.952	0.0012	0.0008	0.936	0.0006	0.0008	0.956
	200	20	4	0.0010	0.0005	0.943	0.0012	0.0005	0.967	0.0016	0.0004	0.946	-0.0006	0.0004	0.947
	200	20	5	0.0011	0.0005	0.944	0.0012	0.0005	0.969	0.0016	0.0004	0.944	-0.0006	0.0004	0.946

				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	0.0021	0.0021	0.939	0.0076	0.0023	0.939	0.0021	0.0017	0.937	-0.0005	0.0016	0.948
	100	10	5	0.0022	0.0021	0.941	0.0077	0.0023	0.940	0.0020	0.0017	0.936	-0.0004	0.0016	0.948
	200	10	4	0.0025	0.0009	0.958	0.0027	0.0010	0.945	0.0007	0.0008	0.957	0.0006	0.0008	0.952
	200	10	5	0.0025	0.0009	0.956	0.0029	0.0010	0.944	0.0007	0.0008	0.952	0.0005	0.0008	0.952
	100	20	4	0.0036	0.0009	0.958	0.0032	0.0011	0.939	-0.0001	0.0008	0.945	-0.0008	0.0007	0.954
	100	20	5	0.0035	0.0009	0.959	0.0033	0.0011	0.938	-0.0001	0.0008	0.945	-0.0008	0.0007	0.956
	200	20	4	0.0014	0.0005	0.958	0.0030	0.0005	0.953	0.0008	0.0004	0.944	0.0013	0.0004	0.955
	200	20	5	0.0013	0.0005	0.958	0.0031	0.0005	0.955	0.0008	0.0004	0.945	0.0013	0.0004	0.951
SBM	100	10	4	0.0028	0.0017	0.954	0.0078	0.0022	0.940	0.0012	0.0015	0.959	-0.0001	0.0016	0.955
	100	10	5	0.0030	0.0017	0.955	0.0082	0.0022	0.942	0.0012	0.0015	0.955	-0.0002	0.0016	0.953
	200	10	4	0.0035	0.0009	0.953	0.0030	0.0010	0.949	0.0016	0.0008	0.948	0.0008	0.0008	0.935
	200	10	5	0.0036	0.0009	0.951	0.0031	0.0010	0.949	0.0015	0.0008	0.947	0.0008	0.0008	0.934
	100	20	4	0.0026	0.0009	0.956	0.0037	0.0011	0.944	0.0003	0.0008	0.959	-0.0001	0.0008	0.956
	100	20	5	0.0028	0.0009	0.960	0.0039	0.0011	0.948	0.0003	0.0008	0.960	-0.0001	0.0008	0.955
	200	20	4	0.0014	0.0005	0.935	0.0027	0.0005	0.952	-0.0006	0.0004	0.944	0.0003	0.0004	0.964
	200	20	5	0.0014	0.0005	0.936	0.0028	0.0005	0.952	-0.0006	0.0004	0.944	0.0002	0.0004	0.964

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	0.0248	1.5084	0.0055	0.9033	0.0152	0.5697	0.0009	0.4991
	100	10	5	0.0338	1.9056	0.0066	1.2489	0.0143	0.6718	0.0016	0.5920
	200	10	4	0.0046	0.4864	0.0028	0.3811	0.0033	0.2539	0.0002	0.2210
	200	10	5	0.0050	0.6239	0.0022	0.5506	0.0034	0.2977	0.0000	0.2650
	100	20	4	0.0113	0.6173	0.0029	0.4179	0.0006	0.2634	0.0006	0.2518
	100	20	5	0.0118	0.7642	0.0031	0.5947	0.0007	0.3049	0.0003	0.2983
	200	20	4	0.0021	0.2230	0.0013	0.1936	0.0011	0.1222	0.0001	0.1195
	200	20	5	0.0026	0.2766	0.0016	0.2771	0.0010	0.1420	0.0001	0.1426
SBM	100	10	4	0.0011	0.3751	0.0010	0.1010	0.0000	0.0497	0.0002	0.0456
	100	10	5	0.0009	0.5029	0.0016	0.1413	0.0000	0.0638	0.0001	0.0569
	200	10	4	0.0002	0.0764	0.0020	0.0490	0.0002	0.0277	0.0002	0.0276
	200	10	5	0.0003	0.0967	0.0009	0.0689	0.0002	0.0342	0.0000	0.0331
	100	20	4	0.0006	0.1259	0.0009	0.0453	0.0000	0.0249	0.0003	0.0235
	100	20	5	0.0006	0.1619	0.0007	0.0616	0.0000	0.0311	0.0001	0.0288
	200	20	4	0.0000	0.0347	0.0017	0.0260	0.0000	0.0138	0.0003	0.0140
	200	20	5	0.0000	0.0430	0.0006	0.0344	0.0000	0.0169	0.0001	0.0165

				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	0.0008	0.1003	0.0021	0.0712	0.0001	0.0320	0.0001	0.0312
	100	10	5	0.0009	0.1529	0.0024	0.0898	0.0000	0.0374	0.0000	0.0367
	200	10	4	0.0003	0.0316	0.0003	0.0304	0.0001	0.0164	0.0001	0.0145
	200	10	5	0.0002	0.0388	0.0003	0.0389	0.0000	0.0186	0.0000	0.0169
	100	20	4	0.0006	0.0360	0.0013	0.0359	0.0000	0.0153	0.0001	0.0148
	100	20	5	0.0005	0.0447	0.0014	0.0451	0.0000	0.0175	0.0000	0.0172
	200	20	4	0.0002	0.0156	0.0003	0.0150	0.0001	0.0077	0.0000	0.0070
	200	20	5	0.0002	0.0192	0.0004	0.0189	0.0000	0.0087	0.0000	0.0081
SBM	100	10	4	0.0062	0.1792	0.0046	0.0801	0.0002	0.0299	0.0001	0.0294
	100	10	5	0.0071	0.2269	0.0049	0.0992	0.0002	0.0382	0.0001	0.0351
	200	10	4	0.0007	0.0445	0.0009	0.0337	0.0001	0.0151	0.0001	0.0149
	200	10	5	0.0005	0.0521	0.0010	0.0433	0.0000	0.0178	0.0000	0.0174
	100	20	4	0.0019	0.0713	0.0009	0.0379	0.0001	0.0147	0.0001	0.0150
	100	20	5	0.0017	0.0826	0.0009	0.0472	0.0000	0.0179	0.0000	0.0179
	200	20	4	0.0003	0.0204	0.0002	0.0169	0.0001	0.0078	0.0000	0.0077
	200	20	5	0.0002	0.0234	0.0002	0.0210	0.0000	0.0090	0.0000	0.0089

5.2 Testing linearity

Next, we examine the finite sample performance of the proposed LM and distance-based tests. The DGPs considered 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.

Tables 3-4 report the rejection frequencies of the distance-based and LM tests at the 10%, 5%, and 1% significance levels. We summarize our findings as follows. First, the distance-based test is somewhat undersized when either H_0^n or H_0^m is considered. When the function f_1 (f_2) is highly nonlinear, as in DGPs 2 and 4 (3 and 4), the distance-based test for H_0^n (H_0^m) has a good power property for all choices of J and all sample sizes. Second, as can be seen from Table 4, the size of the LM test is reasonably well controlled at all nominal levels. The LM test exhibits high power when the function f_1 (f_2) is nonlinear, and it also outperforms the distance-based test for all cases under investigation. Third, the network structure has a significant effect on the empirical power of the tests. Specifically, for testing the linearity of f_1, the power of the test decreases significantly when the network is changed from SBM to ER model, although it can be improved by increasing the sample size. This indicates that detecting nonlinear network effect could be harder when the network becomes denser. Finally, the choice of J has a certain influence on the performance of our tests. 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 slightly.

When we increase the sample size T to 100 and 200, the performance of the sieve LS estimators, as well as the tests, improve considerably. Especially, the size distortion of the $\mathbf{T}$ test becomes milder. Due to space limits, the corresponding results are deferred to Section D in the Supplementary Appendix.

6 Empirical application

We analyze the user posting activities on Sina Weibo, a Twitter-type online social network platform in China, using the dataset created by 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 empirical evidence that the activeness of a node is positively related to its connected neighbors. Zhu and Pan (2020) and Zhu et al. (2023) generalized the linear NAR model to the grouped NAR model, where the network effect of posting activities is allowed to be heterogeneous among different groups. The above analysis is only restricted to a linear model.

Table 3: 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	0.063	0.043	0.015	0.522	0.469	0.347	0.031	0.018	0.010	0.085	0.062	0.026
	100	10	5	0.064	0.037	0.018	0.434	0.350	0.248	0.033	0.024	0.007	0.081	0.048	0.022
	200	10	4	0.055	0.037	0.019	0.875	0.834	0.752	0.036	0.021	0.009	0.153	0.096	0.037
	200	10	5	0.064	0.043	0.019	0.793	0.732	0.606	0.037	0.027	0.009	0.159	0.095	0.042
	100	20	4	0.062	0.044	0.019	0.803	0.741	0.627	0.035	0.021	0.008	0.155	0.109	0.051
	100	20	5	0.069	0.052	0.025	0.703	0.640	0.495	0.032	0.020	0.005	0.145	0.105	0.039
	200	20	4	0.073	0.047	0.016	0.990	0.984	0.962	0.033	0.020	0.008	0.366	0.286	0.158
	200	20	5	0.071	0.044	0.019	0.973	0.955	0.911	0.032	0.020	0.006	0.365	0.268	0.144
SBM	100	10	4	0.052	0.034	0.016	0.999	0.999	0.998	0.042	0.030	0.013	0.899	0.856	0.753
	100	10	5	0.056	0.043	0.021	0.997	0.997	0.994	0.057	0.034	0.015	0.868	0.811	0.711
	200	10	4	0.040	0.024	0.009	1	1	1	0.029	0.018	0.010	0.985	0.972	0.935
	200	10	5	0.052	0.033	0.015	1	1	1	0.041	0.025	0.015	0.978	0.967	0.919
	100	20	4	0.035	0.026	0.009	1	1	1	0.039	0.024	0.010	0.999	0.994	0.984
	100	20	5	0.043	0.030	0.012	1	0.999	0.999	0.053	0.031	0.010	0.996	0.993	0.980
	200	20	4	0.046	0.024	0.010	1	1	1	0.037	0.027	0.012	1	1	1
	200	20	5	0.047	0.031	0.014	1	1	1	0.049	0.035	0.012	1	1	1

(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	0.029	0.018	0.011	0.008	0.005	0.001	0.693	0.600	0.429	0.719	0.623	0.459
	100	10	5	0.031	0.020	0.014	0.020	0.013	0.006	0.676	0.587	0.412	0.708	0.618	0.452
	200	10	4	0.014	0.007	0.005	0.008	0.001	0.000	0.970	0.954	0.885	0.988	0.974	0.935
	200	10	5	0.022	0.017	0.006	0.021	0.011	0.003	0.967	0.949	0.876	0.985	0.973	0.928
	100	20	4	0.020	0.014	0.005	0.004	0.004	0.002	0.976	0.956	0.894	0.990	0.971	0.936
	100	20	5	0.030	0.016	0.008	0.020	0.011	0.001	0.972	0.951	0.877	0.986	0.968	0.933
	200	20	4	0.025	0.015	0.005	0.009	0.002	0.000	1	1	1	1	1	1
	200	20	5	0.030	0.015	0.006	0.015	0.006	0.003	1	1	1	1	1	1
SBM	100	10	4	0.015	0.012	0.005	0.013	0.005	0.001	0.918	0.878	0.796	0.851	0.799	0.665
	100	10	5	0.020	0.014	0.004	0.031	0.017	0.009	0.908	0.867	0.766	0.849	0.786	0.645
	200	10	4	0.015	0.007	0.004	0.010	0.005	0.001	0.999	0.998	0.985	0.993	0.981	0.964
	200	10	5	0.018	0.009	0.003	0.029	0.015	0.003	0.998	0.997	0.982	0.990	0.982	0.965
	100	20	4	0.015	0.009	0.002	0.010	0.006	0.001	0.998	0.998	0.996	0.997	0.994	0.976
	100	20	5	0.021	0.010	0.004	0.026	0.015	0.009	0.998	0.998	0.993	0.997	0.993	0.974
	200	20	4	0.013	0.005	0.002	0.008	0.005	0.001	1	1	1	1	1	1
	200	20	5	0.017	0.008	0.001	0.022	0.012	0.002	1	1	1	1	1	1

Table 4: 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	0.092	0.055	0.013	0.668	0.545	0.311	0.112	0.060	0.012	0.289	0.177	0.065
	100	10	5	0.095	0.053	0.019	0.619	0.498	0.268	0.114	0.055	0.008	0.254	0.164	0.059
	200	10	4	0.102	0.045	0.007	0.929	0.883	0.739	0.099	0.049	0.018	0.468	0.360	0.155
	200	10	5	0.095	0.045	0.010	0.917	0.856	0.690	0.092	0.051	0.017	0.439	0.331	0.148
	100	20	4	0.097	0.043	0.012	0.880	0.805	0.618	0.100	0.057	0.015	0.461	0.318	0.142
	100	20	5	0.104	0.054	0.014	0.856	0.766	0.557	0.106	0.048	0.012	0.399	0.282	0.116
	200	20	4	0.116	0.053	0.012	0.999	0.990	0.952	0.091	0.056	0.007	0.726	0.621	0.398
	200	20	5	0.106	0.047	0.010	0.993	0.985	0.942	0.102	0.045	0.009	0.696	0.583	0.354
SBM	100	10	4	0.096	0.051	0.009	1	1	1	0.094	0.048	0.013	0.994	0.989	0.944
	100	10	5	0.104	0.058	0.010	1	1	1	0.087	0.050	0.011	0.991	0.976	0.915
	200	10	4	0.086	0.050	0.009	1	1	1	0.101	0.049	0.010	1	0.998	0.992
	200	10	5	0.099	0.055	0.012	1	1	1	0.114	0.055	0.014	1	0.999	0.987
	100	20	4	0.102	0.052	0.005	1	1	1	0.100	0.046	0.006	1	1	0.999
	100	20	5	0.107	0.049	0.004	1	1	1	0.101	0.047	0.008	1	1	0.999
	200	20	4	0.105	0.049	0.007	1	1	1	0.110	0.052	0.010	1	1	1
	200	20	5	0.088	0.043	0.006	1	1	1	0.113	0.068	0.015	1	1	1
(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	0.121	0.055	0.015	0.083	0.044	0.006	0.965	0.932	0.853	0.971	0.954	0.876
	100	10	5	0.111	0.052	0.011	0.101	0.042	0.011	0.950	0.919	0.822	0.970	0.940	0.829
	200	10	4	0.076	0.038	0.004	0.085	0.036	0.007	1	1	0.994	1	0.999	0.997
	200	10	5	0.081	0.039	0.004	0.086	0.038	0.008	1	1	0.992	1	1	0.996
	100	20	4	0.091	0.044	0.012	0.087	0.043	0.006	1	1	0.995	1	1	0.999
	100	20	5	0.095	0.041	0.011	0.092	0.044	0.004	1	0.999	0.995	1	0.999	0.998
	200	20	4	0.105	0.054	0.011	0.109	0.064	0.011	1	1	1	1	1	1
	200	20	5	0.111	0.059	0.010	0.100	0.055	0.016	1	1	1	1	1	1
SBM	100	10	4	0.100	0.046	0.011	0.102	0.053	0.005	0.999	0.996	0.960	0.992	0.986	0.941
	100	10	5	0.102	0.044	0.008	0.099	0.061	0.010	0.998	0.988	0.948	0.991	0.979	0.904
	200	10	4	0.100	0.057	0.012	0.103	0.044	0.009	1	1	0.999	1	1	0.999
	200	10	5	0.099	0.054	0.008	0.098	0.047	0.005	1	0.999	0.999	1	1	0.997
	100	20	4	0.112	0.051	0.007	0.109	0.057	0.014	0.999	0.999	0.999	1	1	0.998
	100	20	5	0.107	0.044	0.007	0.114	0.058	0.013	0.999	0.999	0.999	1	1	0.998
	200	20	4	0.101	0.052	0.012	0.104	0.051	0.011	1	1	1	1	1	1
	200	20	5	0.091	0.048	0.012	0.108	0.052	0.010	1	1	1	1	1	1

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 is defined by $Y_{it} = \ln(1 + P_{it})$ 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 constructed as follows: $a_{ij} = 1$ if the ith user follows the jth user on Weibo; otherwise, $a_{ij} = 0$. The resulting network density (i.e., $\sum_{i,j} a_{ij}/\{N(N-1)\}$) is around 2.2%. More information about the network structure and the distribution of the response variable can be found in Zhu et al. (2017).

For comparison, we also fit (i) the linear NAR model (13) using 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 sieve LS method. We use the cubic B-spline basis function with the number of internal knots being 2 for approximating f_1 and f_2. This yields $J_1 = J_2 = 5$. The locations of the knots are determined based on the empirical quantiles.

The standardized $\mathbf{T}$ and LM tests for testing the linearity of f_1 are 0.051 and 2.072, respectively, implying that the null hypothesis H_0^n cannot be rejected at 1% significance level for both tests. In contrast, the standardized $\mathbf{T}$ and LM tests for testing the linearity of f_2 are 3.165 and 307.460, respectively. Thus, the null hypothesis H_0^m is rejected at 1% significance level for both tests. The estimation results for f_2 are given in Figure 1. The estimated f_2 monotonically increases with a near piecewise-linear shape. The slope of curve is slightly bigger in the region $\ln(1 + P_{it}) > 6$ than $\ln(1 + P_{it}) < 6$. This implies that a node with a higher activeness level in the past tends to be more active in the future.

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 10% level. Moreover, the estimation results of the coefficient γ suggest that male users with more self-created labels tend to be more active. The estimates are somewhat different among the four specifications regarding 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) for the linear and semiparametric NAR models are shown in Table 6. It is seen that the MAPE for semiparametric NAR model (8) (0.710) is considerably smaller than the MAPE obtained by the linear NAR model (0.785).

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).

For the AR(1) model fitted to each individual node, the predicted RMSE is 3.304, which is substantially larger than that of the NAR models. We conclude that the nonlinear NAR model (8) gives significant accuracy improvement of the prediction and at the same time achieves 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

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 types of 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 other's (one's own) outcome. An empirical illustration using Sina Weibo data indicates the usefulness of the proposed model and test.

There remain some issues not addressed in this study. These include choosing an optimal order of the basis expansion J_1 and J_2, accounting for unobserved individual heterogeneity, and investigating the sieve estimation with shape restrictions (e.g., monotonicity and/or contractivity). Moreover, as pointed out by Armillotta and Fokianos (2023), it is important to study NAR models with discrete response variables, such as binary or count data. We leave these topics to future research.

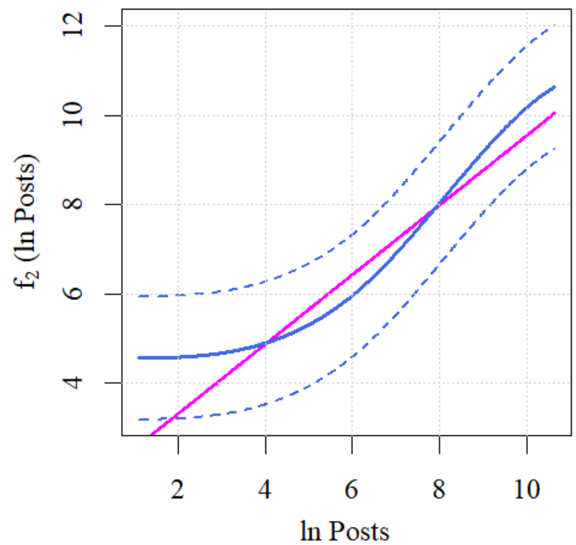

Figure 1: Estimation of f_2 based on (1) for the Sina Weibo dataset. Solid red and blue lines represent the linear and sieve estimates for f_2, respectively, and dashed blue lines represent the 95% pointwise confidence interval.

References

Aleem, A. and Lahiani, A. (2014), 'A threshold vector autoregression model of exchange rate pass-through in Mexico', *Research in International Business and Finance* **30**, 24–33.

Armillotta, M. and Fokianos, K. (2023), 'Count network autoregression', *Journal of time series analysis* .

Balke, N. S. (2000), 'Credit and economic activity: credit regimes and nonlinear propagation of shocks', *The Review of Economics and Statistics* **82**(2), 344–349.

Brana, S., Campmas, A. and Lapteacru, I. (2019), '(Un)Conventional monetary policy and bank risk-taking: A nonlinear relationship', *Economic Modelling* **81**, 576–593.

Chen, C. Y.-H., Härdle, W. K. and Okhrin, Y. (2019), 'Tail event driven networks of sifis', *Journal of Econometrics* **208**(1), 282–298. Special Issue on Financial Engineering and Risk Management.

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.

Chen, X. and Shen, X. (1998), 'Sieve extremum estimates for weakly dependent data', *Econometrica* **66**(2), 289–314.

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.

Erdos, P. L. and Rényi, A. (1959), 'On random graphs. I.', *Publicationes Mathematicae Debrecen* .

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

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

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.

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. Innovations in Multiple Time Series Analysis.

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, Vol. 32, pp. 327–359.

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

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

Knight, M., Leeming, K., Nason, G. and Nunes, M. (2020), 'Generalized network autoregressive processes and the GNAR package', *Journal of Statistical Software* **96**(5), 1–36.

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'.

Li, J. and Liao, Z. (2020), 'Uniform nonparametric inference for time series', *Journal of Econometrics* **219**(1), 38–51.

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

Mitchener, K. J. and Richardson, G. (2019), 'Network contagion and interbank amplification during the great depression', *Journal of Political Economy* **127**(2), 465–507.

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 Jin, S. (2012), ‘Sieve estimation of panel data models with cross section dependence’, *Journal of Econometrics* **169**(1), 34–47. Recent Advances in Panel Data, Nonlinear and Nonparametric Models: A Festschrift in Honor of Peter C.B. Phillips.

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’.

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. (2023), ‘Simultaneous estimation and group identification for network vector autoregressive model with heterogeneous nodes’, *Journal of Econometrics* p. 105564.

Supplementary Appendix to
"Nonparametric Network Autoregression"

This supplementary appendix contains some useful auxiliary results and the proofs for the main theorems in the paper "Nonparametric Network Autoregression". It also includes additional simulation results and theoretical analysis of the proposed nonparametric NAR model under the asymptotic regime where T diverges to infinity and N is fixed.

A Some Lemmas

We first introduce some lemmas that are useful in dealing with convergence in probability of matrices involving 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 $\mathcal{D}_{NT_N} = \{(i,t) : 1 \le i \le N, 1 \le t \le T_N, i,t \in \mathbb{Z}\}$ 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 spatial-time NED process as follows.

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

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

for an array of finite positive constants $\{d_{it} : (i,t) \in \mathcal{D}_{NT_N}; N \ge 1\}$ and some sequence $\upsilon_s \ge 0$ with $\upsilon_s \to 0$ as $s \to \infty$, where $\mathcal{F}_{it}(s)$ is the σ-field generated by the random variables $\{e_{j\tau} : \Delta((i,t),(j,\tau)) \le s\}$ and $\mathbb{Y}_0$. 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.

Lemma 1 establishes the uniform and geometric L^2-NED property of $\{Y_{it}\}$.

Lemma 1. *Let $v_{it} = (Z_i^\top, \epsilon_{it})^\top$. Under Assumptions 1-3, $\{Y_{it}\}$ is uniformly and geometrically L^2-NED on $\{v_{it}\}$.*

Proof. For $A_t = (a_{1t}, \ldots, a_{Nt})^\intercal \in \mathbb{R}^N$, let $\bar{F}(A_t) \equiv WF_1(A_t) + F_2(A_t)$ be a mapping from $\mathbb{R}^N$ to $\mathbb{R}^N$. For $K \geq 1$, define

$$\bar{\mathbb{Y}}_t = \begin{cases} \bar{F}(\bar{\mathbb{Y}}_{t-1}), & t > 0 \\ \mathbb{Y}_0, & t \leq 0 \end{cases}, \quad \hat{\mathbb{Y}}_{t-K}^s = \begin{cases} \bar{F}(\hat{\mathbb{Y}}_{t-K}^{s-1}) + \mathbb{Z}\gamma + \mathcal{E}_t, & \max\{t-K, 0\} < s \leq t \\ \bar{\mathbb{Y}}_s, & s \leq \max\{t-K, 0\} \end{cases}. \quad (17)$$

We first show that under Assumption 2, it holds that

$$\left|\mathbb{Y}_t - \hat{\mathbb{Y}}_{t-K}^t\right|_{vec} \preceq G^K \sum_{j=0}^{t-K-1} G^j \left|\mathbb{Z}\gamma + \mathcal{E}_{t-K-j}\right|_{vec}, \quad (18)$$

where $G \equiv \kappa_1 W + \kappa_2 I_N$. Indeed, for $t \geq 1$,

$$\begin{aligned} |\mathbb{Y}_t - \bar{\mathbb{Y}}_t|_{vec} &= |\bar{F}(\mathbb{Y}_{t-1}) + \mathbb{Z}\gamma + \mathcal{E}_t - \bar{F}(\bar{\mathbb{Y}}_{t-1})|_{vec} \\ &= |W(F_1(\mathbb{Y}_{t-1}) - F_1(\bar{\mathbb{Y}}_{t-1})) + F_2(\mathbb{Y}_{t-1}) - F_2(\bar{\mathbb{Y}}_{t-1})|_{vec} + |\mathbb{Z}\gamma + \mathcal{E}_t|_{vec} \\ &\preceq G|\mathbb{Y}_{t-1} - \bar{\mathbb{Y}}_{t-1}|_{vec} + |\mathbb{Z}\gamma + \mathcal{E}_t|_{vec} \\ &\preceq G\left(G|\mathbb{Y}_{t-2} - \bar{\mathbb{Y}}_{t-2}|_{vec} + |\mathbb{Z}\gamma + \mathcal{E}_{t-1}|_{vec}\right) + |\mathbb{Z}\gamma + \mathcal{E}_t|_{vec} \\ &\preceq G^t|\mathbb{Y}_0 - \bar{\mathbb{Y}}_0|_{vec} + \sum_{l=0}^{t-1} G^l|\mathbb{Z}\gamma + \mathcal{E}_{t-l}|_{vec}. \end{aligned}$$

As $\mathbb{Y}_0 = \bar{\mathbb{Y}}_0$, we have $|\mathbb{Y}_t - \bar{\mathbb{Y}}_t|_{vec} \preceq \sum_{l=0}^{t-1} G^l|\mathbb{Z}\gamma + \mathcal{E}_{t-l}|_{vec}$. For $t - K > 0$,

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

Besides, for $t - K \leq 0$, we can easily show that $|\mathbb{Y}_t - \hat{\mathbb{Y}}_{t-K}^t|_\infty \leq (\kappa_1 + \kappa_2)^t|Y_0 - \bar{Y}_0|_\infty = 0$. Result (18) follows by combining the above results.

Denote the ith element of $\hat{\mathbb{Y}}_{t-K}^t$ by $\hat{Y}_{i,t-K}^t$. Note that $\|G^j\|_\infty \leq \|G\|_\infty^j = (\kappa_1 + \kappa_2)^j$, for $j \in \mathbb{N}$. Then, by (18) and Minkowski's inequality, we have for each $1 \leq i \leq N$,

$$\begin{aligned} \|Y_{it} - \hat{Y}_{i,t-K}^t\|_2 &\leq \sum_{j=0}^{t-K-1} \|G^{j+K}\|_\infty \sup_{i\geq 1} \left\|Z_i^\intercal\gamma + \epsilon_{i,t-K-j}\right\|_2 \\ &\leq \sum_{j=0}^{t-K-1} \|G\|_\infty^{K+j} C_v \leq C_v d^K/(1-d), \end{aligned} \quad (19)$$

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 $\hat{Y}_{i,t-K}^t$ is a $\mathcal{F}_{it}(K)$-measurable approximation to Y_{it}. It follows from Theorem 10.12 of Davidson (1994) that $\|Y_{it} - \mathbb{E}[Y_{it} \mid \mathcal{F}_{it}(K)]\|_2 \leq \|Y_{it} - \hat{Y}_{i,t-K}^t\|_2 \leq C_v d^K/(1-d)$. The lemma immediately follows. $\square$

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 $\mathbb{Q}_{t-1} = (W\mathbb{P}_{t-1}^{J_1}, \mathbb{P}_{t-1}^{J_2})$. 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, $\|\widehat{C}_{NT,ZJ} - C_{N,ZJ}\| = O_p\big(\zeta_{0,J}^{-1}\sqrt{r_{NT,J}}\big) = o_p(1)$, and $\|\widehat{\Psi}_{NT,J} - \Psi_{N,J}\| = O_p\big(\sqrt{r_{NT,J}}\big) = o_p(1)$, where $r_{NT,J}$ is defined in Assumption 5.*

Proof. Hereinafter we suppress the dependence of $\widehat{C}_{NT,ZJ}$, $\widehat{\Psi}_{NT,J}$, $C_{N,ZJ}$ and $\Psi_{N,J}$ on (N,T) for notational simplicity. (i) Recall that $\bar{P}_{it}^{J_1} = \sum_{j=1}^N w_{ij} p^{J_1}(Y_{jt})$ and $P_{it}^{J_2} = p^{J_2}(Y_{it})$. Partition $\widehat{C}_{ZJ}$ as

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

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 $\{v_{j\tau} : (j,\tau) \in \mathcal{D}_{NT_N}, \Delta\big((i,t),(j,\tau)\big) \leq s\}$ and $\mathbb{Y}_0$. Let $\hat{\bar{P}}_{i,t-s}^t \equiv \sum_{j=1}^N w_{ij} p^{J_1}(\hat{Y}_{j,t-s}^t)$ and $\hat{P}_{i,t-s}^t \equiv p^{J_2}(\hat{Y}_{i,t-s}^t)$, where $\hat{Y}_{i,t-s}^t$ is a $\mathcal{F}_{it}(s)$-measurable random variable defined in Lemma 1. Note that

$$\begin{aligned}
\mathbb{E}\big\|Z_i\bar{P}_{i,t-1}^{J_1\top} - Z_i\hat{\bar{P}}_{i,t-s}^{(t-1)\top}\big\|^2 &= \mathbb{E}\|Z_i\|^2 \left\|\sum_{j=1}^N w_{ij}\big(p^{J_1}(Y_{j,t-1}) - p^{J_1}(\hat{Y}_{j,t-s}^{t-1})\big)\right\|^2 \\
&\leq C_z^2 \sum_{j=1}^N w_{ij}\mathbb{E}\left\|p^{J_1}(Y_{j,t-1}) - p^{J_1}(\hat{Y}_{j,t-s}^{t-1})\right\|^2 \\
&\leq C_z^2 \sup_{1\leq i\leq N} \mathbb{E}\left\|p^{J_1}(Y_{i,t-1}) - p^{J_1}(\hat{Y}_{i,t-s}^{t-1})\right\|^2 \\
&\leq C_z^2 \zeta_{1,J_1}^2 \sup_{1\leq i\leq N} \mathbb{E}\big\|Y_{i,t-1} - \hat{Y}_{i,t-s}^{t-1}\big\|^2, \qquad (20)
\end{aligned}$$

where we have used the facts that $\max_{1\leq i\leq N}\|Z_i\| \leq C_z < \infty$ 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 $\mathbb{E}[a_{ij}|\mathcal{F}]$,

$1 \leq i \leq p, 1 \leq j \leq q$. Then,

$$\begin{aligned}\mathbb{E}\big\|\widehat{C}_{ZJ,1} - C_{ZJ,1}\big\|^2 &= \mathbb{E}\Bigg\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\big(Z_i\bar{P}_{i,t-1}^{J_1\intercal} - \mathbb{E}\big(Z_i\bar{P}_{i,t-1}^{J_1\intercal}\big)\big)\Bigg\|^2 \\ &= \frac{1}{(NT)^2}\sum_{(i,t)\in\mathcal{D}_{NT}}\mathbb{E}\big\|Z_i\bar{P}_{i,t-1}^{J_1\intercal} - \mathbb{E}(Z_i\bar{P}_{i,t-1}^{J_1\intercal})\big\|^2 \\ &\quad + \frac{1}{(NT)^2}\sum_{(i,t)\in\mathcal{D}_{NT}}\sum_{(j,\tau)\in\mathcal{D}_{NT}\setminus(i,t)}\mathrm{Cov}\big(Z_i\bar{P}_{i,t-1}^{J_1\intercal}, Z_j\bar{P}_{j,\tau-1}^{J_1\intercal}\big),\end{aligned} \tag{21}$$

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

$$\begin{aligned}\frac{1}{(NT)^2}\sum_{(i,t)\in\mathcal{D}_{NT}}\mathbb{E}\big\|Z_i\bar{P}_{i,t-1}^{J_1\intercal} - \mathbb{E}(Z_i\bar{P}_{i,t-1}^{J_1\intercal})\big\|^2 &\leq \frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}\|Z_i\bar{P}_{i,t-1}^{J_1\intercal}\|^2 \\ &\leq \frac{C_z^2}{NT}\mathrm{tr}\big\{\mathbb{E}\big(\mathbb{P}_{t-1}^{J_1\intercal}W^\intercal W\mathbb{P}_{t-1}^{J_1}/N\big)\big\} \\ &= O\left(J_1(NT)^{-1}\right).\end{aligned} \tag{22}$$

We control the second term on the r.h.s. of (21) using similar arguments as in Lemma A.3 of Jenish and Prucha (2012). For any $(i,t) \in \mathcal{D}_{NT}$ and $m \geq 1$, let

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

By Jansen's inequality, we have for all $(i,t) \in \mathcal{D}_{NT}$ and $m \geq 1$,

$$\begin{aligned}\mathbb{E}\|\xi_{i,t-1}^m\|^2 &= \mathbb{E}\big\|\mathbb{E}\big[Z_i\bar{P}_{i,t-1}^{J_1\intercal}|\mathcal{F}_{i,t-1}(m)\big]\big\|^2 = \mathbb{E}\Bigg\{\sum_{k=1}^{d_Z}\sum_{l=1}^{J_1}\Big|\mathbb{E}\big[Z_{ik}\mathbf{w}_i^\intercal p_l(\mathbb{Y}_{t-1}) \mid \mathcal{F}_{i,t-1}(m)\big]\Big|^2\Bigg\} \\ &\leq \sum_{k=1}^{d_Z}\sum_{l=1}^{J_1}\mathbb{E}\big\{\mathbb{E}\big[\big|Z_{ik}\mathbf{w}_i^\intercal p_l(\mathbb{Y}_{t-1})\big|^2 \mid \mathcal{F}_{i,t-1}(m)\big]\big\} = \mathbb{E}\big\|Z_i\bar{P}_{i,t-1}^{J_1\intercal}\big\|^2,\end{aligned}$$

where $\mathbf{w}_i^\intercal$ denotes the ith row of the matrix W. Thus $\mathbb{E}\|\xi_{i,t-1}^m\|^2 \leq \mathbb{E}\|Z_i\bar{P}_{i,t-1}^{J_1\intercal}\|^2 \lesssim J_1$, and $\mathbb{E}\|\eta_{i,t-1}^m\|^2 \leq 4\mathbb{E}\|Z_i\bar{P}_{i,t-1}^{J_1\intercal}\|^2 \lesssim J_1$ by Assumption 4(i). Also, (19) and (20) imply that for $m \geq 1$,

$$\mathbb{E}\|\eta_{i,t-1}^m\|^2 \leq \mathbb{E}\left\|Z_i\bar{P}_{i,t-1}^{J_1\intercal} - Z_i\hat{\bar{P}}_{i,t-m-1}^{(t-1)\intercal}\right\|^2 \leq C_z\zeta_{1,J_1}^2\sup_{i\geq 1}\mathbb{E}\big\|Y_{i,t-1} - \hat{Y}_{i,t-m-1}^{t-1}\big\|^2 \lesssim \zeta_{1,J_1}^2 d^{2m}.$$

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

$$Z_i\bar{P}_{i,t-1}^{J_1\intercal} = \xi_{i,t-1}^{\lfloor h/3\rfloor} + \eta_{i,t-1}^{\lfloor h/3\rfloor}, \quad Z_j\bar{P}_{j,\tau-1}^{J_1\intercal} = \xi_{j,\tau-1}^{\lfloor h/3\rfloor} + \eta_{j,\tau-1}^{\lfloor h/3\rfloor}, \tag{23}$$

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

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

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:

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

where C is a constant not dependent on i, t, N, T and J_1. Lastly,

$$\left|\mathrm{Cov}\big(\eta_{i,t-1}^{\lfloor h/3 \rfloor}, \eta_{j,\tau-1}^{\lfloor h/3 \rfloor}\big)\right| \leq 2\left\{\mathbb{E}\left\|\eta_{i,t-1}^{\lfloor h/3 \rfloor}\right\|^2 \mathbb{E}\left\|\eta_{j,\tau-1}^{\lfloor h/3 \rfloor}\right\|^2\right\}^{1/2} \leq C\zeta_{1,J_1} J_1^{1/2} d^{\lfloor h/3 \rfloor}.$$

A combination of above results yields that $\left|\mathrm{Cov}\big(Z_i \bar{P}_{i,t-1}^{\mathsf{T}}, Z_j \bar{P}_{j,\tau-1}^{\mathsf{T}}\big)\right| \leq C\zeta_{1,J_1} J_1^{1/2} d^{\lfloor h/3 \rfloor}$.

We now derive the bound for the second term on the r.h.s. of (21). Using the above result, we have for any $(i,t) \in \mathcal{D}_{NT}$,

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

$$\leq C\zeta_{1,J_1} J_1^{1/2} \sum_{s=1}^{N} \min\{s,T\} \big|\mathcal{N}_N^{\partial}(i;s)\big| d^{\lfloor s/3\rfloor}, \tag{24}$$

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

$$\sum_{\tau\neq t}^{T} \mathrm{Cov}\big(Z_i\bar{P}_{i,t-1}^{\mathsf{T}}, Z_i\bar{P}_{i,\tau-1}^{\mathsf{T}}\big) \leq C\zeta_{1,J_1} J_1^{1/2} \sum_{s=1}^{T-1} d^{\lfloor s/3\rfloor} \leq C\zeta_{1,J_1} J_1^{1/2}, \tag{25}$$

and

$$\sum_{j\neq i}^{N} \mathrm{Cov}\big(Z_i\bar{P}_{i,t-1}^{\mathsf{T}}, Z_j\bar{P}_{j,t-1}^{\mathsf{T}}\big) \leq C\zeta_{1,J_1} J_1^{1/2} \sum_{s=1}^{N} \big|\mathcal{N}_N^{\partial}(i;s)\big| d^{\lfloor s/3\rfloor}. \tag{26}$$

Using (24)-(26) 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 (21) is bounded by a multiple of

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

Combining this with (21) and (22), we conclude that $\mathbb{E}\|\widehat{C}_{ZJ,1} - C_{ZJ,1}\|^2 = O(J_1/(NT) + \zeta_{1,J_1} J_1^{1/2} \min\big\{T^{-1}\Xi_{N,1}, \Xi_{N,0}\big\})$. 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 immediately from Markov's inequality.

(ii) Next, for $\widehat{\Psi}_J$, write that

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

where $\widehat{\Psi}_{J,21} = \widehat{\Psi}_{J,12}^{\mathsf{T}}$, $\widehat{\Psi}_{J,11}$, $\widehat{\Psi}_{J,12}$ 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 $\Psi_{J,22} = \mathbb{E}(\widehat{\Psi}_{J,22})$.

We adopt the same notation as the above proof. First, we have $\|\bar{P}_{i,t-1}\|^2 = \|\sum_{j=1}^{N} w_{ij} p^{J_1}(Y_{j,t-1})\|^2 \leq \sum_{j=1}^{N} w_{ij}\|p^{J_1}(Y_{j,t-1})\|^2 \leq \sup_{1\leq i\leq N}\|p^{J_1}(Y_{i,t-1})\|^2 \leq C\zeta_{0,J_1}^2$ and also

$\|\hat{\bar{P}}_{i,t-1}\|^2 \leq C\zeta_{0,J_1}^2$ a.s.. Hence, by triangular and CS inequalities,

$$
\begin{aligned}
&\mathbb{E}\big\|\bar{P}_{i,t-1}\bar{P}_{i,t-1}^{\intercal} - \hat{\bar{P}}_{i,t-s}^{t-1}\hat{\bar{P}}_{i,t-s}^{(t-1)\intercal}\big\|^2 \\
&\leq 2\mathbb{E}\left\|\bar{P}_{i,t-1}\big(\bar{P}_{i,t-1} - \hat{\bar{P}}_{i,t-s}^{t-1}\big)^{\intercal}\right\|^2 + 2\mathbb{E}\left\|\big(\bar{P}_{i,t-1} - \hat{\bar{P}}_{i,t-s}^{t-1}\big)\hat{\bar{P}}_{i,t-s}^{(t-1)\intercal}\right\|^2 \\
&= 2\mathbb{E}\big(\|\bar{P}_{i,t-1}\|^2 + \|\hat{\bar{P}}_{i,t-s}^{t-1}\|^2\big)\left\|\sum_{j=1}^{N} w_{ij}\big(p^{J_1}(Y_{j,t-1}) - p^{J_1}(\hat{Y}_{j,t-s}^{t-1})\big)\right\|^2 \\
&\leq C\zeta_{0,J_1}^2\sum_{j=1}^{N} w_{ij}\mathbb{E}\Big\|p^{J_1}(Y_{j,t-1}) - p^{J_1}(\hat{Y}_{j,t-s}^{t-1})\Big\|^2 \\
&\leq C\zeta_{0,J_1}^2\sup_{1\leq i\leq N}\mathbb{E}\Big\|p^{J_1}(Y_{i,t-1}) - p^{J_1}(\hat{Y}_{i,t-s}^{t-1})\Big\|^2 \\
&\leq C\zeta_{0,J_1}^2\zeta_{1,J_1}^2\sup_{1\leq i\leq N}\mathbb{E}\big\|Y_{i,t-1} - \hat{Y}_{i,t-s}^{t-1}\big\|^2. \qquad (27)
\end{aligned}
$$

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

$$
\begin{aligned}
\mathbb{E}\big\|\widehat{\Psi}_{J,11} - \Psi_{J,11}\big\|^2 &= \mathbb{E}\left\|\frac{1}{NT}\sum_{(i,t)\in\mathcal{D}_{NT}}\big(\bar{P}_{i,t-1}\bar{P}_{i,t-1}^{\intercal} - \mathbb{E}\big(\bar{P}_{i,t-1}\bar{P}_{i,t-1}^{\intercal}\big)\big)\right\|^2 \qquad (28) \\
&= \frac{1}{(NT)^2}\sum_{(i,t)\in\mathcal{D}_{NT}}\mathbb{E}\big\|\bar{P}_{i,t-1}\bar{P}_{i,t-1}^{\intercal} - \mathbb{E}(\bar{P}_{i,t-1}\bar{P}_{i,t-1}^{\intercal})\big\|^2 \\
&\quad + \frac{1}{(NT)^2}\sum_{(i,t)\in\mathcal{D}_{NT}}\sum_{(j,\tau)\in\mathcal{D}_{NT}\setminus(i,t)}\mathrm{Cov}\big(\bar{P}_{i,t-1}\bar{P}_{i,t-1}^{\intercal}, \bar{P}_{j,\tau-1}\bar{P}_{j,\tau-1}^{\intercal}\big).
\end{aligned}
$$

By Assumption 4, we have the first term on the r.h.s. of (28) is less than

$$
\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}\|\bar{P}_{i,t-1}\bar{P}_{i,t-1}^{\intercal}\|^2 \leq \frac{\zeta_{0,J_1}^2}{NT}\mathrm{tr}\big\{\mathbb{E}\big(\mathbb{P}_{t-1}^{J_1\intercal}W^{\intercal}W\mathbb{P}_{t-1}^{J_1}/N\big)\big\} = O\left(\frac{\zeta_{0,J_1}^2 J_1}{NT}\right). \qquad (29)
$$

The rest of the proof proceeds analogously as the proof of (24)-(26). For ease of comparability, we use the same notations as those used there. For all $(i,t)\in\mathcal{D}_{NT}$ and $m\in\mathbb{N}$, denote

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

By Jensen's inequality, we have for all $(i,t)\in\mathcal{D}_{NT}$ and $m\in\mathbb{N}$, $\mathbb{E}\|\xi_{i,t-1}^m\|^2 \leq \mathbb{E}\big\|\bar{P}_{i,t-1}\bar{P}_{i,t-1}^{\intercal}\big\|^2$. It follows from Assumption 4 that $\mathbb{E}\|\xi_{i,t-1}^m\|^2 \leq \mathbb{E}\|\bar{P}_{i,t-1}\bar{P}_{i,t-1}^{\intercal}\|^2 \lesssim J_1\zeta_{0,J_1}^2$, and $\mathbb{E}\|\eta_{i,t-1}^m\|^2 \leq$

$4\mathbb{E}\|\bar{P}_{i,t-1}\bar{P}_{i,t-1}^{\intercal}\|^2 \lesssim J_1\zeta_{0,J_1}^2$. Also, (19) and (27) imply that

$$\mathbb{E}\|\eta_{i,t-1}^{m}\|^2 \le C\zeta_{0,J_1}^2\zeta_{1,J_1}^2 \sup_{1\le i\le N}\mathbb{E}\big\|Y_{i,t-1}-\hat{Y}_{i,t-m-1}^{t-1}\big\|^2 \lesssim \zeta_{0,J_1}^2\zeta_{1,J_1}^2 d^{2m}.$$

Next, by decomposing $\bar{P}_{i,t-1}\bar{P}_{i,t-1}^{\intercal}$ and $\bar{P}_{j,\tau-1}\bar{P}_{j,\tau-1}^{\intercal}$, $(i,t)\neq(j,\tau)$ in the same way as (23), we have from above results that

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

where $h=\Delta((i,t),(j,\tau))$. Therefore, using similar arguments as showing (i), we have

$$\begin{aligned}&\frac{1}{(NT)^2}\sum_{(i,t)\in\mathcal{D}_{NT}}\sum_{(j,\tau)\in\mathcal{D}_{NT}\setminus(i,t)}\big|\mathrm{Cov}\big(\bar{P}_{i,t-1}\bar{P}_{i,t-1}^{\intercal},\bar{P}_{j,\tau-1}\bar{P}_{j,\tau-1}^{\intercal}\big)\big|\\&=O\Big(J_1^{1/2}\zeta_{0,J_1}^2\zeta_{1,J_1}\min\{\Xi_{N,0},\Xi_{N,1}/T\}\Big)\end{aligned}\tag{30}$$

as $N\to\infty$. It follows from (28)-(30) 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 matrix $\widehat{\Psi}_J$ can be shown analogously. We finally have $\mathbb{E}\big\|\widehat{\Psi}_J-\Psi_J\big\|^2=O(\zeta_{0,J}^2(J/(NT)+J^{1/2}\zeta_{1,J}\min\{\Xi_{N,0},T^{-1}\Xi_{N,1}\}))$ as $N\to\infty$. By Markov's inequality, $\big\|\widehat{\Psi}_J-\Psi_J\big\|=O_p(\sqrt{r_{NT,J}})$. □

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,*

$$\big\|A^{\intercal}BA-A'^{\intercal}B'A'\big\|\le\sigma_{max}^2(A)\|B-B'\|+\lambda_{max}(B')\|A-A'\|[\|A-A'\|+2\sigma_{max}(A')].$$

Fact 2. *(Fact A.3 in Hoshino (2022)) Let A and B be symmetric matrices 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 $\widehat{\Sigma}_{NT,Z}=\widehat{\Psi}_{N,Z}-\widehat{C}_{NT,ZJ}\widehat{\Psi}_{NT,J}^{-1}\widehat{C}_{NT,ZJ}^{\top}$, and $\widehat{\Sigma}_{NT,J}=\widehat{\Psi}_{NT,J}-\widehat{C}_{NT,ZJ}^{\top}\widehat{\Psi}_{N,Z}^{-1}\widehat{C}_{NT,ZJ}$, where $\widehat{\Psi}_{N,Z}=\mathbb{Z}^{\top}\mathbb{Z}/N$. Under Assumptions 1-5 and 7, (i) $\|\widehat{\Sigma}_{NT,Z}-\Sigma_{N,Z}\|=O_p\big(N^{-1/2}+\underline{\nu}_{N,J_1J_2}^{-2}r_{NT,J}^{1/2}\big)$; and (ii) $\|\widehat{\Sigma}_{NT,f}-\Sigma_{N,f}\|=O_p\big(N^{-1/2}+r_{NT,J}^{1/2}\big)$.*

Proof. (i) Note that $\|\widehat{\Sigma}_{NT,Z}-\Sigma_{N,Z}\| = \big\|\widehat{\Psi}_{N,Z}+\widehat{C}_{NT,ZJ}\widehat{\Psi}_{NT,J}^{-1}\widehat{C}_{NT,ZJ}^{\intercal}-\Psi_{N,Z}-C_{N,ZJ}\Psi_{N,J}^{-1}C_{N,ZJ}^{\intercal}\big\|\le\big\|\widehat{\Psi}_{N,Z}-\Psi_{N,Z}\big\|+\big\|\widehat{C}_{NT,ZJ}\widehat{\Psi}_{NT,J}^{-1}\widehat{C}_{NT,ZJ}^{\intercal}-C_{N,ZJ}\Psi_{N,J}^{-1}C_{N,ZJ}^{\intercal}\big\|$. First, by Assumption 3 and Markov's inequality, it is easy to see that $\big\|\widehat{\Psi}_{N,Z}-\Psi_{N,Z}\big\|=O_p(1/\sqrt{N})$. Next, using Assumption 4, we have $\lambda_{\min}(\Psi_{N,J})\ge\underline{\nu}_{N,J_1J_2}>0$ and $\sigma_{\max}(C_{N,ZJ})\le\bar{c}<\infty$ for sufficiently large N. These and Lemma 2 together imply that $\sigma_{\max}^2(\widehat{C}_{NT,ZJ})=O_p(1)$

and $\lambda_{\min}(\widehat{\Psi}_{NT,J}) \geq c\underline{\nu}_{N,J_1J_2} > 0$ for some $0 < c < \infty$ with probability approaching one (wpa1). Combining the latter result, Lemma 2 and Fact 2, we obtain that $\|\widehat{\Psi}_{NT,J}^{-1} - \Psi_{N,J}^{-1}\| = O(\underline{\nu}_{N,J_1J_2}^{-2} r_{NT,J}^{1/2})$. Hence,

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

by Lemma 2 and Fact 1. The first part of Lemma 3(i) follows. This and Assumption 7 imply that $\lambda_{\min}(\widehat{\Sigma}_{NT,Z})$ is bounded away from zero for sufficiently large N. Then the second part of result (i) directly follows from Fact 2.

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

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

which is $o_p(1)$. This and Assumption 4 imply that $\lambda_{\min}(\widehat{\Sigma}_{NT,f}) \geq c\underline{\nu}_{N,J_1J_2}$ for some $0 < c < \infty$ wpa1. Then, result (ii) follows from Fact 2. □

B Proof of Main Theorems

B.1 Proof of Proposition 1

Define $|M|_{vec}$ as $|M|_{vec} = (|m_{ij}|) \in \mathbb{R}^{n\times p}$ for any arbitrary matrix $M = (m_{ij}) \in \mathbb{R}^{n\times p}$. 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 $1 \leq j \leq p$. For a random variable X, let $\|X\|_r = \mathbb{E}^{1/r}|X|^r$, and for a random vector $Z = (Z_1, \ldots, Z_k)^\top$, let $\|Z\|_{r,vec} = (\|Z_1\|_r, \ldots, \|Z_k\|_r)^\top$. We first consider 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,

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

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

$$|\mathbb{Y}_t|_{vec} \preceq G|\mathbb{Y}_{t-1}|_{vec} + |\mathbb{Z}\gamma|_{vec} + |\mathcal{E}_t|_{vec} \preceq \sum_{j=0}^{\infty} G^j(|\mathbb{Z}\gamma|_{vec} + |\mathcal{E}_{t-j}|_{vec}). \tag{31}$$

Let $C_0 = \max_{i\ge1} \mathbb{E}|Z_i^\top\gamma| + \mathbb{E}|\epsilon_{it}|$. 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 $\mathbb{E}|\mathbf{w}_N^\top\mathbb{Y}_t| \le \mathbb{E}[|\mathbf{w}_N|_{vec}^\top \sum_{j=0}^\infty G^j(|\mathbb{Z}\gamma|_{vec} + |\mathcal{E}_{t-j}|_{vec})] \le |\mathbf{w}_N|_1 C_0 \sum_{j=0}^\infty(\kappa_1+\kappa_2)^j < \infty$. 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, $\mathbb{E}|\mathbf{w}_N^\top(\mathbb{Y}_t - \tilde{\mathbb{Y}}_t)| \le |\mathbf{w}_N|_{vec}^\top G\mathbb{E}|\mathbb{Y}_{t-1} - \tilde{\mathbb{Y}}_{t-1}|_{vec} = 0$ 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 $a \ge 1$, $\|Y_{it}\|_a \le \|e_i^\top \sum_{j=0}^\infty G^j(|\mathbb{Z}\gamma|_{vec} + |\mathcal{E}_{t-j}|_{vec})\|_a \le \sum_{j=0}^\infty(\kappa_1+\kappa_2)^j \max_{i\ge1}\|Z_i^\top\gamma + \epsilon_{it}\|_a$, where e_i is an $N\times1$ 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 $\max_{i\ge1}\mathbb{E}|Z_i|_1^a < \infty$. The latter condition is implied by the compactness of $\mathcal{R}_Z$. $\square$

B.2 Proof of Theorem 1

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

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

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 $G(\mathbb{Y}_t) = WF_1(\mathbb{Y}_t) + F_2(\mathbb{Y}_t)$.

Under Assumption 7, Lemma 2(i) implies that $\lambda_{\min}(\widehat{\Sigma}_{NT,Z}) \ge C$ for some $0 < C < \infty$

wpa1. As $\mathscr{M}(\mathbf{Q}) = I_{NT} - \mathbf{Q}(\mathbf{Q}^\top\mathbf{Q})^{-1}\mathbf{Q}^\top$ is an idempotent matrix, we have

$$\begin{aligned}\|D_1\|^2 &= [\mathbf{G}(\mathbf{Y}_{-1}) - \mathbf{Q}\beta]^\top \mathscr{M}(\mathbf{Q})\mathbf{Z}\widehat{\Sigma}_{NT,Z}^{-2}\mathbf{Z}^\top \mathscr{M}(\mathbf{Q})[\mathbf{G}(\mathbf{Y}_{-1}) - \mathbf{Q}\beta]/(NT)^2 \\ &\leq O_p(1)\cdot[\mathbf{G}(\mathbf{Y}_{-1}) - \mathbf{Q}\beta]^\top[\mathbf{G}(\mathbf{Y}_{-1}) - \mathbf{Q}\beta]/(NT) = O_p(J_1^{-2\mu} + J_2^{-2\mu}).\end{aligned}$$

Next, for D_2, we write $D_2 = D_{21} + D_{22}$ where $D_{21} = \Sigma_{N,Z}^{-1}\boldsymbol{\xi}^\top\boldsymbol{\mathcal{E}}/(NT)$, $D_{22} = \big(\widehat{\Sigma}_{NT,Z}^{-1}\mathbf{Z}^\top\mathscr{M}(\mathbf{Q}) - \Sigma_{N,Z}^{-1}\boldsymbol{\xi}^\top\big)\boldsymbol{\mathcal{E}}/(NT)$, and $\boldsymbol{\xi} = \mathbf{Z} - \mathbf{Q}\Psi_J^{-1}C_{ZJ}^\top$. As the variance matrix of $\boldsymbol{\xi}^\top\boldsymbol{\mathcal{E}}/(NT)$ equals $\mathrm{Var}\big(\sum_{t=1}^T(\mathbb{Z}^\top - C_{ZJ}\Psi_J^{-1}\mathbb{Q}_{t-1}^\top)\mathcal{E}_t/(NT)\big) = (NT)^{-2}\sum_{t=1}^T \mathrm{Var}\big((\mathbb{Z}^\top - C_{ZJ}\Psi_J^{-1}\mathbb{Q}_{t-1}^\top)\mathcal{E}_t\big) = \sigma^2(NT)^{-1}\Sigma_{N,Z}$, we have $\mathbb{E}\|D_{21}\|^2 = \sigma^2\mathrm{tr}(\Sigma_{N,Z}^{-1}\Sigma_{N,Z}\Sigma_{N,Z}^{-1}/(NT)) = O((NT)^{-1})$. It follows from Markov's inequality that $\|D_{21}\| = O_p((NT)^{-1/2})$. By Lemma 2(ii) and Assumption 4, $\lambda_{\max}(\widehat{\Psi}_J^{-1}) = O_p(\underline{\nu}_{N,J_1J_2}^{-1})$. Using Fact 2 in the Supplementary Appendix, we have $\|\widehat{\Psi}_J^{-1} - \Psi_J^{-1}\| = O_p(\underline{\nu}_{N,J_1J_2}^{-2}r_{NT,J}^{1/2})$. Then, $\|\widehat{C}_{ZJ}\widehat{\Psi}_J^{-1} - C_{ZJ}\Psi_J^{-1}\| = \|C_{ZJ}(\widehat{\Psi}_J^{-1} - \Psi_J^{-1}) + (\widehat{C}_{ZJ} - C_{ZJ})\widehat{\Psi}_J^{-1}\| = O_p\big(\underline{\nu}_{N,J_1J_2}^{-2}r_{NT,J}^{1/2}\big) = o_p(1)$. Accordingly,

$$\begin{aligned}\|(\mathbf{Z}^\top\mathscr{M}(\mathbf{Q})\boldsymbol{\mathcal{E}} - \boldsymbol{\xi}^\top\boldsymbol{\mathcal{E}})/NT\| &= \|(\widehat{C}_{ZJ}\widehat{\Psi}_J^{-1} - C_{ZJ}\Psi_J^{-1})\mathbf{Q}^\top\boldsymbol{\mathcal{E}}/NT\| \\ &= o_p(\|C_{ZJ}\Psi_J^{-1}\mathbf{Q}^\top\boldsymbol{\mathcal{E}}/(NT)\|) = o_p\left((NT)^{-1/2}\right).\end{aligned}$$

This, Assumption 4 and Lemma 3 imply that $\|D_{22}\| = o_p((NT)^{-1/2})$. Then, $\|D_2\| = \Sigma_{N,Z}^{-1}\boldsymbol{\xi}^\top\boldsymbol{\mathcal{E}}/(NT) + o_p((NT)^{-1/2}) = O_p((NT)^{-1/2})$. This completes the proof.

(ii) From the under smoothing condition $(NT)^{1/2}(J_1^{-\mu} + J_2^{-\mu}) = o(1)$ and $\|D_{22}\| = o_p((NT)^{-1/2})$, 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

$$a_{NT} = \boldsymbol{c}^\top\Omega_{N,Z}^{-1/2}\boldsymbol{\xi}^\top\boldsymbol{\mathcal{E}}/\sqrt{NT}. \tag{32}$$

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

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

$$\begin{aligned}\mathcal{F}_{N,i} &= \sigma(\{Y_{j0}\}_{j=1}^N, \{\epsilon_{j1}\}_{j=1}^i), \\ \mathcal{F}_{N,N+i} &= \sigma(\{Y_{j0}\}_{j=1}^N, \{\epsilon_{j1}^o\}_{j=1}^N, \{\epsilon_{j2}\}_{j=1}^i), \\ &\vdots \\ \mathcal{F}_{N,(T-1)N+i} &= \sigma(\{Y_{j0}\}_{j=1}^N, \{\epsilon_{j,T-1}^o\}_{j=1}^N, \{\epsilon_{jT}\}_{j=1}^i),\end{aligned} \tag{33}$$

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 $\mathcal{F}_{N,v} \subset \mathcal{F}_{N,v+1}$, 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 $\{X_{N,v}, \mathcal{F}_{N,v}, 1 \le v \le NT, N \ge 1\}$ is a martingale difference array. Then, it suffices to check the following two conditions for the central limit theorem of (Kuersteiner and Prucha, 2013, Theorem 1):

$$\sum_{v=1}^{k_N} \mathbb{E}[|X_{N,v}|^{2+\delta}] \to 0 \tag{34}$$

for some $\delta > 0$ and

$$\sum_{v=1}^{k_N} \mathbb{E}[X_{N,v}^2 \mid \mathcal{F}_{N,v-1}] \xrightarrow{p} 1. \tag{35}$$

For verification of condition (34), 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 $X_{N2,v} = \boldsymbol{c}^\top \Omega_{N,Z}^{-1/2} C_{ZJ} \Psi_J^{-1} Q_{i,t-1}\epsilon_{it}/\sqrt{NT}$. By Assumptions 3 and 4, we have

$$\begin{aligned}
\mathbb{E}\big(X_{N2,v}^4 \mid \mathcal{F}_{N,v-1}\big) =& \frac{1}{(NT)^2} \mathbb{E}\Big[\big(\boldsymbol{c}^\top \Omega_{N,Z}^{-1/2} C_{ZJ} \Psi_J^{-1} Q_{i,t-1}\epsilon_{it}\big)^4 \big| \mathcal{F}_{N,v-1}\Big] \\
=& \frac{1}{(NT)^2} \big(\boldsymbol{c}^\top \Omega_{N,Z}^{-1/2} C_{ZJ} \Psi_J^{-1} Q_{i,t-1}\big)^4 \cdot \mathbb{E}(\epsilon_{it}^4) \\
\le& \frac{C}{(NT)^2 \underline{\nu}_{N,J_1J_2}^4} \|Q_{i,t-1}\|^4.
\end{aligned}$$

Further, 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

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

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: $\sum_{v=1}^{k_N} \mathbb{E}\big(X_{N,v}^4\big) = \sum_{v=1}^{k_N} \mathbb{E}|X_{N1,v} - X_{N2,v}|^4 \le 8\sum_{v=1}^{k_N} \mathbb{E}|X_{N1,v}|^4 + 8\sum_{v=1}^{k_N} \mathbb{E}|X_{N2,v}|^4 = o(1)$. Thus condition (34) is met.

Next, observe that

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

Using Lemma 2 and Assumptions 4-5, we can easily see that $\big\|1/(NT)\sum_{t=1}^T (\mathbb{Z}^\top - C_{ZJ}\Psi_J^{-1}\mathbb{Q}_{t-1}^\top)(\mathbb{Z} - \mathbb{Q}_{t-1}\Psi_J^{-1}C_{ZJ}^\top) - \Sigma_{N,Z}\big\| = o_p(1)$. 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 (35). □

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

$$\left|p^{J_\ell}(\cdot)^\top \mathcal{S}_\ell[\mathbf{Q}^\top \mathscr{M}(\mathbf{Z})\mathbf{Q}]^- \mathbf{Q}^\top \mathscr{M}(\mathbf{Z})\boldsymbol{\mathcal{E}}\right|_{\infty,\omega} = O_p\left(\zeta_{0,J_\ell}\sqrt{\frac{\ln J_\ell}{\underline{\nu}_{N,J_1J_2}NT}}\right), \tag{36}$$

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

$$\left[\mathbf{Q}^\top \mathscr{M}(\mathbf{Z})\mathbf{Q}\right]^- \mathbf{Q}^\top \mathscr{M}(\mathbf{Z})\boldsymbol{\mathcal{E}} = U_{NT,1} + U_{NT,2},$$

where

$$U_{NT,1} \equiv \Sigma_{N,f}^{-1}\left(\mathbf{Q}^\top\boldsymbol{\mathcal{E}} - C_{N,ZJ}^\top \Psi_{N,Z}^{-1}\mathbf{Z}^\top\boldsymbol{\mathcal{E}}\right)/(NT), \tag{37}$$

$$U_{NT,2} \equiv \left[(\widehat{\Sigma}_{NT,f}^- - \Sigma_{N,f}^{-1})\mathbf{Q}^\top + \left(\widehat{\Sigma}_{NT,f}^-\widehat{C}_{NT,ZJ}^\top\widehat{\Psi}_{N,Z}^{-1} - \Sigma_{N,f}^{-1}C_{N,ZJ}^\top\Psi_{N,Z}^{-1}\right)\mathbf{Z}^\top\right]\boldsymbol{\mathcal{E}}/(NT), \tag{38}$$

with $\Sigma_{N,f} = \Psi_{N,J} - C_{N,ZJ}^\top\Psi_{N,Z}^{-1}C_{N,ZJ}$, and $\widehat{\Sigma}_{NT,f} = \mathbf{Q}^\top\mathscr{M}(\mathbf{Z})\mathbf{Q}/(NT)$. Further, denote $\widetilde{U}_{NT,1} \equiv \mathcal{S}_1 U_{NT,1}$ and $\widetilde{U}_{NT,2} \equiv \mathcal{S}_1 U_{NT,2}$. For the sake of notational simplicity, we will suppress the dependence of $U_{NT,1}$, $U_{NT,2}$, $\widetilde{U}_{NT,1}$ and $\widetilde{U}_{NT,2}$ on N and T in the sequel.

We first examine $|p^{J_1}(\cdot)^\top\widetilde{U}_2|_{\infty,\omega}$. Notice that

$$U_2 = (\widehat{\Sigma}_{NT,f}^- - \Sigma_{N,f}^{-1})\mathbf{H}^\top\boldsymbol{\mathcal{E}}/(NT) - \widehat{\Sigma}_{NT,f}^-\left(\widehat{C}_{NT,ZJ}^\top\widehat{\Psi}_{N,Z}^{-1} - C_{N,ZJ}^\top\Psi_{N,Z}^{-1}\right)\mathbf{Z}^\top\boldsymbol{\mathcal{E}}/(NT),$$

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 $\mathbb{E}\|\mathbf{H}^\top\boldsymbol{\mathcal{E}}/(NT)\|^2 = (NT)^{-1}\mathrm{tr}(\mathbb{E}[\mathbb{H}_{t-1}^\top\mathcal{E}_t\mathcal{E}_t^\top\mathbb{H}_{t-1}/N]) = \sigma^2(NT)^{-1}\mathrm{tr}(\Sigma_{N,f}) = O(J/(NT))$ by Assumption 4. It follows that $\|\mathbf{H}^\top\boldsymbol{\mathcal{E}}/(NT)\| = O_p(\sqrt{J/(NT)})$. According to Lemma 3(ii),

$$\begin{aligned}\left\|\mathcal{S}_1(\widehat{\Sigma}_{NT,f}^- - \Sigma_{N,f}^{-1})\mathbf{H}^\top\boldsymbol{\mathcal{E}}/(NT)\right\| &\leq \|\widehat{\Sigma}_{NT,f}^- - \Sigma_{N,f}^{-1}\|\left\|\mathbf{H}^\top\boldsymbol{\mathcal{E}}/(NT)\right\| \\ &= O_p(\underline{\nu}_{N,J_1J_2}^{-2}(N^{-1/2} + r_{NT,J}^{1/2})\sqrt{J/(NT)}).\end{aligned}$$

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

$$\begin{aligned}&\left\|\mathcal{S}_1\widehat{\Sigma}_{NT,f}^-(\widehat{C}_{NT,ZJ}\widehat{\Psi}_{N,Z}^{-1} - C_{N,ZJ}\Psi_{N,Z}^{-1})\mathbf{Z}^\top\boldsymbol{\mathcal{E}}/(NT)\right\| \\ &= O_p(\underline{\nu}_{N,J_1J_2}^{-1}(N^{-1/2} + \zeta_{0,J}^{-1}r_{NT,J}^{1/2})/\sqrt{NT}).\end{aligned}$$

A combination of above results yields $\|\mathcal{S}_1U_2\| = \|\widetilde{U}_2\| = O_p(\underline{\nu}_{N,J_1J_2}^{-2}(N^{-1/2} +$

$r_{NT,J}^{1/2})\sqrt{J/(NT)})$. As $\sup_{y\in\mathcal{R}_\mathcal{Y}}\|p^{J_1}(y)\| = O(\zeta_{0,J_1})$, we have

$$\left|p^{J_1}(\cdot)^\intercal \widetilde{U}_2\right|_{\infty,\omega} = O_p\left(\underline{\nu}_{N,J_1J_2}^{-2}\zeta_{0,J_1}(N^{-1/2}+r_{NT,J}^{1/2})\sqrt{J/(NT)}\right) = O_p\left(\zeta_{0,J_1}\sqrt{\frac{\ln J_1}{\underline{\nu}_{N,J_1J_2}NT}}\right) \tag{39}$$

by Assumption (c) in Theorem 2(ii).

Next, $\mathbb{E}\|\widetilde{U}_1\|^2 = \sigma^2\mathrm{tr}(\mathcal{S}_1\Sigma_{N,f}^{-1}\mathbb{E}[\mathbf{H}^\intercal\mathbf{H}/(NT)]\Sigma_{N,f}^{-1}\mathcal{S}_1^\intercal/(NT)) = \sigma^2\mathrm{tr}(\mathcal{S}_1\Sigma_{N,f}^{-1}\Sigma_{N,f}\Sigma_{N,f}^{-1}\mathcal{S}_1^\intercal)/(NT) = \sigma^2\mathrm{tr}(\Sigma_{N,f_1}^{-1})/(NT) = O(J_1/(\underline{\nu}_{N,J_1J_2}NT))$ by Assumption 4, whence $\|\widetilde{U}_1\| = O_p(\sqrt{J_1/(\underline{\nu}_{N,J_1J_2}NT)})$. Let c_{NT} be a sequence of positive constants tending to 1 at a 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

$$\begin{aligned}
\left|p^{J_1}(\cdot)^\intercal \widetilde{U}_1\right|_{\infty,\omega} &\leq \sup_{y\in\mathcal{R}_\mathcal{Y}}\left|p^{J_1}(y)^\intercal\widetilde{U}_1\mathbf{1}(|y|\leq c_{NT})\cdot(1+|y|^2)^{-\omega/2}\right| \\
&\quad + \sup_{y\in\mathcal{R}_\mathcal{Y}}\left|p^{J_1}(y)^\intercal\widetilde{U}_1\mathbf{1}(|y|> c_{NT})\cdot(1+|y|^2)^{-\omega/2}\right| \\
&\leq \sup_{y\in\mathcal{R}_\mathcal{Y}:|y|\leq c_{NT}}\left|p^{J_1}(y)^\intercal\widetilde{U}_1\right| + \sup_{y\in\mathcal{R}_\mathcal{Y}}\|p^{J_1}(y)\|\|\widetilde{U}_1\|\cdot\sup_{y:|y|>c_{NT}}(1+|y|^2)^{-\omega/2} \\
&= \sup_{y\in\mathcal{R}_\mathcal{Y}:|y|\leq c_{NT}}\left|p^{J_1}(y)^\intercal\widetilde{U}_1\right| + \underbrace{O_p\left(\frac{\zeta_{0,J_1}\sqrt{J_1}}{c_{NT}^\omega\sqrt{\underline{\nu}_{N,J_1J_2}NT}}\right)}_{=O_p(\zeta_{0,J_1}/\sqrt{\underline{\nu}_{N,J_1J_2}NT})}.
\end{aligned} \tag{40}$$

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 assumption (a) in Theorem 2(ii). Then, similar to Lemma A.9 of Hoshino (2022), we have

$$\begin{aligned}
\sup_{y\in\mathcal{R}_\mathcal{Y}:|y|\leq c_{NT}}\left|p^{J_1}(y)^\intercal\widetilde{U}_1\right| &\leq \max_{a\in\mathcal{T}_{NT}}|p^{J_1}(a)^\intercal\widetilde{U}_1| + \max_{y\in\mathcal{R}_\mathcal{Y}:|y|\leq c_{NT}}|(p^{J_1}(y)-p^{J_1}(a_y))^\intercal\widetilde{U}_1| \\
&\leq \max_{a\in\mathcal{T}_{NT}}|p^{J_1}(a)^\intercal\widetilde{U}_1| + O_p(1)\cdot\|\widetilde{U}_1\| \\
&= \max_{a\in\mathcal{T}_{NT}}|p^{J_1}(a)^\intercal\widetilde{U}_1| + O_p\left(\sqrt{\frac{J_1}{\underline{\nu}_{N,J_1J_2}NT}}\right).
\end{aligned} \tag{41}$$

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

$$\epsilon_{1,it} \equiv \epsilon_{it}\mathbf{1}\{|\epsilon_{it}|\leq M_{NT}\} - \mathbb{E}[\epsilon_{it}\mathbf{1}\{|\epsilon_{it}|\leq M_{NT}\}],$$

$$\epsilon_{2,it} \equiv \epsilon_{it}\mathbf{1}\{|\epsilon_{it}|> M_{NT}\} - \mathbb{E}[\epsilon_{it}\mathbf{1}\{|\epsilon_{it}|> M_{NT}\}],$$

and M_{NT} is a sequence of positive constants tending to ∞. Let $\mathcal{E}_{1,t} = (\epsilon_{1,1t}, \ldots, \epsilon_{1,Nt})^\intercal$, $\mathcal{E}_{2,t} = (\epsilon_{2,1t}, \ldots, \epsilon_{2,Nt})^\intercal$, $\boldsymbol{\mathcal{E}}_1 = (\mathcal{E}_{1,1}^\intercal, \ldots, \mathcal{E}_{1,T}^\intercal)^\intercal$, and $\boldsymbol{\mathcal{E}}_2 = (\mathcal{E}_{2,1}^\intercal, \ldots, \mathcal{E}_{2,T}^\intercal)^\intercal$, so that $\widetilde{U}_1 = \mathcal{S}_1 \Sigma_{N,f}^{-1} \mathbf{H}^\intercal (\boldsymbol{\mathcal{E}}_1 + \boldsymbol{\mathcal{E}}_2)/(NT)$. Further, let $\xi_{N,v}(a) \equiv q_{N,i,t-1}(a)\epsilon_{1,it}$, where $v = (t-1)N + i$ and $q_{N,i,t-1}(a) \equiv p^{J_1}(a)^\intercal \mathcal{S}_1 \Sigma_{N,f}^{-1}(Q_{i,t-1} - C_{N,ZJ}^\intercal \Psi_{N,Z}^{-1} Z_i)/(NT)$. Then

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

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 (33). Furthermore, it is straightforward to see that there exists a positive constant c_1 such that

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

Besides, it is easy to see that $\sum_{v=1}^{k_N} \mathbb{E}[|\xi_{N,v}(a)|^2 | \mathcal{F}_{N,v-1}] = \sigma_1^2 \sum_{i=1}^{N}\sum_{t=1}^{T} |q_{N,i,t-1}(a)|^2 = \sigma_1^2 p^{J_1}(a)^\intercal \mathcal{S}_1 \Sigma_{N,f}^{-1} \mathbf{H}^\intercal \mathbf{H} \Sigma_{N,f}^{-1} \mathcal{S}_1^\intercal p^{J_1}(a)/(NT)^2$ with $\sigma_1^2 \equiv \mathrm{Var}(\epsilon_{1,it})$. Recall that $\mathbb{H}_{t-1} = \mathbb{Q}_{t-1} - \mathbb{Z}\Psi_{N,Z}^{-1} C_{N,ZJ}$. According to Lemma 2 and Assumption 4, we have $\mathbb{E}\big\|(NT)^{-1}\mathbf{H}^\intercal\mathbf{H} - \Sigma_{N,f}\big\|^2 = \mathbb{E}\big\|(NT)^{-1}\sum_{t=1}^{T}\big(\mathbb{H}_{t-1}^\intercal \mathbb{H}_{t-1} - \mathbb{E}\big(\mathbb{H}_{t-1}^\intercal \mathbb{H}_{t-1}\big)\big)\big\|^2 = O_p(r_{NT,J} + N^{-1})$. Therefore, $\|(NT)^{-1}\mathbf{H}^\intercal\mathbf{H} - \Sigma_{N,f}\| = O_p(r_{NT,J}^{1/2} + N^{-1/2})$ by Markov's inequality. As such, $|\sum_{v=1}^{k_N}\mathbb{E}[|\xi_{N,v}(a)|^2|\mathcal{F}_{N,v-1}] - \sigma_1^2 p^{J_1}(a)^\intercal \Sigma_{N,f_1}^{-1} p^{J_1}(a)/(NT)| = (NT)^{-1}\sigma_1^2\big|p^{J_1}(a)^\intercal \mathcal{S}_1\Sigma_{N,f}^{-1}[(NT)^{-1}\mathbf{H}^\intercal\mathbf{H} - \Sigma_{N,f}]\Sigma_{N,f}^{-1}\mathcal{S}_1^\intercal p^{J_1}(a)\big| \leq C\underline{\nu}_{J_1J_2}^{-2}\|p^{J_1}(a)\|^2\|(NT)^{-1}\mathbf{H}^\intercal\mathbf{H} - \Sigma_{N,f}\|/(NT) = O_p(\underline{\nu}_{J_1J_2}^{-2}\zeta_{0,J_1}^2(r_{NT,J}^{1/2} + N^{-1/2})/(NT))$. By triangular inequality, we have

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

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

3.14), we have

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

for some positive constant $\tilde{L}_1$ (depending on η). Then,

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

where $|\mathcal{T}_{NT}|$ denotes the cardinality of the set $\mathcal{T}_{NT}$, and $\mathscr{V}_{NT}$ denotes the event $\{\max_{a\in\mathcal{T}_{NT}}\sum_{v=1}^{k_N}\mathbb{E}[|\xi_{N,v}(a)|^2|\mathcal{F}_{N,v-1}] > L_1\zeta_{0,J_1}^2/(NT\underline{\nu}_{N,J_1J_2})\}$. Then, let $e_{NT} = L_2\zeta_{0,J_1}\sqrt{\ln J_1/(NT\underline{\nu}_{N,J_1J_2})}$ for a large constant $L_2 > 0$. As long as M_{NT} grows sufficiently slowly so that $e_{NT}M_{NT}\zeta_{0,J}/\zeta_{0,J_1} = o(1)$, we have $\ln|\mathcal{T}_{NT}| - e_{NT}^2/[\tilde{L}_1(\zeta_{0,J_1}^2/(\underline{\nu}_{N,J_1J_2}NT))(1+e_{NT}M_{NT}\zeta_{0,J}/\zeta_{0,J_1})] \asymp \ln\big(|\mathcal{T}_{NT}|/J_1^{L_2^2}\big)$. Note that the cardinality of $\mathcal{T}_{NT}$ grows at the rate $J_1^{1/(2\omega)+\xi}$. Thus, for sufficiently large L_2, we have $|\mathcal{T}_{NT}|/J_1^{L_2^2} \to 0$. Recall that for any $\eta > 0$, $\mathbb{P}(\mathscr{V}_{NT}) < \eta$ for sufficiently large N. Combining the above results yields

$$\max_{a\in\mathcal{T}_{NT}}\Big|\sum_{v=1}^{k_N}\xi_{N,v}(a)\Big| = \max_{a\in\mathcal{T}_{NT}}\Big|p^{J_1}(a)^\top\mathcal{S}_1\Sigma_{N,f}^{-1}\mathbf{H}^\top\boldsymbol{\mathcal{E}}_1/(NT)\Big| = O_p\left(\frac{\zeta_{0,J_1}\sqrt{\ln J_1}}{\sqrt{\underline{\nu}_{N,J_1J_2}NT}}\right). \tag{42}$$

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

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

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

$$
\max_{a\in\mathcal{T}_{NT}}\left|p^{J_1}(a)^{\mathsf{T}}\mathcal{S}_1\Sigma_{N,f}^{-1}\mathbf{H}^{\mathsf{T}}\boldsymbol{\mathcal{E}}_2/(NT)\right|=O_p\left(\frac{\zeta_{0,J_1}\sqrt{\ln J_1}}{\sqrt{\underline{\nu}_{N,J_1J_2}NT}}\right). \tag{43}
$$

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 $\zeta_{0,J}\sqrt{NT}/\sqrt{\underline{\nu}_{N,J_1J_2}\ln J_1}=O(M_{NT}^3)$, can be satisfied simultaneously under Assumption (b) of Theorem 2(ii). A combination of (40)-(43) yields

$$
\left|p^{J_1}(\cdot)^{\mathsf{T}}\widetilde{U}_1\right|_{\infty,\omega}=O_p\left(\frac{\zeta_{0,J_1}\sqrt{\ln J_1}}{\sqrt{\underline{\nu}_{N,J_1J_2}NT}}\right)+O_p\left(\frac{\zeta_{0,J_1}+J_1^{1/2}}{\sqrt{\underline{\nu}_{N,J_1J_2}NT}}\right). \tag{44}
$$

Lemma 4 with $\ell=1$ immediately follows from (39) and (44). The case where $\ell=2$ can be similarly derived, and we omit the details. □

B.3 Proof of Theorem 2

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

$$
\widehat{\beta}-\beta=\left[\mathbf{Q}^{\mathsf{T}}\mathscr{M}(\mathbf{Z})\mathbf{Q}\right]^{-}\mathbf{Q}^{\mathsf{T}}\mathscr{M}(\mathbf{Z})\mathbf{Y}-\beta=U_1+U_2+U_3, \tag{45}
$$

where U_1 and U_2 are defined in (37) and (38), and $U_3 \equiv \widehat{\Sigma}_{NT,f}^{-}\mathbf{Q}^{\top}\mathscr{M}(\mathbf{Z})[\mathbf{G}(\mathbf{Y}_{-1})-\mathbf{Q}\beta]/(NT)$. It is shown in the proof of Lemma 4 that $\|\mathcal{S}_1 U_1\| = O_p(\sqrt{J_1/(\underline{\nu}_{N,J_1J_2}NT)})$ and $\|\mathcal{S}_1 U_2\| = O_p(\underline{\nu}_{N,J_1J_2}^{-2}(N^{-1/2}+r_{NT,J}^{1/2})\sqrt{J/(NT)}) = O_p(\sqrt{J_1/(\underline{\nu}_{N,J_1J_2}NT)})$ under the stated conditions. We also have that the bias term $\|U_3\| = O_p(\underline{\nu}_{N,J_1J_2}^{-1/2}(J_1^{-\mu}+J_2^{-\mu}))$. To see this, note that according to Lemma 3, $\|\widehat{\Sigma}_{NT,f}-\Sigma_{N,f}\| = O_p(N^{-1/2}+r_{NT,J}^{1/2}) = o_p(1)$, which implies that $\lambda_{\min}(\widehat{\Sigma}_{NT,f}) \geq c\underline{\nu}_{N,J_1J_2} > 0$ for some $0 < c < \infty$ with probability approaching 1. Moreover, it is straightforward to see that

$$\begin{aligned}\|U_3\|^2 &\leq O_p(\underline{\nu}_{N,J_1J_2}^{-1})\cdot[\mathbf{G}(\mathbf{Y}_{-1})-\mathbf{Q}\beta]^{\top}\mathscr{M}(\mathbf{Z})\mathbf{Q}\widehat{\Sigma}_{NT,f}^{-}\mathbf{Q}^{\top}\mathscr{M}(\mathbf{Z})[\mathbf{G}(\mathbf{Y}_{-1})-\mathbf{Q}\beta]/(NT)^2\\ &\leq O_p(\underline{\nu}_{N,J_1J_2}^{-1})\cdot\left\|\mathbf{G}(\mathbf{Y}_{-1})-\mathbf{Q}\beta\right\|^2/(NT)\\ &= O_p(\underline{\nu}_{N,J_1J_2}^{-1}(J_1^{-2\mu}+J_2^{-2\mu})).\end{aligned} \tag{46}$$

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

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,

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

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 convergence rate of $\widehat{f}_2$ can be shown similarly, and the details are omitted.

(ii) From (45) and Lemma 4, we have

$$\begin{aligned}\left|\widehat{f}_1(\cdot)-f_1(\cdot)\right|_{\infty,\omega} &\leq \left|p^{J_1}(\cdot)^{\top}(\widehat{\beta}_1-\beta_1)\right|_{\infty,\omega}+\left|p^{J_1}(\cdot)^{\top}\beta_1-f_1(\cdot)\right|_{\infty,\omega}\\ &\leq \left|p^{J_1}(\cdot)^{\top}\mathcal{S}_1U_1\right|_{\infty,\omega}+\left|p^{J_1}(\cdot)^{\top}\mathcal{S}_1U_2\right|_{\infty,\omega}+\left|p^{J_1}(\cdot)^{\top}\mathcal{S}_1U_3\right|_{\infty,\omega}+O(J_1^{-\mu})\\ &\leq O_p\left(\zeta_{0,J_1}\sqrt{\ln J_1/(\underline{\nu}_{N,J_1J_2}NT)}\right)+\left|p^{J_1}(\cdot)^{\top}\mathcal{S}_1U_3\right|_{\infty,\omega}+O(J_1^{-\mu}).\end{aligned}$$

Besides, $\left|p^{J_1}(\cdot)^{\top}\mathcal{S}_1U_3\right|_{\infty,\omega} \leq \left|p^{J_1}(\cdot)\right|_{\infty}\|U_3\| = O_p(\zeta_{0,J_1}\underline{\nu}_{N,J_1J_2}^{-1/2}(J_1^{-\mu}+J_2^{-\mu}))$. 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) Noting that $\lambda_{\max}(\Sigma_{N,f}) \leq \bar{c} < \infty$, we have $v_{1N}^2(y) = \sigma^2 p^{J_1}(y)^\intercal \mathcal{S}_1 \Sigma_{N,f}^{-1} \mathcal{S}_1^\intercal p^{J_1}(y) \geq \sigma^2 \|p^{J_1}(y)\|^2 \lambda_{\min}(\mathcal{S}_1 \Sigma_{N,f}^{-1} \mathcal{S}_1^\intercal) \geq \sigma^2 \|p^{J_1}(y)\|^2 \lambda_{\min}(\Sigma_{N,f}^{-1}) \geq \sigma^2 \bar{c}^{-1} \|p^{J_1}(y)\|^2 > 0$. From the proof of part (i), we have

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

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

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

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 $i = 1, \ldots, N$,

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

In a similar manner as 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 (33). It remains to check the conditions (34) and (35). To verify condition (34), note that

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

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

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

Hence, $\sum_{v=1}^{NT} \mathbb{E}\big(X_{N,v}^4\big) = O(\zeta_{0,J}^2 J/(\underline{\nu}_{N,J_1J_2}^4 NT)) = o(1)$. This verifies condition (34).

To verify condition (35), observe that

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

By Lemma 2 and Assumption 4, we have $\|(NT)^{-1}\sum_{t=1}^T \mathbb{H}_{t-1}^\intercal \mathbb{H}_{t-1} - \Sigma_{N,f}\| = \|(NT)^{-1}\sum_{t=1}^T \big(\mathbb{H}_{t-1}^\intercal \mathbb{H}_{t-1} - \mathbb{E}(\mathbb{H}_{t-1}^\intercal \mathbb{H}_{t-1})\big)\| = O_p(1/\sqrt{N} + \sqrt{r_{NT,J}}) = o_p(1)$, implying that $\sum_{v=1}^{k_N} \mathbb{E}\big[\tilde{X}_{N,v}^2 | \mathcal{F}_{N,v-1}\big] - 1 = \sigma^2 v_{1N}^{-2}(y) p^{J_1}(y)^\intercal \mathcal{S}_1 \Sigma_{N,f}^{-1}[(NT)^{-1}\sum_{t=1}^T \mathbb{H}_{t-1}^\intercal \mathbb{H}_{t-1} - \Sigma_{N,f}]\Sigma_{N,f}^{-1}\mathcal{S}_1^\intercal p^{J_1}(y) = O_p(\underline{\nu}_{N,J_1J_2}^{-2}(N^{-1/2} + r_{NT,J}^{1/2})) = o_p(1)$. Thus, condition (35) is met. □

B.4 Proof of Theorem 3

(i) By Assumption 4′, we have

$$\mathbf{s}_{N,1}^2 \ge c \cdot \|\Sigma_{f_1}^{-1}\|^2 \ge c' J_1. \tag{47}$$

Decompose $\mathbf{T}_{NT,1}$ into three parts as follows:

$$\begin{aligned}
\mathbf{T}_{NT,1} &= \sum_{i=1}^N \sum_{t=1}^T \big(\widehat{f}_1(Y_{i,t-1}) - f_1(Y_{i,t-1}) + f_1(Y_{i,t-1}) - \widehat{\rho}_1 Y_{i,t-1}\big)^2 \\
&= \sum_{i=1}^N \sum_{t=1}^T \Big[\big(\widehat{f}_1(Y_{i,t-1}) - f_1(Y_{i,t-1})\big)^2 + \big(f_1(Y_{i,t-1}) - \widehat{\rho}_1 Y_{i,t-1}\big)^2 \\
&\qquad + 2\big(\widehat{f}_1(Y_{i,t-1}) - f_1(Y_{i,t-1})\big)\big(f_1(Y_{i,t-1}) - \widehat{\rho}_1 Y_{i,t-1}\big)\Big] \\
&\equiv \mathbf{T}_{NT,1a} + \mathbf{T}_{NT,1b} + 2\mathbf{T}_{NT,1c}.
\end{aligned} \tag{48}$$

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 (47). In Lemma 5 below, it will be proven 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 (37) and (38), respectively, U_3 is defined below (45), and $u_1(y) \equiv f_1(y) - p^{J_1}(y)^\top \beta_1$. In addition, $\|\widetilde{U}_1\| = O_p(\sqrt{J_1/(\underline{\nu}_{N,J_1J_2}NT)})$, $\|\widetilde{U}_2\| = O_p(\underline{\nu}_{N,J_1J_2}^{-2}(r_{NT,J}^{1/2} + N^{-1/2})\sqrt{J/(NT)})$, and $\|\widetilde{U}_3\| = O_p(\underline{\nu}_{N,J_1J_2}^{-1/2}(J_1^{-\mu} + J_2^{-\mu}))$, as shown in the proof of Lemma 4 and Theorem 2(i). 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

$$\begin{aligned}
\mathbf{T}_{NT,1a} = &\widetilde{U}_1^\top(\mathbf{P}^\top\mathbf{P})\widetilde{U}_1 + \underbrace{\widetilde{U}_2^\top(\mathbf{P}^\top\mathbf{P})\widetilde{U}_2}_{=O_p(\underline{\nu}_{N,J_1J_2}^{-4}(N^{-1}+r_{NT,J})J)} + \underbrace{\widetilde{U}_3^\top(\mathbf{P}^\top\mathbf{P})\widetilde{U}_3}_{=O_p(NT\underline{\nu}_{N,J_1J_2}^{-1}(J_1^{-2\mu}+J_2^{-2\mu}))} \\
&+ \underbrace{2\widetilde{U}_1^\top(\mathbf{P}^\top\mathbf{P})\widetilde{U}_2}_{=O_p((r_{NT,J}^{1/2}+N^{-1/2})\sqrt{J_1J/\underline{\nu}_{N,J_1J_2}^5})} + \underbrace{2\widetilde{U}_1^\top(\mathbf{P}^\top\mathbf{P})\widetilde{U}_3}_{=O_p(\sqrt{J_1NT/\underline{\nu}_{N,J_1J_2}^2}(J_1^{-\mu}+J_2^{-\mu}))} \\
&\underbrace{- 2\sum_{i=1}^N\sum_{t=1}^T p^{J_1}(Y_{i,t-1})^\top\widetilde{U}_1 \cdot u_1(Y_{i,t-1})}_{=O_p(\sqrt{NT}\underline{\nu}_{N,J_1J_2}^{-1/2}\zeta_{0,J_1}J_1^{1/2-\mu})} + \underbrace{2\widetilde{U}_2^\top(\mathbf{P}^\top\mathbf{P})\widetilde{U}_3}_{=O_p(\sqrt{JNT/\underline{\nu}_{N,J_1J_2}^5}(N^{-1/2}+r_{NT,J}^{1/2})(J_1^{-\mu}+J_2^{-\mu}))} \\
&- \underbrace{2\sum_{i=1}^N\sum_{t=1}^T p^{J_1}(Y_{i,t-1})^\top\widetilde{U}_2 \cdot u_1(Y_{i,t-1})}_{=O_p(\sqrt{NT}\underline{\nu}_{N,J_1J_2}^{-2}\zeta_{0,J_1}(N^{-1/2}+r_{NT,J}^{1/2})J^{1/2}J_1^{-\mu})} + O_p(NT\zeta_{0,J_1}J_1^{-\mu}(J_1^{-\mu} + J_2^{-\mu})/\sqrt{\underline{\nu}_{N,J_1J_2}}).
\end{aligned}$$

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,

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

by (47), and Theorem 3(i) follows from Lemma 6.

(ii) Note that $\mathbf{Y} - \mathbf{X}_1\bar{\theta}_1 = \mathscr{M}(\mathbf{X}_1)\mathbf{Y} = \mathscr{M}(\mathbf{X}_1)(\mathbf{X}_1\theta_1 + \mathbf{X}_2\beta_{1,2} + \mathbf{R} + \boldsymbol{\mathcal{E}})$, 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 $\mathbf{Y} - \mathbf{X}_1\bar{\theta}_1 = \mathscr{M}(\mathbf{X}_1)(\mathbf{R}_2 + \boldsymbol{\mathcal{E}})$, 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

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

Denote

$$\mathbf{LM}_{NT,a} \equiv \boldsymbol{\mathcal{E}}^\intercal \mathscr{M}(\mathbf{X}_1)\mathbf{X}_2\big(\mathbf{X}_2^\intercal \mathscr{M}(\mathbf{X}_1)\mathbf{X}_2\big)^- \mathbf{X}_2^\intercal \mathscr{M}(\mathbf{X}_1)\boldsymbol{\mathcal{E}}/\widehat{\sigma}^2.$$

We first show that

$$\frac{\mathbf{LM}_{NT,a} - (J_1 - 1)}{\sqrt{2(J_1 - 1)}} \xrightarrow{d} N(0,1). \tag{49}$$

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

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

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),

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

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

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

Following the lines of Lemmas 2-3, we can easily find that $\|[\mathbf{X}_2^\top \mathbf{X}_1(\mathbf{X}_1^\top \mathbf{X}_1)^{-1} - \Psi_{N,12}^\top \Psi_{N,11}^{-1}]\mathbf{X}_1^\top \boldsymbol{\mathcal{E}}\| \leq \|\mathbf{X}_2^\top \mathbf{X}_1(\mathbf{X}_1^\top \mathbf{X}_1)^{-1} - \Psi_{N,12}^\top \Psi_{N,11}^{-1}\| \cdot \|\mathbf{X}_1^\top \boldsymbol{\mathcal{E}}\| = O_p(\underline{\nu}_{N,J_1 J_2}^{-2}(N^{-1/2} + r_{NT,J}^{1/2})\sqrt{NTJ_2})$. Also,

$$
\begin{aligned}
&\|(\mathbf{X}_2^\top \mathscr{M}(\mathbf{X}_1)\mathbf{X}_2)^{-1}\boldsymbol{\nu}^\top \boldsymbol{\mathcal{E}}\| \\
&\leq \|(\mathbf{X}_2^\top \mathscr{M}(\mathbf{X}_1)\mathbf{X}_2/(NT))^{-1} - \Theta_N^{-1}\|\|\boldsymbol{\nu}^\top \boldsymbol{\mathcal{E}}/(NT)\| + \|\Theta_N^{-1}\boldsymbol{\nu}^\top \boldsymbol{\mathcal{E}}/(NT)\| \\
&= O_p\left(\underline{\nu}_{N,J_1 J_2}^{-2}(r_{NT,J}^{1/2} + N^{-1/2})\sqrt{J_1/(NT)} + \sqrt{J_1/(\underline{\nu}_{N,J_1 J_2} NT)}\right) \\
&= O_p\big(\sqrt{J_1/(\underline{\nu}_{N,J_1 J_2} NT)}\big)
\end{aligned}
$$

by Assumption 5. A combination of above results gives $\|\mathbf{LM}_{NT,a}^{(3)}\| = O_p(\underline{\nu}_{N,J_1 J_2}^{-5/2}(r_{NT,J}^{1/2} + N^{-1/2})\sqrt{J_1 J_2} + \underline{\nu}_{N,J_1 J_2}^{-5}(r_{NT,J} + N^{-1})J) = o_p(J_1^{1/2})$ by Assumption 5′(i). Thus, the convergence in (49) holds.

Next, since $\mathscr{M}(\mathbf{X}_1)$ is idempotent,

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

by Assumption 6 and

$$
\begin{aligned}
|\mathbf{LM}_{NT,c}| \equiv& \left|\boldsymbol{\mathcal{E}}^\top \mathscr{M}(\mathbf{X}_1)\mathbf{X}_2\big(\mathbf{X}_2^\top \mathscr{M}(\mathbf{X}_1)\mathbf{X}_2\big)^{-1}\mathbf{X}_2^\top \mathscr{M}(\mathbf{X}_1)\mathbf{R}_2/\widehat{\sigma}^2\right| \\
\leq& \sqrt{\mathbf{LM}_{NT,a} \cdot \mathbf{LM}_{NT,b}} = O_p(\sqrt{J_1 NT} J_2^{-\mu}),
\end{aligned}
$$

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 (49). □

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 (45), we can write $\mathbf{T}_{NT,1c}$ as

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

$$
\begin{aligned}
&+ O_p\Big(\sqrt{NTJ_1/\underline{\nu}_{N,J_1J_2}}(J_1^{-\mu}+J_2^{-\mu})\Big)\\
&= (\rho_1-\widehat{\rho}_1)\bigg(\sum_{t=1}^T \mathbb{Y}_{t-1}^{\intercal}\mathbb{P}_{t-1}^{J_1}\bigg)\widetilde{U}_1 + o_p(J_1^{1/2}),
\end{aligned}
$$

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

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

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

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

implying that $|\tilde{U}_1^{(1)}| = O_p(J_1^{1/4}\sqrt{NT/\underline{\nu}_{N,J_1J_2}})$. Therefore, $(\widehat{\rho}_1-\rho_1)\tilde{U}_1^{(1)} = O_p(J_1^{1/4}\underline{\nu}_{N,J_1J_2}^{-1/2}) = o_p(J_1^{1/2})$ 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 $\{(Z_j^{\intercal},\epsilon_{j\tau})^{\intercal}:\Delta((i,t),(j,\tau))\le s\}$. In view of Assumption 1′ and (18), we have $\sup_{1\le i\le N}\|Y_{i,t-1}-\hat{Y}_{i,t-s}^{t-1}\|_4 \le C_1 d^{s-1}$ 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, $\mathbb{E}\|Y_{i,t-1}P_{i,t-1}\|^2\le C_2\zeta_{0,J_1}^2$ and

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

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

$$
|\tilde{U}_1^{(2)}| \le \left\|\frac{1}{NT}\sum_{t=1}^T\left[\mathbb{Y}_{t-1}^{\intercal}\mathbb{P}_{t-1}^{J_1}-\mathbb{E}\big(\mathbb{Y}_{t-1}^{\intercal}\mathbb{P}_{t-1}^{J_1}\big)\right]\right\|\cdot\left\|\mathcal{S}_1\Sigma_{N,f}^{-1}\big(\mathbf{Q}-\mathbf{Z}\Psi_{N,Z}^{-1}C_{N,ZJ}\big)^{\intercal}\boldsymbol{\mathcal{E}}\right\|
$$

$$= O_p\big(\sqrt{J_1\zeta_{0,J_1}^2/\underline{\nu}_{N,J_1J_2}} + \sqrt{J_1\zeta_{0,J_1}(\zeta_{0,J_1}+\zeta_{1,J_1})\min\{\Xi_{N,0},\Xi_{N,1}/T\}NT/\underline{\nu}_{N,J_1J_2}}\big). \quad (50)$$

This result, Assumptions 5 and 9 together imply that $(\widehat{\rho}_1 - \rho_1)\widetilde{U}_1^{(2)} = O_p\big(\zeta_{0,J_1}\sqrt{J_1/(\underline{\nu}_{N,J_1J_2}NT)}\big) + O_p\big(\sqrt{J_1\zeta_{0,J_1}(\zeta_{0,J_1}+\zeta_{1,J_1})\min\{\Xi_{N,0},T^{-1}\Xi_{N,1}\}/\underline{\nu}_{N,J_1J_2}}\big) = o_p(J_1^{1/2})$. 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,*

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

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 $H_{i,t-1} \equiv Q_{i,t-1} - C_{N,ZJ}^\intercal\Psi_{N,Z}^{-1}Z_i$, and $Q_{i,t-1} = (\bar{P}_{i,t-1}^{J_1\intercal}, P_{i,t-1}^{J_2\intercal})^\intercal$. For $i = 1,\ldots,N$ and $t = 1,\ldots,T$, denote

$$\pi_{N,(t-1)N+i} \equiv H_{i,t-1}\epsilon_{it} \quad (51)$$

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

$$\begin{aligned}\widetilde{U}_1^\intercal(\mathbf{P}^\intercal\mathbf{P})\widetilde{U}_1 &= \widetilde{U}_1^\intercal\mathbb{E}(\mathbf{P}^\intercal\mathbf{P})\widetilde{U}_1 + \widetilde{U}_1^\intercal\big[\mathbf{P}^\intercal\mathbf{P} - \mathbb{E}(\mathbf{P}^\intercal\mathbf{P})\big]\widetilde{U}_1 \\ &= \frac{2}{NT}\sum_{1\le u<v\le k_N}\pi_{N,u}^\intercal\Omega_{f_1}\pi_{N,v} + \frac{1}{NT}\sum_{v=1}^{k_N}\pi_{N,v}^\intercal\Omega_{f_1}\pi_{N,v} + \underbrace{O_p\big(\zeta_{0,J}r_{NT,J}J_1/\underline{\nu}_{N,J_1J_2}\big)}_{=o_p(J_1^{1/2})},\end{aligned}$$

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

$$\begin{aligned}\left|\frac{1}{NT}\sum_{v=1}^{k_N}\pi_{N,v}^\intercal\Omega_{f_1}\pi_{N,v} - \mathbf{B}_{N,1}\right| &= \left|\mathrm{tr}\left(\Omega_{f_1}\left(\frac{1}{NT}\sum_{v=1}^{k_N}\pi_{N,v}\pi_{N,v}^\intercal - \sigma^2\Sigma_{N,f}\right)\right)\right| \\ &\le \underbrace{\|\Omega_{f_1}\|}_{=O(\underline{\nu}_{N,J_1J_2}^{-2}J_1^{1/2})}\cdot\left\|\frac{1}{NT}\sum_{v=1}^{k_N}\pi_{N,v}\pi_{N,v}^\intercal - \sigma^2\Sigma_{N,f}\right\|.\end{aligned}$$

Also, it follows from Assumption 4, Lemmas 2 and 7 that $\|(NT)^{-1}\sum_{v=1}^{k_N}\pi_{N,v}\pi_{N,v}^\intercal - \sigma^2\Sigma_{N,f}\| = \|(NT)^{-1}\sum_{i=1}^N\sum_{t=1}^T H_{i,t-1}H_{i,t-1}^\intercal\epsilon_{it}^2 - \sigma^2\Sigma_{N,f}\| \le \|(NT)^{-1}\sum_{i=1}^N\sum_{t=1}^T H_{i,t-1}H_{i,t-1}^\intercal(\epsilon_{it}^2 - \sigma^2)\| + \|\sigma^2(NT)^{-1}\sum_{i=1}^N\sum_{t=1}^T[H_{i,t-1}H_{i,t-1}^\intercal - \mathbb{E}(H_{i,t-1}H_{i,t-1}^\intercal)]\| = O_p(\zeta_{0,J}\sqrt{J/NT} + N^{-1/2} + r_{NT,J}^{1/2})$. From this and (47), we have

$$\frac{(NT)^{-1}\sum_{v=1}^{k_N}\pi_{N,v}^\intercal\Omega_{f_1}\pi_{N,v} - \mathbf{B}_{N,1}}{\mathbf{s}_{N,1}} = O_p\left(\underline{\nu}_{N,J_1J_2}^{-2}\big(N^{-1/2} + r_{NT,J}^{1/2}\big)\right) = o_p(1) \quad (52)$$

by Assumption 5. Hence,

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

where $\zeta_{N,1} = 0$ and $\zeta_{N,v} \equiv (NT\mathbf{s}_{N,1})^{-1} 2\pi_{N,v}^{\mathsf{T}} \Omega_{f_1}(\sum_{l=1}^{v-1} \pi_{N,l})$ 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 conditions that are sufficient for the CLT to hold are

$$\sum_{v=1}^{k_N} \mathbb{E}\big[|\zeta_{N,v}|^{2+\eta}\big] \to 0, \tag{53}$$

for some $\eta > 0$, and

$$\sum_{v=1}^{k_N} \mathbb{E}\big[\zeta_{N,v}^2|\mathcal{F}_{N,v-1}\big] \xrightarrow{p} 1. \tag{54}$$

We first verify condition (53). Let $\eta = 2$. By the definition (51), Assumptions 1 and 4,

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

From Lemma 7(ii), $\mathbb{E}\big\| \sum_{j=1}^{i-1} H_{j,t-1}\epsilon_{jt} + \sum_{j=1}^{N}\sum_{s=1}^{t-1} H_{j,s-1}\epsilon_{js} \big\|^4 \leq C(NT)^2 J^2$. Combining these and (47), we have $\sum_{v=1}^{k_N} \mathbb{E}\{\mathbb{E}[\zeta_{N,v}^4|\mathcal{F}_{N,v-1}]\}$ is bounded by

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

which completes the verification of condition (53).

Next, observe that

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

To show (54), it suffices to show that

$$\frac{4}{(NT)^2}\sum_{v=1}^{NT}\mathrm{tr}\Bigg\{\Omega_{f_1}\Bigg(\sum_{u=1}^{v-1}\pi_{N,u}\Bigg)\Bigg(\sum_{u=1}^{v-1}\pi_{N,u}\Bigg)^{\mathsf{T}}\Omega_{f_1}\mathbb{E}[\pi_{N,v}\pi_{N,v}^{\mathsf{T}}|\mathcal{F}_{N,v-1}]\Bigg\}-\mathbf{s}_{N,1}^2=o_p(\mathbf{s}_{N,1}^2). \quad (55)$$

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

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

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

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

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

$$\begin{aligned}
Q_{NT,1a}&\equiv\frac{4}{(NT)^2}\sum_{1\le u<v\le NT}\mathrm{tr}\left\{\Omega_{f_1}\pi_{N,u}\pi_{N,u}^{\mathsf{T}}\Omega_{f_1}\left[\mathbb{E}[\pi_{N,v}\pi_{N,v}^{\mathsf{T}}|\mathcal{F}_{N,v-1}]-\mathbb{E}(\pi_{N,v}\pi_{N,v}^{\mathsf{T}})\right]\right\},\\
Q_{NT,1b}&\equiv\frac{8}{(NT)^2}\sum_{1\le q<u<v\le NT}\mathrm{tr}\left\{\Omega_{f_1}\pi_{N,u}\pi_{N,q}^{\mathsf{T}}\Omega_{f_1}\left[\mathbb{E}[\pi_{N,v}\pi_{N,v}^{\mathsf{T}}|\mathcal{F}_{N,v-1}]-\mathbb{E}(\pi_{N,v}\pi_{N,v}^{\mathsf{T}})\right]\right\}.
\end{aligned}$$

For $Q_{NT,1a}$, write

$$\begin{aligned}
Q_{NT,1a}=&\frac{4}{(NT)^2}\mathrm{tr}\left\{\Omega_{f_1}\sum_{u=1}^{NT}\big(\pi_{N,u}\pi_{N,u}^{\mathsf{T}}\big)\Omega_{f_1}\sum_{v=1}^{NT}\left[\mathbb{E}[\pi_{N,v}\pi_{N,v}^{\mathsf{T}}|\mathcal{F}_{N,v-1}]-\mathbb{E}(\pi_{N,v}\pi_{N,v}^{\mathsf{T}})\right]\right\} &(56)\\
&-\frac{4}{(NT)^2}\sum_{1\le u\le v\le NT}\mathrm{tr}\left\{\Omega_{f_1}\pi_{N,v}\pi_{N,v}^{\mathsf{T}}\Omega_{f_1}\left[\mathbb{E}[\pi_{N,u}\pi_{N,u}^{\mathsf{T}}|\mathcal{F}_{N,u-1}]-\mathbb{E}(\pi_{N,u}\pi_{N,u}^{\mathsf{T}})\right]\right\}. &(57)
\end{aligned}$$

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

$$\begin{aligned}
&\left|\frac{1}{(NT)^2}\sum_{u=1}^{NT}\sum_{v=1}^{NT}\operatorname{tr}\left\{\Omega_{f_1}\pi_{N,u}\pi_{N,u}^{\intercal}\Omega_{f_1}\left[\mathbb{E}[\pi_{N,v}\pi_{N,v}^{\intercal}|\mathcal{F}_{N,v-1}] - \mathbb{E}(\pi_{N,v}\pi_{N,v}^{\intercal})\right]\right\}\right| \\
&\leq \lambda_{\max}^2(\Omega_{f_1})\underbrace{\left\|\frac{1}{NT}\sum_{u=1}^{NT}\pi_{N,u}\pi_{N,u}^{\intercal}\right\|}_{=O_p(J)}\cdot\underbrace{\left\|\frac{1}{NT}\sum_{v=1}^{NT}\left[\mathbb{E}[\pi_{N,v}\pi_{N,v}^{\intercal}|\mathcal{F}_{N,v-1}] - \mathbb{E}(\pi_{N,v}\pi_{N,v}^{\intercal})\right]\right\|}_{=O_p(r_{NT,J}^{1/2}+N^{-1/2})}.
\end{aligned}$$

Therefore, the term (56) is $O_p(J(N^{-1/2} + r_{NT,J}^{1/2})/\underline{\nu}_{N,J_1J_2}^4) = o_p(J_1)$ by Assumption 5.

Next, for $1 \leq v \leq k_N$, denote $\psi_{v,1} \equiv \operatorname{tr}\{\Omega_{f_1}(\pi_{N,v}\pi_{N,v}^{\intercal} - \mathbb{E}[\pi_{N,v}\pi_{N,v}^{\intercal}|\mathcal{F}_{N,v-1}])\Omega_{f_1}\sum_{u=1}^{v}\left[\mathbb{E}[\pi_{N,u}\pi_{N,u}^{\intercal}|\mathcal{F}_{N,u-1}] - \mathbb{E}(\pi_{N,u}\pi_{N,u}^{\intercal})\right]\}$ and $\psi_{v,2} \equiv \sum_{u=1}^{v}\operatorname{tr}\{\Omega_{f_1}\mathbb{E}[\pi_{N,v}\pi_{N,v}^{\intercal}|\mathcal{F}_{N,v-1}]\Omega_{f_1}[\mathbb{E}[\pi_{N,u}\pi_{N,u}^{\intercal}|\mathcal{F}_{N,u-1}] - \mathbb{E}(\pi_{N,u}\pi_{N,u}^{\intercal})]\}$. Then the term (57) can be written as $4\sum_{v=1}^{k_N}\left(\psi_{v,1} + \psi_{v,2}\right)/(NT)^2$. Noting that $\{\psi_{v,1}, \mathcal{F}_{N,v} : 1 \leq v \leq NT, N \geq 1\}$ forms a martingale difference sequence (m.d.s.), we have $\mathbb{E}\left[\frac{1}{(NT)^2}\sum_{v=1}^{k_N}\psi_{v,1}\right]^2 = (NT)^{-4}\sum_{v=1}^{k_N}\mathbb{E}(\psi_{v,1}^2)$, which is further bounded by

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

where $\Upsilon_{it} \equiv \sum_{j=1}^{i}\left[H_{j,t-1}H_{j,t-1}^{\intercal} - \mathbb{E}\left(H_{j,t-1}H_{j,t-1}^{\intercal}\right)\right] + \sum_{j=1}^{N}\sum_{s=1}^{t-1}\left[H_{j,s-1}H_{j,s-1}^{\intercal} - \mathbb{E}\left(H_{j,s-1}H_{j,s-1}^{\intercal}\right)\right]$. 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

$$\mathbb{E}\|\Upsilon_{it}\|^2 \le CNT^2 + C\zeta_{0,J}^2\big(J/(NT) + \zeta_{1,J}J^{1/2}\min\{\Xi_{N,0}, T^{-1}\Xi_{N,1}\}\big)(NT)^2. \tag{58}$$

A combination of above results gives $\mathbb{E}\big[(NT)^{-2}\sum_{v=1}^{k_N}\psi_{v,1}\big]^2 = O(\underline{\nu}_{N,J_1J_2}^{-8}\zeta_{0,J}^4/(N^2T\underline{\nu}_{N,J_1J_2}^8) + \underline{\nu}_{N,J_1J_2}^{-8}\zeta_{0,J}^6(J/(NT)^2+\zeta_{1,J}J^{1/2}\min\{\Xi_{N,0}, T^{-1}\Xi_{N,1}\}/(NT))) = o(J_1^2)$. 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,

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

where $\|A\|_2 = \{\mathbb{E}\|A\|^2\}^{1/2}$. Using Assumption 4 and the definition (51), we have $\max_{1\le v\le NT}\big\|\mathbb{E}\big[\pi_{N,v}\pi_{N,v}^{\intercal}|\mathcal{F}_{N,v-1}\big]\big\|_2 = \sigma^2\max_{i,t}\|H_{i,t-1}H_{i,t-1}^{\intercal}\|_2 = \sigma^2\{\mathbb{E}(H_{i,t-1}^{\intercal}H_{i,t-1})^2\}^{1/2} = O(J^{1/2}\zeta_{0,J})$. Besides, (58) implies that $\big\|\sum_{u=1}^{v}\big[\mathbb{E}[\pi_{N,u}\pi_{N,u}^{\intercal}|\mathcal{F}_{N,u-1}] - \mathbb{E}(\pi_{N,u}\pi_{N,u}^{\intercal})\big]\big\|_2 = O(N^{1/2}T + \zeta_{0,J}\sqrt{JNT + \zeta_{1,J}J^{1/2}\min\{\Xi_{N,0}, T^{-1}\Xi_{N,1}\}(NT)^2})$. Combining the above results, we conclude that $\mathbb{E}|(NT)^{-2}\sum_{v=1}^{NT}\psi_{v,2}| = O(J^{1/2}\zeta_{0,J}/(\underline{\nu}_{N,J_1J_2}^4N^{1/2}) + \zeta_{0,J}^2J/(\underline{\nu}_{N,J_1J_2}^4\sqrt{NT}) + \zeta_{0,J}^2J^{3/4}\sqrt{\zeta_{1,J}\min\{\Xi_{N,0}, T^{-1}\Xi_{N,1}\}}/\underline{\nu}_{N,J_1J_2}^4) = o(J_1)$. Hence, the term (57) is also $o_p(J_1)$. We have shown that $Q_{NT,1a} = o_p(J_1)$.

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

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

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

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

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

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

which is $o_p(J_1)$. Next we examine $Q^{(2)}_{NT,1b}$. Note that $Q^{(2)}_{NT,1b} = (NT)^{-2}\sum_{u=1}^{k_N}\phi_{N,u}$, with $\phi_{N,u} \equiv \sum_{s=1}^{u-1}\pi_{N,s}^\intercal\Omega_{f_1}\sum_{v=1}^{u}\big[\mathbb{E}[\pi_{N,v}\pi_{N,v}^\intercal|\mathcal{F}_{N,v-1}]-\mathbb{E}(\pi_{N,v}\pi_{N,v}^\intercal)\big]\Omega_{f_1}\pi_{N,u}$. By construction, $\phi_{N,u}$ is $\mathcal{F}_{N,u}$-measurable and $\mathbb{E}[\phi_{N,u}|\mathcal{F}_{N,u-1}] = 0$. Therefore,

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

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, $\mathbb{E}\big\|\sum_{v=1}^{u}\big[\mathbb{E}[\pi_{N,v}\pi_{N,v}^\intercal|\mathcal{F}_{N,v-1}] - \mathbb{E}(\pi_{N,v}\pi_{N,v}^\intercal)\big]\big\|^3 \le (CNT\zeta_{0,J}^2)\mathbb{E}\big\|\sum_{v=1}^{u}\big[\mathbb{E}[\pi_{N,v}\pi_{N,v}^\intercal|\mathcal{F}_{N,v-1}] - \mathbb{E}(\pi_{N,v}\pi_{N,v}^\intercal)\big]\big\|^2 \le C\zeta_{0,J}^2N^2T^3 + C\zeta_{0,J}^2 r_{NT,J}(NT)^3$. Thus the $\mathbb{E}|Q_{NT,1b}^{(2)}|^2$ is bounded by $CJ\zeta_{0,J}^{10/3}(N^{-2/3} + r_{NT,J}^{2/3})\underline{\nu}_{N,J_1J_2}^{-8}$, which is $o(J_1^2)$ by Assumption $5'$. The above results show that $Q_{NT,1b}^{(2)} = o_p(J_1)$.

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

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

$$Q_{NT,2b} \equiv \frac{1}{(NT)^2}\sum_{v=1}^{k_N}\sum_{u=1}^{v-1}\sum_{s=1}^{u-1}\mathrm{tr}\left\{\Omega_{f_1}\pi_{N,s}\pi_{N,u}^\intercal\Omega_{f_1}\mathbb{E}\big(\pi_{N,v}\pi_{N,v}^\intercal\big)\right\}.$$

First,

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

From the lines of Lemmas 2 and 7(i), we have $\mathbb{E}\big\|\sum_{s=1}^{v}\big[\pi_{N,s}\pi_{N,s}^\intercal - \mathbb{E}\big(\pi_{N,s}\pi_{N,s}^\intercal\big)\big]\big\| \le \big\|\sum_{s=1}^{v}\big[\pi_{N,s}\pi_{N,s}^\intercal - \mathbb{E}\big(\pi_{N,s}\pi_{N,s}^\intercal\big)\big]\big\|_2 \le \big\|\sum_{s=1}^{v}\big[\pi_{N,s}\pi_{N,s}^\intercal - \mathbb{E}\big[\pi_{N,s}\pi_{N,s}^\intercal|\mathcal{F}_{N,s-1}\big]\big]\big\|_2 + \big\|\sum_{s=1}^{v}\big[\mathbb{E}\big[\pi_{N,s}\pi_{N,s}^\intercal|\mathcal{F}_{N,s-1}\big] - \mathbb{E}\big(\pi_{N,s}\pi_{N,s}^\intercal\big)\big]\big\|_2 = O(N^{1/2}T + NTr_{NT,J}^{1/2})$, where we have used the CS and Minkowski's inequalities. Also, $\big\|\Omega_{f_1}\mathbb{E}(\pi_{N,v}\pi_{N,v}^\intercal)\Omega_{f_1}\big\| = O(\underline{\nu}_{N,J_1J_2}^{-4}J)$. Combining the above results, we have $\mathbb{E}\big|Q_{NT,2a}\big| = O(\underline{\nu}_{N,J_1J_2}^{-4}J(N^{-1/2} + r_{NT,J}^{1/2})) = o(J_1)$. It follows from Markov's inequality and (47) that $|Q_{NT,2a}| = o_p(J_1) = o_p(\mathbf{s}_N^2)$.

It remains to show $|Q_{NT,2b}| = o_p(\mathbf{s}_{N,1}^2)$. Note that $\mathbb{E}(Q_{NT,2b}) = 0$ and $\mathrm{Var}(Q_{NT,2b})$ is

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

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

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

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 $\overline{\Omega}_{f_1,v} \equiv \Omega_{f_1}\mathbb{E}(\pi_{N,v}\pi_{N,v}^{\intercal})\Omega_{f_1}$, we can rewrite $|\mathcal{A}_{NT,1}|$ as

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

where we have made use of triangular and CS inequalities, the fact that $\max_{1\leq v\leq k_N}\|\overline{\Omega}_{f_1,v}\| \leq \lambda_{\max}^2(\Omega_{f_1})\cdot\max_{1\leq v\leq k_N}\|\mathbb{E}(\pi_{N,v}\pi_{N,v}^{\intercal})\| \leq C\underline{\nu}_{N,J_1J_2}^{-4}J$, (58) and Lemma 7. Next, $\mathcal{A}_{NT,2}$ can be written as

$$
\frac{1}{(NT)^3}\sum_{u=1}^{NT}\sum_{v_1=u+1}^{NT}\sum_{v_2=u+1}^{NT}\mathrm{tr}\big\{\Sigma_{N,f}\Omega_{f_1}\mathbb{E}\big(\pi_{N,v_1}\pi_{N,v_1}^{\intercal}\big)\Omega_{f_1}\mathbb{E}\big(\pi_{N,u}\pi_{N,u}^{\intercal}\big)\Omega_{f_1}\mathbb{E}\big(\pi_{N,v_2}\pi_{N,v_2}^{\intercal}\big)\Omega_{f_1}\big\}
$$

$$
\begin{aligned}
&\leq \frac{C}{\underline{\nu}_{N,J_1J_2}(NT)^3}\sum_{u=1}^{NT}\sum_{v_1=u+1}^{NT}\sum_{v_2=u+1}^{NT}\operatorname{tr}\big\{\mathbb{E}\big(\pi_{N,v_1}\pi_{N,v_1}^{\intercal}\big)\Omega_{f_1}\mathbb{E}\big(\pi_{N,u}\pi_{N,u}^{\intercal}\big)\Omega_{f_1}\mathbb{E}\big(\pi_{N,v_2}\pi_{N,v_2}^{\intercal}\big)\Omega_{f_1}\big\}\\
&= \frac{2C}{\underline{\nu}_{N,J_1J_2}(NT)^3}\sum_{1\leq u<v_1<v_2\leq NT}\operatorname{tr}\big\{\mathbb{E}\big(\pi_{N,v_1}\pi_{N,v_1}^{\intercal}\big)\Omega_{f_1}\mathbb{E}\big(\pi_{N,u}\pi_{N,u}^{\intercal}\big)\Omega_{f_1}\mathbb{E}\big(\pi_{N,v_2}\pi_{N,v_2}^{\intercal}\big)\Omega_{f_1}\big\} \qquad (59)\\
&\quad+\underbrace{\frac{C}{\underline{\nu}_{N,J_1J_2}(NT)^3}\sum_{1\leq u<v\leq NT}\operatorname{tr}\big\{\mathbb{E}\big(\pi_{N,v}\pi_{N,v}^{\intercal}\big)\Omega_{f_1}\mathbb{E}\big(\pi_{N,u}\pi_{N,u}^{\intercal}\big)\Omega_{f_1}\mathbb{E}\big(\pi_{N,v}\pi_{N,v}^{\intercal}\big)\Omega_{f_1}\big\}}_{=O(J^3/(\underline{\nu}_{N,J_1J_2}^7NT))}.
\end{aligned}
$$

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

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

Besides, by (47), Assumptions 4′ and 5′,

$$
\operatorname{tr}((\Sigma_{N,f}\Omega_{f_1})^3) \leq C\underline{\nu}_{N,J_1J_2}^{-1}\operatorname{tr}((\Sigma_{N,f}\Omega_{f_1})^2) \leq C'\underline{\nu}_{N,J_1J_2}^{-1}\mathbf{s}_{N,1}^2.
$$

Combining the above results, we have

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

by (47) 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 (54). By Theorem 1 in

Kuersteiner and Prucha (2013), we have $\sum_{v=1}^{k_N} \zeta_{N,v} \xrightarrow{d} N(0,1)$ ($\mathcal{F}_{N,0}$-stably), which implies the desired convergence in distribution. $\square$

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

Proof. Denote the $J\times J$ matrix $\mathscr{G}_{NT} \equiv (NT)^{-1}\sum_{v=1}^{NT}\left(\pi_{N,v}\pi_{N,v}^{\mathsf{T}} - \mathbb{E}[\pi_{N,v}\pi_{N,v}^{\mathsf{T}}|\mathcal{F}_{N,v-1}]\right)$. 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 (33). By Assumption 1′,

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

where the second equality is due to the martingale property, the third is 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 $j=1,\dots,(J_1+J_2)$. 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. $\square$

B.5 Proof of Theorem 4

(i) As in (48), decompose $\mathbf{T}_{NT,1}$ into three parts $\mathbf{T}_{NT,1a} + \mathbf{T}_{NT,1b} + \mathbf{T}_{NT,1c}$. Note that $\mathbf{s}_{N,1}^{-1}(\mathbf{T}_{NT,1} - \mathbf{B}_{N,1}) = \mathbf{s}_{N,1}^{-1}(\mathbf{T}_{NT,1a} - \mathbf{B}_{N,1}) + \mathbf{s}_{N,1}^{-1}\mathbf{T}_{NT,1b} + \mathbf{s}_{N,1}^{-1}\mathbf{T}_{NT,1c}$. The theorem follows if we can show 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 $\mathbb{U}_{t-1} \equiv (\mathbb{P}_{t-1}^{J_2}, \mathbb{Z})$, $\mathbf{U} \equiv (\mathbb{U}_0^\intercal, \ldots, \mathbb{U}_{T-1}^\intercal)^\intercal$, $\mathbf{Y}_{-1} \equiv (\mathbb{Y}_0^\intercal, \ldots, \mathbb{Y}_{T-1}^\intercal)^\intercal$, $\mathbf{P}^J \equiv (\mathbb{P}_0^{J\intercal}, \ldots, \mathbb{P}_{T-1}^{J\intercal})^\intercal$, and $\mathbf{W} \equiv I_T \otimes W$. Moreover, define $\mathbf{R}(\mathbf{Y}_{-1}) \equiv (R(\mathbb{Y}_0)^\intercal, \ldots, R(\mathbb{Y}_{T-1})^\intercal)^\intercal$, and $\mathbf{F}_2(\mathbf{Y}_{-1}) \equiv (F_2(\mathbb{Y}_0)^\intercal, \ldots, F_2(\mathbb{Y}_{T-1})^\intercal)^\intercal$. Then,

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

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

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 $\Psi_{J_2,11} \equiv \mathbb{E}((W\mathbb{Y}_t)^\intercal W\mathbb{Y}_t/N)$, $\Psi_{J_2,12} \equiv \mathbb{E}((W\mathbb{Y}_t)^\intercal \mathbb{U}_t/N)$, $\Psi_{J_2,22} \equiv \mathbb{E}(\mathbb{U}_t^\intercal \mathbb{U}_t/N)$, and $\Sigma_{N,1} \equiv \Psi_{J_2,11} - \Psi_{J_2,12}\Psi_{J_2,22}^{-1}\Psi_{J_2,12}^\intercal$. Then,

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

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

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

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

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

and $\boldsymbol{\eta} \equiv \mathbf{W}\mathbf{Y}_{-1} - \mathbf{U}\Psi_{J_2,22}^{-1}\Psi_{J_2,12}^{\intercal}$. 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 $\widehat{\Psi}_{J_2,12} = (NT)^{-1}\sum_{t=1}^{T}(W\mathbb{Y}_{t-1})^{\intercal}\mathbb{U}_{t-1}$ and $\widehat{\Psi}_{J_2,22} = (NT)^{-1}\sum_{t=1}^{T}\mathbb{U}_{t-1}^{\intercal}\mathbb{U}_{t-1}$. By triangular inequality,

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

where we have used the fact that $\|\widehat{\Psi}_{J_2} - \Psi_{J_2}\| = O_p(N^{-1/2} + r_{NT,J}^{1/2})$, $\sigma_{\max}^2(\widehat{\Psi}_{J_2,12}) = O_p(1)$, $\lambda_{\max}(\widehat{\Psi}_{J_2,22}^{-1}) = O_p(\underline{\nu}_{N,J_1J_2}^{-1})$, $\|\mathbf{U}^{\intercal}\boldsymbol{\mathcal{E}}/(NT)\| = O_p(\sqrt{J_2/(NT)})$ and Assumption 5′. Hence, $\mathcal{Q}_{NT,2} = \Sigma_{N,1}^{-1}\boldsymbol{\eta}^{\intercal}\boldsymbol{\mathcal{E}}/(NT) + o_p((NT)^{-1/2}) = O_p((NT)^{-1/2})$. Finally, using the same arguments as (46), we can easily find 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, $\mathcal{Q}_{NT,1} = \alpha_{NT}e_1^{\intercal}\widehat{\Psi}_{NT,11}^{-1}\widehat{\Gamma}_{NT,1}$, where $\widehat{\Psi}_{NT,11} = (NT)^{-1}\sum_{t=1}^{T}\mathbb{X}_{t-1,1}^{\intercal}\mathbb{X}_{t-1,1}$ and $\widehat{\Gamma}_{NT,1} = (NT)^{-1}\sum_{t=1}^{T}\mathbb{X}_{t-1,1}^{\intercal}WR(\mathbb{Y}_{t-1})$. Then,

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

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 immediately that $\mathcal{Q}_{NT,1}^2 = \alpha_{NT}^2(e_1^{\intercal}\Psi_{N,11}^{-1}\Gamma_{N,1})^2 + o_p(\alpha_{NT}^2)$ under Assumption 5′. Combining the above results, we have

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

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

$$\mathbf{s}_{N,1}^{-1}\mathbf{T}_{NT,1b}$$

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

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 either 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 $H_1^n(\alpha_{NT})$:

$$
\begin{aligned}
&\sum_{i=1}^{N} \sum_{t=1}^{T} \big(\widehat{f}_1(Y_{i,t-1}) - f_1(Y_{i,t-1}) \big) \big(f_1(Y_{i,t-1}) - \widehat{\rho}_1 Y_{i,t-1} \big) \\
&= (\rho_1 - \widehat{\rho}_1) \sum_{i=1}^{N} \sum_{t=1}^{T} \big(p^{J_1}(Y_{i,t-1})^\intercal \widetilde{U}_1 + p^{J_1}(Y_{i,t-1})^\intercal \widetilde{U}_2 + p^{J_1}(Y_{i,t-1})^\intercal \widetilde{U}_3 - u_1(Y_{i,t-1}) \big) Y_{i,t-1} \\
&\quad + \alpha_{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \big(p^{J_1}(Y_{i,t-1})^\intercal \widetilde{U}_1 + p^{J_1}(Y_{i,t-1})^\intercal \widetilde{U}_2 + p^{J_1}(Y_{i,t-1})^\intercal \widetilde{U}_3 - u_1(Y_{i,t-1}) \big) r(Y_{i,t-1}) \\
&= \sum_{t=1}^{T} \big((\rho_1 - \widehat{\rho}_1) \mathbb{Y}_{t-1} + \alpha_{NT} R(\mathbb{Y}_{t-1}) \big)^\intercal \mathbb{P}_{t-1}^{J_1} \widetilde{U}_1 \\
&\quad + \underbrace{\sum_{t=1}^{T} \big((\rho_1 - \widehat{\rho}_1) \mathbb{Y}_{t-1} + \alpha_{NT} R(\mathbb{Y}_{t-1}) \big)^\intercal \mathbb{P}_{t-1}^{J_1} \widetilde{U}_2}_{=O_p\left(\alpha_{NT} \underline{\nu}_{N,J_1J_2}^{-2} (N^{-1/2} + r_{NT,J}^{1/2}) \sqrt{J_1 J N T}\right)} + \underbrace{\sum_{t=1}^{T} \big((\rho_1 - \widehat{\rho}_1) \mathbb{Y}_{t-1} + \alpha_{NT} R(\mathbb{Y}_{t-1}) \big)^\intercal \mathbb{P}_{t-1}^{J_1} \widetilde{U}_3}_{=O_p\left(\alpha_{NT} NT \sqrt{J_1 / \underline{\nu}_{N,J_1J_2}} (J_1^{-\mu} + J_2^{-\mu})\right)} \\
&\quad - \underbrace{\sum_{i=1}^{N} \sum_{t=1}^{T} \big((\rho_1 - \widehat{\rho}_1) Y_{i,t-1} + \alpha_{NT} r(Y_{i,t-1}) \big) u_1(Y_{i,t-1})}_{=O_p(\alpha_{NT} NT J_1^{-\mu})} \\
&= (\rho_1 - \widehat{\rho}_1) \left(\sum_{t=1}^{T} \mathbb{Y}_{t-1}^\intercal \mathbb{P}_{t-1}^{J_1} \right) \widetilde{U}_1 + \alpha_{NT} \left(\sum_{t=1}^{T} R(\mathbb{Y}_{t-1})^\intercal \mathbb{P}_{t-1}^{J_1} \right) \widetilde{U}_1 + o_p(J_1^{1/2}),
\end{aligned}
\tag{60}
$$

where the second equality uses the fact that $\widehat{\rho}_1 - \rho_1 = O_p(\alpha_{NT})$, $\| \sum_{t=1}^{T} \mathbb{P}_{t-1}^{J_1 \intercal} R(\mathbb{Y}_{t-1}) \| = O_p(NT\sqrt{J_1})$, and $\| \sum_{t=1}^{T} \mathbb{P}_{t-1}^{J_1 \intercal} \mathbb{Y}_{t-1} \| = O_p(NT\sqrt{J_1})$, and the last equality is from Assumption 5′. Recall from the proof of Lemma 5 that $\big(\sum_{t=1}^{T} \mathbb{Y}_{t-1}^\intercal \mathbb{P}_{t-1}^{J_1} \big) \widetilde{U}_1 = \tilde{U}_1^{(1)} + \tilde{U}_1^{(2)}$, and $|\tilde{U}_1^{(1)}| = O_p(\sqrt{NT / \underline{\nu}_{N,J_1J_2}})$. Therefore, $(\widehat{\rho}_1 - \rho_1) \tilde{U}_1^{(1)} = O_p(\alpha_{NT} \sqrt{NT / \underline{\nu}_{N,J_1J_2}}) = O_p(\mathbf{s}_{N,1}^{1/2} \underline{\nu}_{N,J_1J_2}^{-1/2}) = o_p(\mathbf{s}_{N,1})$ under Assumption 5′. Moreover, by (50), we have $(\widehat{\rho}_1 - \rho_1) \tilde{U}_1^{(2)} =$

$O_p(\alpha_{NT}\sqrt{J_1\zeta_{0,J_1}^2/\underline{\nu}_{N,J_1J_2}} + \alpha_{NT}\sqrt{J_1\zeta_{0,J_1}(\zeta_{0,J_1}+\zeta_{1,J_1})\min\{\Xi_{N,0},T^{-1}\Xi_{N,1}\}NT/\underline{\nu}_{N,J_1J_2}}) = o_p(\mathbf{s}_{N,1})$. Consequently, the first term on the r.h.s. of (60) is $o_p(\mathbf{s}_{N,1})$. In a similar manner, we can show that the second term on the r.h.s. of (60) 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 $\mathbf{Y} - \mathbf{X}_1\bar{\theta}_1 = \mathscr{M}(\mathbf{X}_1)(\alpha_{NT}\mathbf{W}\mathbf{R}(\mathbf{Y}_{-1}) + \mathbf{R}_2 + \boldsymbol{\mathcal{E}})$, where $\mathbf{R}_2 \equiv (\mathbb{R}_{0,2}^\intercal, \mathbb{R}_{1,2}^\intercal, \ldots, \mathbb{R}_{T-1,2}^\intercal)^\intercal$ 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

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

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 (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 maximum eigenvalue of an idempotent matrix is at most one, we have

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

Moreover, it also holds that

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

To show this, we write $\mathbf{LM}_{NT,2b} = \mathbf{LM}_{NT,2b}^{(1)} + \mathbf{LM}_{NT,2b}^{(2)} + \mathbf{LM}_{NT,2b}^{(3)}$, where

$$\begin{aligned}
\mathbf{LM}_{NT,2b}^{(1)} &\equiv \alpha_{NT}\Lambda_N^\intercal\Theta_N^{-1}\boldsymbol{\nu}^\intercal\boldsymbol{\mathcal{E}}/\widehat{\sigma}^2,\\
\mathbf{LM}_{NT,2b}^{(2)} &\equiv \alpha_{NT}(\widehat{\Lambda}_{NT}^\intercal - \Lambda_N^\intercal)\Theta_N^{-1}\boldsymbol{\nu}^\intercal\boldsymbol{\mathcal{E}}/\widehat{\sigma}^2,\\
\mathbf{LM}_{NT,2b}^{(3)} &\equiv \alpha_{NT}\widehat{\Lambda}_{NT}^\intercal\big(\widehat{\Theta}_{NT}^{-1}\mathbf{X}_2^\intercal\mathscr{M}(\mathbf{X}_1) - \Theta_N^{-1}\boldsymbol{\nu}^\intercal\big)\boldsymbol{\mathcal{E}}/\widehat{\sigma}^2,
\end{aligned}$$

with $\boldsymbol{\nu} = \mathbf{X}_2 - \mathbf{X}_1\Psi_{N,11}^{-1}\Psi_{N,12}$, $\Lambda_N \equiv \Gamma_{N,2} - \Psi_{N,12}^\intercal\Psi_{N,11}^{-1}\Gamma_{N,1}$, $\widehat{\Lambda}_{NT} \equiv (NT)^{-1}\mathbf{X}_2^\intercal\mathscr{M}(\mathbf{X}_1)\mathbf{W}\mathbf{R}(\mathbf{Y}_{-1})$, $\Theta_N \equiv \Psi_{N,22} - \Psi_{N,12}^\intercal\Psi_{N,11}^{-1}\Psi_{N,12}$, and $\widehat{\Theta}_{NT} \equiv (NT)^{-1}\mathbf{X}_2^\intercal\mathscr{M}(\mathbf{X}_1)\mathbf{X}_2$. First, note that $\mathbf{E}|\Lambda_N^\intercal\Theta_N^{-1}\boldsymbol{\nu}^\intercal\boldsymbol{\mathcal{E}}|^2 = \sigma^2(NT)\Lambda_N^\intercal\Theta_N^{-1}\Theta_N\Theta_N^{-1}\Lambda_N = \sigma^2(NT)\Lambda_N^\intercal\Theta_N^{-1}\Lambda_N = O(NT)$. It follows from Markov's inequality that $|\mathbf{LM}_{NT,2b}^{(1)}| =$

$O_p(\alpha_{NT}\sqrt{NT}) = O_p(J_1^{1/4})$. Next, for $\mathbf{LM}_{NT,2b}^{(2)}$, note that

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

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

To show (e), we have

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

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 1

Note that

$$\widehat{\sigma}^2 - \sigma^2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(Y_{it} - \sum_{j=1}^{N}w_{ij}\widehat{f}_1(Y_{j,t-1}) - \widehat{f}_2(Y_{i,t-1}) - Z_i^{\intercal}\widehat{\gamma}\right)^2 - \sigma^2$$

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

where $\phi_1 \equiv (NT)^{-1}\sum_{v=1}^{NT}\epsilon_{it}^2 - \sigma^2$, $\phi_2 \equiv (1/N)\sum_{i=1}^{N}(Z_i^{\mathsf{T}}(\widehat{\gamma} - \gamma))^2$, $\phi_3 \equiv (NT)^{-1}\sum_{v=1}^{NT}(\sum_{j=1}^{N} w_{ij}(\widehat{f}_1(Y_{j,t-1}) - f_1(Y_{j,t-1})) + \widehat{f}_2(Y_{i,t-1}) - f_2(Y_{i,t-1}))^2$, $\phi_4 \equiv (NT)^{-1}\sum_{v=1}^{NT}\epsilon_{it}Z_i^{\mathsf{T}}(\widehat{\gamma} - \gamma)$, $\phi_5 \equiv (NT)^{-1}\sum_{v=1}^{NT}\epsilon_{it}(\sum_{j=1}^{N} w_{ij}(\widehat{f}_1(Y_{j,t-1}) - f_1(Y_{j,t-1})) + \widehat{f}_2(Y_{i,t-1}) - f_2(Y_{i,t-1}))$ and $\phi_6 \equiv (NT)^{-1}\sum_{v=1}^{NT}(\sum_{j=1}^{N} w_{ij}(\widehat{f}_1(Y_{j,t-1}) - f_1(Y_{j,t-1})) + \widehat{f}_2(Y_{i,t-1}) - f_2(Y_{i,t-1}))Z_i^{\mathsf{T}}(\widehat{\gamma} - \gamma)$.

First, it is easy to see from the law of large numbers that $|\phi_1| = o_p(1)$ 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

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

where $\widehat{\Psi}_{NT,J} = \sum_{t=1}^{T}\mathbb{Q}_{t-1}^{\mathsf{T}}\mathbb{Q}_{t-1}/(NT)$ 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 $(\widehat{\beta} - \beta)^{\mathsf{T}}\widehat{\Psi}_{NT,J}(\widehat{\beta} - \beta) \leq O_p(1) \cdot \|\widehat{\beta} - \beta\|^2 = O_p(J/(NT) + J_1^{-2\mu} + J_2^{-2\mu})$. 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 yields $\widehat{\sigma}^2 - \sigma^2 \xrightarrow{p} 0$. $\square$

C.2 Asymptotics with fixed N

The asymptotic results presented in Section 3 are established under the setting where $N \to \infty$ and T can be fixed or tends to ∞. In this section, we discuss another regime, that is, $T \to \infty$ and N is fixed. Redefine $\Sigma_{J_1J_2} \equiv \mathbb{E}(\mathbb{X}_{t-1}^{\mathsf{T}}\mathbb{X}_{t-1}/N \,|\, \mathbb{Z})$, $C_{ZJ} = \mathbb{E}(\mathbb{Z}^{\mathsf{T}}\mathbb{P}_{t-1}/N \,|\, \mathbb{Z})$, and $\Psi_J = \mathbb{E}(\mathbb{P}_{t-1}^{\mathsf{T}}\mathbb{P}_{t-1}/N \,|\, \mathbb{Z})$. We first modify our assumptions as follows.

Assumption 4†. *Assumption 4(i) holds. In addition, let* $\bar{\nu}_{J_1J_2} \equiv \lambda_{\max}(\Sigma_{J_1J_2})$ *and* $\underline{\nu}_{J_1J_2} \equiv$

$\lambda_{\min}(\Sigma_{J_1J_2})$. *For each* J_1 *and* J_2, *there exists a positive constant* $\bar{c} < \infty$ *such that* $\bar{\nu}_{J_1J_2} \leq \bar{c}$ *and* $\underline{\nu}_{J_1J_2} > 0$ *wpa1.*

Assumption 7†. *The matrix* $\Sigma_Z = \widehat{\Psi}_Z - C_{ZJ}\Psi_J^{-1}C_{ZJ}^{\mathsf{T}}$ *is positive definite wpa1.*

Theorems 1† and 2† below establish the asymptotic properties of the sieve LS estimator under the fixed N setting. It can be seen that the convergence rates and the asymptotic distributions of $\widehat{\gamma}$ and $\widehat{f}_\ell(y)$ are basically the same as in Theorems 1 and 2. The major difference is that we now state the results conditional on $\{Z_i : 1 \leq i \leq N\}$.

Theorem 1†. *Suppose that Assumptions 1-3, 4†, 6, 7† and 8 hold. Then, conditional on* $\{Z_i : 1 \leq i \leq N\}$,

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

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

Theorem 2†. *Suppose that Assumptions 1-3, 4†, 6, 7† and 8 hold. Denote* $\tilde{r}_{T,J} \equiv \zeta_{0,J}^2(J + \zeta_{1,J}J^{1/2})/T$. *Then, conditional on* $\{Z_i : 1 \leq i \leq N\}$,

(i) if $\max_{i\geq 1}\lambda_{max}(\mathbb{E}[p^{J_\ell}(Y_{it})p^{J_\ell}(Y_{it})^{\mathsf{T}}]) < \infty$ *and* $J\tilde{r}_{T,J}/(J_\ell \underline{\nu}_{J_1J_2}^3) = O(1)$, $\max_{1\leq i\leq N}\int_{\mathcal{R}_{\mathcal{Y}}}[\widehat{f}_\ell(y) - f_\ell(y)]^2 dF_i(y) = O_p(\underline{\nu}_{J_1J_2}^{-1}(J_\ell/T + J_1^{-2\mu} + J_2^{-2\mu}))$, *where* $F_i(y)$ *is the cumulative distribution function of* Y_{it};

(ii) If (a) $\zeta_{1,J} \lesssim J^{\xi}$ *for some* $\xi > 0$, *(b)* $\zeta_{0,J}^4 \ln J_\ell/(\underline{\nu}_{J_1J_2}^2 T) = o(1)$, *and (c)* $J\tilde{r}_{T,J}/(\underline{\nu}_{J_1J_2}^3 \ln J_\ell) = O(1)$, *we have* $\left|\widehat{f}_\ell(\cdot) - f_\ell(\cdot)\right|_{\infty,\omega} = O_p(\zeta_{0,J_\ell}\sqrt{\ln J_\ell/(\underline{\nu}_{J_1J_2}T)}) + O_p(\zeta_{0,J_\ell}\underline{\nu}_{J_1J_2}^{-1/2}(J_1^{-\mu} + J_2^{-\mu}))$;

(iii) If $\underline{\nu}_{J_1J_2}^{-4}J\tilde{r}_{T,J} = o(1)$ *and* $\sqrt{T}\underline{\nu}_{J_1J_2}^{-1/2}(J_1^{-\mu} + J_2^{-\mu}) = o(1)$, *then for a given finite* $y \in \mathcal{R}_{\mathcal{Y}}$ *such that* $\|p^{J_\ell}(y)\| > 0$, *we have* $\frac{\sqrt{NT}}{v_{\ell N}(y)}(\widehat{f}_\ell(y) - f_\ell(y)) \xrightarrow{d} N(0,1)$ *with* $v_{\ell N}^2(y) \equiv \sigma^2 p^{J_\ell}(y)^{\mathsf{T}}\Sigma_{N,f_\ell}^{-1}p^{J_\ell}(y)$.

Proof of Theorem 1†. Theorem 1† follows from same arguments as those used in Theorem 1. First, when N is fixed, the NED property of $\{\mathbb{Y}_t : t \in \mathbb{Z}\}$ can be considered only along the time dimension. It is also straightforward to see that the arguments in Lemma 2 are valid regardless of which asymptotic regime we are considering. Specifically, we have Lemma 2(i) and (ii) 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 Assumptions 1-4† and 8. Besides, Lemma 3 can be revised as follows.

Lemma 3′. *Under Assumptions 1-5 and 7, conditional on* $\{Z_i : 1 \leq i \leq N\}$,

(i) $\|\widehat{\Sigma}_{NT,Z} - \Sigma_{N,Z}\| = O_p\big(\underline{\nu}_{J_1J_2}^{-2}\tilde{r}_{T,J}^{1/2}\big)$ *and* $\|\widehat{\Sigma}_{NT,Z}^{-1} - \Sigma_{N,Z}^{-1}\| = O_p\big(\underline{\nu}_{J_1J_2}^{-2}\tilde{r}_{T,J}^{1/2}\big)$;

(ii) $\|\widehat{\Sigma}_{NT,f} - \Sigma_{N,f}\| = O_p\big(\tilde{r}_{T,J}^{1/2}\big)$ *and* $\|\widehat{\Sigma}_{NT,f}^{-} - \Sigma_{N,f}^{-1}\| = O_p\big(\underline{\nu}_{J_1J_2}^{-2}\tilde{r}_{T,J}^{1/2}\big)$.

Theorem 1†(i) can be shown in the same manner as Theorem 1(i). To show (ii), note that

$$\sqrt{NT}\Sigma_{N,Z}^{1/2}(\widehat{\gamma} - \gamma)/\sigma = \Omega_{N,Z}^{-1/2}\boldsymbol{\xi}^{\mathsf{T}}\boldsymbol{\mathcal{E}}/\sqrt{NT} + o_p(1),$$

where $\Omega_{N,Z} = \sigma^2 \Sigma_{N,Z}$. Define $a_T = \mathbf{c}^\intercal \Omega_Z^{-1/2} \boldsymbol{\xi}^\intercal \boldsymbol{\mathcal{E}} / \sqrt{NT}$, and write $a_T = \sum_{t=1}^T X_t$, where, with abuse of notation, $X_t \equiv (NT)^{-1/2} \boldsymbol{c}^\intercal \Omega_{N,Z}^{-1/2} (\mathbb{Z}^\intercal - C_{N,ZJ} \Psi_{N,J}^{-1} \mathbb{Q}_{t-1}^\intercal) \mathcal{E}_t$. Define $\mathcal{B}_t$ as the σ-fields generated by the random variables $\{\mathcal{E}_1, \ldots, \mathcal{E}_t\}$. Then, we can easily see that X_t is $\mathcal{B}_t$-measurable and $\mathbb{E}[X_t|\mathcal{B}_{t-1}] = 0$. Then, using the classical CLT for m.d.s., see Corollary 3.1 in Hall and Heyde (1980), we have $a_{NT} \xrightarrow{d} N(0,1)$ provided the following conditions hold: for some $\delta > 0$ $\sum_{t=1}^T \mathbb{E}|X_t|^{2+\delta} \xrightarrow{p} 0$, and $\sum_{t=1}^T \mathbb{E}[|X_t|^2|\mathcal{B}_{t-1}] \xrightarrow{p} 1$. The above conditions can be verified using similar arguments as showing (53) and (54), respectively. The details are omitted for brevity. This proves Theorem $1^\dagger$(ii). □

Theorem $2^\dagger$(ii) follows by slightly modifying the proof of Theorem 2(ii) in the same way as above. The details are omitted for brevity.

D Additional simulation results

In this section, we report the rejection probabilities of the distance-based and LM test statistics under the DGPs considered in Section 5 with large T. We set $N \in \{100, 200\}$ and $T \in \{100, 200\}$. The rest of the settings are the same as Section 5 in the main text. We consider $J \in \{6, 7\}$ for large T setting.

Tables 7 and 8 report the estimation results of sieve estimators $\widehat{\gamma}$ and $(\widehat{f}_1, \widehat{f}_2)$, respectively. It can be seen from Table 7 that the bias and RMSE of γ_1 and γ_2 are quite small. Compared with Table 1, the estimation accuracy of γ significantly improves due to a larger sample size T. The performance of the sieve estimators of f_1 and f_2 exhibits a similar pattern as discussed in Section 5.

Tables 9 and 10 report the rejection frequencies of the LM and distance-based test statistics, respectively. We see that 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 to one. Compared with Table 9, the empirical power of the LM test significantly improves, especially for DGP 4 and ER network model.

Table 7: Estimation results of (γ_1, γ_2) when $T \in \{100, 200\}$.

				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	100	100	6	0.0164	0.0203	0.947	0.1101	0.0225	0.956	-0.0443	0.0147	0.956	0.0984	0.0161	0.950
	100	100	7	0.0152	0.0203	0.949	0.1100	0.0225	0.957	-0.0448	0.0147	0.956	0.0995	0.0161	0.950
	200	100	6	0.0207	0.0100	0.946	0.0547	0.0104	0.950	0.0133	0.0080	0.952	0.0100	0.0079	0.947
	200	100	7	0.0211	0.0100	0.946	0.0544	0.0104	0.951	0.0134	0.0080	0.952	0.0101	0.0079	0.944
	100	200	6	0.0696	0.0106	0.949	0.0554	0.0110	0.943	-0.0349	0.0079	0.950	-0.0196	0.0078	0.952
	100	200	7	0.0700	0.0106	0.950	0.0566	0.0111	0.943	-0.0349	0.0079	0.949	-0.0198	0.0077	0.952
	200	200	6	0.0559	0.0051	0.952	-0.0165	0.0051	0.953	0.0119	0.0037	0.952	-0.0246	0.0040	0.944
	200	200	7	0.0557	0.0051	0.952	-0.0168	0.0051	0.951	0.0118	0.0037	0.952	-0.0246	0.0040	0.944
SBM	100	100	6	0.1073	0.0192	0.946	0.1279	0.0208	0.958	0.0958	0.0141	0.960	0.0004	0.0163	0.949
	100	100	7	0.1078	0.0192	0.945	0.1272	0.0207	0.959	0.0952	0.0141	0.960	0.0002	0.0163	0.949
	200	100	6	0.0468	0.0094	0.950	0.0961	0.0107	0.944	-0.0004	0.0081	0.945	-0.0174	0.0072	0.958
	200	100	7	0.0471	0.0094	0.948	0.0964	0.0107	0.946	-0.0004	0.0080	0.945	-0.0174	0.0072	0.957
	100	200	6	0.0380	0.0102	0.935	0.0521	0.0102	0.955	-0.0228	0.0079	0.947	-0.0492	0.0077	0.960
	100	200	7	0.0380	0.0102	0.937	0.0525	0.0102	0.955	-0.0239	0.0079	0.947	-0.0495	0.0077	0.960
	200	200	6	0.0139	0.0048	0.948	0.0418	0.0053	0.949	0.0434	0.0039	0.946	0.0317	0.0039	0.954
	200	200	7	0.0137	0.0048	0.948	0.0418	0.0053	0.949	0.0432	0.0039	0.944	0.0315	0.0039	0.953

				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	100	6	0.0450	0.0198	0.956	0.1419	0.0220	0.949	0.0621	0.0156	0.951	0.0044	0.0152	0.952
	100	100	7	0.0423	0.0197	0.955	0.1413	0.0220	0.946	0.0620	0.0157	0.952	0.0045	0.0152	0.953
	200	100	6	0.0231	0.0095	0.959	-0.0081	0.0098	0.958	-0.0424	0.0078	0.950	0.0138	0.0079	0.946
	200	100	7	0.0220	0.0095	0.959	-0.0078	0.0098	0.956	-0.0422	0.0078	0.950	0.0133	0.0079	0.946
	100	200	6	0.0500	0.0096	0.953	0.0419	0.0107	0.948	0.0156	0.0083	0.946	0.0478	0.0070	0.960
	100	200	7	0.0495	0.0096	0.953	0.0411	0.0107	0.946	0.0154	0.0083	0.943	0.0478	0.0070	0.961
	200	200	6	0.0140	0.0048	0.947	0.0353	0.0051	0.951	0.0112	0.0037	0.953	0.0025	0.0040	0.942
	200	200	7	0.0144	0.0048	0.947	0.0350	0.0051	0.950	0.0111	0.0037	0.952	0.0024	0.0040	0.942
SBM	100	100	6	0.0641	0.0194	0.953	0.0907	0.0225	0.940	0.0135	0.0142	0.961	0.0519	0.0159	0.949
	100	100	7	0.0650	0.0195	0.954	0.0892	0.0225	0.940	0.0140	0.0141	0.962	0.0504	0.0159	0.951
	200	100	6	0.0205	0.0090	0.940	0.0598	0.0107	0.951	-0.0036	0.0071	0.963	-0.0010	0.0077	0.956
	200	100	7	0.0203	0.0090	0.937	0.0605	0.0107	0.950	-0.0042	0.0071	0.963	-0.0012	0.0077	0.956
	100	200	6	0.0501	0.0104	0.952	0.0290	0.0102	0.954	0.0658	0.0080	0.950	0.0114	0.0077	0.951
	100	200	7	0.0497	0.0104	0.951	0.0296	0.0102	0.952	0.0662	0.0080	0.950	0.0117	0.0077	0.951
	200	200	6	0.0239	0.0041	0.958	-0.0141	0.0054	0.953	-0.0286	0.0036	0.951	0.0026	0.0037	0.947
	200	200	7	0.0238	0.0041	0.959	-0.0140	0.0054	0.952	-0.0288	0.0036	0.951	0.0026	0.0037	0.947

Note. The presented bias and RMSE equal 100 times the true values.

Table 8: Estimation results of f_1 and f_2 when $T \in \{100, 200\}$

				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	100	100	6	0.0715	17.1702	0.0061	14.6970	0.0371	8.1738	0.0065	8.0156
	100	100	7	0.0872	20.0230	0.0065	16.5214	0.0426	10.1197	0.0076	10.1831
	200	100	6	0.0164	6.4786	0.0145	6.5669	0.0211	3.6586	0.0041	3.6281
	200	100	7	0.0159	7.6287	0.0054	7.3156	0.0161	4.7095	0.0092	4.5682
	100	200	6	0.0696	8.4043	0.0190	7.3423	0.0101	3.7631	0.0027	4.2413
	100	200	7	0.0775	9.7522	0.0052	8.1846	0.0070	4.8144	0.0009	5.1982
	200	200	6	0.0041	3.3935	0.0122	3.1141	0.0035	1.7684	0.0011	1.8891
	200	200	7	0.0046	3.8440	0.0031	3.5126	0.0025	2.2016	0.0055	2.3324
SBM	100	100	6	0.0021	2.0043	0.0071	1.5598	0.0003	0.8922	0.0009	0.8493
	100	100	7	0.0022	2.3390	0.0026	1.9387	0.0004	1.0516	0.0011	1.0379
	200	100	6	0.0002	0.8668	0.0187	0.7018	0.0007	0.4217	0.0017	0.4073
	200	100	7	0.0003	1.0138	0.0032	0.8315	0.0010	0.5244	0.0035	0.4922
	100	200	6	0.0014	1.0351	0.0070	0.7646	0.0001	0.4304	0.0008	0.4271
	100	200	7	0.0017	1.1767	0.0013	0.9569	0.0001	0.5158	0.0013	0.5170
	200	200	6	0.0003	0.4737	0.0169	0.3622	0.0000	0.2049	0.0009	0.2055
	200	200	7	0.0003	0.5306	0.0029	0.4134	0.0000	0.2572	0.0020	0.2531
				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	100	6	0.0006	0.9593	0.0132	0.9597	0.0003	0.4234	0.0007	0.4391
	100	100	7	0.0005	1.0911	0.0110	1.0529	0.0008	0.5269	0.0009	0.5352
	200	100	6	0.0005	0.4026	0.0011	0.4481	0.0007	0.2219	0.0004	0.2162
	200	100	7	0.0005	0.4552	0.0009	0.4949	0.0004	0.2736	0.0013	0.2652
	100	200	6	0.0019	0.4941	0.0006	0.4592	0.0008	0.2178	0.0003	0.2155
	100	200	7	0.0018	0.5548	0.0005	0.5033	0.0007	0.2673	0.0002	0.2640
	200	200	6	0.0007	0.2060	0.0003	0.2225	0.0002	0.1056	0.0002	0.1039
	200	200	7	0.0007	0.2287	0.0003	0.2431	0.0002	0.1301	0.0004	0.1264
SBM	100	100	6	0.0011	1.1066	0.0092	0.9919	0.0015	0.4577	0.0008	0.4516
	100	100	7	0.0010	1.2302	0.0065	1.1832	0.0026	0.5282	0.0027	0.5341
	200	100	6	0.0012	0.5019	0.0011	0.4637	0.0011	0.2283	0.0004	0.2171
	200	100	7	0.0018	0.5711	0.0011	0.5203	0.0018	0.2759	0.0003	0.2597
	100	200	6	0.0008	0.4986	0.0008	0.4590	0.0002	0.2413	0.0002	0.2135
	100	200	7	0.0007	0.5539	0.0008	0.5475	0.0002	0.2758	0.0005	0.2507
	200	200	6	0.0006	0.2461	0.0004	0.2351	0.0002	0.1101	0.0004	0.1088
	200	200	7	0.0006	0.2758	0.0002	0.2660	0.0005	0.1342	0.0009	0.1316

Note. The presented ISB and IMSE equal 100 times the true values.

Table 9: Rejection probabilities of the LM test statistic (11) when $T \in \{100, 200\}$

(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	100	100	6	0.105	0.049	0.011	1	1	0.995	0.112	0.057	0.008	0.937	0.896	0.743
	100	100	7	0.108	0.052	0.013	1	1	0.993	0.118	0.057	0.007	0.930	0.879	0.703
	200	100	6	0.109	0.060	0.015	1	1	1	0.091	0.049	0.009	1	0.998	0.991
	200	100	7	0.114	0.061	0.011	1	1	1	0.092	0.047	0.010	0.999	0.995	0.988
	100	200	6	0.113	0.061	0.013	1	1	1	0.103	0.056	0.010	1	0.998	0.979
	100	200	7	0.101	0.052	0.009	1	1	1	0.100	0.052	0.014	1	0.993	0.966
	200	200	6	0.115	0.062	0.014	1	1	1	0.087	0.043	0.007	1	1	1
	200	200	7	0.112	0.057	0.010	1	1	1	0.090	0.044	0.011	1	1	1
SBM	100	100	6	0.099	0.050	0.011	1	1	1	0.109	0.043	0.007	1	1	1
	100	100	7	0.094	0.050	0.010	1	1	1	0.102	0.045	0.008	1	1	1
	200	100	6	0.094	0.050	0.007	1	1	1	0.094	0.056	0.013	1	1	1
	200	100	7	0.090	0.046	0.005	1	1	1	0.100	0.060	0.012	1	1	1
	100	200	6	0.102	0.060	0.014	1	1	1	0.094	0.044	0.015	1	1	1
	100	200	7	0.101	0.057	0.018	1	1	1	0.088	0.048	0.013	1	1	1
	200	200	6	0.117	0.062	0.016	1	1	1	0.104	0.056	0.012	1	1	1
	200	200	7	0.097	0.052	0.018	1	1	1	0.104	0.058	0.013	1	1	1
(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	100	6	0.092	0.048	0.006	0.093	0.045	0.013	1	1	1	1	1	1
	100	100	7	0.098	0.047	0.009	0.102	0.055	0.015	1	1	1	1	1	1
	200	100	6	0.093	0.053	0.009	0.096	0.051	0.012	1	1	1	1	1	1
	200	100	7	0.103	0.053	0.007	0.105	0.057	0.014	1	1	1	1	1	1
	100	200	6	0.089	0.045	0.008	0.093	0.046	0.011	1	1	1	1	1	1
	100	200	7	0.085	0.045	0.005	0.098	0.053	0.011	1	1	1	1	1	1
	200	200	6	0.099	0.050	0.010	0.094	0.038	0.009	1	1	1	1	1	1
	200	200	7	0.103	0.047	0.011	0.096	0.052	0.008	1	1	1	1	1	1
SBM	100	100	6	0.094	0.044	0.007	0.090	0.038	0.010	1	1	1	1	1	1
	100	100	7	0.094	0.042	0.006	0.107	0.047	0.013	1	1	1	1	1	1
	200	100	6	0.090	0.038	0.011	0.095	0.042	0.012	1	1	1	1	1	1
	200	100	7	0.096	0.043	0.010	0.091	0.047	0.009	1	1	1	1	1	1
	100	200	6	0.093	0.045	0.007	0.092	0.052	0.010	1	1	1	1	1	1
	100	200	7	0.087	0.036	0.003	0.108	0.051	0.009	1	1	1	1	1	1
	200	200	6	0.080	0.042	0.009	0.093	0.050	0.006	1	1	1	1	1	1
	200	200	7	0.082	0.044	0.011	0.104	0.056	0.009	1	1	1	1	1	1

Table 10: Rejection probabilities of the distance-based test statistic (10) when $T \in \{100, 200\}$

(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	100	100	6	0.084	0.059	0.026	0.999	0.998	0.994	0.041	0.025	0.010	0.669	0.563	0.371
	100	100	7	0.080	0.060	0.030	0.999	0.998	0.989	0.050	0.035	0.015	0.566	0.443	0.264
	200	100	6	0.077	0.054	0.029	1	1	1	0.042	0.021	0.008	0.979	0.968	0.918
	200	100	7	0.082	0.062	0.037	1	1	1	0.054	0.036	0.010	0.959	0.923	0.817
	100	200	6	0.068	0.045	0.026	1	1	1	0.036	0.026	0.012	0.956	0.931	0.871
	100	200	7	0.063	0.040	0.027	1	1	1	0.052	0.033	0.013	0.921	0.885	0.765
	200	200	6	0.082	0.051	0.025	1	1	1	0.034	0.016	0.003	1	1	1
	200	200	7	0.062	0.043	0.028	1	1	1	0.049	0.035	0.013	1	1	1
SBM	100	100	6	0.066	0.050	0.022	1	1	1	0.055	0.030	0.009	1	1	1
	100	100	7	0.075	0.052	0.022	1	1	1	0.058	0.038	0.013	1	1	0.999
	200	100	6	0.055	0.035	0.020	1	1	1	0.054	0.033	0.014	1	1	1
	200	100	7	0.056	0.037	0.013	1	1	1	0.063	0.042	0.017	1	1	1
	100	200	6	0.078	0.054	0.021	1	1	1	0.046	0.024	0.010	1	1	1
	100	200	7	0.075	0.053	0.020	1	1	1	0.052	0.026	0.010	1	1	1
	200	200	6	0.068	0.043	0.023	1	1	1	0.049	0.032	0.020	1	1	1
	200	200	7	0.071	0.049	0.025	1	1	1	0.065	0.041	0.019	1	1	1
(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	100	6	0.024	0.015	0.005	0.020	0.012	0.006	1	1	1	1	1	1
	100	100	7	0.033	0.023	0.009	0.020	0.014	0.009	1	1	1	1	1	1
	200	100	6	0.036	0.022	0.011	0.015	0.007	0.004	1	1	1	1	1	1
	200	100	7	0.041	0.025	0.013	0.021	0.012	0.007	1	1	1	1	1	1
	100	200	6	0.027	0.016	0.008	0.015	0.009	0.003	1	1	1	1	1	1
	100	200	7	0.038	0.027	0.011	0.022	0.013	0.005	1	1	1	1	1	1
	200	200	6	0.035	0.022	0.008	0.014	0.008	0.003	1	1	1	1	1	1
	200	200	7	0.043	0.026	0.009	0.015	0.009	0.004	1	1	1	1	1	1
SBM	100	100	6	0.034	0.020	0.010	0.015	0.009	0.002	1	1	1	1	1	1
	100	100	7	0.039	0.027	0.013	0.033	0.022	0.007	1	1	1	1	1	1
	200	100	6	0.029	0.021	0.008	0.020	0.013	0.003	1	1	1	1	1	1
	200	100	7	0.040	0.029	0.016	0.015	0.006	0.004	1	1	1	1	1	1
	100	200	6	0.028	0.012	0.005	0.024	0.015	0.005	1	1	1	1	1	1
	100	200	7	0.025	0.016	0.007	0.037	0.025	0.007	1	1	1	1	1	1
	200	200	6	0.026	0.017	0.008	0.022	0.012	0.004	1	1	1	1	1	1
	200	200	7	0.032	0.019	0.013	0.024	0.011	0.002	1	1	1	1	1	1

References

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

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.

Qu, X., fei Lee, L. 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).