THREE ESSAYS ON LARGE PANEL DATA MODELS WITH CROSS-SECTIONAL DEPENDENCE

YONGHUI ZHANG

SINGAPORE MANAGEMENT UNIVERSITY

2013

Three Essays on Large Panel Data Models with Cross-Sectional Dependence

by
Yonghui Zhang

Submitted to School of Economics in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Economics

Dissertation Committee:

Peter C.B. Phillips (Supervisor/Co-Chair)
Sterling Professor of Economics and Statistics
Yale University
Distinguished Term Professor of Economics
Singapore Management University

Liangjun Su (Supervisor/Co-Chair)
Professor of Economics
Singapore Management University

Sainan Jin
Associate Professor of Economics
Singapore Management University

Denis Heng-Yan Leung
Professor of Statistics
Singapore Management University

Singapore Management University
2013

Abstract

Three Essays on Large Panel Data Models with Cross-Sectional Dependence

Yonghui Zhang

My dissertation consists of three essays which contribute new theoretical results to large panel data models with cross-sectional dependence. These essays try to answer or partially answer some prominent questions such as how to detect the presence of cross-sectional dependence and how to capture the latent structure of cross-sectional dependence and estimate parameters efficiently by removing its effects.

Chapter 2 introduces a nonparametric test for cross-sectional contemporaneous dependence in large dimensional panel data models based on the squared distance between the pair-wise joint density and the product of the marginals. The test can be applied to either raw observable data or residuals from local polynomial time series regressions for each individual to estimate the joint and marginal probability density functions of the error terms. In either case, we establish the asymptotic normality of our test statistic under the null hypothesis by permitting both the cross section dimension n and the time series dimension T to pass to infinity simultaneously and relying upon the Hoeffding decomposition of a two-fold U-statistic. We also establish the consistency of our test. A small set of Monte Carlo simulations is conducted to evaluate the finite sample performance of our test and compare it with that of Pesaran (2004) and Chen, Gao, and Li (2009).

Chapter 3 analyzes 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 spe-

cific 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. Following Moon and Weidner's (2010, 2012) 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 consistent bias-corrected estimator is provided. We also propose a consistent specification test for the linearity of the nonparametric 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. To improve the finite sample performance of the test, we also propose a bootstrap procedure to obtain the bootstrap p-values and justify its validity. Monte Carlo simulations are conducted to investigate the finite sample performance of our estimator and test. We apply our model to an economic growth data set to study the relationship between capital accumulation and real GDP growth rate.

Chapter 4 proposes a nonparametric test for common trends in semiparametric panel data models with fixed effects based on a measure of nonparametric goodness-of-fit (R^2). We first estimate the model under the null hypothesis of common trends by the method of profile least squares, and obtain the augmented residual which consistently estimates the sum of the fixed effect and the disturbance under the null. Then we run a local linear regression of the augmented residuals on a time trend and calculate the nonparametric R^2 for each cross section unit. The proposed test statistic is obtained by averaging all cross sectional nonparametric R^2's, which is close to 0 under the null and deviates from 0 under the alternative. We show that after appropriate standardization the test statistic is asymptotically normally distributed under both the null hypothesis and a sequence of Pitman local alternatives. We prove test consistency and propose a bootstrap procedure to obtain p-values. Monte Carlo simulations indicate that the test performs well infinite samples. Empirical

applications are conducted exploring the commonality of spatial trends in UK climate change data and idiosyncratic trends in OECD real GDP growth data. Both applications reveal the fragility of the widely adopted common trends assumption.

Table of Contents

Acknowledgements

First of all, I would like to express my sincerest gratitude to my advisor Professor Liangjun Su, without whose constant help I could not finish my dissertation. In the past few years, he gave me enormous amount of suggestions and guidance to nurture me on research in econometrics. Professor Su led me into the door of academic research and set a role model for me.

I own a special debt to Professor Peter C. B. Phillips for his wonderful courses on Econometrics, valuable feedback on my research, our joint work, and continuously support during my job searching. I am indebted to Professor Sainan Jin for serving in my committee, whose encouragement helps me to survive the hard times. I would like to thank Professor Denis Leung for reviewing my thesis and giving me very valuable and insightful feedback. Thanks also go to Professor Jun Yu, Professor YiuKuen Tse, and Professor Anthony Tay for their kind helps during my stay at SMU.

I wish to thank my friends at SMU. It will be too long to list all of them here. However, I would like to give thanks to Xiaohu Wang for so many interesting discussions and his devotion to our joint works. I also want to give thanks to my dear friends Shouwei Liu, Tao Zeng, Huaxia Zeng, Jing Hu, Yong Li, Qiankun Zhou and Tao Yang for their accompany in the last five years.

Last but not least, I wish to thank my family for all their love and support that helped me so much in my life and during my studies. Special thanks go to my wife Huizhu Xia for her tremendous support, endless care and love.

Chapter 1 Introduction

In recent years one of the most active research areas in the panel data literature has been cross-sectional dependence. The topic has figured prominently in work on economic growth, housing prices, indices of economic activity, asset pricing, and other economic, business and financial activities and decisions. The sources of dependence are manifold and include spatial effects, spillover effects, unobservable common factors and social interactions. Ignoring cross-sectional dependence in panel applications can have serious consequences such as inconsistent estimation, misleading inference and distortions in hypothesis testing. Amidst the ongoing work on cross-sectional dependence, two questions are prominent: (i) how to detect its presence; and (ii) how to capture its latent structure and estimate parameters efficiently by removing its effects. For the first question, many tests have been proposed such as the Breusch-Pagan (1978) LM test and Pesaran (2004) cross-sectional dependence test. For the second question, there are two mainstreams in the literature. In spatial econometric work, one approach is to use additional information such as spatial/economic distance, which has been widely used in research on crime, regional science and social interactions. A second approach is to use factor models which have become popular in empirical finance and macroeconomics.

My dissertation seeks to address these issues: (i) How to test for cross-sectional dependence; (ii) How to estimate and make statistical test for nonparametric dynamic panel data models with interactive fixed effects; (iii) How to test for common trends in semiparametric panel data models with fixed effects where cross-sectional dependence is present in the errors. The contributions of my thesis have twofold. First, it would contribute to theoretical methods, specially to the nonparamteric and semiparametric estimation and testing; Second, my dissertation research would benefit extensive empirical studies, as it could provide methodological approach to help analyze practical real-world questions. We propose a nonparametric test for cross-sectional dependence in Chapter 2. Once we find strong evidence of cross-sectional dependence, we can adopt the nonparametric dynamic panel data models with interactive fixed effects to capture the unknown cross-sectional dependence, which is considered in Chapter 3. Chapter 4 considers a nonparametric test for common trends in semiparametric panel data models with fixed effects and cross-sectional dependence in the errors.

Chapter 2 proposes a test for cross-sectional dependence based on the squared distance between the pair-wise joint density and the product of the marginals. Unlike all available tests which are designed to test for "cross-sectional correlation", the new test is designed for "cross-sectional dependence". Our test has several advantages over traditional tests. First, it can detect cross-sectional dependence in non-Gaussian and nonlinear cases where traditional tests may fail. Second, compared with Pesaran's (2004) and Chen, Gao and Li's (2012) tests, our test is able to detect cross-sectional dependence with multifactor structure errors and zero mean factor loadings. Finally, the test can be applied either to raw observable data or to the residuals from local polynomial time series regression for each cross-sectional unit. This chapter establishes asymptotic normality of the test statistic under the null hypothesis when both the cross-section dimension n and time-series dimension T are large. Test consistency is proved. The theory developed in the chapter relies on the Hoeffding decomposition of a two-fold U-statistic.

Chapter 3 develops a new panel data model with cross-sectional dependence. The model has several desirable features. First, it incorporates time-varying common factors (time fixed effects) and individual-specific factor loadings (individual fixed effects) multiplicatively, which captures heterogeneity in a more flexible way. Second, it does not impose a parametric form for the unknown regression function and is therefore robust to functional misspecification. Finally, just as for the classic dynamic panel data model, predetermined variables may be included in the regressors, which is a desirable feature in empirical research. We propose sieve estimation for the nonparametric function by extending Bai's (2009) principal component analysis (PCA) to the nonparametric framework. Following Moon and Weidner's (2010, 2012) asymptotic expansion of the quasi likelihood function, we derive the convergence rate of the sieve estimator and establish its asymptotic normal distribution. The sources of asymptotic bias are discussed and a consistent bias-corrected estimator is provided. We also introduce a consistent specification test for the linearity of the nonparametric function by comparing its linear estimator with the sieve estimator. We establish the asymptotic distributions of the test statistic both under the null hypothesis and a sequence of Pitman local alternatives. The new model is well suited to macroeconomic and financial applications. An empirical application is conducted with economic growth data from the Penn World Table 7.1 to study the relationship between capital accumulation and real GDP growth rate, showing evidence of a nonlinear relationship.

Chapter 4 considers commonality of slowly changing time trends in panel data models with fixed effects and cross-section dependence in the errors. A nonparametric test for common trends is constructed based on a measure of nonparametric goodness-of-fit (R^2). Under the null hypothesis of common trends, we estimate the model by profile least squares, and obtain the augmented residual as a consistent estimator for the sum of the fixed effect and the disturbance. We then run a local linear regression of the augmented residuals on a time trend and calculate the nonparametric R^2 for each cross-sectional unit. The test statistic is obtained by averaging all cross-section nonparametric R^2's, which is close to 0 under the null and deviates from 0 under the alternative. It is shown that the test statistic is asymptotically normally distributed under both the null hypothesis and a sequence of Pitman local alternatives after appropriate standardization. We prove consistency of the proposed test and introduce a bootstrap procedure to compute p-values. The test is illustrated in two applications, covering the UK climate change data and OECD real GDP growth data. Both examples reveal the fragility of the widely adopted common trends assumption.

Chapter 2 Testing Cross-Sectional Dependence in Nonparametric Panel Data Models

2.1 Introduction

In recent years, there has been a growing literature on large dimensional panel data models with cross-sectional dependence. Cross-sectional dependence may arise due to spatial or spillover effects, or due to unobservable common factors. Much of the recent research on panel data has focused on how to handle cross-sectional dependence. There are two popular approaches in the literature: one is to assume the individuals are spatially dependent, which gives rise to spatial econometrics; and the other is to assume that the disturbances have a factor structure, which gives rise to static or dynamic factor models. For a recent and comprehensive overview of panel data factor model, see the excellent monograph by Bai and Ng (2008).

Traditional panel data models typically assume observations are independent across individuals, which leads to immense simplification to the rules of estimation and inference. Nevertheless, if observations are cross-sectionally dependent, parametric or nonparametric estimators based on the assumption of cross-sectional independence may be inconsistent and statistical inference based on these estimators can generally be misleading. It has been well documented that panel unit root and cointegration tests based on the assumption of cross-sectional independence are generally inadequate and tend to lead to significant size distortions in the presence of cross-sectional dependence; see Chang (2002), Bai and Ng (2004, 2010), Bai and Kao (2006), and Pesaran (2007), among others. Therefore, it is important to test for cross-sectional independence before embarking on estimation and statistical inference.

Many diagnostic tests for cross-sectional dependence in parametric panel data model have been suggested. When the individuals are regularly spaced or ranked by certain rules, several statistics have been introduced to test for spatial dependence, among which the Moran-I test statistic is the most popular one. See Anselin (1988, 2001) and Robinson (2008) for more details. However, economic agents are generally not regularly spaced, and there does not exist a "spatial metric" that can measure the degree of spatial dependence across economic agents effectively. In order to test for cross-sectional dependence in a more general case, Breusch and Pagan (1980) develop a Lagrange multiplier (LM) test statistic to check the diagonality of the error covariance matrix in SURE models. Noticing that Breusch and Pagan's LM test is only effective if the number of time periods T is large relative to the number of cross sectional units n, Frees (1995) considers test for cross-sectional correlation in panel data models when n is large relative to T and show that both the Breusch and Pagan's and his test statistic belong to a general family of test statistics. Noticing that Breusch and Pagan's LM test statistic suffers from huge finite sample bias, Pesaran (2004) proposes a new test for cross-sectional dependence (CD) by averaging all pair-wise correlation coefficients of regression residuals. Nevertheless, Pesaran's CD test is not consistent against all global alternatives. In particular, his test has no power in detecting cross-sectional dependence when the mean of factor loadings is zero. Hence, Ng (2006) employs spacing variance ratio statistics to test cross-sectional correlations, which is more robust and powerful than that of Pesaran (2004). Huang, Kao, and Urga (2008) suggest a copula-based tests for testing cross-sectional dependence of panel data models. Pesaran, Ullah, and Yamagata (2008) improve Pesaran (2004) by considering a bias adjusted LM test in the case of normal errors. Based on the concept of generalized residuals (e.g., Gourieroux et al. (1987)), Hsiao, Pesaran, and Pick (2009) propose a test for cross-sectional dependence in the case of non-linear panel data models. Interestingly, an asymptotic version of their test statistic can be written as the LM test of Breusch and Pagan (1980). Sarafidis, Yamagata, and Robertson (2009) consider tests for cross-sectional dependence in dynamic panel data models.

All the above tests are carried out in the parametric context. They can lead to meaningful interpretations if the parametric models or underlying distributional assumptions are correctly specified, and may yield misleading conclusions otherwise. To avoid the potential misspecification of functional form, Chen, Gao, and Li (2009, CGL hereafter) consider tests for cross-sectional dependence based on nonparametric residuals. Their test is a nonparametric counterpart of Pesaran's (2004) test. So it is constructed by averaging all pair-wise cross-sectional correlations and therefore, like Pesaran's (2004) test, it does not test for "*pair-wise independence*" but "*pair-wise uncorrelation*". It is well known that uncorrelation is generally different from independence in the case of non-Gaussianity or nonlinear dependence (e.g., Granger, Maasoumi, and Racine (2004)). There exist cases where testing for cross-sectional pair-wise independence is more appropriate than testing pair-wise uncorrelation.

Since Hoeffding (1948), there has developed an extensive literature on testing independence or serial independence. See Robinson (1991), Brock et al. (1996), Ahmad and Li (1997), Johnson and McClelland (1998), Pinkse (1998), Hong (1998, 2000), Hong and White (2005), among others. All these tests are based on some measure of deviations from independence. For example, Robinson (1991) and Hong and White (2005) base their tests for serial independence on the Kullback-Leibler information criterion, Ahmad and Li (1997) on an L_2 measure of distance between the joint density and the product of the marginals, and Pinkse (1998) on the distance between the joint characteristic function and the product of the marginal characteristic functions. In addition, Neumeyer (2009) considers a test for independence between regressors and error term in the context of nonparametric regression. Su and White (2003, 2007, 2008) adopt three different methods to test for conditional independence. Except CGL, none of the above nonparametric tests are developed to test for cross-sectional independence in panel data model.

In this chapter, we propose a nonparametric test for contemporary "*pair-wise cross-sectional independence*", which is based on the average of pair-wise L_2 distance between the joint density and the product of pair-wise marginals. Like CGL, we base our test on the residuals from local polynomial regressions. Unlike them, we are interested in the pair-wise independence of the error terms so that our test statistic is based on the comparison of the joint probability density with the product of pair-wise marginal probability densities. We first consider the case where tests for cross-sectional dependence are conducted on raw data so that there is no parameter estimation error involved and then consider the case with parameter estimation error. For both cases, we establish the asymptotic normal distribution of our test statistic under the null hypothesis of cross-sectional independence when $n \to \infty$ and $T \to \infty$ simultaneously. We also show that the test is consistent against global alternatives.

The rest of the chapter is organized as follows. Assuming away parameter estimation error, we introduce our testing statistic in Section 2 and study its asymptotic properties under both the null and the alternative hypotheses in Section 3. In Section 4 we study the asymptotic null distribution of our test statistic when tests are conducted on residuals from heterogeneous nonparametric regressions. In Section 5 we provide a small set of Monte Carlo simulation results to evaluate the finite sample performance of our test. Section 6 concludes. All proofs are relegated to the appendix.

NOTATION. Throughout the chapter we adopt the following notation and conventions. For a matrix A, we denote its transpose as A' and Euclidean norm as $\|A\| \equiv [\operatorname{tr}(AA')]^{1/2}$, where $\equiv$ means "is defined as". When A is a symmetric matrix, we use $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ to denote its minimum and maximum eigenvalues, respectively. The operator $\xrightarrow{p}$ denotes convergence in probability, and $\xrightarrow{d}$ convergence in distribution. Let $P_T^l \equiv T!/(T-l)!$ and $C_T^l \equiv T!/[(T-l)!l!]$ for integers $l \leq T$. We use $(n,T) \to \infty$ to denote the joint convergence of n and T when n and T pass to the infinity simultaneously.

2.2 Hypotheses and test statistics

To fix ideas and avoid distracting complications, we focus on testing pair-wise cross-sectional dependence in observables in this section and the next. The case of testing pair-wise cross-sectional dependence using unobservable error terms is studied in Section 4.

2.2.1 The hypotheses

Consider a nonparametric panel data model of the form

$$y_{it} = g_i(X_{it}) + u_{it},\ i = 1,2,\ldots,n,\ t = 1,2,\ldots,T, \tag{2.2.1}$$

where y_{it} is the dependent variable for individual i at time t, X_{it} is a $d \times 1$ vector of regressors in the ith equation, $g_i(\cdot)$ is unknown smooth regression function, and u_{it} is scalar random error term. We are interested in testing for the cross-sectional dependence in $\{u_{it}\}$. Since it seems impossible to design a test that can detect all kinds of cross-sectional dependence among $\{u_{it}\}$, as a starting point we focus on testing pair-wise cross-sectional dependence among them.

For each i, we assume that $\{u_{it}\}_{t=1}^T$ is a stationary time series process that has a probability density function (PDF) $f_i(\cdot)$. Let $f_{ij}(\cdot,\cdot)$ denote the joint PDF of u_{it} and u_{jt}. We can formulate the null hypothesis of pair-wise cross-sectional independence among $\{u_{it}, i = 1,...,n\}$ as

$$H_0 : f_{ij}\left(u_{it}, u_{jt}\right) = f_i(u_{it}) f_j\left(u_{jt}\right) \text{ almost surely (a.s.) for all } i,\ j = 1,\ldots,n, \text{ and } i \neq j. \tag{2.2.2}$$

That is, under H_0, u_{it} and u_{jt} are pair-wise independent for all $i \neq j$. The alternative hypothesis is

$$H_1: \text{ the negation of } H_0. \tag{2.2.3}$$

2.2.2 The test statistic

For the moment, we assume that $\{u_{it}\}$ is observed and consider a test for the null hypothesis in (2.2.2). Alternatively, one can regard g_i's are identically zero in (2.2.1) and testing for potential cross-sectional dependence among $\{y_{it}\}$. The proposed test is based on the average pair-wise L_2 distance between the joint density and the product of the marginal densities:

$$\Gamma_n = \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} \int \int \left[f_{ij}(u,v) - f_i(u) f_j(v) \right]^2 dudv, \tag{2.2.4}$$

where $\sum_{1 \leq i \neq j \leq n}$ stands for $\sum_{i=1}^n \sum_{j=1, j \neq i}^n$. Obviously, $\Gamma_n = 0$ under H_0 and is nonzero otherwise.

Since the densities are unknown to us, we propose to estimate them by the kernel method. That is, we estimate $f_i(u)$ and $f_{ij}(u,v)$ by

$$\begin{aligned} \widehat{f}_i(u) &\equiv T^{-1} \sum\nolimits_{t=1}^T h^{-1} k\left((u_{it} - u)/h\right), \text{ and} \\ \widehat{f}_{ij}(u,v) &\equiv T^{-1} \sum\nolimits_{t=1}^T h^{-2} k\left((u_{it} - u)/h\right) k\left(\left(u_{jt} - v\right)/h\right), \end{aligned}$$

where h is a bandwidth sequence and $k(\cdot)$ is a symmetric kernel function. Note that we use the same bandwidth and (univariate or product of univariate) kernel functions in estimating both the marginal and joint densities, which can facilitate the asymptotic analysis to a great deal. Then a natural test statistic is given by

$$\widehat{\Gamma}_{1nT} = \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} \int \int \left[\widehat{f}_{ij}(u,v) - \widehat{f}_i(u) \widehat{f}_j(v) \right]^2 dudv. \tag{2.2.5}$$

Let $\overline{k}_{h,ts}^i \equiv h^{-1} \overline{k}\left((u_{it} - u_{is})/h\right)$, where $\overline{k}(\cdot) \equiv \int k(u) k(\cdot - u) du$ is the two-fold convolution of $k(\cdot)$. It is easy to verify that we can rewrite $\widehat{\Gamma}_{1nT}$ as follows:

$$\widehat{\Gamma}_{1nT} = \frac{1}{n(n-1)} \sum_{1 \le i \ne j \le n} \left\{ \frac{1}{T^4} \sum_{1 \le t,s,r,q \le T} \overline{k}^{i}_{h,ts} \left(\overline{k}^{j}_{h,ts} + \overline{k}^{j}_{h,rq} - 2\overline{k}^{j}_{h,tr} \right) \right\}, \quad (2.2.6)$$

where $\sum_{1 \le t,s,r,q \le T} \equiv \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{r=1}^{T} \sum_{q=1}^{T}$.

The above statistic is simple to compute and offers a natural way to test H_0. Nevertheless, we propose a bias-adjusted test statistic, namely

$$\widehat{\Gamma}_{nT} = \frac{1}{n(n-1)} \sum_{1 \le i \ne j \le n} \left\{ \frac{1}{P_T^4} \sum_{1 \le t \ne s \ne r \ne q \le T} \overline{k}^{i}_{h,ts} \left(\overline{k}^{j}_{h,ts} + \overline{k}^{j}_{h,rq} - 2\overline{k}^{j}_{h,tr} \right) \right\}, \quad (2.2.7)$$

where $P_T^4 \equiv T!/[(T-4)!]$ and $\sum_{1 \le t \ne s \ne r \ne q \le T}$ denotes the sum over all different arrangements of the distinct time indices t, s, r, and q. In effect, $\widehat{\Gamma}_{nT}$ removes the the "diagonal" (e.g. $t = s, r = q, t = r$) elements from $\widehat{\Gamma}_{1nT}$, thus reducing the bias of the statistic in finite samples. A similar idea has been used in Lavergne and Vuong (2000), Su and White (2007), and Su and Ullah (2009), to name just a few. We will show that, after being appropriately centered and scaled, $\widehat{\Gamma}_{nT}$ is asymptotically normally distributed under the null hypothesis of cross-sectional independence and some mild conditions.

2.3 Asymptotic distributions of the test statistic

In this section we first present a set of assumptions that are used in deriving the asymptotic distributions of our test statistic. Then we study the asymptotic distribution of our test statistic under the null hypothesis and establish its consistency.

2.3.1 Assumptions

To study the asymptotic distribution of the test statistic with observable "errors" $\{u_{it}\}$, we make the following assumptions.

Assumption A.1 (i) For each i, $\{u_{it},\ t = 1,\ 2, \ldots\}$ is stationary and α-mixing with mixing coefficient $\{\alpha_i(\cdot)\}$ satisfying $\alpha_i(l) = O\left(\rho_i^l\right)$ for some $0 \leq \rho_i < 1$. Let $\overline{\rho} \equiv \max_{1\leq i\leq n} \rho_i$. We further require that $0 \leq \overline{\rho} < 1$.

(ii) For each i and $1 \leq l \leq 8$, the probability density function (PDF) $f_{i,t_1,\ldots,t_l}$ of $(u_{it_1},\ldots,u_{it_l})$ is bounded and satisfies a Lipschitz condition: $|f_{i,t_1,\ldots,t_l}(u_1+v_1,\ldots,u_l+v_l) - f_{i,t_1,\ldots,t_l}(u_1,\ldots,u_l)| \leq D_{i,t_1,\ldots,t_l}(\mathbf{u})||\mathbf{v}||$, where $\mathbf{u} \equiv (u_1,\ldots,u_l)$, $\mathbf{v} \equiv (v_1,\ldots,v_l)$, and $D_{i,t_1,\ldots,t_l}$ is integrable and satisfies the conditions that $\int_{\mathbb{R}^l} D_{i,t_1,\ldots,t_l}(\mathbf{u})||\mathbf{u}||^{2(1+\delta)}d\mathbf{u} < C_1$ and $\int_{\mathbb{R}^l} D_{i,t_1,\ldots,t_l}(\mathbf{u})\, f_{i,t_1,\ldots,t_l}(\mathbf{u})d\mathbf{u} < C_1$ for some $C_1 < \infty$ and $\delta \in (0,1)$. When $l = 1$, we denote the marginal PDF of u_{it} simply as f_i.

Assumption A.2 The kernel function $k : \mathbb{R} \to \mathbb{R}$ is a symmetric, continuous and bounded function such that $k(\cdot)$ is a γth order kernel: $\int k(u)\,du = 1$, $\int u^j k(u)\,du = 0$ for $j = 1,\ldots,\gamma-1$, and $\int u^\gamma k(u)\,du = \kappa_\gamma < \infty$.

Assumption A.3. As $(n,T) \to \infty$, $h \to 0$, $nT^2h^2 \to \infty$, $nh^{\frac{1-\delta}{1+\delta}}/T \to 0$.

Remark 1. Assumption A.1(i) requires that $\{u_{it},\, t = 1,\, 2,\ldots\}$ be a stationary strong mixing process with geometric decay rate. This requirement on the mixing rate is handy for our asymptotic analysis but can be relaxed to the usual algebraic decay rate with more complications involved in the proof. It is also assumed in several early works for stationary β-mixing processes such as Fan and Li (1999), Li (1999), and Su and White (2008), and can be satisfied by many well-known processes such as linear stationary autoregressive moving average (ARMA) processes, and bilinear and nonlinear autoregressive processes. Here we only assume that the stochastic process is strong mixing, which is weaker than β-mixing. Assumption A.1(ii) assumes some standard smooth conditions on the PDF of $(u_{it_1},\ldots,u_{it_l})$. Assumption A.2 imposes conditions on the kernel function which may or may not be a higher order kernel. The use of a higher order kernel typically aims at reducing the bias of kernel estimates, which is common in the nonparametric literature (see Robinson, 1988; Fan and Li, 1996; Li, 1999, and Su and White, 2008). Assumption A.3 imposes restrictions on the bandwidth, n, and T. These restrictions are weak and can be easily met in practice for a wide combinations of n and T. In addition, it is possible to have $n/T \to c \in [0,\infty]$ as $(n,T) \to \infty$.

By the proof of Theorem 2.3.1 below, one can relax Assumption A.1(i) to:

Assumption A.1(i*): For each i, $\{u_{it},\, t = 1, 2, ...\}$ is stationary and α-mixing with mixing coefficient $\alpha_i(\cdot)$. Let $\alpha(s) \equiv \max_{1\le i\le n} \alpha_i(s)$. $\sum_{\tau=1}^{\infty} \alpha^{\frac{\delta}{1+\delta}}(\tau) \le C_2$ for some $C_2 < \infty$ and $\delta \in (0,1)$. There exists $m \equiv m(n,T)$ such that

$$\max\left(n^{-1}T^4h^{\frac{4}{1+\delta}}, T^4h^{\frac{2(2+\delta)}{1+\delta}}, T^2h^{\frac{2}{1+\delta}}\right)\alpha^{\frac{\delta}{1+\delta}}(m) \to 0 \tag{2.3.1}$$

and $\max\left(m^4h^4, m^3h^2\right) \to 0$ as $(n,T) \to \infty$.

For the result in Corollary 2.3.2 to hold, we further need m and $\alpha(\cdot)$ to meet the following condition.

Assumption A.1(i):** For the m and $\alpha(\cdot)$ defined in Assumption A.1(i*), they satisfy that $h^{\frac{2(1-\delta)}{1+\delta}}T^4\alpha^{\frac{\delta}{1+\delta}}(m) + h^2m^4 \to 0$ as $(n,T) \to \infty$.

Clearly, under Assumption A.1(i), we can take $m = \lfloor L\log T\rfloor$ (the integer part of $L\log T$) for a large positive constant L such that both Assumptions A.1(i*) and A.1(i**) are satisfied. For notational simplicity, we continue to apply Assumption A.1(i).

2.3.2 Asymptotic null distributions

To state our main results, we further introduce some notation. Let E_t denote expectation with respect to variables with time indexed by t only. For example, $E_t[\overline{k}^i_{h,ts}] \equiv \int \overline{k}^i_{h,ts} f_i(u_{it})\,du_{it}$, and $E_tE_s[\overline{k}^i_{h,ts}] \equiv \int \left[\int \overline{k}^i_{h,ts} f_i(u_{it})\,du_{is}\right] f_i(u_{it})\,du_{it}$. Let $\varphi_{i,ts} \equiv \overline{k}^i_{h,ts} - E_t[\overline{k}^i_{h,ts}] - E_s[\overline{k}^i_{h,ts}] + E_tE_s[\overline{k}^i_{h,ts}]$. Define[1]

$$B_{nT} \equiv \frac{1}{n-1}\sum_{1\le i\neq j\le n}\frac{h}{T-1}\sum_{1\le t\neq s\le T} E\left[\varphi_{i,ts}\right]E\left[\varphi_{j,ts}\right], \text{ and} \tag{2.3.2}$$

$$\sigma^2_{nT} \equiv \frac{4h^2}{n(n-1)}\sum_{1\le i\neq j\le n}\frac{1}{T(T-1)}\sum_{1\le t\neq s\le T}\mathrm{Var}\left(\overline{k}^i_{h,ts}\right)\mathrm{Var}\left(\overline{k}^j_{h,ts}\right). \tag{2.3.3}$$

[1]The notation can be greatly simplied under identical distributions across individuals. In this case, $B_{nT} = n(T-1)^{-1}h\sum_{1\le t\neq s\le n}[E(\varphi_{1,ts})]^2$, and $\sigma^2_{nT} = 4[T(T-1)]^{-1}h^2\sum_{1\le t\neq s\le n}[\mathrm{Var}(\overline{k}^1_{h,ts})]^2$.

We establish the asymptotic null distribution of the $\widehat{\Gamma}_{nT}$ test statistic in the following theorem.

Theorem 2.3.1 *Suppose Assumptions A.1-A.3 hold. Then under the null of cross-sectional independence we have*

$$nTh\widehat{\Gamma}_{nT} - B_{nT} \xrightarrow{d} N\left(0, \sigma_0^2\right) \text{ as } (n,T) \to \infty,$$

where $\sigma_0^2 \equiv \lim_{(n,T)\to\infty} \sigma_{nT}^2$.

Remark 2. The proof of Theorem 2.3.1 is tedious and is relegated to Appendix A. The idea underlying the proof is simple but the details are quite involved. To see how complications arise, let $\gamma_{nT,ij} \equiv \gamma_{nT}\left(\mathbf{u}_i, \mathbf{u}_j\right) \equiv \frac{1}{P_T^4} \sum_{1\le t\neq s\neq r\neq q\le T} \overline{k}_{h,ts}^{i} (\overline{k}_{h,ts}^{j} + \overline{k}_{h,rq}^{j} - 2\overline{k}_{h,tr}^{j})$ where $\mathbf{u}_i \equiv (u_{i1}, \ldots, u_{iT})'$. Then we have $\widehat{\Gamma}_{nT} = \frac{1}{n(n-1)} \sum_{1\le i\neq j\le n} \gamma_{nT}\left(\mathbf{u}_i, \mathbf{u}_j\right)$. Clearly, for each pair (i,j) with $i \neq j$, $\gamma_{nT,ij}$ is a fourth order U-statistic along the time dimension, and by treating γ_{nT} as a kernel function, $\widehat{\Gamma}_{nT}$ can be regarded as a second order U-statistic along the individual dimension. To the best of our knowledge, there is no literature that treats such a *two-fold* U-statistic, and it is not clear in the first sight how one should pursue in order to yield a useful central limit theorem (CLT) for $\widehat{\Gamma}_{nT}$. Even though it seems apparent for us to apply the idea of Hoeffding decomposition, how to pursue it is still challenging.

In this chapter, we first apply the Hoeffding decomposition on $\gamma_{nT,ij}$ for each pair (i,j) and demonstrate that $\gamma_{nT,ij}$ can be decomposed as follows

$$\gamma_{nT,ij} = 6G_{nT,ij}^{(2)} + 4G_{nT,ij}^{(3)} + G_{nT,ij}^{(4)}$$

where, for $l = 2,3,4$, $G_{nT,ij}^{(l)} \equiv \frac{1}{P_T^l} \sum_{1\le t_1\neq \ldots \neq t_l\le T} \vartheta_{ij}^{(l)}\left(Z_{ij,t_1}, \ldots, Z_{ij,t_l}\right)$ is an l-th order degenerate U-statistic with kernel $\vartheta_{ij}^{(l)}$ being formerly defined in Appendix A, and $Z_{ij,t} \equiv \left(u_{it}, u_{jt}\right)$. Then we can obtain the corresponding decomposition for $\widehat{\Gamma}_{nT}$:

$$\widehat{\Gamma}_{nT} = 6G_{nT}^{(2)} + 4G_{nT}^{(3)} + G_{nT}^{(4)}$$

where $G_{nT}^{(l)} \equiv \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} G_{nT,ij}^{(l)}$ for $l = 2, 3, 4$. Even though for each pair (i, j), $G_{nT,ij}^{(l)}$ is an l-th order degenerate U-statistic with kernel $\vartheta_{ij}^{(l)}$ along the time dimension under H_0, $G_{nT}^{(l)}$ is by no means an l-th order degenerate U-statistic along the individual dimension under H_0. Despite this, we can conjecture as usual that the dominant term in the decomposition of $\widehat{\Gamma}_{nT}$ is given by the first term $6G_{nT}^{(2)}$, and the other two terms $4G_{nT}^{(3)}$ and $G_{nT}^{(4)}$ are asymptotically negligible. So in the second step, we make a decomposition for $6G_{nT}^{(2)} - 6E[G_{nT}^{(2)}]$ and demonstrate that

$$nTh\left\{6G_{nT}^{(2)} - 6E[G_{nT}^{(2)}]\right\} = \sum_{1 \leq i < j \leq n} w_{nT}\left(\mathbf{u}_i, \mathbf{u}_j\right) + o_P(1)$$

where $w_{nT}\left(\mathbf{u}_i, \mathbf{u}_j\right) \equiv \frac{4h}{nT} \sum_{1 \leq t < s \leq T} \varphi_{i,ts}^c \varphi_{j,ts}^c$, and $\varphi_{i,ts}^c = \varphi_{i,ts} - E\left[\varphi_{i,ts}\right]$. Despite the fact that $w_{nT,ij} \equiv w_{nT}\left(\mathbf{u}_i, \mathbf{u}_j\right)$ is a non-degenerate second order U-statistic along the time dimension any more, $\sum_{1 \leq i < j \leq n} w_{nT}\left(\mathbf{u}_i, \mathbf{u}_j\right)$ is a *degenerate* second order U-statistic along the individual dimension. The latter enables us to apply the de Jong's (1987) CLT for second order degenerate U-statistics with independent but non-identical observations. [Under the null hypothesis of cross-sectional independence $\mathbf{u}_i$'s are independent across i but not identically distributed.] The asymptotic variance of $\sum_{1 \leq i < j \leq n} w_{nT}\left(\mathbf{u}_i, \mathbf{u}_j\right)$ is given by σ_0^2 defined in Theorem 2.3.1 and $6nThE[G_{nT}^{(2)}]$ delivers the asymptotic bias B_{nT} to be corrected from the final test statistic. In the third step, for $l = 3, 4$ we demonstrate $nThG_{nT}^{(l)} = o_P(1)$ by using the explicit formula of $\vartheta_{ij}^{(l)}$.

Remark 3. The asymptotic distribution in Theorem 2.3.1 is obtained by letting n and T pass to ∞ simultaneously. Phillips and Moon (1999) introduce three approaches to handle large dimensional panel, namely, sequential limit theory, diagonal path limit theory, and joint limit theory, and discuss relationships between the sequential and joint limit theory. As they remark, the joint limit theory generally requires stronger conditions to establish than the sequential or diagonal path convergence, and by the same token, the results are also stronger and may be expected to be relevant to a wider range of circumstances.

To implement the test, we require consistent estimates of σ_{nT}^2 and B_{nT}. Noting that

$$\begin{aligned}\sigma_{nT}^2 &= \frac{4h^2}{n(n-1)T(T-1)}\sum_{1\le i\neq j\le n}\sum_{1\le t\neq s\le T} E\left[\left(\overline{k}^i_{h,ts}\right)^2\right]E\left[\left(\overline{k}^j_{h,ts}\right)^2\right]+o(1)\\ &= \frac{4R\left(\overline{k}\right)^2}{n(n-1)T(T-1)}\sum_{1\le i\neq j\le n}\sum_{1\le t\neq s\le T}\int f_{i,ts}(u,u)\,du\int f_{j,ts}(v,v)\,dv+o(1),\end{aligned}$$

where $R\left(\overline{k}\right)\equiv\int\overline{k}(u)^2\,du$, then we can estimate σ_{nT}^2 by

$$\widehat{\sigma}_{nT}^2\equiv\frac{4R\left(\overline{k}\right)^2}{n(n-1)}\sum_{1\le i\neq j\le n}\frac{1}{T}\sum_{t=1}^{T}\widehat{f}_{ij,-t}\left(u_{it},u_{jt}\right)$$

where $\widehat{f}_{ij,-t}\left(u_{it},u_{jt}\right)\equiv(T-1)^{-1}\sum_{s=1,s\neq t}^{T}h^{-2}\ k\left((u_{is}-u_{it})/h\right)\ k\left(\left(u_{js}-u_{jt}\right)/h\right)$, i.e., $\widehat{f}_{ij,-t}(u_{it},\ u_{jt})$ is the leave-one-out estimate of $f_{ij}(u_{it},u_{jt})$. One can readily demonstrate $\widehat{\sigma}_{nT}^2$ is a consistent estimate of σ_{nT}^2 under the null. Let

$$\widehat{B}_{nT}\equiv\frac{2}{T-1}\sum_{r=2}^{T}\frac{(T-r+1)h}{n-1}\sum_{1\le i\neq j\le n}\widehat{E}\left[\varphi_{i,1r}\right]\widehat{E}\left[\varphi_{j,1r}\right],$$

where $\widehat{E}\left[\varphi_{i,1r}\right]\equiv(T-r+1)^{-1}\sum_{t=1}^{T-r+1}\overline{k}^i_{h,t,t+r-1}-T^{-1}(T-1)^{-1}\sum_{1\le t\neq s\le T}\overline{k}^i_{h,ts}$. We establish the consistency of $\widehat{B}_{nT}$ for B_{nT} in Appendix B. Then we can define a feasible test statistic:

$$\widehat{I}_{nT}=\frac{nTh\widehat{\Gamma}_{nT}-\widehat{B}_{nT}}{\widehat{\sigma}_{nT}},$$

which is asymptotically distributed as standard normal under the null. We can compare $\widehat{I}_{nT}$ to the one-sided critical value z_α, the upper α percentile from the standard normal distribution, and reject the null if $\widehat{I}_{nT}>z_\alpha$. The following corollary formerly establishes the asymptotic normal distribution of $\widehat{I}_{nT}$ under H_0

Corollary 2.3.2 *Suppose the conditions in Theorem 2.3.1 hold. Then we have*

$$\widehat{I}_{nT}\xrightarrow{d}N(0,1)\ \text{as}\ (n,T)\to\infty.$$

2.3.3 Consistency

To study the consistency of our test, we consider the nontrivial case where $\mu_A \equiv \lim_{n\to\infty}\Gamma_n > 0$, where

$$\Gamma_n \equiv \frac{1}{n(n-1)} \sum_{1\leq i\neq j\leq n} \int\int \left[f_{ij}(u,v) - f_i(u) f_j(v)\right]^2 dudv.$$

We need to add the following assumption that takes into account cross-sectional dependence under the alternative.

Assumption A.4 For each pair (i,j) with $i \neq j$, the joint PDF f_{ij} of u_{it} and u_{jt} is bounded and satisfies a Lipschitz condition: $|f_{ij}(u_1+v_1,u_2+v_2) - f_{ij}(u_1,u_2)| \leq D_{ij}(u_1,u_2)||(v_1,v_2)||$, and D_{ij} is integrable uniformly in (i,j): $\int\int D_{ij}(u,v)f_{ij}(u,v)$ $dudv < C_3$ for some $C_3 < \infty$.

The following theorem establishes the consistency of the test.

Theorem 2.3.3 *Suppose Assumptions A.1-A.4 hold and* $\mu_A > 0$*. Then under* H_1*,* $P\left(\widehat{I}_{nT} > d_{nT}\right) \to 1$ *for any sequence* $d_{nT} = o_P(nTh)$ *as* $(n,T) \to \infty$*.*

Remark 4. Theorem 2.3.3 indicates that under H_1 our test statistic $\widehat{I}_{nT}$ explodes at the rate nTh provided $\mu_A > 0$. This can occur if $f_{ij}(u,v)$ and $f_i(u) f_j(v)$ differ on a set of positive measure for a "large" number of pairs (i,j) where the number of explodes to the infinity at rate n^2. It rules out the case where they differ on a set of positive measure only for a finite fixed number of pairs, or the case where the number of pair-wise joint PDFs that differ from the product of the corresponding marginal PDFs on a set of positive measure is diverging to infinity as $n \to \infty$ but at a slower rate than n^2. In either case, our test statistic $\widehat{I}_{nT}$ cannot explode to the infinity at the rate nTh, but can still be consistent. Specifically, as long as $\lambda_{nT}\Gamma_n \to \mu_A$ and $\lambda_{nT}/(nTh) \to 0$ as $(n,T) \to \infty$ for some diverging sequence $\{\lambda_{nT}\}$, our test is still consistent as $\widehat{I}_{nT}$ now diverges to infinite at rate $(nTh)/\lambda_{nT}$.

Remark 5. We have not studied the asymptotic local power property of our test. Unlike the CGL's test for cross-sectional uncorrelation, it is difficult for us to set up a desirable sequence of Pitman local alternatives that converge to the null at a certain rate and yet enable us to obtain the nontrivial asymptotic power property of our test. Once we deviate from the null hypothesis, all kinds of cross-sectional dependence can arise in the data, which makes the analysis complicated and challenging. See also the remarks in Section 6.

2.4 Tests based on residuals from nonparametric regressions

In this section, we consider tests for cross-sectional dependence among the unobservable error terms in the nonparametric panel data model (2.2.1). We must estimate the error terms from the data before conducting the test.

We assume that the regression functions $g_i(\cdot)$, $i = 1, \ldots, n$, are sufficiently smooth, and consider estimating them by the pth order local polynomial method ($p = 1, 2, 3$ in most applications). See Fan and Gijbels (1996) and Li and Racine (2007) for the advantage of local polynomial estimates over the local constant (Nadaraya-Watson) estimates. If $g_i(\cdot)$ has derivatives up to the pth order at a point x, then for any X_{it} in a neighborhood of x, we have

$$\begin{aligned} g_i(X_{it}) &= g_i(x) + \sum_{1 \le |\mathbf{j}| \le p} \frac{1}{\mathbf{j}!} D^{|\mathbf{j}|} g_i(x) (X_{it} - x)^{\mathbf{j}} + o(\|X_{it} - x\|^p) \\ &\equiv \sum_{0 \le |\mathbf{j}| \le p} \beta_{i,\mathbf{j}}(x; b) ((X_{it} - x)/b)^{\mathbf{j}} + o(\|X_{it} - x\|^p). \end{aligned}$$

Here, we use the notation of Masry (1996a, 1996b): $\mathbf{j} = (j_1, \ldots, j_d)$, $|\mathbf{j}| = \sum_{a=1}^d j_a$, $x^{\mathbf{j}} = \prod_{a=1}^d x_a^{j_a}$, $\sum_{0 \le |\mathbf{j}| \le p} = \sum_{l=0}^p \sum_{\substack{j_1=0 \\ j_1+\ldots+j_d=l}}^l \cdots \sum_{j_d=0}^l$, $D^{|\mathbf{j}|} g_i(x) = \frac{\partial^{|\mathbf{j}|} g_i(x)}{\partial^{j_1} x_1 \ldots \partial^{j_d} x_d}$, $\beta_{i,\mathbf{j}}(x; b) = \frac{b^{|\mathbf{j}|}}{\mathbf{j}!} D^{|\mathbf{j}|} g_i(x)$, where $\mathbf{j}! \equiv \prod_{a=1}^d j_a!$ and $b \equiv b(n, T)$ is a bandwidth parameter that controls how "close" X_{it} is from x. With observations $\{(y_{it}, X_{it})\}_{t=1}^T$, we consider choosing β_i, the stack of $\beta_{i,\mathbf{j}}$ in a lexicographical order, to minimize the following criterion function

$$Q_T(x;\beta_i) \equiv T^{-1}\sum_{t=1}^{T}\left(y_{it} - \sum_{0\le|\mathbf{j}|\le p}\beta_{\mathbf{j}}((X_{it}-x)/b)^{\mathbf{j}}\right)^2 w_b(X_{it}-x), \qquad (2.4.1)$$

where $w_b(x) = b^{-d}w(x/b)$, and w is a symmetric PDF on $\mathbb{R}^d$. The pth order local polynomial estimate of $g_i(x)$ is then defined as the minimizing concept in the above minimization problem.

Let $N_l \equiv (l+d-1)!/(l!(d-1)!)$ be the number of distinct d-tuples $\mathbf{j}$ with $|\mathbf{j}| = l$. It denotes the number of distinct l-th order partial derivatives of $g_i(x)$ with respect to x. Arrange the N_l d-tuples as a sequence in the lexicographical order (with highest priority to last position), so that $\phi_l(1) \equiv (0,0,...,l)$ is the first element in the sequence and $\phi_l(N_l) \equiv (l,0,...,0)$ is the last element, and let ϕ_l^{-1} denote the mapping inverse to ϕ_l. Let $N \equiv \sum_{l=0}^{p} N_l$. Define $\mathbf{S}_{iT}(x)$ and $\mathbf{W}_{iT}(x)$ as a symmetric $N\times N$ matrix and an $N\times 1$ vector, respectively:

$$\mathbf{S}_{iT}(x) \equiv \begin{bmatrix} \mathbf{S}_{iT,0,0}(x) & \mathbf{S}_{iT,0,1}(x) & \cdots & \mathbf{S}_{iT,0,p}(x) \\ \mathbf{S}_{iT,1,0}(x) & \mathbf{S}_{iT,1,1}(x) & \cdots & \mathbf{S}_{iT,1,p}(x) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{S}_{iT,p,0}(x) & \mathbf{S}_{iT,p,1}(x) & \cdots & \mathbf{S}_{iT,p,p}(x) \end{bmatrix}, \ \mathbf{W}_{iT}(x) \equiv \begin{bmatrix} \mathbf{W}_{iT,0}(x) \\ \mathbf{W}_{iT,1}(x) \\ \vdots \\ \mathbf{W}_{iT,p}(x) \end{bmatrix}$$

where $\mathbf{S}_{iT,j,k}(x)$ is an $N_j \times N_k$ submatrix with the (l,r) element given by

$$\left[\mathbf{S}_{iT,j,k}(x)\right]_{l,r} \equiv \frac{1}{T}\sum_{t=1}^{T}\left(\frac{X_{it}-x}{b}\right)^{\phi_j(l)+\phi_k(r)} w_b(X_{it}-x),$$

and $\mathbf{W}_{iT,j}(x)$ is an $N_j \times 1$ subvector whose r-th element is given by

$$\left[\mathbf{W}_{iT,j}(x)\right]_r \equiv \frac{1}{T}\sum_{t=1}^{T} y_{it}\left(\frac{X_{it}-x}{b}\right)^{\phi_j(r)} w_b(X_{it}-x).$$

Then we can denote the pth order local polynomial estimate of $g_i(x)$ as

$$\widetilde{g}_i(x) \equiv e_1'\left[\mathbf{S}_{iT}(x)\right]^{-1}\mathbf{W}_{iT}(x)$$

where $e_1 \equiv (1,0,\cdots,0)'$ is an $N \times 1$ vector.

For each $\mathbf{j}$ with $0 \leq |\mathbf{j}| \leq 2p$, let $\mu_{\mathbf{j}} \equiv \int_{\mathbb{R}^d} x^{\mathbf{j}} w(x) dx$. Define the $N \times N$ dimensional matrix $\mathbb{S}$ by

$$\mathbb{S} \equiv \begin{bmatrix} \mathbb{S}_{0,0} & \mathbb{S}_{0,1} & \dots & \mathbb{S}_{0,p} \\ \mathbb{S}_{1,0} & \mathbb{S}_{1,1} & \dots & \mathbb{S}_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{S}_{p,0} & \mathbb{S}_{p,1} & \dots & \mathbb{S}_{p,p} \end{bmatrix}, \tag{2.4.2}$$

where $\mathbb{S}_{i,j}$ is an $N_i \times N_j$ dimensional matrix whose (l,r) element is $\mu_{\phi_i(l)+\phi_j(r)}$. Note that the elements of the matrix $\mathbb{S}$ are simply multivariate moments of the kernel w. For example, if $p=1$, then

$$\mathbb{S} = \begin{bmatrix} \int w(x)\,dx & \int x'w(x)\,dx \\ \int xw(x)\,dx & \int xx'w(x)\,dx \end{bmatrix} = \begin{bmatrix} 1 & \mathbf{0}_{1\times d} \\ \mathbf{0}_{d\times 1} & \int xx'w(x)\,dx \end{bmatrix},$$

where $\mathbf{0}_{a\times c}$ is an $a \times c$ matrix of zeros.

Let $\widetilde{u}_{it} \equiv y_{it} - \widetilde{g}_i(X_{it})$ for $i=1,\ldots,n$ and $t=1,\ldots,T$. Define $\widetilde{\Gamma}_{nT}$, $\widetilde{B}_{nT}$, and $\widetilde{\sigma}^2_{nT}$ analogously to $\widehat{\Gamma}_{nT}$, $\widehat{B}_{nT}$, $\widehat{\sigma}^2_{nT}$ but with $\{u_{it}\}$ being replaced by $\{\widetilde{u}_{it}\}$. Then we can consider the following "feasible" test statistic

$$\widetilde{I}_{nT} \equiv \frac{nTh\widetilde{\Gamma}_{nT} - \widetilde{B}_{nT}}{\widetilde{\sigma}_{nT}}.$$

To demonstrate the asymptotic equivalence of $\widetilde{I}_{nT}$ and $\widehat{I}_{nT}$, we add the following assumptions.

Assumption A.5 (i) For each $i=1,\ldots,n$, $\{X_{it},\, t=1,\,2,\ldots\}$ is stationary and α-mixing with mixing coefficient $\{a_i(\cdot)\}$ satisfying $\sum_{j=1}^{\infty} j^{\kappa_0} a(j)^{\delta_0/(2+\delta_0)} < C_4$ for some $C_4 < \infty$, $\kappa_0 > \delta_0/(2+\delta_0)$, and $\delta_0 > 0$, where $a(j) \equiv \max_{1\leq i\leq n} a_i(j)$.

(ii) For each $i=1,\ldots,n$, the support $\mathscr{X}_i$ of X_{it} is compact on $\mathbb{R}^d$. The PDF p_i of X_{it} exists, is Lipschitz continuous, and is bounded away from zero on $\mathscr{X}_i$ uniformly in $i: \min_{1\leq i\leq n} \inf_{x_i\in\mathscr{X}_i} p_i(x_i) > C_5$ for some $C_5 > 0$. The joint PDF of X_{it} and X_{is} is uniformly bounded for all $t \neq s$ by a constant that does not depend on i or $|t-s|$.

(iii) $\{u_{it}, i=1,2,\ldots,t=1,2,\ldots\}$ is independent of $\{X_{it}, i=1,2,\ldots,t=1,2,\ldots\}$.

Assumption A.6 (i) For each $i=1,\ldots,n$, the individual regression function $g_i(\cdot)$, is $p+1$ times continuously partially differentiable.

(ii) The $(p+1)$-th order partial derivatives of g_i are Lipschitz continuous on $\mathscr{X}_i$.

Assumption A.7 (i) The kernel function $w:\mathbb{R}^d \to \mathbb{R}^+$ is a continuous, bounded, and symmetric PDF; $\mathbb{S}$ is positive definite (p.d.).

(ii) Let $\mathbf{w}(x) \equiv \|x\|^{2(2+\delta_0)p} w(x)$. $\mathbf{w}$ is integrable with respect to the Lebesgue measure.

(iii) Let $W_{\mathbf{j}}(x) \equiv x^{\mathbf{j}} w(x)$ for all d-tuples $\mathbf{j}$ with $0 \leq |\mathbf{j}| \leq 2p+1$. $W_{\mathbf{j}}(x)$ is Lipschitz continuous for $0 \leq |\mathbf{j}| \leq 2p+1$. For some $C_6 < \infty$ and $C_7 < \infty$, either $w(\cdot)$ is compactly supported such that $w(x)=0$ for $\|x\| > C_6$, and $||W_{\mathbf{j}}(x) - W_{\mathbf{j}}(\widetilde{x})|| \leq C_7 ||x-\widetilde{x}||$ for any x, $\widetilde{x} \in \mathbb{R}^d$ and for all $\mathbf{j}$ with $0 \leq |\mathbf{j}| \leq 2p+1$; or $w(\cdot)$ is differentiable, $\left\|\partial W_{\mathbf{j}}(x)/\partial x\right\| \leq C_6$, and for some $\iota_0 > 1$, $|\partial W_{\mathbf{j}}(x)/\partial x| \leq C_6 \|x\|^{-\iota_0}$ for all $\|x\| > C_7$ and for all $\mathbf{j}$ with $0 \leq |\mathbf{j}| \leq 2p+1$.

Assumption A.8 (i) The kernel function k is second order differentiable with first order derivative k' and second order derivative k''. Both $uk(u)$ and $uk'(u)$ tend to 0 as $|u| \to \infty$. (ii) For some $c_k < \infty$ and $A_k < \infty$, $|k''(u)| \leq c_k$ and for some $\gamma_0 > 1$, $|k''(u)| \leq c_k |u|^{-\gamma_0}$ for all $|u| > A_k$.

Assumption A.9 (i) Let $\eta \equiv T^{-1}b^{-d} + b^{2(p+1)}$. As $(n,T) \to \infty$, $Th^5 \to \infty$, $T^{3/2}b^d h^5 \to \infty$, and $nTh(\eta^2 + h^{-4}\eta^3 + h^{-8}\eta^4) \to 0$.

(ii) For the m defined in Assumption A.1(i*), $\max(nhmb^{2(p+1)}$, $nmT^{-1}b^{-d}$, $n^2T^{-4}m^6h^{-2}$, $n^2m^2h^{-2}b^{4(p+1)}$, nhm^2/T, $nh^{-3}m^3/T^2$, $m^3/T) \to 0$.

Remark 6 Assumptions A.5 (i)-(ii) are subsets of some standard conditions to obtain the uniform convergence of local polynomial regression estimates. Like CGL, we assume the independence of $\{u_{it}\}$ and $\left\{X_{js}\right\}$ for all i, j, t, s in Assumptions A.5(iii), which will greatly facilitate our asymptotic analysis. Assumptions A.6 and A.7 are standard in the literature on local polynomial estimation. In particular, following Hansen (2008), the compact support of the kernel function w in

Masry (1996b) can be relaxed as in Assumption A.7(iii). Assumption A.8 specifies more conditions on the kernel function k used in the estimation of joint and marginal densities of the error terms. They are needed because we need to apply Taylor expansions on functions associated with k. Assumption A.9 imposes further conditions on h, n, and T and their interaction with the smoothing parameter b and the order p of local polynomial used in the local polynomial estimation. If we relax the geometric α-mixing rate in Assumption A.1(i) to the algebraic rate, then we need to add the following condition on the bandwidth parameters, sample sizes, and the choices of m and p :

Assumption A.1(i*):** For the m, $\alpha(\cdot)$, and δ defined in Assumption A.1(i*), they also satisfy that

$$\max\left\{n^2T^2h^{-3-\frac{\delta}{1+\delta}}, T^2h^{-4-\frac{2\delta}{1+\delta}},\ T^2h^{-5-\frac{2\delta}{1+\delta}}b^{4(p+1)}\right\}\alpha^{\frac{\delta}{1+\delta}}(m)\to 0 \text{ as } (n,T)\to\infty.$$

Theorem 2.4.1 *Suppose Assumptions A.1-A.3 and A.5-A.9 hold. Then under the null of cross-sectional independence*

$$\widetilde{I}_{nT}\to N(0,1)$$

as $(n,T)\to\infty$.

Remark 7. The above theorem establishes the asymptotic equivalence of $\widetilde{I}_{nT}$ and $\widehat{I}_{nT}$. That is, the test statistic $\widetilde{I}_{nT}$ that is based on the estimated residuals from heterogeneous local polynomial regressions is asymptotically equivalent to $\widehat{I}_{nT}$ that is constructed from the generally unobservable errors. If evidence suggests that the nonparametric regression relationships are homogeneous, i.e., $g_i(X_{it})=g(X_{it})$ a.s. for some function g on $\mathbb{R}^d$ and for all i, then one can pool the cross section data together and estimate the homogeneous regression function g at a faster rate than estimating each individual regression function g_i by using the time series observations for cross section i only. In this case, we expect that the requirement on the relationship of n,T, h,b, and p becomes less stringent. Similarly, if $g_i(X_{it})=\beta_{0i}+\beta_{1i}'X_{it}$

a.s. for some unknown parameters β_{0i} and β_{1i}, then we can estimate such parametric regression functions at the usual parametric rate $T^{-1/2}$, and it is easy to verify that the result in Theorem 2.4.1 continue to hold by using the residuals from time series parametric regressions for each individual.

The following theorem establishes the consistency of the test.

Theorem 2.4.2 *Suppose Assumptions A.1-A.9 hold and $\mu_A > 0$. Then under H_1, $P\left(\widetilde{I}_{nT} > d_{nT}\right) \to 1$ for any sequence $d_{nT} = o_P(nTh)$ as $(n,T) \to \infty$.*

The proof of the above theorem is almost identical to that of Theorem 2.3.3. The main difference is that one needs to apply Taylor expansions to show that $(nTh)^{-1}\widetilde{I}_{nT}$ is asymptotically equivalent to $(nTh)^{-1}\widehat{I}_{nT}$ under H_1. Remark 4 also holds for the test $\widetilde{I}_{nT}$.

2.5 Monte Carlo simulations

In this section, we conduct a small set of Monte Carlo simulations to evaluate the finite sample performance of our test and compare it with Pesaran's and CGL's tests for cross-sectional uncorrelation .

2.5.1 Data generating processes

We consider the following six data generating processes (DGPs) in our Monte Carlo study. DGPs 1-2 are for size study, and DGPs 3-6 are for power comparisons.

DGP 1:

$$y_{it} = \alpha_i + \beta_i X_{it} + u_{it},$$

where across both i and t, $X_{it} \sim$ IID $U(-3,3)$, $\alpha_i \sim$IID $U(0,1)$, $\beta_i \sim$ IID $N(0,1)$, and they are mutually independent of each other.

DGP 2:

$$y_{it} = (1+\theta_i)\exp(X_{it})/(1+\exp(X_{it})) + u_{it},$$

where across both i and t, $X_{it} \sim$ IID $U(-3,3)$, $\theta_i \sim$ IID $N(0,0.25)$, and they are mutually independent of each other.

In DGPs 1-2, we consider two kinds of error terms: (i) $u_{it} \sim$ IID $N(0,1)$ across both i and t and independent of $\{\alpha_i, \beta_i, X_{it}\}$; and (ii) $\{u_{it}\}$ is IID across i and an AR(1) process over t: $u_{it} = 0.5u_{i,t-1} + \varepsilon_{it}$, where $\varepsilon_{it} \sim$ IID $N(0,0.75)$ across both i and t and independent of $\{\alpha_i, \beta_i, X_{it}\}$. Clearly, there is no cross-sectional dependence in either case.

In terms of conditional mean specification, DGPs 3 and 5 are identical to DGP 1, and DGPs 4 and 6 are identical to DGP2. The only difference lies in the specification of the error term u_{it}. In DGPs 3-4, we consider the following single-factor error structure:

$$u_{it} = 0.5\lambda_i F_t + \varepsilon_{it} \tag{2.5.1}$$

where the factors F_t are IID $N(0,1)$, and the factor loadings λ_i are IID $N(0,1)$ and independent of $\{F_t\}$. We consider two configurations for ε_{it} : (i) ε_{it} are IID $N(0,1)$ and independent of $\{F_t, \lambda_i\}$, and (ii) $\varepsilon_{it} = 0.5\varepsilon_{it-1} + \eta_{it}$ where η_{it} are IID $N(0,0.75)$ across both i and t, and independent of $\{F_t, \lambda_i\}$.

In DGPs 5-6, we consider the following two-factor error structure:

$$u_{it} = 0.3\lambda_{1i}F_{1t} + 0.3\lambda_{2i}F_{2t} + \varepsilon_{it} \tag{2.5.2}$$

where both factors F_{1t} and F_{2t} are IID $N(0,1)$, λ_{1i} are IID $N(0,1)$, λ_{2i} are IID $N(0.5,1)$, F_{1t}, F_{2t}, λ_{1i}, and λ_{2i} are mutually independent of each other, and the error process $\{\varepsilon_{it}\}$ is specified as in DGPs 3-4 with two configurations.

2.5.2 Bootstrap

It is well known that the asymptotic normal distribution typically cannot approximate well the finite sample distribution of many nonparametric test statistics under the null hypothesis. In fact, the empirical level of these tests can be sensitive to the choice of bandwidths or highly distorted in finite samples. So we suggest using a bootstrap method to obtain the bootstrap p-values. Note that we need to estimate $E(\varphi_{ts})$ in B_{nT}, and that the dependent structure in each individual error process $\{u_{it}\}_{t=1}^{T}$ will affect the asymptotic distribution of our test under the null. Like Hsiao and Li (2001), we need to mimic the dependent structure over time. So we propose to apply the stationary bootstrap procedure of Politis and Romano (1994) to each individual i's residual series $\{\widetilde{u}_{it}\}_{t=1}^{T}$. The procedure goes as follows:

1. Obtain the local polynomial regression residuals $\widetilde{u}_{it} = Y_{it} - \widetilde{g}_i(\mathbf{x}_{it})$ for each i and t.

2. For each i, obtain the bootstrap time series sequence $\{u_{it}^*\}_{t=1}^{T}$ by the method of stationary bootstrap. [2]

3. Calculate the bootstrap test statistic $\widetilde{I}_{nT}^* = (nTh\widetilde{\Gamma}_{nT}^* - \widetilde{B}_{nT}^*)/\widetilde{\sigma}_{nT}^*$, where $\widetilde{\Gamma}_{nT}^*$, $\widetilde{B}_{nT}^*$ and $\widetilde{\sigma}_{nT}^*$ are defined analogously to $\widetilde{\Gamma}_{nT}$, $\widetilde{B}_{nT}$ and $\widetilde{\sigma}_{nT}$ but with $\widetilde{u}_{it}$ be replaced by u_{it}^*.

4. Repeat steps 2-3 for B times and index the bootstrap statistics as $\{\widetilde{I}_{nT,j}^*\}_{j=1}^{B}$. Calculate the bootstrap p-value $p^* \equiv B^{-1}\sum_{j=1}^{B}\mathbf{1}(\widetilde{I}_{nT,j}^* > \widetilde{I}_{nT})$ where $\mathbf{1}(\cdot)$ is the usual indicator function, and reject the null hypothesis of cross-sectional independence if p^* is smaller than the prescribed level of significance.

Note that we have imposed the null restriction of cross-sectional independence implicitly because we generate $\{u_{it}^*\}$ independently across all individuals. We conjecture that for sufficiently large B, the empirical distribution of $\{\widetilde{I}_{nT,j}^*\}_{j=1}^{B}$ is able to approximate the finite sample distribution of $\widetilde{I}_{nT}$ under the null hypothesis, but are not sure whether this can have any improvement over the asymptotic normal ap-

[2] A simple description of the resampling algorithm goes as follows. Let p be a fixed number in $(0,1)$. Let u_{i1}^* be picked at random from the original T residuals $\{\widetilde{u}_{i1}, ..., \widetilde{u}_{iT}\}$, so that $u_{i1}^* = \widetilde{u}_{iT_1}$, say, for some $T_1 \in \{1, ..., T\}$. With probability p, let u_{i2}^* be picked at random from the original T residuals $\{\widetilde{u}_{i1}, ..., \widetilde{u}_{iT}\}$; with probability $1-p$, let $u_{i2}^* = \widetilde{u}_{i,T_1+1}$ so that u_{i2}^* would be the "next" observation in the original residual series following $\widetilde{u}_{iT_1}$. In general, given that u_{it}^* is determined by the Jth observation $\widetilde{u}_{iJ}$ in the original residual series, let $u_{i,t+1}^*$ be equal to $\widetilde{u}_{i,J+1}$ with probability $1-p$ and be picked at random from the original T residuals with probability p. We set $p = T^{-1/3}$ in the simulations.

Table 2.1: Finite sample rejection frequency for DGPs 1-2 (size study, nomial level 0.05)

DGP	n	T	(i) $u_{it} \sim$ IID $N(0,1)$			(ii) $u_{it} = 0.5u_{i,t-1} + \varepsilon_{it}$		
			P	CGL	SZ	P	CGL	SZ
1	25	25	0.040	0.044	0.054	0.092	0.060	0.082
		50	0.060	0.044	0.048	0.130	0.062	0.082
		100	0.056	0.058	0.064	0.126	0.080	0.066
	50	25	0.060	0.044	0.062	0.118	0.066	0.128
		50	0.070	0.052	0.080	0.112	0.076	0.074
		100	0.034	0.030	0.048	0.124	0.066	0.064
2	25	25	0.038	0.044	0.052	0.088	0.050	0.090
		50	0.056	0.062	0.060	0.122	0.062	0.082
		100	0.058	0.044	0.064	0.128	0.068	0.070
	50	25	0.054	0.042	0.058	0.076	0.078	0.120
		50	0.064	0.060	0.060	0.110	0.050	0.084
		100	0.038	0.052	0.052	0.108	0.068	0.060

Note: P, CGL, and SZ refer to Pesaran's, CGL's and our tests, respectively.

proximation. The theoretical justification for the validity of our bootstrap procedure goes beyond the scope of this chapter.

2.5.3 Test results

We consider three tests of cross-sectional independence in this section: Pesaran's CD test for cross-sectional dependence, CGL test for cross-sectional uncorrelation, and the $\widetilde{I}_{nT}$ test proposed in this chapter. To conduct our test, we need to choose kernels and bandwidths. To estimate the heterogeneous regression functions, we conduct a third-order local polynomial regression $(p = 3)$ by choosing the second order Gaussian kernel and rule-of-thumb bandwidth: $b = s_X T^{-1/9}$ where s_X denotes the sample standard deviation of $\{X_{it}\}$ across i and t. To estimate the marginal and pairwise joint densities, we choose the second order Gaussian kernel and rule-of-thumb bandwidth $h = s_{\widetilde{u}} T^{-1/6}$, where $s_{\widetilde{u}}$ denotes the sample standard deviation of $\{\widetilde{u}_{it}\}$ across i and t. For the CGL test, we follow their paper and consider a local linear regression to estimate the conditional mean by using the Gaussian kernel and choosing the bandwidth through the leave-one-out cross-validation method. For the Pesaran's test, we estimate the heterogeneous regression functions by using the linear model, and conduct his CD test based on the parametric residuals.

For all tests, we consider $n = 25$, 50, and $T = 25$, 50, 100. For each combination of n and T, we use 500 replications for the level and power study, and 200 bootstrap resamples in each replication.

Table 2.1 reports the finite sample level for Pesaran's CD test, the CGL test and our test (denoted as P, CGL, and SZ, respectively in the table). When the error terms u_{it} are IID across t, all three tests perform reasonably well for all combinations of n and T and both DGPs under investigation in that the empirical levels are close to the nominal level. When $\{u_{it}\}$ follows an AR(1) process along the time dimension, we find out the CGL test outperforms the Pesaran's test in terms of level performance: the latter test tends to have a large size distortion which does not improve when either n or T increases. In contrast, our test can be oversized when n/T is not small (e.g., $n = 50$ and $T = 25)$ so that the parameter estimation error plays a non-negligible role in the finite samples, but the level of our test improves quickly as T increases for fixed n.

Table 2.2 reports the finite sample power performance of all three tests for DGPs 3-6. For DGPs 3-4, we have a single-factor error structure. Noting that the factor loadings λ_i have zero mean in our setup, neither Pesaran's nor CGL's test has power in detecting cross-sectional dependence in this case. This is confirmed by our simulations. In contrast, our tests have power in detecting deviations from cross-sectional dependence. As either n or T increases, the power of our test increases. DGPs 5-6 exhibit a two-factor error structure where one of the two sequences of factor loadings have nonzero mean, and all three tests have power in detecting cross-sectional dependence. As either n or T increases, the powers of all three tests increase quickly and our test tends to more powerful than the Pesaran's and CGL's tests.

2.6 Concluding remarks

In this chapter, we propose a nonparametric test for cross-sectional dependence in large dimensional panel. Our tests can be applied to both raw data and residuals

Table 2.2: Finite sample rejection frequency for DGPs 3-6 (power study, nomial level 0.05)

DGP	n	T	(i) $\varepsilon_{it} \sim$ IID $N(0,1)$			(ii) $\varepsilon_{it} = 0.5\varepsilon_{it-1} + \eta_{it}$		
			P	CGL	SZ	P	CGL	SZ
3	25	25	0.040	0.046	0.446	0.092	0.052	0.590
		50	0.060	0.058	0.778	0.130	0.060	0.860
		100	0.056	0.074	0.950	0.126	0.038	0.984
	50	25	0.060	0.040	0.772	0.118	0.070	0.866
		50	0.070	0.060	0.972	0.112	0.074	0.992
		100	0.034	0.064	0.998	0.124	0.068	1.000
4	25	25	0.038	0.074	0.446	0.098	0.044	0.616
		50	0.056	0.052	0.772	0.206	0.066	0.858
		100	0.058	0.062	0.954	0.234	0.044	0.984
	50	25	0.054	0.046	0.772	0.148	0.086	0.870
		50	0.064	0.068	0.970	0.190	0.072	0.990
		100	0.038	0.062	0.998	0.270	0.068	1.000
5	25	25	0.326	0.248	0.208	0.410	0.304	0.418
		50	0.412	0.332	0.444	0.486	0.350	0.672
		100	0.584	0.446	0.740	0.594	0.424	0.910
	50	25	0.550	0.442	0.456	0.626	0.508	0.680
		50	0.720	0.620	0.812	0.754	0.640	0.918
		100	0.842	0.742	0.988	0.888	0.776	0.996
6	25	25	0.304	0.232	0.250	0.420	0.292	0.406
		50	0.428	0.330	0.424	0.488	0.348	0.634
		100	0.568	0.426	0.762	0.588	0.402	0.908
	50	25	0.548	0.454	0.424	0.624	0.516	0.662
		50	0.724	0.636	0.814	0.760	0.636	0.908
		100	0.838	0.746	0.980	0.888	0.794	1.000

Note: P, CGL, and SZ refer to Pesaran's, CGL's and our tests, respectively.

from heterogenous nonparametric (or parametric) regressions. The requirement on the relative magnitude of n and T is quite weak in the former case, and very strong in the latter case in order to control the asymptotic effect of the parameter estimation error on the test statistic. In both cases, we establish the asymptotic normality of our test statistic under the null hypothesis of cross-sectional independence. The global consistency of our test is also established. Monte Carlo simulations indicate our test performs reasonably well in finite samples and has power in detecting cross-sectional dependence when the Pesaran's and CGL's tests fail.

We have not pursued the asymptotic local power analysis for our nonparametric test in this chapter. It is well known that the study of asymptotic local power is rather difficult in nonparametric testing for serial dependence, see Tjøstheim (1996) and Hong and White (2005). Similar remark holds true for nonparametric testing for cross-sectional dependence. To analyze the local power of their test, Hong and White (2005) consider a class of locally j-dependent processes for which there exists serial dependence at lag *j only*, but j may grow to infinity as the sample size passes to infinity. It is not clear whether one can extend their analysis to our framework since there is no natural ordering along the individual dimensions in panel data models. In addition, it may not be advisable to consider a class of panel data models for which there exists cross-sectional dependence at pairwise level only: if any two of u_{it}, u_{jt}, and u_{kt} $(i \neq j \neq k)$ are dependent, they tend to be dependent on the other one also. Thus we conjecture that it is very challenging to conduct the asymptotic local power analysis for our nonparametric test.

Chapter 3 Nonparametric Dynamic Panel Data Models with Interactive Fixed Effects: Sieve Estimation and Specification Testing

3.1 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 can capture unobserved heterogeneity more flexibly than the traditional ones with additive individual or time fixed effects. As common factors affect all individuals and then form a source of cross-sectional dependence, interactive fixed effects have become a powerful and popular tool to model cross-sectional 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 framework of Gaussian quasi-maximum likelihood estimation (QMLE); they obtain the first order asymptotic theory for the QMLE for linear dynamic panel data models with interactive fixed effects in the first paper 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 in the second paper. 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 and prove the oracle property of their regression coefficient estimator. 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 the 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 of panel factor models with large N and large T by using structural restrictions from economic theory, Pesaran and Tosetti (2011) for estimation of panel data models both with multifactor error structure and spatial correlation, Su and Chen (2013) for testing for slope homogeneity, Su, Jin, and Zhang (2012) for specification test of linearity in panel data models, 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 (2012) 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 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 works for 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 these serious consequences of misspecification, there is a need to develop tests for linear functional forms. To the best of our knowledge, the only available test for linearity in the framework work of dynamic panel data models with interactive fixed effects is due to Su, Jin, and Zhang (2012), who propose a test based on residuals from the estimation under the null hypothesis of linearity. But they do not propose consistent estimates of the regression functions once the null of linearity is rejected.

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

$$Y_{it} = g(X_{it}) + \lambda_i^{0\prime} f_t^0 + e_{it}, \tag{3.1.1}$$

where $i = 1, ..., N$, $t = 1, ..., T$, 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 $X_{1,it}$, $g(\cdot)$ is an unknown smooth function, f_t^0 is an $R \times 1$ vector of common factors, λ_i^0 is an $R \times 1$ vector of factor loadings so that $\lambda_i^{0\prime} f_t^0 = \sum_{l=1}^R \lambda_{li}^0 f_{lt}^0$, 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 chapter but discuss how to determine R in empirical applications.

The model specified in (3.1.1) 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 $t = 1, ..., T$ and some constant vector $\bar{f}$, the interactive fixed effects become commonly-used additive individual fixed effects. When f_t^0 is time-invariant and $g(X_{it}) = X_{it}'\theta^0$, (3.1.1) 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 $X_{it} = Y_{i,t-1}$, (3.1.1) reduces to the nonparametric dynamic panel data model in Lee (2010); when f_t^0 is time-invariant and only exogenous regressors are included in X_{it}, (3.1.1) 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, (3.1.1) becomes the general nonparametric dynamic panel data model, which is investigated by Su and Lu (2013); when f_t^0 is time-invariant and $g(X_{it}) = h(Y_{i,t-1}) + \theta^{0\prime} X_{1,it}$, (3.1.1) becomes the partially linear dynamic panel data model in Baglan (2009); when $g(X_{it}) = X_{it}'\theta^0$, (3.1.1) becomes the model studied by Bai (2009) and Moon and Weidner (2010, 2012). These authors propose various estimators for $g(\cdot)$ (or θ^0) and (λ_i^0, f_t^0) and establish their asymptotic properties.

We are mainly interested in consistent estimation and specification testing for the nonparametric component $g(\cdot)$ in (3.1.1). Noting that $g(\cdot)$ is an unknown smooth function, we combine the method of sieves with the Gaussian QMLE and propose a nonparametric sieve estimator of $g(\cdot)$. Following Moon and Weidner (2010, 2012), we establish its consistency, derive its convergence rates 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 $g(\cdot)$. Using an empirical L_2-distance, we compare two estimators for $g(\cdot)$, 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 chapter 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 in our chapter the correlation between X_{it} and randomly realized fixed effects (f_t^0, λ_i^0) complicates the analysis substantially. To be specific, 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 $\mathscr{D}$ 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 (2011), Su and Chen (2013), and Su, Jin, and Zhang (2012).

The rest of the chapter 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. Throughout the chapter we adopt the following notation. Let $\mu_i(A)$ denote the ith largest eigenvalue (counting multiple eigenvalues 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 X, let $\|X\|_p \equiv [E(\sum_{i=1}^{n} |X_i|^p)]^{1/p}$ denote its L_p-norm, and $\|X\|_{p,\mathscr{D}}$ $\equiv \{E[(\sum_{i=1}^{n} |X_i|^p)|\mathscr{D}]\}^{1/p}$ its L_p-norm conditional on $\mathscr{D}$. 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 ∞ simultaneously.

3.2 Sieve-based quasi-likelihood maximum estimation

Since $g(\cdot)$ is an unknown function in (3.1.1), 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 (3.1.1) very well by $\beta_g' p^K(x)$ for some $K \times 1$ vector β_g under fairly weak conditions. Let $K \equiv K_{NT}$ be some integer such that $K \to \infty$ as $(N,T) \to \infty$. We introduce the following notation: $p_{it,k} \equiv p_k^K(X_{it})$, $p_{it} \equiv p^K(X_{it})$, $P_i \equiv \left(p_{i1}', \cdots, p_{iT}'\right)'$, $P_{i\cdot,k} \equiv \left(p_{i1,k}, \cdots, p_{iT,k}\right)'$, $\mathbf{P}_k \equiv \left(P_{1\cdot,k}, \cdots, P_{N\cdot,k}\right)'$, $Y_i \equiv (Y_{i1}, \cdots, Y_{iT})'$, $\mathbf{Y} \equiv (Y_1, \cdots, Y_N)'$, $f^0 \equiv (f_1^0, \cdots, f_T^0)'$, $\lambda^0 \equiv \left(\lambda_1^0, \cdots, \lambda_N^0\right)'$. We use β^0 to denote the true vector of coefficients β_g in the sieve approximation of $g(x)$ given basis $p^K(x)$. Here we suppress the dependence of p_{it}, β^0, and β_g on K for notational simplicity.

To estimate g, we consider the following approximating linear panel data models with interactive fixed effects:

$$Y_{it} = p_{it}'\beta^0 + \lambda_i^{0\prime} f_t^0 + u_{it} \tag{3.2.1}$$

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 vector and matrix notation, (3.2.1) can be respectively rewritten as

$$Y_i = P_i\beta^0 + f^0\lambda_i^0 + u_i$$

and

$$\mathbf{Y} = \sum_{k=1}^{K} \beta_k^0 \mathbf{P}_k + \lambda^0 f^{0\prime} + \mathbf{u}. \tag{3.2.2}$$

Then we follow Bai (2009) and Moon and Weidner (2010) and estimate the model in (3.2.2) by the Gaussian QMLE method. Specifically, we obtain the estimator $(\hat{\beta}, \hat{\lambda}, \hat{f})$ of $(\beta^0, \lambda^0, f^0)$ as follows

$$(\hat{\beta}, \hat{\lambda}, \hat{f}) = \underset{(\beta,\lambda,f)}{\arg\min} \mathscr{L}(\beta, \lambda, f), \tag{3.2.3}$$

where $\mathscr{L}(\beta, \lambda, f)$ is the approximating negative quasi-log-likelihood function:

$$\mathscr{L}(\beta, \lambda, f) = \frac{1}{NT} \operatorname{tr}\left[\left(\mathbf{Y} - \sum_{k=1}^{K} \beta_k \mathbf{P}_k - \lambda f'\right)' \left(\mathbf{Y} - \sum_{k=1}^{K} \beta_k \mathbf{P}_k - \lambda f'\right)\right], \tag{3.2.4}$$

$\beta = (\beta_1, \cdots, \beta_K)'$, $f \equiv (f_1, \cdots, f_T)'$, and $\lambda \equiv (\lambda_1, \cdots, \lambda_N)'$. In particular, β can be estimated by

$$\hat{\beta} = \underset{\beta \in \mathbb{R}^K}{\arg\min} L_{NT}(\beta) \tag{3.2.5}$$

where $L_{NT}(\beta)$ is the profile approximating negative quasi-log-likelihood function:

$$\begin{aligned} L_{NT}(\beta) &= \min_{\lambda,f} \mathscr{L}_{NT}(\beta, \lambda, f) & (3.2.6) \\ &= \min_{f} \frac{1}{NT} \operatorname{tr}\left[\left(\mathbf{Y} - \sum_{k=1}^{K} \beta_k \mathbf{P}_k\right) M_f \left(\mathbf{Y} - \sum_{k=1}^{K} \beta_k \mathbf{P}_k\right)'\right] & (3.2.7) \\ &= \frac{1}{NT} \sum_{t=R+1}^{T} \mu_t \left[\left(\mathbf{Y} - \sum_{k=1}^{K} \beta_k \mathbf{P}_k\right)' \left(\mathbf{Y} - \sum_{k=1}^{K} \beta_k \mathbf{P}_k\right)\right]. & (3.2.8) \end{aligned}$$

See Moon and Weidner (2010) for the demonstration of equivalence of the above three expressions. Based on (3.2.8), 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

$$\hat{g}(x) = p^K(x)' \hat{\beta}. \tag{3.2.9}$$

The expression in (3.2.8) 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

$$\mathbf{Y}-\sum_{k=1}^{K}\beta_k\mathbf{P}_k=\underbrace{\lambda^{0\prime}f^0}_{\text{leading term}}+\underbrace{\sum_{k=1}^{K}\left(\beta_k^0-\beta_k\right)\mathbf{P}_k+\mathbf{e}+\mathbf{e}_g}_{\text{perturbation terms}} \tag{3.2.10}$$

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), (3.2.10) 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 (3.2.10), $L_{NT}(\beta)$ 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.

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

In this section, we first derive the convergence rate for $\hat{g}(x)$ based on an asymptotic expansion of $L_{NT}(\beta)$, then establish its asymptotic distribution and analyze the sources of asymptotic biases, and finally propose a consistent bias-corrected estimator.

3.3.1 Convergence rate for $\hat{g}(\cdot)$

To estimate the unknown function by the method of sieves, we assume that $g(x)$ is a smooth function. Let $\mathscr{X}\equiv\mathscr{Y}\times\mathscr{X}_1\subset\mathbb{R}^{d_y}\times\mathbb{R}^{d_x}$ be the support of X_{it}. Typical approximation and estimation of regression functions require that $\mathscr{X}$ be compact;

see Newey (1997). In our model, it seems restrictive to impose the compactness of $\mathscr{X}$ because of the presence of lagged dependent variables. To allow for the unboundedness of $\mathscr{X}$, 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

$$\|g\|_{\infty,\omega} \equiv \sup_{x\in\mathscr{X}} |g(x)| \left[1+\|x\|^2\right]^{-\omega/2} \text{ for some } \omega \geq 0. \tag{3.3.1}$$

If $\omega = 0$, the norm defined in (3.3.1) is the usual sup-norm which is suitable for the case of compact support.

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

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

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

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

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

Function $g(\cdot)$ *is said to be* $H(\gamma,\omega)$*-smooth on* $\mathscr{X}$ *if it belongs to a weighted Hölder ball* $\Lambda_c^\gamma(\mathscr{X},\omega)$ *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 $Q_{pp}^{(a)} \equiv E_{\mathscr{D}}[Q_{pp,NT}^{(a)}]$, where $a = (a_1, ..., a_K)'$ with $\|a\| = 1$, and $\mathscr{D} \equiv \sigma\left(f^0, \lambda^0\right)$ 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 $Q_{wpp} \equiv E_{\mathscr{D}}[Q_{wpp,NT}]$, where $w_{it} =$

$w(X_{it})$ and $w(\cdot)$ 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

$$Z_{it} \equiv p_{it} - \frac{1}{N}\sum_{j=1}^{N}\alpha_{ij}p_{jt} - \frac{1}{T}\sum_{s=1}^{T}\eta_{ts}p_{is} + \frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\alpha_{ij}\eta_{ts}p_{js}, \tag{3.3.2}$$

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

We first make some assumptions that are 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 $0 < \underline{c}_\lambda \le \mu_R(\Sigma_\lambda) \le \mu_1(\Sigma_\lambda) \le \overline{c}_\lambda < \infty$;

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

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

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

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

(iii) There exist positive constants $\underline{C}$ and $\overline{C}$ such that $\min_{\{a\in\mathbb{R}^K,\|a\|=1\}}\sum_{l=2R+1}^{N}\mu_l(Q_{pp,NT}^{(a)}) \ge \underline{C} > 0$ and $\mu_1(Q_{pp,NT}^{(a)}) = \left\|\mathbf{P}_{(a)}\right\|/\sqrt{NT} \le \overline{C} < \infty$ 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 $\mathscr{X}$ for some $\gamma > d/2$ and $\omega \ge 0$;

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

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

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

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

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) and 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} as $(N,T) \to \infty$. 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 $g(\cdot)$. Assumption $3(ii)$ quantifies the approximation error of functions in $H(\gamma,\omega)$ by the linear sieve basis functions $p^K(x)$. Assumption $3(iii)$ is used to deal with 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 $(NT)^{-1/2} \left\|\mathbf{e}_g\right\|_F = O_P\left(K^{-\gamma/d}\right)$; 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 (2010, Theorem 3.2). If X_{it} excludes lagged dependent variables, $E_{\mathscr{D}}[Z_i' e_i] = 0$ and then Assumption $3(vi)$ reduces to $(NT)^{-1/2} \sum_{i=1}^N Z_i' e_i = O_P(K^{1/2})$. 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

$$C_{NT,k}^{(1)} \equiv \frac{1}{NT}\mathrm{tr}\left(M_{\lambda^0}\mathbf{P}_k M_{f^0}\mathbf{u}'\right), \tag{3.3.3}$$

$$\begin{aligned} C_{NT,k}^{(2)} &\equiv -\frac{1}{NT}\mathrm{tr}\left(\mathbf{P}_k\Phi'\mathbf{u}M_{f^0}\mathbf{u}'M_{\lambda^0} + \mathbf{P}_k M_{f^0}\mathbf{u}'M_{\lambda^0}\mathbf{u}\Phi' + \mathbf{P}_k M_{f^0}\mathbf{u}'\Phi\mathbf{u}'M_{\lambda^0}\right) \\ &\equiv C_{NT,k}^{(2,a)} + C_{NT,k}^{(2,b)} + C_{NT,k}^{(2,c)}, \end{aligned} \tag{3.3.4}$$

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 3.3.1 *Suppose that Assumptions 1-4 hold. Then*

$$\hat{g}(x) - g(x) = p^K(x)' W_{NT}^{-1}\left(C_{NT}^{(1)} + C_{NT}^{(2)}\right) + \left[p^K(x)'\beta^0 - g(x)\right] + p^K(x)' R_{NT}, \tag{3.3.5}$$

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 $\mu_1\left[\int_{\mathscr{X}} p^K(x)p^K(x)'w(x)\,dx\right] < \infty$ and $\int_{\mathscr{X}}\left(1+\|x\|^2\right)^{\bar{\omega}} w(x)\,dx < \infty$. Then

$$\int_{\mathscr{X}} [\hat{g}(x) - g(x)]^2 w(x)\,dx = O_P\left(\frac{K}{NT} + K\delta_{NT}^{-4} + K^{-\frac{2\gamma}{d}}\right), \tag{3.3.6}$$

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[\hat{g}(X_{it}) - g(X_{it})]^2 w(X_{it}) = O_P\left(\frac{K}{NT} + K\delta_{NT}^{-4} + K^{-\frac{2\gamma}{d}}\right). \tag{3.3.7}$$

Remark 1. In (3.3.5), $\hat{g}(x) - g(x)$ 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 3.3.1 also states the convergence rates for both the weighted integrated mean square error (MSE) and weighted sample mean square error in (3.3.6) and (3.3.7), 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 (3.3.5), respectively.[1] 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

[1] Apparently, $K/(NT) + K\delta_{NT}^{-4} = O\left(K\delta_{NT}^{-4}\right)$. We keep the first term in the expression as it corresponds to the usual variance term for a sieve estimate.

$\mathscr{D}$, then the rates in (3.3.6) and (3.3.7) should be $O_P\left(K^{-2\gamma/d}+K/(NT)\right)$, and K_{opt} would be proportional to $(NT)^{1/[(2\gamma/d)+1]}$.

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

$$|P_{\mathscr{B}}(A\cap B)-P_{\mathscr{B}}(A)P_{\mathscr{B}}(B)|\leq\alpha_{\mathscr{B}}(t)\ \text{a.s.} \tag{3.3.8}$$

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); see also Roussas (2008). When one takes $\alpha_{\mathscr{B}}(t)$ as the supremum of the left hand side object in (3.3.8) 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 it to the $\mathscr{B}$-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 $\tilde{Z}_i\equiv(\tilde{Z}_{i1}',\cdots,\tilde{Z}_{iT}')'=P_i-P_{f^0}E_{\mathscr{D}}(P_i)-N^{-1}\sum_{j=1}^{N}\alpha_{ij}M_{f^0}E_{\mathscr{D}}(P_j)$, $\tilde{Z}_{it}\equiv p_{it}-N^{-1}\sum_{i=1}^{N}\alpha_{ij}E_{\mathscr{D}}(p_{jt})-T^{-1}\sum_{s=1}^{T}\eta_{ts}E_{\mathscr{D}}(p_{is})+(NT)^{-1}\sum_{i=1}^{N}\sum_{s=1}^{T}\alpha_{ij}\eta_{ts}E_{\mathscr{D}}(p_{js})$. Let $\tilde{W}\equiv E_{\mathscr{D}}(\tilde{W}_{NT})$ and $\tilde{\Omega}\equiv E_{\mathscr{D}}(\tilde{\Omega}_{NT})$. We add the following assumptions.

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

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

$\lambda^0, f^0\}$;

(iii) $(e_{it}, X_{it}) \perp (e_{js}, X_{js}) \,|\mathscr{D}$ 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) $\sup_{i,t} E\left|e_{it}\right|^{8+4\delta} < \infty$;

(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) $\sup_k \sup_{i,t} E\left|p_{it,k}\right|^{8+4\delta} < \infty$ 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 $\mathscr{D}$. 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.[2] Assumption 5(ii) imposes a martingale difference sequence (m.d.s.) condition on $\left\{(e_{it}, X_{it}), \mathscr{F}_0^t\right\}_{t=1}^{T}$. Assumption 5$(iii)$ imposes the conditional independence between (e_{it}, X_{it}) and $\left(e_{js}, X_{js}\right)$ for $i \neq j$ given $\mathscr{D}$. 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 $\left\{(X_{1,it}, e_{it})\right\}_{i=1}^{N}$ conditional on $\mathscr{D}$ 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 $\left\|f_t^0\right\|$ and $\left\|\lambda_i^0\right\|$ 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 (2010) where a uniform bound over a truncated support is used. In the case of compact support, it is generally assumed that $\sup_{x\in\mathscr{X}} \left\|p^K(x)\right\| = O_P(\zeta(K))$ 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 $\sup_{i,t} E\left\|p_{it}\right\| = O_P\left(K^{1/2}\right)$, 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

[2]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.

relax this assumption to $\sup_k \sup_{i,t} E\left|p_{it,k}\right|^{8+4\delta} < \zeta_0(K)$ 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., $K \in (N^{d/\gamma}, N^{1/2})$.

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

$$\begin{aligned} b_{1,k} &\equiv \frac{1}{N}tr\left[P_{f^0}E_{\mathscr{D}}\left(\mathbf{e}'\mathbf{P}_k\right)\right], b_{2,k} \equiv \frac{1}{T}tr\left[E_{\mathscr{D}}\left(\mathbf{e}\mathbf{e}'\right)M_{\lambda^0}\mathbf{P}_k\Phi\right], \text{and} \\ b_{3,k} &\equiv \frac{1}{N}tr\left[E_{\mathscr{D}}\left(\mathbf{e}'\mathbf{e}\right)M_{f^0}\mathbf{P}_k'\Phi'\right]. \end{aligned}$$

Define

$$\begin{aligned} B_K(x) &\equiv -A_{NT}p^K(x)'\tilde{W}^{-1}\left(T^{-1}b_1 + N^{-1}b_2 + T^{-1}b_3\right) \\ &\equiv -\kappa_{NT}b_1(x) - \kappa_{NT}^{-1}b_2(x) - \kappa_{NT}b_3(x), \end{aligned} \tag{3.3.9}$$

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 3.3.2 *Suppose that Assumptions 1-8 hold. Then*

$$A_{NT}\left[\hat{g}(x) - g(x)\right] - B_K(x) \xrightarrow{d} N(0,1)$$

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. Let $a_{NT} \equiv A_{NT}p^K(x)'W_{NT}^{-1}$. Theorem 3.3.1 suggests that the leading terms in the expansion of $A_{NT}\left[\hat{g}(x) - g(x)\right]$ 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 term $(-\kappa_{NT}b_1(x))$. The latter also arises in linear dynamic panel data models and is caused by the endogeneity of Z_{it} defined in (3.3.2):

$$E_{\mathscr{D}}(Z_{it}e_{it}) = -\frac{1}{T}\sum_{s=t+1}^{T}\left(1-\frac{1}{N}\alpha_{ii}\right)\eta_{ts}E_{\mathscr{D}}(p_{is}e_{it}) \neq 0$$

by Assumption $5(iii)$. It is easy to see that an equivalent expression for b_1 is

$$b_1 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\eta_{ts}E_{\mathscr{D}}(p_{is}e_{it}). \tag{3.3.10}$$

$a_{NT}C_{NT}^{(2,a)}$ contributes to the second bias term, i.e., $-\kappa_{NT}^{-1}b_2(x)$, and is caused by cross-sectional heteroskedasticity of errors conditional on $\mathscr{D}$; $a_{NT}C_{NT}^{(2,b)}$ contributes to the third bias term, i.e., $-\kappa_{NT}b_3(x)$, and is caused by serial correlation and heteroskedasticity of errors conditional on $\mathscr{D}$. In the special case where e_{it}'s are IID conditional on $\mathscr{D}$ across both i and t, the last two bias terms disappear.

3.3.3 Bias correction

In this section, we propose a bias-corrected estimator for $g(x)$. 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 $A_{ij}\mathbf{1}(i=j)$ with $\mathbf{1}(\cdot)$ being the usual indicator function. Let $\Gamma(\cdot)$ be the truncation kernel: $\Gamma(s) = \mathbf{1}(|s|\leq 1)$. 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):

$$f'f/T = I_R \text{ and } \lambda'\lambda = \text{diagonal matrix}. \tag{3.3.11}$$

Given $\hat{\beta}$, we can obtain $(\hat{\lambda},\hat{f})$ as the solution to the following set of nonlinear equations:

$$\left[\frac{1}{NT}\sum_{i=1}^{N}\left(Y_i - P_i\hat{\beta}\right)\left(Y_i - P_i\hat{\beta}\right)'\right]\hat{f} = \hat{f}\mathbf{V}_{NT}, \tag{3.3.12}$$

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

$$\hat{\lambda} \equiv \left(\hat{\lambda}_1, \cdots, \hat{\lambda}_N\right)' = T^{-1}\left[\hat{f}'\left(Y_1 - P_1\hat{\beta}\right), \cdots, \hat{f}'\left(Y_N - P_N\hat{\beta}\right)\right]'. \tag{3.3.13}$$

The projection matrices P_{f^0} and P_{λ^0} can be estimated as follows

$$P_{\hat{f}} \equiv \hat{f}\hat{f}'/T \text{ and } P_{\hat{\lambda}} \equiv \hat{\lambda}(\hat{\lambda}'\hat{\lambda})^{-1}\hat{\lambda}'. \tag{3.3.14}$$

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. The residuals are given by

$$\hat{e}_{it} \equiv Y_{it} - \hat{g}\left(X_{it}\right) - \hat{\lambda}_i'\hat{f}_t. \tag{3.3.15}$$

Let $\hat{\alpha}_{ij} \equiv \hat{\lambda}_i'(\hat{\lambda}'\hat{\lambda}/N)^{-1}\hat{\lambda}_j$, $\hat{\eta}_{ts} \equiv \hat{f}_t'(\hat{f}'\hat{f}/T)^{-1}\hat{f}_s$, and $\hat{Z}_{it} \equiv p_{it} - \frac{1}{N}\sum_{j=1}^{N}\hat{\alpha}_{ij}p_{jt} - \frac{1}{T}\sum_{s=1}^{T}\hat{\eta}_{ts}p_{is} + \frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\hat{\alpha}_{ij}\hat{\eta}_{ts}p_{js}$. Then we can 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

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

Let

$$\begin{aligned}
\hat{B}_K\left(x\right) &= -\hat{A}_{NT}p^K\left(x\right)'\hat{W}_{NT}^{-1}(T^{-1}\hat{b}_1 + N^{-1}\hat{b}_2 + T^{-1}\hat{b}_3) \\
&\equiv -\kappa_{NT}\hat{b}_1\left(x\right) - \kappa_{NT}^{-1}\hat{b}_2\left(x\right) - \kappa_{NT}\hat{b}_3\left(x\right)
\end{aligned}$$

and

$$\hat{\beta}_{bc} \equiv \hat{\beta} + \hat{W}_{NT}^{-1}(T^{-1}\hat{b}_1 + N^{-1}\hat{b}_2 + T^{-1}\hat{b}_3). \tag{3.3.16}$$

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

$$\hat{g}_{bc}(x) \equiv p^K(x)'\hat{\beta}_{bc} = \hat{g}(x) - \hat{A}_{NT}^{-1}\hat{B}_K(x). \tag{3.3.17}$$

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 $\max\{M_T/T, \sqrt{\frac{NK}{T}}\sum_{\tau=M_T}^{\infty} \alpha_{\mathscr{D}}^{\frac{3+2\delta}{4+2\delta}}(\tau), M_T\sqrt{\frac{NK}{T}}\delta_{NT}^{-1}\} \to 0$;

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

3.4 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.

3.4.1 The hypothesis and test statistic

For the model in (3.1.1), we are interested in testing the null hypothesis:

$$\mathbb{H}_0 : \Pr\left[g\left(X_{it}\right) = X_{it}'\theta^0\right] = 1 \text{ for some } \theta^0 \in \Theta, \tag{3.4.1}$$

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

$$\mathbb{H}_1 : \Pr\left[g\left(X_{it}\right) = X_{it}'\theta\right] < 1 \text{ for all } \theta \in \Theta. \tag{3.4.2}$$

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

$$\mathbb{H}_1\left(\gamma_{NT}\right) : g\left(X_{it}\right) = X_{it}'\theta^0 + \gamma_{NT}\Delta\left(X_{it}\right) \tag{3.4.3}$$

where $\Delta(\cdot) \equiv \Delta_{NT}(\cdot)$

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

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. 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.

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

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

$$D_{NT,k_1k_2} \equiv \frac{1}{NT}\mathrm{tr}\left(M_{\lambda^0}\mathbf{X}_{k_1}M_{f^0}\mathbf{X}_{k_2}'\right). \tag{3.4.4}$$

Let $D \equiv E_{\mathscr{D}}[D_{NT}]$. Let Υ_{NT} be $d \times 1$ vectors whose kth element is given by $\Upsilon_{NT,k} \equiv \frac{1}{NT}\mathrm{tr}\left(M_{\lambda^0}\mathbf{X}_k M_{f^0}\Delta'\right)$. 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 $\|\Delta(\cdot) - p^K(\cdot)'\beta_\Delta^0\|_{\infty,\bar{\omega}} = O\left(K^{-\gamma/d}\right)$.

Assumption 11. (i) $0 < \underline{C}_Q \leq \mu_d(Q_{wxx}) \leq \mu_1(Q_{wxx}) \leq \bar{C}_Q < \infty$ a.s. as $(N,T) \to 0$;

(ii) $\|Q_{wpx}\| \leq C_Q < \infty$ a.s. for all K as $(N,T) \to 0$;

(iii) $0 < \underline{C}_D \leq \mu_d(D) \leq \mu_1(D) \leq \bar{C}_D < \infty$ 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 $(N,T) \to \infty$, $K^3/N \to 0$, $\max\left(\kappa_{NT}, \kappa_{NT}^{-1}\right)K^{-1/4} \to 0$,

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

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 $\|Q_{wpx}\|^2 = \mu_1(Q_{wpx}Q'_{wpx}) \le \mu_1(Q_{wpp})\mu_1(Q_{wxx}) < \infty$. Assumption 12 imposes some requirements on (N, T, K, M_T), which are much weaker than that for the bias-correction of sieve estimator. Note that the case where $N/T = c \in [0, \infty]$ is allowed. If we restrict $c \in (0, \infty)$, Assumption 12 reduces to $K^{1/4} \max\{\sum_{\tau=M_T}^{\infty} \alpha_{\mathscr{D}}^{\frac{3+2\delta}{4+2\delta}}(\tau), \frac{M_T}{\sqrt{N}}\} \to 0$ and $K^3/N \to 0$, $K \in (N^{\gamma_1}, N^{1/3})$, where $\gamma_1 \equiv \max\{\frac{1/2}{\gamma/d-3/2}, \frac{5/8}{\gamma/d-1/4}\}$. The requirement on But it is still necessary to use bias-corrected sieve estimate in specification testing.

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

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

where $A^{\Delta} \equiv p\lim_{(N,T)\to\infty} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\Delta_{it} - X'_{it}D_{NT}^{-1}\Upsilon_{NT}\right)^2 w_{it}$ 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) central limit theorem (CLT) for independent but non-identically distributed (INID) observations. As Su, Jin, and Zhang (2012) 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 $\hat{\mathbb{B}}_{NT}\equiv\mathrm{tr}(\hat{W}_{NT}^{-1}\times Q_{wpp,NT}\hat{W}_{NT}^{-1}\hat{\Omega}_{NT})$ and $\hat{\mathbb{V}}_{NT}\equiv 2\mathrm{tr}(\hat{W}_{NT}^{-1}Q_{wpp,NT}\hat{W}_{NT}^{-1}\hat{\Omega}_{NT}\hat{W}_{NT}^{-1}Q_{wpp,NT}\hat{W}_{NT}^{-1}\hat{\Omega}_{NT})$, respectively, where $\hat{W}_{NT}\equiv\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{Z}_{it}\hat{Z}_{it}'$ and $\hat{\Omega}_{NT}\equiv\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{Z}_{it}\hat{Z}_{it}'\hat{e}_{it}^2$. Then we define a feasible test statistic:

$$\hat{J}_{NT}\equiv\left(NT\Gamma_{NT}-\hat{\mathbb{B}}_{NT}\right)/\sqrt{\hat{\mathbb{V}}_{NT}}. \tag{3.4.5}$$

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

Theorem 3.4.2 *Suppose that Assumptions 1-8 and 10-12 hold. Under* $\mathbb{H}_1(\gamma_{NT})$ *with* $\gamma_{NT}=(NT)^{-1/2}\times\mathbb{V}_{NT}^{1/4}$, $\hat{J}_{NT}\xrightarrow{d} N\left(A^{\Delta},1\right)$.

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 $\Pr\left(\hat{J}_{NT}>z|\mathbb{H}_1(\gamma_{NT})\right)\to 1-\Phi\left(z-A^{\Delta}\right)$ 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 3.4.3 *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 critical values or p-values to conduct an asymptotic test.

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 $\bar{\Delta}_i \equiv \left[\bar{\Delta}(X_{i1}), \cdots, \bar{\Delta}(X_{iT})\right]'$ for $i = 1,...,N$ and $\bar{\Delta} \equiv \left(\bar{\Delta}_1, \cdots, \bar{\Delta}_N\right)'$. With an additional assumption $\left\|\bar{\Delta}\right\| = o_P[(NT)^{1/2}]$,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 whose kth element is given by $\bar{\Upsilon}_{NT,k} \equiv (NT)^{-1}\text{tr}(M_{\lambda^0}\mathbf{X}_k M_{f^0}\bar{\Delta}')$. 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 any 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 $g(X_{it}) = g_1(X_{1,it}) + \theta_2^{0\prime}X_{2,it}$, $X_{it} = \left(X_{1,it}', X_{2,it}'\right)'$, 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 non-parametric 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 $f_{a,t}=1$, $\vec{f}_t^0=\left(1,f_t^{0\prime}\right)'$, and $\vec{\lambda}_i^0=\left(\lambda_{a,i},\lambda_i^{0\prime}\right)'$. 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 factor 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.

3.4.3 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 it is worthwhile to propose a bootstrap procedure to improve the finite sample performance of our test. As Neumann and Paparoditis (2000) note, it is not necessary to reproduce the whole dependence structure of the original data to get a correct estimator of the null distribution of the testing statistic. In the spirit of Hansen (2000), we propose a fixed-regressor wild bootstrap method. The procedure 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 $\hat{g}_{bc}(X_{it}) = \hat{\beta}_{bc}' p^K(X_{it})$, $\hat{\theta}' X_{it}$, $\hat{\lambda}_i$, $\hat{f}_t$, and $\hat{e}_{it}$.

2. For $i = 1, ..., N$, obtain the wild bootstrap errors $\{e_{it}^*\}_{t=1}^T$ as follows: $e_{it}^* = v_{it}\hat{e}_{it}^{(l)}$ where v_{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: $Y_{it}^* = X_{it}'\hat{\theta} + \hat{\lambda}_i^{(l)\prime}\hat{f}_t^{(l)} + e_{it}^*$ 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 $\hat{g}_{bc}^*(X_{it})$, $X_{it}'\hat{\theta}^*$, $\hat{f}_t^*$, $\hat{\lambda}_i^*$, and $\hat{e}_{it}^*$.

4. Repeat Steps 2-3 for B times and index the bootstrap statistics as $\left\{J_{NT,b}^*\right\}_{b=1}^B$. 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 3.4.4 *Suppose that the conditions in Theorem 3.4.2 hold. Then* $\hat{J}_{NT}^* \xrightarrow{d^*} N(0,1)$ *in probability, where* $\xrightarrow{d^*}$ *denotes weak convergence under the bootstrap probability measure conditional on the observed sample* $\mathscr{W}_{NT} \equiv \{(X_{it}, Y_{it}) : i = 1, ..., N, t = 1, ..., T\}$.

The above result holds no matter whether the original sample satisfies the null, the local alternative, or the global alternative hypotheses. If $\mathbb{H}_0$ holds, $\hat{J}_{NT}$ converges in distribution to $N(0,1)$ so that a test based on the bootstrap p-value will have the correct asymptotic level. If $\mathbb{H}_1$ holds for the original sample, $\hat{J}_{NT}$ diverges at $NT/\sqrt{K}$ whereas $\hat{J}_{NT}^{*}$ still converges to $N(0,1)$ with some additional assumptions, which implies that the bootstrap test is consistent.

3.5 Monte Carlo simulations

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

3.5.1 Data generating processes

We consider the following data generating processes (DGPs):

DGP 1: $Y_{it} = \frac{1}{2}Y_{i,t-1} + \lambda_i^{0\prime} f_t^0 + e_{it}$,

DGP 2: $Y_{it} = \frac{1}{2}Y_{i,t-1} + X_{1,it} + \lambda_i^{0\prime} f_t^0 + e_{it}$,

DGP 3: $Y_{it} = \frac{1}{2}Y_{i,t-1} + \frac{1}{2}[\frac{\exp(Y_{i,t-1}-Y_{i,t-1}^2)}{1+\exp(Y_{i,t-1}-Y_{i,t-1}^2)} - \frac{1}{2}] + \lambda_i^{0\prime} f_t^0 + e_{it}$,

DGP 4: $Y_{it} = \frac{1}{2}Y_{i,t-1} + \frac{1}{2}[\Phi\left(Y_{i,t-1} - Y_{i,t-1}^2\right) - \frac{1}{2}] + \lambda_i^{0\prime} f_t^0 + e_{it}$,

DGP 5: $Y_{it} = \frac{1}{2}Y_{i,t-1} + \frac{1}{4}[\phi(Y_{i,t-1}) - \frac{1}{\sqrt{2\pi}}] + \frac{1}{2}[\phi(X_{1,it}) - \frac{1}{\sqrt{2\pi}}] + \lambda_i^{0\prime} f_t^0 + e_{it}$,

DGP 6: $Y_{it} = \frac{1}{2}Y_{i,t-1} + \frac{1}{4}X_{1,it}[\Phi(Y_{i,t-1}) - \frac{1}{2}] + \frac{1}{2}[\phi(X_{1,it}) - \frac{1}{\sqrt{2\pi}}] + \lambda_i^{0\prime} f_t^0 + e_{it}$,

where $\lambda_i^0 = (\lambda_{i1}^0, \lambda_{i2}^0)'$, $f_t^0 = (f_{t1}^0, f_{t2}^0)'$, $i = 1, \ldots, N$, $t = 1, \ldots, T$, $\Phi(\cdot)$ 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_{tt}^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 $e_{it}=\sqrt{h_{it}}\varepsilon_{it}$, $h_{it}=0.1+0.2Y_{i,t-1}^2$, and $\varepsilon_{it}\sim$IID $N(0,1)$. For DGPs 2h, 5h-6h, the errors are generated according to $e_{it}=\sqrt{h_{it}}\varepsilon_{it}$, $h_{it}=0.1+0.1Y_{i,t-1}^2+0.1X_{1,it}^2$, and $\varepsilon_{it}\sim$IID $N(0,1)$.

3.5.2 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 $\left\{f_t^0\right\}_{t=1}^T$ 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 $M_T=\left\lfloor T^{1/7}\right\rfloor$. 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

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

where $\alpha_{j,n}(x) = \frac{x-v_j}{v_{j+n}-v_j}\mathbf{1}\left(v_{j+n} \neq v_j\right)$ and $\left\{v_j\right\}_{j=0}^{J+1}$ 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 $\left\{b_{j,n}(x)\right\}_{j=0}^{J+m-1}$ 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 ($n = 3$)

$$p_Y^{J+4}(y) = \left[b_{0,3}^{(Y)}(y), b_{1,3}^{(Y)}(y), \cdots, b_{J+3,3}^{(Y)}(y)\right]', \tag{3.5.1}$$

where the superscript "(Y)" denotes its correspondence to $\left\{Y_{i,t-1}\right\}$. The knots $\left\{v_{y,j}\right\}_{j=0}^{J+1}$ 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

$$p^K(y,x) = [p_Y^{J+4}(y)', p_X^{J+3}(x)']' \tag{3.5.2}$$

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

$$p^K(y,x) = [p_Y^{J+4}(y) \otimes p_X^{J+4}(x)]', \tag{3.5.3}$$

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]

[3]Alternatively one can follow, e.g., Lee (2010), to use the leave-one-out cross-validation (CV) to

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: $\text{RMSE}(\hat{g}) = \sqrt{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\hat{g}(X_{it}) - g(X_{it})\right]^2 a(X_{it})}$, 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 3.1-3.2 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 3.3 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 sample 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 cor-

choose K adaptively. Another possibility is to apply the Lasso-type techniques to achieve simultaneous variable selection and estimation; see, e.g., Tibshirani (1996) and Fan and Li (2001). We leave these as a future research topic.

rectly 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.

3.5.3 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 $w(X_{it}) = \mathbf{1}(X_{it} \in \mathscr{U})$ where $\mathscr{U}$ 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 3.4-3.5 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 increase very fast as either N or T increases, and it not very sensitive to the choice of J.

Table 3.1: 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 3.2: ARMSE comparison for DGPs 1h-6h: heteroskedastic errors

			C=1			C=1.5			C=2			Linear		
DGP	N	T	$\hat{g}$	$\hat{g}_{bc}$	$\hat{g}_{IF}$	$\hat{g}$	$\hat{g}_{bc}$	$\hat{g}_{IF}$	$\hat{g}$	$\hat{g}_{bc}$	$\hat{g}_{IF}$	$\hat{g}^{(l)}$	$\hat{g}^{(l)}_{bc}$	$\hat{g}^{(l)}_{IF}$
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 3.3: 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	$\hat{g}$	$\hat{g}_{bc}$	$\hat{g}_{IF}$	$\hat{g}$	$\hat{g}_{bc}$	$\hat{g}_{IF}$	$\hat{g}$	$\hat{g}_{bc}$	$\hat{g}_{IF}$	$\hat{g}^{(l)}$	$\hat{g}_{bc}^{(l)}$	$\hat{g}_{IF}^{(l)}$
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.

Table 3.4: Rejection frequency for DGPs 1-6

			C=1			C=1.5			C=2		
DGP	N	T	1%	5%	10%	1%	5%	10%	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 3.5: 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.

3.6 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

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

where $Y_{it} \equiv \log\left(GDP_{it}\right) - \log\left(GDP_{i,t-1}\right)$, 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 $g\left(\cdot\right)$ 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 same set of countries as Bond, Leblebicioglu, and Schiantarelli (2010) but exclude Guyana which does not have observations for the period 1960-1970 for the investment as a share of real GDP. 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 before one tries to minimize any of the above information criteria. 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 $M_T=\left\lfloor T^{1/7.5}\right\rfloor$ 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 3.1 and 3.2, respectively. Figure 3.1 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, Figure 3.1(a)-(c) reports 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 3.2 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 along, 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 3.1(a)-(c) and 3.2(a)-(c) suggest that lagged economic growth rate is globally positively related to the current economic growth rate when investment share and its growth

Table 3.6: 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

are fixed at their sample medians. Third, Figures 3.1(d)-(f) and 3.2(d)-(f) suggest that investment share generally has positive effect on the economic growth rate. Fourth, Figures 3.1(g)-(i) and 3.2(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 3.6 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.

3.7 Conclusion

In this chapter 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 al-

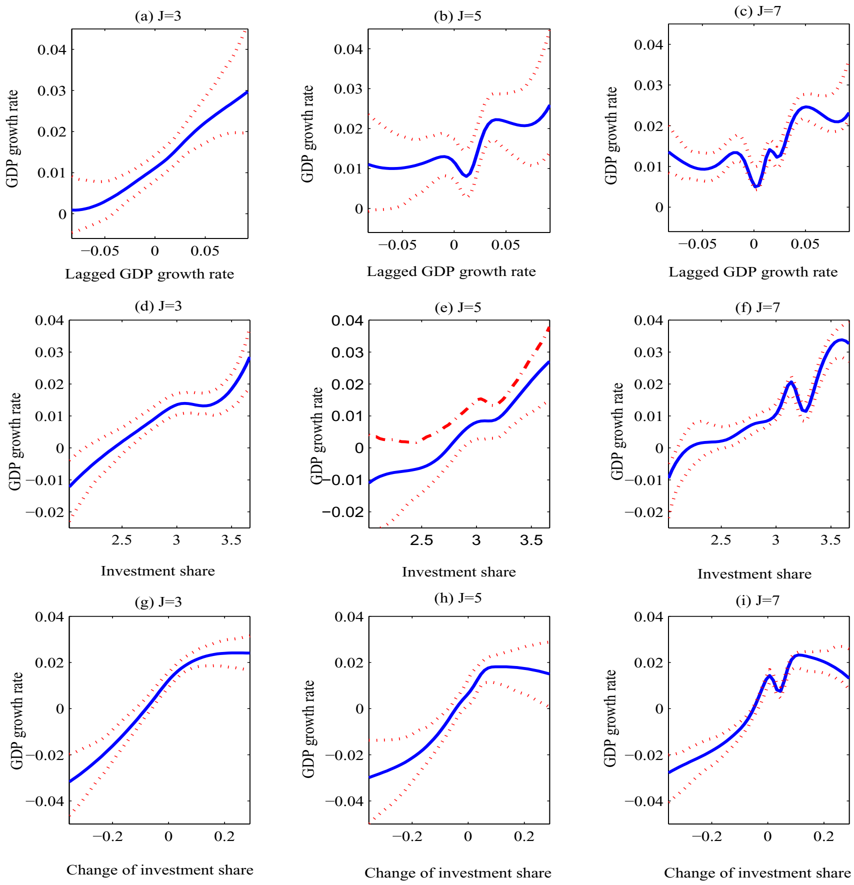

Figure 3.1: Relationship between GDP growth rate and lagged GDP growth rate, investment share, and change of investment share(1960-1985) (solid line: estimated function, dotted lines: 90% bootstrap confidence band)

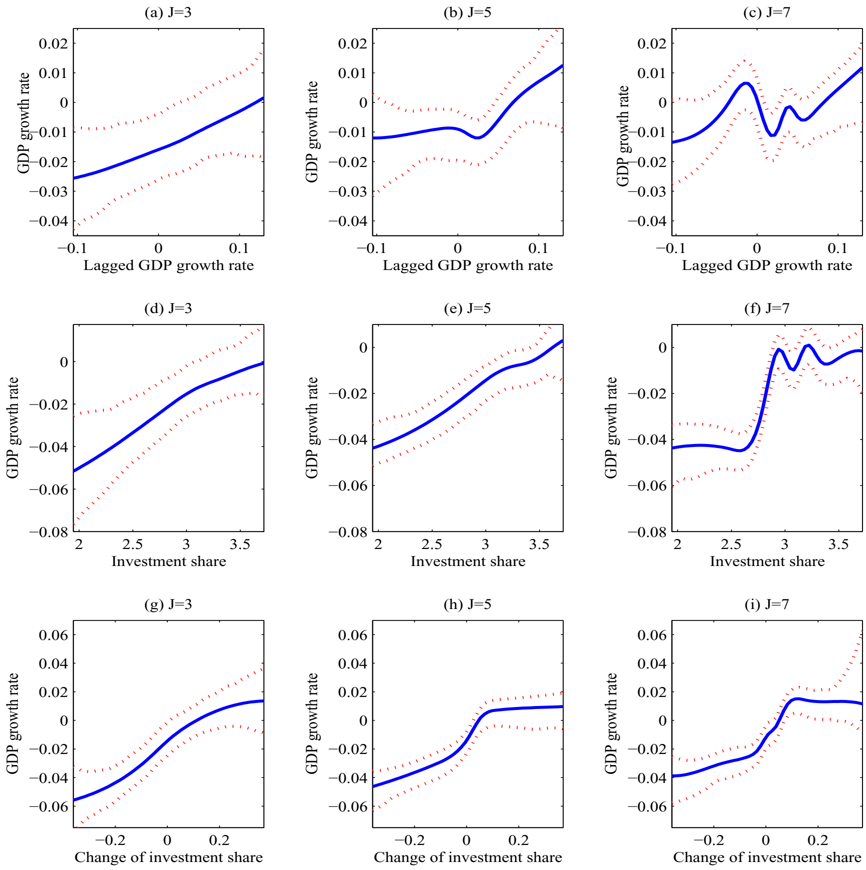

Figure 3.2: Relationship between GDP growth rate and lagged GDP growth rate, investment share, and change of investment share(1986-2010) (solid line: estimated function, dotted lines: 90% bootstrap confidence band)

ternatives. 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.

Chapter 4 Testing for Common Trends in Semiparametric Panel Data Models with Fixed Effects

4.1 Introduction

Modeling trends in time series has a long history. Phillips (2001, 2005, 2010) provides recent overviews covering the development, challenges, and some future directions of trend modeling in time series. White and Granger (2011) offer working definitions of various kinds of trends and invite more discussion on trends in order to facilitate development of increasingly better methods for prediction, estimation and hypothesis testing for non-stationary time-series data. Due to the wide availability of panel data in recent years, research on trend modeling has spread to the panel data models. Most of the literature falls into two categories depending on whether the trends are stochastic or deterministic. But there is also work on evaporating trends (Phillips, 2007) and econometric convergence testing (Phillips and Sul, 2007, 2009). For reviews on stochastic trends in panel data models, see Banerjee (1999) and Breitung and Pesaran (2005).

Recently, some aspects of modeling deterministic time trends in nonparametric and semiparametric settings have attracted interest. Cai (2007) studies a time-varying coefficient time series model with a time trend function and serially correlated errors to characterize the nonlinearity, nonstationarity, and trending phenomenon. Robinson (2010) considers nonparametric trending regression in panel data models with cross-sectional dependence. Atak, Linton, and Xiao (2011) propose a semiparametric panel data model to model climate change in the United

Kingdom (UK hereafter), where seasonal dummies enter the model linearly with heterogeneous coefficients and the time trend enters nonparametrically. Li, Chen, and Gao (2010) extend the work of Cai (2007) to panel data time-varying coefficient models. Most recently, Chen, Gao, and Li (2010, CGL hereafter) extend Robinson's (2010) nonparametric trending panel data models to semiparametric partially linear panel data models with cross-sectional dependence where all individual unit share a common time trend that enters the model nonparametrically. They propose a semiparametric profile likelihood approach to estimate the model.

A conventional feature of work on deterministic trending panel models is the imposition of a common trends assumption, implying that each individual unit follows the same time trend behavior. Such an assumption greatly simplifies the estimation and inference process, and the proposed estimators can be efficient if there is no heterogeneity in individual time trend functions and some other conditions are met. Nevertheless, if the common trends assumption does not stand, the estimates based on nonparametric or semiparametric panel data models with common trends will be generally inefficient and statistical inference will be misleading. It is therefore prudent to test for the common trends assumption before imposing it.

Since Stock and Watson (1988) there has been a large literature on testing for common trends. But to our knowledge, most empirical works have focused on testing for common stochastic trends. Tests for common deterministic trends are far and few between. Vogelsang and Franses (2005) propose tests for common deterministic trend slopes by assuming linear trend functions and a stationary variance process and examining whether two or more trend-stationary time series have the same slopes. Xu (2011) considers tests for multivariate deterministic trend coefficients in the case of nonstationary variance process. Sun (2011) develops a novel testing procedure for hypotheses on deterministic trends in a multivariate trend stationary model where the long run variance is estimated by series method. In all cases, the models are parametric and the asymptotic theory is established by passing the time series dimension T to infinity and keeping the number of cross sectional units n fixed. Empirical applications include Fomby and Vogelsang (2003) and Bacigál (2005), who apply the Vogelsang-Franses test to temperature data and geodetic data, respectively.

This chapter develops a test for common trends in a semiparametric panel data model of the form

$$Y_{it} = \beta' X_{it} + f_i(t/T) + \alpha_i + \varepsilon_{it},\ i = 1, \ldots, n,\ t = 1, \ldots, T, \tag{4.1.1}$$

where β is a $d \times 1$ vector of unknown parameters, X_{it} is a $d \times 1$ vector of regressors, f_i is an unknown smooth time trend function for cross section unit i, the α_i's represent fixed effects that can be correlated with X_{it}, and ε_{it}'s are idiosyncratic errors. The trend functions $f_i(t/T)$ that appear in (4.1.1) provide for idiosyncratic trends for each individual i. For simplicity, we will assume that (i) $\{\varepsilon_{it}\}$ satisfies certain martingale difference conditions along the time dimension but may be correlated across individuals, and (ii) $\{\varepsilon_{it}\}$ are independent of $\{X_{it}\}$. Note that f_i and α_i are not identified in (4.1.1) without further restrictions.

Model (4.1.1) covers and extends some existing models. First, when $f_i \equiv 0$ for all i, (4.1.1) becomes the traditional panel data model with fixed effects. Second, if $n = 1$, then model (4.1.1) reduces to the model discussed in Gao and Hawthorne (2006). Third, when $f_i = f$ for some unknown smooth function f and all i, (4.1.1) becomes the semiparametric trending panel data model of CGL (2010).

The main objective of this chapter is to construct a nonparametric test for common trends. Under the null hypothesis of common trends: $f_i = f$ for all i in (4.1.1), we can pool the observations from both cross section and time dimensions to estimate both the finite dimensional parameter (β) and the infinite dimensional parameter (f) under the single identification restriction $\sum_{i=1}^{n} \alpha_i = 0$ or $f(0) = 0$, whichever is convenient. Let $u_{it} \equiv \alpha_i + \varepsilon_{it}$. Let $\widehat{u}_{it}$ denote the estimate of u_{it} based on the pooled regression. The residuals $\{\widehat{u}_{it}\}$ should not contain any useful trending information in the data. This motivates us to construct a residual-based test for the null hypothesis of common trends. To be concrete, we will propose a test for common trends by averaging the n measures of nonparametric goodness-of-fit $\left(R^2\right)$ from the nonparametric time series regression of $\widehat{u}_{it}$ on the time trend for each cross sectional unit i. Such nonparametric R^2 should tend to zero under the null hypothesis of common trends and diverge from zero otherwise. We show that after being properly centered and scaled, the average nonparametric R^2 is asymptotically normally distributed under the null hypothesis of common trends and a sequence of Pitman local alternatives. We also establish the consistency of the test and propose a bootstrap method to obtain the bootstrap p-values.[1]

To proceed, it is worth mentioning that (4.1.1) complements the model of Atak, Linton, and Xiao (2011) who allow for heterogenous slopes but a single nonparametric common trend across cross sections. As mentioned in the concluding remarks, it is also possible to allow the slope coefficients in (4.1.1) to vary across individuals and consider a joint test for the homogeneity of the slope coefficients and trend components. But this is beyond the scope of the current chapter.

The rest of the chapter is organized as follows. The hypotheses and the test statistic are given in Section 2. We study the asymptotic distributions of the test under the null and a sequence of local alternatives, establish the consistency of the test, and propose a bootstrap procedure to obtain the bootstrap p-values in Section 3. Section 4 conducts a small simulation experiment to evaluate the finite sample performance of our test and reports empirical applications of the test to UK climate change data and OECD economic growth data. Section 5 concludes.

NOTATION. Throughout the chapter we adopt the following notation. For a matrix A, its transpose is A' and Euclidean norm is $\|A\| \equiv [\operatorname{tr}(AA')]^{1/2}$, where $\equiv$ signifies "is defined as". When A is a symmetric matrix, we use $\lambda_{\max}(A)$ to denote its maximum eigenvalue. For a natural number l, we use i_l and I_l to denote the $l \times 1$ vector of ones and the $l \times l$ identity matrix, respectively. For a function f

[1] To the best of our knowledge, Su and Ullah (2011) are the first to suggest applying such a measure of nonparametric R^2 to conduct model specification test based on residuals from restricted parametric, nonparametric, or semiparametric regressions, and apply this idea to test for conditional heteroskedasticity of unknown form. Clearly, the nonparametric R^2 statistic can serve as a useful tool for testing many popular hypotheses in econometrics and statistics by playing a role comparable to the important role that R^2 plays in the parametric setup.

defined on the real line, we use $f^{(a)}$ to denote its a'th derivative whenever it is well defined. The operator $\xrightarrow{p}$ denotes convergence in probability, and $\xrightarrow{d}$ convergence in distribution. We use $(n,T) \to \infty$ to denote the joint convergence of n and T when n and T pass to the infinity simultaneously.

4.2 Basic Framework

In this section, we state the null and alternative hypotheses, introduce the estimation of the restricted model under the null, and then propose a test statistic based on the average of nonparametric goodness-of-fit measures.

4.2.1 Hypotheses

The main objective is to construct a test for common trends in model (4.1.1). We are interested in the null hypothesis that

$$H_0 : f_i(\tau) = f(\tau) \text{ for } \tau \in [0,1] \text{ and some smooth function } f,\ i = 1,\ldots,n, \quad (4.2.1)$$

i.e., all the n cross sectional units share the common trends function f. The alternative hypothesis is

$$H_1 : \text{the negation of } H_0.$$

As mentioned in the introduction, we will propose a residual-based test for the above null hypothesis. To do so, we need to estimate the model under the null hypothesis and obtain the augmented residual, which estimates $\alpha_i + \varepsilon_{it}$. Then for each i, we run the local linear regression of the augmented residuals on t/T, and calculate the nonparametric R^2. Our test statistics is constructed by averaging these n nonparametric R^2's.

4.2.2 Estimation under the null

To proceed, we introduce the following notation.

$$\begin{aligned}
Y_i &\equiv (Y_{i1},\ldots,Y_{iT})',\ Y \equiv \left(Y_1',\ldots,Y_n'\right)',\ X_i \equiv (X_{i1},\ldots,X_{iT})',\ X \equiv \left(X_1',\ldots,X_n'\right)',\\
\varepsilon_i &\equiv (\varepsilon_{i1},\ldots,\varepsilon_{iT})',\ \varepsilon \equiv \left(\varepsilon_1',\ldots,\varepsilon_n'\right)',\ \alpha \equiv (\alpha_2,\ldots,\alpha_n)',\ D \equiv (-i_{n-1},I_{n-1})' \otimes i_T,\\
\mathbf{f}_i &\equiv (f_i(1/T),\ldots,f_i(T/T))',\ \mathbf{F} \equiv (\mathbf{f}_1,\ldots,\mathbf{f}_n)',\ \mathbf{f} \equiv [f(1/T),\ldots,f(T/T)]'.
\end{aligned}$$

Note that under H_0, $\mathbf{F} = i_n \otimes \mathbf{f}$, and we can write the model (4.1.1) as

$$Y_{it} = X_{it}'\beta + f(t/T) + \alpha_i + \varepsilon_{it}, \tag{4.2.2}$$

or in matrix notation as

$$Y = X\beta + i_n \otimes \mathbf{f} + D\alpha + \varepsilon, \tag{4.2.3}$$

provided we impose the identification condition $\sum_{i=1}^n \alpha_i = 0$.

Following Su and Ullah (2006) and CGL (2010), we estimate the model (4.2.2) by using the profile least squares method. Let $k(\cdot)$ denote a univariate kernel function and h a bandwidth. Let $k_h(\cdot) \equiv k(\cdot/h)/h$. For any positive integer p, let $z_{h,t}^{[p]}(\tau) \equiv (1, (t/T-\tau)/h, \ldots, [(t/T-\tau)/h]^p)'$,

$$z_h^{[p]}(\tau) \equiv \left(z_{h,1}^{[p]}(\tau),\ldots,z_{h,T}^{[p]}(\tau)\right)', \text{ and } Z_h^{[p]}(\tau) \equiv i_n \otimes z_h^{[p]}(\tau).$$

We assume that f is $(p+1)$th order continuously differentiable a.e. Let $D_h^p f(\tau) \equiv (f(\tau), hf^{(1)}(\tau), \ldots, h^p f^{(p)}(\tau)/p!)'$. Then for t/T in the neighborhood of $\tau \in (0,1)$, we have by the pth order Taylor expansion that $f\left(\frac{t}{T}\right) = D_h^p f(\tau)' z_{h,t}^{[p]}(\tau) + o\left(\left(\frac{t}{T}-\tau\right)^p\right)$. Let $k_{h,t}(\tau) \equiv k_h(t/T-\tau)$, $K_h(\tau) \equiv \operatorname{diag}\big(k_{h,1}(\tau),\ldots,k_{h,T}(\tau)\big)$, and $\mathbb{K}_h(\tau) \equiv I_n \otimes K_h(\tau)$. Define

$$\begin{aligned}
s(\tau) &\equiv \left(z_h^{[p]}(\tau)' K_h(\tau) z_h^{[p]}(\tau)\right)^{-1} z_h^{[p]}(\tau)' K_h(\tau) \text{ and}\\
S(\tau) &\equiv \left(Z_h^{[p]}(\tau)' \mathbb{K}_h(\tau) Z_h^{[p]}(\tau)\right)^{-1} Z_h^{[p]}(\tau)' \mathbb{K}_h(\tau) = n^{-1} i_n' \otimes s(\tau).
\end{aligned}$$

The profile least squares method is composed of the following three steps:

1. Let $\theta \equiv (\alpha', \beta')'$. For given θ and $\tau \in (0,1)$, we estimate $D_h^p f(\tau)$ by

   $$\widehat{D}_{h,\theta}^{p} f(\tau) \equiv \underset{F \in \mathbb{R}^{p+1}}{\arg\min} (Y - X\beta - D\alpha - Z_h^{[p]}(\tau) F)' \mathbb{K}_h(\tau) (Y - X\beta - D\alpha - Z_h^{[p]}(\tau) F).$$

   Noting that $S(\tau) D = 0$ by straightforward calculations, the estimator $\widehat{D}_{h,\theta}^{p} f(\tau)$ is in fact free of α and its first element is given by

   $$\widehat{f}_\beta(\tau) \equiv e_1' S(\tau)(Y - X\beta - D\alpha) = n^{-1} \sum_{i=1}^{n} e_1' s(\tau)(Y_i - X_i\beta), \qquad (4.2.4)$$

   where $e_1 = (1, 0, \ldots, 0)'$ is a $(p+1) \times 1$ vector. Let $\widehat{\mathbf{f}}_\beta \equiv (\widehat{f}_\beta(1/T), \ldots, \widehat{f}_\beta(T/T))'$, $S_T \equiv ([e_1' S(1/T)]', \cdots, [e_1' S(T/T)]')'$, and $S_{nT} \equiv i_n \otimes S_T$. Then we have

   $$\widehat{\mathbf{F}}_\beta \equiv i_n \otimes \widehat{\mathbf{f}}_\beta = S_{nT}(Y - X\beta). \qquad (4.2.5)$$

2. We estimate (α, β) by

   $$\begin{aligned} \left(\widehat{\alpha}, \widehat{\beta}\right) &\equiv \underset{\alpha,\beta}{\arg\min} \left(Y - X\beta - D\alpha - \widehat{\mathbf{F}}_\beta\right)' \left(Y - X\beta - D\alpha - \widehat{\mathbf{F}}_\beta\right) \\ &= \underset{\alpha,\beta}{\arg\min} (Y^* - X^*\beta - D\alpha)' (Y^* - X^*\beta - D\alpha) \end{aligned}$$

   where $Y^* \equiv (I_{nT} - S_{nT}) Y$ and $X^* \equiv (I_{nT} - S_{nT}) X$. Let $M_D \equiv I_{nT} - D(D'D)^{-1} D'$. Using the formula for partitioned regression, we obtain

   $$\widehat{\beta} = \left(X^{*\prime} M_D X^*\right)^{-1} X^{*\prime} M_D Y^*, \text{ and} \qquad (4.2.6a)$$

   $$\widehat{\alpha} \equiv (\widehat{\alpha}_2, \ldots, \widehat{\alpha}_n) = \left(D'D\right)^{-1} D'(Y^* - X^*\widehat{\beta}). \qquad (4.2.6b)$$

   Then α_1 can be estimated by $\widehat{\alpha}_1 \equiv -\sum_{i=2}^{n} \widehat{\alpha}_i$.

3. Plugging (4.2.6a) into (4.2.4), we obtain the estimator of $f(\tau)$:

   $$\widehat{f}(\tau) = e_1' S(\tau)(Y - X\widehat{\beta}). \qquad (4.2.7)$$

Let

$$\widehat{\mathbf{f}} \equiv \left(\widehat{f}(1/T),\ldots,\widehat{f}(T/T)\right)' \text{ and } \widehat{\mathbf{F}} \equiv S_{nT}\left(Y - X\widehat{\beta}\right) = i_n \otimes \widehat{\mathbf{f}}. \tag{4.2.8}$$

After we obtain estimates of β and $f(t/T)$, we can estimate $u_{it} \equiv \alpha_i + \varepsilon_{it}$ by $\widehat{u}_{it} \equiv Y_{it} - \widehat{\beta}' X_{it} - \widehat{f}(t/T)$ under the null. Let $\widehat{u}_i \equiv (\widehat{u}_{i1},\ldots,\widehat{u}_{iT})'$ and $\widehat{u} \equiv (\widehat{u}_1',\ldots,\widehat{u}_n')'$. Then it is easy to verify that

$$\begin{aligned}
\widehat{u} &= (\varepsilon - S_{nT}\varepsilon) + D\alpha + X^*(\beta - \widehat{\beta}) + \mathbf{F}^*,\\
\widehat{u}_i &= (\varepsilon_i - S_T\varepsilon) + \alpha_i i_T + (X_i - S_T X)(\beta - \widehat{\beta}) + (\mathbf{f}_i - S_T\mathbf{F}),\\
\widehat{u}_{it} &= \alpha_i + \left[\varepsilon_{it} - e_1' S(t/T)\varepsilon\right] + [X_{it} - e_1' S(\tfrac{t}{T})X](\beta - \widehat{\beta}) + [f_i(\tfrac{t}{T}) - e_1' S(\tfrac{t}{T})\mathbf{F}],
\end{aligned}$$

where $\mathbf{F}^* \equiv (I_{nT} - S_{nT})\mathbf{F}$.

4.2.3 A nonparametric R^2-based test for common trends

The idea behind our test is simple. Under H_0, $\widehat{u}_{it}$ is a consistent estimate for $u_{it} = \alpha_i + \varepsilon_{it}$, and there is no time trend in $\{u_{it}\}_{t=1}^T$ for each cross sectional unit i. Nevertheless, under H_1 $\widehat{u}_{it}$ includes an individual-specific time trend component $f_i(t/T) - f^0(t/T)$, where $f^0(\tau) \equiv p\lim \widehat{f}(\tau)$. This motivates us to consider a residual-based test for common trends.

For each i, we propose to run the nonparametric regression of $\{\widehat{u}_{it}\}_{t=1}^T$ on $\{t/T\}_{t=1}^T$:

$$\widehat{u}_{it} = m_i(t/T) + \eta_{it} \tag{4.2.9}$$

where $m_i(\tau) \equiv f_i(\tau) - f^0(\tau)$ and $\eta_{it} = \alpha_i + \varepsilon_{it}^* + (\beta - \widehat{\beta})' X_{it}^* + f^0(t/T) - e_1' S(t/T)\mathbf{F}$ is the new error term in the above regression. Clearly, under H_0 we have $m_i(\tau) = 0$ for $\tau \in [0,1]$. Given observations $\{\widehat{u}_{it}\}_{t=1}^T$, the local linear regression of $\widehat{u}_{it}$ on t/T is fitted by weighted least squares (WLS) as follows

$$\min_{(c_{i0},c_{i1})\in\mathbb{R}^2} \frac{1}{T}\sum_{t=1}^T \left[\widehat{u}_{it} - c_{i0} - c_{i1}\left(\frac{t}{T} - \tau\right)\right]^2 \overline{w}_{b,t}(\tau) \tag{4.2.10}$$

where $b \equiv b(T)$ is a bandwidth parameter such that $b \to 0$ as $T \to \infty$, $\overline{w}_{b,t}(\tau) \equiv$

$w_b(t/T-\tau)/\int_0^1 w_b(t/T-s)ds$, $w_b(\cdot) \equiv w(\cdot/b)/b$, and $w(\cdot)$ is a probability density function (p.d.f.) that has support $[-1,1]$. By the proof of Lemma .0.30 in the appendix, $\lambda_{tT} \equiv \int_0^1 w_b(t/T-s)\,ds = 1$ for $t/T \in [b, 1-b]$ and is larger than 1/2 otherwise. Therefore, $\overline{w}_{b,t}(\tau)$ plays the role of a boundary kernel to ensure that $\int_0^1 \overline{w}_{b,t}(\tau)\,d\tau = 1$ for any $t = 1, ..., T$. [2]

Let $\widetilde{c}_i \equiv (\widetilde{c}_{i0}, \widetilde{c}_{i1})'$ denote the solution to the above minimization problem. Following Su and Ullah (2011), the normal equations for the above regression imply the following local ANOVA decomposition of the total sum of squares (TSS)

$$TSS_i(\tau) = ESS_i(\tau) + RSS_i(\tau) \tag{4.2.11}$$

where

$$\begin{aligned}
TSS_i(\tau) &\equiv \sum_{t=1}^{T} \left(\widehat{u}_{it} - \overline{\widehat{u}}_i\right)^2 \overline{w}_{b,t}(\tau), \\
ESS_i(\tau) &\equiv \sum_{t=1}^{T} \left(\widetilde{c}_{i0} + \widetilde{c}_{i1}(t/T-\tau) - \overline{\widehat{u}}_i\right)^2 \overline{w}_{b,t}(\tau), \\
RSS_i(\tau) &\equiv \sum_{t=1}^{T} \left(\widehat{u}_{it} - \widetilde{c}_{i0} - \widetilde{c}_{i1}(t/T-\tau)\right)^2 \overline{w}_{b,t}(\tau),
\end{aligned}$$

and $\overline{\widehat{u}}_i \equiv T^{-1}\sum_{t=1}^{T} \widehat{u}_{it}$. A global ANOVA decomposition of TSS_i is given by

$$TSS_i = ESS_i + RSS_i \tag{4.2.12}$$

where

$$\begin{aligned}
TSS_i &\equiv \int_0^1 TSS_i(\tau)\,d\tau = \sum_{t=1}^{T} (\widehat{u}_{it} - \overline{\widehat{u}}_i)^2, \; ESS_i \equiv \int_0^1 ESS_i(\tau)\,d\tau, \text{ and} \\
RSS_i &\equiv \int_0^1 RSS_i(\tau)\,d\tau.
\end{aligned} \tag{4.2.13}$$

Then one can define the nonparametric goodness-of-fit $\left(R^2\right)$ for the above local

[2] Alternatively, one can use the standard kernel weight $w_b(t/T-\tau)$ in place of $\overline{w}_{b,t}(\tau)$ in (4.2.10) and decompose $TSS_i(\tau)$ analogously to the decomposition in (4.2.11). But as $\lambda_{tT} \equiv \int_0^1 w_b(t/T-s)\,ds$ is not identically 1 for all t, $\int_0^1 TSS_i(\tau)\,d(\tau)$ in this case does not lead to the simple expression in (4.2.13).

linear regression as

$$R_i^2 \equiv \frac{ESS_i}{TSS_i}.$$

Under H_0, $\{\widehat{u}_{it}\}$ contains no useful trending information so that the above R_i^2 should be close to 0 for each individual i.

Let $W_b(\tau) \equiv$diag$(\overline{w}_{b,1}(\tau), ..., \overline{w}_{b,T}(\tau))$,

$$H(\tau) \equiv W_b(\tau) z_b^{[1]}(\tau) \left(z_b^{[1]}(\tau)' W_b(\tau) z_b^{[1]}(\tau)\right)^{-1} z_b^{[1]}(\tau)' W_b(\tau),$$

and $\bar{H} \equiv \int_0^1 H(\tau) d\tau$. It is easy to show that

$$TSS_i = \widehat{u}_i' M \widehat{u}_i, \ ESS_i = \widehat{u}_i'(\bar{H} - L)\widehat{u}_i, \text{ and } RSS_i = \widehat{u}_i'(I_T - \bar{H})\widehat{u}_i,$$

where $M \equiv I_T - L$ and $L \equiv i_T i_T'/T$. Define the average nonparametric R^2 as

$$\overline{R}^2 \equiv \frac{1}{n}\sum_{i=1}^{n} R_i^2 = \frac{1}{n}\sum_{i=1}^{n}\frac{ESS_i}{TSS_i}.$$

Clearly $0 \leq \overline{R}^2 \leq 1$ by construction. We will show that after being appropriately centered and scaled, $\overline{R}^2$ is asymptotically normally distributed under the null and a sequence of Pitman local alternatives.

Before proceeding further, it is worth mentioning a related test statistic that is commonly used in the literature. Under H_0, the $m_i(\cdot)$ function in (4.2.9) is also common for all i and thus can be written as $m(\cdot)$. $m(t/T) = 0$ for all $t = 1, ..., T$ under H_0 and we can estimate this zero function by pulling all the cross sectional and time series observations together to obtain the estimate $\widehat{m}(\cdot)$, say. Then we can compare this estimate with the nonparametric trend regression estimate $\hat{m}_i(t/T)$ of $m_i(t/T)$ to obtain the following L_2 type of test statistic

$$D_{nT} \equiv \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\left[\hat{m}_i(t/T) - \hat{m}(t/T)\right]^2.$$

Noting that the estimate $\hat{m}(t/T)$ has a faster convergence rate than $\hat{m}_i(t/T)$ to 0 under the null, it is straightforward to show that under suitable conditions this test statistic is asymptotically equivalent to $\bar{D}_{nT} \equiv \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{m}_i(t/T)^2$ under the null.

Further noticing that $\sum_{t=1}^{T} \hat{m}_i (t/T)^2 / TSS_i$ can be regarded as a version of nonparametric *noncentered* R^2 measure for the cross sectional unit i, we can simply interpret $\bar{D}_{nT}$ as a weighted nonparametric noncentered R^2-based test where the weight for cross sectional unit i is given by TSS_i. In this chapter we focus on the test based on $\overline{R}^2$ because it is scale-free and is asymptotically pivotal under the null after bias-correction. See the remark after Theorem 4.3.1 for further discussion.

4.3 Asymptotic Distributions

In this section we first present the assumptions that are used in later analysis and then study the asymptotic distribution of average nonparametric R^2 under both the null hypothesis and a sequence of Pitman local alternatives. We then prove the consistency of the test and propose a bootstrap procedure to obtain bootstrap p-values.

4.3.1 Assumptions

Let $\mathscr{F}_{n,t}(\xi)$ denote the σ-field generated by $(\xi_1, ..., \xi_t)$ for a time series $\{\xi_t\}$. To establish the asymptotic distribution of our test statistic, we make the following assumptions.

Assumption A1. (i) The regressor X_{it} is generated as follows:

$$X_{it} = g_i\left(\frac{t}{T}\right) + v_{it}. \tag{4.3.1}$$

(ii) Let $v_t \equiv (v_{1t}, ..., v_{nt})'$for $t = 1, ..., T$. $\{v_t, \mathscr{F}_{n,t}(v)\}$ is a stationary martingale difference sequence (m.d.s.) of $n \times d$ random matrices.

(iii) $E\left[\|v_{it}\|^2 | \mathscr{F}_{n,t-1}(v)\right] = \sigma_{v,i}^2$ a.s. for each i and $\max_{1 \le i \le n} E\|v_{it}\|^4 < c_v < \infty$. There exist $d \times d$ positive definite matrices Σ_v and Σ_v^* such that

$$\frac{1}{n}\sum_{i=1}^{n} E\left(v_{it}v_{it}'\right) \to \Sigma_v,\ \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} E\left(v_{it}v_{jt}'\right) \to \Sigma_v^*,\ \text{and } E\left\|\sum_{i=1}^{n} v_{it}\right\|^{\delta} = O\left(n^{\delta/2}\right),$$

for some $\delta > 2$.

Assumption A2. (i) Let $\varepsilon_t \equiv (\varepsilon_{1t}, ..., \varepsilon_{nt})'$ for $t = 1, ..., T$. $\{\varepsilon_t, t \geq 1\}$ is a stationary sequence.

(ii) $\{\varepsilon_t, \mathscr{F}_{n,t}(\varepsilon)\}$ is an m.d.s. such that $E(\varepsilon_{it}|\mathscr{F}_{n,t-1}(\varepsilon)) = 0$ a.s. for each i.

(iii) $E\left(\varepsilon_{it}\varepsilon_{jt}|\mathscr{F}_{n,t-1}(\varepsilon)\right) = \omega_{ij}$ for each pair (i,j). Let $\sigma_i^2 \equiv \omega_{ii}$. $0 < \underline{c} \leq \min_{1\leq i\leq n} \sigma_i^2$, $\max_{1\leq i,j\leq n}\left|\omega_{ij}\right| \leq \bar{c} < \infty$, $\max_{1\leq i\leq n} E\left(\varepsilon_{it}^8\right) \leq \bar{c} < \infty$, $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \sum_{j=1}^n \left|\omega_{ij}\right| < \infty$, $\lim_{n\to\infty} \frac{1}{n^2}\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n \left|\varsigma_{ijk}\varsigma_{ijl}\right| < \infty$, and $\lim_{n\to\infty} \frac{1}{n^2}\sum_{1\leq i_1\neq i_2\leq n}\sum_{1\leq i_3\neq i_4\leq n} \left|\kappa_{i_1i_2i_3i_4}\right| < \infty$, where $\varsigma_{ijk} \equiv E\left(\varepsilon_{it}\varepsilon_{jt}\varepsilon_{kt}\right)$ and $\kappa_{i_1i_2i_3i_4} \equiv E\left(\varepsilon_{i_1t}\varepsilon_{i_2t}\varepsilon_{i_3t}\varepsilon_{i_4t}\right)$.

(iv) Let $\xi_{it} \equiv \varepsilon_{it}^2 - \sigma_i^2$. There exists an even number $\lambda \geq 4$ such that $\frac{1}{nT^{\lambda/2}}\sum_{i=1}^n \sum_{1\leq t_1,t_2,...,t_\lambda\leq T} E\left(\xi_{it_1}\xi_{it_2}...\xi_{it_\lambda}\right) < \infty$.

(v) ε_{it} is independent of v_{js} for all i, j, t, s.

(vi) There exists a $d \times d$ positive definite matrix $\Sigma_{v\varepsilon}$ such that as $n \to \infty$,

$$\frac{1}{n}\sum_{i=1}^n \sum_{j=1}^n E\left(v_{i1}v_{j1}'\right) E\left(\varepsilon_{i1}\varepsilon_{j1}\right) \to \Sigma_{v\varepsilon}.$$

Assumption A3. The trend functions $f_i(\cdot)$ and $g_i(\cdot)$ have continuous derivatives up to the $(p+1)$th order.

Assumption A4. The kernel functions $k(\cdot)$ and $w(\cdot)$ are continuous and symmetric p.d.f.'s with compact support $[-1, 1]$.

Assumption A5. As $(n,T) \to \infty$, $b \to 0$, $h \to 0$, $\sqrt{nb^{-1}h^2}/\log(nT) \to \infty$, $\min(Tb, nh^{1/2}) \to \infty$, $n^{1/2}Th^{2p+2} \to 0$, and $n^{1/2+2/\lambda}T^{-1} \to 0$.

Remark 1. A1 is similar is to Assumption A2 in CGL (2010). Like CGL, we allow for cross sectional dependence in $\{v_{it}\}$ and the degree of cross sectional dependence is controlled by the moment conditions in A1(iii). Unlike CGL, we allow $\{X_{it}\}$ to possess heterogeneous time trends $\{g_i\}$ in (4.3.1), and we relax their i.i.d. assumption of v_t to the m.d.s. condition. A2 specifies conditions on $\{\varepsilon_{it}\}$ and their interaction with $\{v_{it}\}$. Note that we allow for cross sectional dependence in $\{\varepsilon_{it}\}$ but rule out serial dependence in A2(ii). To facilitate the derivation of the asymptotic variance of our test statistic, we also impose time-invariant conditional correlations among all cross sectional units in A2(iii). A2(iv) is readily satisfied under suitable mixing conditions together with moment conditions. The independence between $\{\varepsilon_{it}\}$ and $\{v_{it}\}$ in A2(v) can be relaxed by modifying the proofs in CGL (2010) significantly. A3 is standard for local polynomial regressions. A4 is a mild and commonly-used condition in the nonparametrics literature. A5 specifies conditions on the bandwidths h and b and sample sizes n and T. Note that we allow $n/T \to c \in [0,\infty]$ as $(n,T) \to \infty$. If we use the optimal rate of bandwidths, i.e., $h \propto (nT)^{-1/(2p+3)}$ in the p-th order local polynomial regression and $b \propto T^{-1/5}$ in the local linear regression, then A5 requires

$$\frac{n^{4p+5}}{T} \to \infty, \quad \frac{n^{\frac{1}{2}-\frac{1}{2p+3}}T^{\frac{1}{10}-\frac{1}{2p+3}}}{\log(nT)} \to \infty, \quad \frac{(nT)^{\frac{1}{2p+3}}}{n^{1/2}} \to 0, \text{ and } \frac{n^{1/2+2/\lambda}}{T} \to 0.$$

More specifically, if we choose $p = 3$, then A5 implies: $n^{7/18}/(T^{1/90}\log(nT)) \to \infty$, $T/n^{3.5} \to 0$, and $n^{1/2+2/\lambda}/T \to 0$. If $n \propto T^a$, A5 requires $a \in (2/7, 1/(0.5+2/\lambda))$.

4.3.2 Asymptotic null distribution

Let $\bar{H}_{ts}$ denote the (t,s)th element of $\bar{H}$. Let $\alpha_{ts} \equiv T\bar{H}_{ts} - 1$ and $Q \equiv T^{-1}\text{diag}(\alpha_{11}, \ldots, \alpha_{TT})$. Define

$$\begin{aligned}
B_{nT} &\equiv \sqrt{\frac{b}{n}}\sum_{i=1}^{n}\frac{\varepsilon_i' Q \varepsilon_i}{T^{-1}TSS_i}, \\
\Omega_{nT} &\equiv \frac{2b}{T^2}\sum_{1\le t\neq s\le T}\alpha_{ts}^2\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\rho_{ij}^2\right), \text{ where } \rho_{ij} \equiv \omega_{ij}\sigma_i^{-1}\sigma_j^{-1} \\
\Gamma_{nT} &\equiv n^{1/2}Tb^{1/2}\overline{R}^2 - B_{nT} = \sqrt{\frac{b}{n}}\sum_{i=1}^{n}\frac{ESS_i - \varepsilon_i' Q \varepsilon_i}{T^{-1}TSS_i}.
\end{aligned}$$

The following theorem gives the asymptotic null distribution of Γ_{nT}.

Theorem 4.3.1 *Suppose Assumptions A1-A5 hold. Then under H_0,*

$$\Gamma_{nT} \xrightarrow{d} N(0, \Omega_0)$$

where $\Omega_0 \equiv \lim_{(n,T)\to\infty}\Omega_{nT}$.

Remark 2. The proof of the above theorem is lengthy and involves several subsidiary propositions, which are given in Appendix A. Under the null hypothesis, we first demonstrate that $\Gamma_{nT} = \Gamma_{nT,1} + o_P(1)$, where $\Gamma_{nT,1} \equiv \sum_{i=1}^{n} \varphi_i(\varepsilon_i)$ and $\varphi_i(\varepsilon_i) = n^{-1/2}T^{-1}b^{1/2}\sum_{1\le t<s\le T} \alpha_{ts}\varepsilon_{it}\varepsilon_{is}/\sigma_i^2$. Then we apply the martingale central limit theorem (CLT) to show that $\Gamma_{nT,1} \xrightarrow{d} N(0,\Omega_0)$. In general, Γ_{nT} is not asymptotically pivotal as cross sectional dependence enters its asymptotic variance Ω_0. Nevertheless, if cross sectional dependence is absent, then Γ_{nT} is an *asymptotic pivotal* test because now $\Omega_0 = \lim_{(n,T)\to\infty} \frac{2b}{T^2}\sum_{1\le t\ne s\le T} \alpha_{ts}^2$, which is free of nuisance parameters. This is one advantage to base a test on the scale-free nonparametric R^2 measure.

To implement the test, we need to estimate both the asymptotic bias and variance terms. Let

$$\widehat{B}_{nT} \equiv \sqrt{\frac{b}{n}}\sum_{i=1}^{n}\frac{\widehat{u}_i' MQM\widehat{u}_i}{TSS_i/T} \text{ and } \widehat{\Omega}_{nT} \equiv \frac{2b}{T^2}\sum_{1\le t\ne s\le T}\alpha_{ts}^2\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\widehat{\rho}_{ij}^2\right)$$

where $\widehat{\rho}_{ij} \equiv \hat{\omega}_{ij}/(\widehat{\sigma}_i\widehat{\sigma}_j)$, $\hat{\omega}_{ij} \equiv T^{-1}\sum_{t=1}^{T}(\widehat{u}_{it} - \overline{\widehat{u}}_i)(\widehat{u}_{jt} - \overline{\widehat{u}}_j)$, $\widehat{\sigma}_i^2 = T^{-1}\sum_{t=1}^{T}(\widehat{u}_{it} - \overline{\widehat{u}}_i)^2$ and $\overline{\widehat{u}}_i \equiv T^{-1}\sum_{t=1}^{T}\widehat{u}_{it}$. We show in the proof of Corollary 4.3.2 below that $\widehat{B}_{nT} = B_{nT} + o_P(1)$ and $\widehat{\Omega}_{nT} = \Omega_0 + o_P(1)$. Then we obtain a feasible test statistic as

$$\overline{\Gamma}_{nT} = \frac{n^{1/2}Tb^{1/2}\overline{R}^2 - \widehat{B}_{nT}}{\sqrt{\widehat{\Omega}_{nT}}} = \frac{1}{\sqrt{\widehat{\Omega}_{nT}}}\sqrt{\frac{b}{n}}\sum_{i=1}^{n}\frac{ESS_i - \widehat{u}_i' MQM\widehat{u}_i}{TSS_i/T}. \tag{4.3.2}$$

Corollary 4.3.2 *Under Assumptions A1-A5,* $\overline{\Gamma}_{nT} \xrightarrow{d} N(0,1)$.

We then compare $\overline{\Gamma}_{nT}$ with the one-sided critical value z_α, i.e., the upper αth percentile from the standard normal distribution. We reject the null when $\overline{\Gamma}_{nT} > z_\alpha$ at the α significance level.

4.3.3 Asymptotic distribution under local alternatives

To examine the asymptotic local power of our test, we consider the following sequence of Pitman local alternatives:

$$H_1(\gamma_{nT}) : f_i(\tau) = f(\tau) + \gamma_{nT}\Delta_{ni}(\tau) \text{ for all } \tau \in [0,1] \text{ and } i = 1, ..., n \qquad (4.3.3)$$

where $\gamma_{nT} \to 0$ as $(n,T) \to \infty$ and $\Delta_{ni}(\cdot)$ is a continuous function on $[0,1]$. Let $\Delta_{ni} \equiv (\Delta_{ni}(1/T), ..., \Delta_{ni}(T/T))'$. Define

$$\Theta_0 \equiv \lim_{(n,T)\to\infty} \frac{1}{nT} \sum_{i=1}^{n} \Delta_{ni}'(\bar{H} - L)\Delta_{ni}/\sigma_i^2.$$

In the appendix we show that $\Theta_0 = C_w \lim_{n\to\infty}(n^{-1}\sum_{i=1}^n \int_0^1 \Delta_{ni}^2(\tau)\,d\tau/\sigma_i^2)$, where $C_w \equiv \int_{-1}^1 \{\int_{-1}^1 [1+\omega_2^{-1}u(u-v)]\ w(u)w(u-v)\,du\ [\int_{-1}^1 w(z-v)\,dz]^{-1} - 1\}dv$ and $\omega_2 \equiv \int_{-1}^1 w(u)u^2 du$.

To derive the asymptotic property of our test under the alternatives, we add the following assumption.

Assumption A6. $\frac{1}{n}\sum_{i=1}^n \int_0^1 |g_i(\tau) - \overline{g}(\tau)|\,d\tau = o(1)$ where $\overline{g}(\cdot) \equiv \frac{1}{n}\sum_{i=1}^n g_i(\cdot)$.

That is, the nonparametric trending functions $\{g_i(\cdot),\ 1 \le i \le n\}$ that appear in A1 are *asymptotically homogeneous.* This assumption is needed to determine the probability order of $\widehat{\beta} - \beta$ under $H_1(\gamma_{nT})$ and H_1. Without A6, we can only show that $\widehat{\beta} - \beta = O_P(\gamma_{nT})$ under $H_1(\gamma_{nT})$ and that $\widehat{\beta} - \beta = O_P(1)$ under H_1 for γ_{nT} that converges to zero no faster than $n^{-1/2}T^{-1/2}$. With A6, we demonstrate in Lemma .0.35 that $\widehat{\beta} - \beta = o_P(\gamma_{nT})$ under $H_1(\gamma_{nT})$ and that $\widehat{\beta} - \beta = o_P(1)$ under H_1, which are sufficient for us to establish the local power property and the global consistency of our test respectively in Theorems 4.3.3 and 4.3.4 below.

The following theorem establishes the local power property of our test.

Theorem 4.3.3 *Suppose Assumptions A1-A6 hold. Suppose that $\Delta_{ni}(\cdot)$ is a continuous function such that $\sum_{i=1}^n \Delta_{ni}(\tau) = 0$ for $\tau \in [0,1]$ and $\sup_{n\ge1}\max_{1\le i\le n}\int_0^1 \Delta_{ni}^2(\tau)\,d\tau < \infty$. Then with $\gamma_{nT} = n^{-1/4}T^{-1/2}b^{-1/4}$ in (4.3.3) the local power of our test satisfies*

$$P\left(\overline{\Gamma}_{nT} > z_\alpha | H_1(\gamma_{nT})\right) \to 1 - \Phi\left(z_\alpha - \Theta_0/\sqrt{\Omega_0}\right),$$

where $\Phi(\cdot)$ is the cumulative distribution function (CDF) of the standard normal distribution.

Remark 3. Theorem 4.3.3 implies that our test has nontrivial asymptotic power against alternatives that diverge from the null at the rate $n^{-1/4}T^{-1/2}b^{-1/4}$. The power increases with the magnitude of Θ_0. Clearly, as either n or T increases, the power of our test will increase but it increases faster as $T \to \infty$ than as $n \to \infty$ for the same choice of b.

4.3.4 Consistency of the test

To study the consistency of our test, we take $\gamma_{nT} = 1$ and $\Delta_{ni}(\tau) = \Delta_i(\tau)$ in (4.3.3), where $\Delta_i(\cdot)$ is a continuous function on $[0,1]$ such that $\underline{c}_\Delta \le n^{-1}\sum_{i=1}^n \int_0^1 \Delta_i(\tau)^2 d\tau \le \overline{c}_\Delta$ for some $0 < \underline{c}_\Delta < \overline{c}_\Delta < \infty$. Let $\Delta_i \equiv (\Delta_i(1/T), ..., \Delta_i(T/T))'$. Define

$$\Theta_A \equiv \lim_{(n,T)\to\infty} \frac{1}{nT}\sum_{i=1}^n \Delta_i'(\bar{H} - L)\Delta_i / \overline{\sigma}_i^2.$$

where $\overline{\sigma}_i^2 \equiv \sigma_i^2 + \int_0^1 \Delta_i(\tau)^2 d\tau - (\int_0^1 \Delta_i(\tau) d\tau)^2$. The following theorem establishes the consistency of the test.

Theorem 4.3.4 *Suppose Assumptions A1-A6 hold. Under H_1,*

$$n^{-1/2}T^{-1}b^{-1/2}\overline{\Gamma}_{nT} = \Theta_A + o_P(1).$$

Theorem 4.3.4 implies that under H_1, $P\left(\overline{\Gamma}_{nT} > d_{nT}\right) \to 1$ as $(n,T) \to \infty$ for any sequence $d_{nT} = o\left(n^{1/2}Tb^{1/2}\right)$ provided $\Theta_A > 0$, thus establishing the global consistency of the test.

4.3.5 A bootstrap version of the test

It is well known that asymptotic normal distribution of many nonparametric tests may not approximate their finite sample distributions well in practice. Therefore we now propose a fixed-regressor bootstrap method [e.g., Hansen (2000)] to obtain the bootstrap approximation to the finite sample distribution of our test statistic under the null.

We propose to generate the bootstrap version of our test statistic $\overline{\Gamma}_{nT}$ as follows:

1. Obtain the augmented residuals $\widehat{u}_{it} = Y_{it} - \widehat{f}(t/T) - X_{it}\widehat{\beta}$, where $\widehat{f}$ and $\widehat{\beta}$ are obtained by the profile least squares estimation of the restricted model. Calculate the test statistic $\overline{\Gamma}_{nT}$.

2. Let $\overline{\hat{u}}_i \equiv T^{-1}\sum_{t=1}^{T}\widehat{u}_{it}$ and $\widehat{u}_t \equiv (\widehat{u}_{1t} - \overline{\hat{u}}_1, ..., \widehat{u}_{nt} - \overline{\hat{u}}_n)'$. Obtain the bootstrap error u_t^* by random sampling with replacement from $\{\widehat{u}_s, s = 1, 2, ..., T\}$. Generate the bootstrap analog of Y_{it} by holding X_{it} as fixed: $Y_{it}^* = \widehat{f}(t/T) + X_{it}\widehat{\beta} + \overline{\hat{u}}_i + u_{it}^*$ for $i = 1, ..., n$ and $t = 1, \ldots, T$, where u_{it}^* is the ith element in the n-vector u_t^*.

3. Based on the bootstrap resample $\{Y_{it}^*,\ X_{it}\}$, run the profile least squares estimation of the restricted model to obtain the bootstrap augmented residuals $\{\widehat{u}_{it}^*\}$.

4. Based on $\{\widehat{u}_{it}^*\}$, compute the bootstrap test statistic $\overline{\Gamma}_{nT}^* \equiv (Tn^{1/2}b^{1/2}\overline{R}^{2*} - \widehat{B}_{nT}^*)/\sqrt{\widehat{\Omega}_{nT}^*}$, where $\overline{R}^{2*}$, $\widehat{B}_{nT}^*$ and $\widehat{\Omega}_{nT}^*$ are defined analogously to $\overline{R}^2$, $\widehat{B}_{nT}$ and $\widehat{\Omega}_{nT}$, respectively, but with $\widehat{u}_{it}$ being replaced by $\widehat{u}_{it}^*$.

5. Repeat Step 2-4 for B times and index the bootstrap statistics as $\{\overline{\Gamma}_{nT,l}^*\}_{l=1}^{B}$. The bootstrap p-value is calculated by $p^* \equiv B^{-1}\sum_{l=1}^{B} 1\{\overline{\Gamma}_{nT,l}^* > \overline{\Gamma}_{nT}\}$, where $1\{\cdot\}$ is the usual indicator function.

Some facts are worth mentioning: (i) Conditionally on the original sample $\mathscr{W} \equiv \{(Y_{it}, X_{it}),\ i = 1, \ldots, n,\ t = 1, \ldots, T\}$, the bootstrap replicates u_{it}^* are dependent among cross sectional units, and i.i.d. across time for fixed i; (ii) the regressor X_{it} is held fixed during the bootstrap procedure; (iii) the null hypothesis of common trends is imposed in Step 2.

4.4 Simulations and Applications

This section conducts a small set of simulations to assess the finite sample performance of the test. We then report empirical applications of the common trend test to UK climate change data and OECD real GDP growth data.

4.4.1 Simulation study

Data generating processes

We generate data according to six data generating processes (DGPs), among which DGPs 1-2 are used for the level study, and DGPs 3-6 are for the power study.

DGP 1:

$$y_{it} = x_{it}\beta + \left[\left(\frac{t}{T}\right)^3 + \frac{t}{T}\right] + \alpha_i + \varepsilon_{it},$$

where $i = 1,...,n, t = 1,...,T$, $\beta = 2$, for each i we generate x_{it} as i.i.d. $U(a_i - 3, a_i + 3)$ across t with a_i being i.i.d. $N(0,1)$, $\alpha_i = T^{-1}\sum_{t=1}^{T} x_{it}$ for $i = 2,...,n$, and $\alpha_1 = -\sum_{i=2}^{n} \alpha_i$.

DGP 2:

$$y_{it} = x_{it,1}\beta_1 + x_{it,2}\beta_2 + \left[2\left(\frac{t}{T}\right)^2 + \frac{t}{T}\right] + \alpha_i + \varepsilon_{it},$$

where $i = 1,...,n, t = 1,...,T$, $\beta_1 = 1$, $\beta_2 = 1/2$, $x_{it,1} = 1 + \sin(\pi t/T) + v_{it,1}$, $x_{it,2} = 0.5t/T + v_{i2,t}$, $v_{it,1}$ and $v_{it,2}$ are each i.i.d. $N(0,1)$ and independent of each other, $\alpha_i = \max(T^{-1}\sum_{t=1}^{T} x_{it,1}, T^{-1}\sum_{t=1}^{T} x_{it,2})$ for $i = 2,...,n$, and $\alpha_1 = -\sum_{i=2}^{n} \alpha_i$.

DGP 3:

$$y_{it} = x_{it}\beta + \left[(1+\delta_{i1})\left(\frac{t}{T}\right)^3 + (1+\delta_{i2})\frac{t}{T}\right] + \alpha_i + \varepsilon_{it},$$

where $i = 1,...,n$, $t = 1,...,T$, β, x_{it}, and α_i are generated as in DGP 1, and δ_{i1} and δ_{i2} are each i.i.d. $U(-1/2, 1/2)$, mutually independent and independent of x_{it} and α_i.

DGP 4:

$$y_{it} = x_{it,1}\beta_1 + x_{it,2}\beta_2 + \left[(2+\delta_{i1})\left(\frac{t}{T}\right)^2 + (1+\delta_{i2})\frac{t}{T}\right] + \alpha_i + \varepsilon_{it},$$

where $i = 1,...,n, t = 1,...,T$, β_1, β_2, $x_{it,1}$, $x_{it,2}$, and α_i are generated as in DGP 2, and δ_{i1} and δ_{i2} are each i.i.d. $U(-1/2, 1/2)$, mutually independent and independent of $(x_{it,1}, x_{it,2}, \alpha_i)$.

DGP 5:

$$y_{it} = x_{it}\beta + \left[(1+\delta_{nT,i1}) \left(\frac{t}{T} \right)^3 + (1+\delta_{nT,i2}) \frac{t}{T} \right] + \alpha_i + \varepsilon_{it},$$

where $i = 1, ..., n$, $t = 1, ..., T$, β, x_{it}, and α_i are generated as in DGP 1, and $\delta_{nT,i1}$ and $\delta_{nT,i2}$ are each i.i.d. $U(-7\gamma_{nT}, 7\gamma_{nT})$, mutually independent, and independent of x_{it} and α_i.

DGP 6:

$$y_{it} = x_{it,1}\beta_1 + x_{it,2}\beta_2 + \left[(1+\delta_{nT,i1}) \left(\frac{t}{T} \right)^2 + (1+\delta_{nT,i2}) \frac{t}{T} \right] + \alpha_i + \varepsilon_{it},$$

where $i = 1, ..., n$, $t = 1, ..., T$, β_1, β_2, $x_{it,1}$, $x_{it,2}$, and α_i are generated as in DGP 2, and $\delta_{nT,i1}$ and $\delta_{nT,i2}$ are each i.i.d. $U(-7\gamma_{nT}, 7\gamma_{nT})$, mutually independent and independent of $(x_{it,1}, x_{it,2}, \alpha_i)$.

Note that DGPs 5-6 are used to examine the finite sample behavior of our test under the sequence of Pitman local alternatives. For both DGPs, we set $\gamma_{nT} = n^{-1/4}T^{-1/2}\left(T^{-1/5}\right)^{-1/4}$ by choosing $b = T^{-1/5}$, and keep $\left\{\delta_{nT,i1}\right\}$ and $\left\{\delta_{nT,i2}\right\}$ fixed through the simulations. Similarly, $\{\delta_{i1}\}$ and $\{\delta_{i2}\}$ are kept fixed through the simulations for DGPs 3-4.

In all of the above DGPs, we generate $\{\varepsilon_{it}\}$ analogously to that in CGL (2010) and independently of all other variables on the right hand side of each DGP. Specifically, we generate ε_t as i.i.d. n-dimensional vector of Gaussian variables with zero mean and covariance matrix $(\omega_{ij})_{n\times n}$. We consider two configurations for $(\omega_{ij})_{n\times n}$:

$$\text{CD (I)} : \omega_{ij} = 0.5^{|j-i|}\sigma_i\sigma_j \text{ and CD (II): } \omega_{ij} = 0.8^{|j-i|}\sigma_i\sigma_j,$$

where i, $j = 1, ..., n$, and σ_i are i.i.d. $U(0,1)$. By construction, $\{\varepsilon_{it}\}$ are independent across t and cross sectionally dependent across i.

Test results

To implement our test, we need to choose two kernel functions and two bandwidth sequences. We choose both k and w to be the Epanechnikov kernel: $k(v) = w(v) =$

$0.75\left(1-v^2\right)1\{|v|\leq 1\}$. To estimate the restricted semiparametric model, we use the third order local polynomial regression and adopt the "leave-one-out" cross validation method to select the bandwidth h. To run the local linear regression of $\widehat{u}_{it}$ on t/T for each cross sectional unit i, we set $b=c\sqrt{\frac{1}{12}}T^{-1/5}$ for $c=0.5,1$ and 1.5 to examine the sensitivity of our test to the choice of bandwidth.[3]

We consider $n,T=25,50,100$. For each combination of n and T, we use 500 replications for both level and power study and 200 bootstrap resamples in each replication.

Table 4.1 reports the finite sample level of our test when the nominal level is 5%. From Table 4.1, we see that the levels of our test behave reasonably well except when n/T is big (e.g., $(n,T)=(50,25)$ or $(100,25)$). In the latter case, our test is undersized. For fixed n, as T increases, the level of our test approaches the nominal level fairly fast. We also note that the size of our test is robust to different choices of bandwidth.

Tables 4.2 reports the finite sample power of our test against global alternatives at the 5% nominal level. There is no time trend in the regressor x_{it} in DGP 3 whereas both regressors $x_{it,1}$ and $x_{it,2}$ contain a time trend component in DGP 4. We summarize some important findings from Table 4.2. First, as either n or T increases, the power of our test generally increases and finally reaches 1, but it increases faster as T increases than as n increases. This is compatible with our asymptotic theory. Secondly, comparing the power behavior of our test under CD (I) and CD (II) indicates that the degree of cross sectional dependence in the error terms has negative impact on the power of our test. This is as expected, as stronger cross sectional dependence implies less information in each additional cross sectional observation. Third, the choice of the bandwidth b has some effect on the power of our test. Surprisingly, a larger value of b is associated with a larger testing power.

Table 4.3 reports the finite sample power of our test against Pitman local alternatives at the 5% nominal level. From the table, we see that our test has nontrivial

[3] Here, the time trend regressor $\{t/T, t=1,2,...,T\}$ can be regarded as uniformly distributed on the interval $(0,1)$ and thus has variance 1/12.

Table 4.1: Finite sample rejection frequency for DGPs 1-2 (nominal level: 0.05)

			CD (I)			CD (II)		
DGP	n	T	$c=0.5$	$c=1$	$c=1.5$	$c=0.5$	$c=1$	$c=1.5$
1	25	25	0.036	0.038	0.038	0.034	0.028	0.032
		50	0.038	0.044	0.036	0.032	0.038	0.030
		100	0.046	0.054	0.052	0.042	0.042	0.056
	50	25	0.014	0.028	0.042	0.030	0.028	0.030
		50	0.034	0.056	0.054	0.038	0.044	0.044
		100	0.056	0.048	0.046	0.042	0.038	0.054
	100	25	0.018	0.024	0.022	0.018	0.028	0.028
		50	0.038	0.030	0.024	0.048	0.052	0.048
		100	0.052	0.038	0.054	0.042	0.050	0.048
2	25	25	0.048	0.050	0.050	0.036	0.022	0.038
		50	0.046	0.040	0.054	0.034	0.026	0.038
		100	0.056	0.064	0.072	0.030	0.038	0.062
	50	25	0.026	0.024	0.036	0.018	0.026	0.042
		50	0.056	0.056	0.062	0.040	0.036	0.046
		100	0.056	0.066	0.054	0.044	0.044	0.058
	100	25	0.014	0.016	0.016	0.020	0.022	0.036
		50	0.044	0.032	0.028	0.022	0.034	0.042
		100	0.042	0.046	0.058	0.032	0.040	0.040

power to detect the local alternatives at the rate $n^{-1/4}T^{-1/2}b^{-1/4}$, which confirms the asymptotic result in Theorem 4.3.3. As either n or T increases, we observe the alteration of the local power, which, unlike the case of global alternatives, does not necessarily increase.

4.4.2 Applications to real data

In this subsection we apply our test to two real data sets to illustrate its power to detect deviations from common trends, one is to UK climate change data and the other is to OECD economic growth data.

UK climate change data

The issue of global warming has received a lot of attention recently. Atak, Linton, and Xiao (2011) develop a semiparametric model to describe the trend in UK

Table 4.2: Finite sample rejection frequency for DGPs 3-4 (nominal level: 0.05)

			CD (I)			CD (II)		
DGP	n	T	$c = 0.5$	$c = 1$	$c = 1.5$	$c = 0.5$	$c = 1$	$c = 1.5$
3	25	25	0.294	0.486	0.650	0.128	0.184	0.336
		50	0.502	0.710	0.840	0.182	0.326	0.454
		100	0.938	0.996	0.998	0.580	0.888	0.980
	50	25	0.196	0.424	0.606	0.072	0.136	0.224
		50	0.700	0.936	0.982	0.268	0.496	0.654
		100	1.000	1.000	1.000	0.924	0.996	1.000
	100	25	0.456	0.806	0.938	0.162	0.336	0.494
		50	0.912	1.000	1.000	0.462	0.756	0.898
		100	1.000	1.000	1.000	0.910	0.998	1.000
4	25	25	0.288	0.530	0.730	0.124	0.206	0.344
		50	0.432	0.674	0.788	0.156	0.308	0.434
		100	0.790	0.948	0.988	0.348	0.656	0.816
	50	25	0.352	0.732	0.900	0.142	0.282	0.424
		50	0.802	0.962	0.988	0.336	0.586	0.776
		100	1.000	1.000	1.000	0.926	0.996	0.998
	100	25	0.334	0.712	0.884	0.126	0.234	0.384
		50	0.972	0.996	1.000	0.500	0.824	0.946
		100	1.000	1.000	1.000	0.926	0.996	1.000

Table 4.3: Finite sample rejection frequency for DGPs 5-6 (nominal level: 0.05)

				CD (I)			CD (II)		
DGP	n	T	γ_{nT}	$c = 0.5$	$c = 1$	$c = 1.5$	$c = 0.5$	$c = 1$	$c = 1.5$
5	25	25	0.1051	0.550	0.862	0.954	0.280	0.532	0.758
		50	0.0769	0.574	0.796	0.876	0.218	0.390	0.542
		100	0.0563	0.884	0.978	0.994	0.532	0.800	0.916
	50	25	0.0883	0.436	0.774	0.928	0.200	0.344	0.530
		50	0.0647	0.662	0.890	0.952	0.234	0.422	0.554
		100	0.0473	0.878	0.976	0.998	0.336	0.556	0.708
	100	25	0.0743	0.410	0.770	0.926	0.146	0.272	0.416
		50	0.0544	0.612	0.884	0.954	0.198	0.332	0.474
		100	0.0398	0.664	0.892	0.960	0.212	0.346	0.516
6	25	25	0.1051	0.570	0.896	0.956	0.288	0.574	0.796
		50	0.0769	0.494	0.764	0.876	0.192	0.354	0.538
		100	0.0563	0.878	0.976	0.994	0.386	0.408	0.770
	50	25	0.0883	0.488	0.836	0.936	0.178	0.366	0.544
		50	0.0647	0.702	0.914	0.980	0.232	0.416	0.580
		100	0.0473	0.886	0.976	0.996	0.352	0.622	0.796
	100	25	0.0743	0.350	0.702	0.902	0.130	0.276	0.422
		50	0.0544	0.640	0.924	0.976	0.282	0.468	0.624
		100	0.0398	0.722	0.918	0.962	0.290	0.472	0.662

regional temperatures and other weather outcomes over the last century, where a single common trend is assumed across all locations.[4] It is interesting to check whether such a common trend restriction is satisfied. To conserve space, in this application we investigate the pattern of climate change in the UK over the last 32 years. The data set contains monthly mean maximum temperature (in Celsius degrees, *Tmax* for short), mean minimum temperature (in Celsius degrees, *Tmin* for short), total rainfall (in millimeters, *Rain* for short) from 37 stations covering the UK (available from the UK Met Office at: www.metoce. gov.uk/climate/uk/stationdata). According to data availability we adopt a *balanced* panel data set that spans from October 1978 to July 2010 for 26 selected stations ($n = 26$, $T = 382$) to see if there exists a single common trend among these selected stations in *Tmax*, *Tmin*, and *Rain*, respectively. Note that the time span for our data set is much shorter than that in Atak, Linton and Xiao (2011).

For each series we consider a model of the following form

$$y_{it} = D_t'\beta + f_i\left(\frac{t}{T}\right) + \alpha_i + \varepsilon_{it},\ i = 1,...,26,\ T = 1,...,382,$$

where y_{it} is *Tmax*, *Tmin*, or *Rain* for station i at time t, $D_t \in \mathbb{R}^{11}$ is a 11-dimensional vector of monthly dummy variables, α_i is the fixed effect for station i, and the time trend function $f_i(\cdot)$ is unknown. We are interested in testing for $f_i = f$ for all $i = 1, 2, ..., n$.

To implement our test, the Epanechnikov kernel is used in both stages. We choose bandwidth h by the "leave-one-out" cross validation method and consider 10 different bandwidths of the form: $b = c\sqrt{\frac{1}{12}}T^{-1/5}$, where $c = 0.6,\ 0.7, ..., 1.5$. 10000 bootstrap resamples are used to construct the bootstrap distribution.

The results are reported in Table 4.4. From the table, we see that the p-values are smaller than 0.05 for *Tmax* and *Tmin* and larger than 0.1 for *Rain* for all choices of b. We can reject the null hypothesis of common trends at 5% level for both *Tmax*

[4] Atak, Linton, and Xiao (2011) study a model that allows for heterogenous effects of seasonal dummy variables and use different data sets than ours. Consequently, our result is not directly comparable with theirs.

Table 4.4: Bootstrap p-values for application to the U.K. climate data

Series (c)	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5
Tmax	0.0060	0.0101	0.0073	0.0078	0.0061	0.0074	0.0091	0.0110	0.0151	0.0235
Tmin	0.0142	0.0160	0.0153	0.0130	0.0097	0.0053	0.0038	0.0029	0.0024	0.0010
Rain	0.8726	0.8163	0.7365	0.6592	0.5915	0.5670	0.5731	0.5890	0.6265	0.6790

Note: bandwidth $b = c\sqrt{1/12}T^{-1/5}$ and bootstrap number $B = 10000$.

and *Tmin* but not for *Rain* even at 10% level.

OECD economic growth data

Economic growth has been a key issue in marcoeconomics over many decades. It is interesting to model the source of economic growth which incorporates a time trend. In this application we consider a model for the OECD economic growth data which incorporates a time trend. The data set consists of four economic variables from 16 OECD countries $(n = 16)$: Gross domestic product (GDP), Capital Stock (K), Labor input (L), and Human capital (H). We download GDP (at 2005 US$), Capital stock (at 2005 US$), and Labor input (Employment, at thousand persons) from http://www.datastream.com, and Human capital (Educational Attainment for Population Aged 25 and Over) from http://www.barrolee.com. The first three variables are seasonally adjusted quarterly data and span from 1975Q4 to 2010Q3 $(T = 140)$. For Human capital, we have only 5-years census data from the Barro-Lee dataset so that we have to use linear interpolation to obtain the quarterly observations.

We consider the following model for growth rates

$$\Delta \ln GDP_{it} = \beta_1 \Delta \ln L_{it} + \beta_2 \Delta \ln K_{it} + \beta_3 \Delta \ln H_{it} + f_i(t/T) + \alpha_i + \varepsilon_{it},$$

$i = 1, ..., 16$, $T = 1, ..., 140$, where α_i is the fixed effect, $f_i(\cdot)$ is unknown smooth time trends function for country i, and $\Delta \ln Z_{it} = \ln Z_{it} - \ln Z_{i,t-1}$ for $Z = GDP$, L, K, and H. We are interested in testing for common time trends for the 16 OECD countries.

The kernels, bandwidths, and number of bootstrap resamples are chosen as in

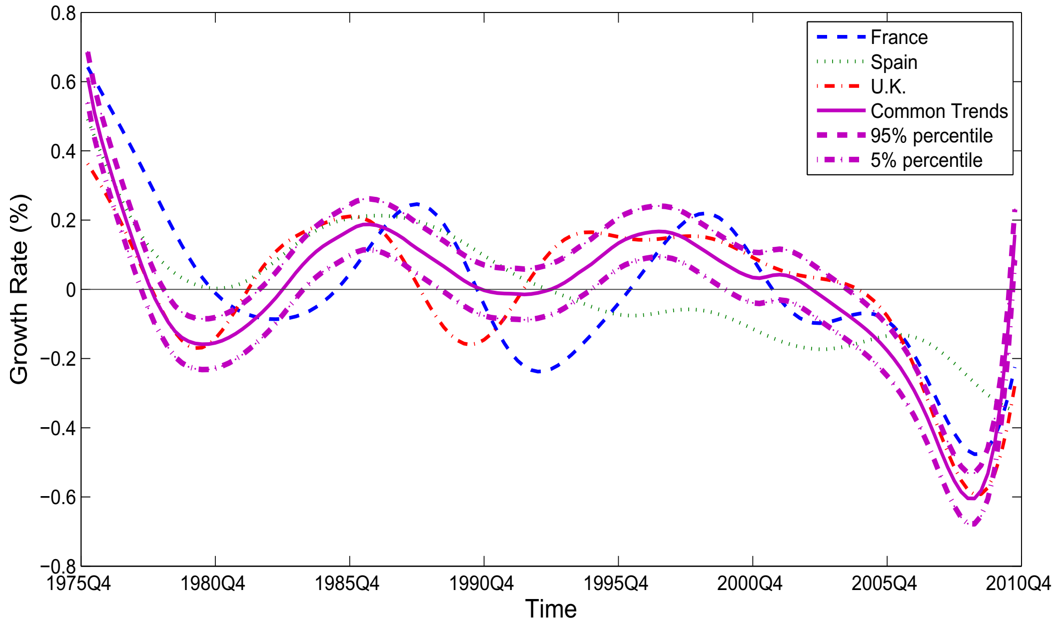

Figure 4.1: Trends in OECD real GDP growth rates from 1975Q4 to 2010Q3

the previous application. In Figure 4.1 we plot the estimated common trends (where we use the recentered trend: $\widehat{f}(\tau)-\int_0^1 \widehat{f}(\tau)\,d\tau$ for comparison) from the restricted semiparametric regression model together with its 90% pointwise confidence bands. Also plotted in Figure 4.1 are three representative individual trend functions for France, Spain, and the UK, which are estimated from the unrestricted semiparametric regression models. For the purpose of comparison, for the unconstrained model we impose the identification condition that the integral of each individual trend function over $(0,1)$ equals zero and use the Silverman rule-of-thumb to choose the bandwidth. Clearly, Figure 4.1 suggests that the estimated common trends function is significantly different from zero over a wide range its support. In addition, the trend functions for the three representative individual countries are obviously different from the estimated common trends, which implies that the widely used common trends assumption may not be plausible at all.

Table 4.5 reports the bootstrap p-values for our test of common trends. From the table, we can see that the p-values for all bandwidths are smaller than 0.1 for all bandwidths under investigation. Then we can reject the null hypothesis of common trends at the 10% level.

Table 4.5: Bootstrap p-values for application to OECD real GDP growth rate data

Series (c)	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5
$\Delta \ln GDP$	0.0001	0.0005	0.0020	0.0063	0.0141	0.0281	0.0336	0.0536	0.0645	0.0820

Note: bandwidth $b = c\sqrt{1/12}T^{-1/5}$ and bootstrap number $B = 10000$.

4.5 Concluding Remarks

In this chapter we propose a nonparametric test for common trends in semiparametric panel data models with fixed effects. We first estimate the restricted semiparametric model to obtain the augmented residuals and then run a local linear regression of the augmented residuals on the time trend for each cross sectional unit to obtain n nonparametric R^2 measures. We construct our test statistic by averaging these individual nonparametric R^2's, and show that after being appropriately centered and scaled, the statistic is asymptotically normally distributed under both the null hypothesis of common trends and a sequence of Pitman local alternatives. We also prove the consistency of the test and propose a bootstrap procedure to obtain the bootstrap p-values. Monte Carlo simulations and applications to both the UK climate change data and the OECD economic growth data are reported, both of which point to the empirical fragility of a common trend assumption.

Some extensions are possible. First, our semiparametric model in (4.1.1) only complements that in Atak, Linton, and Xiao (2011), and it is possible to allow the slope coefficients also to be heterogenous when we test for the null hypothesis of common trends for the nonparametric component. In this case, the profile least squares estimation of Su and Ullah (2006) and Chen, Gao, and Li (2010) and the nonparametric-R^2-based test lose much of their advantage and the heterogenous slope coefficients can only be estimated at a slower convergence rate. It seems straightforward to estimate the unrestricted model for each cross sectional unit to obtain the individual trend function estimates $\widehat{f}_i(\tau)$ and propose an L_2-distance-based test by averaging the squared L_2-distance between $\widehat{f}_i(\tau)$ and $\widehat{f}_j(\tau)$ for all $i \neq j$. It is also possible to test for the homogeneity of the slope coefficients and trend components jointly. Second, to derive the distribution theory of our test statistic, we allow for cross sectional dependence but rule out serial dependence. It is possible to allow the presence of both as in Bai (2009) by imposing some high-level assumptions. Nevertheless, the asymptotic variance of the non-normalized version of test statistic will become complicated and there seems no obvious way to estimate it consistently in order to implement our test in practice.

Chapter 5 Summary of Conclusions

In Chapter 2, we propose a nonparametric test for cross-sectional dependence in large dimensional panel. Our tests can be applied to both raw data and residuals from heterogenous nonparametric (or parametric) regressions. The requirement on the relative magnitude of n and T is quite weak in the former case, and very strong in the latter case in order to control the asymptotic effect of the parameter estimation error on the test statistic. In both cases, we establish the asymptotic normality of our test statistic under the null hypothesis of cross-sectional independence. The global consistency of our test is also established. Monte Carlo simulations indicate our test performs reasonably well in finite samples and has power in detecting cross-sectional dependence when the Pesaran's and CGL's tests fail.

In Chapter 3 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.

In Chapter 4 we propose a nonparametric test for common trends in semiparametric panel data models with fixed effects. We first estimate the restricted semiparametric model to obtain the augmented residuals and then run a local linear regression of the augmented residuals on the time trend for each cross sectional unit to obtain n nonparametric R^2 measures. We construct our test statistic by averaging these individual nonparametric R^2's, and show that after being appropriately centered and scaled, the statistic is asymptotically normally distributed under both the null hypothesis of common trends and a sequence of Pitman local alternatives. We also prove the consistency of the test and propose a bootstrap procedure to obtain the bootstrap p-values. Monte Carlo simulations and applications to both the UK climate change data and the OECD economic growth data are reported, both of which point to the empirical fragility of a common trend assumption.

References

Ahmad, I. A., and Li, Q., 1997. Testing independence by nonparametric kernel method. *Statistics & Probability Letter* 34, 201-210.

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

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

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

Anselin, L., 1998. *Spatial Econometrics: Methods and Models*. Dorddrecht: Kluwer Academic Publishers.

Anselin, L., 2001. Spatial econometrics. B. Baltagi (eds.), *A Companion to Theoretical Econometrics*. Blackwell, Oxford.

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

Atak, A., Linton, O., and Xiao, Z., 2011. A semiparametric panel data model for unbalanced data with application to climate change in the United Kingdom. *Journal of Econometrics* 164, 92-115.

Bacigál, T., 2005. Testing for common deterministic trends in geodetic data. *Journal of Electrical Engineering* 12, 1-5.

Baglan, D., 2010. 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.

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. and Kao, C., 2006. On the estimation inference of a panel cointegration model with cross-sectional dependence. In: Baltagi, Badi (Ed.), *Contributions to Economic Analysis*. Elsevier, pp. 3-30.

Bai, J. and Li K., 2012. Statistical analysis of factor models of high dimension. *Annals of Statistics* 40, 436-465.

Bai, J. and Ng, S., 2002. Determining the number of factors in approximate factor models. *Econometrica* 70, 191-221.

Bai, J. and Ng, S., 2004. A PANIC attack on unit roots and cointegration. *Econometrica* 72, 1127-1177.

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

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

Bai, J. and Ng, S., 2008. *Large Dimensional Factor Analysis*. Foundations and Trends in Econometrics, Vol. 3, No. 2, 89-163.

Bai, J. and Ng, S., 2010. Panel data unit root test with cross-section dependence: a further investigation. *Econometric Theory* 26, 1088-1114.

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

Banerjee, A., 1999. Panel data unit roots and cointegration: an overview. *Oxford Bulletin of Economics and Statistics* 61, 607-629.

Bernstein, D. S., 2005. *Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory*. Princeton University Press, Princeton.

Blundell, R., Chen, X., and Kristensen, D., 2007. Semi-Nonparametric IV estimation of shape invariant Engel curves. *Econometrica* 75, 1613-1670

Bond, S., Leblebicioglu, A., and Schiantarelli, F., 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.

Breitung, J. and Pesaran, M. H., 2008. Unit roots and cointegration in panels. In Matyas, L., and Sevestre, P. (eds.), *The Econometrics of Panel Data* (3rd edition). Springer.

Breusch, T. S. and Pagan, A. R., 1980. The Lagranger multiplier test and its application to model specifications in econometrics. *Review of Economic Studies* 47, 239-253.

Brock, W., Dechert, W., Scheinkman, J., and LeBaron, B., 1996. A test for independence based on the correlation dimension. *Econometric Reviews* 15. 197-235.

Cai, Z., 2007. Trending time-varying coefficients time series models with serially correlated errors. *Journal of Econometrics* 136, 163-188.

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

Chang, Y., 2002. Nonlinear IV unit root tests in panels with cross-sectional dependency. *Journal of Econometrics* 110, 261-292.

Chen, J., Gao, J., and Li, D., 2009. A new diagnostic test for cross-section uncorrelation in nonparametric panel data models. *Working paper Series* No. 0075, University of Adelaide.

Chen, J., Gao, J., and Li, D., 2010. Semiparametric trending panel data models with cross-sectional dependence. *Working paper*, University of Adelaide.

Chen, X., 2007. Large sample sieve estimation of semi-nonparametric models. In J. J. Heckman and 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., Hong, H., and Tamer, E., 2005. Measurement error models with auxiliary data. *Review of Economic Studies* 72, 343-366.

Chen, X., Racine, J., and Swanson, N. R., 2001. Semiparametric ARX neural-network models with an application to forecasting inflation. *Neural Networks, IEEE Transactions on* 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. and Levine R., 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. and Gijbels, I., 1996. *Local Polynomial Modelling and Its Applications*. Chapman & Hall, London.

Fan, J. and Li, R., 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* 96, 1348-1360.

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

Fan, J., Yao, Q., and Cai, Z., 2003. Adaptive varying-coefficient linear models. *Journal of the Royal Statistical Society*: *series B* 65, 57-81.

Fan, Y. and Li, Q., 1996. Consistent model specification tests: omitted variables and semiparametric functional forms. *Econometrica* 64, 865-890.

Fan, Y. and Li, Q., 1999. Central limit theorem for degenerate U-statistics of absolutely regular processed with applications to model specification testing. *Journal of Nonparametric Statistics* 10, 245-271.

Fomby, T. B. and Vogelsang, T. J., 2003. Tests of common deterministic trend slopes applied to quarterly temperature data. *Advances in Econometrics* 17, 29-43.

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

Frees, E. W., 1995. Assessing cross-sectional correlation in panel data. *Journal of Econometrics* 69, 393-414.

Freyberger, J. 2012. Nonparametric panel data models with interactive fixed effects. *Working Paper*, Northwestern University.

Gagliardini, P. and Gourieroux, C., 2011. Efficiency in large dynamic panel models with common factors. *Working Paper*, University of Lugano.

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

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

Gao, J. and Hawthorne, K., 2006. Semiparametric estimation and testing of the trend of temperature series. *Econometrics Journal* 9, 332-355.

Gourieroux, C., Monfort, A., Renault, E., and Trognon, A., 1987. Generalised residuals. *Journal of Econometrics* 34, 5-32.

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

Granger, C., Maasoumi, E., and Racine, J. S., 2004. A dependence metric for possibly nonlinear time series. *Journal of Time Series Analysis* 25, 649-669.

Hahn, J. and Kuersteiner, G., 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.

Hansen, B. E., 2008. Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory* 24, 726-748.

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

Härdle, W. and Mammen, E., 1993. Comparing nonparametric versus parametric regression fits. *Annals of Statistics* 21, 1926-1947.

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

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

Hoeffding, W., 1948. A nonparametric test of independence. *The Annals of Mathematical Statistics* 19, 546-557.

Hong, Y., 1998. Testing for Pairwise Serial Independence via the empirical distribution function. *Journal of the Royal Statistical Society: Series B* 60, part 2, 429-453.

Hong, Y., 2000. Generalized Spectral Tests for Serial Independence. *Journal of the Royal Statistical Society: Series B* 60, part 3, 557-574.

Hong, Y. and White, H., 1995. Consistent specification testing via nonparametric series regression. *Econometrica* 63, 1133-1160.

Hong, Y. and White, H., 2005. Asymptotic Distribution Theory for Nonparametric Entropy Measures of Serial Independence. *Econometrica* 73, 837-901.

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

Hsiao, C. and Li, Q., 2001. A consistent test for conditional heteroskedasticity in time series regression models. *Econometric Theory* 17, 188-221.

Hsiao, C., Pesaran, M. H., and Pick, A., 2009. Diagnostic tests of cross section independence for nonlinear panel data models. *IZA discussion paper* No. 2756.

Huang, H., Kao, C., and Urga, G., 2008. Copula-based tests for cross-sectional independence in panel models. *Economics Letters* 100, 24-228.

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

Johnson, D. and McClelland, R., 1998. A general dependence test and applications. *Journal of Applied Econometrics* 13, 627-644.

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

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

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

Lavergne, P., Vuong, Q., 2000. Nonparametric significance testing. *Econometric Theory* 16, 576-601.

Li, Q., 1999. Consistent model specification tests for time series econometric models. *Journal of Econometrics* 92, 101-147.

Li, Q. and Racine, J., 2007. *Nonparametric Econometrics: Theory and Practice*. Princeton University Press: Princeton.

Lee, A. J., 1990. *U-statistics: Theory and Practice*. Marcel Dekker, Inc.: New York and Basel.

Lee, Y., 2010. 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, D., Chen, J. and Gao, J., 2010. Nonparametric time-varying coefficient panel data models with fixed effects. Forthcoming in *Econometrics Journal*.

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

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

Masry, E., 1996a. Multivariate regression estimation: Local polynomial fitting for time series. *Stochastic processes and Their Application* 65, 81-101.

Masry, E., 1996b. Multivariate local polynomial regression for time series: Uniform strong consistency rates. *Journal of Time Series Analysis* 17, 571-599.

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

Meierrieks, D. and Gries, T., 2012. Economic performance and terrorist activity in Latin America. Forthcoming in *Defence and Peace Economics*.

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

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

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

Neumann, M. H. and Paparoditis, E., 2000. On bootstrapping L_2-type statistics in density testing. *Statistics and Probability Letters* 50, 137-147.

Neumeyer, N., 2009. Testing independence in nonparametric regression. *Journal of Multivariate Analysis* 100, 1551-1566.

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

Ng, S., 2006. Testing cross section correlation in panel data using spacing. *Journal of Business and Economic Statistics* 24, 12-23.

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

Pesaran, M. H., 2004. General diagnostic tests for cross section dependence in panels. Cambridge *Working Paper* in Economics No. 0435.

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

Pesaran, M. H., 2007. A simple panel unit root test in the presence of cross-section dependence. *Journal of Applied Econometrics* 22, 265-312.

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

Pesaran, M. H., Ullah, A. and Yamagata, T., 2008. A bias adjusted LM test of error cross section independence. *Econometrics Journal* 11, 105-127.

Phillips, P. C. B, 2001. Trending time series and macroeconomic activity: Some present and future challenges. *Journal of Econometrics* 100, 21-27.

Phillips, P. C. B. 2005. Challenges of trending time series econometrics. *Mathematics and Computers in Simulation* 68, 401-416.

Phillips, P. C. B. 2007. Regression with slowly varying regressors and nonlinear trends. *Econometric Theory* 23, 557-614.

Phillips, P. C. B., 2010. The mysteries of trend. *Macroeconomic Review* IX, 82-89.

Phillips, P. C. B. and Moon, H., 1999. Linear regression limit theory for nonstationary panel data. *Econometrica* 67, 1057-1111.

Phillips, P. C. B. and Sul, D., 2007. Transition modeling and econometric convergence tests. *Econometrica* 75, 1771-1855.

Phillips, P. C. B. and Sul D., 2009. Economic transition and growth. *Journal of Applied Econometrics* 24, 1153-1185.

Pinkse, J., 1998. A consistent nonparametric test for serial independence. *Journal of Econometrics* 84, 205-231.

Politis, D. and Romano, J. 1994. The stationary bootstrap. *Journal of the American Statistical Association* 89, 1303-1313.

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

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

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

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.

Robinson, P. M., 1988. Root-N-consistent semiparametric regression. *Econometrica* 56, 931-954.

Robinson, P. M., 1991. Consistent nonparametric entropy-based testing. *Review of Economic Studies* 58, 437-453.

Robinson, P. M., 2008. Correlation testing in time series, spatial and cross-sectional data. *Journal of Econometrics* 147, 5-16.

Robinson, P. M., 2010. Nonparametric trending regression with cross-sectional dependence. *Working paper*, LSE.

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

Sarafidis, V., Yamagata, T., and Robertson, D., 2009. A test of cross section dependence for a linear dynamic panel model with regressors. *Journal of Econometrics* 148, 149-161.

Seber, G., 2007. *A Matrix Handbook for Statisticians*. John Wiley & Sons, New Jersey.

Stock, J. H. and Watson, M. W., 1988. Testing for common trends. *Journal of the American Statistical Association* 83, 1097-1107.

Su, L. and Ullah, A., 2006. Profile likelihood estimation of partially linear panel data models with fixed effects. *Economics Letters* 92, 75-81.

Su, L. and Ullah, A., 2011. A nonparametric goodness-of-fit-based test for conditional heteroskedasticity. *Working paper*, Singapore Management University.

Su, L. and Chen, Q., 2013. Testing homogeneity in panel data models with interactive fixed effects. Forthcoming in *Econometric Theory.*

Su, L. and Jin, S., 2012. Sieve estimation of panel data models with cross section dependence. *Journal of Econometrics* 169, 34-47.

Su, L., Jin, S. and Zhang, Y., 2012. Specification test for panel data models with interactive fixed effects. *Working paper*, Singapore Management University.

Su, L. and Lu, X., 2013. Nonparametric dynamic panel data models: kernel estimating and specification testing. *Working paper*, Singapore Management University.

Su, L. and Ullah, A., 2011. Nonparametric and semiparametric panel econometric models: estimation and testing. In A. Ullah and D. E. A. Giles (eds), *Handbook of Empirical Economics and Finance*, 455-497. Taylor & Francis Group, New York.

Su, L. and Ullah, A., 2009. Testing conditional uncorrelatedness. *Journal of Business and Economic Statistics* 27, 18-29.

Su, L. and White, H., 2003. Testing conditional independence via empirical likelihood. *Discussion Paper*, Department of Economics, UCSD.

Su, L. and White, H., 2007. Consistent characteristic function-based test for conditional independence. *Journal of Econometrics* 141, 807-834.

Su, L. and White, H., 2008, Nonparametric Hellinger metric test for conditional independence. *Econometric Theory* 24, 829-864.

Sun, S. and Chiang, C-Y., 1997. Limiting behavior of the perturbed empirical distribution functions evaluated at U-statistics for strongly mixing sequences of random variables. *Journal of Applied Mathematics and Stochastic Analysis* 10, 3-20.

Sun, Y., 2011. Robust trend inference with series variance estimator and testing-optimal smoothing parameter. *Journal of Econometrics* 164, 345-366.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society*: *Series B* 58, 267-288

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

Tjøstheim, D., 1996. Measures and tests of independence: a survey. *Statistics* 28, 249-284.

Vogelsang, T. J. and Franses, P. H., 2005. Testing for common deterministic trend slopes. Journal of Econometrics 126, 1-24.

White, H., 2001. *Asymptotic Theory for Econometricians*. 2nd Ed., Academic Press, San Diego.

White, H. and Granger, C., 2011. Consideration of trends in time series. *Journal of Time Series Econometrics* 3, Iss. 1, Article 2.

Xu, K.-L., 2011. Robustifying multivariate trend tests to nonstationary volatility. Forthcoming in *Journal of Econometrics*.

Yoshihara, K., 1976. Limit behavior of U-statistics for stationary, absolutely regular processes. *Zeitschrift für Wahrscheinlichskeitstheorie und Verwandte Gebiete* 35, 237-252.

Yoshihara, K., 1989. Limiting behavior of generalized quadratic forms generated by absolutely regular processes. In P. Mandl and M. Hušková (eds.), *Proceedings of the Fourth Prague Symposium on Asymptotic Statistics*, pp. 539-547. Charles University Press.

Zafaroni, P., 2010. Generalized least squares estimation of panel with common shocks. *Working paper*, School of Imperial College London.

Appendix

A Proofs in Chapter 2

Throughout this appendix, we use C to signify a generic constant whose exact value may vary from case to case. Recall $P_T^l \equiv T!/(T-l)!$ and $C_T^l \equiv T!/[(T-l)!l!]$ for integers $l \leq T$.

Proof of Theorem 2.3.1

Recall $\varphi_{i,ts} \equiv \overline{k}_{h,ts}^{i} - E_t[\overline{k}_{h,ts}^{i}] - E_s[\overline{k}_{h,ts}^{i}] + E_t E_s[\overline{k}_{h,ts}^{i}]$ where $\overline{k}_{h,ts}^{i} \equiv \overline{k}_h(u_{it} - u_{is})$ and E_s denotes expectation taken only with respect to variables indexed by time s, that is, $E_s(\overline{k}_{h,ts}^{i}) \equiv \int \overline{k}_h(u_{it} - u) f_i(u) du$. Let $c_{i,ts} \equiv E(\varphi_{i,ts})$, and $c_{ts} \equiv (n-1)^{-1} \sum_{i=1}^{n} c_{i,ts}$. We will frequently use the fact that for $t \neq s$,

$$c_{i,ts} \leq Ch^{-\frac{\delta}{1+\delta}} \alpha_i^{\frac{\delta}{1+\delta}}(|t-s|) \qquad (.0.1)$$

as by the law of iterated expectations, the triangle inequality, and Lemma .0.6, we have $|c_{i,ts}|$ $= |E[\overline{k}_{h,ts}^{i}] - E_t E_s[\overline{k}_{h,ts}^{i}]| = |E\{E[\overline{k}_{h,ts}^{i}|u_{it}] - E_s[\overline{k}_{h,ts}^{i}]\}| \leq E|E[\overline{k}_{h,ts}^{i}|u_{it}] - E_s[\overline{k}_{h,ts}^{i}]| \leq Ch^{-\frac{\delta}{1+\delta}}$ $\times \alpha_i^{\frac{\delta}{1+\delta}}(|t-s|)$. Let $\alpha(j) \equiv \max_{1 \leq i \leq n} \alpha_i(j)$. Let $m \equiv \lfloor L \log T \rfloor$ (the integer part of $L \log T$) where L is a large positive constant so that the conditions on m in Assumption A.1(i*) are all met by Assumption A.1(i). In addition, it is obvious that $\sum_{\tau=1}^{\infty} \alpha^{\frac{\delta}{1+\delta}}(\tau) = O(1)$ under Assumption A.1(i).

Let $Z_{ij,t} \equiv (u_{it}, u_{jt})$ and $\varsigma_{ij,tsrq} \equiv \varsigma(Z_{ij,t}, Z_{ij,s}, Z_{ij,r}, Z_{ij,q}) = \overline{k}_{h,ts}^{i}(\overline{k}_{h,ts}^{j} + \overline{k}_{h,rq}^{j} - 2\overline{k}_{h,tr}^{j})$. Let $\overline{\varsigma}_{ij,tsrq} \equiv \overline{\varsigma}(Z_{ij,t}, Z_{ij,s}, Z_{ij,r}, Z_{ij,q}) \equiv \frac{1}{4!} \sum_{4!} \varsigma_{ij,tsrq}$, where $\sum_{4!}$ denotes summation over all 4! different permutations of (t,s,r,q). That is, $\overline{\varsigma}_{ij,tsrq}$ is a symmetric version of $\varsigma_{ij,tsrq}$ by symmetrizing over the four time indices and it is easy to verify that

$$
\begin{aligned}
\bar{\varsigma}_{ij,tsrq} \;=\; & \frac{1}{12}\{\bar{k}^{i}_{h,ts}(2\bar{k}^{j}_{h,ts}+2\bar{k}^{j}_{h,rq}-\bar{k}^{j}_{h,tr}-\bar{k}^{j}_{h,sr}-\bar{k}^{j}_{h,tq}-\bar{k}^{j}_{h,sq}) \\
& +\bar{k}^{i}_{h,tr}(2\bar{k}^{j}_{h,tr}+2\bar{k}^{j}_{h,qs}-\bar{k}^{j}_{h,ts}-\bar{k}^{j}_{h,sr}-\bar{k}^{j}_{h,tq}-\bar{k}^{j}_{h,rq}) \\
& +\bar{k}^{i}_{h,tq}(2\bar{k}^{j}_{h,tq}+2\bar{k}^{j}_{h,sr}-\bar{k}^{j}_{h,tr}-\bar{k}^{j}_{h,qr}-\bar{k}^{j}_{h,ts}-\bar{k}^{j}_{h,sq}) \\
& +\bar{k}^{i}_{h,sr}(2\bar{k}^{j}_{h,sr}+2\bar{k}^{j}_{h,qt}-\bar{k}^{j}_{h,st}-\bar{k}^{j}_{h,rt}-\bar{k}^{j}_{h,sq}-\bar{k}^{j}_{h,rq}) \\
& +\bar{k}^{i}_{h,sq}(2\bar{k}^{j}_{h,sq}+2\bar{k}^{j}_{h,rt}-\bar{k}^{j}_{h,st}-\bar{k}^{j}_{h,qt}-\bar{k}^{j}_{h,sr}-\bar{k}^{j}_{h,qr}) \\
& +\bar{k}^{i}_{h,rq}(2\bar{k}^{j}_{h,rq}+2\bar{k}^{j}_{h,st}-\bar{k}^{j}_{h,rt}-\bar{k}^{j}_{h,qt}-\bar{k}^{j}_{h,rs}-\bar{k}^{j}_{h,qs})\}. \qquad (.0.2)
\end{aligned}
$$

Then we can write $\widehat{\Gamma}_{nT}$ as

$$
\begin{aligned}
\widehat{\Gamma}_{nT} &= \frac{1}{n(n-1)}\sum_{1\le i\ne j\le n}\frac{1}{P_T^4}\sum_{1\le t\ne s\ne r\ne q\le T}\varsigma_{ij,tsrq} \\
&= \frac{1}{n(n-1)}\sum_{1\le i\ne j\le n}\frac{1}{C_T^4}\sum_{1\le t_1<t_2<t_3<t_4\le T}\bar{\varsigma}_{ij,t_1t_2t_3t_4}. \qquad (.0.3)
\end{aligned}
$$

Let $\theta_{ij}=E_1E_2E_3E_4\left[\bar{\varsigma}\left(Z_{ij,1},Z_{ij,2},Z_{ij,3},Z_{ij,4}\right)\right]$ and $\bar{\varsigma}_{ij,c}\left(z_1,\ldots,z_c\right)=E_{c+1}\cdots E_4[\bar{\varsigma}(z_1,\ldots,,$ $z_c,Z_{ij,c+1},\ldots,Z_{ij,4})]$ for nonrandom $z_1,\ldots,z_c$ and $c=1,2,3,4$. Let $\vartheta_{ij}^{(1)}\left(z_1\right)=\bar{\varsigma}_{ij,1}\left(z_1\right)-\theta_{ij}$ and $\vartheta_{ij}^{(c)}\left(z_1,\ldots,z_c\right)=\bar{\varsigma}_{ij,c}\left(z_1,\ldots,z_c\right)-\sum_{k=1}^{c-1}\sum_{(c,k)}\vartheta_{ij}^{(k)}(z_{t_1},\ldots,z_{t_k})-\theta_{ij}$ for $c=2,3,4$, where the sum $\sum_{(c,k)}$ is taken over all subsets $1\le t_1<\cdots<t_k\le c$ of $\{1,2,\ldots,c\}$. It is easy to verify that $\theta_{ij}=0$, $\vartheta_{ij}^{(1)}\left(Z_{ij,t}\right)=0$, and

$$
\vartheta_{ij}^{(2)}\left(Z_{ij,t},Z_{ij,s}\right)=\bar{\varsigma}_{ij,2}\left(Z_{ij,t},Z_{ij,s}\right)=\frac{1}{6}\varphi_{i,ts}\varphi_{j,ts}. \qquad (.0.4)
$$

Similarly, straightforward but tedious calculations show that

$$
\begin{aligned}
& \vartheta_{ij}^{(3)}\left(Z_{ij,t},Z_{ij,s},Z_{ij,r}\right) \\
= \; & \bar{\varsigma}_{ij,3}\left(Z_{ij,t},Z_{ij,s},Z_{ij,r}\right)-\bar{\varsigma}_{ij,2}\left(Z_{ij,t},Z_{ij,s}\right)-\bar{\varsigma}_{ij,2}\left(Z_{ij,t},Z_{ij,r}\right)-\bar{\varsigma}_{ij,2}\left(Z_{ij,s},Z_{ij,r}\right) \\
= \; & -\frac{1}{12}\left[\varphi_{i,ts}\left(\varphi_{j,tr}+\varphi_{j,sr}\right)+\varphi_{i,tr}\left(\varphi_{j,ts}+\varphi_{j,sr}\right)+\varphi_{i,sr}\left(\varphi_{j,st}+\varphi_{j,rt}\right)\right] \qquad (.0.5)
\end{aligned}
$$

and

$$
\begin{aligned}
& \vartheta_{ij}^{(4)}\left(Z_{ij,t}, Z_{ij,s}, Z_{ij,r}, Z_{ij,q}\right) \\
= \; & \overline{\varsigma}\left(Z_{ij,t}, Z_{ij,s}, Z_{ij,r}, Z_{ij,q}\right) - \bar{\varsigma}_{ij,2}\left(Z_{ij,t}, Z_{ij,s}\right) - \bar{\varsigma}_{ij,2}\left(Z_{ij,t}, Z_{ij,r}\right) - \bar{\varsigma}_{ij,2}\left(Z_{ij,t}, Z_{ij,q}\right) \\
& - \bar{\varsigma}_{ij,2}\left(Z_{ij,s}, Z_{ij,r}\right) - \bar{\varsigma}_{ij,2}\left(Z_{ij,s}, Z_{ij,q}\right) - \bar{\varsigma}_{ij,2}\left(Z_{ij,r}, Z_{ij,q}\right) - \bar{\varsigma}_{ij,3}\left(Z_{ij,t}, Z_{ij,s}, Z_{ij,r}\right) \\
& - \bar{\varsigma}_{ij,3}\left(Z_{ij,t}, Z_{ij,s}, Z_{ij,q}\right) - \bar{\varsigma}_{ij,3}\left(Z_{ij,t}, Z_{ij,r}, Z_{ij,q}\right) - \bar{\varsigma}_{ij,3}\left(Z_{ij,s}, Z_{ij,r}, Z_{ij,q}\right) \\
= \; & \frac{1}{6}\left\{\varphi_{i,ts}\varphi_{j,rq} + \varphi_{i,tr}\varphi_{j,sq} + \varphi_{i,rq}\varphi_{j,ts} + \varphi_{i,sq}\varphi_{j,tr} + \varphi_{i,tq}\varphi_{j,sr} + \varphi_{i,sr}\varphi_{j,tq}\right\}, \quad (.0.6)
\end{aligned}
$$

where (.0.5) and (.0.6) will be needed in the proofs of Propositions .0.2 and .0.3, respectively.

Let $G_{nT}^{(k)} \equiv \frac{1}{n(n-1)P_T^k}\sum_{1\le i\neq j\le n}\sum_{(T,k)}\vartheta_{ij}^{(k)}\left(Z_{ij,t_1},\ldots,Z_{ij,t_k}\right)$ for $k = 1,2,3,4$, where $\sum_{(T,k)}$ denotes summation over all P_T^k permutations $(t_1, ..., t_k)$ of distinct integers chosen from $\{1,2,...,T\}$ (See Lee (1990), Ch 1). Then by the Hoeffding decomposition, we have

$$\widehat{\Gamma}_{nT} = 6G_{nT}^{(2)} + 4G_{nT}^{(3)} + G_{nT}^{(4)}. \tag{.0.7}$$

Let $\overline{\Gamma}_{nT} \equiv 6G_{nT}^{(2)}$. Noting that $nThE(\overline{\Gamma}_{nT}) = \frac{2h}{(n-1)(T-1)}\sum_{1\le i\neq j\le n}\sum_{1\le t<s\le T}E\left[\varphi_{i,ts}\varphi_{j,ts}\right] = B_{nT}$ under H_0, we complete the proof of the theorem by showing that: (i) $nTh[\overline{\Gamma}_{nT} - E\left(\overline{\Gamma}_{nT}\right)] \xrightarrow{d} N\left(0,\sigma_0^2\right)$, (ii) $nThG_{nT}^{(3)} = o_P(1)$, and (iii) $nThG_{nT}^{(4)} = o_P(1)$. These results are established respectively in Propositions .0.1, .0.2, and .0.3 below.

Proposition .0.1 $nTh\left[\overline{\Gamma}_{nT} - E\left(\overline{\Gamma}_{nT}\right)\right] \xrightarrow{d} N\left(0,\sigma_0^2\right)$.

Proof. Let $\varphi_{i,ts}^c \equiv \varphi_{i,ts} - E(\varphi_{i,ts})$. Then we have $\overline{\Gamma}_{nT} - E(\overline{\Gamma}_{nT}) = \overline{\Gamma}_{nT,1} + \overline{\Gamma}_{nT,2}$, where

$$
\begin{aligned}
\overline{\Gamma}_{nT,1} &\equiv \frac{2}{n(n-1)}\sum_{1\le i<j\le n}\frac{1}{C_T^2}\sum_{1\le t<s\le T}\varphi_{i,ts}^c\varphi_{j,ts}^c, \text{ and} \\
\overline{\Gamma}_{nT,2} &\equiv \frac{1}{n(n-1)}\sum_{1\le i\neq j\le n}\frac{1}{C_T^2}\sum_{1\le t<s\le T}\left\{\varphi_{i,ts}^c E\left[\varphi_{j,ts}\right] + \varphi_{j,ts}^c E\left[\varphi_{i,ts}\right]\right\}.
\end{aligned}
$$

We prove the proposition by showing that

$$nTh\overline{\Gamma}_{nT,1} = \frac{nT}{(n-1)(T-1)}W_{nT} \xrightarrow{d} N\left(0,\sigma_0^2\right), \tag{.0.8}$$

and

$$nTh\overline{\Gamma}_{nT,2} = o_P(1), \tag{.0.9}$$

where $W_{nT} \equiv \sum_{1\le i<j\le n} w_{ij}$, $w_{ij} \equiv w_{nT,ij} \equiv w_{nT}\left(\mathbf{u}_i, \mathbf{u}_j\right) \equiv \frac{4h}{nT}\sum_{1\le t<s\le T}\varphi_{i,ts}^c\varphi_{j,ts}^c$, and $\mathbf{u}_i \equiv (u_{i1}, ..., u_{iT})'$. Noting that $nT/[(n-1)(T-1)] \to 1$, the proof is completed by Lemmas

.0.1-.0.2 below. ■

Lemma .0.1 $W_{nT} \xrightarrow{d} N\left(0, \sigma_0^2\right)$ *under* H_0.

Proof. W_{nT} is a second order degenerate U-statistic that is "clean" (i.e., $E\left[w_{nT}\left(\mathbf{u}_i, \mathbf{u}_j\right)|\mathbf{u}_i\right]$ $= E\left[w_{nT}\left(\mathbf{u}_i, \mathbf{u}_j\right)|\mathbf{u}_j\right] = 0$ for $i \neq j$) under H_0, we can apply Proposition 3.2 of de Jong (1987) to prove (.0.8) by showing that

$$\begin{aligned}
\overline{\sigma}_{nT}^2 &\equiv \operatorname{Var}\left(W_{nT}\right) = \sigma_{nT}^2 + o\left(1\right), & (.0.10)\\
G_I &\equiv \sum_{1\leq i<j\leq n} E\left[w_{ij}^4\right] = o\left(1\right), & (.0.11)\\
G_{II} &\equiv \sum_{1\leq i<j<k\leq n} E\left[w_{ij}^2 w_{ik}^2 + w_{ji}^2 w_{jk}^2 + w_{ki}^2 w_{kj}^2\right] = o\left(1\right), & (.0.12)\\
G_{IV} &\equiv \sum_{1\leq i<j<k<l\leq n} E\left[w_{ij}w_{ik}w_{lj}w_{lk} + w_{ij}w_{il}w_{kj}w_{kl} + w_{ik}w_{il}w_{jk}w_{jl}\right] = o\left(1\right). & (.0.13)
\end{aligned}$$

Step 1. Proof of (.0.10). First, notice that

$$\begin{aligned}
\overline{\sigma}_{nT}^2 &= \frac{16h^2}{n^2T^2}\operatorname{Var}\left(\sum_{1\leq i<j\leq n}\sum_{1\leq t<s\leq T}\varphi_{i,ts}^c\varphi_{j,ts}^c\right)\\
&= \frac{16h^2}{n^2T^2}\sum_{1\leq i<j\leq n}\sum_{1\leq t_1<t_2\leq T,\ 1\leq t_3<t_4\leq T} E\left[\varphi_{i,t_1t_2}^c\varphi_{i,t_3t_4}^c\right]E\left[\varphi_{j,t_1t_2}^c\varphi_{j,t_3t_4}^c\right].
\end{aligned}$$

We consider three cases for the summation in the last expression: the number of distinct indices in $\{t_1, t_2, t_3, t_4\}$ are 4, 3, and 2, respectively, and use (a), (b), and (c) to denote these three cases in order. In cases (a)-(b), we can apply similar arguments to those used in the proof of (.0.11) below and demonstrate the corresponding sum is $o\left(1\right)$. It follows that

$$\overline{\sigma}_{nT}^2 = \frac{16h^2}{n^2T^2}\sum_{1\leq i<j\leq n}\sum_{1\leq t<s\leq T}\operatorname{Var}\left(\varphi_{i,ts}^c\right)\operatorname{Var}\left(\varphi_{j,ts}^c\right) + o\left(1\right) = \sigma_{nT}^2 + o\left(1\right).$$

Step 2. Proof of (.0.11). We prove a stronger result: $G_I = o(n^{-1})$ by showing that $\max_{1\leq i\neq j\leq n} G_{ijI} = o(n^{-3})$ where $G_{ijI} \equiv E(w_{ij}^4)$. For $i \neq j$, we have that under H_0,

$$G_{ijI} = \frac{256h^4}{n^4T^4}\sum_{1\leq t_{2k-1}<t_{2k}\leq T,\ k=1,2,3,4} E\left[\prod_{l=1}^4\varphi_{i,t_{2l-1}t_{2l}}^c\right]E\left[\prod_{l=1}^4\varphi_{j,t_{2l-1}t_{2l}}^c\right].$$

We consider five cases inside the summation: the number of distinct elements in $\{t_1, t_2, ..., t_8\}$ are 8, 7, 6, 5, and 4 or less. We use (A), (B), (C), (D), and (E) to denote these five cases, respectively, and denote the corresponding sum in G_{ijI} as $G_{ijI,A}$, $G_{ijI,B}$, $G_{ijI,C}$, $G_{ijI,D}$, and $G_{ijI,E}$, respectively (e.g., $G_{ijI,A}$ is defined as G_{ijI} but with the time indices restricted to case (A)).

For case (A), we consider two different subcases: (Aa) there exists $k_0 \in \{1,...,8\}$ such that, $|t_l - t_{k_0}| > m$ for all $l \neq k_0$; (Ab) all the other remaining cases. We use $G_{ijI,Aa}$ and $G_{ijI,Ab}$ to denote $G_{ijI,A}$ but with the time indices restricted to subcases (Aa) and (Ab), respectively. Let $1 \leq r_1 < ... < r_8 \leq T$ be the permutation of $t_1,...,t_8$ in ascending order. Denote $A_i(r_1,...,r_8) \equiv \prod_{l=1}^4 \varphi^c_{i,t_{2l-1}t_{2l}}$. Then it is easy to see that $|E[A_j(r_1,...,r_8)]| \leq C$ uniformly in j.

For subcase (Aa), without loss of generality (WLOG) we assume $t_{k_0} = t_1$. We consider two subsubcases: (Aa1) $t_1 = r_1$, (Aa2) $t_1 = r_{l_0}$ for $l_0 \in \{2,...,7\}$. In subsubcase (Aa1), by splitting variables indexed by t_1 from those indexed by $t_2,\ldots,t_8$, we have by Lemma .0.5 that

$$|E[A_i(r_1,...,r_8)]| \leq \left|E\left\{E_{t_1}\left(\varphi^c_{i,t_1t_2}\right)\varphi^c_{i,t_3t_4}\varphi^c_{i,t_5t_6}\varphi^c_{i,t_7t_8}\right\}\right| + Ch^{-\frac{4\delta}{1+\delta}}\alpha^{\frac{\delta}{1+\delta}}(m).$$

To bound the first term in the last expression, we apply Lemma .0.6 to obtain

$$\begin{aligned} \left|E_{t_1}\left(\varphi^c_{i,t_1t_2}\right)\right| &= \left|E_{t_1}E_{t_2}(\overline{k}^i_{h,t_1t_2}) - E(\overline{k}^i_{h,t_1t_2})\right| = \left|E[E_{t_2}(\overline{k}^i_{h,t_1t_2}) - E(\overline{k}^i_{h,t_1t_2}|u_{it_1})]\right| \\ &\leq E\left|E_{t_2}(\overline{k}^i_{h,t_1t_2}) - E(\overline{k}^i_{h,t_1t_2}|u_{it_1})\right| \leq Ch^{-\frac{\delta}{1+\delta}}\alpha^{\frac{\delta}{1+\delta}}(m). \end{aligned} \quad (.0.14)$$

Consequently, we have $|A_i(t_1,...,t_8)| \leq Ch^{-\frac{4\delta}{1+\delta}}\alpha^{\frac{\delta}{1+\delta}}(m)$. In subsubcase (Aa2), noting that $t_2 \in \{r_{l_0+1},...,r_8\}$ we split first variables indexed by $r_1,...,r_{l_0-1}$ from others and then variables indexed by $r_{l_0}(=t_1)$ from $\{r_{l_0+1},...,r_8\}$ to obtain

$$\begin{aligned} |E[A_i(r_1,...,r_8)]| &\leq |E\{E_{1,...,l_0-1}[A_i(r_1,...,r_8)]\}| + Ch^{-\frac{4\delta}{1+\delta}}\alpha^{\frac{\delta}{1+\delta}}(m) \\ &\leq |E[E_{t_1}\{E_{1,...,l_0-1}[A_i(r_1,...,r_8)]\}]| \\ &\quad + Ch^{-\frac{3\delta}{1+\delta}}\alpha^{\frac{\delta}{1+\delta}}(m) + Ch^{-\frac{4\delta}{1+\delta}}\alpha^{\frac{\delta}{1+\delta}}(m). \end{aligned}$$

Now we can apply Fubini theorem and (.0.14) to bound the first term in the last expression by $Ch^{-\frac{\delta}{1+\delta}}\alpha^{\frac{\delta}{1+\delta}}(m)$. Consequently, we have $|E[A_i(r_1,...,r_8)]| \leq Ch^{-\frac{4\delta}{1+\delta}}\alpha^{\frac{\delta}{1+\delta}}(m)$ uniformly i in case (Aa). It follows that

$$G_{ijI,Aa} \leq \frac{Ch^4}{n^4T^4}T^8h^{-\frac{4\delta}{1+\delta}}\alpha^{\frac{\delta}{1+\delta}}(m) = O\left(n^{-4}T^4h^{\frac{4}{1+\delta}}\alpha^{\frac{\delta}{1+\delta}}(m)\right) = o\left(n^{-3}\right), \quad (.0.15)$$

where here and below $o\left(n^{-3}\right)$ holds uniformly in (i,j). In case (Ab), the number of terms in the summation for $G_{ijI,Ab}$ is of order $O\left(T^4m^4\right)$ and each term is uniformly bounded by a constant C. It follows that

$$G_{ijI,Ab} \leq \frac{Ch^4}{n^4T^4}T^4m^4 = O\left(n^{-4}h^4m^4\right) = o\left(n^{-3}\right). \quad (.0.16)$$

Now, we consider case (B). WLOG we assume $t_8 = t_6$ and consider two subcases for the indices $\{t_1,...,t_7\}$: (Ba) there exist two distinct integers $k_1,k_2 \in \{1,...,7\}$ such that

$|t_l - t_{k_s}| > m$ for all $l \neq k_s$ and $s = 1, 2$; (Bb) all the other remaining cases. We use $G_{ijI,Ba}$ and $G_{ijI,Bb}$ to denote $G_{ijI,B}$ but with the time indices restricted to subcases (Ba) and (Bb), respectively. In case (Ba), at least one (say t_{k_1}) of the two time indices satisfying the condition in (Ba) is not t_6 so that we can apply the same argument as used in case (Aa) to obtain the bound for $G_{I,Ba}$ as

$$G_{ijI,Ba} \leq \frac{Ch^4}{n^4T^4}T^7h^{-\frac{4\delta}{1+\delta}}\alpha^{\frac{\delta}{1+\delta}}(m) = O\left(n^{-4}T^3h^{\frac{4}{1+\delta}}\alpha^{\frac{\delta}{1+\delta}}(m)\right) = o\left(n^{-3}\right). \qquad (.0.17)$$

In case (Bb), the number of terms in the summation for $G_{ijI,Bb}$ is of order $O\left(T^4m^3\right)$ and each term is uniformly bounded by a constant C. It follows that

$$G_{I,Bb} \leq \frac{Ch^4}{n^4T^4}T^4m^3 = O\left(n^{-4}h^4m^3\right) = o\left(n^{-3}\right). \qquad (.0.18)$$

For case (C), we consider two subcases for the indices $\{t_1, ..., t_8\}$: (Ca) there exists four distinct integers $k_1, k_2, k_3, k_4 \in \{1, ..., 8\}$ such that $|t_l - t_{k_s}| > m$ for all $l \neq k_s$ and $s = 1, 2, 3, 4$ (note that some of the t_l indices coincide here so that the total number of distinct indices among $\{t_1, ..., t_8\}$ is six); (Cb) all the other remaining cases. We use $G_{ijI,Ca}$ and $G_{ijI,Cb}$ to denote $G_{I,C}$ but with the time indices restricted to subcases (Ca) and (Cb), respectively. In case (Ca) we can follow the same arguments as used in case (Aa) to bound $G_{ijI,Ca}$ as

$$G_{ijI,Ca} \leq \frac{Ch^4}{n^4T^4}T^6h^{-\frac{2+4\delta}{1+\delta}}\alpha^{\frac{\delta}{1+\delta}}(m) = O\left(n^{-4}T^2h^{\frac{2}{1+\delta}}\alpha^{\frac{\delta}{1+\delta}}(m)\right) = o\left(n^{-3}\right). \qquad (.0.19)$$

In case (Cb), the number of terms in the summation for $G_{ijI,Cb}$ is of order $O\left(T^4m^2\right)$ and each term is uniformly bounded by a constant Ch^{-2}. It follows that

$$G_{ijI,Cb} \leq \frac{Ch^4}{n^4T^4}T^4m^2h^{-2} = O\left(n^{-4}h^2m^2\right) = o\left(n^{-3}\right). \qquad (.0.20)$$

For case (D), we consider two subcases for the indices $\{t_1, ..., t_8\}$: (Da) for all distinct integers $k \in \{1, ..., 8\}$ such that $|t_l - t_k| > m$ for all $l \neq k$ with $t_l \neq t_k$; (Db) all the other remaining cases. We use $G_{ijI,Da}$ and $G_{ijI,Db}$ to denote $G_{ijI,D}$ but with the time indices restricted to subcases (Da) and (Db), respectively. In case (Da) we can follow the same arguments used in cases (Ca), (Ba), and (Aa) to bound $G_{ijI,Da}$ as

$$G_{ijI,Da} \leq \frac{Ch^4}{n^4T^4}T^5h^{-\frac{2+4\delta}{1+\delta}}\alpha^{\frac{\delta}{1+\delta}}(m) = O\left(n^{-4}Th^{\frac{2}{1+\delta}}\alpha^{\frac{\delta}{1+\delta}}(m)\right) = o\left(n^{-3}\right). \qquad (.0.21)$$

In case (Db), the number of terms in the summation for $G_{ijI,Db}$ is of order $O\left(T^4m\right)$ and each term is uniformly bounded by Ch^{-2}. It follows that

$$G_{ijI,Db} \leq \frac{Ch^4}{n^4T^4}T^4mh^{-2} = O\left(n^{-4}h^2m\right) = o\left(n^{-3}\right). \qquad (.0.22)$$

In case (E), it is straightforward to bound $G_{ijI,E}$ as

$$G_{ijI,E} \leq \frac{Ch^4}{n^4T^4}\left(T^4h^{-4}+T^3h^{-4}+T^2h^{-6}\right) = O\left(n^{-4}+n^{-4}T^{-2}h^{-2}\right) = o\left(n^{-3}\right). \quad (.0.23)$$

In sum, combining (.0.15)-(.0.23) yields

$$\max_{1\leq i\neq j\leq n} G_{ijI} = o\left(n^{-3}\right). \quad (.0.24)$$

Step 3. Proof of (.0.12). By the Jensen inequality and (.0.24), $G_{II} \leq \sum_{1\leq i<j<k\leq n}[\{E(w_{ij}^4) \times E(w_{ik}^4)\}^{1/2} + \{E(w_{ji}^4)E(w_{jk}^4)\}^{1/2} + \{E(w_{ki}^4)E(w_{kj}^4)\}^{1/2}] \leq \frac{n^3}{2}\max_{1\leq i\neq j\leq n} E(w_{ij}^4) = o(1)$.

Step 4. Proof of (.0.13). Write $G_{IV} = \sum\limits_{1\leq i<j<k<l\leq n}\{E\left[w_{ij}w_{ik}w_{lj}w_{lk}\right] + E\left[w_{ij}w_{il}w_{kj}w_{kl}\right] + E[w_{ik}\ w_{il}w_{jk}w_{jl}]\} \equiv G_{IV1} + G_{IV2} + G_{IV3}$. Recalling $w_{ij} \equiv \frac{4h}{nT}\sum_{1\leq t<s\leq T}\varphi^c_{i,ts}\varphi^c_{j,ts}$,

$$\begin{aligned} G_{IV1} &= \sum_{1\leq i_1<i_2<i_3<i_4\leq n} E\left[w_{i_1i_2}w_{i_1i_3}w_{i_4i_2}w_{i_4i_3}\right] \\ &= \frac{256h^4}{n^4T^4}\sum_{1\leq i_1<i_2<i_3<i_4\leq n}\ \sum_{1\leq t_{2k-1}<t_{2k}\leq T,\ k=1,2,3,4} E\left[\varphi^c_{i_1,t_1t_2}\varphi^c_{i_1,t_3t_4}\right]E\left[\varphi^c_{i_2,t_1t_2}\varphi^c_{i_2,t_5t_6}\right] \\ &\quad \times E\left[\varphi^c_{i_3,t_3t_4}\varphi^c_{i_3,t_7t_8}\right]E\left[\varphi^c_{i_4,t_5t_6}\varphi^c_{i_4,t_7t_8}\right]. \end{aligned}$$

Like in the analysis of G_I, we consider five cases inside the above summation: the number of distinct elements in $\{t_1,t_2,...,t_8\}$ are 8, 7, 6, 5, and 4 or less. We continue to use (A), (B), (C), (D), and (E) to denote these five cases, respectively, and denote the corresponding sum in G_{IV1} as $G_{IV1,A}$, $G_{IV1,B}$, $G_{IV1,C}$, $G_{IV1,D}$, and $G_{IV1,E}$, respectively (e.g., $G_{IV1,A}$ is defined as G_{IV1} but with the time indices restricted to case (A)). For case (A), we consider two different subcases: (Aa) there exists $k_0 \in \{1,...,8\}$ such that, $|t_l - t_{k_0}| > m$ for all $l \neq k_0$; (Ab) all the other remaining cases. We use $G_{IV1,Aa}$ and $G_{IV1,Ab}$ to denote $G_{IV1,A}$ but with the time indices restricted to subcases (Aa) and (Ab), respectively. In case (Aa) we can follow the same argument as used in case (Aa) in Step 2 to bound $G_{IV1,Aa}$ as $G_{IV1,Aa} \leq \frac{Ch^4}{n^4T^4}n^4T^8h^{-\frac{2\delta}{1+\delta}}\alpha^{\frac{\delta}{1+\delta}}(m) = O(T^4h^{\frac{2(2+\delta)}{1+\delta}}\alpha^{\frac{\delta}{1+\delta}}(m)) = o(1)$. In case (Ab), the number of terms in the summation for $G_{IV1,Ab}$ is of order $O\left(T^4m^4\right)$ and each term is uniformly bounded by a constant C. It follows that $G_{IV1,Ab} \leq \frac{Ch^4}{n^4T^4}n^4T^4m^4 = O\left(h^4m^4\right) = o(1)$.

For case (B), we consider two different subcases: (Ba) there exists $k_0 \in \{1,...,8\}$ such that, $|t_l - t_{k_0}| > m$ for all $l \neq k_0$ with $t_l \neq t_{k_0}$; (Bb) all the other remaining cases. For subcase (Ba), we consider only two representative subcases: (Ba1) $t_8 = t_1$ or $t_8 = t_2$, (Ba2) $t_8 = t_5$

or $t_8 = t_6$ since the other cases are analogous. For subsubcase (Ba1) WLOG we assume $t_8 = t_1$. Noting that all the four time indices in each of the four expectations $E[\varphi^c_{i_1,t_1t_2}\varphi^c_{i_1,t_3t_4}]$, $E[\varphi^c_{i_2,t_1t_2}\varphi^c_{i_2,t_5t_6}]$, $E[\varphi^c_{i_3,t_3t_4}\varphi^c_{i_3,t_7t_1}]$, and $E[\varphi^c_{i_4,t_5t_6}\varphi^c_{i_4,t_7t_1}]$ are different from each other, we can easily get the bound for $G_{IV1,B}$ (with the restriction $t_8 = t_1$) as $O(T^3 h^{\frac{2(2+\delta)}{1+\delta}} \alpha^{\frac{\delta}{1+\delta}}(m)) = o(1)$. For subsubcase (Ba2) we assume $t_8 = t_5$ and consider bounding the following objects: $E[\varphi^c_{i_1,t_1t_2}\varphi^c_{i_1,t_3t_4}]$, $E[\varphi^c_{i_2,t_1t_2}\varphi^c_{i_2,t_5t_6}]$, $E[\varphi^c_{i_3,t_3t_4}\varphi^c_{i_3,t_7t_5}]$, and $E[\varphi^c_{i_4,t_5t_6}\varphi^c_{i_4,t_7t_5}]$. Note that the indices in the last expectation $E[\varphi^c_{i_4,t_5t_6}\varphi^c_{i_4,t_7t_5}]$ are not all distinct. Despite this, since all the four indices in each of the other three expectations are distinct, we can continue to bound $G_{IV1,B}$ (with the restriction $t_8 = t_5$) as $O(T^3 h^{\frac{2(2+\delta)}{1+\delta}} \alpha^{\frac{\delta}{1+\delta}}(m)) = o(1)$. For subcase (Bb), it is easy to tell $G_{IV1,B}$ is bounded by $T^{-4}h^4 O\left(T^4 m^3\right) = O\left(h^4 m^3\right) = o(1)$. It follows that $G_{IV1,B} = o(1)$. For case (C), analogous to the study of case (C) in Step 2, we have

$$G_{IV1,C} = \frac{h^4}{T^4} O\left(T^6 h^{-1-\frac{2\delta}{1+\delta}} \alpha^{\frac{\delta}{1+\delta}}(m) + T^4 m^2 h^{-1}\right) = O\left(T^2 h^{\frac{3+\delta}{1+\delta}} \alpha^{\frac{\delta}{1+\delta}}(m) + h^3 m^2\right) = o(1).$$

Similarly, in case (D) we have

$$G_{IV1,D} \le \frac{h^4}{T^4} O\left(T^5 h^{-1-\frac{2\delta}{1+\delta}} \alpha^{\frac{\delta}{1+\delta}}(m) + T^4 m h^{-1}\right) = O\left(T h^{\frac{3+\delta}{1+\delta}} \alpha^{\frac{\delta}{1+\delta}}(m) + h^3 m\right) = o(1).$$

In case (E), it is straightforward to bound $G_{IV1,E}$ as

$$G_{IV1,E} \le \frac{Ch^4}{n^4 T^4} n^4 \left(T^4 h^{-2} + T^3 h^{-3} + T^2 h^{-4}\right) = O\left(h^2 + T^{-1}h + T^{-2}\right) = o(1).$$

In sum, $G_{IV1} = o(1)$. Similarly we can show that $G_{IVs} = o(1)$ for $s = 2, 3$. ■

Lemma .0.2 $nTh\overline{\Gamma}_{nT,2} = o_P(1)$.

Proof. Let $n_1 \equiv n-1$ and $T_1 \equiv T-1$. Recalling that $c_{i,ts} \equiv E(\varphi_{i,ts})$ and $c_{ts} \equiv n_1^{-1}\sum_{i=1}^n c_{i,ts}$, we have

$$\begin{aligned}
nTh\overline{\Gamma}_{nT,2} &= \frac{2h}{n_1} \sum_{1\le j\ne i\le n} T_1^{-1} \sum_{t=2}^{T}\sum_{s=1}^{t-1} \left[\varphi^c_{i,ts} c_{j,ts} + \varphi^c_{j,ts} c_{i,ts}\right] \\
&= \frac{2h}{n_1} \sum_{i=1}^{n}\sum_{j=1}^{n} T_1^{-1} \sum_{t=2}^{T}\sum_{s=1}^{t-1} \left[\varphi^c_{i,ts} c_{j,ts} + \varphi^c_{j,ts} c_{i,ts}\right] - \frac{4h}{n_1}\sum_{i=1}^{n} T_1^{-1} \sum_{t=2}^{T}\sum_{s=1}^{t-1} \varphi^c_{i,ts} c_{i,ts} \\
&= 4h \sum_{i=1}^{n} T_1^{-1} \sum_{t=2}^{T}\sum_{s=1}^{t-1} \varphi^c_{i,ts} c_{ts} - \frac{4h}{n_1}\sum_{i=1}^{n} T_1^{-1} \sum_{t=2}^{T}\sum_{s=1}^{t-1} \varphi^c_{i,ts} c_{i,ts} \\
&\equiv 4V_{1nT} - 4V_{2nT}, \text{ say.}
\end{aligned}$$

We complete the proof by showing that $V_{1nT} = o_P(1)$ and $V_{2nT} = o_P(1)$. We only prove the first claim since the proof of the second one is similar.

Let $v_{i,t} \equiv \sum_{s=1}^{t-1} h^{1/2} \varphi^c_{i,ts} c_{ts}$ and $v_i \equiv T_1^{-1}\sum_{t=2}^{T} v_{i,t}$. Then we can write $V_{1nT} = h^{1/2}\sum_{i=1}^n v_i$. Note that $E(v_i) = 0$ and $\{v_i\}_{i=1}^n$ are independently distributed under H_0, we have $E[(V_{1nT})^2] =$

$h\sum_{i=1}^{n}\text{Var}(v_i)$. For $\text{Var}(v_i)$, we have

$$\text{Var}(v_i) = E\left[\frac{1}{T_1}\sum_{t=2}^{T} v_{i,t}\right]^2 = \frac{1}{T_1^2}\sum_{t=2}^{T} E\left[v_{i,t}^2\right] + \frac{2}{T_1^2}\sum_{t_1=3}^{T}\sum_{t_2=2}^{t_1-1} E\left[v_{i,t_1}v_{i,t_2}\right] \equiv V_{1i} + V_{2i}, \text{ say.}$$

For V_{1i}, we have

$$V_{1i} = \frac{h}{T_1^2}\sum_{t=2}^{T}\sum_{s=1}^{t-1} E\left[\varphi_{i,ts}^{c2}\right] c_{ts}^2 + \frac{2h}{T_1^2}\sum_{t=3}^{T}\sum_{s=2}^{t-1}\sum_{r=1}^{s-1} E\left[\varphi_{i,ts}^{c}\varphi_{i,tr}^{c}\right] c_{ts}c_{tr} \equiv V_{1i,1} + V_{1i,2}, \text{ say.}$$

By (.0.1) and Assumption A.1, $|c_{ts}| = |n_1^{-1}\sum_{i=1}^{n} E[\varphi_{i,ts}]| \le Ch^{\frac{-\delta}{1+\delta}}\alpha^{\frac{\delta}{1+\delta}}(t-s)$. Thus uniformly in i

$$\begin{aligned} V_{1i,1} &\le \frac{C}{T_1^2}\sum_{t=2}^{T}\sum_{s=1}^{t-1} h^{\frac{-2\delta}{1+\delta}}\alpha^{\frac{2\delta}{1+\delta}}(t-s)\max_{1\le t\ne s\le T}\left\{hE\left[\varphi_{i,ts}^{c2}\right]\right\} \\ &\le \frac{C}{T_1}\max_{1\le i\le n}\max_{1\le t\ne s\le T}\left\{hE\left[\varphi_{i,ts}^{c2}\right]\right\} h^{\frac{-2\delta}{1+\delta}}\sum_{\tau=1}^{T-1}\alpha^{\frac{2\delta}{1+\delta}}(\tau) = O\left(T^{-1}h^{\frac{-2\delta}{1+\delta}}\right). \end{aligned}$$

For $V_{1i,2}$, we have that uniformly in i

$$\begin{aligned} |V_{1i,2}| &= \frac{2h}{T_1^2}\sum_{t=3}^{T}\sum_{s=2}^{t-1}\sum_{r=1}^{s-1}\left|E\left(\varphi_{i,ts}^{c}\varphi_{i,tr}^{c}\right)\right||c_{ts}||c_{tr}| \le \frac{Chh^{\frac{-2\delta}{1+\delta}}}{T_1^2}\sum_{t=3}^{T}\sum_{s=2}^{t-1}\sum_{r=1}^{s-1}\alpha^{\frac{\delta}{1+\delta}}(t-s)\,\alpha^{\frac{\delta}{1+\delta}}(t-r) \\ &\le \frac{Ch^{\frac{1-\delta}{1+\delta}}}{T_1}\sum_{\tau_1=1}^{\infty}\sum_{\tau_2=1}^{\infty}\alpha^{\frac{\delta}{1+\delta}}(\tau_1)\,\alpha^{\frac{\delta}{1+\delta}}(\tau_2) = O\left(T^{-1}h^{\frac{1-\delta}{1+\delta}}\right). \end{aligned}$$

It follows that $V_{1i} = O(T^{-1}h^{\frac{-2\delta}{1+\delta}} + T^{-1}h^{\frac{1-\delta}{1+\delta}})$ uniformly in i.

For V_{2i}, we have

$$\begin{aligned} V_{2i} &= \frac{2h}{T_1^2}\sum_{t_1=3}^{T}\sum_{t_2=2}^{t_1-1}\sum_{t_3=1}^{t_1-1}\sum_{t_4=1}^{t_2-1} E\left[\varphi_{i,t_1t_3}^{c}\varphi_{i,t_2t_4}^{c}\right] c_{t_1t_3}c_{t_2t_4} \\ &= \frac{4h}{T_1^2}\sum_{t_1=3}^{T}\sum_{t_2=2}^{t_1-1}\sum_{t_3=1}^{t_2-1} E\left[\varphi_{i,t_1t_3}^{c}\varphi_{i,t_2t_3}^{c}\right] c_{t_1t_3}c_{t_2t_3} \\ &\quad + \frac{2h}{T_1^2}\sum_{t_1=3}^{T}\sum_{t_2=2}^{t_1-1}\sum_{t_3=1,t_3\ne t_4,t_2}^{t_1-1}\sum_{t_4=1}^{t_2-1} E\left[\varphi_{i,t_1t_3}^{c}\varphi_{i,t_2t_4}^{c}\right] c_{t_1t_3}c_{t_2t_4} \equiv V_{2i,1} + V_{2i,2} \text{ say,} \end{aligned}$$

where the first term is obtained when $t_3 = t_4$ or t_2 as $\varphi_{i,ts} = \varphi_{i,st}$. Following the analysis of $V_{1i,2}$, we can show that $|V_{2i,1}| \le CT_1^{-1}h^{\frac{1-\delta}{1+\delta}}\sum_{\tau_1=1}^{\infty}\sum_{\tau_2=1}^{\infty}\alpha^{\frac{\delta}{1+\delta}}(\tau_1)\,\alpha^{\frac{\delta}{1+\delta}}(\tau_2) = O(T^{-1}h^{\frac{1-\delta}{1+\delta}})$ uniformly in i. For $V_{2i,2}$, we consider three cases: (a) $1 \le t_3 < t_4 < t_2 < t_1 \le T$; (b) $1 \le t_4 < t_3 < t_2 < t_1 \le T$; (c) $1 \le t_4 < t_2 < t_3 < t_1 \le T$, and use $V_{2i,2a}$, $V_{2i,2b}$, and $V_{2i,2c}$ to denote the summation over these three cases of indices, respectively. In case (a), by separating variables indexed by t_3 from those indexed by t_4, t_2, and t_1 and Lemma .0.5, we have

$$\left|E\left[\varphi_{i,t_1t_3}^{c}\varphi_{i,t_2t_4}^{c}\right]\right| \le \left|E\left[E_{t_3}\left(\varphi_{i,t_1t_3}^{c}\right)\varphi_{i,t_2t_4}^{c}\right]\right| + Ch^{\frac{-2\delta}{1+\delta}}\alpha^{\frac{\delta}{1+\delta}}(t_4-t_3) = Ch^{\frac{-2\delta}{1+\delta}}\alpha^{\frac{\delta}{1+\delta}}(t_4-t_3),$$

where the equality follows from the fact that $E_s(\varphi_{i,ts}^{c}) = E_tE_s(\overline{k}_{h,ts}^{i}) - E(\overline{k}_{h,ts}^{i})$ is a constant and that $E(\varphi^c_{i,ts})=0$ for $t\neq s$. It follows that uniformly in i

$$\begin{aligned}
|V_{2i,2a}| &\leq \frac{2h}{T_1^2}\sum_{t_1=3}^{T}\sum_{t_2=2}^{t_1-1}\sum_{t_4=1}^{t_2-1}\sum_{t_3=1}^{t_4-1}\left|E\left[\varphi^c_{i,t_1t_3}\varphi^c_{i,t_2t_4}\right]\right||c_{t_1t_3}||c_{t_2t_4}| \\
&\leq \frac{Chh^{\frac{-4\delta}{1+\delta}}}{T_1^2}\sum_{t_1=3}^{T}\sum_{t_2=2}^{t_1-1}\sum_{t_3=1,t_3\neq t_4,t_2}^{t_1-1}\sum_{t_4=1}^{t_2-1}\alpha^{\frac{\delta}{1+\delta}}(t_4-t_3)\,\alpha^{\frac{\delta}{1+\delta}}(t_1-t_3)\,\alpha^{\frac{\delta}{1+\delta}}(t_2-t_4) \\
&\leq \frac{Ch^{\frac{1-3\delta}{1+\delta}}}{T_1}\sum_{\tau_3=1}^{\infty}\sum_{\tau_2=1}^{\infty}\sum_{\tau_1=1}^{\infty}\alpha^{\frac{\delta}{1+\delta}}(\tau_1)\,\alpha^{\frac{\delta}{1+\delta}}(\tau_2)\,\alpha^{\frac{\delta}{1+\delta}}(\tau_3)=O\left(T^{-1}h^{\frac{1-3\delta}{1+\delta}}\right).
\end{aligned}$$

By the same token, we can show that $\left|V_{2i,2\xi}\right|=O(T^{-1}h^{\frac{1-3\delta}{1+\delta}})$ uniformly in i for $\xi=b,c$. Hence $V_{2i,2}=O(T^{-1}h^{\frac{1-3\delta}{1+\delta}})$ and $V_{2i}=O(T^{-1}h^{\frac{1-\delta}{1+\delta}})+O(T^{-1}h^{\frac{1-3\delta}{1+\delta}})=O(T^{-1}h^{\frac{1-3\delta}{1+\delta}})$ uniformly in i. Consequently

$$E[(V_{1nT})^2]=h\sum_{i=1}^{n}(V_{1i}+V_{2i})=O\left(nh\left(T^{-1}h^{\frac{-2\delta}{1+\delta}}+T^{-1}h^{\frac{1-3\delta}{1+\delta}}\right)\right)=O\left(nh^{\frac{1-\delta}{1+\delta}}/T\right)=o(1).$$

Then $V_{1nT}=o_P(1)$ by the Chebyshev inequality. ■

Proposition .0.2 $nThG^{(3)}_{nT}=o_P(1)$.

Proof. By the definition of $G^{(3)}_{nT}$ and (.0.5), we have

$$\begin{aligned}
-12nThG^{(3)}_{nT} &= \frac{-12nTh}{n(n-1)C_T^3}\sum_{1\leq i\neq j\leq n}\sum_{1\leq t<s<r\leq T}\vartheta^{(3)}_{ij}(Z_{ij,t},Z_{ij,s},Z_{ij,r}) \\
&= \frac{Th}{n_1C_T^3}\sum_{1\leq i\neq j\leq n}\sum_{1\leq t<s<r\leq T}[\varphi_{i,ts}\varphi_{j,tr}+\varphi_{i,ts}\varphi_{j,sr}+\varphi_{i,tr}\varphi_{j,ts}+\varphi_{i,tr}\varphi_{j,sr} \\
&\quad +\varphi_{i,sr}\varphi_{j,st}+\varphi_{i,sr}\varphi_{j,rt}] \\
&\equiv U_{1nT}+U_{2nT}+U_{3nT}+U_{4nT}+U_{5nT}+U_{6nT},\ \text{say},
\end{aligned}$$

where, e.g., $U_{1nT}\equiv\frac{Th}{n_1C_T^3}\sum_{1\leq i\neq j\leq n}\sum_{1\leq t<s<r\leq T}\varphi_{i,ts}\varphi_{j,tr}$. It suffices to show that $U_{rnT}=o_P(1)$ for $r=1,2,...,6$.

For U_{1nT}, we have

$$\begin{aligned}
U_{1nT} &= \frac{Th}{n_1C_T^3}\sum_{1\leq i\neq j\leq n}\sum_{1\leq t<s<r\leq T}\varphi^c_{i,ts}\varphi^c_{j,tr}+\frac{Th}{n_1C_T^3}\sum_{1\leq i\neq j\leq n}\sum_{1\leq t<s<r\leq T}c_{i,ts}\varphi^c_{j,tr} \\
&\quad +\frac{Th}{n_1C_T^3}\sum_{1\leq i\neq j\leq n}\sum_{1\leq t<s<r\leq T}\varphi^c_{i,ts}c_{j,tr}+\frac{Th}{n_1C_T^3}\sum_{1\leq i\neq j\leq n}\sum_{1\leq t<s<r\leq T}c_{i,ts}c_{j,tr} \\
&\equiv U_{1nT,1}+U_{1nT,2}+U_{1nT,3}+U_{1nT,4},\ \text{say},
\end{aligned}$$

where recall $\varphi^c_{i,ts}\equiv\varphi_{i,ts}-E(\varphi_{i,ts})$ and $c_{i,ts}\equiv E(\varphi_{i,ts})$. We further decompose $U_{1nT,1}$ as follows

$$\begin{aligned}
U_{1nT,1} &= \frac{Th}{n_1C_T^3}\sum_{1\leq i<j\leq n}\sum_{1\leq t<s<r\leq T}\varphi^c_{i,ts}\varphi^c_{j,tr}+\frac{Th}{n_1C_T^3}\sum_{1\leq j<i\leq n}\sum_{1\leq t<s<r\leq T}\varphi^c_{i,ts}\varphi^c_{j,tr} \\
&\equiv U_{1nT,1a}+U_{1nT,1b}.
\end{aligned}$$

Noting that $E\left(U_{1nT,1a}\right)=0$ under H_0, we have

$$\begin{aligned}
\operatorname{Var}\left(U_{1nT,1a}\right) &= \frac{T^2h^2}{\left(n_1C_T^3\right)^2}\sum_{1\le i_1<i_2\le n}\sum_{\substack{1\le t_1<t_2<t_3\le T\\1\le t_4<t_5<t_6\le T}} E\left[\varphi^c_{i_1,t_1t_2}\varphi^c_{i_2,t_1t_3}\varphi^c_{i_1,t_4t_5}\varphi^c_{i_2,t_4t_6}\right]\\
&= \frac{T^2h^2}{\left(n_1C_T^3\right)^2}\sum_{1\le i_1<i_2\le n}\sum_{\substack{1\le t_1<t_2<t_3\le T\\1\le t_4<t_5<t_6\le T}} E\left[\varphi^c_{i_1,t_1t_2}\varphi^c_{i_1,t_4t_5}\right]E\left[\varphi^c_{i_2,t_1t_3}\varphi^c_{i_2,t_4t_6}\right].
\end{aligned}$$

Analogously to the proof of (.0.13), we can show

$$\begin{aligned}
\operatorname{Var}\left(U_{1nT,1a}\right) &\le \frac{CT^2h^2}{\left(n_1C_T^3\right)^2}\left\{n^2T^6h^{\frac{-2\delta}{1+\delta}}\alpha^{\frac{\delta}{1+\delta}}(m)+n^2T^3m^3+n^2T^3h^{-2}\right\}\\
&= O\left(T^2h^{\frac{2}{1+\delta}}\alpha^{\frac{\delta}{1+\delta}}(m)+T^{-1}h^2m^3+T^{-1}\right)=o(1).
\end{aligned}$$

Hence $U_{1nT,1a}=o_P(1)$ by the Chebyshev inequality. Similarly, $U_{1nT,1b}=o_P(1)$. It follows that $U_{1nT,1}=o_P(1)$.

For $U_{1nT,2}$, write

$$\begin{aligned}
U_{1nT,2} &= \frac{Th}{n_1C_T^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{1\le t<s<r\le T}c_{j,ts}\varphi^c_{i,tr}-\frac{Th}{n_1C_T^3}\sum_{i=1}^{n}\sum_{1\le t<s<r\le T}c_{i,ts}\varphi^c_{i,tr}\\
&= \frac{Th}{C_T^3}\sum_{i=1}^{n}\sum_{1\le t<s<r\le T}c_{ts}\varphi^c_{i,tr}-\frac{Th}{n_1C_T^3}\sum_{i=1}^{n}\sum_{1\le t<s<r\le T}c_{i,ts}\varphi^c_{i,tr}\equiv U_{1nT,2a}-U_{1nT,2b},
\end{aligned}$$

where recall $c_{ts}\equiv n_1^{-1}\sum_{i=1}^n c_{i,ts}$. Noting $E\left(U_{1nT,2a}\right)=0$, we have

$$\begin{aligned}
\operatorname{Var}\left(U_{1nT,2a}\right) &= \frac{T^2h^2}{\left(C_T^3\right)^2}\sum_{i=1}^{n}\sum_{\substack{1\le t_1<t_2<t_3\le T\\1\le t_4<t_5<t_6\le T}}c_{t_1t_2}c_{t_4t_5}E\left[\varphi^c_{i,t_1t_3}\varphi^c_{i,t_4t_6}\right]\\
&= \frac{T^2h^2}{\left(C_T^3\right)^2}\sum_{i=1}^{n}\sum_{\substack{1\le t_1<t_2<t_3\le T,\\1\le t_4<t_5<t_6\le T,\\t_1,\ldots,t_6\text{ are all distinct}}}c_{t_1t_2}c_{t_4t_5}E\left[\varphi^c_{i,t_1t_3}\varphi^c_{i,t_4t_6}\right]+o(1)\\
&\le \frac{Ch^2h^{\frac{-4\delta}{1+\delta}}}{T}\sum_{i=1}^{n}\sum_{\tau_3=1}^{\infty}\sum_{\tau_2=1}^{\infty}\sum_{\tau_1=1}^{\infty}\alpha^{\frac{\delta}{1+\delta}}(\tau_1)\alpha^{\frac{\delta}{1+\delta}}(\tau_2)\alpha^{\frac{\delta}{1+\delta}}(\tau_3)+o(1)\\
&= O\left(nh^{\frac{2(1-\delta)}{1+\delta}}/T\right)+o(1)=o(1).
\end{aligned}$$

So $U_{1nT,2a}=o_P(1)$. By the same token $U_{1nT,2b}=o_P(1)$. Thus $U_{1nT,2}=o_P(1)$. Similarly we can show that $U_{1nT,3}=o_P(1)$. For $U_{1nT,4}$, we have

$$\begin{aligned}
\left|U_{1nT,4}\right| &\le \frac{Th}{n_1C_T^3}\sum_{1\le i\ne j\le n}\sum_{1\le t<s<r\le T}\left|c_{i,ts}\right|\left|c_{j,tr}\right|\\
&\le \frac{CThh^{\frac{-2\delta}{1+\delta}}}{n_1C_T^3}\sum_{1\le i\ne j\le n}\sum_{1\le t<s<r\le T}\alpha^{\frac{\delta}{1+\delta}}(s-t)\alpha^{\frac{\delta}{1+\delta}}(r-t)\\
&\le \frac{Cnh^{\frac{1-\delta}{1+\delta}}}{T}\sum_{\tau_1=1}^{\infty}\sum_{\tau_2=1}^{\infty}\alpha^{\frac{\delta}{1+\delta}}(\tau_1)\alpha^{\frac{\delta}{1+\delta}}(\tau_2)=O\left(nh^{\frac{1-\delta}{1+\delta}}/T\right)=o(1).
\end{aligned}$$

Consequently, $U_{1nT}=o_P(1)$. Analogously we can show that $U_{rnT}=o_P(1)$ for $r=2,3,...,6$. This completes the proof of the proposition. ■

Proposition .0.3 $nThG_{nT}^{(4)}=o_P(1)$.

Proof. By the definition of $G_{nT}^{(4)}$ and (.0.6), we have

$$\begin{aligned}
6nThG_{nT}^{(4)} &= \frac{6Th}{n_1C_T^4}\sum_{1\leq i\neq j\leq n}\sum_{1\leq t<s<r<q\leq T}\vartheta_{ij}^{(4)}\left(Z_{ij,t},Z_{ij,s},Z_{ij,r},Z_{ij,q}\right)\\
&= \frac{Th}{n_1C_T^4}\sum_{1\leq i\neq j\leq n}\sum_{1\leq t<s<r<q\leq T}\{\varphi_{i,ts}\varphi_{j,rq}+\varphi_{i,tr}\varphi_{j,sq}+\varphi_{i,rq}\varphi_{j,ts}+\varphi_{i,sq}\varphi_{j,tr}\\
&\quad +\varphi_{i,tq}\varphi_{j,sr}+\varphi_{i,sr}\varphi_{j,tq}\}\equiv\sum_{l=1}^{6}Q_{lnT},\text{ say,}
\end{aligned}$$

where e.g., $Q_{1nT}=\frac{Th}{n_1C_T^4}\sum_{1\leq i\neq j\leq n}\sum_{1\leq t<s<r<q}\varphi_{i,ts}\varphi_{j,rq}$. It suffices to show $Q_{lnT}=o_P(1)$ for $l=1,2,...,6$. We only show that $Q_{1nT}=o_P(1)$ since the other cases are similar. Write

$$\begin{aligned}
Q_{1nT} &= \frac{Th}{n_1C_T^4}\sum_{1\leq i\neq j\leq n}\sum_{1\leq t<s<r<q\leq T}\varphi_{i,ts}^c\varphi_{j,rq}^c+\frac{Th}{n_1C_T^4}\sum_{1\leq i\neq j\leq n}\sum_{1\leq t<s<r<q\leq T}c_{i,ts}\varphi_{j,rq}^c\\
&\quad +\frac{Th}{n_1C_T^4}\sum_{1\leq i\neq j\leq n}\sum_{1\leq t<s<r<q\leq T}\varphi_{i,ts}^c c_{j,rq}+\frac{Th}{n_1C_T^4}\sum_{1\leq i\neq j\leq n}\sum_{1\leq t<s<r<q\leq T}c_{i,ts}c_{j,rq}\\
&\equiv Q_{1nT,1}+Q_{1nT,2}+Q_{1nT,3}+Q_{1nT,4},\text{ say.}
\end{aligned}$$

Analogously to the determination of the probability orders of $U_{1nT,1}$, $U_{1nT,2}$, and $U_{1nT,3}$ in the proof of Proposition .0.2, we can show that $Q_{1nT,s}=o_P(1)$ for $s=1,2,3$. For $Q_{1nT,4}$, we have

$$|Q_{1nT,4}|\leq\frac{CThh^{-\frac{2\delta}{1+\delta}}}{n_1C_T^4}\sum_{1\leq i\neq j\leq n}\sum_{1\leq t<s<r<q\leq T}\alpha^{\frac{1}{1+\delta}}(s-t)\,\alpha^{\frac{1}{1+\delta}}(q-r)=O(nh^{\frac{1-\delta}{1+\delta}}/T)=o(1).$$

It follows that $Q_{1nT}=o_P(1)$. ■

.0.1 Proof of Corollary 2.3.2

Given Theorem 2.3.1, it suffices to show: (i) $\widehat{D}_{1nT}\equiv\widehat{\sigma}_{nT}^2-\sigma_{nT}^2=o_P(1)$, and (ii) $\widehat{D}_{2nT}\equiv\widehat{B}_{nT}-B_{nT}=o_P(1)$. For (i), we write

$$\begin{aligned}
\sigma_{nT}^2 &= \frac{4h^2}{n(n-1)T(T-1)}\sum_{1\leq i\neq j\leq n}\sum_{1\leq t\neq s\leq T}E\left[\left(\bar{k}_{h,ts}^i\right)^2\right]E\left[\left(\bar{k}_{h,ts}^j\right)^2\right]+o(1)\\
&= \frac{4R(\bar{k})^2}{n(n-1)T(T-1)}\sum_{1\leq i\neq j\leq n}\sum_{1\leq t\neq s\leq T}\int f_{i,ts}(u,u)\,du\int f_{j,ts}(v,v)\,dv+o(1).
\end{aligned}$$

Then

$$\begin{aligned}
\widehat{D}_{1nT} &= \frac{4R\left(\bar{k}\right)^2}{n(n-1)} \sum_{1\leq i\neq j\leq n} \frac{1}{T}\sum_{t=1}^{T} \widehat{f}_{ij,-t}\left(u_{it},u_{jt}\right) \\
&\quad - \frac{4R\left(\bar{k}\right)^2}{n(n-1)T(T-1)} \sum_{1\leq i\neq j\leq n}\sum_{1\leq t\neq s\leq T} \int f_{i,ts}\left(u,u\right)du \int f_{j,ts}\left(v,v\right)dv - o(1) \\
&= D_{1nT} - o(1).
\end{aligned}$$

where $D_{1nT} \equiv \frac{4R\left(\bar{k}\right)^2}{n(n-1)T(T-1)} \sum_{1\leq i\neq j\leq n}\sum_{1\leq t\neq s\leq T}\{k_{h,ts}^{i}k_{h,ts}^{j} - \int f_{i,ts}\left(u,u\right)du \int f_{j,ts}\left(v,v\right)dv\}$. It is easy to show that $E\left(D_{1nT}\right) = O\left(h^{\gamma}\right) = o(1)$ and $\mathrm{Var}(D_{1nT}) = o(1)$. Consequently, $\widehat{D}_{1nT} = o_P(1)$.

Now we show (ii). Noting that $B_{nT} = \frac{2h}{n_1}\sum_{1\leq i\neq j\leq n}\sum_{r=2}^{T}\frac{T-r+1}{T-1}E\left[\varphi_{i,1r}\right]E\left[\varphi_{j,1r}\right]$, we have

$$\begin{aligned}
\widehat{B}_{nT} - B_{nT} &= \frac{2h}{n_1}\sum_{r=2}^{T}\frac{T-r+1}{T-1}\sum_{1\leq i\neq j\leq n}\{\widehat{E}\left[\varphi_{i,1r}\right]\widehat{E}\left[\varphi_{j,1r}\right] - E\left[\varphi_{i,1r}\right]E\left[\varphi_{j,1r}\right]\} \\
&= \frac{2h}{n_1}\sum_{r=2}^{T}\frac{T-r+1}{T-1}\sum_{1\leq i\neq j\leq n}E\left[\varphi_{i,1r}\right]\left\{\widehat{E}\left[\varphi_{j,1r}\right] - E\left[\varphi_{j,1r}\right]\right\} \\
&\quad + \frac{2h}{n_1}\sum_{r=2}^{T}\frac{T-r+1}{T-1}\sum_{1\leq i\neq j\leq n}\left\{\widehat{E}\left[\varphi_{i,1r}\right] - E\left[\varphi_{i,1r}\right]\right\}E\left[\varphi_{j,1r}\right] \\
&\quad + \frac{2h}{n_1}\sum_{r=2}^{T}\frac{T-r+1}{T-1}\sum_{1\leq i\neq j\leq n}\left\{\widehat{E}\left[\varphi_{i,1r}\right] - E\left[\varphi_{i,1r}\right]\right\}\left\{\widehat{E}\left[\varphi_{j,1r}\right] - E\left[\varphi_{j,1r}\right]\right\} \\
&\equiv 2D_{2nT,1} + 2D_{2nT,2} + 2D_{2nT,3}, \text{ say.}
\end{aligned}$$

Recalling $c_{i,ts} \equiv E\left[\varphi_{i,ts}\right]$ and $c_{ts} \equiv n_1^{-1}\sum_{i=1}^{n}c_{i,ts}$, we have $D_{2nT,1} = h\sum_{r=2}^{T}\frac{T-r+1}{T-1}\sum_{i=1}^{n}c_{1r}\{\widehat{E}\left[\varphi_{i,1r}\right] - E\left[\varphi_{i,1r}\right]\} - \frac{h}{n_1}\sum_{r=2}^{T}\frac{T-r+1}{T-1}\sum_{i=1}^{n}c_{i,1r}\left\{\widehat{E}\left[\varphi_{i,1r}\right] - E\left[\varphi_{j,1r}\right]\right\} \equiv D_{2nT,1a} - D_{2nT,1b}$, say. We only show that $D_{2nT,1a} = o_P(1)$ as the proof that $D_{2nT,1b} = o_P(1)$ is analogous. Noting that

$$\widehat{E}\left[\varphi_{i,1r}\right] - E\left[\varphi_{i,1r}\right] = \frac{1}{T-r+1}\sum_{t=1}^{T-r+1}\{\bar{k}_{h,t,t+r-1}^{i} - E[\bar{k}_{h,t,t+r-1}^{i}]\} - \frac{1}{C_T^2}\sum_{1\leq t<s\leq T}\{\bar{k}_{h,ts}^{i} - E_tE_s[\bar{k}_{h,ts}^{i}]\}, \tag{.0.25}$$

we have

$$\begin{aligned}
D_{2nT,1a} &= h\sum_{i=1}^{n}\sum_{r=2}^{T}\bar{c}_{1r}\frac{1}{T_r+1}\sum_{t=1}^{T_r+1}\left\{\bar{k}_{h,t,t+r-1}^{i} - E[\bar{k}_{h,t,t+r-1}^{i}]\right\} \\
&\quad - h\sum_{i=1}^{n}\sum_{r=2}^{T}\bar{c}_{1r}\frac{1}{C_T^2}\sum_{1\leq t<s\leq T}\left\{\bar{k}_{h,ts}^{i} - E_tE_s[\bar{k}_{h,ts}^{i}]\right\} \\
&\equiv D_{2nT,1a1} - D_{2nT,1a2}, \text{ say,}
\end{aligned} \tag{.0.26}$$

where $\overline{c}_{1r} \equiv c_{1r}(T-r+1)/(T-1)$ and $T_r \equiv T-r$. Noting that $E(D_{2nT,1a1})=0$, we have

$$\begin{aligned}
& \operatorname{Var}(D_{2nT,1a1}) \\
= & \ h^2 \sum_{i=1}^{n} \sum_{r_1=2}^{T} \overline{c}_{1r_1} \sum_{r_2=2}^{T} \overline{c}_{1r_2} \frac{1}{(T_{r_1}+1)(T_{r_2}+1)} \sum_{t=1}^{T_{r_1}+1} \sum_{s=1}^{T_{r_2}+1} \operatorname{Cov}\left(\overline{k}^i_{h,t,t+r_1-1}, \overline{k}^i_{h,s,s+r_2-1}\right) \\
= & \ h^2 \sum_{i=1}^{n} \sum_{r_1=2}^{T} \overline{c}_{1r_1} \sum_{r_2=2}^{T} \overline{c}_{1r_2} \frac{1}{(T_{r_1}+1)(T_{r_2}+1)} \sum_{t=1}^{T_{r_1}+1} \sum_{s=1, s\neq t, s\neq t+r_1-r_2}^{T_{r_2}+1} \operatorname{Cov}\left(\overline{k}^i_{h,t,t+r_1-1}, \overline{k}^i_{h,s,s+r_2-1}\right) \\
& + o(1). \qquad (.0.27)
\end{aligned}$$

We consider three cases for the summation in the last expression: (a) $t<t+r_1-1<s<s+r_2-1$ or $s<s+r_2-1<t<t+r_1-1$, (b) $t<s<s+r_2-1<t+r_1-1$ or $s<t<t+r_1-1<s+r_2-1$, and (c) $t<s<t+r_1-1<s+r_2-1$ or $s<t<s+r_2-1<t+r_1-1$, and use VD_{2nTa}, VD_{2nTb}, and VD_{2nTc} denote the summation in (.0.27) corresponding these three cases, respectively. In case (a) we can apply the fact that $\sum_{r=2}^{T} \overline{c}_{1r} \leq Ch^{-\frac{\delta}{1+\delta}}$ and the Davydov inequality to obtain $VD_{2nTa} \leq Cnh^{2-\frac{4\delta}{1+\delta}}/T = O(nh^{\frac{2(1-\delta)}{1+\delta}}/T) = o(1)$. In case (b), WLOG we assume $t<s<s+r_2-1<t+r_1-1$. Then we apply Lemma .0.5 by first separating t from $(s, s+r_2-1, t+r_1-1)$ and then separating $t+r_1-1$ from $(s, s+r_2-1)$ to obtain

$$\begin{aligned}
& \left|\operatorname{Cov}\left(\overline{k}^i_{h,t,t+r_1-1}, \overline{k}^i_{h,s,s+r_2-1}\right)\right| \\
= & \left|E\left\{\left[\overline{k}^i_{h,t,t+r_1-1} - E(\overline{k}^i_{h,t,t+r_1-1})\right]\left[\overline{k}^i_{h,s,s+r_2-1} - E(\overline{k}^i_{h,s,s+r_2-1})\right]\right\}\right| \\
\leq & \left|E\left\{E_t\left[\overline{k}^i_{h,t,t+r_1-1} - E(\overline{k}^i_{h,t,t+r_1-1})\right]\left[\overline{k}^i_{h,s,s+r_2-1} - E(\overline{k}^i_{h,s,s+r_2-1})\right]\right\}\right| + Ch^{-\frac{2\delta}{1+\delta}} \alpha^{\frac{\delta}{1+\delta}}(s-t) \\
\leq & \ Ch^{-\frac{2\delta}{1+\delta}} \alpha^{\frac{\delta}{1+\delta}}(t+r_1-s-r_2) + Ch^{-\frac{2\delta}{1+\delta}} \alpha^{\frac{\delta}{1+\delta}}(s-t).
\end{aligned}$$

Then we have

$$\begin{aligned}
& h^2 \sum_{i=1}^{n} \sum_{r_1=2}^{T} \overline{c}_{1r_1} \sum_{r_2=2}^{T} \frac{\overline{c}_{1r_2}}{(T_{r_1}+1)(T_{r_2}+1)} \sum_{t=1}^{T_{r_1}+1} \sum_{\substack{s=1, s\neq t, s\neq t+r_1-r_2 \\ t<s<s+r_2-1<t+r_1-1}}^{T_{r_2}+1} \left|\operatorname{Cov}\left(\overline{k}^i_{h,t,t+r_1-1}, \overline{k}^i_{h,s,s+r_2-1}\right)\right| \\
\leq & \ Mh^{\frac{2}{1+\delta}} \sum_{i=1}^{n} \sum_{r_1=2}^{T} \sum_{r_2=2}^{T} \frac{\overline{c}_{1r_1}\overline{c}_{1r_2}}{(T_{r_1}+1)(T_{r_2}+1)} \sum_{t=1}^{T_{r_1}+1} \sum_{\substack{s=1, s\neq t, s\neq t+r_1-r_2 \\ t<s<s+r_2-1<t+r_1-1}}^{T_{r_2}+1} \left\{\alpha^{\frac{\delta}{1+\delta}}(t+r_1-s-r_2) + \alpha^{\frac{\delta}{1+\delta}}(s-t)\right\} \\
= & \ O\left(nh^{\frac{2(1-\delta)}{1+\delta}}/T\right) = o(1).
\end{aligned}$$

It follows that $VD_{2nTb} = o(1)$. Similarly, we have $VD_{2nTc} = o(1)$. Hence $\operatorname{Var}(D_{2nT,1a1}) = o(1)$ and $D_{2nT,1a1} = o_P(1)$ by the Chebyshev inequality.

To study $D_{2nT,1a2}$ in (.0.26), let $\chi_{i,ts} \equiv \overline{k}^i_{h,ts} - E_t E_s[\overline{k}^i_{h,ts}]$, and $\chi^c_{i,ts} \equiv \chi_{i,ts} - E(\chi_{i,ts})$. Noting that $|E(\chi_{i,ts})| \leq Ch^{\frac{-\delta}{1+\delta}} \alpha^{\frac{\delta}{1+\delta}}(|s-t|)$, we can readily show that $D_{2nT,1a2} = \overrightarrow{D}_{2nT,1a2} + o_P(1)$, where $\overrightarrow{D}_{2nT,1a2} = h\sum_{i=1}^{n} \sum_{r=2}^{T} \overline{c}_{1r} \frac{1}{C_T^2} \sum_{1\leq t<s\leq T} \chi^c_{i,ts}$. By construction, $E(\overrightarrow{D}_{2nT,1a2}) =$

0 and

$$E\left[\left(\overrightarrow{D}_{2nT,1a2}\right)^2\right] = h^2\sum_{i=1}^{n}\sum_{r_1=2}^{T}\bar{c}_{1r_1}\sum_{r_2=2}^{T}\bar{c}_{1r_2}\frac{1}{\left(C_T^2\right)^2}\sum_{1\leq t_1<t_2\leq T}\sum_{1\leq t_3<t_4\leq T}E\left(\chi^c_{i,t_1t_2}\chi^c_{i,t_3t_4}\right)$$
$$\leq Cnh^{\frac{2}{1+\delta}}/T = o(1).$$

Consequently, $\overrightarrow{D}_{2nT,1a2} = o_P(1)$ and $D_{2nT,1a2} = o_P(1)$. Hence $D_{2nT,1a} = o_P(1)$. Analogously $D_{2nT,1b} = o_P(1)$ and hence $D_{2nT,1} = o_P(1)$.

By the same token we can show that $D_{2nT,2} = o_P(1)$. To show $D_{2nT,3} = o_P(1)$, by (.0.25) we can decompose $D_{2nT,3}$ as follows

$$\begin{aligned}
D_{2nT,3} &= \frac{h}{n_1}\sum_{r=1}^{T_1}\frac{T_r}{T_1}\sum_{1\leq i\neq j\leq n}^{n}\frac{1}{T_r^2}\sum_{t=1}^{T_r}\sum_{s=1}^{T_r}\left(\bar{k}^i_{h,t,t+r}-E[\bar{k}^i_{h,t,t+r}]\right)\left(\bar{k}^j_{h,s,s+r}-E[\bar{k}^j_{h,s,s+r}]\right)\\
&\quad -\frac{h}{n_1}\sum_{r=1}^{T_1}\frac{T_r}{T_1}\sum_{1\leq i\neq j\leq n}^{n}\frac{1}{T_rC_T^2}\sum_{t_1=1}^{T_r}\sum_{1\leq t_2<t_3\leq T}\left(\bar{k}^i_{h,t_1,t_1+r}-E[\bar{k}^i_{h,t_1,t_1+r}]\right)\chi_{j,t_2t_3}\\
&\quad -\frac{h}{n_1}\sum_{r=1}^{T_1}\frac{T_r}{T_1}\sum_{1\leq i\neq j\leq n}^{n}\frac{1}{T_rC_T^2}\sum_{t_1=1}^{T_r}\sum_{1\leq t_2<t_3\leq T}\chi_{i,t_2t_3}\left(\bar{k}^j_{h,t_1,t_1+r}-E[\bar{k}^j_{h,t_1,t_1+r}]\right)\\
&\quad +\frac{h}{n_1}\sum_{r=1}^{T_1}\frac{T_r}{T_1}\sum_{1\leq i\neq j\leq n}^{n}\frac{1}{\left(C_T^2\right)^2}\sum_{1\leq t_1<t_2\leq T}\sum_{1\leq t_3<t_4\leq T}\chi_{i,t_1t_2}\chi_{j,t_3t_4}\\
&\equiv D_{2nT,3a}-D_{2nT,3b}-D_{2nT,3c}+D_{2nT,3d},\ \text{say}
\end{aligned}$$

It suffices to show $D_{2nT,3\xi} = o_P(1)$ for $\xi = a,b,c,$ and d. We only sketch the proof of $D_{2nT,3d} = o_P(1)$ since the other cases are simpler. First, note that $D_{2nT,3d} = \overrightarrow{D}_{2nT,3d} + o_P(1)$ by a simple application of Lemma .0.5, where

$$\overrightarrow{D}_{2nT,3d} = \frac{2h}{n_1\left(C_T^2\right)^2}\sum_{r=1}^{T_1}\frac{T_r}{T_1}\sum_{1\leq i<j\leq n}\sum_{1\leq t_1<t_2\leq T,1\leq t_3<t_4\leq T}\chi^c_{i,t_1t_2}\chi^c_{j,t_3t_4}.$$

Second, noting that $E(\overrightarrow{D}_{2nT,3d}) = 0$, we can write

$$E\left[\left(\overrightarrow{D}_{2nT,3d}\right)^2\right] = \frac{16h^2\left(\sum_{r=1}^{T_1}\frac{T_r}{T_1}\right)^2}{(n_1)^2\,(C_T^2)^4}\sum_{1\leq i<j\leq n}\sum_{\substack{1\leq t_1<t_2\leq T\\ 1\leq t_3<t_4\leq T}}\sum_{\substack{1\leq t_5<t_6\leq T,\\ 1\leq t_7<t_8\leq T}}E\left[\chi^c_{i,t_1t_2}\chi^c_{i,t_3t_4}\right]E\left[\chi^c_{j,t_5t_6}\chi^c_{j,t_7t_8}\right].$$

Now, following the same arguments as used in the proof of (.0.13) and applying Lemmas .0.5 and .0.6 repeatedly, we can show that $E[(\overrightarrow{D}_{2nT,3d})^2] = O(h^{\frac{2(1-\delta)}{1+\delta}}T^4\alpha^{\frac{\delta}{1+\delta}}(m)+h^2m^4) = o(1)$. Hence $\overrightarrow{D}_{2nT,3d} = o_P(1)$. This completes the proof of the corollary.

.0.2 Proof of Theorem 2.3.3

It suffices to show that under H_1,(i) $\widehat{\Gamma}_{nT} = \mu_A + o_P(1)$, (ii) $(nTh)^{-1}\widehat{B}_{nT} = o_P(1)$, and (iii)

$\widehat{\sigma}_{nT}^2 = \sigma_A^2 + o_P(1)$, because then $(nTh)^{-1}\widehat{I}_{nT} = \frac{\widehat{\Gamma}_{nT}}{\widehat{\sigma}_{nT}} - \frac{(nTh)^{-1}\widehat{B}_{nT}}{\widehat{\sigma}_{nT}} \xrightarrow{p} \frac{\mu_A}{\sigma_A} > 0$. Using the expression of $\widehat{\Gamma}_{nT}$ in (2.2.7), we can easily show that $E[\widehat{\Gamma}_{nT}] = \mu_A + o(1)$ and $\mathrm{Var}(\widehat{\Gamma}_{nT}) = o(1)$. Then (i) follows by the Chebyshev inequality. Next, it is easy to show that $(nTh)^{-1}\widehat{B}_{nT} = O_P\left(T^{-1}\right) = o_P(1)$ and thus (ii) follows. Lastly one can show (iii) by the Chebyshev inequality.

.0.3 Proof of Theorem 2.4.1

Let $\widetilde{\Gamma}_{1nT}$, $\widetilde{\Gamma}_{nT}$, $\widetilde{B}_{nT}$, and $\widetilde{\sigma}_{nT}^2$ be analogously defined as $\widehat{\Gamma}_{1nT}$, $\widehat{\Gamma}_{nT}$, $\widehat{B}_{nT}$, and $\widehat{\sigma}_{nT}^2$ but with $\{u_{it}\}$ being replaced by the residuals $\{\widetilde{u}_{it}\}$ in their definitions. We prove the theorem by showing that: (i) $nTh(\widetilde{\Gamma}_{nT} - \widehat{\Gamma}_{nT}) = o_P(1)$; (ii) $\widetilde{\sigma}_{nT}^2 = \widehat{\sigma}_{nT}^2 + o_P(1)$; and (iii) $\widetilde{B}_{nT} - \widehat{B}_{nT} = o_P(1)$.

To show (i), let $\widehat{\Delta}_{nT} \equiv \widehat{\Gamma}_{1nT} - \widehat{\Gamma}_{nT}$ and $\widetilde{\Delta}_{nT} \equiv \widetilde{\Gamma}_{1nT} - \widetilde{\Gamma}_{nT}$. By straightforward but tedious calculations, we have $\widehat{\Delta}_{nT} = \widehat{\Delta}_{nT,1} + \widehat{\Delta}_{nT,2}$, where $\widehat{\Delta}_{nT,1} = R\left(\overline{k}\right)\left[\frac{1}{Th^2}(1+\frac{1}{T}) - \frac{2}{nTh}\sum_{i=1}^{n}\int \widehat{f}_i^2(u)\,du\right]$, and

$$
\begin{aligned}
\widehat{\Delta}_{nT,2} \quad = \quad & \frac{1}{n(n-1)}\sum_{1\le i\neq j\le n}\left(\frac{1}{T^2} - \frac{1}{P_T^2} + \frac{6}{P_T^4} + \frac{2}{P_T^3}\right)\sum_{1\le t\neq s\le T}\overline{k}_{h,ts}^{i}\overline{k}_{h,ts}^{j} \\
& + \frac{1}{n(n-1)}\sum_{1\le i\neq j\le n}\left(\frac{2}{P_T^3} - \frac{2}{T^3} + \frac{4}{P_T^4}\right)\sum_{1\le t\neq s,t\neq r\le T}\overline{k}_{h,ts}^{i}\overline{k}_{h,tr}^{j} \\
& + \frac{1}{n(n-1)}\sum_{1\le i\neq j\le n}\left(\frac{1}{T^4} - \frac{1}{P_T^4}\right)\sum_{1\le t\neq s,r\neq q\le T}\overline{k}_{h,ts}^{i}\overline{k}_{h,rq}^{j}. \qquad (.0.28)
\end{aligned}
$$

Similarly, $\widetilde{\Delta}_{nT} = \widetilde{\Delta}_{nT,1} + \widetilde{\Delta}_{nT,2}$, where $\widetilde{\Delta}_{nT,1}$ and $\widetilde{\Delta}_{nT,2}$ are analogously defined as $\widehat{\Delta}_{nT,1}$ and $\widehat{\Delta}_{nT,2}$ but with $\{u_{it}\}$ being replaced by $\{\widetilde{u}_{it}\}$ in their definitions. It follows that $nTh(\widetilde{\Gamma}_{nT} - \widehat{\Gamma}_{nT}) = nTh(\widetilde{\Gamma}_{1nT} - \widehat{\Gamma}_{1nT}) - nTh(\Gamma\widetilde{\Delta}_{nT,1} - \widehat{\Delta}_{nT,1}) - nTh(\widetilde{\Delta}_{nT,2} - \widehat{\Delta}_{nT,2})$. We prove (i) by establishing that: (i1) $nTh(\widetilde{\Gamma}_{1nT} - \widehat{\Gamma}_{1nT}) = o_P(1)$, (i2) $nTh(\widetilde{\Delta}_{nT,1} - \widehat{\Delta}_{nT,1}) = o_P(1)$, and (i3) $nTh(\widetilde{\Delta}_{nT,2} - \widehat{\Delta}_{nT,2}) = o_P(1)$, respectively in Propositions .0.4, .0.5 and .0.6 below.

For (ii), we have

$$
\begin{aligned}
\widetilde{\sigma}_{nT}^2 - \widehat{\sigma}_{nT}^2 \quad &= \quad \frac{4R\left(\overline{k}\right)^2}{n(n-1)T}\sum_{1\le i\neq j\le n}\sum_{t=1}^{T}\left[\widetilde{f}_{ij,-t}(\widetilde{u}_{it},\widetilde{u}_{jt}) - \widehat{f}_{ij,-t}(u_{it},u_{jt})\right] \\
&= \quad \frac{4R\left(\overline{k}\right)^2}{n(n-1)T(T-1)}\sum_{1\le i\neq j\le n}\sum_{1\le t\neq s\le T}\left[k_h(\widetilde{u}_{it} - \widetilde{u}_{is})k_h(\widetilde{u}_{jt} - \widetilde{u}_{js}) - k_{h,ts}^{i}k_{h,ts}^{j}\right] \\
&= \quad \frac{4R\left(\overline{k}\right)^2}{n(n-1)T(T-1)}\sum_{1\le i\neq j\le n}\sum_{1\le t\neq s\le T}\{h^{-2}k_{h,ts}^{i}k'_{j,ts}(\Delta u_{jt} - \Delta u_{js}) \\
&\quad\quad + h^{-2}k_{h,ts}^{j}k'_{i,ts}(\Delta u_{it} - \Delta u_{is})\} + o_P(1),
\end{aligned}
$$

where $\widetilde{f}_{ij,-t}$ is analogously defined as $\widehat{f}_{ij,-t}$ with $\{u_{it}\}$ being replaced by $\{\widetilde{u}_{it}\}$, $k'_{i,ts} \equiv k'((u_{it}-u_{is})/h)$ and $\Delta u_{it} \equiv \widetilde{u}_{it} - u_{it}$. Then following the proof of Lemma .0.4 below, one can readily show that the dominant term in the last expression is $o_P(1)$ by the Chebyshev inequality.

For (iii), letting $\widetilde{E}[\varphi_{i,1r}]$ be analogously defined as $\widehat{E}[\varphi_{i,1r}]$ but with $\{u_{it}\}$ being replaced by $\{\widetilde{u}_{it}\}$, we have

$$
\begin{aligned}
\widetilde{B}_{nT}-\widehat{B}_{nT} &= \frac{h}{n_1}\sum_{r=2}^{T}\frac{T-r+1}{T-1}\sum_{1\le i\ne j\le n}\{\widetilde{E}[\varphi_{i,1r}]\widetilde{E}[\varphi_{j,1r}]-\widehat{E}[\varphi_{i,1r}]\widehat{E}[\varphi_{j,1r}]\} \\
&= \frac{h}{n_1}\sum_{r=2}^{T}\frac{T-r+1}{T-1}\sum_{1\le i\ne j\le n}\widehat{E}[\varphi_{i,1r}]\left\{\widetilde{E}[\varphi_{j,1r}]-\widehat{E}[\varphi_{j,1r}]\right\} \\
&\quad +\frac{h}{n_1}\sum_{r=2}^{T}\frac{T-r+1}{T-1}\sum_{1\le i\ne j\le n}\left\{\widetilde{E}[\varphi_{i,1r}]-\widehat{E}[\varphi_{i,1r}]\right\}\widehat{E}[\varphi_{j,1r}] \\
&\quad +\frac{h}{n_1}\sum_{r=2}^{T}\frac{T-r+1}{T-1}\sum_{1\le i\ne j\le n}\left\{\widetilde{E}[\varphi_{i,1r}]-\widehat{E}[\varphi_{i,1r}]\right\}\left\{\widetilde{E}[\varphi_{j,1r}]-\widehat{E}[\varphi_{j,1r}]\right\} \\
&\equiv D_{nT,1}+D_{nT,2}+D_{nT,3},\ \text{say.}
\end{aligned}
$$

Analogously to the proofs of Lemmas .0.3-.0.4 below, we can use the expression $\widetilde{E}[\varphi_{j,1r}] - \widehat{E}[\varphi_{j,1r}] = \frac{1}{T-r}\sum_{t=1}^{T-r}\{\overline{k}_h(\widetilde{u}_{it}-\widetilde{u}_{i,t+r})-\overline{k}^i_{h,t,t+r}\} - \frac{1}{C_T^2}\sum_{1\le t<s\le T}\{\overline{k}_h(\widetilde{u}_{it}-\widetilde{u}_{is})-\overline{k}^i_{h,ts}\}$, the Taylor expansions, and the Chebyshev inequality to show that $D_{nT,s}=o_P(1)$ for $s=1,2,3$.

Proposition .0.4 $nTh(\widetilde{\Gamma}_{1nT}-\widehat{\Gamma}_{1nT}) = o_P(1)$.

Proof. Noting that $x^2-y^2=(x-y)^2+2(x-y)y$, we have

$$
\begin{aligned}
\widetilde{\Gamma}_{1nT}-\widehat{\Gamma}_{1nT} &= \frac{1}{n(n-1)}\sum_{1\le i\ne j\le n}\int R_{ij}(u,v)^2\,dudv \\
&\quad +\frac{2}{n(n-1)}\sum_{1\le i\ne j\le n}\int R_{ij}(u,v)\left[\widehat{f}_{ij}(u,v)-\widehat{f}_i(u)\widehat{f}_j(v)\right]dudv \equiv \Gamma_{nT,1}+\Gamma_{nT,2},
\end{aligned}
$$

where $R_{ij}(u,v)\equiv \widetilde{f}_{ij}(u,v)-\widetilde{f}_i(u)\widetilde{f}_j(v)-\widehat{f}_{ij}(u,v)+\widehat{f}_i(u)\widehat{f}_j(v)$, $\widetilde{f}_i$ and $\widetilde{f}_{ij}$ are analogously defined as $\widehat{f}_i$ and $\widehat{f}_{ij}$ with $\{u_{it},u_{jt}\}_{t=1}^T$ being replaced by $\{\widetilde{u}_{it},\widetilde{u}_{jt}\}_{t=1}^T$. Expanding $k_h(\widetilde{u}_{it}-u) = h^{-1}k((\widetilde{u}_{it}-u)/h)$ in a Taylor series around $u_{it}-u$ with an integral remainder term, we have

$$
k_h(\widetilde{u}_{it}-u) = h^{-1}k_{it}(u)+h^{-2}k'_{it}(u)\Delta u_{it}+h^{-2}\Delta u_{it}\int_0^1 k^+_{it}(u,\lambda)\,d\lambda, \tag{.0.29}
$$

where $\Delta u_{it}\equiv\widetilde{u}_{it}-u_{it}$, $k_{it}(u)\equiv k((u_{it}-u)/h)$, $k'_{it}(u)\equiv k'((u_{it}-u)/h)$, $k^+_{it}(u,\lambda)\equiv k'((u_{it}-$

$u + \lambda\Delta u_{it})/h) - k'_{it}(u)$, and k' denotes the first order derivative of k. It follows that

$$\begin{aligned} R_{ij}(u,v) &= \frac{1}{T}\sum_{t=1}^{T}[k_h(\widetilde{u}_{it}-u)k_h(\widetilde{u}_{jt}-v)-k_h(u_{it}-u)k_h(u_{jt}-v)] \\ &\quad -\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}[k_h(\widetilde{u}_{it}-u)k_h(\widetilde{u}_{js}-v)-k_h(u_{it}-u)k_h(u_{js}-v)] = \sum_{r=1}^{8}R_{rij}(u,v), \end{aligned}$$

where

$$\begin{aligned} R_{1ij}(u,v) &\equiv \frac{1}{T^2h^3}\sum_{t=1}^{T}\sum_{s=1}^{T}[k_{jt}(v)-k_{js}(v)]k'_{it}(u)\Delta u_{it}, \\ R_{2ij}(u,v) &\equiv \frac{1}{T^2h^3}\sum_{t=1}^{T}\sum_{s=1}^{T}[k_{it}(u)-k_{is}(u)]k'_{js}(v)\Delta u_{js}, \\ R_{3ij}(u,v) &\equiv \frac{1}{T^2h^3}\sum_{t=1}^{T}\sum_{s=1}^{T}[k_{jt}(v)-k_{js}(v)]\Delta u_{it}\int_0^1 k_{it}^{+}(u,\lambda)\,d\lambda, \\ R_{4ij}(u,v) &\equiv \frac{1}{T^2h^3}\sum_{t=1}^{T}\sum_{s=1}^{T}[k_{it}(u)-k_{is}(u)]\Delta u_{jt}\int_0^1 k_{jt}^{+}(v,\lambda)\,d\lambda, \\ R_{5ij}(u,v) &\equiv \frac{1}{T^2h^4}\sum_{t=1}^{T}\sum_{s=1}^{T}k'_{it}(u)\Delta u_{it}\left[k'_{jt}(v)\Delta u_{jt}-k'_{js}(v)\Delta u_{js}\right], \\ R_{6ij}(u,v) &\equiv \frac{1}{T^2h^4}\sum_{t=1}^{T}\sum_{s=1}^{T}\left[k'_{jt}(v)\Delta u_{jt}-k'_{js}(v)\Delta u_{js}\right]\Delta u_{it}\int_0^1 k_{it}^{+}(u,\lambda)\,d\lambda, \\ R_{7ij}(u,v) &\equiv \frac{1}{T^2h^4}\sum_{t=1}^{T}\sum_{s=1}^{T}\left[k'_{it}(u)\Delta u_{it}-k'_{is}(u)\Delta u_{is}\right]\Delta u_{jt}\int_0^1 k_{jt}^{+}(v,\lambda)\,d\lambda, \\ R_{8ij}(u,v) &\equiv \frac{1}{T^2h^4}\sum_{t=1}^{T}\sum_{s=1}^{T}\left[\Delta u_{it}\int_0^1 k_{it}^{+}(u,\lambda)\,d\lambda-\Delta u_{is}\int_0^1 k_{is}^{+}(u,\lambda)\,d\lambda\right]\Delta u_{jt}\int_0^1 k_{jt}^{+}(v,\lambda)\,d\lambda. \end{aligned}$$

By the C_r inequality, it suffices to prove the theorem by showing that:

$$R_{rnT} \equiv \frac{Th}{n_1}\sum_{1\le i\neq j\le n}\int R_{rij}(u,v)^2\,dudv = o_P(1) \text{ for } r=1,2,...,8, \tag{.0.30}$$

and

$$S_{rnT} \equiv \frac{Th}{n_1}\sum_{1\le i\neq j\le n}\int R_{rij}(u,v)[\widehat{f}_{ij}(u,v)-\widehat{f}_i(u)\widehat{f}_j(v)]dudv = o_P(1) \text{ for } r=1,2,...,8. \tag{.0.31}$$

We prove (.0.30) in Lemma .0.3 below and (.0.31) in Lemma .0.4 below. ■

To proceed, let $\tau((X_{it}-x)/b)$ be the stack of $((X_{it}-x)/b)^{\mathbf{j}}$, $0 \le |\mathbf{j}| \le p$, in the lexicographical order such that we can write $\mathbf{S}_{iT}(x) = \frac{1}{T}\sum_{t=1}^{T}\tau(\frac{X_{it}-x}{b})\tau(\frac{X_{it}-x}{b})' w_b(X_{it}-x)$. Let $\mathbf{V}_{iT}(x) = \frac{1}{T}\sum_{t=1}^{T}\mathbf{v}_{it}(x)u_{it}$, and $\mathbf{B}_{iT}(x) = \frac{1}{T}\sum_{t=1}^{T}\mathbf{v}_{it}(x)g_i(X_{it}) - g_i(x)$, where $\mathbf{v}_{it}(x) \equiv \tau((X_{it}-x)/b)\ w_b(X_{it}-x)$. By Masry (1996b), we have $\sup_{x\in\mathscr{X}_i}||\mathbf{B}_{iT}(x)|| = O_P\left(b^{p+1}\right)$, $\sup_{x\in\mathscr{X}_i}||\mathbf{V}_{iT}\ (x)|| = O_P(T^{-1/2}b^{-d/2}\sqrt{\log T})$, and $\sup_{x\in\mathscr{X}_i}||\mathbf{S}_{iT}(x) - f_i(x)\mathbb{S}|| = O_P(b+$

$T^{-1/2}b^{-d/2}\sqrt{\log T})$, where $\mathbb{S}$ is defined in (2.4.2). Following Chen, Gao, and Li (2009, Lemma A.1), we can show that

$$\max_{1\leq i\leq n}\sup_{x\in\mathscr{X}_i}||\mathbf{S}_{iT}(x)-f_i(x)\mathbb{S}||=o_P(1). \tag{.0.32}$$

Then by the Slutsky lemma and Assumptions A.5(ii) and A.7(i), we have

$$\max_{1\leq i\leq n}\sup_{x\in\mathscr{X}_i}[\lambda_{\min}(\mathbf{S}_{iT}(x))]^{-1}=\left[\min_{1\leq i\leq n}\min_{x\in\mathscr{X}_i}f_i(x)\right]^{-1}[\lambda_{\min}(\mathbb{S})]^{-1}+o_P(1). \tag{.0.33}$$

By the standard variance and bias decomposition, we have

$$\begin{aligned} u_{it}-\widetilde{u}_{it} &= \widehat{g}_i(X_{it})-g_i(X_{it})=e_1'[\mathbf{S}_{iT}(X_{it})]^{-1}\mathbf{V}_{iT}(X_{it})+e_1'[\mathbf{S}_{iT}(X_{it})]^{-1}\mathbf{B}_{iT}(X_{it})] \\ &\equiv \mathbb{V}_{it}+\mathbb{B}_{it}. \end{aligned} \tag{.0.34}$$

Let

$$\eta_{i,ts}\equiv e_1'[\mathbf{S}_{iT}(X_{it})]^{-1}\mathbf{v}_{is}(X_{it}). \tag{.0.35}$$

We frequently need to evaluate terms associated with $\eta_{i,ts}$ and $\mathbb{B}_{it}$:

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T^2}\sum_{1\leq t,s\leq T}|\eta_{i,ts}|\right)^q=O_P(1),\ q=1,2,3, \tag{.0.36}$$

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T^3}\sum_{1\leq t,s,r\leq T}|\eta_{i,ts}\eta_{i,tr}|\right)^q=O_P(1),\ q=1,2, \tag{.0.37}$$

and

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}|\mathbb{B}_{it}|\right)^q=O_P\left(b^{q(p+1)}\right),\ q=1,2,3,4. \tag{.0.38}$$

(.0.36) and (.0.37) can be proved by using (.0.33) and the Markov inequality. For (.0.38), we first need to apply the fact that $[\mathbf{S}_{iT}(X_{it})]^{-1}\mathbf{S}_{iT}(X_{it})=I_N$ and expanding $g_i(X_{is})$ in a Taylor series around X_{it} with an integral remainder to obtain

$$\mathbb{B}_{it}=e_1'[\mathbf{S}_{iT}(X_{it})]^{-1}\frac{1}{T}\sum_{s=1}^{T}\mathbf{v}_{is}(X_{it})\Delta_{is}(X_{it})$$

where $\Delta_{is}(x)\equiv g_i(X_{is})-g_i(x)-\sum_{|\mathbf{j}|=1}^{p}\frac{1}{\mathbf{j}!}D^{\mathbf{j}}g_i(x)(X_{is}-x)^{\mathbf{j}}=\sum_{|\mathbf{j}|=p+1}\frac{1}{\mathbf{j}!}D^{\mathbf{j}}g_i(x)(X_{is}-x)^{\mathbf{j}}$ $+(p+1)\sum_{|\mathbf{j}|=p+1}\frac{1}{\mathbf{j}!}(X_{is}-x)^{\mathbf{j}}\int\left[\left(D^{\mathbf{j}}g_i\right)(x+\lambda(X_{is}-x))-D^{\mathbf{j}}g_i(x)\right](1-\lambda)^p d\lambda$. Then we can apply (.0.33), the dominated convergence theorem, and the Markov inequality to show that (.0.38) holds. Let $\mathbb{X}\equiv\{X_{it},i=1,...,n,t=1,...,T\}$ and $E^{\mathbb{X}}(\cdot)$ denote expectation conditional on X.

Lemma .0.3 $R_{rnT} \equiv \frac{Th}{n_1}\sum_{1\leq i\neq j\leq n}\int R_{rij}(u,v)^2\,dudv = o_P(1)$ *for* $r = 1,2,...,8$.

Proof. We only prove the lemma for the cases where r = 1, 3, 5, 6, and 8 as the other cases can be proved analogously. By (.0.34) and the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
R_{1nT} &\leq \frac{2}{n_1T^3h^5}\sum_{1\leq i\neq j\leq n}\int\left[\sum_{1\leq t\neq s\leq T}^{T}\left[k_{jt}(v)-k_{js}(v)\right]k'_{it}(u)\mathbb{V}_{it}\right]^2 dudv \\
&\quad + \frac{2}{n_1T^3h^5}\sum_{1\leq i\neq j\leq n}\int\left[\sum_{1\leq t\neq s\leq T}^{T}\left[k_{jt}(v)-k_{js}(v)\right]k'_{it}(u)\mathbb{B}_{it}\right]^2 dudv \\
&= \frac{2}{n_1T^5h^3}\sum_{1\leq i\neq j\leq n}\sum_{1\leq t_1\neq t_2\leq T}\sum_{t_3=1}^{T}\sum_{1\leq t_4\neq t_5\leq T}\sum_{t_6=1}^{T}\kappa_{j,t_1t_2t_4t_5}\overline{k'}_{i,t_1t_4}u_{it_3}u_{it_6}\eta_{i,t_1t_3}\eta_{i,t_4t_6} \\
&\quad + \frac{2}{n_1T^3h^3}\sum_{1\leq i\neq j\leq n}\sum_{1\leq t_1\neq t_2\leq T}\sum_{1\leq t_3\neq t_4\leq T}\kappa_{j,t_1t_2t_3t_4}\overline{k'}_{i,t_1t_3}\mathbb{B}_{it_1}\mathbb{B}_{it_3} \\
&\equiv 2R_{1nT,1} + 2R_{1nT,2},
\end{aligned}$$

where $\overline{k'}$ is a two-fold convolution of k', and

$$\kappa_{j,tsrq} \equiv \overline{k}_{j,tr} - \overline{k}_{j,tq} - \overline{k}_{j,sr} + \overline{k}_{j,sq}. \tag{.0.39}$$

Noting that $R_{1nT,r}$, $r = 1,2$, are nonnegative, it suffices to prove $R_{1nT,r} = o_P(1)$ by showing that $E^{\mathbb{X}}[R_{1nT,r}] = o_P(1)$ by the conditional Markov inequality. For $R_{1nT,1}$, we can easily verify that $E^{\mathbb{X}}[R_{1nT,1}] = \overrightarrow{R}_{1nT,1} + o_P(1)$, where

$$\begin{aligned}
\overrightarrow{R}_{1nT,1} &\equiv \frac{1}{n_1T^5h^3}\sum_{1\leq i\neq j\leq n}\sum_{t_1,t_2,t_3\text{ are distinct}}\sum_{t_4,t_5,t_6\text{ are distinct}}E\left(\kappa_{j,t_1t_2t_4t_5}\right) \\
&\quad \times E(\overline{k'}_{i,t_1t_4}u_{it_3}u_{it_6})\eta_{i,t_1t_3}\eta_{i,t_4t_6}.
\end{aligned} \tag{.0.40}$$

We consider two different cases for the time indices $\{t_1,...,t_6\}$ in the above summation: (a) for at least four different k's in $\{1,...,6\}$, $|t_l - t_k| > m$ for all $l \neq k$; (b) all the other remaining cases. We use $\overrightarrow{R}_{1nT,1a}$ and $\overrightarrow{R}_{1nT,1b}$ to denote $\overrightarrow{R}_{1nT,1}$ when the summation over the time indices are restricted to these two cases, respectively. In case (a) we can apply Lemmas .0.5 and .0.6 repeatedly and show that either $\left|h^{-1}E\left(\kappa_{j,t_1t_2t_4t_5}\right)\right| \leq Ch^{\frac{-\delta}{1+\delta}}\alpha^{\frac{\delta}{1+\delta}}(m)$ or $\left|h^{-1}E(\overline{k'}_{i,t_1t_4}u_{it_3}u_{it_6})\right| \leq Ch^{\frac{-\delta}{1+\delta}}\alpha^{\frac{\delta}{1+\delta}}(m)$ must hold. It follows that

$$\begin{aligned}
\overrightarrow{R}_{1nT,1a} &\leq \frac{Ch^{-\frac{\delta}{1+\delta}}\alpha^{\frac{\delta}{1+\delta}}(m)}{n_1T^5h}\sum_{1\leq i\neq j\leq n}\sum_{t_1,t_2,t_3\text{ are distinct}}\sum_{t_4,t_5,t_6\text{ are distinct}}\left|\eta_{i,t_1t_3}\eta_{i,t_4t_6}\right| \\
&\leq CnTh^{-\frac{1+2\delta}{1+\delta}}\alpha^{\frac{\delta}{1+\delta}}(m)\left[n^{-1}\sum_{i=1}^{n}\left(T^{-2}\sum_{1\leq t,s\leq T}|\eta_{i,ts}|\right)^2\right] \\
&= O_P\left(nTh^{-\frac{1+2\delta}{1+\delta}}\alpha^{\frac{\delta}{1+\delta}}(m)\right) = o_P(1),
\end{aligned}$$

where we have used the result in (.0.36). In case (b) noting that we have $O(n^2T^4m^2)$ terms in the summation in (.0.40) and $h^{-1}E\left(\kappa_{j,t_1t_2t_4t_5}\right)$ and $h^{-3}E(\overline{k'}_{i,t_1t_4}u_{it_3}u_{it_6})$ are bounded uniformly in all indices (as $\overline{k'}$ behaves like a second order kernel by Lemma .0.7), we can apply (.0.36) and show that $\overrightarrow{R}_{1nT,1b} = O_P\left(nhm^2/T\right) = o_P(1)$.

For $R_{1nT,2}$, we can show that $E^{\mathbb{X}}[R_{1nT,2}] = \overrightarrow{R}_{1nT,2} + o_P(1)$, where

$$\overrightarrow{R}_{1nT,2} = \frac{1}{n_1 T^3 h^3} \sum_{1 \le i \ne j \le n} \sum_{t_1, t_2, t_3, t_4 \text{ are distinct}} E\left(\kappa_{j,t_1t_2t_3t_4}\right) E(\overline{k'}_{i,t_1t_3}) \mathbb{B}_{it_1} \mathbb{B}_{it_3}.$$

We consider two cases for the time indices $\{t_1, ..., t_4\}$ in the above summation: (a) for all k's in $\{1, ..., 4\}$, $|t_l - t_k| > m$ for all $l \ne k$; (b) all the other remaining cases. We use $\overrightarrow{R}_{1nT,2a}$, and $\overrightarrow{R}_{1nT,2b}$ to denote $\overrightarrow{R}_{1nT,2}$ when the summation over the time indices are restricted to these cases, respectively. In case (a) we can use the fact that $\left|h^{-1}E\left(\kappa_{j,t_1t_2t_3t_4}\right)\right| \le Ch^{\frac{-\delta}{1+\delta}} \alpha^{\frac{\delta}{1+\delta}}(m)$, the fact that $h^{-1}E(\overline{k'}_{i,t_1t_3}) \le Ch^2$ (by Lemma .0.7) and (.0.38) to obtain $\overrightarrow{R}_{1nT,2a} = O_P(nTh^{\frac{1}{1+\delta}} b^{2(p+1)} \alpha^{\frac{\delta}{1+\delta}}(m)) = o_P(1)$. In case (b), note that $E\left(\kappa_{j,t_1t_2t_3t_4}\right)$ cannot be bounded by a term proportional to $h^{\frac{-\delta}{1+\delta}} \alpha^{\frac{\delta}{1+\delta}}(m)$ in the cases where one of the index-pair $\{(t_1,t_3), (t_1,t_4), (t_2,t_3), (t_2,t_4)\}$ has elements that do not fall from each other at least m-apart. But we can apply the fact that $\left|h^{-1}E\left(\kappa_{j,t_1t_2t_3t_4}\right)\right| \le C$, $|h^{-1}E(\overline{k'}_{i,t_1t_3})| \le Ch^2$, and (.0.38) to obtain $\overrightarrow{R}_{1nT,2b} = O_P(nmhb^{2(p+1)}) = o_P(1)$. Hence we have $E^{\mathbb{X}}[R_{1nT,2}] = o_P(1)$. Consequently, $R_{1nT} = o_P(1)$.

For R_{3nT}, write

$$\begin{aligned} R_{3nT} &= \frac{1}{n_1 T^3 h^4} \sum_{1 \le i \ne j \le n} \sum_{1 \le t_1 \ne t_2 \le T} \sum_{1 \le t_3 \ne t_4 \le T} \kappa_{j,t_1t_2t_3t_4} \Delta u_{it_1} \Delta u_{it_3} \\ &\quad \times \int \int_0^1 \int_0^1 k^+_{it_1}(u, \lambda_1) d\lambda_1 k^+_{it_3}(u, \lambda_2) d\lambda_2 du. \end{aligned}$$

As argued by Hansen (2008, pp.740-741), under Assumption A.8 there exists an integrable function k^* such that

$$\left|k^+_{it}(u, \lambda)\right| = \left|k'\left((u_{it} - u + \lambda \Delta u_{it})/h\right) - k'_{it}(u)\right| \le \lambda h^{-1} |\Delta u_{it}| k^*\left((u_{it} - u)/h\right). \quad (.0.41)$$

It follows that

$$\begin{aligned} E^{\mathbb{X}}(R_{3nT}) &\le \frac{1}{4n_1 T^3 h^5} \sum_{1 \le i \ne j \le n} \sum_{1 \le t_1 \ne t_2 \le T} \sum_{1 \le t_3 \ne t_4 \le T} \left|E\left(\kappa_{j,t_1t_2t_3t_4}\right)\right| E^{\mathbb{X}}\{\overline{k^*}_{i,t_1t_3} (\Delta u_{it_1})^2 (\Delta u_{it_3})^2\} \\ &\le \frac{1}{n_1 T^3 h^5} \sum_{1 \le i \ne j \le n} \sum_{1 \le t_1 \ne t_2 \le T} \sum_{1 \le t_3 \ne t_4 \le T} \left|E\left(\kappa_{j,t_1t_2t_3t_4}\right)\right| E^{\mathbb{X}}\{\overline{k^*}_{i,t_1t_3} [\mathbb{V}^2_{it_1} \mathbb{V}^2_{it_3} + \mathbb{B}^2_{it_1} \mathbb{B}^2_{it_3} \\ &\quad + \mathbb{V}^2_{it_1} \mathbb{B}^2_{it_3} + \mathbb{B}^2_{it_1} \mathbb{V}^2_{it_3}]\} \\ &\equiv ER_{3nT,1} + ER_{3nT,2} + ER_{3nT,3} + ER_{3nT,4}, \end{aligned}$$

where $\overline{k^*}_{i,ts} \equiv \overline{k^*}\left((u_{it} - u_{is})/h\right)$ and $\overline{k^*}$ is the two-fold convolution of k^*. It is easy to show that $ER_{3nT,1}=\overrightarrow{ER}_{3nT,1}+o_P(1)$, where $\overrightarrow{ER}_{3nT,1}=\frac{1}{n_1T^7h^5}\sum_{1\leq i\neq j\leq n}\sum_{t_1,...,t_8\text{ are all distinct}}\left|E\left(\kappa_{j,t_1t_2t_3t_4}\right)\right|$ $E(\overline{k^*}_{i,t_1t_3}u_{it_5}u_{it_6}u_{it_7}u_{it_8})\ \eta_{i,t_1t_5}\eta_{i,t_1t_6}\eta_{i,t_3t_7}\eta_{i,t_3t_8}$. We consider two cases for the time indices $\{t_1,...,t_8\}$ in the last summation: (a) for at least 4 distinct k's in $\{1,...,8\}$, $|t_l-t_k|>m$ for all $l\neq k$; (b) all the other remaining cases. We use $ER_{3nT,1a}$, and $ER_{3nT,1b}$ to denote $ER_{3nT,1}$ when the summation over the time indices are restricted to these cases, respectively. In case (a), we have $\left|h^{-1}E\left(\kappa_{j,t_1t_2t_3t_4}\right)\right|\leq Ch^{\frac{-\delta}{1+\delta}}\alpha^{\frac{\delta}{1+\delta}}(m)$ or $\left|h^{-1}E(\overline{k^*}_{i,t_1t_3}u_{it_5}u_{it_6}u_{it_7}u_{it_8})\right|\leq Ch^{\frac{-\delta}{1+\delta}}\alpha^{\frac{\delta}{1+\delta}}(m)$, and thus by (.0.37)

$$\begin{aligned}|ER_{3nT,1a}| &\leq \frac{CTh^{\frac{-\delta}{1+\delta}}\alpha^{\frac{\delta}{1+\delta}}(m)}{h^3}\sum_{i=1}^n\left\{\frac{1}{T^3}\sum_{1\leq t_1,t_5,t_6\leq T}|\eta_{i,t_1t_5}\eta_{i,t_1t_6}|\right\}^2\\ &\leq O_P\left(nTh^{-3-\frac{\delta}{1+\delta}}\alpha^{\frac{\delta}{1+\delta}}(m)\right)=o_P(1).\end{aligned}$$

In case (b), noting that $h^{-1}\left|E\left(\kappa_{j,t_1t_2t_3t_4}\right)\right|\leq C$ and $h^{-1}\left|E(\overline{k^*}_{i,t_1t_3}u_{it_5}u_{it_6}u_{it_7}u_{it_8})\right|\leq C$, we have by (.0.37)

$$|ER_{3nT,1b}|\leq\frac{m^3}{T^2h^3}\sum_{j=1}^n\left\{\frac{1}{T^3}\sum_{1\leq t_1,t_5,t_6\leq T}\left|\eta_{j,t_1t_5}\eta_{j,t_1t_6}\right|\right\}^2=O_P\left(nm^3h^{-3}/T^2\right)=o_P(1).$$

Consequently $ER_{3nT,1}=o_P(1)$. Next, it is easy to show that $ER_{3nT,2}=\overrightarrow{ER}_{3nT,2}+o_P(1)$, where $\overrightarrow{ER}_{3nT,2}=\frac{1}{n_1T^3h^5}\sum_{1\leq i\neq j\leq n}\sum_{t_1,...,t_4\text{ are all distinct}}\left|E\left(\kappa_{j,t_1t_2t_3t_4}\right)\right|\ E(\overline{k^*}_{i,t_1t_3})\mathbb{B}^2_{it_1}\mathbb{B}^2_{it_3}$. Then we can show that

$$\overrightarrow{ER}_{3nT,2}=O_P\left(nTh^{-3-\frac{\delta}{1+\delta}}\alpha^{\frac{\delta}{1+\delta}}(m)b^{4(p+1)}+nTh^{-3}b^{4(p+1)}\right)=o_P(1).$$

Hence $ER_{3nT,2}=o_P(1)$. Similarly, we can show that $ER_{3nT,r}=o_P(1)$ for $r=3,4$.

For R_{5nT}, note that

$$\begin{aligned}R_{5nT} &\leq 2n_1^{-1}Th\sum_{1\leq i\neq j\leq n}\int\int\left[\frac{1}{Th^4}\sum_{t=1}^T k'_{it}(u)\Delta u_{it}k'_{jt}(v)\Delta u_{jt}\right]^2dudv\\ &\quad+2n_1^{-1}Th\sum_{1\leq i\neq j\leq n}\int\int\left[\frac{1}{T^2h^4}\sum_{1\leq t,s\leq T}k'_{it}(u)\Delta u_{it}k'_{js}(v)\Delta u_{js}\right]^2dudv\\ &\equiv R_{5nT,1}+R_{5nT,2}.\end{aligned}$$

By (.0.36) and (.0.38) and the fact that $\overline{k'}$ behaves like second order kernel (see Lemma .0.7), we can show that

$$\begin{aligned}E^{\mathbb{X}}(R_{5nT,1}) &= \frac{2}{n_1Th^5}\sum_{1\leq i\neq j\leq n}\sum_{1\leq t_1,t_2\leq T}E^{\mathbb{X}}\left[\overline{k'}_{i,t_1t_2}\Delta u_{it_1}\Delta u_{it_2}\right]E^{\mathbb{X}}\left[\overline{k'}_{j,t_1t_2}\Delta u_{jt_1}\Delta u_{jt_2}\right]\\ &= O_P\left(nTh\left(T^{-2}b^{-2d}+b^{4(p+1)}\right)\right)=o_P(1).\end{aligned}$$

It follows that $R_{5nT,1}=o_P(1)$. By the same token, $R_{5nT,2}=o_P(1)$. Consequently $R_{5nT}=$

$o_P(1)$.

For R_{6nT}, write $R_{6ij}(u,v) = \frac{1}{Th^4}\sum_{t=1}^T k'_{jt}(v)\Delta u_{jt}\Delta u_{it}\int_0^1 k_{it}^+(u,\lambda)d\lambda - \frac{1}{T^2h^4}\sum_{t=1}^T\sum_{s=1}^T k'_{js}(v)$ $\Delta u_{js}\Delta u_{it}\int_0^1 k_{it}^+(u,\lambda)d\lambda \equiv R_{6ij,1}(u,v) - R_{6ij,2}(u,v)$. Define $R_{6nT,1}$ and $R_{6nT,2}$ analogously as R_{6nT} but with $R_{6ij}(u,v)$ being replaced by $R_{6ij,1}(u,v)$ and $R_{6ij,2}(u,v)$, respectively. Then

$$R_{6nT,1} = \frac{1}{n_1Th^6}\sum_{1\le i\ne j\le n}\sum_{1\le t,s\le T}\overline{k'}_{j,ts}\Delta u_{jt}\Delta u_{js}\Delta u_{it}\Delta u_{is}\int\int_0^1 k_{it}^+(u,\lambda_1)d\lambda_1\int_0^1 k_{is}^+(u,\lambda_2)d\lambda_2du$$

Using (.0.36) and (.0.38), we have

$$\begin{aligned}
E^{\mathbb{X}}(R_{6nT,1}) &= \frac{1}{n_1Th^6}\sum_{1\le i\ne j\le n}\sum_{1\le t,s\le T}E^{\mathbb{X}}\left[\overline{k'}_{j,ts}\Delta u_{jt}\Delta u_{js}\right] \\
&\quad\times E^{\mathbb{X}}\left[\Delta u_{it}\Delta u_{is}\int\int_0^1 k_{it}^+(u,\lambda_1)d\lambda_1\int_0^1 k_{is}^+(u,\lambda_2)d\lambda_2du\right] \\
&\le \frac{1}{4n_1Th^7}\sum_{1\le i\ne j\le n}\sum_{1\le t,s\le T}\left|E^{\mathbb{X}}\left[\overline{k'}_{j,ts}\Delta u_{jt}\Delta u_{js}\right]\right|E^{\mathbb{X}}\{\overline{k^*}_{i,ts}(\Delta u_{it}\Delta u_{is})^2\} \\
&= O_P\left(nTh^{-3}\left(T^{-3}b^{-3d}+b^{6(p+1)}\right)\right) = o_P(1).
\end{aligned}$$

Similarly, we can show that $E^{\mathbb{X}}(R_{8nT,1}) = O_P\left(nTh^{-7}\left(T^{-4}b^{-4d}+b^{8(p+1)}\right)\right) = o_P(1)$. ■

Lemma .0.4 *$S_{rnT} \equiv \frac{Th}{n_1}\sum_{1\le i\ne j\le n}\int R_{rij}(u,v)[\widehat{f}_{ij}(u,v) - \widehat{f}_i(u)\widehat{f}_j(v)]dudv = o_P(1)$ for $r = 1,2,...,8$.*

Proof. We only prove the lemma for the cases where $r = 1$, 3, and 5 as the other cases can be proved analogously. Decompose

$$\begin{aligned}
S_{1nT} &= \frac{1}{T^3n_1h^4}\sum_{1\le i\ne j\le n}\sum_{1\le t_1\ne t_2\le T}\sum_{1\le t_3\ne t_4\le T}\int\int[k_{jt_1}(v)-k_{jt_2}(v)][k_{jt_3}(v)-k_{jt_4}(v)] \\
&\quad\times k'_{it_1}(u)k_{it_3}(u)\Delta u_{it_1}dudv \\
&= \frac{1}{T^3n_1h^2}\sum_{1\le i\ne j\le n}\sum_{1\le t_1\ne t_2\le T}\sum_{1\le t_3\ne t_4\le T}\kappa_{j,t_1t_2t_3t_4}k^+_{i,t_1t_3}\Delta u_{it_1} \\
&= \frac{1}{T^4n_1h^2}\sum_{1\le i\ne j\le n}\sum_{1\le t_1\ne t_2\le T}\sum_{1\le t_3\ne t_4\le T}\sum_{t_5=1}^T\kappa_{j,t_1t_2t_3t_4}k^+_{i,t_1t_3}u_{it_5}\eta_{i,t_1t_5} \\
&\quad+\frac{1}{T^3n_1h^2}\sum_{1\le i\ne j\le n}\sum_{1\le t_1\ne t_2\le T}\sum_{1\le t_3\ne t_4\le T}\kappa_{j,t_1t_2t_3t_4}k^+_{i,t_1t_3}\mathbb{B}_{it_1} \\
&\equiv S_{1nT,1}+S_{1nT,2},
\end{aligned}$$

where $k^+_{i,ts} \equiv k^+((u_{it}-u_{is})/h) \equiv h^{-1}\int k'_{it}(u)k_{is}(u)du$, and $\kappa_{j,tsrq}$ is defined in (.0.39). To show $S_{1nT,1} = o_P(1)$, we can first show that $S_{1nT,1} = \overrightarrow{S}_{1nT,1} + o_P(1)$, where $\overrightarrow{S}_{1nT,1}$ is analogously defined as $S_{1nT,1}$ but with all distinct time indices inside the summation. Second, we can decompose $\overrightarrow{S}_{1nT,1}$ as $\overrightarrow{S}_{1nT,11}+\overrightarrow{S}_{1nT,12}$ where $\overrightarrow{S}_{1nT,11}$ is analogously defined as $\overrightarrow{S}_{1nT,1}$ but with only $i<j$ terms in the summation and $\overrightarrow{S}_{1nT,12} \equiv \overrightarrow{S}_{1nT,1} - \overrightarrow{S}_{1nT,11}$. Let $e_{i,tsr} \equiv k^+_{i,ts}u_{ir}$, $e^c_{i,tsr} \equiv e_{i,tsr} - E(e_{i,tsr})$, and $\kappa^c_{j,t_1t_2t_3t_4} \equiv \kappa_{j,t_1t_2t_3t_4} - E(\kappa_{j,t_1t_2t_3t_4})$. Then we can

decompose $\overrightarrow{S}_{1nT,11}$ as follows

$$\begin{aligned}
\overrightarrow{S}_{1nT,11} &= \frac{1}{T^4 n_1 h^2} \sum_{1\leq i\neq j\leq n} \sum_{t_1,\ldots,t_5 \text{ are all distinct}} \kappa_{j,t_1t_2t_3t_4} e_{i,t_1t_3t_5} \eta_{i,t_1t_5} + o_P(1) \\
&= \frac{1}{T^4 n_1 h^2} \sum_{1\leq i\neq j\leq n} \sum_{t_1,\ldots,t_5 \text{ are all distinct}} \{\kappa^c_{j,t_1t_2t_3t_4} e^c_{i,t_1t_3t_5} \eta_{i,t_1t_5} + \kappa^c_{j,t_1t_2t_3t_4} E(e_{i,t_1t_3t_5}) \eta_{i,t_1t_5} \\
&\quad + E(\kappa_{j,t_1t_2t_3t_4}) e^c_{i,t_1t_3t_5} \eta_{i,t_1t_5} + E(\kappa_{j,t_1t_2t_3t_4}) E(e_{i,t_1t_3t_5}) \eta_{i,t_1t_5}\} + o_P(1) \\
&\equiv \overrightarrow{S}_{1nT,111} + \overrightarrow{S}_{1nT,112} + \overrightarrow{S}_{1nT,113} + \overrightarrow{S}_{1nT,114} + o_P(1).
\end{aligned}$$

For $\overrightarrow{S}_{1nT,111}$, we have

$$\begin{aligned}
E^{\mathbb{X}}\left[(\overrightarrow{S}_{1nT,111})^2\right] &= \frac{1}{T^8 n_1^2 h^4} \sum_{1\leq i<j\leq n} \sum_{t_1,\ldots,t_5 \text{ are all distinct}} \sum_{t_6,\ldots,t_{10} \text{ are all distinct}} \eta_{i,t_1t_5} \eta_{i,t_6t_{10}} \\
&\quad \times E\left[e^c_{i,t_1t_3t_5} e^c_{i,t_6t_8t_{10}}\right] E[\kappa^c_{j,t_1t_2t_3t_4} \kappa^c_{j,t_6t_7t_8t_9}].
\end{aligned}$$

We consider two cases for the time indices $\{t_1,\ldots,t_{10}\}$: (a) for at least six different k's, $|t_l - t_k| > m$ for all $l \neq k$; (b) all the other remaining cases. We use $ES_{1,111a}$ and $ES_{1,111b}$ to denote the summation corresponding to these two cases, respectively. In the first case, $ES_{1,111a} \leq CT^2 h^{\frac{-2\delta}{1+\delta}} \alpha^{\frac{\delta}{1+\delta}}(m) \sum_{i=1}^n \{T^{-2} \sum_{1\leq t\neq s\leq T} |\eta_{i,ts}|\}^2 = O_P(T^2 h^{\frac{-2\delta}{1+\delta}} \alpha^{\frac{\delta}{1+\delta}}(m)) = o_P(1)$. In the second case,

$$ES_{1,111b} \leq CT^{-1} n_1^{-1} m^3 \sum_{i=1}^n \left(T^{-2} \sum_{1\leq t\neq s\leq T} |\eta_{i,ts}|\right)^2 = O_P\left(m^3/T\right) = o_P(1).$$

It follows that $\overrightarrow{S}_{1nT,111} = o_P(1)$. Analogously, we can show that $\overrightarrow{S}_{1nT,11r} = o_P(1)$ for $r = 2,3,4$. So $\overrightarrow{S}_{1nT,11} = o_P(1)$. Also $\overrightarrow{S}_{1nT,12} = o_P(1)$ by the same argument. Thus $\overrightarrow{S}_{1nT,1} = o_P(1)$ and $S_{1nT,1} = o_P(1)$. Analogously, we can show that $S_{1nT,2} = o_P(1)$. Consequently, $S_{1nT} = o_P(1)$.

For S_{3nT}, we have

$$\begin{aligned}
S_{3nT} &= \frac{1}{n_1 T^3 h^4} \sum_{1\leq i\neq j\leq n} \sum_{1\leq t_1\neq t_2\leq T} \sum_{1\leq t_3\neq t_4\leq T} \int [k_{jt_1}(v) - k_{jt_2}(v)][k_{jt_3}(v) - k_{jt_4}(v)] dv \Delta u_{it_1} \\
&\quad \times \int k_{it_3}(u) \int_0^1 k^+_{it_1}(u,\lambda) d\lambda du \\
&= \frac{1}{n_1 T^3 h^3} \sum_{1\leq i\neq j\leq n} \sum_{1\leq t_1\neq t_2\leq T} \sum_{1\leq t_3\neq t_4\leq T} \kappa_{j,t_1t_2t_3t_4} \Delta u_{it_1} \int k_{it_3}(u) \int_0^1 k^+_{it_1}(u,\lambda) d\lambda du \\
&= \frac{1}{n_1 T^3 h^3} \sum_{1\leq i\neq j\leq n} \sum_{1\leq t_1,t_2\leq T} \sum_{1\leq t_3,t_4\leq T} E[\kappa_{j,t_1t_2t_3t_4}] \Delta u_{it_1} \int k_{it_3}(u) \int_0^1 k^+_{it_1}(u,\lambda) d\lambda du \\
&\quad + \frac{1}{n_1 T^3 h^3} \sum_{1\leq i\neq j\leq n} \sum_{1\leq t_1,t_2\leq T} \sum_{1\leq t_3,t_4\leq T} \kappa^c_{j,t_1t_2t_3t_4} \Delta u_{it_1} \int k_{it_3}(u) \int_0^1 k^+_{it_1}(u,\lambda) d\lambda du \\
&\equiv S_{3nT,1} + S_{3nT,2}.
\end{aligned}$$

Noting that $h^{-1}\kappa_{j,t_1t_2t_3t_4} = \varphi_{j,t_1t_3} - \varphi_{j,t_1t_4} - \varphi_{j,t_2t_3} + \varphi_{j,t_2t_4}$, we can decompose $S_{3nT,r} = S_{3nT,r1} -$

$S_{3nT,r2}-S_{3nT,r3}+S_{3nT,r4}$, where $S_{3nT,r1}$, $S_{3nT,r2}$, $S_{3nT,r3}$, and $S_{3nT,r4}$ are defined analogously as $S_{3nT,r}$ with $E\left[\kappa_{j,t_1t_2t_3t_4}\right]$ (for $r=1$) or $\kappa^c_{j,t_1t_2t_3t_4}$ (for $r=2$) being respectively replaced by $hE\left[\varphi_{j,t_1t_3}\right]$, $hE\left[\varphi_{j,t_1t_4}\right]$, $hE\left[\varphi_{j,t_2t_3}\right]$, and $hE\left[\varphi_{j,t_2t_4}\right]$ (for $r=1$), or by $h\varphi^c_{j,t_1t_3}$, $h\varphi^c_{j,t_1t_4}$, $h\varphi^c_{j,t_2t_3}$, and $h\varphi^c_{j,t_2t_4}$ (for $r=2$). WLOG we prove $S_{3nT,r}=o_P(1)$ by showing that $S_{3nT,r1}=o_P(1)$ for $r=1,2$. For $S_{3nT,11}$, noting that

$$\left|\int k_{is}(u)\int_0^1 k^+_{it}(u,\lambda)\,d\lambda du\right|\leq\frac{1}{2}\left|\Delta u_{it}\right|h^{-1}\int\left|k\left(\left(u_{is}-u\right)/h\right)\right|k^*\left(\left(u_{it}-u\right)/h\right)du=\frac{1}{2}\left|\Delta u_{it}\right|k^{\ddagger}_{i,ts},$$

where $k^{\ddagger}_{i,ts}\equiv k^{\ddagger}\left(\left(u_{it}-u_{is}\right)/h\right)$ and $k^{\ddagger}(u)\equiv\int k^*(u-v)\left|k(v)\right|dv$, we have

$$\begin{aligned}
\left|S_{3nT,11}\right| &\leq \frac{1}{2n_1Th}\sum_{1\leq i\neq j\leq n}\sum_{1\leq t,s\leq T}\left|E(\varphi_{j,ts})\right|\left(\Delta u_{it}\right)^2k^{\ddagger}_{i,ts}\\
&= \frac{1}{2n_1Th}\left\{\sum_{1\leq i\neq j\leq n}\sum_{|t-s|\geq m}+\sum_{1\leq i\neq j\leq n}\sum_{0<|t-s|<m}\right\}\left|E(\varphi_{j,ts})\right|\left(\Delta u_{it}\right)^2k^{\ddagger}_{i,ts}\\
&\equiv S_{3nT,11a}+S_{3nT,11b}.
\end{aligned}$$

By the fact that $\left|E(\varphi_{j,ts})\right|\leq Ch^{-\frac{\delta}{1+\delta}}\alpha^{\frac{\delta}{1+\delta}}\left(|t-s|\right)$ (see (.0.1)), we have

$$\begin{aligned}
S_{3nT,11a} &\leq Ch^{-1-\frac{\delta}{1+\delta}}\alpha^{\frac{\delta}{1+\delta}}(m)\sum_{i=1}^n\sum_{t=1}^T\left(\Delta u_{it}\right)^2k^{\ddagger}_{i,ts}\\
&= h^{-1-\frac{\delta}{1+\delta}}\alpha^{\frac{\delta}{1+\delta}}(m)\,O_P\left(nT\left(T^{-1}b^{-d}+b^{2(p+1)}\right)\right)=o_P(1).
\end{aligned}$$

For $S_{3nT,11b}$ we can apply (.0.36) and (.0.38) and the Markov inequality to show that $S_{3nT,11b}=O_P(nm\left(T^{-1}b^{-d}+b^{2(p+1)}\right))=o_P(1)$. It follows that $S_{3nT,11}=o_P(1)$.

For $S_{3nT,21}$, write $S_{3nT,21}=\frac{1}{n_1Th^2}\{\sum_{1\leq i<j\leq n}+\sum_{1\leq j<i\leq n}\}\sum_{1\leq t_1,t_2\leq T}\varphi^c_{j,t_1t_2}\Delta u_{it_1}\int k_{it_2}(u)\int_0^1 k^+_{it_1}(u,\lambda)d\lambda du\equiv S_{3nT,211}+S_{3nT,212}$. Note that $E^{\mathbb{X}}\left[S_{3nT,211}\right]=0$, and $E^{\mathbb{X}}\left[\left(S_{3nT,211}\right)^2\right]=S_3+o_P(1)$, where

$$\begin{aligned}
S_3 &\equiv \frac{1}{\left(n_1Th^2\right)^2}\sum_{1\leq i_1\neq i_2<j\leq n}\sum_{1\leq t_1,t_2\leq T}\sum_{1\leq t_3,t_4\leq T}E\left(\varphi^c_{j,t_1t_2}\varphi^c_{j,t_3t_4}\right)\\
&\quad\times E^{\mathbb{X}}\left[\Delta u_{i_1t_1}\int k_{i_1t_2}(u)\int_0^1 k^+_{i_1t_1}(u,\lambda)\,d\lambda du\right]\\
&\quad\times E^{\mathbb{X}}\left[\Delta u_{i_2t_3}\int k_{i_2t_4}(u)\int_0^1 k^+_{i_2t_3}(u,\lambda)\,d\lambda du\right]\\
&\leq \frac{1}{4\left(n_1Th^2\right)^2}\sum_{1\leq i_1\neq i_2<j\leq n}\sum_{1\leq t_1,t_2\leq T}\sum_{1\leq t_3,t_4\leq T}\left|E\{\varphi^c_{j,t_1t_2}\varphi^c_{j,t_3t_4}\}\right|E^{\mathbb{X}}\left[\left(\Delta u_{i_1t_1}\right)^2k^{\ddagger}_{i_1,t_1t_2}\right]\\
&\quad\times E^{\mathbb{X}}\left[\left(\Delta u_{i_2t_3}\right)^2k^{\ddagger}_{i_2,t_3t_4}\right].
\end{aligned}$$

It is easy to show that the dominant term on the r.h.s. of the last equation is given by

$\overline{S}_3 = \left(n_1 T h^2\right)^{-2} \sum_{1 \le i_1 \neq i_2 < j \le n} \sum_{t_1, t_2, t_3, t_4 \text{ are all distinct}} \left| E(\varphi^c_{j,t_1 t_2} \varphi^c_{j,t_3 t_4}) \right| E^{\mathbb{X}}[(\Delta u_{i_1 t_1})^2 k^{+}_{i_1, t_1 t_2}]$ $\times E^{\mathbb{X}}[(\Delta u_{i_2 t_3})^2 \; k^{+}_{i_2, t_3 t_4}]$. We consider two cases for the time indices $\{t_1, ..., t_4\}$ in the last summation: (a) there exists at least an integer $k \in \{1, ..., 4\}$, $|t_l - t_k| > m$ for all $l \neq k$; (b) all the other remaining cases. We use $\overline{S}_{3a}$, and $\overline{S}_{3b}$ to denote $\overline{S}_3$ when the summation over the time indices are restricted to these cases, respectively. In case (a), WLOG we assume that t_1 lies at least m-apart from $\{t_2, t_3, t_4\}$. Then by Lemma .0.5, $E\{\varphi^c_{j,t_1 t_2} \varphi^c_{j,t_3 t_4}\} \le \left| E\{E_{t_1}(\varphi^c_{j,t_1 t_2}) \varphi^c_{j,t_3 t_4}\} \right| + C h^{\frac{-2\delta}{1+\delta}} \alpha^{\frac{\delta}{1+\delta}}(m) = C h^{\frac{-2\delta}{1+\delta}} \alpha^{\frac{\delta}{1+\delta}}(m)$ as $E_{t_1}(\varphi^c_{j,t_1 t_2})$ is nonrandom.

$$\begin{aligned}
\overline{S}_{3a} &\le \frac{C h^{\frac{-2\delta}{1+\delta}} \alpha^{\frac{\delta}{1+\delta}}(m)}{n_1 T^2 h^2} \left\{ \sum_{i=1}^{n} \sum_{1 \le t_1 \neq t_2 \le T} E^{\mathbb{X}} \left\{ (\Delta u_{it_1})^2 h^{-1} k^{\ddagger}_{i,t_1 t_2} \right\} \right\}^2 \\
&= n T^2 h^{-2 - \frac{2\delta}{1+\delta}} \alpha^{\frac{\delta}{1+\delta}}(m) O_P\left(T^{-2} b^{-2d} + b^{4(p+1)}\right) = o_P(1).
\end{aligned}$$

In case (b), noting that the total number of terms in the summation is of order $O\left(n^3 T^2 m^2\right)$, we can easily obtain $\left|\overline{S}_{3b}\right| = O\left(n m^2 h^{-2}\right) O_P\left(T^{-2} b^{-2d} + b^{4(p+1)}\right) = O_P\left(n m^2 h^{-2} T^{-2} b^{-2d} \; + n m^2 h^{-2} b^{4(p+1)})\right) = o_P(1)$. Consequently $S_3 = o_P(1)$ and $S_{3nT,211} = o_P(1)$ by the conditional Chebyshev inequality.Next we study S_{5nT}. Write

$S_{5nT} = \frac{Th}{n_1} \left(\sum_{1 \le i < j \le n} + \sum_{1 \le j < i \le n}\right) \int\int R_{5ij}(u,v) [\widehat{f}_{ij}(u,v) - \widehat{f}_i(u) \widehat{f}_j(v)] \, du dv \equiv S_{5nT,1} + S_{5nT,2}$. It suffices to show that $S_{5nT,1} = o_P(1)$ and $S_{5nT,2} = o_P(1)$. We only prove the former claim as the latter one can be proved analogously. It is easy to show that

$$\begin{aligned}
S_{5nT,1} &= \frac{1}{n_1 T^3 h^5} \sum_{1 \le i < j \le n} \int\int \sum_{1 \le t_1 \neq t_2 \le T} \sum_{1 \le t_3 \neq t_4 \le T} k'_{it_1}(u) \Delta u_{it_1} [k'_{jt_1}(v) \Delta u_{jt_1} - k'_{jt_2}(v) \Delta u_{jt_2}] \\
&\quad \times k_{jt_3}(v) [k_{it_3}(u) - k_{it_4}(u)] \, du dv \\
&= \frac{1}{n_1 T^3 h^5} \sum_{1 \le i < j \le n} \sum_{1 \le t_1 \neq t_2 \le T} \sum_{1 \le t_3 \neq t_4 \le T} (k^{\dagger}_{i,t_1 t_3} - k^{\dagger}_{i,t_1 t_4}) \Delta u_{it_1} (k^{\dagger}_{j,t_1 t_3} \Delta u_{jt_1} - k^{\dagger}_{j,t_2 t_3} \Delta u_{jt_2}) \\
&= \overrightarrow{S}_{5nT,1} + o_P(1),
\end{aligned}$$

where $k^{\dagger}_{i,ts} \equiv k^{\dagger}((u_{it} - u_{is})/h)$, $k^{\dagger}(u) \equiv \int k'(u - v) k(v) dv$,

$$\overrightarrow{S}_{5nT,1} = \frac{1}{n_1 T^3 h^5} \sum_{1 \le i < j \le n} \sum_{t_1 \ldots t_4 \text{ are all distinct}} (k^{\dagger}_{i,t_1 t_3} - k^{\dagger}_{i,t_1 t_4}) \Delta u_{it_1} (k^{\dagger}_{j,t_1 t_3} \Delta u_{jt_1} - k^{\dagger}_{j,t_2 t_3} \Delta u_{jt_2}),$$

and the $o_P(1)$ terms arises when the cardinality of the set $\{t_1, t_2, t_3, t_4\}$ is 3 or 2. In particular, by the standard bias-variance decomposition (for Δu_{it_1} and Δu_{jt_2}) and the conditional

Chebyshev inequality, we can show that

$$\begin{aligned}
&\frac{1}{n_1T^3h^5}\sum_{1\le i<j\le n}\sum_{\substack{t_1\ne t_2,t_3\ne t_4\\ \#\{t_1\ldots t_4\}=3\text{ or }2}}(k^\dagger_{i,t_1t_3}-k^\dagger_{i,t_1t_4})\Delta u_{it_1}(k^\dagger_{j,t_1t_3}\Delta u_{jt_1}-k^\dagger_{j,t_2t_3}\Delta u_{jt_2})\\
=\;& O_P\left(h^{-5}\left(T^{-1}+T^{-3/2}b^{-d}\right)+nhb^{2(p+1)}\right)=o_P(1).
\end{aligned}$$

Decompose $\overrightarrow{S}_{5nT,1}=\overrightarrow{S}_{5nT,11}+\overrightarrow{S}_{5nT,12}$, where

$$\begin{aligned}
\overrightarrow{S}_{5nT,11}&\equiv\frac{1}{n_1T^3h^5}\sum_{1\le i<j\le n}\sum_{t_1\ldots t_4\text{ are all distinct}}(k^\dagger_{i,t_1t_3}-k^\dagger_{i,t_1t_4})\Delta u_{it_1}(k^\dagger_{j,t_1t_3}-k^\dagger_{j,t_2t_3})\Delta u_{jt_1},\text{ and}\\
\overrightarrow{S}_{5nT,12}&\equiv\frac{1}{n_1T^3h^5}\sum_{1\le i<j\le n}\sum_{t_1\ldots t_4\text{ are all distinct}}(k^\dagger_{i,t_1t_3}-k^\dagger_{i,t_1t_4})\Delta u_{it_1}k^\dagger_{j,t_2t_3}(\Delta u_{jt_1}-\Delta u_{jt_2}).
\end{aligned}$$

We prove $\overrightarrow{S}_{5nT,1}=o_P(1)$ by showing that $\overrightarrow{S}_{5nT,11}=o_P(1)$ and $\overrightarrow{S}_{5nT,12}=o_P(1)$. We only prove the former claim as the latter can be proved analogously. Let

$$\mathscr{S}(A,B)\equiv\frac{1}{n_1T^3h^5}\sum_{1\le i<j\le n}\sum_{1\le t_1\ne t_2\le T}\sum_{1\le t_3\ne t_4\le T}\left(k^\dagger_{i,t_1t_3}-k^\dagger_{i,t_1t_4}\right)A_{it_1}\left(k^\dagger_{j,t_1t_3}-k^\dagger_{j,t_2t_3}\right)B_{jt_2}.$$

By (.0.34), we have $\overrightarrow{S}_{5nT,11}=\mathscr{S}(\Delta u,\Delta u)=\mathscr{S}(\mathbb{V},\mathbb{V})+\mathscr{S}(\mathbb{B},\mathbb{B})+\mathscr{S}(\mathbb{V},\mathbb{B})+\mathscr{S}(\mathbb{B},\mathbb{V}).$ It suffices to show that each term in the last expression is $o_P(1)$.

First, we consider $\mathscr{S}(\mathbb{V},\mathbb{V})$. It is easy to verify that

$$\mathscr{S}(\mathbb{V},\mathbb{V})=S_1+o_P(1)$$

where

$$S_1\equiv\frac{1}{n_1T^5h^5}\sum_{1\le i<j\le n}\sum_{t_1\ldots t_6\text{ are distinct}}\left(k^\dagger_{i,t_1t_3}-k^\dagger_{i,t_1t_4}\right)u_{it_5}\left(k^\dagger_{j,t_1t_3}-k^\dagger_{j,t_2t_3}\right)u_{jt_6}\eta_{i,t_1t_5}\eta_{i,t_2t_6}.$$

Let $\varphi^\dagger_{i,ts}\equiv k^\dagger_{i,ts}-E_t(k^\dagger_{i,ts})-E_s(k^\dagger_{i,ts})+E_tE_s(k^\dagger_{i,ts})$. Then $k^\dagger_{j,t_1t_3}-k^\dagger_{j,t_1t_4}=\varphi^\dagger_{j,t_1t_3}-\varphi^\dagger_{j,t_1t_4}+E_{t_1}(k^\dagger_{j,t_1t_3})-E_{t_1}(k^\dagger_{j,t_1t_4})$ and $k^\dagger_{j,t_1t_3}-k^\dagger_{j,t_2t_3}=\varphi^\dagger_{j,t_1t_3}-\varphi^\dagger_{j,t_2t_3}+E_{t_3}(k^\dagger_{j,t_1t_3})-E_{t_3}(k^\dagger_{j,t_2t_3})$. With these we can decompose S_1 as follows:

$$\begin{aligned}
S_1\;=\;&\frac{1}{n_1T^3h^5}\sum_{1\le i<j\le n}\sum_{t_1\ldots t_6\text{ are distinct}}\{[\varphi^\dagger_{i,t_1t_3}-\varphi^\dagger_{i,t_1t_4}][\varphi^\dagger_{j,t_1t_3}-\varphi^\dagger_{j,t_2t_3}]\\
&+[\varphi^\dagger_{i,t_1t_3}-\varphi^\dagger_{i,t_1t_4}][E_{t_3}(k^\dagger_{j,t_1t_3})-E_{t_3}(k^\dagger_{j,t_2t_3})]+[E_{t_1}(k^\dagger_{i,t_1t_3})-E_{t_1}(k^\dagger_{i,t_1t_4})][\varphi^\dagger_{j,t_1t_3}-\varphi^\dagger_{j,t_2t_3}]\\
&+[E_{t_1}(k^\dagger_{i,t_1t_3})-E_{t_1}(k^\dagger_{i,t_1t_4})][E_{t_3}(k^\dagger_{j,t_1t_3})-E_{t_3}(k^\dagger_{j,t_2t_3})]\}u_{it_5}u_{jt_6}\eta_{i,t_1t_5}\eta_{j,t_2t_6}\\
\equiv\;&S_{11}+S_{12}+S_{13}+S_{14},\text{ say,}
\end{aligned}\tag{.0.42}$$

where the definitions of S_{1r}, $r=1,2,3,4$, are self-evident. We further decompose S_{11} as follows:

$$
\begin{aligned}
S_{11} &= \frac{1}{n_1 T^5 h^5} \sum_{1\le i<j\le n} \sum_{t_1\ldots t_6 \text{ are distinct}} \{\varphi^{\dagger}_{i,t_1t_3}\varphi^{\dagger}_{j,t_1t_3} - \varphi^{\dagger}_{i,t_1t_3}\varphi^{\dagger}_{j,t_2t_3} - \varphi^{\dagger}_{i,t_1t_4}\varphi^{\dagger}_{j,t_1t_3} \\
&\quad +\varphi^{\dagger}_{i,t_1t_4}\varphi^{\dagger}_{j,t_2t_3}\} u_{it_5}\varphi^{\dagger}_{j,t_1t_3}u_{jt_6}\eta_{i,t_1t_5}\eta_{j,t_2t_6} \\
&\equiv S_{111} - S_{112} - S_{113} + S_{114}
\end{aligned}
$$

To analyze S_{111}, let $A_{i_1j_1,i_2j_2}(t_1,\ldots,t_{10}) \equiv \varphi^{\dagger}_{i_1,t_1t_3}u_{i_1t_4}\varphi^{\dagger}_{j_1,t_1t_3}u_{j_1t_5}\eta_{i_1,t_1t_4}\eta_{j_1,t_2t_5}\varphi^{\dagger}_{i_2,t_6t_8}u_{i_2t_9}\varphi^{\dagger}_{j_2,t_6t_8}u_{j_2t_{10}}$ $\eta_{i_2,t_6t_9}\eta_{j_2,t_7t_{10}}$. Then

$$
\begin{aligned}
& E^{\mathbb{X}}\left[(S_{111})^2\right] \\
=& \frac{1}{(n_1T^4h^5)^2} \sum_{1\le i_1<j_1\le n} \sum_{1\le i_2<j_2\le n} \sum_{t_1\ldots t_5 \text{ are distinct}} \sum_{t_6\ldots t_{10} \text{ are distinct}} E^{\mathbb{X}}\left[A_{i_1j_1,i_2j_2}(t_1,\ldots,t_{10})\right] \\
=& \frac{1}{(n_1T^4h^5)^2} \sum_{\substack{1\le i_1<j_1\le n, 1\le i_2<j_2\le n, \\ i_1,i_2,j_1,j_2 \text{ are all distinct}}} \sum_{t_1\ldots t_5 \text{ are distinct}} \sum_{t_6\ldots t_{10} \text{ are distinct}} E^{\mathbb{X}}\left[A_{i_1j_1,i_2j_2}(t_1,\ldots,t_{10})\right] \\
& + \frac{1}{(n_1T^4h^5)^2} \sum_{\substack{1\le i_1<j_1\le n, 1\le i_2<j_2\le n, \\ \#\{i_1,i_2,j_1,j_2\}=3}} \sum_{t_1\ldots t_5 \text{ are distinct}} \sum_{t_6\ldots t_{10} \text{ are distinct}} E^{\mathbb{X}}\left[A_{i_1j_1,i_2j_2}(t_1,\ldots,t_{10})\right] \\
& + \frac{1}{(n_1T^4h^5)^2} \sum_{1\le i<j\le n} \sum_{t_1\ldots t_5 \text{ are distinct}} \sum_{t_6\ldots t_{10} \text{ are distinct}} E^{\mathbb{X}}\left[A_{ij,ij}(t_1,\ldots,t_{10})\right] \\
\equiv& ES_{111,1} + ES_{111,2} + ES_{111,3},
\end{aligned}
$$

We prove $E^{\mathbb{X}}[(S_{111})^2] = o_P(1)$ by showing that $ES_{111,r} = o_P(1)$ for $r = 1,3$ as one can analogously show that $ES_{111,2} = o_P(1)$. Write $ES_{111,1}$ as

$$
\begin{aligned}
ES_{111,1} &= \frac{1}{(n_1T^4h^5)^2} \sum_{\substack{1\le i_1<j_1\le n, 1\le i_2<j_2\le n, \\ i_1,i_2,j_1,j_2 \text{ are all distinct}}} \sum_{t_1\ldots t_5 \text{ are distinct}} \sum_{t_6\ldots t_{10} \text{ are distinct}} E\left(\varphi^{\dagger}_{i_1,t_1t_3}u_{i_1t_4}\right) \\
&\quad \times E\left(\varphi^{\dagger}_{j_1,t_1t_3}u_{j_1t_5}\right) E\left(\varphi^{\dagger}_{i_2,t_6t_8}u_{i_2t_9}\right) E\left(\varphi^{\dagger}_{j_2,t_6t_8}u_{j_2t_{10}}\right) \eta_{i_1,t_1t_4}\eta_{j_1,t_2t_5}\eta_{i_2,t_6t_9}\eta_{j_2,t_7t_{10}}
\end{aligned}
$$

Let $\mathscr{G}_1 \equiv \{t_1,t_3,t_4\}$, $\mathscr{G}_2 \equiv \{t_1,t_3,t_5\}$, $\mathscr{G}_3 \equiv \{t_6,t_8,t_9\}$, and $\mathscr{G}_4 \equiv \{t_6,t_8,t_{10}\}$. We consider two cases: (a) there exists at least one time index that belongs to either one of these four groups and lies at least m-apart from all other indices within the same group, (b) all the other remaining cases. Noting that $|E(\varphi^{\dagger}_{i_1,t_1t_3}u_{i_1t_4})E(\varphi^{\dagger}_{j_1,t_1t_3}u_{j_1t_5})\ E(\varphi^{\dagger}_{i_2,t_6t_8}u_{i_2t_9})E(\varphi^{\dagger}_{j_2,t_6t_8}u_{j_2t_{10}})|$ is bounded by $Ch^{7-\frac{\delta}{1+\delta}}\alpha^{\frac{\delta}{1+\delta}}(m)$ in case (a) and by Ch^8 in case (b), and the total number of terms in the summation is of order $O\left(n^4T^4m^6\right)$ in case (b), we can readily obtain $ES_{111,1} = O_P(n^2T^2h^{-3-\frac{\delta}{1+\delta}}\alpha^{\frac{\delta}{1+\delta}}(m) + n^2T^{-4}m^6h^{-2}) = o_P(1)$. So $ES_{111,1} = o_P(1)$.

For $ES_{111,3}$, we have

$$
\begin{aligned}
ES_{111,3} &= \frac{1}{(n_1T^4h^5)^2} \sum_{1\le i<j\le n} \sum_{t_1\ldots t_5 \text{ are distinct}} \sum_{t_6\ldots t_{10} \text{ are distinct}} E\left[\varphi^{\dagger}_{i,t_1t_3}u_{it_4}\varphi^{\dagger}_{i,t_6t_8}u_{it_9}\right] \\
&\quad \times E\left[\varphi^{\dagger}_{j,t_1t_3}u_{jt_5}\varphi^{\dagger}_{j,t_6t_8}u_{jt_{10}}\right]\eta_{i,t_1t_4}\eta_{j,t_2t_5}\eta_{i,t_6t_9}\eta_{j,t_7t_{10}}.
\end{aligned}
$$

Let $\mathscr{G}_5 \equiv \{t_1,t_3,t_4,t_6,t_8,t_9\}$, $\mathscr{G}_6 \equiv \{t_1,t_3,t_5,t_6,t_8,t_{10}\}$ and $\mathscr{G} \equiv \mathscr{G}_5 \cup \mathscr{G}_6$. We can consider five cases: the number of distinct time indices in $\mathscr{G}$ are 8, 7, 6, 5, and 4, respectively, and use (a)-(e) to denote these five cases in order. Also, we use $ES_{111,3\xi}$ to denote $ES_{111,3}$ when the time indices in the summation are restricted to these five cases in order for $\xi = a,...,e$. Following the arguments used in the analysis of $S_{111,1}$, we can show that $ES_{111,3a} = O_P(T^2 h^{-4-\frac{2\delta}{1+\delta}} \alpha^{\frac{\delta}{1+\delta}}(m) + T^{-4} m^6 h^{-2}) = o_P(1)$. Similarly we can show that $ES_{111,3\xi} = o_P(1)$ for $\xi = b,c,d$. For $ES_{111,3e}$, noting that the sets $\{t_1,t_3,t_4,t_5\}$ and $\{t_6,t_8,t_9,t_{10}\}$ must coincide, we have $|ES_{111,3e}| = O_P\left(T^{-2}h^{-8}\right) = o_P(1)$. Hence $ES_{111,3} = o_P(1)$, and we have shown that $E^{\mathbb{X}}[(S_{111})^2] = o_P(1)$, implying that $S_{111} = o_P(1)$. Similarly, we can show that $S_{11r} = o_P(1)$ for $r = 2$, 3, 4. It follows that $S_{11} = o_P(1)$.

For S_{12} defined in (.0.42), we decompose it as follows:

$$\begin{aligned}
S_{12} &= \frac{1}{n_1 T^5 h^5} \sum_{1\le i<j\le n} \sum_{t_1\ldots t_6 \text{ are distinct}} [\varphi^{\dagger}_{i,t_1t_3} - \varphi^{\dagger}_{i,t_1t_4}] u_{it_5} [E_{t_3}(k^{\dagger}_{j,t_1t_3}) - E_{t_3}(k^{\dagger}_{j,t_2t_3})] u_{jt_6} \eta_{i,t_1t_5} \eta_{j,t_2t_6} \\
&= \frac{1}{n_1 T^5 h^5} \sum_{1\le i<j\le n} \sum_{t_1\ldots t_6 \text{ are distinct}} \{\varphi^{\dagger}_{i,t_1t_3}[E_{t_3}(k^{\dagger}_{j,t_1t_3}) - c^{\dagger}_j] - \varphi^{\dagger}_{i,t_1t_4}[E_{t_3}(k^{\dagger}_{j,t_1t_3}) - c^{\dagger}_j] \\
&\quad -\varphi^{\dagger}_{i,t_1t_3} u_{it_5} [E_{t_3}(k^{\dagger}_{j,t_2t_3}) - c^{\dagger}_j] + \varphi^{\dagger}_{i,t_1t_4}[E_{t_3}(k^{\dagger}_{j,t_2t_3}) - c^{\dagger}_j]\} u_{it_5} u_{jt_6} \eta_{i,t_1t_5} \eta_{j,t_2t_6} \\
&\equiv S_{121} - S_{122} - S_{123} + S_{124},
\end{aligned}$$

where $c^{\dagger}_j \equiv E_t E_s(k^{\dagger}_{j,ts})$. Analogously to the analysis of S_{111}, we can show $E^{\mathbb{X}}[(S_{12r})^2] = o_P(1)$ for r =1, 2, 3, 4. It follows that $S_{12} = o_P(1)$. By the same token, $S_{113} = o_P(1)$. For S_{114}, we have

$$\begin{aligned}
S_{14} &= \frac{1}{n_1 T^5 h^5} \sum_{1\le i<j\le n} \sum_{t_1\ldots t_6 \text{ are distinct}} \{[E_{t_1}(k^{\dagger}_{i,t_1t_3}) - c^{\dagger}_i][E_{t_3}(k^{\dagger}_{j,t_1t_3}) - c^{\dagger}_j] \\
&\quad -[E_{t_1}(k^{\dagger}_{i,t_1t_3}) - c^{\dagger}_i][E_{t_3}(k^{\dagger}_{j,t_2t_3}) - c^{\dagger}_j] - [E_{t_1}(k^{\dagger}_{i,t_1t_4}) - c^{\dagger}_i][E_{t_3}(k^{\dagger}_{j,t_1t_3}) - c^{\dagger}_j] \\
&\quad +[E_{t_1}(k^{\dagger}_{i,t_1t_4}) - c^{\dagger}_i][E_{t_3}(k^{\dagger}_{j,t_2t_3}) - c^{\dagger}_j]\} u_{it_5} u_{jt_6} \eta_{i,t_1t_5} \eta_{j,t_2t_6} \\
&\equiv S_{141} - S_{142} - S_{143} + S_{144}.
\end{aligned}$$

Then we can show that $E^{\mathbb{X}}[(S_{14r})^2] = o_P(1)$ for r =1, 2, 3, 4. It follows that $S_{14} = o_P(1)$. Hence we have shown that $\mathscr{S}(\mathbb{V},\mathbb{V}) = S_1 + o_P(1) = o_P(1)$.

Now, we consider $\mathscr{S}(\mathbb{B},\mathbb{B})$. We have

$$\begin{aligned}
\mathscr{S}(\mathbb{B},\mathbb{B}) &= \frac{1}{n_1 T^3 h^5} \sum_{1\le i<j\le n} \sum_{t_1\ldots t_4 \text{ are distinct}} \{(\varphi^{\dagger}_{i,t_1t_3} - \varphi^{\dagger}_{i,t_1t_4})(\varphi^{\dagger}_{j,t_1t_3} - \varphi^{\dagger}_{j,t_2t_3}) \\
&\quad +(\varphi^{\dagger}_{i,t_1t_3} - \varphi^{\dagger}_{i,t_1t_4}) E_{t_3}(k^{\dagger}_{j,t_1t_3} - k^{\dagger}_{j,t_2t_3}) + E_{t_1}(k^{\dagger}_{i,t_1t_3} - k^{\dagger}_{i,t_1t_4})(\varphi^{\dagger}_{j,t_1t_3} - \varphi^{\dagger}_{j,t_2t_3}) \\
&\quad +E_{t_1}(k^{\dagger}_{i,t_1t_3} - k^{\dagger}_{i,t_1t_4}) E_{t_3}(k^{\dagger}_{j,t_1t_3} - k^{\dagger}_{j,t_2t_3})\} \mathbb{B}_{it_1} \mathbb{B}_{jt_2} \\
&\equiv S_{21} + S_{22} + S_{23} + S_{24}, \text{ say.}
\end{aligned}$$

Write $S_{21} = \frac{1}{n_1 T^3 h^5} \sum_{1\le i<j\le n} \sum_{t_1\ldots t_4 \text{ are distinct}} \{\varphi^{\dagger}_{i,t_1t_3}\varphi^{\dagger}_{j,t_1t_3} - \varphi^{\dagger}_{i,t_1t_3}\varphi^{\dagger}_{j,t_2t_3} - \varphi^{\dagger}_{i,t_1t_4}\varphi^{\dagger}_{j,t_1t_3} + \varphi^{\dagger}_{i,t_1t_4} \times\varphi^{\dagger}_{j,t_2t_3}\} \mathbb{B}_{it_1}\mathbb{B}_{jt_2} \equiv S_{211} - S_{212} - S_{213} + S_{214}$. It is easy to show that S_{211} dominates S_{21r} for $r = 2,3,4$ and

$$E^{\mathbb{X}}\left[(S_{211})^2\right] = O_P(n^2 T^2 h^{-3-\frac{\delta}{1+\delta}} \alpha^{\frac{\delta}{1+\delta}}(m) + n^2 m^2 h^{-2} + T^2 h^{-5-\frac{2\delta}{1+\delta}} \alpha^{\frac{\delta}{1+\delta}}(m)) \left(b^{4(p+1)}\right) = o_P(1).$$

Hence $S_{211} = o_P(1)$ and $S_{21} = o_P(1)$. Similarly, by decomposing $E_{t_1}(k^{\dagger}_{i,t_1t_3} - k^{\dagger}_{i,t_1t_4})$ as $[E_{t_1}(k^{\dagger}_{i,t_1t_3}) - c^{\dagger}_i] - [E_{t_1}(k^{\dagger}_{i,t_1t_4}) - c^{\dagger}_i]$ and $E_{t_3}(k^{\dagger}_{j,t_1t_3} - k^{\dagger}_{j,t_2t_3})$ as $[E_{t_3}(k^{\dagger}_{j,t_1t_3}) - c^{\dagger}_j] - [E_{t_3}(k^{\dagger}_{j,t_2t_3}) - c^{\dagger}_j]$, we can show $S_{2r} = o_P(1)$ for $r = 2,3,4$ by the conditional Chebyshev inequality. Consequently, $\mathscr{S}(\mathbb{B},\mathbb{B}) = o_P(1)$. Analogously, we can show that $\mathscr{S}(\mathbb{V},\mathbb{B}) = o_P(1)$ and $\mathscr{S}(\mathbb{B},\mathbb{V}) = o_P(1)$. It follows that $S_{5nT,1} = o_P(1)$. ■

Proposition .0.5 $nTh(\widetilde{\Delta}_{nT,1} - \widehat{\Delta}_{nT,1}) = o_P(1)$.

Proof. By the definitions of $\widehat{\Delta}_{nT,1}$ and $\widetilde{\Delta}_{nT,1}$, we have $-nTh(\widetilde{\Delta}_{nT,1} - \widehat{\Delta}_{nT,1}) / \left[2R\left(\overline{k}\right)\right]$ $= \sum_{i=1}^{n} \int [\widetilde{f}_i^2(u) - \widehat{f}_i^2(u)]du \equiv U_{1nT} + 2U_{2nT}$, where $U_{1nT} \equiv \sum_{i=1}^{n} \int [\widetilde{f}_i(u) - \widehat{f}_i(u)]^2 du$, and $U_{2nT} \equiv \sum_{i=1}^{n} \int [\widetilde{f}_i(u) - \widehat{f}_i(u)]\widehat{f}_i(u)\,du$. Then it is straightforward to show that $U_{1nT} = o_P(1)$ and $U_{2nT} = o_P(1)$ by arguments similar to but simpler than those used in the proof of Proposition .0.4. ■

Proposition .0.6 $nTh(\widetilde{\Delta}_{nT,2} - \widehat{\Delta}_{nT,2}) = o_P(1)$.

Proof. Let $\widehat{\Delta}_{nT,21}$, $\widehat{\Delta}_{nT,22}$, and $\widehat{\Delta}_{nT,23}$ denote the three terms on the right hand side of (.0.28). Define $\widetilde{\Delta}_{nT,21}$, $\widetilde{\Delta}_{nT,22}$, and $\widetilde{\Delta}_{nT,23}$ analogously with the estimated residuals replacing the unobservable error terms. Then it suffices to show that $nTh(\widetilde{\Delta}_{nT,2r} - \widehat{\Delta}_{nT,2r}) = o_P(1)$ for $r = 1,2,3$. Each of them can be proved by the use of Taylor expansions and Chebyshev inequality. We omitted the details to save space. ■

.0.4 Some technical lemmas

This appendix presents some technical lemmas that are used in proving the main results.

Lemma .0.5 *Let $\{W_t\}$ be a strong (α-) mixing process with mixing coefficient $\alpha(t)$. For any integer $l > 1$ and integers $(t_1, ..., t_l)$ such that $1 \le t_1 < t_2 < \cdots < t_l$, let θ be a Borel measurable function such that*

$$\int |\theta(w_1, ..., w_l)|^{1+\delta} dF^{(1)}(w_1, ..., w_j)\, dF^{(2)}(w_{j+1}, ..., w_l) \le M$$

for some $\delta > 0$ and $M > 0$, where $F^{(1)} = F_{t_1,...,t_j}$ and $F^{(2)} = F_{t_{j+1},...,t_l}$ are the distribution functions of $(W_{t_1}, ..., W_{t_j})$ and $(W_{t_{j+1}}, ..., W_{t_l})$, respectively. Let F denote the distribution function of $(W_{t_1}, ..., W_{t_l})$. Then

$$\left| \int \theta(w_1, ..., w_l)\, dF(w_1, ..., w_l) - \int \theta(w_1, ..., w_l)\, dF^{(1)}(w_1, ..., w_j)\, dF^{(2)}(w_{j+1}, ..., w_l) \right|$$
$$\le 4M^{1/(1+\delta)} \alpha(t_{j+1} - t_j)^{\delta/(1+\delta)}.$$

Proof. See Lemma 2.1 of Sun and Chiang (1997). ■

Lemma .0.6 *Let* $\{W_t\}$, θ, δ, *and* M *be defined as above. Let* $V_1 \equiv (W_{t_1}, ..., W_{t_j})$ *and* $V_2 \equiv (W_{t_{j+1}}, ..., W_{t_l})$. *Then* $E|E[\theta(V_1,V_2)|V_1] - \Theta(V_1)| \leq 4M^{1/(1+\delta)}\alpha\,(t_{j+1} - t_j)^{\delta/(1+\delta)}$, *where* $\Theta(v_1) \equiv E[\theta(v_1, V_2)]$.

Proof. See Yoshihara (1989) who proved the above lemma for β-mixing processes by using an inequality in Yoshihara (1976). The analogous result holds for α-mixing processes by using the Davydov inequality or Lemma .0.5. ■

Let $k : \mathbb{R} \to \mathbb{R}$ be a differentiable kernel function, and k' be its first derivative. Define $\overline{k}(v) \equiv \int k(u)\,k(v-u)\,du$, $\overline{k'}(v) \equiv \int k'(u)\,k'(v-u)\,du$, and $k^+(v) \equiv \int k'(u)\,k(v-u)\,du$. The following lemma states some properties of $\overline{k}$, $\overline{k'}$, and k^+ that are used in the proof of our main results.

Lemma .0.7 *Suppose* $k : \mathbb{R} \to \mathbb{R}$ *is a symmetric differential* γ*-th order kernel function such that* $\lim_{v\to\infty} v^l k(v) = 0$ *for* $l = 0, 1$. *Then*

(i) $\int \overline{k}(v)\,dv = 1$, $\int \overline{k}(v)\,v^l dv = 0$ *for* $l = 1, \ldots \gamma - 1$, *and* $\int \overline{k}(v)\,v^\gamma dv = 2\kappa_\gamma$ *where* $\kappa_\gamma = \int k(u)\,u^\gamma du$;

(ii) $\int \overline{k'}(v)\,v^l dv = 0$ *for* $l = 0, 1$ *and* $\int \overline{k'}(v)\,v^2 dv = 2$;

(iii) $\int k^+(v)\,dv = 0$, *and* $\int v k^+(v)\,dv = -1$.

Proof. (i) $\int \overline{k}(v)\,dv = \int\int k(u)\,k(v-u)\,dudv = \int k(u)\,du \int k(s)\,ds = 1$,

$\int \overline{k}(v)\,v^l dv = \sum_{s=0}^{l} C_l^s \int k(u)\,u^s du \int k(t)\,t^{l-s} dt = 0$ for $l = 1, \ldots, \gamma - 1$, and

$\int \overline{k}(v)\,v^\gamma dv = \sum_{s=0}^{\gamma} C_\gamma^s \int k(u)\,u^s du \int k(t)\,t^{\gamma-s} dt = 2\int k(u)\,du \int k(t)\,t^\gamma dt = 2\kappa_\gamma$.

(ii) $\int \overline{k'}(v)\,dv = \int\int k'(u)\,k'(v-u)\,dudv = \int k'(u)\,du \int k'(s)\,ds = 0$,

$\int \overline{k'}(v)\,vdv = 2\int k'(u)\,udu \int k'(t)\,dt = 0$ by the fact $\int k'(u)\,du = 0$, and

$\int \overline{k'}(v)\,v^2 dv = \int\int k'(u)\,k'(t)\,(u^2 + 2ut + t^2)\,dudt = 2\left[\int k'(u)\,udu\right]^2 = 2$.

(iii) $\int k^+(v)\,dv = \int k'(u) \int k(u-v)\,dvdu = \int k'(u)\,du = 0$, and

$\int vk^+(v)\,dv = \int k(u)\,k'(s)\,(s+u)\,dsdu = \int k'(s)\,sds + \int k'(s)\,ds \int uk(u)\,du = -1$. ■

B Proofs in Chapter 3

Throughout the appendix, let C signify a generic constant whose exact value may vary from case to case. Let $E_{\mathscr{D}}(\cdot) \equiv E(\cdot|\mathscr{D})$ and $\text{Var}_{\mathscr{D}}(\cdot) \equiv \text{Var}(\cdot|\mathscr{D})$. Let $E_{(\mathscr{D},S)}(\cdot)$ denote expectation with respect to variables indexed by set S conditional on $\mathscr{D}$. 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 $\varepsilon_k \equiv \beta_k^0 - \beta_k$ for $k = 1,...,K$, $\varepsilon_0 \equiv \|\mathbf{u}\|/\sqrt{NT}$ and $\mathbf{P}_0 \equiv (\sqrt{NT}/\|\mathbf{u}\|)\mathbf{u}$. Let $\vartheta_{NT} \equiv \sum_{k=0}^K \varepsilon_k \mathbf{P}_k$, $d_{\max}\left(\lambda^0, f^0\right) \equiv \sqrt{\mu_1\left(\frac{1}{NT}\lambda^{0\prime}f^0 f^{0\prime}\lambda^0\right)}$, 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

$$r_0\left(\lambda^0, f^0\right) \equiv \left(\frac{4d_{\max}\left(\lambda^0, f^0\right)}{d_{\min}^2\left(\lambda^0, f^0\right)} + \frac{1}{2d_{\max}\left(\lambda^0, f^0\right)}\right)^{-1} \text{ and } \alpha_{NT} \equiv \frac{\|\vartheta_{NT}\|}{\sqrt{NT}} \frac{16d_{\max}\left(\lambda^0, f^0\right)}{d_{\min}^2\left(\lambda^0, f^0\right)}.$$

Below we prove the main results in Sections 3 and 4. The proofs of all technical lemmas and Theorem 3.4.4 are given in the online Supplemental Material which is available on the first author's website.

.0.5 Proofs of the main results in Section 3

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

Lemma .0.8 *Suppose that Assumptions 1-4 hold. Then* $||\hat{\beta} - \beta^0|| = O_P(K^{-\gamma/(2d)} + \delta_{NT}^{-1/2})$.

Proof of Theorem 3.3.1. 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 .0.8, 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 $\frac{\|\vartheta_{NT}\|}{\sqrt{NT}} \leq$ $r_0\left(\lambda^0, f^0\right)$ w.p.a.1 and we can apply Proposition .0.10 in the supplementary appendix to expand $L_{NT}(\beta)$ as follows

$$\begin{aligned} L_{NT}(\beta) &= \frac{1}{NT}\sum_{k_1=0}^K\sum_{k_2=0}^K \varepsilon_{k_1}\varepsilon_{k_2} L^{(2)}\left(\lambda^0, f^0, \mathbf{P}_{k_1}, \mathbf{P}_{k_2}\right) \\ &\quad + \frac{1}{NT}\sum_{k_1=0}^K\sum_{k_2=0}^K\sum_{k_3=0}^K \varepsilon_{k_1}\varepsilon_{k_2}\varepsilon_{k_3} L^{(3)}\left(\lambda^0, f^0, \mathbf{P}_{k_1}, \mathbf{P}_{k_2}, \mathbf{P}_{k_3}\right) + O_P\left(\alpha_{NT}^4\right) \\ &= L_{NT}\left(\beta^0\right) + L_{1,NT}(\beta) + L_{2,NT}(\beta) + L_{R,NT}(\beta) + O_P\left(\alpha_{NT}^4\right) - O_P\left(\varepsilon_0^4\right), \end{aligned}$$

where $L^{(2)}$ and $L^{(3)}$ are defined in Proposition .0.10,

$$
\begin{aligned}
L_{NT}\left(\beta^{0}\right) &= \frac{1}{NT}\varepsilon_{0}^{2}L^{(2)}\left(\lambda^{0},f^{0},\mathbf{P}_{0},\mathbf{P}_{0}\right)+\frac{1}{NT}\varepsilon_{0}^{3}L^{(3)}\left(\lambda^{0},f^{0},\mathbf{P}_{0},\mathbf{P}_{0},\mathbf{P}_{0}\right)+O_{P}\left(\varepsilon_{0}^{4}\right),\\
L_{1,NT}\left(\beta\right) &= \frac{2}{NT}\sum_{k=1}^{K}\varepsilon_{k}\varepsilon_{0}L^{(2)}\left(\lambda^{0},f^{0},\mathbf{P}_{k},\mathbf{P}_{0}\right)+\frac{3}{NT}\sum_{k=1}^{K}\varepsilon_{k}\varepsilon_{0}^{2}L^{(3)}\left(\lambda^{0},f^{0},\mathbf{P}_{k},\mathbf{P}_{0},\mathbf{P}_{0}\right),\\
L_{2,NT}\left(\beta\right) &= \frac{1}{NT}\sum_{k_1=1}^{K}\sum_{k_2=1}^{K}\varepsilon_{k_1}\varepsilon_{k_2}L^{(2)}\left(\lambda^{0},f^{0},\mathbf{P}_{k_1},\mathbf{P}_{k_2}\right),
\end{aligned}
$$

and

$$
\begin{aligned}
L_{R,NT}\left(\beta\right) &= \frac{1}{NT}\sum_{k_2=1}^{K}\sum_{k_2=1}^{K}\sum_{k_3=1}^{K}\varepsilon_{k_1}\varepsilon_{k_2}\varepsilon_{k_3}L^{(3)}\left(\lambda^{0},f^{0},\mathbf{P}_{k_1},\mathbf{P}_{k_2},\mathbf{P}_{k_3}\right)\\
&\quad+\frac{3}{NT}\sum_{k_2=1}^{K}\sum_{k_2=1}^{K}\varepsilon_{k_1}\varepsilon_{k_2}\varepsilon_{0}L^{(3)}\left(\lambda^{0},f^{0},\mathbf{P}_{k_1},\mathbf{P}_{k_2},\mathbf{P}_{0}\right)\\
&\quad+O_{P}[\left(\left\|\beta-\beta^{0}\right\|+\varepsilon_{0}\right)^{4}-\varepsilon_{0}^{4}]\\
&= O_{P}\left(\left\|\beta-\beta^{0}\right\|^{2}\varepsilon_{0}+\left\|\beta-\beta^{0}\right\|^{3}+\left\|\beta-\beta^{0}\right\|\varepsilon_{0}^{3}\right) \qquad (.0.43)
\end{aligned}
$$

Clearly, $L_{1,NT}\left(\beta\right)$ and $L_{2,NT}\left(\beta\right)$ 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

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

where $C_{NT}^{(1)}$ and $C_{NT}^{(2)}$ are defined in Theorem 3.3.1. Then

$$
\begin{aligned}
L_{NT}\left(\beta\right) &= L_{NT}\left(\beta^{0}\right)-2((\beta-\beta^{0})'(C_{NT}^{(1)}+C_{NT}^{(2)})+(\beta-\beta^{0})'W_{NT}(\beta-\beta^{0})\\
&\quad+O_{P}\left\{\left\|\beta-\beta^{0}\right\|^{2}\varepsilon_{0}+\left\|\beta-\beta^{0}\right\|^{3}+\left\|\beta-\beta^{0}\right\|\varepsilon_{0}^{3}\right\}. \qquad (.0.44)
\end{aligned}
$$

Noting that $\operatorname{rank}(\mathbf{P}_k\Phi'\mathbf{u}M_{f^0}\mathbf{u}'M_{\lambda^0}+\mathbf{P}_kM_{f^0}\mathbf{u}'M_{\lambda^0}\mathbf{u}\Phi'+\mathbf{P}_kM_{f^0}\mathbf{u}'\Phi\mathbf{u}'M_{\lambda^0})\leq 3R$ and using the trace inequality $\operatorname{tr}(A)\leq\operatorname{rank}(A)\left\|A\right\|$ for any real square matrix A, we have $C_{NT,k}^{(2)}=\frac{1}{NT}\operatorname{tr}(\mathbf{P}_k\Phi'\mathbf{u}M_{f^0}\mathbf{u}'M_{\lambda^0}+\mathbf{P}_kM_{f^0}\mathbf{u}'M_{\lambda^0}\mathbf{u}\Phi'+\mathbf{P}_kM_{f^0}\mathbf{u}'\Phi\mathbf{u}'M_{\lambda^0})\leq$

$\frac{3R}{NT}\|\mathbf{P}_k\|\|\Phi\|\|\mathbf{u}\|^2\|M_{\lambda^0}\|\left\|M_{f^0}\right\| = \frac{\|\mathbf{P}_k\|}{\sqrt{NT}}O_P\left(K^{-2\gamma/d}+\delta_{NT}^{-2}\right)$. It follows that

$$\begin{aligned}\left\|C_{NT}^{(2)}\right\| &= \left\{\sum_{k=1}^{K}\left[C_{NT,k}^{(2)}\right]^2\right\}^{1/2} = \left\{\sum_{k=1}^{K}\frac{\|\mathbf{P}_k\|^2}{NT}\right\}^{1/2} O_P(K^{-2\gamma/d}+\delta_{NT}^{-2}) \\ &= O_P\left[\sqrt{K}\left(K^{-2\gamma/d}+\delta_{NT}^{-2}\right)\right]. \end{aligned} \tag{.0.45}$$

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 $\overrightarrow{W} \equiv (NT)^{-1}\overrightarrow{Z}'W_{NT}^{-1}\overrightarrow{Z}$. Note that $\overrightarrow{W}$ is a projection matrix with $\mu_1(\overrightarrow{W})=1$. By Assumptions $2(ii)$ and $3(i)$-(iii), we have

$$\begin{aligned}\left\|W_{NT}^{-1}\frac{1}{NT}\sum_{i=1}^{N}Z_i'e_{g,i}\right\|^2 &= \frac{1}{N^2T^2}\mathrm{tr}\left(\overrightarrow{e}_g\overrightarrow{Z}'W_{NT}^{-1}W_{NT}^{-1}\overrightarrow{Z}\overrightarrow{e}_g\right) \\ &\leq [\mu_{\min}(W_{NT})]^{-1}\frac{1}{N^2T^2}\mathrm{tr}\left(\overrightarrow{e}_g\overrightarrow{W}\overrightarrow{e}_g'\right) \\ &\leq \left\{[\mu_{\min}(W)]^{-1}+o_P(1)\right\}\|\overrightarrow{e}_g\|_F^2/(NT) = O_P\left(K^{-2\gamma/d}\right).\end{aligned}$$

Then we have

$$\left\|W_{NT}^{-1}C_{NT}^{(1)}\right\| = O_P\left(\delta_{NT}^{-1}\sqrt{K/T}+K^{-\gamma/d}\right). \tag{.0.46}$$

Let

$$r_{NT} \equiv W_{NT}^{-1}C_{NT}^{(1)} + W_{NT}^{-1}C_{NT}^{(2)} \text{ and } R_{NT} \equiv \hat{\beta}-\beta^0-r_{NT}. \tag{.0.47}$$

From (.0.45) and (.0.46) we have

$$\|r_{NT}\| \leq \left\|W_{NT}^{-1}C_{NT}^{(1)}\right\| + \left\|W_{NT}^{-1}C_{NT}^{(2)}\right\| = O_P\left(\sqrt{K}\delta_{NT}^{-2}+K^{-\gamma/d}\right). \tag{.0.48}$$

Since $L_{NT}(\hat{\beta}) \leq L_{NT}\left(\beta^0 + r_{NT}\right)$, we can apply (.0.44) to the objects on both sides of the last inequality to obtain

$$
\begin{aligned}
\|R_{NT}\|^2 &= \left(\hat{\beta} - \beta^0 - r_{NT}\right)' W_{NT}^{1/2} W_{NT}^{-1} W_{NT}^{1/2} \left(\hat{\beta} - \beta^0 - r_{NT}\right) \\
&\leq [\mu_{\min}(W_{NT})]^{-1} \left(\hat{\beta} - \beta^0 - r_{NT}\right)' W_{NT} \left(\hat{\beta} - \beta^0 - r_{NT}\right) \\
&\leq [\mu_{\min}(W_{NT})]^{-1} \left[L_{R,NT}\left(\beta^0 + r_{NT}\right) - L_{R,NT}(\beta^0 + (\hat{\beta} - \beta^0))\right] \\
&\leq O_P(\|r_{NT}\|^2 \varepsilon_0 + \|r_{NT}\| \varepsilon_0^3 + \|r_{NT}\|^3) \\
&\quad - O_P(||\hat{\beta} - \beta^0||^2 \varepsilon_0 + ||\hat{\beta} - \beta^0|| \varepsilon_0^3 + ||\hat{\beta} - \beta^0||^3) \qquad (.0.49)
\end{aligned}
$$

We now argue that $||\hat{\beta} - \beta^0||$ has the same probability order as $||r_{NT}||$ by contradiction. Suppose $||\hat{\beta} - \beta^0|| = o_P(||r_{NT}||)$. Then by (.0.47) and (.0.49), and the fact that $\varepsilon_0^3/||r_{NT}|| = o_P(1)$, we have $||r_{NT}||^2 = O_P(\|R_{NT}\|^2) \leq o_P\left(||r_{NT}||^2\right)$, a contradiction. Similarly, suppose $||r_{NT}|| = o_P(||\hat{\beta} - \beta^0||)$. Then $||\hat{\beta} - \beta^0||^2 = O_P(\|R_{NT}\|^2) \leq O_P(||\hat{\beta} - \beta^0|| \varepsilon_0^3)$, implying that $||\hat{\beta} - \beta^0|| \leq O_P\left(\varepsilon_0^3\right) = o_P(||r_{NT}||) = o_P(||\hat{\beta} - \beta^0||)$, a contradiction. It follows that

$$
\left\|\hat{\beta} - \beta^0\right\| = O_P(||r_{NT}||) = O_P\left(\sqrt{K}\delta_{NT}^{-2} + K^{-\gamma/d}\right) \qquad (.0.50)
$$

and

$$
R_{NT} = O_P(\|r_{NT}\| \varepsilon_0^{1/2}) = O_P[(\sqrt{K}\delta_{NT}^{-2} + K^{-\gamma/d})(\delta_{NT}^{-1/2} + K^{-2\gamma/d})] \qquad (.0.51)
$$

because $\varepsilon_0^2/||r_{NT}|| = O_P(1)$ and $||r_{NT}||/\varepsilon_0 = O_P(1)$ by Assumption 4.

Now we derive the convergence rate of $\hat{g}(x)$.

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

where $C_{w1} \equiv \int_{\mathscr{X}} (1+\|x\|^2)^{\bar{\omega}} w(x)\, dx < \infty$, $Q_{pp,w} \equiv \int_{\mathscr{X}} p^K(x) p^K(x)' w(x)\, dx$, and $\mu_1(Q_{pp,w}) < \infty$. Similarly,

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

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

Proof of Theorem 3.3.2. 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

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

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 .0.7 and .0.8 below, respectively. For (iii), by Cauchy-Schwarz inequality, (.0.51) and Assumptions 7 and 8, we have

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

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 fixed $x \in \mathscr{X}$

$$
\begin{aligned}
\Pi_{4NT} &= (NT)^{1/2} V_K^{-1/2}(x) \left[g(x) - p^K(x)' \beta^0 \right] \\
&\leq C \left\| p^K(x) \right\|^{-1} \sqrt{NT} \left\| g(x) - p^K(x)' \beta^0 \right\|_{\infty,\bar{\omega}} \left(1 + \|x\|^2 \right)^{\bar{\omega}/2} \\
&= O_P\left(\sqrt{NT} K^{-\gamma/d} \right) = o_P(1)
\end{aligned}
$$

as $\inf_{x \in \mathscr{X}} ||p^K(x)|| \geq C > 0$. ■

Next, we state some lemmas are used in the proofs of Propositions .0.7-.0.8 below.

Lemma .0.9 *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 3.3.2 hold. Then*
(i) $\frac{1}{NT}\sum_{t=1}^T \sum_{i=1}^N \left\| d_{it}^4 \right\|_{2,\mathscr{D}}^2 = O_P\left(K^4\right)$;
(ii) $\frac{1}{N^2 T}\sum_{t=1}^T (\sum_{i=1}^N \left\| d_{it}^2 \right\|_{2,\mathscr{D}}^2)^2 = O_P\left(K^4\right)$.

Lemma .0.10 *Suppose that the assumptions in Theorem 3.3.2 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 .0.11 *Suppose that the assumptions in Theorem 3.3.2 hold. Then*
$\frac{1}{\sqrt{NT}}\sum_{i=1}^N \sum_{t=1}^T v_x^{K\prime}\{(Z_{it} - \tilde{Z}_{it})e_{it} - E_{\mathscr{D}}[(Z_{it} - \tilde{Z}_{it})e_{it}]\} = o_P(1)$.

Lemma .0.12 *Suppose that the assumptions in Theorem 3.3.2 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) $\frac{1}{N\sqrt{T}}\sum_{t=1}^T \sum_{i=1}^N \sum_{j=1}^N A_i \lambda_j^0 \left[e_{jt}e_{it} - E_{\mathscr{D}}\left(e_{jt}e_{it}\right) \right] = O_P(\sqrt{K})$;
$(viii)$ $N^{-1}\sum_{i=1}^N ||T^{-1/2}\sum_{s=1}^T v_x^{K\prime}\left(p_{is}^c - p_{is}^{\lambda c}\right) f_s^0 G^0||^2 = O_P(K)$;
(ix) $N^{-1}\sum_{i=1}^N ||(NT)^{-1/2}\sum_{t=1}^T \sum_{j=1}^N \lambda_j^0 \left[e_{it}e_{jt} - E_{\mathscr{D}}\left(e_{it}e_{jt}\right) \right]||^2 = O_P(1)$;
(x) $\frac{1}{\sqrt{NT}}\sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^T B_t f_s^{0\prime}\left[e_{it}e_{is} - E_{\mathscr{D}}\left(e_{it}e_{is}\right) \right] = O_P(\sqrt{K})$;
(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) $\frac{1}{NT^2}\sum_{t=1}^T ||\sum_{i=1}^N \sum_{s=1}^T f_s^{0\prime}\left[e_{it}e_{is} - E_{\mathscr{D}}\left(e_{it}e_{is}\right) \right]||^2 = O_P(1)$;

where $A_i \equiv v_x^{K\prime}[E_{\mathscr{D}}\left(P_i - P_i^{\lambda}\right)^{\prime}]f^0G^0/T$, $P_i^{\lambda} = N^{-1}\sum_{j=1}^{N}\alpha_{ij}E_{\mathscr{D}}\left(P_j\right)$, $G^0 \equiv (\frac{f^{0\prime}f^0}{T})^{-1} \times (\frac{\lambda^{0\prime}\lambda^0}{N})^{-1}$, $p_{is}^c = p_{is} - E_{\mathscr{D}}\left(p_{is}\right)$, $p_{is}^{\lambda c} \equiv p_{is}^{\lambda} - E_{\mathscr{D}}\left(p_{is}^{\lambda}\right)$, $p_{is}^{\lambda} \equiv N^{-1}\sum_{j=1}^{N}\alpha_{ij}p_{js}$, $B_t \equiv v_x^{K\prime}E_{\mathscr{D}}(P_{\cdot t} - P_{\cdot t}^f)^{\prime}\lambda^0G^0N^{-1}$, $P_{\cdot t}^f \equiv T^{-1}\sum_{l=1}^{T}\eta_{lt}P_{\cdot l}$, *and* $P_{\cdot t} \equiv \left(p_{1t}^{\prime}, \cdots, p_{Nt}^{\prime}\right)^{\prime}$.

Proposition .0.7 *Suppose that the assumptions in Theorem 3.3.2 hold. Then* $\Pi_{1NT} + \kappa_{NT}b_1\left(x\right) \xrightarrow{d} N(0,1)$.

Proof. Recall $v_x^K \equiv V_K^{-1/2}\left(x\right)\tilde{W}^{-1}p^K\left(x\right)$. One can readily show that $\left\|v_x^K\right\| = O_P\left(1\right)$. Note that

$$\begin{aligned}
\Pi_{1NT} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}v_x^{K\prime}Z_{it}u_{it} - V_K^{-1/2}\left(x\right)p^K\left(x\right)^{\prime}\left(\tilde{W}^{-1} - W_{NT}^{-1}\right)\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}Z_{it}u_{it} \\
&\equiv \Pi_{1NT,1} + \Pi_{1NT,2}, \text{ say.}
\end{aligned}$$

We complete the proof by showing that (i) $\Pi_{1NT,1} + \kappa_{NT}b_1\left(x\right) \xrightarrow{d} N(0,1)$ and (ii) $\Pi_{1NT,2} = o_P\left(1\right)$.

First, we consider (ii). By (.0.46), Lemmas .0.10(i)-(ii), and Assumption 8, we have

$$\begin{aligned}
\left|\Pi_{1NT,2}\right| &\leq \left\|V_K^{-1/2}\left(x\right)p^K\left(x\right)^{\prime}\tilde{W}^{-1}\right\|\left\{\sqrt{NT}\left\|W_{NT} - \tilde{W}\right\|_F\right\}\left\|W_{NT}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}Z_{it}u_{it}\right\| \\
&= O_P\left(K\right)O_P\left(\sqrt{\frac{K}{T}}\delta_{NT}^{-1} + K^{-\gamma/d}\right) = O_P\left(\sqrt{\frac{K^3}{T\delta_{NT}^2}} + K^{1-\gamma/d}\right) = o_P\left(1\right).
\end{aligned}$$

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

$$\begin{aligned}
\Pi_{1NT,12} &\leq \sqrt{NT}\left\{v_x^{K\prime}\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}Z_{it}Z_{it}^{\prime}\right)v_x^K\right\}^{1/2}\left\{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}e_{g,it}^2\right\}^{1/2} \\
&\leq \left\|v_x^K\right\|\mu_1^{1/2}\left(W_{NT}\right)\left\|g\left(x\right) - p^K\left(x\right)^{\prime}\beta^0\right\|_{\infty,\bar{\omega}}\left\{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(1 + \left\|X_{it}\right\|^2)^{\bar{\omega}}\right\}^{1/2} \\
&= O_P(\sqrt{NT}K^{-\gamma/d}) = o_P\left(1\right).
\end{aligned}$$

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

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

where $\tilde{Z}_{it}=p_{it}-\frac{1}{N}\sum_{j=1}^{N}\alpha_{ij}E_{\mathscr{D}}[p_{jt}]-\frac{1}{T}\sum_{s=1}^{T}\eta_{ts}E_{\mathscr{D}}[p_{is}]+\frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\alpha_{ij}\eta_{ts}E_{\mathscr{D}}[p_{js}]$. We complete the proof by showing that: (ia) $\Pi_{1NT,11a}\xrightarrow{d}N(0,1)$, (ib) $\Pi_{1NT,11b}=o_{P}(1)$, and (ic) $\Pi_{1NT,11c}=o_{P}(1)$. (ic) follows from Lemma .0.11. 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 $\xi_{NT,t}\equiv\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}v_{x}^{K\prime}\tilde{Z}_{it}e_{it}$. Recall that $\mathscr{F}_{0}^{t-1}=\sigma(\lambda^{0},f^{0},\{X_{it},X_{i,t-1},e_{i,t-1},\cdots\}_{i=1}^{N})$. By Assumption 5$(ii)$, $E\left[\xi_{NT,t}|\mathscr{F}_{0}^{t-1}\right]=\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}v_{x}^{K\prime}\tilde{Z}_{it}E\left[e_{it}|\mathscr{F}_{0}^{t-1}\right]=0$. That is, $\left\{\xi_{NT,t},\mathscr{F}_{0}^{t}\right\}_{t=1}^{T}$ is a martingale difference sequence (m.d.s.). Consequently, we can apply the martingale CLT (e.g., Pollard, 1984, p.171) to prove that $\Pi_{1NT,11a}\xrightarrow{d}N(0,1)$ by verifying that

$$
(ia1)\ \vec{\xi}_{NT}\equiv\sum_{t=1}^{T}E\left[\xi_{NT,t}^{4}|\mathscr{F}_{0}^{t-1}\right]=o_{P}(1)\ \text{ and }(ia2)\ \sum_{t=1}^{T}\xi_{NT,t}^{2}-1=o_{P}(1).
$$

Since $\vec{\xi}_{NT}\geq 0$, we will prove $(ia1)$ by showing that $E_{\mathscr{D}}\{\sum_{t=1}^{T}E\left[\xi_{NT,t}^{4}|\mathscr{F}_{0}^{t-1}\right]\}=o_{P}(1)$. 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 $\mathscr{D}$ and $\left\{\xi_{NT,t},\mathscr{F}_{0}^{t}\right\}_{t=1}^{T}$ is an m.d.s., we have

$$
\begin{aligned}
E_{\mathscr{D}}\left[\vec{\xi}_{NT}\right] &= \frac{1}{N^{2}T^{2}}\sum_{t=1}^{T}\sum_{i_{1}=1}^{N}\sum_{i_{2}=1}^{N}\sum_{i_{3}=1}^{N}\sum_{i_{4}=1}^{N}E_{\mathscr{D}}\left[d_{i_{1}t}d_{i_{2}t}d_{i_{3}t}d_{i_{4}t}e_{i_{1}t}e_{i_{2}t}e_{i_{3}t}e_{i_{4}t}\right] \\
&= \frac{1}{N^{2}T^{2}}\sum_{t=1}^{T}\sum_{i=1}^{N}E_{\mathscr{D}}\left[d_{it}^{4}e_{it}^{4}\right]+\frac{3}{N^{2}T^{2}}\sum_{t=1}^{T}\left\{\sum_{i=1}^{N}E_{\mathscr{D}}\left[d_{it}^{2}e_{it}^{2}\right]\right\}^{2} \\
&\quad-\frac{3}{N^{2}T^{2}}\sum_{t=1}^{T}\sum_{i=1}^{N}\left\{E_{\mathscr{D}}\left[d_{it}^{2}e_{it}^{2}\right]\right\}^{2} \\
&\equiv \vec{\xi}_{NT}(1)+3\vec{\xi}_{NT}(2)-3\vec{\xi}_{NT}(3),\ \text{say.}
\end{aligned}
$$

By Hölder inequality, Lemma .0.9(i)-(ii), and Assumption 6(i), we have

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

and

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

Similarly, we can show that $\vec{\xi}_{NT}(3) = O_P\left(K^2/(NT)\right) = o_P(1)$ by Lemma .0.9(i) and Assumption 6(i). Then $(ia1)$ follows by conditional Markov inequality. Now, note that $\sum_{t=1}^{T}E_{\mathscr{D}}\left[\xi_{NT,t}^2\right] = v_x^{K\prime}\{(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}E_{\mathscr{D}}[\tilde{Z}_{it}\tilde{Z}_{it}'e_{it}^2]\}v_x^K = 1$. By some straightforward moment calculations, we can show that $E_{\mathscr{D}}[(\sum_{t=1}^{T}\xi_{NT,t}^2-1)^2] = o_P(1)$. Thus $(ia2)$ follows.

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

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

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 .0.11. So $\Pi_{1NT,11b} = o_P(1)$. ■

Proposition .0.8 *Suppose that the assumptions in Theorem 3.3.2 hold. Then we have* $\Pi_{2NT} = -\kappa_{NT}^{-1} b_2(x) - \kappa_{NT} b_3(x) + o_P(1)$.

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 $\tilde{\Pi}_{2NT} \equiv \sqrt{NT} v_x^K C_{NT}^{(2)}$. Then we have $\Pi_{2NT} = \sqrt{NT} v_x^K C_{NT}^{(2)} + \sqrt{NT}\left[v_x^K - \vec{v}_x^K\right]' C_{NT}^{(2)} = \tilde{\Pi}_{2NT} + o_P(1)$ where the $o_P(1)$ term comes from the fact that

$$\begin{aligned}
\sqrt{NT}\left|\left[v_x^K - \vec{v}_x^K\right]' C_{NT}^{(2)}\right| &= \left|\sqrt{NT} V_K^{-1/2}(x)\, p^K(x)' \tilde{W}^{-1}\left(W_{NT} - \tilde{W}\right) W_{NT}^{-1} C_{NT}^{(2)}\right| \\
&\le \sqrt{NT}\left\|W_{NT}^{-1}\right\| \|C_{NT}^{(2)}\| V_K^{-1/2}(x)\left\|p^K(x)\tilde{W}^{-1}\right\| \left\|W_{NT} - \tilde{W}\right\|_F \\
&= \sqrt{NT}\, O_P\left(K^{1/2-2\gamma/d} + K^{1/2}\delta_{NT}^{-2}\right) O(1)\, O_P\left(K/\sqrt{NT}\right) \\
&= O_P\left(K^{3/2-2\gamma/d} + K^{3/2}\delta_{NT}^{-2}\right) = o_P(1)
\end{aligned}$$

by (.0.45), Lemma .0.10, 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

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

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). We further decompose $\Pi_{2NT,1}$ as follows

$$\Pi_{2NT,1} = \frac{\left\|v_x^K\right\|}{\sqrt{NT}}\left\{\operatorname{tr}[\mathbf{u} P_{f^0} \mathbf{u}' M_{\lambda^0} \mathbf{P}_{(a)} \Phi] - \operatorname{tr}\left[\mathbf{u}\mathbf{u}' M_{\lambda^0} \mathbf{P}_{(a)} \Phi\right]\right\} \equiv \Pi_{2NT,11} + \Pi_{2NT,12}, \text{ say.}$$

To show (i), it suffices to prove 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

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

For $\Pi_{2NT,11a}$, by Lemmas .0.28(i) and (v) in the supplemental appendix and Lemmas .0.12(i)-(ii),

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

For $\Pi_{2NT,11b}$, by Lemmas .0.28(i) and (v), we have $\Pi_{2NT,11b} \leq \frac{CR}{\sqrt{NT}}\left\|\mathbf{e}_g\right\|^2\left\|\mathbf{P}_{(a)}\right\|\left\|\Phi\right\| = O_P\left(\sqrt{NT}K^{-2\gamma/d}\right) = o_P(1)$. For $\Pi_{2NT,11c}$, by Lemmas .0.12(i) we have

$$
\begin{aligned}
\Pi_{2NT,11c} & \leq \frac{C}{\sqrt{NT}}\left|\operatorname{tr}\left[\lambda^{0\prime}\mathbf{e}f^0\left(f^{0\prime}f^0\right)^{-1}f^0\mathbf{e}_g'M_{\lambda^0}\mathbf{P}_{(a)}f^0\left(f^{0\prime}f^0\right)^{-1}\left(\lambda^{0\prime}\lambda^0\right)^{-1}\right]\right| \\
& \leq \frac{CR}{\sqrt{NT}}\left\|\lambda^{0\prime}\mathbf{e}f^0\right\|\left\|\left(f^{0\prime}f^0\right)^{-1}\right\|^2\left\|\left(\lambda^{0\prime}\lambda^0\right)^{-1}\right\|\left\|f^0\right\|^2\left\|\mathbf{e}_g\right\|\left\|\mathbf{P}_{(a)}\right\| \\
& = O_P\left(K^{-\gamma/d}\right).
\end{aligned}
$$

For $\Pi_{2NT,11d}$, by Lemmas .0.12(ii) and (vi) we have

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

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

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

First, decompose $\Pi_{2NT,12a}$ as follows

$$\begin{aligned}\Pi_{2NT,12a} &= -\frac{\left\|v_x^K\right\|}{\sqrt{NT}}\operatorname{tr}\left[E_{\mathscr{D}}\left(\mathbf{e}\mathbf{e}'\right)M_{\lambda^0}\mathbf{P}_{(a)}\Phi\right]-\frac{\left\|v_x^K\right\|}{\sqrt{NT}}\operatorname{tr}\left\{\left[\mathbf{e}\mathbf{e}'-E_{\mathscr{D}}\left(\mathbf{e}\mathbf{e}'\right)\right]M_{\lambda^0}\mathbf{P}_{(a)}\Phi\right\}\\ &\equiv \Pi_{2NT,12aa}+\Pi_{2NT,12ab},\text{ say.}\end{aligned}$$

Clearly, $\Pi_{2NT,12aa}=-\kappa_{NT}^{-1}b_2(x)$ and $\left|\kappa_{NT}^{-1}b_2(x)\right|\le R\kappa_{NT}^{-1}\left\|v_x^K\right\|\|E_{\mathscr{D}}(\mathbf{e}\mathbf{e}'/T)\|\|M_{\lambda^0}\|$ $\times\left\|\mathbf{P}_{(a)}\right\|\|f^0\left(f^{0\prime}f^0\right)^{-1}\left(\lambda^{0\prime}\lambda^0\right)^{-1}\lambda^0\|=O_P\left(\kappa_{NT}^{-1}\right)$ by Lemmas .0.28(i) and (v). Let $P_i^\lambda\equiv N^{-1}\sum_{l=1}^N\alpha_{li}P_l$, $G^0\equiv(f^{0\prime}f^0/T)^{-1}(\lambda^{0\prime}\lambda^0/N)^{-1}$ and $A_i\equiv v_x^{K\prime}[E_{\mathscr{D}}(P_i')$-$E_{\mathscr{D}}(P_i^{\lambda\prime})]\times f^0G^0T^{-1}$. Then $\Pi_{2NT,12ab}$ can be decomposed as follows

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

The first term is $O_P\left(K^{1/2}/N^{1/2}\right)$ by Lemma .0.12(vii), and the second term in the above expression is bounded by $\frac{1}{\sqrt{T}}\{\frac{1}{NT}\sum_{i=1}^N||\sum_{s=1}^T v_x^{K\prime}\left(p_{is}^c-p_{is}^{\lambda c}\right)f_s^{0\prime}G^0||^2\}^{1/2}\times$ $\{\frac{1}{N^2T}\sum_{i=1}^N||\sum_{j=1}^N\sum_{t=1}^T\lambda_j^0\left[e_{it}e_{jt}-E_{\mathscr{D}}\left(e_{it}e_{jt}\right)\right]||^2\}^{1/2}$, which is of order $O_P(K^{1/2}T^{-1/2})$ by Lemma .0.12$(viii)$ and (ix). It follows that $\Pi_{2NT,12ab}=O_P(K^{1/2}(N^{-1/2}+T^{-1/2}))$.

For $\Pi_{2NT,12b}$, we have $\left|\Pi_{2NT,12b}\right|\le\frac{CR}{\sqrt{NT}}\left\|\mathbf{e}_g\right\|^2\left\|\mathbf{P}_{(a)}\right\|\|\Phi\|=O_P(\sqrt{NT}K^{-2\gamma/d})$ $=o_P(1)$. For $\Pi_{2NT,12c}$ and $\Pi_{2NT,12d}$, we can show that they are both bounded from above by $CR(NT)^{-1/2}\left\|\mathbf{e}_g\right\|\|\mathbf{e}\|\left\|\mathbf{P}_{(a)}\right\|\|\Phi\|=O_P(\sqrt{NT}K^{-\gamma/d}\delta_{NT}^{-1})=o_P(1)$. It follows that $\Pi_{2NT,12}=-\kappa_{NT}^{-1}b_2(x)+o_P(1)$.

Now we consider (ii)**.** Noting that $M_{\lambda^0} = I_N - P_{\lambda^0}$, we have

$$\begin{aligned}
\Pi_{2NT,2} &= -\frac{\left\|v_x^K\right\|}{\sqrt{NT}}\left\{\operatorname{tr}\left[\mathbf{u}'\mathbf{u}M_{f^0}\mathbf{P}'_{(a)}\Phi'\right] - \operatorname{tr}\left[\mathbf{u}'P_{\lambda^0}\mathbf{u}M_{f^0}\mathbf{P}'_{(a)}\Phi'\right]\right\} \\
&\equiv \Pi_{2NT,21} + \Pi_{2NT,22}, \text{ say.}
\end{aligned}$$

Noting that $\mathbf{u} = \mathbf{e} + \mathbf{e}_g$ and $\left\|v_x^K\right\| = O_P(1)$, we have

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

For $\Pi_{2NT,22a}$, by Lemma .0.12(i) and (v) we have

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

Similar to the study of $\Pi_{1NT,12}$, we can show that $\Pi_{2NT,22s} = o_P(1)$ for $s = b, c, d$. It follows that $\Pi_{2NT,22} = o_P(1)$.

For $\Pi_{2NT,21}$, we have

$$\begin{aligned}
&\Pi_{2NT,21} \\
&= -\sqrt{N/T}\left\|v_x^K\right\|\operatorname{tr}\left[E_{\mathscr{D}}\left(\mathbf{e}'\mathbf{e}/N\right)M_{f^0}\mathbf{P}'_{(a)}\Phi'\right] - \tfrac{1}{\sqrt{NT}}\operatorname{tr}\left\{\left[\mathbf{e}'\mathbf{e} - \mathbf{E}_{\mathscr{D}}\left(\mathbf{e}'\mathbf{e}\right)\right]M_{f^0}\mathbf{P}'_{(a)}\Phi'\right\} \\
&\equiv -\kappa_{NT}b_3(x) - \Pi_{2NT,21a}, \text{ say.}
\end{aligned}$$

It is easy to show that $|\kappa_{NT}b_3(x)| = O_P(\kappa_{NT})$ by Lemmas .0.28 (i) and (v). For

$\Pi_{2NT,21a}$, by Lemmas .0.12(x) and (xi), we have

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

where $B_t \equiv v_x^{K\prime}[E_{\mathscr{D}}(P_{\cdot t}-P_{\cdot t}^{f})'\lambda^0 G^0 N^{-1}$, $P_{\cdot t}^{f}\equiv T^{-1}\sum_{l=1}^{T}\eta_{lt}P_{\cdot l}$, $p_{jt}^{fc}=p_{jt}^{f}-E_{\mathscr{D}}(p_{jt}^{f})$. By Lemma .0.12$(x)$, the first term is $O_P\left(K^{1/2}/T^{1/2}\right)$. By Cauchy-Schwarz inequlaty, the second term is bounded by $\frac{1}{\sqrt{N}}\{\frac{1}{NT}\sum_{t=1}^{T}||\sum_{j=1}^{N}v_x^{K\prime}[p_{jt}^{c}-p_{jt}^{fc}]\lambda_j^{0\prime}G^0||^2\}^{1/2}$ $\times\{\frac{1}{NT^2}\sum_{t=1}^{T}||\sum_{i=1}^{N}\sum_{s=1}^{T}f_s^{0\prime}\left[e_{it}e_{is}-E_{\mathscr{D}}\left(e_{it}e_{is}\right)\right]||^2\}^{1/2}$,which is $O_P(\sqrt{K/N})$ by Lemmas .0.12(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 .0.12(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 .0.12(iv), we have

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

This completes the proof of the proposition. ■

Bias-corrected estimator

Lemma .0.13 *Suppose that the assumptions in Theorem 3.3.3 hold. Then we have*
(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 .0.14 *Suppose that the assumptions in Theorem 3.3.3 hold. Then we have*
(i) $||\hat{b}_1-b_1|| = O_P(\sqrt{K}\sum_{\tau=M_T}^{T}\alpha_{\mathscr{D}}^{(3+2\delta)/(4+2\delta)}(\tau)+M_T\sqrt{K}\delta_{NT}^{-1})$;
(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 3.3.3. We first make the following decomposition:

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

Noting that $DB_1 \xrightarrow{d} N(0,1)$ by Theorem 3.3.2, 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)**.** Recall that $\hat{B}_K(x) = -\kappa_{NT}\hat{b}_1(x)-\kappa_{NT}^{-1}\hat{b}_2(x)-\kappa_{NT}\hat{b}_3(x)$ where $\hat{b}_s(x) = \hat{V}_K^{-1/2}(x)\,p^K(x)\times\hat{W}_{NT}^{-1}\hat{b}_s$. It follows that

$$
\begin{aligned}
DB_2 &= \kappa_{NT}\left[\hat{b}_1(x)-b_1(x)\right]+\kappa_{NT}^{-1}\left[\hat{b}_2(x)-b_2(x)\right]+\kappa_{NT}\left[\hat{b}_3(x)-b_3(x)\right] \\
&\equiv DB_{21}+DB_{22}+DB_{23}, \text{ say.}
\end{aligned}
$$

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

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

Recalling that $v_x^K = V^K(x)^{-1/2}\,\tilde{W}^{-1}p^K(x)$ with $\left\|v_x^K\right\| = O_P(1)$, by Lemma .0.14$(i)$ and Assumption 9 we have

$$
|DB_{21a}| \le \kappa_{NT}\left\|v_x^K\right\|\left\|\hat{b}_1-b_1\right\| = O_P\left[\kappa_{NT}\sqrt{K}\left(\sum_{\tau=M_T}^{\infty}\alpha_{\mathscr{D}}^{\frac{3+2\delta}{4+2\delta}}(\tau)+M_T\delta_{NT}^{-1}\right)\right] = o_P(1).
$$

Lemmas .0.10 and .0.13 and Minkowski inequality, $||\tilde{W}$-$\hat{W}_{NT}||_F$=$O_P[K(K^{-\gamma/d}$+$\delta_{NT}^{-1})]$. This, in conjunction with Assumption 7, implies that $||\hat{W}_{NT}^{-1}|| = O_P(1)$. Then by Lemma .0.14$(i)$ and Assumption 9, we have

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

Similarly,

$$
\begin{aligned}
DB_{21c} &\le \kappa_{NT}\left\|v_x^K\right\|\left\|\tilde{W}-\hat{W}_{NT}\right\|_F\left\|\hat{W}_{NT}^{-1}\right\|\left\|b_1\right\| \\
&= \kappa_{NT}O_P(1)\,O_P\left[K\left(K^{-\gamma/d}+\delta_{NT}^{-1}\right)\right]O_P(\sqrt{K}) = o_P(1).
\end{aligned}
$$

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 .0.13(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

$$\begin{aligned} \left|V_K^{1/2}(x)/\hat{V}_K^{1/2}(x)-1\right| &= \left|\frac{\hat{V}_K(x)-V_K(x)}{\hat{V}_K^{1/2}(x)\left[\hat{V}_K^{1/2}(x)+V_K^{1/2}(x)\right]}\right| \\ &= O_P\left[K\delta_{NT}^{-1}+(NT)^{1/4}K\left(\delta_{NT}^{-2}+K^{-\gamma/d}\right)\right] \quad (.0.52) \end{aligned}$$

Consequently, $\left|DB_{21d,1}\right|\leq\kappa_{NT}\left|V_K^{1/2}(x)/\hat{V}_K^{1/2}(x)-1\right|\left\|v_x^K\right\|\left\|\hat{b}_1\right\|=\kappa_{NT}O_P[K\delta_{NT}^{-1}+(NT)^{1/4}K(\delta_{NT}^{-2}+K^{-\gamma/d})]O_P\left(K^{1/2}\right)=o_P(1)$ .Similarly, we can show that $\left|DB_{21d,2}\right|=o_P(1)$. Then $(i1)$ follows. Analogously, we can show $(i2)$ and $(i3)$ by Lemmas .0.13 and .0.14.

Proof of (ii). By (.0.52), $\left|\hat{A}_{NT}/A_{NT}-1\right|=\left|\frac{\hat{V}_K(x)-V_K(x)}{\hat{V}_K^{1/2}(x)\left[\hat{V}_K^{1/2}(x)+V_K^{1/2}(x)\right]}\right|=O_P[K\delta_{NT}^{-1}+(NT)^{1/4}K(\delta_{NT}^{-2}+K^{-\gamma/d})]=o_P(1)$. 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 $|DB_4|\leq\left|\hat{A}_{NT}/A_{NT}-1\right||B_K(x)|=O_P[K\delta_{NT}^{-1}+(NT)^{1/4}K\left(\delta_{NT}^{-2}+K^{-\gamma/d}\right)][O_P(\kappa_{NT}\sqrt{K})+O_P\left(\kappa_{NT}^{-1}\right)+O_P(\kappa_{NT})]=o_P(1)$.

■

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

$$\begin{aligned} b_{1,k}^{(l)} &\equiv \frac{1}{N}\mathrm{tr}\left[P_{f^0}E_{\mathscr{D}}\left(\mathbf{e}'\mathbf{X}_k\right)\right], b_{2,k}^{(l)} \equiv \frac{1}{T}\mathrm{tr}\left[E_{\mathscr{D}}\left(\mathbf{e}\mathbf{e}'\right)M_{\lambda^0}\mathbf{X}_k\Phi\right], \text{ and} \\ b_{3,k}^{(l)} &\equiv \frac{1}{N}\mathrm{tr}\left[E_{\mathscr{D}}\left(\mathbf{e}'\mathbf{e}\right)M_{f^0}\mathbf{X}_k'\Phi'\right]. \end{aligned} \tag{.0.53}$$

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

Lemma .0.15 *Suppose that the assumptions in Theorem 3.4.1 hold. Then*
(i) $\left\|Q_{wpp,NT} - Q_{wpp}\right\|_F = O_P(K/(NT)^{1/2})$;
(ii) $\left\|Q_{wpx,NT} - Q_{wpx}\right\|_F = O_P(K^{1/2}/(NT)^{1/2})$;
(iii) $\left\|D_{NT} - D\right\|_F = O_P((NT)^{-1/2})$;
(iv) $\left\|\Omega_{xz,NT} - \tilde{\Omega}_{xz}\right\|_F = O_P((NT)^{-1/2})$.

Lemma .0.16 *Suppose that the assumptions in Theorem 3.4.1 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* $\left\|R_{\beta,NT}\right\| = o_P(\gamma_{NT})$.

Lemma .0.17 *Suppose that the assumptions in Theorem 3.4.1 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 3.4.1. 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 $\Gamma_{NT} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[p_{it}'\hat{\beta}_{bc} - X_{it}'\hat{\theta}]^2 w_{it}$ as follows

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

where

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

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) $\gamma_{NT}^{-2}(\Gamma_{NT2}+\Gamma_{NT3}-2\Gamma_{NT9}) = A^{\Delta}+o_P(1)$, (iii) $\gamma_{NT}^{-2}\Gamma_{NTs} = o_P(1)$ for $s=4,..,8,10$. We prove (i) in Proposition .0.9 below.

For (ii), by Lemma .0.17

$$\begin{aligned}\hat{\theta}-\theta^0 &= \gamma_{NT}D_{NT}^{-1}\Upsilon_{NT}+D^{-1}\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\tilde{X}_{it}e_{it}+B_{\theta,NT}+R_{\theta,NT}\\ &= \gamma_{NT}D_{NT}^{-1}\Upsilon_{NT}+O_P[\delta_{NT}^{-2}+(NT)^{-1/2}]\\ &= \gamma_{NT}D_{NT}^{-1}\Upsilon_{NT}+o_P(\gamma_{NT}) \qquad (.0.54)\end{aligned}$$

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 $2\gamma_{NT}^{-2}\Gamma_{NT9}$=$2\gamma_{NT}^{-2}\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\gamma_{NT}w_{it}\Delta_{it}X_{it}'$ $\times[\gamma_{NT}D_{NT}^{-1}\Upsilon_{NT}\ +o_P(\gamma_{NT})] = \frac{2}{NT}\sum_{i=1}^N\sum_{t=1}^T w_{it}\ \Delta_{it}X_{it}'D_{NT}^{-1}\Upsilon_{NT}+o_P(1)$. 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 (.0.54). 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

.0.16-.0.17, we have

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

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

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

For $\tilde{\Gamma}_{NT51a}$, we have

$$\left|\tilde{\Gamma}_{NT51a}\right|\leq\gamma_{NT}^{-1}\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{Z}_{it}' e_{it}\tilde{W}^{-1}Q_{wpx}D^{-1}\right\|\|\Upsilon_{NT}\|=O_P\left(\gamma_{NT}^{-1}(NT)^{-1/2}\right)=o_P(1)$$

as $||(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{Z}_{it}' e_{it}\tilde{W}^{-1}Q_{px}D^{-1}||=O_P[(NT)^{-1/2}]$ by Chebyshev inequality and the fact that $E_{\mathscr{D}}||\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{Z}_{it}' e_{it}\mathscr{H}_{px}||^2=\frac{1}{NT}\text{tr}(\tilde{\Omega}\tilde{W}^{-1}Q_{wpx}D^{-2}Q_{wpx}'\tilde{W}^{-1})$ $\leq d/(NT)\,\mu_1(\tilde{\Omega})\mu_1(D^{-2})[\mu_1(\tilde{W}^{-1})]^2\ \mu_1\left(Q_{wpx}'Q_{wpx}\right)=O_P((NT)^{-1})$ by Assumption 11 and Lemma .0.28(vi). By the fact that $||\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}e_{it}\tilde{Z}_{it}'\tilde{W}^{-1}||=O_P(\sqrt{\frac{K}{NT}})$, Lemma .0.15, and Assumption $11(ii)$, we have

$$\begin{aligned}\left|\tilde{\Gamma}_{NT51b}\right| &\leq \gamma_{NT}^{-1}\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}e_{it}\tilde{Z}_{it}'\tilde{W}^{-1}\right\|\left\|Q_{wpx,NT}\right\|\left\|D_{NT}^{-1}-D^{-1}\right\|_{F}\left\|\Upsilon_{NT}\right\| \\ &= \gamma_{NT}^{-1}O_{P}\left(K^{1/2}(NT)^{-1/2}\right)O_{P}\left((NT)^{-1/2}\right)O_{P}(1)=o_{P}(1)\end{aligned}$$

and

$$\begin{aligned}\left|\tilde{\Gamma}_{NT51c}\right| &\leq \gamma_{NT}^{-1}\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}e_{it}\tilde{Z}_{it}'\tilde{W}^{-1}\right\|\left\|Q_{wpx,NT}-Q_{wpx}\right\|_{F}\left\|D^{-1}\right\|\left\|\Upsilon_{NT}\right\| \\ &= \gamma_{NT}^{-1}O_{P}\left(K^{1/2}(NT)^{-1/2}\right)O_{P}\left(K^{1/2}(NT)^{-1/2}\right)O_{P}(1)=o_{P}(1).\end{aligned}$$

It follows that $\tilde{\Gamma}_{NT51}=o_{P}(1)$. For $\tilde{\Gamma}_{NT52}$, we decompose it as follows:

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

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

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

So $\tilde{\Gamma}_{NT52a} = o_P(1)$ by Chebyshev inequality. For $\tilde{\Gamma}_{NT52b}$, we have $\tilde{\Gamma}_{NT52b} = \frac{\mathrm{tr}\left(\mathscr{H}_{px}\Omega_{\tilde{x}\tilde{z}}\right)}{\sqrt{\mathbb{V}_{NT}}}$ $+\frac{1}{\sqrt{\mathbb{V}_{NT}}}\mathrm{tr}\left\{\mathscr{H}_{px}\left(\Omega_{\tilde{x}\tilde{z},NT} - \Omega_{\tilde{x}\tilde{z}}\right)\right\} \equiv \tilde{\Gamma}_{NT52b,1} + \tilde{\Gamma}_{NT52b,2}$, say. For $\tilde{\Gamma}_{NT52b,1}$, using Lemma .0.15(v), we have

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

where we use the fact $\|\Omega_{\tilde{x}\tilde{z}}\|^2 \leq \mu_1(\tilde{\Omega})\mu_1\left(\Omega_{\tilde{x}\tilde{x}}\right) = O_P(1)$ 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

$$
\begin{aligned}
\left|\tilde{\Gamma}_{NT52b,2}\right| &\leq \mathbb{V}_{NT}^{-1/2}\left\|\tilde{W}^{-1}Q_{wpx}D^{-1}\right\|_F\left\|\Omega_{\tilde{x}\tilde{z},NT} - \Omega_{\tilde{x}\tilde{z}}\right\|_F \\
&\leq \mathbb{V}_{NT}^{-1/2}\left\|D^{-1}\right\|_F\left\|\tilde{W}^{-1}\right\|\left\|Q_{wpx}\right\|\left\|\Omega_{\tilde{x}\tilde{z},NT} - \Omega_{\tilde{x}\tilde{z}}\right\|_F \\
&= \mathbb{V}_{NT}^{-1/2}O_P(K^{1/2}N^{-1/2}T^{-1/2}) = o_P(1).
\end{aligned}
$$

Similarly, we can show that $\tilde{\Gamma}_{NT52s} = o_P(1)$ for $s = c, d$. For, $\tilde{\Gamma}_{NT52e}$, we have

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

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 .0.16, we have

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

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 $\left\|\beta^{\Delta}\right\|\leq C_{\Delta}<\infty$ 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

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

Analogously to the study of $\left|\gamma_{NT}^{-2}\Gamma_{NT6}\right|$, we have $\left|\tilde{\Gamma}_{NT7b}\right|\leq C\gamma_{NT}^{-1}||\hat{\beta}_{bc}-\beta^{0}||O_{P}(K^{-\gamma/d})$ $=o_{P}(1)$. For $\tilde{\Gamma}_{NT7a}$, by Lemma .0.15 we have

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

where $\tilde{\Gamma}_{NT7a1} \equiv \frac{1}{\gamma_{NT}NT}\sum_{i=1}^{N}\sum_{t=1}^{T} e_{it}\tilde{Z}_{it}'\tilde{W}^{-1}Q_{wpp}\beta_{\Delta}^{0}$. For $\tilde{\Gamma}_{NT7a1}$, we have $E_{\mathscr{D}}[\tilde{\Gamma}_{NT7a1}] = 0$ and

$$\begin{aligned}
E_{\mathscr{D}}\left[\tilde{\Gamma}_{NT7a1}^{2}\right] &= \frac{1}{NT\gamma_{NT}^{2}}\mathrm{tr}\left\{\tilde{\Omega}W^{-1}Q_{wpp}\beta_{\Delta}^{0}\beta_{\Delta}^{0\prime}Q_{wpp}W^{-1}\right\} \\
&\leq \frac{1}{NT}\mu_{1}(\tilde{\Omega})\mu_{1}^{2}\left(W^{-1}\right)\gamma_{NT}^{-2}\mu_{1}^{2}\left(Q_{wpp}\right)\left\|\beta_{\Delta}^{0}\right\|^{2} = O_{P}\left(\frac{\gamma_{NT}^{2}}{NT}\right) = o_{P}(1)
\end{aligned}$$

Then $\tilde{\Gamma}_{NT7a1} = o_{P}(1)$ by Chebyshev inequality. It follows that $\gamma_{NT}^{-2}\Gamma_{NT7} = o_{P}(1)$.

■

Proposition .0.9 *Suppose that the assumptions in Theorem 3.4.1 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 $\left\|Q_{wpp,NT} - Q_{wpp}\right\| = O_{P}(\frac{K}{\sqrt{NT}})$ and $||\hat{\beta}_{bc} - \beta^{0}|| = O_{P}(\sqrt{\frac{K}{NT}})$, we have $\gamma_{NT}^{-2}\Gamma_{NT1} = \Gamma_{NT1,1} + \gamma_{NT}^{-2}O_{P}(\frac{K}{\sqrt{NT}})O_{P}(\sqrt{\frac{K}{NT}}) = o_{P}(1)$, where $\Gamma_{NT1,1} \equiv \gamma_{NT}^{-2}(\hat{\beta}_{bc} - \beta^{0})'Q_{pp}(\hat{\beta}_{bc} - \beta^{0})$. We are left to show that $J_{NT1} \equiv \frac{NT\Gamma_{NT1,1} - \mathbb{B}_{NT}}{\sqrt{\mathbb{V}_{NT}}} \xrightarrow{d} N(0,1)$.

Let $\bar{Q}_{pp} \equiv \tilde{W}^{-1}Q_{wpp}\tilde{W}^{-1}$, $H_{ij,ts} \equiv \tilde{Z}_{it}'\bar{Q}_{pp}\tilde{Z}_{js}$, and $H_{ij} \equiv \tilde{Z}_{i}'\bar{Q}_{pp}\tilde{Z}_{j}$. Decompose $\gamma_{NT}^{-2}\Gamma_{NT1,1}$ as follows

$$\begin{aligned}
\gamma_{NT}^{-2}\Gamma_{NT1,1} &= \frac{1}{N^{2}T^{2}\gamma_{NT}^{2}}\sum_{i=1}^{N}\sum_{j=1}^{N} e_{i}'H_{ij}e_{j} + 2\gamma_{NT}^{-2}R_{\beta}'Q_{wpp}\left(\hat{\beta}_{bc} - \beta^{0}\right) + \gamma_{NT}^{-2}R_{\beta}'Q_{wpp}R_{\beta} \\
&\equiv \tilde{J}_{NT} + o_{P}(1), \text{ say.}
\end{aligned}$$

For $\tilde{J}_{NT}$, we have

$$\begin{aligned}
\tilde{J}_{NT} - \frac{\mathbb{B}_{NT}}{\sqrt{\mathbb{V}_{NT}}} &= \frac{1}{N^{2}T^{2}\gamma_{NT}^{2}}\sum_{1\leq i\neq j\leq N} e_{i}'H_{ij}e_{j} + \frac{1}{N^{2}T^{2}\gamma_{NT}^{2}}\sum_{i=1}^{N}\sum_{1\leq t\neq s\leq T} H_{ii,ts}e_{it}e_{is}] \\
&\equiv \tilde{J}_{NT,1} + \tilde{J}_{NT,2}, \text{say.}
\end{aligned}$$

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). We rewrite $\tilde{J}_{NT,1}$ as follows

$$\tilde{J}_{NT,1} \equiv \frac{1}{NT\mathbb{V}_{NT}^{1/2}}\sum_{1\leq i\neq j\leq N} e_{i}'H_{ij}e_{j} = \sum_{1\leq i<j\leq N} W_{ij},$$

where $W_{ij} \equiv W_{NT}\left(u_i, u_j\right) \equiv 2\left(NT\right)^{-1} \mathbb{V}_{NT}^{-1/2} \sum_{1\leq s,t\leq T} H_{ij,ts} e_{it} e_{js}$ and $u_i \equiv (\tilde{Z}_i', e_i)'$. Noting that $\tilde{J}_{NT,1}$ is a second order degenerate U-statistic which is "clean" since $E_{\mathscr{D}}[W_{NT}\left(u_i, u\right)] = E_{\mathscr{D}}[W_{NT}\left(u, u_j\right)] = 0$ 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 $(i1)$ $\mathrm{Var}_{\mathscr{D}}(\tilde{J}_{NT,1})$ $= 1 + o_P\left(1\right)$, $(i2)$ $G_I \equiv \sum_{1\leq i<j<N} E_{\mathscr{D}}(W_{ij}^4) = o_P\left(1\right)$, $(i3)$ $G_{II} \equiv \sum_{1\leq i<j<l\leq N} E_{\mathscr{D}}(W_{il}^2 W_{jl}^2$ $+ W_{ij}^2 W_{il}^2 + W_{ij}^2 W_{lj}^2) = o_P\left(1\right)$, and $(i4)$ $G_{III} \equiv \sum_{1\leq i<j<r<l\leq N} E_{\mathscr{D}}(W_{ij}W_{ir}W_{lj}W_{lr} + W_{ij}W_{il}$ $\times W_{rj}W_{rl} + W_{ir}W_{il}W_{jr}W_{jl}) = o_P\left(1\right)$.

For $(i1)$, noting that $E_{\mathscr{D}}(\tilde{J}_{NT,1}) = 0$ by Assumption 5(ii) and by the same assumption we have

$$
\begin{aligned}
\mathrm{Var}_{\mathscr{D}}\left(\tilde{J}_{NT,1}\right) &= \frac{4}{N^2T^2\mathbb{V}_{NT}} \sum_{1\leq i<j\leq N} \sum_{t_1=1}^{T} \sum_{t_2=1}^{T} \sum_{s_1=1}^{T} \sum_{s_2=1}^{T} E_{\mathscr{D}}\left(H_{ij,t_1s_1} H_{ij,t_2s_2} e_{it_1} e_{js_1} e_{it_2} e_{js_2}\right) \\
&= \frac{4}{N^2T^2\mathbb{V}_{NT}} \sum_{1\leq i<j\leq N} \sum_{t=1}^{T} \sum_{s=1}^{T} E_{\mathscr{D}}\left(H_{ij,ts}^2 e_{it}^2 e_{js}^2\right) \\
&= \frac{4}{N^2T^2\mathbb{V}_{NT}} \sum_{1\leq i<j\leq N} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathrm{tr}\left\{\bar{Q}_{pp} E_{\mathscr{D}}\left(\tilde{Z}_{js}\tilde{Z}_{js}' e_{js}^2\right) \bar{Q}_{pp} E_{\mathscr{D}}\left(\tilde{Z}_{it}\tilde{Z}_{it}' e_{it}^2\right)\right\} \\
&= \frac{2}{\mathbb{V}_{NT}} \mathrm{tr}\left(\bar{Q}_{pp}\tilde{\Omega}\bar{Q}_{pp}\tilde{\Omega}\right) - \frac{1}{N^2\mathbb{V}_{NT}} \sum_{i=1}^{N} \mathrm{tr}\left(\bar{Q}_{pp}\tilde{\Omega}_i\bar{Q}_{pp}\tilde{\Omega}_i\right) \\
&= 1 - O_P\left(N^{-1}\right) = 1 + o_P\left(1\right)
\end{aligned}
$$

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

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

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

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

First, note that the term inside the last summation takes values 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 .0.26) in the supplementary appendix, we have

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

where $\Phi_{1,it,k} \equiv \left\|\phi_{it,k}\right\|_{8+4\delta,\mathscr{D}}^4$. Let $C_{1\alpha}(T) \equiv \sum_{m=1}^T \alpha_{\mathscr{D}}^{(1+\delta)/(2+\delta)}(m)$. Then

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

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

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

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

For $G_{II,11}$, we have

$$G_{II,11} \le \frac{192}{N^2T^2\mathbb{V}_{NT}^2}\sum_{l=1}^{N}\sum_{1\le t_3<t_4<t_6\le T} E_{\mathscr{D}}\left\{\operatorname{tr}\left[e_{lt_4}\tilde{Z}_{lt_4}'\bar{Q}_{pp}\tilde{\Omega}\bar{Q}_{pp}\tilde{Z}_{lt_3}e_{lt_3}\right]\operatorname{tr}\left[\bar{Q}_{pp}\tilde{\Omega}\bar{Q}_{pp}\tilde{Z}_{lt_6}\tilde{Z}_{lt_6}'e_{lt_6}^2\right]\right\}.$$

Noting that $E_{\mathscr{D}}[e_{lt_4}\tilde{Z}_{lt_4}'\bar{Q}_{pp}\tilde{\Omega}\bar{Q}_{pp}\tilde{Z}_{lt_3}e_{lt_3}] = 0$, by the conditional Davydov inequality we have

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

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

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

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

by Assumption A5(i), where $C_{2\alpha}(T) \equiv T^{-1}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T} \alpha_{\mathscr{D}}^{\frac{1+\delta}{2+\delta}}(s-t) < \infty$. Similarly,

$$\begin{aligned} & G_{II,12} \\ \leq\; & \frac{48}{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} E_{\mathscr{D}}\left\{ \operatorname{tr}\left[E_{\mathscr{D}}(e^2_{it_1}\tilde{Z}_{it_1}\tilde{Z}'_{it_1})\bar{Q}_{pp}e^2_{lt_3}\tilde{Z}_{lt_3}\tilde{Z}'_{lt_3}\bar{Q}_{pp}\right] \right. \\ & \quad \left. \times \operatorname{tr}\left[E_{\mathscr{D}}(e^2_{jt_2}\tilde{Z}_{jt_2}\tilde{Z}'_{jt_2})\bar{Q}_{pp}e^2_{lt_6}\tilde{Z}_{lt_6}\tilde{Z}'_{lt_6}\bar{Q}_{pp}\right]\right\} \\ \leq\; & \frac{8}{N^2T^2\mathbb{V}^2_{NT}} \sum_{l=1}^{N} \sum_{1\leq t_3\neq t_6\leq T} E_{\mathscr{D}}\left\{ \operatorname{tr}\left[\tilde{\Omega}\bar{Q}_{pp}e^2_{lt_3}\tilde{Z}_{lt_3}\tilde{Z}'_{lt_3}\bar{Q}_{pp}\right] \operatorname{tr}\left[\tilde{\Omega}\bar{Q}_{pp}e^2_{lt_6}\tilde{Z}_{lt_6}\tilde{Z}'_{lt_6}\bar{Q}_{pp}\right]\right\} \\ \leq\; & \frac{8}{N^2T^2\mathbb{V}^2_{NT}} \sum_{l=1}^{N} \sum_{1\leq t_3\neq t_6\leq T} \left\{ E_{\mathscr{D}}\left[\operatorname{tr}\left(\tilde{\Omega}\bar{Q}_{pp}e^2_{lt_3}\tilde{Z}_{lt_3}\tilde{Z}'_{lt_3}\bar{Q}_{pp}\right)\right] E_{\mathscr{D}}\left[\operatorname{tr}\left(\tilde{\Omega}\bar{Q}_{pp}e^2_{lt_6}\tilde{Z}_{lt_6}\tilde{Z}'_{lt_6}\bar{Q}_{pp}\right)\right] \right. \\ & \quad \left. + 8\left\|\operatorname{tr}\left[\tilde{\Omega}\bar{Q}_{pp}e^2_{lt_3}\tilde{Z}_{lt_3}\tilde{Z}'_{lt_3}\bar{Q}_{pp}\right]\right\|_{4+2\delta,\mathscr{D}} \left\|\operatorname{tr}\left[\tilde{\Omega}\bar{Q}_{pp}e^2_{lt_6}\tilde{Z}_{lt_6}\tilde{Z}'_{lt_6}\bar{Q}_{pp}\right]\right\|_{4+2\delta,\mathscr{D}} \alpha_{\mathscr{D}}^{\frac{1+\delta}{2+\delta}}(|t_6-t_3|)\right\} \\ =\; & \frac{8\mu_1^2(\tilde{\Omega})\mu_1^4(\bar{Q}_{pp})}{N^2T^2\mathbb{V}^2_{NT}} \sum_{l=1}^{N} \sum_{1\leq t_3\neq t_6\leq T} E_{\mathscr{D}}\left[e^2_{lt_3}\left\|\tilde{Z}_{lt_3}\right\|^2\right] E_{\mathscr{D}}\left[e^2_{lt_6}\left\|\tilde{Z}_{lt_6}\right\|^2\right] \\ & + \frac{64\mu_1^2(\tilde{\Omega})\mu_1^4(\bar{Q}_{pp})}{N^2T^2\mathbb{V}^2_{NT}} \sum_{l=1}^{N} \sum_{1\leq t_3\neq t_6\leq T} \left\|e^2_{lt_3}\left\|\tilde{Z}_{lt_3}\right\|^2\right\|_{4+2\delta,\mathscr{D}} \left\|e^2_{lt_6}\left\|\tilde{Z}_{lt_6}\right\|^2\right\|_{4+2\delta,\mathscr{D}} \alpha_{\mathscr{D}}^{\frac{1+\delta}{2+\delta}}(|t_6-t_3|) \\ =\; & 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 $G_{III} \equiv \sum_{1\leq i<j<r<l\leq N} E_{\mathscr{D}}(W_{ij}W_{ir}W_{lj}W_{lr}+W_{ij}W_{il}W_{rj}W_{rl}+ W_{ir}W_{il}W_{jr}W_{jl})$ $\equiv \sum_{s=1}^{4} G_{III,s}$, say. By Assumptions 5$(ii)$, we have

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

where we use the facts that $\operatorname{tr}\left(\bar{Q}_{pp}\tilde{\Omega}\bar{Q}_{pp}\tilde{\Omega}\bar{Q}_{pp}\tilde{\Omega}\bar{Q}_{pp}\tilde{\Omega}\right) \leq \mu_1^4\left(\bar{Q}_{pp}\right)\mu_1^3(\tilde{\Omega})\operatorname{tr}(\tilde{\Omega}) = O_P(K)$ 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 detail is omitted to save space. ■

Proof of Theorem 3.4.2. 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 3.4.1 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).

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

Following the proof of Lemma .0.14(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 we have $\mathbb{DB}_{3NT} = \text{tr}[(\hat{w}-\tilde{w})\hat{\Omega}_{NT}]$. By Minkowski inequality,

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

By the matrix version of Cauchy-Schwarz inequality, the fact that $\text{tr}(AB) \leq \mu_1(B)\text{tr}(A)$ for any symmetric matrix A and p.s.d. matrix B, and Lemma .0.15, we have

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

Similarly, we can show that $w_{3NT,2} = O_P(K^2(K^{-2\gamma/d}+\delta_{NT}^{-2}))$ and $w_{3NT,3} = O_P(K(K^{-\gamma/d}$ $+\delta_{NT}^{-1}))$ by Lemmas .0.13 and .0.15. It follows that

$$\|\hat{w}-\tilde{w}\|_F = O_P(K\left(K^{-\gamma/d}+\delta_{NT}^{-1}\right)), \tag{.0.55}$$

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

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

Observe that $|\mathbb{D}\mathbb{V}_{1NT}|\leq[\mu_1(\tilde{w})]^2\,||\hat{\Omega}_{NT}-\tilde{\Omega}||^2=O_P(K^2/(NT))=o_P(1)$, and by (.0.55)

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

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

C Supplementary Material on Chapter 3

.0.7 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 (3.1.1) as (3.2.1). Compared with Moon and Weidner's (2010) linear model, the number of regressors increases as sample size (N, T) tends to infinity in (3.2.1) 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

$$\begin{aligned}
\Phi_1 &\equiv f^0\left(f^{0\prime}f^0\right)^{-1}\left(\lambda^{0\prime}\lambda^0\right)^{-1}\left(f^{0\prime}f^0\right)^{-1}f^{0\prime}, \text{ and} \\
\Phi_2 &\equiv \lambda^0\left(\lambda^{0\prime}\lambda^0\right)^{-1}\left(f^{0\prime}f^0\right)^{-1}\left(\lambda^{0\prime}\lambda^0\right)^{-1}\lambda^{0\prime}.
\end{aligned} \tag{.0.56}$$

Recall that $\Phi = f^0(f^{0\prime}f^0)^{-1}(\lambda^{0\prime}\lambda^0)^{-1}\lambda^{0\prime}$ and $\vartheta_{NT} = \sum_{k=0}^{K}\varepsilon_k\mathbf{P}_k$, where $\varepsilon_k = \beta_k - \beta_k^0$ for $k = 1, ..., K$, $\varepsilon_0 = \|\mathbf{u}\|/\sqrt{NT}$, 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 .0.10 *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 (3.2.6). 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.,*

$$\begin{aligned}
\mathscr{L}_{NT}^0(\beta) \equiv\ & \frac{1}{NT}\sum_{k_1=0}^{K}\sum_{k_2=0}^{K}\varepsilon_{k_1}\varepsilon_{k_2}L^{(2)}\left(\lambda^0, f^0, \mathbf{P}_{k_1}, \mathbf{P}_{k_2}\right) \\
& + \frac{1}{NT}\sum_{k_1=0}^{K}\sum_{k_2=0}^{K}\sum_{k_3=0}^{K}\varepsilon_{k_1}\varepsilon_{k_2}\varepsilon_{k_3}L^{(3)}\left(\lambda^0, f^0, \mathbf{P}_{k_1}, \mathbf{P}_{k_2}, \mathbf{P}_{k_3}\right) \\
& + O_P\left(\alpha_{NT}^4\right)
\end{aligned} \tag{.0.57}$$

where

$$\begin{aligned} L^{(2)}\left(\lambda^0, f^0, \mathbf{P}_{k_1}, \mathbf{P}_{k_2}\right) &\equiv tr\left(M_{\lambda^0}\mathbf{P}_{k_1}M_{f^0}\mathbf{P}'_{k_2}\right) \\ L^{(3)}\left(\lambda^0, f^0, \mathbf{P}_{k_1}, \mathbf{P}_{k_2}, \mathbf{P}_{k_3}\right) &\equiv -\frac{1}{3!}\sum_{all\ 6\ permutations\ for\ (k_1,k_2,k_3)} tr\left(M_{\lambda^0}\mathbf{P}_{k_1}M_{f^0}\mathbf{P}'_{k_2}\Phi\mathbf{P}'_{k_3}\right); \end{aligned}$$

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

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

where

$$\begin{aligned} M_{\lambda}^{(1)}\left(\lambda^0, f^0, \mathbf{P}_k\right) &= -M_{\lambda^0}\mathbf{P}_k\Phi - \Phi'\mathbf{P}'_k M_{\lambda^0} \\ M_{\lambda}^{(2)}\left(\lambda^0, f^0, \mathbf{P}_{k_1}, \mathbf{P}_{k_2}\right) &= M_{\lambda^0}\mathbf{P}_{k_1}\Phi\mathbf{P}_{k_2}\Phi + \Phi'\mathbf{P}'_{k_2}\Phi'\mathbf{P}'_{k_1}M_{\lambda^0} - M_{\lambda^0}\mathbf{P}_{k_1}M_{f^0}\mathbf{P}'_{k_2}\Phi_2 \\ &\quad -\Phi_2\mathbf{P}_{k_2}M_{f^0}\mathbf{P}'_{k_1}M_{\lambda^0} - M_{\lambda^0}\mathbf{P}_{k_1}\Phi_1\mathbf{P}'_{k_2}M_{\lambda^0} + \Phi'\mathbf{P}'_{k_1}M_{\lambda^0}\mathbf{P}_{k_2}\Phi; \end{aligned}$$

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

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

where

$$\begin{aligned} M_{f}^{(1)}\left(\lambda^0, f^0, \mathbf{P}_k\right) &= -M_{f^0}\mathbf{P}'_k\Phi' - \Phi\mathbf{P}_k M_{f^0} \\ M_{f}^{(2)}\left(\lambda^0, f^0, \mathbf{P}_{k_1}, \mathbf{P}_{k_2}\right) &= M_{f^0}\mathbf{P}'_{k_1}\Phi'\mathbf{P}'_{k_2}\Phi' + \Phi\mathbf{P}_{k_2}\Phi\mathbf{P}_{k_1}M_{f^0} - M_{f^0}\mathbf{P}'_{k_1}M_{\lambda^0}\mathbf{P}_{k_2}\Phi_1 \\ &\quad -\Phi_1\mathbf{P}'_{k_2}M_{\lambda^0}\mathbf{P}_{k_1}M'_{f^0} - M_{f^0}\mathbf{P}'_{k_1}\Phi_2\mathbf{P}_{k_2}M_{f^0} + \Phi\mathbf{P}_{k_1}M_{f^0}\mathbf{P}'_{k_2}\Phi'; \end{aligned}$$

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

$$\mathbf{Y} - \sum_{k=1}^{K}\beta_k\mathbf{P}_k = \lambda^0 f^{0\prime} + \varepsilon_0\mathbf{P}_0 + \varepsilon_1\mathbf{P}_1 + \cdots + \varepsilon_K\mathbf{P}_K, \qquad (.0.58)$$

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 (3.2.8) as follows:

$$\frac{1}{NT}\sum_{R+1}^{T}\mu_t\left[\left(\mathbf{Y}-\sum_{k=1}^{K}\beta_k\mathbf{P}_k\right)'\left(\mathbf{Y}-\sum_{k=1}^{K}\beta_k\mathbf{P}_k\right)\right]=\frac{1}{NT}\sum_{R+1}^{T}\mu_t\left[\mathscr{T}(1)\right] \quad (.0.59)$$

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

$$\mathscr{T}^0\equiv f^0\lambda^{0\prime}\lambda^0 f^{0\prime},\ \mathscr{T}^{(1)}\equiv\vartheta_{NT}\left(\lambda^0 f^{0\prime}+f^0\lambda^{0\prime}\right),\text{ and }\mathscr{T}^{(2)}\equiv\vartheta_{NT}\vartheta_{NT}. \quad (.0.60)$$

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

Since $\mathscr{T}(1)\equiv\mathscr{T}^0+\mathscr{T}^{(1)}+\mathscr{T}^{(2)}$, under some conditions to be specified later (see (.0.66) and (.0.67) below), we can expand the weighted mean $\hat{\lambda}(1)$ of the λ-group eigenvalues ($\lambda=0$ in this case) as

$$\hat{\lambda}(1)\equiv\frac{1}{T-R}\sum_{R+1}^{T}\mu_t\left[\mathscr{T}(1)\right]=0+\sum_{g=0}^{\infty}1^g\hat{\lambda}^{(g)}, \quad (.0.61)$$

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

$$\hat{\lambda}^{(g)} \equiv \frac{1}{T-R}\sum_{p=1}^{g}(-1)^{p+1}\sum_{\substack{v_1+\cdots+v_p=g,\\ m_1+\cdots+m_{p+1}=p-1\\ 2\geq v_j\geq 1, m_j\geq 0}}\operatorname{tr}\left(S^{(k_1)}\mathscr{T}^{(v_1)}S^{(k_2)}\cdots S^{(k_p)}\mathscr{T}^{(v_p)}S^{(k_{p+1})}\right), \quad (.0.62)$$

$$S^{(0)} \equiv -M_{\lambda^0},\ S^{(k)}\equiv\left[\lambda^0\left(\lambda^{0\prime}\lambda^0\right)^{-1}\left(f^{0\prime}f^0\right)^{-1}\left(\lambda^{0\prime}\lambda^0\right)^{-1}\lambda^{0\prime}\right]^k, \quad (.0.63)$$

and $\mathscr{T}^{(s)}$ ($s=1,2$) are defined in (.0.60). Note that $2\geq v_j$ comes from the facts that $\mathscr{T}^{(g)}\equiv 0$ 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 (.0.60) for $\mathscr{T}^{(1)}$ and $\mathscr{T}^{(2)}$ we have

$$
\begin{aligned}
& \frac{1}{NT}\sum_{R+1}^{T}\mu_t\left[\mathscr{T}(1)\right] \\
= & \frac{1}{NT}\sum_{g=1}^{\infty}\sum_{p=1}^{g}(-1)^{p+1}\sum_{\substack{v_1+\cdots+v_p=s \\ m_1+\cdots+m_{p+1}=p-1 \\ 2\geq v_j\geq 1,k_j\geq 0}}\operatorname{tr}(S^{(m_1)}\mathscr{T}^{(v_1)}S^{(m_2)}\cdots S^{\left(m_p\right)}\mathscr{T}^{\left(v_p\right)}S^{(m_{p+1})}) \\
= & \frac{1}{NT}\sum_{g=2}^{\infty}\sum_{k_1=0}^{K}\sum_{k_2=0}^{K}\cdots\sum_{k_g=0}^{K}\varepsilon_{k_1}\varepsilon_{k_2}\cdots\varepsilon_{k_g}L^{(g)}\left(\lambda^0,f^0,\mathbf{P}_{k_1},\cdots,\mathbf{P}_{k_g}\right) \qquad (.0.64)
\end{aligned}
$$

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

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

$$
\begin{aligned}
& \tilde{L}^{(g)}\left(\lambda^0,f^0,\mathbf{P}_{k_1},\cdots\mathbf{P}_{k_g}\right) \\
\equiv & \sum_{p=1}^{g}(-1)^{p+1}\sum_{\substack{v_1+\cdots+v_p=s, \\ m_1+\cdots+m_{p+1}=p-1, \\ 2\geq v_j\geq 1,k_j\geq 0}}\operatorname{tr}\left(S^{(m_1)}\mathscr{T}_{k_1\ldots}^{(v_1)}S^{(m_2)}\cdots S^{\left(m_p\right)}\mathscr{T}_{\ldots k_g}^{\left(v_p\right)}S^{\left(m_{p+1}\right)}\right),
\end{aligned}
\qquad (.0.65)
$$

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

To ensure that $\mathscr{T}(\varkappa)$ can be expanded at $\varkappa=1$ in (.0.64), 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

$$\left\|\mathscr{T}^{(1)}+\mathscr{T}^{(2)}\right\|\leq\frac{NT}{2}d_{\min}^2\left(\lambda^0,f^0\right). \qquad (.0.66)$$

2. Convergence of the expansion in eqn. (.0.64) in an infinite sequence with $\varkappa=1$ requires that the convergence radius is at least 1. Define $a\equiv\sqrt{NT}\left\|\vartheta_{NT}\right\|$ $2d_{\max}\left(\lambda^0,f^0\right)$, $c\equiv\left\|\vartheta_{NT}\right\|\left[2\sqrt{NT}d_{\max}\left(\lambda^0,f^0\right)\right]^{-1}$. It is straightforward to

show that

$$\left\|\mathscr{T}^{(1)}\right\| \leq a,\ \left\|\mathscr{T}^{(2)}\right\| \leq ac \text{ and } \left\|\mathscr{T}^{(s)}\right\| \equiv 0 \leq ac^{s-1} \text{ for } s = 3,4,\cdots. \tag{.0.67}$$

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

$$\frac{\|\vartheta_{NT}\|}{\sqrt{NT}} \leq r_0\left(\lambda^0, f^0\right) \equiv \left(\frac{4d_{\max}\left(\lambda^0, f^0\right)}{d_{\min}^2\left(\lambda^0, f^0\right)} + \frac{1}{2d_{\max}\left(\lambda^0, f^0\right)}\right)^{-1}. \tag{.0.68}$$

Step 2. Finite order truncation of the quasi-log-likelihood function. To conduct the asymptotic analysis, we need to truncate the expansion in (.0.64) to a finite order. Noting that $\left\|S^{(g)}\right\| = \left[NTd_{\min}^2\left(\lambda^0, f^0\right)\right]^{-g}$, $\left\|\mathscr{T}^{(1)}\right\| \leq a$, and $\left\|\mathscr{T}^{(2)}\right\| \leq ac$, we have

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

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

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

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

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

The infinite summation is convergent given $\alpha_{NT} < 1$, which is implied by $r_0(\lambda^0, f^0) > 1$. 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. ■

.0.8 Proofs of the technical lemmas

Convergence rate

Lemma .0.18 *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

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

by Assumptions $1(iii)$-(iv), $2(ii)$, and 5, Lemmas .0.28(vi), (i), and (xi), and the fact rank$(P_f) \leq R$.

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

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

by Assumption $2(i)$, Lemma .0.28(i), and the fact that

$$\frac{\left\|\mathbf{e}_g'\right\|_F^2}{NT} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}e_{g,i}^2 \leq \left\|g(x) - p^K(x)'\beta^0\right\|_{\infty,\bar{\omega}}^2\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(1+\|X_{it}\|^2\right)^{\bar{\omega}} = O_P\left(K^{-2\gamma/d}\right)$$

by Assumptions $3(i)$ and $4(i)$.

(iii) By Lemmas .0.28 (ii) and (iv), we have

$$\left|\frac{1}{NT}\mathrm{tr}\left(\lambda^0 f^{0\prime}M_f\mathbf{u}'\right)\right| \leq \mathrm{rank}\left(\lambda^0 f^{0\prime}M_f\mathbf{u}'\right)\frac{\left\|\lambda^0\right\|}{\sqrt{N}}\frac{\left\|f^0\right\|}{\sqrt{T}}\frac{\|\mathbf{e}\|+\left\|\mathbf{e}_g\right\|_F}{\sqrt{NT}} = O_P\left(\delta_{NT}^{-1} + K^{-\gamma/d}\right).$$

(iv) By Lemmas .0.28 (ii) and (iv), we have $\frac{1}{NT}\left|\mathrm{tr}\left(\mathbf{u}P_f\mathbf{u}'\right)\right| \leq \mathrm{rank}\left(\mathbf{u}P_f\mathbf{u}'\right)\frac{\|\mathbf{u}\|^2}{NT}\left\|P_f\right\|$ $= O_P\left(\delta_{NT}^{-2} + K^{-2\gamma/d}\right) = o_P(1)$. ■

Proof of Lemma .0.8. 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

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

where $\tilde{S}_{NT}(\beta, f) \equiv \frac{1}{NT}\mathrm{tr}\left[\left(\lambda^{0\prime}f^0 + \left\|\beta^0 - \beta\right\|\mathbf{P}_{(a)}\right)M_f\left(\lambda^0 f^{0\prime} + \left\|\beta^0 - \beta\right\|\mathbf{P}_{(a)}\right)'\right]$. 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 .0.9. Recall $V_K(x) \equiv p^K(x)'\tilde{W}^{-1}\tilde{\Omega}\tilde{W}^{-1}p^K(x)$ and $v_x^K \equiv \frac{\tilde{W}^{-1}p^K(x)}{\sqrt{V_K(x)}}$. 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 $\tilde{Z}_{it} = p_{it} - \frac{1}{N}\sum_{j=1}^{N}\alpha_{ij}E_{\mathscr{D}}\left(p_{jt}\right) - \frac{1}{T}\sum_{s=1}^{T}\eta_{ts}E_{\mathscr{D}}\left(p_{is}\right) + \frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T} \alpha_{ij}\eta_{ts}E_{\mathscr{D}}\left(p_{js}\right) \equiv p_{it}+\zeta_{it}$. Note that ζ_{it} is a $K\times 1$ $\mathscr{D}$-measurable vector, and

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

where we use the fact that $\left|\alpha_{ij}\right| \leq \varsigma_N^{-1}\left\|\lambda_i^0\right\|\left\|\lambda_j^0\right\|$ 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

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

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 $\left(\frac{a+b+c}{3}\right)^{8}$ $\leq\left(a^{8}+b^{8}+c^{8}\right)/3$, 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

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

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

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

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

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

Asymptotic normality for the sieve estimator

Proof of Lemma .0.10. (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 $\tilde{Z}_{it}=p_{it}+E_{\mathscr{D}}[\bar{p}_{it}]$. We have

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

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

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

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 $Z_{it}=\tilde{Z}_{it}+(\bar{p}_{it}-E_{\mathscr{D}}[\bar{p}_{it}])$, 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

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

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

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

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

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

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

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

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 $E_{\mathscr{D}}(\Psi_{NT,1}) = 0$ and

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

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

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

where we use the fact $E_{\mathscr{D}}(p_{is}^c e_{it}) = E_{\mathscr{D}}(p_{is} e_{it}) = 0$ 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, $E_{\mathscr{D}}\left[\Psi_{NT,22}\right]=0$. By Assumption 5$(iii)$ and conditional Jensen inequality,

$$\begin{aligned} E_{\mathscr{D}}[\Psi_{NT,22}^2] &= \mathrm{Var}_{\mathscr{D}}\left(\Psi_{NT,22}\right)=\frac{1}{NT^3}\sum_{i=1}^{N}\mathrm{Var}_{\mathscr{D}}\left(\sum_{1\leq t<s\leq T}\eta_{ts}E_{\mathscr{D}}\left(e_{it}v_x^{K\prime}p_{is}\right)\right) \\ &\leq \frac{1}{NT^3}\sum_{i=1}^{N}\sum_{1\leq t_1<t_2\leq T}\sum_{1\leq t_3<t_4\leq T}\eta_{t_1t_2}\eta_{t_3t_4}v_x^{K\prime}E_{\mathscr{D}}(e_{it_1}p_{it_2}^{c}e_{it_3}p_{it_4}^{c\prime})v_x^{K} \\ &\equiv E\Psi_{NT,22}. \end{aligned} \tag{.0.69}$$

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 (.0.69) 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 proof of $E\Psi_{NT,22a2}=o_P(1)$ and $E\Psi_{NT,22a3}=o_P(1)$ is similar. For subcase $(a1)$, by the symmetry of $(t_1,t_2)\longleftrightarrow(t_3,t_4)$, we have

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

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 $E\Pi_{\mathscr{D}214,a12}$ 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 $\varphi_{is,q}^{c}\equiv K^{-1/q}\left\|p_{it}^{c}\right\|_{q,\mathscr{D}}$ for $0<q\leq 8+4\delta$. By the conditional

Davydov inequality (see Lemma .0.26 in the supplementary appendix) and Hölder inequality, we have

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

and

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

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

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

For subsubcase $(a12)$, we have

$$
\begin{aligned}
E\Psi_{NT,22a11} &\le \frac{1}{NT^3}\sum_{i=1}^{N}\sum_{1\le t_1<t_2<t_3<t_4\le T,\, d_2>d_1\ge d_3}\Big\{\|f_{t_1}^0\|\,\|f_{t_2}^0\|\,\|f_{t_3}^0\|\,\|f_{t_4}^0\| \\
&\qquad \times\left|E_{\mathscr{D}}\left(e_{it_1}v_x^{K\prime}p_{it_2}^{c}e_{it_3}p_{it_4}^{c\prime}v_x^{K}\right)\right|\Big\} \\
&\quad +\frac{1}{NT^3}\sum_{i=1}^{N}\sum_{1\le t_1<t_2<t_3<t_4\le T,\, d_2>d_3>d_1}\Big\{\|f_{t_1}^0\|\,\|f_{t_2}^0\|\,\|f_{t_3}^0\|\,\|f_{t_4}^0\| \\
&\qquad \times\left|E_{\mathscr{D}}\left(e_{it_1}v_x^{K\prime}p_{it_2}^{c}e_{it_3}p_{it_4}^{c\prime}v_x^{K}\right)\right|\Big\} \\
&\equiv E\Psi_{NT,22a11}(1)+E\Psi_{NT,22a11}(2), \text{ say.}
\end{aligned}
$$

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

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

and

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

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

$$
\begin{aligned}
&E\Psi_{NT,22a11}(1) \\
\le\ & \tfrac{CK}{NT^3}\sum_{i=1}^{N}\sum_{1\le t_1<t_2<t_3<t_4\le T,\, d_2>d_1\ge d_3}\left(C_{1,it_1,e}+C_{1,it_2,p}+C_{1,it_3,e}+C_{1,it_4,p}\right)\alpha_{\mathscr{D}}^{\frac{1+\delta}{2+\delta}}(t_2-t_1) \\
=\ & \tfrac{CK}{NT^3}\sum_{i=1}^{N}\Bigg\{\sum_{t_1=1}^{T-3}C_{1,it_1,e}\sum_{d_2=2}^{T-3-t_1}\sum_{d_1=1}^{d_2-1}\alpha_{\mathscr{D}}^{\frac{1+\delta}{2+\delta}}(d_1)\,d_1+\sum_{t_2=2}^{T-2}C_{1,it_2,p}\sum_{d_2=2}^{T-3-t_2}\sum_{d_1=1}^{d_2-1}\alpha_{\mathscr{D}}^{\frac{1+\delta}{2+\delta}}(d_1)\,d_1 \\
&+\sum_{t_3=3}^{T-1}C_{1,it_3,e}\sum_{d_2=2}^{t_3-1}\sum_{d_1=1}^{d_{2-1}}\alpha_{\mathscr{D}}^{\frac{1+\delta}{2+\delta}}(d_1)\,d_1+\sum_{t_4=4}^{T}C_{1,it_4,p}\sum_{d_2=2}^{t_4-2}\sum_{d_1=1}^{d_2-1}\alpha_{\mathscr{D}}^{\frac{1+\delta}{2+\delta}}(d_1)\,d_1\Bigg\} \\
\le\ & \left\{T\sum_{m=1}^{T-1}\left(1-\frac{m}{T}\right)m\alpha_{\mathscr{D}}^{\frac{1+\delta}{2+\delta}}(m)\right\}\frac{CK}{NT^3}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(C_{1,it,p}+C_{1,it,e}\right) \\
=\ & O_P(T)\frac{CK}{NT^3}O_P(NT)=O_P(K/T)=o_P(1).
\end{aligned}
$$

Similarly, we can show that $E\Psi_{NT,22a11}(2) = O_P(K/T) = o_P(1)$. Consequently $E\Psi_{NT,22a11} = O_P(K/T)$. By the same token we can show that $E\Psi_{NT,22a12} = O_P(K/T)$. 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 $t_2 = t_4$, $(b2)$ $t_1 = t_4$ 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 subcases $(b1)$, wlog we assume $t_1 = t_3$. By the conditional Davydov inequality, we have $\left|E_{\mathscr{D}}(e_{it_1}^2 v_x^{K\prime} p_{it_2}^c p_{it_4}^{c\prime} v_x^K)\right| \leq 8\left\|e_{it_1}^2 v_x^{K\prime} p_{it_2}^c\right\|_{(8+4\delta)/3,\mathscr{D}}\left\|p_{it_4}^{c\prime} v_x^K\right\|_{8+4\delta,\mathscr{D}} \alpha_{\mathscr{D}}^{\frac{1+\delta}{2+\delta}}(t_4 - t_2)$ when $t_4 > t_2$ and $\left|E_{\mathscr{D}}(e_{it_1}^2 v_x^{K\prime} p_{it_2}^c p_{it_4}^{c\prime} v_x^K)\right| \leq 8\left\|e_{it_1}^2 v_x^{K\prime} p_{it_4}^c\right\|_{(8+4\delta)/3,\mathscr{D}}\left\|p_{it_2}^{c\prime} v_x^K\right\|_{8+4\delta,\mathscr{D}}$ $\alpha_{\mathscr{D}}^{\frac{1+\delta}{2+\delta}}(t_2 - t_4)$ when $t_2 > t_4$. If $t_4 > t_2$, by Hölder and Jensen inequalities, each term inside the summation is bounded by

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

Similarly, each term inside the summation is bounded by

$$
2K\left\|v_x^K\right\|^2 \left(2C_{1,it_1,e} + C_{1,it_2,p} + C_{1,it_4,p}\right) \times \alpha_{\mathscr{D}}^{\frac{1+\delta}{2+\delta}}(t_2 - t_4)
$$

when if $t_2 > t_4$. It follows that

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

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

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

Proof of Lemma .0.12. 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

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

Note that $T_{1NT}\leq\|a\|^2\sum_{i=1}^{N}\sum_{1\leq s,t\leq T}\left\|f_t^0\right\|^2E_{\mathscr{D}}\left[\|p_{is}\|^2e_{it}^2\right]=O_P\left(NT^2K\right)$ by Markov inequality. For T_{2NT} and T_{3NT}, following the proof of Proposition .0.7 and by the conditional Davydov and Jensen inequalities we have

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

and

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

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

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

(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

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

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

(vi) By (v), we have $\left\|P_{\lambda^{0}}\mathbf{e}\mathbf{P}_{(a)}'\right\|_{F}\leq\left\|\lambda^{0}\left(\lambda^{0\prime}\lambda^{0}\right)^{-1}\right\|_{F}\left\|\lambda^{0\prime}\mathbf{e}'\mathbf{P}_{(a)}\right\|_{F}=O_{P}\left(N^{-1/2}\right)$

$\times O_P\left((NTK)^{1/2}\delta_{NT}\right) = O_P(\sqrt{NTK})$.

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

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

Then (vii) follows by Chebyshev inequality.

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

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

The first term is bounded by

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

by the conditional Davydov inequality. Similarly, we can show that the second term is also $O_P(K)$. 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 $E_{\mathscr{D}}\{\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}B_t f_s^{0\prime}\left[e_{it}e_{is}-E_{\mathscr{D}}\left(e_{it}e_{is}\right)\right]\}^2$ is bounded by

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

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

(xi) Note that

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

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. ■

Bias correction

Let $\hat{\mathbf{e}}(\beta)\equiv\mathbf{Y}-\sum_{k=1}^{K}\beta_k\mathbf{P}_k-\hat{\lambda}(\beta)\hat{f}(\beta)^{\prime}$. 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 .0.13 and .0.14.

Lemma .0.19 *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 $\hat{\mathbf{e}}_k^{(1)} = M_{\lambda^0}\mathbf{P}_k M_{f^0}$, $\hat{\mathbf{e}}_e^{(1)} = -M_{\lambda^0}\mathbf{u}M_{f^0}\mathbf{u}'\Phi' - \Phi'\mathbf{u}'M_{\lambda^0}\mathbf{u}M_{f^0} - M_{\lambda^0}\mathbf{u}\Phi\mathbf{u}M_{f^0}$,
the expansion coefficients of $M_{\hat{\lambda}}(\beta)$ *are given by*

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

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

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

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 proof of $M_{\hat{f}}(\beta)$ and $\hat{\mathbf{e}}(\beta)$.

Expansion of $M_{\hat{f}}(\beta)$. By Proposition .0.10 (*iii*) and the fact $\mathbf{u} = \varepsilon_0\mathbf{P}_0$, we have

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

Following the proof in Proposition .0.10, we can show that

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

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

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

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

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

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 $\hat{\mathbf{e}}^{(rem)}(\beta) = A_1 + A_2 + A_3$,and $\text{rank}(A_1) \le 2R$, $\text{rank}(A_2) \le 2R$, and $\text{rank}(A_3) \le 3R$. It follows that $\text{rank}(\hat{\mathbf{e}}^{(rem)}(\beta)) \le 7R$. ■

Lemma .0.20 *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 .0.20$(ii)$

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

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 .0.21 *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 rank$(P_{\hat{f}}P_{f^0}) = R$, i.e., 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) Recall that $\hat{\lambda}\hat{f}'\hat{f} = \left(\mathbf{Y} - \sum_{k=1}^{K}\hat{\beta}_k\mathbf{P}_k\right)\hat{f}$. Then

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

First,

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

Noting that

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

we have

$$\begin{aligned}\Lambda_{2NT} &\leq \|\lambda^0\|\|f^{0\prime}f^0/T\|\left[(f^{0\prime}P_{\hat{f}}f^0/T)^{-1}-\left(f^{0\prime}f^0/T\right)^{-1}\right]\left\|(H')^{-1}\right\| \\ &= \sqrt{N}O_P\left(\delta_{NT}^{-1}+K^{-\gamma/d}\right).\end{aligned}$$

Now,

$$\begin{aligned}\|\Lambda_{3NT}\| &\leq \frac{1}{T}\left[\left\|\beta^0-\hat{\beta}\right\|\|\mathbf{P}_{(a)}\|+\|\mathbf{u}\|\right]\left\|P_{\hat{f}}\right\|\|f^0\|\left\|\left(f^{0\prime}P_{\hat{f}}f^0/T\right)^{-1}\right\|\|H^{-1}\| \\ &= O_P\left[\sqrt{N}\left(\left\|\beta^0-\hat{\beta}\right\|+\delta_{NT}^{-1}+K^{-\gamma/d}\right)\right]=O_P\left[\sqrt{N}\left(\delta_{NT}^{-1}+K^{-\gamma/d}\right)\right].\end{aligned}$$

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

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

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

$$
\begin{aligned}
& \sqrt{NT}||\hat{\Phi} - \Phi|| \\
= & \left\| \frac{\hat{\lambda}}{\sqrt{N}} \left(\frac{\hat{\lambda}'\hat{\lambda}}{N} \right)^{-1} \left(\frac{\hat{f}'\hat{f}}{T} \right)^{-1} \frac{\hat{f}'}{\sqrt{T}} - \frac{\lambda^0}{\sqrt{N}} \left(\frac{\lambda^{0\prime}\lambda^0}{N} \right)^{-1} \left(\frac{f^{0\prime}f^0}{T} \right)^{-1} \frac{f^{0\prime}}{\sqrt{T}} \right\| \\
= & \left\| \frac{\hat{\lambda}}{\sqrt{N}} \left(\frac{\hat{\lambda}'\hat{\lambda}}{N} \right)^{-1} \left(\frac{\hat{f}'\hat{f}}{T} \right)^{-1} \frac{\hat{f}'}{\sqrt{T}} - \frac{\lambda^0 (H')^{-1}}{\sqrt{N}} \left(\frac{H^{-1}\lambda^{0\prime}\lambda^0 (H')^{-1}}{N} \right)^{-1} \left(\frac{H' f^{0\prime} f^0 H}{T} \right)^{-1} \frac{H' f^{0\prime}}{\sqrt{T}} \right\| \\
= & \; O_P\left(\delta_{NT}^{-1} + K^{-\gamma/d} \right).
\end{aligned}
$$

■

Lemma .0.22 *Suppose that the conditions in Theorem 3.3.3 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 $B_{2,it} = e_{it}^2 Z_{it}$. Then

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

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 $A_{11,k_1k_2}^{(a)} = \frac{1}{NT}\mathrm{tr}(\mathbf{B}_{1,k_1}\mathbf{B}_{2,k_2}')+$ $\frac{1}{NT}\mathrm{tr}\left(\mathbf{B}_{2,k_1}\mathbf{B}_{1,k_2}' \right)$. Note that $\mathbf{B}_{1,k} = \left(M_{\hat{\lambda}} - M_{\lambda^0} \right)\mathbf{P}_k M_{f^0} +$ $M_{\hat{\lambda}}\mathbf{P}_k\left(M_{\hat{f}} - M_{f^0} \right)$ and

$\left\|\mathbf{B}_{1,k}\right\| = O_P\left(K^{-\gamma/d} + \delta_{NT}^{-1}\right)\left\|\mathbf{P}_k\right\|$. For $\mathbf{B}_{2,k}$, we have

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

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 $\left\|\mathbf{B}_{2,k}\right\| = O_P[(NT)^{1/2}]Z_k^{(4)}$,

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

and

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

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

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

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

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

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

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 .0.13, 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

$$\begin{aligned}
\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\hat{e}_{it}^2-e_{it}^2\right)\hat{Z}_{it}\hat{Z}_{it}' &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\hat{e}_{it}-e_{it}\right)^2\hat{Z}_{it}\hat{Z}_{it}'+\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\hat{e}_{it}-e_{it}\right)e_{it}\hat{Z}_{it}\hat{Z}_{it}' \\
&\equiv A_{21}+2A_{22},\text{ say.}
\end{aligned}$$

For A_{22}, we have $A_{22,k_1k_2}=\frac{1}{NT}\text{tr}(M_{\hat{f}}\tilde{\mathbf{P}}_{k_1}^{(e)\prime}M_{\hat{\lambda}}\mathbf{P}_{k_2}^{(e)})$, 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

$$\left|A_{22,k_1k_2}\right|\leq\frac{1}{NT}\left\|M_{\hat{f}}\tilde{\mathbf{P}}_{k_1}^{(e)\prime}\right\|_F\left\|M_{\hat{\lambda}}\mathbf{P}_{k_2}^{(e)}\right\|_F\leq\frac{1}{NT}\left\|\tilde{\mathbf{P}}_{k_1}^{(e)}\right\|_F\left\|\mathbf{P}_{k_2}^{(e)}\right\|_F,$$

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

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

Now we consider the key term $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\hat{e}_{it}-e_{it}\right)^4$. By Lemma .0.19, we have

$$\hat{e}_{it}-e_{it} = \left(\beta^0-\hat{\beta}\right)'Z_{it}+\overrightarrow{e}_{it}+r_{it}$$

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

$$\begin{aligned}
\left\|\hat{\mathbf{e}}_e^{(1)}\right\|_F^2 &\le R\left\|\hat{\mathbf{e}}_e^{(1)}\right\|^2 = O_P\left(NT\left(\delta_{NT}^{-4}+K^{-4\gamma/d}\right)\right), && (.0.70)\\
\left\|\hat{\mathbf{e}}^{(rem)}\right\|_F^2 &\le O_P(NT\|\hat{\beta}-\beta^0\|^2\left(\delta_{NT}^{-2}+K^{-2\gamma/d}\right)), && (.0.71)\\
\left\|M_{f^0}\mathbf{e}_g'M_{\lambda^0}\right\|_F^2 &= O_P\left(NTK^{-2\gamma/d}\right), && (.0.72)
\end{aligned}$$

by Lemma .0.19, where we use the fact that $\text{rank}\left(\hat{\mathbf{e}}^{(rem)}(\beta)\right) \le 7R$, $\left\|\hat{\beta}-\beta^0\right\| = o_P\left(\delta_{NT}^{-1}+K^{-\gamma/d}\right)$ and $\delta_{NT}^{-2}+K^{-2\gamma/d} = o_P(\|\hat{\beta}-\beta^0\|)$ in the second line. Then

$$\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\hat{e}_{it}-e_{it}\right)^4 \\
\le\ & 9\left(\left\|\beta^0-\hat{\beta}\right\|^4\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\|Z_{it}\|^4+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\overrightarrow{e}_{it}^{\,4}+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}r_{it}^4\right). \qquad (.0.73)
\end{aligned}$$

It is easy to see that the first term in (.0.73) 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 $\mathscr{D}$, 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 $\mathscr{D}$. For the third term in (.0.73), 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 (.0.70)-(.0.72). In sum, we have

$$
\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\hat{e}_{it}-e_{it}\right)^4=O_P\left(\delta_{NT}^{-4}+NT\delta_{NT}^{-8}+NTK^{-4\gamma/d}\right). \tag{.0.74}
$$

It follows that

$$
\|A_{21}\|_F=O_P\left(K\delta_{NT}^{-2}+(NT)^{1/2}K\delta_{NT}^{-4}+(NT)^{1/2}K^{1-2\gamma/d}\right)
$$

and

$$
\|A_{22}\|_F=O_P\left(K\delta_{NT}^{-1}+(NT)^{1/4}K\delta_{NT}^{-2}+(NT)^{1/4}K^{1-\gamma/d}\right).
$$

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

Lemma .0.23 *Suppose that the conditions in Theorem 3.3.3 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 .0.19. By Lemma .0.12(ii), $\left\|P_{\lambda^0}\mathbf{e}P_{f^0}\right\|_F = O_P(1)$. By Chebyshev inequality, one can readily show that $\|P_{\lambda^0}\mathbf{e}\|_F = O_P\left(T^{1/2}\right)$ and $\left\|\mathbf{e}P_{f^0}\right\|_F = O_P\left(N^{1/2}\right)$. By (.0.70)-(.0.72), we have $\left\|\hat{\mathbf{e}}_e^{(1)}\right\|_F \leq O_P[\sqrt{NT}\left(\delta_{NT}^{-2}+K^{-2\gamma/d}\right)]$, $\left\|\hat{\mathbf{e}}^{(rem)}\right\|_F \leq O_P[\sqrt{NT}\left\|\hat{\beta}-\beta^0\right\|\left(\delta_{NT}^{-1}+K^{-\gamma/d}\right)]$, 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

$$\frac{1}{NT}\left\|M_{\lambda^0}\mathbf{P}_{(a)}M_{f^0}\right\|_F^2 \leq \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}Z_{it,(a)}^2 = a'W_{NT}a \leq \mu_1\left(W_{NT}\right)\|a\|^2 = 1,$$

we have $\left\|\hat{\beta}-\beta^0\right\|\left\|M_{\lambda^0}\mathbf{P}_{(a)}M_{f^0}\right\|_F = O_P\left(\sqrt{K}\delta_{NT}^{-2}+K^{-\gamma/d}\right)O_P(\sqrt{NT}) = o_P(\sqrt{N}+\sqrt{T})$ by (.0.50). Consequently, $||\hat{\mathbf{e}}-\mathbf{e}||_F = O_P(\sqrt{N}+\sqrt{T})$. ■

Lemma .0.24 *Suppose that the conditions in Theorem 3.3.3 hold. Then we have*
(i) $N^{-1}||E_{\mathscr{D}}[\mathbf{e}'M_{\lambda^0}\mathbf{e}]-(\hat{\mathbf{e}}'\hat{\mathbf{e}})^{truncD}|| = o_P[T^{5/8}(K^{-\gamma/d}+\sqrt{K}\delta_{NT}^{-2})+T^{-1/4}]$;
(ii) $T^{-1}||E_{\mathscr{D}}[\mathbf{e}M_{f^0}\mathbf{e}']-\left(\hat{\mathbf{e}}\hat{\mathbf{e}}'\right)^{truncD}|| = o_P[N^{5/8}(K^{-\gamma/d}+\sqrt{K}\delta_{NT}^{-2})+N^{-1/4}]$.

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

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

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

$$\begin{aligned}&\frac{1}{N}\left\|E_{\mathscr{D}}\left(\mathbf{e}'M_{\lambda^0}\mathbf{e}\right)-\left(\hat{\mathbf{e}}'\hat{\mathbf{e}}\right)^{\text{truncD}}\right\| \\ \leq\ &\frac{1}{N}\left\|\left[\mathbf{e}'M_{\lambda^0}\mathbf{e}-E_{\mathscr{D}}\left(\mathbf{e}'M_{\lambda^0}\mathbf{e}\right)\right]^{\text{truncD}}\right\| + \frac{1}{N}\left\|\left[\mathbf{e}'M_{\lambda^0}\mathbf{e}-\hat{\mathbf{e}}'\hat{\mathbf{e}}\right]^{\text{truncD}}\right\| \qquad (.0.75)\end{aligned}$$

For the first term in (.0.75), noting the tth diagonal element of $\mathbf{e}'M_{\lambda^0}\mathbf{e}-E_{\mathscr{D}}\left(\mathbf{e}'M_{\lambda^0}\mathbf{e}\right)$ is given by $\left[e_{it}^2-E_{\mathscr{D}}\left(e_{it}^2\right)\right]-\frac{2}{N}\sum_{j=1}^N\alpha_{ij}\left[e_{jt}e_{it}-E_{\mathscr{D}}\left(e_{jt}e_{it}\right)\right]+\frac{1}{N^2}\sum_{j_1=1}^N\sum_{j_2=1}^N\{\alpha_{ij_1}\alpha_{ij_2}$ $\times\left[e_{j_1t}e_{j_2t}-E_{\mathscr{D}}\left(e_{j_1t}e_{j_2t}\right)\right]\}$, we have

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

Noting that $E|N^{-1/2}\sum_{i=1}^N\left[e_{it}^2-E_{\mathscr{D}}\left(e_{it}^2\right)\right]|^4<\infty$, we have $\max_{1\leq t\leq T}C_{1t}=o_P\left(N^{-1/2}T^{1/4}\right)$ by Lemma .0.27. For the second term, we have

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

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

$$
\frac{1}{N}\left\|\left[\mathbf{e}'M_{\lambda^0}\mathbf{e}-E_{\mathscr{D}}\left(\mathbf{e}'M_{\lambda^0}\mathbf{e}\right)\right]^{\text{truncD}}\right\|=o_P\left(N^{-1/2}T^{1/4}\right). \qquad (.0.76)
$$

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

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

For the second term in (.0.75), we have

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

Let c_{tt} be the (t,t)th element of $N^{-1}P_{f^0}\mathbf{e}'M_{\lambda^0}\mathbf{e}P_{f^0}$. 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

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

because $E\left\|f_t^0\right\|^8<\infty$. Similar, we can show that $\max_{1\le t\le T}\left|c_{tt,2}\right| = o_P[(NT)^{1/4}]$. By Lemma .0.28$(viii)$,

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

Similarly, we have

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

where the first term comes from Assumption 6(i) and Lemma .0.27, and the second term comes from

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

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

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

Note that

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

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

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

■

Lemma .0.25 *Suppose that the conditions in Theorem 3.3.3 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 .0.28(iv), (vii), and .0.23, we have

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

■

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

Proof of Lemma .0.13. (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

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

we have

$$
\begin{aligned}
& \left\|\hat{W}_{NT}-W_{NT}\right\|_F \\
= & \left[\sum_{k_1=1}^{K}\sum_{k_2=1}^{K}\left(\hat{W}_{NT,k_1k_2}-W_{NT,k_1k_2}\right)^2\right]^{1/2} \\
\leq & \; 2R\left(\left\|M_{\hat{\lambda}}-M_{\lambda^0}\right\|+\left\|M_{\hat{f}}-M_{f^0}\right\|\right)\left\{\sum_{k_1=1}^{K}\sum_{k_2=1}^{K}\left[\frac{1}{NT}\left\|\mathbf{P}_{k_1}\right\|\left\|\mathbf{P}_{k_2}\right\|\right]^2\right\}^{1/2} \\
= & \; O_P\left(K\left(\delta_{NT}^{-1}+K^{-\gamma/d}\right)\right) \text{ by Lemma .0.20.}
\end{aligned}
$$

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

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

By Lemmas .0.22(i)-(ii), we have $\left\|D\Omega_{NT,1}+D\Omega_{NT,2}\right\|_F=O_P(K\delta_{NT}^{-1}+(NT)^{1/4}K\delta_{NT}^{-2}+(NT)^{1/4}K^{1-\gamma/d})$. Following the study of $\left\|\tilde{W}_{NT}-W_{NT}\right\|_F$, we can show that $\left\|D\Omega_{NT,3}\right\|_F=O_P(K/\sqrt{NT})$. It follows that $\left\|\hat{\Omega}_{NT}-\Omega_{NT}\right\|_F=O_P((NT)^{1/4}K\delta_{NT}^{-2}+(NT)^{1/4}K^{1-\gamma/d}+K\delta_{NT}^{-1})$.

(iii) By Minkowski inequality

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

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 .0.14. (i) Note that b_1 can be rewritten as follows

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

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

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

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

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

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

$$\begin{aligned}
& E_{\mathscr{D}}\left[\|Db_2\|^2\right] \\
= & \frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{1\le t_1<s_1\le\min(t_1+M_T,T)}\sum_{1\le t_2<s_2\le\min(t_2+M_T,T)}\eta_{t_1s_1}\eta_{t_2s_2}E_{\mathscr{D}}\left(\zeta_{i,t_1s_1}'\zeta_{i,t_2s_2}\right).
\end{aligned}$$

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 $E_{\mathscr{D}}[\|Db_2\|^2]$ when the summation is restricted to the time indices in these two cases, respectively. Then $E_{\mathscr{D}}[\|Db_2\|^2] = EDb_{21a} + EDb_{21b}$. 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

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

It follows that

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

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 .0.11, we can show that

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

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

For Db_1, we have

$$\begin{aligned}
Db_1 &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{1\leq t<s\leq\min(t+M_T,T)}\Big\{\left(\hat{\eta}_{ts}-\eta_{ts}\right)p_{is}e_{it}+\eta_{ts}p_{is}\left(\hat{e}_{it}-e_{it}\right) \\
&\qquad +\left(\hat{\eta}_{ts}-\eta_{ts}\right)p_{is}\left(\hat{e}_{it}-e_{it}\right)\Big\} \\
&\equiv Db_{11}+Db_{12}+Db_{13},\text{ say.}
\end{aligned}$$

For Db_{11}, we have by Cauchy-Schwarz inequality and Lemma .0.20(ii),

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

Similarly, by Cauchy-Schwarz inequality and Lemmas .0.23 and .0.20(ii) , we have

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

and

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

Consequently, $\|Db_1\| = O_P(M_T\sqrt{K}\delta_{NT}^{-1})$ and

$$\begin{aligned}
\left\|\hat{b}_1 - b_1^{(2)}\right\| &\leq \|Db_1\| + \|Db_2\| \\
&= O_P\left(M_T\sqrt{K}\delta_{NT}^{-1}\right) + O_P\left(\sqrt{KM_T^3/(NT)}\right) = O_P\left(M_T\sqrt{K}\delta_{NT}^{-1}\right).
\end{aligned}$$

This completes the proof of (i).

(ii) Recall that $b_{2,k}=T^{-1}\mathrm{tr}[E_{\mathscr{D}}(\mathbf{ee}')M_{\lambda^0}\mathbf{P}_k\Phi]$ and $\hat{b}_{2,k}=T^{-1}\mathrm{tr}\left[(\hat{\mathbf{e}}\hat{\mathbf{e}}')^{\mathrm{truncD}}M_{\hat{\lambda}}\mathbf{P}_k\hat{\Phi}\right]$. Then by Lemmas .0.19, .0.20, .0.24, and .0.14, we have

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

$$\begin{aligned}
\left\|E_{\mathscr{D}}\left(\mathbf{e}P_{f^0}\mathbf{e}'\right)^{\mathrm{truncD}}\right\| &\leq \max_{1\leq i\leq N}\left|\frac{1}{T}\sum_{t=1}^{T}E_D\left[\frac{1}{T}\sum_{s=1}^{T}\eta_{ts}e_{is}\right]^2\right| \\
&= \frac{1}{T}\max_{1\leq i\leq N}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\eta_{ts}^2E_D\left(e_{is}^2\right)=O_P\left(T^{-1}N^{1/2}\right)
\end{aligned}$$

because $E|T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}\eta_{ts}^2E_D\left(e_{is}^2\right)|^2<\infty$. It follow that

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

(iii) The proof is analogous to that of (ii) by using Lemmas .0.20, .0.21, .0.13, and .0.14. ■

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 $\Upsilon_{NT,k} \equiv \frac{1}{NT}\mathrm{tr}\left(M_{\lambda^0}\mathbf{X}_k M_{f^0}\Delta\right)$ and D_{NT} is defined in (3.4.4). Let $C^{(1)}_{l,NT}$, and $C^{(2)}_{l,NT}$ be $d \times 1$ vectors whose kth elements are respectively given by

$$C^{(1)}_{l,NT,k} \equiv \frac{1}{NT}\mathrm{tr}\left(M_{\lambda^0}\mathbf{X}_k M_{f^0}\varepsilon'\right), \tag{.0.77}$$

$$C^{(2)}_{l,NT,k} \equiv -\frac{1}{NT}\mathrm{tr}\left(\mathbf{X}_k\Phi'\varepsilon M_{f^0}\varepsilon' M_{\lambda^0} + \mathbf{X}_k M_{f^0}\varepsilon' M_{\lambda^0}\varepsilon\Phi' + \mathbf{X}_k M_{f^0}\varepsilon'\Phi\varepsilon M_{\lambda^0}\right) \tag{.0.78}$$

$$\equiv C^{(2,a)}_{l,NT,k} + C^{(2,b)}_{l,NT,k} + C^{(2,b)}_{l,NT,k}, \text{ say,} \tag{.0.79}$$

where ε 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 (2012), we can show that under $\mathbb{H}_1(\gamma_{NT})$ with $\gamma_{NT} = O(K^{1/4}/\sqrt{NT})$

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

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 3.3.2 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 (.0.53), $D = E_{\mathscr{D}}[D_{NT}]$, 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 $B^{(l)} = o_P\left(K^{1/4}\right)$. Under Assumption 12, we have $B^{(l)} = O_P\{\max\left(\kappa_{NT}, \kappa_{NT}^{-1}\right)\} = o_P\left(K^{1/4}\right)$. 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 $\left(\lambda^0, f^0, \mathbf{e}\right)$ 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 $\hat{D}_{NT,k_1k_2} \equiv \frac{1}{NT}\operatorname{tr}(M_{\hat{\lambda}^{(l)}}\mathbf{X}_{k_1}M_{\hat{f}^{(l)}}\mathbf{X}_{k_2}')$. 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 .0.15. The proof is similar to that of Lemma .0.13. ■

Proof of Lemma .0.16. 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 (3.3.16). By (.0.47) and (**??**)-(3.3.4),

$$\hat{\beta} - \beta^0 = W_{NT}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} Z_{it}u_{it} + W_{NT}^{-1}\left[C_{NT}^{(2,a)} + C_{NT}^{(2,b)} + C_{NT}^{(2,c)}\right] + R_{NT}.$$

Decompose $\hat{\beta}_{bc} - \beta^0$ as follows

$$\begin{aligned}
\hat{\beta}_{bc} - \beta^0 \quad = \quad & \left\{W_{NT}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} Z_{it}e_{it} + \frac{1}{T}\hat{W}_{NT}^{-1}\hat{b}_1\right\} + \left\{W_{NT}^{-1}C_{NT}^{(2,a)} + \frac{1}{N}\hat{W}_{NT}^{-1}\hat{b}_2\right\} \\
& + \left\{W_{NT}^{-1}C_{NT}^{(2,b)} + \frac{1}{T}\hat{W}_{NT}^{-1}\hat{b}_3\right\} + \left\{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} W_{NT}^{-1}Z_{it}e_{g,it} + W_{NT}^{-1}C_{NT}^{(2,c)} + R_{NT}\right\} \\
\equiv \quad & \mathscr{B}_{NT1} + \mathscr{B}_{NT2} + \mathscr{B}_{NT3} + \mathscr{B}_{NT4}, \text{ say.}
\end{aligned}$$

We complete the proof by showing that (i) $\mathscr{B}_{NT1} = \tilde{W}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{Z}_{it}e_{it} + o_P(\gamma_{NT})$, and (ii) $\mathscr{B}_{NTs} = o_P(\gamma_{NT})$ for $s = 2,3,4$. We first study $\mathscr{B}_{NT1}$. Note that

$$\begin{aligned}
& \mathscr{B}_{NT1} - \tilde{W}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{Z}_{it}e_{it} \\
= \quad & W_{NT}^{-1}\left(\tilde{W} - W_{NT}\right)\tilde{W}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{Z}_{it}e_{it} + \left\{W_{NT}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(Z_{it} - \tilde{Z}_{it}\right)e_{it} + \frac{1}{T}\hat{W}_{NT}^{-1}\hat{b}_1\right\} \\
\equiv \quad & \mathscr{B}_{NT11} + \mathscr{B}_{NT12}, \text{ say.}
\end{aligned}$$

By Lemma .0.28(iii) and Assumption 7, we have

$$\begin{aligned}\|\mathscr{B}_{NT11}\| &= \left\|W_{NT}^{-1}\right\| \left\|\tilde{W}-W_{NT}\right\|_F \left\|\tilde{W}^{-1}\right\| \left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{Z}_{it}'e_{it}\right\| \\ &= O_P\left(\frac{K}{\sqrt{NT}}\sqrt{\frac{K}{NT}}\right) = o_P(\gamma_{NT}).\end{aligned}$$

For $\mathscr{B}_{NT12}$, we have

$$\begin{aligned}\mathscr{B}_{NT12} &= W_{NT}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\left(Z_{it}-\tilde{Z}_{it}\right)e_{it}-E_{\mathscr{D}}\left[\left(Z_{it}-\tilde{Z}_{it}\right)e_{it}\right]\right\} \\ &\quad+\left\{W_{NT}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}E_{\mathscr{D}}\left(Z_{it}e_{it}\right)+\frac{1}{T}\hat{W}_{NT}^{-1}\hat{b}_1\right\} \equiv \mathscr{B}_{NT12a}+\mathscr{B}_{NT12b}, \text{ say.}\end{aligned}$$

Following the proof of Lemma .0.11, we can readily show that $\mathscr{B}_{NT12a}=O_P\left(\sqrt{\frac{K}{NT}}\delta_{NT}^{-1}\right)$ $=o_P(\gamma_{NT})$. By Lemmas .0.13, .0.14, and (.0.77), we have

$$\begin{aligned}\mathscr{B}_{NT12b} &= \frac{1}{T}\left(\hat{W}_{NT}^{-1}-W_{NT}^{-1}\right)b_1+\frac{1}{T}\hat{W}_{NT}^{-1}\left(\hat{b}_1-b_1\right) \\ &\quad+W_{NT}^{-1}\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{1\le t<s\le T}\alpha_{ii}\eta_{ts}E_{\mathscr{D}}\left(p_{is}e_{it}\right) \\ &= \frac{1}{T}O_P\left(K^{3/2}\left(\delta_{NT}^{-1}+K^{-\gamma/d}\right)\right) \\ &\quad+\frac{1}{T}O_P\left(\sqrt{K}\sum_{\tau=M_T}^{T}\alpha_{\mathscr{D}}^{\frac{3+2\delta}{4+2\delta}}(\tau)+M_T\sqrt{K}\delta_{NT}^{-1}\right)+O_P\left(\frac{\sqrt{K}}{NT}\right) \\ &= o_P(\gamma_{NT})\end{aligned}$$

under Assumption 12. Consequently, $\mathscr{B}_{NT12}=o_P(\gamma_{NT})$ and (i) follows.

For $\mathscr{B}_{NT2}$, we decompose it as follows:

$$\mathscr{B}_{NT2}=\frac{1}{N}(\hat{W}_{NT}^{-1}\hat{b}_2-W_{NT}^{-1}b_2)+W_{NT}^{-1}[C_{NT}^{(2,a)}-\frac{1}{N}b_2]\equiv\mathscr{B}_{NT2a}+\mathscr{B}_{NT2b}, \text{say.}$$

As in the study of $\mathscr{B}_{NT12b}$,

$$\begin{aligned}\|\mathscr{B}_{NT2a}\| &\le \frac{1}{N}\left\|\hat{W}_{NT}^{-1}-W_{NT}^{-1}\right\|\|b_2\|+\frac{1}{N}\left\|\hat{W}_{NT}^{-1}\right\|\left\|\hat{b}_2-b_2\right\| \\ &= \frac{1}{N}O_P\left[K\left(K^{-\gamma/d}+\delta_{NT}^{-1}\right)\right]+\frac{1}{N}o_P\left(\kappa_{NT}K^{1/4}\right)=o_P(\gamma_{NT})\end{aligned}$$

by Lemmas .0.13 and .0.14, and Assumption 12. For $\mathscr{B}_{NT2b}$, recall that

$$\begin{aligned}
C_{NT,k}^{(2,a)}+\frac{1}{N}b_{2,k} &= -\frac{1}{NT}\operatorname{tr}\left(\mathbf{u}\mathbf{u}'M_{\lambda^0}\mathbf{P}_{(k)}\Phi\right)+\frac{1}{N}b_{2,k}+\frac{1}{NT}\operatorname{tr}\left(\mathbf{u}P_{f^0}\mathbf{u}'M_{\lambda^0}\mathbf{P}_{(k)}\Phi\right)\\
&= -\frac{1}{NT}\operatorname{tr}\left\{\left[\mathbf{e}\mathbf{e}'-E_{\mathscr{D}}\left(\mathbf{e}\mathbf{e}'\right)\right]M_{\lambda^0}\mathbf{P}_{(k)}\Phi\right\}-\frac{1}{NT}\operatorname{tr}\left(\mathbf{e}_g\mathbf{e}_g'M_{\lambda^0}\mathbf{P}_{(k)}\Phi\right)\\
&\quad +\frac{1}{NT}\operatorname{tr}\left(\mathbf{e}\mathbf{e}_g'M_{\lambda^0}\mathbf{P}_{(k)}\Phi\right)+\frac{1}{NT}\operatorname{tr}\left(\mathbf{e}_g\mathbf{e}'M_{\lambda^0}\mathbf{P}_{(k)}\Phi\right)\\
&\quad +\frac{1}{NT}\operatorname{tr}\left(\mathbf{u}P_{f^0}\mathbf{u}'M_{\lambda^0}\mathbf{P}_{(k)}\Phi\right)\\
&\equiv -C_{2a1,k}-C_{2a2,k}+C_{2a3,k}+C_{2a4,k}+C_{2a5,k}, \text{ say.}
\end{aligned}$$

Denote C_{2as} as a $K\times 1$ vector whose kth element is $C_{2as,k}$, for $s=1,\ldots,5$. Following the study of $\Pi_{2NT,1}$ in Proposition .0.8 we have $\|\mathscr{B}_{NT2b}\|\leq\left\|W_{NT}^{-1}\right\|\left\|C_{NT}^{(2,a)}-\frac{1}{N}b_2\right\|\leq$ $\left\|W_{NT}^{-1}\right\|\sum_{s=1}^{5}\|C_{2as}\|=O_P\{\sqrt{\frac{K}{NT}}\left(\delta_{NT}^{-1}+K^{-\gamma/d}\right)\}=o_P\left(\gamma_{NT}\right)$. It follows that $\|\mathscr{B}_{NT2}\|$ $=o_P\left(\gamma_{NT}\right)$. Analogously, we can show that $\|\mathscr{B}_{NT3}\|=o_P\left(\gamma_{NT}\right)$. $\mathscr{B}_{NT12}=o_P\left(\gamma_{NT}\right)$.

Now we consider $\mathscr{B}_{NT4}$. Following the study of $\Pi_{2NT,3}$ in Theorem 3.3.2 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 $R_{NT}=O_P(\|r_{NT}\|\varepsilon_0^{1/2})$, we have $\mathscr{B}_{NT4}=\sqrt{\frac{K}{NT}}O_P\left(\delta_{NT}^{-1}+K^{-\gamma/d}\right)+$ $O_P\left(K^{-\gamma/d}\right)+O_P[(\sqrt{K}\delta_{NT}^{-2}+K^{-\gamma/d})(\delta_{NT}^{-1/2}+K^{-\gamma/2d})]=o_P\left(\gamma_{NT}\right)$. ■

Proof of Lemma .0.17. Let $\varepsilon\equiv\mathbf{e}+\gamma_{NT}\Delta$ and $\tilde{\varepsilon}_0\equiv\|\varepsilon\|/\sqrt{NT}\leq(\|\mathbf{e}\|+\gamma_{NT}\|\Delta\|)/\sqrt{NT}$ $=O_P\left(\delta_{NT}^{-1}+\gamma_{NT}\right)$. 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 (.0.77)-(.0.79). Noting that

$$\begin{aligned}
C_{l,NT,k}^{(1)} &= \frac{1}{NT}\operatorname{tr}\left(M_{f^0}\mathbf{e}'M_{\lambda^0}\mathbf{X}_k\right)+\gamma_{NT}\frac{1}{NT}\operatorname{tr}\left(M_{f^0}\Delta'M_{\lambda^0}\mathbf{X}_k\right)\\
&= O_P\left(T^{-1}+(NT)^{-1/2}+\gamma_{NT}\right)
\end{aligned}$$

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 .0.10 and following the proof of Theorem 3.3.1, 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 $\tilde{R}_{NT} = O_P[(\|\tilde{r}_{NT}\|^2 \tilde{\varepsilon}_0 + \|\tilde{r}_{NT}\| \tilde{\varepsilon}_0^3 + \|\tilde{r}_{NT}\|^3)^{1/2}] = O_P(\|\tilde{r}_{NT}\| \tilde{\varepsilon}_0^{1/2})$; see Su, Jin, and Zhang (2012) for details. Following the proof of Lemma .0.16, with some minor modifications[2] 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

$$\begin{aligned} R_{\theta,NT} &\equiv \left(D_{NT}^{-1} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \bar{X}_{it} e_{it} - D^{-1} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{X}_{it} e_{it} + \frac{1}{T} D^{-1} b_1^{(l)} \right) \\ &\quad + \left(D_{NT}^{-1} C_{l,NT}^{(2,a)} + \frac{1}{N} D^{-1} b_2^{(l)} \right) + \left(D_{NT}^{-1} C_{l,NT}^{(2,b)} + \frac{1}{T} D^{-1} b_3^{(l)} \right) \\ &\quad + D_{NT}^{-1} C_{l,NT}^{(2,c)} + \tilde{R}_{NT} \\ &\equiv R_{\theta,NT}^{(1)} + R_{\theta,NT}^{(2)} + R_{\theta,NT}^{(3)} + D_{NT}^{-1} C_{l,NT}^{(2,c)} + \tilde{R}_{NT}, \text{ say.} \end{aligned}$$

Clearly, $\tilde{R}_{NT} = O_P(\|\tilde{r}_{NT}\| \tilde{\varepsilon}_0^{1/2}) = O_P[(\delta_{NT}^{-2} + \gamma_{NT})(\delta_{NT}^{-1/2} + \gamma_{NT}^{1/2})] = o_P(\gamma_{NT})$. Following the study of $\Pi_{2NT,3}$ in Proposition .0.8 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

$$\begin{aligned} R_{\theta,NT}^{(1)} &= D_{NT}^{-1} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\{ \left(\bar{X}_{it} - \tilde{X}_{it} \right) e_{it} - E_{\mathscr{D}} \left[\left(\bar{X}_{it} - \tilde{X}_{it} \right) e_{it} \right] \right\} \\ &\quad + \left\{ D_{NT}^{-1} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} E_{\mathscr{D}}(\bar{X}_{it} e_{it}) + \frac{1}{T} D_{NT}^{-1} b_1^{(l)} \right\} \\ &\quad + \frac{1}{T} \left(D^{-1} - D_{NT}^{-1} \right) b_1^{(l)} + \left(D_{NT}^{-1} - D^{-1} \right) \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{X}_{it} e_{it} \\ &\equiv R_{\theta,NT}^{(1,a)} + R_{\theta,NT}^{(1,b)} + R_{\theta,NT}^{(1,c)} + R_{\theta,NT}^{(1,d)}, \text{ say.} \end{aligned}$$

Following the proof of Lemma .0.11, we have $R_{\theta,NT}^{(1,a)} = O_P(\delta_{NT}^{-1}/\sqrt{NT})$. Analogously to the proof of (ib) in Proposition .0.7, $R_{\theta,NT}^{(1,b)} = O_P((NT)^{-1})$. By Lemma .0.15 (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

[2]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.

$R^{(1)}_{\theta,NT} = O_P(\delta^{-1}_{NT}/\sqrt{NT}) = o_P(\gamma_{NT})$. For $R^{(2)}_{\theta,NT}$, we have

$$R^{(2)}_{\theta,NT} = D^{-1}_{NT}\left(C^{(2,a)}_{l,NT} + \frac{1}{N}b^{(l)}_2\right) + \frac{1}{N}\left(D^{-1} - D^{-1}_{NT}\right)b^{(l)}_2 \equiv R^{(2,a)}_{\theta,NT} + R^{(2,b)}_{\theta,NT}, \text{ say.}$$

It is easy to show that $R^{(2,b)}_{\theta,NT} = O_P\left(T^{-1/2}N^{-3/2}\right)$ by .0.15 (iii) and the fact that $b^{(l)}_2 = O_P(1)$. Following the proof of (i) in Proposition .0.8, 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 3.4.4. Let P^* denote the probability measure induced by the wild bootstrap conditional on the original sample $\mathscr{W}_{NT} \equiv \{(X_{it}, Y_{it}) : i = 1,...,N, t = 1,...,T\}$. 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 $\varepsilon > 0$, $P^*(\|b_{NT}\| > \varepsilon) = o_P(1)$. 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 $\mathscr{W}_{NT}$, 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 $\mathscr{F}^*_t$ be the σ-field generated by $\{e^*_{it},...,e^*_{i1}\}^N_{i=1}$. For each i, $\{e^*_{it}, \mathscr{F}^*_t\}$ is an m.d.s. such that $E^*\left(e^*_{it}|\mathscr{F}^*_{t-1}\right) = \hat{e}^{(l)}_{it}E(v_{it}) = 0$ and $E^*[(e^*_{it})^2|\mathscr{F}^*_{t-1}] = [\hat{e}^{(l)}_{it}]^2 E\left(v^2_{it}\right) = [\hat{e}^{(l)}_{it}]^2$. These observations greatly simplify the proofs in the bootstrap world. In particular, we can show that: (i) $\hat{\beta}^*_{bc} - \beta^{0*} = \tilde{W}^{-1}\frac{1}{NT}\sum^N_{i=1}\sum^T_{t=1}\tilde{Z}_{it}e^*_{it} + R^*_{\beta,NT}$, where $\left\|R^*_{\beta,NT}\right\| = o_{P^*}(\frac{K^{1/4}}{\sqrt{NT}})$ 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) $\hat{\theta}^* - \theta^0 = D^{-1}\frac{1}{NT}\sum^N_{i=1}\sum^T_{t=1}\tilde{X}'_{it}e^*_{it} + B^*_{\theta,NT} + R^*_{\theta,NT}$, 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

$$\mathbb{B}_{NT}^{*} \equiv \mathrm{tr}(\tilde{W}^{-1}Q_{wpp}\tilde{W}^{-1}\tilde{\Omega}^{*}) \text{ and } \mathbb{V}_{NT}^{*} \equiv 2\mathrm{tr}(\tilde{W}^{-1}Q_{wpp}\tilde{W}^{-1}\tilde{\Omega}^{*}\tilde{W}^{-1}Q_{wpp}\tilde{W}^{-1}\tilde{\Omega}^{*}),$$

where $\tilde{\Omega}^{*} \equiv \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}E^{*}(\tilde{Z}_{it}\tilde{Z}_{it}' e_{it}^{*2}) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\{\tilde{Z}_{it}\tilde{Z}_{it}' \times [\hat{e}_{it}^{(l)}]^{2}\}$. Following the proof of Theorem 3.4.2, 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 $\mathbb{H}_{0}$. 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 3.4.1, 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) $\tilde{J}_{NT}^{*} \equiv \left(NT\Gamma_{NT1}^{*} - \mathbb{B}_{NT}^{*}\right)/\sqrt{\mathbb{V}_{NT}^{*}} \xrightarrow{d^{*}} N(0,1)$, (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 3.4.1 and 3.4.2 to show (ii)-(iv). Analogously to the proof of Proposition .0.9, we can show that $\tilde{J}_{NT}^{*} = \sum_{1\leq i\neq j\leq N}W_{ij}^{*} + o_{P^{*}}(1)$, where $W_{ij}^{*} \equiv W_{NT}^{*}(u_{i}^{*}, u_{j}^{*}) \equiv \frac{1}{NT\mathbb{V}_{NT}^{*}}\sum_{1\leq t,s\leq T} e_{it}^{*}H_{ij,ts}e_{js}^{*}$, $u_{i}^{*} \equiv (\tilde{Z}_{i}, e_{i}^{*})'$, 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 $(i1)$ $\mathrm{Var}^{*}(\tilde{J}_{NT}^{*}) = 1 + o_{P^{*}}(1)$, $(i2)$ $G_{I}^{*} \equiv \sum_{1\leq i<j<N}E^{*}[(W_{ij}^{*})^{4}] = o_{P^{*}}(1)$, $(i3)$ $G_{II}^{*} \equiv \sum_{1\leq i<j<l\leq N}E^{*}(W_{il}^{*2}W_{jl}^{*2} + W_{ij}^{*2}W_{il}^{*2} + W_{ij}^{*2}W_{lj}^{*2}) = o_{P^{*}}(1)$, and $(i4)$ $G_{III}^{*} \equiv \sum_{1\leq i<j<r<l\leq N}E^{*}(W_{ij}^{*}W_{ir}^{*}W_{lj}^{*}W_{lr}^{*} + W_{ij}^{*}W_{il}^{*}W_{rj}^{*}W_{rl}^{*} + W_{ir}^{*}W_{il}^{*} \times W_{jr}^{*}W_{jl}^{*}) = o_{P^{*}}(1)$. Note that v_{it} is IID across i and t, $E^{*}[(e_{it}^{*})] = 0$, $E^{*}[(e_{it}^{*})^{2}] = [\hat{e}_{it}^{(l)}]^{2}$, and $E^{*}[(e_{it}^{*})^{4}] = 3[\hat{e}_{it}^{(l)}]^{4}$.

For $(i1)$, using the IID property of $\{v_{it}\}$, we can readily show that

$$\begin{aligned}
&\mathrm{Var}^*(\tilde{J}^*_{NT}) \\
= \;& \tfrac{4}{N^2T^2\mathbb{V}^*_{NT}} \sum_{1\le i<j\le N} \sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\sum_{s_1=1}^{T}\sum_{s_2=1}^{T} H_{ij,t_1s_1}H_{ij,t_2s_2}\hat{e}^{(l)}_{it_1}\hat{e}^{(l)}_{js_1}\hat{e}^{(l)}_{it_2}\hat{e}^{(l)}_{js_2}E^*\left(v_{it_1}v_{js_1}v_{it_2}v_{js_2}\right) \\
= \;& \tfrac{4}{N^2T^2\mathbb{V}^*_{NT}} \sum_{1\le i<j\le N} \sum_{t=1}^{T}\sum_{s=1}^{T} H^2_{ij,ts}[\hat{e}^{(l)}_{it}]^2[\hat{e}^{(l)}_{js}]^2 \\
= \;& 1-\frac{2}{N^2T^2\mathbb{V}^*_{NT}} \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} H^2_{ij,ts}[\hat{e}^{(l)}_{it}]^2[\hat{e}^{(l)}_{js}]^2 \\
= \;& 1+O_P\left(N^{-1}\right) = 1+o_{P^*}(1),
\end{aligned}$$

where we follow the proof of Theorem 3.4.2 and show the term $O_P\left(N^{-1}\right)$ in the last line. For $(i2)$, recall that $\bar{q}_{k_1k_2}$ is the (k_1,k_2)th element of $\bar{Q}_{pp}$, and $H_{ij,ts} = \sum_{k_1=1}^{K}\sum_{k_2=1}^{K}\bar{q}_{k_1k_2}\tilde{Z}_{it,k_1}\tilde{Z}_{js,k_2}$. Let $\phi^*_{it,k}\equiv\tilde{Z}_{it,k}e^*_{it}$. Then we have

$$\begin{aligned}
G^*_I \;=\; & \tfrac{16}{N^4T^4\mathbb{V}^2_{NT}} \sum_{1\le k_1,\ldots,k_8\le K} \bar{q}_{k_1k_2}\bar{q}_{k_3k_4}\bar{q}_{k_5k_6}\bar{q}_{k_7k_8} \\
& \times \sum_{1\le i<j<N}\sum_{1\le t_1,\ldots,t_8\le T} E^*\left(\phi^*_{it_1,k_1}\phi^*_{it_3,k_3}\phi^*_{it_5,k_5}\phi^*_{it_7,k_7}\right)E^*\left(\phi^*_{jt_2,k_2}\phi^*_{jt_4,k_4}\phi^*_{jt_6,k_6}\phi^*_{jt_8,k_8}\right)
\end{aligned}$$

First, note that the term inside the last summation takes values 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 $S=\{t_1,\ldots,t_8\}$: (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

$$\begin{aligned}
|G_{Ia}^*| &\leq \tfrac{16}{N^4T^4\mathbb{V}_{NT}^{*2}} \sum_{1\leq k_1,\ldots,k_8\leq K} \left|\bar{q}_{k_1k_2}\right|\left|\bar{q}_{k_3k_4}\right|\left|\bar{q}_{k_5k_6}\right|\left|\bar{q}_{k_7k_8}\right| \\
&\quad \times \left\{\sum_{i=1}^{N}\sum_{1\leq t_5<t_1\leq T} \tilde{Z}_{it_1,k_1}\tilde{Z}_{it_1,k_3}(\hat{e}_{it_1}^{(l)})^2\tilde{Z}_{it_5,k_5}\tilde{Z}_{it_5,k_7}(\hat{e}_{it_5}^{(l)})^2\right\} \\
&\quad \times \left\{\sum_{j=1}^{N}\sum_{1\leq t_6<t\leq T} \tilde{Z}_{it_2,k_2}\tilde{Z}_{it_2,k_4}(\hat{e}_{it_2}^{(l)})^2\tilde{Z}_{it_6,k_6}\tilde{Z}_{it_6,k_8}(\hat{e}_{it_6}^{(l)})^2\right\} \\
&= \tfrac{64}{N^4T^4\mathbb{V}_{NT}^{*2}} O_P\left(K^8N^2T^4\right) = O_P\left(K^6/N^2\right) = O_{P^*}\left(K^6/N^2\right).
\end{aligned}$$

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 $G_{II}^* \equiv \sum_{1\leq i<j<l\leq N} E^*(W_{il}^{*2}W_{jl}^{*2} + W_{ij}^{*2}W_{il}^{*2} + W_{ij}^{*2}W_{lj}^{*2}) = G_{II,1}^* + G_{II,2}^* + G_{II,3}^*$. By the IID property of v_{it}, we have

$$\begin{aligned}
&G_{II,1}^* \\
&\equiv \tfrac{16}{N^4T^4\mathbb{V}_{NT}^{*2}} \sum_{1\leq i<j<l\leq N}\sum_{1\leq t_1,\ldots,t_6\leq T} E^*\left[e_{it_1}^{*2}e_{jt_2}^{*2}H_{il,t_1t_3}H_{il,t_1t_4}H_{jl,t_2t_5}H_{jl,t_2t_6}e_{lt_3}^*e_{lt_4}^*e_{lt_5}^*e_{lt_6}^*\right] \\
&= \tfrac{16}{N^4T^4\mathbb{V}_{NT}^{*2}} \sum_{1\leq i<j<l\leq N}\sum_{1\leq t_1,t_2\leq T}\sum_{1\leq t_3\neq t_6\leq T} H_{il,t_1t_3}^2H_{jl,t_2t_6}^2[\hat{e}_{it_1}^{(l)}]^2[\hat{e}_{jt_2}^{(l)}]^2[\hat{e}_{lt_3}^{(l)}]^2[\hat{e}_{lt_6}^{(l)}]^2 \\
&\quad + \tfrac{48}{N^4T^4\mathbb{V}_{NT}^{2}} \sum_{1\leq i<j<l\leq N}\sum_{1\leq t_1,t_2\leq T}\sum_{1\leq t_3\leq T} H_{il,t_1t_3}^2H_{jl,t_2t_3}^2[\hat{e}_{it_1}^{(l)}]^2[\hat{e}_{jt_2}^{(l)}]^2[\hat{e}_{lt_3}^{(l)}]^4 \\
&= G_{II,11}^* + G_{II,12}^*, \text{ say.}
\end{aligned}$$

For $G_{II,11}^*$, we have

$$\begin{aligned}
&G_{II,11}^* \\
&\leq \tfrac{16}{N^4T^4\mathbb{V}_{NT}^{*2}} \sum_{1\leq i<j<l\leq N}\sum_{1\leq t_1,t_2\leq T}\sum_{1\leq t_3\neq t_6\leq T} \Big\{\operatorname{tr}\left[(\hat{e}_{it_1}^{(l)})^2\tilde{Z}_{it_1}\tilde{Z}_{it_1}'\bar{Q}_{pp}[(\hat{e}_{lt_3}^{(l)})^2\tilde{Z}_{lt_3}\tilde{Z}_{lt_3}'\bar{Q}_{pp}\right] \\
&\qquad \times \operatorname{tr}\left[(\hat{e}_{jt_2}^{(l)})^2\tilde{Z}_{jt_2}\tilde{Z}_{jt_2}'\bar{Q}_{pp}(\hat{e}_{it_6}^{(l)})^2\tilde{Z}_{lt_6}\tilde{Z}_{lt_6}'\bar{Q}_{pp}\right]\Big\} \\
&\leq \tfrac{8}{3N^2T^2\mathbb{V}_{NT}^{*2}} \sum_{l=1}^{N}\sum_{1\leq t_3\neq t_6\leq T} \operatorname{tr}\left[\tilde{\Omega}^*\bar{Q}_{pp}(\hat{e}_{lt_3}^{(l)})^2\tilde{Z}_{lt_3}\tilde{Z}_{lt_3}'\bar{Q}_{pp}\right]\operatorname{tr}\left[\tilde{\Omega}^*\bar{Q}_{pp}(\hat{e}_{it_6}^{(l)})^2\tilde{Z}_{lt_6}\tilde{Z}_{lt_6}'\bar{Q}_{pp}\right] \\
&\leq \tfrac{8\mu_1^2(\tilde{\Omega}^*)\mu_1^4(\bar{Q}_{pp})}{3N^2T^2\mathbb{V}_{NT}^{*2}} \sum_{l=1}^{N}\sum_{1\leq t_3\neq t_6\leq T} (\hat{e}_{lt_3}^{(l)})^2\left\|\tilde{Z}_{lt_3}\right\|_F^2(\hat{e}_{it_6}^{(l)})^2\left\|\tilde{Z}_{lt_6}\right\|_F^2 \\
&= \tfrac{8\left[\mu_1^2(\tilde{\Omega}_{NT})+o_P(1)\right]\mu_1^4(\bar{Q}_{pp})}{3N^2T^2\mathbb{V}_{NT}^{*2}} \left\{\sum_{l=1}^{N}\sum_{1\leq t_3\neq t_6\leq T} e_{lt_3}^2\left\|\tilde{Z}_{lt_3}\right\|_F^2 e_{it_6}^2\left\|\tilde{Z}_{lt_6}\right\|_F^2 + o_P\left(NT^2K^2\right)\right\} \\
&= O_P\left(N^{-2}T^{-2}K^{-2}\right)O_P\left(NT^2K^2\right) = O_P\left(N^{-1}\right) = O_{P^*}\left(N^{-1}\right).
\end{aligned}$$

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 .0.9, we have

$$
\begin{aligned}
G^*_{III,1} &= \tfrac{16}{N^4T^4\mathbb{V}^{*2}_{NT}} \sum_{1\leq i<j<r<l\leq N} \sum_{1\leq t_1,\ldots,t_8\leq T} \big\{E^*(H_{ij,t_1t_2}e^*_{it_1}e^*_{jt_2}H_{ir,t_3t_4}e^*_{it_3}e^*_{rt_4} \\
&\qquad \times H_{lj,t_5t_6}e^*_{lt_5}e^*_{jt_6}H_{lr,t_7t_8}e^*_{lt_7}e^*_{rt_8})\big\} \\
&= \tfrac{16}{N^4T^4\mathbb{V}^{*2}_{NT}} \sum_{1\leq i<j<r<l\leq N} \sum_{1\leq t,s,p,q\leq T} \big\{\mathrm{tr}[E^*(\bar{Q}_{pp}\tilde{Z}_{it}\tilde{Z}'_{it}e^{*2}_{it}\bar{Q}_{pp}\tilde{Z}_{rs}\tilde{Z}'_{rs}e^{*2}_{rs}\bar{Q}_{pp} \\
&\qquad \times \tilde{Z}_{lp}\tilde{Z}'_{lp}e^{*2}_{lp}\bar{Q}_{pp}\tilde{Z}_{jq}\tilde{Z}'_{jq}e^{*2}_{jq})]\big\} \\
&= \tfrac{2}{3\mathbb{V}^{*2}_{NT}}\mathrm{tr}\left(\bar{Q}_{pp}\tilde{\Omega}^*\bar{Q}_{pp}\tilde{\Omega}^*\bar{Q}_{pp}\tilde{\Omega}^*\bar{Q}_{pp}\tilde{\Omega}^*\right) = O_P\left(\frac{1}{K}\right) = o_{P^*}(1)
\end{aligned}
$$

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.

■

.0.9 Some technical lemmas

Let $\{\xi_t, t\geq 1\}$ be a $\mathscr{D}$-strong mixing process with mixing coefficient $\alpha_{\mathscr{D}}(\cdot)$. We will use the following lemmas frequently.

Lemma .0.26 *(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 $\|A_1\|_{p,\mathscr{D}}$ and $\|A_2\|_{q,\mathscr{D}}$ are bounded in probability, where $p,q>1$ and $p^{-1}+q^{-1}<1$. Then*

$$|E_{\mathscr{D}}(A_1A_2) - E_{\mathscr{D}}(A_1)E_{\mathscr{D}}(A_2)| \leq 8\|A_1\|_{p,\mathscr{D}}\|A_2\|_{q,\mathscr{D}}\alpha_{\mathscr{D}}^{1-p^{-1}-q^{-1}}(\tau).$$

Lemma .0.27 *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

$$\begin{aligned}
\Pr\left(\max_{1\le t\le T}|A_t| > \varepsilon_T\right) &\le \sum_{t=1}^{T}\Pr(|A_t|>\varepsilon_T) = \sum_{t=1}^{T} E\left[1\left(|A_t|>\varepsilon_T\right)\right] \\
&\le \sum_{t=1}^{T} E\left[\frac{|A_t|^q}{\varepsilon_T^q} 1\left(|A_t|>\varepsilon_T\right)\right] = \varepsilon_T^{-q}\sum_{t=1}^{T} E\left[|A_t|^q 1\left(|A_t|>\varepsilon_T\right)\right] \\
&\le \max_{1\le t\le T} E\left[|A_t|^q 1\left(|A_t|>\varepsilon_T\right)\right] \to 0.
\end{aligned}$$

It follows that $\max_{1\le t\le T}|A_t| = o_P\left(T^{1/q}\right)$. ■

Lemma .0.28 *Let A be an $n\times m$ matrices, B and C be $m\times p$ matrices, and D be an $n\times n$ matrix. Then we have*

(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;

$(viii)$ $\|D\| \le \max_{1\le i\le n}|D_{ii}|$ *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). ■

.0.10 Data appendix

Countries listed for the application of economic growth: Argentina, Australia, Austria, Bangladesh, Belgium, Benin, Bolivia, Brazil, Burkina Faso, Cameroon, Canada, Central African Rep., Chile, China, Colombia, Congo, Rep., Costa Rica, Cote d'Ivoire, Denmark, Dominican Republic, Ecuador, Egypt, El Salvador, Finland, France, Ghana, Greece, Guatemala, Honduras, Hong Kong, Iceland, India, Indonesia, Ireland, Italy, Jamaica, Japan, Kenya, Korea, Republic of, Luxemburg, Madagascar, Malawi, Malysia, Mali, Mauritania, Mexico, Morocco, Netherlands, Niger, Nigeria, Norway, Pakistan, Papua New Guinea, Paraguay, Peru, Philippines, Rwanda, Senegal, South Africa, Spain, Sri Lanka, Sweden, Switzerland, Syria, Thailand, Trinidad and Tobago, Turkey, United Kingdom, United States of America, Uruguay, Zambia.

D Proofs in Chapter 4

.0.11 Proof of Theorem 4.3.1

Noting that

$$\begin{aligned}\Gamma_{nT} &= \sqrt{\frac{b}{n}}\sum_{i=1}^{n}\frac{(ESS_i-\varepsilon_i'Q\varepsilon_i)}{\sigma_i^2}+\sqrt{\frac{b}{n}}\sum_{i=1}^{n}\left(ESS_i-\varepsilon_i'Q\varepsilon_i\right)\left(\frac{1}{TSS_i/T}-\frac{1}{\sigma_i^2}\right)\\ &\equiv \Gamma_{nT,1}+\Gamma_{nT,2},\ \text{say},\end{aligned}$$

we complete the proof by showing that (i) $\Gamma_{nT,1}\xrightarrow{d}N(0,\Omega_0)$, and (ii) $\Gamma_{nT,2}=o_P(1)$. These results are established in Propositions .0.11 and .0.12, respectively.

Proposition .0.11 $\Gamma_{nT,1}\xrightarrow{d}N(0,\Omega_0)$.

Proof. Decompose

$$\Gamma_{nT,1}=\sqrt{\frac{b}{n}}\sum_{i=1}^{n}\frac{\widehat{u}_i'(\bar{H}-L)\widehat{u}_i}{\sigma_i^2}-\sqrt{\frac{b}{n}}\sum_{i=1}^{n}\frac{\varepsilon_i'Q\varepsilon_i}{\sigma_i^2}\equiv\Gamma_{nT,11}-\Gamma_{nT,12}. \tag{.0.80}$$

Let $X_i^*\equiv X_i-S_TX$ and $\varepsilon_i^*\equiv\varepsilon_i-S_T\varepsilon$. Define

$$\bar{\mathbf{f}}\equiv\left(\overline{f}(1/T),...,\overline{f}(T/T)\right)'\text{ and }\bar{\mathbf{f}}^*\equiv\bar{\mathbf{f}}-S_T\mathbf{F}, \tag{.0.81}$$

where $\overline{f}(\tau)\equiv n^{-1}\sum_{i=1}^{n}f_i(\tau)$. Noting that

$$\widehat{u}_i=\varepsilon_i^*-X_i^*(\widehat{\beta}-\beta)+\bar{\mathbf{f}}^*+(\mathbf{f}_i-\bar{\mathbf{f}})+\alpha_i i_T \tag{.0.82}$$

and $Mi_T=0$, we have

$$\Gamma_{nT,11}=\sqrt{\frac{b}{n}}\sum_{i=1}^{n}\frac{\widehat{u}_i'(\bar{H}-L)\widehat{u}_i}{\sigma_i^2}=\sum_{l=1}^{10}D_{nTl} \tag{.0.83}$$

where

$$
\begin{array}{ll}
D_{nT1} \equiv \sqrt{\frac{b}{n}} \sum_{i=1}^{n} \varepsilon_i^{*\prime} (\bar{H} - L) \varepsilon_i^{*} / \sigma_i^2, & D_{nT2} \equiv \sqrt{\frac{b}{n}} \sum_{i=1}^{n} (\mathbf{f}_i - \bar{\mathbf{f}})' (\bar{H} - L) (\mathbf{f}_i - \bar{\mathbf{f}}) / \sigma_i^2, \\
D_{nT3} \equiv \sqrt{\frac{b}{n}} \sum_{i=1}^{n} (\widehat{\beta} - \beta)' X_i^{*\prime} (\bar{H} - L) X_i^{*} (\widehat{\beta} - \beta) / \sigma_i^2, & D_{nT4} \equiv \sqrt{\frac{b}{n}} \sum_{i=1}^{n} \bar{\mathbf{f}}^{*\prime} (\bar{H} - L) \bar{\mathbf{f}}^{*} / \sigma_i^2, \\
D_{nT5} \equiv -2\sqrt{\frac{b}{n}} \sum_{i=1}^{n} \varepsilon_i^{*\prime} (\bar{H} - L) X_i^{*} (\widehat{\beta} - \beta) / \sigma_i^2, & D_{nT6} \equiv 2\sqrt{\frac{b}{n}} \sum_{i=1}^{n} \varepsilon_i^{*\prime} (\bar{H} - L) \bar{\mathbf{f}}^{*} / \sigma_i^2, \\
D_{nT7} \equiv -2\sqrt{\frac{b}{n}} \sum_{i=1}^{n} (\widehat{\beta} - \beta)' X_i^{*\prime} (\bar{H} - L) \bar{\mathbf{f}}^{*} / \sigma_i^2, & D_{nT8} \equiv 2\sqrt{\frac{b}{n}} \sum_{i=1}^{n} \varepsilon_i^{*\prime} (\bar{H} - L) (\mathbf{f}_i - \bar{\mathbf{f}}) / \sigma_i^2, \\
D_{nT9} \equiv -2\sqrt{\frac{b}{n}} \sum_{i=1}^{n} (\widehat{\beta} - \beta)' X_i^{*\prime} (\bar{H} - L) (\mathbf{f}_i - \bar{\mathbf{f}}) / \sigma_i^2, & D_{nT10} \equiv 2\sqrt{\frac{b}{n}} \sum_{i=1}^{n} \bar{\mathbf{f}}^{*\prime} (\bar{H} - L) (\mathbf{f}_i - \bar{\mathbf{f}}) / \sigma_i^2.
\end{array}
$$

Under H_0, $D_{nTs} = 0$ for $s = 2, 8, 9, 10$. We complete the proof of the proposition by showing that:

$$
\begin{aligned}
\mathscr{D}_{nT1} &\equiv D_{nT1} - \Gamma_{nT,12} \xrightarrow{d} N(0, \Omega_0), \text{ and} && (.0.84) \\
D_{nTs} &= o_P(1), \ s = 3, \ldots, 7. && (.0.85)
\end{aligned}
$$

Step 1. We first prove (.0.84). Noting that $\varepsilon_i^* \equiv \varepsilon_i - S_T \varepsilon$, we can decompose $\mathscr{D}_{nT1}$ as:

$$
\begin{aligned}
\mathscr{D}_{nT1} &= \sqrt{\frac{b}{n}} \sum_{i=1}^{n} \frac{\varepsilon_i^{*\prime} (\bar{H} - L) \varepsilon_i^{*}}{\sigma_i^2} - \sqrt{\frac{b}{n}} \sum_{i=1}^{n} \frac{\varepsilon_i' Q \varepsilon_i}{\sigma_i^2} \\
&= \sqrt{\frac{b}{n}} \sum_{i=1}^{n} \frac{\varepsilon_i' (\bar{H} - L - Q) \varepsilon_i}{\sigma_i^2} + \sqrt{\frac{b}{n}} \varepsilon' S_T' (\bar{H} - L) S_T \varepsilon \sum_{i=1}^{n} \frac{1}{\sigma_i^2} \\
&\quad -2\sqrt{\frac{b}{n}} \sum_{i=1}^{n} \frac{\varepsilon_i' (\bar{H} - L) S_T \varepsilon}{\sigma_i^2} \\
&\equiv \mathscr{D}_{nT11} + \mathscr{D}_{nT12} - 2\mathscr{D}_{nT13}.
\end{aligned}
$$

We prove (.0.84) by showing that $\mathscr{D}_{nT11} \xrightarrow{d} N(0, \Omega_0)$ and $\mathscr{D}_{nT1s} = o_P(1)$ for $s = 2, 3$. The former claim follows from Lemma .0.29 below. We now prove the latter claim. Let $\overline{\mathscr{D}}_{nT12} \equiv \sqrt{nb} \varepsilon' S_T' (\bar{H} - L) S_T \varepsilon$. By Lemmas .0.31(ii) and .0.34, we have

$$
\begin{aligned}
\overline{\mathscr{D}}_{nT12} &= \sqrt{nb} \sum_{t=1}^{T} \sum_{s=1}^{T} \left(e_1' S(t/T) \varepsilon \right) \left(\bar{H}_{ts} - T^{-1} \right) \left(e_1' S(s/T) \varepsilon \right) \\
&\leq \sqrt{nb} \max_{1 \leq t \leq T} \left| e_1' S(t/T) \varepsilon \right|^2 \sum_{t=1}^{T} \sum_{s=1}^{T} \left| \bar{H}_{ts} - T^{-1} \right| \\
&= \sqrt{nb} O_P \left(\frac{\log(nT)}{nTh} \right) O(T) = O_P \left(\frac{\log(nT)}{\sqrt{nb^{-1}h^2}} \right) = o_P(1).
\end{aligned}
$$

Then $\mathscr{D}_{nT12} = o_P(1)$ by Assumption A2(iii).

For $\mathscr{D}_{nT13}$, we have $\mathscr{D}_{nT13} = n^{-1/2}b^{1/2}\sum_{i=1}^{n} \varepsilon_i' (\bar{H} - L) S_T \varepsilon/\sigma_i^2 = \mathscr{D}_{nT131} + \mathscr{D}_{nT132}$, where

$$\begin{aligned}
\mathscr{D}_{nT131} &\equiv \sqrt{\frac{b}{n}} \sum_{i=1}^{n} \sum_{t=1}^{T} a_{tt} \varepsilon_{it} e_1' S(t/T) \varepsilon \sigma_i^{-2}, \\
\mathscr{D}_{nT132} &\equiv \sqrt{\frac{b}{n}} \sum_{i=1}^{n} \sum_{1 \le s \ne t \le T} a_{ts} \varepsilon_{it} e_1' S(s/T) \varepsilon \sigma_i^{-2},
\end{aligned}$$

and $a_{ts} \equiv \bar{H}_{ts} - T^{-1}$. For $\mathscr{D}_{nT131}$, write

$$\begin{aligned}
\mathscr{D}_{nT131} &= \frac{b^{1/2}}{n^{3/2}} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{t=1}^{T} a_{tt} \varepsilon_{it} e_1' s(t/T) \varepsilon_j \sigma_i^{-2} \\
&= \frac{b^{1/2}}{Tn^{3/2}} \sum_{1 \le i,j \le n} \sum_{1 \le t,s \le T} a_{tt} c_{ts} k_{h,ts} \varepsilon_{it} \varepsilon_{js} \sigma_i^{-2} \\
&= \frac{b^{1/2}}{Tn^{3/2}} \sum_{i=1}^{n} \sum_{t=1}^{T} a_{tt} c_{tt} k_{h,tt} \varepsilon_{it}^2 \sigma_i^{-2} \\
&\quad + \frac{b^{1/2}}{Tn^{3/2}} \sum_{i=1}^{n} \sum_{1 \le t < s \le T} (a_{tt} c_{ts} + a_{ss} c_{st}) k_{h,ts} \varepsilon_{it} \varepsilon_{is} \sigma_i^{-2} \\
&\quad + \frac{b^{1/2}}{Tn^{3/2}} \sum_{1 \le i \ne j \le n} \sum_{t=1}^{T} a_{tt} c_{tt} k_{h,tt} \varepsilon_{it} \varepsilon_{jt} \sigma_i^{-2} \\
&\quad + \frac{b^{1/2}}{Tn^{3/2}} \sum_{1 \le i \ne j \le n} \sum_{1 \le t < s \le T} (a_{tt} c_{ts} + a_{ss} c_{st}) k_{h,ts} \varepsilon_{it} \varepsilon_{js} \sigma_i^{-2} \\
&\equiv \mathscr{D}_{nT131a} + \mathscr{D}_{nT131b} + \mathscr{D}_{nT131c} + \mathscr{D}_{nT131d},
\end{aligned}$$

where $c_{ts} \equiv e_1' [T^{-1} z_h^{[p]}(t/T)' K_h(t/T) z_h^{[p]}(t/T)]^{-1} z_{h,s}^{[p]}(t/T)$. By Lemmas .0.31 and .0.33(iii) and Assumption A5, we have

$$E|\mathscr{D}_{nT131a}| \le \tfrac{k(0)b^{1/2}}{n^{1/2}h} \max_{1 \le t \le n} |a_{tt}| \left(\frac{1}{T} \sum_{t=1}^{T} |c_{tt}| \right) = n^{-1/2} b^{1/2} h^{-1} O\left(T^{-1}b^{-1}\right) O(1) = o(1).$$

So $\mathscr{D}_{nT131a} = o_P(1)$ by the Markov inequality. For $\mathscr{D}_{nT131b}$, we have by Lemmas .0.31 and .0.33(ii)

$$\begin{aligned}
& E\left(\mathscr{D}_{nT131b}^2\right) \\
= \;& \frac{b}{T^2 n^3} \sum_{i=1}^n \sum_{j=1}^n \sum_{1 \leq t_1 < t_2 \leq T} \sum_{1 \leq t_3 < t_4 \leq T} e_{t_1 t_2} k_{h,t_1 t_2} e_{t_3 t_4} k_{h,t_3 t_4} E\left(\varepsilon_{it_1} \varepsilon_{it_2} \varepsilon_{jt_3} \varepsilon_{jt_4}\right) \sigma_i^{-2} \sigma_j^{-2} \\
= \;& \frac{b}{T^2 n^3} \sum_{i=1}^n \sum_{j=1}^n \sum_{1 \leq t_1 < t_2 \leq T} \left(e_{t_1 t_2} k_{h,t_1 t_2}\right)^2 E\left(\varepsilon_{it_1} \varepsilon_{it_2} \varepsilon_{jt_1} \varepsilon_{jt_2}\right) \sigma_i^{-2} \sigma_j^{-2} \\
\leq \;& \frac{2b}{T^2 n^3} \sum_{i=1}^n \sum_{j=1}^n \sum_{1 \leq t_1 < t_2 \leq T} \left(a_{t_1 t_1}^2 c_{t_1 t_2}^2 + a_{t_2 t_2}^2 c_{t_2 t_1}^2\right) k_{h,t_1 t_2}^2 \left|E\left(\varepsilon_{it_1} \varepsilon_{it_2} \varepsilon_{jt_1} \varepsilon_{jt_2}\right)\right| \sigma_i^{-2} \sigma_j^{-2} \\
\leq \;& \frac{2b}{T^2 n^2} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \rho_{ij}^2\right) \sum_{1 \leq t_1 < t_2 \leq T} \left(a_{t_1 t_1}^2 c_{t_1 t_2}^2 + a_{t_2 t_2}^2 c_{t_2 t_1}^2\right) k_{h,t_1 t_2}^2 \\
\leq \;& \frac{2b}{n^2 h} \left(\max_{1 \leq t \leq T} a_{tt}^2\right) \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \rho_{ij}^2\right) \left(\frac{h}{T^2} \sum_{1 \leq t_1 \neq t_2 \leq T} c_{t_1 t_2}^2 k_{h,t_1 t_2}^2\right) \\
= \;& \frac{2b}{n^2 h} O\left(T^{-2} b^{-2}\right) O(1) = O\left(n^{-2} T^{-2} b^{-1} h^{-1}\right) = o(1),
\end{aligned}$$

where $e_{ts} \equiv a_{tt} c_{ts} + a_{ss} c_{st}$, $\rho_{ij} \equiv \omega_{ij} \sigma_i^{-1} \sigma_j^{-1}$, and the second equality follows from the fact that $E(\varepsilon_{it_1} \varepsilon_{it_2} \varepsilon_{jt_1} \varepsilon_{jt_3}) = 0$ and $E(\varepsilon_{it_1} \varepsilon_{it_2} \varepsilon_{jt_3} \varepsilon_{jt_4}) = 0$ when t_1, t_2, t_3, and t_4 are all distinct by Assumptions A2(ii)-(iii). It follows that $\mathscr{D}_{nT131b} = o_P(1)$ by the Chebyshev inequality. For $\mathscr{D}_{nT131c}$, we have by Lemma .0.31 and Assumptions A2 and A5

$$
\begin{aligned}
& E\left[\mathscr{D}_{nT131c}^2\right] \\
= & \frac{b}{T^2n^3}\sum_{1\leq i_1\neq i_2\leq n}\sum_{1\leq i_3\neq i_4\leq n}\sum_{t=1}^{T}\sum_{s=1}^{T} a_{tt}c_{tt}k_{h,tt}a_{ss}c_{ss}k_{h,ss}E\left(\varepsilon_{i_1t}\varepsilon_{i_2t}\varepsilon_{i_3s}\varepsilon_{i_4s}\right)\sigma_{i_1}^{-2}\sigma_{i_3}^{-2} \\
= & \frac{bk^2(0)}{T^2n^3h^2}\sum_{1\leq i_1\neq i_2\leq n}\sum_{1\leq i_3\neq i_4\leq n}\sum_{1\leq t\neq s\leq T} a_{tt}c_{tt}a_{ss}c_{ss}\omega_{i_1i_2}\omega_{i_3i_4}\sigma_{i_1}^{-2}\sigma_{i_3}^{-2} \\
& +\frac{bk^2(0)}{T^2n^3h^2}\sum_{t=1}^{T}\left[a_{tt}^2c_{tt}^2\sum_{1\leq i_1\neq i_2\leq n}\sum_{1\leq i_3\neq i_4\leq n} E\left(\varepsilon_{i_1t}\varepsilon_{i_2t}\varepsilon_{i_3t}\varepsilon_{i_4t}\right)\sigma_{i_1}^{-2}\sigma_{i_3}^{-2}\right] \\
\leq & \frac{b}{nh^2}\left(\max_{1\leq t\leq T}a_{tt}^2\right)\left(\frac{1}{n}\sum_{1\leq i_1\neq i_2\leq n}\omega_{i_1i_2}\sigma_{i_1}^{-2}\right)^2\left(\frac{1}{T}\sum_{t=1}^{T}|c_{tt}|\right)^2 \\
& +\frac{b}{Tnh^2}\left(\max_{1\leq t\leq T}a_{tt}^2\right)\left|\frac{1}{n^2}\sum_{1\leq i_1\neq i_2\leq n}\sum_{1\leq i_3\neq i_4\leq n} E\left(\varepsilon_{i_1t}\varepsilon_{i_2t}\varepsilon_{i_3t}\varepsilon_{i_4t}\right)\sigma_{i_1}^{-2}\sigma_{i_3}^{-2}\right|\left(\frac{1}{T}\sum_{t=1}^{T}c_{tt}^2\right) \\
= & \frac{b}{nh^2}O\left(T^{-2}b^{-2}\right)O(1)O(1)+\frac{b}{Tnh^2}O\left(T^{-2}b^{-2}\right)O(1)O(1) \\
= & O\left(n^{-1}T^{-2}h^{-2}b^{-1}+n^{-1}T^{-3}b^{-1}h^{-2}\right)=o(1).
\end{aligned}
$$

It follows that $\mathscr{D}_{nT131c}=o_P(1)$ by the Chebyshev inequality. Similarly, $\mathscr{D}_{nT131d}=o_P(1)$ because

$$
\begin{aligned}
& E\left(\mathscr{D}_{nT131d}\right)^2 \\
= & \frac{4b}{T^2n^3}\sum_{1\leq i_1\neq i_2\leq n}\sum_{1\leq i_3\neq i_4\leq n}\sum_{1\leq t_1<t_2\leq T} a_{t_1t_1}^2c_{t_1t_2}^2k_{h,t_1t_2}^2E\left(\varepsilon_{i_1t_1}\varepsilon_{i_2t_2}\varepsilon_{i_3t_1}\varepsilon_{i_4t_2}\right)\sigma_{i_1}^{-2}\sigma_{i_3}^{-2} \\
= & \frac{4b}{T^2n^3}\sum_{1\leq i_1\neq i_2\leq n}\sum_{1\leq i_3\neq i_4\leq n}\sum_{1\leq t_1<t_2\leq T} a_{t_1t_1}^2c_{t_1t_2}^2k_{h,t_1t_2}^2\omega_{i_1i_3}\omega_{i_2i_4}\sigma_{i_1}^{-2}\sigma_{i_3}^{-2} \\
\leq & \frac{4\underline{c}^{-2}b}{nh}\left(\max_{1\leq t\leq T}a_{tt}^2\right)\left(\frac{h}{T^2}\sum_{1\leq t_1<t_2\leq T}c_{t_1t_2}^2k_{h,t_1t_2}^2\right)\left(\frac{1}{n}\sum_{1\leq i_1,i_2\leq n}|\omega_{i_1i_2}|\right)^2 \\
= & \frac{b}{nh}O\left(T^{-2}b^{-2}\right)O(1)O(1)=O\left(n^{-1}T^{-2}h^{-1}b^{-1}\right)=o(1).
\end{aligned}
$$

In sum, we have shown that $\mathscr{D}_{nT131}=o_P(1)$.

For $\mathscr{D}_{nT132}$, we have

$$\begin{aligned}
\mathscr{D}_{nT132} &= \frac{b^{1/2}}{n^{3/2}} \sum_{1\leq i,j\leq n} \sum_{1\leq s\neq t\leq T} a_{ts}\varepsilon_{it}e_1' s\left(s/T\right)\varepsilon_j \sigma_i^{-2} \\
&= \frac{b^{1/2}}{Tn^{3/2}} \sum_{1\leq i,j\leq n} \sum_{1\leq s\neq t\leq T} \sum_{r=1}^{T} a_{ts}c_{sr}k_{h,sr}\varepsilon_{it}\varepsilon_{jr}\sigma_i^{-2} \\
&= \frac{b^{1/2}}{Tn^{3/2}} \sum_{1\leq i\neq j\leq n} \sum_{1\leq s\neq t\neq r\leq T} a_{ts}c_{sr}k_{h,sr}\varepsilon_{it}\varepsilon_{jr}\sigma_i^{-2} + o_P(1) \\
&\equiv \mathscr{D}_{nT132a} + o_P(1).
\end{aligned}$$

Following the same arguments as used in the proof of $\mathscr{D}_{nT131a} = o_P(1)$, we can show that $E\left(\mathscr{D}_{nT132a}\right)^2 = o(1)$. It follows that $\mathscr{D}_{nT132a} = o_P(1)$ and $\mathscr{D}_{nT132} = o_P(1)$.

Step 2. We now prove (.0.85). For D_{nT3}, by Assumption A2(iii), and Lemmas .0.32, .0.35(i) and .0.36, we have

$$\begin{aligned}
\left|D_{nT3}\right| &\leq \underline{c}^{-1} n^{-1/2} b^{1/2} \left\|\bar{H}-L\right\| \left\|\widehat{\beta}-\beta\right\|^2 \sum_{i=1}^{n} \left\|X_i - S_T X\right\|^2 \\
&= \underline{c}^{-1} n^{-1/2} \left(b^{1/2}\left\|\bar{H}-L\right\|\right) \left\|\widehat{\beta}-\beta\right\|^2 \left\|X - S_{nT}X\right\|^2 \\
&= n^{-1/2} O(1)\, O_P\left(n^{-1}T^{-1}\right) O_P(nT) = O_P\left(n^{-1/2}\right) = o_P(1).
\end{aligned}$$

For D_{nT4}, noting that $\max_{1\leq t\leq T} \left|\overline{f}(t) - e_1' S(t/T)\mathbf{F}\right| = O\left(h^{p+1}\right)$ by analysis analogous to CGL (2010), by Lemma .0.32 and Assumption A5 we have

$$\begin{aligned}
\left|D_{nT4}\right| &\leq \underline{c}^{-1} n^{1/2} \left(b^{1/2}\left\|\bar{H}-L\right\|\right) ||\bar{\mathbf{f}}^*||^2 = n^{1/2} O(1)\, O\left(Th^{2p+2}\right) \\
&= O\left(n^{1/2}Th^{2p+2}\right) = o(1).
\end{aligned}$$

Now decompose D_{nT5} as follows

$$\begin{aligned}
D_{nT5} &= -2\left[\sqrt{\frac{b}{n}} \sum_{i=1}^{n} \varepsilon_i' \left(\bar{H}-L\right) X_i^* \sigma_i^{-2} - \sqrt{\frac{b}{n}} \sum_{i=1}^{n} \left(S_T\varepsilon\right)' \left(\bar{H}-L\right) X_i^* \sigma_i^{-2}\right] (\widehat{\beta}-\beta) \\
&\equiv -2\left(D_{nT51} - D_{nT52}\right)(\widehat{\beta}-\beta), \text{ say.}
\end{aligned}$$

Noting that $D_{nT51} = \sqrt{\frac{b}{n}} \sum_{i=1}^{n} \sigma_i^{-2} \varepsilon_i' (\bar{H} - L)(X_i - S_T X) = \sqrt{\frac{b}{n}} \sum_{i=1}^{n} \sum_{t=1}^{T} \sum_{s=1}^{T} \sigma_i^{-2} \varepsilon_{it} a_{ts}$ $\times [X_{is} - e_1' S(s/T) X]$, by Assumption A2, the Cauchy inequality, and Lemma .0.32(ii),

$$
\begin{aligned}
& E \|D_{nT51}\|^2 \\
= \;& \frac{b}{n} \sum_{1 \leq i,j \leq n1} \sum_{1 \leq t,s,r \leq T} a_{ts} a_{tr} E \left\{ \operatorname{tr}[(X_{is} - e_1' S(s/T) X)(X_{jr} - e_1' S(r/T) X)'] \right\} \omega_{ij} \sigma_i{}^{-2} \sigma_j{}^{-2} \\
\leq \;& Tb \max_{1 \leq i \leq n} \max_{1 \leq s \leq T} \left(E \left\| X_{is} - e_1' S(s/T) X \right\|^2 \right) \left(\frac{1}{n} \sum_{1 \leq i,j \leq n1} \left| \rho_{ij} \right| \right) \left(\frac{1}{T} \sum_{1 \leq t,s,r \leq T} \left| a_{ts} a_{tr} \right| \right) \\
= \;& TbO(1) O(1) O(1) = O(Tb).
\end{aligned}
$$

For D_{nT52} we have

$$
\begin{aligned}
\|D_{nT52}\|^2 &= \frac{b}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \operatorname{tr} \left[(\bar{H} - L) X_i^* X_j^{*\prime} (\bar{H} - L) S_T \varepsilon \varepsilon' S_T' \right] \sigma_i^{-2} \sigma_j^{-2} \\
&= \frac{b}{n} \operatorname{tr} \left[\left(\sum_{i=1}^{n} \sum_{j=1}^{n} X_i^* X_j^{*\prime} \sigma_i^{-2} \sigma_j^{-2} \right) (\bar{H} - L) S_T \varepsilon \varepsilon' S_T' (\bar{H} - L) \right] \\
&\leq \frac{\underline{c}^{-2}}{n} \left(\sum_{i=1}^{n} \|X_i^*\| \right)^2 \left(b \|\bar{H} - L\|^2 \right) \|S_T \varepsilon\|^2 \\
&= \frac{1}{n} O_P\left(Tn^2\right) O(1) O_P(1/(nh)) = O(T/h).
\end{aligned}
$$

It follows that $D_{nT5} = O_P(T^{1/2} b^{1/2} + T^{1/2} h^{-1/2}) O_P((nT)^{-1/2}) = O_P(n^{-1/2}(b^{1/2} + h^{-1/2})) = o_P(1)$.

For D_{nT6}, we write

$$
\begin{aligned}
D_{nT6} &= 2\sqrt{\frac{b}{n}} \sum_{i=1}^{n} \sigma_i^{-2} \varepsilon_i' (\bar{H} - L) \left(\bar{\mathbf{f}} - S_T \overline{\mathbf{F}} \right) - 2\sqrt{\frac{b}{n}} \sum_{i=1}^{n} \sigma_i^{-2} (S_T \varepsilon)' (\bar{H} - L) \left(\bar{\mathbf{f}} - S_T \overline{\mathbf{F}} \right) \\
&\equiv 2D_{nT61} - 2D_{nT62},
\end{aligned}
$$

where $\overline{\mathbf{F}} \equiv i_n \otimes \bar{\mathbf{f}} = i_n \otimes \mathbf{f}$ under H_0. Noting that $D_{nT61} = n^{-1/2} b^{1/2} \sum_{i=1}^n \sum_{t=1}^T \sum_{s=1}^T \sigma_i^{-2} \varepsilon_{it} a_{ts}$ $\times [\overline{f}(s/T) - e_1' S(s/T) \overline{\mathbf{F}}]$, by Assumptions A2 and A5 and Lemma .0.32(ii), we have

$$
\begin{aligned}
& E\left(D_{nT61}^2\right) \\
= & \frac{b}{n} \sum_{1 \leq i,j \leq n1} \sum_{1 \leq t,s,r \leq T} \omega_{ij} a_{ts} a_{tr} \left[\overline{f}(s/T) - e_1' S(s/T) \overline{\mathbf{F}}\right] \left[\overline{f}(r/T) - e_1' S(r/T) \overline{\mathbf{F}}\right] \sigma_i^{-2} \sigma_j^{-2} \\
\leq & \underline{c}^{-2} T b \max_{1 \leq s \leq T} \left| \overline{f}\left(\frac{s}{T}\right) - e_1' S(\frac{s}{T}) \overline{\mathbf{F}} \right|^2 \left(\frac{1}{n} \sum_{1 \leq i,j \leq n1} |\omega_{ij}| \right) \left(\frac{1}{T} \sum_{1 \leq t,s,r \leq T} |a_{ts} a_{tr}| \right) \\
= & TbO\left(h^{2p+2}\right) O(1) O(1) = O\left(Tbh^{2p+2}\right) = o(1).
\end{aligned}
$$

It follows that $D_{nT61} = o_P(1)$ by the Chebyshev inequality. For D_{nT62}, we can follow the proof of D_{nT52} and show that $D_{nT62} = o_P(1)$. Consequently, $D_{nT6} = o_P(1)$. Now write $D_{nT7} \equiv -2\sqrt{b/n} \sum_{i=1}^n \sigma_i^{-2} (\widehat{\beta} - \beta)' X_i^{*\prime} \bar{H} \bar{\mathbf{f}}^* + 2(b/n)^{1/2} \sum_{i=1}^n \sigma_i^{-2} (\widehat{\beta} - \beta)' X_i^{*\prime} L \bar{\mathbf{f}}^* \equiv -2D_{nT71} + 2D_{nT72}$. By the Cauchy-Schwarz inequality, we have

$$
\begin{aligned}
D_{nT71} & \leq \left(\sqrt{\frac{b}{n}} \left\| \widehat{\beta} - \beta \right\|^2 \sum_{i=1}^n \sigma_i^{-4} \left\| X_i^{*\prime} \bar{H} X_i^* \right\| \right)^{1/2} \left(\sqrt{nb} \bar{\mathbf{f}}^{*\prime} \bar{H} \bar{\mathbf{f}}^* \right)^{1/2} \\
& = \left[O_P\left(n^{-1/2}\right) O\left(Tn^{1/2} h^{2(p+1)}\right) \right]^{1/2} = O_P\left(T^{1/2} h^{p+1}\right) = o_P(1).
\end{aligned}
$$

Similarly, we have $D_{nT72} = o_P(1)$. Thus $D_{nT7} = o_P(1)$. ■

Lemma .0.29 $\mathscr{D}_{nT11} = \frac{b^{1/2}}{\sqrt{n}} \sum_{i=1}^n \varepsilon_i' (\bar{H} - L - Q) \varepsilon_i / \sigma_i^2 \xrightarrow{d} N(0, \Omega_0)$.

Proof. Write $\mathscr{D}_{nT11} = \frac{1}{\sqrt{T}} \sum_{t=2}^T Z_{nT,t}$, where $Z_{nT,t} \equiv \frac{2b^{1/2}}{\sqrt{nT}} \sum_{s=1}^{t-1} \sum_{i=1}^n \alpha_{ts} \sigma_i^{-2} \varepsilon_{it} \varepsilon_{is}$ and $\alpha_{ts} \equiv T\bar{H}_{ts} - 1 = Ta_{ts}$. Noting that $\{Z_{nT,t}, \mathscr{F}_{n,t}(\varepsilon)\}$ is an m.d.s., we prove the lemma by applying the martingale CLT. By Corollary 5.26 of White (2001) it suffices to show that: (i) $E\left(Z_{nT,t}^4\right) < C$ for all t and (n,T) for some $C < \infty$, and (ii) $T^{-1} \sum_{t=2}^T Z_{nT,t}^2 - \Omega_0 = o_P(1)$.

We first prove (i). For $2 \leq t \leq T$, decompose

$$\begin{aligned}
Z_{nT,t}^2 &= \frac{4b}{nT}\sum_{s_1=1}^{t-1}\sum_{s_2=1}^{t-1}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\alpha_{ts_1}\alpha_{ts_2}\sigma_{i_1}^{-2}\sigma_{i_2}^{-2}\varepsilon_{i_1t}\varepsilon_{i_1s_1}\varepsilon_{i_2t}\varepsilon_{i_2s_2} \\
&= \frac{4b}{nT}\sum_{s=1}^{t-1}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\alpha_{ts}^2\sigma_{i_1}^{-2}\sigma_{i_2}^{-2}\varepsilon_{i_1t}\varepsilon_{i_1s}\varepsilon_{i_2t}\varepsilon_{i_2s} \\
&\quad +\frac{4b}{nT}\sum_{1\leq s_1<s_2\leq t-1}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\alpha_{ts_1}\alpha_{ts_2}\sigma_{i_1}^{-2}\sigma_{i_2}^{-2}\varepsilon_{i_1t}\varepsilon_{i_1s_1}\varepsilon_{i_2t}\varepsilon_{i_2s_2} \\
&\quad +\frac{4b}{nT}\sum_{1\leq s_2<s_1\leq t-1}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\alpha_{ts_1}\alpha_{ts_2}\sigma_{i_1}^{-2}\sigma_{i_2}^{-2}\varepsilon_{i_1t}\varepsilon_{i_1s_1}\varepsilon_{i_2t}\varepsilon_{i_2s_2} \\
&\equiv z_{1t}+z_{2t}+z_{3t}, \text{ say.} \qquad (.0.86)
\end{aligned}$$

Then $E\left(Z_{nT,t}^4\right) = E\left(z_{1t}+z_{2t}+z_{3t}\right)^2 \leq 3\{E\left(z_{1t}^2\right)+E\left(z_{2t}^2\right)+E\left(z_{3t}^2\right)\} \equiv 3\{\mathscr{Z}_{1t}+\mathscr{Z}_{2t}+\mathscr{Z}_{3t}\}$, say.

$$\begin{aligned}
&\mathscr{Z}_{1t} \\
&= \tfrac{16b^2}{n^2T^2}\sum_{1\leq s_1,s_2\leq t-1}\sum_{1\leq i_1,i_2\leq n}\sum_{1\leq i_3,i_4\leq n}\alpha_{ts_1}^2\alpha_{ts_2}^2\sigma_{i_1}^{-2}\sigma_{i_2}^{-2}\sigma_{i_3}^{-2}\sigma_{i_4}^{-2}E\left(\varepsilon_{i_1t}\varepsilon_{i_2t}\varepsilon_{i_3t}\varepsilon_{i_4t}\varepsilon_{i_1s_1}\varepsilon_{i_2s_1}\varepsilon_{i_3s_2}\varepsilon_{i_4s_2}\right) \\
&= \tfrac{16b^2}{n^2T^2}\sum_{1\leq s_1,s_2\leq t-1}\sum_{1\leq i_1,i_2\leq n}\sum_{1\leq i_3,i_4\leq n}\alpha_{ts_1}^2\alpha_{ts_2}^2\sigma_{i_1}^{-2}\sigma_{i_2}^{-2}\sigma_{i_3}^{-2}\sigma_{i_4}^{-2}\kappa_{i_1i_2i_3i_4}E\left(\varepsilon_{i_1s_1}\varepsilon_{i_2s_1}\varepsilon_{i_3s_2}\varepsilon_{i_4s_2}\right) \\
&= \tfrac{16b^2}{n^2T^2}\sum_{s=1}^{t-1}\sum_{1\leq i_1,i_2\leq n}\sum_{1\leq i_3,i_4\leq n}\alpha_{ts}^4\sigma_{i_1}^{-2}\sigma_{i_2}^{-2}\sigma_{i_3}^{-2}\sigma_{i_4}^{-2}\kappa_{i_1i_2i_3i_4}^2 \\
&\quad +\tfrac{16b^2}{n^2T^2}\sum_{1\leq s_1\neq s_2\leq t-1}\sum_{1\leq i_1,i_2\leq n}\sum_{1\leq i_3,i_4\leq n}\alpha_{ts_1}^2\alpha_{ts_2}^2\sigma_{i_1}^{-2}\sigma_{i_2}^{-2}\sigma_{i_3}^{-2}\sigma_{i_4}^{-2}\kappa_{i_1i_2i_3i_4}\omega_{i_1i_2}\omega_{i_3i_4} \\
&\leq \tfrac{Cb^2}{T^2}\sum_{s=1}^{t-1}\alpha_{ts}^4+C\left(\tfrac{b}{T}\sum_{s=1}^{t-1}\alpha_{ts}^2\right)^2 \leq \tfrac{C}{Tb}+C\leq 2C.
\end{aligned}$$

Similarly,

$$\begin{aligned}
\mathscr{Z}_{2t} &= \frac{16b^2}{n^2T^2}\sum_{1\leq s_1<s_2\leq t-1}\sum_{1\leq s_3<s_4\leq t-1}\sum_{1\leq i_1,i_2\leq n}\sum_{1\leq i_3,i_4\leq n}\alpha_{ts_1}\alpha_{ts_2}\alpha_{ts_3}\alpha_{ts_4}\sigma_{i_1}^{-2}\sigma_{i_2}^{-2}\sigma_{i_3}^{-2}\sigma_{i_4}^{-2} \\
&\quad \times E\left(\varepsilon_{i_1t}\varepsilon_{i_2t}\varepsilon_{i_3t}\varepsilon_{i_4t}\varepsilon_{i_1s_1}\varepsilon_{i_2s_2}\varepsilon_{i_3s_3}\varepsilon_{i_4s_4}\right) \\
&= \frac{16b^2}{n^2T^2}\sum_{1\leq s_1<s_2\leq t-1}\sum_{1\leq i_1,i_2\leq n}\sum_{1\leq i_3,i_4\leq n}\alpha_{ts_1}^2\alpha_{ts_2}^2\sigma_{i_1}^{-2}\sigma_{i_2}^{-2}\sigma_{i_3}^{-2}\sigma_{i_4}^{-2}\kappa_{i_1i_2i_3i_4}\omega_{i_1i_2}\omega_{i_3i_4} \\
&\leq \frac{Cb^2}{T^2}\sum_{1\leq s_1<s_2\leq t-1}\alpha_{ts_1}^2\alpha_{ts_2}^2\leq C,
\end{aligned}$$

where we have used the fact that $T^{-1}b\sum_{s=1}^{t}\alpha_{ts}^{2}\leq C$ uniformly in t and C may vary across lines. By the same token $\mathscr{Z}_{3t}\leq C$ for all t. Consequently, $E\left(Z_{nT,t}^{4}\right)<C$ for all t and some large enough constant C.

Now we prove (ii) by the Chebyshev inequality. First, by Assumption A2(ii)-(iii),

$$E\left(\frac{1}{T}\sum_{t=2}^{T}Z_{nT,t}^{2}\right)=\frac{4b}{nT^{2}}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\sum_{1\leq i,j\leq n}\alpha_{ts}^{2}\sigma_{i}^{-2}\sigma_{j}^{-2}\omega_{ij}^{2}=\frac{2b}{nT^{2}}\sum_{1\leq t\neq s\leq T}\alpha_{ts}^{2}\sum_{1\leq i,j\leq n}\rho_{ij}^{2},$$

where $\rho_{ij}=\omega_{ij}/(\sigma_{i}\sigma_{j})$ by Assumption A2. Second, decompose

$$E\left[\left(\frac{1}{T}\sum_{t=2}^{T}Z_{nT,t}^{2}\right)^{2}\right]=\frac{1}{T^{2}}\sum_{t=2}^{T}E\left(Z_{nT,t}^{4}\right)+\frac{2}{T^{2}}\sum_{2\leq t<s\leq T}E\left(Z_{nT,t}^{2}Z_{nT,s}^{2}\right)\equiv\mathbb{Z}_{1nT}+\mathbb{Z}_{2nT}.$$

By the proof of (i), $\mathbb{Z}_{1nT}=T^{-2}\sum_{t=2}^{T}E\left(Z_{nT,t}^{4}\right)=O(1/T)=o(1)$. For $\mathbb{Z}_{2nT}$, by (.0.86) we have $\mathbb{Z}_{2nT}=2T^{-2}\sum_{2\leq t<s\leq T}E(z_{1t}z_{1s}+z_{1t}z_{2s}+z_{1t}z_{3s}+z_{2t}z_{1s}+z_{2t}z_{2s}+z_{2t}z_{3s}+z_{3t}z_{1s}+z_{3t}z_{2s}+z_{3t}z_{3s})\equiv\sum_{j=1}^{9}\mathbb{Z}_{2nTj}$, say, where, e.g., $\mathbb{Z}_{2nT1}=2T^{-2}\sum_{2\leq t<s\leq T}E\left(z_{1t}z_{1s}\right)$. For $\mathbb{Z}_{2nT1}$, we have

$$\begin{aligned}
\mathbb{Z}_{2nT1} &= \frac{32b^{2}}{n^{2}T^{4}}\sum_{2\leq t_{1}<t_{2}\leq T}\sum_{s_{1}=1}^{t_{1}-1}\sum_{s_{2}=1}^{t_{2}-1}\sum_{1\leq i_{1},i_{2}\leq n}\sum_{1\leq i_{3},i_{4}\leq n}\alpha_{t_{1}s_{1}}^{2}\alpha_{t_{2}s_{2}}^{2}\sigma_{i_{1}}^{-2}\sigma_{i_{2}}^{-2}\sigma_{i_{3}}^{-2}\sigma_{i_{4}}^{-2}\\
&\quad\times\omega_{i_{3}i_{4}}E\left(\varepsilon_{i_{1}t_{1}}\varepsilon_{i_{2}t_{1}}\varepsilon_{i_{1}s_{1}}\varepsilon_{i_{2}s_{1}}\varepsilon_{i_{3}s_{2}}\varepsilon_{i_{4}s_{2}}\right)\\
&= \frac{32b^{2}}{n^{2}T^{4}}\sum_{2\leq t_{1}<t_{2}\leq T}\sum_{s_{1}=1}^{t_{1}-1}\sum_{s_{2}=1}^{t_{2}-1}\sum_{1\leq i_{1},i_{2}\leq n}\sum_{1\leq i_{3},i_{4}\leq n}\alpha_{t_{1}s_{1}}^{2}\alpha_{t_{2}s_{2}}^{2}\sigma_{i_{1}}^{-2}\sigma_{i_{2}}^{-2}\sigma_{i_{3}}^{-2}\sigma_{i_{4}}^{-2}\omega_{i_{1}i_{2}}^{2}\omega_{i_{3}i_{4}}^{2}\\
&\quad+O\left(\tfrac{1}{T}\right)\\
&= \frac{16b^{2}}{n^{2}T^{4}}\sum_{t_{1}=1}^{T}\sum_{t_{2}=1}^{T}\sum_{s_{1}=1}^{t_{1}-1}\sum_{s_{2}=1}^{t_{2}-1}\sum_{1\leq i_{1},i_{2}\leq n}\sum_{1\leq i_{3},i_{4}\leq n}\alpha_{t_{1}s_{1}}^{2}\alpha_{t_{2}s_{2}}^{2}\rho_{i_{1}i_{2}}^{2}\rho_{i_{3}i_{4}}^{2}+O(1/T)\\
&= \left(\frac{2b}{nT^{2}}\sum_{1\leq t\neq s\leq T}\alpha_{ts}^{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\rho_{ij}^{2}\right)^{2}+O(1/T).
\end{aligned}$$

Similarly, by Assumption A2 and Lemmas .0.31 and .0.32(ii)

$$\begin{aligned}
\mathbb{Z}_{2nT2} &= \frac{32b^2}{n^2T^4}\sum_{2\le t_1<t_2\le T}\sum_{s=1}^{t_1-1}\sum_{1\le s_1<s_2\le t_2-1}\sum_{1\le i_1,i_2\le n}\sum_{1\le i_3,i_4\le n}\Big\{\alpha_{t_1s}^2\alpha_{t_2s_1}\alpha_{t_2s_2}\sigma_{i_1}^{-2}\sigma_{i_2}^{-2}\sigma_{i_3}^{-2}\sigma_{i_4}^{-2} \\
&\quad\times\varsigma_{i_2i_3i_4}E\left(\varepsilon_{i_1t_1}\varepsilon_{i_1s}\varepsilon_{i_2s}\varepsilon_{i_3s_1}\varepsilon_{i_4s_2}\right)\Big\} \\
&= \frac{32b^2}{n^2T^4}\sum_{2\le t_1<t_2\le T}\sum_{s=1}^{t_1-1}\sum_{1\le i_1,i_2\le n}\sum_{1\le i_3,i_4\le n}\alpha_{t_1s}^2\alpha_{t_2s}\alpha_{t_2t_1}\sigma_{i_1}^{-2}\sigma_{i_2}^{-2}\sigma_{i_3}^{-2}\sigma_{i_4}^{-2}\varsigma_{i_2i_3i_4}\omega_{i_1i_4}\varsigma_{i_1i_2i_3} \\
&\le C\left(b^2\max_{1\le t\neq s\le T}a_{ts}^2\right)\left(\sum_{2\le t_1<t_2\le T}\sum_{s=1}^{t_1-1}\left|a_{t_2s}a_{t_2t_1}\right|\right)\left(\frac{1}{n^2}\sum_{1\le i_1,i_2\le n}\sum_{1\le i_3,i_4\le n}\left|\varsigma_{i_2i_3i_4}\varsigma_{i_1i_2i_3}\right|\right) \\
&= O\left(T^{-2}\right)O(T)O(1)=o(1),
\end{aligned}$$

where recall $\varsigma_{ijk}\equiv E\left(\varepsilon_{it}\varepsilon_{jt}\varepsilon_{kt}\right)$. Analogously we can show that $\mathbb{Z}_{2nTl}=o(1)$ for $l=3,4,...,9$. It follows that

$$E\left[\left(\frac{1}{T}\sum_{t=2}^{T}Z_{nT,t}^2\right)^2\right]=\left(\frac{2b}{nT^2}\sum_{1\le t\neq s\le T}\alpha_{ts}^2\sum_{i=1}^{n}\sum_{j=1}^{n}\rho_{ij}^2\right)^2+o(1),$$

and

$$\operatorname{Var}\left(\frac{1}{T}\sum_{t=2}^{T}Z_{nT,t}^2\right)=E\left[\left(\frac{1}{T}\sum_{t=2}^{T}Z_{nT,t}^2\right)^2\right]-\left[E\left(\frac{1}{T}\sum_{t=2}^{T}Z_{nT,t}^2\right)\right]^2=o(1).$$

Consequently, $\frac{1}{T}\sum_{t=2}^{T}Z_{nT,t}^2-\frac{2b}{nT^2}\sum_{1\le t\neq s\le T}\alpha_{ts}^2\sum_{i=1}^{n}\sum_{j=1}^{n}\rho_{ij}^2=o_P(1)$ and (ii) follows by the definition of Ω_0. ■

Proposition .0.12 $\Gamma_{nT,2}=o_P(1)$.

Proof. Let $\widehat{\sigma}_i^2\equiv TSS_i/T$. By a geometric expansion, $1/\widehat{\sigma}_i^2-1/\sigma_i^2=-(\widehat{\sigma}_i^2-\sigma_i^2)/\sigma_i^4+(\widehat{\sigma}_i^2-\sigma_i^2)^2/(\sigma_i^4\widehat{\sigma}_i^2)$. It follows that

$$\begin{aligned}
\Gamma_{nT,2} &= -\sqrt{\frac{b}{n}}\sum_{i=1}^{n}\left(ESS_i-\varepsilon_i'Q\varepsilon_i\right)\frac{\widehat{\sigma}_i^2-\sigma_i^2}{\sigma_i^4}+\sqrt{\frac{b}{n}}\sum_{i=1}^{n}\left(ESS_i-\varepsilon_i'Q\varepsilon_i\right)\frac{\left(\widehat{\sigma}_i^2-\sigma_i^2\right)^2}{\sigma_i^4\widehat{\sigma}_i^2} \\
&\equiv -\Gamma_{nT,21}+\Gamma_{nT,22},\ \text{say.}
\end{aligned}$$

Noting that $\widehat{u}_i=\varepsilon_i^*-X_i^*(\widehat{\beta}-\beta)+\overline{\mathbf{f}}^*+(\mathbf{f}_i-\overline{\mathbf{f}})+\alpha_i i_T$ and $Mi_T=0$ where $\overline{\mathbf{f}}$ and $\overline{\mathbf{f}}^*$ are defined in (.0.81), we have

$$\widehat{\sigma}_i^2=TSS_i/T=\widehat{u}_i'M\widehat{u}_i/T=\sum_{l=1}^{10}TSS_{il}/T, \tag{.0.87}$$

where

$$
\begin{aligned}
&TSS_{i1} \equiv \varepsilon_i^{*\prime} M \varepsilon_i^{*}, && TSS_{i2} \equiv (\widehat{\beta}-\beta)' X_i^{*\prime} M X_i^{*} (\widehat{\beta}-\beta), \\
&TSS_{i3} \equiv \bar{\mathbf{f}}^{*\prime} M \bar{\mathbf{f}}^{*}, && TSS_{i4} \equiv -2\varepsilon_i^{*\prime} M X_i^{*} (\widehat{\beta}-\beta), \\
&TSS_{i5} \equiv 2\varepsilon_i^{*\prime} M \bar{\mathbf{f}}^{*}, && TSS_{i6} \equiv -2\bar{\mathbf{f}}^{*\prime} M X_i^{*} (\widehat{\beta}-\beta), \\
&TSS_{i7} \equiv 2\varepsilon_i^{*\prime} M (\mathbf{f}_i-\bar{\mathbf{f}}), && TSS_{i8} \equiv (\mathbf{f}_i-\bar{\mathbf{f}})' M (\mathbf{f}_i-\bar{\mathbf{f}}), \\
&TSS_{i9} \equiv 2\bar{\mathbf{f}}^{*\prime} M (\mathbf{f}_i-\bar{\mathbf{f}}), && TSS_{i10} \equiv -2(\widehat{\beta}-\beta)' X_i^{*\prime} M (\mathbf{f}_i-\bar{\mathbf{f}}).
\end{aligned}
$$

Under H_0, we have $\mathbf{f}_i-\bar{\mathbf{f}}=0$. Thus $TSS_{il}=0$ for $l=7,\ldots,10$. We want to show that

$$
\max_{1\le i\le n}\left|T^{-1}TSS_{i1}-\sigma_i^2\right| = O_P(\upsilon_{nT}), \text{ and } \max_{1\le i\le n} T^{-1}TSS_{il} = o_P(\upsilon_{nT}) \text{ for } l=2,...,6, \tag{.0.88}
$$

where $\upsilon_{nT}\equiv n^{1/\lambda}T^{-1/2}$.

For TSS_{i1}, we have

$$
T^{-1}TSS_{i1}-\sigma_i^2 = \left(T^{-1}\varepsilon_i' M\varepsilon_i-\sigma_i^2\right) - 2T^{-1}\varepsilon_i' M S_T\varepsilon + T^{-1}\left(S_T\varepsilon\right)' M S_T \varepsilon. \tag{.0.89}
$$

We first bound the last term in (.0.89). By the idempotence of M and the Markov inequality, $T^{-1}\left(S_T\varepsilon\right)' M S_T\varepsilon \le T^{-1}\left\|S_T\varepsilon\right\|^2 = O_P\left(n^{-1}T^{-1}h^{-1}\right)$. For the first term in (.0.89), we want to show that $\max_{1\le i\le n}|\varepsilon_i' M\varepsilon_i/T-\sigma_i^2| = O_P(\upsilon_{nT})$. Write $\varepsilon_i' M\varepsilon_i/T = T^{-1}\sum_{t=1}^T(\varepsilon_{it}-\bar{\varepsilon}_i)^2 = T^{-1}\sum_{t=1}^T\varepsilon_{it}^2-\bar{\varepsilon}_i^2$. Let $\xi_{it}\equiv\varepsilon_{it}^2-\sigma_i^2$. Then by Assumption A2(iv) and the Chebyshev inequality, for any $\varepsilon>0$

$$
P\left(\max_{1\le i\le n}\frac{1}{T}\sum_{t=1}^T\xi_{it}\ge\varepsilon v_{nT}\right) \le \varepsilon^{-\lambda}\upsilon_{nT}^{-\lambda}\sum_{i=1}^n E\left(\frac{1}{T}\sum_{t=1}^T\xi_{it}\right)^{\lambda} = O\left(nT^{-\lambda/2}\upsilon_{nT}^{-\lambda}\right) = O(1).
$$

It follows that $\max_{1\le i\le n}|T^{-1}\sum_{t=1}^T\varepsilon_{it}^2-\sigma_i^2| = O_P(\upsilon_{nT})$. Similarly, $\max_{1\le i\le n}|\bar{\varepsilon}_i| = O_P(\upsilon_{nT}^2) = o_P(\upsilon_{nT})$. It follows that $\varepsilon_i' M\varepsilon_i/T = \sigma_i^2+O_P(\upsilon_{nT})$ uniformly in i. Then by the Cauchy-Schwarz inequality, we can readily show that the second term in (.0.89) is $O_P\left(n^{-1/2}T^{-1/2}h^{-1/2}\right) = o_P(\upsilon_{nT})$. Consequently, the first result in (.0.88) follows and $\max_{1\le i\le n}T^{-1}TSS_{i1} = O_P(1)$.

For TSS_{i2}, we have

$$\max_{1\le i\le n}\left\{T^{-1}TSS_{i2}\right\}\le C\left\|\widehat{\beta}-\beta\right\|^2\max_{1\le i\le n}\left\{T^{-1}\left\|X_i-S_TX\right\|^2\right\}=O_P\left(\frac{1}{nT}\right)O_P(\sqrt{\frac{n}{T}}+1),$$

where we use the fact that $\max_{1\le i\le n}T^{-1}\left\|X_i-S_TX\right\|^2=O_P(\sqrt{n/T}+1)$ under our moment conditions. For TSS_{i3}, noting that $||\bar{\mathbf{f}}^*||=\left\|\bar{\mathbf{f}}-S_T\bar{\mathbf{F}}\right\|=O\left(T^{1/2}h^{p+1}\right)$, we have $T^{-1}TSS_{i3}\le T^{-1}\left\|\bar{\mathbf{f}}-S_T\bar{\mathbf{F}}\right\|^2=O\left(h^{2p+2}\right)$. By the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
\max_{1\le i\le n}T^{-1}\left|TSS_{i4}\right| &\le \max_{1\le i\le n}\left(T^{-1}TSS_{i1}\right)^{1/2}\left(T^{-1}TSS_{i2}\right)^{1/2}\\
&= O_P\left(n^{-1/4}T^{-3/4}+n^{-1/2}T^{-1/2}\right)=o_P\left(\upsilon_{nT}\right),\\
\max_{1\le i\le n}T^{-1}\left|TSS_{i5}\right| &\le \max_{1\le i\le n}\left(T^{-1}TSS_{i1}\right)^{1/2}\left(T^{-1}TSS_{i3}\right)^{1/2}\\
&= O_P\left(h^{p+1}\right)=o_P\left(\upsilon_{nT}\right),\text{ and}\\
\max_{1\le i\le n}T^{-1}\left|TSS_{i6}\right| &\le \max_{1\le i\le n}\left(T^{-1}TSS_{i2}\right)^{1/2}\left(T^{-1}TSS_{i3}\right)^{1/2}=o_P\left(\upsilon_{nT}\right).
\end{aligned}$$

Consequently, we have $\max_{1\le i\le n}|\widehat{\sigma}_i^2-\sigma_i^2|=O_P\left(\upsilon_{nT}\right)$. Then by Assumption A5

$$\begin{aligned}
\Gamma_{nT,22} &\le \frac{\max_{1\le i\le n}|\widehat{\sigma}_i^2-\sigma_i^2|^2}{\min_{1\le i\le n}\sigma_i^4\widehat{\sigma}_i^2}\frac{b^{1/2}}{\sqrt{n}}\sum_{i=1}^n\left|ESS_i-\varepsilon_i'Q\varepsilon_i\right|\\
&\le \frac{\sqrt{n}\max_{1\le i\le n}|\widehat{\sigma}_i^2-\sigma_i^2|^2}{\min_{1\le i\le n}\sigma_i^4\widehat{\sigma}_i^2}\left(\frac{b}{n}\sum_{i=1}^n\left(ESS_i-\varepsilon_i'Q\varepsilon_i\right)^2\right)^{1/2}\\
&= \sqrt{n}O_P\left(\upsilon_{nT}^2\right)O_P\left(1\right)=O_P\left(n^{1/2+2/\lambda}T^{-1}\right)=o\left(1\right),
\end{aligned}$$

because one can easily show that $\frac{b}{n}\sum_{i=1}^n\left(ESS_i-\varepsilon_i'Q\varepsilon_i\right)^2=O_P\left(1\right)$.

For $\Gamma_{nT,21}$, we have $\Gamma_{nT,21}=\sum_{l=1}^6\Gamma_{nT,21l}$, where

$$\begin{aligned}
\Gamma_{nT,211} &\equiv \sqrt{\frac{b}{n}}\sum_{i=1}^n\sigma_i^{-4}\left(ESS_i-\varepsilon_i'Q\varepsilon_i\right)\left(T^{-1}TSS_{i1}-\sigma_i^2\right),\text{ and}\\
\Gamma_{nT,21l} &\equiv \sqrt{\frac{b}{n}}\sum_{i=1}^n\sigma_i^{-4}\left(ESS_i-\varepsilon_i'Q\varepsilon_i\right)\left(T^{-1}TSS_{il}\right)\text{ for }l=2,...,6.
\end{aligned}$$

Following the proof of Proposition .0.11 and the above analysis for TSS_{il}, we can show that $\Gamma_{nT,21l}=o_P\left(1\right)$ for $l=1,...,6$. ■

.0.12 Proof of Corollary 4.3.2

Given Theorem 4.3.1, it suffices to show that: (i) $\widehat{B}_{nT} = B_{nT} + o_P(1)$, and (ii) $\widehat{\Omega}_{nT} = \Omega_0 + o_P(1)$. We first prove (i). By (.0.82) and the fact that $Mi_T = 0$, we have

$$\widehat{u}_i' \bar{Q} \widehat{u}_i = \sum_{l=1}^{10} B_{nT,il}, \tag{.0.90}$$

where

$$\begin{aligned}
&B_{nT,i1} \equiv \varepsilon_i^{*\prime} \bar{Q} \varepsilon_i^{*}, && B_{nT,i2} \equiv (\widehat{\beta}-\beta)' X_i^{*\prime} \bar{Q} X_i^{*} (\widehat{\beta}-\beta), \\
&B_{nT,i3} \equiv \bar{\mathbf{f}}^{*\prime} \bar{Q} \bar{\mathbf{f}}^{*}, && B_{nT,i4} \equiv -2\varepsilon_i^{*\prime} \bar{Q} X_i^{*} (\widehat{\beta}-\beta), \\
&B_{nT,i5} \equiv 2\varepsilon_i^{*\prime} \bar{Q} \bar{\mathbf{f}}^{*}, && B_{nT,i6} \equiv -2\bar{\mathbf{f}}^{*\prime} \bar{Q} X_i^{*} (\widehat{\beta}-\beta), \\
&B_{nT,i7} \equiv 2\bar{\mathbf{f}}^{*\prime} \bar{Q} (\mathbf{f}_i - \bar{\mathbf{f}}) && B_{nT,i8} \equiv -2(\widehat{\beta}-\beta)' X_i^{*\prime} \bar{Q} (\mathbf{f}_i - \bar{\mathbf{f}}), \\
&B_{nT,i9} \equiv 2\varepsilon_i^{*\prime} \bar{Q} (\mathbf{f}_i - \bar{\mathbf{f}}), && B_{nT,i10} \equiv (\mathbf{f}_i - \bar{\mathbf{f}})' \bar{Q} (\mathbf{f}_i - \bar{\mathbf{f}}),
\end{aligned}$$

$\bar{Q} \equiv MQM$, and $\bar{\mathbf{f}}$ and $\bar{\mathbf{f}}^{*}$ are defined in (.0.81). Under H_0, we have $\mathbf{f}_i - \bar{\mathbf{f}} = 0$. Thus $B_{nT,il} = 0$ for $l = 7, \ldots, 10$. By (4.3.2) and (.0.90), it suffices to show that

$$\begin{aligned}
\mathscr{B}_{nT,1} &\equiv \sqrt{\frac{b}{n}} \sum_{i=1}^{n} \widehat{\sigma}_i^{-2} (B_{nT,i1} - B_{nT}) = \sqrt{\frac{b}{n}} \sum_{i=1}^{n} \widehat{\sigma}_i^{-2} \left[\varepsilon_i^{*\prime} \bar{Q} \varepsilon_i^{*} - \varepsilon_i' Q \varepsilon_i \right] = o_P(1) \\
B_{nT,l} &\equiv n^{-1/2} b^{1/2} \textstyle\sum_{i=1}^{n} \widehat{\sigma}_i^{-2} B_{nT,il} = o_P(1) \text{ for } l = 2, \ldots, 6.
\end{aligned}$$

Recalling $\varepsilon_i^{*} \equiv \varepsilon_i - S_T \varepsilon$, we decompose $\mathscr{B}_{nT,1}$ as follows

$$\begin{aligned}
\mathscr{B}_{nT,1} &= n^{-1/2} b^{1/2} \textstyle\sum_{i=1}^{n} \widehat{\sigma}_i^{-2} \left[(\varepsilon_i - S_T \varepsilon)' \bar{Q} (\varepsilon_i - S_T \varepsilon) - \varepsilon_i' Q \varepsilon_i \right] \\
&= n^{-1/2} b^{1/2} \textstyle\sum_{i=1}^{n} \widehat{\sigma}_i^{-2} \left[\varepsilon_i' \bar{Q} \varepsilon_i - \varepsilon_i' Q \varepsilon_i \right] - 2n^{-1/2} b^{1/2} \textstyle\sum_{i=1}^{n} \widehat{\sigma}_i^{-2} \varepsilon_i' \bar{Q} S_T \varepsilon \\
&\quad + n^{-1/2} b^{1/2} \textstyle\sum_{i=1}^{n} \widehat{\sigma}_i^{-2} (S_T \varepsilon)' \bar{Q} S_T \varepsilon \\
&\equiv \mathscr{B}_{nT,11} - 2\mathscr{B}_{nT,12} + \mathscr{B}_{nT,13}.
\end{aligned}$$

Noting that $\bar{Q} - Q = (I_T - L) Q (I_T - L) - Q = LQL - QL - LQ$ and both Q and L are symmetric, we have

$$\begin{aligned}
\mathscr{B}_{nT,11} &= n^{-1/2} b^{1/2} \textstyle\sum_{i=1}^{n} \widehat{\sigma}_i^{-2} \varepsilon_i' LQL \varepsilon_i - 2n^{-1/2} b^{1/2} \textstyle\sum_{i=1}^{n} \widehat{\sigma}_i^{-2} \varepsilon_i' QL \varepsilon_i \\
&\equiv \mathscr{B}_{nT,11a} - 2\mathscr{B}_{nT,11b}, \text{ say.}
\end{aligned}$$

Following the proof of Proposition .0.12, we can show that $\mathscr{B}_{nT,11a} = B_{nT,11a} + o_P(1)$, where $B_{nT,11a} = n^{-1/2}b^{1/2}\sum_{i=1}^{n}\sigma_i^{-2}\varepsilon_i' LQL\varepsilon_i$. Even though Q is not positive semidefinite (p.s.d.), it can be written as the difference between two p.s.d. matrices: $Q = Q^* - T^{-1}I_T$, where $Q^* =$diag$(\bar{H}_{11},...,\bar{H}_{TT})$. So we can write $B_{nT,11a} = n^{-1/2}b^{1/2}\sum_{i=1}^{n}\sigma_i^{-2}\varepsilon_i' LQ^*L\varepsilon_i - n^{-1/2}T^{-1}b^{1/2}\sum_{i=1}^{n}\sigma_i^{-2}\varepsilon_i' LL\varepsilon_i = B_{nT,11a1} - B_{nT,11a2}$. Noting that

$$
\begin{aligned}
E\left|B_{nT,11a1}\right| &= n^{-1/2}b^{1/2}\sum_{i=1}^{n}\sigma_i^{-2}E\left(\varepsilon_i' LQ^*L\varepsilon_i\right) = T^{-2}n^{-1/2}b^{1/2}\sum_{i=1}^{n}\sum_{t=1}^{T} i_T' Q^* i_T \\
&= O\left(T^{-1}n^{1/2}b^{1/2}\right)\mathrm{tr}\left(Q^*\right) = O\left(T^{-1}n^{1/2}b^{1/2}\right)O\left(b^{-1}\right) = o(1),
\end{aligned}
$$

and similarly $E\left|B_{nT,11a2}\right| = O\left(T^{-1}n^{1/2}b^{1/2}\right) = o(1)$, we have $\mathscr{B}_{nT,11a} = o_P(1)$ by the Markov inequality. Similarly, $\mathscr{B}_{nT,11b} = o_P(1)$. Consequently $\mathscr{B}_{nT,11} = o_P(1)$. Analogously, we can show that $\mathscr{B}_{nT,1l} = o_P(1)$ for $l = 2, 3$. It follows that $\mathscr{B}_{nT,1} = o_P(1)$.

Using the fact that $|\mathrm{tr}(AB)| \leq \lambda_{\max}(A)\mathrm{tr}(B)$ for any conformable p.s.d. matrix B and symmetric matrix A (see, e.g., Bernstein, 2005, p. 309) and that $\lambda_{\max}(M) = 1$, we can show that $\left\|X_i^{*\prime}\bar{Q}X_i^*\right\|^2 =tr(MQMX_i^*X_i^{*\prime}MQMX_i^*X_i^{*\prime}) \leq \|X_i^{*\prime}QX_i^*\|^2$. It follows that

$$
\begin{aligned}
B_{nT,2} &= n^{-1/2}b^{1/2}\sum_{i=1}^{n}\widehat{\sigma}_i^{-2}(\widehat{\beta}-\beta)X_i^{*\prime}\bar{Q}X_i^*(\widehat{\beta}-\beta) \\
&\leq n^{-1/2}b^{1/2}\left\|\widehat{\beta}-\beta\right\|^2\sum_{i=1}^{n}\widehat{\sigma}_i^{-2}\left\|X_i^{*\prime}QX_i^*\right\| \\
&= n^{-1/2}b^{1/2}O_P\left((nT)^{-1}\right)O_P\left(nb^{-1}\right) = O_P\left(n^{-1/2}T^{-1}b^{-1/2}\right) = o_P(1)
\end{aligned}
$$

where we use the fact that $\sum_{i=1}^{n}\widehat{\sigma}_i^{-2}\|X_i^{*\prime}QX_i^*\| = O_P\left(nb^{-1}\right)$. Similarly, we have

$$
\begin{aligned}
B_{nT,3} &= n^{-1/2}b^{1/2}\sum_{i=1}^{n}\widehat{\sigma}_i^{-2}\bar{\mathbf{f}}^{*\prime}\bar{Q}\bar{\mathbf{f}}^* \leq n^{-1/2}b^{1/2}\sum_{i=1}^{n}\widehat{\sigma}_i^{-2}\left\|\bar{\mathbf{f}}^{*\prime}Q\bar{\mathbf{f}}^*\right\| \\
&= n^{-1/2}b^{1/2}\left\|\sum_{t=1}^{T}\left(\bar{H}_{tt}-T^{-1}\right)\left[\overline{f}(t/T) - e_1'S(t/T)\mathbf{F}\right]^2\right\|^2\sum_{i=1}^{n}\widehat{\sigma}_i^{-2} \\
&= n^{-1/2}b^{1/2}O_P\left(b^{-1}h^{2p+2}\right)O_P(n) = O_P\left(n^{1/2}h^{2p+2}b^{-1/2}\right) = o_P(1).
\end{aligned}
$$

By the repeated use of the Cauchy-Schwarz inequality, we can show that $B_{nT,il} =$

$o_P(1)$ for $l = 4, 5$, and 6.

To show (ii), it suffices to show that $DV_{nT} \equiv n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}(\widehat{\rho}_{ij}^2 - \rho_{ij}^2) = o_P(1)$. Noting that $x^2 - y^2 = (x-y)^2 + 2(x-y)y$, we can decompose DV_{nT} as follows

$$DV_{nT} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}(\widehat{\rho}_{ij} - \rho_{ij})^2 + \frac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}(\widehat{\rho}_{ij} - \rho_{ij})\rho_{ij} \equiv DV_{nT1} + 2DV_{nT2}.$$

Following the argument in the proof of Proposition .0.12, we can show that

$$\begin{aligned} DV_{nT1} &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{\widehat{u}_i'M\widehat{u}_j}{\widehat{\sigma}_i\widehat{\sigma}_j} - \frac{\omega_{ij}}{\sigma_i\sigma_j}\right)^2 = \overline{DV}_{nT1} + o_P(1), \text{ and} \\ DV_{nT2} &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{\widehat{u}_i'M\widehat{u}_j}{\widehat{\sigma}_i\widehat{\sigma}_j} - \frac{\omega_{ij}}{\sigma_i\sigma_j}\right)\rho_{ij} = \overline{DV}_{nT2} + o_P(1). \end{aligned}$$

where $\overline{DV}_{nT1} \equiv n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_i^{-2}\sigma_j^{-2}\left(\widehat{u}_i'M\widehat{u}_j - \omega_{ij}\right)^2$ and $\overline{DV}_{nT2} \equiv n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}$ $\rho_{ij}\sigma_i^{-1}\sigma_j^{-1}(\widehat{u}_i'M\widehat{u}_j - \omega_{ij})$.

By (.0.82) and the fact that $Mi_T = 0$, we have that under H_0, $\widehat{u}_i'M\widehat{u}_j = \varepsilon_i^{*\prime}M\varepsilon_j^* + (\widehat{\beta} - \beta)'X_i^{*\prime}MX_j^*(\widehat{\beta} - \beta) + \bar{\mathbf{f}}^{*\prime}M\bar{\mathbf{f}}^* - (\varepsilon_i^{*\prime}MX_j^* + \varepsilon_j^{*\prime}MX_i^*)(\widehat{\beta} - \beta) + \left(\varepsilon_i^* + \varepsilon_j^*\right)'M\bar{\mathbf{f}}^* - \bar{\mathbf{f}}^{*\prime}M\left(X_i^* + X_j^*\right)(\widehat{\beta} - \beta) \equiv \sum_{l=1}^{6} DV_{nT,ijl}$. We can prove that $\overline{DV}_{nT1} = o_P(1)$ by showing that

$$\begin{aligned} \overline{DV}_{nT1,1} &\equiv \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_i^{-2}\sigma_j^{-2}\left(DV_{nT,ij1} - \omega_{ij}\right)^2 = o_P(1), \text{ and} \\ \overline{DV}_{nT1,l} &\equiv \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_i^{-2}\sigma_j^{-2}\left(DV_{nT,ijl}\right)^2 = o_P(1) \text{ for } l = 2, \ldots, 6. \end{aligned}$$

Similarly we can prove $\overline{DV}_{nT2} = o_P(1)$ by using the above decomposition for $\widehat{u}_i'M\widehat{u}_j$. The details are omitted for brevity.

.0.13 Proof of Theorem 4.3.3

By (4.3.2) we have

$$\begin{aligned}
\sqrt{\widehat{\Omega}_{nT}}\bar{\Gamma}_{nT} &= \frac{b^{1/2}}{n^{1/2}}\sum_{i=1}^{n}\widehat{\sigma}_i^{-2}\left(ESS_i-\widehat{u}_i'\bar{Q}\widehat{u}_i\right) \\
&= \sqrt{\frac{b}{n}}\sum_{i=1}^{n}\sigma_i^{-2}\left(ESS_i-\varepsilon_i'Q\varepsilon_i\right)-\sqrt{\frac{b}{n}}\sum_{i=1}^{n}\left(ESS_i-\varepsilon_i'Q\varepsilon_i\right)\left(\frac{1}{\widehat{\sigma}_i^2}-\frac{1}{\sigma_i^2}\right) \\
&\quad -\sqrt{\frac{b}{n}}\sum_{i=1}^{n}\sigma_i^{-2}\left(\widehat{u}_i'\bar{Q}\widehat{u}_i-\varepsilon_i'Q\varepsilon_i\right)+\sqrt{\frac{b}{n}}\sum_{i=1}^{n}\left(\widehat{u}_i'\bar{Q}\widehat{u}_i-\varepsilon_i'Q\varepsilon_i\right)\left(\frac{1}{\widehat{\sigma}_i^2}-\frac{1}{\sigma_i^2}\right) \\
&\equiv \Gamma_{nT,1}-\Gamma_{nT,2}-\Gamma_{nT,3}+\Gamma_{nT,4},\ \text{say,} \qquad (.0.91)
\end{aligned}$$

where $\Gamma_{nT,1}$ and $\Gamma_{nT,2}$ are as defined in the proof of Theorem 4.3.1, and $\widehat{\sigma}_i^2 \equiv TSS_i/T$. It is easy to show that $\widehat{\Omega}_{nT}=\Omega_0+o_P(1)$ under $H_1(\gamma_{nT})$ with $\gamma_{nT}=n^{-1/4}T^{-1/2}b^{-1/4}$. It suffices to show that: (i) $\Gamma_{nT,1}\xrightarrow{d}N(\Theta_0,\Omega_0)$, (ii) $\Gamma_{nT,2}=o_P(1)$, (iii) $\Gamma_{nT,3}=o_P(1)$, and (iv) $\Gamma_{nT,4}=o_P(1)$. We complete the proof by Propositions .0.13-.0.16 below.

Proposition .0.13 $\Gamma_{nT,1}\xrightarrow{d}N(\Theta_0,\Omega_0)$ *under* $H_1(\gamma_{nT})$.

Proof. Decompose $\Gamma_{nT,1}=\Gamma_{nT,11}-\Gamma_{nT,12}$ where $\Gamma_{nT,11}$ and $\Gamma_{nT,12}$ are defined in (.0.80). Using the notation defined in the proof of Proposition .0.11, it suffices to show: (i) $\mathscr{D}_{nT1}\equiv D_{nT1}-\Gamma_{nT,12}\xrightarrow{d}N(0,\Omega_0)$, (ii) $D_{nT2}=\Theta_0+o_P(1)$, and (iii) $D_{nTs}=o_P(1)$ for $s=3,...,10$, where $\Theta_0=\lim_{(n,T)\to\infty}\Theta_{nT}$ and $\Theta_{nT}\equiv n^{-1/2}b^{1/2}\gamma_{nT}^2\sum_{i=1}^{n}\sigma_i^{-2}\Delta_{ni}'(\bar{H}-L)\Delta_{ni}=n^{-1}T^{-1}\sum_{i=1}^{n}\sigma_i^{-2}\Delta_{ni}'(\bar{H}-L)\Delta_{ni}$. (i) follows the proof of Proposition .0.11. We are left to prove (ii) and (iii).

For (ii), letting ω_2 and $\mathbb{S}$ be as defined in the proof of Lemma .0.31, by (.0.95) we have

$$\begin{aligned}
D_{nT2} &= \gamma_{nT}^2\sqrt{\tfrac{b}{n}}\sum_{i=1}^{n}\sigma_i^{-2}\Delta_{ni}'(\bar{H}-L)\Delta_{ni} \\
&= \tfrac{1}{nT}\sum_{i=1}^{n}\sigma_i^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\bar{H}_{ts}-T^{-1}\right)\Delta_{ni}\left(\tfrac{t}{T}\right)\Delta_{ni}\left(\tfrac{s}{T}\right) \\
&= \tfrac{1}{nT^2}\sum_{i=1}^{n}\sigma_i^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left\{\int_0^1 w_{b,t}(\tau)\, z_{b,t}^{[1]}(\tau)'\mathbb{S}^{-1}z_{b,s}^{[1]}(\tau)\, w_{b,s}(\tau)\, d\tau\right. \\
&\quad \left.\times\left[\int_0^1 w_{b,t}(\tau)\, d\tau\int_0^1 w_{b,s}(\tau)\, d\tau\right]^{-1}-1\right\}\Delta_{ni}\left(\tfrac{t}{T}\right)\Delta_{ni}\left(\tfrac{s}{T}\right)+o(1) \\
&= \tfrac{1}{nT^2b}\sum_{i=1}^{n}\sigma_i^{-2}\sum_{t=\lfloor Tb\rfloor+1}^{\lfloor T(1-b)\rfloor-1}\sum_{s=1}^{T}\left\{\int_{-\frac{t}{Tb}}^{\frac{1}{b}-\frac{t}{Tb}}\left[1+\omega_2^{-1}u\left(u-\tfrac{s-t}{Tb}\right)\right]w(u)\, w\left(u-\tfrac{s-t}{Tb}\right)du\right. \\
&\quad \left.\times\left[\int_0^{1/b} w\left(z-\tfrac{t}{Tb}\right)dz\int_0^{1/b} w\left(\tfrac{s-t}{Tb}-\left(z'-\tfrac{t}{Tb}\right)\right)dz'\right]^{-1}-1\right\}\Delta_{ni}\left(\tfrac{t}{T}\right)\Delta_{ni}\left(\tfrac{s}{T}\right) \\
&\quad +o(1) \\
&= \tfrac{1}{nT}\sum_{i=1}^{n}\sigma_i^{-2}\sum_{t=\lfloor Tb\rfloor+1}^{\lfloor T(1-b)\rfloor-1}\int_{\frac{-t}{Tb}}^{\frac{T-t}{Tb}}\left\{\int_{-1}^{1}\left[1+\omega_2^{-1}u(u-v)\right]w(u)\, w(u-v)\, du\right. \\
&\quad \left.\times\left[\int_{\frac{-t}{Tb}}^{\frac{T-t}{Tb}} w(z)\, dz\int_{\frac{-t}{Tb}}^{\frac{T-t}{Tb}} w\left(z'-v\right)dz'\right]^{-1}-1\right\}\Delta_{ni}\left(\tfrac{t}{T}\right)\Delta_{ni}\left(\tfrac{t}{T}+vb\right)dv+o(1) \\
&= \frac{1}{n}\sum_{i=1}^{n}\sigma_i^{-2}\int_0^1\Delta_{ni}(\tau)^2\, d\tau\, C_w+o(1),
\end{aligned}$$

where $C_w\equiv\int_{-1}^{1}\left\{\int_{-1}^{1}\left[1+\omega_2^{-1}u(u-v)\right]w(u)\, w(u-v)\, du[\int_{-1}^{1}w(z-v)\, dz]^{-1}-1\right\}dv$. That is, $D_{nT2}=\Theta_{nT}=\Theta_0+o(1)$.

For (iii), following the proof of Proposition .0.11, we can show that $D_{nTl}=o_P(1)$ under $H_1(\gamma_{nT})$ for $l=3,...,7$. It suffices to prove (iii) by showing that $D_{nTl}=o_P(1)$ under $H_1(\gamma_{nT})$ for $l=8,...,10$. For D_{nT8}, write

$$\begin{aligned}
D_{nT8} &\equiv 2\sqrt{\frac{b}{n}}\sum_{i=1}^{n}\varepsilon_i'(\bar{H}-L)(\mathbf{f}_i-\bar{\mathbf{f}})/\sigma_i^2-2\sqrt{\frac{b}{n}}\sum_{i=1}^{n}(S_T\varepsilon)'(\bar{H}-L)(\mathbf{f}_i-\bar{\mathbf{f}})/\sigma_i^2 \\
&\equiv 2D_{nT8,1}-2D_{nT8,2}.
\end{aligned}$$

It is easy to show that

$$\begin{aligned} D_{nT8,1} &= (b/n)^{1/2} O_P(\gamma_{nT}(n^{1/2}T^{-1/2}b^{-1}+n^{1/2}T^{1/2})) \\ &= O_P(n^{-1/4}T^{-1}b^{-3/4}+n^{-1/4}b^{1/4}) = o_P(1), \end{aligned}$$

and $D_{nT8,2} = O_P(n^{-1/4}b^{1/4}\sqrt{\log(nT)}) = o_P(1)$. It follows that $D_{nT8} = o_P(1)$. By the Cauchy-Schwarz inequality, $D_{nTl} = o_P(1)$ for $l = 9, 10$. ■

Proposition .0.14 $\Gamma_{nT,2} = o_P(1)$ *under* $H_1(\gamma_{nT})$.

Proof. Analogously to the proof of Proposition .0.12, we can write

$$\begin{aligned} \Gamma_{nT,2} &= -\sqrt{\frac{b}{n}}\sum_{i=1}^{n}\left(ESS_i - \varepsilon_i' Q\varepsilon_i\right)\frac{\widehat{\sigma}_i^2-\sigma_i^2}{\sigma_i^4} + \sqrt{\frac{b}{n}}\sum_{i=1}^{n}\left(ESS_i - \varepsilon_i' Q\varepsilon_i\right)\frac{\left(\widehat{\sigma}_i^2-\sigma_i^2\right)^2}{\sigma_i^4\widehat{\sigma}_i^2} \\ &\equiv -\Gamma_{nT,21}+\Gamma_{nT,22}, \text{ say.} \end{aligned}$$

Note that $\widehat{\sigma}_i^2 = \sum_{l=1}^{10} TSS_{il}/T$ by (.0.87). First, we want to show that

$$\max_{1\le i\le n}\left|T^{-1}TSS_{i1}-\sigma_i^2\right| = O_P(\upsilon_{nT}) \text{ and } \max_{1\le i\le n} T^{-1}TSS_{il} = o_P(\upsilon_{nT}) \text{ for } l=2,\ldots,10, \tag{.0.92}$$

where $\upsilon_{nT}\equiv n^{1/\lambda}T^{-1/2}$. By (.0.88), it suffices to show that $\max_{1\le i\le n}T^{-1}TSS_{il} = o_P(\upsilon_{nT})$, for $l = 7,\ldots,10$. In the sequel, we will use the fact that $\max_{1\le i\le n}\sup_{\tau\in[0,1]}\left|f_i(\tau)-\overline{f}(\tau)\right| = O(\gamma_{nT})$ and $\widehat{\beta}-\beta = o_P(\gamma_{nT})$ under $H_1(\gamma_{nT})$ by Lemma .0.35(ii). Following the study of TSS_{i2} in Proposition .0.12, we can show that $\max_{1\le i\le n}T^{-1}TSS_{i7} = o_P(\upsilon_{nT})$. For TSS_{i8} we have

$$\begin{aligned} T^{-1}TSS_{i8} &= T^{-1}\gamma_{nT}^2\Delta_{ni}'M\Delta_{ni} \le T^{-1}\gamma_{nT}^2\left\|\Delta_{ni}\right\|^2 \\ &= n^{-1/2}T^{-2}b^{-1/2}\sum_{t=1}^{T}\Delta_{ni}^2\left(\frac{t}{T}\right) = O\left(n^{-1/2}T^{-1}b^{-1/2}\right) = o(v_{nT}) \end{aligned}$$

uniformly in i. By the Cauchy-Schwarz inequality, $\max_{1\le i\le n}T^{-1}TSS_{il} = o_P(\upsilon_{nT})$ for $l = 9, 10$. Consequently, we have $\max_{1\le i\le n}|\widehat{\sigma}_i^2-\sigma_i^2| = O_P(\upsilon_{nT})$. By the proof of Proposition .0.12, $\frac{b}{n}\sum_{i=1}^{n}(ESS_i-\varepsilon_i'Q\varepsilon_i)^2 = O_P(1)$. It follows that

$$\Gamma_{nT,22} \le \frac{n^{1/2}\max_{1\le i\le n}|\widehat{\sigma}_i^2-\sigma_i^2|^2}{\min_{1\le i\le n}\sigma_i^4\widehat{\sigma}_i^2}\left[\frac{b}{n}\sum_{i=1}^{n}\left(ESS_i-\varepsilon_i'Q\varepsilon_i\right)^2\right]^{1/2} = n^{1/2}O_P\left(\upsilon_{nT}^2\right) = o_P(1).$$

To analyze $\Gamma_{nT,21}$, using (.0.87) we can write

$$\Gamma_{nT,21} = \sqrt{\frac{b}{n}} \sum_{i=1}^{n} \left(ESS_i - \varepsilon_i' Q \varepsilon_i \right) \frac{\widehat{\sigma}_i^2 - \sigma_i^2}{\sigma_i^4} = \sum_{l=1}^{10} \Gamma_{nT,21l},$$

where $\Gamma_{nT,211} \equiv (b/n)^{1/2} \sum_{i=1}^{n} \sigma_i^{-4} (ESS_i - \varepsilon_i' Q \varepsilon_i)(T^{-1} TSS_{i1} - \sigma_i^2)$, and $\Gamma_{nT,21l} \equiv (b/n)^{1/2} \sum_{i=1}^{n} \sigma_i^{-4} \, (ESS_i - \varepsilon_i' Q \varepsilon_i) T^{-1} TSS_{il}$ for $l = 2, ..., 10$. Following the proof of Proposition .0.11 and the analysis for TSS_{il} in the proof of Corollary 4.3.2, we can show that $\Gamma_{nT,21l} = o_P(1)$ for $l = 1, ..., 10$. It follows that $\Gamma_{nT,21} = o_P(1)$. ■

Proposition .0.15 $\Gamma_{nT,3} = o_P(1)$ *under* $H_1(\gamma_{nT})$.

Proof. By the proof of Corollary 4.3.2, we can write

$$\Gamma_{nT,3} = \sqrt{\frac{b}{n}} \sum_{i=1}^{n} \sigma_i^{-2} \left(\widehat{u}_i' \bar{Q} \widehat{u}_i - \varepsilon_i' Q \varepsilon_i \right) = \sum_{l=1}^{10} \overline{B}_{nT,l}$$

where $\overline{B}_{nT1} = (b/n)^{1/2} \sum_{i=1}^{n} \sigma_i^{-2} (B_{nT,i1} - \varepsilon_i' Q \varepsilon_i)$, and $\overline{B}_{nTl} = (b/n)^{1/2} \sum_{i=1}^{n} \sigma_i^{-2} B_{nT,il}$ for $l = 2, ..., 10$. Following the argument in the proof of Corollary 4.3.2, we can readily show that $\overline{B}_{nTl} = o_P(1)$ for $l = 1, 2, ..., 6$ as in the case when H_0 holds. It remains to prove that $\overline{B}_{nTl} = o_P(1)$ for $l = 7, ..., 10$ under $H_1(\gamma_{nT})$. Noting that $\lambda_{\max}(M) = 1$, we have

$$\begin{aligned} \overline{B}_{nT10} &= \sqrt{\frac{b}{n}} \sum_{i=1}^{n} \sigma_i^{-2} (\mathbf{f}_i - \bar{\mathbf{f}})' \bar{Q} (\mathbf{f}_i - \bar{\mathbf{f}}) \leq \frac{b^{1/2} \gamma_{nT}^2}{\sqrt{n}} \sum_{i=1}^{n} \sigma_i^{-2} \Delta_{ni}' Q \Delta_{ni} \\ &= n^{-1} T^{-1} \sum_{i=1}^{n} \sigma_i^{-2} \sum_{t=1}^{T} \Delta_{ni}^2 (t/T) \left(\bar{H}_{tt} - T^{-1} \right) = O\left(T^{-1} b^{-1} \right) = o(1). \end{aligned}$$

By the Cauchy-Schwarz inequality, we have $\overline{B}_{nT7} = o(1)$ and $\overline{B}_{nT8} = o_P(1)$. Decompose $\overline{B}_{nT9} = 2n^{-1/2} b^{1/2} \sum_{i=1}^{n} \sigma_i^{-2} \varepsilon_i' \bar{Q} \; \bar{\mathbf{f}}_i^* - 2n^{-1/2} b^{1/2} \; \sum_{i=1}^{n} \sigma_i^{-2} (S_T \varepsilon)' \bar{Q} \; \bar{\mathbf{f}}_i^* \equiv 2\overline{B}_{nT9,1} - 2\overline{B}_{nT9,2}$. By moments calculation and the Chebyshev inequality, we can show that $\overline{B}_{nT9,1} = O_P\left(T^{1/2} h^{p+1} b^{1/2} \right) = o_P(1)$, and $\overline{B}_{nT9,2} = O_P\left(T^{1/2} h^{p+1} b^{1/2} \right) = o_P(1)$. Consequently $\overline{B}_{nT9} = o_P(1)$. ■

Proposition .0.16 $\Gamma_{nT,4} = o_P(1)$ *under* $H_1(\gamma_{nT})$.

Proof. Analogously to the proof of Proposition .0.12, we can write

$$\begin{aligned}\Gamma_{nT,4} &= -\sqrt{\frac{b}{n}}\sum_{i=1}^{n}\left(\widehat{u}_i'\bar{Q}\widehat{u}_i-\varepsilon_i'Q\varepsilon_i\right)\frac{\widehat{\sigma}_i^2-\sigma_i^2}{\sigma_i^4}+\sqrt{\frac{b}{n}}\sum_{i=1}^{n}\left(\widehat{u}_i'\bar{Q}\widehat{u}_i-\varepsilon_i'Q\varepsilon_i\right)\frac{\left(\widehat{\sigma}_i^2-\sigma_i^2\right)^2}{\sigma_i^4\widehat{\sigma}_i^2}\\ &\equiv -\Gamma_{nT,41}+\Gamma_{nT,42},\text{ say.}\end{aligned}$$

We prove the proposition by showing that $\Gamma_{nT,4l}=o_P(1)$ for $l=1,2$. For $\Gamma_{nT,41}$, write $\Gamma_{nT,41}=\sum_{l=1}^{10}\Gamma_{nT,41}(l)$, where

$$\begin{aligned}\Gamma_{nT,41}(1) &= \sqrt{\frac{b}{n}}\sum_{i=1}^{n}\sigma_i^{-4}\left(B_{nT,i1}-\varepsilon_i'Q\varepsilon_i\right)\left(\widehat{\sigma}_i^2-\sigma_i^2\right),\\ \Gamma_{nT,41}(l) &= \sqrt{\frac{b}{n}}\sum_{i=1}^{n}\sigma_i^{-4}B_{nT,il}\left(\widehat{\sigma}_i^2-\sigma_i^2\right)\text{ for } l=2,...,10,\end{aligned}$$

and $B_{nT,il}$ are defined after (.0.90). Further decompose $\Gamma_{nT,41}(1)=\sum_{m=1}^{10}\Gamma_{nT,41}(1,m)$ by using the decomposition $\widehat{\sigma}_i^2=\sum_{l=1}^{10}TSS_{il}/T$ in (.0.87), where $\Gamma_{nT,41}(1,1)=\sqrt{\frac{b}{n}}\sum_{i=1}^{n}\sigma_i^{-4}\left(B_{nT,i1}-\varepsilon_i'Q\varepsilon_i\right)\left(T^{-1}TSS_{i1}-\sigma_i^2\right)$ and $\Gamma_{nT,41}(1,m)=\sqrt{\frac{b}{n}}\sum_{i=1}^{n}\sigma_i^{-4}\frac{TSS_{im}}{T}$ $\times\left(B_{nT,i1}-\varepsilon_i'Q\varepsilon_i\right)$ for $m=2,...,10$. It is easy to show that $\Gamma_{nT,41}(1,m)=o_P(1)$ for $m=1,...,10$. Consequently $\Gamma_{nT,41}(1)=o_P(1)$. Similarly, we can show $\Gamma_{nT,41}(l)=(b/n)^{1/2}\sum_{i=1}^{n}\sigma_i^{-4}B_{nT,il}(\widehat{\sigma}_i^2-\sigma_i^2)$ for $l=2,...,10$ by using the decomposition of $\widehat{\sigma}_i^2$ in (.0.87). It follows that $\Gamma_{nT,41}=o_P(1)$.

For $\Gamma_{nT,42}$, we can apply the decomposition of $\widehat{u}_i'\bar{Q}\widehat{u}_i$ in (.0.90) to demonstrate that $(b/n)^{1/2}\sum_{i=1}^{n}|\widehat{u}_i'\bar{Q}\widehat{u}_i-\varepsilon_i'Q\varepsilon_i|=o_P\left(n^{1/2}\right)$. Then $\Gamma_{nT,42}=o_P\left(n^{1/2}\upsilon_{nT}^2\right)=o_P(\frac{n}{T})=o_P(1)$ by (.0.92). ■

.0.14 Proof of Theorem 4.3.4

As in the proof of Theorem 4.3.3, we have the decomposition

$$\sqrt{\widehat{\Omega}_{nT}}\Gamma_{nT}=\overline{\Gamma}_{nT1}-\overline{\Gamma}_{nT2}-\overline{\Gamma}_{nT3}+\overline{\Gamma}_{nT4}, \tag{.0.93}$$

where $\overline{\Gamma}_{nTl}$, $l=1, 2, 3, 4$, are defined analogously to Γ_{nTl} in (.0.91) with σ_i^2 being replaced by $\overline{\sigma}_i^2\equiv\sigma_i^2+\Upsilon_{i0}$, $\Upsilon_{i0}\equiv\int_0^1\Delta_i^2(\tau)d\tau-[\int_0^1\Delta_i(\tau)d\tau]^2$, and recall $\Delta_i(\tau)\equiv f_i(\tau)-f(\tau)$ under H_1. By (.0.87), $\widehat{\sigma}_i^2=T^{-1}\sum_{l=1}^{10}TSS_{il}$. Under H_1, by Lemma

.0.35(iii) the results in (.0.88) become

$$\max_{1\le i\le n}\left|T^{-1}TSS_{i1}-\sigma_i^2\right|=o_P(1)\text{ and }\max_{1\le i\le n}T^{-1}TSS_{il}=o_P(1)\text{ for }l=2,\ldots,6.$$

We can also show that $T^{-1}TSS_{il}=o_P(1)$ uniformly in i for $l=7$, 9, and 10. For TSS_{i8}, we have uniformly in i,

$$\frac{1}{T}TSS_{i8}=\frac{1}{T}\sum_{t=1}^{T}[\Delta_i(t/T)-\overline{\Delta}_i]^2=\int_0^1\Delta_i^2(\tau)\,d\tau-\left(\int_0^1\Delta_i(\tau)\,d\tau\right)^2+o(1)=\Upsilon_{i0}+o(1),$$

where $\overline{\Delta}_i\equiv T^{-1}\sum_{t=1}^{T}\Delta_i(t/T)$. It follows that uniformly in i

$$\widehat{\sigma}_i^2=\sigma_i^2+\Upsilon_{i0}+o_P(1)=\overline{\sigma}_i^2+o_P(1). \qquad (.0.94)$$

That is, $\overline{\sigma}_i^2$ is the probability limit of $\widehat{\sigma}_i^2$ under H_1. We prove the theorem by showing that (i) $\Lambda_{nT1}\equiv(n^{1/2}Tb^{1/2})^{-1}\overline{\Gamma}_{nT1}=\Xi_A+o_P(1)$, and (ii) $\Lambda_{nTl}\equiv(n^{1/2}Tb^{1/2})^{-1}\overline{\Gamma}_{nTl}=o_P(1)$ for $l=2,3,4$.

Following the proof of Propositions .0.11 and .0.13, we can show that $\Lambda_{nT1}=\left(n^{1/2}Tb^{1/2}\right)^{-1}\overline{\Gamma}_{nT1}=\overline{\Lambda}_{nT1}+o_P(1)$, where $\overline{\Lambda}_{nT1}\equiv(n^{1/2}Tb^{1/2})^{-1}D_{nT2}$. Following the analysis of D_{nT2} in the proof of Proposition .0.13, we have

$$\overline{\Lambda}_{nT1}=\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T}(\bar{H}_{ts}-T^{-1})\Delta_i(t/T)\Delta_i(s/T)/\overline{\sigma}_i^2=\Theta_A+o(1),$$

where Θ_A is defined analogously to Θ_0 with (σ_i^2,Δ_{ni}) being replaced by $(\overline{\sigma}_i^2,\Delta_i)$. This proves (i). Following the proof of Propositions .0.12 and .0.14-.0.16, we can show that $\Lambda_{nTl}=o_P(1)$ for $l=2,3,4$.

.0.15 Some Useful Lemmas

In this Appendix, we present some technical lemmas that are used in the proofs of the main results in the chapter.

Lemma .0.30 *Let $\lambda_{tT}\equiv\int_0^1 w_b\left(\frac{t}{T}-\tau\right)d\tau$. Then $\frac{1}{2}\le\min_{1\le t\le T}\lambda_{tT}\le\max_{1\le t\le T}\lambda_{tT}=1$.*

Proof. First, write $\lambda_{tT}=\int_0^1 w\left(\frac{\tau}{b}-\frac{t}{Tb}\right)d\left(\frac{\tau}{b}\right)=\int_0^{1/b}w\left(u-\frac{t}{Tb}\right)du=\int_{\frac{-t}{Tb}}^{\frac{1}{b}-\frac{t}{Tb}}w(u)\,du.$ Clearly, $\max_{1\le t\le T}\lambda_{tT}=1$. If $Tb\le t\le T(1-b)$, then $\lambda_{tT}=\int_{-1}^{1}w(u)\,du=1$. If

$1 \leq t = T\varepsilon < Tb$ for some $\varepsilon \in (0,b)$, then

$$\lambda_{tT} = \int_{-t/(Tb)}^{1/b-t/(Tb)} w(s)\,ds = \int_{-\varepsilon}^{1} w(u)\,du \geq \int_{0}^{1} w(u)\,du = \frac{1}{2}$$

where the last equality follows from the symmetry of w and the fact that $\int_{-1}^{1} w(u)\,du = 1$. Similarly, if $T(1-b) < t = T\varepsilon \leq T$ for some $\varepsilon \in (1-b,1)$, then we have $\int_0^1 w_b\left(\frac{t}{T}-\tau\right)d\tau = \int_{-t/(Tb)}^{1/b-t/(Tb)} w(u)\,du = \int_{-1}^{\varepsilon} w(u)\,du \geq \int_{-1}^{0} w(u)\,du = \frac{1}{2}$. This proves the lemma. ■

Lemma .0.31 $\max_{1\leq t,s\leq T}|\bar{H}_{ts}| \leq C_1 (Tb)^{-1}$ *for some constant* $C_1 < \infty$ *where* $\bar{H}_{ts}$ *denote the* (t,s)*th element of* $\bar{H}$, $\bar{H} \equiv \int_0^1 H(\tau)\,d\tau$, *and*

$$H(\tau) \equiv W_b(\tau)\, z_b^{[1]}(\tau)\left(z_b^{[1]}(\tau)' W_b(\tau)\, z_b^{[1]}(\tau)\right)^{-1} z_b^{[1]}(\tau)' W_b(\tau).$$

Proof. Let $S_b(\tau) \equiv T^{-1} z_b^{[1]}(\tau)' W_b(\tau)\, z_b^{[1]}(\tau)$. Then

$$S_b(\tau) = \mathbb{S} + o(1) \text{ uniformly in } \tau \in (0,1), \tag{.0.95}$$

where $\mathbb{S} \equiv \begin{pmatrix} 1 & 0 \\ 0 & \omega_2 \end{pmatrix}$ and $\omega_2 = \int_{-1}^{1} w(u)\,u^2 du$. By (.0.95), Lemma .0.30, and Assumption A4, we have

$$\begin{aligned}
|\bar{H}_{ts}| &= \left|T^{-1}\int_0^1 z_{b,t}^{[1]}(\tau)' [S_b(\tau)]^{-1} z_{b,s}^{[1]}(\tau)\,\overline{w}_{b,t}(\tau)\,\overline{w}_{b,s}(\tau)\,d\tau\right| \\
&\approx \left|T^{-1}\int_0^1 z_{b,t}^{[1]}(\tau)' \mathbb{S}^{-1} z_{b,s}^{[1]}(\tau)\,\overline{w}_{b,t}(\tau)\,\overline{w}_{b,s}(\tau)\,d\tau\right| \\
&\leq \left|T^{-1}\int_0^1 w_b\left(\frac{t}{T}-\tau\right) w_b\left(\frac{s}{T}-\tau\right) d\tau (\lambda_{tT}\lambda_{sT})^{-1}\right| \\
&\quad + \left|\omega_2^{-1} T^{-1}\int_0^1 \left(\frac{t-\tau T}{Tb}\right)\left(\frac{s-\tau T}{Tb}\right) w_b\left(\frac{t}{T}-\tau\right) w_b\left(\frac{s}{T}-\tau\right) d\tau (\lambda_{tT}\lambda_{sT})^{-1}\right| \\
&\leq C(Tb)^{-1}\int_0^1 w_b\left(\frac{t}{T}-\tau\right) d\tau/\lambda_{tT} + C(Tb)^{-1}\int \frac{|t-\tau T|}{Tb} w_b\left(\frac{t}{T}-\tau\right) d\tau \\
&\leq C(Tb)^{-1}\left(1+\int_{-1}^{1}|u|\,w(u)\,d\tau\right) \leq C_1 (Tb)^{-1},
\end{aligned}$$

where $A \approx B$ denotes $A = B(1+o(1))$. ■

Lemma .0.32 *(i)* $A_{T1} \equiv b\sum_{1\leq t\neq s\leq T} a_{ts}^2 = O(1)$, *(ii)* $A_{T2} \equiv T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}|a_{ts}a_{tr}| = O(1)$, *and (iii)* $A_{T3} \equiv \|\bar{H}-L\| = O\left(b^{-1/2}\right)$, *where recall* $a_{ts} \equiv \bar{H}_{ts} - T^{-1}$ *denotes the* (t,s)*th element of* $\bar{H}-L$, *and* $L \equiv T^{-1} i_T i_T'$.

Proof. For (i) it is easy to show that $A_{T1} = \overline{A}_{T1} + O(b)$, where $\overline{A}_{T1} \equiv b\sum_{1\le t\neq s\le T} \bar{H}_{ts}^2$. By (.0.95),

$$
\begin{aligned}
&\overline{A}_{T1} \\
\approx\ & \tfrac{b}{T^2} \sum_{1\le t\neq s\le T} \left\{ \int_0^1 z_{b,t}^{[1]}(\tau)\mathbb{S}^{-1} z_{b,s}^{[1]}(\tau)\overline{w}_{b,t}(\tau)\overline{w}_{b,s}(\tau)d\tau \right\}^2 \\
=\ & \tfrac{b}{T^2} \sum_{1\le t\neq s\le T} \left\{ \int_0^1 \left[1+\omega_2^{-1}\left(\tfrac{T\tau-t}{Tb}\right)\left(\tfrac{T\tau-s}{Tb}\right)\right] \frac{1}{b^2} w\left(\tfrac{T\tau-s}{Tb}\right) w\left(\tfrac{T\tau-t}{Tb}\right) d\tau \right\}^2 \lambda_{tT}^{-1}\lambda_{sT}^{-1} \\
=\ & \tfrac{b}{T^2} \sum_{1\le t\neq s\le T} \left\{ \int_{-t/(Tb)}^{1/b-t/(Tb)} \left[1+\omega_2^{-1}u\left(u+\tfrac{t\text{-}s}{Tb}\right)\right] \frac{1}{b} w(u) w\left(u+\tfrac{t\text{-}s}{Tb}\right) du \right\}^2 (\lambda_{tT}\lambda_{sT})^{-2} \\
=\ & \tfrac{b}{T^2} \sum_{t=\lfloor Tb\rfloor+1}^{\lfloor T(1-b)\rfloor-1} \sum_{s=1}^{T} \left\{ \int_{-1}^1 \left[1+\omega_2^{-1}u\left(u+\tfrac{t\text{-}s}{Tb}\right)\right] \frac{1}{b} w(u) w\left(u+\tfrac{t\text{-}s}{Tb}\right) du \right\}^2 \\
& \times \left\{ \int_0^{1/b} w\left(z\text{-}\tfrac{t}{Tb}\right) dz \int_0^{1/b} w\left(\tfrac{s\text{-}t}{Tb}\text{-}(z'\text{-}\tfrac{t}{Tb})\right) dz' \right\}^{-2} + O(b) \\
=\ & \frac{1}{T} \sum_{t=\lfloor Tb\rfloor+1}^{\lfloor T(1-b)\rfloor-1} \int_{-t/(Tb)}^{(T-t)/(Tb)} \left(\int_{-1}^1 \left[1+\omega_2^{-1}u(u\text{-}v)\right] w(u) w(u\text{-}v) du \right)^2 \\
& \times \left(\int_{-t/(Tb)}^{1/b-t/(Tb)} w(z) dz \int_{-t/(Tb)}^{1/b-t/(Tb)} w\left(z'\text{-}v\right) dz' \right)^{-2} dv + o(1) \\
=\ & \int_b^{1-b} \int_{-1}^1 \left\{ \left(\int_{-1}^1 \left[1+\omega_2^{-1}u(u\text{-}v)\right] w(u) w(u\text{-}v) du \right)^2 \right. \\
& \qquad \left. \times \left(\int_{-1}^1 w(z) dz \int_{-1}^1 w\left(z'\text{-}v\right) dz' \right)^{-2} \right\} dv dv' + o(1) \\
=\ & \int_{-1}^1 \left(\int_{-1}^1 \left[1+\omega_2^{-1}u(u\text{-}v)\right] w(u) w(u\text{-}v) du \right)^2 \left(\int_{-1}^1 w(z\text{-}v) dz \right)^{-2} dv + o(1) = O(1).
\end{aligned}
$$

By the same token, we can show (ii). For (iii), noting that $\|\bar{H}-L\|^2 = \sum_{1\le t\neq s\le T} a_{ts}^2 + \sum_{t=1}^T a_{tt}^2 = O(b^{-1}) + O(T^{-1}b^{-2})$, $\|\bar{H}-L\| = O\left(b^{-1/2}\right)$ as $T^{-1}b^{-1} = o(1)$. ■

Lemma .0.33 *Let $c_{ts} \equiv e_1'[T^{-1}z_h^{[p]}(t/T)'K_h(t/T)z_h^{[p]}(t/T)]^{-1}z_{h,s}^{[p]}(t/T)$. Then (i) $C_{T1} \equiv T^{-2}\sum_{1\le t\neq s\le T} |c_{ts}|k_{h,ts} = O(1)$; (ii) $C_{T2} \equiv T^{-2}h\sum_{1\le t\neq s\le T} c_{ts}^2 k_{h,ts}^2 = O(1)$, (iii) $C_{T3} \equiv T^{-1}\sum_{t=1}^T |c_{tt}| = O(1)$; (iv) $C_{T4} \equiv T^{-1}\sum_{t=1}^T c_{tt}^2 = O(1)$.*

Proof. (i) Let $S_{p,h}(\tau) \equiv T^{-1}z_h^{[p]}(t/T)'K_h(t/T)z_h^{[p]}(t/T)$. The (j,l)th element of $S_{p,h}(\tau)$ is $s_{jl}(\tau) = \frac{1}{Th}\sum_{s=1}^T \left(\frac{s-\tau T}{Th}\right)^{j+l-2} k\left(\frac{s-\tau T}{Th}\right)$. For any $\tau \in (0,1)$, we have by the definition of Riemann integral that

$$\begin{aligned} s_{jl}(\tau) &= \frac{1}{Th}\sum_{r=1}^{T}\left(\frac{r}{Th}-\frac{\tau}{h}\right)^{j+l-2}k\left(\frac{r}{Th}-\frac{\tau}{h}\right)=\int_{-\tau/(Th)}^{1/h-\tau/(Th)}u^{j+l-2}k(u)\,du+o(1) \\ &= \int_{-1}^{1}u^{j+l-2}k(u)\,du+o(1). \end{aligned}$$

That is, $S_{p,h}(\tau)=\mathbb{S}_p+o(1)$ for any $\tau\in(0,1)$, where

$$\mathbb{S}_p=\begin{pmatrix} \mu_0 & \mu_1 & \cdots & \mu_p \\ \mu_1 & \mu_2 & \cdots & \mu_{p+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_p & \mu_{p+1} & \cdots & \mu_{2p} \end{pmatrix},$$

and $\mu_j\equiv\int_{-1}^{1}v^jk(v)\,dv$ for $j=0,1,...,2p$. It follows that

$$\begin{aligned} C_{T1} &= \frac{1}{T^2h}\sum_{t=1}^{T}\sum_{s=1}^{T}\left|e_1'\mathbb{S}_p^{-1}\left[1,\frac{s-t}{Th},...,\left(\frac{s-t}{Th}\right)^p\right]\right|k\left(\frac{s-t}{Th}\right)+o(1) \\ &= \frac{1}{T}\sum_{t=1}^{T}\int_{-t/(Th)}^{(T-t)/(Th)}\left|e_1'\mathbb{S}_p^{-1}[1,v,...,v^p]\right|k(v)\,dv+o(1) \\ &= \frac{1}{T}\sum_{t=\lfloor Th\rfloor+1}^{\lfloor T(1-h)\rfloor-1}\int_{-t/(Th)}^{(T-t)/(Th)}\left|e_1'\mathbb{S}_p^{-1}[1,v,...,v^p]\right|k(v)\,dv+o(1) \\ &= \int_{-1}^{1}\left|e_1'\mathbb{S}_p^{-1}[1,v,...,v^p]\right|k(v)\,dv+o(1)=O(1). \end{aligned}$$

This proves (i). By the same token,

$$\begin{aligned} C_{T2} &= \frac{1}{T^2h}\sum_{t=1}^{T}\sum_{s=1}^{T}\left|e_1'\mathbb{S}_p^{-1}\left[1,\frac{s-t}{Th},...,\left(\frac{s-t}{Th}\right)^p\right]\right|^2k\left(\frac{s-t}{Th}\right)^2+o(1) \\ &= \int_{-1}^{1}\left|e_1'\mathbb{S}_p^{-1}[1,v,...,v^p]\right|^2k(v)^2\,dv+o(1)=O(1). \end{aligned}$$

Similarly, we can prove (iii)-(iv). ■

Lemma .0.34 $\sup_{\tau\in(0,1)}e_1'S(\tau)\varepsilon=O_P\left(\sqrt{\log(nT)/(nTh)}\right)$.

Proof. The proof is analogous to that of (A.11) in Chen, Gao, and Li (2010, pp. 27-30). ■

Lemma .0.35 *Suppose Assumptions A1-A5 hold. Recall* $\gamma_{nT}=n^{-1/4}T^{-1/2}b^{-1/2}$ *in* $H_1(\gamma_{nT})$. *Then as* $(n,T)\to\infty$,

(i) $\widehat{\beta}-\beta=O_P\left(n^{-1/2}T^{-1/2}\right)$ *under* H_0;
(ii) $\widehat{\beta}-\beta=o_P(\gamma_{nT})$ *under* $H_1(\gamma_{nT})$ *provided that A6 also holds;*
(iii) $\widehat{\beta}-\beta=o_P(1)$ *under* H_1 *provided that A6 also holds.*

Proof. (i) This can be done by following the proof of Theorem 3.1 in CGL (2010). Note that CGL also proves the asymptotic normality under the independence of $\{(\varepsilon_{it},v_{it})\}$ across t and the assumption that g_i in Assumption A1 is the same for all i ($g_i=g$, say). One can verify that the above probability order can be attained even if we relax their independence condition to our m.d.s. condition and their homogenous trending assumption on g to our heterogeneous case.

(ii) Recalling that $\overline{\mathbf{F}}\equiv i_n\otimes\bar{\mathbf{f}}$ and $S_{nT}\mathbf{F}=S_{nT}\overline{\mathbf{F}}$, we have

$$\widehat{\beta}-\beta=\left(X^{*\prime}M_DX^*\right)^{-1}X^{*\prime}M_D(\varepsilon^*+\overline{\mathbf{F}}^*)+\left(X^{*\prime}M_DX^*\right)^{-1}X^{*\prime}M_D(\mathbf{F}-\overline{\mathbf{F}})\equiv d_1+d_2,\text{ say.}\tag{.0.96}$$

The first term also appears under H_0 and thus $d_1=O_P\left(n^{-1/2}T^{-1/2}\right)$. The second term vanishes under H_0 and plays asymptotically non-negligible role under $H_1(\gamma_{nT})$. Let $\overline{d}_2\equiv X^{*\prime}M_D(\mathbf{F}-\overline{\mathbf{F}})$. Note that

$$\overline{d}_2=X^{*\prime}(\mathbf{F}-\overline{\mathbf{F}})-X^{*\prime}D\left(D'D\right)^{-1}D(\mathbf{F}-\overline{\mathbf{F}}).\tag{.0.97}$$

Similarly to the proof in CGL (2010), we can show that the leading term on the right hand side of the above equation is $X^{*\prime}(\mathbf{F}-\overline{\mathbf{F}})$. Noting that $X_{it}=g_i(t/T)+v_{it}$ and $X^*=(I-S_{nT})X$, we have

$$\begin{aligned}
& X^{*\prime}(\mathbf{F}-\overline{\mathbf{F}})\\
= & \sum_{i=1}^{n}\sum_{t=1}^{T}\left[X_{it}-e_1'S(t/T)X\right]\left[f_i(t/T)-\overline{f}(t/T)\right]\\
= & \sum_{i=1}^{n}\sum_{t=1}^{T}v_{it}\left[f_i(t/T)-\overline{f}(t/T)\right]-\sum_{i=1}^{n}\sum_{t=1}^{T}\left\{e_1'S(t/T)V\right\}\left[f_i(t/T)-\overline{f}(t/T)\right]\\
& +\sum_{i=1}^{n}\sum_{t=1}^{T}\left[g_i(t/T)-\overline{g}(t/T)\right]\left[f_i(t/T)-\overline{f}(t/T)\right]\\
& +\sum_{i=1}^{n}\sum_{t=1}^{T}\left[\overline{g}(t/T)-e_1'S(t/T)\mathbf{G}\right]\left[f_i(t/T)-\overline{f}(t/T)\right]\\
\equiv & \Psi_{nT1}-\Psi_{nT2}+\Psi_{nT3}+\Psi_{nT4},
\end{aligned}\tag{.0.98}$$

where $V \equiv \left(v_{11}', ..., v_{1T}', ..., v_{n1}', ..., v_{nT}'\right)'$, $\overline{g}(\frac{t}{T}) \equiv n^{-1}\sum_{i=1}^{n} g_i(\frac{t}{T})$, $\mathbf{g}_i \equiv (g_i(\frac{1}{T})', ..., g_i(\frac{T}{T})')'$ and $\mathbf{G} \equiv (\mathbf{g}_1', ..., \mathbf{g}_n')'$. Clearly $\Psi_{nTl} = 0$ for $l = 2,4$ by the definition of $\overline{f}$. Noting that $\max_{1\leq i\leq n}\sup_{0\leq\tau\leq 1}\left|f_i(\tau) - \overline{f}(\tau)\right| = O(\gamma_{nT})$, we have

$$\begin{aligned} E\|\Psi_{nT1}\|^2 &= \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T} E\left(v_{it}'v_{jt}\right)\left[f_i(t/T) - \overline{f}(t/T)\right]\left[f_j(t/T) - \overline{f}(t/T)\right] \\ &\leq \left(\max_{1\leq i\leq n}\sup_{0\leq\tau\leq 1}\left|f_i(\tau) - \overline{f}(\tau)\right|\right)^2 \left(T\sum_{i=1}^{n}\sum_{j=1}^{n}\left|E\left(v_{i1}'v_{j1}\right)\right|\right) \\ &= O\left(\gamma_{nT}^2\right)O(nT) = o(nT), \end{aligned}$$

implying that $\Psi_{nT1} = o_P(\sqrt{nT})$. For Ψ_{nT3}, we have

$$\begin{aligned} |\Psi_{nT3}| &\leq \max_{1\leq i\leq n}\sup_{0\leq\tau\leq 1}\left|f_i(\tau) - \overline{f}(\tau)\right|\sum_{i=1}^{n}\sum_{t=1}^{T}|g_i(t/T) - \overline{g}(t/T)| \\ &= O(\gamma_{nT})T\sum_{i=1}^{n}\left(\int_0^1 |g_i(\tau) - \overline{g}(\tau)|d\tau + O(1/T)\right) \\ &= O(\gamma_{nT})o(nT) = o(\gamma_{nT}nT). \end{aligned}$$

Consequently, we have shown that $X^{*\prime}(\mathbf{F} - \overline{\mathbf{F}}) = O_P(\sqrt{nT}) + o(\gamma_{nT}nT)$. It follows $X^{*\prime}M_D(\mathbf{F} - \overline{\mathbf{F}}) = O_P(\sqrt{nT})$. Noting that $(nT)^{-1}X^{*\prime}M_D X^* = O_P(1)$, we have $(X^{*\prime}M_D X^*)^{-1}X^{*\prime}M_D(\mathbf{F} - \overline{\mathbf{F}}) = o_P(\gamma_{nT})$. Thus $\widehat{\beta} - \beta = o_P(\gamma_{nT})$ under $H_1(\gamma_{nT})$.

(iii) Using the notation above, we continue to have $d_1 = O_P(n^{-1/2}T^{-1/2})$ and $(nT)^{-1}X^{*\prime}M_D X^* = O_P(1)$ under H_1. For $\overline{d}_2$, we analyze the dominant term $X^{*\prime}(\mathbf{F} - \overline{\mathbf{F}})$ by using the same decomposition in (.0.98). Clearly, we still have $\Psi_{nT2} = 0$, $\Psi_{nT3} = o_P(nT)$ and $\Psi_{nT4} = 0$. For Ψ_{nT1}, since $\max_{1\leq i\leq n}\sup_{0\leq\tau\leq 1}\left|f_i(\tau) - \overline{f}(\tau)\right| = O(1)$ under H_1, we have $E(\|\Psi_{nT1}\|^2) = O(nT)$, which implies that $\Psi_{nT1} = O_P(\sqrt{nT})$. Thus $X^{*\prime}(\mathbf{F} - \overline{\mathbf{F}}) = o_P(nT)$ and $\widehat{\beta} - \beta = o_P(1)$ under H_1. ■

Remark. If $g_i(\tau) - \overline{g}(\tau) = 0$ for all $\tau \in [0,1]$, then from the proof of (ii) and (iii) we can see that $\widehat{\beta} - \beta = O_P\left(n^{-1/2}T^{-1/2}\right)$ also holds under $H_1(\gamma_{nT})$ and $H_1(1)$ as $\Psi_{nT3} = 0$ in this case.

Lemma .0.36 $\|X - S_{nT}X\|^2 = O_P(nT)$.

Proof. Recall $\mathbf{g}_i \equiv (g_i(1/T), ..., g_i(T/T))'$ and $\mathbf{G} \equiv (\mathbf{g}_1', ..., \mathbf{g}_n')'$. Noting that $X_{it} = g_i(t/T) + v_{it}$, we have

$$
\begin{aligned}
& \|X - S_{nT}X\|^2 \\
= \; & \sum_{i=1}^{n}\sum_{t=1}^{T} \|X_{it} - e_1 S(t/T) X\|^2 \\
= \; & \sum_{i=1}^{n}\sum_{t=1}^{T} \|v_{it} - e_1 S(t/T) V + [g_i(t/T) - \overline{g}(t/T)] + [\overline{g}(t/T) - e_1 S(t/T)\mathbf{G}]\|^2 \\
= \; & \sum_{i=1}^{n}\sum_{t=1}^{T} v_{it}'v_{it} + \sum_{i=1}^{n}\sum_{t=1}^{T} \|e_1 S(t/T) V\|^2 + \sum_{i=1}^{n}\sum_{t=1}^{T} \|g_i(t/T) - \overline{g}(t/T)\|^2 \\
& + \sum_{i=1}^{n}\sum_{t=1}^{T} \|\overline{g}(t/T) - e_1 S(t/T)\mathbf{G}\|^2 + 2\sum_{i=1}^{n}\sum_{t=1}^{T} v_{it}' e_1 S(t/T) V \\
& + 2\sum_{i=1}^{n}\sum_{t=1}^{T} v_{it}'(g_i(t/T) - \overline{g}(t/T)) + 2\sum_{i=1}^{n}\sum_{t=1}^{T} v_{it}'(\overline{g}(t/T) - e_1 S(t/T)\mathbf{G}) \\
& + 2\sum_{i=1}^{n}\sum_{t=1}^{T} (e_1 S(t/T) V)'(\overline{g}(t/T) - e_1 S(t/T)\mathbf{G}) \\
& + 2\sum_{i=1}^{n}\sum_{t=1}^{T} (e_1 S(t/T) V)'(g_i(t/T) - \overline{g}(t/T)) \\
& + 2\sum_{i=1}^{n}\sum_{t=1}^{T} (g_i(t/T) - \overline{g}(t/T))'(\overline{g}(t/T) - e_1 S(t/T)\mathbf{G}) \\
\equiv \; & \sum_{r=1}^{10} \Pi_{nT,r}, \text{ say.}
\end{aligned}
$$

It is easy to show that: $\Pi_{nT,1}=O_P(nT)$ by the Markov inequality, $\Pi_{nT,2}=O_P(\frac{nT\log(nT)}{nTh})$ $= o_P(nT)$, $\Pi_{nT,3} = O(nT)$ by the property of Riemann integral, $\Pi_{nT,4} = O\left(nTh^{2p+2}\right)$ $= o(nT)$ by the Taylor expansion. For the remaining terms, it is clear that $\Pi_{nT,r} = 0$ for $r = 9, 10$, and we can show that $\sum_{r=6}^{8} \Pi_{nT,r} = O_P(nT)$ by the Cauchy-Schwarz inequality. ■