5-2013

Nonparametric Dynamic Panel Data Models with Interactive Fixed Effects: Sieve Estimation and Specification Testing

Liangjun SU
Singapore Management University, ljsu@smu.edu.sg

Yonghui ZHANG
Singapore Management University, yhzhang.2008@phdecons.smu.edu.sg

Nonparametric Dynamic Panel Data Models with Interactive Fixed Effects: Sieve Estimation and Specification Testing *

Liangjun Su, Yonghui Zhang
School of Economics, Singapore Management University

May 5, 2013

Abstract

In this paper we analyze nonparametric dynamic panel data models with interactive fixed effects, where the predetermined regressors enter the models nonparametrically and the common factors enter the models linearly but with individual specific factor loadings. We consider the issues of estimation and specification testing when both the cross-sectional dimension N and the time dimension T are large. We propose sieve estimation for the nonparametric function by extending Bai's (2009) principal component analysis (PCA) to our nonparametric framework. Based on the asymptotic expansion of the Gaussian quasi-log-likelihood function, we derive the convergence rate for the sieve estimator and establish its asymptotic normality. The sources of asymptotic biases are discussed and a bias-corrected estimator is provided. We also propose a consistent specification test for the linearity of the functional form by comparing the linear and sieve estimators. We establish the asymptotic distributions of the test statistic under both the null hypothesis and a sequence of Pitman local alternatives. A bootstrap procedure is proposed to obtain the bootstrap p-values and its asymptotic validity is justified. Monte Carlo simulations are conducted to investigate the finite sample performance of our estimator and test. We apply our method to an economic growth data set to study the relationship between capital accumulation and real GDP growth rate.

Key Words: Common factors; Cross section dependence; Interactive fixed effects; Linearity; Nonparametric dynamic panel; Sieve method; Specification test

JEL Classifications: C14, C33, C36

1 Introduction

Recently there has been a growing literature on large dimensional panel data models with interactive fixed effects where both the individual dimension N and time dimension T pass to infinity. By the adoption of time-varying common factors that affect the cross-sectional units with individual specific factor loadings, these models allow individual and time effects to enter the models multiplicatively and

*The authors gratefully thank Sainan Jin, Denis Leung, and Peter C. B. Phillips for their valuable comments on the paper. Su gratefully acknowledges Singapore Management University for financial support under the grant number 12-C244-SMU-008 and the Singapore Ministry of Education for Academic Research Fund under grant number MOE2012-T2-2-021. Address Correspondence to: Liangjun Su, School of Economics, Singapore Management University, 90 Stamford Road, Singapore, 178903; E-mail: ljsu@smu.edu.sg, Phone: (+65) 6828 0386.

can capture unobserved heterogeneity more flexibly than the traditional ones with additive individual or time fixed effects. As common factors affect all individuals, interactive fixed effects have become a powerful and popular tool to model cross section dependence in economics and finance. See Bai and Ng (2008) for an overview.

Most of the literature on panel data models with interactive fixed effects falls into two categories depending on whether the model includes additional regressors or not. The first category focuses on the estimation of the common components (factors and factor loadings) or the determination of the number of factors; see Bai (2003), Bai and Ng (2006a), Bai and Li (2012), and Choi (2012) for estimation, and Bai and Ng (2002) and Onatski (2009) for the determination of the number of factors. The second category concentrates on the consistent estimation of the regression coefficients. Pesaran (2006) proposes a common correlated estimator (CCE) for linear static panel data models with homogeneous or heterogeneous coefficients. Bai (2009) proposes a principal component analysis (PCA) estimator for the same model but with homogeneous coefficients and establishes its limiting distribution. Moon and Weidner (2010, 2012) reinvestigate Bai's (2009) PCA estimator and put it in the Gaussian quasi-maximum likelihood estimation (QMLE) framework; they allow dynamics in the model and show that the limiting distribution of the QMLE is independent of the number of factors used in the estimation as long as the number of factors does not fall below the true number of factors. Lu and Su (2013) propose an adaptive group Lasso method for simultaneous selection of regressors and factors and estimation in linear dynamic panel data models with interactive fixed effects. For more developments on panel data models with interactive fixed effects, see Ahn, Lee, and Schmidt (2001, 2013) for GMM approach with fixed T and large N, Zaffaroni (2010) for generalized least squares (GLS) estimation, Kapetanios and Pesaran (2007) and Greenaway-McGrevy, Han, and Sul (2012) for factor-augmented panel regression, Harding (2009) for estimation under structural restrictions from economic theory, Pesaran and Tosetti (2011) for models with both multifactor error structure and spatial correlation, Su and Chen (2013) for testing for slope homogeneity, Su, Jin, and Zhang (2013) for testing for linear functional form, among others.

Note that almost all of the above works are carried out in the parametric framework. Although economic theory dictates that some economic variables are important for the causal effects of the others, rarely does it state exactly how the variables enter an econometric model. Models derived from first principles such as utility maximization or profit maximization have particular parametric relationship under some narrow functional form restrictions. So it is not only meaningful but also necessary to extend some commonly used parametric models to the nonparametric framework. Recently, Su and Jin (2012) consider the sieve estimation of nonparametric *static* panel data models with multifactor error terms, which is a nonparametric extension of Pesaran's (2006) models; for the same models Jin and Su (2013) propose a poolability test of nonparametric functions. Freyberger (2012) studies nonparametric panel data models with multidimensional unobserved individual effects. He focuses on identification and estimation when the unobservables have a factor structure and enter an unknown structural function non-additively under fixed T and large N. However, there is still no work on the estimation of nonparametric *dynamic* panel data models where interactive fixed effects and idiosyncratic errors enter the model additively.

Linearity assumption is widely adopted in empirical works for its convenience and interpretability. A correctly specified linear model may afford precise inference whereas a badly misspecified one may lead to seriously misleading inference. So it is important to test for the correct specification of linear functional form. Recently several specification tests for linearity have been proposed in panel data models with fixed effects. Lee (2011) proposes a residual-based test to check the validity of linear dynamic models with both large N and large T; Li and Sun (2011) propose a test for *static* panel data models with both large N and large T based on an integrated squared difference between a parametric and a nonparametric estimate; Su and Lu (2013) propose a linearity test based on the comparison of the restricted estimate under the linear assumption and the unrestricted nonparametric estimate for *dynamic* panel data models with large N and fixed T. But none of these tests are applicable to panel data models with interactive fixed effects. The linear estimators for the regression coefficients and factor space generally cannot be consistent when the underlying functional form is nonlinear, and the tests on the coefficients or the number of factors based on the linear estimators could be invalid. To avoid the consequences of misspecification, Su, Jin, and Zhang (2013) propose a residual-based test for linearity that works for panel data models with *interactive* fixed effects. But they do not propose consistent estimates of the regression functions.

Based on the above observations, we consider the following nonparametric dynamic panel data models with interactive fixed effects

$$Y_{it} = g\left(X_{it}\right) + \lambda_i^{0\prime} f_t^0 + e_{it},\ i = 1, ..., N,\ t = 1, ..., T, \tag{1.1}$$

where X_{it} is a $d \times 1$ vector of observable regressors which may contain d_y lagged dependent variables $Y_{i,t-1}, ..., Y_{i,t-d_y}$ and $d_x \times 1$ vector of exogenous variables $X_{1,it}$, $g\left(\cdot\right)$ is an unknown smooth function, f_t^0 and λ_i^0 are $R \times 1$ vectors of common factors and factor loadings, respectively, and e_{it}'s are idiosyncratic error terms. Note that λ_i^0, f_t^0 and e_{it} are all unobserved. The superscript "0" in λ_i^0 and f_t^0 indicates the true parameters. We will assume that the true number of factors R is known for the theoretical part of the paper but discuss how to determine R in empirical applications.

The model specified in (1.1) is fairly general and encompasses various panel data models as special cases. If $f_t^0 = (1, \tilde{f}_t^0)'$ and $\lambda_i^0 = (\tilde{\lambda}_i^0, 1)'$ where both $\tilde{f}_t^0$ and $\tilde{\lambda}_i^0$ are scalars, the interactive fixed effects reduce to the traditional two-way fixed effects; if f_t^0 is time-invariant, i.e., $f_t^0 = \bar{f}$ for some constant vector $\bar{f}$, the interactive fixed effects become the commonly-used additive individual fixed effects. When f_t^0 is time-invariant and $g\left(X_{it}\right) = X_{it}'\theta^0$, (1.1) becomes the classical dynamic linear panel data models with individual fixed effects given by $\lambda_i^{0\prime}\bar{f}$; when f_t^0 is time-invariant and $X_{it} = Y_{i,t-1}$, (1.1) reduces to the nonparametric dynamic panel data model in Lee (2013); when f_t^0 is time-invariant and only exogenous regressors are included in X_{it}, (1.1) becomes the fixed effects nonparametric panel data model in Henderson, Carroll, and Li (2008); when f_t^0 is time-invariant and X_{it} includes both $Y_{i,t-1}$ and exogenous regressors, (1.1) becomes the general nonparametric dynamic panel data model investigated by Su and Lu (2013); when f_t^0 is time-invariant and $g\left(X_{it}\right) = h\left(X_{1,it}\right) + \theta^0 Y_{i,t-1}$, (1.1) becomes the partially linear dynamic panel data model in Baglan (2009); when $g\left(X_{it}\right) = X_{it}'\theta^0$, (1.1) becomes the model studied by Bai (2009) and Moon and Weidner (2010, 2012). These authors propose various estimators for $g\left(\cdot\right)$ (or θ^0) and $\left(\lambda_i^0, f_t^0\right)$ and establish their asymptotic properties.

Here we are mainly interested in consistent estimation and specification testing for the unknown function $g\left(\cdot\right)$ in (1.1). By combining the method of sieves with the Gaussian QMLE, we propose a nonparametric sieve estimator of $g\left(\cdot\right)$. Following Moon and Weidner (2010, 2012), we establish its consistency, derive its convergence rate based on the perturbation theory of matrix operator in Kato (1980), and establish its asymptotic normal distribution. We also discuss different sources of biases and propose a bias-corrected estimator. In addition, we consider the specification test for the commonly used linear functional form for $g\left(\cdot\right)$. Using an empirical L_2-distance, we compare two estimators for $g\left(\cdot\right)$, the linear estimator under the null hypothesis and the sieve estimator under the alternative. We establish the asymptotic distributions for the proposed test statistic under both the null hypothesis and a sequence of Pitman local alternatives. To improve the finite sample performance of the test, we also propose a bootstrap procedure to obtain the bootstrap p-values and justify its asymptotic validity.

The paper also contributes to the literature on nonlinear dynamic panel data models. Many asymptotic theories for traditional dynamic panel data models are established with large N and small T; see Arellano (2003), Baltagi (2008), and Hsiao (2003). By contrast, we derive the asymptotic results when both N and T tend to infinity simultaneously. With large T, we need to investigate the properties of (X_{it}, e_{it}) along the time dimension. Stationarity and mixing conditions are usually imposed on the observed data and the error terms. But the correlation between X_{it} and randomly realized fixed effects $\left(f_t^0, \lambda_i^0\right)$ complicates the analysis substantially. Specifically, the randomness of λ_i^0 leads to the persistence of Y_{it} along the time dimension such that we cannot directly assume mixing conditions on $\{X_{it}, e_{it}\}_{t=1}^T$, and the randomness of f_t^0 gives rise to cross-sectional dependence among $\{Y_{it}\}_{i=1}^N$. Following the idea of Hahn and Kuersteiner (2011), we adopt the concept of *conditional mixing* as defined and discussed by Prakasa Rao (2009) and Roussas (2008). We assume that $\{X_{it}, e_{it}\}_{t=1}^T$ is strong mixing conditional on the σ-field $\mathcal{D}$ generated by the factors and factor loadings and then establish the asymptotic properties of our estimator and test statistic. The concept of conditional mixing is also used in Ahn and Moon (2001), Gagliardini and Gourieroux (2012), Su and Chen (2013), and Su, Jin, and Zhang (2013).

The paper is organized as follows. In Section 2, we propose a sieve estimator for $g(\cdot)$. In Section 3, based on the asymptotic expansion of the Gaussian quasi-log-likelihood function, we prove the consistency of the sieve estimator, derive its convergence rate, establish its asymptotic normality, and provide a bias-corrected estimator. We propose a specification test statistic for linearity and study its asymptotic properties in Section 4. In Section 5, Monte Carlo simulations are conducted to investigate the finite sample performance of our estimator and test statistic. In Section 6, we apply our model to a set of real data. Section 7 concludes. All the proofs of the main theorems are relegated to the appendix. Additional proofs for the technical lemmas are provided in the online supplementary material.

NOTATION. Let $\mu_i(A)$ denote the ith largest eigenvalue (counting eigenvalues of multiplicity multiple times) of a symmetric matrix A. For an $m \times n$ matrix B, let $\|B\|_F \equiv \sqrt{\operatorname{tr}(B'B)}$ denote its Frobenius norm and $\|B\| = \sqrt{\mu_1(B'B)}$ its spectral norm. For an $n \times 1$ random vector $\xi = (\xi_1, ..., \xi_n)'$, let $\|\xi\|_p \equiv [E(\sum_{i=1}^n |\xi_i|^p)]^{1/p}$ denote its L_p-norm, and $\|\xi\|_{p,\mathcal{D}} \equiv \{E[(\sum_{i=1}^n |\xi_i|^p)|\mathcal{D}]\}^{1/p}$ its L_p-norm conditional on $\mathcal{D}$. For an $n \times m$ matrix A, let $P_A = A(A'A)^{-1}A'$ and $M_A = I_n - P_A$, where I_n is an $n \times n$ identity matrix, and $(A'A)^{-1}$ denotes some generalized inverse if A does not full column rank. For any real square matrices A and B, we use $A < B$ (or $A \leq B$) to signify that $B - A$ is positive definite (or positive semi-definite). For a positive definite symmetric matrix A, we use $A^{1/2}$ and $A^{-1/2}$ to stand for the unique symmetric matrices that satisfy $A^{1/2}A^{1/2} = A$ and $A^{-1/2}A^{-1/2} = A^{-1}$. For a real number a, let $\lfloor a \rfloor$ denote its integer part and $\lceil a \rceil$ be the largest integer that is strictly smaller than a. We use "a.s." to denote "almost surely". The operators $\xrightarrow{P}$ and $\xrightarrow{d}$ denote convergence in probability and distribution, respectively. $(N, T) \to \infty$ denotes N and T passing to infinity simultaneously.

2 Sieve-based quasi-likelihood maximum estimation

Since $g(\cdot)$ is an unknown function in (1.1), we propose to estimate $g(\cdot)$ by the method of sieves. For some excellent reviews on sieve methods, see Chen (2007, 2011). To proceed, let $p^K(x) \equiv (p_1(x), \cdots, p_K(x))'$ denote a sequence of basis functions that can approximate any square-integrable function of x very well (to be more precise later). Then we can approximate $g(x)$ in (1.1) very well by $\beta' p^K(x)$ for some $K \times 1$ vector β under some conditions. Let $K \equiv K_{NT}$ be some integer such that $K \to \infty$ as $(N, T) \to \infty$. We introduce the following notation: $p_{it,k} \equiv p_k(X_{it})$, $p_{it} \equiv p^K(X_{it})$, $P_i \equiv (p_{i1}, \cdots, p_{iT})'$, $P_{i\cdot,k} \equiv$

$(p_{i1,k},\cdots,p_{iT,k})'$, $\mathbf{P}_k \equiv (P_{1\cdot,k},\cdots,P_{N\cdot,k})'$, $Y_i \equiv (Y_{i1},\cdots,Y_{iT})'$, $\mathbf{Y} \equiv (Y_1,\cdots,Y_N)'$, $f^0 \equiv (f_1^0,\cdots,f_T^0)'$, $\lambda^0 \equiv \left(\lambda_1^0,\cdots,\lambda_N^0\right)'$. We use β^0 to denote the true vector of coefficients β in the sieve approximation of $g(x)$ given basis $p^K(x)$. Here we suppress the dependence of p_{it}, β^0, and β on K for notational simplicity.

To estimate g, we consider the following approximating linear panel data models:

$$Y_{it} = p_{it}'\beta^0 + \lambda_i^{0\prime} f_t^0 + u_{it} \tag{2.1}$$

where $u_{it} \equiv e_{it} + e_{g,it}$ is the new error term, and $e_{g,it} \equiv g(X_{it}) - p_{it}'\beta^0$ represents the sieve approximation error. Let $u_i \equiv (u_{i1},\cdots,u_{iT})'$ and $\mathbf{u} \equiv (u_1,\cdots,u_N)'$. In matrix notation, (2.1) can be rewritten as

$$\mathbf{Y} = \sum_{k=1}^{K} \beta_k^0 \mathbf{P}_k + \lambda^0 f^{0\prime} + \mathbf{u}. \tag{2.2}$$

Following Bai (2009) and Moon and Weidner (2010) we propose to estimate the model in (2.2) by the Gaussian QMLE method. Specifically, the QMLE estimator of $\left(\beta^0,\lambda^0,f^0\right)$ is given by $(\hat{\beta},\hat{\lambda},\hat{f}) = \arg\min_{(\beta,\lambda,f)} \mathcal{L}(\beta,\lambda,f)$, where $\mathcal{L}(\beta,\lambda,f)$ is the approximating negative quasi-log-likelihood function:

$$\mathcal{L}(\beta,\lambda,f) = \frac{1}{NT}\operatorname{tr}\left[\left(\mathbf{Y} - \sum_{k=1}^{K}\beta_k\mathbf{P}_k - \lambda f'\right)'\left(\mathbf{Y} - \sum_{k=1}^{K}\beta_k\mathbf{P}_k - \lambda f'\right)\right], \tag{2.3}$$

$\beta = (\beta_1,\cdots,\beta_K)'$, $f \equiv (f_1,\cdots,f_T)'$, and $\lambda \equiv (\lambda_1,\cdots,\lambda_N)'$. In particular, $\hat{\beta} = \arg\min_{\beta\in\mathbb{R}^K} L_{NT}(\beta)$ where $L_{NT}(\beta)$ is the profile approximating negative quasi-log-likelihood function:

$$\begin{aligned} L_{NT}(\beta) &= \min_{\lambda,f} \mathcal{L}_{NT}(\beta,\lambda,f) && (2.4)\\ &= \min_{f} \frac{1}{NT}\operatorname{tr}\left[\left(\mathbf{Y} - \sum_{k=1}^{K}\beta_k\mathbf{P}_k\right) M_f \left(\mathbf{Y} - \sum_{k=1}^{K}\beta_k\mathbf{P}_k\right)'\right] && (2.5)\\ &= \frac{1}{NT}\sum_{t=R+1}^{T}\mu_t\left[\left(\mathbf{Y} - \sum_{k=1}^{K}\beta_k\mathbf{P}_k\right)'\left(\mathbf{Y} - \sum_{k=1}^{K}\beta_k\mathbf{P}_k\right)\right]. && (2.6) \end{aligned}$$

See Moon and Weidner (2010) for the demonstration of equivalence of the above three expressions. Based on (2.6), one only needs to calculate the $T-R$ smallest eigenvalues of a $T\times T$ matrix at each step of the numerical optimization over β. Note that the objective function $L_{NT}(\beta)$ is neither convex nor differentiable with respect to β. Multiple starting values for numerical optimization should be used to find the global minimum. After obtaining $\hat{\beta}$, one estimates $g(x)$ by

$$\hat{g}(x) = p^K(x)'\hat{\beta}. \tag{2.7}$$

The expression in (2.6) is our starting point to establish the asymptotic theory. Following Moon and Weidner (2010), we also adopt the perturbation theory for linear operator in Kato (1980) to derive the asymptotic expansion of $L_{NT}(\beta)$ around β^0. The key idea is to form the following decomposition

$$\mathbf{Y} - \sum_{k=1}^{K}\beta_k\mathbf{P}_k = \underbrace{\lambda^{0\prime} f^0}_{\text{leading term}} + \underbrace{\sum_{k=1}^{K}\left(\beta_k^0 - \beta_k\right)\mathbf{P}_k + \mathbf{e} + \mathbf{e}_g}_{\text{perturbation terms}} \tag{2.8}$$

where $\mathbf{e}_g$ is an $N\times T$ matrix whose (i,t)th element is $g(X_{it}) - p_{it}'\beta^0$. Compared with the decomposition in eqn. (3.1) in Moon and Weidner (2010), (2.8) has a diverging number of perturbation terms (as $K\to\infty$)

and includes the additional sieve approximation error term. If there were no perturbation term in (2.8), $L_{NT}(\beta)$ would be equal to zero. By the continuity of the eigenvalue operator, $L_{NT}(\beta)$ should be close to zero when these perturbation terms are small enough. Using the perturbation theory of linear operators, we can work out an expansion of $L_{NT}(\beta)$ in the perturbation terms and show that this expansion is convergent as long as the spectral norm of the perturbation terms is sufficiently small. Based on the first order asymptotic theory for QMLE, we show the consistency of $\hat{g}(x)$ and establish its asymptotic normality under suitable conditions.

3 Asymptotic properties of $\hat{g}(\cdot)$

In this section, we first derive the convergence rate for $\hat{g}(x)$ and then establish its asymptotic distribution. We also analyze the sources of asymptotic biases and propose a bias-corrected estimator.

3.1 Convergence rate of $\hat{g}(\cdot)$

To estimate the unknown function by the method of sieves, we assume that $g(x)$ is a smooth function. Let $\mathcal{X} \equiv \mathcal{Y} \times \mathcal{X}_1 \subset \mathbb{R}^{d_y} \times \mathbb{R}^{d_x}$ be the support of X_{it}. Typical approximation and estimation of regression functions require that $\mathcal{X}$ be compact; see Newey (1997). In our model, it seems restrictive to impose the compactness of $\mathcal{X}$ because of the presence of lagged dependent variables. To allow for the unboundedness of $\mathcal{X}$, we follow Chen, Hong, and Tamer (2005), Blundell, Chen, and Kristensen (2007), and Su and Jin (2012) and use a weighted sup-norm metric defined as

$$\|g\|_{\infty,\omega} \equiv \sup_{x\in\mathcal{X}} |g(x)| \left[1+\|x\|^2\right]^{-\omega/2} \text{ for some } \omega \geq 0. \tag{3.1}$$

If $\omega = 0$, the norm defined in (3.1) is the usual sup-norm which is suitable for the case of compact support.

Recall that a typical smoothness assumption requires that a function $g : \mathcal{X} \to \mathbb{R}$ belong to a Hölder space. Let $\boldsymbol{\alpha} \equiv (\alpha_1, \cdots, \alpha_d)'$ denote a d-vector of non-negative integers and $|\boldsymbol{\alpha}| \equiv \sum_{l=1}^{d} \alpha_l$. For any $x = (x_1, \cdots, x_d)$, the $|\boldsymbol{\alpha}|$th derivative of $g : \mathcal{X} \to \mathbb{R}$ is denoted as $\nabla^{\boldsymbol{\alpha}} g(x) \equiv \partial^{|\boldsymbol{\alpha}|} g(x)/(\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d})$. The Hölder space $\Lambda^\gamma(\mathcal{X})$ of order $\gamma > 0$ is a space of functions $g : \mathcal{X} \to \mathbb{R}$ such that the first $\lceil\gamma\rceil$ derivatives are bounded, and the $\lceil\gamma\rceil$th derivatives are Hölder continuous with the exponent $\gamma - \lceil\gamma\rceil \in (0, 1]$. Define the Hölder norm:

$$\|g\|_{\Lambda^\gamma} \equiv \sup_{x\in\mathcal{X}} |g(x)| + \max_{|\boldsymbol{\alpha}|=\lceil\gamma\rceil} \sup_{x\neq x^*} \frac{|\nabla^{\boldsymbol{\alpha}} g(x) - \nabla^{\boldsymbol{\alpha}} g(x^*)|}{\|x-x^*\|^{\gamma-\lceil\gamma\rceil}}.$$

The following definition is adopted from Chen, Hong, and Tamer (2005).

Definition 1. *Let $\Lambda^\gamma(\mathcal{X},\omega) \equiv \left\{g : \mathcal{X} \to \mathbb{R} \text{ such that } g(\cdot)[1+||\cdot||^2]^{-\omega/2} \in \Lambda^\gamma(\mathcal{X})\right\}$ denote a weighted Hölder space of functions. A weighted Hölder ball with radius c is*

$$\Lambda_c^\gamma(\mathcal{X},\omega) \equiv \left\{g \in \Lambda^\gamma(\mathcal{X},\omega) : \left\|g(\cdot)[1+||\cdot||^2]^{-\omega/2}\right\|_{\Lambda^\gamma} \leq c < \infty\right\}.$$

Function $g(\cdot)$ is said to be $H(\gamma,\omega)$-smooth on $\mathcal{X}$ if it belongs to a weighted Hölder ball $\Lambda_c^\gamma(\mathcal{X},\omega)$ for some $\gamma > 0$, $c > 0$ and $\omega \geq 0$.

Let $\mathbf{P}_{(a)} \equiv \sum_{k=1}^{K} a_k \mathbf{P}_k$, $Q_{pp,NT}^{(a)} \equiv (NT)^{-1} \mathbf{P}_{(a)} \mathbf{P}_{(a)}'$, and $Q_{pp}^{(a)} \equiv E_{\mathcal{D}}[Q_{pp,NT}^{(a)}]$, where $a = (a_1, ..., a_K)'$ with $\|a\| = 1$, and $\mathcal{D} \equiv \sigma\left(f^0, \lambda^0\right)$ is the σ-field generated by f^0 and λ^0. Let $Q_{wpp,NT} \equiv \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} w_{it} p_{it} p_{it}'$

and $Q_{wpp} \equiv E_{\mathcal{D}}\left[Q_{wpp,NT}\right]$, where $w_{it} = w\left(X_{it}\right)$ and $w\left(\cdot\right)$ is some nonnegative integrable function. Let $W_{NT} \equiv \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} Z_{it}Z_{it}'$, where

$$Z_{it} \equiv p_{it} - \frac{1}{N}\sum_{j=1}^{N}\alpha_{ij}p_{jt} - \frac{1}{T}\sum_{s=1}^{T}\eta_{ts}p_{is} + \frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\alpha_{ij}\eta_{ts}p_{js}, \tag{3.2}$$

$\alpha_{ij} \equiv \lambda_i^{0\prime}(\frac{1}{N}\lambda^{0\prime}\lambda^0)^{-1}\lambda_j^0$, and $\eta_{ts} \equiv f_t^{0\prime}(\frac{1}{T}f^{0\prime}f^0)^{-1}f_s^0$. Let $W \equiv E_{\mathcal{D}}(W_{NT})$ and $Z_i \equiv (Z_{i1}, ..., Z_{iT})' \equiv M_{f^0}P_i - N^{-1}\sum_{j=1}^{N}\alpha_{ij}M_{f^0}P_j$.

We first state assumptions to be used in the derivation of convergence rate for the sieve estimator.

Assumption 1. (i) $\lambda^{0\prime}\lambda^0/N \xrightarrow{P} \Sigma_\lambda$ as $N \to \infty$ and $0 < \underline{c}_\lambda \leq \mu_R\left(\Sigma_\lambda\right) \leq \mu_1\left(\Sigma_\lambda\right) \leq \overline{c}_\lambda < \infty$;

(ii) $f^{0\prime}f^0/T \xrightarrow{P} \Sigma_f$ as $T \to \infty$ and $0 < \underline{c}_f \leq \mu_R\left(\Sigma_f\right) \leq \mu_1\left(\Sigma_f\right) \leq \overline{c}_f < \infty$;

(iii) $\|\mathbf{e}\|/\sqrt{NT} = O_P(\delta_{NT}^{-1})$ where $\delta_{NT} \equiv \sqrt{\min\left(N,T\right)}$.

Assumption 2.(i) $\|Q_{wpp,NT} - Q_{wpp}\| = o_P\left(1\right)$ and $0 < \underline{c}_Q \leq \mu_K\left(Q_{wpp}\right) \leq \mu_1\left(Q_{wpp}\right) \leq \overline{c}_Q < \infty$ a.s. for given $w\left(\cdot\right)$ and all K as $(N,T) \to \infty$;

(ii) $\|W_{NT} - W\| = o_P\left(1\right)$ and $0 < \underline{c}_W \leq \mu_K\left(W\right) \leq \mu_1\left(W\right) \leq \overline{c}_W < \infty$ a.s. for all K as $(N,T) \to \infty$;

(iii) There exist positive constants $\underline{C}$ and $\overline{C}$ such that $\min_{\{a\in\mathbb{R}^K,\|a\|=1\}}\sum_{l=2R+1}^{N}\mu_l(Q_{pp,NT}^{(a)}) \geq \underline{C} > 0$ and $\mu_1(Q_{pp,NT}^{(a)}) = \left\|\mathbf{P}_{(a)}\right\|/\sqrt{NT} \leq \overline{C} < \infty$ for any $a \in \mathbb{R}^K$ with $\|a\| = 1$ as $(N,T) \to \infty$.

Assumption 3. (i) $g(\cdot)$ is $H(\gamma,\omega)$-smooth on $\mathcal{X}$ for some $\gamma > d/2$ and $\omega \geq 0$;

(ii) For any $H(\gamma,\omega)$-smooth function $g\left(x\right)$, there exists a linear combination of basis functions $\Pi_{\infty,K}g \equiv \beta_g'p^K\left(\cdot\right)$ in the sieve space $\mathcal{G}_K \equiv \left\{g\left(\cdot\right) = \beta'p^K\left(\cdot\right)\right\}$ such that $\|g\left(\cdot\right) - \Pi_{\infty,K}g\|_{\infty,\bar{\omega}} = O\left(K^{-\gamma/d}\right)$;

(iii) $\text{plim}_{(N,T)\to\infty}\left(NT\right)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(1 + ||X_{it}||^2\right)^{\bar{\omega}}\left(w_{it} + 1\right) < \infty$ for some $\bar{\omega} > \omega + \gamma$;

(v) $||\sum_{i=1}^{N}\sum_{t=1}^{T}p_{it}e_{it}|| = O_P(\sqrt{NTK})$;

(vi) $||\sum_{i=1}^{N}\left[Z_i'e_i - E_{\mathcal{D}}\left(Z_i'e_i\right)\right]|| = O_P(\sqrt{NTK})$ and $||\sum_{i=1}^{N}E_{\mathcal{D}}(Z_i'e_i)|| = O_P(\sqrt{NK/T})$.

Assumption 4. As $(N,T) \to \infty$, $K \to \infty$ and $K\delta_{NT}^{-2} \to 0$.

Assumptions 1(i)-(ii) are widely used in the literature on panel data models with interactive fixed effects; see Bai (2009), Moon and Weidner (2010, 2012), and Su and Chen (2013). Assumption 1(iii) is also adopted by Moon and Weidner (2010) and can be verified for various error processes; see the supplementary material in Moon and Weidner (2010). Assumptions 2(i)-(ii) impose restrictions on the eigenvalues of conditional probability limits of $Q_{wpp,NT}$ and W_{NT}. Assumption 2(iii) is essential for the consistency and it requires that $\mathbf{P}_{(a)}$ be still full rank after one projects the sieve terms onto the factor space (f^0) and factor loading space (λ^0). In other words, we need that the sieve terms are all high rank regressors as defined by Moon and Weidner (2010). The low rank regressors such as time-invariant or individual-invariant regressors deserve special attention. Assumption 2(iii) implies that $\left\|\mathbf{P}_{(a)}\right\|/\sqrt{NT}$ is uniformly bounded. Assumption 3(i) imposes smooth conditions on $g\left(\cdot\right)$. Assumption 3(ii) quantifies the approximation error of functions in $H(\gamma,\omega)$ by a linear combination of basis functions. Assumption 3(iii) is used to handle unbounded support, which can be replaced by some conditions on the tail behavior of the marginal density of X_{it} as in Chen, Hong, and Tamer (2005) and Su and Jin (2012). Assumptions 3(ii)-(iii) jointly imply that $\left(NT\right)^{-1/2}\|\mathbf{e}_g\|_F = O_P\left(K^{-\gamma/d}\right)$; see Lemma A.2 in Su and Jin (2012). Assumptions 3(v)-(vi) can be verified for various data generating processes (DGPs) and various sieve bases. The second part of (vi) is similar to the assumption on Φ_K in Lee (2013). If X_{it} excludes lagged dependent variables, $E_{\mathcal{D}}(Z_i'e_i) = 0$ and then Assumption 3(vi) reduces to $\left(NT\right)^{-1/2}\sum_{i=1}^{N}Z_i'e_i = O_P(K^{1/2})$. In the next section, we will provide primitive conditions on the DGPs and sieve bases. Assumption 4 imposes conditions on K.

Let $\Phi \equiv \lambda^0 \left(\lambda^{0\prime}\lambda^0\right)^{-1} \left(f^{0\prime}f^0\right)^{-1} f^{0\prime}$. Let $C_{NT}^{(1)}$ and $C_{NT}^{(2)}$ be $K \times 1$ vectors whose kth elements are respectively given by

$$\begin{aligned} C_{NT,k}^{(1)} &\equiv \frac{1}{NT}\mathrm{tr}\left(M_{\lambda^0}\mathbf{P}_k M_{f^0}\mathbf{u}'\right), & (3.3)\\ C_{NT,k}^{(2)} &\equiv -\frac{1}{NT}\mathrm{tr}\left(\mathbf{P}_k\Phi'\mathbf{u}M_{f^0}\mathbf{u}'M_{\lambda^0} + \mathbf{P}_k M_{f^0}\mathbf{u}'M_{\lambda^0}\mathbf{u}\Phi' + \mathbf{P}_k M_{f^0}\mathbf{u}'\Phi\mathbf{u}'M_{\lambda^0}\right) & (3.4)\\ &\equiv C_{NT,k}^{(2,a)} + C_{NT,k}^{(2,b)} + C_{NT,k}^{(2,c)}, & (3.5)\end{aligned}$$

where $C_{NT,k}^{(2,s)}$ denotes the kth element of $C_{NT}^{(2,s)}$ for $s = a$, b, and c. We derive an asymptotic expansion for $\hat{g}(x)$ and establish its convergence rate in the following theorem.

Theorem 3.1 *Suppose that Assumptions 1-4 hold. Then*

$$\hat{g}(x) - g(x) = p^K(x)' W_{NT}^{-1}\left(C_{NT}^{(1)} + C_{NT}^{(2)}\right) + \left[p^K(x)'\beta^0 - g(x)\right] + p^K(x)' R_{NT}, \tag{3.6}$$

where R_{NT} is a $K \times 1$ vector with $\|R_{NT}\| = O_P[(K^{-\gamma/d} + \sqrt{K}\delta_{NT}^{-2})(\delta_{NT}^{-1/2} + K^{-\gamma/(2d)})]$. Further, suppose $\mu_1[\int_{\mathcal{X}} p^K(x) p^K(x)' w(x)\,dx] < \infty$ and $\int_{\mathcal{X}}\left(1 + \|x\|^2\right)^{\bar{\omega}} w(x)\,dx < \infty$. Then

$$\begin{aligned} \int_{\mathcal{X}} \left[\hat{g}(x) - g(x)\right]^2 w(x)\,dx &= O_P\left(K/(NT) + K\delta_{NT}^{-4} + K^{-2\gamma/d}\right), & (3.7)\\ \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\hat{g}(X_{it}) - g(X_{it})\right]^2 w(X_{it}) &= O_P\left(K/(NT) + K\delta_{NT}^{-4} + K^{-2\gamma/d}\right). & (3.8)\end{aligned}$$

Remark 1. In (3.6), $\hat{g}(x) - g(x)$ is decomposed into three parts: the first part contributes to the asymptotic variance and bias, the second part signals the sieve approximation error, and the third part summarizes higher order terms from the asymptotic expansion of $L_{NT}(\hat{\beta})$. Theorem 3.1 also states the convergence rates for both the weighted integrated mean square error (MSE) and weighted sample mean square error in (3.7) and (3.8), respectively. $O_P\left(K/(NT) + K\delta_{NT}^{-4}\right)$ and $O_P\left(K^{-2\gamma/d}\right)$ come from the first and second terms in (3.6), respectively. Apparently, $K/(NT) + K\delta_{NT}^{-4} = O(K\delta_{NT}^{-4})$, but we keep the first term in the expression as it corresponds to the usual variance term for a sieve estimate. It is easy to show that the optimal choice of K, say K_{opt}, to minimize the integrated or sample MSE is of order $\delta_{NT}^{4/[(2\gamma/d)+1]}$, yielding the minimized integrated or sample MSE of order $O_P(\delta_{NT}^{-4/[d/(2\gamma)+1]})$. If there were no lagged dependent variables in X_{it} and no cross-sectional heteroskedasticity and serial correlation in the error terms conditional on $\mathcal{D}$, then the rates in (3.7) and (3.8) should be $O_P\left(K^{-2\gamma/d} + K/(NT)\right)$, and K_{opt} would be proportional to $(NT)^{1/[(2\gamma/d)+1]}$.

3.2 Asymptotic distribution of $\hat{g}(x)$

To study the asymptotic distribution of $\hat{g}(x)$, we introduce the concept of conditional strong mixing.

Definition 2. *Let $(\Omega, \mathcal{A}, P)$ be a probability space and $\mathcal{B}$ be a sub-σ-algebra of $\mathcal{A}$. Let $P_{\mathcal{B}}(\cdot) \equiv P(\cdot|\mathcal{B})$. Let $\{\xi_t, t \geq 1\}$ be a sequence of random variables defined on $(\Omega, \mathcal{A}, P)$. A sequence $\{\xi_t, t \geq 1\}$ is said to be conditionally strong mixing given $\mathcal{B}$ (or $\mathcal{B}$-strong-mixing) if there exists a nonnegative $\mathcal{B}$-measurable random variable $\alpha_{\mathcal{B}}(t)$ converging to 0 a.s. as $t \to \infty$ such that*

$$\left|P_{\mathcal{B}}(A \cap B) - P_{\mathcal{B}}(A) P_{\mathcal{B}}(B)\right| \leq \alpha_{\mathcal{B}}(t) \text{ a.s.} \tag{3.9}$$

for all $A \in \sigma(\xi_1, ..., \xi_k)$, $B \in \sigma(\xi_{k+t}, \xi_{k+t+1}, ...)$ and $k \geq 1$, $t \geq 1$.

The above definition is due to Prakasa Rao (2009). When one takes $\alpha_{\mathcal{B}}(t)$ as the supremum of the left hand side object in (3.9) over the set $\{A \in \sigma(\xi_1, ..., \xi_k), B \in \sigma(\xi_{k+t}, \xi_{k+t+1}, ...), k \geq 1\}$, we refer to it as the $\mathcal{B}$-strong-mixing coefficient.

Define

$$\tilde{W}_{NT} \equiv \frac{1}{NT}\sum_{i=1}^{N}\tilde{Z}_i'\tilde{Z}_i \equiv \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{Z}_{it}\tilde{Z}_{it}' \text{ and } \tilde{\Omega}_{NT} \equiv \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{Z}_{it}\tilde{Z}_{it}'e_{it}^2,$$

where $\tilde{Z}_i \equiv (\tilde{Z}_{i1}', \cdots, \tilde{Z}_{iT}')' = P_i - P_{f^0}E_{\mathcal{D}}(P_i) - N^{-1}\sum_{j=1}^{N}\alpha_{ij}M_{f^0}E_{\mathcal{D}}(P_j)$, $\tilde{Z}_{it} \equiv p_{it} - N^{-1}\sum_{j=1}^{N}\alpha_{ij}E_{\mathcal{D}}(p_{jt})$ $-T^{-1}\sum_{s=1}^{T}\eta_{ts}E_{\mathcal{D}}(p_{is}) + (NT)^{-1}\sum_{j=1}^{N}\sum_{s=1}^{T}\alpha_{ij}\eta_{ts}E_{\mathcal{D}}(p_{js})$. Let $\tilde{W} \equiv E_{\mathcal{D}}(\tilde{W}_{NT})$ and $\tilde{\Omega} \equiv E_{\mathcal{D}}(\tilde{\Omega}_{NT})$. We add the following assumptions.

Assumption 5. (i) For each $i = 1, ..., N$, $\{(X_{it}, \varepsilon_{it}) : t = 1, 2, ...\}$ is $\mathcal{D}$-strong-mixing with mixing coefficients $\{\alpha_{\mathcal{D},i}(t), 1 \leq t \leq T-1\}$. $\alpha_{\mathcal{D}}(\cdot) \equiv \max_{1\leq i\leq N}\alpha_{\mathcal{D},i}(\cdot)$ satisfies $\sum_{s=1}^{\infty}s^2\alpha_{\mathcal{D}}^{(1+\delta)/(2+\delta)}(s) < \infty$ where δ is given in Assumption 6;

(ii) $E\left[e_{it}|\mathcal{F}_0^{t-1}\right] = 0$ a.s. where $\mathcal{F}_0^{t-1} \equiv \sigma\{\lambda^0, f^0, (X_{it}, X_{i,t-1}, e_{i,t-1}, X_{i,t-2}, e_{i,t-1}, \cdots)_{i=1}^N\}$;

(iii) $(e_{it}, X_{it}) \perp (e_{js}, X_{js})\,|\mathcal{D}$ for all $i \neq j$ and all t, $s = 1, ..., T$, where $A \perp B|C$ denotes independence between A and B given C.

Assumption 6. There exists $\delta > 0$ such that

(i) $\sup_{i,t}E|e_{it}|^{8+4\delta} < \infty$;

(ii) $\sup_i E\left\|\lambda_i^0\right\|^{8+4\delta} < \infty$ and $\sup_t E\left\|f_t^0\right\|^{8+4\delta} < \infty$;

(iii) $\sup_k\sup_{i,t}E|p_{it,k}|^{8+4\delta} < \infty$ and $\sup_k\sup_{i,t}E|\tilde{Z}_{it,k}|^{8+4\delta} < \infty$, where $\tilde{Z}_{it,k}$ is the kth element of $\tilde{Z}_{it}$.

Assumption 7. There exist constants $\underline{c}_w$, $\bar{c}_w$, $\underline{c}_\Omega$, and $\bar{c}_\Omega$ that do not depend on K, N, and T such that $0 < \underline{c}_w \leq \mu_K(\tilde{W}) \leq \mu_1(\tilde{W}) \leq \bar{c}_w < \infty$ a.s. and $0 < \underline{c}_\Omega \leq \mu_K(\tilde{\Omega}) \leq \mu_1(\tilde{\Omega}) \leq \bar{c}_\Omega < \infty$ a.s. for all K as $(N, T) \to \infty$.

Assumption 8. As $(N, T) \to \infty$, $K \to \infty$ and $\max\{\sqrt{NT}K^{-\gamma/d}, K\delta_{NT}^{-1}, \sqrt{NTK}\delta_{NT}^{-5/2}\} \to 0$.

Assumptions 5(i) imposes strong mixing on $\{(X_{it}, e_{it})\}_{t=1}^T$ conditional on $\mathcal{D}$. Its unconditional version is widely used in the time series literature; see, e.g., Bosq (1998) and Fan and Yao (2003). In the time series literature, one can find various sufficient conditions for the strong mixing property of a nonlinear autoregressive (AR) process with identically and independently distributed (IID) errors or nonlinear ARCH/GARCH type of errors; see Tjøstheim (1990) and Doukhan (1994) for nonlinear AR process with IID errors, Fan, Yao, and Cai (2003) for functional coefficient AR processes, and Meitz and Saikkonen (2010) for nonlinear AR-ARCH/GARCH processes. When the nonlinear time series contains exogenous regressors, sufficient conditions are also available for the strong mixing property; see Doukhan (1994) and Chen, Racine, and Swanson (2001) for nonlinear ARX processes where exogenous variables and errors are both IID, Franke and Diagne (2006) for nonlinear ARX-ARCHX processes but the exogenous variables are lagged exogenous variables, and Hahn and Kuersteiner (2010) for dynamic Tobit models with mixing exogenous regressors which follow an AR process. Similar tools used in the time series literature can be used to establish the conditional strong mixing property for $\{Y_{it}\}_{t=1}^T$ in our framework. On the other hand, if one assumes that the interactive fixed effects are not random (which is analogous to treating the individual fixed effects as nonrandom in a classical linear panel data model), it suffices to use the concept of strong mixing.[1] Assumption 5(ii) imposes a martingale difference sequence (m.d.s.) condition on $\{(e_{it}, X_{it}), \mathcal{F}_0^t\}_{t=1}^T$. Assumption 5$(iii)$ imposes the conditional independence between (e_{it}, X_{it}) and (e_{js}, X_{js}) for $i \neq j$ given $\mathcal{D}$. This assumption implies that all the cross-sectional dependence comes from the common factor f_t^0. We can relax this assumption to allow for weak cross-sectional dependence among $\{(X_{1,it}, e_{it})\}_{i=1}^N$ conditional on $\mathcal{D}$ at the cost of more complicated proofs.

Assumption 6 imposes moment conditions on e_{it}, λ_i^0, f_t^0, and $p_{it,k}$. Assumption 6(ii) imposes the existence of $(8+4\delta)$th moments for the factors and factor loadings and thus relaxes the uniform boundedness of $\|f_t^0\|$ and $\|\lambda_i^0\|$ in Moon and Weidner (2010, 2012). Assumption 6(iii) is a little stronger than what is typically assumed for sieve estimation in the IID framework (e.g., Newey, 1997), but is more general than that in Lee (2013) where a uniform bound over a truncated support is used. In the case of compact support, it is generally assumed that $\sup_{x \in \mathcal{X}} \|p^K(x)\| = O_P(\zeta(K))$ for a non-decreasing function $\zeta(\cdot)$. But for the case of infinite support, this assumption is not reasonable for general sieves except for some special sieves (e.g., Fourier series and Hermite polynomials) that can automatically deal with the tail behavior or are uniformly bounded over the infinite support. For this reason, we impose moment conditions on $p_{it,k}$ instead. One direct implication of Assumption 6(iii) is that $\sup_{i,t} E\|p_{it}\| = O_P(K^{1/2})$, which allows for cubic splines or trigonometric series, but excludes polynomial functions. See Newey (1997) for more discussions on sieves. In addition, we remark that it is possible to relax this assumption to $\sup_k \sup_{i,t} E|p_{it,k}|^{8+4\delta} < \zeta_0(K)$ for some non-decreasing function $\zeta_0(\cdot)$ to include more sieve bases. Assumption 7 imposes some restrictions on the eigenvalues of $\tilde{W}$ and $\tilde{\Omega}$. Assumption 8 specifies the relative rates at which N, T, and K pass to infinity. Note that we allow for $N/T = c \in [0, \infty]$. When $N/T \in (0, \infty)$, the assumption reduces to $N/K^{\gamma/d} + K^2/N \to 0$, i.e., $K \in (N^{d/\gamma}, N^{1/2})$.

3.2.1 Asymptotic distribution

Let $V_K(x) \equiv p^K(x)' \tilde{W}^{-1}\tilde{\Omega}\tilde{W}^{-1}p^K(x)$ and $A_{NT} \equiv \sqrt{NT}V_K^{-1/2}(x)$. Let b_1, b_2, and b_3 denote $K \times 1$ vectors whose kth elements are respectively given by $b_{1,k} \equiv \frac{1}{N}\mathrm{tr}\left[P_{f^0}E_{\mathcal{D}}(\mathbf{e}'\mathbf{P}_k)\right]$, $b_{2,k} \equiv \frac{1}{T}\mathrm{tr}[E_{\mathcal{D}}(\mathbf{e}\mathbf{e}')M_{\lambda^0}\mathbf{P}_k\Phi]$, and $b_{3,k} \equiv \frac{1}{N}\mathrm{tr}\left[E_{\mathcal{D}}(\mathbf{e}'\mathbf{e})M_{f^0}\mathbf{P}_k'\Phi'\right]$. Define

$$B_K(x) \equiv -A_{NT}p^K(x)' \tilde{W}^{-1}\left(T^{-1}b_1 + N^{-1}b_2 + T^{-1}b_3\right) \equiv -\kappa_{NT}b_1(x) - \kappa_{NT}^{-1}b_2(x) - \kappa_{NT}b_3(x), \quad (3.10)$$

where $\kappa_{NT} \equiv \sqrt{N/T}$. Clearly, $b_s(x) = V_K^{-1/2}(x)p^K(x)'\tilde{W}^{-1}b_s$ for $s = 1, 2, 3$. We establish the asymptotic normality of $\hat{g}(x)$ in the following theorem.

Theorem 3.2 *Let Assumptions 1-8 hold. Then* $A_{NT}[\hat{g}(x) - g(x)] - B_K(x) \xrightarrow{d} N(0,1)$ *as* $(N,T) \to \infty$.

Remark 2. The proof of the above theorem is quite complicated despite the fact that we establish the asymptotic normality by a version of martingale central limit theorem (CLT). Let $a_{NT} \equiv A_{NT}p^K(x)'W_{NT}^{-1}$. Theorem 3.1 suggests that the leading terms in the expansion of $A_{NT}[\hat{g}(x) - g(x)]$ are given by $a_{NT}C_{NT}^{(1)}$, $a_{NT}C_{NT}^{(2,a)}$, and $a_{NT}C_{NT}^{(2,b)}$. $a_{NT}C_{NT}^{(1)}$ contributes to both the asymptotic variance and asymptotic bias $(-\kappa_{NT}b_1(x))$. The latter also arises in linear dynamic panel data models and is caused by the endogeneity of Z_{it} defined in (3.2): $E_{\mathcal{D}}(Z_{it}e_{it}) = -T^{-1}\sum_{s=t+1}^{T}\left(1 - N^{-1}\alpha_{ii}\right)\eta_{ts}E_{\mathcal{D}}(p_{is}e_{it}) \neq 0$ by Assumption

[1] An alternative for strong mixing is *Near Epoch Dependence* (NED), which is a much weaker condition and easily verified for many DGPs; see Gallant (1987), Gallant and White (1988), Davidson (1994), Pötscher and Prucha (1997), and de Jong (2009). However, there are no works on the sufficient conditions for the NED of $\{Y_{it}\}_{t=1}^T$ when the models include both nonlinear ARX and nonlinear ARCHX/GARCHX error. We conjecture that one can apply NED to study our model but the proofs are much more complicated in various places. For this reason, we adopt the notion of conditional strong mixing.

5(iii). It is easy to see that an equivalent expression for b_1 is $b_1 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\eta_{ts}E_{\mathcal{D}}\left(p_{is}e_{it}\right)$. $a_{NT}C_{NT}^{(2,a)}$ contributes to the second bias term, i.e., $-\kappa_{NT}^{-1}b_2\left(x\right)$, and is caused by cross-sectional heteroskedasticity of errors conditional on $\mathcal{D}$; $a_{NT}C_{NT}^{(2,b)}$ contributes to the third bias term, i.e., $-\kappa_{NT}b_3\left(x\right)$, and is caused by serial correlation and heteroskedasticity of errors conditional on $\mathcal{D}$. In the special case where e_{it}'s are IID conditional on $\mathcal{D}$ across both i and t, the last two bias terms disappear.

3.3 Bias correction

In this section, we propose a bias-corrected estimator for $g\left(x\right)$. Let $\mathbf{i}_t$ be a $T\times 1$ unit vector that has unity at position t. For an $N\times N$ matrix A, define the diagonal truncation of A as $A^{\text{truncD}} = \text{diag}(A)$, whose (i,j)th element is given by $A_{ij}\mathbf{1}\left(i=j\right)$ with $\mathbf{1}(\cdot)$ being the usual indicator function. Let $\Gamma\left(\cdot\right)$ be the truncation kernel: $\Gamma\left(s\right) = \mathbf{1}\left(|s|\le 1\right)$. Let M_T be a bandwidth parameter such that $M_T/T + 1/M_T \to 0$ as $T\to\infty$. The right truncation of matrix B is defined by $B^{\text{truncR}} = \sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\Gamma((s-t)/M_T)\mathbf{i}_t\mathbf{i}_t'B\mathbf{i}_s\mathbf{i}_s'$.

To construct consistent estimates for the asymptotic bias and variance, we need consistent estimates of λ^0 and f^0 under suitable identification restrictions. We use the same identification restrictions as Bai (2009): $f'f/T = I_R$ and $\lambda'\lambda$ =diagonal matrix. Given $\hat{\beta}$, we can obtain $(\hat{\lambda},\hat{f})$ as the solution to the following set of nonlinear equations:

$$\left[\frac{1}{NT}\sum_{i=1}^{N}\left(Y_i - P_i\hat{\beta}\right)\left(Y_i - P_i\hat{\beta}\right)'\right]\hat{f} = \hat{f}\mathbf{V}_{NT}, \tag{3.11}$$

where $\mathbf{V}_{NT}$ is a diagonal matrix that consists of the largest R eigenvalues of the matrix in the above bracket, arranged in descending order, and

$$\hat{\lambda} \equiv \left(\hat{\lambda}_1,\cdots,\hat{\lambda}_N\right)' = T^{-1}\left[\hat{f}'\left(Y_1 - P_1\hat{\beta}\right),\cdots,\hat{f}'\left(Y_N - P_N\hat{\beta}\right)\right]'. \tag{3.12}$$

The projection matrices P_{f^0} and P_{λ^0} can be estimated respectively by $P_{\hat{f}} \equiv \hat{f}\hat{f}'/T$ and $P_{\hat{\lambda}} \equiv \hat{\lambda}(\hat{\lambda}'\hat{\lambda})^{-1}\hat{\lambda}'$. Then $M_{\hat{f}} \equiv I_T - P_{\hat{f}}$, $M_{\hat{\lambda}} \equiv I_N - P_{\hat{\lambda}}$ and $\hat{\Phi} \equiv \hat{f}(\hat{f}'\hat{f})^{-1}(\hat{\lambda}'\hat{\lambda})^{-1}\hat{\lambda}'$ are estimators of M_{f^0}, M_{λ^0}, and Φ, respectively. Let $\hat{e}_{it} \equiv Y_{it} - \hat{g}\left(X_{it}\right) - \hat{\lambda}_i'\hat{f}_t$, $\hat{\alpha}_{ij} \equiv \hat{\lambda}_i'(\hat{\lambda}'\hat{\lambda}/N)^{-1}\hat{\lambda}_j$, $\hat{\eta}_{ts} \equiv \hat{f}_t'(\hat{f}'\hat{f}/T)^{-1}\hat{f}_s$, and $\hat{Z}_{it} \equiv p_{it} - \frac{1}{N}\sum_{j=1}^{N}\hat{\alpha}_{ij}p_{jt} - \frac{1}{T}\sum_{s=1}^{T}\hat{\eta}_{ts}p_{is} + \frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\hat{\alpha}_{ij}\hat{\eta}_{ts}p_{js}$. Define

$$\begin{aligned}\hat{W}_{NT} &\equiv \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{Z}_{it}\hat{Z}_{it}',\ \hat{\Omega}_{NT} \equiv \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{Z}_{it}\hat{Z}_{it}'\hat{e}_{it}^2,\\ \hat{V}_K\left(x\right) &\equiv p^K\left(x\right)'\hat{W}_{NT}^{-1}\hat{\Omega}_{NT}\hat{W}_{NT}^{-1}p^K\left(x\right),\ \text{and } \hat{A}_{NT} \equiv \sqrt{NT/\hat{V}_K\left(x\right)},\end{aligned}$$

which are estimators of W_{NT}, Ω_{NT}, $V_K\left(x\right)$ and A_{NT}, respectively. For b_1, b_2, and b_3, define their corresponding estimates as $\hat{b}_1$, $\hat{b}_2$, and $\hat{b}_3$ whose kth elements are respectively given by

$$\hat{b}_{1,k} \equiv \frac{1}{N}\text{tr}\left[\left(\hat{\mathbf{e}}'\mathbf{P}_k\right)^{\text{truncR}}P_{\hat{f}}\right],\ \hat{b}_{2,k} \equiv \frac{1}{T}\text{tr}\left[\left(\hat{\mathbf{e}}\hat{\mathbf{e}}'\right)^{\text{truncD}}M_{\hat{\lambda}}\mathbf{P}_k\hat{\Phi}\right] \text{ and } \hat{b}_{3,k} \equiv \frac{1}{N}\text{tr}\left[\left(\hat{\mathbf{e}}'\hat{\mathbf{e}}\right)^{\text{truncD}}M_{\hat{f}}\mathbf{P}_k'\hat{\Phi}'\right],$$

where $\hat{\mathbf{e}}$ is an $N\times T$ matrix with (i,t)th element $\hat{e}_{it}$. Let $\hat{B}_K\left(x\right) = -\hat{A}_{NT}p^K\left(x\right)'\hat{W}_{NT}^{-1}(T^{-1}\hat{b}_1 + N^{-1}\hat{b}_2 + T^{-1}\hat{b}_3) \equiv -\kappa_{NT}\hat{b}_1\left(x\right) - \kappa_{NT}^{-1}\hat{b}_2\left(x\right) - \kappa_{NT}\hat{b}_3\left(x\right)$ and

$$\hat{\beta}_{bc} \equiv \hat{\beta} + \hat{W}_{NT}^{-1}(T^{-1}\hat{b}_1 + N^{-1}\hat{b}_2 + T^{-1}\hat{b}_3). \tag{3.13}$$

The bias-corrected estimator of $g\left(x\right)$ is given by

$$\hat{g}_{bc}\left(x\right) \equiv p^K\left(x\right)'\hat{\beta}_{bc} = \hat{g}\left(x\right) - \hat{A}_{NT}^{-1}\hat{B}_K\left(x\right). \tag{3.14}$$

To estimate the asymptotic bias and variance consistently, we add the following assumption.

Assumption 9. (i) As $(N,T)\to\infty$, $M_T\to\infty$ and $\max\{M_T/T,\ \sqrt{NK/T}\sum_{\tau=M_T}^{\infty}\alpha_{\mathcal{D}}^{(3+2\delta)/(4+2\delta)}(\tau),\ M_T\sqrt{NK/T}\delta_{NT}^{-1}\}\to 0$;

(ii) As $(N,T)\to\infty$,

$$\begin{aligned}
\max\left(\kappa_{NT},\kappa_{NT}^{-1}\right)\left[K^{3/2}\left(K^{-\gamma/d}+\delta_{NT}^{-1}\right)\right] &\to 0,\\
\max\left(\kappa_{NT}K^{1/2},\kappa_{NT}^{-1}\right)(NT)^{1/4}K\left(K^{-\gamma/d}+\delta_{NT}^{-2}\right) &\to 0,\\
\kappa_{NT}^{-1}\sqrt{K}[N^{-1/4}+N^{5/8}(K^{-\gamma/d}+\sqrt{K}\delta_{NT}^{-2})+T^{-1}N^{1/2}] &\to 0,\\
\kappa_{NT}\sqrt{K}[T^{-1/4}+T^{5/8}(K^{-\gamma/d}+\sqrt{K}\delta_{NT}^{-2})+N^{-1}T^{1/2}] &\to 0.
\end{aligned}$$

Assumption 9(i) imposes conditions on the bandwidth parameter M_T. Assumption 9(ii) seems quite complicated but can be simplified under some extra conditions. If we assume $\kappa_{NT}\to c\in(0,\infty)$, then Assumption 9$(ii)$ reduces to $K/N^{1/3}\to 0$, $K^{3/2-\gamma/d}N^{1/2}\to 0$, $K^{1/2-\gamma/d}N^{5/8}\to 0$, which, in conjunction with Assumption 8 and the additional requirement $\gamma/d>3/2$, implies that $K\in(N^{\gamma_0},N^{1/3})$, where $\gamma_0\equiv\max\{\frac{1/2}{\gamma/d-3/2},\frac{5/8}{\gamma/d-1/2}\}$.

The following theorem establishes the asymptotic distribution for the bias-corrected estimator $\hat{g}_{bc}(x)$.

Theorem 3.3 *Let Assumptions 1-9 hold. Then* $\hat{A}_{NT}[\hat{g}_{bc}(x)-g(x)]\xrightarrow{d}N(0,1)$ *as* $(N,T)\to\infty$.

4 A specification test for linearity

In this section, we consider a specification test for the commonly used linear dynamic panel data models with interactive fixed effects. We propose a test statistic based on the comparison of the linear estimator under the null hypothesis and the sieve estimator under the alternative.

4.1 The hypothesis and test statistic

For the model in (1.1), we are interested in testing the null hypothesis:

$$\mathbb{H}_0:\Pr\left[g(X_{it})=X_{it}'\theta^0\right]=1\,\text{for some }\theta^0\in\Theta, \tag{4.1}$$

where Θ is a compact subset of $\mathbb{R}^d$. The alternative hypothesis is

$$\mathbb{H}_1:\Pr\left[g(X_{it})=X_{it}'\theta\right]<1\text{ for all }\theta\in\Theta. \tag{4.2}$$

To facilitate the asymptotic local power analysis, we shall consider the following sequence of Pitman local alternatives:

$$\mathbb{H}_1(\gamma_{NT}):g(X_{it})=X_{it}'\theta^0+\gamma_{NT}\Delta(X_{it}) \tag{4.3}$$

where $\Delta(\cdot)\equiv\Delta_{NT}(\cdot)$ is a measurable nonlinear function and $\gamma_{NT}\to 0$ as $(N,T)\to\infty$. Let $\Delta_i\equiv(\Delta(X_{i1}),\cdots,\Delta(X_{iT}))'$ and $\mathbf{\Delta}\equiv(\Delta_1,\cdots,\Delta_N)'$.

We propose a test for $\mathbb{H}_0$ versus $\mathbb{H}_1$ by comparing the L_2-distance between two estimators of $g(\cdot)$, i.e., the linear and sieve estimators. Intuitively, both estimators are consistent under the null hypothesis of linearity while only the sieve estimator is consistent under the alternative. So if there is any deviation from the null, the L_2-distance between two estimators will signal it out asymptotically. This motivates us to consider the following test statistic

$$\Gamma_{NT} \equiv \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\hat{g}_{bc} \left(X_{it} \right) - \hat{g}^{(l)} \left(X_{it} \right) \right]^2 w \left(X_{it} \right),$$

where $\hat{g}^{(l)} (x) = x' \hat{\theta}$, $\hat{\theta}$ is Moon and Weidner's (2010, 2012) linear estimator of the coefficient θ under $\mathbb{H}_0$, and $w(x)$ is a user-specified nonnegative weighting function.[2] Similar test statistics have been proposed in various other contexts in the literature; see, e.g., Härdle and Mammen (1993) and Hong and White (1995). We will show that after being appropriately centered and scaled, Γ_{NT} is asymptotically normally distributed under the null hypothesis of linearity.

4.2 The asymptotic distribution under $\mathbb{H}_1(\gamma_{NT})$

Let $Q_{wxx,NT} \equiv \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} w_{it} X_{it} X_{it}'$, $Q_{wxx} \equiv E_{\mathcal{D}}[Q_{wxx,NT}]$, $Q_{wpx,NT} \equiv \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} w_{it} p_{it} X_{it}'$, and $Q_{wpx} \equiv E_{\mathcal{D}} [Q_{wpx,NT}]$. Let D_{NT} be a $d \times d$ matrix with its (k_1, k_2)th element given by

$$D_{NT,k_1k_2} \equiv \frac{1}{NT} \operatorname{tr} \left(M_{\lambda^0} \mathbf{X}_{k_1} M_{f^0} \mathbf{X}_{k_2}' \right). \tag{4.4}$$

Let $D \equiv E_{\mathcal{D}} [D_{NT}]$. Let Υ_{NT} be a $d \times 1$ vector whose kth element is given by $\Upsilon_{NT,k} \equiv \frac{1}{NT} \operatorname{tr}\left(M_{\lambda^0} \mathbf{X}_k M_{f^0} \mathbf{\Delta}' \right)$. We add the following assumptions.

Assumption 10. $\Delta(x)$ is $H(\gamma, \omega)$-smooth, and there exists $\beta_\Delta^0 \in \mathbb{R}^K$ such that $\|\beta_\Delta^0\| < \infty$ and $\left\| \Delta(\cdot) - p^K(\cdot)' \beta_\Delta^0 \right\|_{\infty,\bar{\omega}} = O\left(K^{-\gamma/d} \right)$.

Assumption 11. (i) $0 < \underline{C}_Q \le \mu_d(Q_{wxx}) \le \mu_1(Q_{wxx}) \le \bar{C}_Q < \infty$ a.s. as $(N,T) \to 0$;
(ii) $\|Q_{wpx}\| \le C_Q < \infty$ a.s. for all K as $(N,T) \to 0$;
(iii) $0 < \underline{C}_D \le \mu_d(D) \le \mu_1(D) \le \bar{C}_D < \infty$ a.s. as $(N,T) \to 0$;
where $\underline{C}_Q$, $\bar{C}_Q$, C_Q, $\underline{C}_D$, and $\bar{C}_D$ are constants that do not depend on K, N, or T.

Assumption 12. As $(N,T) \to \infty$, $K^3/N \to 0$, $\max\left(\kappa_{NT}, \kappa_{NT}^{-1} \right) K^{-1/4} \to 0$,

$$\begin{aligned}
K^{1/4} \sqrt{N/T} \textstyle\sum_{\tau=M_T}^{\infty} \alpha_{\mathcal{D}}^{(3+2\delta)/(4+2\delta)}(\tau) + K^{1/4} \sqrt{N/T} M_T \delta_{NT}^{-1} &\to 0, \\
\max\left(\kappa_{NT}, \kappa_{NT}^{-1} \right) \left[K^{5/4} \left(K^{-\gamma/d} + \delta_{NT}^{-1} \right) \right] &\to 0, \\
\kappa_{NT}^{-1} K^{1/4} [N^{-1/4} + N^{5/8} (K^{-\gamma/d} + \sqrt{K} \delta_{NT}^{-2}) + T^{-1} N^{1/2}] &\to 0, \\
\kappa_{NT} K^{1/4} [T^{-1/4} + T^{5/8} (K^{-\gamma/d} + \sqrt{K} \delta_{NT}^{-2}) + N^{-1} T^{1/2}] &\to 0.
\end{aligned}$$

Assumption 11 imposes some restrictions on the eigenvalues of certain matrices. Assumptions 11(i) and (iii) are reasonable as both Q_{wxx} and D are $d \times d$ matrices. Assumption 11(ii) is a high-level assumption. Let $Q_w \equiv \begin{pmatrix} Q_{wpp} & Q_{wpx} \\ Q_{wpx}' & Q_{wxx} \end{pmatrix}$, an augmented version of Q_{wpp}. In the literature on sieve estimation, it is commonly assumed that $\mu_1(Q_{wpp})$ is bounded above from infinity and below from 0 uniformly in K in large samples. Under this condition and Assumption 11(i), if one further requires that $\mu_1(Q_w) < C < \infty$, then one can readily demonstrate that $\|Q_{wpx}\|^2 = \mu_1\left(Q_{wpx} Q_{wpx}' \right) \le \mu_1(Q_{wpp}) \mu_1(Q_{wxx}) < \infty$. Note that

[2] In theory, the restricted parametric estimator $\hat{\theta}$ can be bias corrected or not. Intuitively, the asymptotic bias of $\hat{\theta}$ is of order δ_{NT}^{-2}, which is of smaller order than $(NT)^{-1/2} K^{1/4}$. The latter is the rate at which the nontrivial local alternatives our test has power to detect converge to the null. Of course, in practice a bias corrected parameter estimator is recommended.

Assumption 12 imposes much weaker requirement on (N, T, K, M_T) than that for the bias-correction of sieve estimator. But it is still necessary to use bias-corrected sieve estimate in specification testing. Assumption 12 also allows for the case where $N/T = c \in [0, \infty]$. If we restrict $c \in (0, \infty)$, Assumption 12 reduces to $K^{1/4} \max\{\sum_{\tau=M_T}^{\infty} \alpha_{\mathcal{D}}^{\frac{3+2\delta}{4+2\delta}}(\tau), \frac{M_T}{\sqrt{N}}\} \to 0$ and $K^3/N \to 0$, $K \in (N^{\gamma_1}, N^{1/3})$, where $\gamma_1 \equiv \max\{\frac{1/2}{\gamma/d-3/2}, \frac{5/8}{\gamma/d-1/4}\}$.

We define the asymptotic bias and variance terms as follows

$$\mathbb{B}_{NT} \equiv \operatorname{tr}\left(\tilde{W}^{-1} Q_{wpp} \tilde{W}^{-1} \tilde{\Omega}\right) \text{ and } \mathbb{V}_{NT} \equiv 2\operatorname{tr}\left(\tilde{W}^{-1} Q_{wpp} \tilde{W}^{-1} \tilde{\Omega} \tilde{W}^{-1} Q_{wpp} \tilde{W}^{-1} \tilde{\Omega}\right).$$

The following theorem establishes the asymptotic distribution of our test statistic under $\mathbb{H}_1(\gamma_{NT})$.

Theorem 4.1 *Suppose that Assumptions 1-8 and 10-12 hold. Under* $\mathbb{H}_1(\gamma_{NT})$ *with* $\gamma_{NT} \equiv (NT)^{-1/2} \mathbb{V}_{NT}^{1/4}$,

$$J_{NT} \equiv (NT\Gamma_{NT} - \mathbb{B}_{NT}) / \sqrt{\mathbb{V}_{NT}} \xrightarrow{d} N(A^{\Delta}, 1),$$

where $A^{\Delta} \equiv \operatorname{plim}_{(N,T)\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\Delta_{it} - X_{it}' D_{NT}^{-1} \Upsilon_{NT}\right)^2 w_{it}$ *is assumed to exist and be finite.*

Remark 3. The proof of the above theorem is tedious and is relegated to Appendix B. The idea is to express J_{NT} as a degenerate second order U-statistic plus some smaller order terms and then apply de Jong's (1987) CLT for independent but non-identically distributed (INID) observations. As Su, Jin, and Zhang (2013) notice, even though the CLT in de Jong (1987) works for second order U-statistics associated with INID observations, a close examination of his proof shows that it also works for conditionally independent but nonidentically distributed (CINID) observations. Noting that $A^{\Delta} = 0$ under $\mathbb{H}_0$, an immediate consequence of the above theorem is that $(NT\Gamma_{NT} - \mathbb{B}_{NT}) / \sqrt{\mathbb{V}_{NT}} \xrightarrow{d} N(0,1)$ under the null. In view of the fact that $\mathbb{V}_{NT} = O_P(K)$, we have $\gamma_{NT} = (NT)^{-1/2} \mathbb{V}_{NT}^{1/4} = O_P((NT)^{-1/2} K^{1/4})$. This indicates that J_{NT} has power to detect local alternatives that converge to the null hypothesis at the rate $(NT)^{-1/2} K^{1/4}$ provided that $A^{\Delta} > 0$. This is the rate we can obtain even if f_t^0 and λ_i^0 are observable. We obtain this rate despite the fact that the unobserved factors f_t^0 and factor loadings λ_i^0 can be only estimated at slower rates ($N^{-1/2}$ for the former and $T^{-1/2}$ for the latter, subject to certain matrix rotation), which suggests that the slower convergence rates of the estimates of f_t^0 and λ_i^0 do not have adverse first-order asymptotic effects on the asymptotic distribution of J_{NT}.

To implement the test, we propose to estimate $\mathbb{B}_{NT}$ and $\mathbb{V}_{NT}$ by $\hat{\mathbb{B}}_{NT} \equiv \operatorname{tr}(\hat{W}_{NT}^{-1} Q_{wpp,NT} \hat{W}_{NT}^{-1} \hat{\Omega}_{NT})$ and $\hat{\mathbb{V}}_{NT} \equiv 2\operatorname{tr}(\hat{W}_{NT}^{-1} Q_{wpp,NT} \hat{W}_{NT}^{-1} \hat{\Omega}_{NT} \hat{W}_{NT}^{-1} Q_{wpp,NT} \hat{W}_{NT}^{-1} \hat{\Omega}_{NT})$, respectively, where $\hat{W}_{NT} \equiv (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{Z}_{it} \hat{Z}_{it}'$ and $\hat{\Omega}_{NT} \equiv (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{Z}_{it} \hat{Z}_{it}' \hat{e}_{it}^2$. Then we define a feasible test statistic:

$$\hat{J}_{NT} \equiv \left(NT\Gamma_{NT} - \hat{\mathbb{B}}_{NT}\right) / \sqrt{\hat{\mathbb{V}}_{NT}}. \tag{4.5}$$

The following theorem establishes the asymptotic distribution of $\hat{J}_{NT}$ under $\mathbb{H}_1(\gamma_{NT})$.

Theorem 4.2 *Suppose that Assumptions 1-8 and 10-12 hold. Under* $\mathbb{H}_1(\gamma_{NT})$ *with* $\gamma_{NT} = (NT)^{-1/2} \times \mathbb{V}_{NT}^{1/4}$, $\hat{J}_{NT} \xrightarrow{d} N(A^{\Delta}, 1)$.

Remark 4. The above theorem implies that $\hat{J}_{NT}$ has nontrivial asymptotic power against local alternatives that converges to the null at the rate $(NT)^{-1/2} K^{1/4}$. The asymptotic local power function satisfies $\Pr\left(\hat{J}_{NT} > z | \mathbb{H}_1(\gamma_{NT})\right) \to 1 - \Phi(z - A^{\Delta})$ as $(N,T) \to \infty$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function (CDF).

Under $\mathbb{H}_0$, $A^{\Delta} = 0$, and $\hat{J}_{NT}$ is asymptotically distributed $N(0,1)$. This is stated in the following corollary.

Corollary 4.3 *Suppose that Assumptions 1-8 and 11-12 hold. Then under* $\mathbb{H}_0$, $\hat{J}_{NT} \xrightarrow{d} N(0,1)$.

Remark 5. In principle, one can compare $\hat{J}_{NT}$ with the one-sided critical value z_α, the upper αth percentile from the standard normal distribution, and reject the null when $\hat{J}_{NT} > z_\alpha$ at the α significant level. An alternative approach is to use bootstrap p-values.

Remark 6. To understand the asymptotic behavior of $\hat{J}_{NT}$ under global alternatives, we need to study the asymptotic property of $\hat{\theta}$ under $\mathbb{H}_1$. In this case, we define a pseudo-true parameter θ^* as the probability limit of $\hat{\theta}$. Then $\bar{\Delta}(X_{it}) \equiv g(X_{it}) - X_{it}'\theta^*$ is not equal to 0 a.s. Let $\bar{\Delta}_i \equiv [\bar{\Delta}(X_{i1}), \cdots, \bar{\Delta}(X_{iT})]'$ for $i = 1, ..., N$ and $\bar{\mathbf{\Delta}} \equiv (\bar{\Delta}_1, \cdots, \bar{\Delta}_N)'$. With the additional assumption $\|\bar{\mathbf{\Delta}}\| = o_P[(NT)^{1/2}]$, we can show that $\hat{\theta} - \theta^* = D_{NT}^{-1}\bar{\Upsilon}_{NT} + o_P(1)$, where $\bar{\Upsilon}_{NT}$ is a $d \times 1$ vector with kth element $\bar{\Upsilon}_{NT,k} \equiv (NT)^{-1}\mathrm{tr}(M_{\lambda^0}\mathbf{X}_k M_{f^0}\bar{\mathbf{\Delta}}')$. By some calculations, we can show that $\Gamma_{NT} = \frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T \bar{\Delta}(X_{it})^2 w_{it} + o_P(1) = O_P(1)$. This, together with the fact that $\hat{\mathbb{B}}_{NT} = O_P(K)$ and $\hat{\mathbb{V}}_{NT} = O_P(\sqrt{K})$ under $\mathbb{H}_1$, implies that our test statistic $\hat{J}_{NT}$ diverges at the rate $O_P(NT/\sqrt{K})$ under $\mathbb{H}_1$. That is, $\Pr(\hat{J}_{NT} > b_{NT}|\mathbb{H}_1) \to 1$ as $(N,T) \to \infty$ under $\mathbb{H}_1$ for any nonstochastic sequence $b_{NT} = o(NT/\sqrt{K})$. So our test achieves consistency against global alternatives.

Remark 7. With a little modification, our test can also be applied to testing for the specification of various other models with interactive fixed effects. First, one can consider a partially linear panel data model with interactive fixed effects where $g(X_{it}) = g_1(X_{1,it}) + \theta_2^{0\prime}X_{2,it}$, $X_{it} = (X_{1,it}', X_{2,it}')'$, and $g_1(\cdot)$ is an unknown smooth function. In this case, the hypotheses are $\mathbb{H}_0' : \Pr[g_1(X_{1,it}) = \theta_1^{0\prime}X_{1,it}] = 1$ for some $\theta_1^0 \in \Theta_1$ v.s. $\mathbb{H}_1' : \Pr[g_1(X_{1,it}) \neq \theta_1' X_{1,it}] < 1$ for all $\theta_1 \in \Theta_1$. One can continue to apply our test by estimating the model under the null and under the general nonparametric alternative for $g(\cdot)$ without imposing its partially linear structure. But this test may suffer some loss of efficiency as it does not impose the partially linear structure under the alternative. Alternatively, one can establish the asymptotic distribution theory for the sieve estimator for the partially linear model and compare it with the linear estimator under the null. The asymptotic distribution theory for the resulting test statistic is similar to what we have above. We omit the details to save space. Second, our test can also be applied to models that include both additive and multiplicative fixed effects. Let $(\lambda_{a,1}, ..., \lambda_{a,N})$ be the N individual fixed effects. We can write the common component as $\lambda_{a,i}f_{a,t} + \lambda_i^{0\prime}f_t^0 = \vec{\lambda}_i^{0\prime}\vec{f}_t^0$ for individual i at time period t, where $f_{a,t} = 1$, $\vec{f}_t^0 = (1, f_t^{0\prime})'$, and $\vec{\lambda}_i^0 = (\lambda_{a,i}, \lambda_i^{0\prime})'$. In this case, $f_{a,t}$ is known. We can obtain the sieve QMLE without estimating $f_{a,t}$ in the optimization process. With some minor modifications, we can establish the asymptotic distributions for the resulting estimator and test statistic. Third, we can also modify our test statistic to test for the hypotheses: $\mathbb{H}_0'' : \Pr[g(X_{it}) = 0] = 1$ v.s. $\mathbb{H}_1'' : \Pr[g(X_{it}) = 0] < 1$. This testing problem is particularly important in the nonlinear autoregressive panel data models (e.g., $Y_{it} = g(Y_{i,t-1}) + \lambda_i^{0\prime}f_t^0 + e_{it}$) because it is equivalent to testing for the presence of dynamic effects. It is also important to test the presence of anomaly effects in the asset pricing literature. Apparently we can compare the sieve estimate of $g(\cdot)$ with 0 to construct a test statistic, which is a special case of our test.

4.3 A bootstrap version of the test

Despite the fact that $\hat{J}_{NT}$ is asymptotically $N(0,1)$ under the null, it is not wise to rely on the asymptotic normal critical values to make statistical inference in finite samples because of the nonparametric nature of our test. In addition, even though the slow convergence rates of our factors and factor loadings estimates do not affect the asymptotic normal distribution of our test statistic, they tend to have adverse effects in finite samples (see, Su and Chen, 2013). As a result, tests based on standard normal critical values tend to suffer severe size distortions in finite samples. Therefore we propose a bootstrap procedure instead to obtain the bootstrap p values. The procedure is in the spirit of Hansen's (2000) fixed-regressor bootstrap and goes as follows:

1. Under $\mathbb{H}_0$, obtain the linear estimators $\hat{\theta}$, $\hat{f}_t^{(l)}$, $\hat{\lambda}_i^{(l)}$, and $\hat{e}_{it}^{(l)}$, where the superscript "(l)" denotes estimates under the null hypothesis of linearity; under $\mathbb{H}_1$, obtain the bias-corrected sieve estimators: $\hat{\beta}_{bc}$, $\hat{f}_t$, $\hat{\lambda}_i$, and $\hat{e}_{it}$. Calculate the test statistic $\hat{J}_{NT}$ based on $\hat{g}_{bc}(X_{it}) = \hat{\beta}_{bc}^{\prime} p^K(X_{it})$, $\hat{\theta}^{\prime} X_{it}$, $\hat{\lambda}_i$, $\hat{f}_t$, and $\hat{e}_{it}$.

2. For $i = 1, ..., N$, obtain the wild bootstrap errors $\{e_{it}^*\}_{t=1}^T$ as follows: $e_{it}^* = \nu_{it}\hat{e}_{it}^{(l)}$ where ν_{it} are IID $N(0,1)$. Then generate the bootstrap analogue Y_{it}^* of Y_{it} by holding $(X_{it}, \hat{f}_t^{(l)}, \hat{\lambda}_i^{(l)})$ as fixed: $Y_{it}^* = X_{it}^{\prime}\hat{\theta} + \hat{\lambda}_i^{(l)\prime}\hat{f}_t^{(l)} + e_{it}^*$ for $i = 1, ..., N$ and $t = 1, ..., T$.

3. Given the bootstrap resample $\{Y_{it}^*, X_{it}\}$, obtain the sieve QMLEs $\hat{g}_{bc}^*(X_{it})$, $\hat{\lambda}_i^*$, $\hat{f}_t^*$ and $\hat{e}_{it}^*$, and the linear estimators $\hat{\theta}^*$, $\hat{\lambda}_i^{(l)*}$, $\hat{f}_t^{(l)*}$ and $\hat{e}_{it}^{(l)*}$. Calculate the bootstrap test statistic $\hat{J}_{NT}^*$ based on $\hat{g}_{bc}^*(X_{it})$, $X_{it}^{\prime}\hat{\theta}^*$, $\hat{f}_t^*$, $\hat{\lambda}_i^*$, and $\hat{e}_{it}^*$.

4. Repeat Steps 2-3 for B times and index the bootstrap statistics as $\{\hat{J}_{NT,b}^*\}_{b=1}^B$. Calculate the bootstrap p-value: $p^* = B^{-1}\sum_{b=1}^B \mathbf{1}(\hat{J}_{NT,b}^* \geq \hat{J}_{NT})$.

It is straightforward to implement the above bootstrap procedure. Note that we impose the null hypothesis of linearity in Step 2. Since the regressors are treated as fixed, there is no dynamic structure in the bootstrap world. The next theorem implies the asymptotic validity of the above bootstrap procedure.

Theorem 4.4 *Suppose that the conditions in Theorem 4.2 hold. Then* $\hat{J}_{NT}^* \overset{d^*}{\to} N(0,1)$ *in probability, where* $\overset{d^*}{\to}$ *denotes weak convergence under the bootstrap probability measure conditional on the observed sample* $\mathcal{W}_{NT} \equiv \{(X_{it}, Y_{it}) : i = 1, ..., N,\ t = 1, ..., T\}$.

5 Monte Carlo simulations

In this section, we conduct Monte Carlo simulations to evaluate the finite sample performance of our estimators and test.

5.1 Data generating processes

We consider the following data generating processes (DGPs):
DGP 1: $Y_{it} = 0.5Y_{i,t-1} + \lambda_i^{0\prime} f_t^0 + e_{it}$,
DGP 2: $Y_{it} = 0.5Y_{i,t-1} + X_{1,it} + \lambda_i^{0\prime} f_t^0 + e_{it}$,
DGP 3: $Y_{it} = 0.5Y_{i,t-1} + 0.5\left[\frac{\exp(Y_{i,t-1} - Y_{i,t-1}^2)}{1+\exp(Y_{i,t-1} - Y_{i,t-1}^2)} - 0.5\right] + \lambda_i^{0\prime} f_t^0 + e_{it}$,
DGP 4: $Y_{it} = 0.5Y_{i,t-1} + 0.5\left[\Phi\left(Y_{i,t-1} - Y_{i,t-1}^2\right) - 0.5\right] + \lambda_i^{0\prime} f_t^0 + e_{it}$,
DGP 5: $Y_{it} = 0.5Y_{i,t-1} + 0.25\left[\phi(Y_{i,t-1}) - 1/\sqrt{2\pi}\right] + 0.5\left[\phi(X_{1,it}) - 1/\sqrt{2\pi}\right] + \lambda_i^{0\prime} f_t^0 + e_{it}$,
DGP 6: $Y_{it} = 0.5Y_{i,t-1} + 0.25X_{1,it}[\Phi(Y_{i,t-1}) - 0.5] + 0.5\left[\phi(X_{1,it}) - 1/\sqrt{2\pi}\right] + \lambda_i^{0\prime} f_t^0 + e_{it}$,
where $\lambda_i^0 = \left(\lambda_{i1}^0, \lambda_{i2}^0\right)^{\prime}$, $f_t^0 = \left(f_{t1}^0, f_{t2}^0\right)^{\prime}$, $i = 1, ..., N$, $t = 1, ..., T$, $\Phi(\cdot)$ and $\phi(\cdot)$ are the standard normal CDF and PDF, respectively. The regressors $X_{1,it}$ in DGPs 2, 5, and 6 are generated according to $X_{1,it} = 0.5\alpha_{i,x} + 0.5\lambda_{x,i1}^0 f_{t1}^0 + 0.5\lambda_{x,i2}^0 f_{t2}^0 + \varepsilon_{it}$, where $\lambda_{i1}^0, \lambda_{i2}^0$, $\lambda_{x,i1}^0$, $\lambda_{x,i2}^0$, and ε_{it} are IID $N(0,1)$, f_{t1}^0, f_{t2}^0, and e_{it} are IID $N(0, 0.25)$, $\alpha_{i,x}$ are IID $U[-0.25, 0.25]$, and they are mutually independent of each

other. Clearly, the exogenous regressor $X_{1,it}$ has a factor structure and is correlated with the common factors f_{t1}^0 and f_{t2}^0. All the above six DGPs are used to evaluate the finite sample performance of our estimator and test statistic. In the specification testing for linearity, DGPs 1-2 and 3-6 are used for level and power studies, respectively. For all DGPs, we discard the first 200 observations along the time dimension when generating the data.

Note that the idiosyncratic error terms in the above six DGPs are all homoskedastic (conditionally and unconditionally). To investigate the effect of conditional heteroskedasticity for the estimation and testing, we consider another set of DGPs, namely, DGPs 1h-6h, which are identical to DGPs 1-6, respectively, in the mean regression components but different from the latter in error terms. For DGPs 1h, 3h-4h, we generate the errors as follows $e_{it} = \sqrt{h_{it}}\epsilon_{it}$, $h_{it} = 0.1 + 0.2Y_{i,t-1}^2$, and ϵ_{it} ~IID $N(0,1)$ across both i and t. For DGPs 2h, 5h-6h, the errors are generated according to $e_{it} = \sqrt{h_{it}}\epsilon_{it}$, $h_{it} = 0.1 + 0.1Y_{i,t-1}^2 + 0.1X_{1,it}^2$, and ϵ_{it} ~IID $N(0,1)$ across both i and t.

5.2 Estimation: implementation and evaluation

In each DGP, we compute six estimators. We first compute the sieve estimate $\hat{g}(x)$ and its bias-corrected version $\hat{g}_{bc}(x)$. Then we compute the bias-corrected infeasible estimate $\hat{g}_{IF}(x)$ which is obtained by treating $\{f_t^0\}_{t=1}^T$ as observables. We also calculate another three estimates by pretending the regression function takes the commonly assumed linear functional form and term them as the linear QMLE $\hat{g}^{(l)}(x)$, its bias-corrected version $\hat{g}_{bc}^{(l)}(x)$, and the infeasible linear estimate $\hat{g}_{IF}^{(l)}(x)$ by treating the factors as observables, respectively. The infeasible estimates $\hat{g}_{IF}^{(l)}(x)$ and $\hat{g}_{IF}(x)$ provide a reference for efficiency comparison in DGPs 1-2 (or 1h-2h) and 3-6 (or 3h-6h), respectively. Compared with the sieve estimates $(\hat{g}(x), \hat{g}_{bc}(x))$, the linear estimates $(\hat{g}^{(l)}(x), \hat{g}_{bc}^{(l)}(x))$ signify the bias due to functional form misspecification in DGPs 3-6 or 3h-6h. Although there is no conditional heteroskedasticity across i, or serial correlation or heteroskedasticity across t for some DGPs (e.g., DGPs 1-6), we correct all three bias terms to obtain $\hat{g}_{bc}(x)$ and $\hat{g}_{bc}^{(l)}(x)$.

To obtain these estimates, we need to choose the bandwidth M_T for the bias correction. Throughout the simulation, we use $M_T = \lfloor T^{1/7} \rfloor$. The cubic B-spline is adopted as the sieve basis in all DGPs. The basis $b_{i,n}$ of a B-spline of degree $n \geq 1$ (of order $m = n+1$) is given recursively by

$$\begin{aligned} b_{j,n}(x) &= \alpha_{j,n}(x) b_{j,n-1}(x) + [1 - \alpha_{j+1,n}(x)] b_{j+1,n-1}(x), \\ b_{j,0}(x) &= \mathbf{1}(x \in [v_j, v_{J+1})), \end{aligned}$$

where $\alpha_{j,n}(x) = \frac{x - v_j}{v_{j+n} - v_j}\mathbf{1}(v_{j+n} \neq v_j)$ and $\{v_j\}_{j=0}^{J+1}$ is a sequence of non-decreasing real numbers (i.e., knots). We can approximate any smooth scalar function $B(x)$ by a linear combination of $\{b_{j,n}(x)\}_{j=0}^{J+m-1}$ for $x \in [v_0, v_{J+1}]$. For more details on the recursive construction of B-spline basis, see Racine (2012). In DGPs 1, 3, 4, 1h, 3h, and 4h where $g(x)$ is a univariate function, we use the cubic B-spline basis $(n = 3)$

$$p_Y^{J+4}(y) = \left[b_{0,3}^{(Y)}(y), b_{1,3}^{(Y)}(y), \cdots, b_{J+3,3}^{(Y)}(y)\right]', \tag{5.1}$$

where the superscript "(Y)" denotes its correspondence to $\{Y_{i,t-1}\}$. The knots $\{v_{y,j}\}_{j=0}^{J+1}$ are chosen as the empirical quantiles of $\{Y_{i,t-1}, i = 1, ..., N,\ t = 2, ..., T\}$, i.e., $v_{y,j}$ denotes the $j/(J+1)$th sample quantile of $\{Y_{i,t-1}\}$. So the total number of approximating terms in the sieve basis is given by $K = J+4$. In DGPs 2, 5, 6, 2h, 5h, and 6h, we consider two choices of sieve bases depending on whether we impose additivity on $g(y,x)$ or not. When we impose additivity, i.e., $g(y,x) = g_1(y) + g_2(x)$, the basis can be chosen as follows

$$p^K(y,x) = [p_Y^{J+4}(y)', p_X^{J+3}(x)']' \tag{5.2}$$

where $p_X^{J+3}(x) = [b_{0,3}^{(X)}(x), b_{1,3}^{(X)}(x), \cdots, b_{J+2,3}^{(X)}(x)]'$ with $b_{j,3}^{(X)}(x)$ being analogously defined as $b_{j,3}^{(Y)}(x)$. For convenience, we adopt the same number of knots for different regressors. Note that we leave the last element $b_{J+3,3}^{(X)}(x)$ out of $p_X^{J+3}(x)$ to avoid perfect multicollinearity as $\sum_{j=0}^{J+3} b_{j,3}^{(X)}(x) = 1$. For this case, the total number of approximating terms is $K = 2J + 7$. When we do not impose additivity, the basis is chosen as follows

$$p^K(y,x) = [p_Y^{J+4}(y) \otimes p_X^{J+4}(x)]', \tag{5.3}$$

where $\otimes$ denotes the tensor product. Then the total number of approximating terms is $K = (J+4)^2$. Even for as small values as $J = 3$, 4, and 5, we have $K = 49$, 64, and 81 terms in the sieve estimation, respectively. In all cases, to evaluate how the estimators are sensitive to the choice of J, we consider choosing $J = \lfloor C(NT)^{1/7.5} \rfloor$ for $C = 1$, 1.5, and 2.[3]

We consider the (N, T) pairs with N, $T = 20$, 40, and 60. To evaluate the finite sample performance of different estimators, we first calculate the root mean squared error (RMSE) for each replication: RMSE$(\hat{g}) = \sqrt{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[\hat{g}(X_{it}) - g(X_{it})]^2 a(X_{it})}$, where $a(\cdot)$ is used to trim out 2.5% tail observations along each tail of each dimension of X_{it}. Then we obtain the average RMSE (ARMSE) by averaging RMSE$(\hat{g})$ across 2000 replications, where $\hat{g}$ is a generic estimator of g. Other evaluation criteria like the median of RMSE, the average or median mean absolute deviation are also considered and they tend to yield qualitatively similar behavior for various estimators considered here. We only report the results based on the ARMSE to conserve space.

Tables 1-2 report the estimation results for homoskedastic or heteroskedastic errors, respectively, when we do not impose additivity for the bivariate regressions in DGPs 2, 5, 6, 2h, 5h, and 6h. Table 3 reports the estimation results for the latter six DGPs when we impose additivity. We summarize some important findings. First, for all DGPs, the ARMSEs for $\hat{g}$, $\hat{g}_{bc}$ and $\hat{g}_{IF}$ decrease as either N or T increases. The results for homoskedastic and heteroskedastic errors are similar. Second, as expected, when the regression functions are linear in DGPs 1, 2, 1h, and 2h, the linear estimate is more efficient than sieve estimate; when the regression functions are nonlinear, the sieve estimates (bias-corrected or not) outperform the linear estimates in terms ARMSE significantly, and the ARMSEs of the linear estimates tend to be stabilized at some large constant due to their inconsistency in the case of misspecification of functional form. Third, the bias correction works well for almost all DGPs and combinations (N, T) under investigation. The reduction of the percentage of ARMSE due to the bias correction is diminishing as T increases, which is consistent with our asymptotic result that the dominant first bias term is of order $O_P(\sqrt{K}/T)$. Fourth, the infeasible estimates always beat the feasible ones but the differences in ARMSEs for different types of estimates are shrinking as either N or T increases. Fifth, when additivity is correctly imposed for the bivariate regressions in DGPs 2, 5, 2h, and 5h, a comparison across the three tables suggests it leads to more precise estimation and significant reductions of ARMSEs for all estimates under investigation when compared with the case it is not imposed. When additivity is not correctly imposed for DGPs 6 and 6h, it generally results in large ARMSEs in large samples; exceptions may occur when there are too many sieve approximation terms that tend to result in large variance. Lastly, the above results are kind of robust for the three choices of J for both univariate regressions and additive bivariate regressions.

[3] Alternatively one can follow, e.g., Lee (2013), to use the leave-one-out cross-validation (CV) to choose K adaptively. Another possibility is to apply the Lasso-type techniques to achieve simultaneous variable selection and estimation. We leave these as a future research topic.

Table 1: ARMSE comparison for DGPs 1-6: homoskedastic errors

			$C=1$			$C=1.5$			$C=2$			Linear		
DGP	N	T	$\hat{g}$	$\hat{g}_{bc}$	$\hat{g}_{IF}$	$\hat{g}$	$\hat{g}_{bc}$	$\hat{g}_{IF}$	$\hat{g}$	$\hat{g}_{bc}$	$\hat{g}_{IF}$	$\hat{g}^{(l)}$	$\hat{g}_{bc}^{(l)}$	$\hat{g}_{IF}^{(l)}$
1	20	20	0.0575	0.0559	0.0453	0.0639	0.0625	0.0520	0.0688	0.0675	0.0572	0.0304	0.0277	0.0135
		40	0.0384	0.0380	0.0310	0.0408	0.0406	0.0342	0.0475	0.0474	0.0410	0.0206	0.0199	0.0105
		60	0.0307	0.0303	0.0248	0.0364	0.0361	0.0309	0.0388	0.0385	0.0337	0.0157	0.0152	0.0085
	40	20	0.0401	0.0384	0.0317	0.0439	0.0422	0.0358	0.0511	0.0497	0.0440	0.0240	0.0212	0.0107
		40	0.0268	0.0262	0.0216	0.0319	0.0314	0.0272	0.0344	0.0339	0.0296	0.0147	0.0140	0.0072
		60	0.0230	0.0227	0.0195	0.0248	0.0245	0.0215	0.0289	0.0287	0.0258	0.0117	0.0113	0.0061
	60	20	0.0347	0.0322	0.0268	0.0401	0.0379	0.0331	0.0424	0.0403	0.0356	0.0209	0.0175	0.0085
		40	0.0230	0.0226	0.0197	0.0253	0.0249	0.0222	0.0289	0.0285	0.0261	0.0115	0.0105	0.0059
		60	0.0181	0.0178	0.0159	0.0195	0.0192	0.0174	0.0224	0.0222	0.0204	0.0088	0.0082	0.0046
2	20	20	0.1107	0.1102	0.0844	0.1312	0.1312	0.1025	0.1480	0.1472	0.1194	0.0297	0.0294	0.0251
		40	0.0843	0.0841	0.0566	0.0932	0.0931	0.0675	0.1076	0.1072	0.0913	0.0187	0.0186	0.0158
		60	0.0732	0.0731	0.0459	0.0772	0.0772	0.0652	0.0844	0.0842	0.0747	0.0156	0.0156	0.0133
	40	20	0.0860	0.0858	0.0594	0.0960	0.0959	0.0709	0.1142	0.1128	0.0947	0.0192	0.0190	0.0170
		40	0.0679	0.0679	0.0402	0.0685	0.0685	0.0572	0.0729	0.0726	0.0658	0.0127	0.0125	0.0113
		60	0.0583	0.0581	0.0394	0.0630	0.0629	0.0462	0.0659	0.0657	0.0605	0.0100	0.0100	0.0094
	60	20	0.0756	0.0755	0.0480	0.0828	0.0824	0.0681	0.0912	0.0904	0.0778	0.0156	0.0154	0.0141
		40	0.0592	0.0591	0.0405	0.0643	0.0643	0.0477	0.0676	0.0673	0.0623	0.0110	0.0108	0.0100
		60	0.0511	0.0511	0.0322	0.0544	0.0543	0.0381	0.0566	0.0566	0.0500	0.0084	0.0084	0.0077
3	20	20	0.0590	0.0576	0.0468	0.0647	0.0634	0.0523	0.0686	0.0673	0.0563	0.0963	0.0956	0.1017
		40	0.0398	0.0395	0.0326	0.0426	0.0424	0.0359	0.0490	0.0488	0.0429	0.0928	0.0929	0.1036
		60	0.0308	0.0305	0.0259	0.0371	0.0368	0.0321	0.0392	0.0390	0.0344	0.0923	0.0924	0.1046
	40	20	0.0410	0.0397	0.0336	0.0443	0.0431	0.0371	0.0511	0.0501	0.0442	0.0934	0.0933	0.1038
		40	0.0276	0.0271	0.0230	0.0317	0.0313	0.0274	0.0339	0.0336	0.0297	0.0905	0.0906	0.1033
		60	0.0245	0.0243	0.0214	0.0261	0.0259	0.0231	0.0294	0.0293	0.0264	0.0912	0.0913	0.1045
	60	20	0.0346	0.0326	0.0278	0.0405	0.0386	0.0340	0.0423	0.0406	0.0361	0.0902	0.0899	0.1016
		40	0.0245	0.0241	0.0217	0.0264	0.0260	0.0236	0.0297	0.0293	0.0272	0.0900	0.0902	0.1035
		60	0.0192	0.0190	0.0173	0.0203	0.0201	0.0183	0.0232	0.0230	0.0213	0.0895	0.0897	0.1031
4	20	20	0.0591	0.0576	0.0472	0.0645	0.0632	0.0523	0.0687	0.0674	0.0566	0.0869	0.0861	0.0892
		40	0.0404	0.0401	0.0336	0.0424	0.0422	0.0360	0.0486	0.0484	0.0425	0.0831	0.0832	0.0905
		80	0.0324	0.0321	0.0278	0.0373	0.0370	0.0323	0.0394	0.0391	0.0345	0.0825	0.0825	0.0912
	40	20	0.0417	0.0403	0.0346	0.0445	0.0432	0.0373	0.0509	0.0498	0.0440	0.0838	0.0836	0.0905
		40	0.0293	0.0288	0.0253	0.0322	0.0318	0.0280	0.0343	0.0340	0.0300	0.0808	0.0809	0.0901
		60	0.0252	0.0250	0.0223	0.0263	0.0262	0.0234	0.0293	0.0291	0.0262	0.0814	0.0815	0.0911
	60	20	0.0358	0.0338	0.0294	0.0405	0.0386	0.0340	0.0424	0.0406	0.0361	0.0809	0.0805	0.0888
		40	0.0254	0.0250	0.0227	0.0268	0.0264	0.0241	0.0300	0.0296	0.0274	0.0804	0.0805	0.0903
		60	0.0203	0.0201	0.0185	0.0209	0.0207	0.0190	0.0232	0.0230	0.0213	0.0798	0.0800	0.0898
5	20	20	0.1176	0.1132	0.0831	0.1403	0.1344	0.0990	0.1623	0.1552	0.1145	0.0893	0.0872	0.0785
		40	0.0742	0.0723	0.0537	0.0893	0.0864	0.0655	0.1224	0.1182	0.0899	0.0803	0.0799	0.0768
		60	0.0594	0.0586	0.0435	0.0854	0.0834	0.0628	0.0989	0.0965	0.0721	0.0787	0.0784	0.0760
	40	20	0.0842	0.0786	0.0576	0.1024	0.0951	0.0688	0.1374	0.1276	0.0929	0.0825	0.0809	0.0762
		40	0.0536	0.0520	0.0382	0.0783	0.0753	0.0555	0.0911	0.0877	0.0645	0.0776	0.0773	0.0755
		60	0.0504	0.0493	0.0378	0.0629	0.0611	0.0449	0.0831	0.0807	0.0590	0.0780	0.0778	0.0760
	60	20	0.0677	0.0638	0.0467	0.0996	0.0928	0.0668	0.1135	0.1059	0.0769	0.0798	0.0791	0.0752
		40	0.0521	0.0503	0.0383	0.0655	0.0629	0.0456	0.0862	0.0827	0.0598	0.0774	0.0771	0.0753
		60	0.0419	0.0410	0.0313	0.0522	0.0507	0.0372	0.0704	0.0683	0.0491	0.0773	0.0771	0.0755
6	20	20	0.1164	0.1121	0.0832	0.1400	0.1343	0.0988	0.1611	0.1542	0.1144	0.0885	0.0867	0.0792
		40	0.0732	0.0713	0.0540	0.0886	0.0859	0.0660	0.1220	0.1180	0.0907	0.0802	0.0798	0.0771
		60	0.0585	0.0577	0.0433	0.0850	0.0830	0.0626	0.0981	0.0957	0.0718	0.0781	0.0780	0.0761
	40	20	0.0835	0.0781	0.0575	0.1010	0.0940	0.0688	0.1354	0.1260	0.0928	0.0820	0.0804	0.0765
		40	0.0524	0.0510	0.0381	0.0777	0.0748	0.0555	0.0904	0.0869	0.0645	0.0776	0.0773	0.0761
		60	0.0495	0.0485	0.0377	0.0619	0.0602	0.0447	0.0820	0.0797	0.0586	0.0778	0.0777	0.0765
	60	20	0.0664	0.0627	0.0466	0.0983	0.0916	0.0668	0.1121	0.1048	0.0772	0.0790	0.0784	0.0755
		40	0.0512	0.0496	0.0384	0.0648	0.0623	0.0456	0.0854	0.0820	0.0599	0.0771	0.0769	0.0757
		60	0.0402	0.0394	0.0312	0.0510	0.0496	0.0370	0.0691	0.0671	0.0487	0.0770	0.0769	0.0761

Table 2: ARMSE comparison for DGPs 1h-6h: heteroskedastic errors

DGP	N	T	$C=1$			$C=1.5$			$C=2$			Linear		
			$\hat{g}$	$\hat{g}_{bc}$	$\hat{g}_{IF}$	$\hat{g}$	$\hat{g}_{bc}$	$\hat{g}_{IF}$	$\hat{g}$	$\hat{g}_{bc}$	$\hat{g}_{IF}$	$\hat{g}^{(l)}$	$\hat{g}_{bc}^{(l)}$	$\hat{g}_{IF}^{(l)}$
1h	20	20	0.0724	0.0693	0.0531	0.0765	0.0733	0.0558	0.0802	0.0770	0.0596	0.0527	0.0488	0.0299
		40	0.0488	0.0480	0.0381	0.0517	0.0510	0.0406	0.0560	0.0554	0.0449	0.0346	0.0326	0.0216
		60	0.0389	0.0385	0.0314	0.0429	0.0426	0.0348	0.0446	0.0443	0.0363	0.0249	0.0235	0.0179
	40	20	0.0492	0.0474	0.0384	0.0528	0.0511	0.0415	0.0575	0.0559	0.0462	0.0381	0.0333	0.0219
		40	0.0334	0.0329	0.0271	0.0368	0.0365	0.0302	0.0385	0.0381	0.0317	0.0228	0.0211	0.0141
		60	0.0290	0.0288	0.0242	0.0304	0.0302	0.0256	0.0327	0.0324	0.0278	0.0211	0.0203	0.0141
	60	20	0.0454	0.0420	0.0329	0.0509	0.0477	0.0371	0.0525	0.0494	0.0390	0.0340	0.0288	0.0177
		40	0.0299	0.0293	0.0248	0.0314	0.0308	0.0262	0.0332	0.0326	0.0280	0.0199	0.0185	0.0128
		60	0.0234	0.0230	0.0194	0.0244	0.0239	0.0204	0.0261	0.0256	0.0220	0.0156	0.0150	0.0101
2h	20	20	0.1447	0.1450	0.1161	0.1685	0.1682	0.1317	0.1806	0.1791	0.1481	0.0483	0.0474	0.0453
		40	0.1050	0.1053	0.0777	0.1164	0.1161	0.0890	0.1274	0.1267	0.1124	0.0345	0.0342	0.0327
		60	0.0899	0.0898	0.0620	0.0928	0.0926	0.0806	0.1003	0.0997	0.0898	0.0264	0.0262	0.0245
	40	20	0.1054	0.1051	0.0794	0.1161	0.1158	0.0911	0.1337	0.1320	0.1151	0.0340	0.0326	0.0310
		40	0.0802	0.0802	0.0549	0.0825	0.0824	0.0720	0.0860	0.0856	0.0807	0.0230	0.0228	0.0220
		60	0.0695	0.0695	0.0521	0.0755	0.0755	0.0589	0.0764	0.0761	0.0724	0.0194	0.0192	0.0180
	60	20	0.0910	0.0910	0.0659	0.0976	0.0971	0.0861	0.1076	0.1062	0.0960	0.0269	0.0267	0.0253
		40	0.0688	0.0686	0.0513	0.0736	0.0735	0.0580	0.0761	0.0758	0.0726	0.0195	0.0192	0.0180
		60	0.0598	0.0598	0.0433	0.0630	0.0630	0.0489	0.0676	0.0676	0.0605	0.0166	0.0165	0.0159
3h	20	20	0.0813	0.0777	0.0612	0.0850	0.0815	0.0634	0.0873	0.0838	0.0660	0.1139	0.1119	0.1087
		40	0.0545	0.0542	0.0461	0.0576	0.0573	0.0479	0.0613	0.0610	0.0509	0.1018	0.1023	0.1080
		60	0.0453	0.0449	0.0382	0.0493	0.0490	0.0409	0.0504	0.0502	0.0417	0.1013	0.1015	0.1089
	40	20	0.0566	0.0547	0.0455	0.0596	0.0576	0.0476	0.0634	0.0617	0.0511	0.1024	0.1017	0.1077
		40	0.0399	0.0396	0.0347	0.0422	0.0418	0.0359	0.0430	0.0426	0.0364	0.0970	0.0975	0.1056
		60	0.0356	0.0354	0.0309	0.0365	0.0362	0.0315	0.0368	0.0365	0.0317	0.0976	0.0981	0.1073
	60	20	0.0520	0.0494	0.0408	0.0562	0.0534	0.0433	0.0577	0.0550	0.0444	0.0989	0.0981	0.1045
		40	0.0350	0.0346	0.0307	0.0360	0.0356	0.0314	0.0375	0.0370	0.0320	0.0954	0.0963	0.1057
		60	0.0299	0.0297	0.0267	0.0301	0.0298	0.0266	0.0303	0.0300	0.0263	0.0948	0.0953	0.1041
4h	20	20	0.0788	0.0754	0.0598	0.0815	0.0783	0.0611	0.0837	0.0805	0.0638	0.1023	0.1002	0.0956
		40	0.0543	0.0541	0.0466	0.0559	0.0556	0.0464	0.0596	0.0592	0.0489	0.0914	0.0914	0.0940
		80	0.0461	0.0458	0.0402	0.0476	0.0473	0.0396	0.0485	0.0483	0.0402	0.0899	0.0899	0.0944
	40	20	0.0565	0.0548	0.0468	0.0581	0.0564	0.0470	0.0611	0.0596	0.0496	0.0921	0.0912	0.0936
		40	0.0413	0.0410	0.0372	0.0403	0.0400	0.0346	0.0410	0.0407	0.0348	0.0866	0.0867	0.0915
		60	0.0349	0.0347	0.0306	0.0352	0.0350	0.0307	0.0357	0.0354	0.0304	0.0866	0.0869	0.0928
	60	20	0.0515	0.0490	0.0417	0.0539	0.0512	0.0414	0.0552	0.0524	0.0423	0.0888	0.0877	0.0909
		40	0.0347	0.0343	0.0305	0.0350	0.0345	0.0305	0.0356	0.0351	0.0305	0.0851	0.0856	0.0915
		60	0.0296	0.0294	0.0265	0.0291	0.0288	0.0257	0.0287	0.0284	0.0248	0.0841	0.0845	0.0899
5h	20	20	0.1213	0.1200	0.0889	0.1444	0.1380	0.1042	0.1627	0.1520	0.1184	0.0937	0.0915	0.0833
		40	0.0777	0.0773	0.0623	0.0878	0.0876	0.0716	0.1102	0.1088	0.0924	0.0843	0.0836	0.0804
		60	0.0649	0.0641	0.0499	0.0816	0.0802	0.0661	0.0939	0.0919	0.0738	0.0808	0.0806	0.0788
	40	20	0.0855	0.0826	0.0615	0.0992	0.0953	0.0726	0.1293	0.1210	0.0942	0.0846	0.0826	0.0786
		40	0.0548	0.0542	0.0439	0.0745	0.0728	0.0580	0.0842	0.0822	0.0663	0.0794	0.0789	0.0774
		60	0.0512	0.0505	0.0407	0.0591	0.0582	0.0462	0.0773	0.0751	0.0584	0.0776	0.0774	0.0769
	60	20	0.0706	0.0674	0.0506	0.1006	0.0931	0.0683	0.1129	0.1042	0.0768	0.0830	0.0815	0.0785
		40	0.0514	0.0505	0.0408	0.0609	0.0587	0.0473	0.0783	0.0752	0.0596	0.0775	0.0772	0.0763
		60	0.0432	0.0422	0.0338	0.0510	0.0502	0.0384	0.0670	0.0651	0.0487	0.0772	0.0770	0.0766
6h	20	20	0.1229	0.1193	0.0904	0.1423	0.1368	0.1040	0.1598	0.1526	0.1177	0.0931	0.0908	0.0846
		40	0.0813	0.0796	0.0603	0.0935	0.0907	0.0711	0.1208	0.1164	0.0915	0.0825	0.0819	0.0798
		60	0.0649	0.0643	0.0487	0.0853	0.0833	0.0652	0.0965	0.0941	0.0732	0.0795	0.0793	0.0776
	40	20	0.0895	0.0849	0.0635	0.1048	0.0974	0.0732	0.1351	0.1250	0.0947	0.0872	0.0841	0.0796
		40	0.0563	0.0552	0.0428	0.0785	0.0755	0.0575	0.0890	0.0854	0.0650	0.0790	0.0785	0.0775
		60	0.0530	0.0521	0.0409	0.0618	0.0602	0.0472	0.0789	0.0767	0.0594	0.0784	0.0783	0.0774
	60	20	0.0730	0.0692	0.0513	0.1005	0.0933	0.0691	0.1127	0.1048	0.0782	0.0818	0.0803	0.0776
		40	0.0553	0.0539	0.0419	0.0650	0.0625	0.0478	0.0825	0.0792	0.0598	0.0781	0.0777	0.0766
		60	0.0436	0.0429	0.0345	0.0507	0.0492	0.0392	0.0660	0.0639	0.0494	0.0775	0.0774	0.0767

Table 3: ARMSE comparison for DGPs 2 , 5, 6, 2h, 5h, and 6h: additivity is imposed

			$C=1$			$C=1.5$			$C=2$			Linear		
DGP	N	T	$\hat{g}$	$\hat{g}_{bc}$	$\hat{g}_{IF}$	$\hat{g}$	$\hat{g}_{bc}$	$\hat{g}_{IF}$	$\hat{g}$	$\hat{g}_{bc}$	$\hat{g}_{IF}$	$\hat{g}^{(l)}$	$\hat{g}^{(l)}_{bc}$	$\hat{g}^{(l)}_{IF}$
2	20	20	0.1105	0.0933	0.0636	0.1034	0.1032	0.0715	0.1024	0.1022	0.0785	0.0297	0.0294	0.0251
		40	0.0547	0.0546	0.0426	0.0668	0.0668	0.0473	0.0734	0.0734	0.0561	0.0187	0.0186	0.0158
		60	0.0445	0.0444	0.0358	0.0599	0.0599	0.0430	0.0607	0.0606	0.0467	0.0156	0.0156	0.0133
	40	20	0.0554	0.0553	0.0463	0.0679	0.0677	0.0514	0.0755	0.0756	0.0614	0.0192	0.0190	0.0170
		40	0.0377	0.0377	0.0311	0.0523	0.0523	0.0384	0.0533	0.0533	0.0415	0.0127	0.0125	0.0113
		60	0.0417	0.0417	0.0293	0.0434	0.0434	0.0324	0.0437	0.0437	0.0374	0.0100	0.0100	0.0094
	60	20	0.0446	0.0445	0.0374	0.0610	0.0610	0.0451	0.0606	0.0606	0.0490	0.0156	0.0154	0.0141
		40	0.0427	0.0427	0.0291	0.0434	0.0434	0.0316	0.0432	0.0432	0.0365	0.0110	0.0108	0.0100
		60	0.0357	0.0357	0.0243	0.0370	0.0370	0.0263	0.0357	0.0357	0.0298	0.0084	0.0084	0.0077
5	20	20	0.0762	0.0748	0.0627	0.0853	0.0839	0.0706	0.0921	0.0909	0.0770	0.0893	0.0872	0.0785
		40	0.0465	0.0460	0.0400	0.0514	0.0509	0.0455	0.0605	0.0601	0.0546	0.0803	0.0799	0.0768
		60	0.0390	0.0388	0.0343	0.0469	0.0467	0.0421	0.0506	0.0505	0.0461	0.0787	0.0784	0.0760
	40	20	0.0517	0.0499	0.0441	0.0567	0.0551	0.0495	0.0650	0.0634	0.0586	0.0825	0.0809	0.0762
		40	0.0314	0.0311	0.0281	0.0382	0.0379	0.0347	0.0413	0.0410	0.0381	0.0776	0.0773	0.0755
		60	0.0289	0.0287	0.0267	0.0320	0.0317	0.0299	0.0373	0.0371	0.0354	0.0780	0.0778	0.0760
	60	20	0.0401	0.0387	0.0347	0.0481	0.0468	0.0431	0.0518	0.0505	0.0471	0.0798	0.0791	0.0752
		40	0.0293	0.0289	0.0267	0.0319	0.0316	0.0295	0.0365	0.0362	0.0347	0.0774	0.0771	0.0753
		60	0.0225	0.0223	0.0207	0.0248	0.0246	0.0230	0.0289	0.0288	0.0273	0.0773	0.0771	0.0755
6	20	20	0.0916	0.0900	0.0796	0.0976	0.0961	0.0853	0.1034	0.1019	0.0912	0.0885	0.0867	0.0792
		40	0.0675	0.0669	0.0620	0.0708	0.0703	0.0652	0.0792	0.0787	0.0732	0.0802	0.0798	0.0771
		60	0.0605	0.0604	0.0574	0.0658	0.0657	0.0625	0.0683	0.0681	0.0650	0.0781	0.0780	0.0761
	40	20	0.0716	0.0697	0.0648	0.0745	0.0726	0.0677	0.0825	0.0808	0.0758	0.0820	0.0804	0.0765
		40	0.0567	0.0564	0.0548	0.0611	0.0608	0.0591	0.0628	0.0626	0.0610	0.0776	0.0773	0.0761
		60	0.0552	0.0551	0.0536	0.0566	0.0565	0.0550	0.0595	0.0594	0.0580	0.0778	0.0777	0.0765
	60	20	0.0622	0.0613	0.0583	0.0674	0.0665	0.0636	0.0700	0.0691	0.0662	0.0790	0.0784	0.0755
		40	0.0549	0.0548	0.0534	0.0564	0.0562	0.0548	0.0592	0.0591	0.0579	0.0771	0.0769	0.0757
		60	0.0519	0.0518	0.0512	0.0528	0.0528	0.0520	0.0548	0.0548	0.0541	0.0770	0.0769	0.0761
2h	20	20	0.1101	0.1101	0.0920	0.1240	0.1239	0.1010	0.1324	0.1325	0.1091	0.0483	0.0474	0.0453
		40	0.0715	0.0714	0.0613	0.0820	0.0820	0.0656	0.0915	0.0915	0.0760	0.0345	0.0342	0.0327
		60	0.0584	0.0584	0.0503	0.0723	0.0723	0.0584	0.0752	0.0752	0.0625	0.0264	0.0262	0.0245
	40	20	0.0748	0.0747	0.0644	0.0866	0.0866	0.0701	0.0952	0.0951	0.0813	0.0340	0.0326	0.0310
		40	0.0496	0.0496	0.0436	0.0624	0.0623	0.0513	0.0650	0.0650	0.0552	0.0230	0.0228	0.0220
		60	0.0502	0.0501	0.0393	0.0527	0.0525	0.0422	0.0542	0.0541	0.0478	0.0194	0.0192	0.0180
	60	20	0.0602	0.0600	0.0534	0.0739	0.0738	0.0618	0.0770	0.0769	0.0662	0.0269	0.0267	0.0253
		40	0.0525	0.0525	0.0407	0.0539	0.0538	0.0436	0.0552	0.0551	0.0485	0.0195	0.0192	0.0180
		60	0.0435	0.0435	0.0326	0.0441	0.0441	0.0351	0.0459	0.0459	0.0403	0.0166	0.0165	0.0159
5h	20	20	0.0898	0.0875	0.0723	0.0956	0.0937	0.0798	0.1018	0.1001	0.0855	0.0937	0.0915	0.0833
		40	0.0567	0.0558	0.0485	0.0614	0.0606	0.0534	0.0700	0.0692	0.0622	0.0843	0.0836	0.0804
		60	0.0444	0.0443	0.0386	0.0515	0.0514	0.0456	0.0551	0.0550	0.0492	0.0808	0.0806	0.0788
	40	20	0.0606	0.0585	0.0505	0.0649	0.0628	0.0548	0.0731	0.0711	0.0641	0.0846	0.0826	0.0786
		40	0.0377	0.0372	0.0339	0.0448	0.0442	0.0411	0.0480	0.0474	0.0444	0.0794	0.0789	0.0774
		60	0.0322	0.0319	0.0297	0.0347	0.0345	0.0323	0.0393	0.0391	0.0369	0.0776	0.0774	0.0769
	60	20	0.0488	0.0470	0.0418	0.0556	0.0540	0.0490	0.0589	0.0573	0.0526	0.0830	0.0815	0.0785
		40	0.0338	0.0333	0.0303	0.0364	0.0359	0.0329	0.0403	0.0399	0.0372	0.0775	0.0772	0.0763
		60	0.0274	0.0272	0.0254	0.0294	0.0293	0.0274	0.0332	0.0331	0.0316	0.0772	0.0770	0.0766
6h	20	20	0.1014	0.0997	0.0886	0.1065	0.1046	0.0932	0.1116	0.1099	0.0983	0.0931	0.0908	0.0846
		40	0.0739	0.0732	0.0672	0.0773	0.0767	0.0702	0.0843	0.0837	0.0774	0.0825	0.0819	0.0798
		60	0.0651	0.0649	0.0612	0.0705	0.0703	0.0660	0.0727	0.0725	0.0684	0.0795	0.0793	0.0776
	40	20	0.0788	0.0765	0.0698	0.0821	0.0798	0.0731	0.0903	0.0881	0.0811	0.0872	0.0841	0.0796
		40	0.0604	0.0599	0.0578	0.0650	0.0646	0.0619	0.0666	0.0662	0.0639	0.0790	0.0785	0.0775
		60	0.0573	0.0572	0.0556	0.0587	0.0587	0.0570	0.0618	0.0617	0.0602	0.0784	0.0783	0.0774
	60	20	0.0675	0.0658	0.0615	0.0728	0.0712	0.0669	0.0750	0.0735	0.0692	0.0818	0.0803	0.0776
		40	0.0577	0.0574	0.0555	0.0590	0.0587	0.0569	0.0618	0.0616	0.0599	0.0781	0.0777	0.0766
		60	0.0539	0.0538	0.0529	0.0548	0.0547	0.0537	0.0570	0.0569	0.0558	0.0775	0.0774	0.0767

Note: Here the additivity of functional form is imposed in the estimation, which is correct for DGPs 2, 5, 2h and 5h, but incorrect for DGPs 6 and 6h.

5.3 Testing: implementation and evaluation

To conduct the specification test, we choose the same M_T, J, and basis functions as in the estimation stage. We use $w(X_{it}) = \mathbf{1}(X_{it} \in \mathcal{U})$ where $\mathcal{U}$ is chosen to trim out 2.5% tail observations along each tail of each dimension of X_{it}. For the bivariate regression function g in DGPs 2, 5, 6, 2h, 5h, and 6h, we only consider the test by imposing additivity of g although g has nonadditive nonlinear component in DGPs 6 and 6h. For each scenario, we consider 250 replications and adopt 200 bootstrap resamples in each replication for both the size and power studies.

Tables 4-5 report the empirical rejection frequencies of our test at 1%, 5%, and 10% nominal levels for the case of homoskedastic and heteroskedastic errors, respectively. We summarize some important findings from these tables. First, when the null hypothesis of linearity holds in DGPs 1, 2, 1h, and 2h, these tables suggest that the level of our test behaves reasonably well for almost all DGPs, sample sizes, and all choices of J under investigation despite the fact that slight to moderate size distortions may occur in the case of heteroskedastic errors terms. Second, the power of our test generally increases very fast as either N or T increases, and it not very sensitive to the choice of J.

6 An application to the economic growth data

The relationship between the long-run economic growth and investment in physical capital has been studied extensively and has played a crucial role in the evaluation of different growth theories. A positive association between the investment as a share of gross domestic product (GDP) and per capita GDP growth rate is supported by the early endogenous growth models such as the AK model. However, the exogenous growth theories such as the Solow model assert that an increase in investment can only raise the level of per capita GDP, but have no effect on the steady-state growth rate. Many empirical studies show that there is little or no association between the investment and the long-run growth rate; see Jones (1995) and Easterly and Levine (2001). Recently, Bond, Leblebicioglu, and Schiantarelli (2010) reassess the relationship between these two by using a panel data of 71 countries covering 41 years (1960-2000). By estimating a dynamic panel data model with both individual and time fixed effects they find strong evidence of a positive relationship between the investment as a share of real GDP and the long-run growth rate of GDP per worker.

Note that most empirical works are carried out under the linear framework and only include additive fixed effects to control unobservable heterogeneity. In this section, we re-investigate the problem using the following nonparametric dynamic panel data model with interactive fixed effects

$$Y_{it} = g(Y_{i,t-1}, I_{it}, \Delta I_{it}) + \lambda_i' f_t + e_{it}$$

where $Y_{it} \equiv \log(GDP_{it}) - \log(GDP_{i,t-1})$, GDP_{it} is the real GDP per worker for country i in year t, I_{it} is the logarithm of the investment as a share of real GDP, $\Delta I_{it} \equiv I_{it} - I_{i,t-1}$, and the multi-factor error structure $\lambda_i' f_t + e_{it}$ is used to control for heterogeneity and capture the unobservable common shocks. $Y_{i,t-1}$ is included in the unknown function $g(\cdot)$ to partially control serial correlation; see some recent empirical studies on growth such as Chambers and Guo (2009) and Meierrieks and Gries (2012) that consider dynamic panel data models. Su and Lu (2013) also consider nonparametric dynamic panel growth regressions but with individual fixed effects only.

The data set is from the Penn World Tables (PTW7.1); see Heston, Summers, and Aten (2009). We use the almost same set of countries as Bond, Leblebicioglu, and Schiantarelli (2010) but exclude

Table 4: Rejection frequency for DGPs 1-6

DGP	N	T	$C=1$ 1%	5%	10%	$C=1.5$ 1%	5%	10%	$C=2$ 1%	5%	10%
1	20	20	0.016	0.064	0.128	0.012	0.068	0.124	0.008	0.040	0.100
		40	0.016	0.044	0.108	0.016	0.052	0.108	0.012	0.048	0.116
		60	0.004	0.052	0.100	0.016	0.040	0.112	0.012	0.056	0.100
	40	20	0.010	0.060	0.096	0.012	0.052	0.088	0.016	0.060	0.104
		40	0.012	0.052	0.096	0.012	0.036	0.100	0.012	0.044	0.104
		60	0.008	0.056	0.096	0.016	0.044	0.088	0.012	0.048	0.092
	60	20	0.010	0.072	0.116	0.010	0.050	0.100	0.010	0.040	0.096
		40	0.008	0.036	0.072	0.012	0.036	0.080	0.012	0.040	0.096
		60	0.016	0.048	0.108	0.012	0.040	0.104	0.016	0.056	0.112
2	20	20	0.016	0.048	0.080	0.008	0.068	0.100	0.008	0.060	0.096
		40	0.016	0.056	0.100	0.008	0.056	0.088	0.012	0.072	0.104
		60	0.020	0.056	0.088	0.012	0.052	0.096	0.008	0.044	0.096
	40	20	0.032	0.088	0.132	0.032	0.060	0.136	0.012	0.076	0.120
		40	0.012	0.084	0.116	0.004	0.064	0.100	0.012	0.048	0.112
		60	0.024	0.064	0.096	0.024	0.068	0.116	0.008	0.056	0.104
	60	20	0.008	0.048	0.124	0.012	0.048	0.108	0.008	0.052	0.112
		40	0.004	0.052	0.104	0.000	0.044	0.104	0.016	0.052	0.092
		60	0.020	0.060	0.100	0.016	0.052	0.120	0.020	0.068	0.100
3	20	20	0.248	0.460	0.616	0.184	0.432	0.568	0.176	0.372	0.532
		40	0.740	0.888	0.932	0.676	0.848	0.904	0.572	0.764	0.852
		60	0.904	0.964	0.984	0.832	0.912	0.960	0.808	0.904	0.944
	40	20	0.656	0.820	0.908	0.608	0.784	0.888	0.536	0.752	0.840
		40	0.984	1.000	1.000	0.976	0.996	1.000	0.972	0.996	1.000
		60	1.000	1.000	1.000	1.000	1.000	1.000	0.996	1.000	1.000
	60	20	0.848	0.948	0.984	0.748	0.876	0.940	0.716	0.864	0.916
		40	1.000	1.000	1.000	0.996	1.000	1.000	0.996	1.000	1.000
		60	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
4	20	20	0.248	0.488	0.620	0.224	0.436	0.592	0.180	0.408	0.548
		40	0.740	0.888	0.944	0.688	0.864	0.912	0.608	0.796	0.872
		60	0.908	0.976	0.988	0.848	0.924	0.964	0.824	0.912	0.956
	40	20	0.684	0.864	0.928	0.664	0.848	0.912	0.596	0.776	0.872
		40	0.992	1.000	1.000	0.984	1.000	1.000	0.976	1.000	1.000
		60	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	60	20	0.920	0.972	0.988	0.852	0.952	0.964	0.848	0.944	0.956
		40	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		60	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
5	20	20	0.440	0.632	0.716	0.396	0.564	0.668	0.352	0.484	0.644
		40	0.844	0.924	0.968	0.796	0.908	0.940	0.696	0.872	0.924
		60	0.968	0.988	0.992	0.948	0.980	0.988	0.932	0.980	0.992
	40	20	0.860	0.928	0.948	0.836	0.900	0.936	0.736	0.860	0.904
		40	0.992	1.000	1.000	0.992	0.996	0.996	0.988	0.992	0.996
		60	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	60	20	0.972	0.992	0.992	0.936	0.984	0.992	0.892	0.952	0.980
		40	1.000	1.000	1.000	0.996	1.000	1.000	0.988	0.992	1.000
		60	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
6	20	20	0.246	0.400	0.516	0.208	0.388	0.472	0.196	0.312	0.448
		40	0.572	0.740	0.852	0.492	0.692	0.776	0.368	0.576	0.708
		60	0.828	0.928	0.972	0.744	0.880	0.920	0.728	0.872	0.900
	40	20	0.580	0.752	0.848	0.488	0.712	0.804	0.440	0.628	0.712
		40	0.944	0.988	0.992	0.912	0.952	0.976	0.884	0.936	0.972
		60	0.996	1.000	1.000	0.996	1.000	1.000	0.988	0.996	1.000
	60	20	0.780	0.900	0.952	0.716	0.864	0.912	0.664	0.836	0.884
		40	0.988	1.000	1.000	0.984	1.000	1.000	0.980	0.996	1.000
		60	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Note: $J = \lfloor C(NT)^{1/7.5} \rfloor$ where $C = 1$, 1.5, and 2.

Table 5: Rejection frequency for DGPs 1h-6h

			$C = 1$			$C = 1.5$			$C = 2$		
DGP	N	T	1%	5%	10%	1%	5%	10%	1%	5%	10%
1h	20	20	0.024	0.060	0.112	0.024	0.080	0.136	0.028	0.072	0.124
		40	0.020	0.076	0.136	0.020	0.084	0.128	0.024	0.088	0.144
		60	0.032	0.076	0.124	0.028	0.056	0.108	0.024	0.056	0.112
	40	20	0.032	0.064	0.144	0.036	0.072	0.136	0.028	0.068	0.120
		40	0.040	0.080	0.128	0.036	0.076	0.136	0.040	0.080	0.132
		60	0.028	0.064	0.128	0.024	0.064	0.128	0.020	0.064	0.108
	60	20	0.024	0.072	0.124	0.032	0.068	0.116	0.032	0.064	0.116
		40	0.016	0.056	0.096	0.016	0.052	0.100	0.020	0.056	0.096
		60	0.012	0.060	0.100	0.012	0.060	0.088	0.008	0.056	0.092
2h	20	20	0.020	0.052	0.120	0.016	0.040	0.120	0.028	0.076	0.128
		40	0.024	0.060	0.136	0.016	0.056	0.136	0.032	0.076	0.120
		60	0.028	0.068	0.124	0.016	0.064	0.124	0.020	0.068	0.132
	40	20	0.020	0.076	0.124	0.016	0.076	0.124	0.004	0.072	0.128
		40	0.012	0.064	0.108	0.016	0.056	0.100	0.012	0.044	0.104
		60	0.008	0.048	0.096	0.008	0.052	0.096	0.012	0.056	0.100
	60	20	0.016	0.056	0.104	0.016	0.060	0.104	0.012	0.052	0.104
		40	0.008	0.044	0.096	0.012	0.036	0.092	0.016	0.056	0.104
		60	0.016	0.064	0.132	0.012	0.056	0.120	0.016	0.064	0.124
3h	20	20	0.140	0.296	0.448	0.152	0.292	0.436	0.140	0.288	0.396
		40	0.372	0.588	0.680	0.352	0.560	0.652	0.336	0.472	0.612
		60	0.532	0.684	0.772	0.504	0.652	0.796	0.484	0.664	0.780
	40	20	0.348	0.508	0.672	0.348	0.500	0.680	0.308	0.488	0.620
		40	0.616	0.816	0.872	0.620	0.828	0.912	0.628	0.812	0.896
		60	0.808	0.936	0.956	0.800	0.948	0.960	0.808	0.948	0.964
	60	20	0.400	0.556	0.656	0.368	0.568	0.684	0.368	0.556	0.692
		40	0.760	0.904	0.932	0.760	0.912	0.928	0.748	0.908	0.964
		60	0.996	1.000	1.000	0.992	0.996	1.000	0.996	1.000	1.000
4h	20	20	0.148	0.300	0.424	0.168	0.324	0.436	0.144	0.276	0.400
		40	0.380	0.600	0.672	0.404	0.612	0.684	0.360	0.536	0.660
		60	0.524	0.676	0.768	0.548	0.724	0.824	0.536	0.740	0.832
	40	20	0.364	0.536	0.676	0.392	0.572	0.724	0.348	0.520	0.672
		40	0.604	0.820	0.856	0.712	0.852	0.928	0.708	0.852	0.932
		60	0.876	0.972	0.988	0.868	0.972	0.984	0.868	0.968	0.988
	60	20	0.460	0.676	0.780	0.548	0.736	0.808	0.528	0.696	0.800
		40	0.824	0.948	0.980	0.820	0.948	0.976	0.808	0.944	0.976
		60	0.988	0.996	1.000	0.984	0.992	1.000	0.980	0.988	0.996
5h	20	20	0.344	0.516	0.616	0.316	0.504	0.616	0.284	0.484	0.592
		40	0.744	0.848	0.916	0.660	0.820	0.876	0.604	0.796	0.840
		60	0.920	0.964	0.976	0.892	0.940	0.972	0.864	0.940	0.960
	40	20	0.756	0.880	0.896	0.716	0.848	0.900	0.620	0.784	0.832
		40	0.976	0.996	1.000	0.956	0.988	0.996	0.936	0.984	0.992
		60	0.996	1.000	1.000	0.996	0.996	1.000	0.996	0.996	1.000
	60	20	0.892	0.944	0.972	0.840	0.924	0.944	0.804	0.896	0.944
		40	0.996	1.000	1.000	0.992	0.996	1.000	0.992	0.996	1.000
		60	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
6h	20	20	0.228	0.384	0.464	0.204	0.336	0.456	0.188	0.296	0.408
		40	0.400	0.612	0.708	0.356	0.552	0.712	0.332	0.476	0.588
		60	0.692	0.824	0.896	0.584	0.772	0.840	0.596	0.764	0.840
	40	20	0.416	0.632	0.756	0.416	0.584	0.688	0.412	0.556	0.664
		40	0.848	0.932	0.972	0.800	0.896	0.948	0.772	0.892	0.940
		60	0.964	0.984	0.996	0.952	0.976	0.992	0.944	0.980	0.992
	60	20	0.580	0.736	0.828	0.556	0.696	0.792	0.520	0.664	0.764
		40	0.964	0.984	0.992	0.948	0.976	0.988	0.924	0.976	0.988
		60	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Note: $J = \lfloor C(NT)^{1/7.5} \rfloor$ where $C = 1$, 1.5, and 2.

Guyana and include other four countries according to the data availability. The number of countries is 74 ($N = 74$) and the time period is 1960-2010 ($T = 51$).

We use the cubic B-spline to approximate the unknown function g. Note that g has three variables. Without imposing any structure on g, we need to use the tensor product of the sieve bases for each variable to approximate the unknown function. Then the total number of sieve approximation terms is $K = (J+4)^3$. Even for a small number of knots $J = 1, 2$, or 3, we have $K = 125, 216$, or 343, respectively. This is the notorious "curse of dimensionality" in nonparametric regression. For this reason, we only allow bivariate interactions and a single trivariate interaction term in our sieve estimation. Specifically, our sieve approximate terms are comprised of $p_Y^{J+4}(Y_{i,t-1}) \otimes p_I^{J+4}(I_{it})$, $p_Y^{J+4}(Y_{i,t-1}) \otimes p_{\Delta I}^{J+3}(\Delta I_{it})$, $p_{\Delta I}^{J+3}(\Delta I_{it}) \otimes p_I^{J+3}(I_{it})$, and $Y_{i,t-1} I_{it} \Delta I_{it}$ where we have avoided perfect multicollinearity. In this case, the total number of sieve approximating terms is $(J+4)^2 + (J+4)(J+3) + (J+3)^2 + 1$. To choose the number of factors, we follow Bai and Ng (2002) and adopt the following information criteria:

$$
\begin{aligned}
PC_1(R) &= V\left(R, \hat{f}^R\right) + R\hat{\sigma}^2 \left(\frac{N+T}{NT}\right) \ln\left(\frac{NT}{N+T}\right), \\
PC_2(R) &= V\left(R, \hat{f}^R\right) + R\hat{\sigma}^2 \left(\frac{N+T}{NT}\right) \ln\left[\min(N,T)\right], \\
IC_1(R) &= \ln\left[V\left(R, \hat{f}^R\right)\right] + R\left(\frac{N+T}{NT}\right) \ln\left(\frac{NT}{N+T}\right), \\
IC_2(R) &= \ln\left[V\left(R, \hat{f}^R\right)\right] + R\left(\frac{N+T}{NT}\right) \ln\left[\min(N,T)\right],
\end{aligned}
$$

where $V(R, \hat{f}^R) = (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T \left(\hat{e}_{it}^R\right)^2$, $\hat{e}_{it}^R = Y_{it} - \hat{g}^R(X_{it}) - \hat{\lambda}_i^{R\prime} \hat{f}_t^R$, $\hat{g}^R(\cdot)$, $\hat{f}_t^R$ and $\hat{\lambda}_i^R$ are estimates when R factors are used, and $\hat{\sigma}^2$ is a consistent estimate for $(NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T E(e_{it}^2)$ and is replaced by $V(R_{\max}, \hat{f}^{R_{\max}})$ in applications. Here $R_{\max}$ denotes the maximum number of factors under consideration and has to be specified in advance. In simulations we find that IC_1 and IC_2 work fairly well in finite samples for different choices of knots in cubic B splines, but PC_1 and PC_2 tend to choose a larger number of factors, which may be close to the largest upper bound sometimes. When this occurs, we use the number of factors recommended by IC_1 and IC_2. We follow Bai and Ng (2006b) and set $R_{\max} = 8$ throughout. For both estimation and testing, we use $M_T = \lfloor T^{1/7.5} \rfloor$ for bias correction as in the simulations and consider a sequence of knots in the cubic B-spline: $J = 3, 4, ..., 8$.

To reduce the risk of structural change, we partition the full sample (1960-2010) into two subsamples (1960-1985 and 1986-2010). For both the full sample and two subsamples, IC_1 and IC_2 recommend $1 \backsim 2$ factors both for linear estimation and sieve estimations with different choices of J. So we set $R = 2$ for all samples. We first consider the problem of estimation and report the estimation results for the two subsamples in Figures 1 and 2, respectively. Figure 1 plots the estimation of $g(\cdot,\cdot,\cdot)$ against each of its three arguments when the other two are fixed at their sample medians. For example, Figures 1(a)-(c) report the estimates of $g(\cdot, \bar{I}, \bar{\Delta} I)$ together with their bootstrap-based 90% pointwise confidence bands for $J = 3, 5$, and 7, respectively, where $\bar{I}$ and $\bar{\Delta} I$ are the respective sample medians of I_{it}'s and ΔI_{it}'s in the first subsample (1960-1985). Figure 2 repeats the above exercises for the second subsample (1986-2010). We summarize some important findings from these figures. First, as expected, the fitted curves tend to be smooth for a small value of J and rough for a large value of J. By looking at those plots alone, whether one can conclude a regressor (e.g., lagged economic growth rate) has significant nonlinear effect on the economic growth rate simply depends on the choice of J. This calls upon a formal test for the linear functional form. Second, Figures 1(a)-(c) and 2(a)-(c) suggest that lagged economic growth rate

Table 6: Bootstrap p-values for testing the linear economic growth model

Subsamples\\J	3	4	5	6	7	8
1960 − 1985 (T=26, N=74)	0.0000	0.0001	0.0001	0.0002	0.0003	0.0000
1986 − 2010 (T=25, N=74)	0.0030	0.0028	0.0022	0.0019	0.0021	0.0019
1960 − 2010 (T=51, N=74)	0.0498	0.0427	0.0390	0.0338	0.0299	0.0261

is globally positively related to the current economic growth rate when investment share and its growth are fixed at their sample medians. Third, Figures 1(d)-(f) and 2(d)-(f) suggest that investment share generally has positive effect on the economic growth rate. Fourth, Figures 1(g)-(i) and 2(g)-(i) indicate that the effect of the change of investment on the economic growth rate is nonlinear and non-monotone, and the effect tends to vary across subsamples. This suggests that some sort of structural change may occur during the full sample period.

Table 6 reports the bootstrap p-values for the specification test of linearity for both subsamples and the full sample based on 10000 bootstrap resamples. The p-values are smaller than 0.05 across all J's for both subsamples and the full sample as well. This suggests a strong degree of nonlinearity in the data.

7 Conclusion

In this paper we consider the estimation and testing for large dimensional nonparametric dynamic panel data models with interactive fixed effects. A sieve-based QMLE is proposed to estimate the nonparametric function and common components jointly. Following Moon and Weidner (2010, 2012), we derive the convergence rate for the sieve estimator and establish its asymptotic distribution. The sources of different asymptotic biases are discussed in detail and a consistent bias-corrected estimator is provided. We also propose a consistent specification test for the commonly used linear dynamic panel data models based on the L_2 distance between the linear and sieve estimators. We establish the asymptotic distributions of the test statistic under both the null hypothesis and a sequence of Pitman local alternatives. To improve the finite sample performance of the test, we also propose a bootstrap procedure to obtain the bootstrap p-values and justify its asymptotic validity. Through Monte Carlo simulations, we investigate the finite sample performance of our estimator and test statistic. We apply the model to an economic growth data set and demonstrate that lagged economic growth rate, investment share and its change have significant nonlinear effect on the economic growth rate.

REFERENCES

Ahn, S.C., Y.H. Lee, & P. Schmidt (2001) GMM estimation of linear panel data models with time-varying individuals effects. *Journal of Econometrics* 101, 219–255.

Ahn, S.C., Y.H. Lee, & P. Schmidt (2013) Panel data models with multiple time-varying individual effects. *Journal of Econometrics* 174, 1–14.

Ahn, S.C. & H.R. Moon (2001) Large-N and large-T properties of panel data estimators and the Hausman test. Mimeo, Arizona State University.

Arellano, M. (2003) *Panel Data Econometrics*. Oxford: Oxford University Press.

Baglan, D. (2009) Efficient estimation of a partially linear dynamic panel data model with fixed effects: application to unemployment dynamics in the U.S.. Working paper, Howard University.

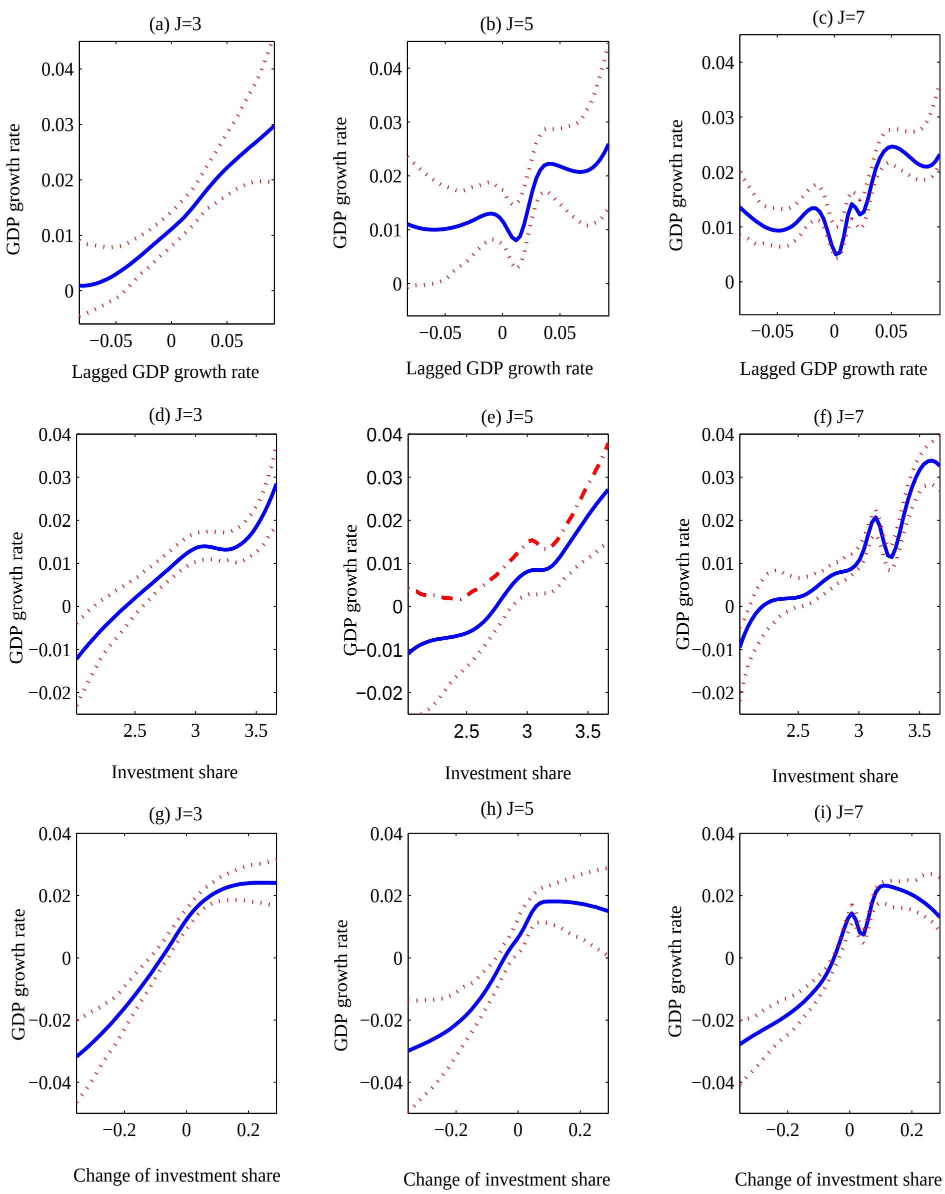

Figure 1: Relationship between GDP growth rate and lagged GDP growth rate, investment share, and change of investment share(1960-1985) (solid line: estimated function, dotted lines: 90% bootstrap confidence band)

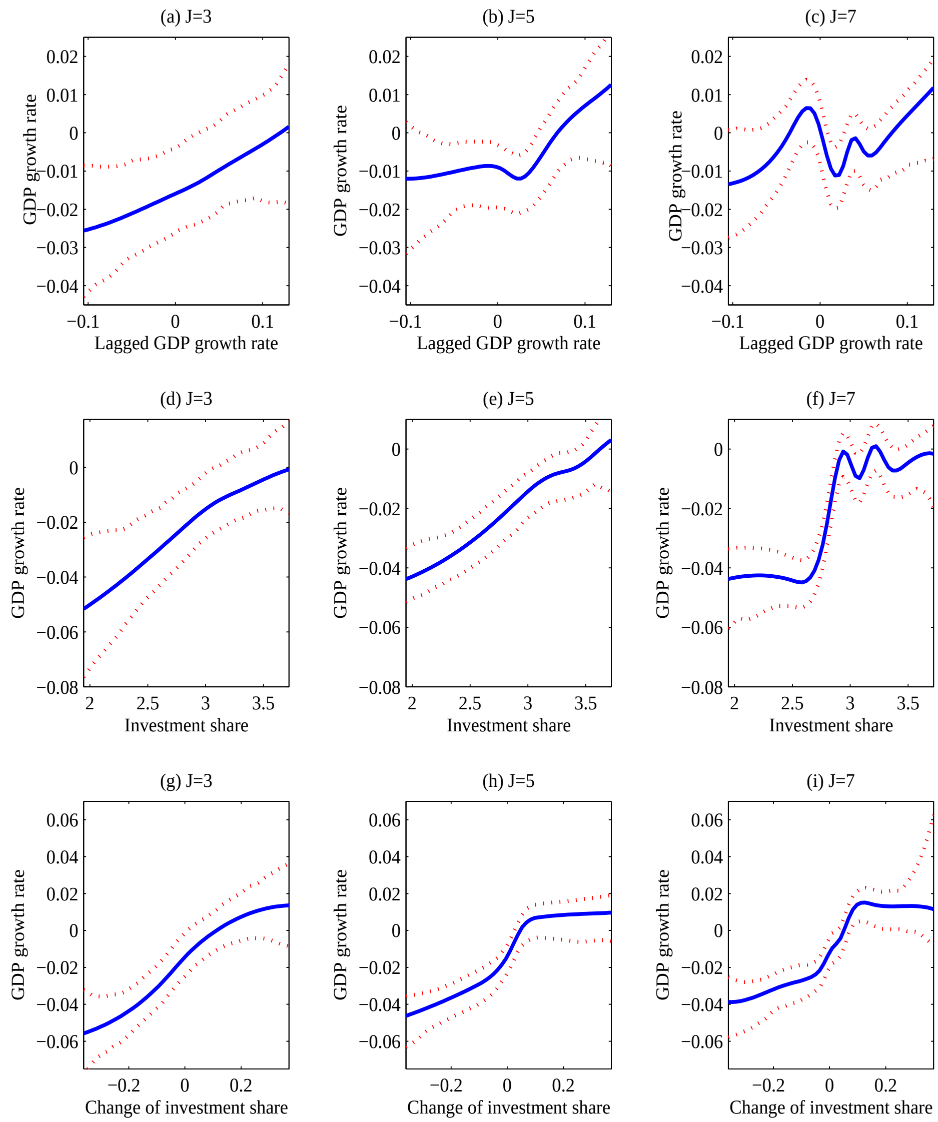

Figure 2: Relationship between GDP growth rate and lagged GDP growth rate, investment share, and change of investment share(1986-2010) (solid line: estimated function, dotted lines: 90% bootstrap confidence band)

Bai, J. (2003) Inferential theory for factor models of large dimensions. *Econometrica* 71, 135–172.

Bai, J. (2009) Panel data model with interactive fixed effects. *Econometrica* 77, 1229–1279.

Bai, J. & K. Li (2012) Statistical analysis of factor models of high dimension. *Annals of Statistics* 40, 436–465.

Bai, J. & S. Ng (2002) Determining the number of factors in approximate factor models. *Econometrica* 70, 191–221.

Bai, J. & S. Ng (2006a) Confidence intervals for diffusion index forecasts and inference for factor-augmented regression. *Econometrica* 74, 1133–1150.

Bai, J. & S. Ng (2006b) Evaluating latent and observed factors in macroeconomics and finance. *Journal of Econometrics* 131, 507–537.

Bai, J. & S. Ng (2008) Large dimensional factor analysis. *Foundations and Trends in Econometrics*, 3:2, 89–163.

Baltagi, B.H. (2008) *Econometric Analysis of Panel Data*, 4th ed. West Sussex: John Wiley & Sons.

Blundell, R., X. Chen, & D. Kristensen (2007) Semi-nonparametric IV estimation of shape invariant Engel curves. *Econometrica* 75, 1613–1670

Bond, S., A. Leblebicioglu, & F. Schiantarelli (2010) Capital accumulation and growth: a new look at the empirical evidence. *Journal of Applied Econometrics* 25, 1073–1099.

Bosq, D. (1998) *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction*. Springer Verlag.

Chambers, D. & J.-T. Guo (2009) Natural resources and economic growth: some theory and evidence. *Annals of Economics and Finance* 10, 367–389.

Chen, X. (2007) Large sample sieve estimation of semi-nonparametric models. In J.J. Heckman & E. Leamer (eds), *Handbook of Econometrics* 6, 5549–5632, North Holland, Amsterdam.

Chen, X. (2011) Penalized sieve estimation and inference of semi-nonparametric dynamic models: a selective review. Forthcoming in *Advances in Economics and Econometrics*, Cambridge University Press.

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

Chen, X., J. Racine, & N.R. Swanson (2001) Semiparametric ARX neural-network models with an application to forecasting inflation. *IEEE Transactions on Neural Networks* 12, 674–683.

Choi, I. (2012) Efficient estimation of factor models. *Econometric Theory* 28, 274–308.

Davidson, J. (1994) *Stochastic Limit Theory*. Oxford: Oxford University Press.

de Jong, P. (1987) A central limit theorem for generalized quadratic forms. *Probability Theory and Related Fields* 75, 261–277.

de Jong, R. (2009) Nonlinear time series models and weakly dependent innovations. Working paper, Ohio State University.

Doukhan, P. (1994) *Mixing: Properties and Examples*. New York: Springer-Verlag.

Easterly, W. & R. Levine (2001) What have we learned from a decade of empirical research on growth? It's not factor accumulation: stylized facts and growth models. *World Bank Economic Review* 15, 177–219.

Fan, J. & Q. Yao (2003) *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer Verlag.

Fan, J., Q. Yao & Z. Cai (2003) Adaptive varying-coefficient linear models. *Journal of the Royal Statistical Society: Series B* 65, 57–81.

Freyberger, J. (2012) Nonparametric panel data models with interactive fixed effects. Working paper, Northwestern University.

Franke, J. & M. Diagne (2006) Estimating market risk with neural networks. *Statistics & Decisions* 24, 233–253.

Gallant, A.R. (1987) *Nonlinear Statistical Models*. New York: Wiley.

Gallant, A.R. & H. White (1988) *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. New York, Basil Blackwell.

Gagliardini, P. & C. Gourieroux (2012) Efficiency in large dynamic panel models with common factors. Working paper, University of Lugano.

Greenaway-McGrevy, R., C. Han, & D. Sul (2012) Asymptotic distribution of factor augmented estimators for panel regression. *Journal of Econometrics* 169, 48–53.

Hahn, J. & G. Kuersteiner (2010) Stationarity and mixing properties of the dynamic Tobit model. *Economics Letters* 107, 105–111.

Hahn, J. & G. Kuersteiner (2011) Bias reduction for dynamic nonlinear panel models with fixed effects. *Econometric Theory* 27, 1152–1191.

Hansen, B.E. (2000) Testing for structural change in conditional models. *Journal of Econometrics* 97, 93–115.

Harding, M. (2009) Structural estimation of high-dimensional factor models. Working paper, Stanford University.

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

Henderson, D.J., R.J. Carroll, & Q. Li (2008) Nonparametric estimation and testing of fixed effects panel data models. *Journal of Econometrics* 144, 257–275.

Heston, A., R. Summers, & B. Aten (2012) *Penn World Table Version* 7.1. Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania.

Hong, Y. & H. White (1995) Consistent specification testing via nonparametric series regression. *Econometrica* 63, 1133–1160.

Hsiao, C. (2003) *Analysis of Panel Data*. Cambridge and New York: Cambridge University Press.

Jin, S. & L. Su (2013) Nonparametric tests for poolability in panel data models with cross section dependence. *Econometric Reviews* 32, 469–512

Jones, C.I. (1995) Time series tests of endogenous growth models. *Quarterly Journal of Economics* 110, 495–525.

Kapetanios, G. & M.H. Pesaran (2007) Alternative approaches to estimation and inference in large multifactor panels: small sample results with an application to modelling of asset returns. In G. Phillips & E. Tzavalis (eds.), *The Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis*, CUP, Ch.11, 239–281.

Kato, T. (1980) *Perturbation Theory for Linear Operators*. Springer-Verlag.

Lee, Y. (2013) Nonparametric estimation of dynamic panel models with fixed effects. Working paper, Michigan University.

Lee, Y.-J. (2011) Testing a linear dynamic panel data model against nonlinear alternatives. Working paper, Indiana University.

Li, Q. & Y. Sun (2011) A consistent nonparametric test of parametric regression functional form in fixed effects panel data models. Working paper, Texas A&M University.

Lu, X. & L. Su (2013) Shrinkage estimation of dynamic panel data models with interactive fixed effects. Working paper, Hong Kong University of Science & Technology.

Masry, E. & D. Tjøtheim (1997) Additive nonlinear ARX time series and projection estimates. *Econometric Theory* 13, 214–252.

Meierrieks, D. & T. Gries (2012) Economic performance and terrorist activity in Latin America. *Defence and Peace Economics* 23, 447–470.

Meitz, M. & P. Saikkonen (2010) A note on the geometric ergodicity of a nonlinear AR-ARCH model. *Statistics & Probability Letters* 80, 631–638.

Moon, H.R. & M. Weidner (2010) Dynamic linear panel data regression models with interactive fixed effects. Working paper, University of Southern California.

Moon, H.R. & M. Weidner (2012) Linear regression for panel with unknown number of factors as interactive fixed effects. Working paper, University of Southern California.

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

Onatski, A. (2009) Testing hypotheses about the number of factors in large factor models. *Econometrica* 77, 1447–1479.

Prakasa Rao, B.L.S. (2009) Conditional independence, conditional mixing and conditional association. *Annals of the Institute of Statistical Mathematics* 61, 441–460.

Pesaran, M.H. (2006) Estimation and inference in large heterogeneous panels with multifactor error. *Econometrica* 74, 967–1012.

Pesaran, M.H. & E. Tosetti (2011) Large panels with common factors and spatial correlation. *Journal of Econometrics* 161, 182–202.

Pollard, D. (1984) *Convergence of Stochastic Processes.* Springer-Verlag, New York.

Pötscher, B.M. & I.R. Prucha (1997) *Dynamic Nonlinear Econometric Models: Asymptotic Theory.* Springer-Verlag Berlin Heidelberg.

Racine, J.S. (2012) A primer on regression splines. CRAN.R-Project. Available online: http://cran.r-project.org/web/packages/crs/vignettes/spline_primer.pdf.

Roussas, G.G. (2008) On conditional independence, conditional mixing and conditional association. *Annals of the Institute of Statistical Mathematics* 61, 441–460.

Su, L. & Q. Chen (2013) Testing homogeneity in panel data models with interactive fixed effects. *Econometric Theory,* forthcoming.

Su, L. & S. Jin (2012) Sieve estimation of panel data models with cross section dependence. *Journal of Econometrics* 169, 34–47.

Su, L., S. Jin, & Y. Zhang (2013) Specification test for panel data models with interactive fixed effects. Working paper, Singapore Management University.

Su, L. & X. Lu (2013) Nonparametric dynamic panel data models: kernel estimation and specification testing. *Journal of Econometrics*, forthcoming.

Tjøstheim, D. (1990) Non-linear time series and Markov chains. *Advances in Applied Probability* 22, 587–611.

Zaffaroni, P. (2010) Generalized least squares estimation of panel with common shocks. Working paper, School of Imperial College London.

APPENDIX

Throughout the appendix, let C signify a generic constant whose exact value may vary from case to case. Let $E_{\mathcal{D}}(\cdot) \equiv E(\cdot|\mathcal{D})$ and $\text{Var}_{\mathcal{D}}(\cdot) \equiv \text{Var}(\cdot|\mathcal{D})$. Let $E_{(\mathcal{D},S)}(\cdot)$ denote expectation with respect to variables indexed by set S conditional on $\mathcal{D}$. Let $\varsigma_N \equiv \mu_{\min}\left(\lambda^{0\prime}\lambda^0/N\right)$ and $\varsigma_T \equiv \mu_{\min}\left(f^{0\prime}f^0/T\right)$ where $\mu_{\min}(A)$ denotes the minimum eigenvalue of A. Let $\epsilon_k \equiv \beta_k^0 - \beta_k$ for $k = 1, ..., K$, $\epsilon_0 \equiv \|\mathbf{u}\|/\sqrt{NT}$ and $\mathbf{P}_0 \equiv (\sqrt{NT}/\|\mathbf{u}\|)\mathbf{u}$. Let $\vartheta_{NT} \equiv \sum_{k=0}^{K} \epsilon_k \mathbf{P}_k$, $d_{\max}\left(\lambda^0, f^0\right) \equiv \sqrt{\mu_1\left(\frac{1}{NT}\lambda^{0\prime}f^0 f^{0\prime}\lambda^0\right)}$, and $d_{\min}\left(\lambda^0, f^0\right) \equiv \sqrt{\mu_R\left(\frac{1}{NT}\lambda^{0\prime}f^0 f^{0\prime}\lambda^0\right)}$. Define $r_0\left(\lambda^0, f^0\right) \equiv \left(\frac{4d_{\max}(\lambda^0,f^0)}{d_{\min}^2(\lambda^0,f^0)} + \frac{1}{2d_{\max}(\lambda^0,f^0)}\right)^{-1}$ and $\alpha_{NT} \equiv \frac{\|\vartheta_{NT}\|}{\sqrt{NT}} \frac{16 d_{\max}(\lambda^0,f^0)}{d_{\min}^2(\lambda^0,f^0)}$. Below we prove the main results in Sections 3 and 4. The proofs of all technical lemmas and Theorem 4.4 are given in the online Supplementary Material which is available on the first author's website.

A Proofs of the main results in Section 3

A.1 Convergence rate of $\hat{g}(x)$

Lemma A.1 *Suppose that Assumptions 1-4 hold. Then* $||\hat{\beta} - \beta^0|| = O_P(K^{-\gamma/(2d)} + \delta_{NT}^{-1/2})$.

Proof of Theorem 3.1. Let $a_k \equiv (\hat{\beta}_k - \beta_k^0)/||\hat{\beta} - \beta^0||$ and $\mathbf{P}_{(a)} \equiv \sum_{k=1}^{K} a_k \mathbf{P}_k$ with $\|a\| = 1$. By Lemma A.1, Assumptions 1(*iii*), 2(*iii*), 3(*i*)-(*iii*), and 4, we have $\frac{\|\vartheta_{NT}\|}{\sqrt{NT}} \leq \frac{\|\mathbf{e}\| + \|\mathbf{e}_g\|}{\sqrt{NT}} + ||\hat{\beta} - \beta^0|| \frac{\|\mathbf{P}_{(a)}\|}{\sqrt{NT}} = O_P\left(\delta_{NT}^{-1} + K^{-\gamma/d}\right) + O_P(K^{-\gamma/(2d)} + \delta_{NT}^{-1/2}) = o_P(1)$. By Assumptions 1(*i*)-(*ii*), $r_0\left(\lambda^0, f^0\right) = O_P(1)$. It follows that $\|\vartheta_{NT}\|/\sqrt{NT} \leq r_0\left(\lambda^0, f^0\right)$ w.p.a.1. and we can apply Proposition C.1 in the supplementary appendix to expand $L_{NT}(\beta)$ as follows

$$
\begin{aligned}
L_{NT}(\beta) &= \frac{1}{NT}\sum_{k_1=0}^{K}\sum_{k_2=0}^{K}\epsilon_{k_1}\epsilon_{k_2}L^{(2)}\left(\lambda^0, f^0, \mathbf{P}_{k_1}, \mathbf{P}_{k_2}\right) \\
&\quad + \frac{1}{NT}\sum_{k_1=0}^{K}\sum_{k_2=0}^{K}\sum_{k_3=0}^{K}\epsilon_{k_1}\epsilon_{k_2}\epsilon_{k_3}L^{(3)}\left(\lambda^0, f^0, \mathbf{P}_{k_1}, \mathbf{P}_{k_2}, \mathbf{P}_{k_3}\right) + O_P\left(\alpha_{NT}^4\right) \\
&= L_{NT}\left(\beta^0\right) + L_{1,NT}(\beta) + L_{2,NT}(\beta) + L_{R,NT}(\beta) + O_P\left(\alpha_{NT}^4\right) - O_P\left(\epsilon_0^4\right),
\end{aligned}
$$

where $L^{(2)}$ and $L^{(3)}$ are defined in Proposition C.1,

$$
\begin{aligned}
L_{NT}\left(\beta^0\right) &= \frac{1}{NT}\epsilon_0^2 L^{(2)}\left(\lambda^0, f^0, \mathbf{P}_0, \mathbf{P}_0\right) + \frac{1}{NT}\epsilon_0^3 L^{(3)}\left(\lambda^0, f^0, \mathbf{P}_0, \mathbf{P}_0, \mathbf{P}_0\right) + O_P\left(\epsilon_0^4\right), \\
L_{1,NT}(\beta) &= \frac{2}{NT}\sum_{k=1}^{K}\epsilon_k\epsilon_0 L^{(2)}\left(\lambda^0, f^0, \mathbf{P}_k, \mathbf{P}_0\right) + \frac{3}{NT}\sum_{k=1}^{K}\epsilon_k\epsilon_0^2 L^{(3)}\left(\lambda^0, f^0, \mathbf{P}_k, \mathbf{P}_0, \mathbf{P}_0\right), \\
L_{2,NT}(\beta) &= \frac{1}{NT}\sum_{k_1=1}^{K}\sum_{k_2=1}^{K}\epsilon_{k_1}\epsilon_{k_2}L^{(2)}\left(\lambda^0, f^0, \mathbf{P}_{k_1}, \mathbf{P}_{k_2}\right), \text{ and} \\
L_{R,NT}(\beta) &= \frac{1}{NT}\sum_{k_2=1}^{K}\sum_{k_2=1}^{K}\sum_{k_3=1}^{K}\epsilon_{k_1}\epsilon_{k_2}\epsilon_{k_3}L^{(3)}\left(\lambda^0, f^0, \mathbf{P}_{k_1}, \mathbf{P}_{k_2}, \mathbf{P}_{k_3}\right) \\
&\quad + \frac{3}{NT}\sum_{k_2=1}^{K}\sum_{k_2=1}^{K}\epsilon_{k_1}\epsilon_{k_2}\epsilon_0 L^{(3)}\left(\lambda^0, f^0, \mathbf{P}_{k_1}, \mathbf{P}_{k_2}, \mathbf{P}_0\right) + O_P\left[\left(\|\beta - \beta^0\| + \epsilon_0\right)^4 - \epsilon_0^4\right] \\
&= O_P\left(\|\beta - \beta^0\|^2 \epsilon_0 + \|\beta - \beta^0\|^3 + \|\beta - \beta^0\|\epsilon_0^3\right) \qquad \text{(A.1)}
\end{aligned}
$$

Clearly, $L_{1,NT}(\beta)$ and $L_{2,NT}(\beta)$ are linear and quadratic in ϵ_k, $k=1,...,K$, respectively, and $L_{R,NT}(\beta)$ includes the third and higher order asymptotically negligible terms in the likelihood expansion. Noting that $L^{(s)}\left(\lambda^0, f^0, \mathbf{P}_{k_1}, \cdots, \mathbf{P}_{k_s}\right)$ is linear in the last s arguments, we have

$$L_{1,NT}(\beta) = -2(\beta-\beta^0)'(C_{NT}^{(1)} + C_{NT}^{(2)}) \text{ and } L_{2,NT}(\beta) = (\beta-\beta^0)'W_{NT}(\beta-\beta^0),$$

where $C_{NT}^{(1)}$ and $C_{NT}^{(2)}$ are defined in Theorem 3.1. Then

$$\begin{aligned} L_{NT}(\beta) &= L_{NT}\left(\beta^0\right) - 2(\beta-\beta^0)'(C_{NT}^{(1)} + C_{NT}^{(2)}) + (\beta-\beta^0)'W_{NT}(\beta-\beta^0) \\ &\quad + O_P\left\{\left\|\beta-\beta^0\right\|^2 \epsilon_0 + \left\|\beta-\beta^0\right\|^3 + \left\|\beta-\beta^0\right\| \epsilon_0^3\right\}. \end{aligned} \tag{A.2}$$

Noting that rank$(\mathbf{P}_k\Phi'\mathbf{u}M_{f^0}\mathbf{u}'M_{\lambda^0} + \mathbf{P}_kM_{f^0}\mathbf{u}'M_{\lambda^0}\mathbf{u}\Phi' + \mathbf{P}_kM_{f^0}\mathbf{u}'\Phi\mathbf{u}'M_{\lambda^0}) \leq 3R$ and using the trace inequality tr$(A) \leq$rank$(A)\|A\|$ for any real square matrix A, we have $C_{NT,k}^{(2)} = \frac{1}{NT}\text{tr}(\mathbf{P}_k\Phi'\mathbf{u}M_{f^0}\mathbf{u}'M_{\lambda^0} + \mathbf{P}_kM_{f^0}\mathbf{u}'M_{\lambda^0}\mathbf{u}\Phi' + \mathbf{P}_kM_{f^0}\mathbf{u}'\Phi\mathbf{u}'M_{\lambda^0}) \leq \frac{3R}{NT}\|\mathbf{P}_k\|\|\Phi\|\|\mathbf{u}\|^2\|M_{\lambda^0}\|\left\|M_{f^0}\right\| = \frac{\|\mathbf{P}_k\|}{\sqrt{NT}}O_P\left(K^{-2\gamma/d} + \delta_{NT}^{-2}\right)$. It follows that

$$\left\|C_{NT}^{(2)}\right\| = \left\{\sum_{k=1}^K \frac{\|\mathbf{P}_k\|^2}{NT}\right\}^{1/2} O_P(K^{-2\gamma/d} + \delta_{NT}^{-2}) = O_P\left[\sqrt{K}(K^{-2\gamma/d} + \delta_{NT}^{-2})\right]. \tag{A.3}$$

For $C_{NT}^{(1)}$, we have $||W_{NT}^{-1}C_{NT}^{(1)}|| = ||W_{NT}^{-1}(NT)^{-1}\sum_{i=1}^N Z_i'e_i|| + ||W_{NT}^{-1}(NT)^{-1}\sum_{i=1}^N Z_i'e_{g,i}||$. By Assumption 3$(v)$, the first term is $O_P(\delta_{NT}^{-1}K^{1/2}/T^{1/2})$. Let $\overrightarrow{e}_g \equiv (e_{g,1}', \cdots, e_{g,N}')'$, $\overrightarrow{Z} \equiv (Z_1', \cdots, Z_N')'$ and $\overrightarrow{W} \equiv (NT)^{-1}\overrightarrow{Z}'W_{NT}^{-1}\overrightarrow{Z}$. Noting that $\overrightarrow{W}$ is a projection matrix with $\mu_1(\overrightarrow{W}) = 1$ and by Assumptions 2(ii) and 3(i)-(iii), $\left\|W_{NT}^{-1}\frac{1}{NT}\sum_{i=1}^N Z_i'e_{g,i}\right\|^2 = \frac{1}{N^2T^2}\text{tr}(\overrightarrow{e}_g\overrightarrow{Z}'W_{NT}^{-1}W_{NT}^{-1}\overrightarrow{Z}\overrightarrow{e}_g) \leq [\mu_{\min}(W_{NT})]^{-1}\frac{1}{N^2T^2}\text{tr}(\overrightarrow{e}_g\overrightarrow{W}\overrightarrow{e}_g')$ $\leq O_P(1)\|\overrightarrow{e}_g\|_F^2/(NT) = O_P\left(K^{-2\gamma/d}\right)$. It follows that

$$\left\|W_{NT}^{-1}C_{NT}^{(1)}\right\| = O_P\left(\delta_{NT}^{-1}\sqrt{K/T} + K^{-\gamma/d}\right). \tag{A.4}$$

Let

$$r_{NT} \equiv W_{NT}^{-1}C_{NT}^{(1)} + W_{NT}^{-1}C_{NT}^{(2)} \text{ and } R_{NT} \equiv \hat{\beta} - \beta^0 - r_{NT}. \tag{A.5}$$

From (A.3) and (A.4) we have

$$\|r_{NT}\| \leq \left\|W_{NT}^{-1}C_{NT}^{(1)}\right\| + \left\|W_{NT}^{-1}C_{NT}^{(2)}\right\| = O_P\left(\sqrt{K}\delta_{NT}^{-2} + K^{-\gamma/d}\right). \tag{A.6}$$

Since $L_{NT}(\hat{\beta}) \leq L_{NT}\left(\beta^0 + r_{NT}\right)$, we can apply (A.2) to the objects on both sides of the last inequality to obtain

$$\begin{aligned} & \|R_{NT}\|^2 \\ \leq\ & [\mu_{\min}(W_{NT})]^{-1}\left(\hat{\beta} - \beta^0 - r_{NT}\right)' W_{NT}\left(\hat{\beta} - \beta^0 - r_{NT}\right) \\ \leq\ & [\mu_{\min}(W_{NT})]^{-1}\left[L_{R,NT}\left(\beta^0 + r_{NT}\right) - L_{R,NT}(\beta^0 + (\hat{\beta} - \beta^0))\right] \\ \leq\ & O_P\left(\|r_{NT}\|^2\epsilon_0 + \|r_{NT}\|\epsilon_0^3 + \|r_{NT}\|^3\right) + O_P\left(||\hat{\beta}-\beta^0||^2\epsilon_0 + ||\hat{\beta}-\beta^0||\epsilon_0^3 + ||\hat{\beta}-\beta^0||^3\right). \end{aligned} \tag{A.7}$$

We now argue that $||\hat{\beta}-\beta^0|| = O_P(||r_{NT}||)$ by contradiction. Suppose $||r_{NT}|| = o_P(||\hat{\beta}-\beta^0||)$. Then by (A.5) and (A.7), $||\hat{\beta}-\beta^0||^2 = O_P(\|R_{NT}\|^2) \leq O_P(||\hat{\beta}-\beta^0||\epsilon_0^3)$, implying that $||\hat{\beta}-\beta^0|| \leq O_P\left(\epsilon_0^3\right)$. Noting that $\epsilon_0^3 = O_P\left(\delta_{NT}^{-3} + K^{-3\gamma/d}\right) = o_P(||r_{NT}||)$, this further implies that $||\hat{\beta}-\beta^0|| \leq o_P(||r_{NT}||)$, a contradiction. It follows that

$$\left\|\hat{\beta}-\beta^0\right\| = O_P(||r_{NT}||) = O_P\left(\sqrt{K}\delta_{NT}^{-2} + K^{-\gamma/d}\right) \tag{A.8}$$

and

$$R_{NT} = O_P\left(\|r_{NT}\|\,\epsilon_0^{1/2}\right) = O_P\left[\left(\sqrt{K}\delta_{NT}^{-2} + K^{-\gamma/d}\right)\left(\delta_{NT}^{-1/2} + K^{-\gamma/(2d)}\right)\right] \tag{A.9}$$

because $\epsilon_0^2/||r_{NT}|| = O_P(1)$ and $||r_{NT}||/\epsilon_0 = O_P(1)$ by Assumption 4.

Now we derive the convergence rate of $\hat{g}(x)$. By the C_r inequality, (A.8) and Assumption 3(ii)

$$\begin{aligned}
\int_{\mathcal{X}} [\hat{g}(x) - g(x)]^2 w(x)\,dx &= \int_{\mathcal{X}} \left\{p^K(x)'\left(\hat{\beta}-\beta^0\right) + \left[p^K(x)'\beta^0 - g(x)\right]\right\}^2 w(x)\,dx \\
&\leq 2\int_{\mathcal{X}}\left[p^K(x)'\left(\hat{\beta}-\beta^0\right)\right]^2 w(x)\,dx + 2\int_{\mathcal{X}}\left[g(x) - p^K(x)'\beta^0\right]^2 w(x)\,dx \\
&\leq 2\mu_1(Q_{pp,w})\left\|\hat{\beta}-\beta^0\right\|^2 + 2C_{w1}\left\|g(x) - p^K(x)'\beta^0\right\|^2_{\infty,\bar{\omega}} \\
&= O_P\left(\left\|\hat{\beta}-\beta^0\right\|^2 + K^{-2\gamma/d}\right) = O_P\left(K\delta_{NT}^{-4} + K^{-2\gamma/d}\right),
\end{aligned}$$

where $Q_{pp,w} \equiv \int_{\mathcal{X}} p^K(x)\,p^K(x)'\,w(x)\,dx$ with $\mu_1(Q_{pp,w}) < \infty$ and $C_{w1} \equiv \int_{\mathcal{X}}\left(1+\|x\|^2\right)^{\bar{\omega}} w(x)\,dx < \infty$. Similarly, using $\hat{g}(X_{it}) - g(X_{it}) = p_{it}'(\hat{\beta}-\beta^0) + [g(X_{it}) - p_{it}'\beta^0]$, the C_r inequality and Assumptions 3(ii)-(iii),

$$\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\hat{g}(X_{it}) - g(X_{it})\right]^2 w_{it} \\
\leq\;& \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[p_{it}'\left(\hat{\beta}-\beta^0\right)\right]^2 w_{it} + \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[g(X_{it}) - p_{it}'\beta^0\right]^2 w_{it} \\
\leq\;& 2\mu_1(Q_{wpp,NT})\left\|\hat{\beta}-\beta^0\right\|^2 + 2\left\|g(x) - p^K(x)'\beta^0\right\|^2_{\infty,\bar{\omega}}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(1+\|X_{it}\|^2\right)^{\bar{\omega}} w_{it} \\
=\;& O_P\left(\left\|\hat{\beta}-\beta^0\right\|^2\right) + O_P\left(K^{-2\gamma/d}\right) = O_P\left(K\delta_{NT}^{-4} + K^{-2\gamma/d}\right). \blacksquare
\end{aligned}$$

A.2 Asymptotic normality of $\hat{g}(x)$

Proof of Theorem 3.2. Recall that $V_K(x) = p^K(x)'\,\tilde{W}^{-1}\tilde{\Omega}\tilde{W}^{-1}p^K(x)$ and $A_{NT} = (NT)^{1/2}V_K^{-1/2}(x)$. Write

$$\begin{aligned}
&A_{NT}\left[\hat{g}(x) - g(x)\right] \\
=\;& A_{NT}p^K(x)'\left(\hat{\beta}-\beta^0\right) + A_{NT}\left[g(x) - p^K(x)'\beta^0\right] \\
=\;& A_{NT}p^K(x)'\,W_{NT}^{-1}C_{NT}^{(1)} + A_{NT}p^K(x)'\,W_{NT}^{-1}C_{NT}^{(2)} + A_{NT}p^K(x)'\,R_{NT} + A_{NT}\left[g(x) - p^K(x)'\beta^0\right] \\
\equiv\;& \Pi_{1NT} + \Pi_{2NT} + \Pi_{3NT} + \Pi_{4NT}, \text{ say.}
\end{aligned}$$

It suffices to show that: (i) $\Pi_{1NT} + \kappa_{NT}b_1(x) \xrightarrow{d} N(0,1)$, ($ii$) $\Pi_{2NT} = -\kappa_{NT}^{-1}b_2(x) - \kappa_{NT}b_3(x) + o_P(1)$, ($iii$) $\Pi_{3NT} = o_P(1)$, and (iv) $\Pi_{4NT} = o_P(1)$. We prove (i) and (ii) in Propositions A.6 and A.7 below, respectively. For (iii), by Cauchy-Schwarz inequality, (A.9) and Assumptions 7 and 8, we have

$$\begin{aligned}
\Pi_{3NT} &\leq \sqrt{\frac{NT}{V_K(x)}}\left\|p^K(x)\right\|\left\|R_{NT}\right\| \leq \mu_K^{-1/2}(\tilde{\Omega})\mu_1(\tilde{W})\sqrt{NT}\left\|R_{NT}\right\| \\
&= O_P\left[\sqrt{NT}\left(\sqrt{K}\delta_{NT}^{-2} + K^{-\gamma/d}\right)\left(\delta_{NT}^{-1/2} + K^{-\gamma/(2d)}\right)\right] = o_P(1),
\end{aligned}$$

as $V_K(x) = p^K(x)'\tilde{W}^{-1}\tilde{\Omega}\tilde{W}^{-1}p^K(x) \geq \mu_1^{-2}(\tilde{W})\mu_K(\tilde{\Omega})||p^K(x)||^2$. For ($iv$), by Assumptions 3($i$)-($ii$), and 8, we have for any $x \in \mathcal{X}$, $\Pi_{4NT} = (NT)^{1/2}V_K^{-1/2}(x)\left[g(x) - p^K(x)'\beta^0\right] \leq C\left\|p^K(x)\right\|^{-1}\left(1+\|x\|^2\right)^{\bar{\omega}/2} \times\sqrt{NT}\left\|g(x) - p^K(x)'\beta^0\right\|_{\infty,\bar{\omega}} = O_P(\sqrt{NT}K^{-\gamma/d}) = o_P(1)$ as $\inf_{x\in\mathcal{X}}||p^K(x)|| \geq C > 0$. $\blacksquare$

Next, we state some lemmas are used in the proofs of Propositions A.6-A.7 below.

Lemma A.2 *Let* $v_x^K \equiv V_K^{-1/2}(x)\tilde{W}^{-1}p^K(x)$ *and* $d_{it} \equiv v_x^{K\prime}\tilde{Z}_{it}$. *Suppose that the assumptions in Theorem 3.2 hold. Then*

(i) $\frac{1}{NT}\sum_{t=1}^T\sum_{i=1}^N \left\|d_{it}^4\right\|_{2,\mathcal{D}}^2 = O_P\left(K^4\right)$;

(ii) $\frac{1}{N^2T}\sum_{t=1}^T(\sum_{i=1}^N \left\|d_{it}^2\right\|_{2,\mathcal{D}}^2)^2 = O_P\left(K^4\right)$.

Lemma A.3 *Suppose that the assumptions in Theorem 3.2 hold. Then*

(i) $\left\|\tilde{W}_{NT}-\tilde{W}\right\|_F = O_P(K/\sqrt{NT})$;

(ii) $\left\|\tilde{W}_{NT}-W_{NT}\right\|_F = O_P(K/\sqrt{NT})$.

Lemma A.4 *Suppose that the assumptions in Theorem 3.2 hold. Then* $\frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T v_x^{K\prime}\{(Z_{it}-\tilde{Z}_{it})e_{it} - E_{\mathcal{D}}[(Z_{it}-\tilde{Z}_{it})e_{it}]\} = o_P(1)$.

Lemma A.5 *Suppose that the assumptions in Theorem 3.2 hold. Then*

(i) $\left\|\lambda^{0\prime}\mathbf{e}f^0\right\|_F = O_P(\sqrt{NT})$;

(ii) $\left\|P_{\lambda^0}\mathbf{e}P_{f^0}\right\|_F = O_P(1)$;

(iii) $\left\|f^{0\prime}\mathbf{e}'\mathbf{P}_{(a)}\right\|_F = O_P(\sqrt{NTK}\delta_{NT})$;

(iv) $\left\|P_{f^0}\mathbf{e}'\mathbf{P}_{(a)}\right\|_F = O_P(\sqrt{NK}\delta_{NT})$;

(v) $\left\|\lambda^{0\prime}\mathbf{e}\mathbf{P}_{(a)}'\right\|_F = O_P(N\sqrt{TK})$;

(vi) $\left\|P_{\lambda^0}\mathbf{e}\mathbf{P}_{(a)}'\right\|_F = O_P(\sqrt{NTK})$;

(vii) $\frac{1}{N\sqrt{T}}\sum_{t=1}^T\sum_{i=1}^N\sum_{j=1}^N A_i\lambda_j^0\left[e_{jt}e_{it}-E_{\mathcal{D}}\left(e_{jt}e_{it}\right)\right] = O_P(\sqrt{K})$;

(viii) $N^{-1}\sum_{i=1}^N ||T^{-1/2}\sum_{s=1}^T v_x^{K\prime}\left(p_{is}^c - p_{is}^{\lambda c}\right)f_s^0 G^0||^2 = O_P(K)$;

(ix) $N^{-1}\sum_{i=1}^N ||\left(NT\right)^{-1/2}\sum_{t=1}^T\sum_{j=1}^N \lambda_j^0\left[e_{it}e_{jt}-E_{\mathcal{D}}\left(e_{it}e_{jt}\right)\right]||^2 = O_P(1)$;

(x) $\frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T\sum_{s=1}^T B_t f_s^{0\prime}\left[e_{it}e_{is}-E_{\mathcal{D}}\left(e_{it}e_{is}\right)\right] = O_P(\sqrt{K})$;

(xi) $\frac{1}{NT}\sum_{t=1}^T ||\sum_{j=1}^N v_x^{K\prime}[p_{jt}^c - T^{-1}\sum_{l=1}^T \eta_{tl}p_{jl}^c]\lambda_j^{0\prime}G^0||^2 = O_P(K)$;

(xii) $\frac{1}{NT^2}\sum_{t=1}^T ||\sum_{i=1}^N\sum_{s=1}^T f_s^{0\prime}\left[e_{it}e_{is}-E_{\mathcal{D}}\left(e_{it}e_{is}\right)\right]||^2 = O_P(1)$;

where $A_i \equiv v_x^{K\prime}[E_{\mathcal{D}}\left(P_i - P_i^\lambda\right)']f^0G^0/T$, $P_i^\lambda = N^{-1}\sum_{j=1}^N \alpha_{ij}E_{\mathcal{D}}\left(P_j\right)$, $G^0 \equiv \left(f^{0\prime}f^0/T\right)^{-1}\left(\lambda^{0\prime}\lambda^0/N\right)^{-1}$, $p_{is}^c = p_{is} - E_{\mathcal{D}}\left(p_{is}\right)$, $p_{is}^{\lambda c} \equiv p_{is}^\lambda - E_{\mathcal{D}}\left(p_{is}^\lambda\right)$, $p_{is}^\lambda \equiv N^{-1}\sum_{j=1}^N \alpha_{ij}p_{js}$, $B_t \equiv v_x^{K\prime}E_{\mathcal{D}}(P_{\cdot t} - P_{\cdot t}^f)'\lambda^0G^0N^{-1}$, $P_{\cdot t}^f \equiv T^{-1}\sum_{l=1}^T \eta_{lt}P_{\cdot l}$, *and* $P_{\cdot t} \equiv (p_{1t}', \cdots, p_{Nt}')'$.

Proposition A.6 *Let the conditions in Theorem 3.2 hold. Then* $\Pi_{1NT} + \kappa_{NT}b_1(x) \xrightarrow{d} N(0,1)$.

Proof. Recall $v_x^K = V_K^{-1/2}(x)\tilde{W}^{-1}p^K(x)$. One can readily show that $\left\|v_x^K\right\| = O_P(1)$. Note that $\Pi_{1NT} = \frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T v_x^{K\prime}Z_{it}u_{it} - V_K^{-1/2}(x)p^K(x)'(\tilde{W}^{-1} - W_{NT}^{-1})\frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T Z_{it}u_{it} \equiv \Pi_{1NT,1} + \Pi_{1NT,2}$, say. We complete the proof by showing that (*i*) $\Pi_{1NT,1} + \kappa_{NT}b_1(x) \xrightarrow{d} N(0,1)$ and (*ii*) $\Pi_{1NT,2} = o_P(1)$.

First, we consider (*ii*). By (A.4), Lemmas A.3(*i*)-(*ii*), and Assumption 8, we have

$$\begin{aligned}
|\Pi_{1NT,2}| &\leq \left\|V_K^{-1/2}(x)p^K(x)'\tilde{W}^{-1}\right\|\left\{\sqrt{NT}\left\|W_{NT}-\tilde{W}\right\|_F\right\}\left\|W_{NT}^{-1}\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T Z_{it}u_{it}\right\| \\
&= O_P(1)O_P(K)O_P\left(\sqrt{K/T}\delta_{NT}^{-1} + K^{-\gamma/d}\right) = O_P\left(K\sqrt{K/T}\delta_{NT}^{-1} + K^{1-\gamma/d}\right) = o_P(1).
\end{aligned}$$

Now, we consider (i). Using $u_{it}=e_{it}+e_{g,it}$, we decompose $\Pi_{1NT,1}$ as follows: $\Pi_{1NT,1}=\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T} v_x^{K\prime}Z_{it}e_{it}+\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}v_x^{K\prime}Z_{it}e_{g,it}\equiv\Pi_{1NT,11}+\Pi_{1NT,12}$, say. By Cauchy-Schwarz inequality and Assumptions 3(i)-(iii) and 2(ii) we have

$$\begin{aligned}\Pi_{1NT,12} &\leq \sqrt{NT}\left\{v_x^{K\prime}\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}Z_{it}Z_{it}'\right)v_x^K\right\}^{1/2}\left\{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}e_{g,it}^2\right\}^{1/2}\\ &\leq \left\|v_x^K\right\|\mu_1^{1/2}\left(W_{NT}\right)\left\|g\left(x\right)-p^K\left(x\right)'\beta^0\right\|_{\infty,\bar{\omega}}\left\{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(1+\|X_{it}\|^2\right)^{\bar{\omega}}\right\}^{1/2}\\ &= O_P(\sqrt{NT}K^{-\gamma/d})=o_P\left(1\right).\end{aligned}$$

We are left to show $\Pi_{1NT,11}+\kappa_{NT}b_1\left(x\right)\xrightarrow{d}N(0,1)$. We further decompose $\Pi_{1NT,11}$ as follows

$$\begin{aligned}\Pi_{1NT,11}+\kappa_{NT}b_1\left(x\right) &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}v_x^{K\prime}\tilde{Z}_{it}e_{it}+\left\{\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}v_x^{K\prime}E_{\mathcal{D}}\left(Z_{it}e_{it}\right)+\kappa_{NT}b_1\left(x\right)\right\}\\ &\quad+\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}v_x^{K\prime}\left\{\left(Z_{it}-\tilde{Z}_{it}\right)e_{it}-E_{\mathcal{D}}\left[\left(Z_{it}-\tilde{Z}_{it}\right)e_{it}\right]\right\}\\ &\equiv \Pi_{1NT,11a}+\Pi_{1NT,11b}+\Pi_{1NT,11c},\text{ say,}\end{aligned}$$

where $\tilde{Z}_{it}=p_{it}-N^{-1}\sum_{j=1}^{N}\alpha_{ij}E_{\mathcal{D}}\left[p_{jt}\right]-T^{-1}\sum_{s=1}^{T}\eta_{ts}E_{\mathcal{D}}\left[p_{is}\right]+(NT)^{-1}\sum_{j=1}^{N}\sum_{s=1}^{T}\alpha_{ij}\eta_{ts}E_{\mathcal{D}}\left[p_{js}\right].$ We complete the proof by showing that: (ia) $\Pi_{1NT,11a}\xrightarrow{d}N\left(0,1\right)$, (ib) $\Pi_{1NT,11b}=o_P\left(1\right)$, and (ic) $\Pi_{1NT,11c}=o_P\left(1\right)$. (ic) follows from Lemma A.4. We are left to show (ia) and (ib).

Proof of (ia). Note that $\Pi_{1NT,11a}=\sum_{t=1}^{T}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}v_x^{K\prime}\tilde{Z}_{it}e_{it}=\sum_{t=1}^{T}\xi_{NT,t}$ where $\xi_{NT,t}\equiv(NT)^{-1/2}\sum_{i=1}^{N}v_x^{K\prime}\tilde{Z}_{it}e_{it}$. Recall that $\mathcal{F}_0^{t-1}=\sigma(\lambda^0,f^0,\{X_{it},X_{i,t-1},e_{i,t-1},\cdots\}_{i=1}^N)$. By Assumption 5$(ii)$, $E\left[\xi_{NT,t}|\mathcal{F}_0^{t-1}\right]=(NT)^{-1/2}\sum_{i=1}^{N}v_x^{K\prime}\tilde{Z}_{it}E\left[e_{it}|\mathcal{F}_0^{t-1}\right]=0$. That is, $\left\{\xi_{NT,t},\mathcal{F}_0^t\right\}_{t=1}^T$ is a martingale difference sequence (m.d.s.). Consequently, we can apply the martingale CLT (e.g., Pollard, 1984, p.171) to prove that $\Pi_{1NT,11a}\xrightarrow{d}N\left(0,1\right)$ by verifying that $(ia1)$ $\vec{\xi}_{NT}\equiv\sum_{t=1}^{T}E\left[\xi_{NT,t}^4|\mathcal{F}_0^{t-1}\right]=o_P\left(1\right)$ and $(ia2)$ $\sum_{t=1}^{T}\xi_{NT,t}^2-1=o_P\left(1\right)$.

Since $\vec{\xi}_{NT}\geq 0$, we will prove $(ia1)$ by showing that $E_{\mathcal{D}}\{\sum_{t=1}^{T}E\left[\xi_{NT,t}^4|\mathcal{F}_0^{t-1}\right]\}=o_P\left(1\right)$. Let $d_{it}\equiv v_x^{K\prime}\tilde{Z}_{it}$. Noting that $\{(p_{it},e_{it})\}_{t=1}^T$ are independent across i conditional $\mathcal{D}$ and $\left\{\xi_{NT,t},\mathcal{F}_0^t\right\}_{t=1}^T$ is an m.d.s., we have

$$\begin{aligned}E_{\mathcal{D}}\left[\vec{\xi}_{NT}\right] &= \frac{1}{N^2T^2}\sum_{t=1}^{T}\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\sum_{i_3=1}^{N}\sum_{i_4=1}^{N}E_{\mathcal{D}}\left[d_{i_1t}d_{i_2t}d_{i_3t}d_{i_4t}e_{i_1t}e_{i_2t}e_{i_3t}e_{i_4t}\right]\\ &= \frac{1}{N^2T^2}\sum_{t=1}^{T}\sum_{i=1}^{N}E_{\mathcal{D}}\left[d_{it}^4e_{it}^4\right]+\frac{3}{N^2T^2}\sum_{t=1}^{T}\left\{\sum_{i=1}^{N}E_{\mathcal{D}}\left[d_{it}^2e_{it}^2\right]\right\}^2-\frac{3}{N^2T^2}\sum_{t=1}^{T}\sum_{i=1}^{N}\left\{E_{\mathcal{D}}\left[d_{it}^2e_{it}^2\right]\right\}^2\\ &\equiv \vec{\xi}_{NT}\left(1\right)+3\vec{\xi}_{NT}\left(2\right)-3\vec{\xi}_{NT}\left(3\right),\text{ say.}\end{aligned}$$

By Hölder inequality, Lemmas A.2(i)-(ii), and Assumption 6(i), we have

$$\begin{aligned}\vec{\xi}_{NT}\left(1\right) &\leq \frac{1}{N^2T^2}\sum_{t=1}^{T}\sum_{i=1}^{N}\left\|e_{it}^4\right\|_{2,\mathcal{D}}\left\|d_{it}^4\right\|_{2,\mathcal{D}}\\ &\leq \frac{1}{NT}\left\{\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\left\|e_{it}^4\right\|_{2,\mathcal{D}}^2\right\}^{1/2}\left\{\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\left\|d_{it}^4\right\|_{2,\mathcal{D}}^2\right\}^{1/2}=O_P\left(\frac{K^2}{NT}\right)=o_P\left(1\right),\end{aligned}$$

and

$$\begin{aligned}
\vec{\xi}_{NT}(2) &\leq \frac{1}{N^2T^2}\sum_{t=1}^{T}\left\{\sum_{i=1}^{N}\left\|d_{it}^2\right\|_{2,\mathcal{D}}\left\|e_{it}^2\right\|_{2,\mathcal{D}}\right\}^2 \leq \frac{1}{N^2T^2}\sum_{t=1}^{T}\left\{\sum_{i=1}^{N}\left\|d_{it}^2\right\|_{2,\mathcal{D}}^2\right\}\left\{\sum_{i=1}^{N}\left\|e_{it}^2\right\|_{2,\mathcal{D}}^2\right\} \\
&\leq \frac{1}{N^2T^2}\left\{\sum_{t=1}^{T}\left[\sum_{i=1}^{N}\left\|d_{it}^2\right\|_{2,\mathcal{D}}^2\right]^2\right\}^{1/2}\left\{\sum_{t=1}^{T}\left[\sum_{i=1}^{N}\left\|e_{it}^2\right\|_{2,\mathcal{D}}^2\right]^2\right\}^{1/2} \\
&= (NT)^{-2} O_p\left[\left(TN^2K^4\right)^{1/2}\right] O_P\left[\left(TN^2\right)^{1/2}\right] = O_P\left(K^2/T\right) = o_P(1).
\end{aligned}$$

Similarly, we can show that $\vec{\xi}_{NT}(3) = O_P\left(K^2/(NT)\right) = o_P(1)$ by Lemma A.2(i) and Assumption 6(i). Then $(ia1)$ follows by conditional Markov inequality. Now, note that $\sum_{t=1}^{T} E_{\mathcal{D}}\left[\xi_{NT,t}^2\right] = v_x^{K\prime}\{(NT)^{-1} \sum_{i=1}^{N}\sum_{t=1}^{T} E_{\mathcal{D}}[\tilde{Z}_{it}\tilde{Z}_{it}' e_{it}^2]\} v_x^K = 1$. By some straightforward moment calculations, we can show that $E_{\mathcal{D}}[(\sum_{t=1}^{T}\xi_{NT,t}^2 - 1)^2] = o_P(1)$. Thus $(ia2)$ follows.

Proof of (ib). Noting that $E_{\mathcal{D}}(p_{js}e_{it}) = 0$ for $s \leq t$, we have

$$\begin{aligned}
\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T} v_x^{K\prime} E_{\mathcal{D}}\left(Z_{it}e_{it}\right) &= -\frac{\kappa_{NT}}{NT}\sum_{i=1}^{N}\sum_{1\leq t<s\leq T}\eta_{ts}E_{\mathcal{D}}\left[v_x^{K\prime}p_{is}e_{it}\right] \\
&\quad +\frac{\kappa_{NT}}{N^2T^2}\sum_{i=1}^{N}\sum_{1\leq t<s\leq T}\eta_{ts}\alpha_{ii}E_{\mathcal{D}}\left[v_x^{K\prime}p_{is}e_{it}\right] \\
&= -\kappa_{NT}b_1(x) + O_P\left(K^{1/2}/(NT)^{1/2}\right) = -\kappa_{NT}b_1(x) + o_P(1),
\end{aligned}$$

where the term $O_P\left(K^{1/2}/(NT)^{1/2}\right)$ is obtained by similar arguments as used in the proof of Lemma A.4. So $\Pi_{1NT,11b} = o_P(1)$. ■

Proposition A.7 *Let the conditions in Theorem 3.2 hold. Then* $\Pi_{2NT} = -\kappa_{NT}^{-1}b_2(x) - \kappa_{NT}b_3(x) + o_P(1)$.

Proof. Let $\vec{v}_x^K \equiv V_K^{-1/2}(x) W_{NT}^{-1} p^K(x)$ and $\vec{v}_{x,k}^K$ be its kth element. Let $\tilde{\Pi}_{2NT} \equiv \sqrt{NT} v_x^{K\prime} C_{NT}^{(2)}$. Then we have $\Pi_{2NT} = \sqrt{NT} v_x^{K\prime} C_{NT}^{(2)} + \sqrt{NT}\left[v_x^K - \vec{v}_x^K\right]' C_{NT}^{(2)} = \tilde{\Pi}_{2NT} + o_P(1)$ where the $o_P(1)$ term comes from the fact that

$$\begin{aligned}
\sqrt{NT}\left|\left[v_x^K - \vec{v}_x^K\right]' C_{NT}^{(2)}\right| &= \left|\sqrt{NT} V_K^{-1/2}(x) p^K(x)' \tilde{W}^{-1}\left(W_{NT} - \tilde{W}\right) W_{NT}^{-1} C_{NT}^{(2)}\right| \\
&\leq \sqrt{NT}\left\|W_{NT}^{-1}\right\|\left\|C_{NT}^{(2)}\right\| V_K^{-1/2}(x)\left\|p^K(x)\tilde{W}^{-1}\right\|\left\|W_{NT} - \tilde{W}\right\|_F \\
&= \sqrt{NT} O_P\left(K^{1/2-2\gamma/d} + K^{1/2}\delta_{NT}^{-2}\right) O(1)\, O_P\left(K/\sqrt{NT}\right) = o_P(1)
\end{aligned}$$

by (A.3), Lemma A.3, and Assumption 5. Let $a = v_x^K/\left\|v_x^K\right\|$ and $\mathbf{P}_{(a)} = \sum_{k=1}^{K} a_k \mathbf{P}_k$. We decompose $\tilde{\Pi}_{2NT}$ as follows

$$\begin{aligned}
\tilde{\Pi}_{2NT} &= -\frac{\left\|v_x^K\right\|}{\sqrt{NT}}\left\{\operatorname{tr}\left[\mathbf{u} M_{f^0}\mathbf{u}' M_{\lambda^0}\mathbf{P}_{(a)}\Phi\right] + \operatorname{tr}\left[\mathbf{u}' M_{\lambda^0}\mathbf{u} M_{f^0}\mathbf{P}_{(a)}'\Phi'\right] + \operatorname{tr}\left[\mathbf{u}' M_{\lambda^0}\mathbf{P}_{(a)} M_{f^0}\mathbf{u}'\Phi'\right]\right\} \\
&\equiv \Pi_{2NT,1} + \Pi_{2NT,2} + \Pi_{2NT,3}, \text{ say.}
\end{aligned}$$

We complete the proof by showing that (i) $\Pi_{2NT,1} = -\kappa_{NT}^{-1}b_2(x) + o_P(1)$, (ii) $\Pi_{2NT,2} = -\kappa_{NT}b_3(x) + o_P(1)$, and (iii) $\Pi_{2NT,3} = o_P(1)$.

First, we consider (i). Observing that $\Pi_{2NT,1}=\frac{\|v_x^K\|}{\sqrt{NT}}\left\{\operatorname{tr}\left[\mathbf{u}P_{f^0}\mathbf{u}'M_{\lambda^0}\mathbf{P}_{(a)}\Phi\right]-\operatorname{tr}\left[\mathbf{u}\mathbf{u}'M_{\lambda^0}\mathbf{P}_{(a)}\Phi\right]\right\}\equiv\Pi_{2NT,11}+\Pi_{2NT,12}$, we prove (i) by showing that: (ia) $\Pi_{2NT,11}=o_P(1)$ and (ib) $\Pi_{2NT,12}=-\kappa_{NT}^{-1}b_2(x)+o_P(1)$. We first consider (ia). Using $M_{\lambda^0}=I_N-\mathbf{P}_{\lambda^0}$ and $\mathbf{u}=\mathbf{e}+\mathbf{e}_g$, we have

$$\begin{aligned}|\Pi_{2NT,11}| &\le (NT)^{-1/2}\|v_x^K\|\left|\operatorname{tr}\left[\mathbf{e}P_{f^0}\mathbf{e}'M_{\lambda^0}\mathbf{P}_{(a)}\Phi\right]\right|+(NT)^{-1/2}\|v_x^K\|\left|\operatorname{tr}\left[\mathbf{e}_gP_{f^0}\mathbf{e}_g'M_{\lambda^0}\mathbf{P}_{(a)}\Phi\right]\right| \\ &\quad+(NT)^{-1/2}\|v_x^K\|\left|\operatorname{tr}\left[\mathbf{e}P_{f^0}\mathbf{e}_g'M_{\lambda^0}\mathbf{P}_{(a)}\Phi\right]\right|+(NT)^{-1/2}\|v_x^K\|\left|\operatorname{tr}\left[\mathbf{e}_gP_{f^0}\mathbf{e}'M_{\lambda^0}\mathbf{P}_{(a)}\Phi\right]\right| \\ &\equiv \Pi_{2NT,11a}+\Pi_{2NT,11b}+\Pi_{2NT,11c}+\Pi_{2NT,11d},\text{ say.}\end{aligned}$$

By Lemmas E.3(i) and (v) in the supplemental appendix and Lemmas A.5(i)-(ii),

$$\begin{aligned}|\Pi_{2NT,11a}| &= (NT)^{-1/2}\|v_x^K\|\left|\operatorname{tr}\left[\mathbf{e}P_{f^0}\mathbf{e}'\mathbf{P}_{(a)}\Phi\right]\right|+(NT)^{-1/2}\|v_x^K\|\left|\operatorname{tr}\left[P_{f^0}\mathbf{e}'P_{\lambda^0}\mathbf{P}_{(a)}\Phi\mathbf{e}\right]\right| \\ &\le C(NT)^{-1/2}\left[\left|\operatorname{tr}\left[\left(f^{0\prime}f^0\right)^{-1}f^{0\prime}\mathbf{e}'\mathbf{P}_{(a)}\Phi\mathbf{e}f^0\right]\right|+C(NT)^{-1/2}\left|\operatorname{tr}\left[P_{f^0}\mathbf{e}'P_{\lambda^0}\mathbf{P}_{(a)}\Phi\mathbf{e}\right]\right|\right] \\ &\le C(NT)^{-1/2}R\left\|\left(f^{0\prime}f^0\right)^{-2}\right\|\left\|\left(\lambda^{0\prime}\lambda^0\right)^{-1}\right\|\|f^0\|\|\lambda^{0\prime}\mathbf{e}f^0\|\|f^{0\prime}\mathbf{e}'\mathbf{P}_{(a)}\| \\ &\quad+C(NT)^{-1/2}R\|\Phi\|\|\mathbf{e}\|\|\mathbf{P}_{(a)}\|\|P_{f^0}\mathbf{e}'P_{\lambda^0}\| \\ &= CR(NT)^{-1/2}O_P\left(T^{-2}\right)O_P\left(N^{-1}\right)O_P\left(T^{1/2}\right)O_P\left(\sqrt{NT}\right)O_P\left(\sqrt{NTK}\delta_{NT}\right) \\ &\quad+RC(NT)^{-1/2}O_P\left((NT)^{-1/2}\right)O_P\left(\sqrt{NT}\delta_{NT}^{-1}\right)O_P\left(\sqrt{NT}\right)O_P(1) \\ &= O_P\left(K^{1/2}T^{-1/2}\delta_{NT}^{-1}+\delta_{NT}^{-1}\right)=O_P\left(\delta_{NT}^{-1}\right)=o_P(1).\end{aligned}$$

By Lemmas E.3(i) and (v), $\Pi_{2NT,11b}\le C(NT)^{-1/2}R\|\mathbf{e}_g\|^2\|\mathbf{P}_{(a)}\|\|\Phi\|=O_P(\sqrt{NT}K^{-2\gamma/d})=o_P(1)$. By Lemma A.5$(i)$, $\Pi_{2NT,11c}\le C(NT)^{-1/2}\left|\operatorname{tr}[\lambda^{0\prime}\mathbf{e}f^0(f^{0\prime}f^0)^{-1}f^0\mathbf{e}_g'M_{\lambda^0}\mathbf{P}_{(a)}f^0(f^{0\prime}f^0)^{-1}(\lambda^{0\prime}\lambda^0)^{-1}]\right|\le C(NT)^{-1/2}R\|\lambda^{0\prime}\mathbf{e}f^0\|\|(f^{0\prime}f^0)^{-1}\|^2\ \|(\lambda^{0\prime}\lambda^0)^{-1}\|\|f^0\|^2\|\mathbf{e}_g\|\|\mathbf{P}_{(a)}\|=O_P\left(K^{-\gamma/d}\right)$. For $\Pi_{2NT,11d}$, by Lemmas A.5(ii) and (vi) we have

$$\begin{aligned}\Pi_{2NT,11d} &\le C(NT)^{-1/2}\left|\operatorname{tr}\left[\mathbf{e}_gP_{f^0}\mathbf{e}'M_{\lambda^0}\mathbf{P}_{(a)}\Phi\right]\right| \\ &\le C(NT)^{-1/2}\left\{\left|\operatorname{tr}\left[\mathbf{e}_gP_{f^0}\mathbf{e}'\mathbf{P}_{(a)}\Phi\right]\right|+\left|\operatorname{tr}\left[\mathbf{e}_gP_{f^0}\mathbf{e}'P_{\lambda^0}\mathbf{P}_{(a)}\Phi\right]\right|\right\} \\ &\le R(NT)^{-1/2}\|\mathbf{e}_g\|\|\Phi\|\left[\|P_{f^0}\mathbf{e}'\mathbf{P}_{(a)}\|+\|P_{f^0}\mathbf{e}'P_{\lambda^0}\|\|\mathbf{P}_{(a)}\|\right] \\ &= (NT)^{-1/2}O_P\left(K^{-\gamma/d}\right)\left[O_P\left(\sqrt{NTK}\right)+O_P\left(\sqrt{NT}\right)\right]=O_P\left(K^{1/2-\gamma/d}\right)=o_P(1).\end{aligned}$$

It follows that $\Pi_{2NT,11}=o_P(1)$.

Now we consider (ib). Noting that $\mathbf{u}=\mathbf{e}+\mathbf{e}_g$, we rewrite $\Pi_{2NT,12}$ as follows

$$\begin{aligned}\Pi_{2NT,12} &= -(NT)^{-1/2}\|v_x^K\|\operatorname{tr}\left[\mathbf{e}\mathbf{e}'M_{\lambda^0}\mathbf{P}_{(a)}\Phi\right]-(NT)^{-1/2}\|v_x^K\|\operatorname{tr}\left[\mathbf{e}_g\mathbf{e}_g'M_{\lambda^0}\mathbf{P}_{(a)}\Phi\right] \\ &\quad-(NT)^{-1/2}\|v_x^K\|\operatorname{tr}\left[\mathbf{e}\mathbf{e}_g'M_{\lambda^0}\mathbf{P}_{(a)}\Phi\right]-(NT)^{-1/2}\|v_x^K\|\operatorname{tr}\left[\mathbf{e}_g\mathbf{e}'M_{\lambda^0}\mathbf{P}_{(a)}\Phi\right] \\ &\equiv \Pi_{2NT,12a}+\Pi_{2NT,12b}+\Pi_{2NT,12c}+\Pi_{2NT,12d},\text{ say.}\end{aligned}$$

First, $\Pi_{2NT,12a}=-(NT)^{-1/2}\|v_x^K\|\operatorname{tr}\left[E_{\mathcal{D}}(\mathbf{e}\mathbf{e}')M_{\lambda^0}\mathbf{P}_{(a)}\Phi\right]-(NT)^{-1/2}\|v_x^K\|\operatorname{tr}\left\{\left[\mathbf{e}\mathbf{e}'-E_{\mathcal{D}}(\mathbf{e}\mathbf{e}')\right]M_{\lambda^0}\mathbf{P}_{(a)}\Phi\right\}$ $\equiv\Pi_{2NT,12aa}+\Pi_{2NT,12ab}$, say. Clearly, $\Pi_{2NT,12aa}=-\kappa_{NT}^{-1}b_2(x)$ and $\left|\kappa_{NT}^{-1}b_2(x)\right|\le R\kappa_{NT}^{-1}\|v_x^K\|\|E_{\mathcal{D}}(\mathbf{e}\mathbf{e}'/T)\|$ $\|M_{\lambda^0}\|\|\mathbf{P}_{(a)}\|\ ||f^0\left(f^{0\prime}f^0\right)^{-1}\ \left(\lambda^{0\prime}\lambda^0\right)^{-1}\lambda^0||=O_P\left(\kappa_{NT}^{-1}\right)$ by Lemmas E.3(i) and (v). Recall $P_i^\lambda=N^{-1}\sum_{l=1}^N\alpha_{li}P_l$, $G^0=(f^{0\prime}f^0/T)^{-1}(\lambda^{0\prime}\lambda^0/N)^{-1}$ and $A_i=v_x^{K\prime}\left[E_{\mathcal{D}}(P_i')-E_{\mathcal{D}}(P_i^{\lambda\prime})\right]f^0G^0T^{-1}$. Then $\Pi_{2NT,12ab}$ can be decomposed as follows

$$\begin{aligned}\Pi_{2NT,12ab} &= \frac{1}{N^{1/2}}\frac{1}{N\sqrt{T}}\sum_{t=1}^T\sum_{i=1}^N\sum_{j=1}^N A_i\lambda_j^0\left[e_{jt}e_{it}-E_{\mathcal{D}}(e_{jt}e_{is})\right] \\ &\quad+\frac{1}{N^{3/2}T^{3/2}}\sum_{i=1}^N\left\{\left[\sum_{s=1}^T v_x^{K\prime}\left(p_{is}^c-p_{is}^{\lambda c}\right)f_s^{0\prime}G^0\right]\left[\sum_{j=1}^N\sum_{t=1}^T\lambda_j^0\left[e_{it}e_{jt}-E_{\mathcal{D}}(e_{it}e_{jt})\right]\right]\right\}.\end{aligned}$$

The first term is $O_P\left(K^{1/2}/N^{1/2}\right)$ by Lemma A.5(vii), and the second term in the above expression is bounded by

$$\frac{1}{\sqrt{T}}\left\{\frac{1}{NT}\sum_{i=1}^{N}\left\|\sum_{s=1}^{T}v_x^{K\prime}\left(p_{is}^c-p_{is}^{\lambda c}\right)f_s^{0\prime}G^0\right\|^2\right\}^{1/2}\left\{\frac{1}{N^2T}\sum_{i=1}^{N}\left\|\sum_{j=1}^{N}\sum_{t=1}^{T}\lambda_j^0\left[e_{it}e_{jt}-E_{\mathcal{D}}\left(e_{it}e_{jt}\right)\right]\right\|^2\right\}^{1/2}$$

which is of order $T^{-1/2}O_P\left(K^{1/2}\right)O_P\left(1\right)=O_P\left(K^{1/2}T^{-1/2}\right)$ by Lemmas A.5($viii$) and (ix). So $\Pi_{2NT,12ab}=O_P\left(K^{1/2}(N^{-1/2}+T^{-1/2})\right)$. For $\Pi_{2NT,12b}$, we have $|\Pi_{2NT,12b}|\leq CR\left(NT\right)^{-1/2}\left\|\mathbf{e}_g\right\|^2\left\|\mathbf{P}_{(a)}\right\|\left\|\Phi\right\|=O_P((NT)^{1/2}K^{-2\gamma/d})=o_P\left(1\right)$. For $\Pi_{2NT,12c}$ and $\Pi_{2NT,12d}$, we can show that they are both bounded from above by $CR\left(NT\right)^{-1/2}\left\|\mathbf{e}_g\right\|\left\|\mathbf{e}\right\|\,\left\|\mathbf{P}_{(a)}\right\|\left\|\Phi\right\|=O_P(\sqrt{NT}K^{-\gamma/d}\delta_{NT}^{-1})=o_P\left(1\right)$. It follows that $\Pi_{2NT,12}=-\kappa_{NT}^{-1}b_2\left(x\right)+o_P\left(1\right)$.

Now we consider (ii)**.** Noting that $M_{\lambda^0}=I_N-P_{\lambda^0}$, we have $\Pi_{2NT,2}=-\left(NT\right)^{-1/2}\left\|v_x^K\right\|\{\mathrm{tr}[\mathbf{u}'\mathbf{u}M_{f^0}\mathbf{P}_{(a)}'\Phi'] -\mathrm{tr}[\mathbf{u}'P_{\lambda^0}\mathbf{u}M_{f^0}\mathbf{P}_{(a)}'\Phi']\}\equiv\Pi_{2NT,21}+\Pi_{2NT,22}$, say. Noting that $\mathbf{u}=\mathbf{e}+\mathbf{e}_g$ and $\left\|v_x^K\right\|=O_P\left(1\right)$, we have

$$\begin{aligned}\Pi_{2NT,22}\quad&\leq\quad\left(NT\right)^{-1/2}\left\|v_x^K\right\|\left|\mathrm{tr}\left[\mathbf{e}'P_{\lambda^0}\mathbf{e}M_{f^0}\mathbf{P}_{(a)}'\Phi'\right]\right|+\left(NT\right)^{-1/2}\left\|v_x^K\right\|\left|\mathrm{tr}\left[\mathbf{e}'P_{\lambda^0}\mathbf{e}_gM_{f^0}\mathbf{P}_{(a)}'\Phi'\right]\right|\\&\quad+\left(NT\right)^{-1/2}\left\|v_x^K\right\|\left|\mathrm{tr}\left[\mathbf{e}_g'P_{\lambda^0}\mathbf{e}M_{f^0}\mathbf{P}_{(a)}'\Phi'\right]\right|+\left(NT\right)^{-1/2}\left\|v_x^K\right\|\left|\mathrm{tr}\left[\mathbf{e}_g'P_{\lambda^0}\mathbf{e}_gM_{f^0}\mathbf{P}_{(a)}'\Phi'\right]\right|\\&\equiv\quad\Pi_{2NT,22a}+\Pi_{2NT,22b}+\Pi_{2NT,22c}+\Pi_{2NT,22d},\ \text{say.}\end{aligned}$$

For $\Pi_{2NT,22a}$, by Lemmas A.5(i) and (v) we have

$$\begin{aligned}\Pi_{2NT,22a}\quad&\leq\quad\left(NT\right)^{-1/2}\left\|v_x^K\right\|\left|\mathrm{tr}\left[\mathbf{e}'P_{\lambda^0}\mathbf{e}\mathbf{P}_{(a)}'\Phi'\right]\right|+\left(NT\right)^{-1/2}\left\|v_x^K\right\|\left|\mathrm{tr}\left[\mathbf{e}'P_{\lambda^0}\mathbf{e}P_{f^0}\mathbf{P}_{(a)}'\Phi'\right]\right|\\&\leq\quad\left(NT\right)^{-1/2}\left\|v_x^K\right\|R\left\|\lambda^{0\prime}\mathbf{e}\mathbf{P}_{(a)}'\right\|\left\|\lambda^0\right\|\left\|\left(\lambda^{0\prime}\lambda^0\right)^{-2}\right\|\left\|\left(f^{0\prime}f^0\right)^{-1}\right\|\left\|f^{0\prime}\mathbf{e}'\lambda^0\right\|\\&\quad+\left(NT\right)^{-1/2}\left\|v_x^K\right\|R\left|\left\|\lambda^{0\prime}\mathbf{e}P_{f^0}\right\|\left\|\mathbf{P}_{(a)}'\right\|\left\|\lambda^0\right\|\left\|\left(\lambda^{0\prime}\lambda^0\right)^{-2}\right\|\left\|\left(f^{0\prime}f^0\right)^{-1}\right\|\left\|f^{0\prime}\mathbf{e}'\lambda^0\right\|\right|\\&=\quad C\left(NT\right)^{-1/2}O_P(N\sqrt{TK})O_P\left(\sqrt{N}\right)O_P\left(N^{-2}\right)O_P\left(T^{-1}\right)O_P\left(\sqrt{NT}\right)\\&\quad+\left(NT\right)^{-1/2}O_P\left(\sqrt{N}\right)O_P\left(\sqrt{NT}\right)O_P\left(\sqrt{N}\right)O_P\left(N^{-2}\right)O_P\left(T^{-1}\right)O_P\left(\sqrt{NT}\right)=o_P\left(1\right).\end{aligned}$$

As in the study of $\Pi_{1NT,12}$, we can show that $\Pi_{2NT,22s}=o_P\left(1\right)$ for $s=b,c,d.$ Thus $\Pi_{2NT,22}=o_P\left(1\right)$.

For $\Pi_{2NT,21}$, we have $\Pi_{2NT,21}=-\sqrt{N/T}\left\|v_x^K\right\|\mathrm{tr}[E_{\mathcal{D}}\left(\mathbf{e}'\mathbf{e}/N\right)M_{f^0}\mathbf{P}_{(a)}'\Phi']-\left(NT\right)^{-1/2}\mathrm{tr}\{[\mathbf{e}'\mathbf{e}-\mathbf{E}_{\mathcal{D}}\left(\mathbf{e}'\mathbf{e}\right)] M_{f^0}\mathbf{P}_{(a)}'\Phi'\}\equiv-\kappa_{NT}b_3\left(x\right)-\Pi_{2NT,21a}$, say. It is easy to show that $|\kappa_{NT}b_3\left(x\right)|=O_P\left(\kappa_{NT}\right)$ by Lemmas E.3(i) and (v). For $\Pi_{2NT,21a}$, by Lemmas A.5(x) and (xi), we have

$$\begin{aligned}\Pi_{2NT,21a}\quad&=\quad\left(NT\right)^{-1/2}\left\|v_x^K\right\|\mathrm{tr}\left\{\left[\mathbf{e}'\mathbf{e}-\mathbf{E}_{\mathcal{D}}\left(\mathbf{e}'\mathbf{e}\right)\right]M_{f^0}\mathbf{P}_{(a)}'\Phi'\right\}\\&=\quad\frac{1}{\sqrt{T}}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}B_tf_s^{0\prime}\left[e_{it}e_{is}-E_{\mathcal{D}}\left(e_{it}e_{is}\right)\right]\\&\quad+\frac{1}{\sqrt{N}}\frac{1}{NT^{3/2}}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}v_x^{K\prime}\left[p_{jt}^c-p_{jt}^{fc}\right]\lambda_j^{0\prime}G^0f_s^{0\prime}\left[e_{it}e_{is}-E_{\mathcal{D}}\left(e_{it}e_{is}\right)\right],\end{aligned}$$

where $B_t=v_x^{K\prime}[E_{\mathcal{D}}(P_{\cdot t}-P_{\cdot t}^f)'\lambda^0G^0N^{-1}$, $P_{\cdot t}^f=T^{-1}\sum_{l=1}^T\eta_{lt}P_{\cdot l}$, and $p_{jt}^{fc}=p_{jt}^f-E_{\mathcal{D}}(p_{jt}^f)$. By Lemma A.5($x$), the first term is $O_P\left(K^{1/2}/T^{1/2}\right)$. The second term is bounded by

$$\frac{1}{\sqrt{N}}\left\{\frac{1}{NT}\sum_{t=1}^{T}\left\|\sum_{j=1}^{N}v_x^{K\prime}\left[p_{jt}^c-p_{jt}^{fc}\right]\lambda_j^{0\prime}G^0\right\|^2\right\}^{1/2}\left\{\frac{1}{NT^2}\sum_{t=1}^{T}\left\|\sum_{i=1}^{N}\sum_{s=1}^{T}f_s^{0\prime}\left[e_{it}e_{is}-E_{\mathcal{D}}\left(e_{it}e_{is}\right)\right]\right\|^2\right\}^{1/2},$$

which is $O_P\left(K^{1/2}/N^{1/2}\right)$ by Lemmas A.5(xi)-(xii).

Last, we consider (iii). For the first term, using $\Phi' P_{f^0} = \Phi'$ and $M_{\lambda^0} = I_N - P_{\lambda^0}$, we have

$$\begin{aligned}
\Pi_{2NT,3} &= -(NT)^{-1/2}\left\|v_x^K\right\| \operatorname{tr}\left[\mathbf{u}' M_{\lambda^0}\mathbf{P}_{(a)} M_{f^0}\mathbf{u}'\Phi'\right] \\
&= (NT)^{-1/2}\left\|v_x^K\right\| \left\{\operatorname{tr}\left[P_{f^0}\mathbf{u}' P_{\lambda^0}\mathbf{P}_{(a)} M_{f^0}\mathbf{u}'\Phi'\right] - \operatorname{tr}\left[P_{f^0}\mathbf{u}'\mathbf{P}_{(a)} M_{f^0}\mathbf{u}'\Phi'\right]\right\} \\
&\equiv \Pi_{2NT,31} + \Pi_{2NT,32}, \text{ say.}
\end{aligned}$$

By Lemma A.5(ii), we have

$$\begin{aligned}
\left|\Pi_{2NT,31}\right| &\leq (NT)^{-1/2}\left\|v_x^K\right\| \left|\operatorname{tr}\left[P_{f^0}\mathbf{u}' P_{\lambda^0}\mathbf{P}_{(a)} M_{f^0}\mathbf{u}'\Phi'\right]\right| \\
&\leq R(NT)^{-1/2}\left\|P_{f^0}\mathbf{u}' P_{\lambda^0}\right\| \left\|\mathbf{P}_{(a)}\right\| \left\|M_{f^0}\right\| \|\mathbf{u}\| \|\Phi\| \\
&= (NT)^{-1/2} O_P\left(1 + \sqrt{NT}K^{-\gamma/d}\right) O_P\left(\sqrt{NT}\right) O_P\left(\delta_{NT}^{-1} + K^{-\gamma/d}\right) \\
&= O_P\left[\left(1 + \sqrt{NT}K^{-\gamma/d}\right)\left(\delta_{NT}^{-1} + K^{-\gamma/d}\right)\right] = o_P(1).
\end{aligned}$$

By Lemma A.5(iv), we have

$$\begin{aligned}
\left|\Pi_{2NT,32}\right| &\leq CR(NT)^{-1/2}\left(\left\|P_{f^0}\mathbf{e}'\mathbf{P}_{(a)}\right\| + \left\|P_{f^0}\right\| \|\mathbf{e}_g\| \left\|\mathbf{P}_{(a)}\right\|\right) \|\mathbf{u}\| \|\Phi\| \\
&= CR(NT)^{-1/2} O_P\left(\sqrt{NK}\delta_{NT} + NTK^{-\gamma/d}\right) O_P\left(\delta_{NT}^{-1} + K^{-\gamma/d}\right) = o_P(1).
\end{aligned}$$

This completes the proof of the proposition. ■

A.3 Bias-corrected estimator

Lemma A.8 *Suppose that the assumptions in Theorem 3.3 hold. Then*

(i) $\left\|\hat{W}_{NT} - W_{NT}\right\|_F = O_P\left[K(K^{-\gamma/d} + \delta_{NT}^{-1})\right]$;

(ii) $||\hat{\Omega}_{NT} - \tilde{\Omega}||_F = O_P\left[K\delta_{NT}^{-1} + (NT)^{1/4} K(\delta_{NT}^{-2} + K^{-\gamma/d})\right]$;

(iii) $\left\|\hat{W}_{NT}^{-1}\hat{\Omega}_{NT}\hat{W}_{NT}^{-1} - \tilde{W}^{-1}\tilde{\Omega}\tilde{W}^{-1}\right\|_F = O_P\left[K\delta_{NT}^{-1} + (NT)^{1/4} K(\delta_{NT}^{-2} + K^{-\gamma/d})\right]$.

Lemma A.9 *Suppose that the assumptions in Theorem 3.3 hold. Then*

(i) $||\hat{b}_1 - b_1|| = O_P(\sqrt{K}\sum_{\tau=M_T}^{T} \alpha_{\mathcal{D}}^{(3+2\delta)/(4+2\delta)}(\tau) + M_T\sqrt{K}\delta_{NT}^{-1})$;

(ii) $||\hat{b}_2 - b_2|| = O_P\{\sqrt{K}[N^{-1/4} + N^{5/8}(K^{-\gamma/d} + \sqrt{K}\delta_{NT}^{-2}) + T^{-1}N^{1/2}]\}$;

(iii) $||\hat{b}_3 - b_3|| = O_P\{\sqrt{K}[T^{-1/4} + T^{5/8}(K^{-\gamma/d} + \sqrt{K}\delta_{NT}^{-2}) + N^{-1}T^{1/2}]\}$.

Proof of Theorem 3.3. We first make the following decomposition:

$$\begin{aligned}
\hat{A}_{NT}\left[\hat{g}_{bc}(x) - g(x)\right] &= \left\{A_{NT}\left[\hat{g}(x) - g(x)\right] - B_K(x)\right\} - \left[\hat{B}_K(x) - B_K(x)\right] \\
&\quad + \left(\hat{A}_{NT}/A_{NT} - 1\right)\left\{A_{NT}\left[\hat{g}(x) - g(x)\right] - B_K(x)\right\} + \left(\hat{A}_{NT}/A_{NT} - 1\right) B_K(x) \\
&\equiv DB_1 - DB_2 + DB_3 + DB_4, \text{ say.}
\end{aligned}$$

Noting that $DB_1 \xrightarrow{d} N(0,1)$ by Theorem 3.2, it suffices to show that (i) $DB_2 = o_P(1)$; (ii) $DB_3 = o_P(1)$; and $DB_4 = o_P(1)$.

Proof of (i)**.** Using $\hat{B}_K(x) = -\kappa_{NT}\hat{b}_1(x) - \kappa_{NT}^{-1}\hat{b}_2(x) - \kappa_{NT}\hat{b}_3(x)$ where $\hat{b}_s(x) = \hat{V}_K^{-1/2}(x) p^K(x)' \times \hat{W}_{NT}^{-1}\hat{b}_s$, we have $DB_2 = \kappa_{NT}[\hat{b}_1(x) - b_1(x)] + \kappa_{NT}^{-1}[\hat{b}_2(x) - b_2(x)] + \kappa_{NT}[\hat{b}_3(x) - b_3(x)] \equiv DB_{21} + DB_{22} + DB_{23}$. We prove that $DB_2 = o_P(1)$ by showing that

$$\begin{aligned}
(i1)\ DB_{21} &= \kappa_{NT}\left(\hat{V}_K^{-1/2}(x) p^K(x)' \hat{W}_{NT}^{-1}\hat{b}_1 - V_K^{1/2}(x) p^K(x)' \tilde{W}^{-1} b_1\right) = o_P(1), \\
(i2)\ DB_{22} &= \kappa_{NT}^{-1}\left(\hat{V}_K^{-1/2}(x) p^K(x)' \hat{W}_{NT}^{-1}\hat{b}_2 - V_K^{1/2}(x) p^K(x)' \tilde{W}^{-1} b_2\right) = o_P(1), \\
(i3)\ DB_{23} &= \kappa_{NT}\left(\hat{V}_K^{-1/2}(x) p^K(x)' \hat{W}_{NT}^{-1}\hat{b}_3 - V_K^{1/2}(x) p^K(x)' \tilde{W}^{-1} b_3\right) = o_P(1).
\end{aligned}$$

Note that

$$
\begin{aligned}
DB_{21} &= \kappa_{NT}\left[\hat{V}_K^{-1/2}(x)\,p^K(x)'\,\hat{W}_{NT}^{-1}\hat{b}_1 - V_K^{1/2}(x)\,p^K(x)'\,\tilde{W}^{-1}b_1\right] \\
&= \kappa_{NT}V_K^{-1/2}(x)\,p^K(x)'\,\tilde{W}^{-1}\left(\hat{b}_1-b_1\right) + \kappa_{NT}V_K^{-1/2}(x)\,p^K(x)'\left(\hat{W}_{NT}^{-1}-\tilde{W}^{-1}\right)\left(\hat{b}_1-b_1\right) \\
&\quad +\kappa_{NT}V_K^{-1/2}(x)\,p^K(x)'\left(\hat{W}_{NT}^{-1}-\tilde{W}^{-1}\right)b_1 + \kappa_{NT}\left[\hat{V}_K^{-1/2}(x)-V_K^{-1/2}(x)\right]p^K(x)'\,\hat{W}_{NT}^{-1}\hat{b}_1 \\
&\equiv DB_{21a}+DB_{21b}+DB_{21c}+DB_{21d}, \text{ say.}
\end{aligned}
$$

Recalling that $v_x^K = V_K^{-1/2}(x)\,\tilde{W}^{-1}p^K(x)$ with $\left\|v_x^K\right\| = O_P(1)$, by Lemma A.9$(i)$ and Assumption 9 we have $|DB_{21a}| \leq \kappa_{NT}\left\|v_x^K\right\|\,||\hat{b}_1-b_1|| = O_P[\kappa_{NT}\sqrt{K}(\sum_{\tau=M_T}^{\infty}\alpha_{\mathcal{D}}^{\frac{3+2\delta}{4+2\delta}}(\tau)+M_T\delta_{NT}^{-1})] = o_P(1)$. By Lemmas A.3 and A.8 and Minkowski inequality, $||\tilde{W}-\hat{W}_{NT}||_F = O_P\left[K\left(K^{-\gamma/d}+\delta_{NT}^{-1}\right)\right]$. This, in conjunction with Assumption 7, implies that $||\hat{W}_{NT}^{-1}|| = O_P(1)$. Then by Lemma A.9$(i)$ and Assumption 9, we have

$$
\begin{aligned}
|DB_{21b}| &= \left|\kappa_{NT}V_K^{-1/2}(x)\,p^K(x)'\,\tilde{W}^{-1}\left(\tilde{W}-\hat{W}_{NT}\right)\hat{W}_{NT}^{-1}\left(\hat{b}_1-b_1\right)\right| \\
&\leq \kappa_{NT}\left\|v_x^K\right\|\left\|\tilde{W}-\hat{W}_{NT}\right\|_F\left\|\hat{W}_{NT}^{-1}\right\|\left\|\hat{b}_1-b_1\right\| \\
&= \kappa_{NT}O_P(1)\,O_P\left[K\left(K^{-\gamma/d}+\delta_{NT}^{-1}\right)\right]O_P(1)\,O_P\left[\sqrt{K}\left(\sum_{\tau=M_T}^{\infty}\alpha_{\mathcal{D}}^{\frac{3+2\delta}{4+2\delta}}(\tau)+M_T\delta_{NT}^{-1}\right)\right] = o_P(1).
\end{aligned}
$$

Similarly, $DB_{21c} \leq \kappa_{NT}\left\|v_x^K\right\|\,||\tilde{W}-\hat{W}_{NT}||_F||\hat{W}_{NT}^{-1}||\,\|b_1\| = \kappa_{NT}O_P(1)\,O_P\left[K\left(K^{-\gamma/d}+\delta_{NT}^{-1}\right)\right]O_P(\sqrt{K})$ $= o_P(1)$. Now, we decompose DB_{21d} as follows

$$
\begin{aligned}
DB_{21d} &= \kappa_{NT}\left[V_K^{1/2}(x)/\hat{V}_K^{1/2}(x)-1\right]V_K^{-1/2}(x)\,p^K(x)'\,\hat{W}_{NT}^{-1}\hat{b}_1 \\
&= \kappa_{NT}\left[V_K^{1/2}(x)/\hat{V}_K^{1/2}(x)-1\right]V_K^{-1/2}(x)\,p^K(x)'\,\tilde{W}^{-1}\hat{b}_1 \\
&\quad +\kappa_{NT}\left[V_K^{1/2}(x)/\hat{V}_K^{1/2}(x)-1\right]V_K^{-1/2}(x)\,p^K(x)'\left(\hat{W}_{NT}^{-1}-\tilde{W}^{-1}\right)\hat{b}_1 \\
&\equiv DB_{21d,1}+DB_{21d,2}.
\end{aligned}
$$

By Lemma A.8(iii),

$$
\begin{aligned}
\left|\hat{V}_K(x)-V_K(x)\right| &= \left|p^K(x)'\left[\hat{W}_{NT}^{-1}\hat{\Omega}_{NT}\hat{W}_{NT}^{-1}-\tilde{W}^{-1}\tilde{\Omega}\tilde{W}^{-1}\right]p^K(x)\right| \\
&\leq \left\|p^K(x)\right\|^2 O_P\left[K\delta_{NT}^{-1}+(NT)^{1/4}K\left(\delta_{NT}^{-2}+K^{-\gamma/d}\right)\right].
\end{aligned}
$$

This, in conjunction with the fact that $V_K(x) \geq \left\|p^K(x)\right\|^2\mu_{\min}(\tilde{W}^{-1}\tilde{\Omega}\tilde{W}^{-1}) \geq C\left\|p^K(x)\right\|^2$, implies that

$$
\left|V_K^{1/2}(x)/\hat{V}_K^{1/2}(x)-1\right| = \left|\frac{\hat{V}_K(x)-V_K(x)}{\hat{V}_K^{1/2}(x)\left[\hat{V}_K^{1/2}(x)+V_K^{1/2}(x)\right]}\right| = O_P\left[K\delta_{NT}^{-1}+(NT)^{1/4}K\left(\delta_{NT}^{-2}+K^{-\gamma/d}\right)\right]. \tag{A.10}
$$

Consequently, $|DB_{21d,1}| \leq \kappa_{NT}\left|V_K^{1/2}(x)/\hat{V}_K^{1/2}(x)-1\right|\left\|v_x^K\right\|\left\|\hat{b}_1\right\| = \kappa_{NT}O_P[K\delta_{NT}^{-1}+(NT)^{1/4}K(\delta_{NT}^{-2}$ $+K^{-\gamma/d})]O_P\left(K^{1/2}\right) = o_P(1)$. Similarly, we can show that $|DB_{21d,2}| = o_P(1)$. Then $(i1)$ follows. Analogously, we can show $(i2)$ and $(i3)$ by Lemmas A.8 and A.9.

Proof of (ii). By (A.10), $\left|\hat{A}_{NT}/A_{NT}-1\right| = \left|\frac{\hat{V}_K(x)-V_K(x)}{\hat{V}_K^{1/2}(x)\left[\hat{V}_K^{1/2}(x)+V_K^{1/2}(x)\right]}\right| = O_P[K\delta_{NT}^{-1}+(NT)^{1/4}K(\delta_{NT}^{-2}$ $+K^{-\gamma/d})] = o_P(1)$. It follows that $|DB_3| \leq |\hat{A}_{NT}/A_{NT}-1|\,|A_{NT}[\hat{g}(x)-g(x)]-B_K(x)| = o_P(1)\,O_P(1)$ $= o_P(1)$.

Proof of (iii). Noting that $|B_K(x)| \leq |\kappa_{NT}b_1(x)| + \left|\kappa_{NT}^{-1}b_2(x)\right| + |\kappa_{NT}b_3(x)| = O_P\left(\kappa_{NT}K^{1/2}\right) +$ $O_P\left(\kappa_{NT}^{-1}\right) + O_P(\kappa_{NT})$, we have $|DB_4| \leq \left|\hat{A}_{NT}/A_{NT}-1\right||B_K(x)| = O_P[K\delta_{NT}^{-1}+(NT)^{1/4}K\left(\delta_{NT}^{-2}+K^{-\gamma/d}\right)]$ $\times\left[O_P\left(\kappa_{NT}K^{1/2}\right)+O_P\left(\kappa_{NT}^{-1}\right)+O_P(\kappa_{NT})\right] = o_P(1)$. ■

B Proofs of main results for specification test

Let $\psi_{it} \equiv \frac{1}{N}\sum_{j=1}^{N}\alpha_{ij}X_{jt} + \frac{1}{T}\sum_{s=1}^{T}\eta_{ts}X_{is} - \frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\alpha_{ij}\eta_{ts}X_{js}$ and $\tilde{X}_{it} \equiv X_{it} - E_{\mathcal{D}}(\psi_{it})$. Let $\tilde{\Omega}_{\tilde{x}\tilde{x},NT} \equiv \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}' e_{it}^2$, $\tilde{\Omega}_{\tilde{x}\tilde{z},NT} \equiv \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{Z}_{it}' e_{it}^2$, $\tilde{\Omega}_{\tilde{x}\tilde{x}} \equiv E_{\mathcal{D}}[\tilde{\Omega}_{xx,NT}]$, $\tilde{\Omega}_{\tilde{x}\tilde{z}} \equiv E_{\mathcal{D}}[\tilde{\Omega}_{\tilde{x}\tilde{z},NT}]$, $\mathcal{H}_{px} \equiv \tilde{W}^{-1}Q_{wpx}D^{-1}$, and $h_{it,js} \equiv \tilde{Z}_{it}'\mathcal{H}_{px}\tilde{X}_{js}$. Let $b_1^{(l)}$, $b_2^{(l)}$, $b_3^{(l)}$ denote $d\times 1$ vectors whose kth elements are respectively given by

$$b_{1,k}^{(l)} \equiv \frac{1}{N}\mathrm{tr}\left[P_{f^0}E_{\mathcal{D}}\left(\mathbf{e}'\mathbf{X}_k\right)\right], \; b_{2,k}^{(l)} \equiv \frac{1}{T}\mathrm{tr}\left[E_{\mathcal{D}}\left(\mathbf{e}\mathbf{e}'\right)M_{\lambda^0}\mathbf{X}_k\Phi\right], \text{ and } b_{3,k}^{(l)} \equiv \frac{1}{N}\mathrm{tr}\left[E_{\mathcal{D}}\left(\mathbf{e}'\mathbf{e}\right)M_{f^0}\mathbf{X}_k'\Phi'\right]. \tag{B.1}$$

The following lemmas are needed in the proofs of the main results in Section 4.

Lemma B.1 *Suppose that the assumptions in Theorem 4.1 hold. Then*
(*i*) $\|Q_{wpp,NT} - Q_{wpp}\|_F = O_P(K/(NT)^{1/2})$;
(*ii*) $\|Q_{wpx,NT} - Q_{wpx}\|_F = O_P(K^{1/2}/(NT)^{1/2})$;
(*iii*) $\|D_{NT} - D\|_F = O_P((NT)^{-1/2})$;
(*iv*) $\left\|\Omega_{xz,NT} - \tilde{\Omega}_{xz}\right\|_F = O_P(K^{1/2}/(NT)^{1/2})$.

Lemma B.2 *Suppose that the assumptions in Theorem 4.1 hold. Then* $\hat{\beta}_{bc} - \beta^0 = \tilde{W}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{Z}_{it}e_{it} + R_{\beta,NT}$, *where* $\|R_{\beta,NT}\| = o_P(\gamma_{NT})$.

Lemma B.3 *Suppose that the assumptions in Theorem 4.1 hold. Then under* $\mathbb{H}_1(\gamma_{NT})$ *we have* $\hat{\theta} - \theta^0 = \gamma_{NT}D_{NT}^{-1}\Upsilon_{NT} + D^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{it}e_{it} + B_{\theta,NT} + R_{\theta,NT}$, *where* $R_{\theta,NT} = o_P(\gamma_{NT})$ *and* $B_{\theta,NT} \equiv -T^{-1}D^{-1}b_1^{(l)} - N^{-1}D^{-1}b_2^{(l)} - T^{-1}D^{-1}b_3^{(l)}$.

Proof of Theorem 4.1. Recall that $e_{g,it} = g(X_{it}) - p_{it}'\beta^0$ and $g(X_{it}) - X_{it}'\theta^0 = \gamma_{NT}\Delta_{it}$ under $\mathbb{H}_1(\gamma_{NT})$. We can decompose Γ_{NT} as follows $\Gamma_{NT} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[p_{it}'\hat{\beta}_{bc} - X_{it}'\hat{\theta}]^2 w_{it} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[p_{it}'(\hat{\beta}_{bc} - \beta^0) - e_{g,it} + \gamma_{NT}\Delta(X_{it}) - X_{it}'(\hat{\theta} - \theta^0)]^2 w_{it} = \Gamma_{NT1} + \Gamma_{NT2} + \Gamma_{NT3} + \Gamma_{NT4} - 2\Gamma_{NT5} - 2\Gamma_{NT6} + 2\Gamma_{NT7} + 2\Gamma_{NT8} - 2\Gamma_{NT9} - 2\Gamma_{NT10}$, where

$$\begin{aligned}
&\Gamma_{NT1} \equiv \left(\hat{\beta}_{bc} - \beta^0\right)' Q_{wpp,NT}\left(\hat{\beta}_{bc} - \beta^0\right), && \Gamma_{NT2} \equiv (\hat{\theta} - \theta^0)' Q_{wxx,NT}(\hat{\theta} - \theta^0),\\
&\Gamma_{NT3} \equiv \gamma_{NT}^2\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}w_{it}\Delta_{it}^2, && \Gamma_{NT4} \equiv (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}w_{it}e_{g,it}^2,\\
&\Gamma_{NT5} \equiv \left(\hat{\beta}_{bc} - \beta^0\right)' Q_{wpx,NT}(\hat{\theta} - \theta^0), && \Gamma_{NT6} \equiv \left(\hat{\beta}_{bc} - \beta^0\right)'\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}w_{it}p_{it}e_{g,it},\\
&\Gamma_{NT7} \equiv \gamma_{NT}\left(\hat{\beta}_{bc} - \beta^0\right)'\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}w_{it}p_{it}\Delta_{it}, && \Gamma_{NT8} \equiv \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}w_{it}e_{g,it}X_{it}'(\hat{\theta} - \theta^0),\\
&\Gamma_{NT9} \equiv \gamma_{NT}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}w_{it}\Delta_{it}X_{it}'(\hat{\theta} - \theta^0), && \Gamma_{NT10} \equiv \gamma_{NT}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}w_{it}e_{g,it}\Delta_{it}.
\end{aligned}$$

We complete the proof by showing that under $\mathbb{H}_1(\gamma_{NT})$, (*i*) $(NT\Gamma_{NT1} - \mathbb{B}_{NT})/\mathbb{V}_{NT}^{1/2} \xrightarrow{d} N(0,1)$; (*ii*) $\gamma_{NT}^{-2}(\Gamma_{NT2} + \Gamma_{NT3} - 2\Gamma_{NT9}) = A^{\Delta} + o_P(1)$; and (*iii*) $\gamma_{NT}^{-2}\Gamma_{NTs} = o_P(1)$ for $s = 4,..,8,10$. We prove (*i*) in Proposition B.4 below.

For (*ii*), by Lemma B.3

$$\begin{aligned}
\hat{\theta} - \theta^0 &= \gamma_{NT}D_{NT}^{-1}\Upsilon_{NT} + D^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{it}e_{it} + B_{\theta,NT} + R_{\theta,NT}\\
&= \gamma_{NT}D_{NT}^{-1}\Upsilon_{NT} + O_P[\delta_{NT}^{-2} + (NT)^{-1/2}] = \gamma_{NT}D_{NT}^{-1}\Upsilon_{NT} + o_P(\gamma_{NT}). \qquad \text{(B.2)}
\end{aligned}$$

Then we have $\gamma_{NT}^{-2}\Gamma_{NT2} = \gamma_{NT}^{-2}[\gamma_{NT}D_{NT}^{-1}\Upsilon_{NT} + o_P(\gamma_{NT})]' Q_{wxx,NT}[\gamma_{NT}D_{NT}^{-1}\Upsilon_{NT} + o_P(\gamma_{NT})] = \Upsilon_{NT}' D_{NT}^{-1} Q_{wxx,NT}D_{NT}^{-1}\Upsilon_{NT} + o_P(1)$, and $2\gamma_{NT}^{-2}\Gamma_{NT9} = 2\gamma_{NT}^{-2}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\gamma_{NT}w_{it}\Delta_{it}X_{it}'[\gamma_{NT}D_{NT}^{-1}\Upsilon_{NT} + o_P(\gamma_{NT})] = \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}w_{it}\Delta_{it}X_{it}'D_{NT}^{-1}\Upsilon_{NT} + o_P(1)$. It follows that $\gamma_{NT}^{-2}(\Gamma_{NT2} + \Gamma_{NT3} - 2\Gamma_{NT9}) = (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}w_{it}(\Delta_{it} - X_{it}'D_{NT}^{-1}\Upsilon_{NT})^2 = A^{\Delta} + o_P(1)$.

For (iii), it is clear that $\gamma_{NT}^{-2}\Gamma_{NT4}=O_P\left(\gamma_{NT}^{-2}K^{-2\gamma/d}\right)=o_P(1)$ and $\gamma_{NT}^{-2}\Gamma_{NT10}=O_P\left(\gamma_{NT}^{-1}K^{-\gamma/d}\right)=o_P(1)$ by Assumption 4 and (B.2). We complete the proof of (iii) by showing that $(iii1)$ $\gamma_{NT}^{-2}\Gamma_{NT5}=o_P(1)$, $(iii2)$ $\gamma_{NT}^{-2}\Gamma_{NT6}=o_P(1)$, $(iii3)$ $\gamma_{NT}^{-2}\Gamma_{NT7}=o_P(1)$, and $(iii4)$ $\gamma_{NT}^{-2}\Gamma_{NT8}=o_P(1)$. We first show $(iii1)$. By Lemmas B.2-B.3, we have

$$
\begin{aligned}
\gamma_{NT}^{-2}\Gamma_{NT5} &= \gamma_{NT}^{-2}\left(\hat{\beta}_{bc}-\beta^0\right)' Q_{wpx,NT}(\hat{\theta}-\theta^0)\\
&= \gamma_{NT}^{-2}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{Z}_{it}'e_{it}\tilde{W}^{-1}Q_{wpx,NT}\left(\gamma_{NT}D_{NT}^{-1}\Upsilon_{NT}\right)\\
&\quad+\gamma_{NT}^{-2}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{Z}_{it}'e_{it}\tilde{W}^{-1}Q_{wpx,NT}D^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{it}e_{it}\\
&\quad+\gamma_{NT}^{-2}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{Z}_{it}'e_{it}\tilde{W}^{-1}Q_{wpx,NT}B_{\theta,NT}+\gamma_{NT}^{-2}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{Z}_{it}'e_{it}\tilde{W}^{-1}Q_{wpx,NT}R_{\theta,NT}\\
&\quad+\gamma_{NT}^{-2}R_{\beta,NT}'Q_{wpx,NT}(\hat{\theta}-\theta^0)\\
&\equiv \tilde{\Gamma}_{NT51}+\tilde{\Gamma}_{NT52}+\tilde{\Gamma}_{NT53}+\tilde{\Gamma}_{NT54}+\tilde{\Gamma}_{NT55}, \text{ say.}
\end{aligned}
$$

Recall that $\mathcal{H}_{px}=\tilde{W}^{-1}Q_{wpx}D^{-1}$. We further decompose $\tilde{\Gamma}_{NT51}$ as follows

$$
\begin{aligned}
\tilde{\Gamma}_{NT51} &= \gamma_{NT}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{Z}_{it}'e_{it}\mathcal{H}_{px}\Upsilon_{NT}+\gamma_{NT}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{Z}_{it}'e_{it}\tilde{W}^{-1}Q_{wpx,NT}\left[D_{NT}^{-1}-D^{-1}\right]\Upsilon_{NT}\\
&\quad+\gamma_{NT}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{Z}_{it}'e_{it}\tilde{W}^{-1}\left[Q_{wpx,NT}-Q_{wpx}\right]D^{-1}\Upsilon_{NT}\\
&\equiv \tilde{\Gamma}_{NT51a}+\tilde{\Gamma}_{NT51b}+\tilde{\Gamma}_{NT51c}, \text{ say.}
\end{aligned}
$$

Note that $||(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{Z}_{it}'e_{it}\tilde{W}^{-1}Q_{wpx}D^{-1}||=O_P[(NT)^{-1/2}]$ by Chebyshev inequality and the fact that $E_{\mathcal{D}}||\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{Z}_{it}'e_{it}\mathcal{H}_{px}||^2=\frac{1}{NT}\text{tr}(\tilde{\Omega}\tilde{W}^{-1}Q_{wpx}D^{-2}Q_{wpx}'\tilde{W}^{-1})\leq d/(NT)\,\mu_1(\tilde{\Omega})\mu_1(D^{-2})$ $\times[\mu_1(\tilde{W}^{-1})]^2\mu_1\left(Q_{wpx}'Q_{wpx}\right)=O_P((NT)^{-1})$ by Assumption 11 and Lemma E.3(vi). It follows that $|\tilde{\Gamma}_{NT51a}|\leq\gamma_{NT}^{-1}||\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{Z}_{it}'e_{it}\tilde{W}^{-1}Q_{wpx}D^{-1}||\,\|\Upsilon_{NT}\|=O_P\left(\gamma_{NT}^{-1}(NT)^{-1/2}\right)=o_P(1)$. By the fact that $||\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}e_{it}\tilde{Z}_{it}'\tilde{W}^{-1}||=O_P\left(K^{1/2}(NT)^{-1/2}\right)$, Lemma B.1, and Assumption 11(ii), we have $|\tilde{\Gamma}_{NT51b}|\leq\gamma_{NT}^{-1}\|Q_{wpx,NT}\|\,||\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}e_{it}\tilde{Z}_{it}'\tilde{W}^{-1}||\,\left\|D_{NT}^{-1}-D^{-1}\right\|_F\|\Upsilon_{NT}\|=\gamma_{NT}^{-1}O_P\left((NT)^{-1/2}\right)$ $O_P\left(K^{1/2}(NT)^{-1/2}\right)O_P(1)=o_P(1)$ and $|\tilde{\Gamma}_{NT51c}|\leq\gamma_{NT}^{-1}||\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}e_{it}\tilde{Z}_{it}'\tilde{W}^{-1}||\,\|Q_{wpx,NT}-Q_{wpx}\|_F$ $\left\|D^{-1}\right\|\|\Upsilon_{NT}\|=\gamma_{NT}^{-1}O_P\left(K^{1/2}(NT)^{-1/2}\right)O_P\left(K^{1/2}(NT)^{-1/2}\right)O_P(1)=o_P(1)$. It follows that $\tilde{\Gamma}_{NT51}=o_P(1)$. For $\tilde{\Gamma}_{NT52}$, we decompose it as follows:

$$
\begin{aligned}
\tilde{\Gamma}_{NT52} &= \gamma_{NT}^{-2}\frac{1}{N^2T^2}\sum_{1\leq i\neq j\leq N}\sum_{1\leq t\neq s\leq T}\tilde{Z}_{it}'\mathcal{H}_{px}\tilde{X}_{js}e_{js}e_{it}+\gamma_{NT}^{-2}\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{Z}_{it}'\mathcal{H}_{px}\tilde{X}_{its}e_{it}^2\\
&\quad+\gamma_{NT}^{-2}\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{1\leq t\neq s\leq T}\tilde{Z}_{it}'\mathcal{H}_{px}\tilde{X}_{is}e_{is}e_{it}+\gamma_{NT}^{-2}\frac{1}{N^2T^2}\sum_{t=1}^{T}\sum_{1\leq i\neq j\leq N}\tilde{Z}_{it}'\mathcal{H}_{px}\tilde{X}_{jt}e_{it}e_{jt}\\
&\quad+\gamma_{NT}^{-2}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{Z}_{it}'e_{it}\tilde{W}^{-1}\left[Q_{wpx,NT}-Q_{wpx}\right]D^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{it}'e_{it}\\
&\equiv \tilde{\Gamma}_{NT52a}+\tilde{\Gamma}_{NT52b}+\tilde{\Gamma}_{NT52c}+\tilde{\Gamma}_{NT52d}+\tilde{\Gamma}_{NT52e}, \text{ say.}
\end{aligned}
$$

Recall that $h_{it,js} = \tilde{Z}_{it}'\mathcal{H}_{px}\tilde{X}_{js}$. Apparently, $E_{\mathcal{D}}[\tilde{\Gamma}_{NT52a}] = 0$ and

$$\begin{aligned}
E_{\mathcal{D}}[\tilde{\Gamma}_{NT52a}^2] &= \frac{1}{\gamma_{NT}^4 N^4 T^4}\sum_{1\le i_1\ne j_1\le N}\sum_{1\le i_2\ne j_2\le N}\sum_{1\le t_1\ne s_1\le T}\sum_{1\le t_2\ne s_2\le T} E_{\mathcal{D}}\left[h_{i_1t_1,j_1s_1}e_{j_1s_1}e_{i_1t_1}h_{i_2t_2,j_2s_2}e_{j_2s_2}e_{i_2t_2}\right] \\
&= \frac{2}{\gamma_{NT}^4 N^4 T^4}\sum_{1\le i\ne j\le N}\sum_{1\le t\ne s\le T} E_{\mathcal{D}}\left[h_{it,js}^2 e_{js}^2 e_{it}^2\right] \\
&\le \frac{2}{\gamma_{NT}^4 N^4 T^4}\sum_{1\le i,j\le N}\sum_{1\le t,s\le T} \mathrm{tr}\left[\mathcal{H}_{px}E_{\mathcal{D}}\left(\tilde{X}_{js}\tilde{X}_{js}'e_{js}^2\right)\mathcal{H}_{px}'E_{\mathcal{D}}\left(\tilde{Z}_{it}\tilde{Z}_{it}'e_{it}^2\right)\right] \\
&= \frac{2}{\gamma_{NT}^4 N^2 T^2}\mathrm{tr}\left[\tilde{W}^{-1}Q_{px}D^{-1}\tilde{\Omega}_{xx}D^{-1}Q_{px}'\tilde{W}^{-1}\tilde{\Omega}_{zz}\right] \\
&= \frac{2}{\gamma_{NT}^4 N^2 T^2}\mu_1(\tilde{\Omega}_{zz})\left[\mu_1\left(D^{-1}\right)\right]^2\mu_1(\tilde{\Omega}_{xx})\mu_1\left(Q_{px}'Q_{px}\right)\mu_1(\tilde{W}^{-1})\left\|\tilde{W}^{-1}\right\|_F = O_P\left(\frac{1}{K}\right).
\end{aligned}$$

So $\tilde{\Gamma}_{NT52a} = o_P(1)$ by Chebyshev inequality. For $\tilde{\Gamma}_{NT52b}$, we have $\tilde{\Gamma}_{NT52b} = \frac{1}{\sqrt{\mathbb{V}_{NT}}}\mathrm{tr}(\mathcal{H}_{px}\Omega_{\tilde{x}\tilde{z}}) + \frac{1}{\sqrt{\mathbb{V}_{NT}}}\mathrm{tr}\{\mathcal{H}_{px}\left(\Omega_{\tilde{x}\tilde{z},NT} - \Omega_{\tilde{x}\tilde{z}}\right)\} \equiv \tilde{\Gamma}_{NT52b,1} + \tilde{\Gamma}_{NT52b,2}$, say. For $\tilde{\Gamma}_{NT52b,1}$, using Lemma B.1(v), we have

$$\begin{aligned}
\tilde{\Gamma}_{NT52b,1} &\le \mathbb{V}_{NT}^{-1/2}\mathrm{tr}\left(\Omega_{\tilde{x}\tilde{z}}\tilde{W}^{-1}Q_{wpx}D^{-1}\right) \\
&\le \mathbb{V}_{NT}^{-1/2}\left[\mathrm{tr}\left(\Omega_{\tilde{x}\tilde{z}}\tilde{W}^{-1}Q_{wpx}D^{-1}Q_{wpx}'\tilde{W}^{-1}\Omega_{\tilde{x}\tilde{z}}'\right)\right]^{1/2}\left[\mathrm{tr}\left(D^{-1}\right)\right]^{1/2} \\
&\le \mathbb{V}_{NT}^{-1/2}\left[\mu_1\left(Q_{wpx}'\tilde{W}^{-1}\Omega_{\tilde{x}\tilde{z}}'\Omega_{\tilde{x}\tilde{z}}\tilde{W}^{-1}Q_{wpx}\right)\right]^{1/2}\mathrm{tr}\left(D^{-1}\right) \\
&\le \mathbb{V}_{NT}^{-1/2}\mu_1(\tilde{W}^{-1})\left\|Q_{wpx}\right\|\left\|\Omega_{\tilde{x}\tilde{z}}\right\|O_P(1) = O_P\left(K^{-1/2}\right),
\end{aligned}$$

where we use the fact $\left\|\Omega_{\tilde{x}\tilde{z}}\right\|^2 \le \mu_1(\tilde{\Omega})\mu_1\left(\Omega_{\tilde{x}\tilde{x}}\right) = O_P(1)$ by Assumption 7 and additional assumption that $\mu_1\left(\Omega_{\tilde{x}\tilde{x}}\right) = O_P(1)$. For $\tilde{\Gamma}_{NT52b,2}$, we have $|\tilde{\Gamma}_{NT52b,2}| \le \mathbb{V}_{NT}^{-1/2}\left\|D^{-1}\right\|_F ||\tilde{W}^{-1}||\left\|Q_{wpx}\right\|\left\|\Omega_{\tilde{x}\tilde{z},NT} - \Omega_{\tilde{x}\tilde{z}}\right\|_F$ $= \mathbb{V}_{NT}^{-1/2}O_P(K^{1/2}N^{-1/2}T^{-1/2}) = o_P(1)$. Similarly, we can show that $\tilde{\Gamma}_{NT52s} = o_P(1)$ for $s = c, d$. For, $\tilde{\Gamma}_{NT52e}$, we have

$$\begin{aligned}
|\tilde{\Gamma}_{NT52e}| &\le \gamma_{NT}^{-2}\left\|\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\tilde{Z}_{it}e_{it}\right\|\left\|\tilde{W}^{-1}\right\|\left\|Q_{wpx,NT} - Q_{wpx}\right\|\left\|D^{-1}\right\|\left\|\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\tilde{X}_{it}e_{it}\right\| \\
&= \gamma_{NT}^{-2}O_P(K^{1/2}N^{-1/2}T^{-1/2})O_P(1)\,O_P(K^{1/2}N^{-1/2}T^{-1/2})O_P(1)\,O_P[(NT)^{-1/2}] = O_P(1/\sqrt{NT}).
\end{aligned}$$

Consequently, $\tilde{\Gamma}_{NT52} = o_P(1)$.

Following the proof of $\tilde{\Gamma}_{NT51} = o_P(1)$, we can show that $\tilde{\Gamma}_{NT53} = o_P(1)$. In addition, it is straightforward to show that $\tilde{\Gamma}_{NT5s} = o_P(1)$ for $s = 4, 5$ by using the rough probability bound for the remainder terms $R_{\beta,NT}$ and $R_{\theta,NT}$. It follows that $\gamma_{NT}^{-2}\Gamma_{NT5} = o_P(1)$.

For $(iii2)$, by Cauchy-Schwarz inequality and Lemma B.2, we have

$$\begin{aligned}
\left|\gamma_{NT}^{-2}\Gamma_{NT6}\right| &= \gamma_{NT}^{-2}\left|\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\left(\hat{\beta}_{bc} - \beta^0\right)'w_{it}p_{it}e_{g,it}\right| \\
&\le \gamma_{NT}^{-2}\left\{\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\left(\hat{\beta}_{bc} - \beta^0\right)'w_{it}p_{it}p_{it}'\left(\hat{\beta}_{bc} - \beta^0\right)\right\}^{1/2}\left\{\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T e_{g,it}^2\right\}^{1/2} \\
&\le \gamma_{NT}^{-2}\left\|\hat{\beta}_{bc} - \beta^0\right\|\left[\mu_1\left(Q_{wpp,NT}\right)\right]^{1/2}O_P\left(K^{-\gamma/d}\right) \\
&\le \gamma_{NT}^{-2}O_P\left(K^{1/2}/\sqrt{NT}\right)O_P\left(K^{-\gamma/d}\right) = O_P\left(\sqrt{NT}K^{-\gamma/d}\right) = o_P(1).
\end{aligned}$$

Similarly, $\gamma_{NT}^{-2}\Gamma_{NT8} = O_P\left(\gamma_{NT}^{-1}K^{-\gamma/d}\right) = o_P(1)$, proving $(iii4)$.

We now show $(iii3)$. By Assumption 10, there exists a $K\times 1$ vector $\beta_\Delta^0 \in \mathbb{R}^K$ satisfying $\left\|\beta^\Delta\right\| \le C_\Delta < \infty$ and $\left\|\Delta(x) - p^K(x)'\beta_\Delta^0\right\|_{\infty,\bar{\omega}} = O\left(K^{-\gamma/d}\right)$ for as $K\to\infty$. Using $\Delta_{it} = p_{it}'\beta_\Delta^0 + \left(\Delta_{it} - p_{it}'\beta_\Delta^0\right) = p_{it}'\beta_\Delta^0 + e_{\Delta,it}$, we have

$$\begin{aligned}
\gamma_{NT}^{-2}\Gamma_{NT7} &= \gamma_{NT}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\hat{\beta}_{bc}-\beta^0\right)' w_{it}p_{it}p_{it}'\beta_\Delta^0 + \gamma_{NT}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\hat{\beta}_{bc}-\beta^0\right)' p_{it}e_{\Delta,it}w_{it} \\
&\equiv \tilde{\Gamma}_{NT7a} + \tilde{\Gamma}_{NT7b}, \text{ say.}
\end{aligned}$$

Analogously to the study of $\left|\gamma_{NT}^{-2}\Gamma_{NT6}\right|$, we have $\left|\tilde{\Gamma}_{NT7b}\right| \le C\gamma_{NT}^{-1}\left\|\hat{\beta}_{bc}-\beta^0\right\| O_P\left(K^{-\gamma/d}\right) = o_P(1)$. For $\tilde{\Gamma}_{NT7a}$, by Lemma B.1 we have

$$\begin{aligned}
\tilde{\Gamma}_{NT7a} &= \tilde{\Gamma}_{NT7a1} + \frac{1}{\gamma_{NT}NT}\sum_{i=1}^{N}\sum_{t=1}^{T} e_{it}\tilde{Z}_{it}'\tilde{W}^{-1}\left(Q_{wpp,NT} - Q_{wpp}\right)\beta_\Delta^0 + \gamma_{NT}^{-1}R_{\beta,NT}Q_{wpp,NT}\beta_\Delta^0 \\
&= \tilde{\Gamma}_{NT7a1} + \gamma_{NT}^{-1}O_P\left(K^{1/2}(NT)^{-1/2}\right)O_P\left(K(NT)^{-1/2}\right)O_P(1) + O_P\left(\gamma_{NT}^{-1}R_{\beta,NT}\right) \\
&= \tilde{\Gamma}_{NT7a1} + o_P(1),
\end{aligned}$$

where $\tilde{\Gamma}_{NT7a1} \equiv \frac{1}{\gamma_{NT}NT}\sum_{i=1}^{N}\sum_{t=1}^{T} e_{it}\tilde{Z}_{it}'\tilde{W}^{-1}Q_{wpp}\beta_\Delta^0$. Noting $E_{\mathcal{D}}[\tilde{\Gamma}_{NT7a1}] = 0$ and $E_{\mathcal{D}}[\tilde{\Gamma}_{NT7a1}^2] = (NT)^{-1}\gamma_{NT}^{-2}\,\mathrm{tr}[\tilde{\Omega}W^{-1}Q_{wpp}\beta_\Delta^0\beta_\Delta^{0\prime}Q_{wpp}W^{-1}] \le (NT)^{-1}\mu_1(\tilde{\Omega})\mu_1^2\left(W^{-1}\right)\gamma_{NT}^{-2}\mu_1^2\left(Q_{wpp}\right)\left\|\beta_\Delta^0\right\|^2 = O_P[\gamma_{NT}^2 / (NT)] = o_P(1)$, $\tilde{\Gamma}_{NT7a1} = o_P(1)$ by Chebyshev inequality. It follows that $\gamma_{NT}^{-2}\Gamma_{NT7} = o_P(1)$. ■

Proposition B.4 *Suppose that the assumptions in Theorem 4.1 hold. Then* $\left(NT\Gamma_{NT1} - \mathbb{B}_{NT}\right)/\sqrt{\mathbb{V}_{NT}} \xrightarrow{d} N(0,1)$ *under* $\mathbb{H}_1(\gamma_{NT})$.

Proof. Noting that $\|Q_{wpp,NT} - Q_{wpp}\| = O_P[K/(NT)^{1/2}]$ and $\left\|\hat{\beta}_{bc}-\beta^0\right\| = O_P\left(K^{1/2}/(NT)^{1/2}\right)$, we have $\gamma_{NT}^{-2}\Gamma_{NT1} = \Gamma_{NT1,1} + \gamma_{NT}^{-2}O_P(K/\sqrt{NT})O_P\left(K/(NT)\right) = o_P(1)$, where $\Gamma_{NT1,1} \equiv \gamma_{NT}^{-2}(\hat{\beta}_{bc}-\beta^0)'$ $Q_{wpp}(\hat{\beta}_{bc}-\beta^0)$. We are left to show that $J_{NT1} \equiv \left(NT\Gamma_{NT1,1} - \mathbb{B}_{NT}\right)/\mathbb{V}_{NT}^{1/2} \xrightarrow{d} N(0,1)$.

Let $\bar{Q}_{pp} \equiv \tilde{W}^{-1}Q_{wpp}\tilde{W}^{-1}$, $H_{ij,ts} \equiv \tilde{Z}_{it}'\bar{Q}_{pp}\tilde{Z}_{js}$, $H_{ij} \equiv \tilde{Z}_i'\bar{Q}_{pp}\tilde{Z}_j$, and $\tilde{J}_{NT} = \frac{1}{N^2T^2\gamma_{NT}^2}\sum_{i=1}^{N}\sum_{j=1}^{N} e_i'H_{ij}e_j$. Note that $\gamma_{NT}^{-2}\Gamma_{NT1,1} = \frac{1}{N^2T^2\gamma_{NT}^2}\sum_{i=1}^{N}\sum_{j=1}^{N} e_i'H_{ij}e_j + 2\gamma_{NT}^{-2}R_\beta'Q_{wpp}(\hat{\beta}_{bc}-\beta^0) + \gamma_{NT}^{-2}R_\beta'Q_{wpp}R_\beta = \tilde{J}_{NT} + o_P(1)$. Further, $\tilde{J}_{NT} - \frac{\mathbb{B}_{NT}}{\sqrt{\mathbb{V}_{NT}}} = \frac{1}{N^2T^2\gamma_{NT}^2}\sum_{1\le i\ne j\le N} e_i'H_{ij}e_j + \frac{1}{N^2T^2\gamma_{NT}^2}\sum_{i=1}^{N}\sum_{1\le t\ne s\le T} H_{ii,ts}e_{it}e_{is} \equiv \tilde{J}_{NT,1} + \tilde{J}_{NT,2}$. We complete the proof by showing that: (i) $\tilde{J}_{NT,1} \xrightarrow{d} N(0,1)$ and (ii) $\tilde{J}_{NT,2} = o_P(1)$.

Proof of (i). Note that $\tilde{J}_{NT,1} \equiv \frac{1}{NT\mathbb{V}_{NT}^{1/2}}\sum_{1\le i\ne j\le N} e_i'H_{ij}e_j = \sum_{1\le i<j\le N} W_{ij}$ where $W_{ij} \equiv W_{NT}(u_i,u_j)$ $\equiv 2(NT)^{-1}\mathbb{V}_{NT}^{-1/2}\sum_{1\le s,t\le T} H_{ij,ts}e_{it}e_{js}$ and $u_i \equiv (\tilde{Z}_i', e_i)'$. Noting that $\tilde{J}_{NT,1}$ is a second order degenerate U-statistic that is "clean" ($E_{\mathcal{D}}[W_{NT}(u_i,u)] = E_{\mathcal{D}}[W_{NT}(u,u_j)] = 0$ a.s. for any nonrandom u), we apply Proposition 3.2 in de Jong (1987) to prove the CLT for $\tilde{J}_{NT,1}$ by showing that $(i1)$ $\mathrm{Var}_{\mathcal{D}}(\tilde{J}_{NT,1}) = 1 + o_P(1)$, $(i2)$ $G_I \equiv \sum_{1\le i<j<N} E_{\mathcal{D}}(W_{ij}^4) = o_P(1)$, $(i3)$ $G_{II} \equiv \sum_{1\le i<j<l\le N} E_{\mathcal{D}}(W_{il}^2W_{jl}^2 + W_{ij}^2W_{il}^2 + W_{ij}^2W_{lj}^2) = o_P(1)$, and $(i4)$ $G_{III} \equiv \sum_{1\le i<j<r<l\le N} E_{\mathcal{D}}(W_{ij}\,W_{ir}W_{lj}W_{lr} + W_{ij}W_{il}W_{rj}W_{rl} + W_{ir}W_{il}W_{jr}W_{jl})$ $= o_P(1)$. For $(i1)$, by Assumption $5(ii)$, $E_{\mathcal{D}}(\tilde{J}_{NT,1}) = 0$ and

$$\begin{aligned}
\mathrm{Var}_{\mathcal{D}}\left(\tilde{J}_{NT,1}\right) &= \frac{4}{N^2T^2\mathbb{V}_{NT}}\sum_{1\le i<j\le N}\sum_{t=1}^{T}\sum_{s=1}^{T} E_{\mathcal{D}}\left(H_{ij,ts}^2e_{it}^2e_{js}^2\right) \\
&= \frac{4}{N^2T^2\mathbb{V}_{NT}}\sum_{1\le i<j\le N}\sum_{t=1}^{T}\sum_{s=1}^{T} \mathrm{tr}\left\{\bar{Q}_{pp}E_{\mathcal{D}}\left(\tilde{Z}_{js}\tilde{Z}_{js}'e_{js}^2\right)\bar{Q}_{pp}E_{\mathcal{D}}\left(\tilde{Z}_{it}\tilde{Z}_{it}'e_{it}^2\right)\right\} \\
&= \frac{2}{\mathbb{V}_{NT}}\mathrm{tr}\left(\bar{Q}_{pp}\tilde{\Omega}\bar{Q}_{pp}\tilde{\Omega}\right) - \frac{1}{N^2\mathbb{V}_{NT}}\sum_{i=1}^{N}\mathrm{tr}\left(\bar{Q}_{pp}\tilde{\Omega}_i\bar{Q}_{pp}\tilde{\Omega}_i\right) = 1 - O_P\left(N^{-1}\right)
\end{aligned}$$

where $\tilde{\Omega}_i \equiv T^{-1}\sum_{t=1}^{T} E_{\mathcal{D}}(\tilde{Z}_{it}\tilde{Z}_{it}' e_{it}^2)$ with $N^{-1}\sum_{i=1}^{N}\mu_1(\tilde{\Omega}_i)^2 \le C < \infty$, and we use the fact

$$\frac{1}{N^2\mathbb{V}_{NT}}\sum_{i=1}^{N}\mathrm{tr}\left(\bar{Q}_{pp}\tilde{\Omega}_i\bar{Q}_{pp}\tilde{\Omega}_i\right) \le \left\{N^{-1}\sum_{i=1}^{N}\mu_1(\tilde{\Omega}_i)^2\right\}\frac{\mu_1(\bar{Q}_{pp})\mathrm{tr}\left(\bar{Q}_{pp}\right)}{N\mathbb{V}_{NT}} = O_P(1)\,O_P\left(\frac{1}{N}\right) = o_P(1).$$

Proof of $(i2)$. Let $\bar{q}_{k_1k_2}$ be the (k_1,k_2)th element of $\bar{Q}_{pp}$. Let $\phi_{it,k} = \tilde{Z}_{it,k}e_{it}$. Noting that $H_{ij,ts} = \tilde{Z}_{it}'\bar{Q}\tilde{Z}_{js} = \sum_{k_1=1}^{K}\sum_{k_2=1}^{K}\bar{q}_{k_1k_2}\tilde{Z}_{it,k_1}\tilde{Z}_{js,k_2}$, we have

$$\begin{aligned} G_I &= \frac{16}{N^4T^4\mathbb{V}_{NT}^2}\sum_{1\le k_1,\dots,k_8\le K}\bar{q}_{k_1k_2}\bar{q}_{k_3k_4}\bar{q}_{k_5k_6}\bar{q}_{k_7k_8} \\ &\quad\times\sum_{1\le i<j<N}\sum_{1\le t_1,\dots,t_8\le T}E_{\mathcal{D}}\left(\phi_{it_1,k_1}\phi_{it_3,k_3}\phi_{it_5,k_5}\phi_{it_7,k_7}\right)E_{\mathcal{D}}\left(\phi_{jt_2,k_2}\phi_{jt_4,k_4}\phi_{jt_6,k_6}\phi_{jt_8,k_8}\right). \end{aligned}$$

First, note that the term inside the last summation takes value 0 if either $\#\{t_1,t_3,t_5,t_7\} = 4$ or $\#\{t_2,t_4,t_6,t_8\} = 4$. So it suffices to consider three cases according to the number of distinct time indices in the set $S = \{t_1,...,t_8\}$: (a) $\#S = 6$, (b) $\#S = 5$, and (c) $\#S < 5$. We use G_{Ia}, G_{Ib}, and G_{Ic} to denote the corresponding summations when the time indices are restricted to cases (a), (b), and (c), respectively. Then $G_I = G_{Ia} + G_{Ib} + G_{Ic}$. For G_{Ia}, we must have $\#\{t_1,t_3,t_5,t_7\} = 3$ and $\#\{t_2,t_4,t_6,t_8\} = 3$. Without loss of generality, assume that $t_1 = t_3 > t_5 > t_7$ and $t_2 = t_4 > t_6 > t_8$. By the conditional Davydov inequality (see Lemma E.1) in the supplementary appendix, we have

$$\begin{aligned} E_{\mathcal{D}}\left(\phi_{it_1,k_1}\phi_{it_1,k_3}\phi_{it_5,k_5}\phi_{it_7,k_7}\right) &\le 8\left\|\phi_{it_1,k_1}\phi_{it_1,k_3}\phi_{it_5,k_5}\right\|_{(8+4\delta)/3,\mathcal{D}}\left\|\phi_{it_7,k_7}\right\|_{8+4\delta,\mathcal{D}}\alpha_{\mathcal{D}}^{\frac{1+\delta}{2+\delta}}(t_7-t_5) \\ &\le 8\left\|\phi_{it_1,k_1}\right\|_{8+4\delta,\mathcal{D}}\left\|\phi_{it_1,k_3}\right\|_{8+4\delta,\mathcal{D}}\left\|\phi_{it_5,k_5}\right\|_{8+4\delta,\mathcal{D}}\left\|\phi_{it_7,k_7}\right\|_{8+4\delta,\mathcal{D}}\alpha_{\mathcal{D}}^{\frac{1+\delta}{2+\delta}}(t_7-t_5) \\ &\le 2\left(\Phi_{1,it_1,k_1}+\Phi_{1,it_1,k_3}+\Phi_{1,it_5,k_5}+\Phi_{1,it_7,k_7}\right)\alpha_{\mathcal{D}}^{\frac{1+\delta}{2+\delta}}(t_7-t_5) \end{aligned}$$

where $\Phi_{1,it,k} \equiv \left\|\phi_{it,k}\right\|_{8+4\delta,\mathcal{D}}^4$. Let $C_{1\alpha}(T) \equiv \sum_{m=1}^{T}\alpha_{\mathcal{D}}^{(1+\delta)/(2+\delta)}(m)$ and $C_{2\alpha}(T) \equiv \sum_{m}^{T}\left(1-\frac{m}{T}\right)\alpha_{\mathcal{D}}^{\frac{1+\delta}{2+\delta}}(m)$. Clearly, $\max\{C_{1\alpha}(T), C_{2\alpha}(T)\} < \infty$ by Assumption 5(i). Then

$$\begin{aligned} |G_{Ia}| &\le \frac{64}{N^4T^4\mathbb{V}_{NT}^2}\sum_{1\le k_1,\dots,k_8\le K}|\bar{q}_{k_1k_2}||\bar{q}_{k_3k_4}||\bar{q}_{k_5k_6}||\bar{q}_{k_7k_8}| \\ &\quad\times\left\{\sum_{i=1}^{N}\sum_{1\le t_7<t_5<t_1\le T}\left(\Phi_{1,it_1,k_1}+\Phi_{1,it_1,k_3}+\Phi_{1,it_5,k_5}+\Phi_{1,it_7,k_7}\right)\alpha_{\mathcal{D}}^{\frac{1+\delta}{2+\delta}}(t_7-t_5)\right\} \\ &\quad\times\left\{\sum_{j=1}^{N}\sum_{1\le t_8<t_6<t_2\le T}\left(\Phi_{1,jt_2,k_2}+\Phi_{1,jt_2,k_4}+\Phi_{1,it_6,k_6}+\Phi_{1,jt_8,k_8}\right)\alpha_{\mathcal{D}}^{\frac{1+\delta}{2+\delta}}(t_8-t_6)\right\} \\ &\le \frac{64}{N^4T^2\mathbb{V}_{NT}^2}\sum_{1\le k_1,\dots,k_8\le K}|\bar{q}_{k_1k_2}||\bar{q}_{k_3k_4}||\bar{q}_{k_5k_6}||\bar{q}_{k_7k_8}| \\ &\quad\times\left\{C_{1\alpha}(T)\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\Phi_{1,it,k_1}+\Phi_{1,it,k_3}\right)+C_{2\alpha}(T)\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\Phi_{1,it,k_5}+\Phi_{1,it,k_7}\right)\right\} \\ &\quad\times\left\{C_{1\alpha}(T)\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\Phi_{1,it,k_2}+\Phi_{1,it,k_4}\right)+C_{2\alpha}(T)\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\Phi_{1,it,k_6}+\Phi_{1,it,k_8}\right)\right\} \\ &= \frac{64C}{N^4T^2\mathbb{V}_{NT}^2}O_P\left(K^8N^2T^2\right) = O_P\left(K^6/N^2\right). \end{aligned}$$

Similarly, we can show that $G_{Is} = O_P\left(K^6/N^2\right) = o_P(1)$. It follows that $G_I = O_P\left(K^6/N^2\right) = o_P(1)$.[4]

[4]This is a rough bound but it suffices for our proof. With more complicated arguments, we can show that $G_1 = O_P\left(K^2/N^2\right)$.

For $(i3)$, we write $G_{II} \equiv \sum_{1\leq i<j<l\leq N} E_{\mathcal{D}}(W_{il}^2 W_{jl}^2 + W_{ij}^2 W_{il}^2 + W_{ij}^2 W_{lj}^2) = G_{II,1} + G_{II,2} + G_{II,3}$. By Assumptions 5$(ii)$, we have

$$\begin{aligned}
G_{II,1} &\equiv \frac{16}{N^4T^4\mathbb{V}_{NT}^2} \sum_{1\leq i<j<l\leq N} \sum_{1\leq t_1,\ldots,t_6\leq T} E_{\mathcal{D}}\left[e_{it_1}^2 e_{jt_2}^2 H_{il,t_1t_3} H_{il,t_1t_4} H_{jl,t_2t_5} H_{jl,t_2t_6} e_{lt_3} e_{lt_4} e_{lt_5} e_{lt_6}\right] \\
&= \frac{192}{N^4T^4\mathbb{V}_{NT}^2} \sum_{1\leq i<j<l\leq N} \sum_{1\leq t_1,t_2\leq T} \sum_{1\leq t_3<t_4<t_6\leq T} E_{\mathcal{D}}\left[e_{it_1}^2 e_{jt_2}^2 H_{il,t_1t_3} H_{il,t_1t_4} H_{jl,t_2t_6}^2 e_{lt_3} e_{lt_4} e_{lt_6}^2\right] \\
&\quad + \frac{48}{N^4T^4\mathbb{V}_{NT}^2} \sum_{1\leq i<j<l\leq N} \sum_{1\leq t_1,t_2\leq T} \sum_{1\leq t_3\neq t_6\leq T} E_{\mathcal{D}}\left[e_{it_1}^2 e_{jt_2}^2 H_{il,t_1t_3}^2 H_{jl,t_2t_6}^2 e_{lt_3}^2 e_{lt_6}^2\right] \\
&\equiv G_{II,11} + G_{II,12}, \text{ say.}
\end{aligned}$$

Observe that $G_{II,11} \leq \frac{192}{N^2T^2\mathbb{V}_{NT}^2} \sum_{l=1}^N \sum_{1\leq t_3<t_4<t_6\leq T} E_{\mathcal{D}}\{\mathrm{tr}[e_{lt_4}\tilde{Z}_{lt_4}'\bar{Q}_{pp}\tilde{\Omega}\bar{Q}_{pp}\tilde{Z}_{lt_3}e_{lt_3}]\mathrm{tr}[\bar{Q}_{pp}\tilde{\Omega}\bar{Q}_{pp}\tilde{Z}_{lt_6}\tilde{Z}_{lt_6}'e_{lt_6}^2]\}$. Noting that $E_{\mathcal{D}}[e_{lt_4}\tilde{Z}_{lt_4}'\bar{Q}_{pp}\tilde{\Omega}\bar{Q}_{pp}\tilde{Z}_{lt_3}e_{lt_3}] = 0$, by the conditional Davydov inequality we have

$$\begin{aligned}
&\left|E_{\mathcal{D}}\left\{\mathrm{tr}\left[e_{lt_4}\tilde{Z}_{lt_4}'\bar{Q}_{pp}\tilde{\Omega}\bar{Q}_{pp}\tilde{Z}_{lt_3}e_{lt_3}\right]\mathrm{tr}\left[\bar{Q}_{pp}\tilde{\Omega}\bar{Q}_{pp}\tilde{Z}_{lt_6}\tilde{Z}_{lt_6}'e_{lt_6}^2\right]\right\}\right| \\
\leq\ & 8\left\|\mathrm{tr}\left[e_{lt_4}\tilde{Z}_{lt_4}'\bar{Q}_{pp}\tilde{\Omega}\bar{Q}_{pp}\tilde{Z}_{lt_3}e_{lt_3}\right]\right\|_{4+2\delta,\mathcal{D}}\left\|\mathrm{tr}\left[\bar{Q}_{pp}\tilde{\Omega}\bar{Q}_{pp}\tilde{Z}_{lt_6}\tilde{Z}_{lt_6}'e_{lt_6}^2\right]\right\|_{4+2\delta,\mathcal{D}} \alpha_{\mathcal{D}}^{\frac{1+\delta}{2+\delta}}(t_6-t_4) \\
\leq\ & 8\mu_1^2\left(\bar{Q}_{pp}\tilde{\Omega}\bar{Q}_{pp}\right)\left\|\left\|e_{lt_4}\tilde{Z}_{lt_4}'\right\|_F\left\|\tilde{Z}_{lt_3}e_{lt_3}\right\|_F\right\|_{4+2\delta,\mathcal{D}}\left\|\left\|\tilde{Z}_{lt_6}\right\|_F^2 e_{lt_6}^2\right\|_{4+2\delta}\alpha_{\mathcal{D}}^{\frac{1+\delta}{2+\delta}}(t_6-t_4) \\
\leq\ & CK^2\left(\|e_{lt_4}\|_{8+4\delta,\mathcal{D}}\tilde{\varphi}_{lt_4,8+4\delta}\right)\left(\|e_{lt_3}\|_{8+4\delta,\mathcal{D}}\tilde{\varphi}_{lt_3,8+4\delta}\right)\left(\|e_{lt_6}\|_{8+4\delta,\mathcal{D}}^2\tilde{\varphi}_{lt_6,8+4\delta}^2\right)\alpha_{\mathcal{D}}^{\frac{1+\delta}{2+\delta}}(t_6-t_4) \\
\leq\ & CK^2\left(C_{3,lt_4,e}+C_{3,lt_4,p}+C_{3,lt_3,e}+C_{3,lt_3,p}+2C_{3,lt_6,e}+2C_{3,lt_6,p}\right)\alpha_{\mathcal{D}}^{\frac{1+\delta}{2+\delta}}(t_6-t_4)
\end{aligned}$$

where $C_{3,lt,e} \equiv \|e_{lt}\|_{8+4\delta,\mathcal{D}}^8$, $C_{3,lt,p} \equiv \tilde{\varphi}_{lt,8+4\delta}^8$, $\tilde{\varphi}_{is,q} \equiv K^{-1/q}\left\|\tilde{Z}_{is}\right\|_{q,\mathcal{D}}$, and $E\left|\tilde{\varphi}_{is,q}\right|^{8+4\delta} < \infty$ by Assumption 6(iii). Then

$$\begin{aligned}
G_{II,11} &\leq \frac{192CK^2}{N^2T^2\mathbb{V}_{NT}^2}\sum_{l=1}^N\sum_{1\leq t_3<t_4<t_6\leq T}\left\{\alpha_{\mathcal{D}}^{(1+\delta)/(2+\delta)}(t_6-t_4)\right. \\
&\qquad \left.\times\left(C_{3,lt_4,e}+C_{3,lt_4,p}+C_{3,lt_3,e}+C_{3,lt_3,p}+2C_{3,lt_6,e}+2C_{3,lt_6,p}\right)\right\} \\
&= \frac{CK^2}{N^2T^2\mathbb{V}_{NT}^2}\left[TC_{2\alpha}(T)+3TC_{1\alpha}(T)\right]\sum_{i=1}^N\sum_{t=1}^T\left(C_{3,it,e}+C_{3,it,p}\right) = O_P\left(N^{-1}\right)
\end{aligned}$$

by Assumption A5(i). Similarly,

$$\begin{aligned}
G_{II,12} &\leq \frac{48}{N^4T^4\mathbb{V}_{NT}^2}\sum_{1\leq i<j<l\leq N}\sum_{1\leq t_1,t_2\leq T}\sum_{1\leq t_3\neq t_6\leq T}E_{\mathcal{D}}\left\{\mathrm{tr}\left[E_{\mathcal{D}}(e_{it_1}^2\tilde{Z}_{it_1}\tilde{Z}_{it_1}')\bar{Q}_{pp}e_{lt_3}^2\tilde{Z}_{lt_3}\tilde{Z}_{lt_3}'\bar{Q}_{pp}\right]\right. \\
&\qquad \left.\times\ \mathrm{tr}\left[E_{\mathcal{D}}(e_{jt_2}^2\tilde{Z}_{jt_2}\tilde{Z}_{jt_2}')\bar{Q}_{pp}e_{lt_6}^2\tilde{Z}_{lt_6}\tilde{Z}_{lt_6}'\bar{Q}_{pp}\right]\right\} \\
&\leq \frac{8}{N^2T^2\mathbb{V}_{NT}^2}\sum_{l=1}^N\sum_{1\leq t_3\neq t_6\leq T}E_{\mathcal{D}}\left\{\mathrm{tr}\left[\tilde{\Omega}\bar{Q}_{pp}e_{lt_3}^2\tilde{Z}_{lt_3}\tilde{Z}_{lt_3}'\bar{Q}_{pp}\right]\mathrm{tr}\left[\tilde{\Omega}\bar{Q}_{pp}e_{lt_6}^2\tilde{Z}_{lt_6}\tilde{Z}_{lt_6}'\bar{Q}_{pp}\right]\right\} \\
&\leq \frac{8}{N^2T^2\mathbb{V}_{NT}^2}\sum_{l=1}^N\sum_{1\leq t_3\neq t_6\leq T}\left\{E_{\mathcal{D}}\left[\mathrm{tr}\left(\tilde{\Omega}\bar{Q}_{pp}e_{lt_3}^2\tilde{Z}_{lt_3}\tilde{Z}_{lt_3}'\bar{Q}_{pp}\right)\right]E_{\mathcal{D}}\left[\mathrm{tr}\left(\tilde{\Omega}\bar{Q}_{pp}e_{lt_6}^2\tilde{Z}_{lt_6}\tilde{Z}_{lt_6}'\bar{Q}_{pp}\right)\right]\right. \\
&\qquad \left.+\ 8\left\|\mathrm{tr}\left[\tilde{\Omega}\bar{Q}_{pp}e_{lt_3}^2\tilde{Z}_{lt_3}\tilde{Z}_{lt_3}'\bar{Q}_{pp}\right]\right\|_{4+2\delta,\mathcal{D}}\left\|\mathrm{tr}\left[\tilde{\Omega}\bar{Q}_{pp}e_{lt_6}^2\tilde{Z}_{lt_6}\tilde{Z}_{lt_6}'\bar{Q}_{pp}\right]\right\|_{4+2\delta,\mathcal{D}}\alpha_{\mathcal{D}}^{\frac{1+\delta}{2+\delta}}(|t_6-t_3|)\right\}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{8\mu_1^2\left(\tilde{\Omega}\right)\mu_1^4\left(\bar{Q}_{pp}\right)}{N^2T^2\mathbb{V}_{NT}^2}\sum_{l=1}^{N}\sum_{1\leq t_3\neq t_6\leq T}E_{\mathcal{D}}\left[e_{lt_3}^2\left\|\tilde{Z}_{lt_3}\right\|^2\right]E_{\mathcal{D}}\left[e_{lt_6}^2\left\|\tilde{Z}_{lt_6}\right\|^2\right] \\
&\quad +\frac{64\mu_1^2\left(\tilde{\Omega}\right)\mu_1^4\left(\bar{Q}_{pp}\right)}{N^2T^2\mathbb{V}_{NT}^2}\sum_{l=1}^{N}\sum_{1\leq t_3\neq t_6\leq T}\left\|e_{lt_3}^2\left\|\tilde{Z}_{lt_3}\right\|^2\right\|_{4+2\delta,\mathcal{D}}\left\|e_{lt_6}^2\left\|\tilde{Z}_{lt_6}\right\|^2\right\|_{4+2\delta,\mathcal{D}}\alpha_{\mathcal{D}}^{\frac{1+\delta}{2+\delta}}\left(|t_6-t_3|\right) \\
&= O_P\left(\frac{1}{N^2T^2K^2}\right)\left[O_P\left(NT^2K^2\right)+O_P\left(NTK^2\right)\right]=O_P\left(N^{-1}\right).
\end{aligned}
$$

Thus $G_{II,1}=o_P(1)$. Similarly, we can show that $G_{II,2}=o_P(1)$ and $G_{II,3}=o_P(1)$. It follows that $G_{II}=o_P(1)$.

For $(i4)$, we write $G_{III}\equiv\sum_{1\leq i<j<r<l\leq N}E_{\mathcal{D}}\left(W_{ij}W_{ir}W_{lj}W_{lr}+W_{ij}W_{il}W_{rj}W_{rl}+W_{ir}W_{il}W_{jr}W_{jl}\right)\equiv\sum_{s=1}^{4}G_{III,s}$, say. By Assumptions 5$(ii)$, we have

$$
\begin{aligned}
G_{III,1} &= \frac{16}{N^4T^4\mathbb{V}_{NT}^2}\sum_{1\leq i<j<r<l\leq N}\sum_{1\leq t_1,\ldots,t_8\leq T}E_{\mathcal{D}}\left[H_{ij,t_1t_2}e_{it_1}e_{jt_2}H_{ir,t_3t_4}e_{it_3}e_{rt_4}H_{lj,t_5t_6}e_{lt_5}e_{jt_6}H_{lr,t_7t_8}e_{lt_7}e_{rt_8}\right] \\
&= \frac{16}{N^4T^4\mathbb{V}_{NT}^2}\sum_{1\leq i<j<r<l\leq N}\sum_{1\leq t_1,t_2,t_4,t_5\leq T}E_{\mathcal{D}}\left[e_{it_1}^2e_{jt_2}^2e_{lt_5}^2e_{rt_4}^2H_{ij,t_1t_2}H_{ir,t_1t_4}H_{lj,t_5t_2}H_{lr,t_5t_4}\right] \\
&= \frac{16}{N^4T^4\mathbb{V}_{NT}^2}\sum_{1\leq i<j<r<l\leq N}\sum_{1\leq t,s,p,q\leq T}\mathrm{tr}\left[E_{\mathcal{D}}\left(\bar{Q}_{pp}\tilde{Z}_{it}\tilde{Z}_{it}'e_{it}^2\bar{Q}_{pp}\tilde{Z}_{rs}\tilde{Z}_{rs}'e_{rs}^2\bar{Q}_{pp}\tilde{Z}_{lp}\tilde{Z}_{lp}'e_{lp}^2\bar{Q}_{pp}\tilde{Z}_{jq}\tilde{Z}_{jq}'e_{jq}^2\right)\right] \\
&= \frac{16}{24N^4\mathbb{V}_{NT}^2}\sum_{1\leq i\neq j\neq r\neq l\leq N}\mathrm{tr}\left(\bar{Q}_{pp}\tilde{\Omega}_i\bar{Q}_{pp}\tilde{\Omega}_r\bar{Q}_{pp}\tilde{\Omega}_l\bar{Q}_{pp}\tilde{\Omega}_j\right) \\
&= \frac{2}{3\mathbb{V}_{NT}^2}\mathrm{tr}\left(\bar{Q}_{pp}\tilde{\Omega}\bar{Q}_{pp}\tilde{\Omega}\bar{Q}_{pp}\tilde{\Omega}\bar{Q}_{pp}\tilde{\Omega}\right)+O_P\left(\frac{1}{NK}\right)=O_P\left(\frac{1}{K}\right),
\end{aligned}
$$

where we use the facts that $\mathrm{tr}\left(\bar{Q}_{pp}\tilde{\Omega}\bar{Q}_{pp}\tilde{\Omega}\bar{Q}_{pp}\tilde{\Omega}\bar{Q}_{pp}\tilde{\Omega}\right)\leq\mu_1^4\left(\bar{Q}_{pp}\right)\mu_1^3(\tilde{\Omega})\mathrm{tr}(\tilde{\Omega})=O_P(K)$ and $N^{-1}\sum_{i=1}^{N}\tilde{\Omega}_i$ $=\tilde{\Omega}$ in the last line.

For (ii), we can easily show that $\tilde{J}_{NT2}=O_P\left(N^{-1/2}\right)=o_P(1)$ by conditional Chebyshev inequality. The details are omitted to save space. ■

Proof of Theorem 4.2. Note that $\hat{J}_{NT}=\frac{NT\Gamma_{NT}-\hat{\mathbb{B}}_{NT}}{\sqrt{\hat{\mathbb{V}}_{NT}}}=J_{NT}\left(\frac{\mathbb{V}_{NT}}{\hat{\mathbb{V}}_{NT}}\right)^{1/2}+\frac{\mathbb{B}_{NT}-\hat{\mathbb{B}}_{NT}}{\sqrt{\hat{\mathbb{V}}_{NT}}}$ and $\mathbb{V}_{NT}^{-1}=O_P\left(K^{-1}\right)$, by Theorem 4.1 it suffices to show that (i) $\hat{\mathbb{B}}_{NT}-\mathbb{B}_{NT}=o_P\left(K^{1/2}\right)$ and (ii) $\hat{\mathbb{V}}_{NT}-\mathbb{V}_{NT}=o_P(K)$. We first prove (i).

$$
\begin{aligned}
\hat{\mathbb{B}}_{NT}-\mathbb{B}_{NT} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\hat{e}_{it}^2\hat{Z}_{it}'\hat{W}_{NT}^{-1}Q_{wpp,NT}\hat{W}_{NT}^{-1}\hat{Z}_{it}-e_{it}^2\tilde{Z}_{it}'\tilde{W}^{-1}Q_{wpp}\tilde{W}^{-1}\tilde{Z}_{it}\right) \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{e}_{it}^2\left[\hat{Z}_{it}'\tilde{W}^{-1}Q_{wpp}\tilde{W}^{-1}\hat{Z}_{it}-\tilde{Z}_{it}'\tilde{W}^{-1}Q_{wpp}\tilde{W}^{-1}\tilde{Z}_{it}\right] \\
&\quad +\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\hat{e}_{it}^2-e_{it}^2\right)\tilde{Z}_{it}'\tilde{W}^{-1}Q_{wpp}\tilde{W}^{-1}\tilde{Z}_{it} \\
&\quad +\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{e}_{it}^2\left[\hat{Z}_{it}'\left(\hat{W}_{NT}^{-1}Q_{wpp,NT}\hat{W}_{NT}^{-1}-\tilde{W}^{-1}Q_{wpp}\tilde{W}^{-1}\right)\hat{Z}_{it}\right] \\
&\equiv \mathbb{DB}_{1NT}+\mathbb{DB}_{2NT}+\mathbb{DB}_{3NT},\text{ say.}
\end{aligned}
$$

Following the proof of Lemma A.9(i), we can readily show that $\mathbb{DB}_{sNT}=o_P(1)$ for $s=1,2$ because $Q_{wpp}^{1/2}\tilde{W}^{-1}\tilde{Z}_{it}$ and $Q_{wpp}^{1/2}\tilde{W}^{-1}\hat{Z}_{it}$ behave similarly to $\tilde{Z}_{it}$ and $\hat{Z}_{it}$, respectively. Let $\hat{w}\equiv\hat{W}_{NT}^{-1}Q_{wpp,NT}\hat{W}_{NT}^{-1}$

and $\tilde{w} \equiv \tilde{W}^{-1} Q_{wpp} \tilde{W}^{-1}$. Then $\mathbb{DB}_{3NT}$ =tr$[(\hat{w}-\tilde{w})\,\hat{\Omega}_{NT}]$. By Minkowski inequality,

$$\begin{aligned}\|\hat{w}-\tilde{w}\|_F &\leq \left\|\tilde{W}^{-1}\left(Q_{wpp,NT}-Q_{wpp}\right)\tilde{W}^{-1}\right\|_F + \left\|(\hat{W}_{NT}^{-1}-\tilde{W}^{-1})Q_{wpp,NT}(\hat{W}_{NT}^{-1}-\tilde{W}^{-1})\right\|_F \\ &\quad +2\left\|\tilde{W}^{-1}Q_{wpp,NT}(\hat{W}_{NT}^{-1}-\tilde{W}^{-1})\right\|_F \\ &= w_{1NT}+w_{2NT}+2w_{3NT}, \text{ say.}\end{aligned}$$

By the matrix version of Cauchy-Schwarz inequality, the fact that tr$(AB) \leq \mu_1(B)$tr(A) for any symmetric matrix A and p.s.d. matrix B, and Lemma B.1, we have

$$\begin{aligned}w_{1NT} &\leq \left\{\mathrm{tr}\left[\tilde{W}^{-1}\left(Q_{wpp,NT}-Q_{wpp}\right)\tilde{W}^{-1}\tilde{W}^{-1}\left(Q_{wpp,NT}-Q_{wpp}\right)\tilde{W}^{-1}\right]\right\}^{1/2} \\ &\leq \mu_1\left(\tilde{W}^{-1}\right)\left\{\mathrm{tr}\left[\tilde{W}^{-1}\left(Q_{wpp,NT}-Q_{wpp}\right)\left(Q_{wpp,NT}-Q_{wpp}\right)\tilde{W}^{-1}\right]\right\}^{1/2} \\ &\leq \left[\mu_1\left(\tilde{W}^{-1}\right)\right]^2\|Q_{wpp,NT}-Q_{wpp}\|_F = O_P(1)\,O_P(K/(NT)^{1/2}) = O_P(K/(NT)^{1/2}).\end{aligned}$$

Similarly, we can apply Lemmas A.8 and B.1 and show that $w_{3NT,2} = O_P(K^2(K^{-2\gamma/d}+\delta_{NT}^{-2}))$ and $w_{3NT,3} = O_P(K\left(K^{-\gamma/d}+\delta_{NT}^{-1}\right))$. It follows that

$$\|\hat{w}-\tilde{w}\|_F = O_P\left(K\left(K^{-\gamma/d}+\delta_{NT}^{-1}\right)\right), \tag{B.3}$$

and $|\mathbb{DB}_{3NT}| \leq \|\hat{w}-\tilde{w}\|_F\,||\hat{\Omega}_{NT}||_F = O_P[K(K^{-\gamma/d}+\delta_{NT}^{-1})]O_P(K^{1/2}) = o_P(K^{1/2})$. Thus $\hat{\mathbb{B}}_{NT}-\mathbb{B}_{NT} = o_P(K^{1/2})$.

(ii) Using the notation $\hat{w}$ and $\tilde{w}$, we can decompose $\hat{\mathbb{V}}_{NT}-\mathbb{V}_{NT}$ as follows

$$\begin{aligned}\hat{\mathbb{V}}_{NT}-\mathbb{V}_{NT} &= 2\mathrm{tr}\left(\hat{w}\hat{\Omega}_{NT}\hat{w}\hat{\Omega}_{NT}-\tilde{w}\tilde{\Omega}\tilde{w}\tilde{\Omega}\right) \\ &= 2\mathrm{tr}\left(\tilde{w}\hat{\Omega}_{NT}\tilde{w}\hat{\Omega}_{NT}-\tilde{w}\tilde{\Omega}\tilde{w}\tilde{\Omega}\right) \\ &\quad +2\mathrm{tr}\left[\left(\hat{w}-\tilde{w}\right)\hat{\Omega}_{NT}\left(\hat{w}-\tilde{w}\right)\hat{\Omega}_{NT}\right]+4\mathrm{tr}\left[\left(\hat{w}-\tilde{w}\right)\hat{\Omega}_{NT}\tilde{w}\hat{\Omega}_{NT}\right] \\ &= 2\mathrm{tr}\left[\tilde{w}(\hat{\Omega}_{NT}-\tilde{\Omega})\tilde{w}(\hat{\Omega}_{NT}-\tilde{\Omega})\right]+4\mathrm{tr}\left[\tilde{w}(\hat{\Omega}_{NT}-\tilde{\Omega})\tilde{w}\tilde{\Omega}\right] \\ &\quad +2\mathrm{tr}\left[\left(\hat{w}-\tilde{w}\right)\hat{\Omega}_{NT}\left(\hat{w}-\tilde{w}\right)\hat{\Omega}_{NT}\right]+4\mathrm{tr}\left[\left(\hat{w}-\tilde{w}\right)\hat{\Omega}_{NT}\tilde{w}\hat{\Omega}_{NT}\right] \\ &\equiv 2\mathbb{DV}_{1NT}+4\mathbb{DV}_{2NT}+2\mathbb{DV}_{3NT}+4\mathbb{DV}_{3NT}.\end{aligned}$$

Observe that $|\mathbb{DV}_{1NT}| \leq [\mu_1(\tilde{w})]^2\,||\hat{\Omega}_{NT}-\tilde{\Omega}||^2 = O_P(K^2/(NT)) = o_P(1)$, and by (B.3)

$$\begin{aligned}|\mathbb{DV}_{2NT}| &\leq \left\{\mathrm{tr}\left[\left(\hat{w}-\tilde{w}\right)\hat{\Omega}_{NT}\hat{\Omega}_{NT}\left(\hat{w}-\tilde{w}\right)\right]\right\}^{1/2}\left[\mathrm{tr}\left(\tilde{w}\hat{\Omega}_{NT}\hat{\Omega}_{NT}\tilde{w}\right)\right]^{1/2} \\ &\leq [\mu_1(\hat{\Omega}_{NT})]^2\,\|\hat{w}-\tilde{w}\|_F\,\|\tilde{w}\|_F = O_P(1)\,O_P(K\left(K^{-\gamma/d}+\delta_{NT}^{-1}\right))O_P\left(K^{1/2}\right) = o_P\left(K^{1/2}\right).\end{aligned}$$

Similarly, we can show that $|\mathbb{DV}_{3NT}| = O_P(K^2(K^{-2\gamma/d}+\delta_{NT}^{-2})) = o_P(1)$ and $|\mathbb{DV}_{4NT}| = O_P(K^{3/2}(K^{-\gamma/d}+\delta_{NT}^{-1})) = o_P\left(K^{1/2}\right)$. Consequently, $\hat{\mathbb{V}}_{NT}-\mathbb{V}_{NT} = o_P\left(K^{1/2}\right) = o_P(K)$. ■

Supplementary Material On

"Nonparametric Dynamic Panel Data Models with Interactive Fixed Effects: Sieve Estimation and Specification Testing"

Liangjun Su, Yonghui Zhang

School of Economics, Singapore Management University

THIS APPENDIX PROVIDES PROOFS FOR SOME TECHNICAL LEMMAS AND THEOREM 4.4 IN THE ABOVE PAPER.

C Expansion of the quasi-log-likelihood function

We extend the expansion of the (negative) quasi-log-likelihood function of Moon and Weidner (2010) to our nonparametric framework. This expansion is the starting point of our asymptotic analysis. Given the sieve basis $\{p_k(x), k = 1, ..., K\}$, we can linearize model (1.1) as (2.1). Compared with Moon and Weidner's (2010) linear model, the number of regressors increases as sample size (N, T) tends to infinity in (2.1) and the new error term includes an extra component, i.e., the sieve approximation error. We can modify the proof in Moon and Weidner (2010) and still resort to the perturbation theory of operator in Kato (1980) to establish the first order expansion of approximating quasi-log-likelihood function.

Define

$$\Phi_1 \equiv f^0 \left(f^{0\prime} f^0\right)^{-1} \left(\lambda^{0\prime}\lambda^0\right)^{-1} \left(f^{0\prime} f^0\right)^{-1} f^{0\prime} \text{ and } \Phi_2 \equiv \lambda^0 \left(\lambda^{0\prime}\lambda^0\right)^{-1} \left(f^{0\prime} f^0\right)^{-1} \left(\lambda^{0\prime}\lambda^0\right)^{-1} \lambda^{0\prime}. \tag{C.1}$$

Recall that $\Phi = f^0(f^{0\prime}f^0)^{-1}(\lambda^{0\prime}\lambda^0)^{-1}\lambda^{0\prime}$ and $\vartheta_{NT} = \sum_{k=0}^{K} \epsilon_k \mathbf{P}_k$, where $\epsilon_k = \beta_k - \beta_k^0$ for $k = 1, ..., K$, $\epsilon_0 = \|\mathbf{u}\| / \sqrt{NT}$, and $\mathbf{P}_0 = (\sqrt{NT}/\|\mathbf{u}\|)\, \|\mathbf{u}\|$. Let $d_{\max}\left(\lambda^0, f^0\right)$, $d_{\min}\left(\lambda^0, f^0\right)$, $r_0\left(\lambda^0, f^0\right)$, and α_{NT} be as defined at the beginning of the Appendix.

Proposition C.1 *Suppose that $\|\vartheta_{NT}\| \leq \sqrt{NT} r_0\left(\lambda^0, f^0\right)$. Let $\hat{\lambda}(\beta)$ and $\hat{f}(\beta)$ be the minimizing parameters in (2.4). Let $M_{\hat{\lambda}}(\beta) \equiv M_{\hat{\lambda}(\beta)}$ and $M_{\hat{f}}(\beta) \equiv M_{\hat{f}(\beta)}$. Then*

(i) the profile quasi-log-likelihood function can be written as a power series in the $K+1$ parameters ϵ_k $(k = 0, 1, ..., K)$, i.e.,

$$\begin{aligned}\mathcal{L}_{NT}^0(\beta) \;\equiv\; & \frac{1}{NT} \sum_{k_1=0}^{K} \sum_{k_2=0}^{K} \epsilon_{k_1}\epsilon_{k_2} L^{(2)}\left(\lambda^0, f^0, \mathbf{P}_{k_1}, \mathbf{P}_{k_2}\right) \\ & + \frac{1}{NT} \sum_{k_1=0}^{K} \sum_{k_2=0}^{K} \sum_{k_3=0}^{K} \epsilon_{k_1}\epsilon_{k_2}\epsilon_{k_3} L^{(3)}\left(\lambda^0, f^0, \mathbf{P}_{k_1}, \mathbf{P}_{k_2}, \mathbf{P}_{k_3}\right) + O_P\left(\alpha_{NT}^4\right) \end{aligned} \tag{C.2}$$

where $L^{(2)}\left(\lambda^0, f^0, \mathbf{P}_{k_1}, \mathbf{P}_{k_2}\right) \equiv tr\left(M_{\lambda^0}\mathbf{P}_{k_1} M_{f^0} \mathbf{P}_{k_2}'\right)$ and

$$L^{(3)}\left(\lambda^0, f^0, \mathbf{P}_{k_1}, \mathbf{P}_{k_2}, \mathbf{P}_{k_3}\right) \equiv -\frac{1}{3!} \sum_{\text{all 6 permutations for } (k_1,k_2,k_3)} tr\left(M_{\lambda^0}\mathbf{P}_{k_1} M_{f^0} \mathbf{P}_{k_2}' \Phi \mathbf{P}_{k_3}'\right);$$

(ii) the projector $M_{\hat{\lambda}}(\beta)$ can be written as a power series in the parameters ϵ_k $(k = 0, 1, ..., K)$, i.e.,

$$M_{\hat{\lambda}}(\beta) = M_{\lambda^0} + \sum_{k=0}^{K} \epsilon_k M_{\lambda}^{(1)}\left(\lambda^0, f^0, \mathbf{P}_k\right) + \sum_{k_1=0}^{K} \sum_{k_2=0}^{K} \epsilon_{k_1}\epsilon_{k_2} M_{\lambda}^{(2)}\left(\lambda^0, f^0, \mathbf{P}_{k_1}, \mathbf{P}_{k_2}\right) + O_P\left(\alpha_{NT}^3\right)$$

where $M_{\lambda}^{(1)}\left(\lambda^{0},f^{0},\mathbf{P}_{k}\right)=-M_{\lambda^{0}}\mathbf{P}_{k}\Phi-\Phi'\mathbf{P}_{k}'M_{\lambda^{0}}$ *and*

$$\begin{aligned} M_{\lambda}^{(2)}\left(\lambda^{0},f^{0},\mathbf{P}_{k_1},\mathbf{P}_{k_2}\right) &= M_{\lambda^{0}}\mathbf{P}_{k_1}\Phi\mathbf{P}_{k_2}\Phi+\Phi'\mathbf{P}_{k_2}'\Phi'\mathbf{P}_{k_1}'M_{\lambda^{0}}-M_{\lambda^{0}}\mathbf{P}_{k_1}M_{f^{0}}\mathbf{P}_{k_2}'\Phi_2 \\ &\quad -\Phi_2\mathbf{P}_{k_2}M_{f^{0}}\mathbf{P}_{k_1}'M_{\lambda^{0}}-M_{\lambda^{0}}\mathbf{P}_{k_1}\Phi_1\mathbf{P}_{k_2}'M_{\lambda^{0}}+\Phi'\mathbf{P}_{k_1}'M_{\lambda^{0}}\mathbf{P}_{k_2}\Phi; \end{aligned}$$

(iii) the projector $M_{\hat{f}}(\beta)$ *can be written as a power series in the parameters* $\epsilon_k\ (k=0,1,...,K)$*, i.e.,*

$$M_{\hat{f}}(\beta)=M_{f^{0}}+\sum_{k=0}^{K}\epsilon_k M_{f}^{(1)}\left(\lambda^{0},f^{0},\mathbf{P}_{k}\right)+\sum_{k_1=0}^{K}\sum_{k_2=0}^{K}\epsilon_{k_1}\epsilon_{k_2}M_{f}^{(2)}\left(\lambda^{0},f^{0},\mathbf{P}_{k_1},\mathbf{P}_{k_2}\right)+O_P\left(\alpha_{NT}^{3}\right)$$

where $M_{f}^{(1)}\left(\lambda^{0},f^{0},\mathbf{P}_{k}\right)=-M_{f^{0}}\mathbf{P}_{k}'\Phi'-\Phi\mathbf{P}_{k}M_{f^{0}}$ *and*

$$\begin{aligned} M_{f}^{(2)}\left(\lambda^{0},f^{0},\mathbf{P}_{k_1},\mathbf{P}_{k_2}\right) &= M_{f^{0}}\mathbf{P}_{k_1}'\Phi'\mathbf{P}_{k_2}'\Phi'+\Phi\mathbf{P}_{k_2}\Phi\mathbf{P}_{k_1}M_{f^{0}}-M_{f^{0}}\mathbf{P}_{k_1}'M_{\lambda^{0}}\mathbf{P}_{k_2}\Phi_1 \\ &\quad -\Phi_1\mathbf{P}_{k_2}'M_{\lambda^{0}}\mathbf{P}_{k_1}M_{f^{0}}'-M_{f^{0}}\mathbf{P}_{k_1}'\Phi_2\mathbf{P}_{k_2}M_{f^{0}}+\Phi\mathbf{P}_{k_1}M_{f^{0}}\mathbf{P}_{k_2}'\Phi'; \end{aligned}$$

Proof. (i) The proof follows the proofs of Theorems 2.1 and 3.1 in Moon and Weidner (2010) closely, and is composed of two steps.

Step 1. We expand the quasi-log-likelihood function into the summation of an infinite sequence. Observe that

$$\mathbf{Y}-\sum_{k=1}^{K}\beta_k\mathbf{P}_k=\lambda^{0}f^{0\prime}+\epsilon_0\mathbf{P}_0+\epsilon_1\mathbf{P}_1+\cdots+\epsilon_K\mathbf{P}_K, \tag{C.3}$$

where we can view the last $K+1$ terms as perturbations to the leading term $\lambda^{0}f^{0\prime}$. Now we rewrite the profile quasi-log-likelihood function in (2.6) as follows:

$$\frac{1}{NT}\sum_{R+1}^{T}\mu_t\left[\left(\mathbf{Y}-\sum_{k=1}^{K}\beta_k\mathbf{P}_k\right)'\left(\mathbf{Y}-\sum_{k=1}^{K}\beta_k\mathbf{P}_k\right)\right]=\frac{1}{NT}\sum_{R+1}^{T}\mu_t\left[\mathcal{T}(1)\right] \tag{C.4}$$

where $\mathcal{T}(\varkappa)\equiv\mathcal{T}^{0}+\varkappa\mathcal{T}^{(1)}+\varkappa^{2}\mathcal{T}^{(2)}$,

$$\mathcal{T}^{0}\equiv f^{0}\lambda^{0\prime}\lambda^{0}f^{0\prime},\ \mathcal{T}^{(1)}\equiv\vartheta_{NT}\left(\lambda^{0}f^{0\prime}+f^{0}\lambda^{0\prime}\right),\text{ and }\mathcal{T}^{(2)}\equiv\vartheta_{NT}\vartheta_{NT}. \tag{C.5}$$

Clearly, if $\epsilon_k=0$ for $k=0,1,\cdots,K$, then the $T-R$ smallest eigenvalues of $\mathcal{T}^{0}$ are all equal to zero.

Since $\mathcal{T}(1)\equiv\mathcal{T}^{0}+\mathcal{T}^{(1)}+\mathcal{T}^{(2)}$, under some conditions to be specified later (see (C.11) and (C.12) below), we can expand the weighted mean $\hat{\lambda}(1)$ of the λ-group eigenvalues ($\lambda=0$ in this case) as

$$\hat{\lambda}(1)\equiv\frac{1}{T-R}\sum_{R+1}^{T}\mu_t\left[\mathcal{T}(1)\right]=0+\sum_{g=0}^{\infty}1^{g}\hat{\lambda}^{(g)}, \tag{C.6}$$

where the coefficients $\hat{\lambda}^{(g)}$ are given by

$$\hat{\lambda}^{(g)}\equiv\frac{1}{(T-R)}\sum_{p=1}^{g}(-1)^{p+1}\sum_{\substack{v_1+\cdots+v_p=g,\\ m_1+\cdots+m_{p+1}=p-1\\ 2\geq v_j\geq1,m_j\geq0}}\operatorname{tr}\left(S^{(k_1)}\mathcal{T}^{(v_1)}S^{(k_2)}\cdots S^{(k_p)}\mathcal{T}^{(v_p)}S^{(k_{p+1})}\right), \tag{C.7}$$

$$S^{(0)}\equiv-M_{\lambda^{0}},\ S^{(k)}\equiv\left[\lambda^{0}\left(\lambda^{0\prime}\lambda^{0}\right)^{-1}\left(f^{0\prime}f^{0}\right)^{-1}\left(\lambda^{0\prime}\lambda^{0}\right)^{-1}\lambda^{0\prime}\right]^{k}, \tag{C.8}$$

and $\mathcal{T}^{(s)}$ $(s=1,2)$ are defined in (C.5). Note that $2\geq v_j$ comes from the facts that $\mathcal{T}^{(g)}\equiv 0$ for $g=3,4,\cdots$, $k_j\geq 0$ and requirement $-T+R+1<0$. See (2.12) in p. 76, (2.18) in p. 77, and (2.22) in p. 78 in Kato (1980) for more details. Using the expressions in (C.5) for $\mathcal{T}^{(1)}$ and $\mathcal{T}^{(2)}$ we have

$$\begin{aligned}\frac{1}{NT}\sum_{R+1}^{T}\mu_t\left[\mathcal{T}(1)\right] &= \frac{1}{NT}\sum_{g=1}^{\infty}\sum_{p=1}^{g}(-1)^{p+1}\sum_{\substack{v_1+\cdots+v_p=s\\ m_1+\cdots+m_{p+1}=p-1\\ 2\geq v_j\geq 1,k_j\geq 0}}\operatorname{tr}\left(S^{(m_1)}\mathcal{T}^{(v_1)}S^{(m_2)}\cdots S^{(m_p)}\mathcal{T}^{(v_p)}S^{(m_{p+1})}\right)\\ &= \frac{1}{NT}\sum_{g=2}^{\infty}\sum_{k_1=0}^{K}\sum_{k_2=0}^{K}\cdots\sum_{k_g=0}^{K}\epsilon_{k_1}\epsilon_{k_2}\cdots\epsilon_{k_g}L^{(g)}\left(\lambda^0,f^0,\mathbf{P}_{k_1},\cdots,\mathbf{P}_{k_g}\right) \qquad \text{(C.9)}\end{aligned}$$

by noting that the term with $g=1$ is equal to zero, and where

$$L^{(g)}\left(\lambda^0,f^0,\mathbf{P}_{k_1},\cdots,\mathbf{P}_{k_g}\right)\equiv\frac{1}{g!}\left[\tilde{L}^{(g)}\left(\lambda^0,f^0,\mathbf{P}_{k_1},\cdots,\mathbf{P}_{k_g}\right)+\text{all permutations of }\left(k_1,\cdots,k_g\right)\right],$$

$$\tilde{L}^{(g)}\left(\lambda^0,f^0,\mathbf{P}_{k_1},\cdots\mathbf{P}_{k_g}\right)\equiv\sum_{p=1}^{g}(-1)^{p+1}\sum_{\substack{v_1+\cdots+v_p=s,\\ m_1+\cdots+m_{p+1}=p-1,\\ 2\geq v_j\geq 1,k_j\geq 0}}\operatorname{tr}\left(S^{(m_1)}\mathcal{T}^{(v_1)}_{k_1\ldots}S^{(m_2)}\cdots S^{(m_p)}\mathcal{T}^{(v_p)}_{\ldots k_g}S^{(m_{p+1})}\right), \qquad \text{(C.10)}$$

$\mathcal{T}_k^{(1)}\equiv\lambda^0 f^{0\prime}\mathbf{P}_k'+\mathbf{P}_k f^0\lambda^{0\prime}$, and $\mathcal{T}_{k_1k_2}^{(2)}\equiv\mathbf{P}_{k_1}\mathbf{P}_{k_2}'$.

To ensure that $\mathcal{T}(\varkappa)$ can be expanded at $\varkappa=1$ in (C.9), we need the following conditions:

1. The perturbation terms must be small enough so that the quasi-log-likelihood function can be expanded. The separating distance of eigenvalue 0 (with multiplicity $T-R$) is defined as $d_{sp}\equiv NTd_{\min}^2\left(\lambda^0,f^0\right)$. Then it requires that

$$\left\|\mathcal{T}^{(1)}+\mathcal{T}^{(2)}\right\|\leq\frac{NT}{2}d_{\min}^2\left(\lambda^0,f^0\right). \qquad \text{(C.11)}$$

2. Convergence of the expansion in eqn. (C.9) in an infinite sequence with $\varkappa=1$ requires that the convergence radius is at least 1. Let $a\equiv\sqrt{NT}\left\|\vartheta_{NT}\right\|2d_{\max}\left(\lambda^0,f^0\right)$, $c\equiv\left\|\vartheta_{NT}\right\|\left[2\sqrt{NT}d_{\max}\left(\lambda^0,f^0\right)\right]^{-1}$. It is straightforward to show that

$$\left\|\mathcal{T}^{(1)}\right\|\leq a,\ \left\|\mathcal{T}^{(2)}\right\|\leq ac\text{ and }\left\|\mathcal{T}^{(s)}\right\|\equiv 0\leq ac^{s-1}\text{ for }s=3,4,\cdots. \qquad \text{(C.12)}$$

Then by (3.51) in Kato (1980, p.95), the sum of the power series for $L_{NT}(\beta)$ is convergent if $1\leq\left(\frac{2a}{d_{sp}}+c\right)^{-1}$, i.e., if

$$\frac{\left\|\vartheta_{NT}\right\|}{\sqrt{NT}}\leq r_0\left(\lambda^0,f^0\right)\equiv\left(\frac{4d_{\max}\left(\lambda^0,f^0\right)}{d_{\min}^2\left(\lambda^0,f^0\right)}+\frac{1}{2d_{\max}\left(\lambda^0,f^0\right)}\right)^{-1}. \qquad \text{(C.13)}$$

Step 2. Finite order truncation of the quasi-log-likelihood function. To conduct the asymptotic analysis, we need to truncate the expansion in (C.9) to a finite order. Noting that $\left\|S^{(g)}\right\|=\left[NTd_{\min}^2\left(\lambda^0,f^0\right)\right]^{-g}$, $\left\|\mathcal{T}^{(1)}\right\|\leq a$, and $\left\|\mathcal{T}^{(2)}\right\|\leq ac$, we have

$$\left\|S^{(h_1)}\mathcal{T}^{(v_1)}S^{(h_2)}\cdots S^{(h_p)}\mathcal{T}^{(v_p)}S^{(h_{p+1})}\right\|\leq\left[NTd_{\min}^2\left(\lambda^0,f^0\right)\right]^{-\sum h_j}\left(2\sqrt{NT}d_{\max}\left(\lambda^0,f^0\right)\right)^{2p-\sum v_j}\left\|\vartheta_{NT}\right\|^g.$$

Using $\sum_{v_1+\cdots+v_p=g, 2\geq v_j\geq 1} 1 \leq 2^g$ and $\sum_{h_1+\cdots+h_{p+1}=p-1,\ h_j\geq 0} \leq 4^p$, we have

$$
\begin{aligned}
& \frac{1}{NT} \sum_{\substack{v_1+\cdots+v_p=g,\ h_1+\cdots+h_{p+1}=p-1 \\ 2\geq v_j\geq 1,\ h_j\geq 0}} \left| \mathrm{tr}\left(S^{(h_1)} \mathcal{T}^{(v_1)} S^{(h_2)} \cdots S^{(h_p)} \mathcal{T}^{(v_p)} S^{(h_{p+1})} \right) \right| \\
\leq \;& R d_{\min}^2\left(\lambda^0, f^0\right) \left(2\sqrt{NT} d_{\max}\left(\lambda^0, f^0\right) \right)^{-g} \|\vartheta_{NT}\|^g \sum_{p=\lceil g/2 \rceil}^{g} \left(\frac{32 d_{\max}^2\left(\lambda^0, f^0\right)}{d_{\min}^2\left(\lambda^0, f^0\right)} \right)^p \\
\leq \;& \frac{R g d_{\min}^2\left(\lambda^0, f^0\right)}{2} \left\| \frac{\vartheta_{NT}}{\sqrt{NT}} \right\|^g \left(\frac{16 d_{\max}\left(\lambda^0, f^0\right)}{d_{\min}^2\left(\lambda^0, f^0\right)} \right)^g
\end{aligned}
$$

for $g \geq 3$. Recalling that $\alpha_{NT} \equiv \left\| \frac{1}{\sqrt{NT}} \vartheta_{NT} \right\| \frac{16 d_{\max}\left(\lambda^0, f^0\right)}{d_{\min}^2\left(\lambda^0, f^0\right)}$, we have

$$
\begin{aligned}
& \left| L_{NT}^0(\beta) - \frac{1}{NT} \sum_{g=2}^{G} \sum_{k_1=0}^{K} \cdots \sum_{k_g=0}^{K} \epsilon_{k_1} \cdots \epsilon_{k_g} L^{(g)}\left(\lambda^0, f^0, \mathbf{P}_{k_1}, \cdots, \mathbf{P}_{k_g}\right) \right| \\
= \;& \frac{1}{NT} \sum_{g=G+1}^{\infty} \sum_{k_1=0}^{K} \cdots \sum_{k_g=0}^{K} \epsilon_{k_1} \cdots \epsilon_{k_g} L^{(g)}\left(\lambda^0, f^0, \mathbf{P}_{k_1}, \cdots, \mathbf{P}_{k_g}\right) \\
\leq \;& \sum_{g=G+1}^{\infty} \frac{R g \alpha_{NT}^g d_{\min}^2\left(\lambda^0, f^0\right)}{2} \leq \frac{R(G+1) \alpha_{NT}^{G+1} d_{\min}^2\left(\lambda^0, f^0\right)}{2\left(1-\alpha_{NT}\right)^2}.
\end{aligned}
$$

The infinite summation is convergent given $\alpha_{NT} < 1$, which is implied by $r_0\left(\lambda^0, f^0\right) > 1$. Letting $G = 3$, we complete the proof of (i).

(ii)-(iii) Following the proof of (i) and that of Theorems 2.1 and 3.1 in Moon and Weidner (2010), we can prove (ii)-(iii) analogously. ■

D Proofs of the technical lemmas

D.1 Convergence rate

Lemma D.1 *Suppose that Assumptions 1-4 hold. Then for any $f \in \mathbb{R}^{T\times R}$ satisfying rank$(f) = R$, we have*

(i) $\sup_f \left| \frac{1}{NT} tr\left(\mathbf{P}_{(a)} M_f \mathbf{e}\right) \right| = O_P\left(\delta_{NT}^{-1}\right)$ *for any $a \in \mathbb{R}^K$ with $\|a\| = 1$;*
(ii) $\sup_f \left| \frac{1}{NT} tr\left(\mathbf{P}_{(a)} M_f \mathbf{e}_g\right) \right| = O_P\left(K^{-\gamma/d}\right)$ *for any $a \in \mathbb{R}^K$ with $\|a\| = 1$;*
(iii) $\sup_f \left| \frac{1}{NT} tr\left(\lambda^0 f^{0\prime} M_f \mathbf{u}'\right) \right| = O_P\left(\delta_{NT}^{-1} + K^{-\gamma/d}\right)$;
(iv) $\sup_f \left| \frac{1}{NT} tr\left(\mathbf{u} P_f \mathbf{u}'\right) \right| = O_P\left(\delta_{NT}^{-2} + K^{-2\gamma/d}\right)$.

Proof. (i) Using $M_f = I_T - P_f$, we have

$$
\begin{aligned}
\frac{1}{NT} \left| \mathrm{tr}\left[\mathbf{P}_{(a)} M_f \mathbf{e}'\right] \right| &\leq \left| \frac{1}{NT} \mathrm{tr}\left[\mathbf{P}_{(a)} \mathbf{e}'\right] \right| + \left| \frac{1}{NT} \mathrm{tr}\left[\mathbf{P}_{(a)} P_f \mathbf{e}'\right] \right| \\
&= \left| a' \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} p_{it} e_{it} \right| + \frac{1}{NT} \mathrm{rank}(\mathbf{P}_{(a)} P_f \mathbf{e}') \|P_f\| \left\|\mathbf{P}_{(a)}\right\| \|\mathbf{e}\| \\
&\leq \|a\| \left\| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} p_{it} e_{it} \right\| + R \frac{\left\|\mathbf{P}_{(a)}\right\|}{\sqrt{NT}} \frac{\|\mathbf{e}\|}{\sqrt{NT}} \\
&= O_P\left(K^{1/2}/(NT)^{1/2}\right) + O_P\left(\delta_{NT}^{-1}\right) = O_P\left(\delta_{NT}^{-1}\right)
\end{aligned}
$$

by Assumptions 1(iii)-(iv), 2(ii), and 5, Lemmas E.3(vi), (i), and (xi), and the fact rank$(P_f) = R$.

(ii) Using $M_f = I_T - P_f$, we have

$$\begin{aligned}
\left|\frac{1}{NT}\text{tr}\left[\mathbf{P}_{(a)} M_f \mathbf{e}_g'\right]\right| &\leq \left|\frac{1}{NT}\text{tr}\left[\mathbf{P}_{(a)}\mathbf{e}_g'\right]\right| + \left|\frac{1}{NT}\text{tr}\left[\mathbf{P}_{(a)} P_f \mathbf{e}_g'\right]\right| \\
&\leq \frac{1}{NT}\left\{a'\sum_{i=1}^{N}\sum_{t=1}^{T} p_{it}p_{it}' a\right\}^{1/2}\left\{\sum_{i=1}^{N}\sum_{t=1}^{T} e_{g,it}^2\right\}^{1/2} + \frac{R}{NT}\left\|\mathbf{P}_{(a)}\right\|\left\|P_f\right\|\left\|\mathbf{e}_g'\right\| \\
&\leq \mu_1\left(Q_{pp,NT}\right)^{1/2}\|a\|\frac{\left\|\mathbf{e}_g'\right\|_F}{\sqrt{NT}} + C\frac{\left\|\mathbf{P}_{(a)}\right\|}{\sqrt{NT}}\frac{\left\|\mathbf{e}_g'\right\|_F}{\sqrt{NT}} = O_P\left(K^{-\gamma/d}\right)
\end{aligned}$$

by Assumption 2(i), Lemma E.3(i), and the fact that $\frac{1}{NT}\left\|\mathbf{e}_g'\right\|_F^2 = \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T e_{g,it}^2 \leq \left\|g(x) - p^K(x)'\beta^0\right\|_{\infty,\bar{\omega}}^2$ $\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\left(1+\|X_{it}\|^2\right)^{\bar{\omega}} = O_P\left(K^{-2\gamma/d}\right)$ by Assumptions 3(i) and 4(i).

(iii) By Lemmas E.3 (ii) and (iv), $\frac{1}{NT}\left|\text{tr}\left(\lambda^0 f^{0\prime} M_f \mathbf{u}'\right)\right| \leq$rank$\left(\lambda^0 f^{0\prime} M_f \mathbf{u}'\right)$ $\frac{\|\lambda^0\|}{\sqrt{N}}\frac{\|f^0\|}{\sqrt{T}}\frac{\|\mathbf{e}\|+\|\mathbf{e}_g\|_F}{\sqrt{NT}} = O_P\left(\delta_{NT}^{-1} + K^{-\gamma/d}\right)$.

(iv) By Lemmas E.3 (ii) and (iv), $\frac{1}{NT}\left|\text{tr}\left(\mathbf{u}P_f\mathbf{u}'\right)\right| \leq$rank$(\mathbf{u}P_f\mathbf{u}')\frac{\|\mathbf{u}\|^2}{NT}\|P_f\| = O_P\left(\delta_{NT}^{-2} + K^{-2\gamma/d}\right)$ $= o_P(1)$. ■

Proof of Lemma A.1. Let $\mathbf{P}_{(a)} \equiv \sum_{k=1}^K a_k\mathbf{P}_k$ and $a_k \equiv \left(\beta_k^0 - \beta_k\right)/\left\|\beta^0 - \beta\right\|$. We first give a lower bound for $S_{NT}(\beta, f)$. Since $\mathbf{Y} - \sum_{k=1}^K \beta_k\mathbf{P}_k = \lambda^0 f^{0\prime} + \sum_{k=1}^K\left(\beta_k^0 - \beta_k\right)\mathbf{P}_k + \mathbf{u}$, we have

$$\begin{aligned}
S_{NT}(\beta, f) &= \frac{1}{NT}\text{tr}\left\{\left[\lambda^0 f^{0\prime} + \sum_{k=1}^K\left(\beta_k^0 - \beta_k\right)\mathbf{P}_k + \mathbf{u}\right] M_f\left[\lambda^0 f^{0\prime} + \sum_{k=1}^K\left(\beta_k^0 - \beta_k\right)\mathbf{P}_k + \mathbf{u}\right]'\right\} \\
&= S_{NT}\left(\beta^0, f^0\right) + \tilde{S}_{NT}(\beta, f) \\
&\quad + \frac{2}{NT}\text{tr}\left\{\left[\lambda^0 f^{0\prime} + \left\|\beta^0 - \beta\right\|\mathbf{P}_{(a)}\right] M_f\mathbf{u}'\right\} + \frac{1}{NT}\text{tr}\{\mathbf{u}(P_{f^0} - P_f)\mathbf{u}'\} \\
&\geq S_{NT}\left(\beta^0, f^0\right) + \tilde{S}_{NT}(\beta, f) - \left(\left\|\beta^0 - \beta\right\|\right) O_P\left(K^{-\gamma/d} + \delta_{NT}^{-1}\right) - O_P\left(K^{-\gamma/d} + \delta_{NT}^{-1}\right)
\end{aligned}$$

where $\tilde{S}_{NT}(\beta, f) \equiv \frac{1}{NT}\text{tr}\Big[\left(\lambda^{0\prime} f^0 + \left\|\beta^0 - \beta\right\|\mathbf{P}_{(a)}\right) M_f\left(\lambda^0 f^{0\prime} + \left\|\beta^0 - \beta\right\|\mathbf{P}_{(a)}\right)'\Big]$. It is obvious that

$$\begin{aligned}
\tilde{S}_{NT}(\beta, f) &\geq \min_f \tilde{S}_{NT}(\beta, f) = \left\|\beta^0 - \beta\right\|^2\sum_{i=2R+1}^N \mu_i\left(Q_{pp,NT}^{(a)}\right) \\
&\geq \left\|\beta^0 - \beta\right\|^2\min_{\|a\|=1, a\in\mathbb{R}^K}\sum_{i=2R+1}^N \mu_i\left[Q_{pp,NT}^{(a)}\right] = b\left\|\beta^0 - \beta\right\|^2
\end{aligned}$$

by Assumption 2(iii). It follows that $S_{NT}(\beta, f) \geq S_{NT}\left(\beta^0, f^0\right) + b\left\|\beta^0 - \beta\right\|^2 - o_P\left(\left\|\beta^0 - \beta\right\|\right) - o_P(1)$. Since $S_{NT}(\hat{\beta}, \hat{f}) = \min_{\beta, f} S_{NT}(\beta, f) \leq S_{NT}\left(\beta^0, f^0\right)$, we have

$$b\left\|\beta^0 - \hat{\beta}\right\|^2 \leq \left\|\beta^0 - \beta\right\| O_P\left(K^{-\gamma/d} + \delta_{NT}^{-1}\right) + O_P\left(K^{-\gamma/d} + \delta_{NT}^{-1}\right)$$

Then we get $||\beta^0 - \hat{\beta}|| = O_P(K^{-\gamma/2d} + \delta_{NT}^{-1/2}) = o_P(1)$. ■

Proof of Lemma A.2. Recall $V_K(x) \equiv p^K(x)'\tilde{W}^{-1}\tilde{\Omega}\tilde{W}^{-1}p^K(x)$ and $v_x^K \equiv V_K^{-1/2}(x)\tilde{W}^{-1}p^K(x)$. By

Cauchy-Schwarz inequality, we have

$$\begin{aligned}
|d_{it}| &= V_K^{-1/2}(x)\left|p^K(x)'\tilde{W}^{-1}\tilde{Z}_{it}\right| \\
&\leq \frac{\left\{p^K(x)'\tilde{W}^{-1}p^K(x)\right\}^{1/2}\left\{\tilde{Z}_{it}'\tilde{W}^{-1}\tilde{Z}_{it}\right\}^{1/2}}{\left\{p^K(x)'\tilde{W}^{-1}\tilde{\Omega}\tilde{W}^{-1}p^K(x)\right\}^{1/2}} \leq \frac{\left\|p^K(x)\right\|\mu_1\left(\tilde{W}^{-1}\right)\left\|\tilde{Z}_{it}\right\|}{\left\|p^K(x)\right\|\mu_{\min}^{1/2}\left(\tilde{\Omega}\right)\mu_1\left(\tilde{W}^{-1}\right)} \\
&= \mu_{\min}^{-1/2}\left(\tilde{\Omega}\right)\left\|\tilde{Z}_{it}\right\|.
\end{aligned}$$

Recall that $\tilde{Z}_{it} = p_{it} - \frac{1}{N}\sum_{j=1}^N \alpha_{ij}E_{\mathcal{D}}(p_{jt}) - \frac{1}{T}\sum_{s=1}^T \eta_{ts}E_{\mathcal{D}}(p_{is}) + \frac{1}{NT}\sum_{j=1}^N\sum_{s=1}^T \alpha_{ij}\eta_{ts}E_{\mathcal{D}}(p_{js}) \equiv p_{it} + \zeta_{it}$. Note that ζ_{it} is a $K\times 1$ $\mathcal{D}$-measurable vector, and

$$\begin{aligned}
\|\zeta_{it}\| \leq\; & \left\|\lambda_i^0\right\|\frac{\varsigma_N^{-1}}{N}\sum_{j=1}^N\left\|\lambda_j^0\right\|\left\|E_{\mathcal{D}}(p_{jt})\right\| + \left\|f_t^0\right\|\frac{\varsigma_T^{-1}}{T}\sum_{s=1}^T\left\|f_s^0\right\|\left\|E_{\mathcal{D}}(p_{is})\right\| \\
& + \left\|\lambda_i^0\right\|\left\|f_t^0\right\|\frac{\varsigma_N^{-1}\varsigma_T^{-1}}{NT}\sum_{j=1}^N\sum_{s=1}^T\left\|\lambda_j^0\right\|\left\|f_s^0\right\|\left\|E_{\mathcal{D}}(p_{js})\right\|
\end{aligned}$$

where we use the fact that $|\alpha_{ij}| \leq \varsigma_N^{-1}\left\|\lambda_i^0\right\|\left\|\lambda_j^0\right\|$ and $|\eta_{ts}| \leq \varsigma_T^{-1}\left\|f_t^0\right\|\left\|f_s^0\right\|$. For (i), noting that $\left\|\tilde{Z}_{it}\right\|^4 \leq (\|p_{it}\| + \|\zeta_{it}\|)^4 \leq 2^3\left(\|p_{it}\|^4 + \|\zeta_{it}\|^4\right)$ and $\mu_{\min}^{-2}\left(\tilde{\Omega}\right) = O_P(1)$, we have

$$\begin{aligned}
\frac{1}{NT}\sum_{t=1}^T\sum_{i=1}^N\left\|d_{it}^4\right\|_{2,\mathcal{D}}^2 &\leq \frac{2^3\mu_{\min}^{-2}\left(\tilde{\Omega}\right)}{NT}\sum_{t=1}^T\sum_{i=1}^N\left\|\|p_{it}\|^4 + \|\zeta_{it}\|^4\right\|_{2,\mathcal{D}}^2 \\
&\leq 2^4\mu_{\min}^{-2}\left(\tilde{\Omega}\right)\left\{\frac{1}{NT}\sum_{t=1}^T\sum_{i=1}^N\left[E_{\mathcal{D}}\left(\|p_{it}\|^8\right) + \|\zeta_{it}\|^8\right]\right\} = O_P\left(K^4\right),
\end{aligned}$$

where we use the fact that $\frac{1}{NT}\sum_{t=1}^T\sum_{i=1}^N\|\zeta_{it}\|^8 = O_P\left(K^4\right)$. To see this, using $[(a+b+c)/3]^8 \leq \left(a^8+b^8+c^8\right)/3$, we have $\frac{1}{NT}\sum_{t=1}^T\sum_{i=1}^N\|\zeta_{it}\|^8 \leq \zeta_{NT}(4,a) + \zeta_{NT}(4,b) + \zeta_{NT}(4,c)$, where

$$\begin{aligned}
\zeta_{NT}(4,a) &\equiv \frac{3^7}{NT}\sum_{t=1}^T\sum_{i=1}^N\left(\left\|\lambda_i^0\right\|\frac{\varsigma_N^{-1}}{N}\sum_{j=1}^N\left\|\lambda_j^0\right\|\left\|E_{\mathcal{D}}(p_{jt})\right\|\right)^8, \\
\zeta_{NT}(4,b) &\equiv \frac{3^7}{NT}\sum_{t=1}^T\sum_{i=1}^N\left(\left\|f_t^0\right\|\frac{\varsigma_T^{-1}}{T}\sum_{s=1}^T\left\|f_s^0\right\|\left\|E_{\mathcal{D}}(p_{is})\right\|\right)^8, \text{ and} \\
\zeta_{NT}(4,c) &\equiv \frac{3^7}{NT}\sum_{t=1}^T\sum_{i=1}^N\left(\left\|\lambda_i^0\right\|\left\|f_t^0\right\|\frac{\varsigma_N^{-1}\varsigma_T^{-1}}{NT}\sum_{j=1}^N\sum_{s=1}^T\left\|\lambda_j^0\right\|\left\|f_s^0\right\|\left\|E_{\mathcal{D}}(p_{js})\right\|\right)^4.
\end{aligned}$$

For $\zeta_{NT}(4,a)$, by Cauchy-Schwarz inequality

$$\begin{aligned}
\zeta_{NT}(4,a) &\leq 3^7\varsigma_N^{-8}\left\{\frac{1}{N}\sum_{i=1}^N\left\|\lambda_i^0\right\|^8\right\}\left\{\frac{1}{N}\sum_{j=1}^N\left\|\lambda_j^0\right\|^2\right\}^4\frac{1}{T}\sum_{t=1}^T\left\{E_{\mathcal{D}}\left(\frac{1}{N}\sum_{j=1}^N\|p_{jt}\|^2\right)\right\}^4 \\
&= O_P(1)\,O_P(1)\,O_P\left(K^4\right) = O_P\left(K^4\right).
\end{aligned}$$

Similarly, we can show that $\zeta_{NT}(4,b) = O_P\left(K^4\right)$ and $\zeta_{NT}(4,c) = O_P\left(K^4\right)$.

For (ii), following the study of (i) and Jensen inequality, we have

$$\begin{aligned}
\frac{1}{N^2T}\sum_{t=1}^{T}\left(\sum_{i=1}^{N}\left\|d_{it}^2\right\|_{2,\mathcal{D}}^2\right)^2 &\leq \mu_{\min}^{-2}\left(\tilde{\Omega}\right)\frac{1}{N^2T}\sum_{t=1}^{T}\left(\sum_{i=1}^{N}\left\|\left\|\tilde{Z}_{it}\right\|^2\right\|_{2,\mathcal{D}}^2\right)^2 \\
&\leq \frac{4\mu_{\min}^{-2}\left(\tilde{\Omega}\right)}{N^2T}\sum_{t=1}^{T}\left(\sum_{i=1}^{N}\left\|\left\|p_{it}\right\|^2\right\|_{2,\mathcal{D}}^2+\sum_{i=1}^{N}\left\|\zeta_{it}\right\|^4\right)^2 \\
&\leq 8\mu_{\min}^{-2}\left(\tilde{\Omega}\right)\frac{1}{T}\sum_{t=1}^{T}\left\{E_{\mathcal{D}}\left[\left(\frac{1}{N}\sum_{i=1}^{N}\left\|p_{it}\right\|^4\right)^2\right]+\frac{1}{N}\sum_{i=1}^{N}\left\|\zeta_{it}\right\|^8\right\} \\
&= O_P(1)\,O_P\left(K^4\right)=O_P\left(K^4\right). \blacksquare
\end{aligned}$$

D.2 Asymptotic normality for the sieve estimator

Proof of Lemma A.3. (i) Let $\bar{p}_{it} \equiv -\frac{1}{N}\sum_{j=1}^{N}\alpha_{ij}p_{jt}-\frac{1}{T}\sum_{s=1}^{T}\eta_{ts}p_{is}+\frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\alpha_{ij}\eta_{ts}p_{js}$.Then $\tilde{Z}_{it}=p_{it}+E_{\mathcal{D}}\left[\bar{p}_{it}\right]$. We have

$$\begin{aligned}
\tilde{W}_{NT}-\tilde{W} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\tilde{Z}_{it}\tilde{Z}_{it}'-E_{\mathcal{D}}\left(\tilde{Z}_{it}\tilde{Z}_{it}'\right)\right] \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\left[p_{it}p_{it}'-E_{\mathcal{D}}\left(p_{it}p_{it}'\right)\right]+\left[\left(p_{it}-E_{\mathcal{D}}\left(p_{it}\right)\right)E_{\mathcal{D}}\left(\bar{p}_{it}\right)'\right]+\left[E_{\mathcal{D}}\left(\bar{p}_{it}\right)\left(p_{it}'-E_{\mathcal{D}}\left(p_{it}'\right)\right)\right]\right\} \\
&\equiv D\tilde{W}_{1NT}+D\tilde{W}_{2NT}+D\tilde{W}_{3NT}, \text{ say.}
\end{aligned}$$

For $D\tilde{W}_{1NT}$, we have

$$\begin{aligned}
&E_{\mathcal{D}}\left[\left\|D\tilde{W}_{1NT}\right\|_F^2\right] \\
=\;& \frac{1}{N^2}\sum_{l=1}^{K}\sum_{k=1}^{K}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{1\leq t\neq s\leq T}\operatorname{Cov}_{\mathcal{D}}\left(p_{it,l}p_{it,k},p_{is,k}p_{is,l}\right)+\frac{1}{N^2T^2}\sum_{l=1}^{K}\sum_{k=1}^{K}\sum_{i=1}^{N}\sum_{t=1}^{T}\operatorname{Var}_{\mathcal{D}}\left(p_{it,l}p_{it,k}\right) \\
\leq\;& \frac{8}{N^2T^2}\sum_{l=1}^{K}\sum_{k=1}^{K}\sum_{i=1}^{N}\sum_{1\leq t\neq s\leq T}\left\|p_{it,l}\right\|_{8+4\delta,\mathcal{D}}\left\|p_{it,k}\right\|_{8+4\delta,\mathcal{D}}\left\|p_{is,k}\right\|_{8+4\delta,\mathcal{D}}\left\|p_{is,l}\right\|_{8+4\delta,\mathcal{D}}\alpha_{\mathcal{D}}^{\frac{1+\delta}{2+\delta}}(t-s)+O_P\left(\frac{K^2}{NT}\right) \\
=\;& O_P\left(\frac{K^2}{NT}\right)+O_P\left(\frac{K^2}{NT}\right)=O_P\left(\frac{K^2}{NT}\right).
\end{aligned}$$

Then $\left\|D\tilde{W}_{1NT}\right\|_F=O_P(K/\sqrt{NT})=o_P(1)$. Similarly, we can show that $D\tilde{W}_{sNT}\equiv O_P\left(K/\sqrt{NT}\right)$ for $s=2,3$. Then (i) follows.

(ii) Noting that $Z_{it}=\tilde{Z}_{it}+(\bar{p}_{it}-E_{\mathcal{D}}\left[\bar{p}_{it}\right])$, we can decompose $W_{NT}-\tilde{W}_{NT}=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[Z_{it}Z_{it}'-\tilde{Z}_{it}\tilde{Z}_{it}']$ as follows

$$\begin{aligned}
W_{NT}-\tilde{W}_{NT} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}p_{it}\left(\bar{p}_{it}-E_{\mathcal{D}}\left[\bar{p}_{it}\right]\right)'+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\bar{p}_{it}-E_{\mathcal{D}}\left[\bar{p}_{it}\right]\right)p_{it}' \\
&\quad+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}E_{\mathcal{D}}\left[\bar{p}_{it}\right]\left(\bar{p}_{it}-E_{\mathcal{D}}\left[\bar{p}_{it}\right]\right)'+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\bar{p}_{it}-E_{\mathcal{D}}\left[\bar{p}_{it}\right]\right)E_{\mathcal{D}}\left[\bar{p}_{it}\right]' \\
&\quad+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\bar{p}_{it}-E_{\mathcal{D}}\left[\bar{p}_{it}\right]\right)\left(\bar{p}_{it}-E_{\mathcal{D}}\left[\bar{p}_{it}\right]\right)' \\
&\equiv DW_{1NT}+DW_{2NT}+DW_{3NT}+DW_{4NT}+DW_{5NT}, \text{ say.}
\end{aligned}$$

It is easy to see that $\left\|W_{NT}-\tilde{W}_{NT}\right\|_F \leq \sum_{s=1}^{5}\|DW_{sNT}\|_F = 2\|DW_{1NT}\|_F + 2\|DW_{3NT}\|_F + \|DW_{5NT}\|_F$.

For DW_{1NT}, using the expression for $\bar{p}_{it}$ and by Minkowski inequality, we have

$$\begin{aligned}
\|DW_{1NT}\|_F &\leq \left\|\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\alpha_{ij}p_{it}\left(p_{jt}-E_{\mathcal{D}}\left[p_{jt}\right]\right)'\right\|_F + \left\|\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\eta_{ts}p_{it}\left(p_{is}-E_{\mathcal{D}}\left[p_{is}\right]\right)'\right\|_F \\
&\quad + \left\|\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\alpha_{ij}\eta_{ts}p_{it}\left(p_{js}-E_{\mathcal{D}}\left[p_{js}\right]\right)'\right\|_F \\
&\equiv DW_{1NT,1}+DW_{1NT,2}+DW_{1NT,3}, \text{ say.}
\end{aligned}$$

For $DW_{1NT,1}$, we have

$$\begin{aligned}
DW_{1NT,1} &= \left\|\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{t=1}^{T}\alpha_{ii}p_{it}\left[p_{it}-E_{\mathcal{D}}\left(p_{it}\right)\right]'\right\|_F + \left\|\frac{1}{N^2T}\sum_{1\leq i\neq j\leq N}\sum_{t=1}^{T}\alpha_{ij}p_{it}\left(p_{jt}-E_{\mathcal{D}}\left[p_{jt}\right]\right)'\right\|_F \\
&= O_P\left(\frac{K}{\sqrt{N^3T}}\right)+O_P\left(\frac{K}{\sqrt{NT}}\right)=O_P\left(\frac{K}{\sqrt{NT}}\right)
\end{aligned}$$

by Chebyshev's inequality. Similarly, we can show that $DW_{1s}=O_P(K/\sqrt{NT})$ for $s=2,3$. Hence $\|DW_{1NT}\|_F = O_P(K/\sqrt{NT})$.

Analogously, we can show that $\|DW_{sNT}\|_F = O_P(K/\sqrt{NT})$ for $s=3,5$. Thus (ii) follow. ■

Proof of Lemma A.4. Let $\Psi_{NT} \equiv \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}v_x^{K\prime}\{(Z_{it}-\tilde{Z}_{it})e_{it} - E_{\mathcal{D}}[(Z_{it}-\tilde{Z}_{it})e_{it}]\}$. Let $p_{is}^c \equiv p_{is}-E_{\mathcal{D}}\left(p_{is}\right)$. We first make the following decomposition:

$$\begin{aligned}
\Psi_{NT} &= -\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}v_x^{K\prime}\left\{\frac{1}{N}\sum_{j=1}^{N}\alpha_{ij}p_{jt}^c\right\}e_{it} \\
&\quad -\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}v_x^{K\prime}\left\{\frac{1}{T}\sum_{s=1}^{T}\eta_{ts}\left[p_{is}^c e_{it}-E_{\mathcal{D}}\left(p_{is}^c e_{it}\right)\right]\right\} \\
&\quad +\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}v_x^{K\prime}\left\{\frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\alpha_{ij}\eta_{ts}\left[p_{js}^c e_{it}-E_{\mathcal{D}}\left(p_{js}^c e_{it}\right)\right]\right\} \\
&\equiv -\Psi_{NT,1}-\Psi_{NT,2}+\Psi_{NT,3}, \text{ say.}
\end{aligned}$$

We want to show that: (i) $\Psi_{NT,1}=o_P(1)$, (ii) $\Psi_{NT,2}=o_P(1)$, and (iii) $\Psi_{NT,3}=o_P(1)$.

First, we consider (i). Note that $E_{\mathcal{D}}(\Psi_{NT,1})=0$ and

$$\begin{aligned}
E_{\mathcal{D}}\left(\Psi_{NT,1}^2\right) &= \frac{1}{N^3T}\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\sum_{j_1=1}^{N}\sum_{j_2=1}^{N}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\alpha_{i_1j_1}\alpha_{i_2j_2}v_x^{K\prime}E_{\mathcal{D}}\left(p_{j_1t_1}^c p_{j_2t_2}^{c\prime}e_{i_1t_1}e_{i_2t_2}\right)v_x^K \\
&= \frac{1}{N^3T}\sum_{j=1}^{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\alpha_{ij}^2 v_x^{K\prime}E_{\mathcal{D}}\left(p_{jt}^c p_{jt}^{c\prime}e_{it}^2\right)v_x^K \\
&\leq \frac{\zeta_N^{-2}\left\|v_x^K\right\|^2}{N}\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\left\|\lambda_i^0\right\|\left\|\lambda_j^0\right\|\left\|E_{\mathcal{D}}\left(p_{jt}^c p_{jt}^{c\prime}e_{it}^2\right)\right\|_F \\
&\leq \frac{\zeta_N^{-2}\left\|v_x^K\right\|^2}{N}\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\left\|\lambda_i^0\right\|\left\|\lambda_j^0\right\|\sum_{t=1}^{T}\left\|p_{jt}^c\right\|_{2,\mathcal{D}}^2\left\|e_{it}^2\right\|_{2,\mathcal{D}} = O_P\left(K/N\right).
\end{aligned}$$

It follows that $\Pi_{1NT,121} = O_P(K^{1/2}/N^{1/2}) = o_P(1)$ by conditional Chebyshev inequality.

Next, we consider (ii). We decompose $\Psi_{NT,2}$ as follows

$$\begin{aligned}\Psi_{NT,2} &= \frac{1}{\sqrt{NT^3}}\sum_{i=1}^{N}\sum_{1\leq s\leq t\leq T}\eta_{ts}v_x^{K\prime}p_{is}^c e_{it} + \frac{1}{\sqrt{NT^3}}\sum_{i=1}^{N}\sum_{1\leq t<s\leq T}\eta_{ts}v_x^{K\prime}\left[p_{is}^c e_{it} - E_{\mathcal{D}}\left(p_{is}^c e_{it}\right)\right] \\ &\equiv \Psi_{NT,21} + \Psi_{NT,22}, \text{ say,}\end{aligned}$$

where we use the fact $E_{\mathcal{D}}\left(p_{is}^c e_{it}\right) = E_{\mathcal{D}}\left(p_{is}e_{it}\right) = 0$ for $s \leq t$ in the first term. Following the study of $\Psi_{NT,1}$, we can show that $\Psi_{NT,21} = O_P\left(K^{1/2}/T^{1/2}\right) = o_P(1)$ by conditional Chebyshev inequality. We are left to show that $\Psi_{NT,22} = o_P(1)$. By construction, $E_{\mathcal{D}}[\Psi_{NT,22}] = 0$. By Assumption 5$(iii)$ and conditional Jensen inequality,

$$\begin{aligned}E_{\mathcal{D}}[\Psi_{NT,22}^2] &= \mathrm{Var}_{\mathcal{D}}\left(\Psi_{NT,22}\right) = \frac{1}{NT^3}\sum_{i=1}^{N}\mathrm{Var}_{\mathcal{D}}\left(\sum_{1\leq t<s\leq T}\eta_{ts}E_{\mathcal{D}}\left(e_{it}v_x^{K\prime}p_{is}\right)\right) \\ &\leq \frac{1}{NT^3}\sum_{i=1}^{N}\sum_{1\leq t_1<t_2\leq T}\sum_{1\leq t_3<t_4\leq T}\eta_{t_1t_2}\eta_{t_3t_4}v_x^{K\prime}E_{\mathcal{D}}(e_{it_1}p_{it_2}^c e_{it_3}p_{it_4}^{c\prime})v_x^K \equiv E\Psi_{NT,22} \quad \text{(D.1)}\end{aligned}$$

There are three cases according to the number of distinct time indices in the set $S = \{t_1, t_2, t_3, t_4\}$: (a) $\#S = 4$, (b) $\#S = 3$, and (c) $\#S = 2$. We use $E\Psi_{NT,22a}$, $E\Psi_{NT,22b}$ and $E\Psi_{NT,22c}$ to denote the summation when the time indices in (D.1) are restricted to these three cases, respectively. Then $E\Psi_{NT,22} = E\Psi_{NT,22a} + E\Psi_{NT,22b} + E\Psi_{NT,22c}$. It suffices to prove $\Psi_{NT,22} = o_P(1)$ by showing that $E\Psi_{NT,22s} = o_P(1)$ for $s = a, b, c$.

We dispense with the easiest term first. In case (c), we must have $t_1 = t_3$ and $t_2 = t_4$. By direct moment calculations, we can readily show that $E\Psi_{NT,22c} = O_P(K/T)$.

Now we consider $E\Psi_{NT,22a}$. There are three subcases: $(a1)$ $t_1 < t_2 < t_3 < t_4$ or $t_3 < t_4 < t_1 < t_2$; $(a2)$ $t_1 < t_3 < t_2 < t_4$ or $t_3 < t_1 < t_4 < t_2$; $(a3)$ $t_1 < t_3 < t_4 < t_2$ or $t_3 < t_1 < t_2 < t_4$. Let $E\Psi_{NT,22a1}$, $E\Psi_{NT,22a2}$, and $E\Psi_{NT,22a3}$ denote the corresponding summation when the time indices are are restricted to subcases $(a1)$, $(a2)$, and $(a3)$, respectively, in the definition of $E\Psi_{NT,22a}$. We only prove that $E\Psi_{NT,22a1} = o_P(1)$ as the proofs of $E\Psi_{NT,22a2} = o_P(1)$ and $E\Psi_{NT,22a3} = o_P(1)$ are similar. For subcase $(a1)$, by the symmetry of $(t_1, t_2) \longleftrightarrow (t_3, t_4)$, we have

$$E\Psi_{NT,22a1} = \frac{2}{NT^3}\sum_{i=1}^{N}\sum_{1\leq t_1<t_2<t_3<t_4\leq T}\eta_{t_1t_2}\eta_{t_3t_4}v_x^{K\prime}E_{\mathcal{D}}\left(p_{it_2}^c e_{it_1}p_{it_4}^{c\prime}e_{it_3}\right)v_x^K$$

Let $d_l = t_{l+1} - t_l$, for $l = 1, 2, 3$. Let $d_{l_{\max}}$ be the largest increment, i.e., $t_{l_{\max}} - t_{l_{\max}-1} = \max_{s=2,3,4}(t_s - t_{s-1})$. We consider two subsubcases for $(a1)$: $(a11)$ $l_{\max} = 2$ or $l_{\max} = 4$; $(a12)$ $l_{\max} = 3$. Let $E\Psi_{NT,22a11}$ and $E\Psi_{NT,22a12}$ denote the corresponding summation when the time indices restricted to subsubcases $(a11)$ and $(a12)$, respectively. For subsubcase $(a11)$, without loss of generality (wlog) assume $l_{\max} = 2$. Let $\varphi_{is,q}^c \equiv K^{-1/q}\left\|p_{it}^c\right\|_{q,\mathcal{D}}$ for $0 < q \leq 8 + 4\delta$. By the conditional Davydov inequality (see Lemma E.1 in the supplementary appendix) and Hölder inequality, we have

$$\begin{aligned}\left|E_{\mathcal{D}}\left(e_{it_1}v_x^{K\prime}p_{it_2}^c e_{it_3}p_{it_4}^{c\prime}v_x^K\right)\right| &\leq 8\left\|e_{it_1}\right\|_{8+4\delta,\mathcal{D}}\left\|v_x^{K\prime}p_{it_2}^c e_{it_3}p_{it_4}^{c\prime}v_x^K\right\|_{(8+4\delta)/3,\mathcal{D}}\alpha_{\mathcal{D}}^{\frac{1+\delta}{2+\delta}}(t_2 - t_1) \\ &\leq 8K\left\|v_x^K\right\|^2\left\|e_{it_1}\right\|_{8+4\delta,\mathcal{D}}\varphi_{it_2,8+4\delta}^c\left\|e_{it_3}\right\|_{8+4\delta,\mathcal{D}}\varphi_{it_4,8+4\delta}^c\alpha_{\mathcal{D}}^{\frac{1+\delta}{2+\delta}}(t_2 - t_1),\end{aligned}$$

and

$$\begin{aligned}
& \left|E_{\mathcal{D}}\left(e_{it_1} v_x^{K\prime} p_{it_2}^c e_{it_3} p_{it_4}^{c\prime} v_x^K\right)\right| \left\|f_{t_1}^0\right\| \left\|f_{t_2}^0\right\| \left\|f_{t_3}^0\right\| \left\|f_{t_4}^0\right\| \\
\leq\ & 8K \left\|v_x^K\right\|^2 \left(\left\|f_{t_1}^0\right\| \left\|e_{it_1}\right\|_{8+4\delta,\mathcal{D}}\right) \left(\left\|f_{t_2}^0\right\| \varphi_{it_2,8+4\delta}^c\right) \left(\left\|f_{t_3}^0\right\| \left\|e_{it_3}\right\|_{8+4\delta,\mathcal{D}}\right) \left(\left\|f_{t_4}^0\right\| \varphi_{it_4,8+4\delta}^c\right) \alpha_{\mathcal{D}}^{\frac{1+\delta}{2+\delta}}\left(t_2-t_1\right) \\
\leq\ & 2K \left\|v_x^K\right\|^2 \left(C_{1,it_1,e}+C_{1,it_2,p}+C_{1,it_3,e}+C_{1,it_4,p}\right) \alpha_{\mathcal{D}}^{\frac{1+\delta}{2+\delta}}\left(t_2-t_1\right),
\end{aligned}$$

where $C_{1,it,e} = \left\|f_t^0\right\|^4 \left\|e_{it}\right\|_{8+4\delta,\mathcal{D}}^4$ and $C_{1,it,p} = \left\|f_t^0\right\|^4 (\varphi_{it,8+4\delta}^c)^4$. It follows that

$$\begin{aligned}
E\Psi_{NT,22a11} &\leq \frac{2\varsigma_T^{-2}}{NT^3} \sum_{i=1}^N \sum_{\substack{1\leq t_1<t_2<t_3<t_4\leq T, \\ l_{\max}=2 \text{ or } l_{\max}=4}} \left|E_{\mathcal{D}}\left(e_{it_1} v_x^{K\prime} p_{it_2}^c e_{it_3} p_{it_4}^{c\prime} v_x^K\right)\right| \left\|f_{t_1}^0\right\| \left\|f_{t_2}^0\right\| \left\|f_{t_3}^0\right\| \left\|f_{t_4}^0\right\| \\
&\leq \frac{4\varsigma_T^{-2}K}{NT^3} \sum_{i=1}^N \sum_{1\leq t_1<t_2<t_3<t_4\leq T, l_{\max}=2} \left(C_{1,it_1,e}+C_{1,it_2,p}+C_{1,it_3,e}+C_{1,it_4,p}\right) \alpha_{\mathcal{D}}^{\frac{1+\delta}{2+\delta}}\left(t_2-t_1\right) \\
&\leq \frac{CK}{NT^3} \left\{\sum_{i=1}^N \sum_{t=1}^T \left(C_{1,it,e}+C_{1,it,p}\right)\right\} \left\{\sum_{m=1}^T m^2 \alpha_{\mathcal{D}}^{(1+\delta)/(2+\delta)}(m)\right\} \\
&\leq \frac{CK}{NT^3} \sqrt{N} \left\{\sum_{t=1}^T \left\|f_t^0\right\|^4\right\}^{1/2} \left\{\left[\sum_{t=1}^T \sum_{i=1}^N \left\|e_{it}\right\|_{8+4\delta,\mathcal{D}}^8\right]^{1/2} + \left[\sum_{t=1}^T \sum_{i=1}^N (\varphi_{it,8+4\delta}^c)^8\right]^{1/2}\right\} \\
&= \frac{CK}{NT^3} \sqrt{N} O_P\left(\sqrt{T}\right) O_P\left(\sqrt{NT}\right) = O_P\left(\frac{K}{T^2}\right).
\end{aligned}$$

For subsubcase $(a12)$, we have

$$\begin{aligned}
E\Psi_{NT,22a12} &\leq \frac{1}{NT^3} \sum_{i=1}^N \sum_{1\leq t_1<t_2<t_3<t_4\leq T,\ d_2>d_1\geq d_3} \left\|f_{t_1}^0\right\| \left\|f_{t_2}^0\right\| \left\|f_{t_3}^0\right\| \left\|f_{t_4}^0\right\| \left|E_{\mathcal{D}}\left(e_{it_1} v_x^{K\prime} p_{it_2}^c e_{it_3} p_{it_4}^{c\prime} v_x^K\right)\right| \\
&\quad + \frac{1}{NT^3} \sum_{i=1}^N \sum_{1\leq t_1<t_2<t_3<t_4\leq T,\ d_2>d_3>d_1} \left\|f_{t_1}^0\right\| \left\|f_{t_2}^0\right\| \left\|f_{t_3}^0\right\| \left\|f_{t_4}^0\right\| \left|E_{\mathcal{D}}\left(e_{it_1} v_x^{K\prime} p_{it_2}^c e_{it_3} p_{it_4}^{c\prime} v_x^K\right)\right| \\
&\equiv E\Psi_{NT,22a12}(1) + E\Psi_{NT,22a12}(2), \text{ say.}
\end{aligned}$$

By the conditional Davydov inequality, Hölder and Jensen inequalities, we have

$$\begin{aligned}
\left|E_{\mathcal{D}}\left(e_{it_1} v_x^{K\prime} p_{it_2}^c e_{it_3} p_{it_4}^{c\prime} v_x^K\right)\right| &\leq 8 \left\|e_{it_1}\right\|_{8+4\delta,\mathcal{D}} \left\|v_x^{K\prime} p_{it_2}^c e_{it_3} p_{it_4}^{c\prime} v_x^K\right\|_{(8+4\delta)/3,\mathcal{D}} \alpha_{\mathcal{D}}^{\frac{1+\delta}{2+\delta}}\left(t_2-t_1\right) \\
&\leq 8K \left\|v_x^K\right\|^2 \left\|e_{it_1}\right\|_{8+4\delta,\mathcal{D}} \varphi_{it_2,8+4\delta}^c \left\|e_{it_3}\right\|_{8+4\delta,\mathcal{D}} \varphi_{it_4,8+4\delta}^c \alpha_{\mathcal{D}}^{\frac{1+\delta}{2+\delta}}\left(t_2-t_1\right)
\end{aligned}$$

and $\left|E_{\mathcal{D}}\left(e_{it_1} v_x^{K\prime} p_{it_2}^c e_{it_3} p_{it_4}^{c\prime} v_x^K\right)\right| \left\|f_{t_1}^0\right\| \left\|f_{t_2}^0\right\| \left\|f_{t_3}^0\right\| \left\|f_{t_4}^0\right\| \leq 2K \left\|v_x^K\right\|^2 \left(C_{1,it_1,e}+C_{1,it_2,p}+C_{1,it_3,e}+C_{1,it_4,p}\right)$ $\times\alpha_{\mathcal{D}}^{\frac{1+\delta}{2+\delta}}\left(t_2-t_1\right)$. It is easy to verify that $\sum_{i=1}^N \sum_{t=1}^T (C_{1,it,e}+C_{1,it,p}) = O_P(NT)$. It follows that

$$\begin{aligned}
E\Psi_{NT,22a12}(1) &\leq \frac{CK}{NT^3} \sum_{i=1}^N \sum_{1\leq t_1<t_2<t_3<t_4\leq T,\ d_2>d_1\geq d_3} \left(C_{1,it_1,e}+C_{1,it_2,p}+C_{1,it_3,e}+C_{1,it_4,p}\right) \alpha_{\mathcal{D}}^{\frac{1+\delta}{2+\delta}}\left(t_2-t_1\right) \\
&= \frac{CK}{NT^3} \sum_{i=1}^N \left\{\sum_{t_1=1}^{T-3} C_{1,it_1,e} \sum_{d_2=2}^{T-3-t_1} \sum_{d_1=1}^{d_2-1} \alpha_{\mathcal{D}}^{\frac{1+\delta}{2+\delta}}(d_1)\, d_1 + \sum_{t_2=2}^{T-2} C_{1,it_2,p} \sum_{d_2=2}^{T-3-t_2} \sum_{d_1=1}^{d_2-1} \alpha_{\mathcal{D}}^{\frac{1+\delta}{2+\delta}}(d_1)\, d_1 \right. \\
&\quad \left. + \sum_{t_3=3}^{T-1} C_{1,it_3,e} \sum_{d_2=2}^{t_3-1} \sum_{d_1=1}^{d_2-1} \alpha_{\mathcal{D}}^{\frac{1+\delta}{2+\delta}}(d_1)\, d_1 + \sum_{t_4=4}^{T} C_{1,it_4,p} \sum_{d_2=2}^{t_4-2} \sum_{d_1=1}^{d_2-1} \alpha_{\mathcal{D}}^{\frac{1+\delta}{2+\delta}}(d_1)\, d_1 \right\}
\end{aligned}$$

$$
\begin{aligned}
&\leq \left\{ T\sum_{m=1}^{T-1}\left(1-\frac{m}{T}\right) m\alpha_{\mathcal{D}}^{\frac{1+\delta}{2+\delta}}(m)\right\} \frac{CK}{NT^3}\sum_{i=1}^{N}\sum_{t=1}^{T}(C_{1,it,p}+C_{1,it,e}) \\
&= O_P(T)\frac{CK}{NT^3}O_P(NT)=O_P(K/T)=o_P(1).
\end{aligned}
$$

Similarly, we can show that $E\Psi_{NT,22a12}(2)=O_P(K/T)=o_P(1)$. Consequently $E\Psi_{NT,22a12}=O_P(K/T)$. Thus $E\Psi_{NT,22a1}=o_P(1)$. As remarked early on, one analogously show that $E\Psi_{NT,22as}=o_P(1)$ for $s=2,3$. Consequently, we have $E\Psi_{NT,22a}=o_P(1)$.

Now we study $E\Psi_{NT,22b}$. We consider two subcases: $(b1)$ $t_1=t_3$ or $t_2=t_4$ and $(b2)$ $t_1=t_4$ or $t_2=t_3$. Let $E\Psi_{NT,22b1}$ and $E\Psi_{NT,22b2}$ denote the corresponding summation when the time indices are restricted to subcases $(b1)$ and $(b2)$, respectively. For subcase $(b1)$, wlog we assume $t_1=t_3$. By the conditional Davydov inequality,

$$
\left|E_{\mathcal{D}}\left(e_{it_1}^2 v_x^{K\prime} p_{it_2}^c p_{it_4}^{c\prime} v_x^K\right)\right| \leq \begin{cases} 8\left\|e_{it_1}^2 v_x^{K\prime} p_{it_2}^c\right\|_{(8+4\delta)/3,\mathcal{D}} \left\|p_{it_4}^{c\prime} v_x^K\right\|_{8+4\delta,\mathcal{D}} \alpha_{\mathcal{D}}^{\frac{1+\delta}{2+\delta}}(t_4-t_2) & \text{if } t_4>t_2 \\ 8\left\|e_{it_1}^2 v_x^{K\prime} p_{it_4}^c\right\|_{(8+4\delta)/3,\mathcal{D}} \left\|p_{it_2}^{c\prime} v_x^K\right\|_{8+4\delta,\mathcal{D}} \alpha_{\mathcal{D}}^{\frac{1+\delta}{2+\delta}}(t_2-t_4) & \text{if } t_2>t_4 \end{cases}.
$$

If $t_4>t_2$, by Hölder and Jensen inequalities, each term inside the summation is bounded by

$$
\begin{aligned}
&\left|E_{\mathcal{D}}\left(e_{it_1}^2 v_x^{K\prime} p_{it_2}^c p_{it_4}^{c\prime} v_x^K\right)\right| \left\|f_{t_1}^0\right\|^2 \left\|f_{t_3}^0\right\| \left\|f_{t_4}^0\right\| \\
\leq\ & 8\left\|e_{it_1}^2 v_x^{K\prime} p_{it_2}^c\right\|_{(8+4\delta)/3,\mathcal{D}} \left\|p_{it_4}^{c\prime} v_x^K\right\|_{8+4\delta,\mathcal{D}} \alpha_{\mathcal{D}}^{\frac{1+\delta}{2+\delta}}(t_4-t_2)\left\|f_{t_1}^0\right\|^2 \left\|f_{t_2}^0\right\| \left\|f_{t_4}^0\right\| \\
\leq\ & 8K\left\|v_x^K\right\|^2 \left\|e_{it_1}\right\|_{8+4\delta,\mathcal{D}}^2 \varphi_{it_2,8+4\delta}^c \varphi_{it_4,8+4\delta}^c \alpha_{\mathcal{D}}^{\frac{1+\delta}{2+\delta}}(t_4-t_2)\left\|f_{t_1}^0\right\|^2 \left\|f_{t_2}^0\right\| \left\|f_{t_4}^0\right\| \\
\leq\ & 2K\left\|v_x^K\right\|^2 (2C_{1,it_1,e}+C_{1,it_2,p}+C_{1,it_4,p})\alpha_{\mathcal{D}}^{\frac{1+\delta}{2+\delta}}(t_4-t_2).
\end{aligned}
$$

Similarly, if $t_2>t_4$, each term inside the summation is bounded by $2K\left\|v_x^K\right\|^2(2C_{1,it_1,e}+C_{1,it_2,p}+C_{1,it_4,p})$ $\times\alpha_{\mathcal{D}}^{\frac{1+\delta}{2+\delta}}(t_2-t_4)$. It follows that

$$
\begin{aligned}
\left|E\Psi_{NT,22b1}\right| &\leq \frac{2}{NT^3}\sum_{i=1}^{N}\left\{\sum_{1\leq t_1<t_2<t_4\leq T}+\sum_{1\leq t_1<t_4<t_2\leq T}\right\}\left|\eta_{t_1t_2}\eta_{t_3t_4} v_x^{K\prime} E_{\mathcal{D}}\left(p_{it_2}^c e_{it_1} p_{it_4}^{c\prime} e_{it_3}\right) v_x^K\right| \\
&\leq \frac{4K\left\|v_x^K\right\|^2}{NT^3}\sum_{i=1}^{N}\sum_{1\leq t_1<t_2<t_4\leq T}(2C_{1,it_1,e}+C_{1,it_2,p}+C_{1,it_4,p})\alpha_{\mathcal{D}}^{\frac{1+\delta}{2+\delta}}(t_4-t_2) \\
&\leq \frac{CK}{NT^3}\left\{\sum_{i=1}^{N}\sum_{t_1=1}^{T}C_{1,it_1,e}\right\}\left\{\sum_{1\leq t_2<t_4\leq T}\alpha_{\mathcal{D}}^{\frac{1+\delta}{2+\delta}}(t_4-t_2)\right\} \\
&\quad +\frac{CKT}{NT^3}\left\{\sum_{i=1}^{N}\sum_{t=1}^{T}(C_{1,it,p}+C_{1,it,p})\right\}\left\{\sum_{m=1}^{T}\alpha_{\mathcal{D}}^{\frac{1+\delta}{2+\delta}}(m)\right\} \\
&= O_P(K/T)+O_P(K/T)=O_P(K/T).
\end{aligned}
$$

Similarly, we can show that $E\Psi_{NT,22b2}=O_P(K/T)$. Thus $E\Psi_{NT,22b}=O_P(K/T)$. In sum, we have shown that $E\Psi_{NT,22}=O_P(K/T)$, implying that $\Psi_{NT,22}=o_P(1)$ by Chebyshev inequality.

Using similar arguments as used in the study of $\Psi_{NT,22}$, we can show that $\Psi_{NT,23}=o_P(1)$. ■

Proof of Lemma A.5. By straightforward moment calculations and Chebyshev inequality, one can prove (i)-(ii); see also Moon and Weidner (2010, S.4 p.14).

(iii) Noting that the (r,s)th element of $f^{0\prime}\mathbf{e}'\mathbf{P}_{(a)}$ is given by $\sum_{i=1}^{N}\sum_{t=1}^{T} f_{tr}^0 e_{it} a' p_{is}$, we have

$$\begin{aligned}
E_{\mathcal{D}}\left[\left\|f^{0\prime}\mathbf{e}'\mathbf{P}_{(a)}\right\|_F^2\right] &= E_{\mathcal{D}}\left[\sum_{r=1}^{R}\sum_{s=1}^{T}\left(\sum_{i=1}^{N}\sum_{t=1}^{T} f_{tr}^0 e_{it} a' p_{is}\right)^2\right] \\
&= \sum_{r=1}^{R}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s=1}^{T}\sum_{t=1}^{T}\sum_{q=1}^{T} f_{tr}^0 f_{qr}^0 E_{\mathcal{D}}\left[a' p_{is} a' p_{js} e_{it} e_{jq}\right] \\
&= \sum_{r=1}^{R}\sum_{i=1}^{N}\sum_{1\le t,s\le T}\left(f_{tr}^0\right)^2 E_{\mathcal{D}}\left[\left(a' p_{is}\right)^2 e_{it}^2\right] + \sum_{r=1}^{R}\sum_{i=1}^{N}\sum_{1\le t\neq q<s\le T} f_{tr}^0 f_{qr}^0 E_{\mathcal{D}}\left[\left(a' p_{is}\right)^2 e_{it} e_{iq}\right] \\
&\quad +\sum_{r=1}^{R}\sum_{1\le i\neq j\le N}\sum_{s=1}^{T}\sum_{t=1}^{T}\sum_{q=1}^{T} f_{tr}^0 f_{qr}^0 E_{\mathcal{D}}\left[a' p_{is} a' p_{js} e_{it} e_{jq}\right] \\
&\equiv T_{1NT}+T_{2NT}+T_{3NT}, \text{ say.}
\end{aligned}$$

Note that $T_{1NT}\le\|a\|^2\sum_{i=1}^{N}\sum_{1\le s,t\le T}\left\|f_t^0\right\|^2 E_{\mathcal{D}}\left[\|p_{is}\|^2 e_{it}^2\right] = O_P\left(NT^2K\right)$ by Markov inequality. For T_{2NT} and T_{3NT}, following the proof of Proposition A.6 and by the conditional Davydov and Jensen inequalities we have

$$\begin{aligned}
\left|T_{2NT}\right| &\le \sum_{r=1}^{R}\sum_{i=1}^{N}\sum_{1\le t\neq q<s\le T}\left|f_{tr}^0\right|\left|f_{qr}^0\right|\left|E_{\mathcal{D}}\left[\left(a' p_{is}\right)^2 e_{it} e_{iq}\right]\right| \\
&\le 16\|a\|^2 K\sum_{i=1}^{N}\sum_{1\le t<q<s\le T}\left\|f_t^0\right\|\left\|f_q^0\right\|\|e_{it}\|_{8+4\delta,\mathcal{D}}\,\varphi_{is,8+4\delta}^2\|e_{iq}\|_{8+4\delta,\mathcal{D}}\,\alpha_{\mathcal{D}}^{\frac{1+\delta}{2+\delta}}(q-t) \\
&= O_P\left(NT^2K\right),
\end{aligned}$$

and

$$\begin{aligned}
\left|T_{3NT}\right| &\le \sum_{r=1}^{R}\sum_{1\le i,j\le N}\sum_{1\le t,q<s\le T}\left|f_{tr}^0\right|\left|f_{qr}^0\right|\left|E_{\mathcal{D}}\left(e_{it} a' p_{is}\right)\right|\left|E_{\mathcal{D}}\left(a' p_{js} e_{jq}\right)\right| \\
&\le \|a\|^2\sum_{s=2}^{T}\left\{8K^{1/2}\sum_{i=1}^{N}\sum_{t=1}^{s-1}\left\|f_t^0\right\|\|e_{it}\|_{8+4\delta,\mathcal{D}}\,\varphi_{is,8+4\delta}\,\alpha_{\mathcal{D}}^{\frac{3+2\delta}{4+2\delta}}(s-t)\right\}^2 \\
&= O_P\left(N^2TK\right).
\end{aligned}$$

It follows that $||f^{0\prime}\mathbf{e}'\mathbf{P}_{(a)}||_F = O_P\left((NTK)^{1/2}\delta_{NT}\right)$.

(iv) By (iii), $\left\|P_{f^0}\mathbf{e}'\mathbf{P}_{(a)}\right\||_F\le\left\|f^0\left(f^{0\prime}f^0\right)^{-1}\right\|_F\left\|f^{0\prime}\mathbf{e}'\mathbf{P}_{(a)}\right\|_F = O_P\left(T^{-1/2}\right)O_P\left((NTK)^{1/2}\delta_{NT}\right) = O_P(\sqrt{NK}\delta_{NT})$.

(v) Noting that (r,j)th element of $\lambda^{0\prime}\mathbf{e}\mathbf{P}_{(a)}'$ is given by $\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_{ir}^0 e_{it} a' p_{jt}$, we have

$$\begin{aligned}
E_{\mathcal{D}}\left[\left\|\lambda^{0\prime}\mathbf{e}\mathbf{P}_{(a)}'\right\|_F^2\right] &= E_{\mathcal{D}}\left[\sum_{r=1}^{R}\sum_{j=1}^{N}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_{ir}^0 e_{it} a' p_{jt}\right)^2\right] \\
&= \sum_{r=1}^{R}\sum_{1\le i\neq j\le N}\sum_{t=1}^{T}\left(\lambda_{ir}^0\right)^2 E_{\mathcal{D}}\left[e_{it}^2\right] E_{\mathcal{D}}\left(a' p_{jt}\right)^2 + \sum_{r=1}^{R}\sum_{j=1}^{N}\sum_{t=1}^{T}\left(\lambda_{jr}^0\right)^2 E_{\mathcal{D}}\left[e_{jt}^2\left(a' p_{jt}\right)^2\right] \\
&= \sum_{1\le i\neq j\le N}\sum_{t=1}^{T}\left\|\lambda_i^0\right\|^2 E_{\mathcal{D}}\left[e_{it}^2\right] E_{\mathcal{D}}\left(a' p_{jt}\right)^2 + \sum_{j=1}^{N}\sum_{t=1}^{T}\left\|\lambda_j^0\right\|^2 E_{\mathcal{D}}\left[e_{jt}^2\left(a' p_{jt}\right)^2\right] \\
&= O_P\left(N^2TK\right) + O_P\left(NTK\right) = O_P\left(N^2TK\right).
\end{aligned}$$

It follows that $\left\|\lambda^{0\prime}\mathbf{e}\mathbf{P}_{(a)}^{\prime}\right\|_F = O_P(N\sqrt{TK})$.

(vi) By (v), $\left\|P_{\lambda^0}\mathbf{e}\mathbf{P}_{(a)}^{\prime}\right\|_F \leq \left\|\lambda^0\left(\lambda^{0\prime}\lambda^0\right)^{-1}\right\|_F \left\|\lambda^{0\prime}\mathbf{e}^{\prime}\mathbf{P}_{(a)}\right\|_F = O_P\left(N^{-1/2}\right)O_P\left((NTK)^{1/2}\delta_{NT}\right) = O_P(\sqrt{NTK})$.

(vii) Noting that $A_i \equiv T^{-1}v_x^{K\prime}[E_{\mathcal{D}}\left(P_i - P_i^{\lambda}\right)^{\prime}]f^0G^0$ is a $1\times R$ vector and $\mathcal{D}$-measurable, we have

$$\begin{aligned}
& E_{\mathcal{D}}\left\{\frac{1}{N\sqrt{T}}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}A_i\lambda_j^0\left[e_{jt}e_{it} - E_{\mathcal{D}}\left(e_{jt}e_{it}\right)\right]\right\}^2 \\
\leq\ & 2E_{\mathcal{D}}\left\{\frac{2}{N\sqrt{T}}\sum_{t=1}^{T}\sum_{1\leq i\neq j\leq N}A_i\lambda_j^0 e_{jt}e_{it}\right\}^2 + 2E_{\mathcal{D}}\left\{\frac{1}{N\sqrt{T}}\sum_{t=1}^{T}\sum_{i=1}^{N}A_i\lambda_i^0\left[e_{it}^2 - E_{\mathcal{D}}\left(e_{it}^2\right)\right]\right\}^2 \\
=\ & \frac{4}{N^2T}\sum_{t=1}^{T}\sum_{1\leq i\neq j\leq N}\left\|A_i\right\|^2\left\|\lambda_j^0\right\|^2 E_{\mathcal{D}}\left(e_{jt}^2\right)E_{\mathcal{D}}\left(e_{it}^2\right) \\
& +\frac{2}{N^2T}\sum_{t=1}^{T}\sum_{i=1}^{N}\left\|A_i\right\|^2\left\|\lambda_i^0\right\|^2\left[E_{\mathcal{D}}\left(e_{it}^4\right) - E_{\mathcal{D}}\left(e_{it}^2\right)E_{\mathcal{D}}\left(e_{it}^2\right)\right] \\
=\ & O_P\left(K\right) + O_P\left(K/N\right) = O_P\left(K\right) \text{ by Assumption 6.}
\end{aligned}$$

Then (vii) follows by Chebyshev inequality.

$(viii)$ Note that $\frac{1}{NT}\sum_{i=1}^{N}E_{\mathcal{D}}\left(\left\|\sum_{s=1}^{T}v_x^{K\prime}\left(p_{is}^c - p_{is}^{\lambda c}\right)f_s^0G^0\right\|^2\right)$ is bounded by

$$\frac{2}{N}\sum_{i=1}^{N}E_{\mathcal{D}}\left\|\frac{1}{\sqrt{T}}\sum_{s=1}^{T}v_x^{K\prime}p_{is}^c f_s^0G^0\right\|^2 + \frac{2}{N}\sum_{i=1}^{N}E_{\mathcal{D}}\left\|\frac{1}{\sqrt{T}N}\sum_{s=1}^{T}\sum_{j=1}^{N}\alpha_{ij}v_x^{K\prime}p_{js}^c f_s^0G^0\right\|^2.$$

The first term is bounded by

$$\begin{aligned}
& \frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}\sum_{t=1}^{T}\left\|f_s^0\right\|\left\|f_t^0\right\|\left|E_{\mathcal{D}}\left[v_x^{K\prime}p_{is}^c p_{it}^{c\prime}v_x^K\right]\right|\left\|G^0\right\| \\
\leq\ & 8\left\|v_x^{K\prime}\right\|^2 K\left\|G^0\right\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}\sum_{t=1}^{T}\left\|f_s^0\right\|\left\|f_t^0\right\|\varphi_{is,8+4\delta}^c\varphi_{it,8+4\delta}^c\alpha_{\mathcal{D}}^{\frac{3+2\delta}{4+2\delta}}\left(|s-t|\right) = O_P\left(K\right)
\end{aligned}$$

by the conditional Davydov inequality. Similarly, we can show that the second term is also $O_P\left(K\right)$. Thus $(viii)$ follows by Markov inequality.

(ix) Using similar arguments as used in the proof of (vii), one can prove (iv) by Markov inequality.

(x) Note that $E_{\mathcal{D}}\{\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}B_t f_s^{0\prime}\left[e_{it}e_{is} - E_{\mathcal{D}}\left(e_{it}e_{is}\right)\right]\}^2$ is bounded by

$$\begin{aligned}
& 2E_{\mathcal{D}}\left\{\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}B_t f_t^{0\prime}\left[e_{it}^2 - E_{\mathcal{D}}\left(e_{it}^2\right)\right]\right\}^2 + 2E_{\mathcal{D}}\left\{\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{1\leq t\neq s\leq T}B_t f_s^{0\prime}e_{it}e_{is}\right\}^2 \\
=\ & \frac{2}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}B_t f_t^{0\prime}B_s f_s^{0\prime}E_{\mathcal{D}}\left[e_{it}^2e_{is}^2 - E_{\mathcal{D}}\left(e_{it}^2\right)E_{\mathcal{D}}\left(e_{is}^2\right)\right] + \frac{4}{NT^2}\sum_{i=1}^{N}\sum_{1\leq t\neq s\leq T}\left\|B_t\right\|^2\left\|f_s\right\|^2 E_{\mathcal{D}}\left(e_{it}^2e_{is}^2\right) \\
=\ & O_P\left(K\right) + O_P\left(K\right) = O_P\left(K\right)
\end{aligned}$$

by Assumption 9. Then (x) follows by Chebyshev inequality.

(xi) Note that

$$
\begin{aligned}
& E_{\mathcal{D}}\left\{\frac{1}{NT}\sum_{t=1}^{T}\left\|\sum_{j=1}^{N}v_x^{K\prime}\left(p_{jt}^{c}-p_{jt}^{fc}\right)\lambda_j^{0\prime}G^0\right\|^2\right\} \\
\leq\ & \frac{2}{NT}\sum_{t=1}^{T}E_{\mathcal{D}}\left\{\left\|\sum_{j=1}^{N}v_x^{K\prime}p_{jt}^{c}\lambda_j^{0\prime}G^0\right\|^2\right\}+\frac{2}{NT}\sum_{t=1}^{T}E_{\mathcal{D}}\left\{\left\|\frac{1}{T}\sum_{j=1}^{N}\sum_{s=1}^{T}\eta_{ts}v_x^{K\prime}p_{js}^{c}\lambda_j^{0\prime}G^0\right\|^2\right\} \\
=\ & \frac{2}{NT}\sum_{t=1}^{T}\sum_{j=1}^{N}E_{\mathcal{D}}\left(v_x^{K\prime}p_{jt}^{c}\right)^2\alpha_j^0+\frac{2}{NT}\sum_{t=1}^{T}\frac{1}{T^2}\sum_{j=1}^{N}\sum_{s=1}^{T}\alpha_j^0\eta_{ts}^2v_x^{K\prime}E_{\mathcal{D}}\left(p_{js}^{c}p_{js}^{c\prime}\right)v_x^{K} \\
& +\frac{2K\left\|v_x^K\right\|^2}{NT}\sum_{t=1}^{T}\frac{1}{T^2}\sum_{j=1}^{N}\sum_{1\leq s\neq r\leq T}\alpha_j^0\eta_{ts}\eta_{tr}\varphi_{js,8+4\delta}^{c}\varphi_{jr,8+4\delta}^{c}\alpha_{\mathcal{D}}^{\frac{3+2\delta}{4+2\delta}}\left(|r-s|\right) \\
=\ & O_P(K)+O_P(K)+O_P(K)=O_P(K),
\end{aligned}
$$

where $\alpha_j^0\equiv\lambda_j^{0\prime}G^0G^{0\prime}\lambda_j^0$. Then (xi) follows by Chebyshev inequality.

(xii) The proof is similar to that of (x) and thus omitted. ■

D.3 Bias correction

Let $\hat{\mathbf{e}}(\beta)\equiv\mathbf{Y}-\sum_{k=1}^{K}\beta_k\mathbf{P}_k-\hat{\lambda}(\beta)\hat{f}(\beta)'$. Following Moon and Weidner (2010, 2012), we first derive the asymptotic expansions for the projectors $M_{\hat{f}}(\beta)$ and $M_{\hat{\lambda}}(\beta)$, and the residual matrix $\hat{\mathbf{e}}(\beta)$, and then establish some lemmas that are used to prove Lemmas A.8 and A.9.

Lemma D.2 *Under Assumptions 1-4, we have the following expansions*

(i) $M_{\hat{\lambda}}(\beta)=M_{\lambda^0}+M_{\hat{\lambda},\mathbf{u}}^{(1)}+M_{\hat{\lambda},\mathbf{u}}^{(2)}+\sum_{k=1}^{K}\left(\beta_k^0-\beta_k\right)M_{\hat{\lambda},k}^{(1)}+M_{\hat{\lambda}}^{(rem)}(\beta)$,

(ii) $M_{\hat{f}}(\beta)=M_{f^0}+M_{\hat{f},\mathbf{u}}^{(1)}+M_{\hat{f},\mathbf{u}}^{(2)}+\sum_{k=1}^{K}\left(\beta_k^0-\beta_k\right)M_{\hat{f},k}^{(1)}+M_{\hat{f}}^{(rem)}(\beta)$,

(iii) $\hat{\mathbf{e}}(\beta)=M_{\lambda^0}\mathbf{u}M_{f^0}+\hat{\mathbf{e}}_e^{(1)}+\sum_{k=1}^{K}\left(\beta_k^0-\beta_k\right)\hat{\mathbf{e}}_k^{(1)}+\hat{\mathbf{e}}^{(rem)}(\beta)$,

where $\hat{\mathbf{e}}_k^{(1)}=M_{\lambda^0}\mathbf{P}_kM_{f^0}$, $\hat{\mathbf{e}}_e^{(1)}=-M_{\lambda^0}\mathbf{u}M_{f^0}\mathbf{u}'\Phi'-\Phi'\mathbf{u}'M_{\lambda^0}\mathbf{u}M_{f^0}-M_{\lambda^0}\mathbf{u}\Phi\mathbf{u}M_{f^0}$, *the expansion coefficients of* $M_{\hat{\lambda}}(\beta)$ *are given by*

$$
\begin{aligned}
M_{\hat{\lambda},\mathbf{u}}^{(1)} &= -M_{\lambda^0}\mathbf{u}\Phi-\Phi'\mathbf{u}'M_{\lambda^0}, \\
M_{\hat{\lambda},k}^{(1)} &= -M_{\lambda^0}\mathbf{P}_k\Phi-\Phi'\mathbf{P}_k'M_{\lambda^0}, \\
M_{\hat{\lambda},\mathbf{u}}^{(2)} &= M_{\lambda^0}\mathbf{u}\Phi\mathbf{u}\Phi+\Phi'\mathbf{u}'\Phi'\mathbf{u}'M_{\lambda^0}-M_{\lambda^0}\mathbf{u}M_{f^0}\mathbf{u}'\Phi_2-\Phi_2\mathbf{u}M_{f^0}\mathbf{u}'M_{\lambda^0}-M_{\lambda^0}\mathbf{u}\Phi_1\mathbf{u}'M_{\lambda^0}+\Phi'\mathbf{u}'M_{\lambda^0}\mathbf{u}\Phi,
\end{aligned}
$$

and, analogously, the expansion coefficients of $M_{\hat{f}}(\beta)$ *are given by*

$$
\begin{aligned}
M_{\hat{f},\mathbf{u}}^{(1)} &= -M_{f^0}\mathbf{u}\Phi'-\Phi\mathbf{u}'M_{f^0}, \\
M_{\hat{f},k}^{(1)} &= -M_{f^0}\mathbf{P}_k'\Phi'-\Phi\mathbf{P}_kM_{f^0}, \\
M_{\hat{f},\mathbf{u}}^{(2)} &= M_{f^0}\mathbf{u}'\Phi'\mathbf{u}'\Phi'+\Phi\mathbf{u}\Phi\mathbf{u}M_{f^0}-M_{f^0}\mathbf{u}'M_{\lambda^0}\mathbf{u}\Phi_1-\Phi_1\mathbf{u}'M_{\lambda^0}\mathbf{u}M_{f^0}-M_{f^0}\mathbf{u}'\Phi_2\mathbf{u}M_{f^0}+\Phi\mathbf{u}M_{f^0}\mathbf{u}'\Phi'.
\end{aligned}
$$

For the remainder terms, we have

$$
\begin{aligned}
||M_{\hat{\lambda}}^{(rem)}(\beta)|| &= O_P[(\delta_{NT}^{-1}+K^{-\gamma/d})\left\|\beta^0-\beta\right\|+\left\|\beta^0-\beta\right\|^2+(\delta_{NT}^{-3}+K^{-3\gamma/d})], \\
||M_{\hat{f}}^{(rem)}(\beta)|| &= O_P[(\delta_{NT}^{-1}+K^{-\gamma/d})\left\|\beta^0-\beta\right\|+\left\|\beta^0-\beta\right\|^2+(\delta_{NT}^{-3}+K^{-3\gamma/d})], \\
||\hat{\mathbf{e}}^{(rem)}(\beta)|| &= O_P\{\sqrt{NT}[\left\|\beta-\beta^0\right\|^2+(\delta_{NT}^{-1}+K^{-\gamma/d})\left\|\beta-\beta^0\right\|+(\delta_{NT}^{-3}+K^{-3\gamma/d})]\},
\end{aligned}
$$

and rank$\left(\hat{\mathbf{e}}^{(rem)}(\beta)\right) \leq 7R$.

Proof. Since the symmetry of $N \leftrightarrow T$, $\lambda \leftrightarrow f$, $\mathbf{u} \leftrightarrow \mathbf{u}'$, and $\mathbf{P}_k \leftrightarrow \mathbf{P}_k'$, the proofs for $M_{\hat{f}}(\beta)$ and $M_{\hat{\lambda}}(\beta)$ are similar. So we only consider the proofs of $M_{\hat{f}}(\beta)$ and $\hat{\mathbf{e}}(\beta)$.

Expansion of $M_{\hat{f}}(\beta)$. By Proposition C.1 (iii) and the fact $\mathbf{u} = \epsilon_0 \mathbf{P}_0$, we have

$$
\begin{aligned}
M_{\hat{f}}(\beta) &= M_{f^0} + M_f^{(1)}\left(\lambda^0, f^0, \mathbf{u}\right) + M_f^{(1)}\left(\lambda^0, f^0, \sum_{k=1}^{K} \epsilon_k \mathbf{P}_k\right) + M_f^{(2)}\left(\lambda^0, f^0, \mathbf{u}, \mathbf{u}\right) \\
&\quad + \left\{ M_f^{(2)}\left(\lambda^0, f^0, \sum_{k=1}^{K} \epsilon_k \mathbf{P}_k, \sum_{k=1}^{K} \epsilon_k \mathbf{P}_k\right) + O_P\left(a_{NT}^3\right) \right\} \\
&= M_{f^0} + M_{\hat{f},\mathbf{u}}^{(1)} + \sum_{k=1}^{K}\left(\beta_k^0 - \beta_k\right) M_{\hat{f},k}^{(1)} + M_{\hat{f},\mathbf{u}}^{(2)} + M_{\hat{f}}^{(rem)}(\beta)
\end{aligned}
$$

Following the proof in Proposition C.1, we can show that

$$
M_{\hat{f}}^{(\text{rem})}(\beta) = O_P\left[\left(\delta_{NT}^{-1} + K^{-\gamma/d}\right)\left\|\beta^0 - \beta\right\| + \left\|\beta^0 - \beta\right\|^2 + \left(\delta_{NT}^{-3} + K^{-3\gamma/d}\right)\right].
$$

Expansion of $\hat{\mathbf{e}}(\beta)$. By the definition of $\hat{\mathbf{e}}(\beta)$ and using the expansions of $M_{\hat{\lambda}}$ and $M_{\hat{f}}$, we have

$$
\begin{aligned}
\hat{\mathbf{e}}(\beta) &= \mathbf{Y} - \sum_{k=1}^{K} \beta_k \mathbf{P}_k - \hat{\lambda}(\beta)\hat{f}(\beta)' = M_{\hat{\lambda}}\left[\mathbf{Y} - \sum_{k=1}^{K} \beta_k \mathbf{P}_k\right] M_{\hat{f}} \\
&= M_{\hat{\lambda}}\left[\mathbf{u} - \sum_{k=1}^{K}\left(\beta_k - \beta_k^0\right)\mathbf{P}_k + \lambda^0 f^{0\prime}\right] M_{\hat{f}} \\
&= M_{\lambda^0}\mathbf{u}M_{f^0} - \left\|\beta - \beta^0\right\| M_{\lambda^0}\mathbf{P}_{(a)}M_{f^0} - M_{\lambda^0}\mathbf{u}M_{f^0}\mathbf{u}\Phi' - M_{\lambda^0}\mathbf{u}\Phi'\mathbf{u}'M_{f^0} - \Phi'\mathbf{u}'M_{\lambda^0}\mathbf{u}M_{f^0} + \hat{\mathbf{e}}^{(rem)}(\beta).
\end{aligned}
$$

Noting that $\left\|M_{\hat{f},\mathbf{u}}^{(1)}\right\| = O_P\left(\delta_{NT}^{-1} + K^{-\gamma/d}\right)$, $\left\|M_{\hat{\lambda},\mathbf{u}}^{(1)}\right\| = O_P\left(\delta_{NT}^{-1} + K^{-\gamma/d}\right)$, $\left\|M_{\hat{f},\mathbf{u}}^{(2)}\right\| = O_P\left(\delta_{NT}^{-2} + K^{-2\gamma/d}\right)$, $\left\|M_{\hat{\lambda},\mathbf{u}}^{(2)}\right\| = O_P\left(\delta_{NT}^{-2} + K^{-2\gamma/d}\right)$, $\left\|\sum_{k=1}^{K}\left(\beta_k - \beta_k^0\right) M_{\hat{f},k}^{(1)}\right\| = O_P\left(\left\|\beta - \beta^0\right\|\right)$, and $\left\|\sum_{k=1}^{K}\left(\beta_k - \beta_k^0\right) M_{\hat{\lambda},k}^{(1)}\right\| = O_P\left(\left\|\beta - \beta^0\right\|\right)$, we have

$$
\left\|\hat{\mathbf{e}}^{(rem)}(\beta)\right\| = O_P\left(\sqrt{NT}\left[\left\|\beta - \beta^0\right\|^2 + \left(\delta_{NT}^{-1} + K^{-\gamma/d}\right)\left\|\beta - \beta^0\right\| + \left(\delta_{NT}^{-3} + K^{-3\gamma/d}\right)\right]\right).
$$

Let $A_0 = \mathbf{u} - \sum_{k=1}^{K}\left(\beta_k - \beta^0\right)\mathbf{P}_k$, $A_1 = A_0 - M_{\lambda^0}A_0M_{f^0}$, $A_2 = \lambda^{0\prime}f^0 - \hat{\lambda}(\beta)'\hat{f}(\beta)$, and $A_3 = -\hat{\mathbf{e}}_e^{(1)}$, where $\hat{\lambda}(\beta) = P_{\hat{\lambda}}(\beta)\lambda^0$ and $\hat{f}(\beta) = P_{\hat{f}}(\beta) f^0$. Note that $\hat{\mathbf{e}}^{(rem)}(\beta) = A_1 + A_2 + A_3$, rank$(A_1) \leq 2R$, rank$(A_2) \leq 2R$, and rank$(A_3) \leq 3R$. It follows that rank$(\hat{\mathbf{e}}^{(rem)}(\beta)) \leq 7R$. ■

Lemma D.3 *Under Assumptions 1-4, we have*

(i) $||P_{\hat{\lambda}} - P_{\lambda^0}|| = ||M_{\hat{\lambda}} - M_{\lambda^0}|| = O_P(\delta_{NT}^{-1} + K^{-\gamma/d})$,

(ii) $||P_{\hat{f}} - P_{f^0}|| = ||M_{\hat{f}} - M_{f^0}|| = O_P(\delta_{NT}^{-1} + K^{-\gamma/d})$.

Proof. Noting that $\|\mathbf{u}\|/\sqrt{NT} = O_P\left(\delta_{NT}^{-1} + K^{-\gamma/d}\right)$, $\left\|\mathbf{P}_{(a)}\right\|/\sqrt{NT} = O_P(1)$, and $\left\|\beta^0 - \hat{\beta}\right\| = O_P\left(K^{1/2}\delta_{NT}^{-2} + K^{-\gamma/d}\right)$, we have by D.3$(ii)$

$$
\begin{aligned}
\left\|P_{\hat{f}} - P_{f^0}\right\| &\leq \left\|M_{\hat{f},\mathbf{u}}^{(1)}\right\| + \left\|M_{\hat{f},\mathbf{u}}^{(2)}\right\| + \left\|\sum_{k=1}^{K}\left(\beta_k^0 - \hat{\beta}_k\right) M_{\hat{f},k}^{(1)}\right\| + \left\|M_{\hat{f}}^{(rem)}(\beta)\right\| \\
&= O_P\left(\delta_{NT}^{-1} + K^{-\gamma/d}\right) + O_P\left(\delta_{NT}^{-2} + K^{-2\gamma/d}\right) + O_P\left(\left\|\beta^0 - \hat{\beta}\right\|\right) \\
&\quad + O_P\left[\left(\delta_{NT}^{-1} + K^{-\gamma/d}\right)\left\|\beta^0 - \beta\right\| + \left\|\beta^0 - \beta\right\|^2 + \left(\delta_{NT}^{-3} + K^{-3\gamma/d}\right)\right] \\
&= O_P\left(\delta_{NT}^{-1} + K^{-\gamma/d}\right).
\end{aligned}
$$

Similarly, we can show that $\left\|P_{\hat{\lambda}} - P_{\hat{\lambda}^0}\right\| = O_P\left(\delta_{NT}^{-1} + K^{-\gamma/d}\right)$. ■

Lemma D.4 *Under Assumptions 1-4, there exists an $R \times R$ matrix $H = H_{NT}$ such that*
(i) $||\hat{f} - f^0 H||/\sqrt{T} = O_P(\delta_{NT}^{-1} + K^{-\gamma/d})$;
(ii) $||\hat{\lambda} - \lambda^0 (H')^{-1}||/\sqrt{N} = O_P(\delta_{NT}^{-1} + K^{-\gamma/d})$;
(iii) $\sqrt{NT}||\hat{\Phi} - \Phi|| = O_P(\delta_{NT}^{-1} + K^{-\gamma/d})$.

Proof. *(i)* Noting that $||P_{\hat{f}} - P_{f^0}|| = o_P(1)$, we have $\text{rank}(P_{\hat{f}} P_{f^0}) = R$, i.e., $\text{rank}(P_{\hat{f}} f^0) = R$ as $(N, T) \to \infty$. Write $\hat{f} = P_{\hat{f}} f^0 H$ with some non-singular $R \times R$ matrix $H = H_{NT}$. It is easy to see that $H = (\hat{f}' P_{\hat{f}} f^0/T)^{-1}(\hat{f}'\hat{f}/T) = (\hat{f}' f^0/T)^{-1}$ and $||H^{-1}|| \leq T^{-1}||\hat{f}' f^0|| = O_P(1)$. Note that $\hat{f} = f^0 H + (P_{\hat{f}} - P_{f^0}) f^0 H$ and $H = (f^{0\prime} f^0/T)^{-1} f^{0\prime}\hat{f}/T - (f^{0\prime} f^0/T)^{-1} f^{0\prime}(P_{\hat{f}} - P_{f^0}) f^0 H/T$. It follows that $||H|| \leq O_P(1) + ||H|| O_P(\delta_{NT}^{-1} + K^{-\gamma/d})$, which implies that $||H|| = O_P(1)$. Noting that $\hat{f} = P_{\hat{f}} f^0 H$, we have $\left\|\hat{f} - f^0 H\right\| = \left\|(P_{\hat{f}} - P_{f^0}) f^0 H\right\| \leq R\left\|P_{\hat{f}} - P_{f^0}\right\| \|f^0\| \|H\| = O_P[\sqrt{T}\left(\delta_{NT}^{-1} + K^{-\gamma/d}\right)]$.

(ii) Noting that $\hat{\lambda}\hat{f}'\hat{f} = \left(\mathbf{Y} - \sum_{k=1}^K \hat{\beta}_k \mathbf{P}_k\right)\hat{f}$, we have

$$\begin{aligned}
\hat{\lambda} - \lambda^0 (H')^{-1} &= \left[\lambda^0 f^{0\prime} + \sum_{k=1}^K \left(\beta_k^0 - \hat{\beta}_k\right)\mathbf{P}_k + \mathbf{u}\right]\hat{f}\left(\hat{f}'\hat{f}\right)^{-1} - \lambda^0 (H')^{-1} \\
&= \lambda^0 f^{0\prime}\left(P_{\hat{f}} - P_{f0}\right) f^0 \left(f^{0\prime} P_{\hat{f}} f^0\right)^{-1} (H')^{-1} + \lambda^0 f^{0\prime} f^0 \left[\left(f^{0\prime} P_{\hat{f}} f^0\right)^{-1} - \left(f^{0\prime} f^0\right)^{-1}\right] (H')^{-1} \\
&\quad + \left[\sum_{k=1}^K \left(\beta_k^0 - \hat{\beta}_k\right)\mathbf{P}_k + \mathbf{u}\right] P_{\hat{f}} f^0 \left(f^{0\prime} P_{\hat{f}} f^0\right)^{-1} (H')^{-1} \\
&\equiv \Lambda_{1NT} + \Lambda_{2NT} + \Lambda_{3NT}, \text{ say.}
\end{aligned}$$

First, $\Lambda_{1NT} \leq \frac{2R}{T}\left\|\lambda^0\right\|\left\|f^0\right\|^2 ||(H')^{-1}||||P_{\hat{f}} - P_{f0}||||(f^{0\prime} P_{\hat{f}} f^0/T)^{-1}|| = O_P[\sqrt{N}(\delta_{NT}^{-1} + K^{-\gamma/d})]$. Noting that

$$\begin{aligned}
\left\|\left(f^{0\prime} P_{\hat{f}} f^0/T\right)^{-1} - \left(f^{0\prime} f^0/T\right)^{-1}\right\| &\leq \left\|f^{0\prime}\left(P_{\hat{f}} - P_{f^0}\right) f^0/T\right\| \left\|\left(f^{0\prime} f^0/T\right)^{-1}\right\| \left\|\left(f^{0\prime} P_{\hat{f}} f^0/T\right)^{-1}\right\| \\
&= \left\|P_{\hat{f}} - P_{f^0}\right\| \left\|f^0\right\|^2/T \left\|\left(f^{0\prime} f^0/T\right)^{-1}\right\| \left\|\left(f^{0\prime} P_{\hat{f}} f^0/T\right)^{-1}\right\| \\
&= O_P\left(\delta_{NT}^{-1} + K^{-\gamma/d}\right),
\end{aligned}$$

we have $\Lambda_{2NT} \leq \left\|\lambda^0\right\| \left\|f^{0\prime} f^0/T\right\| [(f^{0\prime} P_{\hat{f}} f^0/T)^{-1} - \left(f^{0\prime} f^0/T\right)^{-1}]||\,(H')^{-1}|| = \sqrt{N} O_P\left(\delta_{NT}^{-1} + K^{-\gamma/d}\right)$. Now, $\|\Lambda_{3NT}\| \leq \frac{1}{T}(||\beta^0 - \hat{\beta}|| \left\|\mathbf{P}_{(a)}\right\| + \|\mathbf{u}\|)||P_{\hat{f}}|| \left\|f^0\right\| ||(f^{0\prime} P_{\hat{f}} f^0/T)^{-1}|| \left\|H^{-1}\right\| = O_P[\sqrt{N}(||\beta^0 - \hat{\beta}|| + \delta_{NT}^{-1} + K^{-\gamma/d})] = O_P[\sqrt{N}\left(\delta_{NT}^{-1} + K^{-\gamma/d}\right)]$. Consequently, $\left\|\hat{\lambda} - \lambda^0 (H')^{-1}\right\| = O_P\left[\sqrt{N}\left(\delta_{NT}^{-1} + K^{-\gamma/d}\right)\right]$.

(iii) Noting that $||\hat{\lambda}'\hat{\lambda}/N - H^{-1}\lambda^{0\prime}\lambda^0(H')^{-1}/N|| = ||N^{-1}(\hat{\lambda}' - H^{-1}\lambda^{0\prime})(\hat{\lambda} + \lambda^0(H')^{-1})|| \leq N^{-1}||\hat{\lambda} - H^{-1}\lambda^0||||\hat{\lambda}||/\sqrt{N} + ||\lambda^0||/\sqrt{N}\left\|(H')^{-1}\right\| = O_P(\delta_{NT}^{-1} + K^{-\gamma/d})$, we have

$$\begin{aligned}
&\left\|\left(\hat{\lambda}'\hat{\lambda}/N\right)^{-1} - \left(H^{-1}\lambda^{0\prime}\lambda^0(H')^{-1}/N\right)^{-1}\right\| \\
\leq\ &\left\|\hat{\lambda}'\hat{\lambda}/N\right\| \left\|\hat{\lambda}'\hat{\lambda}/N - H^{-1}\lambda^{0\prime}\lambda^0(H')^{-1}/N\right\| \left\|H^{-1}\lambda^{0\prime}\lambda^0(H')^{-1}/N\right\| = O_P\left(\delta_{NT}^{-1} + K^{-\gamma/d}\right).
\end{aligned}$$

Similarly, $||(\hat{f}'\hat{f}/T)^{-1} - \left(H' f^{0\prime} f^0 H/T\right)^{-1}|| = O_P\left(\delta_{NT}^{-1} + K^{-\gamma/d}\right)$. Combining these results, we have $\sqrt{NT}||\hat{\Phi} - \Phi|| = ||(\hat{\lambda}/\sqrt{N})(\hat{\lambda}'\hat{\lambda}/N)^{-1}(\hat{f}'\hat{f}/T)^{-1}\hat{f}'/\sqrt{T} - (\lambda^0/\sqrt{N})(H')^{-1}[H^{-1}\lambda^{0\prime}\lambda^0(H')^{-1}/N]^{-1}\left(H' f^{0\prime} f^0 H/T\right)^{-1} H' f^{0\prime}/\sqrt{T}|| = O_P\left(\delta_{NT}^{-1} + K^{-\gamma/d}\right)$. ■

Lemma D.5 *Suppose that the conditions in Theorem 3.3 hold. Then we have*

(*i*) $(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}e_{it}^2(\hat{Z}_{it}\hat{Z}_{it}'-\tilde{Z}_{it}\tilde{Z}_{it}')=O_P\left(K^{1-\gamma/d}+K\delta_{NT}^{-1}\right)$;

(*ii*) $(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}(e_{it}^2-\hat{e}_{it}^2)\hat{Z}_{it}\hat{Z}_{it}'=O_P(K\delta_{NT}^{-1}+(NT)^{1/4}K\delta_{NT}^{-2}+(NT)^{1/4}K^{1-\gamma/d})$.

Proof. (*i*) Note that $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}e_{it}^2(\hat{Z}_{it}\hat{Z}_{it}'-\tilde{Z}_{it}\tilde{Z}_{it}')=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[e_{it}^2(\hat{Z}_{it}\hat{Z}_{it}'-Z_{it}Z_{it}')\right]$ $+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[e_{it}^2(Z_{it}Z_{it}'-\tilde{Z}_{it}\tilde{Z}_{it}')\right]\equiv A_{11}+A_{12}$, say. Let $B_{1,it}=\hat{Z}_{it}-Z_{it}$ and $B_{2,it}=e_{it}^2Z_{it}$. Then

$$
\begin{aligned}
A_{11} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}e_{it}^2\left(\hat{Z}_{it}\hat{Z}_{it}'-Z_{it}Z_{it}'\right)\\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\left[\left(\hat{Z}_{it}-Z_{it}\right)Z_{it}'e_{it}^2+Z_{it}e_{it}^2\left(\hat{Z}_{it}-Z_{it}\right)'\right]+e_{it}^2\left(\hat{Z}_{it}-Z_{it}\right)\left(\hat{Z}_{it}-Z_{it}\right)'\right\}\\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(B_{1,it}B_{2,it}'+B_{2,it}B_{1.it}'\right)+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}e_{it}^2B_{1,it}B_{1,it}'=A_{11}^{(a)}+A_{11}^{(b)},\text{ say.}
\end{aligned}
$$

Define $N\times T$ matrices $\mathbf{B}_{1,k}$ and $\mathbf{B}_{2,k}$ with their (i,t)th elements given by the kth elements of $B_{1,it}$ and $B_{2,it}$, respectively. Then we have $A_{11,k_1k_2}^{(a)}=\frac{1}{NT}\text{tr}\left(\mathbf{B}_{1,k_1}\mathbf{B}_{2,k_2}'+\mathbf{B}_{2,k_1}\mathbf{B}_{1,k_2}'\right)$. Note that $\mathbf{B}_{1,k}=\left(M_{\hat{\lambda}}-M_{\lambda^0}\right)\mathbf{P}_kM_{f^0}+M_{\hat{\lambda}}\mathbf{P}_k\left(M_{\hat{f}}-M_{f^0}\right)$ and $\|\mathbf{B}_{1,k}\|=O_P\left(K^{-\gamma/d}+\delta_{NT}^{-1}\right)\|\mathbf{P}_k\|$. For $\mathbf{B}_{2,k}$, we have

$$
\|\mathbf{B}_{2,k}\|^2\leq\|\mathbf{B}_{2,k}\|_F^2=\sum_{i=1}^{N}\sum_{t=1}^{T}e_{it}^4Z_{it,k}^2\leq\left\{\sum_{i=1}^{N}\sum_{t=1}^{T}e_{it}^8\right\}^{1/2}\left\{\sum_{i=1}^{N}\sum_{t=1}^{T}Z_{it,k}^4\right\}^{1/2}=O_P\left(NT\right)\left[Z_k^{(4)}\right]^2
$$

where $Z_k^{(4)}=\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}Z_{it,k}^4\right)^{1/4}$. It follows that $\|\mathbf{B}_{2,k}\|=O_P[(NT)^{1/2}]Z_k^{(4)}$,

$$
\begin{aligned}
A_{11,k_1k_2}^{(a)} &\leq \frac{6R}{NT}\left[\|\mathbf{B}_{1,k_1}\|\,\|\mathbf{B}_{2,k_2}\|+\|\mathbf{B}_{2,k_1}\|\,\|\mathbf{B}_{1,k_2}\|\right]\\
&= O_P\left(K^{-\gamma/d}+\delta_{NT}^{-1}\right)\left[Z_{k_2}^{(4)}\|\mathbf{P}_{k_1}\|+Z_{k_1}^{(4)}\|\mathbf{P}_{k_2}\|\right]/\left(NT\right)^{1/2},
\end{aligned}
$$

and

$$
\begin{aligned}
\left\|A_{11}^{(a)}\right\|_F^2 &= \sum_{k_1=1}^{K}\sum_{k_2=1}^{K}\left[A_{11,k_1k_2}^{(a)}\right]^2=O_P\left(K^{-2\gamma/d}+\delta_{NT}^{-2}\right)\sum_{k_1=1}^{K}\sum_{k_2=1}^{K}(NT)^{-1}\left[Z_{k_2}^{(4)}\|\mathbf{P}_{k_1}\|+Z_{k_1}^{(4)}\|\mathbf{P}_{k_2}\|\right]^2\\
&\leq O_P\left(K^{-2\gamma/d}+\delta_{NT}^{-2}\right)2\sum_{k=1}^{K}\left[Z_k^{(4)}\right]^2\sum_{k=1}^{K}(NT)^{-1}\|\mathbf{P}_k\|^2\\
&\leq O_P\left(K^{-2\gamma/d}+\delta_{NT}^{-2}\right)\sqrt{K}\left\{\sum_{k=1}^{K}\left[Z_k^{(4)}\right]^4\right\}^{1/2}\sum_{k=1}^{K}\|\mathbf{P}_k\|^2/\left(NT\right)\\
&= O_P\left[K^2\left(K^{-2\gamma/d}+\delta_{NT}^{-2}\right)\right],
\end{aligned}
$$

where we use $\sum_{k=1}^{K}[Z_k^{(4)}]^4=\frac{1}{NT}\sum_{k=1}^{K}\sum_{i=1}^{N}\sum_{t=1}^{T}Z_{it,k}^4=O_P\left(K\right)$ by Assumption 6.

For $A_{11}^{(b)}$, its (k_1,k_2)th element is given by

$$\begin{aligned}
A_{11,k_1k_2}^{(b)} &= \frac{1}{NT}\mathrm{tr}\left(\mathbf{B}_{1,k_1}^{(e)}\mathbf{B}_{1,k_2}^{(e)}\right)=\frac{1}{NT}\mathrm{tr}\left[\left(M_{\hat{\lambda}}\mathbf{P}_{k_1}^{(e)}M_{\hat{f}}-M_{\lambda^0}\mathbf{P}_{k_1}^{(e)}M_{f^0}\right)\left(M_{\hat{\lambda}}\mathbf{P}_{k_2}^{(e)}M_{\hat{f}}-M_{\lambda^0}\mathbf{P}_{k_2}^{(e)}M_{f^0}\right)\right]\\
&\le (NT)^{-1}\left[\left\|\left(P_{\lambda^0}-P_{\hat{\lambda}}\right)\mathbf{P}_{k_1}^{(e)}M_{\hat{f}}\right\|_F+\left\|M_{\hat{\lambda}}\mathbf{P}_{k_1}^{(e)}\left(P_{f^0}-P_{\hat{f}}\right)\right\|_F\right]\\
&\quad\times\left[\left\|\left(P_{\lambda^0}-P_{\hat{\lambda}}\right)\mathbf{P}_{k_2}^{(e)}M_{\hat{f}}\right\|_F+\left\|M_{\lambda}\mathbf{P}_{k_2}^{(e)}\left(P_{f^0}-P_{\hat{f}}\right)\right\|_F\right]\\
&\le (NT)^{-1}\left(\left\|P_{\lambda^0}-P_{\hat{\lambda}}\right\|^2+\left\|P_{f^0}-P_{\hat{f}}\right\|^2\right)\left\|\mathbf{P}_{k_1}^{(e)}\right\|_F\left\|\mathbf{P}_{k_2}^{(e)}\right\|_F\\
&= O_P\left(K^{-2\gamma/d}+\delta_{NT}^{-2}\right)\left[(NT)^{-1}\left\|\mathbf{P}_{k_1}^{(e)}\right\|_F\left\|\mathbf{P}_{k_2}^{(e)}\right\|_F\right]
\end{aligned}$$

where $\mathbf{B}_{1,k}^{(e)}$ is an $N\times T$ matrix with its (i,t)th element given by the kth element of $e_{it}B_{1,it}$ and $\mathbf{P}_k^{(e)}$ is an $N\times T$ matrix with its (i,t)th element $p_{it,k}e_{it}$. Then we have

$$\begin{aligned}
\left\|A_{11}^{(b)}\right\|_F^2 &= \sum_{k_1=1}^{K}\sum_{k_2=1}^{K}\left[A_{11,k_1k_2}^{(b)}\right]^2=O_P\left(K^{-4\gamma/d}+\delta_{NT}^{-4}\right)\sum_{k_1=1}^{K}\sum_{k_2=1}^{K}\left[(NT)^{-1}\left\|\mathbf{P}_{k_1}^{(e)}\right\|_F\left\|\mathbf{P}_{k_2}^{(e)}\right\|_F\right]^2\\
&\le O_P\left(K^{-4\gamma/d}+\delta_{NT}^{-4}\right)\left\{\frac{1}{NT}\sum_{k=1}^{K}\left\|\mathbf{P}_k^{(e)}\right\|_F^2\right\}^2\\
&= O_P\left(K^{-4\gamma/d}+\delta_{NT}^{-4}\right)O_P\left(K^2\right)=o_P\left(K^2\left(K^{-2\gamma/d}+\delta_{NT}^{-2}\right)\right),
\end{aligned}$$

where we use the fact that $\sum_{k=1}^K||\mathbf{P}_k^{(e)}||_F^2=O_P(NTK)$ because $\sum_{k=1}^K E[||\mathbf{P}_k^{(e)}||_F^2]=\sum_{k=1}^K\sum_{i=1}^N\sum_{t=1}^T E\left(p_{it,k}^2e_{it}^2\right)=O(NTK)$ by Assumptions 6(i) and (iii). It follows that $\|A_{11}\|_F=O_P\left(K^{1-\gamma/d}+K\delta_{NT}^{-1}\right)=o_P(1)$.

Following the study of $\left\|\tilde{W}_{NT}-W_{NT}\right\|$ in Lemma A.8, we can show that $\|A_{12}\|_F=O_P(K/\sqrt{NT})$. Consequently, $\left\|\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T e_{it}^2(\hat{Z}_{it}\hat{Z}_{it}'-\tilde{Z}_{it}\tilde{Z}_{it}')\right\|_F=O_P\left(K^{1-\gamma/d}+K\delta_{NT}^{-1}\right)$.

(ii) Write $\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T(\hat{e}_{it}^2-e_{it}^2)\hat{Z}_{it}\hat{Z}_{it}'=\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T(\hat{e}_{it}-e_{it})^2\hat{Z}_{it}\hat{Z}_{it}'+\frac{2}{NT}\sum_{i=1}^N\sum_{t=1}^T(\hat{e}_{it}-e_{it})\times e_{it}\hat{Z}_{it}\hat{Z}_{it}'\equiv A_{21}+2A_{22}$. For A_{22}, we have $A_{22,k_1k_2}=\frac{1}{NT}\mathrm{tr}(M_{\hat{f}}\tilde{\mathbf{P}}_{k_1}^{(e)\prime}M_{\hat{\lambda}}\mathbf{P}_{k_2}^{(e)})$, where $\tilde{\mathbf{P}}_{k_1}^{(e)}$ and $\mathbf{P}_{k_2}^{(e)}$ are $N\times T$ matrices with their (i,t)th elements given by $p_{it,k_1}(e_{it}-\hat{e}_{it})$ and $p_{it,k_2}e_{it}$, respectively. Noting that $|A_{22,k_1k_2}|\le\frac{1}{NT}||\tilde{\mathbf{P}}_{k_1}^{(e)}||_F||\mathbf{P}_{k_2}^{(e)}||_F$, we have

$$\begin{aligned}
\|A_{22}\|_F^2 &\le \frac{1}{N^2T^2}\sum_{k=1}^K\left\|\mathbf{P}_k^{(e)}\right\|_F^2\sum_{k=1}^K\left\|\tilde{\mathbf{P}}_k^{(e)}\right\|_F^2\\
&\le \frac{1}{N^2T^2}\left\{\sum_{k=1}^K\sum_{i=1}^N\sum_{t=1}^T p_{it,k}^2e_{it}^2\right\}\left\{\sum_{k=1}^K\sum_{i=1}^N\sum_{t=1}^T p_{it,k}^2\left(\hat{e}_{it}-e_{it}\right)^2\right\}\\
&\le \left\{\frac{1}{NT}\sum_{k=1}^K\sum_{i=1}^N\sum_{t=1}^T p_{it,k}^2e_{it}^2\right\}\left\{\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\|p_{it}\|^4\right\}^{1/2}\left\{\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\left(\hat{e}_{it}-e_{it}\right)^4\right\}^{1/2}\\
&\le O_P(K)\,O_P(K)\left\{\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\left(\hat{e}_{it}-e_{it}\right)^4\right\}^{1/2}.
\end{aligned}$$

For A_{21}, we have $\|A_{21}\|_F^2\le\frac{1}{N^2T^2}\{\sum_{k=1}^K\sum_{i=1}^N\sum_{t=1}^T p_{it,k}^2(\hat{e}_{it}-e_{it})^2\}^2\le O_P\left(K^2\right)\{\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T(\hat{e}_{it}-e_{it})^4\}$.

Now we consider the key term $\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T(\hat{e}_{it}-e_{it})^4$. By Lemma D.2, we have $\hat{e}_{it}-e_{it}=(\beta^0-\hat{\beta})'Z_{it}+\overrightarrow{e}_{it}+r_{it}$ where $\overrightarrow{e}_{it}\equiv\frac{1}{N}\sum_{j=1}^N\alpha_{ij}e_{jt}+\frac{1}{T}\sum_{s=1}^T\eta_{ts}e_{jt}-\frac{1}{NT}\sum_{j=1}^N\sum_{s=1}^T\eta_{ts}\alpha_{ij}e_{js}$, and $r_{it}\equiv$

$(\hat{\mathbf{e}}_e^{(1)})_{it} + (\hat{\mathbf{e}}^{(rem)})_{it} + (M_{\lambda^0}\mathbf{e}_g M_{f^0})_{it}$. Note that

$$\begin{aligned}
\left\|\hat{\mathbf{e}}_e^{(1)}\right\|_F^2 &\leq R\left\|\hat{\mathbf{e}}_e^{(1)}\right\|^2 = O_P\left(NT\left(\delta_{NT}^{-4} + K^{-4\gamma/d}\right)\right), && \text{(D.2)}\\
\left\|\hat{\mathbf{e}}^{(rem)}\right\|_F^2 &\leq \operatorname{rank}(\hat{\mathbf{e}}^{(rem)})\left\|\hat{\mathbf{e}}^{(rem)}\right\|^2 = O_P\left\{NT\left\|\hat{\beta} - \beta^0\right\|^2\left(\delta_{NT}^{-2} + K^{-2\gamma/d}\right)\right\}, && \text{(D.3)}\\
\left\|M_{f^0}\mathbf{e}_g' M_{\lambda^0}\right\|_F^2 &= O_P\left(NTK^{-2\gamma/d}\right), && \text{(D.4)}
\end{aligned}$$

by Lemma D.2, where we use the facts that $\left\|\hat{\beta} - \beta^0\right\| = o_P\left(\delta_{NT}^{-1} + K^{-\gamma/d}\right)$ and that $\delta_{NT}^{-2} + K^{-2\gamma/d} = o_P(||\hat{\beta} - \beta^0||)$ in the second line. Then

$$\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T(\hat{e}_{it} - e_{it})^4 \leq 9\left(\left\|\beta^0 - \hat{\beta}\right\|^4 \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\|Z_{it}\|^4 + \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \overrightarrow{e}_{it}^4 + \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T r_{it}^4\right). \tag{D.5}$$

It is easy to see that the first term in (D.5) is $O_P\left(\left\|\beta^0 - \hat{\beta}\right\|^4 K^2\right)$. For the second term, we have

$$\begin{aligned}
\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \overrightarrow{e}_{it}^4 &\leq \frac{9}{N^2}\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\left\{\frac{1}{\sqrt{N}}\sum_{j=1}^N\alpha_{ij}e_{jt}\right\}^4 + \frac{9}{T^2}\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\left\{\frac{1}{\sqrt{T}}\sum_{s=1}^T\eta_{ts}e_{jt}\right\}^4\\
&\quad + \frac{9}{N^2T^2}\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\left\{\frac{1}{\sqrt{NT}}\sum_{j=1}^N\sum_{s=1}^T\eta_{ts}\alpha_{ij}e_{js}\right\}^4\\
&= O_P\left(N^{-2}\right) + O_P\left(T^{-2}\right) + O_P\left(N^{-2}T^{-2}\right) = O_P\left(N^{-2} + T^{-2}\right),
\end{aligned}$$

where $O_P\left(N^{-2}\right)$ comes from Markov inequality and cross-sectional independence across i for e_{it} conditional on $\mathcal{D}$, and the $O_P\left(T^{-2}\right)$ and $O_P\left(N^{-2}T^{-2}\right)$ terms can be obtained by Markov inequality and the strong mixing property of $\{e_{it}, t = 1, ..., T\}$ conditional on $\mathcal{D}$. For the third term in (D.5), we use a rough bound:

$$\begin{aligned}
\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T r_{it}^4 &\leq \frac{1}{NT}\left\{\sum_{i=1}^N\sum_{t=1}^T r_{it}^2\right\}^2 \leq \frac{9}{NT}\left(\left\|M_{f^0}\mathbf{e}_g' M_{\lambda^0}\right\|_F^2 + \left\|\hat{\mathbf{e}}_e^{(1)}\right\|_F^2 + \left\|\hat{\mathbf{e}}^{(rem)}\right\|_F^2\right)^2\\
&\leq \frac{27}{NT}\left(\left\|M_{f^0}\mathbf{e}_g' M_{\lambda^0}\right\|_F^4 + \left\|\hat{\mathbf{e}}_e^{(1)}\right\|_F^4 + \left\|\hat{\mathbf{e}}^{(rem)}\right\|_F^4\right)\\
&= O_P\left[NT\left(\delta_{NT}^{-8} + K^{-8\gamma/d}\right)\right] + O_P\left[NT\left(K^{-4\gamma/d}\right)\right]\\
&\quad + O_P\left\{\left[NT\left(K^2\delta_{NT}^{-8} + K^{-4\gamma/d}\right)\left(\delta_{NT}^{-4} + K^{-4\gamma/d}\right)\right]\right\}\\
&= O_P\left(NT\delta_{NT}^{-8} + NTK^{-4\gamma/d}\right)
\end{aligned}$$

by (D.2)-(D.4). In sum, we have

$$\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T(\hat{e}_{it} - e_{it})^4 = O_P\left(\delta_{NT}^{-4} + NT\delta_{NT}^{-8} + NTK^{-4\gamma/d}\right). \tag{D.6}$$

It follows that $\|A_{21}\|_F = O_P[K\delta_{NT}^{-2} + (NT)^{1/2}K\delta_{NT}^{-4} + (NT)^{1/2}K^{1-2\gamma/d}]$ and $\|A_{22}\|_F = O_P[K\delta_{NT}^{-1} + (NT)^{1/4}K\delta_{NT}^{-2} + (NT)^{1/4}K^{1-\gamma/d}]$. Consequently, $\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\left(\hat{e}_{it}^2 - e_{it}^2\right)\hat{Z}_{it}\hat{Z}_{it}' = O_P[K\delta_{NT}^{-1} + (NT)^{1/4}K\delta_{NT}^{-2} + (NT)^{1/4}K^{1-\gamma/d}]$. ■

Lemma D.6 *Suppose that the conditions in Theorem 3.3 hold. Then we have* $\|\hat{\mathbf{e}}-\mathbf{e}\|_F = O_P(N^{1/2}+T^{1/2})$.

Proof. Note that

$$\begin{aligned}\|\hat{\mathbf{e}}-\mathbf{e}\|_F &\leq \left\|P_{\lambda^0}\mathbf{e}P_{f^0}\right\|_F + \|P_{\lambda^0}\mathbf{e}\|_F + \left\|\mathbf{e}P_{f^0}\right\|_F \\ &\quad + \left\|\hat{\mathbf{e}}_e^{(1)}\right\|_F + \left\|\hat{\beta}-\beta^0\right\| \left\|M_{\lambda^0}\mathbf{P}_{(a)}M_{f^0}\right\|_F + \left\|\hat{\mathbf{e}}^{(rem)}\right\|_F + \left\|M_{f^0}\mathbf{e}_g' M_{\lambda^0}\right\|_F\end{aligned}$$

by Lemma D.2. By Lemma A.5(ii), $\left\|P_{\lambda^0}\mathbf{e}P_{f^0}\right\|_F = O_P(1)$. By Chebyshev inequality, one can readily show that $\|P_{\lambda^0}\mathbf{e}\|_F = O_P(T^{1/2})$ and $\left\|\mathbf{e}P_{f^0}\right\|_F = O_P(N^{1/2})$. By (D.2)-(D.4), we have $\left\|\hat{\mathbf{e}}_e^{(1)}\right\|_F = O_P[\sqrt{NT}(\delta_{NT}^{-2}+K^{-2\gamma/d})]$, $\left\|\hat{\mathbf{e}}^{(rem)}\right\|_F \leq O_P[\sqrt{NT}\left\|\hat{\beta}-\beta^0\right\|(\delta_{NT}^{-1}+K^{-\gamma/d})]$, and $\left\|M_{f^0}\mathbf{e}_g' M_{\lambda^0}\right\|_F = O_P(\sqrt{NT}K^{-\gamma/d})$. In view of the fact that $\frac{1}{NT}\left\|M_{\lambda^0}\mathbf{P}_{(a)}M_{f^0}\right\|_F^2 \leq \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T Z_{it,(a)}^2 = a'W_{NT}a \leq \mu_1(W_{NT})\|a\|^2 = 1$, we have $||\hat{\beta}-\beta^0||\left\|M_{\lambda^0}\mathbf{P}_{(a)}M_{f^0}\right\|_F = O_P(\sqrt{K}\delta_{NT}^{-2}+K^{-\gamma/d})O_P(\sqrt{NT}) = o_P(\sqrt{N}+\sqrt{T})$ by (A.8). Consequently, $\|\hat{\mathbf{e}}-\mathbf{e}\|_F = O_P(\sqrt{N}+\sqrt{T})$. ■

Lemma D.7 *Suppose that the conditions in Theorem 3.3 hold. Then we have*
(i) $N^{-1}||E_{\mathcal{D}}[\mathbf{e}'M_{\lambda^0}\mathbf{e}] - (\hat{\mathbf{e}}'\hat{\mathbf{e}})^{truncD}|| = o_P[T^{5/8}(K^{-\gamma/d}+\sqrt{K}\delta_{NT}^{-2})+T^{-1/4}]$;
(ii) $T^{-1}||E_{\mathcal{D}}[\mathbf{e}M_{f^0}\mathbf{e}'] - (\hat{\mathbf{e}}\hat{\mathbf{e}}')^{truncD}|| = o_P[N^{5/8}(K^{-\gamma/d}+\sqrt{K}\delta_{NT}^{-2})+N^{-1/4}]$.

Proof. We only prove (i) as the proof of (ii) is analogous. Note that the (t,s)th element of $E_{\mathcal{D}}(\mathbf{e}'M_{\lambda^0}\mathbf{e})$ is given by

$$\sum_{i=1}^N E_{\mathcal{D}}\left\{\left(e_{it}-\frac{1}{N}\sum_{j=1}^N\alpha_{ij}e_{jt}\right)\left(e_{is}-\frac{1}{N}\sum_{j=1}^N\alpha_{ij}e_{js}\right)\right\} = 0$$

because $E_{\mathcal{D}}[e_{it}e_{js}] = 0$ for $t \neq s$, we have $E_{\mathcal{D}}(\mathbf{e}'M_{\lambda^0}\mathbf{e}) = [E_{\mathcal{D}}(\mathbf{e}'M_{\lambda^0}\mathbf{e})]^{\text{truncD}}$. Then

$$\frac{1}{N}\left\|E_{\mathcal{D}}(\mathbf{e}'M_{\lambda^0}\mathbf{e}) - (\hat{\mathbf{e}}'\hat{\mathbf{e}})^{\text{truncD}}\right\| \leq \frac{1}{N}\left\|[\mathbf{e}'M_{\lambda^0}\mathbf{e} - E_{\mathcal{D}}(\mathbf{e}'M_{\lambda^0}\mathbf{e})]^{\text{truncD}}\right\| + \frac{1}{N}\left\|[\mathbf{e}'M_{\lambda^0}\mathbf{e} - \hat{\mathbf{e}}'\hat{\mathbf{e}}]^{\text{truncD}}\right\|. \tag{D.7}$$

For the first term in (D.7), noting the tth diagonal element of $\mathbf{e}'M_{\lambda^0}\mathbf{e} - E_{\mathcal{D}}(\mathbf{e}'M_{\lambda^0}\mathbf{e})$ is given by $\left[e_{it}^2 - E_{\mathcal{D}}(e_{it}^2)\right] - \frac{2}{N}\sum_{j=1}^N\alpha_{ij}[e_{jt}e_{it} - E_{\mathcal{D}}(e_{jt}e_{it})] + \frac{1}{N^2}\sum_{j_1=1}^N\sum_{j_2=1}^N\alpha_{ij_1}\alpha_{ij_2}[e_{j_1t}e_{j_2t} - E_{\mathcal{D}}(e_{j_1t}e_{j_2t})]$, we have

$$\begin{aligned}\frac{1}{N}\left\|[\mathbf{e}'M_{\lambda^0}\mathbf{e} - E_{\mathcal{D}}(\mathbf{e}'M_{\lambda^0}\mathbf{e})]^{\text{truncD}}\right\| &\leq \max_{1\leq t\leq T}\left|\frac{1}{N}\sum_{i=1}^N\left[e_{it}^2 - E_{\mathcal{D}}(e_{it}^2)\right]\right| \\ &\quad +2\max_{1\leq t\leq T}\left|\frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N\alpha_{ij}[e_{jt}e_{it} - E_{\mathcal{D}}(e_{jt}e_{it})]\right| \\ &\quad +\max_{1\leq t\leq T}\left|\frac{1}{N^3}\sum_{i=1}^N\sum_{j_1=1}^N\sum_{j_2=1}^N\alpha_{ij_1}\alpha_{ij_2}[e_{j_1t}e_{j_2t} - E_{\mathcal{D}}(e_{j_1t}e_{j_2t})]\right| \\ &\equiv \max_{1\leq t\leq T}C_{1t} + 2\max_{1\leq t\leq T}C_{2t} + \max_{1\leq t\leq T}C_{3t},\text{ say.}\end{aligned}$$

Noting that $E|N^{-1/2}\sum_{i=1}^{N}\left[e_{it}^{2}-E_{\mathcal{D}}\left(e_{it}^{2}\right)\right]|^{4}<\infty$, we have $\max_{1\leq t\leq T}C_{1t}=o_{P}\left(N^{-1/2}T^{1/4}\right)$ by Lemma E.2. For the second term, we have

$$
\begin{aligned}
\max_{1\leq t\leq T}C_{2t} &\leq \max_{1\leq t\leq T}\left|\frac{1}{N^{2}}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_{ij}\left[e_{jt}e_{it}\right]\right|+\max_{1\leq t\leq T}\left|\frac{1}{N^{2}}\sum_{i=1}^{N}\alpha_{ii}E_{\mathcal{D}}\left(e_{it}^{2}\right)\right| \\
&\leq \frac{1}{N}\max_{1\leq t\leq T}\left|\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_{i}^{0\prime}e_{it}\right)\left(\frac{\lambda^{0\prime}\lambda^{0}}{N}\right)^{-1}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_{i}^{0\prime}e_{it}\right)\right| \\
&\quad+\left\{\frac{1}{N}\sum_{i=1}^{N}\alpha_{ii}^{2}\right\}^{1/2}\frac{1}{N}\max_{1\leq t\leq T}\left\{\frac{1}{N}\sum_{i=1}^{N}\left[E_{\mathcal{D}}\left(e_{it}^{2}\right)\right]^{2}\right\}^{1/2} \\
&\leq \frac{\varsigma_{N}^{-1}}{N}\max_{1\leq t\leq T}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_{i}^{0\prime}e_{it}\right|^{2}+o_{P}\left(N^{-1}T^{1/4}\right)=o_{P}\left(N^{-1}T^{1/2}\right)+o_{P}\left(N^{-1}T^{1/4}\right) \\
&= o_{P}\left(N^{-1}T^{1/2}\right)
\end{aligned}
$$

by the fact that $E[|N^{1/2}\sum_{i=1}^{N}\lambda_{i}^{0\prime}e_{it}|^{4}]<\infty$ and that $E\left(e_{it}^{8}\right)<\infty$. Similarly, we can show that $\max_{1\leq t\leq T}C_{3t}=o_{P}\left(N^{-1}T^{1/4}\right)$. Then we have

$$
\frac{1}{N}\left\|\left[\mathbf{e}^{\prime}M_{\lambda^{0}}\mathbf{e}-E_{\mathcal{D}}\left(\mathbf{e}^{\prime}M_{\lambda^{0}}\mathbf{e}\right)\right]^{\text{truncD}}\right\|=o_{P}\left(N^{-1/2}T^{1/4}\right). \tag{D.8}
$$

Write $\hat{\mathbf{e}}=M_{\lambda^{0}}\mathbf{e}-M_{\lambda^{0}}\mathbf{e}P_{f^{0}}+\mathbf{e}^{(REM)}$, where $\hat{\mathbf{e}}^{(REM)}=\hat{\mathbf{e}}_{e}^{(1)}+\sum_{k=1}^{K}\left(\beta_{k}^{0}-\hat{\beta}_{k}\right)\hat{\mathbf{e}}_{k}^{(1)}+\hat{\mathbf{e}}^{(rem)}+M_{\lambda^{0}}\mathbf{e}_{g}M_{f^{0}}$. Note that

$$
\begin{aligned}
\left\|\hat{\mathbf{e}}^{(REM)}\right\|_{F} &\leq \left\|\hat{\mathbf{e}}_{e}^{(1)}\right\|_{F}+\left\|\sum_{k=1}^{K}\left(\beta_{k}^{0}-\hat{\beta}_{k}\right)\hat{\mathbf{e}}_{k}^{(1)}\right\|_{F}+\left\|M_{\lambda^{0}}\mathbf{e}_{g}M_{f^{0}}\right\|_{F}+\left\|\hat{\mathbf{e}}^{(rem)}\right\|_{F} \\
&\leq R\left\|\hat{\mathbf{e}}_{e}^{(1)}\right\|+\left\|\beta^{0}-\hat{\beta}\right\|\left\|M_{\lambda^{0}}\mathbf{P}_{(a)}M_{f^{0}}\right\|_{F}+\left\|M_{\lambda^{0}}\mathbf{e}_{g}M_{f^{0}}\right\|_{F}+\left\|\hat{\mathbf{e}}^{(rem)}\right\| \\
&= O_{P}\left(\sqrt{NT}\left(K^{-2\gamma/d}+\delta_{NT}^{-2}\right)\right)+O_{P}\left(\sqrt{NT}\left(K^{1/2}\delta_{NT}^{-2}+K^{-\gamma/d}\right)\right) \\
&\quad+O_{P}\left(\sqrt{NT}K^{-\gamma/d}\right)+O_{P}\left(\sqrt{NT}\left[\left(\delta_{NT}^{-2}\sqrt{K}+K^{-\gamma/d}\right)\left(K^{1/2}\delta_{NT}^{-2}+K^{-\gamma/d}\right)\right]\right) \\
&= O_{P}\left[\sqrt{NT}\left(K^{-\gamma/d}+K^{1/2}\delta_{NT}^{-2}\right)\right].
\end{aligned}
$$

For the second term in (D.7), we have

$$
\begin{aligned}
&N^{-1}\left\|\left[\mathbf{e}^{\prime}M_{\lambda^{0}}\mathbf{e}-\hat{\mathbf{e}}^{\prime}\hat{\mathbf{e}}\right]^{\text{truncD}}\right\| \\
\leq\ &N^{-1}\left\|\left[P_{f^{0}}\mathbf{e}^{\prime}M_{\lambda^{0}}\mathbf{e}P_{f^{0}}\right]^{\text{truncD}}\right\|+2N^{-1}\left\|\left[\mathbf{e}^{\prime}M_{\lambda^{0}}\mathbf{e}P_{f^{0}}\right]^{\text{truncD}}\right\|+N^{-1}\left\|\left[\mathbf{e}^{(REM)\prime}\mathbf{e}^{(REM)}\right]^{\text{truncD}}\right\| \\
&+2N^{-1}\left\|\left[\mathbf{e}^{(REM)\prime}M_{\lambda^{0}}\mathbf{e}P_{f^{0}}\right]^{\text{truncD}}\right\|+2N^{-1}\left\|\left[\mathbf{e}^{(REM)\prime}M_{\lambda^{0}}\mathbf{e}\right]^{\text{truncD}}\right\|.
\end{aligned}
$$

Let c_{tt} be the (t,t)th element of $N^{-1}P_{f^{0}}\mathbf{e}^{\prime}M_{\lambda^{0}}\mathbf{e}P_{f^{0}}$. We have

$$
\begin{aligned}
c_{tt} &= \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{s=1}^{T}\eta_{ts}e_{is}-\frac{1}{NT}\sum_{s=1}^{T}\sum_{j=1}^{N}\alpha_{ij}\eta_{ts}e_{js}\right)^{2} \\
&\leq \frac{2}{T}\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\eta_{ts}e_{is}\right)^{2}+\frac{2}{NT}\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{\sqrt{NT}}\sum_{s=1}^{T}\sum_{j=1}^{N}\alpha_{ij}\eta_{ts}e_{js}\right)^{2}\equiv\frac{2}{T}c_{tt,1}+\frac{2}{NT}c_{tt,2},\text{ say.}
\end{aligned}
$$

For $c_{tt,1}$, we have

$$\begin{aligned}\max_{1\le t\le T}|c_{tt,1}| &= \max_{1\le t\le T}\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{s_1=1}^{T}\sum_{s_2=1}^{T} f_{s_1}^{0}\left(\frac{f^{0\prime}f^0}{T}\right)^{-1} f_t^{0\prime}f_t^0\left(\frac{f^{0\prime}f^0}{T}\right)^{-1} f_{s_2}^{0\prime}e_{is_1}e_{is_2}\right| \\ &= \max_{1\le t\le T}\operatorname{tr}\left\{f_t^{0\prime}f_t^0\left(\frac{f^{0\prime}f^0}{T}\right)^{-1}\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{s_1=1}^{T}\sum_{s_2=1}^{T} f_{s_2}^{0\prime}f_{s_1}^{0}e_{is_1}e_{is_2}\right)\left(\frac{f^{0\prime}f^0}{T}\right)^{-1}\right\} \\ &\le \varsigma_T^{-2}\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}\sum_{q=1}^{T} e_{is}f_s^{0\prime}f_q^0 e_{iq}\right)\max_{1\le t\le T}\left\{\left\|f_t^0\right\|^2\right\} \\ &= O_P(1)\,o_P\left(T^{1/4}\right)=o_P\left(T^{1/4}\right)\end{aligned}$$

because $E\left\|f_t^0\right\|^8<\infty$. Similar, we can show that $\max_{1\le t\le T}|c_{tt,2}|=o_P[(NT)^{1/4}]$. By Lemma E.3$(viii)$,

$$N^{-1}\left\|\left[P_{f^0}\mathbf{e}'M_{\lambda^0}\mathbf{e}P_{f^0}\right]^{\text{truncD}}\right\|\le\max_{1\le t\le T}|c_{tt}|\le\frac{2}{T}\max_{1\le t\le T}|c_{tt,1}|+\frac{2}{NT}\max_{1\le t\le T}|c_{tt,2}|=o_P\left(T^{-3/4}\right).$$

Similarly, we have

$$\begin{aligned}N^{-1}\left\|[\mathbf{e}'M_{\lambda^0}\mathbf{e}]^{\text{truncD}}\right\| &\le \max_{1\le t\le T}\left|\frac{1}{N}\sum_{i=1}^{N}\left(e_{it}-\frac{1}{N}\sum_{j=1}^{N}\alpha_{ij}e_{jt}\right)^2\right| \\ &\le \max_{1\le t\le T}\left|\frac{2}{N}\sum_{i=1}^{N}e_{it}^2\right|+\frac{2}{N}\max_{1\le t\le T}\left|\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\alpha_{ij}e_{jt}\right)^2\right| \\ &= o_P\left(T^{1/4}\right)+o_P\left(N^{-1}T^{1/2}\right)=o_P\left(T^{1/4}\right)\end{aligned}$$

where the first term comes from Assumption 6(i) and Lemma E.2, and the second term comes from

$$\begin{aligned}\max_{1\le t\le T}\left|\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\alpha_{ij}e_{jt}\right)^2\right| &= \max_{1\le t\le T}\left|\frac{1}{N}\sum_{j_1=1}^{N}\sum_{j_2=1}^{N}\alpha_{j_1j_2}e_{j_1t}e_{j_2t}\right| \\ &= \max_{1\le t\le T}\left(\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\lambda_j^0e_{jt}\right)\left(\frac{\lambda^{0\prime}\lambda^0}{N}\right)^{-1}\left(\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\lambda_j^0e_{jt}\right)' \\ &\le \varsigma_N^{-1}\max_{1\le t\le T}\left\|\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\lambda_j^0e_{jt}\right\|^2=o_P\left(T^{1/2}\right)\end{aligned}$$

because $E_{\mathcal{D}}\left(\left\|N^{-1/2}\sum_{j=1}^{N}\lambda_j^0e_{jt}\right\|^4\right)<\infty$. By Cauchy-Schwarz inequality, we have

$$\begin{aligned}N^{-1}\left\|\left[\mathbf{e}'M_{\lambda^0}\mathbf{e}P_{f^0}\right]^{\text{truncD}}\right\| &\le \left\{N^{-1}\left\|[\mathbf{e}'M_{\lambda^0}\mathbf{e}]^{\text{truncD}}\right\|\right\}^{1/2}\left\{N^{-1}\left\|\left[P_{f^0}\mathbf{e}'M_{\lambda^0}\mathbf{e}P_{f^0}\right]^{\text{truncD}}\right\|\right\}^{1/2} \\ &= o_P\left(T^{1/8}\right)o_P\left(T^{-3/8}\right)=o_P\left(T^{-1/4}\right).\end{aligned}$$

Note that

$$\begin{aligned}N^{-1}\left\|\left[\mathbf{e}^{(REM)\prime}\mathbf{e}^{(REM)}\right]^{\text{truncD}}\right\| &\le \frac{1}{N}\max_t\left(\sum_{i=1}^{N}\left[e_{it}^{(REM)}\right]^2\right)\le\frac{1}{N}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\left[e_{it}^{(REM)}\right]^2\right) \\ &\le \frac{1}{N}\left\|\mathbf{e}^{(REM)}\right\|_F^2\le O_P\left[T\left(K^{-2\gamma/d}+K\delta_{NT}^{-4}\right)\right]\end{aligned}$$

By Cauchy-Schwarz inequality, we have

$$
\begin{aligned}
N^{-1}\left\|\left[\mathbf{e}^{(REM)\prime}M_{\lambda^0}\mathbf{e}P_{f^0}\right]^{\text{truncD}}\right\| &\leq \left\{N^{-1}\left\|\left[\mathbf{e}^{(REM)\prime}\mathbf{e}^{(REM)}\right]^{\text{truncD}}\right\|\right\}^{1/2}\left\{N^{-1}\left\|\left[P_{f^0}\mathbf{e}'M_{\lambda^0}\mathbf{e}P_{f^0}\right]^{\text{truncD}}\right\|\right\}^{1/2} \\
&= o_P\left[\left(K^{-\gamma/d}+\sqrt{K}\delta_{NT}^{-2}\right)T^{-1/8}\right]
\end{aligned}
$$

and

$$
\begin{aligned}
N^{-1}\left\|\left[\mathbf{e}^{(REM)\prime}M_{\lambda^0}\mathbf{e}\right]^{\text{truncD}}\right\| &\leq 2\left\{N^{-1}\left\|\left[\mathbf{e}^{(REM)\prime}\mathbf{e}^{(REM)\prime}\right]^{\text{truncD}}\right\|\right\}^{1/2}\left\{N^{-1}\left\|\left[\mathbf{e}'M_{\lambda^0}\mathbf{e}\right]^{\text{truncD}}\right\|\right\}^{1/2} \\
&= o_P\left[T^{5/8}\left(K^{-\gamma/d}+\sqrt{K}\delta_{NT}^{-2}\right)\right].
\end{aligned}
$$

Finally, we have

$$
\begin{aligned}
\frac{1}{N}\left\|E_{\mathcal{D}}\left(\mathbf{e}'M_{\lambda^0}\mathbf{e}\right)-\left(\hat{\mathbf{e}}'\hat{\mathbf{e}}\right)^{\text{truncD}}\right\| &= o_P\left(T^{-3/4}\right)+o_P\left(T^{-1/4}\right)+o_P\left[T\left(K^{-2\gamma/d}+K\delta_{NT}^{-4}\right)\right] \\
&\quad +o_P\left[\left(K^{-\gamma/d}+\sqrt{K}\delta_{NT}^{-2}\right)T^{-1/8}\right]+o_P\left[T^{5/8}\left(K^{-\gamma/d}+\sqrt{K}\delta_{NT}^{-2}\right)\right] \\
&= o_P\left[T^{5/8}\left(K^{-\gamma/d}+\sqrt{K}\delta_{NT}^{-2}\right)+T^{-1/4}\right].
\end{aligned}
$$

■

Lemma D.8 *Suppose that the conditions in Theorem 3.3 hold. Then we have* $N^{-1}||\left(\hat{\mathbf{e}}'\hat{\mathbf{e}}\right)^{truncD}|| = O_P\left(T\delta_{NT}^{-2}\right)$.

Proof. By Lemmas E.3(iv), (vii), and D.6, we have

$$
\begin{aligned}
N^{-1}\left\|\left[\hat{\mathbf{e}}'\hat{\mathbf{e}}\right]^{\text{truncD}}\right\| &\leq \max_t\left|N^{-1}\sum_{i=1}^N\hat{e}_{it}^2\right| = N^{-1}\max_t\|\hat{\mathbf{e}}_{\cdot t}\|^2 \leq N^{-1}\|\hat{\mathbf{e}}'\hat{\mathbf{e}}\| \\
&\leq N^{-1}\|\hat{\mathbf{e}}\|^2 = N^{-1}\left(\|\mathbf{e}\|^2+\|\hat{\mathbf{e}}-\mathbf{e}\|^2\right) \leq N^{-1}\left(\|\mathbf{e}\|^2+\|\hat{\mathbf{e}}-\mathbf{e}\|_F^2\right) \\
&= O_P\left(T\delta_{NT}^{-2}\right)+N^{-1}O_P\left(N^{1/2}+T^{1/2}\right) = O_P\left(T\delta_{NT}^{-2}\right).
\end{aligned}
$$

■

Now we prove the main lemmas used in the proof of consistency of bias-corrected estimator.

Proof of Lemma A.8. (i) We use $\hat{W}_{NT,k_1k_2}-W_{NT,k_1k_2}$ to denote the (k_1,k_2)th element of $\hat{W}_{NT}-W_{NT}$. Noting that

$$
\begin{aligned}
\left|\hat{W}_{NT,k_1k_2}-W_{NT,k_1k_2}\right| &= \left|\frac{1}{NT}\text{tr}\left(M_{\hat{\lambda}}\mathbf{P}_{k_1}M_{\hat{f}}\mathbf{P}_{k_2}'\right)-\frac{1}{NT}\text{tr}\left(M_{\lambda^0}\mathbf{P}_{k_1}M_{f^0}\mathbf{P}_{k_2}'\right)\right| \\
&\leq \left|\frac{1}{NT}\text{tr}\left[\left(M_{\hat{\lambda}}-M_{\lambda^0}\right)\mathbf{P}_{k_1}M_{\hat{f}}\mathbf{P}_{k_2}'\right]\right|+\left|\frac{1}{NT}\text{tr}\left[M_{\lambda^0}\mathbf{P}_{k_1}\left(M_{\hat{f}}-M_{f^0}\right)\mathbf{P}_{k_2}'\right]\right| \\
&\leq \frac{2R}{NT}\left\|M_{\hat{\lambda}}-M_{\lambda^0}\right\|\|\mathbf{P}_{k_1}\|\|\mathbf{P}_{k_2}\|+\frac{2R}{NT}\left\|M_{\hat{f}}-M_{f^0}\right\|\|\mathbf{P}_{k_1}\|\|\mathbf{P}_{k_2}\| \\
&= \frac{2R}{NT}\left(\left\|M_{\hat{\lambda}}-M_{\lambda^0}\right\|+\left\|M_{\hat{f}}-M_{f^0}\right\|\right)\|\mathbf{P}_{k_1}\|\|\mathbf{P}_{k_2}\|,
\end{aligned}
$$

we have

$$\begin{aligned}\left\|\hat{W}_{NT}-W_{NT}\right\|_F &= \left[\sum_{k_1=1}^{K}\sum_{k_2=1}^{K}\left(\hat{W}_{NT,k_1k_2}-W_{NT,k_1k_2}\right)^2\right]^{1/2}\\ &\leq 2R\left(\left\|M_{\hat{\lambda}}-M_{\lambda^0}\right\|+\left\|M_{\hat{f}}-M_{f^0}\right\|\right)\left\{\sum_{k_1=1}^{K}\sum_{k_2=1}^{K}\left[\frac{1}{NT}\left\|\mathbf{P}_{k_1}\right\|\left\|\mathbf{P}_{k_2}\right\|\right]^2\right\}^{1/2}\\ &= O_P\left(K\left(\delta_{NT}^{-1}+K^{-\gamma/d}\right)\right)\text{ by Lemma D.3.}\end{aligned}$$

(ii) We decompose $\hat{\Omega}_{NT}-\tilde{\Omega}$ as follows:

$$\begin{aligned}\hat{\Omega}_{NT}-\tilde{\Omega} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{Z}_{it}\hat{Z}_{it}'\hat{e}_{it}^2-E_{\mathcal{D}}\left(\tilde{Z}_{it}\tilde{Z}_{it}'e_{it}^2\right)\\ &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\hat{Z}_{it}\hat{Z}_{it}'\left(\hat{e}_{it}^2-e_{it}^2\right)+\left(\hat{Z}_{it}\hat{Z}_{it}-\tilde{Z}_{it}\tilde{Z}_{it}'\right)e_{it}^2+\left[\tilde{Z}_{it}\tilde{Z}_{it}'e_{it}^2-E_{\mathcal{D}}\left(\tilde{Z}_{it}\tilde{Z}_{it}'e_{it}^2\right)\right]\right\}\\ &\equiv D\Omega_{NT,1}+D\Omega_{NT,2}+D\Omega_{NT,3},\text{ say.}\end{aligned}$$

By Lemmas D.5(i)-(ii), we have $\|D\Omega_{NT,1}+D\Omega_{NT,2}\|_F=O_P(K\delta_{NT}^{-1}+(NT)^{1/4}K\delta_{NT}^{-2}+(NT)^{1/4}K^{1-\gamma/d})$. Following the study of $\left\|\tilde{W}_{NT}-W_{NT}\right\|_F$, we can show that $\|D\Omega_{NT,3}\|_F=O_P(K/\sqrt{NT})$. It follows that $\left\|\hat{\Omega}_{NT}-\Omega_{NT}\right\|_F=O_P(K\delta_{NT}^{-1}+(NT)^{1/4}K\delta_{NT}^{-2}+(NT)^{1/4}K^{1-\gamma/d})$.

(iii) By Minkowski inequality

$$\begin{aligned}\left\|\hat{W}_{NT}^{-1}\hat{\Omega}_{NT}\hat{W}_{NT}^{-1}-\tilde{W}^{-1}\tilde{\Omega}\tilde{W}^{-1}\right\|_F &\leq \left\|\left(\hat{W}_{NT}^{-1}-\tilde{W}^{-1}\right)\hat{\Omega}_{NT}\hat{W}_{NT}^{-1}\right\|_F+\left\|\tilde{W}^{-1}\left(\hat{\Omega}_{NT}-\tilde{\Omega}\right)\hat{W}_{NT}^{-1}\right\|_F\\ &\quad+\left\|\tilde{W}^{-1}\tilde{\Omega}\left(\hat{W}_{NT}^{-1}-\tilde{W}^{-1}\right)\right\|_F\\ &\equiv \Pi_1+\Pi_2+\Pi_3,\text{ say.}\end{aligned}$$

By $(i)-(ii)$,

$$\begin{aligned}\Pi_1^2 &= \left\|\hat{W}_{NT}^{-1}\left(\hat{W}_{NT}-\tilde{W}\right)\tilde{W}^{-1}\hat{\Omega}_{NT}\hat{W}_{NT}^{-1}\right\|_F^2\\ &= \operatorname{tr}\left\{\tilde{W}^{-1}\left(\hat{W}_{NT}-\tilde{W}\right)\hat{W}_{NT}^{-1}\hat{\Omega}_{NT}\hat{W}_{NT}^{-1}\ \hat{W}_{NT}^{-1}\hat{\Omega}_{NT}\hat{W}_{NT}^{-1}\left(\hat{W}_{NT}-\tilde{W}\right)\tilde{W}^{-1}\right\}\\ &\leq \mu_1\left(\hat{W}_{NT}^{-1}\hat{\Omega}_{NT}\hat{W}_{NT}^{-1}\ \hat{W}_{NT}^{-1}\hat{\Omega}_{NT}\hat{W}_{NT}^{-1}\right)\operatorname{tr}\left\{\tilde{W}^{-1}\left(\hat{W}_{NT}-\tilde{W}\right)\left(\hat{W}_{NT}-\tilde{W}\right)\tilde{W}^{-1}\right\}\\ &\leq \mu_1\left(\hat{W}_{NT}^{-1}\hat{\Omega}_{NT}\hat{W}_{NT}^{-1}\ \hat{W}_{NT}^{-1}\hat{\Omega}_{NT}\hat{W}_{NT}^{-1}\right)\left[\mu_{\min}\left(\tilde{W}\right)\right]^{-2}\left\|\hat{W}_{NT}-\tilde{W}\right\|_F^2\\ &= O_P(1)\,O_P(1)\left[O_P\left(K\left(\delta_{NT}^{-1}+K^{-\gamma/d}\right)\right)\right]^2.\end{aligned}$$

So $\Pi_1=O_P\left(K\left(\delta_{NT}^{-1}+K^{-\gamma/d}\right)\right)$. Analogously, we can show that $\Pi_2=O_P(K\delta_{NT}^{-1}+(NT)^{1/4}K\delta_{NT}^{-2}+(NT)^{1/4}K^{1-\gamma/d})$ and $\Pi_3=O_P\left(K\left(\delta_{NT}^{-1}+K^{-\gamma/d}\right)\right)$. It follows that $\left\|\hat{W}_{NT}^{-1}\hat{\Omega}_{NT}\hat{W}_{NT}^{-1}-\tilde{W}^{-1}\tilde{\Omega}\tilde{W}^{-1}\right\|_F=O_P(K\delta_{NT}^{-1}+(NT)^{1/4}K\delta_{NT}^{-2}+(NT)^{1/4}K^{1-\gamma/d})$. ■

Proof of Lemma A.9. (i) Note that b_1 can be rewritten as follows

$$\begin{aligned}b_1 &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{1\leq t<s\leq T}^{T}\eta_{ts}E_{\mathcal{D}}\left(p_{is}e_{it}\right)\\ &= \frac{1}{NT}\sum_{i=1}^{N}\left\{\sum_{1\leq t<s\leq T,s-t>M_T}+\sum_{1\leq t<s\leq\min(t+M_T,T)}\right\}\eta_{ts}E_{\mathcal{D}}\left(p_{is}e_{it}\right)=b_1^{(1)}+b_1^{(2)},\text{ say.}\end{aligned}$$

Noting that $\|E_{\mathcal{D}}(p_{is}e_{it})\| \leq 8K^{1/2}\varphi_{is,8+4\delta}\|e_{it}\|_{8+4\delta,\mathcal{D}}\alpha_{\mathcal{D}}^{(3+2\delta)/(4+2\delta)}(s-t)$ by the conditional Davydov inequality where $\varphi_{is,q} \equiv K^{-1/q}\|p_{is}\|_{q,\mathcal{D}}$, we have

$$\begin{aligned}
\left\|b_1^{(1)}\right\| &\leq \frac{8\varsigma_T^{-1}}{NT}\sum_{i=1}^{N}\sum_{1\leq t<s\leq T, s-t>M_T}\left\|f_t^0\right\|\left\|f_s^0\right\|K^{1/2}\varphi_{is,8+4\delta}\|e_{it}\|_{8+4\delta,\mathcal{D}}\,\alpha_{\mathcal{D}}^{\frac{3+2\delta}{4+2\delta}}(s-t) \\
&\leq \frac{4\varsigma_T^{-1}K^{1/2}}{NT}\sum_{i=1}^{N}\sum_{1\leq t<s\leq T, s-t>M_T}\left(\|e_{it}\|_{8+4\delta,\mathcal{D}}^2\left\|f_t^0\right\|^2+\varphi_{is,8+4\delta}^2\left\|f_s^0\right\|^2\right)\alpha_{\mathcal{D}}^{\frac{3+2\delta}{4+2\delta}}(s-t) \\
&\leq \frac{4\varsigma_T^{-1}K^{1/2}}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|f_t^0\right\|^2\left(\|e_{it}\|_{8+4\delta,\mathcal{D}}^2+\varphi_{it,8+4\delta}^2\right)\sum_{m=M_T+1}^{T-1}\alpha_{\mathcal{D}}^{\frac{3+2\delta}{4+2\delta}}(m) \\
&= O_P\left(K^{1/2}\sum_{m=M_T+1}^{\infty}\alpha_{\mathcal{D}}^{\frac{3+2\delta}{4+2\delta}}(m)\right) = o_P\left(K^{1/4}\right).
\end{aligned}$$

Now, we decompose $\hat{b}_1 - b_1^{(2)}$ as follows:

$$\begin{aligned}
\hat{b}_1 - b_1^{(2)} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{1\leq t<s\leq \min(t+M_T,T)}\left[\hat{\eta}_{ts}p_{is}\hat{e}_{it}-\eta_{ts}E_{\mathcal{D}}(p_{is}e_{it})\right] \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{1\leq t<s\leq \min(t+M_T,T)}\left\{(\hat{\eta}_{ts}p_{is}\hat{e}_{it}-\eta_{ts}p_{is}e_{it})+\eta_{ts}\left[p_{is}e_{it}-E_{\mathcal{D}}(p_{is}e_{it})\right]\right\} \\
&\equiv Db_1 + Db_2, \text{ say.}
\end{aligned}$$

For Db_2, let $\zeta_{i,ts} \equiv p_{is}e_{it}$ and $\zeta_{i,ts}^c \equiv p_{is}e_{it} - E_{\mathcal{D}}(p_{is}e_{it})$. Then $E_{\mathcal{D}}(Db_2) = 0$ and

$$E_{\mathcal{D}}\left[\|Db_2\|^2\right] = \frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{1\leq t_1<s_1\leq \min(t_1+M_T,T)}\sum_{1\leq t_2<s_2\leq \min(t_2+M_T,T)}\eta_{t_1s_1}\eta_{t_2s_2}E_{\mathcal{D}}\left(\zeta_{i,t_1s_1}^{c\prime}\zeta_{i,t_2s_2}^{c}\right).$$

We consider two cases for the time indices $\{t_1, s_1, t_2, s_2\}$ inside the last summation: (a) $s_1 < t_2$ or $s_2 < t_1$; (b) all the remaining cases. Let EDb_{21a} and EDb_{21b} denote $E_{\mathcal{D}}[\|Db_2\|^2]$ when the summation is restricted to the time indices in these two cases, respectively. Then $E_{\mathcal{D}}[\|Db_2\|^2] = EDb_{21a} + EDb_{21b}$. For case (a), the two intervals (t_1, s_1) and (t_2, s_2) are separated from each other. Wlog we assume that $s_1 < t_2$. Then by the conditional Davydov and Jensen inequalities, we have

$$\begin{aligned}
\left|E_{\mathcal{D}}\left(\zeta_{i,t_1s_1}^{c\prime}\zeta_{i_2,t_2s_2}^{c}\right)\right| &\leq 8\left\|\zeta_{i,t_1s_1}^{c}\right\|_{4+2\delta,\mathcal{D}}\left\|\zeta_{i_2,t_2s_2}^{c}\right\|_{4+2\delta,\mathcal{D}}\alpha_{\mathcal{D}}^{\frac{1+\delta}{2+\delta}}(t_2-s_1) \\
&\leq 32\|p_{is_1}e_{it_1}\|_{4+2\delta,\mathcal{D}}\|p_{i_2s_2}e_{i_2t_2}\|_{4+2\delta,\mathcal{D}}\alpha_{\mathcal{D}}^{\frac{1+\delta}{2+\delta}}(t_2-s_1) \\
&\leq 32K\varphi_{is_1,8+4\delta}\|e_{it_1}\|_{8+4\delta,\mathcal{D}}\varphi_{i_2s_2,8+4\delta}\|e_{i_2t_2}\|_{8+4\delta,\mathcal{D}}\alpha_{\mathcal{D}}^{\frac{1+\delta}{2+\delta}}(t_2-s_1).
\end{aligned}$$

It follows that

$$\begin{aligned}
&\left|E_{\mathcal{D}}\left(\zeta_{i,t_1s_1}^{c\prime}\zeta_{i_2,t_2s_2}^{c}\right)\right|\left|\eta_{t_1s_1}\right|\left|\eta_{t_2s_2}\right| \\
\leq\;& 32\varsigma_T^{-2}K\left\|f_{t_1}^0\right\|\left\|f_{t_2}^0\right\|\left\|f_{s_1}^0\right\|\left\|f_{s_2}^0\right\|\varphi_{is_1,8+4\delta}\|e_{it_1}\|_{8+4\delta,\mathcal{D}}\varphi_{i_2s_2,8+4\delta}\|e_{i_2t_2}\|_{8+4\delta,\mathcal{D}}\alpha_{\mathcal{D}}^{\frac{1+\delta}{2+\delta}}(t_2-s_1) \\
\leq\;& 8\varsigma_T^{-2}K\left(C_{1,it_1,e}+C_{2,it_2,e}+\tilde{C}_{1,is_1,p}+\tilde{C}_{1,is_2,p}\right)\alpha_{\mathcal{D}}^{\frac{1+\delta}{2+\delta}}(t_2-s_1)
\end{aligned}$$

where $\tilde{C}_{1,is,p} \equiv \left\|f_s^0\right\|^4 \varphi_{is,8+4\delta}^4$. Then similarly to the proof of Lemma A.4, we can show that

$$\begin{aligned}
|E_{\mathcal{D}} Db_{21a}| &\leq 2\frac{8\varsigma_T^{-2}K}{N^2T^2}\sum_{i=1}^{N}\sum_{\substack{1\leq t_1<s_1\leq\min(t_1+M_T,T)\\1\leq t_2<s_2\leq\min(t_2+M_T,T)}}\left(C_{1,it_1,e}+C_{2,it_2,e}+\tilde{C}_{1,is_1,p}+\tilde{C}_{1,is_2,p}\right)\alpha_{\mathcal{D}}^{\frac{1+\delta}{2+\delta}}(t_2-s_1)\\
&= O_P\left(KM_T^2/(NT)\right).
\end{aligned}$$

For case (b), it is easy to see that $\max(s_1,s_2)-\min(t_1,t_2)\leq 3M_T$. Each term in the summation is bounded by $\frac{1}{N^2T^2}\left|\eta_{t_1s_1}\right|\left|\eta_{t_2s_2}\right|\mathrm{Var}_{\mathcal{D}}^{1/2}(p_{is_1}e_{it_1})\mathrm{Var}_{\mathcal{D}}^{1/2}(p_{is_2}e_{it_2})$, and the number of such terms is of order $O\left(TM_T^3\right)$. By Markov inequality, $EDb_{21b}=O_P\left(TM_T^3K/\left(NT^2\right)\right)=O_P\left(M_T^3K/(NT)\right)$. Consequently, $E_{\mathcal{D}}[\|Db_2\|^2]=O_P(M_T^2K/(NT)+M_T^3K/(NT))=O_P(M_T^3K/(NT))$ and $\|Db_2\|=O_P(\sqrt{M_T^3K/(NT)})$ by Chebyshev inequality.

For Db_1, we have $Db_1=\frac{1}{NT}\sum_{i=1}^{N}\sum_{1\leq t<s\leq\min(t+M_T,T)}[(\hat{\eta}_{ts}-\eta_{ts})p_{is}e_{it}+\eta_{ts}p_{is}(\hat{e}_{it}-e_{it})+(\hat{\eta}_{ts}-\eta_{ts})\times p_{is}(\hat{e}_{it}-e_{it})]\equiv Db_{11}+Db_{12}+Db_{13}$, say. For Db_{11}, we have by Cauchy-Schwarz inequality and Lemma D.3(ii),

$$\begin{aligned}
\|Db_{11}\| &\leq \left\{\frac{1}{NT}\sum_{i=1}^{N}\sum_{1\leq t<s\leq\min(t+M_T,T)}(\hat{\eta}_{ts}-\eta_{ts})^2\right\}^{1/2}\left\{\frac{1}{NT}\sum_{i=1}^{N}\sum_{1\leq t<s\leq\min(t+M_T,T)}\|p_{is}e_{it}\|^2\right\}^{1/2}\\
&\leq \left\{\frac{1}{T}\sum_{1\leq t,s\leq T}\sum(\hat{\eta}_{ts}-\eta_{ts})^2\right\}^{1/2}\left\{\frac{1}{NT}\sum_{i=1}^{N}\sum_{1\leq t<s\leq\min(t+M_T,T)}\|p_{is}e_{it}\|^2\right\}^{1/2}\\
&= \left\|P_{\hat{f}}-P_{f^0}\right\|_F O_P\left[(M_TK)^{1/2}\right]\leq\sqrt{\mathrm{rank}\left(P_{\hat{f}}-P_{f^0}\right)}\left\|P_{\hat{f}}-P_{f^0}\right\|O_P\left[(M_TK)^{1/2}\right]\\
&= O_P\left(\left(\delta_{NT}^{-1}+K^{-\gamma/d}\right)\sqrt{M_TK}\right).
\end{aligned}$$

Similarly, by Cauchy-Schwarz inequality and Lemmas D.6 and D.3(ii) , we have

$$\begin{aligned}
\|Db_{12}\| &\leq \frac{1}{NT}\sum_{i=1}^{N}\sum_{1\leq t<s\leq\min(t+M_T,T)}\|\eta_{ts}p_{is}\|\,|\hat{e}_{it}-e_{it}|\\
&\leq \left\{\frac{1}{NT}\sum_{i=1}^{N}\sum_{1\leq t<s\leq\min(t+M_T,T)}\eta_{ts}^2\|p_{is}\|^2\right\}^{1/2}\left\{\frac{1}{NT}\sum_{i=1}^{N}\sum_{1\leq t<s\leq\min(t+M_T,T)}(\hat{e}_{it}-e_{it})^2\right\}^{1/2}\\
&= O_P\left(\sqrt{M_TK}\right)\left\{\frac{M_T}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\hat{e}_{it}-e_{it})^2\right\}^{1/2}\\
&= O_P\left(\sqrt{M_TK}\right)\sqrt{M_T/(NT)}\,\|\hat{\mathbf{e}}-\mathbf{e}\|_F=O_P\left(M_T\sqrt{K}\delta_{NT}^{-1}\right),
\end{aligned}$$

and

$$\begin{aligned}
\|Db_{13}\| &\leq \frac{1}{NT}\sum_{i=1}^{N}\sum_{1\leq t<s\leq\min(t+M_T,T)}|\hat{\eta}_{ts}-\eta_{ts}|\,\|p_{is}\|\,|\hat{e}_{it}-e_{it}|\\
&\leq \max_{i,s}\|p_{is}\|\left\{\frac{1}{T}\sum_{1\leq t<s\leq\min(t+M_T,T)}|\hat{\eta}_{ts}-\eta_{ts}|^2\right\}^{1/2}\left\{\frac{1}{NT}\sum_{i=1}^{N}\sum_{1\leq t<s\leq\min(t+M_T,T)}(\hat{e}_{it}-e_{it})^2\right\}^{1/2}\\
&\leq O_P\left[(NT)^{1/8}\sqrt{K}\right]\left\|P_{\hat{f}}-P_{f^0}\right\|_F\sqrt{\frac{M_T}{NT}}\,\|\hat{\mathbf{e}}-\mathbf{e}\|_F\\
&= O_P\left[(NT)^{1/8}\sqrt{K}\right]O_P\left(\delta_{NT}^{-1}+K^{-\gamma/d}\right)O_P\left(\sqrt{M_T}\delta_{NT}^{-1}\right)=o_P\left(M_T\sqrt{K}\delta_{NT}^{-1}\right).
\end{aligned}$$

Consequently, $\|Db_1\| = O_P(M_T\sqrt{K}\delta_{NT}^{-1})$ and $||\hat{b}_1 - b_1^{(2)}|| \leq \|Db_1\| + \|Db_2\| = O_P(M_T\sqrt{K}\delta_{NT}^{-1})$. This completes the proof of (i).

(ii) Recall that $b_{2,k} = T^{-1}\text{tr}[E_{\mathcal{D}}(\mathbf{ee}')\,M_{\lambda^0}\mathbf{P}_k\Phi]$ and $\hat{b}_{2,k} = T^{-1}\text{tr}\Big[(\hat{\mathbf{e}}\hat{\mathbf{e}}')^{\text{truncD}}M_{\hat{\lambda}}\mathbf{P}_k\hat{\Phi}\Big]$. Then by Lemmas D.2, D.3, D.7, and A.9, we have

$$\begin{aligned}
\left|\hat{b}_{2,k} - b_{2,k}\right| &= \frac{1}{T}\text{tr}\left[\left(\hat{\mathbf{e}}\hat{\mathbf{e}}'\right)^{\text{truncD}} M_{\hat{\lambda}}\mathbf{P}_k\hat{\Phi}\right] - \frac{1}{T}\text{tr}\left[E_{\mathcal{D}}\left(\mathbf{ee}'\right)M_{\lambda^0}\mathbf{P}_k\Phi\right] \\
&= \frac{1}{T}\text{tr}\left[\left(\hat{\mathbf{e}}\hat{\mathbf{e}}'\right)^{\text{truncD}} M_{\hat{\lambda}}\mathbf{P}_k\left(\hat{\Phi}-\Phi\right)\right] + \frac{1}{T}\text{tr}\left[\left(\hat{\mathbf{e}}\hat{\mathbf{e}}'\right)^{\text{truncD}}\left(M_{\hat{\lambda}} - M_{\lambda^0}\right)\mathbf{P}_k\Phi\right] \\
&\quad + \frac{1}{T}\text{tr}\left\{\left[\left(\hat{\mathbf{e}}\hat{\mathbf{e}}'\right) - E_{\mathcal{D}}\left(\mathbf{e}M_{f^0}\mathbf{e}'\right)\right]^{\text{truncD}} M_{\lambda^0}\mathbf{P}_k\Phi\right\} \\
&\quad + \frac{1}{T}\text{tr}\left\{\left[E_{\mathcal{D}}\left(\mathbf{ee}'\right) - E_{\mathcal{D}}\left(\mathbf{e}M_{f^0}\mathbf{e}'\right)^{\text{truncD}}\right] M_{\lambda^0}\mathbf{P}_k\Phi\right\} \\
&\leq \frac{R}{T}\|\mathbf{P}_k\|\left[\left\|M_{\hat{\lambda}}\right\|\left\|\hat{\Phi}-\Phi\right\| + \left\|P_{\hat{\lambda}} - P_{\lambda^0}\right\|\|\Phi\|\right]\left\|\left(\hat{\mathbf{e}}\hat{\mathbf{e}}'\right)^{\text{truncD}}\right\| \\
&\quad + R\|M_{\lambda^0}\|\|\mathbf{P}_k\|\|\Phi\|\frac{1}{T}\left\{\left\|\left[E_{\mathcal{D}}\left(\mathbf{ee}'\right) - E_{\mathcal{D}}\left(\mathbf{e}M_{f^0}\mathbf{e}'\right)^{\text{truncD}}\right]\right\| + \left\|E_{\mathcal{D}}\left(\mathbf{e}P_{f^0}\mathbf{e}'\right)^{\text{truncD}}\right\|\right\} \\
&= \frac{\|\mathbf{P}_k\|}{\sqrt{NT}}O_P\left\{N\delta_{NT}^{-2}\left(K^{-\gamma/d} + \delta_{NT}^{-1}\right) + N^{-1/4} + N^{5/8}\left(K^{-\gamma/d} + \sqrt{K}\delta_{NT}^{-2}\right) + T^{-1}N^{1/2}\right\} \\
&= \frac{\|\mathbf{P}_k\|}{\sqrt{NT}}O_P\left\{N^{-1/4} + N^{5/8}\left(K^{-\gamma/d} + \sqrt{K}\delta_{NT}^{-2}\right) + T^{-1}N^{1/2}\right\}
\end{aligned}$$

where we also use the fact that $\left\|E_{\mathcal{D}}\left(\mathbf{e}P_{f^0}\mathbf{e}'\right)^{\text{truncD}}\right\| \leq \max_{1\leq i\leq N}\left|\frac{1}{T}\sum_{t=1}^T E_D\left[\frac{1}{T}\sum_{s=1}^T \eta_{ts}e_{is}\right]^2\right| = \frac{1}{T}\max_{1\leq i\leq N}\frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T \eta_{ts}^2 E_D\left(e_{is}^2\right) = O_P\left(T^{-1}N^{1/2}\right)$ because $E\left|T^{-2}\sum_{t=1}^T\sum_{s=1}^T \eta_{ts}^2 E_D\left(e_{is}^2\right)\right|^2 < \infty$. It follow that

$$\begin{aligned}
\left\|\hat{b}_2 - b_2\right\| &= \left\{\sum_{k=1}^K\left|\hat{b}_{2,k} - b_{2,k}\right|^2\right\}^{1/2} \\
&= \left\{\frac{1}{NT}\sum_{k=1}^K\|\mathbf{P}_k\|^2\right\}^{1/2} O_P\left\{N^{-1/4} + N^{5/8}\left(K^{-\gamma/d} + \sqrt{K}\delta_{NT}^{-2}\right) + T^{-1}N^{1/2}\right\} \\
&= O_P\left\{\sqrt{K}\left[N^{-1/4} + N^{5/8}\left(K^{-\gamma/d} + \sqrt{K}\delta_{NT}^{-2}\right) + T^{-1}N^{1/2}\right]\right\}.
\end{aligned}$$

(iii) The proof is analogous to that of (ii) by using Lemmas D.3, D.4, A.8, and A.9. ■

D.4 Specification test

To establish the asymptotic distribution of our test statistic, we need to study the behavior of the linear estimator $\hat{g}^{(l)}(x)$ under $\mathbb{H}_1(\gamma_{NT})$. Recall Υ_{NT} is a $d\times 1$ vector whose kth element is given by $\Upsilon_{NT,k} \equiv \frac{1}{NT}\text{tr}\big(M_{\lambda^0}\mathbf{X}_k M_{f^0}\mathbf{\Delta}'\big)$ and D_{NT} is defined in (4.4). Let $C_{l,NT}^{(1)}$, and $C_{l,NT}^{(2)}$ be $d\times 1$ vectors whose kth elements are respectively given by

$$C_{l,NT,k}^{(1)} \equiv \frac{1}{NT}\text{tr}\left(M_{\lambda^0}\mathbf{X}_k M_{f^0}\boldsymbol{\varepsilon}'\right), \tag{D.9}$$

$$\begin{aligned}
C_{l,NT,k}^{(2)} &\equiv -\frac{1}{NT}\text{tr}\left(\mathbf{X}_k\Phi'\boldsymbol{\varepsilon}M_{f^0}\boldsymbol{\varepsilon}'M_{\lambda^0} + \mathbf{X}_k M_{f^0}\boldsymbol{\varepsilon}'M_{\lambda^0}\boldsymbol{\varepsilon}\Phi' + \mathbf{X}_k M_{f^0}\boldsymbol{\varepsilon}'\Phi\boldsymbol{\varepsilon}M_{\lambda^0}\right) && \text{(D.10)} \\
&\equiv C_{l,NT,k}^{(2,a)} + C_{l,NT,k}^{(2,b)} + C_{l,NT,k}^{(2,b)}, \text{ say,} && \text{(D.11)}
\end{aligned}$$

where $\boldsymbol{\varepsilon}$ is an $N \times T$ matrix whose (i,t)th element is $\varepsilon_{it} = e_{it} + \gamma_{NT}\Delta(X_{it})$. Let $\hat{\theta}$ be Moon and Weidner's (2010, 2012) estimate for θ^0 without bias-correction. Following Su, Jin, and Zhang (2013), we can show that under $\mathbb{H}_1(\gamma_{NT})$ with $\gamma_{NT} = O(K^{1/4}/\sqrt{NT})$

$$\hat{\theta} - \theta^0 = \gamma_{NT}^{-1} D_{NT}^{-1} \Upsilon_{NT} + D_{NT}^{-1}\left(C_{l,NT}^{(1)} + C_{l,NT}^{(2)}\right) + \tilde{R}_{NT},$$

where $\tilde{R}_{NT} = O_P[(\gamma_{NT} + \delta_{NT}^{-2})(\gamma_{NT}^{1/2} + \delta_{NT}^{-1/2})] = o_P((NT)^{-1/2})$. Further, we can modify the proof of Theorem 3.2 to show that

$$\sqrt{NT}\left(\hat{\theta} - \theta^0 - \gamma_{NT}^{-1} D_{NT}^{-1} \Upsilon_{NT}\right) - B^{(l)} \xrightarrow{d} N(0, V_{\theta^0})$$

where $B^{(l)} \equiv -D^{-1}(\kappa_{NT} b_1^{(l)} + \kappa_{NT}^{-1} b_2^{(l)} + \kappa_{NT} b_3^{(l)})$, $b_1^{(l)}$, $b_2^{(l)}$, and $b_3^{(l)}$ are all $d \times 1$ vectors and their kth elements are defined in (B.1), $D = E_{\mathcal{D}}[D_{NT}]$, and V_{θ^0} is positive definite.

Our asymptotic analysis indicates it is not necessary to use the bias-corrected linear estimator for θ. In order for this term related to $B^{(l)}$ to be asymptotically negligible under both $\mathbb{H}_0$ and $\mathbb{H}_1(\gamma_{NT})$, we need $B^{(l)} = o_P(K^{1/4})$. Under Assumption 12, we have $B^{(l)} = O_P\{\max(\kappa_{NT}, \kappa_{NT}^{-1})\} = o_P(K^{1/4})$. But if we make bias correction, $B^{(l)}$ can be corrected up to order $o_P(1)$ and then the finite sample performance of our test can be improved. After obtaining $\hat{\theta}$, we obtain the estimators $\hat{f}_{(l)}$, $\hat{\lambda}_{(l)}$ and $\hat{\mathbf{e}}^{(l)}$ under the same identification restrictions as Bai (2009), and then use them to obtain estimates of the three bias terms, i.e., $\hat{b}_1^{(l)}$, $\hat{b}_2^{(l)}$, and $\hat{b}_3^{(l)}$, which are analogously defined as $\hat{b}_1$, $\hat{b}_2$, and $\hat{b}_3$ but with the sieve estimates of $(\lambda^0, f^0, \mathbf{e})$ being replaced by Moon and Weidner's (2010) linear estimates. Let $\hat{D}_{NT}$ be a $d \times d$ matrix whose (k_1, k_2)th element is given by $\hat{D}_{NT,k_1k_2} \equiv \frac{1}{NT}\text{tr}(M_{\hat{\lambda}^{(l)}} \mathbf{X}_{k_1} M_{\hat{f}^{(l)}} \mathbf{X}_{k_2}')$. Define the bias-corrected estimator $\hat{\theta}_{bc} \equiv \hat{\theta} + \hat{D}_{NT}^{-1}(T^{-1}\hat{b}_1^{(l)} + N^{-1}\hat{b}_2^{(l)} + T^{-1}\hat{b}_3^{(l)})$.

Proof of Lemma B.1. The proof is similar to that of Lemma A.8. ■

Proof of Lemma B.2. Recall that $\hat{\beta}_{bc} = \hat{\beta} + \hat{W}_{NT}^{-1}(T^{-1}\hat{b}_1 + N^{-1}\hat{b}_2 + T^{-1}\hat{b}_3)$ by (3.13). By (A.5) and (3.3)-(3.5), $\hat{\beta} - \beta^0 = W_{NT}^{-1}\frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T Z_{it} u_{it} + W_{NT}^{-1}[C_{NT}^{(2,a)} + C_{NT}^{(2,b)} + C_{NT}^{(2,c)}] + R_{NT}$. Decompose $\hat{\beta}_{bc} - \beta^0$ as follows

$$\begin{aligned}
\hat{\beta}_{bc} - \beta^0 &= \left\{W_{NT}^{-1}\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T Z_{it}e_{it} + \frac{1}{T}\hat{W}_{NT}^{-1}\hat{b}_1\right\} + \left\{W_{NT}^{-1}C_{NT}^{(2,a)} + \frac{1}{N}\hat{W}_{NT}^{-1}\hat{b}_2\right\} \\
&\quad + \left\{W_{NT}^{-1}C_{NT}^{(2,b)} + \frac{1}{T}\hat{W}_{NT}^{-1}\hat{b}_3\right\} + \left\{\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T W_{NT}^{-1}Z_{it}e_{g,it} + W_{NT}^{-1}C_{NT}^{(2,c)} + R_{NT}\right\} \\
&\equiv \mathcal{B}_{NT1} + \mathcal{B}_{NT2} + \mathcal{B}_{NT3} + \mathcal{B}_{NT4}, \text{ say.}
\end{aligned}$$

We complete the proof by showing that (i) $\mathcal{B}_{NT1} = \tilde{W}^{-1}\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \tilde{Z}_{it}e_{it} + o_P(\gamma_{NT})$, and (ii) $\mathcal{B}_{NTs} = o_P(\gamma_{NT})$ for $s = 2, 3, 4$. We first study $\mathcal{B}_{NT1}$. Note that

$$\begin{aligned}
&\mathcal{B}_{NT1} - \tilde{W}^{-1}\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \tilde{Z}_{it}e_{it} \\
=\ & W_{NT}^{-1}\left(\tilde{W} - W_{NT}\right)\tilde{W}^{-1}\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \tilde{Z}_{it}e_{it} + \left\{W_{NT}^{-1}\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\left(Z_{it} - \tilde{Z}_{it}\right)e_{it} + \frac{1}{T}\hat{W}_{NT}^{-1}\hat{b}_1\right\} \\
\equiv\ & \mathcal{B}_{NT11} + \mathcal{B}_{NT12}, \text{ say.}
\end{aligned}$$

By Lemma E.3(iii) and Assumption 7, we have

$$\|\mathcal{B}_{NT11}\| = \left\|W_{NT}^{-1}\right\|\left\|\tilde{W} - W_{NT}\right\|_F\left\|\tilde{W}^{-1}\right\|\left\|\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \tilde{Z}_{it}'e_{it}\right\| = O_P\left(\frac{K}{\sqrt{NT}}\sqrt{\frac{K}{NT}}\right) = o_P(\gamma_{NT}).$$

For $\mathcal{B}_{NT12}$, we have

$$\begin{aligned}\mathcal{B}_{NT12} &= W_{NT}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\left(Z_{it}-\tilde{Z}_{it}\right)e_{it}-E_{\mathcal{D}}\left[\left(Z_{it}-\tilde{Z}_{it}\right)e_{it}\right]\right\}\\ &\quad+\left\{W_{NT}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}E_{\mathcal{D}}\left(Z_{it}e_{it}\right)+\frac{1}{T}\hat{W}_{NT}^{-1}\hat{b}_1\right\}\equiv\mathcal{B}_{NT12a}+\mathcal{B}_{NT12b},\text{ say.}\end{aligned}$$

Following the proof of Lemma A.4, we can readily show that $\mathcal{B}_{NT12a}=O_P\left(\sqrt{\frac{K}{NT}}\delta_{NT}^{-1}\right)=o_P\left(\gamma_{NT}\right)$. By Lemmas A.8, A.9, and (D.9), we have

$$\begin{aligned}\mathcal{B}_{NT12b} &= \frac{1}{T}\left(\hat{W}_{NT}^{-1}-W_{NT}^{-1}\right)b_1+\frac{1}{T}\hat{W}_{NT}^{-1}\left(\hat{b}_1-b_1\right)+W_{NT}^{-1}\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{1\le t<s\le T}\alpha_{ii}\eta_{ts}E_{\mathcal{D}}\left(p_{is}e_{it}\right)\\ &= \frac{1}{T}O_P\left(K^{3/2}\left(\delta_{NT}^{-1}+K^{-\gamma/d}\right)\right)+\frac{1}{T}O_P\left(\sqrt{K}\sum_{\tau=M_T}^{T}\alpha_{\mathcal{D}}^{\frac{3+2\delta}{4+2\delta}}(\tau)+M_T\sqrt{K}\delta_{NT}^{-1}\right)+O_P\left(\frac{\sqrt{K}}{NT}\right)\\ &= o_P\left(\gamma_{NT}\right)\end{aligned}$$

under Assumption 12. Consequently, $\mathcal{B}_{NT12}=o_P\left(\gamma_{NT}\right)$ and (i) follows.

For $\mathcal{B}_{NT2}$, we decompose it as follows: $\mathcal{B}_{NT2}=\frac{1}{N}(\hat{W}_{NT}^{-1}\hat{b}_2-W_{NT}^{-1}b_2)+W_{NT}^{-1}[C_{NT}^{(2,a)}-\frac{1}{N}b_2]\equiv\mathcal{B}_{NT2a}+\mathcal{B}_{NT2b}$, say. As in the study of $\mathcal{B}_{NT12b}$,

$$\begin{aligned}\left\|\mathcal{B}_{NT2a}\right\| &\le \frac{1}{N}\left\|\hat{W}_{NT}^{-1}-W_{NT}^{-1}\right\|\left\|b_2\right\|+\frac{1}{N}\left\|\hat{W}_{NT}^{-1}\right\|\left\|\hat{b}_2-b_2\right\|\\ &= \frac{1}{N}O_P\left[K\left(K^{-\gamma/d}+\delta_{NT}^{-1}\right)\right]+\frac{1}{N}o_P\left(\kappa_{NT}K^{1/4}\right)=o_P\left(\gamma_{NT}\right)\end{aligned}$$

by Lemmas A.8 and A.9, and Assumption 12. For $\mathcal{B}_{NT2b}$, recall that

$$\begin{aligned}C_{NT,k}^{(2,a)}+\frac{1}{N}b_{2,k} &= -\frac{1}{NT}\operatorname{tr}\left(\mathbf{u}\mathbf{u}'M_{\lambda^0}\mathbf{P}_{(k)}\Phi\right)+\frac{1}{N}b_{2,k}+\frac{1}{NT}\operatorname{tr}\left(\mathbf{u}P_{f^0}\mathbf{u}'M_{\lambda^0}\mathbf{P}_{(k)}\Phi\right)\\ &= -\frac{1}{NT}\operatorname{tr}\left\{\left[\mathbf{e}\mathbf{e}'-E_{\mathcal{D}}\left(\mathbf{e}\mathbf{e}'\right)\right]M_{\lambda^0}\mathbf{P}_{(k)}\Phi\right\}-\frac{1}{NT}\operatorname{tr}\left(\mathbf{e}_g\mathbf{e}_g'M_{\lambda^0}\mathbf{P}_{(k)}\Phi\right)\\ &\quad+\frac{1}{NT}\operatorname{tr}\left(\mathbf{e}\mathbf{e}_g'M_{\lambda^0}\mathbf{P}_{(k)}\Phi\right)+\frac{1}{NT}\operatorname{tr}\left(\mathbf{e}_g\mathbf{e}'M_{\lambda^0}\mathbf{P}_{(k)}\Phi\right)+\frac{1}{NT}\operatorname{tr}\left(\mathbf{u}P_{f^0}\mathbf{u}'M_{\lambda^0}\mathbf{P}_{(k)}\Phi\right)\\ &\equiv -C_{2a1,k}-C_{2a2,k}+C_{2a3,k}+C_{2a4,k}+C_{2a5,k},\text{ say.}\end{aligned}$$

Denote C_{2as} as a $K\times 1$ vector whose kth element is $C_{2as,k}$, for $s=1,...,5$. Following the study of $\Pi_{2NT,1}$ in Proposition A.7 we have $\left\|\mathcal{B}_{NT2b}\right\|\le\left\|W_{NT}^{-1}\right\|\left\|C_{NT}^{(2,a)}-\frac{1}{N}b_2\right\|\le\left\|W_{NT}^{-1}\right\|\sum_{s=1}^{5}\left\|C_{2as}\right\|=O_P\left\{\sqrt{\frac{K}{NT}}\left(\delta_{NT}^{-1}+K^{-\gamma/d}\right)\right\}=o_P\left(\gamma_{NT}\right)$. It follows that $\left\|\mathcal{B}_{NT2}\right\|=o_P\left(\gamma_{NT}\right)$. Analogously, we can show that $\left\|\mathcal{B}_{NT3}\right\|=o_P\left(\gamma_{NT}\right)$.

Now we consider $\mathcal{B}_{NT4}$. Following the study of $\Pi_{2NT,3}$ in Theorem 3.2 we can show that $W_{NT}^{-1}C_{NT}^{(2,c)}=\sqrt{\frac{K}{NT}}O_P(\delta_{NT}^{-1}+K^{-\gamma/d})$. Noting that $W_{NT}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}Z_{it}'e_{g,it}=O_P(K^{-\gamma/d})$ and $R_{NT}=O_P(||r_{NT}||\epsilon_0^{1/2})$, we have $\mathcal{B}_{NT4}=O_P\left(K^{-\gamma/d}\right)+\sqrt{\frac{K}{NT}}O_P\left(\delta_{NT}^{-1}+K^{-\gamma/d}\right)+O_P[(\sqrt{K}\delta_{NT}^{-2}+K^{-\gamma/d})(\delta_{NT}^{-1/2}+K^{-\gamma/2d})]=o_P\left(\gamma_{NT}\right)$. ■

Proof of Lemma B.3. Let $\boldsymbol{\varepsilon}\equiv\mathbf{e}+\gamma_{NT}\boldsymbol{\Delta}$ and $\tilde{\epsilon}_0\equiv\left\|\boldsymbol{\varepsilon}\right\|/\sqrt{NT}\le\left(\left\|\mathbf{e}\right\|+\gamma_{NT}\left\|\boldsymbol{\Delta}\right\|\right)/\sqrt{NT}=O_P\left(\delta_{NT}^{-1}+\gamma_{NT}\right)$. Let $\tilde{r}_{NT}=D_{NT}^{-1}\left[C_{l,NT}^{(1)}+C_{l,NT}^{(2,a)}+C_{l,NT}^{(2,b)}+C_{l,NT}^{(2,c)}\right]$, where $C_{l,NT}^{(1)}$, $C_{l,NT}^{(2,a)}$, $C_{l,NT}^{(2,b)}$, and $C_{l,NT}^{(2,c)}$ are defined in (D.9)-(D.11). Noting that

$$C_{l,NT,k}^{(1)}=\frac{1}{NT}\operatorname{tr}\left(M_{f^0}\mathbf{e}'M_{\lambda^0}\mathbf{X}_k\right)+\gamma_{NT}\frac{1}{NT}\operatorname{tr}\left(M_{f^0}\boldsymbol{\Delta}'M_{\lambda^0}\mathbf{X}_k\right)=O_P\left(T^{-1}+(NT)^{-1/2}+\gamma_{NT}\right)$$

and $D_{NT}^{-1}C_{l,NT}^{(2)} = D_{NT}^{-1}[C_{l,NT}^{(2,a)} + C_{l,NT}^{(2,b)} + C_{l,NT}^{(2,c)}] = O_P\left(\delta_{NT}^{-2} + \gamma_{NT}^2\right)$, we have

$$\|\tilde{r}_{NT}\| = \gamma_{NT} D_{NT}^{-1}\Upsilon_{NT} + O_P\left(T^{-1/2}\delta_{NT}^{-1}\right) + O_P\left(\delta_{NT}^{-2} + \gamma_{NT}^2\right) = O_P\left(\gamma_{NT} + \delta_{NT}^{-2}\right).$$

Using Proposition C.1 and following the proof of Theorem 3.1, we can show that

$$\hat{\theta} - \theta^0 = D_{NT}^{-1}C_{l,NT}^{(1)} + D_{NT}^{-1}\left[C_{l,NT}^{(2,a)} + C_{l,NT}^{(2,b)} + C_{l,NT}^{(2,c)}\right] + \tilde{R}_{NT},$$

where $\tilde{R}_{NT} = O_P[(\|\tilde{r}_{NT}\|^2 \tilde{\epsilon}_0 + \|\tilde{r}_{NT}\| \tilde{\epsilon}_0^3 + \|\tilde{r}_{NT}\|^3)^{1/2}] = O_P(\|\tilde{r}_{NT}\| \tilde{\epsilon}_0^{1/2})$; see Su, Jin, and Zhang (2013) for details. Following the proof of Lemma B.2, with some minor modifications[5] we can easily show that under $\mathbb{H}_1(\gamma_{NT})$

$$\hat{\theta} - \theta^0 = \gamma_{NT} D_{NT}^{-1}\Upsilon_{NT} + D^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{it}e_{it} - D^{-1}\left[\frac{1}{T}b_1^{(l)} + \frac{1}{N}b_2^{(l)} + \frac{1}{T}b_3^{(l)}\right] + R_{\theta,NT}$$

where

$$\begin{aligned} R_{\theta,NT} &\equiv \left(D_{NT}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\bar{X}_{it}e_{it} - D^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{it}e_{it} + \frac{1}{T}D^{-1}b_1^{(l)}\right) \\ &\quad + \left(D_{NT}^{-1}C_{l,NT}^{(2,a)} + \frac{1}{N}D^{-1}b_2^{(l)}\right) + \left(D_{NT}^{-1}C_{l,NT}^{(2,b)} + \frac{1}{T}D^{-1}b_3^{(l)}\right) + D_{NT}^{-1}C_{l,NT}^{(2,c)} + \tilde{R}_{NT} \\ &\equiv R_{\theta,NT}^{(1)} + R_{\theta,NT}^{(2)} + R_{\theta,NT}^{(3)} + D_{NT}^{-1}C_{l,NT}^{(2,c)} + \tilde{R}_{NT}, \text{ say.} \end{aligned}$$

Clearly, $\tilde{R}_{NT} = O_P(\|\tilde{r}_{NT}\| \tilde{\epsilon}_0^{1/2}) = O_P[(\delta_{NT}^{-2} + \gamma_{NT})(\delta_{NT}^{-1/2} + \gamma_{NT}^{1/2})] = o_P(\gamma_{NT})$. Following the study of $\Pi_{2NT,3}$ in Proposition A.7 we have $D_{NT}^{-1}C_{l,NT}^{(2,c)} = O_P\{[(NT)^{-1/2} + T^{-1} + \gamma_{NT}](\delta_{NT}^{-1} + \gamma_{NT})\} = o_P(\gamma_{NT})$. To complete the proof of the lemma, it suffices to show that $R_{\theta,NT}^{(s)} = o_P(\gamma_{NT})$ for $s = 1, 2, 3$. For $R_{\theta,NT}^{(1)}$, we have

$$\begin{aligned} R_{\theta,NT}^{(1)} &= D_{NT}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\left(\bar{X}_{it} - \tilde{X}_{it}\right)e_{it} - E_{\mathcal{D}}\left[\left(\bar{X}_{it} - \tilde{X}_{it}\right)e_{it}\right]\right\} \\ &\quad + \left\{D_{NT}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}E_{\mathcal{D}}(\bar{X}_{it}e_{it}) + \frac{1}{T}D_{NT}^{-1}b_1^{(l)}\right\} \\ &\quad + \frac{1}{T}\left(D^{-1} - D_{NT}^{-1}\right)b_1^{(l)} + \left(D_{NT}^{-1} - D^{-1}\right)\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{it}e_{it} \\ &\equiv R_{\theta,NT}^{(1,a)} + R_{\theta,NT}^{(1,b)} + R_{\theta,NT}^{(1,c)} + R_{\theta,NT}^{(1,d)}, \text{ say.} \end{aligned}$$

Following the proof of Lemma A.4, we have $R_{\theta,NT}^{(1,a)} = O_P(\delta_{NT}^{-1}/\sqrt{NT})$. Analogously to the proof of (*ib*) in Proposition A.6, $R_{\theta,NT}^{(1,b)} = O_P((NT)^{-1})$. By Lemma B.1(*iii*) and the facts that $b_1^{(l)} = O_P(1)$ and $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{it}e_{it} = O_P((NT)^{-1/2})$, we have $R_{\theta,NT}^{(1,c)} = O_P\left(N^{-1/2}T^{-3/2}\right)$ and $R_{\theta,NT}^{(1,d)} = O_P((NT)^{-1})$. It follows that $R_{\theta,NT}^{(1)} = O_P(\delta_{NT}^{-1}/\sqrt{NT}) = o_P(\gamma_{NT})$. For $R_{\theta,NT}^{(2)}$, we have

$$R_{\theta,NT}^{(2)} = D_{NT}^{-1}\left(C_{l,NT}^{(2,a)} + \frac{1}{N}b_2^{(l)}\right) + \frac{1}{N}\left(D^{-1} - D_{NT}^{-1}\right)b_2^{(l)} \equiv R_{\theta,NT}^{(2,a)} + R_{\theta,NT}^{(2,b)}, \text{ say.}$$

[5] There are two main differences. The first one is $\left\|\hat{\theta} - \theta^0\right\| = O_P\left(\gamma_{NT} + \delta_{NT}^{-2}\right)$ under $\mathbb{H}_1(\gamma_{NT})$, compared with $\left\|\hat{\beta} - \beta^0\right\| = O_P\left(K^{-\gamma/d} + \sqrt{K}\delta_{NT}^{-2}\right)$ in sieve QMLE framework; the second one is the dimension d of unknown parameter $\hat{\theta}$ is fixed.

It is easy to show that $R^{(2,b)}_{\theta,NT} = O_P\left(T^{-1/2}N^{-3/2}\right)$ by Lemma B.1(*iii*) and the fact that $b_2^{(l)} = O_P(1)$. Following the proof of (*i*) in Proposition A.7, we can show that

$$R^{(2,a)}_{\theta,NT} = O_P((NT)^{-1/2}\,\delta^{-1}_{NT} + \gamma^2_{NT} + (NT)^{-1/2}\,\gamma_{NT}).$$

It follows $R^{(2)}_{\theta,NT} = o_P(\gamma_{NT})$. Similarly, we can show $R^{(3)}_{\theta,NT} = o_P(\gamma_{NT})$. The details are omitted for saving space. ■

Proof of Theorem 4.4. Let P^* denote the probability measure induced by the wild bootstrap conditional on the original sample $\mathcal{W}_{NT} \equiv \{(X_{it}, Y_{it}) : i = 1, ..., N,\ t = 1, ..., T\}$. Let E^* and Var^* denote the expectation and variance with respect to P^*. Let $O_{P^*}(\cdot)$ and $o_{P^*}(\cdot)$ denote the probability order under P^*; e.g., $b_{NT} = o_{P^*}(1)$ if for any $\epsilon > 0$, $P^*(\|b_{NT}\| > \epsilon) = o_P(1)$. We will use the fact that $b_{NT} = o_P(1)$ implies that $b_{NT} = o_{P^*}(1)$.

Observing that $Y^*_{it} = \hat{\theta}' X_{it} + \hat{\lambda}^{(l)\prime}_i \hat{f}^{(l)}_t + e^*_{it}$, the null hypothesis is maintained in the bootstrap world. Given $\mathcal{W}_{NT}$, e^*_{it} are independent across i and t, and independent of X_{js}, $\hat{\lambda}^{(l)}_j$ and $\hat{f}^{(l)}_s$ for all i, t, j, and s, because the latter objects are fixed in the fixed-design bootstrap world. Let $\mathcal{F}^*_t$ be the σ-field generated by $\{e^*_{it}, ..., e^*_{i1}\}_{i=1}^N$. For each i, $\{e^*_{it}, \mathcal{F}^*_t\}$ is an m.d.s. such that $E^*\left(e^*_{it}|\mathcal{F}^*_{t-1}\right) = \hat{e}^{(l)}_{it} E(v_{it}) = 0$ and $E^*[(e^*_{it})^2|\mathcal{F}^*_{t-1}] = [\hat{e}^{(l)}_{it}]^2 E\left(v^2_{it}\right) = [\hat{e}^{(l)}_{it}]^2$. These observations greatly simplify the proofs in the bootstrap world. In particular, we can show that: (*i*) $\hat{\beta}^*_{bc} - \beta^{0*} = \tilde{W}^{-1}\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \tilde{Z}_{it}e^*_{it} + R^*_{\beta,NT}$, where $||R^*_{\beta,NT}|| = o_{P^*}(K^{1/4}/\sqrt{NT})$ and $\beta^{0*} \equiv \left(\beta^{0*}_1, ..., \beta^{0*}_K\right)'$ satisfying $||\hat{\theta}' x - p^K(x)'\beta^{0*}||_{\infty,\varpi} = O_P\left(K^{-\gamma/d}\right)$; and (*ii*) $\hat{\theta}^* - \theta^0 = D^{-1}\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \tilde{X}'_{it}e^*_{it} + B^*_{\theta,NT} + R^*_{\theta,NT}$, where $R^*_{\theta,NT} = o_{P^*}[\delta^{-2}_{NT} + (NT)^{-1/2}]$, $B^*_{\theta,NT} \equiv -N^{-1}D^{-1}b^{(l)*}_2 - T^{-1}D^{-1}b^{(l)*}_3$ and $b^{(l)*}_2, b^{(l)*}_3$ are the bootstrap analogues of $b^{(l)}_2, b^{(l)}_3$, respectively.

Let Γ^*_{NT}, $\mathbb{B}^*_{NT}$, $\mathbb{V}^*_{NT}$, $\hat{\mathbb{B}}^*_{NT}$, and $\hat{\mathbb{V}}^*_{NT}$ be the bootstrap analogues of Γ_{NT}, $\mathbb{B}_{NT}$, $\mathbb{V}_{NT}$, $\hat{\mathbb{B}}_{NT}$, and $\hat{\mathbb{V}}_{NT}$, respectively. Noting that v_{it} are IID $N(0,1)$, we have $\mathbb{B}^*_{NT} \equiv \mathrm{tr}(\tilde{W}^{-1}Q_{wpp}\tilde{W}^{-1}\tilde{\Omega}^*)$ and $\mathbb{V}^*_{NT} \equiv 2\mathrm{tr}(\tilde{W}^{-1}Q_{wpp}\tilde{W}^{-1}\tilde{\Omega}^*\tilde{W}^{-1}Q_{wpp}\tilde{W}^{-1}\tilde{\Omega}^*)$, where $\tilde{\Omega}^* \equiv E^*(\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \tilde{Z}_{it}\tilde{Z}'_{it}e^{*2}_{it}) = \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\{\tilde{Z}_{it}\tilde{Z}'_{it} \times[\hat{e}^{(l)}_{it}]^2\}$. Following the proof of Theorem 4.2, we can show that $\mathbb{V}^*_{NT} = \mathbb{V}_{NT} + o_P(K)$ and $\mathbb{B}^*_{NT} = \mathbb{B}_{NT} + o_P\left(K^{1/2}\right)$ under $\mathbb{H}_1(\gamma_{NT})$. Let $J^*_{NT} \equiv (NT\Gamma^*_{NT} - \mathbb{B}^*_{NT})/\sqrt{\mathbb{V}^*_{NT}}$ and $\hat{J}^*_{NT} \equiv (NT\Gamma^*_{NT} - \hat{\mathbb{B}}^*_{NT})/\sqrt{\hat{\mathbb{V}}^*_{NT}}$. Similar to γ_{NT}, we define $\gamma^*_{NT} \equiv (\mathbb{V}^*_{NT})^{1/4}/\sqrt{NT}$. Let Γ^*_{NTs} denote the bootstrap analogue of Γ_{NTs} for $s \in S^* \equiv \{1, 2, 4, 5, 6, 8\}$. Note that $\Gamma^*_{NTs} = 0$ for $s \in \{3, 7, 9, 10\}$ because the null is explicitly imposed in the bootstrap world. As in the proof of Theorem 4.1, we have

$$\begin{aligned} J^*_{NT} &\equiv (NT\Gamma^*_{NT} - \mathbb{B}^*_{NT})/\sqrt{\mathbb{V}^*_{NT}} \\ &= (NT\Gamma^*_{NT1} - \mathbb{B}^*_{NT})/\sqrt{\mathbb{V}^*_{NT}} + \gamma^*_{NT}(\Gamma^*_{NT2} + \Gamma^*_{NT4} - 2\Gamma^*_{NT5} - 2\Gamma^*_{NT6} + 2\Gamma^*_{NT8}). \end{aligned}$$

We prove the theorem by showing that: (*i*) $\tilde{J}^*_{NT} \equiv (NT\Gamma^*_{NT1} - \mathbb{B}^*_{NT})/\sqrt{\mathbb{V}^*_{NT}} \xrightarrow{d^*} N(0,1)$, (*ii*) $\gamma^*_{NT}\Gamma^*_{NTs} = o_{P^*}(1)$ for $s \in \{2, 4, 5, 6, 8\}$, (*iii*) $\hat{\mathbb{B}}^*_{NT} = \mathbb{B}^*_{NT} + o_{P^*}\left(K^{1/2}\right)$, and (*iv*) $\hat{\mathbb{V}}^*_{NT} = \mathbb{V}^*_{NT} + o_{P^*}(K)$.

We only outline the proof of (*i*) as we can follow the proofs of Theorems 4.1 and 4.2 to show (*ii*)-(*iv*). Analogously to the proof of Proposition B.4, we can show that $\tilde{J}^*_{NT} = \sum_{1\le i<j\le N} W^*_{ij} + o_{P^*}(1)$, where $W^*_{ij} \equiv W^*_{NT}(u^*_i, u^*_j) \equiv \frac{2}{NT\mathbb{V}^*_{NT}}\sum_{1\le t,s\le T} e^*_{it}H_{ij,ts}e^*_{js}$, $u^*_i \equiv (\tilde{Z}_i, e^*_i)'$, and e^*_i is the bootstrap analogue of e_i. Noting that $\tilde{J}^*_{NT}$ is a second order degenerate U-statistic that is "clean" ($E^*[W^*_{NT}(u^*_i, u)] = E^*[W^*_{NT}(u, u^*_j)] = 0$ a.s. for any nonrandom u), we can still apply Proposition 3.2 in de Jong (1987) to prove the CLT for $\tilde{J}^*_{NT}$ by showing that (*i*1) $\mathrm{Var}^*(\tilde{J}^*_{NT}) = 1 + o_{P^*}(1)$, (*i*2) $G^*_I \equiv \sum_{1\le i<j<N} E^*[(W^*_{ij})^4] = o_{P^*}(1)$, (*i*3) $G^*_{II} \equiv \sum_{1\le i<j<l\le N} E^*(W^{*2}_{il}W^{*2}_{jl} + W^{*2}_{ij}W^{*2}_{il} + W^{*2}_{ij}W^{*2}_{lj}) = o_{P^*}(1)$, and (*i*4) $G^*_{III} \equiv \sum_{1\le i<j<r<l\le N} E^*(W^*_{ij}W^*_{ir}\ W^*_{lj}W^*_{lr} + W^*_{ij}W^*_{il}W^*_{rj}W^*_{rl} + W^*_{ir}W^*_{il}W^*_{jr}W^*_{jl}) = o_{P^*}(1)$. Note that v_{it} is IID across i and t, $E^*[(e^*_{it})] = 0$, $E^*[(e^*_{it})^2] = [\hat{e}^{(l)}_{it}]^2$, and $E^*[(e^*_{it})^4] = 3[\hat{e}^{(l)}_{it}]^4$.

For $(i1)$, using the IID property of $\{v_{it}\}$, we can readily show that

$$\begin{aligned}
\mathrm{Var}^*(\tilde{J}_{NT}^*) &= \frac{4}{N^2T^2\mathbb{V}_{NT}^*}\sum_{1\le i<j\le N}\sum_{t_1=1}^T\sum_{t_2=1}^T\sum_{s_1=1}^T\sum_{s_2=1}^T H_{ij,t_1s_1}H_{ij,t_2s_2}\hat{e}_{it_1}^{(l)}\hat{e}_{js_1}^{(l)}\hat{e}_{it_2}^{(l)}\hat{e}_{js_2}^{(l)}E^*\left(v_{it_1}v_{js_1}v_{it_2}v_{js_2}\right) \\
&= \frac{4}{N^2T^2\mathbb{V}_{NT}^*}\sum_{1\le i<j\le N}\sum_{t=1}^T\sum_{s=1}^T H_{ij,ts}^2[\hat{e}_{it}^{(l)}]^2[\hat{e}_{js}^{(l)}]^2 \\
&= 1-\frac{2}{N^2T^2\mathbb{V}_{NT}^*}\sum_{i=1}^N\sum_{t=1}^T\sum_{s=1}^T H_{ii,ts}^2[\hat{e}_{it}^{(l)}]^2[\hat{e}_{is}^{(l)}]^2 \\
&= 1+O_P\left(N^{-1}\right)=1+o_{P^*}(1),
\end{aligned}$$

where we follow the proof of Theorem 4.2 and show the term $O_P\left(N^{-1}\right)$ in the last line. For $(i2)$, recall that $\overline{q}_{k_1k_2}$ is the (k_1,k_2)th element of $\bar{Q}_{pp}$, and $H_{ij,ts}=\sum_{k_1=1}^K\sum_{k_2=1}^K\overline{q}_{k_1k_2}\tilde{Z}_{it,k_1}\tilde{Z}_{js,k_2}$. Let $\phi_{it,k}^*\equiv\tilde{Z}_{it,k}e_{it}^*$. Then we have

$$\begin{aligned}
G_I^* &= \tfrac{16}{N^4T^4\mathbb{V}_{NT}^2}\sum_{1\le k_1,\ldots,k_8\le K}\overline{q}_{k_1k_2}\bar{q}_{k_3k_4}\bar{q}_{k_5k_6}\bar{q}_{k_7k_8} \\
&\quad\times\sum_{1\le i<j<N}\sum_{1\le t_1,\ldots,t_8\le T}E^*\left(\phi_{it_1,k_1}^*\phi_{it_3,k_3}^*\phi_{it_5,k_5}^*\phi_{it_7,k_7}^*\right)E^*\left(\phi_{jt_2,k_2}^*\phi_{jt_4,k_4}^*\phi_{jt_6,k_6}^*\phi_{jt_8,k_8}^*\right).
\end{aligned}$$

First, note that the term inside the last summation takes value 0 if either $\#\{t_1,t_3,t_5,t_7\}>2$ or $\#\{t_2,t_4,t_6,t_8\}>2$. So it suffices to consider three cases according to the number of distinct time indices in the set $S=\{t_1,...,t_8\}$: (a) $\#S=4$, (b) $\#S=3$, and (c) $\#S\le 2$. We use G_{Ia}^*, G_{Ib}^*, and G_{Ic}^* to denote the corresponding summations when the time indices are restricted to cases (a), (b) and (c), respectively. Then $G_I^*=G_{Ia}^*+G_{Ib}^*+G_{Ic}^*$. For G_{Ia}^*, we must have $\#\{t_1,t_3,t_5,t_7\}=2$ and $\#\{t_2,t_4,t_6,t_8\}=2$. Without loss of generality, assume that $t_1=t_3>t_5=t_7$ and $t_2=t_4>t_6=t_8$. By the IID property of v_{it}, $|E^*(\phi_{it_1,k_1}^*\phi_{it_3,k_3}^*\phi_{it_5,k_5}^*\phi_{it_7,k_7}^*)|=\tilde{Z}_{it_1,k_1}\tilde{Z}_{it_1,k_3}[\hat{e}_{it_1}^{(l)}]^2\tilde{Z}_{it_5,k_5}\tilde{Z}_{it_5,k_6}[\hat{e}_{it_5}^{(l)}]^2$. Then

$$\begin{aligned}
|G_{Ia}^*| &\le \tfrac{16}{N^4T^4\mathbb{V}_{NT}^{*2}}\sum_{1\le k_1,\ldots,k_8\le K}|\bar{q}_{k_1k_2}|\,|\bar{q}_{k_3k_4}|\,|\bar{q}_{k_5k_6}|\,|\bar{q}_{k_7k_8}| \\
&\quad\times\left\{\sum_{i=1}^N\sum_{1\le t_5<t_1\le T}\tilde{Z}_{it_1,k_1}\tilde{Z}_{it_1,k_3}(\hat{e}_{it_1}^{(l)})^2\tilde{Z}_{it_5,k_5}\tilde{Z}_{it_5,k_7}(\hat{e}_{it_5}^{(l)})^2\right\} \\
&\quad\times\left\{\sum_{j=1}^N\sum_{1\le t_6<t_2\le T}\tilde{Z}_{it_2,k_2}\tilde{Z}_{it_2,k_4}(\hat{e}_{it_2}^{(l)})^2\tilde{Z}_{it_6,k_6}\tilde{Z}_{it_6,k_8}(\hat{e}_{it_6}^{(l)})^2\right\} \\
&= \tfrac{64}{N^4T^4\mathbb{V}_{NT}^{*2}}O_P\left(K^8N^2T^4\right)=O_P\left(K^6/N^2\right)=O_{P^*}\left(K^6/N^2\right).
\end{aligned}$$

Similarly, we can show that $G_{Is}^*=O_{P^*}\left(K^6/N^2\right)=o_{P^*}(1)$ for $s=b,c$. It follows that $G_I^*=o_{P^*}(1)$. For $(i3)$, we write $G_{II}^*\equiv\sum_{1\le i<j<l\le N}E^*(W_{il}^{*2}W_{jl}^{*2}+W_{ij}^{*2}W_{il}^{*2}+W_{ij}^{*2}W_{lj}^{*2})=G_{II,1}^*+G_{II,2}^*+G_{II,3}^*$. By the IID property of v_{it}, we have

$$\begin{aligned}
G_{II,1}^* &\equiv \frac{16}{N^4T^4\mathbb{V}_{NT}^{*2}}\sum_{1\le i<j<l\le N}\sum_{1\le t_1,\ldots,t_6\le T}E^*\left[e_{it_1}^{*2}e_{jt_2}^{*2}H_{il,t_1t_3}H_{il,t_1t_4}H_{jl,t_2t_5}H_{jl,t_2t_6}e_{lt_3}^*e_{lt_4}^*e_{lt_5}^*e_{lt_6}^*\right] \\
&= \frac{48}{N^4T^4\mathbb{V}_{NT}^{*2}}\sum_{1\le i<j<l\le N}\sum_{1\le t_1,t_2\le T}\sum_{1\le t_3\ne t_6\le T}H_{il,t_1t_3}^2H_{jl,t_2t_6}^2[\hat{e}_{it_1}^{(l)}]^2[\hat{e}_{jt_2}^{(l)}]^2[\hat{e}_{lt_3}^{(l)}]^2[\hat{e}_{lt_6}^{(l)}]^2 \\
&\quad+\frac{16}{N^4T^4\mathbb{V}_{NT}^2}\sum_{1\le i<j<l\le N}\sum_{1\le t_1,t_2\le T}\sum_{t_3=1}^T H_{il,t_1t_3}^2H_{jl,t_2t_3}^2[\hat{e}_{it_1}^{(l)}]^2[\hat{e}_{jt_2}^{(l)}]^2[\hat{e}_{lt_3}^{(l)}]^4 \\
&= G_{II,11}^*+G_{II,12}^*,\text{ say.}
\end{aligned}$$

For $G^*_{II,11}$, we have

$$\begin{aligned}
G^*_{II,11} &\leq \frac{16}{N^4T^4\mathbb{V}^{*2}_{NT}} \sum_{1\leq i<j<l\leq N} \sum_{1\leq t_1,t_2\leq T} \sum_{1\leq t_3\neq t_6\leq T} \Big\{ \mathrm{tr}\left[(\hat{e}^{(l)}_{it_1})^2 \tilde{Z}_{it_1}\tilde{Z}'_{it_1}\bar{Q}_{pp}[(\hat{e}^{(l)}_{lt_3})^2\tilde{Z}_{lt_3}\tilde{Z}'_{lt_3}\bar{Q}_{pp}\right] \\
&\qquad \times \mathrm{tr}\left[(\hat{e}^{(l)}_{jt_2})^2\tilde{Z}_{jt_2}\tilde{Z}'_{jt_2}\bar{Q}_{pp}(\hat{e}^{(l)}_{it_6})^2\tilde{Z}_{lt_6}\tilde{Z}'_{lt_6}\bar{Q}_{pp}\right]\Big\} \\
&\leq \frac{8}{3N^2T^2\mathbb{V}^{*2}_{NT}} \sum_{l=1}^{N} \sum_{1\leq t_3\neq t_6\leq T} \mathrm{tr}\left[\tilde{\Omega}^*\bar{Q}_{pp}(\hat{e}^{(l)}_{lt_3})^2\tilde{Z}_{lt_3}\tilde{Z}'_{lt_3}\bar{Q}_{pp}\right] \mathrm{tr}\left[\tilde{\Omega}^*\bar{Q}_{pp}(\hat{e}^{(l)}_{it_6})^2\tilde{Z}_{lt_6}\tilde{Z}'_{lt_6}\bar{Q}_{pp}\right] \\
&\leq \frac{8\mu_1^2(\tilde{\Omega}^*)\mu_1^4\left(\bar{Q}_{pp}\right)}{3N^2T^2\mathbb{V}^{*2}_{NT}} \sum_{l=1}^{N} \sum_{1\leq t_3\neq t_6\leq T} (\hat{e}^{(l)}_{lt_3})^2 \left\|\tilde{Z}_{lt_3}\right\|_F^2 (\hat{e}^{(l)}_{it_6})^2 \left\|\tilde{Z}_{lt_6}\right\|_F^2 \\
&= \frac{8\left[\mu_1^2(\tilde{\Omega}_{NT}) + o_P(1)\right]\mu_1^4\left(\bar{Q}_{pp}\right)}{3N^2T^2\mathbb{V}^{*2}_{NT}} \left\{ \sum_{l=1}^{N} \sum_{1\leq t_3\neq t_6\leq T} e^2_{lt_3}\left\|\tilde{Z}_{lt_3}\right\|_F^2 e^2_{it_6}\left\|\tilde{Z}_{lt_6}\right\|_F^2 + o_P\left(NT^2K^2\right)\right\} \\
&= O_P\left(N^{-2}T^{-2}K^{-2}\right)O_P\left(NT^2K^2\right) = O_P\left(N^{-1}\right) = O_{P^*}\left(N^{-1}\right).
\end{aligned}$$

Then $G^*_{II,11} = o_{P^*}(1)$. With the same method we can show that $G^*_{II,12} = o_{P^*}(1)$. Thus $G^*_{II,1} = o_{P^*}(1)$. Similarly, we can show that $G^*_{II,2} = o_{P^*}(1)$ and $G^*_{II,3} = o_{P^*}(1)$. It follows that $G^*_{II} = o_{P^*}(1)$.

For $(i4)$, we write $G^*_{III} \equiv \sum_{1\leq i<j<r<l\leq N} E^*(W^*_{ij}W^*_{ir}W^*_{lj}W^*_{lr} + W^*_{ij}W^*_{il}W^*_{rj}W^*_{rl} + W^*_{ir}W^*_{il}W^*_{jr}W^*_{jl}) \equiv \sum_{s=1}^4 G^*_{III,s}$, say. Following the proof of $G_{III,1} = o_P(1)$ in Proposition B.4, we have

$$\begin{aligned}
G^*_{III,1} &= \frac{16}{N^4T^4\mathbb{V}^{*2}_{NT}} \sum_{1\leq i<j<r<l\leq N} \sum_{1\leq t_1,\ldots,t_8\leq T} E^*(H_{ij,t_1t_2}e^*_{it_1}e^*_{jt_2}H_{ir,t_3t_4}e^*_{it_3}e^*_{rt_4}H_{lj,t_5t_6}e^*_{lt_5}e^*_{jt_6}H_{lr,t_7t_8}e^*_{lt_7}e^*_{rt_8}) \\
&= \frac{16}{N^4T^4\mathbb{V}^{*2}_{NT}} \sum_{1\leq i<j<r<l\leq N} \sum_{1\leq t,s,p,q\leq T} \mathrm{tr}[E^*(\bar{Q}_{pp}\tilde{Z}_{it}\tilde{Z}'_{it}e^{*2}_{it}\bar{Q}_{pp}\tilde{Z}_{rs}\tilde{Z}'_{rs}e^{*2}_{rs}\bar{Q}_{pp}\tilde{Z}_{lp}\tilde{Z}'_{lp}e^{*2}_{lp}\bar{Q}_{pp}\tilde{Z}_{jq}\tilde{Z}'_{jq}e^{*2}_{jq})] \\
&= \frac{2}{3\mathbb{V}^{*2}_{NT}}\mathrm{tr}\left(\bar{Q}_{pp}\tilde{\Omega}^*\bar{Q}_{pp}\tilde{\Omega}^*\bar{Q}_{pp}\tilde{\Omega}^*\bar{Q}_{pp}\tilde{\Omega}^*\right) = O_P\left(\frac{1}{K}\right) = o_{P^*}(1)
\end{aligned}$$

where we use the facts that $\mathrm{tr}\left(\bar{Q}_{pp}\tilde{\Omega}^*\bar{Q}_{pp}\tilde{\Omega}^*\bar{Q}_{pp}\tilde{\Omega}^*\bar{Q}_{pp}\tilde{\Omega}^*\right) = \mathrm{tr}\left(\bar{Q}_{pp}\tilde{\Omega}\bar{Q}_{pp}\tilde{\Omega}\bar{Q}_{pp}\tilde{\Omega}\bar{Q}_{pp}\tilde{\Omega}\right) + o_P(1)$ and $\mathrm{tr}\left(\bar{Q}_{pp}\tilde{\Omega}\bar{Q}_{pp}\tilde{\Omega}\bar{Q}_{pp}\tilde{\Omega}\bar{Q}_{pp}\tilde{\Omega}\right) \leq \mu_1^4\left(\bar{Q}_{pp}\right)\mu_1^3(\tilde{\Omega})\mathrm{tr}(\tilde{\Omega}) = O_P(K)$ in the last line. ■

E Some technical lemmas

Let $\{\xi_t, t\geq 1\}$ be a $\mathcal{D}$-strong mixing process with mixing coefficient $\alpha_{\mathcal{D}}(\cdot)$. We will use the following lemmas frequently.

Lemma E.1 *(Conditional Davydov Inequality) Suppose that* A_1 *and* A_2 *are random variables which are measurable with respect to* $\sigma(\xi_1,\ldots,\xi_s)$ *and* $\sigma(\xi_{s+\tau},\ldots,\xi_T)$*, respectively, and that both* $\|A_1\|_{p,\mathcal{D}}$ *and* $\|A_2\|_{q,\mathcal{D}}$ *are bounded in probability, where* $p,q>1$ *and* $p^{-1}+q^{-1}<1$*. Then* $|E_{\mathcal{D}}(A_1A_2) - E_{\mathcal{D}}(A_1)E_{\mathcal{D}}(A_2)| \leq 8\|A_1\|_{p,\mathcal{D}}\|A_2\|_{q,\mathcal{D}}\alpha_{\mathcal{D}}^{1-p^{-1}-q^{-1}}(\tau)$.

Lemma E.2 *Suppose* $\max_{1\leq t\leq T} E|A_t|^q < \infty$*. Then* $\max_{1\leq t\leq T}|A_t| = o_P\left(T^{1/q}\right)$.

Proof. Let $\varepsilon_T \equiv T^{1/q}$. We have

$$\begin{aligned}
\Pr\left(\max_{1\le t\le T}|A_t|>\varepsilon_T\right) &\le \sum_{t=1}^T \Pr\left(|A_t|>\varepsilon_T\right)=\sum_{t=1}^T E\left[1\left(|A_t|>\varepsilon_T\right)\right]\le\sum_{t=1}^T E\left[\frac{|A_t|^q}{\varepsilon_T^q}1\left(|A_t|>\varepsilon_T\right)\right]\\
&= \varepsilon_T^{-q}\sum_{t=1}^T E\left[|A_t|^q\,1\left(|A_t|>\varepsilon_T\right)\right]\le \max_{1\le t\le T}E\left[|A_t|^q\,1\left(|A_t|>\varepsilon_T\right)\right]\to 0.
\end{aligned}$$

It follows that $\max_{1\le t\le T}|A_t| = o_P\left(T^{1/q}\right)$. ■

Lemma E.3 *Let A be an $n\times m$ matrix, B and C be $m\times p$ matrices, and D be an $n\times n$ matrix. Then*
(i) $\|A\| \le \|A\|_F \le \|A\|\sqrt{rank(A)}$;
(ii) $\|AB\| \le \|A\|\,\|B\|$;
(iii) $\|AB\|_F \le \|A\|\,\|B\|_F \le \|A\|_F\,\|B\|_F$;
(iv) $\max\{\|A\|_1, \|A\|_{\max}\} \le \|A\| \le \sqrt{nm}\,\|A\|$, *where* $\|A\|_1 \equiv \max_j \sum_{i=1}^n |A_{ij}|$ *and* $\|A\|_\infty \equiv \max_i \sum_{j=1}^n |A_{ij}|$;
(v) $tr(AB) \le \|A\|_F\,\|B\|_F$;
(vi) $tr\,(D) \le rank(D)\,\|D\|$;
(vii) $\|D\| \le tr\,(D)$ *for any p.s.d. diagonal matrix* D;
(viii) $\|D\| \le \max_{1\le i\le n}|D_{ii}|$ *for any diagonal matrix* D;
(ix) $\|A\|_F = \|vec(A)\|$;
(x) $\mu_1(A'A) = \mu_1(AA')$;
(xi) $rank(AB) \le \min\{rank(A), rank(B)\}$;
(xii) $rank(B+C) \le rank(B)+rank(C)$.

Proof. For the proofs of (i)-(vii), see Theorem S.3.1 in Moon and Weidner (2010). For the proofs of $(viii)$-(xi), see Bernstein (2005) or Seber (2007). ■

REFERENCES

Bai, J. (2009) Panel data model with interactive fixed effects. *Econometrica* 77, 1229–1279.

Bernstein, D.S. (2005) *Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory*. Princeton University Press, Princeton.

Kato, T. (1980) *Perturbation Theory for Linear Operators*. Springer-Verlag.

Moon, H.R. & M. Weidner (2010) Dynamic linear panel data regression models with interactive fixed effects. Working paper, University of Southern California.

Moon, H.R. & M. Weidner (2012) Linear regression for panel with unknown number of factors as interactive fixed effects. Working paper, University of Southern California.

Seber, G. (2007) *A Matrix Handbook for Statisticians*. John Wiley & Sons, New Jersey.

Su, L., S. Jin, & Y. Zhang (2013) Specification test for panel data models with interactive fixed effects. Working paper, Singapore Management University.