2010

Nonparametric Tests for Poolability in Panel Data Models with Cross Section Dependence

Sainan JIN
Singapore Management University, snjin@smu.edu.sg

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

A Nonparametric Poolability Test for Panel Data Models with Cross Section Dependence*

Sainan Jin[a], Liangjun Su[a]
[a]School of Economics, Singapore Management University, Singapore

August 8, 2010

Abstract

In this paper we propose a nonparametric test for poolability in large dimensional semiparametric panel data models with cross-section dependence based on the sieve estimation technique. To construct the test statistic, we only need to estimate the model under the alternative. We establish the asymptotic normal distributions of our test statistic under the null hypothesis of poolability and a sequence of local alternatives, and prove the consistency of our test. We also suggest a bootstrap method as an alternative way to obtain the critical values and justify its validity. A small set of Monte Carlo simulations indicate the test performs reasonably well in finite samples.

JEL Classifications: C13, C14, C33

Key Words: Common factor; Cross-section dependence; Poolability; Semiparametric panel data model; Sieve estimation; Test

1 Introduction

Recently there has been a growing interest in the estimation of panel data models with cross-section dependence. See Bai (2003, 2009), Greenaway-McGrevy, Han, and Sul (2009), Harding (2007), Kapetanios and Pesaran (2005), Pesaran (2006), Pesaran and Tosetti (2007), Phillips and Sul (2003, 2007), among others, for an overview. All of these papers focus on the linear specification of regression relationship. More recently, Su and Jin (2009, SJ hereafter)

*Address Correspondence to: Sainan Jin, School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903; E-mail: snjin@smu.edu.sg, Phone: +65 6828 0431. The first author gratefully acknowledges financial support from the SMU research grant (Grant number: C244/MSS8E002).

have extended the linear model of Pesaran (2006) to the following semiparametric panel data model with multi-factor error structure

$$y_{it} = g_i(x_{it}) + \gamma_{1i}' f_{1t} + e_{it},\ e_{it} = \gamma_{2i}' f_{2t} + \varepsilon_{it}, \tag{1.1}$$

$$x_{it} = \Gamma_{1i}' f_{1t} + \Gamma_{2i}' f_{2t} + v_{it}, i = 1, \cdots, n,\ t = 1, \cdots, T \tag{1.2}$$

where $x_{it} \in \mathcal{X}_i \subset \mathbb{R}^d$ is a vector of observed individual-specific regressors on the ith cross section unit at time t, $g_i(\cdot) \in \mathcal{G}_i$, $\mathcal{G}_i$ is a specified class of continuous function from $\mathcal{X}_i$ to $\mathbb{R}$, f_{1t} is a $q_1 \times 1$ vector of observed common factors that contains the constant term 1, f_{2t} is a $q_2 \times 1$ vector of unobserved common factors, γ_{1i} and γ_{2i}, are factor loadings, ε_{it} is the individual-specific (idiosyncratic) errors assumed to be independently distributed of (x_{it}, f_{1t}, f_{2t}), and γ_{2i}, Γ_{1i} and Γ_{2i} are $q_1 \times d$ and $q_2 \times d$ factor loading matrices, and v_{it} is a $d \times 1$ vector of individual-specific components of x_{it}. [1] If the regression function $g_i(\cdot)$ is not identical across i, then we have heterogenous regressions; otherwise we have homogenous regression relationship that is denoted as $g(\cdot)$.

SJ considered sieve estimation of both heterogeneous and homogenous nonparametric regressions when both the cross-section dimension (n) and the time dimension (T) are large and find that significant gains can be achieved when the regression relationship is homogenous and such knowledge is employed in the estimation procedure. [2]This is as expected. Nevertheless, in practice economic theory usually cannot tell whether the regression relationship is homogenous or not. So it is worthwhile to consider a test for homogenous relationships. If we fail to reject the null of homogenous relationship, then we can pool the cross section data together and estimate a single homogenous relationship more effectively.

In this paper, we consider a nonparametric test for poolability in the model (1.1)-(1.2). Testing for poolability can be traced back to Chow (1960) in econometrics. Since then a large literature has been developed to test structural stability of economic relationships over time or equality of regression functions over individuals. These tests were soon generalized to the nonparametric context for curve comparison. See Baltagi, Hidalgo, and Li (1996), Criado (2008), Hall and Hart (1990), Koul and Schick (1997), Lavergne (2001), Neumeyer and

[1]Write $f_{1t} = (1, f_{1t}^{*\prime})'$. As SJ remarked, we can allow f_{1t}^* to enter (1.1) nonparametrically, in which case (1.1) will become $y_{it} = g_i(x_{it}, f_{1t}^*) + \gamma_{1i} + e_{it}$. Our asymptotic theory allows some component of x_{it} in (1.1) not to vary across i, and thus this specification can be treated as a special case of (1.1), where in (1.1) x_{it} includes some observable common factors and $f_{1t} \equiv 1$.

[2]If $f_{1t} \equiv 1$ and $f_{2t} \equiv 0$ for all t, and $g_i \equiv g$ for all i, then the model in (1.1) becomes the typical "fixed-effect" model. In this case one can follow Baltagi and Li (2002) and take a first difference of the data before using series estimation. In our paper the presence of nonconstant elements in the observable factors and unobservable factors complicates the asymptotic analysis to a great deal, and simple first difference cannot yield the desirable model to be estimated consistently by series method.

Dette (2003), Vilar-Fernández and González-Manteiga (2004), and Vilar-Fernández, Vilar-Fernández, and González-Manteiga (2007), among others. Nevertheless, to the best of our knowledge, these tests are only designed to test for the equality of a fixed number of nonparametric regression curves. It is not clear whether they continue to be valid when the number of regression curves is increasing over the sample size.

There are several key features that distinguish our tests from the existing literature on curve comparisons. First, unlike the large number of parametric tests for slope homogeneity, our test is a nonparametric test for homogeneity or poolability of nonparametric regression relationships. This is important since few economic theories suggest exact functional forms and nonparametric poolability test can effectively avoid lack of robustness to correct functional specification. Second, our test is designed to test for poolability in large dimensional panel data models with cross-section dependence. In the absence of cross-section dependence, Pesaran and Yamagata (2008) proposed a test for slope homogeneity in large panels when the functional relationship is assumed to be linear. In comparison with their test, our task is complicated significantly by the presence of both cross-section dependence and the unknown smooth nonparametric functional relationship. Third, given the large dimensional panel setup the number of regression curves (n here) passes to infinity. Since our test is based on SJ's semiparametric common correlated estimator (CCE) which requires the use of cross-section sample mean of (x_{it}, y_{it}) as a proxy for the unobservable common factor f_{2t}, n must tend to infinity sufficiently fast to ensure that the proxy error is asymptotically negligible in our test.

The rest of the paper is structured as follows. Section 2 introduces the hypothesis and test statistic. In Section 3 we study the asymptotic distributions of the test statistic under the null, a sequence of local alternatives, and global alternatives, where we also propose a bootstrap version of the test. A small set of Monte Carlo simulation results is reported in Section 4. Final remarks are contained in Section 5. All technical details are relegated to the Appendix.

NOTATION. Throughout the paper we adopt the following notation and conventions. For a matrix A, we denote its Euclidean norm as $\|A\| = [\mathrm{tr}\,(AA')]^{1/2}$ and its generalized inverse as A^{-}. When A is a square matrix, we use $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ to its minimum and maximum eigenvalues respectively, and $\mathrm{diag}(A)$ to denote the diagonal matrix formed from the diagonal elements of A. I_T denotes a $T \times T$ identity matrix. For a vector $a \equiv (a_1, ..., a_T)'$, $\mathrm{diag}(a)$ denotes a diagonal matrix with a_i as a typical diagonal element. The operator $\xrightarrow{p}$ denotes convergence in probability, and $\xrightarrow{d}$ convergence in distributions. We use $(n, T) \to \infty$ to denote the joint convergence of n and T.

2 Hypotheses and test statistic

In this section we first state the hypotheses and then introduce the test statistic.

2.1 The hypotheses

We consider testing possible homogenous regression relationships in model (1.1). The null hypothesis is

$$H_0 : g_i(x) = g_j(x) \text{ a.e. on the joint support of } g_i \text{ and } g_j \text{ and for all } i, j = 1, \cdots, n, \quad (2.1)$$

where a.e. is the abbreviation for almost everywhere. The alternative hypothesis is the negation of H_0 :

$$H_1 : g_i(x) \neq g_j(x) \text{ for some } i \neq j \text{ with probability greater than zero.} \quad (2.2)$$

Let $g(x) \equiv n^{-1} \sum_{i=1}^{n} g_i(x)$. We can rewrite the null and alternative hypotheses equivalently as

$$H_0 : g_i(x) = g(x) \text{ a.e. for all } i = 1, \cdots, n, \quad (2.3)$$

$$H_1 : g_i(x) \neq g(x) \text{ a.e. for some } i \text{ with probability greater than zero.} \quad (2.4)$$

To facilitate the local power analysis, we also define a sequence of Pitman local alternatives:

$$H_1(\gamma_{nT}) : g_i(x) = g(x) + \gamma_{nT} \Delta_{in}(x) \text{ for all } i = 1, \cdots n \quad (2.5)$$

where $\Delta_{in}(x)$ is uniformly bounded measurable functions, $\gamma_{nT} \to 0$ as $(n, T) \to \infty$, and the exact rate of γ_{nT} is specified in Theorem 3.3 below.

In this paper, we consider a test of poolability based on the null hypothesis H_0 in (2.1). In fact we can construct consistent tests of H_0 versus H_1 using various distance measures. A convenient choice is to use the measure

$$\Gamma = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \int (g_i(x) - g_j(x))^2 w(x) dx, \quad (2.6)$$

where $w(x)$ is a nonnegative weight function that has support on $\mathbb{R}^d$ and could be allowed to depend on (i, j), i.e., $w(x) = w_{ij}(x)$. But for notational simplicity, we restrict ourselves to the employment of same weight functions for different pairs (i, j). Note that $\Gamma = 0$ if and only if H_0 holds and we can propose a test statistic based upon consistent estimation of Γ.

2.2 Estimation and test statistic

To estimate Γ, we follow SJ and estimate the unknown functions $g_i(\cdot)$, $i = 1, \cdots, n$, by sieve method. Let $\{p_l(x),\ l = 1, 2, \cdots\}$ denote a sequence of known basis functions that can approximate any square-integrable function. Let $K \equiv K(n, T)$ be some integer such that $K \to \infty$ as $(n, T) \to \infty$. [3]Let $p^K(x) = (p_1(x), p_2(x), \cdots, p_K(x))'$, $p_{it} = p^K(x_{it})$, and $p_i = (p_{i1}, p_{i2}, \cdots, p_{iT})'$. Obviously we have suppressed the dependence of p_{it}, and p_i, on K and T. In particular, p_i is a $T \times K$ matrix.

Under fairly weak conditions, we can approximate $g_i(x)$ very well by $\alpha_{g_i}' p^K(x)$ for some $K \times 1$ vector α_{g_i}. Let $\overline{x}_t \equiv n^{-1} \sum_{i=1}^n x_{it}$, $\overline{y}_t \equiv n^{-1} \sum_{i=1}^n y_{it}$, and $h_t \equiv (f_{1t}', \overline{x}_t', \overline{y}_t)'$. As SJ argued, we can use h_t as an observable proxy for the unobservable factor f_{2t}. This motivates them to estimate $g_i(\cdot)$ by augmenting the sieve regression of y_{it} on x_{it} with h_t :

$$y_{it} = \alpha_{g_i}' p^K(x_{it}) + \vartheta_i' h_t + u_{it} \tag{2.7}$$

where u_{it} is the new error term. By the formula for partitioned regression, the estimate of α_{g_i} is given by

$$\widehat{\alpha}_{g_i} = \left(p_i' m_h p_i\right)^- p_i' m_h y_i, \tag{2.8}$$

where $h \equiv (h_1, h_2, \cdots, h_T)'$, $y_i \equiv (y_{i1}, y_{i2}, \cdots, y_{iT})'$, $m_h \equiv I_T - h(h'h)^- h$, and $(\cdot)^-$ denotes any symmetric generalized inverse. Then we estimate $g_i(\cdot)$ by

$$\widehat{g}_i(x) = p^K(x)' \, \widehat{\alpha}_{g_i}. \tag{2.9}$$

With $\widehat{g}_i(x)$, we then estimate Γ by the following functional:

$$\Gamma_{nT} = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \int \left(\widehat{g}_i(x) - \widehat{g}_j(x)\right)^2 w(x)\, dx. \tag{2.10}$$

This statistic is simple to compute and offers a natural way to test the null hypothesis. In the following, we limit ourselves to the case where $w(x)$ is a probability density function (PDF) chosen by the researchers. We will show that, after being appropriately normalized, Γ_n is asymptotically normally distributed under suitable assumptions and have power to detect sequences of Pitman local alternatives at certain rate.

3 The asymptotic distributions of the test statistic

In this section, we first present a set of assumptions that are used in the asymptotic analysis. Then we study the asymptotic distribution of our test statistic under the null hypothesis, a

[3]In theory one can choose $K \equiv K(n, T)$ to balance the size and power of our test. But this will require higher order theory, which is beyond the scope of the paper. In practice we recommend the use of least-squares cross validation (LSCV) to choose K which seems to work very well in our simulation study.

sequence of Pitman local alternatives, and fixed alternatives. We also propose a bootstrap version of the test.

3.1 Assumptions

To proceed, let $\varepsilon_i \equiv (\varepsilon_{i1}, \varepsilon_{i2}, \cdots, \varepsilon_{iT})'$ and $v_i \equiv (v_{i1}, v_{i2}, \cdots, v_{iT})'$. Let $\overline{P}_w \equiv \int p^K(x) p^K(x)' w(x) dx$. Define

$$Q_{ipp} \equiv E[p_{it}p_{it}'],\ Q_{iph} \equiv E[p_{it}h_t'],\ Q_{hh} \equiv E[h_t h_t'],\ \text{and}\ Q_i \equiv Q_{ipp} - Q_{iph}Q_{hh}^{-1}Q_{iph}', \quad (3.1)$$

where we have suppressed the dependence of $Q_{hh} (\equiv Q_{n,hh})$ on n through h_t. The $K \times K$ matrices Q_{ipp} and Q_i play an important role in this paper. We make the following assumptions.

Assumption 1. (i) For each i, ε_{it} are independent and identically distributed (IID) with mean 0 and variance σ_i^2, and the process $\{v_{it} : t \geq 1\}$ is a strictly stationary and α-mixing process with mixing coefficient $\alpha_i(j)$ such that $\sum_{j=1}^{\infty} j^2 \alpha_i(j)^{\eta/(4+\eta)} \leq C_1 < \infty$ for some $\eta > 0$. $\underline{c} \leq \min_{1 \leq i \leq n} \sigma_i^2 \leq \max_{1 \leq i \leq n} \sigma_i^2 \leq \overline{c}$ for some $\underline{c} > 0$ and $\overline{c} < \infty$. (ii) The common factor process $\{(f_{1t}, f_{2t}) : t \geq 1\}$ is a strictly stationary and α-mixing process with mixing coefficient $\alpha_0(j)$ such that $\sum_{j=1}^{\infty} j^2 \alpha_0(j)^{\eta/(4+\eta)} \leq C_2 < \infty$. (iii) (f_{1t}, f_{2t}) is distributed independently of the individual-specific errors ε_{is} and v_{is} for all i, t, and s. $E[(f_{1t}', f_{2t}')'(f_{1t}', f_{2t}')]$ is positive definite. (iv) The individual-specific errors ε_{it} and v_{js} are distributed independently for all i, j, t, and s. (v) (ε_i, v_i) are independently distributed across i with zero means. (vi) $E\left[\varepsilon_{i1}^8\right] < \infty$, and $\sup_{n \geq 1} \max_{1 \leq i \leq n} E|\zeta_i|^{4+\eta} \leq \overline{\mu}_{4+\eta} < \infty$ for $\zeta_i = v_{i1}$, $g_i(x_{i1})$, f_{11}, and f_{21}. (vii) Let $\alpha(j) \equiv \sup_{n \geq 1} \max_{0 \leq i \leq n} \alpha_i(j)$. $\sum_{j=1}^{\infty} j^2\ \alpha(j)^{\eta/(4+\eta)} \leq C_3 < \infty$. (vii) $E[g_i(x_{it})] = 0$ for all i.

Assumption 2. (i) The unobserved factor loadings γ_{2i} and Γ_{2i} are IID. γ_{2i} and Γ_{2i} are independent of the individual-specific errors ε_{jt} and v_{jt}, and the common factors (f_{1t}, f_{2t}) for all j and t. The $(4+\eta)$-th moment of Γ_{2i} is finite. (ii) Γ_{1i} are either fixed factor loadings that are uniformly bounded or random factor loadings that are IID across i with finite $(4+\eta)$-th moments and are independent of $\Gamma_{2j}, \varepsilon_{jt}$, v_{jt}, f_{1t} and f_{2t} for all j and t. (iii) Let $\Gamma_2^* \equiv E(\Gamma_{2i}^*)$ where $\Gamma_{2i}^* \equiv (\Gamma_{2i}, \gamma_{2i})$. rank$(\Gamma_2^*) = q_2 \leq d+1$. [4]

Assumption 3. (i) For each i, $g_i(\cdot)$ is $H(\lambda_i, \omega_i)$-smooth on $\mathcal{X}_i$ for some $\lambda_i > d/2, \omega_i \geq 0$. (See SJ for the definition of $H(\cdot,\cdot)$-smoothness.) (ii) For each i, $\int \left(1 + ||x||^2\right)^{\overline{\omega}_i} dF_i(x) < C < \infty$ for some $\overline{\omega}_i > \omega_i + \lambda_i$, where $dF_i(x) = f_i(x) dx$, and $f_i(x)$ is the probability density function of x_{it}. (iii) For any $H(\lambda_i, \omega_i)$-smooth $g_i(\cdot)$ on $\mathcal{X}_i$, there is a function $\Pi_{\infty K} g_i \equiv \alpha_{g_i}' p^K(\cdot)$

[4] It is desirable to have a statistical test to check such a rank condition. For a recent review on the methods to test the rank of a general matrix, see Camba-Mendeza and Kapetanios (2009). Nevertheless, it seems to diffuclt to apply none of the reviewed method to our case because we do not have repeated panel observations to estimate the expected value of the matrix of factor loadings, i.e., $E(\Gamma_{2i}^*)$.

in the sieve space $\mathcal{G}_K \equiv \{f(\cdot) = a'p^K(\cdot)\}$ such that $||g_i(\cdot) - \Pi_{\infty K} g_i(\cdot)||_{\infty,\overline{\omega}_i} = O\left(K^{-\lambda_i/d}\right)$. (iv) For each i, $\sup_{1\leq j\leq K} E|p_j(x_{i1})|^{4+\eta} < \infty$ for the same η defined in Assumption 1(i), Q_i has the smallest eigenvalues bounded away from zero, Q_{ipp} has bounded largest eigenvalues uniformly in K, and $Q_{hh} \equiv Q_{n,hh}$ tends to a positive definite matrix as $n \to \infty$.

Assumption 4. (i) The nonnegative weight function $w(\cdot)$ is a PDF. (ii) $\int \left(1 + ||x||^2\right)^{\overline{\omega}} w(x)\, dx < \infty$ where $\overline{\omega} \equiv \max_{1\leq i\leq n} \overline{\omega}_i$. (iii) For each K, the smallest and largest eigenvalues of $\overline{P}_w$ are bounded away from zero and infinity, respectively.

Assumption 5. $K^2/T \to 0$, $KT^2/n \to 0$, $\max\left(nK^{-2\lambda/d},\ TK^{-2\lambda/d+1}, TK^{-2\lambda/d-1}\zeta(K)^2\right) \to 0$ as $(n,T) \to \infty$, where $\lambda \equiv \min_{1\leq i\leq n} \lambda_i$ and $\zeta(K) \equiv \sup_x \left\|p^K(x)\right\|$.

Under a set of conditions that are weaker than Assumptions 1-3 and 5 above, SJ establish the consistency and asymptotic normality of the sieve estimator $\widehat{g}_i(x)$. In comparison with SJ's conditions, our assumptions are stronger in three aspects. First, to facilitate the establishment of asymptotic distributions of our test statistic, we strengthen the strong mixing condition of SJ on the process $\{\varepsilon_{it}, t \geq 1\}$ to the IID condition in Assumption 1. This greatly simplifies the application of a central limit theorem (CLT) for the summation of quadratic forms and the estimation of its asymptotic variance. We conjecture that the latter condition can be relaxed at the cost of lengthy and complicated arguments. Second, the moment condition on ε_{it} is also strengthened from the existence of $(4+\eta)$th finite moments to the existence of 8th finite moments because we need to verify that the 4th moments of a quadratic form of ε_i is finite. Third, the conditions on (K, n, T) in Assumption 5 are much stronger than those in SJ in order to ensure some terms are asymptotically negligible. For example, SJ only requires that $KT/n \to 0$ as $(n,T) \to \infty$ but we need $KT^2/n \to 0$ as $(n,T) \to \infty$. The latter condition corresponds to and is much stronger than Pesaran's (2006) requirement that $\sqrt{T}/n \to 0$ when g_i is assumed to be linear. In particular, it means that our test is mainly applicable in large dimensional panel where the number of cross-sectional units is much larger than the number of time periods. [5] See SJ for remarks on other parts of Assumptions 1-3 and 5.

In addition, Assumption 4 is new. The requirement that $w(\cdot)$ be a PDF is innocuous

[5] As a referee remarked, one can interprete this from the standpoint of kernel-based specification tests. In Fan and Li (1996), they choose bandwidth a to estimate the restricted model with q_1 regressors and bandwidth h to construct the residual-based test statistic that requires a higher dimension regression with q_1+q_2 regressors under the alternative. When $q_2 \leq q_1$, their condition implies that $h/a \to 0$, implying that they smooth the alternative model less than the null-resticted model. Here, the condition $KT^2/n \to 0$ suggests that the value of K used in our test must be smaller than the value of K used in estimating a correctly-specified homogeneous model which only requires that $KT/n \to 0$. Smaller K is anaglogous to undersmoothing for the kernel-based test.

and can be relaxed. Assumptions 4(ii)-(iii) parallel Assumptions 3(ii) and (iv). If x_{it} are identically distributed across i, then in principle one can choose $w(\cdot)$ as the PDF of x_{it} (or its consistent estimate in practice). In this case, $\overline{P}_w$ reduces to Q_{ipp} which is the same across i.

3.2 Asymptotic distributions

We first study the asymptotic distributions of Γ_{nT} under the null hypothesis. Let $z_t \equiv (f_{1t}', f_{2t}')'$. Then

$$h_t = \Gamma' z_t + \overline{v}_t^*, \tag{3.2}$$

where

$$\underset{(q_1+q_2)\times(q_1+d+1)}{\Gamma} = \begin{pmatrix} I_{q_1} & \underset{q_1\times d}{\overline{\Gamma}_1} & \underset{q_1\times 1}{\overline{\gamma}_1} \\ \underset{q_2\times q_1}{0} & \underset{q_2\times d}{\overline{\Gamma}_2} & \underset{q_2\times 1}{\overline{\gamma}_2} \end{pmatrix}, \quad \underset{1\times(q_1+d+1)}{\overline{v}_t^{*\prime}} = \begin{pmatrix} \underset{1\times q_1}{0'} & \underset{1\times d}{\overline{v}_t'} & \underset{1\times 1}{\overline{g}_t + \overline{\varepsilon}_t} \end{pmatrix}, \tag{3.3}$$

$\overline{\Gamma}_1$, $\overline{\Gamma}_2$, $\overline{\gamma}_1$, $\overline{\gamma}_2$, $\overline{v}_t$, $\overline{\varepsilon}_t$, and $\overline{g}_t$ are sample averages of Γ_{1i}, Γ_{1i}, γ_{1i}, γ_{2i}, v_{it}, ε_{it}, and $g_i(x_{it})$ over i, respectively. (3.2) justifies the use of h_t as the proxy for z_t. Let $b_t \equiv \Gamma' z_t$, $b \equiv (b_1, \cdots, b_T)'$, and $m_b \equiv I_T - b(b'b)^- b$. Define

$$B_{nT} \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sigma_i^2 \mathrm{tr}\left(\left(p_i' m_b p_i / T\right)^- \overline{P}_w \right), \tag{3.4}$$

and

$$V_{nT} \equiv \frac{2}{n} \sum_{i=1}^{n} \sigma_i^4 \mathrm{tr}\left(\left(\left(p_i' m_b p_i / T\right)^- \overline{P}_w \right)^2 \right), \tag{3.5}$$

The following theorem establishes the asymptotic normality of Γ_{nT} after being appropriately scaled and centered.

Theorem 3.1 *Under Assumptions 1-5 and under H_0,*

$$\frac{c_{nT}\Gamma_{nT} - B_{nT}}{\sqrt{V_{nT}}} \xrightarrow{d} N(0,1),$$

where $c_{nT} = \frac{T}{\sqrt{n}(n-1)}$.

The proof of Theorem 3.1 is given in Appendix A1. The idea underlying the proof is very simple. Let $\overrightarrow{\varepsilon}_i = (p_i' m_b p_i)^- p_i' m_b \varepsilon_i$. We first demonstrate that

$$c_{nT}\Gamma_{nT} = c_{nT} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (\overrightarrow{\varepsilon}_i - \overrightarrow{\varepsilon}_j)' \overline{P}_w (\overrightarrow{\varepsilon}_i - \overrightarrow{\varepsilon}_j) + o_p(1),$$

that is, $c_{nT}\Gamma_{nT}$ can be written as a second-order U-statistic. Next, we apply the Hoeffding decomposition and show that

$$\frac{c_{nT}\Gamma_{nT}-B_{nT}}{\sqrt{V_{nT}}}=\frac{TV_{nT}^{-1/2}}{\sqrt{n}}\sum_{i=1}^{n}\left(\overrightarrow{\varepsilon}_i'\overline{P}_w\overrightarrow{\varepsilon}_i-\sigma_i^2\text{tr}\left(\left(p_i'm_bp_i/T\right)^{-}\overline{P}_w\right)\right)+o_p\left(1\right).$$

Then we apply the CLT for independent but non-identically distributed (INID) variables to obtain the desired result.

Note that the test "statistic" in Theorem 3.1 is not feasible as it depends on the unknown objects B_{nT} and V_{nT}. To implement the test, we need to estimate both the "bias" B_{nT} and the variance V_{nT} consistently. It turns out that consistent estimation of V_{nT} is straightforward whereas that of B_{nT} is not (as V_{nT} is diverging to ∞ at the rate $K^{1/2}$, slower than the $(nK)^{1/2}$-rate at which B_{nT} is diverging to ∞). Let $\widehat{\mathbf{g}}_i\equiv(\widehat{g}_i(x_{i1}),\cdots,\widehat{g}_i(x_{iT}))'$ and $\widehat{e}_i\equiv m_h(y_i-\widehat{\mathbf{g}}_i)$. Denote the tth element of $\widehat{e}_i$ as $\widehat{e}_{it}$. We propose to estimate V_{nT} by

$$\widehat{V}_{nT}\equiv\frac{2}{n}\sum_{i=1}^{n}\widehat{\sigma}_i^4\text{tr}\left(\left(\left(p_i'm_hp_i/T\right)^{-}\overline{P}_w\right)^2\right)$$

where $\widehat{\sigma}_i^2\equiv\frac{1}{T}\sum_{t=1}^{n}\widehat{e}_{it}^2$. It is straightforward to demonstrate that $\widehat{V}_{nT}=V_{nT}+o_p\left(1\right).$ For B_{nT}, a simple replacement of m_b and σ_i^2 by m_h and $\widehat{\sigma}_i^2$ won't deliver a consistent estimate. In fact, we can show that

$$\begin{aligned}&B_{nT}-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widehat{\sigma}_i^2\text{tr}\left(\left(p_i'm_hp_i/T\right)^{-}\overline{P}_w\right)\\=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widehat{\sigma}_i^2\text{tr}\left\{\left(\left(p_i'm_bp_i/T\right)^{-}-\left(p_i'm_hp_i/T\right)^{-1}\right)\overline{P}_w\right\}+o_p\left(1\right).\end{aligned}$$

The dominant term in the last expression is $O_p(K)$ by Lemma **??**(ix) in the appendix, which also needs to be estimated. Let $\widehat{\overline{g}}_t\equiv\frac{1}{n}\sum_{i=1}^{n}\widehat{g}_i\left(x_{it}\right)$, $\widehat{\overline{g}}_t^*\equiv(0_{1\times(q_1+d)},\widehat{\overline{g}}_t)'$, and $\widehat{\overline{g}}^*\equiv(\widehat{\overline{g}}_1^*,\cdots,\widehat{\overline{g}}_T^*)'$. Define

$$\begin{aligned}\widehat{b}_{nT}&=\frac{T}{\sqrt{n}}\sum_{i=1}^{n}\widehat{\sigma}_i^2\text{tr}\left\{\left(p_i'm_hp_i\right)^{-}\widehat{b}_{inT}\left(p_i'm_hp_i\right)^{-}\overline{P}_w\right\},\text{ and}\\\widehat{B}_{nT}&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widehat{\sigma}_i^2\text{tr}\left(\left(p_i'm_hp_i/T\right)^{-}\overline{P}_w\right)+\widehat{b}_{nT},\end{aligned}$$

where $\widehat{b}_{inT}\equiv p_i'(-\widehat{\overline{g}}^*\left(h'h\right)^{-1}h'-h\left(h'h\right)^{-1}\widehat{\overline{g}}^*+h\left(h'h\right)^{-1}\widehat{\overline{g}}^{*\prime}\widehat{\overline{g}}^*\left(h'h\right)^{-1}h')p_i$. As we demonstrate in the appendix, $\widehat{B}_{nT}-B_{nT}=o_p\left(\sqrt{V_{nT}}\right).$ Thus we have the following corollary.

Corollary 3.2 *Let $D_{nT}\equiv(c_{nT}\Gamma_{nT}-\widehat{B}_{nT})/\sqrt{\widehat{V}_{nT}}$. Suppose Assumptions 1-5 hold with Assumption 3(ii) being strengthened to: for all i, $\int\left(1+||x||^2\right)^{2\overline{\omega}_i}dF_i\left(x\right)<C<\infty$ for some $\overline{\omega}_i>\omega_i+\lambda_i$. Then*

$$D_{nT}\overset{d}{\to}N\left(0,1\right).$$

Corollary 3.2 indicates that we can compare D_{nT} to z_α, the αth upper percentile from the standard normal distribution, and we reject the null hypothesis when $D_{nT} > z_\alpha$.

Next, we study the asymptotic distribution of D_{nT} under the Pitman local alternative in (2.5). Let

$$\Delta \equiv \lim_{(n,T)\to\infty} \sqrt{\frac{\text{tr}\left(\overline{P}_w\right)}{V_{nT}}} \frac{1}{n(n-1)} \sum_{1\leq i<j\leq n} \int \left(\Delta_{in}(x) - \Delta_{jn}(x)\right)^2 w(x)\, dx.$$

Then we have the following theorem.

Theorem 3.3 *Suppose Assumptions 1-5 hold. Then under* $H_1(\gamma_{nT})$ *with* $\gamma_{nT} = n^{-1/4}T^{-1/2}$ $tr\left(\overline{P}_w\right)^{1/4}$, $D_{nT} \xrightarrow{d} N(\Delta, 1)$.

Theorem 3.3 suggests that our test has nontrivial power to detect Pitman local alternatives at the rate $n^{-1/4}T^{-1/2}\text{tr}\left(\overline{P}_w\right)^{1/4}$, which is slower than the rate $n^{-1/4}T^{-1/2}$ as $\text{tr}\left(\overline{P}_w\right) = O(K)$. The latter rate was obtained by Pesaran and Yamagata (2008) in the case of testing slope homogeneity in large linear panel data panels. In addition, Theorem 3.3 indicates the power of the test satisfies $P\left(D_{nT} > z_\alpha | H_1\left(n^{-1/4}T^{-1/2}tr\left(\overline{P}_w\right)^{1/4}\right)\right) \to 1 - \Phi(z_\alpha - \Delta)$ as $(n,T) \to \infty$, where $\Phi(\cdot)$ is the cumulative distribution function of standard normal.

The following theorem establishes the consistency of the test.

Theorem 3.4 *Let* $\Delta_{gn} \equiv (n(n-1))^{-1} \sum_{1\leq i<j\leq n} \int [g_i(x) - g_j(x)]^2\ w(x)\, dx$. *Suppose* $\Delta_g \equiv \lim_{n\to\infty} \Delta_{gn} > 0$ *and Assumptions 1-5 hold. Then under* H_1, $P(D_{nT} > d_{nT}) \to 1$ *for any sequence* $d_{nT} = o_p\left(n^{1/4}T^{1/2}tr\left(\overline{P}_w\right)^{-1/4}\right)$ *as* $(n,T) \to \infty$.

Theorem 3.4 indicates that under H_1 our test statistic D_{nT} explodes at the rate $n^{1/4}T^{1/2}$ $tr\left(\overline{P}_w\right)^{-1/4}$ provided $\Delta_g > 0$. This can occur if $g_i(\cdot)$ and $g_j(\cdot)$ differ on a set of positive measure for a "large" number of pairs (i,j) with $i \neq j$. It rules out the case where only a finite fixed number of functions among $\{g_i(\cdot)\}_{i=1}^n$ are distinct from others on a set of positive measure (e.g., only $g_1(\cdot)$ is different from others), or the case where the cardinality of the set of functions among $\{g_i(\cdot)\}_{i=1}^n$ that are distinct from others on a set of positive measure is diverging to infinity as $n \to \infty$ but at a slower rate than n. In the former case, Δ_{gn} shrinks to 0 at rate n^{-2} so that the D_{nT} test cannot be consistent; in the latter case case, our test can still be consistent as long as Δ_{gn} does not shrink to 0 too fast so that D_{nT} is still diverging to infinity as $(n,T) \to \infty$.

3.3 A bootstrap version of our test

We now propose a bootstrap version of the test. Similar to Li, Hsiao and Zinn (2003), one may want to propose a bootstrap version of the test using a residual-based wild bootstrap method.

To construct the bootstrap analog of y_{it}, one needs not only estimates of $g(x_{it})$ but also estimates of the non-observable factor f_{2t} and the factor loadings γ_{1i} and γ_{2i}. Unfortunately, consistency of estimates for the latter objects has not been established, not to mention the consistency rate. So Li, Hsiao and Zinn's method cannot be applied in our framework.

Nevertheless, the results in the last subsection suggest that under the null hypothesis

$$D_{nT} = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\overrightarrow{\varepsilon}_i'\overline{P}_w\overrightarrow{\varepsilon}_i - \sigma_i^2\text{tr}\left(\left(p_i'm_bp_i/T\right)^{-}\overline{P}_w\right)\right)}{\sqrt{\frac{2}{n}\sum_{i=1}^{n}\widehat{\sigma}_i^4\text{tr}\left(\left(\left(p_i'm_hp_i/T\right)^{-}\overline{P}_w\right)^2\right)}} + o_p(1) \quad (3.6)$$

$$\xrightarrow{d} N(0,1).$$

Therefore we can continue to obtain a bootstrapped version of D_{nT} by mimicking the leading term on the right hand side of (3.6). Let $e_{it}^* = \eta_{it}\widehat{e}_{it}$ where η_{it} are IID across i and t with mean zero and variance 1. We define the bootstrap analog of D_{nT} as

$$D_{nT}^* = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\overrightarrow{\varepsilon}_i^{*\prime}\overline{P}_w\overrightarrow{\varepsilon}_i^* - T\text{tr}\left(\overline{A}_{ih}\Sigma_i\right)\right)}{\sqrt{\frac{2}{n}\sum_{i=1}^{n}\text{tr}\left(\overline{A}_{ih}\Sigma_i\overline{A}_{ih}\Sigma_i\right)}}$$

where $\overrightarrow{\varepsilon}_i^* \equiv (p_i'm_hp_i)^{-}p_i'm_he_{it}^*$, $e_i^* \equiv (e_{i1}^*,\cdots,e_{iT}^*)'$, $\Sigma_i \equiv \text{diag}(\widehat{e}_{it}^2,\cdots,\widehat{e}_{iT}^2)$, and $\overline{A}_{ih} \equiv m_hp_i\,(p_i'm_hp_i)^{-}\,\overline{P}_w\,(p_i'm_hp_i)^{-}\,p_i'm_h$. Note that $T\text{tr}\left(\overline{A}_{ih}\Sigma_i\right)$ is the conditional expectation of $\overrightarrow{\varepsilon}_i'\overline{P}_w\overrightarrow{\varepsilon}_i$ given the data, and $2\text{tr}\left(\overline{A}_{ih}\Sigma_i\overline{A}_{ih}\Sigma_i\right)$ is the conditional variance of $\overrightarrow{\varepsilon}_i\overline{P}_w\overrightarrow{\varepsilon}_i$ given the data, ignoring asymptotically negligible terms. It is straightforward to show that D_{nT}^* converges to the standard normal distribution no matter whether H_0 holds or not. Therefore, we have the following theorem.

Theorem 3.5 *Suppose Assumptions 1-5 hold. Suppose that η_{it} are IID across i and t such that $E(\eta_{it}) = 0$, $E\left(\eta_{it}^2\right) = 1$, and $E\left(\eta_{it}^8\right) < \infty$. Then $D_{nT}^* \xrightarrow{d} N(0,1)$ conditionally on the data, and $P\left(D_{nT} > D_{nT}^*\right) \to 1$ under H_1.*

The conditions on η_{it} can easily be met in practice. In the following simulations, we simply draw η_{it} from the standard normal distributions. The first part of theorem 3.5 shows that the proposed bootstrap provides an asymptotically valid approximation to the null limit distribution of D_{nT}. The second part implies that the test D_{nT} based upon the bootstrap critical value is consistent against all global alternatives. We will compare the finite-sample performance of the bootstrap test with that of the asymptotic normal approximation in our simulation.

4 Simulation

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 some other tests. We consider three data generating processes (DGPs) that are different only in the specification of the regression functions g_i of interest. In all three DGPs, we generate y_{it} and $x_{it,s}$ $(s = 1, 2)$ according to:

$$\begin{aligned} y_{it} &= g_i\left(x_{it,1}, x_{it,2}\right) + \gamma_{1i} + \gamma_{2i,1} f_{2t,1} + \gamma_{2i,2} f_{2t,2} + \varepsilon_{it}, \\ x_{it,s} &= \Gamma_{1i,s} + \Gamma_{2i,s1} f_{2t,1} + \Gamma_{2i,s2} f_{2t,2} + v_{it,s}, \ \ s = 1, 2, \end{aligned}$$

for $i = 1, 2, \cdots, n$, and $t = 1, 2, \cdots, T$. Clearly, we have two individual-specific regressors $(x_{it} \equiv (x_{it,1}, x_{it,2})')$, one observed common factor $(f_{1t} = 1)$, and two unobserved common factors $(f_{2t} \equiv (f_{2t,1}, f_{2t,2})')$. We generate $v_{it} \equiv (v_{it,1}, v_{it,2})'$, x_{it}, the unobserved factors (f_{2t}), and the factor loadings $(\Gamma_{1i} \equiv (\Gamma_{1i,1}, \Gamma_{1i,2})$, $\Gamma_{2i} \equiv (\Gamma_{2i,1}, \Gamma_{2i,2})$, $\gamma_{2i} \equiv (\gamma_{2i,1}, \gamma_{2i,2})')$ according to the full rank regression case of SJ. For each i, ε_{it} are IID drawn from $N\left(0, \sigma_i^2\right)$ where σ_i^2 are IIDU$[0.5, 1]$ across i's.

Following Pesaran (2006), DGP 1 considers a linear specification for $g_i (= g)$ under the null:

$$g\left(x_{it,1}, x_{it,2}\right) = 0.5 x_{it,1} + 0.5 x_{it,1} + \delta_i x_{it,2}.$$

In contrast, both DGPs 2 and 3 consider a nonlinear specification for $g_i (= g)$ under the null:

$$g\left(x_{it,1}, x_{it,2}\right) = \exp\left(x_{it,1}\right) / \left(\exp\left(x_{it,1}\right) + 1\right) + \left(0.5 x_{it,2} - 0.25 x_{it,2}^2\right).$$

Under the alternative, we consider $g_i\left(x_{it,1}, x_{it,2}\right) = g\left(x_{it,1}, x_{it,2}\right) + \delta_i \cos(\pi x_{it,2})$ where g is specified as above, δ_i's are IIDU$[0, c]$ in DGPs 1 and 2, $\delta_i = 0$ for all $i = 1, 2, \cdots, n/2$, and generate δ_i's are IIDU$[0, c]$ for $i = n/2{+}1, n/2{+}2, \cdots, n$ in DGP3 under the alternative. Here $c > 0$ is a parameter that controls the degree of heterogeneity in the DGPs: the larger value of c, the greater degree of heterogeneity; we will set $c = 1$ and 2 for our power study. Clearly, DGP 3 is identical to DGP 2 under the null and different from it under the alternative.

We consider three tests of poolability in this paper. The first one is our D_{nT} test which takes into account the unobservable common factors and does not assume known functional relationship. The second one is a variant of our D_{nT} test that does not assume known functional relationship but neglects the presence of unobservable common factors. The third one is the test of Pesaran and Yamagata (2008, PY hereafter) that assumes linear functional relationship and neglects the unobservable common factors. We denote the second and third tests as $D_{NT}^{(N)}$ and $D_{NT}^{(PY)}$, respectively, where the superscript N stands for "naive" and it indicates that we construct $D_{NT}^{(N)}$ based on the naive estimators of $g_i\left(x_{it,1}, x_{it,2}\right)$ in SJ by augmenting the sieve regression of y_{it} on x_{it} with only observed common factor f_{1t} (i.e., $h_t = f_{1t}$ in (2.7)). The third test was calculated according to eq. (54) in PY.

To conduct the D_{nT} and $D_{NT}^{(N)}$ tests, we need to estimate the model under the alternative. Since $g_i\left(x_1, x_2\right)$ has the additive structure and can be written as the sum of $g_{i1}\left(x_1\right)$ and $g_{i2}\left(x_2\right)$, [e.g., $g_{i1}\left(x_1\right) \equiv \exp\left(x_{it,1}\right) / \left(\exp\left(x_{it,1}\right) + 1\right)$ and $g_{i2}\left(x_2\right) \equiv \left(0.5 x_{it,2} - 0.25 x_{it,2}^2\right)$

$+\delta_i \cos(\pi x_{it,2})$ in DGPs 2-3], we approximate each component by J terms of Hermite polynomials, where J is chosen by the least square cross-validation method. That is, we choose J to minimize the criterion function[6]

$$CV(J) \equiv \sum_{i=1}^{n} \sum_{t=1}^{T} \left\{ y_{it} - \widehat{g}_{-i}^{(J)}(x_{it}) \right\}^2$$

where $\widehat{g}_{-i}^{(J)}(x_{it})$, $t = 1, \cdots, T$, are the restricted semiparametric CCE estimate (for the D_{NT} test) or SJ's naive estimate (for the $D_{NT}^{(N)}$ test) of $g(x_{it})$ under the null by deleting the T observations corresponding to individual i and using the $2J$ terms of Hermite polynomials to approximate the restricted homogeneous regression function $g(\cdot)$.

It is well known that the asymptotic normal distribution typically cannot approximate the finite sample distribution of many nonparametric test statistics. So we suggest using a conditional bootstrap method to obtain the bootstrap p-values. Unfortunately, simulations indicate the bootstrap test procedure proposed in the previous section is sensitive to the specifications of DGPs and combinations of n and T. So we now propose an alternative procedure for the D_{nT} test:

1. Obtain the semiparametric CCE pooled estimate $\widehat{g}(x_{it,1}, x_{it,2})$ under the null. Let $\widehat{u}_{it} \equiv y_{it} - \widehat{g}(x_{it,1}, x_{it,2})$. Estimate the unobserved common factor f_{2t} and factor loadings γ_{2i} by the principal components method. Denote the estimates as $\widehat{f}_{2t}$ and $\widehat{\gamma}_{2i}$, respectively. Estimate γ_{1i} by $\widehat{\gamma}_{1i} \equiv T^{-1} \sum_{t=1}^{T} \{y_{it} - \widehat{g}(x_{it,1}, x_{it,2}) - \widehat{\gamma}_{2i}' \widehat{f}_{2t}\}$ for $i = 1, 2, \cdots, n$. Let $\widehat{\varepsilon}_{it} \equiv \widehat{u}_{it} - \widehat{\gamma}_{1i} - \widehat{\gamma}_{2i}' \widehat{f}_{2t}$ and $\widehat{\varepsilon}_i \equiv (\widehat{\varepsilon}_{i1}, \widehat{\varepsilon}_{i2}, \cdots \widehat{\varepsilon}_{iT})'$.

2. For $i = 1, \cdots, n$ and $t = 1, \cdots, T$, generate

$$y_{it}^* = \widehat{g}(x_{it,1}, x_{it,2}) + \widehat{\gamma}_{1i} + \widehat{\gamma}_{2i}' \widehat{f}_{2t} + \varepsilon_{it}^*,$$

   where ε_{it}^* is the tth element of $\varepsilon_i^* = (\varepsilon_{i1}^*, \varepsilon_{i2}^*, \cdots, \varepsilon_{iT}^*)'$, and ε_i^* is a random drawn from $\{\widehat{\varepsilon}_1, \widehat{\varepsilon}_2, \cdots, \widehat{\varepsilon}_n\}$ with replacement.

3. Compute the bootstrap test statistic D_{nT}^* in the same way as D_{nT} by using $\{(y_{it}^*, x_{it}, f_{1t}), i = 1, \cdots, n, t = 1, \cdots, T\}$ instead.

4. Repeat steps 2-3 B times to obtain B bootstrap test statistic $\left\{D_{nT,j}^*\right\}_{j=1}^{B}$. Calculate the bootstrap p-values $p^* \equiv B^{-1} \sum_{j=1}^{B} \mathbf{1}\left(D_{nT,j}^* \geq D_{nT}\right)$ and reject the null hypothesis of conditional independence if $p^*$is smaller then the prescribed level of significance.

[6] A better strategy is to allow J to depend on the additive component to approximate, i.e., we can choose J_s terms of Hermite polynomials to approximate $g_s(x) = g_{is}(x)$ under the null for $s = 1, 2$.

Clearly the above procedure imposes the null hypothesis when we generate the bootstrap data $\{y_{it}^*\}$ in step 2. Following Bai (2009), we can justify that under H_0, $\widehat{f}_{2t}$ and $\widehat{\gamma}_{2i}$ consistently estimate f_{2t} and γ_{2i} subject to certain normalization restrictions. Nevertheless the rigorous justification for the validity of the above bootstrap method is beyond the scope of the paper. We only demonstrate that it works effectively through simulations.

To obtain a bootstrap analog of $D_{NT}^{(N)}$, one can readily modify the above procedure. For example, in step 1 $\widehat{g}\,(x_{it,1}, x_{it,2})$ is now replaced by SJ's naive estimate $\widehat{g}^{(N)}\,(x_{it,1}, x_{it,2})$ of $g\,(x_{it,1}, x_{it,2})$ under the null, there is no need to estimate f_{2t} and γ_{2i}, and one can estimate γ_{1i} by $\widehat{\gamma}_{1i} \equiv T^{-1} \sum_{t=1}^{T} \{y_{it} - \widehat{g}^{(N)}\,(x_{it,1}, x_{it,2})\}$. Similarly, one can obtain the bootstrap analog of $D_{NT}^{(PY)}$ by imposing linearity in the pooled regression under the null and neglecting the presence of unobserved common factors.

We consider two sample sizes for n : $n = 50$, and 100. We consider $T = 20$, 30, 40, 50 when $n = 50$, and $T = 25$, 50, 75, 100 when $n = 100$. In each scenario, the number of replications in the Monte Carlo study is 1000 for the size study and 500 for the power study. For the bootstrap version of the test, we use B =200 bootstrap resamples for each replication.

Table 1 reports the level performance of the three tests for poolability using the above DGPs. The upper and lower panels of Table 1 summarize the rejection frequency of our tests based on the asymptotic normal critical values and the bootstrap p-values, respectively. For the normal-critical-values-based tests, we find that the D_{nT} test seems to perform reasonably well despite the fact that it tends to be oversized for smaller values of T in DGP 1 and can be somewhat undersized for some values of n and T in DGP 2. Similar phenomenon occurs for the $D_{nT}^{(N)}$ test, and the noticeable main difference is that this test tends to be more oversized for small values of T in DGP 1 than the D_{nT} test. In contrast, the $D_{nT}^{(PY)}$ test, which ignores both unknown functional form and the existence of unobserved common factors, always mistakenly rejects the null hypothesis of homogeneous regression functional relationship. For the bootstrap version of the three tests, we find that the bootstrap-p-value-based D_{nT} test outperforms the normal-critical-values-based test in that the empirical level of the bootstrapped D_{nT} test is quite close to the nominal level for both DGPs 1 and 2, the bootstrapped $D_{nT}^{(N)}$ test tends to be oversized for both DGPs, and the bootstrapped $D_{nT}^{(PY)}$ test is not stable at all across different DGPs and sample sizes. Interestingly, the empirical level of the bootstrapped $D_{nT}^{(PY)}$ test seems fine for DGP 1 when $n = 50$ but it is identically zero when $n = 100$. Nevertheless, for DGP 2 the bootstrapped $D_{nT}^{(PY)}$ test is severely oversized for all combinations of n and T. The last two columns in Table 1 report the average number of series terms J chosen by the LSCV method in the construction of the (bootstrapped and non-bootstrapped) D_{nT} and $D_{nT}^{(N)}$ tests, respectively. As either n or T increases, we see that

Table 1: Finite sample rejection frequency under the null (nominal level: 0.05 and 0.10)

DGP	n	T	5% tests			10% tests				
			D_{nT}	$D_{nT}^{(N)}$	$D_{nT}^{(PY)}$	D_{nT}	$D_{nT}^{(N)}$	$D_{nT}^{(PY)}$	J_{cv1}	J_{cv2}
			Tests based on asymptotic normal critical values							
1	50	20	0.131	0.112	1.000	0.176	0.145	1.000	5.03	4.63
		30	0.103	0.135	1.000	0.138	0.176	1.000	5.23	4.66
		40	0.080	0.164	1.000	0.111	0.207	1.000	5.48	4.60
		50	0.045	0.053	1.000	0.069	0.074	1.000	6.00	5.63
	100	25	0.148	0.145	1.000	0.183	0.182	1.000	5.74	4.53
		50	0.090	0.078	1.000	0.104	0.105	1.000	6.92	5.44
		75	0.027	0.019	1.000	0.045	0.040	1.000	7.49	6.61
		100	0.054	0.056	1.000	0.066	0.071	1.000	7.76	6.70
2	50	20	0.110	0.069	1.000	0.159	0.093	1.000	5.32	4.90
		30	0.079	0.095	1.000	0.102	0.128	1.000	5.48	5.03
		40	0.079	0.104	1.000	0.102	0.142	1.000	5.67	5.12
		50	0.073	0.061	1.000	0.100	0.088	1.000	5.87	5.59
	100	25	0.069	0.096	1.000	0.105	0.126	1.000	5.73	5.06
		50	0.043	0.069	1.000	0.059	0.085	1.000	6.58	5.93
		75	0.028	0.017	1.000	0.040	0.031	1.000	7.05	6.63
		100	0.046	0.041	1.000	0.060	0.054	1.000	7.19	6.71
			Tests based on bootstrap p-values							
1	50	20	0.052	0.217	0.056	0.116	0.274	0.142		
		30	0.077	0.310	0.035	0.137	0.374	0.118		
		40	0.060	0.332	0.018	0.106	0.386	0.113		
		50	0.063	0.274	0.013	0.117	0.320	0.091		
	100	25	0.074	0.358	0.000	0.139	0.406	0.000		
		50	0.077	0.406	0.000	0.134	0.451	0.000		
		75	0.062	0.375	0.000	0.118	0.423	0.000		
		100	0.061	0.366	0.000	0.113	0.428	0.000		
2	50	20	0.067	0.190	0.347	0.121	0.254	0.528		
		30	0.067	0.238	0.287	0.123	0.308	0.540		
		40	0.078	0.284	0.300	0.128	0.344	0.568		
		50	0.050	0.197	0.298	0.096	0.253	0.576		
	100	25	0.068	0.317	0.768	0.122	0.386	0.916		
		50	0.067	0.341	0.864	0.107	0.389	0.979		
		75	0.060	0.296	0.913	0.113	0.338	0.995		
		100	0.049	0.280	0.924	0.090	0.316	0.997		

Note: J_{cv1} and J_{cv2} denote the average values of the number of series terms J chosen by the LSCV method for the tests D_{nT} and $D_{nT}^{(N)}$, respectively.

Table 2: Finite sample rejection frequency under the alternative (nominal level: 0.05)

DGP	n	T		$c=1$		J_{cv1}	J_{cv2}		$c=2$		J_{cv1}	J_{cv2}
			D_{nT}	$D_{nT}^{(N)}$	$D_{nT}^{(PY)}$			D_{nT}	$D_{nT}^{(N)}$	$D_{nT}^{(PY)}$		
						Tests based on asymptotic normal critical values						
1	50	20	0.298	0.172	1.000	4.94	4.55	0.652	0.400	1.000	4.76	4.57
		30	0.308	0.336	1.000	5.22	4.53	0.614	0.520	1.000	5.14	4.60
		40	0.338	0.376	1.000	5.35	4.49	0.680	0.624	1.000	5.13	4.54
		50	0.332	0.264	1.000	5.95	5.66	0.732	0.578	1.000	5.72	5.57
	100	25	0.326	0.236	1.000	5.44	4.69	0.656	0.516	1.000	5.14	4.86
		50	0.346	0.226	1.000	6.48	5.67	0.756	0.588	1.000	6.24	5.74
		75	0.336	0.148	1.000	7.50	6.73	0.880	0.556	1.000	7.08	6.78
		100	0.524	0.252	1.000	7.47	6.71	0.886	0.792	1.000	7.17	6.67
2	50	20	0.216	0.130	1.000	5.26	4.89	0.558	0.338	1.000	5.09	4.86
		30	0.262	0.232	1.000	5.37	4.93	0.592	0.500	1.000	5.29	4.96
		40	0.302	0.290	1.000	5.57	5.02	0.660	0.596	1.000	5.48	5.04
		50	0.378	0.262	1.000	5.79	5.58	0.720	0.614	1.000	5.65	5.50
	100	25	0.170	0.124	1.000	5.88	5.10	0.558	0.294	1.000	5.65	5.07
		50	0.226	0.152	1.000	6.53	6.06	0.596	0.398	1.000	6.31	6.00
		75	0.164	0.102	1.000	7.28	6.69	0.644	0.378	1.000	7.12	6.73
		100	0.234	0.182	1.000	7.42	6.77	0.748	0.538	1.000	7.27	6.70
3	50	20	0.134	0.088	1.000	5.27	4.88	0.282	0.152	1.000	5.12	4.87
		30	0.140	0.152	1.000	5.44	4.95	0.344	0.286	1.000	5.30	4.95
		40	0.186	0.190	1.000	5.56	5.06	0.380	0.344	1.000	5.50	5.03
		50	0.208	0.148	1.000	5.80	5.57	0.380	0.290	1.000	5.77	5.57
	100	25	0.100	0.096	1.000	5.93	5.11	0.308	0.194	1.000	5.82	5.08
		50	0.116	0.100	1.000	6.57	6.13	0.338	0.220	1.000	6.50	6.13
		75	0.136	0.076	1.000	7.35	6.71	0.350	0.234	1.000	7.16	6.73
		100	0.120	0.116	1.000	7.50	6.73	0.408	0.294	1.000	7.44	6.68
							Tests based on bootstrap p-values					
1	50	20	0.208	0.298	0.028			0.548	0.472	0.028		
		30	0.294	0.436	0.036			0.656	0.602	0.036		
		40	0.428	0.486	0.026			0.764	0.648	0.008		
		50	0.540	0.462	0.020			0.876	0.696	0.016		
	100	25	0.262	0.370	0.000			0.696	0.500	0.000		
		50	0.556	0.472	0.000			0.912	0.744	0.000		
		75	0.728	0.518	0.000			0.970	0.792	0.000		
		100	0.794	0.568	0.000			0.974	0.884	0.000		
2	50	20	0.178	0.274	0.308			0.504	0.508	0.270		
		30	0.280	0.376	0.246			0.698	0.648	0.232		
		40	0.444	0.456	0.266			0.794	0.710	0.230		
		50	0.498	0.424	0.270			0.858	0.738	0.232		
	100	25	0.316	0.438	0.704			0.694	0.628	0.614		
		50	0.472	0.440	0.798			0.844	0.658	0.728		
		75	0.538	0.498	0.880			0.892	0.736	0.796		
		100	0.628	0.542	0.896			0.892	0.788	0.814		
3	50	20	0.106	0.204	0.296			0.238	0.302	0.212		
		30	0.144	0.290	0.260			0.380	0.404	0.206		
		40	0.204	0.314	0.244			0.488	0.462	0.180		
		50	0.216	0.274	0.250			0.502	0.414	0.184		
	100	25	0.174	0.392	0.722			0.458	0.480	0.652		
		50	0.244	0.378	0.800			0.652	0.508	0.776		
		75	0.322	0.414	0.880			0.726	0.614	0.826		
		100	0.364	0.454	0.906			0.738	0.634	0.850		

Note: J_{cv1} and J_{cv2} denote the average values of the number of series terms J chosen by the LSCV method for the tests D_{nT} and $D_{nT}^{(N)}$, respectively.

the average number of series terms increase slowly and steadily.

Table 2 reports the power performance of the three tests for poolability. Like Table 1, the upper and lower panels of Table 2 summarize the rejection frequency of the tests based on the asymptotic normal critical values and the bootstrap p-values, respectively. Due to the size distortion of the $D_{nT}^{(N)}$ and $D_{nT}^{(PY)}$ tests, we focus on the D_{nT} test and summarize some main findings. First, the bootstrap version of the D_{nT} test tends to be more powerful than the normal-critical-values-based D_{nT} test. Second, the power of the D_{nT} test is mainly driven by the increase of T. As T increases, the power of the test tends to increase. For fixed T, the power is not necessarily increasing when n increases (see DGP 2, $T = 50$). This is in line with our theoretical findings in the last section because the effect of the increase of n on the power may be cancelling the effect of the increase of $\mathrm{tr}(\overline{P}_w)$ on the power. Intuitively speaking, the larger n is, the more heterogenous regression relationships that need to be estimated under the alternative. This may have adverse effect on the power performance of the test. Similar phenomenon has been found in Pesaran and Yamagata (2008) when they only consider linear functional relationship without unobserved common factors. Third, as the degree of heterogeneity increases (c increases from 1 to 2), the power of the D_{nT} test increases rapidly. For the other two tests, we notice that surprisingly the normal-critical-value-based $D_{nT}^{(PY)}$ test always reject the null across different DGPs, but its bootstrap version can behave quite differently in different DGPs; the $D_{nT}^{(N)}$ test (both bootstrapped and non-bootstrapped versions) has some power but it is less powerful than the D_{nT} test.

Columns 7-8 (resp. 11-12) in Table 2 report the average number of series terms J chosen by the LSCV method in the construction of the (bootstrapped and non-bootstrapped) D_{nT} and $D_{nT}^{(N)}$ tests, respectively, for the case $c = 1$ (resp. 2). As expected, the average number of series terms increase slowly and steadily as either n or T increases.

5 Concluding remarks

In this paper we propose a nonparametric poolability test for semiparametric panel data models with multi-factor error structure. We establish the asymptotic distributions of our test statistics under both the null hypothesis of poolability and a sequence of Pitman local alternatives. In addition, we prove the consistency of the test and justify the validity of a bootstrap method as an alternative to obtain the critical values. Simulations suggest that the proposed test works fairly well in finite samples.

Our test requires sieve estimation of the heterogeneous regression relationships under the alternative. Alternatively, we can propose a test that compares the homogeneous regression estimate with the heterogeneous regression estimate, which requires the selection of two sieve approximating terms. It is also possible to propose a test that only requires estimation under the null hypothesis. In addition, other types of testing procedure are possible. For example, one can extend the specification test of Li, Hsiao, and Zinn (2003) to our framework which relies on the application of empirical process theory, or one can apply the kernel method as Baltagi, Hidalgo, and Li (1996).

Appendix

Let C signify a generic constant whose exact value may vary from case to case. Let $\mathcal{D} = \{(x_{it}, f_{1t}, f_{2t}) : i = 1, \cdots, n, t = 1, \cdots, T\}$. Let $E_{\mathcal{D}}(\cdot)$ and $\mathrm{Var}_{\mathcal{D}}(\cdot)$ denote the conditional expectation and variance given $\mathcal{D}$, respectively. Let $\sum_{1\leq i<j\leq n} \equiv \sum_{i=1}^{n-1}\sum_{j=i+1}^{n}$. Recall $y_i \equiv (y_{i1}, y_{i2}, \cdots, y_{iT})'$, $\varepsilon_i \equiv (\varepsilon_{i1}, \varepsilon_{i2}, \cdots, \varepsilon_{iT})'$, $\overrightarrow{\varepsilon}_i \equiv (p_i' m_b p_i)^- p_i' m_b \varepsilon_i$, and $\overline{P}_w \equiv \int p^K(x) p^K(x)' w(x)\, dx$.

A Proof of results in Section 3

In this appendix, we prove the main results in Section 3. In the next appendix we state and prove several lemmas that are used in the proof of these main results.

Proof of Theorem 3.1

Let $f_1 \equiv (f_{11}, f_{12}, \cdots, f_{1T})'$, $f_2 \equiv (f_{21}, f_{22}, \cdots, f_{2T})'$, and $\mathbf{g}_i \equiv (g_i(x_{i1}), g_i(x_{i2}), \cdots, g_i(x_{iT}))'$. Using (1.1) and the sieve approximation for $g_i(\cdot)$ we have

$$y_i = p_i \alpha_{g_i} + f_1 \gamma_{1i} + f_2 \gamma_{2i} + \varepsilon_i + (\mathbf{g}_i - p_i \alpha_{g_i}). \tag{A.1}$$

Therefore by (2.8), $\widehat{\alpha}_{g_i} = \alpha_{g_i} + \widetilde{\varepsilon}_i + \widetilde{r}_i$, where $\widetilde{\varepsilon}_i \equiv (p_i' m_h p_i)^- p_i' m_h \varepsilon_i$, $\widetilde{r}_i \equiv (p_i' m_h p_i)^- p_i' m_h r_i$, and $r_i \equiv f_2 \gamma_{2i} + (\mathbf{g}_i - p_i \alpha_{g_i})$. Then by (2.9)-(2.10), we have

$$\begin{aligned}
& c_{nT}\Gamma_{nT} \\
= \;& c_{nT} \sum_{1\leq i<j\leq n} \left(\widehat{\alpha}_{g_i} - \widehat{\alpha}_{g_j}\right)' \overline{P}_w \left(\widehat{\alpha}_{g_i} - \widehat{\alpha}_{g_j}\right) \\
= \;& c_{nT} \sum_{1\leq i<j\leq n} \left\{(\alpha_{g_i} - \alpha_{g_j}) + (\widetilde{\varepsilon}_i - \widetilde{\varepsilon}_j) + (\widetilde{r}_i - \widetilde{r}_j)\right\}' \overline{P}_w \left\{(\alpha_{g_i} - \alpha_{g_j}) + (\widetilde{\varepsilon}_i - \widetilde{\varepsilon}_j) + (\widetilde{r}_i - \widetilde{r}_j)\right\} \\
= \;& A_1 + A_2 + A_3 + A_4 + A_5 + A_6,
\end{aligned} \tag{A.2}$$

where the A's are defined as follows:

$$\begin{aligned}
&A_1 \equiv c_{nT} \textstyle\sum_{1\leq i<j\leq n} (\alpha_{g_i} - \alpha_{g_j})' \overline{P}_w (\alpha_{g_i} - \alpha_{g_j}), && A_2 \equiv c_{nT} \textstyle\sum_{1\leq i<j\leq n} (\widetilde{\varepsilon}_i - \widetilde{\varepsilon}_j)' \overline{P}_w (\widetilde{\varepsilon}_i - \widetilde{\varepsilon}_j), \\
&A_3 \equiv c_{nT} \textstyle\sum_{1\leq i<j\leq n} (\widetilde{r}_i - \widetilde{r}_j)' \overline{P}_w (\widetilde{r}_i - \widetilde{r}_j), && A_4 \equiv 2c_{nT} \textstyle\sum_{1\leq i<j\leq n} (\alpha_{g_i} - \alpha_{g_j})' \overline{P}_w (\widetilde{\varepsilon}_i - \widetilde{\varepsilon}_j), \\
&A_5 \equiv 2c_{nT} \textstyle\sum_{1\leq i<j\leq n} \left(\alpha_{g_i} - \alpha_{g_j}\right)' \overline{P}_w (\widetilde{r}_i - \widetilde{r}_j), && A_6 \equiv 2c_{nT} \textstyle\sum_{1\leq i<j\leq n} (\widetilde{\varepsilon}_i - \widetilde{\varepsilon}_j)' \overline{P}_w (\widetilde{r}_i - \widetilde{r}_j).
\end{aligned} \tag{A.3}$$

Under H_0, $A_l = 0$ for $l = 1, 4$, and 5. So it suffices to show that

$$\frac{A_2 - B_{nT}}{\sqrt{V_{nT}}} \xrightarrow{d} N(0,1), \tag{A.4}$$

$$\frac{A_3}{\sqrt{V_{nT}}} = o_p(1), \text{ and } \frac{A_6}{\sqrt{V_{nT}}} = o_p(1). \tag{A.5}$$

We first show (A.4). Note that $E_{\mathcal{D}}(\widetilde{\varepsilon}_i)$ is generally not 0 since the sample mean $\overline{y}_t$ enters the definition of h_t and thus h. By Assumptions 1(i), 3(iv) and 4(iii), and Lemmas B.1(i)-(ii),

$$\begin{aligned} V_{nT} &= \frac{2}{n}\sum_{i=1}^{n}\sigma_i^4 \mathrm{tr}\left(\left(\left(p_i' m_b p_i/T\right)^{-}\overline{P}_w\right)^2\right) \\ &\geq \frac{2\lambda_{\min}\left(\overline{P}_w\right)}{n}\sum_{i=1}^{n}\left[\lambda_{\max}\left(p_i' m_b p_i/T\right)\right]^{-2}\sigma_i^4 \mathrm{tr}(\overline{P}_w) > c > 0. \end{aligned} \quad (A.6)$$

By (A.6), it suffices to prove (A.4) by first establishing that

$$A_2 = c_{nT}\sum_{1\leq i<j\leq n}\left(\overrightarrow{\varepsilon}_i - \overrightarrow{\varepsilon}_j\right)'\overline{P}_w\left(\overrightarrow{\varepsilon}_i - \overrightarrow{\varepsilon}_j\right) + o_p(1) \equiv \overrightarrow{A}_2 + o_p\left(V_{nT}^{1/2}\right) \quad (A.7)$$

and then showing that

$$\frac{\overrightarrow{A}_2 - B_{nT}}{\sqrt{V_{nT}}} \xrightarrow{d} N(0,1). \quad (A.8)$$

where $\overrightarrow{A}_2 \equiv c_{nT}\sum_{1\leq i<j\leq n}\left(\overrightarrow{\varepsilon}_i - \overrightarrow{\varepsilon}_j\right)'\overline{P}_w\left(\overrightarrow{\varepsilon}_i - \overrightarrow{\varepsilon}_j\right).$

To prove (A.7), write

$$\begin{aligned} A_2 - \overrightarrow{A}_2 &= \frac{c_{nT}}{2}\sum_{1\leq i\neq j\leq n}\left\{\left(\widetilde{\varepsilon}_i - \overrightarrow{\varepsilon}_i\right)'\overline{P}_w\left(\widetilde{\varepsilon}_i - \overrightarrow{\varepsilon}_i\right) + \left(\widetilde{\varepsilon}_j - \overrightarrow{\varepsilon}_j\right)'\overline{P}_w\left(\widetilde{\varepsilon}_j - \overrightarrow{\varepsilon}_j\right)\right\} \\ &\quad - c_{nT}\sum_{1\leq i\neq j\leq n}\left(\widetilde{\varepsilon}_i - \overrightarrow{\varepsilon}_i\right)'\overline{P}_w\left(\widetilde{\varepsilon}_j - \overrightarrow{\varepsilon}_j\right) \\ &\quad + 2c_{nT}\sum_{1\leq i\neq j\leq n}\left(\widetilde{\varepsilon}_i - \overrightarrow{\varepsilon}_i\right)'\overline{P}_w\left(\overrightarrow{\varepsilon}_i - \overrightarrow{\varepsilon}_j\right) \\ &\equiv A_{21} - A_{22} + A_{23}. \end{aligned}$$

By Lemma B.4(i) and the Cauchy-Schwarz inequality, $|A_{22}| \leq A_{21} \leq nc_{nT}\sum_{i=1}^{n}\left(\widetilde{\varepsilon}_i - \overrightarrow{\varepsilon}_i\right)'\overline{P}_w(\widetilde{\varepsilon}_i - \overrightarrow{\varepsilon}_i) = O_p\left(K^2/n^{1/2}\right)$. By Lemma B.4(ii), $A_{23} = O_p\left(K^{1/2}/T^{1/2}\right) = o_p\left(K^{1/2}\right)$. Then (A.7) follows by noticing that $V_{nT}^{1/2} = O_p\left(K^{1/2}\right)$ and $K^3/n = o(1)$.

To prove (A.8), let $\varphi\left(\overrightarrow{\varepsilon}_i, \overrightarrow{\varepsilon}_j\right) \equiv T\left(\overrightarrow{\varepsilon}_i - \overrightarrow{\varepsilon}_j\right)'\overline{P}_w\left(\overrightarrow{\varepsilon}_i - \overrightarrow{\varepsilon}_j\right)$. Then we can write $\overrightarrow{A}_2 = \sqrt{n}U_{nT}/2$, where

$$U_{nT} = \frac{2}{n(n-1)}\sum_{1\leq i<j\leq n}\varphi\left(\overrightarrow{\varepsilon}_i, \overrightarrow{\varepsilon}_j\right)$$

is a standard second-order U-statistic with symmetric kernel $\varphi(\cdot,\cdot)$. By the idea of Hoeffding decomposition, we have

$$U_{nT} = \theta + H_{nT}^{(1)} + H_{nT}^{(2)}$$

where $\theta \equiv \frac{2}{n(n-1)}\sum_{1\leq i<j\leq n}\theta_{ij}$, $\theta_{ij} \equiv E_i E_j\left[\varphi\left(\overrightarrow{\varepsilon}_i, \overrightarrow{\varepsilon}_j\right)\right]$,

$$\begin{aligned} H_{nT}^{(1)} &\equiv \frac{2}{n(n-1)}\sum_{1\leq i<j\leq n}\left\{E_i\left[\varphi\left(\overrightarrow{\varepsilon}_i, \overrightarrow{\varepsilon}_j\right)\right] + E_j\left[\varphi\left(\overrightarrow{\varepsilon}_i, \overrightarrow{\varepsilon}_j\right)\right] - 2\theta_{ij}\right\} \\ H_{nT}^{(2)} &\equiv \frac{2}{n(n-1)}\sum_{1\leq i<j\leq n}\left\{\varphi\left(\overrightarrow{\varepsilon}_i, \overrightarrow{\varepsilon}_j\right) - E_j\left[\varphi\left(\overrightarrow{\varepsilon}_i, \overrightarrow{\varepsilon}_j\right)\right] - E_i\left[\varphi\left(\overrightarrow{\varepsilon}_i, \overrightarrow{\varepsilon}_j\right)\right] + \theta_{ij}\right\}, \end{aligned}$$

and $E_j[\cdot]$ denotes expectation taken with respect to $\overrightarrow{\varepsilon}_j$ conditional on $\mathcal{D}$. Straightforward calculations show that $\theta_{ij} = T[E_{\mathcal{D}}(\overrightarrow{\varepsilon}_i'\overline{P}_w\overrightarrow{\varepsilon}_i)+E_{\mathcal{D}}(\overrightarrow{\varepsilon}_j'\overline{P}_w\overrightarrow{\varepsilon}_j)]$ and hence $\theta = \frac{2T}{n}\sum_{i=1}^n E_{\mathcal{D}}(\overrightarrow{\varepsilon}_i'\overline{P}_w\overrightarrow{\varepsilon}_i)$. Noting that conditional on $\mathcal{D}$, $\overrightarrow{\varepsilon}_i$ are INID by Assumptions 1(i), (v) and 2, by the standard U-statistic theory (e.g., Lee, 1990), it is easy to show that $H_{nT}^{(2)} = o_p\left(n^{-1/2}\right)$. It follows that

$$\begin{aligned}
&\overrightarrow{A}_2 - \frac{T}{\sqrt{n}}\sum_{i=1}^n E_{\mathcal{D}}(\overrightarrow{\varepsilon}_i'\overline{P}_w\overrightarrow{\varepsilon}_i) \\
= &\frac{1}{\sqrt{n}(n-1)}\sum_{1\leq i<j\leq n}\{E_i[\varphi(\overrightarrow{\varepsilon}_i,\overrightarrow{\varepsilon}_j)] + E_j[\varphi(\overrightarrow{\varepsilon}_i,\overrightarrow{\varepsilon}_j)] - 2\theta_{ij}\} + o_p(1) \\
= &\frac{T}{\sqrt{n}}\sum_{i=1}^n\left\{\overrightarrow{\varepsilon}_i'\overline{P}_w\overrightarrow{\varepsilon}_i - E_{\mathcal{D}}(\overrightarrow{\varepsilon}_i'\overline{P}_w\overrightarrow{\varepsilon}_i)\right\} + o_p(1).
\end{aligned}$$

Noting that $\frac{T}{\sqrt{n}}\sum_{i=1}^n E_{\mathcal{D}}(\overrightarrow{\varepsilon}_i'\overline{P}_w\overrightarrow{\varepsilon}_i) = B_{nT}$, we have

$$\frac{\overrightarrow{A}_2 - B_{nT}}{\sqrt{V_{nT}}} = \frac{T(V_{nT})^{-1/2}}{\sqrt{n}}\sum_{i=1}^n\left\{\overrightarrow{\varepsilon}_i'\overline{P}_w\overrightarrow{\varepsilon}_i - E_{\mathcal{D}}(\overrightarrow{\varepsilon}_i'\overline{P}_w\overrightarrow{\varepsilon}_i)\right\} + o_p(1).$$

Thus it suffices to prove (A.8) by showing that

$$\mathrm{Var}_{\mathcal{D}}\left(\frac{T(V_{nT})^{-1/2}}{\sqrt{n}}\sum_{i=1}^n\overrightarrow{\varepsilon}_i'\overline{P}_w\overrightarrow{\varepsilon}_i\right) = 1 + o_p(1) \tag{A.9}$$

and verifying the Liapounov condition for the central limit theorem. Let $\mu_{4i} \equiv E\left(\varepsilon_{it}^4\right)$ and $\overline{A}_{ib} \equiv m_b p_i\,(p_i'm_bp_i)^-\,\overline{P}_w\,(p_i'm_bp_i)^-\,p_i'm_b$. Standard variance calculations show that

$$\begin{aligned}
&\mathrm{Var}_{\mathcal{D}}\left(\frac{T}{\sqrt{n}}\sum_{i=1}^n\overrightarrow{\varepsilon}_i'\overline{P}_w\overrightarrow{\varepsilon}_i\right) \\
= &\frac{2}{n}\sum_{i=1}^n\sigma_i^4\mathrm{tr}\left\{\left((p_i'm_bp_i/T)^-\,\overline{P}_w\right)^2\right\} + \frac{1}{n}\sum_{i=1}^n\left(\mu_{4i} - 3\sigma_i^4\right)T^2\mathrm{tr}\left(\overline{A}_{ib}\mathrm{diag}\left(\overline{A}_{ib}\right)\right).
\end{aligned}$$

Let $m_{b,ts}$ and $\overline{a}_{ib,ts}$ denote the (t,s)th element of m_b and $\overline{A}_{ib}$, respectively. Let $\boldsymbol{\iota}_t$ denote the T-vector with one in its tth place and zeros elsewhere. Noting that

$$\begin{aligned}
\sum_{s=1}^T\|p_{is}m_{b,ts}\| &= \sum_{s=1}^T\left\|p_{is}\left(\mathbf{1}(s=t) - b_t'\left(b'b\right)^{-1}b_s\right)\right\| \\
&\leq \|p_{it}\| + T^{-1}\sum_{s=1}^T\left\|b_t'\left(b'b/T\right)^{-1}b_sp_{is}'\right\| \leq \|p_{it}\| + \alpha_T\|b_t\|,
\end{aligned}$$

where $\mathbf{1}(\cdot)$ is the usual indicator function and $\alpha_T \equiv ||\left(b'b/T\right)^{-1}||T^{-1}\sum_{s=1}^T\|p_{is}b_s'\| = O_p(\sqrt{K})$, we have

$$\begin{aligned}
\overline{a}_{ib,tt} &= \boldsymbol{\iota}_t' m_b p_i \left(p_i' m_b p_i\right)^- \overline{P}_w \left(p_i' m_b p_i\right)^- p_i' m_b \boldsymbol{\iota}_t \\
&\leq \lambda_{\max}\left(\overline{P}_w\right) T^{-2} \left[\lambda_{\min}\left(p_i' m_b p_i / T\right)\right]^{-2} \left\|p_i' m_b \boldsymbol{\iota}_t\right\|^2 \\
&= T^{-2} \left[\lambda_{\min}\left(p_i' m_b p_i / T\right)\right]^{-2} \left\|\sum_{s=1}^{T} p_{is} m_{b,ts}\right\| \\
&= T^{-2} \left[\lambda_{\min}\left(p_i' m_b p_i / T\right)\right]^{-2} \left[\left\|p_{it}\right\| + \alpha_T \left\|b_t\right\|\right],
\end{aligned}$$

and $T^2 \text{tr}\left(\overline{A}_{ib} \text{diag}\left(\overline{A}_{ib}\right)\right) = T^2 \sum_{t=1}^{T} \overline{a}_{ib,tt}^2 = T^{-2}\left[\lambda_{\min}\left(p_i' m_b p_i / T\right)\right]^{-4} \sum_{t=1}^{T} \left[\left\|p_{it}\right\| + \alpha_T \left\|b_t\right\|\right]^2$ $= O_p\left(K/T\right) = o_p\left(1\right)$. Thus $\text{Var}_{\mathcal{D}}\left(\frac{T}{\sqrt{n}} \sum_{i=1}^{n} \overrightarrow{\varepsilon}_i' \overline{P}_w \overrightarrow{\varepsilon}_i\right) = V_{nT} + o_p\left(1\right)$. This, together with (A.6), implies (A.9). Now, let

$$\xi_{iT} \equiv \frac{T[\overrightarrow{\varepsilon}_i' \overline{P}_w \overrightarrow{\varepsilon}_i - E_{\mathcal{D}}(\overrightarrow{\varepsilon}_i' \overline{P}_w \overrightarrow{\varepsilon}_i)]}{\text{tr}(\overline{P}_w)} = \frac{T\left[\varepsilon_i' \overline{A}_{ib} \varepsilon_i - E_{\mathcal{D}}\left(\varepsilon_i' \overline{A}_{ib} \varepsilon_i\right)\right]}{\text{tr}(\overline{P}_w)}.$$

Then by the C_r and Jensen inequalities and Theorem 2 of Bao and Ullah (2009),

$$\begin{aligned}
E_{\mathcal{D}}\left[\xi_{iT}^4\right] &\leq 16\sigma_i^4 T^4 [\text{tr}\overline{P}_w]^{-4} \{(\text{tr}\overline{A}_{ib})^4 + 12\left(\text{tr}\overline{A}_{ib}\right)^2 \text{tr}(\overline{A}_{ib}^2) + 12(\text{tr}(\overline{A}_{ib}^2))^2 \\
&\quad +32\left(\text{tr}\overline{A}_{ib}\right) \text{tr}(\overline{A}_{ib}^3) + 48\text{tr}(\overline{A}_{ib}^4) + \text{remainder terms}\},
\end{aligned} \tag{A.10}$$

where the expression of the remainder terms (which vanish if ε_{it} is normally distributed) is tedious and can be found from Bao and Ullah. By some tedious algebra, we can show that each term on the right hand side of (A.10) is of order $O_p(1)$ or smaller by using the fact that $\text{tr}\left(\overline{A}_{ib}\right) = \text{tr}((p_i' m_b p_i)^- \overline{P}_w) \leq T^{-1}[\lambda_{\min}\left(p_i' m_b p_i / T\right)]^{-1} \text{tr}\left(\overline{P}_w\right) = O_p\left(T^{-1} \text{tr}\overline{P}_w\right)$ and that all elements of $\overline{A}_{ib}$ are of order $O_p(K^{3/2} T^{-2})$. On the other hand, $\sum_{i=1}^{n} E_{\mathcal{D}}\left[\xi_{iT}^2\right] = \frac{n}{2} V_{nT} / [\text{tr}\overline{P}_w]^2$. Then by (A.6) and Assumption 5, we have

$$\frac{\sum_{i=1}^{n} E_{\mathcal{D}}\left[\xi_{iT}^4\right]}{\left\{\sum_{i=1}^{n} E_{\mathcal{D}}\left[\xi_{iT}^2\right]\right\}^2} = \frac{O_p\left(n\right)}{\left[\frac{n}{2} V_{nT} / (\text{tr}\overline{P}_w)^2\right]^2} = O_p\left(\frac{(\text{tr}\overline{P}_w)^4}{n V_{nT}^2}\right) = O_p\left(\frac{K^2}{n}\right) = o_p\left(1\right).$$

This verifies the Liapounov condition.

We now show (A.5). Let $\overrightarrow{A}_3 \equiv c_{nT} \sum_{1 \leq i < j \leq n} \left(\overrightarrow{r}_i - \overrightarrow{r}_j\right)' \overline{P}_w \left(\overrightarrow{r}_i - \overrightarrow{r}_j\right)$. We prove (iii) by showing that $A_3 - \overrightarrow{A}_3 = o_p(V_{nT}^{1/2})$, and $\overrightarrow{A}_3 = o_p(V_{nT}^{1/2})$. First, we decompose $A_3 - \overrightarrow{A}_3$ as follows

$$\begin{aligned}
A_3 - \overrightarrow{A}_3 &= \frac{c_{nT}}{2} \sum_{1 \leq i \neq j \leq n} \left\{\left(\widetilde{r}_i - \overrightarrow{r}_i\right)' \overline{P}_w \left(\widetilde{r}_i - \overrightarrow{r}_i\right) + \left(\widetilde{r}_j - \overrightarrow{r}_j\right)' \overline{P}_w \left(\widetilde{r}_j - \overrightarrow{r}_j\right)\right\} \\
&\quad - c_{nT} \sum_{1 \leq i \neq j \leq n} \left(\widetilde{r}_i - \overrightarrow{r}_i\right)' \overline{P}_w \left(\widetilde{r}_j - \overrightarrow{r}_j\right) \\
&\quad + 2c_{nT} \sum_{1 \leq i \neq j \leq n} \left(\widetilde{r}_i - \overrightarrow{r}_i\right)' \overline{P}_w \left(\overrightarrow{r}_i - \overrightarrow{r}_j\right) \\
&\equiv A_{31} - A_{32} + A_{33}.
\end{aligned}$$

By Lemma B.4(iii) and the Cauchy-Schwarz inequality, $|A_{32}| \leq A_{31} \leq nc_{nT}\sum_{i=1}^{n}(\widetilde{r}_i - \overrightarrow{r}_i)'\overline{P}_w(\widetilde{r}_i - \overrightarrow{r}_i) = O_p\left(Tn^{-1/2}K^{1-2\lambda/d} + KT/n^{1/2}\right) = o_p(K^{1/2})$. By Lemma B.4(iv), $A_{33} = O_p(Tn^{1/2}K^{1-2\lambda/d}) = o_p(K^{1/2})$. It follows that $A_3 = o_p(V_{nT}^{1/2})$.

Now, we decompose A_6 as follows

$$
\begin{aligned}
A_6 &= 2c_{nT}\sum_{1\leq i<j\leq n}\{(\widetilde{\varepsilon}_i - \overrightarrow{\varepsilon}_i - \widetilde{\varepsilon}_j + \overrightarrow{\varepsilon}_j)'\overline{P}_w(\widetilde{r}_i - \overrightarrow{r}_i - \widetilde{r}_j + \overrightarrow{r}_j) \\
&\quad + (\widetilde{\varepsilon}_i - \overrightarrow{\varepsilon}_i - \widetilde{\varepsilon}_j + \overrightarrow{\varepsilon}_j)'\overline{P}_w(\overrightarrow{r}_i - \overrightarrow{r}_j) + (\overrightarrow{\varepsilon}_i - \overrightarrow{\varepsilon}_j)'\overline{P}_w(\widetilde{r}_i - \overrightarrow{r}_i - \widetilde{r}_j + \overrightarrow{r}_j) \\
&\quad + (\overrightarrow{\varepsilon}_i - \overrightarrow{\varepsilon}_j)'\overline{P}_w(\overrightarrow{r}_i - \overrightarrow{r}_j)\} \\
&\equiv A_{61} + A_{62} + A_{63} + A_{64}.
\end{aligned}
$$

where, e.g., $A_{61} \equiv 2c_{nT}\sum_{1\leq i<j\leq n}(\widetilde{\varepsilon}_i - \overrightarrow{\varepsilon}_i - \widetilde{\varepsilon}_j + \overrightarrow{\varepsilon}_j)'\overline{P}_w(\widetilde{r}_i - \overrightarrow{r}_i - \widetilde{r}_j + \overrightarrow{r}_j)$. Then by the repeated use of the Cauchy-Schwarz inequality and Lemmas B.4(i) and (iii) ,

$$
\begin{aligned}
|A_{61}| &\leq 8c_{nT}\sum_{i=1}^{n}(\widetilde{\varepsilon}_i - \overrightarrow{\varepsilon}_i)'\overline{P}_w(\widetilde{r}_i - \overrightarrow{r}_i) \\
&\leq 8\left\{c_{nT}\sum_{i=1}^{n}(\widetilde{\varepsilon}_i - \overrightarrow{\varepsilon}_i)'\overline{P}_w(\widetilde{\varepsilon}_i - \overrightarrow{\varepsilon}_i)\right\}^{1/2}\left\{c_{nT}\sum_{i=1}^{n}(\widetilde{r}_i - \overrightarrow{r}_i)'\overline{P}_w(\widetilde{r}_i - \overrightarrow{r}_i)\right\}^{1/2} \\
&= \left\{O_p\left(K^2/n^{1/2}\right)O_p\left(Tn^{-1/2}K^{1-2\lambda/d} + KT/n^{1/2}\right)\right\}^{1/2} = o_p\left(K^{1/2}\right).
\end{aligned}
$$

Similarly, by Lemmas B.4(i) and (iv), we have $|A_{62}| = \left\{O_p\left(K^2/n^{1/2}\right)O_p\left(Tn^{1/2}K^{1-2\lambda/d}\right)\right\}^{1/2} = o_p\left(K^{1/2}\right)$. By Lemmas B.4(v) and (vi), $A_{6s} = o_p\left(K^{1/2}\right)$ for $s = 3, 4$. This completes the proof of the theorem. ■

Proof of Corollary 3.2

Noting that $V_{nT} \geq C\mathrm{tr}(\overline{P}_w) = O(K)$ for some $C > 0$ by (A.6), it suffices to prove the corollary by showing that

$$
\widehat{V}_{nT} = V_{nT} + o_p(V_{nT}), \text{ and } \widehat{B}_{nT} - B_{nT} = o_p\left(K^{1/2}\right).
$$

First, we write

$$
\begin{aligned}
\widehat{V}_{nT} - V_{nT} &= \frac{2}{n}\sum_{i=1}^{n}\widehat{\sigma}_i^4\mathrm{tr}\left(\left((p_i'm_hp_i/T)^{-}\overline{P}_w\right)^2\right) - \frac{2}{n}\sum_{i=1}^{n}\sigma_i^4\mathrm{tr}\left(\left((p_i'm_bp_i/T)^{-}\overline{P}_w\right)^2\right) \\
&= \frac{2}{n}\sum_{i=1}^{n}\left(\widehat{\sigma}_i^4 - \sigma_i^4\right)\mathrm{tr}\left(\left((p_i'm_hp_i/T)^{-}\overline{P}_w\right)^2\right) \\
&\quad + \frac{2}{n}\sum_{i=1}^{n}\sigma_i^4\mathrm{tr}\left(\left((p_i'm_hp_i/T)^{-}\overline{P}_w\right)^2 - \left((p_i'm_bp_i/T)^{-}\overline{P}_w\right)^2\right) \\
&\equiv 2V_{1nT} + 2V_{2nT}.
\end{aligned}
$$

We want to show $V_{snT} = o_p(K)$ for $s = 1, 2$. Let $\overline{m}_{ih} \equiv m_h - m_h p_i (p_i' m_h p_i)^- p_i' m_h$ and $\overline{m}_{ib} \equiv m_b - m_b p_i (p_i' m_b p_i)^- p_i' m_b$. Then $\widehat{e}_i = m_h (y_i - \widehat{\mathbf{g}}_i) = \overline{m}_{ih} (\varepsilon_i + r_i)$ where recall $r_i \equiv f_2 \gamma_{2i} + (\mathbf{g}_i - p_i \alpha_{g_i})$. It follows that

$$\widehat{\sigma}_i^2 = \frac{1}{T} \widehat{e}_i' \widehat{e}_i = \frac{1}{T} \varepsilon_i' \overline{m}_{ih} \varepsilon_i + \frac{1}{T} r_i' \overline{m}_{ih} r_i + \frac{2}{T} r_i' \overline{m}_{ih} \varepsilon_i, \tag{A.11}$$

and

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^{n} \left(\widehat{\sigma}_i^2 - \sigma_i^2 \right)^2 &\leq \frac{3}{n} \sum_{i=1}^{n} \left(\frac{1}{T} \varepsilon_i' \overline{m}_{ih} \varepsilon_i - \sigma_i^2 \right)^2 + \frac{3}{n} \sum_{i=1}^{n} \left(\frac{1}{T} r_i' \overline{m}_{ih} r_i \right)^2 + \frac{12}{n} \sum_{i=1}^{n} \left(\frac{1}{T} r_i' \overline{m}_{ih} \varepsilon_i \right)^2 \\ &\equiv 3\delta_{1nT} + 3\delta_{2nT} + 12\delta_{3nT}. \end{aligned}$$

For δ_{1nT}, we can use the decomposition in (B.1) as in the proof of Lemma B.4(ii) to show that $\delta_{1nT} = \overline{\delta}_{1nT} + O_p(K/T)$, where $\overline{\delta}_{1nT} \equiv \frac{1}{n} \sum_{i=1}^{n} \left(\frac{1}{T} \varepsilon_i' \overline{m}_{ib} \varepsilon_i - \sigma_i^2 \right)^2$. Let $\overline{m}_{ib,ts}$ denote the (t, s) element of $\overline{m}_{ib}$. Noting that $\text{tr}(\overline{m}_{ib}) = T - K - q_1 - q_2$, and $\frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} (\overline{m}_{ib,ts})^2 = 1 + O_p(K/T)$ uniformly in i, we have

$$\begin{aligned} & E_{\mathcal{D}} \left(\overline{\delta}_{1nT} \right) \\ = & \frac{1}{n} \sum_{i=1}^{n} E_{\mathcal{D}} \left(\frac{1}{T^2} \varepsilon_i' \overline{m}_{ib} \varepsilon_i \varepsilon_i' \overline{m}_{ib} \varepsilon_i - \frac{2}{T} \varepsilon_i' \overline{m}_{ib} \varepsilon_i \sigma_i^2 + \sigma_i^4 \right) \\ = & \frac{1}{n} \sum_{i=1}^{n} \left[\frac{1}{T^2} \left\{ \sum_{t=1}^{T} \sum_{s=1}^{T} (\overline{m}_{ib,ts})^2 \sigma_i^4 + \sum_{t=1}^{T} (\overline{m}_{ib,tt})^2 \left[E\left(\varepsilon_{it}^4 \right) - 3\sigma_i^4 \right] \right\} - \frac{2\text{tr}(\overline{m}_{ib})}{T} \sigma_i^4 + \sigma_i^4 \right] \\ = & O_p(K/T). \end{aligned}$$

Hence $\delta_{1nT} = O_p(K/T)$ by the conditional Markov inequality. For δ_{2nT}, noting that $\overline{m}_{ih} = m_h \overline{m}_{ih} = \overline{m}_{ih} m_h$ and that both m_h and $\overline{m}_{ih}$ are projection matrices, we have

$$\frac{1}{T} r_i' \overline{m}_{ih} r_i \leq \frac{1}{T} \| m_h r_i \|^2 \leq \frac{2}{T} \| m_h f_2 \gamma_{2i} \|^2 + \frac{2}{T} \| m_h (\mathbf{g}_i - p_i \alpha_{g_i}) \|^2 .$$

It follows that $\delta_{2nT} \leq \frac{8}{nT^2} \sum_{i=1}^{n} \| (m_h - m_b) f_2 \gamma_{2i} \|^4 + \frac{8}{nT^2} \sum_{i=1}^{n} \| m_h (\mathbf{g}_i - p_i \alpha_{g_i}) \|^4 = O_p(n^{-1}$ $+ K^{-4\lambda/d})$ where $n^{-1} T^{-2} \sum_{i=1}^{n} \| (m_h - m_b) f_2 \gamma_{2i} \|^4 = O_p\left(n^{-1}\right)$ can be proved analogously to the proof of Lemma A5(v) of Su and Jin (2010), and $\frac{1}{nT^2} \sum_{i=1}^{n} \| (\mathbf{g}_i - p_i \alpha_{g_i}) \|^4 = O_p(K^{-4\lambda/d})$ by arguments similar to the proof of Lemma A.2 of Su and Jin (2010) under the strengthened condition given in the corollary. For δ_{3nT}, we can show that $\delta_{3nT} = \overline{\delta}_{3nT} + O_p(K/T)$, where $\overline{\delta}_{3nT} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{1}{T} r_i' \overline{m}_{ib} \varepsilon_i \right)^2$. Then noting that $\overline{m}_{ib} r_i = \overline{m}_{ib} (\mathbf{g}_i - p_i \alpha_{g_i})$ and $\overline{m}_{ib}$ is a projection matrix, by Lemma (B.2)(ii) we have

$$E_{\mathcal{D}} \left(\overline{\delta}_{3nT} \right) = \frac{1}{nT^2} \sum_{i=1}^{n} r_i' \overline{m}_{ib} E\left(\varepsilon_i \varepsilon_i' \right) \overline{m}_{ib} r_i \leq \frac{C}{nT^2} \sum_{i=1}^{n} \| \mathbf{g}_i - p_i \alpha_{g_i} \|^2 = O_p\left(T^{-1} K^{-2\lambda/d} \right),$$

Consequently, $\delta_{3nT} = O_p\left(K/T + T^{-1} K^{-2\lambda/d} \right) = O_p(K/T)$, and

$$\frac{1}{n} \sum_{i=1}^{n} \left(\widehat{\sigma}_i^2 - \sigma_i^2 \right)^2 = O_p(K/T).$$

Similarly, we can show that $\frac{1}{n}\sum_{i=1}^{n}(\widehat{\sigma}_i^2+\sigma_i^2)^2=O_p(1)$. It follows that

$$\begin{aligned} V_{1nT} &= \frac{1}{n}\sum_{i=1}^{n}\left(\widehat{\sigma}_i^2-\sigma_i^2\right)\left(\widehat{\sigma}_i^2+\sigma_i^2\right)\mathrm{tr}\left(\left(\left(p_i'm_hp_i/T\right)^{-}\overline{P}_w\right)^2\right) \\ &\leq (c_{2\lambda})^{-2}\lambda_{\max}\left(\overline{P}_w\right)\mathrm{tr}\left(\overline{P}_w\right)\left\{\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{\sigma}_i^2-\sigma_i^2\right)^2\right\}^{1/2}\left\{\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{\sigma}_i^2+\sigma_i^2\right)^2\right\}^{1/2} \\ &= O(K)\,O_p(\sqrt{K/T})O_p(1)=o_p(K). \end{aligned}$$

Noting that σ_i^2 is uniformly bounded, $V_{2nT}=O_p(K/\sqrt{n})=o_p(K)$ by Lemma B.4(iii). Consequently, $\widehat{V}_{nT}-V_{nT}=o_p(V_{nT})$ as $V_{nT}=O_p(K)$.

Next, write

$$\begin{aligned} &\widehat{B}_{nT}-B_{nT} \\ =\;& \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widehat{\sigma}_i^2\mathrm{tr}\left(\left(p_i'm_hp_i/T\right)^{-}\overline{P}_w\right)+\widehat{b}_{nT}-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sigma_i^2\mathrm{tr}\left(\left(p_i'm_bp_i/T\right)^{-}\overline{P}_w\right)\equiv B_1+B_2, \end{aligned}$$

where $B_1\equiv\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\widehat{\sigma}_i^2-\sigma_i^2)\mathrm{tr}\Big(\left(p_i'm_bp_i/T\right)^{-}\overline{P}_w\Big)$, and $B_2\equiv\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widehat{\sigma}_i^2\mathrm{tr}\{[(p_i'm_hp_i/T)^{-}-(p_i'm_bp_i/T)^{-}]\overline{P}_w\}+\widehat{b}_{nT}$. Using (A.11) we can show that $B_1=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{1}{T}\varepsilon_i'\overline{m}_{ib}\varepsilon_i-\sigma_i^2\right)\times\mathrm{tr}\Big(\left(p_i'm_bp_i/T\right)^{-}\overline{P}_w\Big)+o_p(1)=\overrightarrow{B}_1+o_p(1)$, where $\overrightarrow{B}_1\equiv\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\frac{1}{T-q_1-d-1}\varepsilon_i'\overline{m}_{ib}\varepsilon_i-\sigma_i^2)\mathrm{tr}\Big(\left(p_i'm_bp_i/T\right)^{-}\overline{P}_w\Big)$. Simple calculations reveal that $E_{\mathcal{D}}(\overrightarrow{B}_1)=0$ and

$$\begin{aligned} \mathrm{Var}_{\mathcal{D}}(\overrightarrow{B}_1) &= \frac{1}{n\left(T-q_1-d-1\right)^2}\sum_{i=1}^{n}\mathrm{Var}_{\mathcal{D}}(\varepsilon_i'\overline{m}_{ib}\varepsilon_i)\left\{\mathrm{tr}\left(\left(p_i'm_bp_i/T\right)^{-}\overline{P}_w\right)\right\}^2 \\ &= O_p\left(K^2/T\right)=o_p(1) \end{aligned}$$

as $\mathrm{Var}_{\mathcal{D}}(\varepsilon_i'\overline{m}_{ib}\varepsilon_i)\leq\mathrm{Var}(\varepsilon_i'\varepsilon_i)=O(T)$. Hence $B_1=o_p(1)$.

Now we show that $B_2=o_p\left(K^{1/2}\right)$. Note that

$$\begin{aligned} B_2 &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widehat{\sigma}_i^2T\mathrm{tr}\left(\left(p_i'm_hp_i\right)^{-}\left[p_i'm_bp_i-p_i'm_hp_i\right]\left(p_i'm_bp_i\right)^{-}\overline{P}_w\right)+\widehat{b}_{nT} \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widehat{\sigma}_i^2T^{-1}\mathrm{tr}\left(\left[p_i'\left(m_b-m_h\right)p_i+\widehat{b}_{inT}\right]\beta_i\right) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sigma_i^2T^{-1}\mathrm{tr}\left(\left[p_i'\left(m_b-m_h\right)p_i+\widehat{b}_{inT}\right]\beta_i\right) \\ &\quad+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\widehat{\sigma}_i^2-\sigma_i^2\right)T^{-1}\mathrm{tr}\left(\left[p_i'\left(m_b-m_h\right)p_i+\widehat{b}_{inT}\right]\beta_i\right) \\ &\equiv B_{21}+B_{22}, \end{aligned}$$

where $\beta_i\equiv\left(p_i'm_hp_i/T\right)^{-}\overline{P}_w\left(p_i'm_bp_i/T\right)^{-}$. It suffices to prove $B_{21}=o_p\left(K^{1/2}\right)$ and $B_{22}=o_p\left(K^{1/2}\right)$. We only prove the former result since the latter can be established analogously based on the decomposition of $\widehat{\sigma}_i^2$ in (A.11). Let $\overline{\varepsilon} \equiv (\overline{\varepsilon}_1, \cdots, \overline{\varepsilon}_T)'$, $\overline{v} \equiv (\overline{v}_1, \cdots, \overline{v}_T)'$, $\overline{g} \equiv (\overline{g}_1, \cdots, \overline{g}_T)'$, and $\widehat{\overline{g}} \equiv (\widehat{\overline{g}}_t, \cdots, \widehat{\overline{g}}_T)'$. Then

$$\begin{aligned}
B_{21} &= \frac{1}{n^{1/2}T}\sum_{i=1}^{n}\sigma_i^2 \mathrm{tr}\left(\left[p_i'\left(m_b - m_h\right)p_i + \widehat{b}_{inT}\right]\beta_i\right) \\
&= \frac{1}{n^{1/2}T}\sum_{i=1}^{n}\sigma_i^2 \mathrm{tr}\left\{p_i'[h\left(h'h\right)^{-1}h - b\left(b'b\right)^{-1}b]p_i\beta_i\right. \\
&\quad \left. + p_i'[-\widehat{\overline{g}}^*\left(h'h\right)^{-1}h' - h\left(h'h\right)^{-1}\widehat{\overline{g}}^* + h\left(h'h\right)^{-1}\widehat{\overline{g}}^{*\prime}\widehat{\overline{g}}^*\left(h'h\right)^{-1}h']p_i\beta_i\right\} \\
&= \frac{1}{n^{1/2}T}\sum_{i=1}^{n}\sigma_i^2 \mathrm{tr}\left\{p_i'[(h-b) - \widehat{\overline{g}}^*]\left(h'h\right)^{-1}h'p_i\beta_i\right\} \\
&\quad + \frac{1}{n^{1/2}T}\sum_{i=1}^{n}\sigma_i^2 \mathrm{tr}\left\{p_i'h\left(h'h\right)^{-1}[(b'b - h'h) + \widehat{\overline{g}}^{*\prime}\widehat{\overline{g}}^*]\left(h'h\right)^{-1}h'p_i\beta_i\right\} \\
&\quad + \frac{1}{n^{1/2}T}\sum_{i=1}^{n}\sigma_i^2 \mathrm{tr}\left\{p_i'h\left(h'h\right)^{-1}[(h-b) - \widehat{\overline{g}}^*]'p_i\beta_i\right\} + O_p(K/T^{1/2}) \\
&= B_{21,1} + B_{21,2} + B_{21,3} + O_p(K/T^{1/2}), \text{ say,}
\end{aligned}$$

where the $O_p\left(K/T^{1/2}\right)$ term comes from the replacement of $b'b/T$ and $h'p_i/T$ by $h'h/T$ and $b'p_i$, respectively. We only show that $B_{21,1} = o_p(\sqrt{K})$ as one can prove $B_{21,s} = o_p(\sqrt{K/n})$, $s = 2, 3$, analogously. By the proof of Lemma A.5(i) in Su and Jin (2010), $T^{-1}\sum_{i=1}^{n}(\|p_i'\overline{\varepsilon}\| + \|p_i'\overline{v}]\|) = O_p(\sqrt{K} + \sqrt{Kn/T})$. It follows that

$$\begin{aligned}
|B_{21}| &\leq \frac{\max_{1\leq i\leq n}\sigma_i^2}{n^{1/2}T}\sum_{i=1}^{n}\left|\mathrm{tr}\left\{p_i'\left(\widehat{\overline{g}} - \overline{g}\right)\left(h'h\right)^{-1}h'p_i\beta_i\right\}\right| + O_p(K/n^{1/2} + K/T^{1/2}) \\
&\leq C\overline{B}_{21,1} + o_p(\sqrt{K}),
\end{aligned}$$

where $\overline{B}_{21,1} \equiv \frac{1}{n^{1/2}T}\sum_{i=1}^{n}|\mathrm{tr}\{p_i'(\widehat{\overline{g}} - \overline{g})\left(h'h\right)^{-1}h'p_i\beta_i\}|$. By Theorem 4.2 of Su and Jin (2010), we can show that $\widehat{\overline{g}}_t - \overline{g}_t = \frac{1}{n}\sum_{j=1}^{n}[\widehat{g}_j\left(x_{jt}\right) - g_j\left(x_{jt}\right)] = O_p(\sqrt{K/nT})$. By the Cauchy-Schwarz inequality and the fact that $\|\beta_i\| = O_p(\sqrt{K})$ uniformly in i, we have

$$\begin{aligned}
\overline{B}_{21,1} &\leq \frac{1}{n^{1/2}T}\sum_{i=1}^{n}\left\{\mathrm{tr}\left(p_i'\left(\widehat{\overline{g}} - \overline{g}\right)\left(h'h\right)^{-1}\left(\widehat{\overline{g}} - \overline{g}\right)'p_i\right)\right\}^{1/2}\left\{\mathrm{tr}\left(\beta_i'p_i'h\left(h'h\right)^{-1}h'p_i\beta_i\right)\right\}^{1/2} \\
&\leq \frac{1}{n^{1/2}T}\sum_{i=1}^{n}\left\{\mathrm{tr}\left(p_i'\left(\widehat{\overline{g}} - \overline{g}\right)\left(h'h\right)^{-1}\left(\widehat{\overline{g}} - \overline{g}\right)'p_i\right)\right\}^{1/2}\left\{\mathrm{tr}\left(\beta_i'p_i'p_i\beta_i\right)\right\}^{1/2} \\
&\leq \left[\lambda_{\min}\left(h'h/T\right)\right]^{-1}\max_{1\leq i\leq n}\lambda_{\max}\left(p_i'p_i/T\right)\frac{1}{n^{1/2}T}\sum_{i=1}^{n}\left\|p_i'\left(\widehat{\overline{g}} - \overline{g}\right)\right\|\|\beta_i\| \\
&= O_p\left(\sqrt{K}\right)\left\|\widehat{\overline{g}} - \overline{g}\right\|\left\{\frac{1}{n^{1/2}T}\sum_{i=1}^{n}\|p_i\|\right\} = O_p\left(\sqrt{K}\right)O_p\left(\sqrt{K/n}\right)O_p\left(\sqrt{nK/T}\right) \\
&= O_p\left(\sqrt{K^3/T}\right) = o_p\left(\sqrt{K}\right).
\end{aligned}$$

Consequently, $B_{21,1} = o_p\left(\sqrt{K}\right)$. This completes the proof. ■

Proof of Theorem 3.3

The proof follows closely from that of Theorems 3.1 and 3.2, now keeping the additional terms that do not vanish under $H_1\left(n^{-1/4}T^{-1/2}\text{tr}\left(\overline{P}_w\right)^{1/4}\right)$. Since the proof in Theorem 3.2 does not impose the null hypothesis, by (A.6) it suffices to show that under $H_1\left(n^{-1/4}T^{-1/2}\text{tr}\left(\overline{P}_w\right)^{1/4}\right)$,

$$A_1/\sqrt{V_{nT}} \xrightarrow{p} \Delta, \tag{A.12}$$

$$A_4 = o_p\left(\text{tr}(\overline{P}_w)^{1/2}\right), \text{ and } A_5 = o_p\left(\text{tr}(\overline{P}_w)^{1/2}\right). \tag{A.13}$$

where A_1, A_4, and A_5 are defined in (A.3). Under $H_1\left(n^{-1/4}T^{-1/2}\text{tr}\left(\overline{P}_w\right)^{1/4}\right)$, $g_i(x) - g_j(x) = \gamma_{nT}\Delta_{ij,n}(x)$ with $\Delta_{ij,n}(x) \equiv \Delta_{in}(x) - \Delta_{jn}(x)$. It follows that

$$\begin{aligned}
A_1 &= c_{nT}\sum_{1\leq i<j\leq n}\int\left(p^K(x)'\left(\alpha_{g_i} - \alpha_{g_j}\right)\right)^2 w(x)\,dx \\
&= c_{nT}\sum_{1\leq i<j\leq n}\int\left\{\gamma_{nT}\Delta_{ij,n}(x) + \left[d_{g_j}(x) - d_{g_i}(x)\right]\right\}^2 w(x)\,dx \\
&= A_{11} + A_{12} + A_{13} - A_{14} - A_{15},
\end{aligned}$$

where $d_{g_i}(x) \equiv g_i(x) - p^K(x)'\alpha_{g_i}$, $A_{11} \equiv c_{nT}\gamma_{nT}^2\sum_{1\leq i<j\leq n}\int\Delta_{ij,n}^2(x)\,w(x)\,dx$, $A_{12} \equiv \frac{T}{\sqrt{n}}\sum_{i=1}^n\int d_{g_i}^2(x)\,w(x)\,dx$, $A_{13} \equiv 2\gamma_{nT}c_{nT}\sum_{1\leq i<j\leq n}\int\Delta_{ij,n}(x)\,d_{g_j}(x)\,w(x)\,dx$, $A_{14} \equiv 2\gamma_{nT}c_{nT}\sum_{1\leq i<j\leq n}\int\Delta_{ij,n}(x)\,d_{g_i}(x)\,w(x)\,dx$, and $A_{15} \equiv 2c_{nT}\sum_{1\leq i<j\leq n}\int d_{g_i}(x)\,d_{g_j}(x)\,w(x)\,dx$.

First, $A_{11}/\sqrt{V_{nT}} = \left(\text{tr}\left(\overline{P}_w\right)/V_{nT}\right)^{1/2}\frac{1}{n(n-1)}\sum_{1\leq i<j\leq n}\int\left(\Delta_{in}(x) - \Delta_{jn}(x)\right)^2 w(x)\,dx \to \Delta$. By Assumptions 3(iii), 3(vi) and 5,

$$A_{12} \leq n^{-1/2}T\sum_{i=1}^n ||d_{g_i}(\cdot)||_{\infty,\overline{\omega}_i}\int\left(1 + \|x\|^2\right)^{\overline{\omega}_i} w(x)\,dx = O\left(n^{1/2}TK^{-2\lambda/d}\right) = o(1).$$

Similarly, $A_{13} = O\left(n^2\gamma_{nT}c_{nT}K^{-\lambda/d}\right) = o(1)$, $A_{14} = O(n^2\gamma_{nT}\ c_{nT}K^{-\lambda/d}) = o(1)$, and $A_{15} = O\left(n^2c_{nT}K^{-2\lambda/d}\right) = o(1)$. Consequently, (A.12) follows.

Now write

$$A_4 = 2c_{nT}\sum_{1\leq i<j\leq n}\int\left\{\gamma_{nT}\Delta_{ij,n}(x) - \left[d_{g_i}(x) - d_{g_j}(x)\right]\right\}p^K(x)'\left(\widetilde{\varepsilon}_i - \widetilde{\varepsilon}_j\right)w(x)\,dx = A_{41} - A_{42},$$

where $A_{41} \equiv 2c_{nT}\gamma_{nT}\sum_{1\leq i<j\leq n}\int\Delta_{ij,n}(x)\,p^K(x)'\left(\widetilde{\varepsilon}_i - \widetilde{\varepsilon}_j\right)w(x)\,dx$, and $A_{42} \equiv 2c_{nT}\sum_{1\leq i<j\leq n}\int\left(d_{g_i} - d_{g_j}\right)p^K(x)'\left(\widetilde{\varepsilon}_i - \widetilde{\varepsilon}_j\right)w(x)\,dx$. Let $\overline{\gamma}_{nT} \equiv \gamma_{nT}/\text{tr}(\overline{P}_w)^{1/2} = n^{-1/4}T^{-1/2}\text{tr}(\overline{P}_w)^{-1/4}$.

Analogously to the proof of Lemma B.4(i) by replacing $\widetilde{\varepsilon}_i$ with $\overrightarrow{\varepsilon}_i$, we can show that

$$
\begin{aligned}
& A_{41}/\mathrm{tr}(\overline{P}_w)^{1/2} \\
= \; & c_{nT}\overline{\gamma}_{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\int\left(\Delta_{in}(x)-\Delta_{jn}(x)\right)p^K(x)'\left(\widetilde{\varepsilon}_i-\widetilde{\varepsilon}_j\right)w(x)\,dx \\
= \; & 2nc_{nT}\overline{\gamma}_{nT}\sum_{i=1}^{n}\int\Delta_{in}(x)\,p^K(x)'\widetilde{\varepsilon}_iw(x)\,dx-2c_{nT}\overline{\gamma}_{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\int\Delta_{in}(x)\,p^K(x)'\widetilde{\varepsilon}_jw(x)\,dx \\
= \; & 2nc_{nT}\overline{\gamma}_{nT}\sum_{i=1}^{n}\int\Delta_{in}(x)\,p^K(x)'\overrightarrow{\varepsilon}_iw(x)\,dx-2c_{nT}\gamma_{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\int\Delta_{in}(x)\,p^K(x)'\overrightarrow{\varepsilon}_jw(x)\,dx \\
& +n^2c_{nT}\overline{\gamma}_{nT}O_p(K/\sqrt{nT}) \\
= \; & \overrightarrow{A}_{41a}-\overrightarrow{A}_{41b}+o_p(1),
\end{aligned}
$$

where $\overrightarrow{A}_{41a}\equiv 2nc_{nT}\overline{\gamma}_{nT}\sum_{i=1}^{n}\int\Delta_{in}(x)\,p^K(x)'\overrightarrow{\varepsilon}_iw(x)\,dx$ and $\overrightarrow{A}_{41b}\equiv 2c_{nT}\overline{\gamma}_{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}$ $\int\Delta_{in}(x)\,p^K(x)'\overrightarrow{\varepsilon}_jw(x)\,dx$. Noting that $E_{\mathcal{D}}(\overrightarrow{A}_{41a})=0$ and

$$
\begin{aligned}
& \mathrm{Var}_{\mathcal{D}}(\overrightarrow{A}_{41a}) \\
= \; & 4n^2c_{nT}^2\overline{\gamma}_{nT}^2\sum_{i=1}^{n}\sigma_i^2\int\Delta_{in}(x)\,p^K(x)'w(x)\,dx\left(p_i'm_bp_i\right)^{-1}\int p^K(\widetilde{x})\,\Delta_{in}(\widetilde{x})\,w(\widetilde{x})\,d\widetilde{x} \\
\leq \; & 4[\lambda_{\min}\left(T^{-1}p_i'm_bp_i\right)]^{-1}T^{-1}n^2c_{nT}^2\overline{\gamma}_{nT}^2\sum_{i=1}^{n}\sigma_i^2\int\int w(x)\,\Delta_{in}(x)\,p^K(x)'p^K(\widetilde{x})\,\Delta_{in}(\widetilde{x})\,w(\widetilde{x})\,dxd\widetilde{x} \\
= \; & O_p(T^{-1}n^3c_{nT}^2\overline{\gamma}_{nT}^2K)=O_p\left(n^{-1/2}K\mathrm{tr}(\overline{P}_w)^{-1/2}\right)=o_p(1),
\end{aligned}
$$

it follows that $\overrightarrow{A}_{41a}=o_p(1)$. Similarly, we can show that $\overrightarrow{A}_{41b}=o_p(1)$ and thus $A_{41}=o_p\left(\mathrm{tr}(\overline{P}_w)^{1/2}\right)$. Analogously,

$$
\begin{aligned}
& A_{42}/\mathrm{tr}(\overline{P}_w)^{1/2} \\
= \; & 2n\overline{c}_{nT}\sum_{i=1}^{n}\int d_{g_i}(x)\,p^K(x)'\widetilde{\varepsilon}_iw(x)\,dx-2\overline{c}_{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\int d_{g_i}(x)\,p^K(x)'\widetilde{\varepsilon}_jw(x)\,dx \\
= \; & A_{42a}+A_{42b}+o_p(1)
\end{aligned}
$$

where $\overline{c}_{nT}\equiv c_{nT}/\mathrm{tr}(\overline{P}_w)^{1/2}$, $A_{42a}\equiv 2n\overline{c}_{nT}\sum_{i=1}^{n}\int d_{g_i}(x)\,p^K(x)'\overrightarrow{\varepsilon}_iw(x)\,dx$ and $A_{42b}\equiv$

$2\overline{c}_{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\int d_{g_i}(x)\ p^K(x)'\ \overrightarrow{\varepsilon}_j w(x)\,dx$. Noting that $E_{\mathcal{D}}(\overrightarrow{A}_{42a})=0$ and

$$
\begin{aligned}
&\mathrm{Var}_{\mathcal{D}}(\overrightarrow{A}_{42a})\\
= \;& 4n^2\overline{c}_{nT}^2\sum_{i=1}^{n}\sigma_i^2\int d_{g_i}(x)\,p^K(x)'\,w(x)\,dx\,\left(p_i'm_bp_i\right)^{-1}\int p^K(\widetilde{x})\,d_{g_i}(\widetilde{x})\,w(\widetilde{x})\,d\widetilde{x}\\
\leq \;& 4[\lambda_{\min}\left(T^{-1}p_i'm_bp_i\right)]^{-1}T^{-1}n^2\overline{c}_{nT}^2\sum_{i=1}^{n}\sigma_i^2\int\int w(x)\,d_{g_i}(x)\,p^K(x)'\,p^K(\widetilde{x})\,d_{g_i}(\widetilde{x})\,w(\widetilde{x})\,dxd\widetilde{x}\\
\leq \;& 4[\lambda_{\min}\left(T^{-1}p_i'm_bp_i\right)]^{-1}T^{-1}n^2\overline{c}_{nT}^2\sum_{i=1}^{n}\sigma_i^2\left\|d_{g_i}(\cdot)\right\|_{\infty,\overline{\omega}_i}^2\int\left(1+\|x\|^2\right)^{\overline{\omega}_i}w(x)^2\left\|p^K(x)\right\|^2dx\\
= \;& O_p(T^{-1}n^3\overline{c}_{nT}^2K^{-2\lambda/d}\zeta(K)^2)=O_p(TK^{-2\lambda/d}\zeta(K)^2\,\mathrm{tr}(\overline{P}_w)^{-1})=o_p(1)\,,
\end{aligned}
$$

it follows that $\overrightarrow{A}_{42a}=o_p(1)$. Similarly, we can show that $\overrightarrow{A}_{42b}=o_p(1)$ and thus $A_{42}=o_p\left(\mathrm{tr}(\overline{P}_w)^{1/2}\right)$. Lastly, $|A_5|=o_p\left(\mathrm{tr}(\overline{P}_w)^{1/2}\right)$ by the determination of the order of A_1 and A_4, and the Cauchy-Schwarz inequality. This completes the proof. ■

Proof of Theorem 3.4

The proof follows closely from that of Theorems 3.1-3.3. Now, by (A.2) and the proof of Theorems 3.1 and 3.3, we can show that

$$
\begin{aligned}
&V_{nT}^{1/2}n^{-1/2}T^{-1}D_{nT}\\
= \;& V_{nT}^{1/2}n^{-1/2}T^{-1}(c_{nT}\Gamma_{nT}-\widehat{B}_{nT})/\sqrt{\widehat{V}_{nT}}\\
= \;& n^{-1/2}T^{-1}\left((c_{nT}\Gamma_{nT}-B_{nT})\right)\{1+o_p(1)\}+n^{-1/2}T^{-1}\{\widehat{B}_{nT}-B_{nT}\}\{1+o_p(1)\}\\
= \;& n^{-1/2}T^{-1}c_{nT}\sum_{1\leq i<j\leq n}\left(\alpha_{g_i}-\alpha_{g_j}\right)'\overline{P}_w\left(\alpha_{g_i}-\alpha_{g_j}\right)\{1+o_p(1)\}+o_p(1)\,.
\end{aligned}
$$

Next, $n^{-1/2}T^{-1}c_{nT}\sum_{1\leq i<j\leq n}\left(\alpha_{g_i}-\alpha_{g_j}\right)'\overline{P}_w\left(\alpha_{g_i}-\alpha_{g_j}\right)=(n(n-1))^{-1}\sum_{1\leq i<j\leq n}\int\{g_i(x)-g_j(x)\}^2w(x)\,dx+o(1)\rightarrow\Delta_g$, where $\Delta_g=\lim_{n\rightarrow\infty}(n(n-1))^{-1}\sum_{1\leq i<j\leq n}\int\{g_i(x)-g_j(x)\}^2w(x)\,dx\neq 0$. The result follows because $V_{nT}=O_p\left(\mathrm{tr}(\overline{P}_w)\right)$ by (A.6).■

Proof of Theorem 3.5

Let E^* and Var^* denote expectation and variance conditional on the original sample. The proof follows closely from the same argument as used in proving $\frac{T}{\sqrt{nV_{nT}}}\sum_{i=1}^{n}\{\overrightarrow{\varepsilon}_i'\overline{P}_w\overrightarrow{\varepsilon}_i-E_{\mathcal{D}}(\overrightarrow{\varepsilon}_i'\overline{P}_w\overrightarrow{\varepsilon}_i)\}\overset{d}{\rightarrow}N(0,1)$ in the proof of Theorem 3.1. Note that $E^*(\widetilde{\varepsilon}_i^{*\prime}\overline{P}_w\widetilde{\varepsilon}_i)=\mathrm{tr}(A_{ih}\Sigma_i)$ and $\mathrm{Var}^*(\widetilde{\varepsilon}_i^{*\prime}\overline{P}_w\widetilde{\varepsilon}_i)=2\mathrm{tr}(A_{ih}\Sigma_iA_{ih}\Sigma_i)+o_p(1)$, where $o_p(1)$ becomes 0 if $\eta_{it}\backsim N(0,1)$. Since $E(\eta_{it}^8)<\infty$, one can check the Liapounov condition as in the proof of Theorem 3.1. ■

B Some technical lemmas

In this appendix we list some technical lemmas that are used in the proof of the main results in Section 3. Note that all lemmas hold without imposing the null restriction. For notational simplicity, let $c_{1\lambda} \equiv \min_{1\leq i\leq n}\{\lambda_{\min}(T^{-1}p_i'm_b p_i)\}$ and $c_{2\lambda} \equiv \min_{1\leq i\leq n}\{\lambda_{\min}(T^{-1}p_i'm_h p_i)\}$.

Lemma B.1 *Suppose Assumptions 1-2 and 3(iv) hold, then*

(i) $||T^{-1}p_i'm_h p_i - Q_i|| = O_p(K/\sqrt{T})$;

(ii) $\left\|T^{-1}(p_i'(m_h - m_b)p_i)\right\| = O_p(K/\sqrt{n})$;

(iii) $m_b f_2 = 0$ *w.p.a.1 as* $n \to \infty$.

Proof. (i) and (iii) follow from Lemmas A.1(iii) and A.5(iv) of Su and Jin (2010), respectively. The proof of (ii) is analogous to that of Lemma A.5(vi) of Su and Jin (2010) by using the decomposition for $m_h - m_b$ in (B.1) below. ■

Lemma B.2 *(i) Suppose Assumptions 1(i), (v), (vi), and (viii) hold, then* $nE\left[\overline{v}\overline{v}'\right] \leq CI_T$ *and* $nE\left[\overline{\varepsilon}\overline{\varepsilon}'\right] \leq CI_T$ *for some* $C < \infty$, *where* $\overline{v} = (\overline{v}_1, \overline{v}_2, \cdots, \overline{v}_T)'$ *and* $\overline{\varepsilon} = (\overline{\varepsilon}_1, \overline{\varepsilon}_2, \cdots, \overline{\varepsilon}_T)'$.

(ii) Suppose Assumptions 3(i)-(iii) hold, then $(nT)^{-1}\sum_{i=1}^n E\left\|\mathbf{g}_i - p_i a_{g_i}\right\|^2 = O(K^{-2\lambda/d})$.

Proof. See Lemmas A.3 and B.2 of Su and Jin (2010). ■

Lemma B.3 *Suppose Assumptions 1-2 and 3(iv) hold, then*

(i) $\sum_{i=1}^n ||T^{-1}p_i'(m_h - m_b)\varepsilon_i||^2 = O_p(K/T)$;

(ii) $\sum_{i=1}^n \left\|T^{-1}p_i'(m_h - m_b)p_i p_i' m_b \varepsilon_i\right\|^2 = O_p(K^2/T)$;

(iii) $n^{-1}\sum_{i=1}^n tr\left(\left[\left(T^{-1}p_i'm_b p_i\right)^{-}\overline{P}_w\right]^2 - \left[\left(T^{-1}p_i'm_h p_i\right)^{-}\overline{P}_w\right]^2\right) = O_p(K/\sqrt{n})$.

Proof. (i) Using the decomposition

$$
\begin{aligned}
m_h - m_b &= b\left(b'b\right)^{-1}b' - h\left(h'h\right)^{-1}h' \\
&= (b-h)\left(b'b\right)^{-1}b' + h[\left(b'b\right)^{-1} - \left(h'h\right)^{-1}]b' + h\left(h'h\right)^{-1}(b-h)', \quad \text{(B.1)}
\end{aligned}
$$

we have that by the C_r inequality,

$$
\begin{aligned}
\frac{1}{T^2}\sum_{i=1}^n \left\|p_i'(m_h - m_b)\varepsilon_i\right\|^2 &\leq \frac{1}{T^2}\sum_{i=1}^n \left\|p_i'\overline{v}^*\left(b'b\right)^{-1}b'\varepsilon_i\right\|^2 + \frac{1}{T^2}\sum_{i=1}^n \left\|p_i'h[\left(b'b\right)^{-1} - \left(h'h\right)^{-1}]b'\varepsilon_i\right\|^2 \\
&\quad + \frac{1}{T^2}\sum_{i=1}^n \left\|p_i'h\left(h'h\right)^{-1}\overline{v}^{*\prime}\varepsilon_i\right\|^2 \\
&\equiv D_1 + D_2 + D_3, \text{ say,}
\end{aligned}
$$

where $\overline{v}^* \equiv (\overline{v}_1^{*\prime}, ..., \overline{v}_T^{*\prime})' = h - b$. Noting that

$$
\begin{aligned}
T^{-4}E_{\mathcal{D}}\left(\sum_{i=1}^n \left\|p_i\right\|^2 \left\|b'\varepsilon_i\right\|^2\right) &= T^{-4}\sum_{i=1}^n \left\|p_i\right\|^2 \operatorname{tr}\left(b'E\left(\varepsilon_i\varepsilon_i'\right)b\right) \\
&\leq \max_{1\leq i\leq n}\left\{\lambda_{\max}\left(E\left(\varepsilon_i\varepsilon_i'\right)\right)\right\}\left\{T^{-1}\left\|b\right\|^2\right\}\left\{T^{-3}\sum_{i=1}^n \left\|p_i\right\|^2\right\} \\
&= O(1)\,O_p(1)\,O_p\left(nK/T^2\right) = O_p\left(nK/T^2\right), \quad \text{(B.2)}
\end{aligned}
$$

we have

$$\begin{aligned} D_1 &\le \left[\lambda_{\min}\left(T^{-1}b'b\right)\right]^{-2}\|\overline{v}^*\|^2\left\{T^{-4}\sum_{i=1}^n\|p_i\|^2\left\|b'\varepsilon_i\right\|^2\right\} \\ &= O_p(1)\,O_p(T/n)\,O_p\left(nK/T^2\right)=O_p(K/T) \end{aligned}$$

Similarly, $D_2 \le ||h[\left(T^{-1}b'b\right)^{-1}-\left(T^{-1}h'h\right)^{-1}]||^2\{T^{-4}\sum_{i=1}^n\|p_i\|^2\|b'\varepsilon_i\|^2\} = O_p(T/n)\,O_p\left(nK/T^2\right)$ $= O_p(K/T)$, and $D_3 \le ||\left(h'h\right)^{-1}h'||^2\{T^{-2}\sum_{i=1}^n\|\varepsilon_i'\overline{v}^*\|^2\|p_i\|^2\} = O_P\left(T^{-1}\right)O_p(K) = O_p(K/T)$. It follows that $\frac{1}{T^2}\sum_{i=1}^n\|p_i'(m_h-m_b)\varepsilon_i\|^2 = O_p(K/T)$.

For (ii), using the decomposition in (B.1), we have

$$\begin{aligned} &T^{-2}\sum_{i=1}^n\left\|p_i'(m_h-m_b)p_ip_i'm_b\varepsilon_i\right\|^2 \\ \le\;& T^{-2}\sum_{i=1}^n\left\|p_i'(b-h)\left(b'b\right)^{-1}b'p_ip_i'm_b\varepsilon_i\right\|^2+T^{-2}\sum_{i=1}^n\left\|p_i'h[\left(b'b\right)^{-1}-\left(h'h\right)^{-1}]b'p_ip_i'm_b\varepsilon_i\right\|^2 \\ &+T^{-2}\sum_{i=1}^n\left\|p_i'h\left(h'h\right)^{-1}(b-h)'p_ip_i'm_b\varepsilon_i\right\|^2 \\ \equiv\;& D_4+D_5+D_6,\ \text{say.} \end{aligned}$$

Noting that $E_{\mathcal{D}}(T^{-2}\sum_{i=1}^n\|p_i\|^2\|p_ip_i'm_b\varepsilon_i\|^2) = T^{-2}\sum_{i=1}^n\|p_i\|^2\mathrm{tr}(p_ip_i'm_bE\left(\varepsilon_i\varepsilon_i'\right)m_bp_ip_i') \le \max_{1\le i\le n}\lambda_{\max}\left(E\left(\varepsilon_i\varepsilon_i'\right)\right)T^{-2}\sum_{i=1}^n\|p_i\|^2\mathrm{tr}(p_i'm_bp_ip_i'p_i) \le \max_{1\le i\le n}\left[\lambda_{\max}\left(E\left(\varepsilon_i\varepsilon_i'\right)\right)\right]\max_{1\le i\le n}$ $\{\lambda_{\max}\left(T^{-1}p_i'm_bp_i\right)\}\;T^{-1}\sum_{i=1}^n\|p_i\|^4 = O_p\left(nK^2/T\right)$, we have

$$\begin{aligned} D_4 &\le \|\overline{v}^*\|^2\left\|\left(b'b\right)^{-1}b'\right\|^2\left\{T^{-2}\sum_{i=1}^n\|p_i\|^2\left\|p_ip_i'm_b\varepsilon_i\right\|^2\right\} \\ &= O_p(T/n)\,O_p(1/T)\,O_p\left(nK^2/T\right)=O_p\left(K^2/T\right). \end{aligned}$$

Similarly, we can show that $D_s = O_p\left(K^2/T\right)$ for $s=5,6$. Thus $T^{-2}\sum_{i=1}^n\|p_i'(m_h-m_b)p_ip_i'm_b\varepsilon_i\|^2$ $= O_p\left(K^2/T\right)$.

(iii) Let $M_{ih}\equiv\left(T^{-1}p_i'm_hp_i\right)^-\overline{P}_w$, $M_{ib}\equiv\left(T^{-1}p_i'm_bp_i\right)^-\overline{P}_w$, and $D_7\equiv\frac{1}{n}\sum_{i=1}^n\mathrm{tr}\{[(T^{-1}p_i'$ $m_bp_i)^-\overline{P}_w]^2-[\left(T^{-1}p_i'm_hp_i\right)^-\overline{P}_w]^2\}$. Then by the Hölder inequality,

$$\begin{aligned} D_7 &= \frac{1}{n}\sum_{i=1}^n\mathrm{tr}((M_{ib}+M_{ih})(M_{ib}-M_{ih})) \\ &\le \frac{1}{n}\sum_{i=1}^n\left\{\mathrm{tr}\left((M_{ib}+M_{ih})^2\right)\right\}^{1/2}\left\{\mathrm{tr}\left((M_{ib}-M_{ih})^2\right)\right\}^{1/2} \\ &\le 2\lambda_{\max}\left(\overline{P}_w\right)\left[(c_{1\lambda})^{-1}+(c_{2\lambda})^{-1}\right]\overline{D}_7, \end{aligned}$$

where recall $c_{1\lambda}\equiv\min_{1\le i\le n}[\lambda_{\min}(p_i'm_bp_i/T)]$ and $c_{2\lambda}\equiv\min_{1\le i\le n}[\lambda_{\min}(p_i'm_hp_i/T)]$, and $\overline{D}_7\equiv\frac{1}{n}\sum_{i=1}^n\left\{\mathrm{tr}\left((M_{ib}-M_{ih})^2\right)\right\}^{1/2}$. By the Cauchy-Schwarz inequality, $\overline{D}_7\le\overrightarrow{D}_7^{1/2}$, where

$\overrightarrow{D}_7 \equiv \frac{1}{n}\sum_{i=1}^n \mathrm{tr}\big((M_{ib}-M_{ih})^2\big)$. Writing $M_{ib}-M_{ih}=\{T^{-1}p_i'(m_h-m_b)p_i\left(T^{-1}p_i'm_bp_i\right)^-\}'$ $\left(T^{-1}p_i'm_hp_i\right)^-\overline{P}_w$ and using the fact that $\mathrm{tr}\{(A'B)^2\}\leq\mathrm{tr}\{(A'A)(B'B)\}$ (e.g., Magnus and Neudecker, 1999, p.201), we have

$$
\begin{aligned}
\overrightarrow{D}_7 &\leq \frac{1}{n}\sum_{i=1}^n \mathrm{tr}\left\{\left(T^{-1}p_i'(m_h-m_b)p_i\left(T^{-1}p_i'm_bp_i\right)^-\right)'T^{-1}p_i'(m_h-m_b)p_i\left(T^{-1}p_i'm_bp_i\right)^-\right.\\
&\quad \left.\overline{P}_w\left(T^{-1}p_i'm_hp_i\right)^-\left(T^{-1}p_i'm_hp_i\right)^-\overline{P}_w\right\}\\
&\leq \left[\lambda_{\max}\left(\overline{P}_w\right)\right]^2(c_{1\lambda}c_{2\lambda})^{-1}\frac{1}{n}\sum_{i=1}^n\left\|T^{-1}p_i'(m_h-m_b)p_i\left(T^{-1}p_i'm_bp_i\right)^-\right\|^2\\
&\leq \left[\lambda_{\max}\left(\overline{P}_w\right)\right]^2c_{1\lambda}^{-3}c_{2\lambda}^{-1}\frac{1}{n}\sum_{i=1}^n\left\|T^{-1}p_i'(m_h-m_b)p_i\right\|^2.
\end{aligned}
$$

Again, using the decomposition in (B.1), it is straightforward to show that $\frac{1}{n}\sum_{i=1}^n||T^{-1}p_i'(m_h-m_b)p_i||^2=O_p\left(K^2/n\right)$. It follows that $D_7=O_p\left(K/\sqrt{n}\right)$. ■

Lemma B.4 *Recall* $\widetilde{\varepsilon}_i\equiv(p_i'm_hp_i)^-p_i'm_h\varepsilon_i$, $\overrightarrow{\varepsilon}_i\equiv(p_i'm_bp_i)^-p_i'm_b\varepsilon_i$, $\widetilde{r}_i\equiv(p_i'm_hp_i)^-p_i'm_hr_i$, $\overrightarrow{r}_i\equiv(p_i'm_bp_i)^-p_i'm_br_i$, $r_i\equiv f_2\gamma_{2i}+(\mathbf{g}_i-p_i\alpha_{g_i})$, *and* $\lambda\equiv\min_{1\leq i\leq n}\lambda_i$. *Suppose Assumptions 1-2 and 3(iv) hold, then*

(i) $nc_{nT}\sum_{i=1}^n(\widetilde{\varepsilon}_i-\overrightarrow{\varepsilon}_i)'\overline{P}_w(\widetilde{\varepsilon}_i-\overrightarrow{\varepsilon}_i)=O_p\left(K^2/n^{1/2}\right)$;

(ii) $c_{nT}\sum_{1\leq i\neq j\leq n}(\widetilde{\varepsilon}_i-\overrightarrow{\varepsilon}_i)'\overline{P}_w(\overrightarrow{\varepsilon}_i-\overrightarrow{\varepsilon}_j)=O_p\left(K^{1/2}/T^{1/2}\right)$;

(iii) $nc_{nT}\sum_{i=1}^n(\widetilde{r}_i-\overrightarrow{r}_i)'\overline{P}_w(\widetilde{r}_i-\overrightarrow{r}_i)=O_p\left(Tn^{-1/2}K^{1-2\lambda/d}+KT/n^{1/2}\right)$;

(iv) $c_{nT}\sum_{1\leq i\neq j\leq n}(\overrightarrow{r}_i-\overrightarrow{r}_j)'\overline{P}_w(\overrightarrow{r}_i-\overrightarrow{r}_j)=O_p\left(Tn^{1/2}K^{1-2\lambda/d}\right)$;

(v) $c_{nT}\sum_{1\leq i<j\leq n}(\overrightarrow{\varepsilon}_i-\overrightarrow{\varepsilon}_j)'\overline{P}_w(\widetilde{r}_i-\overrightarrow{r}_i)=O_p\left(K/T^{1/2}+T^{1/2}K^{1-\lambda/d}\right)$;

(vi) $c_{nT}\sum_{1\leq i<j\leq n}(\overrightarrow{\varepsilon}_i-\overrightarrow{\varepsilon}_j)'\overline{P}_w(\overrightarrow{r}_i-\overrightarrow{r}_j)\}=O_p\left(T^{1/2}K^{1/2-\lambda/d}\right)$.

Proof. (i) Noting that $\widetilde{\varepsilon}_i-\overrightarrow{\varepsilon}_i=[(p_i'm_hp_i)^--(p_i'm_bp_i)^-]p_i'm_h\varepsilon_i+(p_i'm_bp_i)^-p_i'(m_h-m_b)\varepsilon_i$ and $(p_i'm_hp_i)^--(p_i'm_bp_i)^-=(p_i'm_hp_i)^-p_i'(m_b-m_h)p_i(p_i'm_bp_i)^-$, by the C_r inequality and Lemmas B.3(i)-(ii) we have

$$
\begin{aligned}
&nc_{nT}\sum_{i=1}^n(\widetilde{\varepsilon}_i-\overrightarrow{\varepsilon}_i)'\overline{P}_w(\widetilde{\varepsilon}_i-\overrightarrow{\varepsilon}_i)\\
\leq\ & nc_{nT}\lambda_{\max}\left(\overline{P}_w\right)\sum_{i=1}^n\|\widetilde{\varepsilon}_i-\overrightarrow{\varepsilon}_i\|^2\\
\leq\ & 2nc_{nT}\lambda_{\max}\left(\overline{P}_w\right)\sum_{i=1}^n\left\|[(p_i'm_hp_i)^--(p_i'm_bp_i)^-]p_i'm_h\varepsilon_i\right\|^2\\
&+2nc_{nT}\lambda_{\max}\left(\overline{P}_w\right)\sum_{i=1}^n\left\|\left(T^{-1}p_i'm_hp_i\right)^-T^{-1}p_i'(m_h-m_b)\varepsilon_i\right\|^2\\
\leq\ & 2nc_{nT}\lambda_{\max}\left(\overline{P}_w\right)(c_{1\lambda}c_{2\lambda})^{-2}\sum_{i=1}^n\left\|T^{-1}p_i'(m_h-m_b)p_iT^{-1}p_i'm_b\varepsilon_i\right\|^2
\end{aligned}
$$

$$+2nc_{nT}\lambda_{\max}\left(\overline{P}_w\right)(c_{2\lambda})^{-2}\sum_{i=1}^{n}\left\|T^{-1}p_i'(m_h-m_b)\varepsilon_i\right\|^2$$
$$= nc_{nT}\left(O_p\left(K^2/T\right)+O_p\left(K/T\right)\right)=O_p\left(K^2/n^{1/2}\right).$$

(ii) Noting that $c_{nT}\sum_{1\leq i\neq j\leq n}(\widetilde{\varepsilon}_i-\overrightarrow{\varepsilon}_i)'\overline{P}_w(\overrightarrow{\varepsilon}_i-\overrightarrow{\varepsilon}_j)=Q_{1nT}-Q_{2nT}$, where $Q_{1nT}\equiv nc_{nT}\sum_{i=1}^{n}(\widetilde{\varepsilon}_i-\overrightarrow{\varepsilon}_i)'\overline{P}_w\overrightarrow{\varepsilon}_i$, and $Q_{2nT}\equiv c_{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}(\widetilde{\varepsilon}_i-\overrightarrow{\varepsilon}_i)'\overline{P}_w\overrightarrow{\varepsilon}_j$. It suffices to prove (ii) by showing that $Q_{snT}=o_p\left(K^{1/2}\right)$ for $s=1,2$. We decompose Q_{1nT} as follows

$$\begin{aligned}
Q_{1nT} &= nc_{nT}\sum_{i=1}^{n}\left\{\left(p_i'm_bp_i\right)^{-}p_i'(m_h-m_b)\varepsilon_i\right\}'\overline{P}_w\left(p_i'm_bp_i\right)^{-}p_i'm_b\varepsilon_i\\
&\quad+nc_{nT}\sum_{i=1}^{n}\left\{[\left(p_i'm_hp_i\right)^{-}-\left(p_i'm_bp_i\right)^{-}]p_i'm_h\varepsilon_i\right\}'\overline{P}_w\left(p_i'm_bp_i\right)^{-}p_i'm_b\varepsilon_i\\
&\equiv Q_{1nT,1}+Q_{1nT,2}.
\end{aligned}$$

By (B.1) we can further decompose $Q_{1nT,1}$ as follows

$$\begin{aligned}
Q_{1nT,1} &= -nc_{nT}\sum_{i=1}^{n}\varepsilon_i'b\left(b'b\right)^{-1}\overline{v}^{*\prime}p_i\left(p_i'm_bp_i\right)^{-}\overline{P}_w\left(p_i'm_bp_i\right)^{-}p_i'm_b\varepsilon_i\\
&\quad+nc_{nT}\sum_{i=1}^{n}\varepsilon_i'b[\left(b'b\right)^{-1}-\left(h'h\right)^{-1}]h'p_i\left(p_i'm_bp_i\right)^{-}\overline{P}_w\left(p_i'm_bp_i\right)^{-}p_i'm_b\varepsilon_i\\
&\quad-nc_{nT}\sum_{i=1}^{n}\varepsilon_i'\overline{v}^{*}\left(h'h\right)^{-1}h'p_i\left(p_i'm_bp_i\right)^{-}\overline{P}_w\left(p_i'm_bp_i\right)^{-}p_i'm_b\varepsilon_i\\
&\equiv -Q_{1nT,11}+Q_{1nT,12}-Q_{1nT,13}.
\end{aligned}$$

To analyze $Q_{1nT,11}$, note that $\overline{v}_t^*\equiv(0',\overline{v}_t',\overline{g}_t+\overline{\varepsilon}_t)'$. So we can decompose $\overline{v}^*\equiv(\overline{v}_1^*,\cdots,\overline{v}_T^*)'=\overline{v}^++\overline{g}^*$, where the tth rows of $\overline{v}^+$ and $\overline{g}^*$ are given by $\overline{v}_t^+\equiv(0',\overline{v}_t',\overline{\varepsilon}_t)'$ and $\overline{g}_t^*\equiv(0',0',\overline{g}_t)'$, respectively. Then $Q_{1nT,11}=Q_{1nT,111}+Q_{1nT,112}$, where $Q_{1nT,111}$ and $Q_{1nT,112}$ are analogously defined as $Q_{1nT,11}$ but with $\overline{v}^*$ being replaced by $\overline{v}^+$ and $\overline{g}^*$, respectively. Let $\xi_i\equiv p_i\left(p_i'm_bp_i\right)^{-}\overline{P}_w\left(p_i'm_bp_i\right)^{-}p_i'm_b\varepsilon_i\varepsilon_i'$ and $\xi_i^c\equiv\xi_i-E_{\mathcal{D}}(\xi_i)$. Then we can decompose $Q_{1nT,111}$ as follows

$$\begin{aligned}
Q_{1nT,111} &= nc_{nT}\sum_{i=1}^{n}\mathrm{tr}\left(b\left(b'b\right)^{-1}\overline{v}^{+\prime}E_{\mathcal{D}}(\xi_i)\right)+nc_{nT}\sum_{i=1}^{n}\mathrm{tr}\left(b\left(b'b\right)^{-1}\overline{v}^{+\prime}\xi_i^c\right)\\
&\equiv Q_{1nT,111a}+Q_{1nT,111b}.
\end{aligned}$$

Let $\varsigma_i\equiv(\overline{v}^{+\prime}p_i\left(p_i'm_bp_i\right)^{-}\overline{P}_w\left(p_i'm_bp_i\right)^{-}p_i'm_b)'$. Then by Lemma B.2(i)

$$\begin{aligned}
E_{\mathcal{D}}\left[\mathrm{tr}\left(\varsigma_i'\varsigma_i\right)\right] &= E_{\mathcal{D}}\left[\mathrm{tr}\left\{\overline{v}^{+\prime}p_i\left(p_i'm_bp_i\right)^{-}\overline{P}_w\left(p_i'm_bp_i\right)^{-}\overline{P}_w\left(p_i'm_bp_i\right)^{-}p_i'\overline{v}^{+}\right\}\right]\\
&= \mathrm{tr}\left\{p_i\left(p_i'm_bp_i\right)^{-}\overline{P}_w\left(p_i'm_bp_i\right)^{-}\overline{P}_w\left(p_i'm_bp_i\right)^{-}p_i'E\left(\overline{v}^{+}\overline{v}^{+\prime}\right)\right\}\\
&\leq Cn^{-1}\mathrm{tr}\left\{p_i\left(p_i'm_bp_i\right)^{-}\overline{P}_w\left(p_i'm_bp_i\right)^{-}\overline{P}_w\left(p_i'm_bp_i\right)^{-}p_i'\right\}\\
&\leq Cn^{-1}\left[\lambda_{\max}\left(\overline{P}_w\right)\right]^2(c_{1\lambda})^{-3}T^{-3}\left\|p_i\right\|^2.
\end{aligned}$$

Using this and the fact that $\operatorname{tr}(A'B) \leq \{\operatorname{tr}(A'A)\operatorname{tr}(B'B)\}\}^{1/2}$ we have

$$\begin{aligned} Q_{1nT,111a} &= nc_{nT} \sum_{i=1}^{n} \operatorname{tr}\left(\varsigma_i' E\left(\varepsilon_i \varepsilon_i'\right) b\left(b'b\right)^{-1}\right) \\ &\leq nc_{nT} \sum_{i=1}^{n} \left\{\operatorname{tr}\left(\varsigma_i' \varsigma_i\right)\right\}^{1/2} \left\{\operatorname{tr}\left(\left(b'b\right)^{-1} b' E\left(\varepsilon_i \varepsilon_i'\right) E\left(\varepsilon_i \varepsilon_i'\right) b\left(b'b\right)^{-1}\right)\right\}^{1/2} \\ &\leq Cn^{-1/2} \lambda_{\max}\left(\overline{P}_w\right) \left(c_{1\lambda}\right)^{-3/2} \max_{1\leq i\leq n} \left[\lambda_{\max}\left(E\left(\varepsilon_i \varepsilon_i'\right)\right)\right] \lambda_{\min} \left\|\left(T^{-1} b'b\right)^{-1}\right\|^{1/2} \\ &\quad \times nc_{nT} \left\{T^{-2} \sum_{i=1}^{n} \left\|p_i\right\|\right\} \\ &= n^{1/2} c_{nT} O_p\left(n\sqrt{TK}/T^2\right) = O_p\left(\sqrt{K/T}\right). \end{aligned}$$

Similarly, for $Q_{1nT,111b}$ we have

$$Q_{1nT,111b} = nc_{nT} \operatorname{tr}\left(b\left(b'b\right)^{-1} \overline{v}^{+\prime} \sum_{i=1}^{n} \xi_i^c\right) \leq nc_{nT} \left\{Q_1\right\}^{1/2} \left\{Q_2\right\}^{1/2},$$

where $Q_1 \equiv \operatorname{tr}(b\left(b'b\right)^{-1} \overline{v}^{+\prime} \overline{v}^{+} \left(b'b\right)^{-1} b')$, and $Q_2 \equiv \operatorname{tr}\left(\sum_{i=1}^{n} \sum_{j=1}^{n} \xi_i^{c\prime} \xi_j^c\right)$. Noting that

$$Q_1 = \operatorname{tr}(\left(b'b\right)^{-1} \overline{v}^{+\prime} \overline{v}^{+}) = \left[\lambda_{\min}\left(T^{-1} b'b\right)\right]^{-1} T^{-1} \left\|\overline{v}^{+}\right\|^2 = O_p\left(1/n\right),$$

and

$$\begin{aligned} E_{\mathcal{D}}\left[Q_2\right] &= \sum_{i=1}^{n} \operatorname{tr}\left(E_{\mathcal{D}}\left(\xi_i^c \xi_i^{c\prime}\right)\right) \\ &\leq C \sum_{i=1}^{n} p_i \left(p_i' m_b p_i\right)^{-} \overline{P}_w \left(p_i' m_b p_i\right)^{-} \overline{P}_w \left(p_i' m_b p_i\right)^{-} p_i' \\ &\leq C \left[\lambda_{\max}\left(\overline{P}_w\right)\right]^2 \left(c_{1\lambda}\right)^{-3} T^{-3} \sum_{i=1}^{n} \left\|p_i\right\|^2 = O_p\left(nK/T^2\right), \end{aligned}$$

we have $Q_{1nT,111b} = nc_{nT} O_p\left(1/\sqrt{n}\right) O_p\left(\sqrt{nK}/T\right) = O_p(\sqrt{K/n})$. It follows that $Q_{1nT,111} = O_p(\sqrt{K/T} + \sqrt{K/n}) = O_p(\sqrt{K/T})$.

For $Q_{1nT,112}$, noting that $Q_{1nT,112} = nc_{nT} \sum_{i=1}^{n} \zeta_i$ with $\zeta_i \equiv \varepsilon_i' b\left(b'b\right)^{-1} \overline{g}^{*\prime} p_i \left(p_i' m_b p_i\right)^{-} \overline{P}_w \left(p_i' m_b p_i\right)^{-} p_i' m_b \varepsilon_i$, we have $\left(Q_{1nT,112}\right)^2 = Q_{1nT,112a} + Q_{1nT,112b}$, where $Q_{1nT,112a} \equiv \left(nc_{nT}\right)^2 \sum_{i=1}^{n} \zeta_i^2$, and $Q_{1nT,112b} \equiv \left(nc_{nT}\right)^2 \sum_{1\leq i\neq j\leq n} \zeta_i \zeta_j$. It is easy to show that $E_{\mathcal{D}}\left(Q_{1nT,112a}\right) = O_p\left(K/\left(nT\right)\right)$, implying that $Q_{1nT,112a} = O_p\left(K/\left(nT\right)\right)$ by the Markov inequality. For $Q_{1nT,112b}$, we have

$$\begin{aligned} E_{\mathcal{D}}\left[Q_{1nT,112b}\right] &= \left(nc_{nT}\right)^2 \sum_{1\leq i\neq j\leq n} \operatorname{tr}\left\{b\left(b'b\right)^{-1} \overline{g}^{*\prime} p_i \left(p_i' m_b p_i\right)^{-} \overline{P}_w \left(p_i' m_b p_i\right)^{-} p_i' m_b E\left(\varepsilon_i \varepsilon_i'\right)\right\} \\ &\quad \times \operatorname{tr}\left\{b\left(b'b\right)^{-1} \overline{g}^{*\prime} p_j \left(p_j' m_b p_j\right)^{-} \overline{P}_w \left(p_j' m_b p_j\right)^{-} p_j' m_b E\left(\varepsilon_j \varepsilon_j'\right)\right\} \leq D^2, \end{aligned}$$

where $D \equiv nc_{nT} \sum_{i=1}^{n} \{\mathrm{tr}(b\,(b'b)^{-1}\,\overline{g}^{*\prime} p_i\,(p_i' m_b p_i)^{-}\,\overline{P}_w\,(p_i' m_b p_i)^{-}\,p_i' \overline{g}^{*}\,(b'b)^{-1}\,b')\}^{1/2}\,\{\mathrm{tr}(E\,(\varepsilon_i \varepsilon_i')$ $m_b p_i\,(p_i' m_b p_i)^{-}\,\overline{P}_w\,(p_i' m_b p_i)^{-}\,p_i' m_b E\,(\varepsilon_i \varepsilon_i'))\}^{1/2}$. Note that

$$
\begin{aligned}
D &\leq C\lambda_{\max}\left(\overline{P}_w\right) c_{1\lambda}^{-2} \left\| \left(T^{-1} b'b\right)^{-1} \right\|^{1/2} nc_{nT} \left\{ T^{-5/2} \sum_{i=1}^{n} \left\| p_i' \overline{g}^{*} \right\| \left\| p_i' m_b \right\| \right\} \\
&= nc_{nT} O_p\left(n^{1/2} K / T^{3/2}\right) = O_p\left(K / T^{1/2}\right).
\end{aligned}
$$

So we have $Q_{1nT,112b} = O_p\left(K/T^{1/2}\right)$ and $Q_{1nT,112} = O_p\left(K/T^{1/2}\right)$. Consequently $Q_{1nT,11} = O_p\left(K^{1/2} + K/T^{1/2}\right) = O_p\left(K^{1/2}\right)$. Analogously, we can show that $Q_{1nT,1s} = O_p\left(K^{1/2}\right)$ for $s = 2, 3$. It follows that $Q_{1nT,1} = O_p\left(K^{1/2}\right)$.

For $Q_{1nT,2}$, it is easy to show that $Q_{1nT,2} = nc_{nT} \sum_{i=1}^{n} \varepsilon_i' m_h p_i\,(p_i' m_b p_i)^{-}\,p_i'(m_b - m_h) p_i\,(p_i' m_h p_i)^{-}$ $\overline{P}_w\,(p_i' m_b p_i)^{-}\,p_i' m_b \varepsilon_i = \overline{Q}_{1nT,2} + o_p(1)$ where $\overline{Q}_{1nT,2} \equiv nc_{nT} \sum_{i=1}^{n} \varepsilon_i' m_b p_i\,(p_i' m_b p_i)^{-}\,p_i'(m_b - m_h) p_i\,(p_i' m_b p_i)^{-}\,\overline{P}_w\,(p_i' m_b p_i)^{-}\,p_i' m_b \varepsilon_i$. Then using the decomposition in (B.1) and the proof strategy for $Q_{1nT,1}$, we can show that $\overline{Q}_{1nT,2} = O_p\left(K^{1/2}\right)$. Consequently, $Q_{1nT} = O_p\left(K^{1/2}\right)$. Analogously we can prove $Q_{2nT} \equiv c_{nT} \sum_{i=1}^{n} \sum_{j=1}^{n} (\widetilde{\varepsilon}_i - \overrightarrow{\varepsilon}_i)'\,\overline{P}_w\,\overrightarrow{\varepsilon}_j = O_p\left(K^{1/2}\right)$.

(iii) By the definition of r_i and the fact $m_b f_2 = 0$ w.p.a.1 as $n \to \infty$, we have

$$
\begin{aligned}
\widetilde{r}_i - \overrightarrow{r}_i &= \left(p_i' m_h p_i\right)^{-} p_i' \left(m_h - m_b\right) r_i + \left[\left(p_i' m_h p_i\right)^{-} - \left(p_i' m_b p_i\right)^{-}\right] p_i' m_b r_i \\
&= \left(p_i' m_h p_i\right)^{-} p_i' \left(m_h - m_b\right) f_2 \gamma_{2i} + \left(p_i' m_h p_i\right)^{-} p_i' \left(m_h - m_b\right) \mathbf{d}_i \\
&\quad + \left[\left(p_i' m_h p_i\right)^{-} - \left(p_i' m_b p_i\right)^{-}\right] p_i' m_b \mathbf{d}_i,
\end{aligned}
\tag{B.3}
$$

where $\mathbf{d}_i \equiv \mathbf{g}_i - p_i \alpha_{g_i}$. It follows that

$$
\begin{aligned}
& nc_{nT} \sum_{i=1}^{n} (\widetilde{r}_i - \overrightarrow{r}_i)'\,\overline{P}_w\,(\widetilde{r}_i - \overrightarrow{r}_i) \\
\leq\;& 3nc_{nT} \sum_{i=1}^{n} \left(\left(p_i' m_h p_i\right)^{-} p_i' \left(m_h - m_b\right) \mathbf{d}_i\right)' \overline{P}_w \left(\left(p_i' m_h p_i\right)^{-} p_i' \left(m_h - m_b\right) \mathbf{d}_i\right) \\
&+ 3nc_{nT} \sum_{i=1}^{n} \left(\left(p_i' m_h p_i\right)^{-} p_i' \left(m_h - m_b\right) f_2 \gamma_{2i}\right)' \overline{P}_w \left(p_i' m_h p_i\right)^{-} p_i' \left(m_h - m_b\right) f_2 \gamma_{2i} \\
&+ 3nc_{nT} \sum_{i=1}^{n} \left\{\left[\left(p_i' m_h p_i\right)^{-} - \left(p_i' m_b p_i\right)^{-}\right] p_i' m_b \mathbf{d}_i\right\}' \overline{P}_w \{[\left(p_i' m_h p_i\right)^{-} - \left(p_i' m_b p_i\right)^{-}] p_i' m_b \mathbf{d}_i\} \\
\equiv\;& 3U_{1nT} + 3U_{2nT} + 3U_{3nT}.
\end{aligned}
$$

For U_{1nT}, we have

$$
U_{1nT} \leq \lambda_{\max}\left(\overline{P}_w\right) nc_{nT} \sum_{i=1}^{n} \left\| \left(p_i' m_h p_i\right)^{-} p_i' \left(m_h - m_b\right) \mathbf{d}_i \right\|^2 \leq \lambda_{\max}\left(\overline{P}_w\right) c_{2\lambda}^{-2} \overline{U}_{1nT}
$$

where $\overline{U}_{1nT} \equiv nc_{nT} \sum_{i=1}^{n} \left\| T^{-1} p_i' \left(m_h - m_b\right) \mathbf{d}_i \right\|^2$. Using (B.1), we have

$$\begin{aligned}
\overline{U}_{1nT} &\leq 3nc_{nT} \sum_{i=1}^{n} \left\| T^{-1} p_i' \left(b-h\right) \left(b'b\right)^{-1} b' \mathbf{d}_i \right\|^2 + 3nc_{nT} \sum_{i=1}^{n} \left\| T^{-1} p_i' h [\left(b'b\right)^{-1} - \left(h'h\right)^{-1}] b' \mathbf{d}_i \right\|^2 \\
&\quad + 3nc_{nT} \sum_{i=1}^{n} \left\| T^{-1} p_i' h \left(h'h\right)^{-1} \left(b-h\right)' \mathbf{d}_i \right\|^2 \\
&\equiv 3\overline{U}_{1nT,1} + 3\overline{U}_{1nT,2} + 3\overline{U}_{1nT,3}, \text{ say.}
\end{aligned}$$

To proceed, we notice that by Lemma B.2

$$\begin{aligned}
T^{-2} \sum_{i=1}^{n} \left\| p_i \right\|^2 \left\| \mathbf{d}_i \right\|^2 &\leq nK \max_{1 \leq i \leq n} \left[\lambda_{\max} \left(T^{-1} p_i' p_i \right) \right] \left\{ \left(nT\right)^{-1} \sum_{i=1}^{n} \left\| \mathbf{d}_i \right\|^2 \right\} \\
&= nK O_p\left(K^{-2\lambda/d}\right) = O_p\left(nK^{1-2\lambda/d}\right). \qquad \text{(B.4)}
\end{aligned}$$

For $\overline{U}_{1nT,1}$, we have

$$\overline{U}_{1nT,1} \leq nc_{nT} T^{-2} \left\| \left(b'b\right)^{-1} b' \right\|^2 \sum_{i=1}^{n} \left\| p_i'(b-h) \right\|^2 \left\| \mathbf{d}_i \right\|^2 = \operatorname{tr}\left(\left(T^{-1} b'b\right)^{-1} \right) \overrightarrow{U}_{1nT,1}$$

where $\overrightarrow{U}_{1nT,1} \equiv nc_{nT} T^{-3} \sum_{i=1}^{n} \left\| p_i' \left(b-h\right) \right\|^2 \left\| \mathbf{d}_i \right\|^2$. Recall that $\overline{v}_t^*$ denotes the tth row of $h - b$: $\overline{v}_t^* = \left(0', \overline{v}_t', \overline{\varepsilon}_t + \overline{g}_t\right)$. Let $\overline{v} \equiv \left(\overline{v}_1', ..., \overline{v}_T'\right)'$. Similarly define $\overline{\varepsilon}$ and $\overline{g}$. Noting that $nE\left(\overline{v}\overline{v}'\right) \leq CI_T$ and $nE\left(\overline{\varepsilon}\overline{\varepsilon}'\right) \leq CI_T$ for some $C > 0$ by Lemma A.3 of Su and Jin (2010), we have by (B.4)

$$\begin{aligned}
E_{\mathcal{D}}(\overrightarrow{U}_{1nT,1}) &= nc_{nT} T^{-3} \sum_{i=1}^{n} \operatorname{tr}\left(p_i' E_{\mathcal{D}} \left(b-h\right) \left(b-h\right)' p_i \right) \left\| \mathbf{d}_i \right\|^2 \\
&= Cc_{nT} T^{-3} \sum_{i=1}^{n} \operatorname{tr}\left(p_i' p_i \right) \left\| \mathbf{d}_i \right\|^2 + Cnc_{nT} T^{-3} \sum_{i=1}^{n} \operatorname{tr}\left(p_i' \overline{g}\overline{g}' p_i \right) \left\| \mathbf{d}_i \right\|^2 \\
&\leq Cc_{nT} T^{-1} \left(1 + n \left\| \overline{g} \right\|^2\right) \left\{ T^{-2} \sum_{i=1}^{n} \left\| p_i \right\|^2 \left\| \mathbf{d}_i \right\|^2 \right\} \\
&= c_{nT} T^{-1} O_p\left(T\right) O_p\left(nK^{1-2\lambda/d}\right) = O_p\left(Tn^{-1/2} K^{1-2\lambda/d}\right).
\end{aligned}$$

Consequently, $\overrightarrow{U}_{1nT,1} = O_p\left(Tn^{-1/2} K^{1-2\lambda/d}\right)$ and $\overline{U}_{1nT,1} = O_p\left(Tn^{-1/2} K^{1-2\lambda/d}\right)$. Analogously, we can show that $\overline{U}_{1nT,3} = O_p\left(Tn^{-1/2} K^{1-2\lambda/d}\right)$. For $\overrightarrow{U}_{1nT,2}$, we have

$$\begin{aligned}
\overline{U}_{1nT,2} &\leq \left\| h[\left(b'b\right)^{-1} - \left(h'h\right)^{-1}] b' \right\|^2 nc_{nT} T^{-2} \sum_{i=1}^{n} \left\| p_i \right\|^2 \left\| \mathbf{d}_i \right\|^2 \\
&= \operatorname{tr}\left(h'h[\left(b'b\right)^{-1} - \left(h'h\right)^{-1}] b'b [\left(b'b\right)^{-1} - \left(h'h\right)^{-1}] \right) nc_{nT} T^{-2} \sum_{i=1}^{n} \left\| p_i \right\|^2 \left\| \mathbf{d}_i \right\|^2 \\
&\leq \lambda_{\max}\left(T^{-1} h'h\right) \lambda_{\max}\left(T^{-1} b'b\right) \left\| \left(T^{-1} b'b\right)^{-1} - \left(T^{-1} h'h\right)^{-1} \right\|^2 nc_{nT} \left\{ T^{-2} \sum_{i=1}^{n} \left\| p_i \right\|^2 \left\| \mathbf{d}_i \right\|^2 \right\} \\
&= O_p\left(n^{-1}\right) nc_{nT} O_p\left(nK^{1-2\lambda/d}\right) = O_p\left(Tn^{-1/2} K^{1-2\lambda/d}\right),
\end{aligned}$$

where we have repeatedly used the fact that $\mathrm{tr}(AB) \leq \lambda_{\max}(A)\mathrm{tr}(B)$ for symmetric A and p.s.d. B. It follows $U_{1nT} = O_p\left(Tn^{-1/2}K^{1-2\lambda/d}\right)$.

For U_{2nT}, notice that

$$U_{2nT} \leq \lambda_{\max}\left(\overline{P}_w\right) nc_{nT} \sum_{i=1}^{n} \left\| \left(p_i' m_h p_i\right)^{-} p_i' \left(m_h - m_b\right) f_2 \gamma_{2i} \right\|^2 \leq \lambda_{\max}\left(\overline{P}_w\right) \left(c_{2\lambda}\right)^{-2} \overline{U}_{2nT}$$

where $\overline{U}_{2nT} \equiv nc_{nT} \sum_{i=1}^{n} \left\| T^{-1} p_i' \left(m_h - m_b\right) f_2 \gamma_{2i} \right\|^2$. By (B.1) we have

$$\begin{aligned}
\overline{U}_{2nT} &\leq 3nc_{nT} \sum_{i=1}^{n} \left\| T^{-1} p_i' \overline{v}^* \left(b'b\right)^{-1} b' f_2 \gamma_{2i} \right\|^2 + 3nc_{nT} \sum_{i=1}^{n} \left\| T^{-1} p_i' h [\left(b'b\right)^{-1} - \left(h'h\right)^{-1}] b' f_2 \gamma_{2i} \right\|^2 \\
&\quad + 3nc_{nT} \sum_{i=1}^{n} \left\| T^{-1} p_i' h \left(h'h\right)^{-1} \overline{v}^* f_2 \gamma_{2i} \right\|^2 \\
&\equiv 3\overline{U}_{2nT,1} + 3\overline{U}_{2nT,2} + 3\overline{U}_{2nT,3}.
\end{aligned}$$

Noting that

$$\begin{aligned}
E_{\mathcal{D}}\left(\overline{U}_{2nT,1}\right) &= nc_{nT} \sum_{i=1}^{n} E_{\mathcal{D}} \left\| T^{-1} p_i' \overline{v}^* \left(b'b\right)^{-1} b' f_2 \gamma_{2i} \right\|^2 \\
&\leq Cnc_{nT} \left(n^{-1} + \|\overline{g}\|^2\right) \left[\lambda_{\min}\left(T^{-1} b'b\right)\right]^{-2} \left\{ T^{-1} \sum_{i=1}^{n} \mathrm{tr}\left(T^{-1} p_i' p_i\right) E_{\mathcal{D}} \left\| T^{-1} b' f_2 \gamma_{2i} \right\|^2 \right\} \\
&= nc_{nT} O_p\left(T/n\right) O_p\left(Kn/T\right) = O_p\left(KT/n^{1/2}\right),
\end{aligned}$$

we have $\overline{U}_{2nT,1} = O_p\left(KT/n^{1/2}\right)$. Similarly, $\overline{U}_{2nT,2} = nc_{nT} O_p\left(1/n\right) O_p\left(Kn/T\right) = O_p\left(K/n^{1/2}\right)$ as $\left(T^{-1}h'h\right)^{-1} - \left(T^{-1}b'b\right)^{-1} = O_p\left(n^{-1/2}\right)$, and $\overline{U}_{2nT,3} = O_p\left(KT/n^{1/2}\right)$. It follows that $U_{2nT} = O_p\left(KT/n^{1/2}\right)$.

For U_{3nT}, using the expression $\left(p_i' m_h p_i\right)^{-} - \left(p_i' m_b p_i\right)^{-} = \left(p_i' m_b p_i\right)^{-} p_i' (m_b - m_h) p_i \left(p_i' m_h p_i\right)^{-}$, the decomposition in (B.1), and analogous arguments as used in the determination of the probability order of U_{1nT}, we can show that $U_{3nT} = O_p\left(Tn^{-1/2}K^{1-2\lambda/d}\right)$. Consequently, $nc_{nT} \sum_{i=1}^{n} \left(\widetilde{r}_i - \overrightarrow{r}_i\right)' \overline{P}_w \left(\widetilde{r}_i - \overrightarrow{r}_i\right) = O_p\left(Tn^{-1/2}K^{1-2\lambda/d} + KT/n^{1/2}\right)$.

(iv) Noting that by (B.4)

$$\begin{aligned}
nc_{nT} \sum_{i=1}^{n} \overrightarrow{r}_i' \overline{P}_w \overrightarrow{r}_i &= nc_{nT} \lambda_{\max}\left(\overline{P}_w\right) \sum_{i=1}^{n} \left\| \left(p_i' m_b p_i\right)^{-} p_i' m_b \mathbf{d}_i \right\|^2 \\
&\leq nc_{nT} \lambda_{\max}\left(\overline{P}_w\right) \left(c_{1\lambda}\right)^{-2} T^{-2} \sum_{i=1}^{n} \|p_i\|^2 \|\mathbf{d}_i\|^2 \\
&= nc_{nT} O_p\left(nK^{1-2\lambda/d}\right) = O_p\left(Tn^{1/2}K^{1-2\lambda/d}\right),
\end{aligned}$$

we have by the Cauchy-Schwarz inequality

$$\begin{aligned}\left|c_{nT}\sum_{1\le i\ne j\le n}\left(\overrightarrow{r}_i-\overrightarrow{r}_j\right)'\overline{P}_w\left(\overrightarrow{r}_i-\overrightarrow{r}_j\right)\right| &= \left|c_{nT}\sum_{1\le i,j\le n}\left\{\overrightarrow{r}_i'\overline{P}_w\overrightarrow{r}_i+\overrightarrow{r}_j'\overline{P}_w\overrightarrow{r}_j-2\overrightarrow{r}_i'\overline{P}_w\overrightarrow{r}_j\right\}\right| \\ &\le 4nc_{nT}\sum_{i=1}^n\overrightarrow{r}_i'\overline{P}_w\overrightarrow{r}_i=O_p\left(Tn^{1/2}K^{1-2\lambda/d}\right).\end{aligned}$$

(v) Write $c_{nT}\sum_{1\le i<j\le n}\left(\overrightarrow{\varepsilon}_i-\overrightarrow{\varepsilon}_j\right)'\overline{P}_w\left(\widetilde{r}_i-\overrightarrow{r}_i\right)=W_{1nT}-W_{2nT}$, where $W_{1nT}\equiv nc_{nT}\sum_{i=1}^n\overrightarrow{\varepsilon}_i'\overline{P}_w\left(\widetilde{r}_i-\overrightarrow{r}_i\right)$ and $W_{2nT}\equiv c_{nT}\sum_{1\le i,j\le n}\overrightarrow{\varepsilon}_j'\overline{P}_w\left(\widetilde{r}_i-\overrightarrow{r}_i\right)$. For W_{1nT}, by (B.3)

$$\begin{aligned}W_{1nT} &= nc_{nT}\sum_{i=1}^n\varepsilon_i'm_bp_i\left(p_i'm_bp_i\right)^-\overline{P}_w\left(p_i'm_hp_i\right)^-p_i'\left(m_h-m_b\right)f_2\gamma_{2i} \\ &\quad +nc_{nT}\sum_{i=1}^n\varepsilon_i'm_bp_i\left(p_i'm_bp_i\right)^-\overline{P}_w\left(T^{-1}p_i'm_hp_i\right)^-T^{-1}p_i'\left(m_h-m_b\right)\mathbf{d}_i \\ &\quad +nc_{nT}\sum_{i=1}^n\varepsilon_i'm_bp_i\left(p_i'm_bp_i\right)^-\overline{P}_w\left[\left(T^{-1}p_i'm_hp_i\right)^--\left(T^{-1}p_i'm_bp_i\right)^-\right]T^{-1}p_i'm_b\mathbf{d}_i \\ &\equiv W_{1nT,1}+W_{1nT,2}+W_{1nT,3}.\end{aligned}$$

Using (B.1) we further decompose $W_{1nT,1}$ as follows:

$$\begin{aligned}W_{1nT,1} &= -nc_{nT}\sum_{i=1}^n\varepsilon_i'm_bp_i\left(p_i'm_bp_i\right)^-\overline{P}_w\left(p_i'm_hp_i\right)^-p_i'\overline{v}^*\left(b'b\right)^{-1}b'f_2\gamma_{2i} \\ &\quad +nc_{nT}\sum_{i=1}^n\varepsilon_i'm_bp_i\left(p_i'm_bp_i\right)^-\overline{P}_w\left(p_i'm_hp_i\right)^-p_i'h[\left(b'b\right)^{-1}-\left(h'h\right)^{-1}]b'f_2\gamma_{2i} \\ &\quad -nc_{nT}\sum_{i=1}^n\varepsilon_i'm_bp_i\left(p_i'm_bp_i\right)^-\overline{P}_w\left(p_i'm_hp_i\right)^-p_i'h\left(h'h\right)^{-1}\overline{v}^*f_2\gamma_{2i} \\ &\equiv -W_{1nT,11}+W_{1nT,12}-W_{1nT,13}.\end{aligned}$$

It is easy to show that $W_{1nT,11}=\overline{W}_{1nT,11}+o_p(1)$, where $\overline{W}_{1nT,11}\equiv nc_{nT}\sum_{i=1}^n\varepsilon_i'm_bp_i(p_i'm_b p_i)^-\overline{P}_w\left(p_i'm_bp_i\right)^-p_i'\overline{v}^*\left(b'b\right)^{-1}b'f_2\gamma_{2i}$. Using the decomposition $\overline{v}^*=\overline{v}^++\overline{g}^*$, we can further decompose $\overline{W}_{1nT,11}=\overline{W}_{1nT,111}+\overline{W}_{1nT,112}$, where $W_{1nT,111}$ and $W_{1nT,112}$ are analogously defined as $W_{1nT,11}$ but with $\overline{v}^*$ being replaced by $\overline{v}^+$ and $\overline{g}^*$, respectively. It is easy to show $\left|\overline{W}_{1nT,111}\right|=O_p\left(K/T^{1/2}\right)$. For $\overline{W}_{1nT,112}$, noting that $E_{\mathcal{D}}\left(\overline{W}_{1nT,112}\right)=0$ and $E_{\mathcal{D}}[\left(\overline{W}_{1nT,112}\right)^2]=\left(nc_{nT}\right)^2O_p(K/T^2)=O_p(K/n)$. So $\overline{W}_{1nT,112}=O_p\left(K^{1/2}/n^{1/2}\right)$ and $W_{1nT,11}=O_p\left(K/T^{1/2}\right)$. Similarly, we can show that $W_{1nT,1s}=O_p\left(K/T^{1/2}\right)$ for $s=2,3$. It follows that $W_{1nT,1}=O_p\left(K/T^{1/2}\right)$. For $W_{1nT,2}$, we have

$$\begin{aligned}\left|W_{1nT,2}\right| &\le \left\{nc_{nT}\sum_{i=1}^n\varepsilon_i'm_bp_i\left(p_i'm_bp_i\right)^-\overline{P}_w\left(p_i'm_hp_i\right)^-p_i'm_b\varepsilon_i\right\}^{1/2}\left\{U_{1nT}\right\}^{1/2} \\ &= \left\{O_p\left(Kn^{1/2}\right)O_p\left(Tn^{-1/2}K^{1-2\lambda/d}\right)\right\}^{1/2}=O_p\left(T^{1/2}K^{1-\lambda/d}\right).\end{aligned}$$

Similarly, $W_{1nT,3} = O_p\left(T^{1/2}K^{1-\lambda/d}\right)$. Consequently $W_{1nT} = O_p\left(K/T^{1/2} + T^{1/2}K^{1-\lambda/d}\right)$. Analogously we can show that $W_{2nT} = O_p\left(K/T^{1/2} + T^{1/2}K^{1-\lambda/d}\right)$.

(vi) By the Cauchy-Schwarz inequality, it suffices to prove (vi) by showing that $W_{3nT} \equiv nc_{nT}\sum_{i=1}^{n} \overrightarrow{\varepsilon}'_i \overline{P}_w \overrightarrow{r}_i = O_p\left(T^{1/2}K^{1/2-\lambda/d}\right)$. Note that $E_{\mathcal{D}}\left(W_{3nT}\right) = 0$, and

$$\begin{aligned} E_{\mathcal{D}}\left[\left(W_{3nT}\right)^2\right] &= \left(nc_{nT}\right)^2 \sum_{i=1}^{n} \mathbf{d}'_i m_b p_i \left(p'_i m_b p_i\right)^- \overline{P}_w \left(p'_i m_b p_i\right)^- p'_i m_b E\left(\varepsilon_i \varepsilon'_i\right) \\ &\quad \times m_b p_i \left(p'_i m_b p_i\right)^- \overline{P}_w \left(p'_i m_b p_i\right)^- p'_i m_b \mathbf{d}_i \\ &\leq C\left(nc_{nT}\right)^2 \sum_{i=1}^{n} \mathbf{d}'_i m_b p_i \left(p'_i m_b p_i\right)^- \overline{P}_w \left(p'_i m_b p_i\right)^- \overline{P}_w \left(p'_i m_b p_i\right)^- p'_i m_b \mathbf{d}_i \\ &\leq C\left(nc_{nT}\right)^2 \left(\lambda_{\max}\left(\overline{P}_w\right)\right)^2 \left(c_{1\lambda}\right)^{-3} \left\{T^{-3}\sum_{i=1}^{n} \mathbf{d}'_i m_b p_i p'_i m_b \mathbf{d}_i\right\} \\ &= \left(nc_{nT}\right)^2 O_p\left(nK^{1-2\lambda/d}/T\right) = O_p\left(TK^{1-2\lambda/d}\right). \end{aligned}$$

It follows that $W_{3nT} = O_p\left(T^{1/2}K^{1/2-\lambda/d}\right)$. ■

References

Bai, J., 2003. Inferential theory for factor models of large dimension. Econometrica 71, 135-171.

Bai, J., 2009. Panel data models with interactive fixed effects. Econometrica 77, 1229-1279.

Baltagi, B.H., Hidalgo, J., Li, Q., 1996. A nonparametric test for poolability using panel data. Journal of Econometrics 75, 345-367.

Baltagi, B. H., Li, D. 2002. Series estimation of partially linear panel data models with fixed effects. Annals of Economics and Finance 3, 103-116.

Bao, Y., Ullah, A., 2009. Expectation of quadratic forms in normal and nonnormal variables with econometric applications. Working paper, Dept. of Economics, Univ. of California, Riverside.

Bun, M. J. G., 2004. Testing poolability in a system of dynamic regressions with nonspherical disturbances. Empirical Economics 29, 89-106.

Camba-Mendeza, G., Kapetanios, G., 2009. Statistical tests and estimators of the rank of a matrix and their applications in econometric modelling. Econometric Reviews 28, 581-611.

Chow, G. C., 1960. Tests of equality between sets of coefficients in two linear regressions. Econometrica 28, 591-605.

Cridao, C. O., 2008. Temporal and spatial homogeneity in air pollutants panel EKC estimations: two nonparametric tests applied to Spanish Provinces. Environmental and Resource Economics 40, 265-283.

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

Greenaway-McGrevy, R., Han, C., Sul, D., 2009. Estimating and testing idiosyncratic equations using cross-section dependent panel data. Working paper, Dept. of Economics, Univ. of Auckland.

Hall, P., Hart, J. D., 1990. Bootstrap test for difference between means in nonparametric regression. Journal of the American Statistical Association 85, 1039-1049.

Harding, M. C., 2007. Structural estimation of high-dimensional factor models: uncovering the effect of global factors on the US economy. Mimeo, Dept. of Economics, MIT.

Kapetanios, G., Pesaran, M. H., 2005. Alternative approaches to estimation and inference in large multifactor panels: small sample results with an application to modelling of asset returns. Working paper, Cambridge Univ.

Koul, H. L., Schick, A., 1997. Testing for the equality of two nonparametric regression curves. Journal of Statistical Planning and Inference 65, 293-314.

Lavergne, P., 2001. An equality test across nonparametric regressions. Journal of Econometrics 103, 307-344.

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

Li, Q., 2000. Efficient estimation of additive partially linear models. International Economic Review 41, 1073-1092.

Li, Q., Hsiao, C., Zinn, J., 2003. Consistent specification tests for semiparametric/nonparametric models based on series estimation methods. Journal of Econometrics 112, 295-325.

Magnus, J. R., Neudecker, H., 1999. Matrix Differential Calculus with Applications in Statistics and Econometrics. John Wiley & Sons, New York.

Neumeyer, N., Dette, H., 2003. Nonparametric comparison of regression curves: an empirical process approach. Annals of Statistics 31, 880-920.

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

Pesaran, M. H., Tosetti, E., 2007. Large panels with common factors and spatial correlations. Working paper no. 2103, Cambridge Univ.

Pesaran, M. H., Yamagata T., 2008. Testing slope homogeneity in large panels. Journal of Econometrics 142, 50-93.

Phillips, P. C. B., Sul, D., 2003. Dynamic panel estimation and homogeneity testing under cross sectional dependence. The Econometrics Journal 6, 217-259.

Phillips, P. C. B., Sul, D., 2007. Bias in dynamic panel estimation with fixed effects, incidental trends and cross section dependence. Journal of Econometrics 137, 162-188.

Su, L., Jin, S., 20010. Sieve estimation of panel data models with cross section dependence. Working paper, School of Economics, Singapore Management Univ.

Vilar-Fernández, J. M., González-Manteiga W., 2004. Nonparametric comparison of curves with dependent errors. Statistics 38, 81-99.

Vilar-Fernández, J. M., Vilar-Fernández, J. M., González-Manteiga W., 2007. Bootstrap tests for nonparametric comparison of regression curves with dependent errors. Test 16, 123-144.