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

Sainan Jin\textsuperscript{a}, Liangjun Su\textsuperscript{a}

\textsuperscript{a}School of Economics, Singapore Management University, Singapore

September 17, 2011

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. A small set of Monte Carlo simulations indicate the test performs reasonably well in finite samples.

\textbf{JEL Classifications:} C13, C14, C33

\textbf{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 (2011), Harding (2007), Kapetanios and Pesaran (2005), Moon and Weidner (2010a, 2010b), Pesaran (2006), Pesaran and Tosetti (2011), Phillips and Sul (2003, 2007), Su and Chen (2011), among others, for an overview. All of these papers focus on the linear specification of regression relationship. More recently, Su and Jin (2011, SJ hereafter) have extended the linear model of Pesaran (2006) to the following semiparametric panel data model

\textsuperscript{*}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 authors gratefully thank the editor, associate editor, and three anonymous referees for their constructive suggestions and comments. The first author acknowledges financial support from the SMU research grant (Grant number: C244/MSS8E002).
with multi-factor error structure

\[
\begin{align*}
y_{it} &= g_i(x_{it}) + \gamma_{1i} f_{1i} t + \epsilon_{it}, \\
x_{it} &= \Gamma_{1i} f_{1i} t + \Gamma_{2i} f_{2i} t + \nu_{it},
\end{align*}
\]

where \( x_{it} \in X_i \subset \mathbb{R}^d \) is a vector of observed individual-specific regressors on the \( i \)th cross section unit at time \( t \), \( g_i(\cdot) \in G_i \), \( G_i \) is a specified class of continuous function from \( X_i \) to \( \mathbb{R} \), \( f_{1i} \) is a \( q_1 \times 1 \) vector of observed common factors that contains the constant term 1, \( f_{2i} \) is a \( q_2 \times 1 \) vector of unobserved common factors, \( \gamma_{1i} \) and \( \gamma_{2i} \), are factor loadings, \( \epsilon_{it} \) is the individual-specific (idiosyncratic) errors assumed to be independently distributed of \( (x_{it}, f_{1i}, f_{2i}) \), and \( \nu_{it} \), \( \Gamma_{1i} \) and \( \Gamma_{2i} \) are \( q_1 \times d \) and \( q_2 \times d \) factor loading matrices, and \( \nu_{it} \) is a \( d \times 1 \) vector of individual-specific components of \( x_{it} \). If the regression function \( g_i(\cdot) \) is not identical across \( i \), then we have heterogeneous 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. 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 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

1 Write \( f_{1i} = (1, f_{1i}')' \). As SJ remarked, we can allow \( f_{1i} ' \) to enter \((1.1)\) nonparametrically, in which case \((1.1)\) will become
\[
y_{it} = g_i(x_{it}, f_{1i}' + \gamma_{1i} + \epsilon_{it}.
\]

Out 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 \( f_{1i} \equiv 1 \).

2 If \( f_{1i} \equiv 1 \) and \( f_{2i} \equiv 0 \) 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.
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. 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 $\|A\| = [\text{tr}(AA^\prime)]^{1/2}$ and its generalized inverse as $A^\prime$. When $A$ is a symmetric matrix, we use $\lambda_{\text{min}}(A)$ and $\lambda_{\text{max}}(A)$ to its minimum and maximum eigenvalues respectively, and $\text{diag}(A)$ to denote the diagonal matrix formed from the diagonal elements of $A$. $I_T$ denotes a $T \times T$ identity matrix. For a vector $a = (a_1, \ldots, a_T)'$, $\text{diag}(a)$ denotes a diagonal matrix with $a_i$ as a typical diagonal element. The operator $p \to$ denotes convergence in probability, and $d \to$ 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

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

(2.1)

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

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

(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

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

(2.3)

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

(2.4)
To facilitate the local power analysis, we also define a sequence of Pitman local alternatives:

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

where $\Delta_n (x)$ is uniformly bounded measurable functions, $\gamma_{nT} \rightarrow 0$ as $(n, T) \rightarrow \infty$, and the exact rate of $\gamma_{nT}$ 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

$$\Gamma_n = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \int [g_i (x) - g_j (x)]^2 w (x) \, dx,$$

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 $\Gamma_n = 0$ if and only if $H_0$ holds and we can propose a test statistic based upon consistent estimation of $\Gamma_n$.

To proceed, it is worth mentioning that a test statistic based on $\Gamma_n$ in (2.6) cannot distinguish all deviations from $H_0$ in (2.1). As a referee kindly suggests, we may regard our test as testing

$$\Pi_0 : \Delta_{\theta} = 0 \text{ versus } \Pi_1 : \Delta_{\theta} > 0$$  \hspace{1cm} (2.7)

where $\Delta_{\theta} = \lim_{n \rightarrow \infty} \Delta_{gn}$ and $\Delta_{gn} = \lfloor n (n-1) \rfloor^{-1} \Gamma_n$. $\Pi_0$ allows for some pair $(i, j)$ with $i \neq j$, $g_i (x) \neq g_j (x)$ with probability greater than zero, but the count measure of such pair has to be of smaller order than $n (n-1)$. In contrast, $H_0$ seems too restrictive to be useful in practice because it requires $g_i (x) = g_j (x)$ for all pairs $(i, j)$ with $i \neq j$. From this point of view, testing $\Pi_0$ against $\Pi_1$ seems more interesting than testing $H_0$ against $H_1$. The formulation in (2.7) is extremely useful when we consider the global power behavior of our test statistic in Theorem 3.4 below. See also the remark after it.  \(^3\)

2.2 Estimation and test statistic

To estimate $\Gamma_n$, 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 $K \rightarrow \infty$ as $(n, T) \rightarrow \infty$. \(^4\) Let $p^K (x) = (p_1 (x), p_2 (x), \cdots, p_K (x))'$, $p_{it} = p^K (x_{it})$, and $p_t = (p_{t1}, p_{t2}, \cdots, p_{tT})'$. Obviously we have suppressed the dependence of $p_{it}$, and $p_t$, on $K$ and $T$. In particular, $p_t$ is a $T \times K$ matrix.

Under fairly weak conditions, we can approximate $g_i (x)$ very well by $\alpha'_{gi} p^K (x)$ for some $K \times 1$ vector $\alpha_{gi}$. Let $\tilde{x}_l \equiv n^{-1} \sum_{i=1}^{n} x_{it}$, $\tilde{y}_l \equiv n^{-1} \sum_{i=1}^{n} y_{it}$, and $h_l \equiv (f'_{it}, \tilde{x}'_t, \tilde{y}'_t)'. \hspace{1cm}$ As SJ argued, we can use $h_l$ as an

\(^3\)In the proof of our asymptotic results, nevertheless, we find it is easy to impose $H_0$ to derive the asymptotic null distribution of our test statistic defined below.

\(^4\)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.
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 \( h_t \):

\[
y_{it} = \alpha^t_g p^K(x_{it}) + \tilde{\nu}^t_i h_t + u_{it}
\]

where \( u_{it} \) is the new error term. By the formula for partitioned regression, the estimate of \( \alpha_g \) is given by

\[
\hat{\alpha}_g_i = (p'_i m_h p_i)^{-} p'_i m_h y_i,
\]

where \( h \equiv (h_1, h_2, \cdots, h_T)' \), \( y_i \equiv (y_{i1}, y_{i2}, \cdots, y_{iT})' \), \( m_h \equiv I_T - h (h' h)^{-1} h \), and \((\cdot)^-\) denotes any symmetric generalized inverse. Then we estimate \( g_i(\cdot) \) by

\[
\hat{g}_i(x) = p^K(x)' \hat{\alpha}_g_i.
\]

With \( \hat{g}_i(x) \), we then estimate \( \Gamma_n \) by the following functional:

\[
\hat{\Gamma}_n = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \int [\hat{g}_i(x) - \hat{g}_j(x)]^2 w(x) dx.
\]

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, \( \hat{\Gamma}_n \) 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 sequence of Pitman local alternatives, and fixed alternatives.

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 \( \bar{P}_w \equiv \int p^K(x) p^K(x)' w(x) dx \).

Define

\[
Q_{ipp} \equiv E[p_{ih} p_{ih}'] , \quad Q_{iph} \equiv E[p_{ih} h_i'] , \quad Q_{hh} \equiv E[h_i h_i'] , \text{ and } Q_i \equiv Q_{ipp} - Q_{iph} Q_{hh}^{-1} Q_{iph} ',
\]

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 \( i \), \( \varepsilon_{it} \) are independently and identically distributed (IID) with mean 0 and variance \( \sigma^2_i \), and the process \( \{v_{it} : t \geq 1\} \) is a strictly stationary and \( \alpha \)-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 \( \eta > 0 \). \( \xi \leq \min_{1 \leq i \leq n} \sigma^2_i \leq \max_{1 \leq i \leq n} \sigma^2_i \leq \kappa \) for some \( \xi > 0 \) and \( \kappa < \infty \). (ii) The common factor process \( \{(f_{1t}, f_{2t}) : t \geq 1\} \) is a
strictly stationary and $\alpha$-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 $\varepsilon_{1s}$ and $v_{1s}$ for all $i$, $t$, and $s$. $E[(f_{1t}', f_{2t}')']$ is positive definite. (iv) The individual-specific errors $\varepsilon_{it}$ and $v_{js}$ are distributed independently for all $i$, $j$, $t$, and $s$. (v) $(\varepsilon_{it}, v_{jt})$ are independently distributed across $i$ with zero means. (vi) $E[\varepsilon_{it}^2] < \infty$, and $\sup_{n \geq 1} \max_{1 \leq i \leq n} E[|\varepsilon_{it}|^{4+\eta}] \leq \mathcal{P}_{4+\eta} < \infty$ for $\zeta_1 = \varepsilon_{11}$, $g_1(x_{11})$, $f_{11}$, and $f_{21}$. (vii) Let $\alpha(j) \equiv \sup_{n \geq 1} \max_{1 \leq i \leq n} \alpha_i(j)$. $\sum_{j=1}^{\infty} j^2 \alpha(j)^{\eta/(4+\eta)} \leq C_3 < \infty$. (viii) $E[g_i(x_{it})] = 0$ for all $i$.

**Assumption 2.** (i) The unobserved factor loadings $\gamma_{2i}$ and $\Gamma_{2i}$ are IID. $\gamma_{2i}$ and $\Gamma_{2i}$ are independent of the individual-specific errors $\varepsilon_{jt}$ and $v_{jt}$, and the common factors $(f_{1t}, f_{2t})$ for all $j$ and $t$. The $(4 + \eta)$-th moment of $\Gamma_{2i}$ is finite. (ii) $\Gamma_{1i}$ 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 $\Gamma_2^* = E (\Gamma_{2i})$ where $\Gamma_2^* \equiv (\Gamma_{21}, \gamma_{2i})$. rank($\Gamma_2^*$) = $q_2 \leq d + 1$.

**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 $i$, $\int (1 + ||x||^2)^{\overline{\omega}_i} dF_i(x) < C < \infty$ 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 $\Pi_{\infty,K} g_i \equiv \alpha_i' p^K(\cdot)$ in the sieve space $\mathcal{G}_K \equiv \{ f(\cdot) = \alpha' p^K(\cdot) \}$ such that $||g_i(\cdot) - \Pi_{\infty,K} g_i(\cdot)||_{\infty, \mathcal{X}_i} = O \left( K^{-\lambda_i/d} \right)$. (iv) For each $i$, $\sup_{1 \leq j \leq K} E|p_j(x_{i1})|^4 + \eta < \infty$ for the same $\eta$ defined in Assumption 1(i), $Q_1$ 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) $\int (1 + ||x||^2)^{\overline{\omega}_i} w(x) dx < \infty$ where $\overline{\omega}_i \equiv \max_{1 \leq i \leq n} \omega_i$. (iii) For each $K$, the smallest and largest eigenvalues of $\overline{\mathcal{P}}_w$ 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 $\zeta(K) \equiv \sup_x \| p^K(x) \|$. 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 $\hat{g}_i(x)$. 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 $\varepsilon_{it}$ 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

\[ it \text{ is desirable to have a statistical test to check such a rank condition. For a recent review on the methods to test the rank of a general matrix, see Camba-Mendez and Kapetanios (2009). Nevertheless, it seems too difficult to apply any of the reviewed method to our case because we do not have repeated panel observations to estimate the expected value of the matrix of factor loadings, i.e., } E(\Gamma_{2i}^*). \]
form of \( \varepsilon_i \) 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.\(^6\) 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 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, \( \mathbf{P}_w \) reduces to \( Q_{ipp} \) which is the same across \( i \).

### 3.2 Asymptotic distributions

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

\[
h_t = \Gamma'_{2t} + \mathbf{v}_t, \tag{3.2}
\]

where

\[
\Gamma = \begin{pmatrix} I_{q_1} \\ 0 \end{pmatrix} \quad \Gamma_1 = \begin{pmatrix} \gamma_{1,1} \\ \gamma_{1,2} \end{pmatrix}, \quad \mathbf{v}_t = \begin{pmatrix} \mathbf{v}_t^1 \\ \mathbf{v}_t^2 \end{pmatrix}, \quad \mathbf{v}_t = \begin{pmatrix} \mathbf{v}_t^1 \\ \mathbf{v}_t^2 \end{pmatrix}, \quad \gamma_{1,1} = \gamma_{1,2} = \gamma_{2,1} = \gamma_{2,2} = \gamma, \quad \mathbf{v}_t = \begin{pmatrix} \mathbf{v}_t^1 \\ \mathbf{v}_t^2 \end{pmatrix}, \quad \mathbf{v}_t = \begin{pmatrix} \mathbf{v}_t^1 \\ \mathbf{v}_t^2 \end{pmatrix}, \quad \gamma_{1,1} = \gamma_{1,2} = \gamma_{2,1} = \gamma_{2,2} = \gamma.
\]

(3.2) justifies the use of \( h_t \) as the proxy for \( z_t \). Let \( b_t \equiv \Gamma' z_t, b \equiv (b_1, \ldots, b_T)' \), and \( m_b \equiv I_T - b(b'b)^{-1} b \). Define

\[
B_{nT} \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sigma^2_i \text{tr} \left( (p_i m_b p_i/T)^{-} \mathbf{P}_w \right), \tag{3.4}
\]

and

\[
V_{nT} \equiv \frac{2}{\sqrt{n}} \sum_{i=1}^{n} \sigma^2_i \text{tr} \left( (p_i m_b p_i/T)^{-} \mathbf{P}_w \right)^2, \tag{3.5}
\]

The following theorem establishes the asymptotic normality of \( \hat{\Gamma}_{nT} \) after being appropriately scaled and centered.

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

\[
c_{nT} \hat{\Gamma}_{nT} - B_{nT} \xrightarrow{d} \mathcal{N}(0,1),
\]

\(^6\) 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-restricted 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 analogous to undersmoothing for the kernel-based test.
where $c_{nT} = \frac{T}{\sqrt{n(n-1)}}$.

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

$$c_{nT} \hat{\Gamma}_{nT} = c_{nT} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left( \overline{\varepsilon}_i - \overline{\varepsilon}_j \right)^T \mathcal{P}_w \left( \overline{\varepsilon}_i - \overline{\varepsilon}_j \right) + o_p(1),$$

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

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

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 $\infty$ at the rate $K$, slower than the $(nK)^{1/2}$-rate at which $B_{nT}$ is diverging to $\infty$). Let $\hat{g}_i(\equiv (\hat{g}_i(x_{i1}), \ldots, \hat{g}_i(x_{iT}))' \equiv m_b(y_i - \hat{\mu}_i)$. Denote the $t$th element of $\hat{\varepsilon}_i$ as $\hat{\varepsilon}_{it}$. We propose to estimate $V_{nT}$ by

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

where $\hat{\sigma}_i^2 \equiv \frac{1}{n} \sum_{t=1}^{n} \hat{\varepsilon}_{it}^2$. It is straightforward to demonstrate that $\hat{V}_{nT}/V_{nT} = 1 + o_p(1)$. For $B_{nT}$, a simple replacement of $m_b$ and $\sigma_i^2$ by $m_b$ and $\hat{\sigma}_i^2$ won’t deliver a consistent estimate. In fact, we can show that

$$B_{nT} - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\sigma}_i^2 \text{tr} \left( (p_i'm_b p_i/T)^{-} \mathcal{P}_w \right)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\sigma}_i^2 \text{tr} \left\{ \left( (p_i'm_b p_i/T)^{-} - (p_i'm_b p_i/T)^{-1} \right) \mathcal{P}_w \right\} + o_p(1).$$

By using Lemma B.3(iv) in the appendix and the decomposition for $\hat{\sigma}_i^2$ in (A.11), we can also show that the dominant term in the last expression is $O_p(K)$, which also needs to be estimated. Let $\hat{\mathbf{G}}_i \equiv \frac{1}{n} \sum_{t=1}^{n} \hat{g}_i(x_{it}), \hat{\mathbf{G}}_i^* \equiv (0_{1 \times (q_1+d)}, \hat{\mathbf{G}}_i)'$, and $\hat{\mathbf{G}}^* \equiv (\hat{\mathbf{G}}_1, \ldots, \hat{\mathbf{G}}_T)'.$ Define

$$\hat{b}_{nT} = \frac{T}{\sqrt{n}} \sum_{i=1}^{n} \hat{\sigma}_i^2 \text{tr} \left\{ (p_i'm_b p_i/T)^{-} \mathcal{P}_w \right\},$$

and

$$\hat{B}_{nT} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\sigma}_i^2 \text{tr} \left( (p_i'm_b p_i/T)^{-} \mathcal{P}_w \right) \hat{b}_{nT},$$

where $\hat{b}_{nT} \equiv p_i'(\mathbf{H}^*)^{-} \mathbf{H}^* \hat{P}^* + \hat{\mathbf{P}}^* \hat{\mathbf{P}}^* \mathbf{H} \mathbf{H}^{-1} h' p_i$. As we demonstrate in the appendix, $\hat{B}_{nT} - B_{nT} = o_p(\sqrt{\hat{V}_{nT}})$. Thus we have the following corollary.
Corollary 3.2 Let $D_{nT} \equiv (c_n \hat{\Gamma}_{nT} - \hat{B}_{nT})/\sqrt{V_{nT}}$. Suppose Assumptions 1-5 hold with Assumption 3(ii) being strengthened to: $\int (1 + ||x||^2)z^2 dF_i(x) < C < \infty$ for some $\overline{\lambda}_i > \omega_i + \lambda_i$ and for all $i$. Then

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

Corollary 3.2 indicates that we can compare $D_{nT}$ to $z_\alpha$, the $\alpha$th 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

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

Then we have the following theorem.

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

Theorem 3.3 suggests that our test has nontrivial power to detect Pitman local alternatives at the rate $n^{-1/4}T^{-1/2} \text{tr}(\bar{P}_w)^{1/4}$, which is slower than the rate $n^{-1/4}T^{-1/2}$ as $\text{tr}(\bar{P}_w) = O(K)$. 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 $P \left( D_{nT} > z_\alpha | H_1(n^{-1/4}T^{-1/2} \text{tr}(\bar{P}_w)^{1/4}) \right) \to 1 - \Phi(z_\alpha - \Delta)$ 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 Suppose Assumptions 1-5 hold. Then under $H_1$ in (2.7), $P(D_{nT} > d_{nT}) \to 1$ for any sequence $d_{nT} = o_p(n^{1/2}T \text{tr}(\bar{P}_w)^{-1/2})$ as $(n,T) \to \infty$.

Theorem 3.4 indicates that under $H_1$ our test statistic $D_{nT}$ explodes at the rate $n^{1/2}T \text{tr}(\bar{P}_w)^{-1/2}$ provided $\Delta_g > 0$, where $\Delta_g$ is defined below (2.7). 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 a finite number of others on a set of positive measure (e.g., only $g_1(\cdot)$ is different from a finite number of others), or the case where the cardinality of the set of functions among $\{g_i(\cdot)\}_{i=1}^n$ that are distinct from a diverging number of 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, $\Delta_g$ shrinks to 0 at rate $n^{-2}$ so that the $D_{nT}$ test cannot be consistent; in the latter case, our test can still be consistent as long as $\Delta_g$ does not shrink to 0 too fast so that $D_{nT}$ is still diverging to infinity (now at a rate slower than $n^{1/2}T \text{tr}(\bar{P}_w)^{-1/2}$) as $(n,T) \to \infty$.

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:

$$
y_{it} = g_i(x_{it,1}, x_{it,2}) + \gamma_{i1} + \gamma_{2i,1}f_{2i,1} + \gamma_{2i,2}f_{2i,2} + \varepsilon_{it},$$

$$
x_{it,s} = \Gamma_{1i,s} + \Gamma_{2s,1}f_{2i,1} + \Gamma_{2s,2}f_{2i,2} + \nu_{it,s}, \ s = 1, 2,$$

for $i = 1, 2, \ldots, n$, and $t = 1, 2, \ldots, T$. Clearly, we have two individual-specific regressors ($x_{it} \equiv (x_{it,1}, x_{it,2})'$), one observed common factor ($f_{1i} = 1$), and two unobserved common factors ($f_{2i} \equiv (f_{2i,1}, f_{2i,2})'$). We generate $\nu_{it} \equiv (\nu_{it,1}, \nu_{it,2})'$, $x_{it}$, the unobserved factors ($f_{2i}$), 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.

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

$$
g(x_{it,1}, x_{it,2}) = 0.5x_{it,1} + 0.5x_{it,2}.
$$

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

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

Under the alternative, we consider $g_i(x_{it,1}, x_{it,2}) = g(x_{it,1}, x_{it,2}) + \delta_i \cos(\pi x_{it,2})$ where $g$ is specified as above, $\delta_i$'s are IID$[0, c]$ in DGPs 1 and 2, $\delta_i = 0$ for all $i = 1, 2, \ldots, n/2$, and $\delta_i$'s are IID$[0, c]$ for $i = n/2 + 1, n/2 + 2, \ldots, n$ in DGP 3 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 $g_i(x_{it,1}, x_{it,2})$ in SJ by augmenting the sieve regression of $y_{it}$ on $x_{it}$ with only observed common factor $f_{1i}$ (i.e., $h_t = f_{1i}$ in (2.8)). 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 $g_i(x_1, x_2)$ has the additive structure and can be written as the sum of $g_{i1}(x_1)$ and $g_{i2}(x_2)$, [e.g., $g_{i1}(x_1) \equiv \exp(x_{it,1})/\exp(x_{it,1}) + 1$ and $g_{i2}(x_2) \equiv (0.5x_{it,2} - 0.25x_{it,2}^2) + \delta_i \cos(\pi x_{it,2})$ 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:

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

$^7$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 $s = 1, 2$. 

10
where \( \hat{g}_{-i}^{(J)}(x_{it}) \), \( t = 1, \cdots, T \), 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. Our bootstrap procedure is in the spirit of Hansen’s (2000) fixed-regressor bootstrap and goes as follows:

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

2. For \( i = 1, \cdots, n \) and \( t = 1, \cdots, T \), generate \(^8\)

\[
y_{it}^* = \hat{g}(x_{i1}, x_{i2}, \hat{\gamma}_{1i} + \hat{\gamma}_{2i} f_{2t} + \hat{e}_{it})
\]

where \( \hat{e}_{it}^* \) is the \( t \)-th element of \( \hat{e}_{it}^* = (\hat{e}_{i1}^*, \hat{e}_{i2}^*, \cdots, \hat{e}_{iT}^*)' \), and \( \hat{e}_{it}^* \) is a random drawn from \( \{\tilde{e}_{1}, \tilde{e}_{2}, \cdots, \tilde{e}_{n}\} \) with replacement.

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

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

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 \( H_0 \), \( \hat{f}_{2t} \) and \( \hat{\gamma}_{2i} \) consistently estimate \( f_{2t} \) and \( \gamma_{2i} \) 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 \( \hat{g}(x_{i1}, x_{i2}) \) is now replaced by SJ’s naive estimate \( \hat{g}^{(N)}(x_{i1}, x_{i2}) \) of \( g(x_{i1}, x_{i2}) \) under the null, there is no need to estimate \( f_{2t} \) and \( \gamma_{2i} \), and one can estimate \( \gamma_{1i} \) by \( \hat{\gamma}_{1i} \equiv T^{-1} \sum_{t=1}^{T} [y_{it} - \hat{g}^{(N)}(x_{i1}, x_{i2})] \). 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

\(^8\)Following Hansen (2000), we can treat \( (x_{i1}, x_{i2}, \hat{f}_{2t}) \) as the fixed-regressor and \( (\hat{g}, \hat{\gamma}_{1i}, \hat{\gamma}_{2i}) \) as the parameter to be estimated in the bootstrap world.
Figure 1: Finite sample distribution of the test statistic when the null hypothesis is satisfied in DGPs 1 and 2.

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. But the empirical levels of the test at the 5% and 10% nominal levels tend to be close to each other when $n$ and $T$ are large, indicating the asymptotic normal null distribution cannot well approximate the finite sample distribution of the test statistic for such sample sizes. Figure 1 reports the finite sample density of $D_{nT}$ in DGPs 1-2 for $(n,T) = (50,20)$ and $(100,100)$ based on 1000 replications where the density is estimated by using the Gaussian kernel and Silverman’s rule of...
Table 1: Finite sample rejection frequency under the null (nominal level: 0.05 and 0.10)

<table>
<thead>
<tr>
<th>DGP</th>
<th>n</th>
<th>T</th>
<th>5% tests</th>
<th>10% tests</th>
<th>Tests based on asymptotic normal critical values</th>
<th>J_{cv1}</th>
<th>J_{cv2}</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>D_{nT}</td>
<td>D_{nT}^{(N)}</td>
<td>D_{nT}^{(PY)}</td>
</tr>
<tr>
<td>1</td>
<td>50</td>
<td>20</td>
<td>0.131</td>
<td>0.112</td>
<td>1.000</td>
<td>0.176</td>
<td>0.145</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.103</td>
<td>0.135</td>
<td>1.000</td>
<td>0.138</td>
<td>0.176</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.080</td>
<td>0.164</td>
<td>1.000</td>
<td>0.111</td>
<td>0.207</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.045</td>
<td>0.053</td>
<td>1.000</td>
<td>0.069</td>
<td>0.074</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>25</td>
<td>0.148</td>
<td>0.145</td>
<td>1.000</td>
<td>0.183</td>
<td>0.182</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.090</td>
<td>0.078</td>
<td>1.000</td>
<td>0.104</td>
<td>0.105</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.027</td>
<td>0.019</td>
<td>1.000</td>
<td>0.045</td>
<td>0.040</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.054</td>
<td>0.056</td>
<td>1.000</td>
<td>0.066</td>
<td>0.071</td>
</tr>
<tr>
<td>2</td>
<td>50</td>
<td>20</td>
<td>0.110</td>
<td>0.069</td>
<td>1.000</td>
<td>0.159</td>
<td>0.093</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.079</td>
<td>0.095</td>
<td>1.000</td>
<td>0.102</td>
<td>0.128</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.079</td>
<td>0.104</td>
<td>1.000</td>
<td>0.102</td>
<td>0.142</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.073</td>
<td>0.061</td>
<td>1.000</td>
<td>0.100</td>
<td>0.088</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>25</td>
<td>0.069</td>
<td>0.096</td>
<td>1.000</td>
<td>0.105</td>
<td>0.126</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.043</td>
<td>0.069</td>
<td>1.000</td>
<td>0.059</td>
<td>0.085</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.028</td>
<td>0.017</td>
<td>1.000</td>
<td>0.040</td>
<td>0.031</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.046</td>
<td>0.041</td>
<td>1.000</td>
<td>0.060</td>
<td>0.054</td>
</tr>
<tr>
<td>1</td>
<td>50</td>
<td>20</td>
<td>0.052</td>
<td>0.217</td>
<td>0.056</td>
<td>0.116</td>
<td>0.274</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.077</td>
<td>0.310</td>
<td>0.035</td>
<td>0.137</td>
<td>0.374</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.060</td>
<td>0.332</td>
<td>0.018</td>
<td>0.106</td>
<td>0.386</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.063</td>
<td>0.274</td>
<td>0.013</td>
<td>0.117</td>
<td>0.320</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>25</td>
<td>0.074</td>
<td>0.358</td>
<td>0.000</td>
<td>0.139</td>
<td>0.406</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.077</td>
<td>0.406</td>
<td>0.000</td>
<td>0.134</td>
<td>0.451</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.062</td>
<td>0.375</td>
<td>0.000</td>
<td>0.118</td>
<td>0.423</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.061</td>
<td>0.366</td>
<td>0.000</td>
<td>0.113</td>
<td>0.428</td>
</tr>
<tr>
<td>2</td>
<td>50</td>
<td>20</td>
<td>0.067</td>
<td>0.190</td>
<td>0.347</td>
<td>0.121</td>
<td>0.254</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.067</td>
<td>0.238</td>
<td>0.287</td>
<td>0.123</td>
<td>0.308</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.078</td>
<td>0.284</td>
<td>0.300</td>
<td>0.128</td>
<td>0.344</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.050</td>
<td>0.197</td>
<td>0.298</td>
<td>0.096</td>
<td>0.253</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>25</td>
<td>0.068</td>
<td>0.317</td>
<td>0.768</td>
<td>0.122</td>
<td>0.386</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.067</td>
<td>0.341</td>
<td>0.864</td>
<td>0.107</td>
<td>0.389</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.060</td>
<td>0.296</td>
<td>0.913</td>
<td>0.113</td>
<td>0.338</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.049</td>
<td>0.280</td>
<td>0.924</td>
<td>0.090</td>
<td>0.316</td>
</tr>
</tbody>
</table>

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)

<table>
<thead>
<tr>
<th>DGP</th>
<th>nT</th>
<th>Tests based on asymptotic normal critical values</th>
<th>Tests based on bootstrap p-values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>c = 1</td>
<td>D_{nT}</td>
<td>D_{nT}^{(N)}</td>
</tr>
<tr>
<td></td>
<td>c = 2</td>
<td>D_{nT}</td>
<td>D_{nT}^{(N)}</td>
</tr>
<tr>
<td>1</td>
<td>50</td>
<td>0.298</td>
<td>0.172</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.308</td>
<td>0.336</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>0.338</td>
<td>0.376</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.332</td>
<td>0.264</td>
</tr>
<tr>
<td>100</td>
<td>25</td>
<td>0.326</td>
<td>0.236</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.346</td>
<td>0.226</td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>0.336</td>
<td>0.148</td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>0.524</td>
<td>0.252</td>
</tr>
<tr>
<td>2</td>
<td>50</td>
<td>0.216</td>
<td>0.130</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.262</td>
<td>0.232</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>0.302</td>
<td>0.290</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.378</td>
<td>0.262</td>
</tr>
<tr>
<td>100</td>
<td>25</td>
<td>0.170</td>
<td>0.124</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.226</td>
<td>0.152</td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>0.164</td>
<td>0.102</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.234</td>
<td>0.182</td>
</tr>
<tr>
<td>3</td>
<td>50</td>
<td>0.134</td>
<td>0.088</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.140</td>
<td>0.152</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>0.186</td>
<td>0.190</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.208</td>
<td>0.148</td>
</tr>
<tr>
<td>100</td>
<td>25</td>
<td>0.100</td>
<td>0.096</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.116</td>
<td>0.100</td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>0.136</td>
<td>0.076</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.120</td>
<td>0.116</td>
</tr>
</tbody>
</table>

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.
to 2), the power of the number of series terms non-bootstrapped) tends to be more oversized for small values of the power may be cancelling the effect of the increase of tr(\(P_w\)) 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) where 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 \(D_{nT}^{(N)}\) test increases rapidly. For the other two tests, we notice that surprisingly the normal-critical-value-based \(D_{nT}^{(N)}\) 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 \(D_{nT}^{(N)}\) test.

Columns 7-8 (resp. 12-13) 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.

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 tr(\(P_w\)) 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) where 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 \(D_{nT}^{(N)}\) 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 \(D_{nT}^{(N)}\) test.

Columns 7-8 (resp. 12-13) 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. 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 $D = \{ (x_{it}, f_{1t}, f_{2t}) : i = 1, \ldots, n, t = 1, \ldots, T \}$. Let $E_D(\cdot)$ and $\text{Var}_D(\cdot)$ denote the conditional expectation and variance given $D$, respectively. Let $\sum_{1 \leq i < j \leq n} = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n}$. Recall $y_i \equiv (y_{i1}, y_{i2}, \ldots, y_{iT})'$, $\varepsilon_i \equiv (\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iT})'$, $\mathbb{E}_i \equiv (p'_i m_b p_i)^{-} p'_i m_b \varepsilon_i$, and $\mathcal{P}_w \equiv \int p^K(x) p^K(x)' w(x) dx$.

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

$$ y_i = p_i \alpha g_i + f_1 \gamma_{1i} + f_2 \gamma_{2i} + \varepsilon_i + (g_i - p_i \alpha g_i). \quad (A.1) $$

Therefore by (2.9), $\widehat{\alpha}_g = \alpha_g + \bar{\varepsilon}_i + \bar{r}_i$, where $\bar{\varepsilon}_i \equiv (p'_i m_b p_i)^{-} p'_i m_b \varepsilon_i$, $\bar{r}_i \equiv (p'_i m_b p_i)^{-} p'_i m_b r_i$, and $r_i \equiv f_2 \gamma_{2i} + (g_i - p_i \alpha g_i)$. Then by (2.10)-(2.11), we have

$$ c_{nT} \hat{\Gamma}_{nT} = c_{nT} \sum_{1 \leq i < j \leq n} (\widehat{\alpha}_g - \alpha_g)' \mathcal{P}_w (\widehat{\alpha}_g - \alpha_g) $$

$$ = c_{nT} \sum_{1 \leq i < j \leq n} \left\{ (\alpha_{g_i} - \alpha_{g_j}) + (\bar{\varepsilon}_i - \bar{\varepsilon}_j) + (\bar{r}_i - \bar{r}_j) \right\}' \mathcal{P}_w \left\{ (\alpha_{g_i} - \alpha_{g_j}) + (\bar{\varepsilon}_i - \bar{\varepsilon}_j) + (\bar{r}_i - \bar{r}_j) \right\} $$

$$ = A_1 + A_2 + A_3 + A_4 + A_5 + A_6, \quad (A.2) $$

where the $A$'s are defined as follows:

$$ A_1 \equiv c_{nT} \sum_{1 \leq i < j \leq n} (\alpha_{g_i} - \alpha_{g_j})' \mathcal{P}_w (\alpha_{g_i} - \alpha_{g_j}) $$

$$ A_2 \equiv c_{nT} \sum_{1 \leq i < j \leq n} (\bar{\varepsilon}_i - \bar{\varepsilon}_j)' \mathcal{P}_w (\bar{\varepsilon}_i - \bar{\varepsilon}_j), $$

$$ A_3 \equiv c_{nT} \sum_{1 \leq i < j \leq n} (\bar{r}_i - \bar{r}_j)' \mathcal{P}_w (\bar{r}_i - \bar{r}_j), $$

$$ A_4 \equiv 2 c_{nT} \sum_{1 \leq i < j \leq n} (\alpha_{g_i} - \alpha_{g_j})' \mathcal{P}_w (\bar{\varepsilon}_i - \bar{\varepsilon}_j), $$

$$ A_5 \equiv 2 c_{nT} \sum_{1 \leq i < j \leq n} (\alpha_{g_i} - \alpha_{g_j})' \mathcal{P}_w (\bar{r}_i - \bar{r}_j), $$

$$ A_6 \equiv 2 c_{nT} \sum_{1 \leq i < j \leq n} (\bar{\varepsilon}_i - \bar{\varepsilon}_j)' \mathcal{P}_w (\bar{r}_i - \bar{r}_j). \quad (A.3) $$

Under $H_0$, $A_1 = 0$ for $l = 1, 4$, and 5. So it suffices to show that

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

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

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

$$ V_{nT} = \frac{2}{n} \sum_{i=1}^{n} \sigma_i^4 \text{tr} \left( \left((p'_i m_b p_i T)^{-} \mathcal{P}_w \right)^2 \right) \geq \frac{2 \lambda_{\text{min}}(\mathcal{P}_w)}{n} \sum_{i=1}^{n} [\lambda_{\text{max}}(p'_i m_b p_i T)]^{-2} \sigma_i^4 \text{tr}(\mathcal{P}_w) > c > 0. \quad (A.6) $$

17
By (A.6), it suffices to prove (A.4) by first establishing that

\[
A_2 = c_{nT} \sum_{1 \leq i < j \leq n} (\varepsilon_i - \varepsilon_j)^P w(\varepsilon_i - \varepsilon_j) + o_p(1) \equiv A_2'(1) + o_p\left(V_{nT}^{1/2}\right) \tag{A.7}
\]

and then showing that

\[
\frac{A_2 - B_{nT}}{\sqrt{V_{nT}}} \xrightarrow{d} N(0,1),
\]

where \(A_2 = c_{nT} \sum_{1 \leq i < j \leq n} (\varepsilon_i - \varepsilon_j)^P w(\varepsilon_i - \varepsilon_j)\).

To prove (A.7), write

\[
A_2 - A_2' = \frac{c_{nT}}{2} \sum_{1 \leq i < j \leq n} \left\{ (\varepsilon_i - \varepsilon_j)^P w(\varepsilon_i - \varepsilon_j) + (\varepsilon_j - \varepsilon_i)^P w(\varepsilon_j - \varepsilon_i) \right\}
\]

\[
- c_{nT} \sum_{1 \leq i < j \leq n} (\varepsilon_i - \varepsilon_j)^P w(\varepsilon_j - \varepsilon_i)
\]

\[
+ 2c_{nT} \sum_{1 \leq i < j \leq n} (\varepsilon_i - \varepsilon_j)^P w(\varepsilon_i - \varepsilon_j)
\]

\[
= A_{21} - 2A_{22} + A_{23}.
\]

By Lemma B.4(i) and the Cauchy-Schwarz inequality, \(|A_{22}| \leq A_{21} \leq n c_{nT} \sum_{i=1}^n (\varepsilon_i - \varepsilon_i)^P w(\varepsilon_i - \varepsilon_i) = O_p(K^2/n^{1/2})\). By Lemma B.4(ii), \(A_{23} = o_p(K/T^{1/2}) = o_p(K^{1/2})\). Then (A.7) follows by noticing that \(V_{nT}^{1/2} = O_p(K^{1/2})\) and \(K^3/n = o(1)\) under Assumption 5.

To prove (A.8), let \(\varphi(\varepsilon_i, \varepsilon_j) = 1(\varepsilon_i - \varepsilon_j)^P w(\varepsilon_i - \varepsilon_j)\). Then we can write \(A_2 = \sqrt{n}U_{nT}/2\), where

\[
U_{nT} = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \varphi(\varepsilon_i, \varepsilon_j)
\]

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 \(\theta \equiv \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \theta_{ij}, \theta_{ij} \equiv E_iE_j[\varphi(\varepsilon_i, \varepsilon_j)]\),

\[
H_{nT}^{(1)} = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \left\{ E_i[\varphi(\varepsilon_i, \varepsilon_j)] + E_j[\varphi(\varepsilon_i, \varepsilon_j)] - 2\theta_{ij} \right\}
\]

\[
H_{nT}^{(2)} = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \left\{ \varphi(\varepsilon_i, \varepsilon_j) - E_j[\varphi(\varepsilon_i, \varepsilon_j)] - E_i[\varphi(\varepsilon_i, \varepsilon_j)] + \theta_{ij} \right\},
\]

and \(E_j[\cdot]\) denotes expectation taken with respect to \(\varepsilon_j\) conditional on \(\mathcal{D}\). Straightforward calculations show that \(\theta_{ij} = T[\mathbb{E}(\varepsilon_j^P w_i) + \mathbb{E}(\varepsilon_i^P w_\varepsilon)]\) and hence \(\theta = \frac{2n}{n} \sum_{i=1}^n \mathbb{E}(\varepsilon_i^P w_i \varepsilon_i)\). Noting that conditional on \(\mathcal{D}\), \(\varepsilon_i\) 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 \( H_{nT}^{(2)} = o_p (n^{-1/2}) \). It follows that

\[
\bar{A}_2 - \frac{T}{\sqrt{n}} \sum_{i=1}^{n} E_D (\bar{\varepsilon}_i \bar{p}_w \bar{\varepsilon}_i) = \frac{1}{\sqrt{n} (n-1)} \sum_{1 \leq i < j \leq n} \{ E_i [\varphi (\bar{\varepsilon}_i, \bar{\varepsilon}_j)] + E_j [\varphi (\bar{\varepsilon}_i, \bar{\varepsilon}_j)] - 2 \theta_{ij} \} + o_p (1) \]

\[
= \frac{T}{\sqrt{n}} \sum_{i=1}^{n} \{ \bar{\varepsilon}_i \bar{p}_w \bar{\varepsilon}_i - E_D (\bar{\varepsilon}_i \bar{p}_w \bar{\varepsilon}_i) \} + o_p (1) .
\]

Noting that \( \frac{T}{\sqrt{n}} \sum_{i=1}^{n} E_D (\bar{\varepsilon}_i \bar{p}_w \bar{\varepsilon}_i) = B_{nT} \), we have

\[
\frac{\bar{A}_2 - B_{nT}}{\sqrt{n} \sqrt{\text{Var}(A_{nT}, \bar{p}_w \bar{\varepsilon}_i)}} = \frac{T (V_{nT})^{1/2}}{\sqrt{n}} \sum_{i=1}^{n} \{ \bar{\varepsilon}_i \bar{p}_w \bar{\varepsilon}_i - E_D (\bar{\varepsilon}_i \bar{p}_w \bar{\varepsilon}_i) \} + o_p (1) .
\]

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

\[
\text{Var}_D \left( \frac{T (V_{nT})^{1/2}}{\sqrt{n}} \sum_{i=1}^{n} \bar{\varepsilon}_i \bar{p}_w \bar{\varepsilon}_i \right) = 1 + o_p (1) \hspace{1cm} (A.9)
\]

and verifying the Laplounov condition for the central limit theorem. Let \( \mu_{4i} \equiv E (\varepsilon_i^4) \) and \( \bar{A}_{ib} \equiv m \bar{p}_i (p_i' m b) - p_i' m b \). Standard variance calculations show that

\[
\text{Var}_D \left( \frac{T}{\sqrt{n}} \sum_{i=1}^{n} \bar{\varepsilon}_i \bar{p}_w \bar{\varepsilon}_i \right) = \frac{2}{n} \sum_{i=1}^{n} \sigma_i^4 \left( \left( p_i' m b \bar{t}_i / T \right) - p_i' m b \bar{t}_i \right)^2 + \frac{1}{n} \sum_{i=1}^{n} (\mu_{4i} - 3 \sigma_i^4) T^2 \text{tr} (\bar{A}_{ib} \text{diag} (\bar{A}_{ib})) .
\]

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

\[
\sum_{s=1}^{T} \| p_{is} m_{b, ts} \| = \sum_{s=1}^{T} \| p_{is} \left( \mathbf{1} (s = t) - b_t (b'b)^{-1} b_s \right) \| \leq \| p_{it} \| + T^{-1} \sum_{s=1}^{T} \| b_t (b'b)^{-1} b_s p_{is} \| \leq \| p_{it} \| + \alpha_T \| b_t \| ,
\]

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

\[
\pi_{ib, tt} = \sum_{s=1}^{T} \| p_{is} m_{b, ts} \| = \sum_{s=1}^{T} \| p_{is} \left( \mathbf{1} (s = t) - b_t (b'b)^{-1} b_s \right) \| \leq \| p_{it} \| + T^{-1} \sum_{s=1}^{T} \| b_t (b'b)^{-1} b_s p_{is} \| \leq \| p_{it} \| + \alpha_T \| b_t \| ,
\]

and \( T^2 \text{tr} (\bar{A}_{ib} \text{diag} (\bar{A}_{ib})) = T^2 \sum_{i=1}^{n} \pi_{ib, tt} = T^2 \left( \sum_{i=1}^{n} \pi_{ib, tt} \right) = T^2 \left( \sum_{i=1}^{n} \pi_{ib, tt} \right) \leq T^2 \sum_{i=1}^{n} \| \bar{A}_{ib} \| + T^2 \| b_t \| \| p_{it} \| .
\]

Thus \( \text{Var}_D \left( \frac{T}{\sqrt{n}} \sum_{i=1}^{n} \bar{\varepsilon}_i \bar{p}_w \bar{\varepsilon}_i \right) = V_{nT} + o_p (1) \). This, together with \((A.6)\), implies \((A.9)\). Now, let

\[
\xi_{it} = \frac{T \left( \bar{\varepsilon}_i \bar{p}_w \bar{\varepsilon}_i - E_D (\bar{\varepsilon}_i \bar{p}_w \bar{\varepsilon}_i) \right) \text{tr}(\bar{p}_w)}{\text{tr}(\bar{p}_w)} = \frac{T \left[ \bar{\varepsilon}_i \bar{A}_{ib} \bar{\varepsilon}_i - E_D (\bar{\varepsilon}_i \bar{A}_{ib} \bar{\varepsilon}_i) \right]}{\text{tr}(\bar{p}_w)} .
\]
Then by the $C_r$ and Jensen inequalities and Theorem 2 of Bao and Ullah (2010),

\[
E_D \left[ \xi_{iT}^4 \right] \leq 16\sigma_w^4 T^4 |\text{tr}\mathbf{P}_w|^{-4} \left\{ (\text{tr}\mathbf{A}_{ib})^4 + 12 (\text{tr}\mathbf{A}_{ib})^2 \text{tr}(\mathbf{A}_{ib})^2 + 12 (\text{tr}\mathbf{A}_{ib})^2 \right\}
+ 32 (\text{tr}\mathbf{A}_{ib}) \text{tr}(\mathbf{A}_{ib})^3 + 48\text{tr}(\mathbf{A}_{ib}) + \text{remainder terms},
\]

(A.10)

where the expression of the remainder terms (which vanish if $\varepsilon_{it}$ 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 $\text{tr}(\mathbf{A}_{ib}) = \text{tr}(\mathbf{p}_i\mathbf{p}_i^T - \mathbf{P}_w) \leq T^{-1} \text{tr}(\mathbf{A}_{ib})$ and that all elements of $\mathbf{A}_{ib}$ are of order $O_p(K^{3/2}T^{-2})$.

On the other hand, \( \sum_{i=1}^n E_D [\xi_{iT}^4] = \frac{2}{n} V_{nT}^2 / (|\text{tr}\mathbf{P}_w|^2) \). Then by (A.6) and Assumption 5, we have

\[
\sum_{i=1}^n E_D [\xi_{iT}^4] = \sum_{i=1}^n \frac{O_p(n)}{[\frac{2}{n} V_{nT}^2 / (|\text{tr}\mathbf{P}_w|^2)]^2} = O_p \left( \frac{(\text{tr}\mathbf{P}_w)^4}{n V_{nT}^2} \right) = O_p \left( \frac{K^2}{n} \right) = o_p(1).
\]

This verifies the Liapounov condition.

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

\[
A_3 - \overline{A}_3 = \frac{c_{nT}}{2} \sum_{1 \leq i \neq j \leq n} \left\{ (\overline{r}_i - \overline{r}_j)^\prime \mathbf{P}_w (\overline{r}_i - \overline{r}_j) + (\overline{r}_j - \overline{r}_j)^\prime \mathbf{P}_w (\overline{r}_j - \overline{r}_j) \right\}
- c_{nT} \sum_{1 \leq i \neq j \leq n} (\overline{r}_i - \overline{r}_i)^\prime \mathbf{P}_w (\overline{r}_i - \overline{r}_j)
+ 2c_{nT} \sum_{1 \leq i \neq j \leq n} (\overline{r}_i - \overline{r}_i)^\prime \mathbf{P}_w (\overline{r}_i - \overline{r}_j)
\equiv A_{31} - A_{32} + A_{33}.
\]

By Lemma B.4(iii) and the Cauchy-Schwarz inequality, $|A_{32}| \leq n c_{nT} \sum_{i=1}^n (\overline{r}_i - \overline{r}_i)^\prime \mathbf{P}_w (\overline{r}_i - \overline{r}_i) = o_p(\text{tr}(\mathbf{P}_w) + K T/n^{1/2}) = o_p(K^{1/2})$ under Assumption 5. By Lemma B.4(iv), $A_{33} = O_p(\text{tr}(\mathbf{P}_w) + K T/n^{1/2}) = o_p(K^{1/2})$. It follows that $A_3 = o_p(V_{nT}^{1/2})$.

Now, we decompose $A_6$ as follows

\[
A_6 = 2c_{nT} \sum_{1 \leq i < j \leq n} \left\{ (\overline{r}_i - \overline{r}_i - \overline{r}_j + \overline{r}_j)^\prime \mathbf{P}_w (\overline{r}_i - \overline{r}_i - \overline{r}_j + \overline{r}_j)
+ (\overline{r}_i - \overline{r}_i - \overline{r}_j + \overline{r}_j)^\prime \mathbf{P}_w (\overline{r}_i - \overline{r}_i - \overline{r}_j + \overline{r}_j)
+ (\overline{r}_i - \overline{r}_i - \overline{r}_j + \overline{r}_j)^\prime \mathbf{P}_w (\overline{r}_i - \overline{r}_i - \overline{r}_j + \overline{r}_j)ight\}
\equiv A_{61} + A_{62} + A_{63} + A_{64}.
\]

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

\[
\left| A_{61} \right| \leq 8 c_{nT} \sum_{i=1}^n (\overline{r}_i - \overline{r}_i)^\prime \mathbf{P}_w (\overline{r}_i - \overline{r}_i)
\leq 8 \left\{ c_{nT} \sum_{i=1}^n (\overline{r}_i - \overline{r}_i)^\prime \mathbf{P}_w (\overline{r}_i - \overline{r}_i) \right\}^{1/2} \left\{ c_{nT} \sum_{i=1}^n (\overline{r}_i - \overline{r}_i)^\prime \mathbf{P}_w (\overline{r}_i - \overline{r}_i) \right\}^{1/2}
= \left\{ O_p(K^{1/2}) \right\}^{1/2} = o_p(K^{1/2}).
\]

20
Similarly, by Lemmas B.4(i) and (iv), we have $|A_{62}| = \{O_p \left( \frac{K^2}{n^{1/2}} \right) O_p \left( T n^{1/2} K^{1-2 \lambda/4} \right) \}^{1/2} = o_p \left( K^{1/2} \right)$. 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 $V_{nT} \geq \text{Ctr}(\overline{\mathbf{p}}_w) = O(K)$ for some $C > 0$ by (A.6), it suffices to prove the corollary by showing that

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

First, we write

$$\hat{V}_{nT} - V_{nT} = \frac{2}{n} \sum_{i=1}^{n} \hat{\sigma}_i^4 \text{tr} \left( \left( (p'_i m_{ih} p_i / T - \overline{\mathbf{p}}_w) \right)^2 \right) - \frac{2}{n} \sum_{i=1}^{n} \sigma_i^4 \text{tr} \left( \left( (p'_i m_{ih} p_i / T - \overline{\mathbf{p}}_w) \right)^2 \right)$$

$$= \frac{2}{n} \sum_{i=1}^{n} \left( \hat{\sigma}_i^4 - \sigma_i^4 \right) \text{tr} \left( \left( (p'_i m_{ih} p_i / T - \overline{\mathbf{p}}_w) \right)^2 \right)$$

$$+ \frac{2}{n} \sum_{i=1}^{n} \sigma_i^4 \text{tr} \left( \left( (p'_i m_{ih} p_i / T - \overline{\mathbf{p}}_w) \right)^2 \right) - \left( (p'_i m_{ih} p_i / T - \overline{\mathbf{p}}_w) \right)^2 \right)$$

$$= 2V_{1nT} + 2V_{2nT} .$$

We want to show $V_{snT} = o_p (K)$ for $s = 1, 2$. Let $\overline{m}_{ih} \equiv m_{ih} - m_{ih} (p'_i m_{ih} p_i / T - p'_i m_{ih})$ and $\overline{m}_{ib} \equiv m_{ib} - m_{ib} (p'_i m_{ib} p_i / T - p'_i m_{ib})$. Then $\hat{\varepsilon}_i = m_{ih} (y_i - \overline{\mathbf{g}}_i) = \overline{m}_{ih} (\varepsilon_i + r_i)$ where recall $r_i \equiv f_2 \gamma_{2i} + (\mathbf{g}_i - p_i \alpha g_i)$. It follows that

$$\hat{\sigma}_i^2 = \frac{1}{T} \hat{\varepsilon}_i^2 \hat{\varepsilon}_i = \frac{1}{T} \varepsilon_i^2 \overline{m}_{ih} \varepsilon_i + \frac{1}{T} r'_i \overline{m}_{ih} r_i + \frac{2}{T} r'_i \overline{m}_{ih} \varepsilon_i , \quad (A.11)$$

and

$$\frac{1}{n} \sum_{i=1}^{n} \left( \hat{\sigma}_i^2 - \sigma_i^2 \right)^2 \leq \frac{3}{n} \sum_{i=1}^{n} \left( \frac{1}{T} \varepsilon_i^2 \overline{m}_{ih} \varepsilon_i - \sigma_i^2 \right)^2 + \frac{3}{n} \sum_{i=1}^{n} \left( \frac{1}{T} r'_i \overline{m}_{ih} r_i \right)^2 + \frac{12}{n} \sum_{i=1}^{n} \left( \frac{1}{T} r'_i \overline{m}_{ih} \varepsilon_i \right)^2$$

$$= 3\delta_{1nT} + 3\delta_{2nT} + 12\delta_{3nT} .$$

For $\delta_{1nT}$, we can use the decomposition in (B.1) as in the proof of Lemma B.4(ii) to show that $\delta_{1nT} = a_{1nT} + O_p (K / T)$, where $a_{1nT} = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{1}{T} \overline{m}_{ih} \varepsilon_i - \sigma_i^2 \right)^2$. Let $\overline{m}_{ib,t,s}$ denote the $(t, s)$ element of $\overline{m}_{ib}$. Noting that $\text{tr}(\overline{m}_{ib}) = T - K - q_1 - q_2$, and $\frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} (\overline{m}_{ib,t,s})^2 = 1 + O_p (K / T)$ uniformly in $i$, we have

$$E_D \left( \hat{\sigma}_{1nT} \right)$$

$$= \frac{1}{n} \sum_{i=1}^{n} E_D \left( \frac{1}{T} \varepsilon_i^2 \overline{m}_{ih} \varepsilon_i - \frac{2}{T} \varepsilon_i \overline{m}_{ih} \varepsilon_i \sigma_i^2 + \sigma_i^4 \right)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left[ \frac{1}{T^2} \left( \sum_{t=1}^{T} \sum_{s=1}^{T} (\overline{m}_{ib,t,s})^2 \sigma_i^4 + \sum_{t=1}^{T} (\overline{m}_{ib,t})^2 \left[ E (\varepsilon_i^4) - 3 \sigma_i^4 \right] - 2 \frac{\text{tr} (\overline{m}_{ib})}{T} \sigma_i^4 + \sigma_i^4 \right) \right]$$

$$= O_p (K / T) .$$

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

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

21
It follows that $\delta_{2nT} \leq \frac{n^{1/2}}{\sqrt{n}} \sum_{i=1}^{n} \|(m_i - b_i) f_2 \gamma_2 \|^2 + \frac{n^{1/2}}{\sqrt{n}} \sum_{i=1}^{n} \|(m_i - b_i) f_2 \gamma_2 \|^2 = O_p(n^{-1} + K^{-4\lambda/d})$

where $n^{-1}T^{-2} \sum_{i=1}^{n} \|(m_i - b_i) f_2 \gamma_2 \|^2 = O_p(n^{-1})$ can be proved analogously to the proof of Lemma A5(v) of Su and Jin (2011), and $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \|(g_i - p_i \alpha_g)\|^4 = O_p(K^{-4\lambda/d})$ by arguments similar to the proof of Lemma A.2 of Su and Jin (2011) under the strengthened condition given in the corollary. For $\delta_{3nT}$, we can show that $\delta_{3nT} = \delta_{3nT} + O_p(K/T)$, where $\delta_{3nT} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{1}{T} r_i' \bar{m}_b \varepsilon_i \right)^2$. Then noting that $\bar{m}_b r_i = \bar{m}_b (g_i - p_i \alpha_g)$ and $\bar{m}_b$ is a projection matrix, by Lemma B.2(ii) we have

$$E_D(\tilde{\delta}_{3nT}) = \frac{1}{nT^2} \sum_{i=1}^{n} r_i' \bar{m}_b E(\varepsilon_i' \varepsilon_i) \bar{m}_b r_i \leq \frac{C}{nT^2} \sum_{i=1}^{n} \|g_i - p_i \alpha_g\|^2 = O_p(T^{-1} K^{-2\lambda/d}).$$

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

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

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

$$V_{nT} = \frac{1}{n} \sum_{i=1}^{n} \left(\hat{\sigma}^2 - \sigma^2 \right) \left(\hat{\sigma}^2 + \sigma^2 \right) \text{tr} \left(\left(\left[p_i' m_k p_i / T \right] - \mathcal{P}_w \right)^2 \right)$$

$$\leq (\sigma_{2\lambda})^{-2} \lambda_{\max} (\mathcal{P}_w) \text{tr} (\mathcal{P}_w) \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(\hat{\sigma}^2 - \sigma^2 \right)^2 \right\}^{1/2} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(\hat{\sigma}^2 + \sigma^2 \right)^2 \right\}^{1/2}$$

$$= O(K) O_p(\sqrt{K/T}) O_p(1) = o_p(K).$$

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

$$\hat{B}_{nT} = B_{nT}$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\sigma}^2_i \text{tr} (\left[p_i' m_k p_i / T \right] - \mathcal{P}_w) + \hat{\delta}_{nT} - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sigma^2_i \text{tr} (\left[p_i' m_k p_i / T \right] - \mathcal{P}_w) \equiv B_1 + B_2,$$

where $B_1 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\hat{\sigma}^2_i - \sigma^2_i) \text{tr} (\left[p_i' m_k p_i / T \right] - \mathcal{P}_w)$, and $B_2 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sigma^2_i \text{tr} (\left[p_i' m_k p_i / T \right] - \mathcal{P}_w) + o_p(1) = \overline{B}_1 + o_p(1)$, where $\overline{B}_1 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\frac{1}{T} r_i' \bar{m}_b \varepsilon_i - \sigma^2_i) \text{tr} (\left[p_i' m_k p_i / T \right] - \mathcal{P}_w)$. Simple calculations reveal that $E_D(\overline{B}_1) = 0$ and

$$\text{Var}_D(\overline{B}_1) = \frac{1}{\sqrt{n}} \left(\frac{1}{T-q_1 - d - 1} \right) \sum_{i=1}^{n} \text{Var}_D(\varepsilon_i' \bar{m}_b \varepsilon_i) \left\{ \text{tr} (\left[p_i' m_k p_i / T \right] - \mathcal{P}_w) \right\}^2$$

$$= O_p(K^2/T) = o_p(1)$$

as $\text{Var}_D(\varepsilon_i' \bar{m}_b \varepsilon_i) \leq \text{Var}(\varepsilon_i' \varepsilon_i) = O(T)$. Hence $B_1 = o_p(1)$. 

22
Now we show that $B_2 = o_p \left( K^{1/2} \right)$. Note that

$$B_2 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sigma_i^2 \frac{1}{T} \text{tr} \left( (p_i' m_b p_i) - (p_i' m_b p_i - p_i' m_b p_i) (p_i' m_b p_i)^{-1} P_w \right) + \tilde{b}_{inT}$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sigma_i^2 \frac{1}{T} \text{tr} \left( \left( p_i' (m_b - m_h) \right) p_i + \tilde{b}_{inT} \right) \beta_i$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sigma_i^2 \frac{1}{T} \text{tr} \left( \left( p_i' (m_b - m_h) \right) p_i + \tilde{b}_{inT} \right) \beta_i$$

$$+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left( \sigma_i^2 - \sigma_i^2 \right) T^{-1} \text{tr} \left( \left( p_i' (m_b - m_h) \right) p_i + \tilde{b}_{inT} \right) \beta_i$$

$$= B_{21} + B_{22},$$

where $\beta_i \equiv (p_i' m_b p_i / T)^{-1} P_w (p_i' m_b p_i / T)^{-1}$. 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 $\hat{\sigma}_i^2$ in (A.11). Let $\sigma \equiv (\sigma_1, \ldots, \sigma_T)'$, $\sigma \equiv (\sigma_1, \ldots, \sigma_T)'$, $g \equiv (g_1, \ldots, g_T)'$, and $\tilde{g} \equiv (\tilde{g}_1, \ldots, \tilde{g}_T)'$. Then

$$B_{21} = \frac{1}{n^{1/2} T} \sum_{i=1}^{n} \sigma_i^2 \text{tr} \left( \left( p_i' (m_b - m_h) \right) p_i + \tilde{b}_{inT} \right) \beta_i$$

$$= \frac{1}{n^{1/2} T} \sum_{i=1}^{n} \sigma_i^2 \text{tr} \left\{ p_i' \left( h (h')^{-1} h - b (b')^{-1} b \right) \right\} p_i \beta_i$$

$$+ p_i' \left( \frac{1}{\sigma_i^2} \right) \left( h (h')^{-1} h - b (b')^{-1} b \right) p_i \beta_i$$

$$= \frac{1}{n^{1/2} T} \sum_{i=1}^{n} \sigma_i^2 \text{tr} \left\{ p_i' \left( h (h')^{-1} h - b (b')^{-1} b \right) \right\} p_i \beta_i$$

$$+ \frac{1}{n^{1/2} T} \sum_{i=1}^{n} \sigma_i^2 \text{tr} \left\{ p_i' \left( h (h')^{-1} h - b (b')^{-1} b \right) \right\} p_i \beta_i$$

$$= B_{21,1} + B_{21,2} + B_{21,3} + O_p(K/T^{1/2}),$$

where the $O_p(K/T^{1/2})$ 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 $B_{21,1} = o_p(K)$ as we 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 (2011), $T^{-1} \sum_{i=1}^{n} (||p_i' \sigma|| + ||p_i' \sigma||) = O_p(\sqrt{K} + \sqrt{Kn/T})$. It follows that

$$|B_{21}| \leq \max_{1 \leq i \leq n} \frac{\sigma_i^2}{n^{1/2} T} \sum_{i=1}^{n} | \text{tr} \left\{ p_i' \left( \frac{1}{\sigma_i^2} \right) \left( h (h')^{-1} h - b (b')^{-1} b \right) \right\} p_i \beta_i | + O_p(K/n^{1/2} + K/T^{1/2})$$

$$\leq C T_{21,1} + o_p(\sqrt{K}),$$

where $T_{21,1} \equiv \frac{1}{n^{1/2} T} \sum_{i=1}^{n} | \text{tr} \left\{ p_i' \left( \frac{1}{\sigma_i^2} \right) \left( h (h')^{-1} h - b (b')^{-1} b \right) \right\} |$. By Theorem 4.2 of Su and Jin (2011), we can show that $\tilde{y}_i - \bar{y}_i = \frac{1}{n} \sum_{j=1}^{n} [\tilde{y}_j (x_{ji}) - \bar{y}_j (x_{ji})] = O_p(\sqrt{K/nT})$. By the Cauchy-Schwarz inequality and
the fact that $\|\beta_i\| = O_p(\sqrt{K})$ uniformly in $i$, we have

$$
\overline{B}_{21,1} \leq \frac{1}{n^{1/2}T} \sum_{i=1}^{n} \left\{ \text{tr} \left( p_i' \left( \tilde{\gamma} - \bar{\gamma} \right) (h' h)^{-1} \left( \tilde{\gamma} - \bar{\gamma} \right)' p_i \right) \right\}^{1/2} \left\{ \text{tr} \left( \beta_i' p_i' (h' h)^{-1} h' p_i \beta_i \right) \right\}^{1/2}
$$

$$
\leq \frac{1}{n^{1/2}T} \sum_{i=1}^{n} \left\{ \text{tr} \left( p_i' \left( \tilde{\gamma} - \bar{\gamma} \right) (h' h)^{-1} \left( \tilde{\gamma} - \bar{\gamma} \right)' p_i \right) \right\}^{1/2} \left\{ \text{tr} (\beta_i' p_i \beta_i) \right\}^{1/2}
$$

$$
\leq \left[ \lambda_{\min} (h' h/T) \right]^{-1} \max_{1 \leq i \leq n} \lambda_{\max} (p'_i p_i / T) \frac{1}{n^{1/2}T} \sum_{i=1}^{n} \left\| p_i' \left( \tilde{\gamma} - \bar{\gamma} \right) \right\| \|\beta_i\|
$$

$$
= O_p \left( \sqrt{K} \right) \left\| \tilde{\gamma} - \bar{\gamma} \right\| \left\{ \frac{1}{n^{1/2}T} \sum_{i=1}^{n} \|p_i\| \right\} = O_p \left( \sqrt{K} \right) O_p \left( \sqrt{K/n} \right) O_p \left( \sqrt{nK/T} \right)
$$

$$
= O_p \left( \sqrt{K^3/T} \right) = o_p \left( \sqrt{K} \right). \quad \text{This completes the proof.} \quad \blacksquare
$$

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 $H_1 \left( n^{-1/4}T^{-1/2} \text{tr} \left( \overline{B}_w \right)^{1/4} \right)$. Since the proof in Theorem 3.2 does not impose the null hypothesis, by (A.6) it suffices to show that under $H_1 \left( n^{-1/4}T^{-1/2} \text{tr} \left( \overline{B}_w \right)^{1/4} \right)$,

$$
A_1 / \sqrt{V_{nT}} \xrightarrow{P} \Delta, \quad \text{(A.12)}
$$

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

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

$$
A_1 = c_{nT} \sum_{1 \leq i < j \leq n} \int \left( p_i^K (x)' (\alpha_{g_i} - \alpha_{g_j}) \right)^2 w(x) dx
$$

$$
A_4 = c_{nT} \sum_{1 \leq i < j \leq n} \int \left\{ \gamma_{nT} \Delta_{ij,n}(x) + [d_{g_i}(x) - d_{g_j}(x)] \right\}^2 w(x) dx
$$

$$
A_5 = A_{11} + A_{12} + A_{13} - A_{14} - A_{15},
$$

where $d_{g_i}(x) \equiv g_i(x) - p^K(x)' \alpha_{g_i}$, $A_{11} \equiv c_{nT} \gamma_{nT} \sum_{1 \leq i < j \leq n} \int \Delta_{ij,n}(x) w(x) dx$, $A_{12} \equiv \frac{\sqrt{n}}{n \epsilon} \sum_{i=1}^{n} \int d_{g_i}(x) w(x) dx$, $A_{13} = \gamma_{nT} c_{nT} \sum_{1 \leq i < j \leq n} \int \Delta_{ij,n}(x) d_{g_i}(x) w(x) dx$, $A_{14} = \gamma_{nT} c_{nT} \sum_{1 \leq i < j \leq n} \int \Delta_{ij,n}(x) d_{g_j}(x) w(x) dx$, and $A_{15} = 2 \gamma_{nT} \sum_{1 \leq i < j \leq n} \int \Delta_{ij,n}(x) w(x) dx$.

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

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

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

Now write

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

where $A_{41} = \frac{\sqrt{n}}{n \epsilon} \sum_{i=1}^{n} \int d_{g_i}(x) w(x) dx$ and $A_{42} = c_{nT} \sum_{1 \leq i < j \leq n} \int \Delta_{ij,n}(x) d_{g_j}(x) w(x) dx$.

24
where \( A_{41} \equiv 2 c_n T \gamma_{nT} \sum_{1 \leq i < j \leq n} \int \Delta_{ij,n}(x) p^K(x)'(\bar{e}_i - \bar{e}_j) w(x) \, dx \), and \( A_{42} \equiv 2 c_n T \sum_{1 \leq i < j \leq n} \int (d_{gi} - d_{gj}) p^K(x)'(\bar{e}_i - \bar{e}_j) w(x) \, dx \). Let \( \bar{\gamma}_{nT} \equiv \gamma_{nT} / \text{tr}(\bar{F}_w)^{1/2} = n^{-1/4}T^{-1/2}\text{tr}(\bar{F}_w)^{-1/4} \). Analogously to the proof of Lemma B.4(i) by replacing \( \bar{e}_i \) with \( \bar{\gamma}_i \), we can show that

\[
\begin{align*}
A_{41}/\text{tr}(\bar{F}_w)^{1/2} &= c_n T \bar{\gamma}_{nT} \sum_{i=1}^{n} \int \Delta_{in}(x) \, p^K(x)' \bar{e}_i \, w(x) \, dx \\
&= 2 n c_n T \bar{\gamma}_{nT} \sum_{i=1}^{n} \int \Delta_{in}(x) p^K(x)' \bar{e}_i w(x) \, dx - 2 c_n T \bar{\gamma}_{nT} \sum_{i=1}^{n} \sum_{j=1}^{n} \int \Delta_{in}(x) p^K(x)' \bar{e}_j w(x) \, dx \\
&= 2 n c_n T \bar{\gamma}_{nT} \sum_{i=1}^{n} \int \Delta_{in}(x) p^K(x)' \bar{\gamma}_i w(x) \, dx - 2 c_n T \gamma_{nT} \sum_{i=1}^{n} \sum_{j=1}^{n} \int \Delta_{in}(x) p^K(x)' \bar{\gamma}_j w(x) \, dx + n^2 c_n T \gamma_{nT} O_p(K/\sqrt{nT}) \\
&= \overline{A}_{41a} - \overline{A}_{41b} + o_p(1),
\end{align*}
\]

where \( \overline{A}_{41a} \equiv 2 n c_n T \bar{\gamma}_{nT} \sum_{i=1}^{n} \int \Delta_{in}(x) p^K(x)' \bar{\gamma}_i w(x) \, dx \) and \( \overline{A}_{41b} \equiv 2 c_n T \gamma_{nT} \sum_{i=1}^{n} \sum_{j=1}^{n} \int \Delta_{in}(x) p^K(x)' \bar{\gamma}_j w(x) \, dx \). Noting that \( E_D(\overline{A}_{41a}) = 0 \) and

\[
\begin{align*}
\text{Var}_D(\overline{A}_{41a}) &= 4 n^2 c_n^2 T^2 \gamma_{nT} \sum_{i=1}^{n} \sigma_i^2 \int \Delta_{in}(x) p^K(x)' w(x) \, dx (p_i'm_{bi})^{-1} \int p^K(\bar{\gamma}_i) \Delta_{in}(\bar{\gamma}) \, w(\bar{\gamma}) \, d\bar{\gamma} \\
&\leq 4 [\lambda_{\min}(T^{-1}p_i'm_{bi})]^{-1} T^{-1} n^2 c_n^2 T^2 \gamma_{nT} \sum_{i=1}^{n} \sigma_i^2 \int \int w(x) \Delta_{in}(x) p^K(x)' p^K(\bar{\gamma}_i) \Delta_{in}(\bar{\gamma}) \, w(\bar{\gamma}) \, dx \, d\bar{\gamma} \\
&= O_p(T^{-1} n^3 c_n^2 T^2 \gamma_{nT} K) = O_p(T^{-1/2} K \text{tr}(\bar{F}_w)^{-1/2}) = o_p(1),
\end{align*}
\]

it follows that \( \overline{A}_{41a} = o_p(1) \). Similarly, we can show that \( \overline{A}_{41b} = o_p(1) \) and thus \( A_{41} = o_p(\text{tr}(\bar{F}_w)^{1/2}) \) .

Analogously,

\[
\begin{align*}
A_{42}/\text{tr}(\bar{F}_w)^{1/2} &= 2 n \bar{\gamma}_{nT} \sum_{i=1}^{n} \int d_{gi}(x) p^K(x)' \bar{\gamma}_i w(x) \, dx - 2 \bar{\gamma}_{nT} \sum_{i=1}^{n} \sum_{j=1}^{n} \int d_{gi}(x) p^K(x)' \bar{\gamma}_j w(x) \, dx \\
&= A_{42a} + A_{42b} + o_p(1)
\end{align*}
\]

where \( \bar{\gamma}_{nT} \equiv c_n T / \text{tr}(\bar{F}_w)^{1/2} \), \( A_{42a} \equiv 2 n \bar{\gamma}_{nT} \sum_{i=1}^{n} \int d_{gi}(x) p^K(x)' \bar{\gamma}_i w(x) \, dx \) and \( A_{42b} \equiv 2 \bar{\gamma}_{nT} \sum_{i=1}^{n} \sum_{j=1}^{n} \int d_{gi}(x) p^K(x)' \bar{\gamma}_j w(x) \, dx \). Noting that \( E_D(\overline{A}_{42a}) = 0 \) and

\[
\begin{align*}
\text{Var}_D(\overline{A}_{42a}) &= 4 n^2 c_n^2 T^2 \gamma_{nT} \sum_{i=1}^{n} \sigma_i^2 \int d_{gi}(x) p^K(x)' w(x) \, dx (p_i'm_{bi})^{-1} \int p^K(\bar{\gamma}_i) d_{gi}(\bar{\gamma}) \, w(\bar{\gamma}) \, d\bar{\gamma} \\
&\leq 4 [\lambda_{\min}(T^{-1}p_i'm_{bi})]^{-1} T^{-1} n^2 c_n^2 T^2 \gamma_{nT} \sum_{i=1}^{n} \sigma_i^2 \int \int w(x) d_{gi}(x) p^K(x)' p^K(\bar{\gamma}_i) d_{gi}(\bar{\gamma}) \, w(\bar{\gamma}) \, dx \, d\bar{\gamma} \\
&\leq 4 [\lambda_{\min}(T^{-1}p_i'm_{bi})]^{-1} T^{-1} n^2 c_n^2 T^2 \gamma_{nT} \sum_{i=1}^{n} \sigma_i^2 \|d_{gi}(\cdot)\|_{\infty, \bar{\gamma}}^2 \int \left( 1 + \|x\|^2 \right)^{\bar{\gamma}_i} w(x)^2 \|p^K(x)\|^2 \, dx \\
&= O_p(T^{-1} n^3 c_n^2 T^{-2\lambda/d} \zeta(K)^2) = O_p(TK^{-2\lambda/d} \zeta(K)^2 \text{tr}(\bar{F}_w)^{-1}) = o_p(1),
\end{align*}
\]
it follows that \( \overline{A}_{42a} = o_p(1) \). Similarly, we can show that \( \overline{A}_{42b} = o_p(1) \) and thus \( A_{42} = o_p(\min(\text{tr}(\overline{P}_w)^{1/2})) \).

Lastly, \( |A_5| = o_p(\min(\text{tr}(\overline{P}_w)^{1/2})) \) 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

\[
V_{nT}^{1/2} n^{-1/2} T^{-1} D_{nT} = V_{nT}^{1/2} n^{-1/2} T^{-1} (c_{nT} \overline{B}_{nT} - \overline{B}_{nT}) / \sqrt{V_{nT}} = n^{-1/2} T^{-1} \left( (c_{nT} \overline{H}_{nT} - B_{nT}) \right) \{ 1 + o_p(1) \} + \overline{B}_{nT} \{ 1 + o_p(1) \} = n^{-1/2} T^{-1} c_{nT} \sum_{1 \leq i < j \leq n} (\alpha_{gi} - \alpha_{gj})^T \overline{P}_w (\alpha_{gi} - \alpha_{gj}) \{ 1 + o_p(1) \} + o_p(1).
\]

Next, \( n^{-1/2} T^{-1} c_{nT} \sum_{1 \leq i < j \leq n} (\alpha_{gi} - \alpha_{gj})^T \overline{P}_w (\alpha_{gi} - \alpha_{gj}) = (n(n-1))^{-1} \sum_{1 \leq i < j \leq n} \{ g_i(x) - g_j(x) \}^2 w(x) \ dx + o(1) \rightarrow \Delta_{g} > 0 \) under \( \mathbb{H}_1 \). The result follows because \( V_{nT} = O_p(\min(\text{tr}(\overline{P}_w))) \) by (A.6).

**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 \( c_{1\lambda} \equiv \min_{1 \leq i \leq n} \{ \lambda_{\min}(T^{-1} p_i^T m_i b_i) \} \) and \( c_{2\lambda} \equiv \min_{1 \leq i \leq n} \{ \lambda_{\min}(T^{-1} p_i^T m_i b_i) \} \).

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

(i) \( \| T^{-1} p_i^T m_i b_i - Q_i \| = O_p(K/\sqrt{T}) \);
(ii) \( \| T^{-1} (p_i^T (m_i - m_\bar{b}) b_i) \| = O_p(K/\sqrt{n}) \);
(iii) \( m_i b_i f_2 = 0 \) w.p.a.1 as \( n \rightarrow \infty \).

**Proof.** (i) and (iii) follow from Lemmas A.1(iii) and A.5(iv) of Su and Jin (2011), respectively. The proof of (ii) is analogous to that of Lemma A.5(vi) of Su and Jin (2011) by using the decomposition for \( m_i - m_\bar{b} \) in (B.1) below. ■

**Lemma B.2** (i) Suppose Assumptions 1(i), (v), (vi), and (viii) hold, then \( n E \| \mathbb{V} \mathbb{V}^T \| = CI_T \) and \( n E \| \mathbb{V} \| \leq CI_T \) for some \( C < \infty \), where \( \mathbb{V} = (\mathbb{v}_1, \mathbb{v}_2, \ldots, \mathbb{v}_T)' \) and \( \mathbb{V} = (\mathbb{v}_1, \mathbb{v}_2, \ldots, \mathbb{v}_T)' \).

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

**Proof.** See Lemmas A.3 and B.2 of Su and Jin (2011). ■

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

(i) \( \sum_{i=1}^n \| T^{-1} p_i^T (m_i - m_\bar{b}) \mathbb{v}_i \|^2 = O_p(K/T) \);
(ii) \( \sum_{i=1}^n \| T^{-1} p_i^T (m_i - m_\bar{b}) p_i m_i \mathbb{v}_i \|^2 = O_p(K^2/T) \);
(iii) \( n^{-1} \sum_{i=1}^n \text{tr} \left( \left( T^{-1} p_i^T m_i b_i \right)^2 - \left( T^{-1} p_i^T m_i b_i \right)^T \overline{P}_w \right) \right) = O_p(K/\sqrt{n}) \);
(iv) \( n^{-1} \sum_{i=1}^n \text{tr} \left( \left( \left( p_i^T m_i b_i / T \right)^2 - \left( p_i^T m_i b_i / T \right)^{-1} \right) \overline{P}_w \right) = O_p(K/\sqrt{n}) \).
Proof. (i) Using the decomposition
\[ m_h - m_b = b (b'b)^{-1} b' - h (h'h)^{-1} h' \]
we have that by the $C_r$ inequality,
\[
\frac{1}{T^2} \sum_{i=1}^{n} \| p'_i (m_h - m_b) \varepsilon_i \|^2 \leq \frac{1}{T^2} \sum_{i=1}^{n} \left\| p'_i (b'b)^{-1} b' \varepsilon_i \right\|^2 + \frac{1}{T^2} \sum_{i=1}^{n} \left\| p'_i h [(b'b)^{-1} - (h'h)^{-1}] b' \varepsilon_i \right\|^2 \\
+ \frac{1}{T^2} \sum_{i=1}^{n} \left\| p'_i h (h'h)^{-1} \varepsilon_i \right\|^2
\]
\[ \equiv D_1 + D_2 + D_3, \text{ say}, \]
where $\varepsilon = (\varepsilon_1', ..., \varepsilon_T') = h - b$. Noting that
\[
T^{-4} E_P \left( \sum_{i=1}^{n} \left\| p_i \right\|^2 \left\| b' \varepsilon_i \right\|^2 \right) = T^{-4} \sum_{i=1}^{n} \left\| p_i \right\|^2 \text{tr} (b'E (\varepsilon_i \varepsilon_i^T) b) \leq \max_{1 \leq i \leq n} \{ \lambda_{\max} (E (\varepsilon_i \varepsilon_i^T)) \} \left\{ T^{-1} \left\| b \right\|^2 \right\} \left\{ T^{-3} \sum_{i=1}^{n} \left\| p_i \right\|^2 \right\}
\]
\[ = O_p (1) O_p (1) O_p (nK/T^2) = O_p (nK/T^2), \] (B.2)
we have
\[
D_1 \leq \left[ \lambda_{\min} (T^{-1} b'b) \right]^{-2} \left\| \varepsilon \right\|^2 \left\{ T^{-4} \sum_{i=1}^{n} \left\| p_i \right\|^2 \left\| b' \varepsilon_i \right\|^2 \right\}
\]
\[ = O_p (1) O_p (T/n) O_p (nK/T^2) = O_p (K/T) \]
Similarly, $D_2 \leq \| h [(T^{-1} b'b)^{-1} - (T^{-1} h'h)^{-1}] \|^2 \left\{ T^{-4} \sum_{i=1}^{n} \left\| p_i \right\|^2 \left\| b' \varepsilon_i \right\|^2 \right\} = O_p (T/n) O_p (nK/T^2) = O_p (K/T)$, and $D_3 \leq \| (h'h)^{-1} h' \| \left\{ T^{-2} \sum_{i=1}^{n} \left\| \varepsilon_i \varepsilon_i^T \right\|^2 \left\| p_i \right\|^2 \right\} = O_p (T^{-1}) O_p (K) = O_p (K/T)$. It follows that $\frac{1}{T^2} \sum_{i=1}^{n} \left\| p'_i (m_h - m_b) \varepsilon_i \right\|^2 = O_p (K/T)$.

For (ii), using the decomposition in (B.1), we have
\[
T^{-2} \sum_{i=1}^{n} \left\| p'_i (m_h - m_b) p_i p'_i m_b \varepsilon_i \right\|^2 \leq T^{-2} \sum_{i=1}^{n} \left\| p'_i (b - h) (b'b)^{-1} b' p_i p'_i m_b \varepsilon_i \right\|^2 + T^{-2} \sum_{i=1}^{n} \left\| p'_i h [(b'b)^{-1} - (h'h)^{-1}] b' p_i p'_i m_b \varepsilon_i \right\|^2 \\
+ T^{-2} \sum_{i=1}^{n} \left\| p'_i h (h'h)^{-1} (b - h)' p_i p'_i m_b \varepsilon_i \right\|^2
\]
\[ \equiv D_4 + D_5 + D_6, \text{ say}. \]
Noting that $E_P (T^{-2} \sum_{i=1}^{n} \left\| p_i \right\|^2 \left\| p'_i m_b \varepsilon_i \right\|^2) = T^{-2} \sum_{i=1}^{n} \left\| p_i \right\|^2 \text{tr} (p_i p'_i m_b E (\varepsilon_i \varepsilon_i^T) m_b p'_i) \leq \max_{1 \leq i \leq n} \lambda_{\max} (E (\varepsilon_i \varepsilon_i^T)) T^{-2} \sum_{i=1}^{n} \left\| p_i \right\|^2 \text{tr} (p'_i m_b p_i) \leq \max_{1 \leq i \leq n} \lambda_{\max} (E (\varepsilon_i \varepsilon_i^T)) \max_{1 \leq i \leq n} \lambda_{\max} (T^{-1} p'_i m_b p_i) T^{-1} \sum_{i=1}^{n} \left\| p_i \right\|^4 = O_p (nK^2/T)$, we have
\[
D_4 \leq \left\| \varepsilon \right\|^2 \left\| (b'b)^{-1} b' \right\|^2 \left\{ T^{-2} \sum_{i=1}^{n} \left\| p_i \right\|^2 \left\| p'_i m_b \varepsilon_i \right\|^2 \right\}
\]
\[ = O_p (T/n) O_p (1/T) O_p (nK^2/T) = O_p (K^2/T). \]
Similarly, we can show that \( D_s = O_p \left( K^2 / T \right) \) for \( s = 5, 6 \). Thus \( T^{-2} \sum_{i=1}^n \| p_i (m_h - m_b) p_i p_i m_b \| ^2 = O_p \left( K^2 / T \right) \).

(iii) Let \( M_{ih} \equiv (T^{-1} p_i m_h p_i) - T^{-1} p_i m_b p_i \), \( M_{ib} \equiv (T^{-1} p_i m_i p_i) - T^{-1} p_i m_b p_i \), and \( D_T = \frac{1}{T} \sum_{i,j=1}^n \text{tr} \left\{ (T^{-1} p_i m_i p_i) - T^{-1} p_i m_b p_i \right\}^2 \). Then by the Hölder inequality,

\[
D_T = \frac{1}{n} \sum_{i=1}^n \text{tr} \left( (M_{ib} + M_{ih}) (M_{ib} - M_{ih}) \right) \\
\leq \frac{1}{n} \sum_{i=1}^n \left\{ \text{tr} \left( (M_{ib} + M_{ih})^2 \right) \right\} \left\{ \text{tr} \left( (M_{ib} - M_{ih})^2 \right) \right\}^{1/2} \\
\leq 2 \lambda_{\text{max}} \left( \mathcal{P}_w \right) \left[ (c_{1\lambda}^{-1} + c_{2\lambda}^{-1}) D_T \right],
\]

where recall \( c_{1\lambda} \equiv \min_{1 \leq i \leq n} [\lambda_{\text{min}}(p_i m_b p_i / T) \) and \( c_{2\lambda} \equiv \min_{1 \leq i \leq n} [\lambda_{\text{min}}(p_i m_b p_i / T) \), and \( D_T = \frac{1}{n} \sum_{i=1}^n \text{tr} \left( (M_{ib} - M_{ih})^2 \right) \). By the Cauchy-Schwarz inequality, \( D_T \leq \overline{D}_T \), where \( \overline{D}_T = \frac{1}{n} \sum_{i=1}^n \text{tr} \left( (M_{ib} - M_{ih})^2 \right) \). Writing \( M_{ib} - M_{ih} = (T^{-1} p_i (m_h - m_b) p_i (T^{-1} p_i m_b p_i) - T^{-1} p_i m_i p_i) - \mathcal{P}_w \) and using the fact that \( \text{tr} \left( (A'B)^2 \right) \leq \text{tr} \left( (A' A) (B'B) \right) \) (e.g., Magnus and Neudecker, 1999, p.201), we have

\[
\overline{D}_T \leq \frac{1}{n} \sum_{i=1}^n \left\{ \left( T^{-1} p_i (m_h - m_b) p_i (T^{-1} p_i m_b p_i) - \mathcal{P}_w \right) \right\}^2 \\
\leq \left[ \lambda_{\text{max}} \left( \mathcal{P}_w \right) \right] \left( c_{1\lambda} c_{2\lambda} \right)^{-1} \frac{1}{n} \sum_{i=1}^n \| T^{-1} p_i (m_h - m_b) p_i (T^{-1} p_i m_b p_i) \|^2 \\
\leq \left[ \lambda_{\text{max}} \left( \mathcal{P}_w \right) \right] \left( c_{1\lambda} c_{2\lambda} \right)^{-1} \frac{1}{n} \sum_{i=1}^n \| T^{-1} p_i (m_h - m_b) p_i \|^2.
\]

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 \( D_T = O_p \left( K / \sqrt{n} \right) \). Finally, (iv) can be proved by using the decomposition in (B.1) and arguments similar to those used above. ■

**Lemma B.4** Recall \( \overline{z_i} \equiv (p_i m_h p_i)^{-} p_i m_h \varepsilon_i \), \( \overline{z_i} \equiv (p_i m_b p_i)^{-} p_i m_b \varepsilon_i \), \( \overline{z_i} \equiv (p_i m_b p_i)^{-} p_i m_b r_i \), \( \overline{r_i} \equiv (p_i m_b p_i)^{-} p_i m_b r_i \), \( \overline{z_i} \equiv (p_i m_b p_i)^{-} p_i m_b r_i \), \( \overline{r_i} \equiv f_2 \gamma_{2i} + (g_i - p_i c_{2i}) \), and \( \lambda \equiv \lambda^{-1} \leq n \lambda_i \). Suppose Assumptions 1-2 and 3(iv) hold, then

(i) \( n_{\text{c}} T \sum_{i=1}^n (\overline{z_i} - \overline{z}_i) \mathcal{P}_w (\overline{z}_i - \overline{z}_i) = O_p \left( K^2 / n^{1/2} \right) \);
(ii) \( n_{\text{c}} T \sum_{1 \leq i \neq j \leq n} (\overline{z}_i - \overline{z}_i) \mathcal{P}_w (\overline{z}_i - \overline{z}_j) = O_p \left( K / T^{1/2} \right) \);
(iii) \( n_{\text{c}} T \sum_{1 \leq i \neq j \leq n} (\overline{r}_i - \overline{r}_i) \mathcal{P}_w (\overline{r}_i - \overline{r}_j) = O_p \left( T n^{-1/2} K^{1 - 2\lambda / d} + KT / n^{1/2} \right) \);
(iv) \( n_{\text{c}} T \sum_{1 \leq i \neq j \leq n} (\overline{z}_i - \overline{z}_j) \mathcal{P}_w (\overline{z}_i - \overline{z}_j) = O_p \left( T n^{-1/2} K^{1 - 2\lambda / d} \right) \);
(v) \( n_{\text{c}} T \sum_{1 \leq i \neq j \leq n} (\overline{z}_i - \overline{z}_j) \mathcal{P}_w (\overline{z}_i - \overline{z}_i) = O_p \left( K / T^{1/2} + T^{1/2} K^{1 - \lambda / d} \right) \);
(vi) \( n_{\text{c}} T \sum_{1 \leq i \neq j \leq n} (\overline{z}_i - \overline{z}_i) \mathcal{P}_w (\overline{z}_i - \overline{z}_j) = O_p \left( T^{1/2} K^{1 - 2\lambda / d} \right) \).

**Proof.** (i) Noting that \( \overline{z}_i - \overline{z}_i = [(p_i m_h p_i)^{-} (p_i m_b p_i)^{-}] p_i (m_h - m_b) \varepsilon_i \) and \( (p_i m_b p_i)^{-} - (p_i m_b p_i)^{-} = (p_i m_h p_i)^{-} p_i (m_h - m_b) p_i (p_i m_b p_i)^{-} \), by the \( C_r \) inequality and Lemmas B.3(i)-
(ii) we have
\[
nc_{nT} \sum_{i=1}^{n} (\bar{e}_i - \bar{e}_j) \cdot P_w (\bar{e}_i - \bar{e}_j)
\]
\[
\leq nc_{nT} \lambda_{\text{max}} (P_w) \sum_{i=1}^{n} \|\bar{e}_i - \bar{e}_j\|^2
\]
\[
\leq 2nc_{nT} \lambda_{\text{max}} (P_w) \sum_{i=1}^{n} \left\| (p'_i m_{h i}) - (p'_i m_{h j}) \right\| \| p'_i m_{h i} \bar{e}_i \|^2
\]
\[
+ 2nc_{nT} \lambda_{\text{max}} (P_w) \sum_{i=1}^{n} \left\| T^{-1} p'_i (m_h - m_b) \bar{e}_i \right\|^2
\]
\[
\leq 2nc_{nT} \lambda_{\text{max}} (P_w) \left( c_1 c_2 \right) \cdot (c_2) ^{-2} \sum_{i=1}^{n} \left\| T^{-1} p'_i (m_h - m_b) p_i T^{-1} p'_i m_b \right\|^2
\]
\[
+ 2nc_{nT} \lambda_{\text{max}} (P_w) \left( c_2 \right) ^{-2} \sum_{i=1}^{n} \left\| T^{-1} p'_i (m_h - m_b) \right\|^2
\]
\[
= nc_{nT} \left( O_p (K^2/T) + O_p (K/T) \right) = O_p \left( K^{2/1/2} \right).
\]

(ii) Noting that \( c_{nT} \sum_{1 \leq i \neq j \leq n} (\bar{e}_i - \bar{e}_j) \cdot P_w (\bar{e}_i - \bar{e}_j) = Q_{1nT} - Q_{2nT} \), where \( Q_{1nT} = nc_{nT} \sum_{i=1}^{n} (\bar{e}_i - \bar{e}_j) \cdot P_w (\bar{e}_i - \bar{e}_j) \), and \( Q_{2nT} = c_{nT} \sum_{i=1}^{n} \sum_{j=1}^{n} (\bar{e}_i - \bar{e}_j) \cdot P_w (\bar{e}_i - \bar{e}_j) \). It suffices to prove (ii) by showing that \( Q_{snT} = O_p (K^{1/2}) \) for \( s = 1, 2 \). We decompose \( Q_{1nT} \) as follows
\[
Q_{1nT} = nc_{nT} \sum_{i=1}^{n} \left\{ (p'_i m_{h i}) - (p'_i m_{h j}) \right\} \cdot P_w (p'_i m_{h i}) - p'_i m_{h i}
\]
\[
+ nc_{nT} \sum_{i=1}^{n} \left\{ (p'_i m_{h i}) - (p'_i m_{h j}) \right\} \cdot P_w (p'_i m_{h i}) - p'_i m_{h i}
\]
\[
= Q_{1nT,1} + Q_{1nT,2}.
\]

By (B.1) we can further decompose \( Q_{1nT,1} \) as follows
\[
Q_{1nT,1} = nc_{nT} \sum_{i=1}^{n} \left\{ (p'_i m_{h i}) - (p'_i m_{h j}) \right\} \cdot P_w (p'_i m_{h i}) - p'_i m_{h i}
\]
\[
+ nc_{nT} \sum_{i=1}^{n} \left\{ (p'_i m_{h i}) - (p'_i m_{h j}) \right\} \cdot P_w (p'_i m_{h i}) - p'_i m_{h i}
\]
\[
= -Q_{1nT,11} + Q_{1nT,12} - Q_{1nT,13}.
\]
To analyze \( Q_{1nT,11} \), note that \( \tilde{\nu}_t^+ = (0', \tilde{\nu}_t', \tilde{\eta}_t) \). So we can decompose \( \tilde{\nu}^+ = (\tilde{\nu}_1^+, \cdots, \tilde{\nu}_T^+)' = \nu^+ + \tilde{\gamma}^+ \), where the \( t \)th rows of \( \nu^+ \) and \( \tilde{\gamma}^+ \) are given by \( \nu_t^+ = (0', \tilde{\nu}_t', \tilde{\eta}_t) \) and \( \tilde{\gamma}_t = (0', 0', \tilde{\eta}_t)' \), 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 \( \tilde{\nu}^+ \) being replaced by \( \nu^+ \) and \( \tilde{\gamma}^+ \), respectively. Let \( \xi_i = p_i (p'_i m_{h i}) - p'_i m_{h i} \) and
\( \xi_i^c \equiv \xi_i - E_D (\xi_i) \). Then we can decompose \( Q_{1nT,111} \) as follows

\[
Q_{1nT,111} = nc_{nT} \sum_{i=1}^n \text{tr} \left( b (b')^{-1} \pi^+ E_D (\xi_i) \right) + nc_{nT} \sum_{i=1}^n \text{tr} \left( b (b')^{-1} \pi^+ \xi_i^c \right)
\]

\[
= Q_{1nT,111a} + Q_{1nT,111b}.
\]

Let \( \varsigma_i \equiv (\pi^+ p_i (p'_i \nu p_i) - p'_i \nu) \equiv (p_i (p'_i \nu p_i) - p'_i E (\pi^+ \nu)) \). Then by Lemma B.2(i)

\[
E_D [\text{tr} (\varsigma_i)] = E_D \left[ \text{tr} \left( \pi^+ p_i (p'_i \nu p_i) - p'_i \nu \right) \right]
\]

\[
= \text{tr} \left\{ p_i (p'_i \nu p_i) - p'_i E (\pi^+ \nu) \right\}
\]

\[
\leq Cn^{-1} \text{tr} \left\{ p_i (p'_i \nu p_i) - p'_i E (\pi^+ \nu) \right\}
\]

\[
\leq Cn^{-1} [\lambda_{\max} (\pi^+ \nu)]^2 (c_{1\lambda})^{-3} T^{-3} \|p_i\|^2.
\]

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

\[
Q_{1nT,111a} = nc_{nT} \sum_{i=1}^n \text{tr} \left( \varsigma_i E (\xi_i \nu_i) b (b')^{-1} \right)
\]

\[
\leq nc_{nT} \sum_{i=1}^n \left\{ \{\text{tr} (\varsigma_i \nu_i)\}^{1/2} \{\text{tr} (b (b')^{-1} b' E (\xi_i \nu_i) b (b')^{-1})\}^{1/2} \right\}
\]

\[
\leq Cn^{-1/2} \lambda_{\max} (\pi^+ \nu) (c_{1\lambda})^{-3/2} \max_{1 \leq i \leq n} [\lambda_{\max} (E (\xi_i \nu_i))] \lambda_{\min} \left\{ (T^{-1} b')^{-1} \right\}^{1/2}
\]

\[
\times nc_{nT} \left\{ T^{-2} \sum_{i=1}^n \|p_i\| \right\}
\]

\[
= n^{1/2} c_{nT} O_p \left( n\sqrt{K/T} \right) = O_p \left( \sqrt{K/T} \right).
\]

Similarly, for \( Q_{1nT,111b} \) we have

\[
Q_{1nT,111b} = nc_{nT} \text{tr} \left( b (b')^{-1} \pi^+ \sum_{i=1}^n \xi_i^c \right) \leq nc_{nT} \{Q_1 \}^{1/2} \{Q_2 \}^{1/2},
\]

where \( Q_1 \equiv \text{tr}(b (b')^{-1} \pi^+ \nu (b')^{-1} b') \), and \( Q_2 \equiv \text{tr} \left( \sum_{i=1}^n \sum_{j=1}^n \xi_i^c \xi_j^c \right) \). Noting that

\[
Q_1 = \text{tr}(b (b')^{-1} \pi^+ \nu) = [\lambda_{\min} (T^{-1} b')^{-1}]^{-1} \|\nu^+\|^2 = O_p (1/n),
\]

and

\[
E_D [Q_2] = \sum_{i=1}^n \text{tr} (E_D (\xi_i^c \nu_i^c)) \leq C \sum_{i=1}^n p_i (p'_i \nu p_i) - p'_i \nu E (\pi^+ \nu) \nu (p'_i \nu p_i) - p'_i \nu
\]

\[
\leq C \left[ \lambda_{\max} (\nu \nu) \right] (c_{1\lambda})^{-3} T^{-3} \sum_{i=1}^n \|p_i\|^2 = O_p (nK/T^2),
\]

we have \( Q_{1nT,111b} = nc_{nT} O_p (1/\sqrt{n}) O_p \left( \sqrt{nK/T} \right) = O_p (\sqrt{K/n}) \). 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 \( Q_{1nT,112} = nc_{nT} \sum_{i=1}^n \xi_i \) with \( \xi_i \equiv \nu_i^c (b (b')^{-1} \pi^+ p_i (p'_i \nu p_i) - p'_i \nu p_i) - p'_i \nu \xi_i \), we have \( (Q_{1nT,112})^2 = Q_{1nT,112a} + Q_{1nT,112b} \), where \( Q_{1nT,112a} = (nc_{nT})^2 \sum_{i=1}^n \xi_i^2 \), and \( Q_{1nT,112b} = \)
$(nc_nT)^2 \sum_{1 \leq i \neq j \leq n} \zeta_i \zeta_j$. It is easy to show that $E_D(Q_{1nT,112a}) = O_p(K/(nt))$, implying that $Q_{1nT,112a} = O_p(K/(nt))$ by the Markov inequality. For $Q_{1nT,112b}$, we have

$$E_D[Q_{1nT,112b}] = (nc_nT)^2 \sum_{1 \leq i \neq j \leq n} \text{tr} \left\{ b(b')^{-1} \mathcal{g}' p_i (p'_i m_b p_i) \mathcal{w}_w (p'_i m_b) - p'_i m_b E(\varepsilon_i') \right\} \times \text{tr} \left\{ b(b')^{-1} \mathcal{g}' p_j (p'_j m_b p_j) \mathcal{w}_w (p'_j m_b) - p'_j m_b E(\varepsilon_j') \right\} \leq D^2,$$  

where $D \equiv nc_nT \sum_{i=1}^n \{ \text{tr}(b(b')^{-1} \mathcal{g}' p_i (p'_i m_b p_i) \mathcal{w}_w (p'_i m_b) - p'_i m_b E(\varepsilon'_i)) \}^{1/2}$. Note that

$$D \leq C \lambda_{\max}(\mathcal{w}_w) c_1 \left\{ (T^{-1/2}) \right\}^{1/2} nc_nT \left\{ T^{-5/2} \sum_{i=1}^n \|p'_i \mathcal{g}' \| \|p'_i m_b\| \right\} = nc_nT O_p\left( (n^{1/2} K T^{3/2}) \right) = O_p\left( K T^{1/2} \right).$$

So we have $Q_{1nT,112b} = O_p(K^{1/2}/T^{1/2})$ and $Q_{1nT,112} = O_p(K^{1/2}/T^{1/2})$. Consequently $Q_{1nT,111} = O_p(K^{1/2}/T^{1/2} + K T^{1/2}) = O_p(K T^{1/2})$. Analogously, we can show that $Q_{1nT,13} = O_p(K T^{1/2})$ for $s = 2, 3$. It follows that $Q_{1nT,1} = O_p(K T^{1/2})$.

For $Q_{1nT,2}$, it is easy to show that $Q_{1nT,2} \equiv nc_nT \sum_{i=1}^n \varepsilon'_i m_h p_i (p'_i m_b p_i) - p'_i (m_h - m_b) p_i (p'_i m_b p_i) \mathcal{w}_w (p'_i m_b) - p'_i m_b \varepsilon_i = Q_{1nT,2} + O_p(1)$ where $Q_{1nT,2} \equiv nc_nT \sum_{i=1}^n \varepsilon'_i m_h p_i (p'_i m_b p_i) - p'_i (m_h - m_b) p_i (p'_i m_b p_i) \mathcal{w}_w (p'_i m_b) - p'_i m_b \varepsilon_i$. Then using the decomposition in (B.1) and the proof strategy for $Q_{1nT,1}$, we can show that $Q_{1nT,2} = O_p(K T^{1/2})$. Consequently, $Q_{1nT} = O_p(K T^{1/2})$. Analogously we can prove $Q_{2nT,3} \equiv c_{nT} \sum_{j=1}^n (\varepsilon_i - \varepsilon'_j) \mathcal{w}_w \varepsilon_j = O_p(K T^{1/2})$.

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

$$\tilde{r}_i - \bar{r}_i = (p'_i m_b p_i) - p'_i (m_h - m_b) r_i + \left[ (p'_i m_b p_i) - (p'_i (m_h - m_b)) \right] p'_i m_b r_i$$

$$= (p'_i m_b p_i) - p'_i (m_h - m_b) f_2 \gamma_i + (p'_i m_b p_i) - p'_i (m_h - m_b) d_i$$

$$+ \left[ (p'_i m_b p_i) - (p'_i (m_h - m_b)) \right] p'_i m_b d_i,$$

where $d_i \equiv g_i - p_i \alpha_i$. It follows that

$$nc_nT \sum_{i=1}^n \left( \tilde{r}_i - \bar{r}_i \right)' \mathcal{w}_w (\tilde{r}_i - \bar{r}_i)$$

$$\leq 3nc_nT \sum_{i=1}^n \left[ (p'_i m_b p_i) - p'_i (m_h - m_b) d_i \right]' \mathcal{w}_w \left[ (p'_i m_b p_i) - p'_i (m_h - m_b) d_i \right]$$

$$+ 3nc_nT \sum_{i=1}^n \left[ (p'_i m_b p_i) - p'_i (m_h - m_b) f_2 \gamma_i \right]' \mathcal{w}_w \left[ (p'_i m_b p_i) - p'_i (m_h - m_b) f_2 \gamma_i \right]$$

$$+ 3nc_nT \sum_{i=1}^n \left\{ \left[ (p'_i m_b p_i) - (p'_i (m_h - m_b)) \right] p'_i m_b d_i \right\}' \mathcal{w}_w \{ \left[ (p'_i m_b p_i) - (p'_i (m_h - m_b)) \right] p'_i m_b d_i \}$$

$$= 3U_{1nT} + 3U_{2nT} + 3U_{3nT}.$$

For $U_{1nT}$, we have

$$U_{1nT} \leq \lambda_{\max}(\mathcal{w}_w) nc_nT \sum_{i=1}^n \left\| (p'_i m_b p_i) - p'_i (m_h - m_b) d_i \right\|^2 \leq \lambda_{\max}(\mathcal{w}_w) c_2^2 U_{1nT}.$$
where \( U_{1nT} \equiv n c_{nT} \sum_{i=1}^{n} \| T^{-1} p_i (m_b - m_b) d_i \|^2 \). Using (B.1), we have

\[
U_{1nT} \leq 3 n c_{nT} \sum_{i=1}^{n} \left\| T^{-1} p_i (b - h) (b' b)^{-1} b' d_i \right\|^2 + 3 n c_{nT} \sum_{i=1}^{n} \left\| T^{-1} p_i h [(b' b)^{-1} - (h' h)^{-1}] b' d_i \right\|^2 + 3 n c_{nT} \sum_{i=1}^{n} \left\| T^{-1} p_i h (h' h)^{-1} (b - h)' d_i \right\|^2
\]

\[= 3 U_{1nT,1} + 3 U_{1nT,2} + 3 U_{1nT,3}, \text{ say.}
\]

To proceed, we notice that by Lemma B.2

\[
T^{-2} \sum_{i=1}^{n} \| p_i \|^2 \| d_i \|^2 \leq n K \max_{1 \leq i \leq n} \left\{ \epsilon_{\max} \left( T^{-1} p_i p_i \right) \right\} \left\{ (nT)^{-1} \sum_{i=1}^{n} \| d_i \|^2 \right\} = n K O_p \left( K^{-2 \lambda/d} \right) = O_p \left( n K^{1-2 \lambda/d} \right).
\]

For \( U_{1nT,1} \), we have

\[
U_{1nT,1} \leq n c_{nT} T^{-2} \left\| (b' b)^{-1} b' \right\|^2 \sum_{i=1}^{n} \| p_i (b - h) \|^2 \| d_i \|^2 = \text{tr} \left( (T^{-1} b' b)^{-1} \right) U_{1nT,1}
\]

where \( U_{1nT,1} \equiv n c_{nT} T^{-3} \sum_{i=1}^{n} \| p_i (b - h) \|^2 \| d_i \|^2 \). Recall that \( \bar{v}_i \) denotes the \( i \)th row of \( h - b : \bar{v}_i = (0, \bar{v}_i^T, \bar{v}_i + \bar{g}_i) \). Let \( \bar{v} = (\bar{v}_1, ..., \bar{v}_p) \). Similarly define \( \bar{s} \) and \( \bar{g} \). Noting that \( n E(\bar{v} \bar{v}') \leq C I_T \) and \( n E(\bar{s} \bar{g}) \leq C I_T \) for some \( C > 0 \) by Lemma B.2(i), we have by (B.4)

\[
E_D(\bar{U}_{1nT,1}) = n c_{nT} T^{-3} \sum_{i=1}^{n} \text{tr} \left( p_i E_D (b - h) (b - h)' p_i \right) \| d_i \|^2 = C_{nT} T^{-3} \sum_{i=1}^{n} \text{tr} \left( p_i p_i \right) \| d_i \|^2 + C n c_{nT} T^{-3} \sum_{i=1}^{n} \text{tr} \left( p_i \bar{g} \bar{g}' p_i \right) \| d_i \|^2 \leq C_{nT} T^{-1} \left( 1 + n \| \bar{g} \|^2 \right) \left\{ T^{-2} \sum_{i=1}^{n} \| p_i \|^2 \| d_i \|^2 \right\}
\]

\[= c_{nT} T^{-1} O_p (T) O_p \left( n K^{1-2 \lambda/d} \right) = O_p \left( T n^{-1/2} K^{1-2 \lambda/d} \right).
\]

Consequently, \( \bar{U}_{1nT,1} = O_p \left( T n^{-1/2} K^{1-2 \lambda/d} \right) \) and \( U_{1nT,1} = O_p \left( T n^{-1/2} K^{1-2 \lambda/d} \right) \). Analogously, we can show that \( U_{1nT,2} = O_p \left( T n^{-1/2} K^{1-2 \lambda/d} \right) \). For \( U_{1nT,2} \), we have

\[
U_{1nT,2} \leq \| h [(b' b)^{-1} - (h' h)^{-1}] b' \|^2 n c_{nT} T^{-2} \sum_{i=1}^{n} \| p_i \|^2 \| d_i \|^2 \leq \text{tr} \left( h' h [(b' b)^{-1} - (h' h)^{-1}] b' (b' b)^{-1} - (h' h)^{-1} \right) n c_{nT} T^{-2} \sum_{i=1}^{n} \| p_i \|^2 \| d_i \|^2 \leq \lambda_{\max} (T^{-1} h' h) \lambda_{\max} (T^{-1} b' b) \| (T^{-1} b' b)^{-1} - (T^{-1} h' h)^{-1} \|^2 n c_{nT} \left\{ T^{-2} \sum_{i=1}^{n} \| p_i \|^2 \| d_i \|^2 \right\}
\]

\[= O_p \left( n^{-1} n c_{nT} T \left( n K^{1-2 \lambda/d} \right) = O_p \left( T n^{-1/2} K^{1-2 \lambda/d} \right),
\]

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

$$U_{2nT} \leq \lambda_{\max} \left( \mathcal{P}_w \right) nc_{nT} \sum_{i=1}^{n} \left\| \left( p'_i m_{ib} p_i \right)^{-1} p'_i \left( m_{ib} - m_b \right) f_2 \gamma_{2i} \right\|^2 \leq \lambda_{\max} \left( \mathcal{P}_w \right) (c_{2\lambda})^{-2} U_{2nT}$$

where $U_{2nT} = nc_{nT} \sum_{i=1}^{n} \left\| T^{-1} p'_i \left( m_{ib} - m_b \right) f_2 \gamma_{2i} \right\|^2$. By (B.1) we have

$$U_{2nT} \leq 3nc_{nT} \sum_{i=1}^{n} \left\| T^{-1} p'_i \left( b' b^{-1} b' f_2 \gamma_{2i} \right) \right\|^2 + 3nc_{nT} \sum_{i=1}^{n} \left\| T^{-1} p'_i h \left( b' b^{-1} - (h'h)^{-1} \right) b' f_2 \gamma_{2i} \right\|^2$$

$$+ 3nc_{nT} \sum_{i=1}^{n} \left\| T^{-1} p'_i h (h'h)^{-1} f_2 \gamma_{2i} \right\|^2$$

$$= 3U_{2nT,1} + 3U_{2nT,2} + 3U_{2nT,3}.$$ 

Noting that

$$E_D (U_{2nT,1}) = nc_{nT} \sum_{i=1}^{n} E_D \left\| T^{-1} p'_i \left( b' b^{-1} b' f_2 \gamma_{2i} \right) \right\|^2$$

$$\leq Cnc_{nT} \left( n^{-1} + \left\| B \right\|^2 \right) \left[ \lambda_{\min} \left( T^{-1} b' b^{-1} \right) \right]^{-2} \left\{ T^{-1} \sum_{i=1}^{n} \text{tr} \left( T^{-1} p'_i p_i \right) E_D \left\| T^{-1} b' f_2 \gamma_{2i} \right\|^2 \right\}$$

$$= nc_{nT} O_p \left( (KT/n)^{1/2} \right),$$

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

For $U_{3nT}$, using the expression $(p'_i m_{ib} p_i)^{-1} - (p'_i m_{ib} p_i)^{-1} p'_i (m_{ib} - m_b) p_i (p'_i m_{ib} p_i)^{-1}$, 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, $nc_{nT} \sum_{i=1}^{n} (\tilde{r}_i - \tilde{r}_i)' \mathcal{P}_w (\tilde{r}_i - \tilde{r}_i) = O_p \left( Tn^{-1/2} K^{1-2\lambda/d} + KT/n^{1/2} \right)$.

(iv) Noting that by (B.4)

$$nc_{nT} \sum_{i=1}^{n} (\tilde{r}_i - \tilde{r}_j)' \mathcal{P}_w (\tilde{r}_i - \tilde{r}_j) = nc_{nT} \lambda_{\max} \left( \mathcal{P}_w \right) \sum_{i=1}^{n} \left\| \left( p'_i m_{ib} p_i \right)^{-1} p'_i m_{ib} d_i \right\|^2$$

$$\leq nc_{nT} \lambda_{\max} \left( \mathcal{P}_w \right) (c_{1\lambda})^{-2} T^{-2} \sum_{i=1}^{n} \left\| p_i \right\|^2 \left\| d_i \right\|^2$$

$$= nc_{nT} O_p \left( n K^{1-2\lambda/d} \right) = O_p \left( T n^{1/2} K^{1-2\lambda/d} \right),$$

we have by the Cauchy-Schwarz inequality

$$\left| \sum_{1 \leq i \neq j \leq n} (\tilde{r}_i - \tilde{r}_j)' \mathcal{P}_w (\tilde{r}_i - \tilde{r}_j) \right| = \left| \sum_{1 \leq i \neq j \leq n} \{ \tilde{r}_i' \mathcal{P}_w \tilde{r}_i + \tilde{r}_j' \mathcal{P}_w \tilde{r}_j - 2 \tilde{r}_i' \mathcal{P}_w \tilde{r}_j \} \right|$$

$$\leq 4nc_{nT} \sum_{i=1}^{n} \tilde{r}_i' \mathcal{P}_w \tilde{r}_i = O_p \left( T n^{1/2} K^{1-2\lambda/d} \right).$$
(v) Write $c_{nT} \sum_{1 \leq i < j \leq n} (\vec{\xi}_i - \vec{\xi}_j)' \mathbf{P}_w (\vec{\xi}_i - \vec{\xi}_j) = W_{1nT} - W_{2nT}$, where $W_{1nT} = nc_{nT} \sum_{i=1}^n \vec{\xi}_i' \mathbf{P}_w (\vec{\xi}_i - \vec{\xi}_i)$ and $W_{2nT} = c_{nT} \sum_{1 \leq i, j \leq n} \vec{\xi}_i' \mathbf{P}_w (\vec{\xi}_i - \vec{\xi}_j)$. For $W_{1nT}$, by (B.3)

$$W_{1nT} = nc_{nT} \sum_{i=1}^n \vec{\xi}_i' m_o (p_i' m_i) - \mathbf{P}_w (p_i' m_i) - p_i' (m_h - m_b) f_2 \gamma_{2i}$$

$$+ nc_{nT} \sum_{i=1}^n \vec{\xi}_i' m_o (p_i' m_i) - \mathbf{P}_w (T^{-1} p_i' m_h) - T^{-1} p_i' (m_h - m_b) \mathbf{d}_i$$

$$+ nc_{nT} \sum_{i=1}^n \vec{\xi}_i' m_o (p_i' m_i) - \mathbf{P}_w \left[ (T^{-1} p_i' m_h) - (T^{-1} p_i' m_h) \right] T^{-1} p_i' m_b \mathbf{d}_i$$

$$\equiv W_{1nT,1} + W_{1nT,2} + W_{1nT,3}.$$

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

$$W_{1nT,1} = -nc_{nT} \sum_{i=1}^n \vec{\xi}_i' m_o (p_i' m_i) - \mathbf{P}_w (p_i' m_i) - p_i' \mathbf{v} \mathbf{b}^{-1} b' f_2 \gamma_{2i}$$

$$+ nc_{nT} \sum_{i=1}^n \vec{\xi}_i' m_o (p_i' m_i) - \mathbf{P}_w (p_i' m_i) - p_i' h (b'h)^{-1} - (h'h)^{-1} b' f_2 \gamma_{2i}$$

$$- nc_{nT} \sum_{i=1}^n \vec{\xi}_i' m_o (p_i' m_i) - \mathbf{P}_w (p_i' m_i) - p_i' h (h'h)^{-1} \mathbf{w} f_2 \gamma_{2i}$$

$$\equiv -W_{1nT,11} + W_{1nT,12} - W_{1nT,13}.$$

It is easy to show that $W_{1nT,11} = \mathbf{W}_{1nT,11} + o_p(1)$, where $\mathbf{W}_{1nT,11} = nc_{nT} \sum_{i=1}^n \vec{\xi}_i' m_o (p_i' m_i) - \mathbf{P}_w (p_i' m_i) - p_i' \mathbf{v} \mathbf{b}^{-1} b' f_2 \gamma_{2i}$. Using the decomposition $\mathbf{v} = \mathbf{v}^+ + \mathbf{v}^+$, we can further decompose $W_{1nT,11} = \mathbf{W}_{1nT,111} + \mathbf{W}_{1nT,112}$, where $W_{1nT,111}$ and $W_{1nT,112}$ are analogously defined as $W_{1nT,11}$ but with $\mathbf{v}^+$ being replaced by $\mathbf{v}^+$ and $\mathbf{v}^+$, respectively. It is easy to show $|\mathbf{W}_{1nT,111}| = O_p(K/T^{1/2})$.

For $\mathbf{W}_{1nT,112}$, noting that $E_D(\mathbf{W}_{1nT,112}) = 0$ and $E_D(|\mathbf{W}_{1nT,112}|^2) = (nc_{nT})^2 O_p(K/T^2) = O_p(K/n)$, So $\mathbf{W}_{1nT,112} = O_p(K^{1/2}/n^{1/2})$ and $W_{1nT,11} = O_p(K/T^{1/2})$. Similarly, we can show that $W_{1nT,1s} = O_p(K/T^{1/2})$ for $s = 2, 3$. It follows that $W_{1nT,1} = O_p(K/T^{1/2})$. For $W_{1nT,2}$, we have

$$|W_{1nT,2}| \leq \left\{ nc_{nT} \sum_{i=1}^n \vec{\xi}_i' m_o (p_i' m_i) - \mathbf{P}_w (p_i' m_i) - p_i' \mathbf{v} \mathbf{z} \right\}^{1/2} U_{1nT}^{1/2}$$

$$= \left\{ O_p \left( K n^{1/2} \right) \right\} = \left\{ \left. \frac{1}{2} \left\{ O_p \left( T^{1/2} K^{1-\lambda/d} \right) \right\} \right\}^{1/2} = O_p \left( T^{1/2} K^{1-\lambda/d} \right).$$

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 $W_{2nT} = O_p \left( K/T^{1/2} + T^{1/2} K^{1-\lambda/d} \right)$. For $W_{1nT}$, by the Cauchy-Schwarz inequality, it suffices to prove (vi) by showing that $W_{3nT} \equiv nc_{nT} \sum_{i=1}^n \vec{\xi}_i' \mathbf{P}_w \mathbf{z} = O_p \left( T^{1/2} K^{1-\lambda/d} \right)$. Note that $E_D(\mathbf{W}_{3nT}) = 0$, and

$$E_D \left[ (W_{3nT})^2 \right] = (nc_{nT})^2 \sum_{i=1}^n d_i' m_o (p_i' m_i) - \mathbf{P}_w (p_i' m_i) - p_i' m_b E (\varepsilon_i \varepsilon_i')$$

$$\times m_o (p_i' m_i) - \mathbf{P}_w (p_i' m_i) - p_i' m_b \mathbf{d}_i$$

$$\equiv \sum_{i=1}^n d_i' m_o (p_i' m_i) - \mathbf{P}_w (p_i' m_i) - p_i' m_b E (\varepsilon_i \varepsilon_i')$$

$$\times m_o (p_i' m_i) - \mathbf{P}_w (p_i' m_i) - p_i' m_b \mathbf{d}_i$$

$$\equiv \sum_{i=1}^n d_i' m_o (p_i' m_i) - \mathbf{P}_w (p_i' m_i) - p_i' m_b E (\varepsilon_i \varepsilon_i')$$

$$\times m_o (p_i' m_i) - \mathbf{P}_w (p_i' m_i) - p_i' m_b \mathbf{d}_i$$

$$\equiv W_{3nT,1} + W_{3nT,2} + W_{3nT,3}.$$

Using (B) we further decompose $W_{3nT,1}$ as follows:

$$W_{3nT,1} = -nc_{nT} \sum_{i=1}^n \vec{\xi}_i' m_o (p_i' m_i) - \mathbf{P}_w (p_i' m_i) - p_i' \mathbf{v} \mathbf{b}^{-1} b' f_2 \gamma_{2i}$$

$$+ nc_{nT} \sum_{i=1}^n \vec{\xi}_i' m_o (p_i' m_i) - \mathbf{P}_w (p_i' m_i) - p_i' h (b'h)^{-1} - (h'h)^{-1} b' f_2 \gamma_{2i}$$

$$- nc_{nT} \sum_{i=1}^n \vec{\xi}_i' m_o (p_i' m_i) - \mathbf{P}_w (p_i' m_i) - p_i' h (h'h)^{-1} \mathbf{w} f_2 \gamma_{2i}$$

$$\equiv -W_{3nT,11} + W_{3nT,12} - W_{3nT,13}.$$
\[ \leq C (nc_{nT})^2 \sum_{i=1}^{n} d_i m_b p_i (p_i' m_b p_i)^{-1} P_w (p_i' m_b p_i)^{-1} P_w (p_i' m_b p_i) - p_i' m_b d_i \]
\[ \leq C (nc_{nT})^2 \left( \lambda_{\max} (P_w) \right)^2 (c_1 \lambda)^{-3} \left\{ T^{-3} \sum_{i=1}^{n} d_i m_b p_i p_i' m_b d_i \right\} \]
\[ = (nc_{nT})^2 O_p \left( nK^{1-2\lambda/d}/T \right) = O_p \left( TK^{1-2\lambda/d} \right) . \]

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

References


Moon, H., Weidner, M., 2010a. Linear regression for panel with unknown number of factors as interactive fixed effects. Working paper, University of Southern California.

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


