Testing Heterogeneity in Panel Data Models with Interactive Fixed Effects*

Liangjun Su, Qihui Chen
School of Economics, Singapore Management University, Singapore

July 14, 2011

Abstract

This paper proposes a test for the slope homogeneity in large dimensional panel data models with interactive fixed effects based on a measure of goodness-of-fit (R^2). We first obtain, for each cross-sectional unit, the R^2 from the time series regression of residuals on the constant and observable regressors and then construct the test statistic $\bar{R}^2$ as an equally weighted average of the cross-sectional R^2's. $\bar{R}^2$ is close to 0 under the null hypothesis of homogenous slopes and deviates away from 0 otherwise. We show that after being appropriately centered and scaled, $\bar{R}^2$ is asymptotically normally distributed under the null and a sequence of Pitman local alternatives. To improve the finite sample performance of the test, we also propose a bootstrap procedure to obtain the bootstrap p-values and justify its validity. Monte Carlo simulations suggest that the test has correct size and satisfactory power, and is superior to a recent test proposed by Pesaran and Yamagata (2008) that neglects cross-sectional dependence in panel data models. We apply our tests to study the OECD economic growth model and the Fama-French three factor model for asset returns.

JEL Classifications: C12, C14, C23

Key Words: Cross-sectional dependence; Goodness-of-fit; Heterogeneity; Interactive fixed effects; Large panels; Principal component analysis

*We sincerely thank the participants at the 2011 Tsinghua International Conference in Econometrics and the seminar at the University of Adelaide who provided valuable suggestions and discussion. Address correspondence to Liangjun Su, School of Economics, Singapore Management University, 90 Stamford Road, Singapore, 178903, E-mail: ljsu@smu.edu.sg; Phone: +65 6828 0386.

1 Introduction

Recently large dimensional panel data models with interactive fixed effects have attracted huge attention in econometrics. Pesaran (2006) proposed a number of estimators [referred to as common correlated effects (CCE) estimators] for heterogeneous panels and derived their asymptotic normal distributions under fairly general conditions. Bai (2009a) studied identification, consistency, and the limiting distribution of the principal component analysis (PCA) estimators and demonstrated that they are $\sqrt{NT}$ consistent, where N and T refer to the individual and time series dimensions, respectively. Kapetanios and Pesaran (2007) proposed a factor-augmented estimator by augmenting a linear panel data model with estimated common factors to account for cross sectional dependence and studied its finite sample properties via Monte Carlo simulations. Greenaway-McGrevy, Han and Sul (2010) formally established the asymptotic distribution of this estimator and provided specific conditions under which the estimated factors can be used in place of the latent factors in the regression. Moon and Weidner (2010b) considered dynamic linear panel regression models with interactive fixed effects, and found that there are two sources of asymptotic biases for the Gaussian quasi maximum likelihood estimator (QMLE): one is due to correlation or heteroscedasticity of the idiosyncratic error term and the other is the presence of predetermined regressor. In addition, Moon and Weidner (2010a) discussed the validity of QMLE method for panel data models when the number of factors as interactive fixed effects is unknown and has to be chosen according to certain information criteria. Pesaran and Tosetti (2011) considered estimation of panel data models with a multifactor error structure and spatial error correlations and found that Pesaran's CCE procedure continues to yield consistent and asymptotically normal estimates of the slope coefficients.

Panel data models with interactive fixed effects are useful modelling paradigm. In macroeconomics, incorporating interactive effects can account for the heterogenous impact of unobservable common shocks, while the regressors can be such input as labor and capital. In finance, combination of unobserved factors and observed covariates can explain the excess returns of assets. In microeconomics, panel data models with interactive fixed effects can incorporate unmeasured skills or unobservable characteristics to study the individual wage rate. Nevertheless, in most empirical studies it is commonly assumed that the coefficients of the observed regressors are homogeneous. In fact, most of the literature reviewed above is developed for homogeneous panel data models with interactive fixed effects. The only exceptions are Pesaran (2006), Kapetanios and Pesaran (2007) and Pesaran and Tosetti (2011) that are applicable to heterogeneous panels but typically require certain rank conditions to satisfy in order to estimate individual slopes. Su and Jin (2010) extended Pesaran (2006) to nonparametric regression with a multi-factor error structure.

Slope homogeneity assumption greatly simplifies the estimation and inference process and the proposed estimator can be efficient if there is no heterogeneity in individual slopes. Nevertheless, if the slope homogeneity assumption is not true, estimates based on panel data models with homogeneous slopes can be inconsistent and lead to misleading statistical inference, see, for example, Hsiao (2003, Chapter 6) and Baltagi, Bresson and Pirotte (2008). So it is necessary and prudent to test for slope homogeneity before imposing it.

There are many studies on testing for slope homogeneity in the panel data literature, see Pesaran, Smith and Im (1996), Phillips and Sul (2003), Pesaran and Yamagata (2008, PY hereafter), Blomquist (2010), Lin (2010), Jin and Su (2011), among others. Pesaran, Smith and Im (1996) proposed a Hausman-type test by comparing the standard fixed effects estimator with the mean group estimator. Phillips and Sul (2003) also proposed a Hausman-type test for slope homogeneity for AR(1) panel data models in the presence of cross-sectional dependence. Recently, PY developed a standardized version of Swamy's test for the slope homogeneity in large panel data model with fixed effects and unconditional heteroscedasticity, and Blomquist (2010) proposed a bootstrap version of PY's Swamy test that is claimed to be robust to general forms of cross-sectional dependence and serial correlation. Lin (2010) proposed a test for slope homogeneity in a linear panel data models with fixed effects and conditional heteroscedasticity. Jin and Su (2011) proposed a nonparametric test for poolability in nonparametric regression models with a multi-factor error structure. Nevertheless, to the best of our knowledge, there is no available test of slope homogeneity for large dimensional panel data models with interactive fixed effects.

In this paper we consider a test of slope homogeneity in large dimensional panel data models with interactive fixed effects based on a measure of goodness-of-fit (R^2). Under the null hypothesis of homogenous slopes, the residuals from Bai's (2009a) PCA estimation should not contain any useful information about the observable regressors. This motivates us to construct a residual-based test. We first estimate a restricted model by imposing slope homogeneity and adopting Bai's estimation procedure. Then we obtain the cross-sectional R^2's by running the time series regression of residuals on the constant and observable regressors for each cross-sectional unit. Our test statistic $\bar{R}^2$ is constructed as a simple average of these cross-sectional R^2's. Under the null, $\bar{R}^2$ should be close to 0 and deviates away from 0 otherwise. We show that after being appropriately standardized, $\bar{R}^2$ is asymptotically normally distributed under the null hypothesis and a sequence of Pitman local alternatives. We also propose a bootstrap method to obtain the bootstrap p-values to improve the finite sample performance of our test and justify its asymptotic validity. In the Monte Carlo experiments, we show that the test has correct size and satisfactory power. We also compare it with a recent test proposed by PY that neglects cross-sectional dependence in panel data models and Blomquist's (2010) bootstrap version of PY's test. Simulations suggest that the latter test, with or without bootstrapping, has huge size distortions in the presence of cross-sectional dependence. We apply our test to the OECD economic growth data and reject the null of homogeneous slopes. We also apply our test to the Fama-French three factor model to assess how well these three factors can approximate the latent factors in that model.

To sum up, our R^2-based test has several advantages. First, the intuition as detailed above is clear. Like many other goodness-of-fit types of tests in the literature, it is a consistent test and has power in detecting local alternatives converging to the null at the usual $N^{-1/4}T^{-1/2}$ rate which is also obtained by PY. Second, unlike PY's test that requires estimation under both the null and alternative, we only require estimation of the panel data models under the null hypothesis. This is extremely important because Bai's (2009a) PCA estimation is only applicable to homogeneous large dimensional panels with interactive fixed effects. The estimation of the model under the alternative would require us to assume certain rank conditions that are not needed here in order to apply Pesaran's (2006) CCE procedure.

Third, the local asymptotic behavior of our test statistic is tractable. In order to analyze the asymptotic local power property of our test, we need to extend Bai's (2009a) asymptotic distribution theory from the case of homogenous slopes to the case where local deviations from the null are allowed [see eq. (3.2) below]. As demonstrated in the appendix, this extension is nontrivial. The local deviations affect the asymptotic behavior of the estimator of the dominant component, i.e., β in eq. (3.2), in the heterogenous slope parameters and the asymptotic mean of our test statistic in a fairly complicated but tractable manner. Fourth, due to the measurement-unit-free and self-normalizing nature of R^2, our non-normalized test statistic has transparent asymptotic bias and variance formulae, which can be easily estimated.

The remainder of the paper is organized as follows. In Section 2, we introduce the hypotheses and the test statistic. In Section 3 we derive the asymptotic distributions of our test statistic under both the null and a sequence of local Pitman alternatives, and propose a bootstrap procedure to obtain the p-values for our test. We also remark on the other potential applications and extensions of our test. In Section 4, we conduct Monte Carlo experiments to evaluate the finite sample performance of our test and apply it to the OECD economic growth data and the Fama-French three factor model and data. Section 5 concludes. All proofs are relegated to the Appendix.

To proceed, we adopt the following notation. For an $m \times n$ matrix A, we use $||A||$ to denote its Frobenius norm, i.e. $[\text{tr}(A'A)]^{1/2}$. Let $P_A \equiv A(A'A)^{-1}A'$ and $M_A \equiv I_m - A(A'A)^{-1}A'$ where $\equiv$ means "is defined as". When A is a symmetric matrix, we use $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ respectively to denote its maximum and minimum eigenvalues and $A > 0$ to denote that A is positive definite. Let $\mathbf{i}_T$ and I_m denote a $T \times 1$ vector of ones and an $m \times m$ identity matrix, respectively. Let $L \equiv \frac{1}{T}\mathbf{i}_T\mathbf{i}_T'$. We use p.s.d. to abbreviate positive semidefinite. Moreover, the operator $\xrightarrow{p}$ denotes convergence in probability, and $\xrightarrow{d}$ convergence in distribution. We use $(N,T) \to \infty$ to denote the joint convergence of N and T when N and T pass to the infinity simultaneously.

2 Basic Framework

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

2.1 The model and hypotheses

Consider the heterogeneous panel data model with interactive fixed effects

$$Y_{it} = \beta_i' X_{it} + \lambda_i' F_t + \varepsilon_{it},\ i = 1, ..., N,\ t = 1, \dots, T, \tag{2.1}$$

where X_{it} is a $K \times 1$ vector of strictly exogenous regressors, β_i is a $K \times 1$ vector of unknown slope coefficients, λ_i is a $r \times 1$ vector of factor loadings, and F_t is a $r \times 1$ vector of common factors, ε_{it} is idiosyncratic error, and β_i, λ_i, F_t and ε_{it} are unobserved. Here $\{\lambda_i\}$ and $\{F_t\}$ may be potentially correlated with $\{X_{it}\}$. For simplicity, we will assume that ε_{it} satisfy certain martingale difference condition along time dimension t and are independent across the cross sectional dimension i.

The null hypothesis of interest is

$$\mathbb{H}_0\colon\ \beta_i = \beta \text{ for some } \beta \in \mathbb{R}^K\ \forall i = 1, \ldots, N. \tag{2.2}$$

The alternative hypothesis is

$$\mathbb{H}_1\colon\ \beta_i \neq \beta_j \text{ for some } i \neq j. \tag{2.3}$$

To construct a residual-based test for the above null hypothesis, we need to estimate the model under the null hypothesis and obtain the residuals from the regression. Then for each cross sectional unit i, we run the linear regression of the residuals on a constant and X_{it}, and calculate R^2. Our test statistic is constructed by averaging these cross sectional R^2's.

2.2 Estimation of the restricted model

To proceed, we introduce the following notation:

$$Y_i \equiv (Y_{i1}, Y_{i2}, \ldots, Y_{iT})',\ X_i \equiv (X_{i1}, X_{i2}, \ldots, X_{iT})',\ \boldsymbol{\varepsilon}_i \equiv (\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iT})',$$
$$F \equiv (F_1, F_2, \ldots, F_T)',\ \text{and } \Lambda \equiv (\lambda_1, \lambda_2, \ldots, \lambda_N)'.$$

Then under $\mathbb{H}_0$ we can write the model (2.1) in vector form as

$$Y_i = X_i\beta + F\lambda_i + \boldsymbol{\varepsilon}_i, i = 1, ..., N. \tag{2.4}$$

For the restricted model in (2.4), Bai (2009a) studied the PCA estimators of the homogeneous slope β, the factor loadings Λ, and the common factors F, which are given by the solutions of the following set of nonlinear equations

$$\hat{\beta} = \left(\sum_{i=1}^{N} X_i' M_{\hat{F}} X_i\right)^{-1} \sum_{i=1}^{N} X_i' M_{\hat{F}} Y_i, \tag{2.5}$$

$$\left[\frac{1}{NT}\sum_{i=1}^{N}(Y_i - X_i\hat{\beta})(Y_i - X_i\hat{\beta})'\right]\hat{F} = \hat{F}V_{NT}, \tag{2.6}$$

and

$$\hat{\Lambda}' = \frac{1}{T}[\hat{F}'(Y_1 - X_1\hat{\beta}), \ldots, \hat{F}'(Y_N - X_N\hat{\beta})], \tag{2.7}$$

where V_{NT} is a diagonal matrix that consists of the r largest eigenvalues of the bracketed matrix in (2.6), arranged in decreasing order. To obtain the above results, we need to impose some identification restrictions:

$$F'F/T = I_r \text{ and } \Lambda'\Lambda = \text{diagonal}.$$

Bai (2009a) suggested a robust iteration scheme to estimate (β, F, Λ). The procedure goes as follows:

1. Obtain an initial estimator $(\hat{F}, \hat{\Lambda})$ of (F, Λ).

2. Given $\hat{F}$ and $\hat{\Lambda}$, compute

$$\hat{\beta}(\hat{F}, \hat{\Lambda}) = \left(\sum_{i=1}^{N} X_i' X_i\right)^{-1} \sum_{i=1}^{N} X_i'(Y_i - \hat{F}\hat{\lambda}_i).$$

3. Given $\hat{\beta}$, compute $\hat{F}$ according to (2.6) (multiplied by $\sqrt{T}$ due to the restriction that $F'F/T = I_r$) and calculate $\hat{\Lambda}$ using formula (2.7).

4. Repeat steps 2 and 3 until $(\hat{\beta}, \hat{F}, \hat{\Lambda})$ satisfies certain convergence criterion.

After obtaining $(\hat{\beta}, \hat{F}, \hat{\Lambda})$, we can estimate $\boldsymbol{\varepsilon}_i$ by $\hat{\boldsymbol{\varepsilon}}_i = Y_i - X_i\hat{\beta} - \hat{F}\hat{\lambda}_i$ under the null, where $\hat{F} = (\hat{F}_1, \hat{F}_2, \ldots \hat{F}_T)'$, $\hat{\Lambda} = (\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_N)'$ and $\hat{\boldsymbol{\varepsilon}}_i = (\hat{\varepsilon}_{i1}, \hat{\varepsilon}_{i2}, \ldots, \hat{\varepsilon}_{iT})'$. Then it is easy to verify that

$$\hat{\boldsymbol{\varepsilon}}_i = M_{\hat{F}}\boldsymbol{\varepsilon}_i + M_{\hat{F}}X_i(\beta - \hat{\beta}) + M_{\hat{F}}F\lambda_i + M_{\hat{F}}X_i(\beta_i - \beta) \tag{2.8}$$

by noting that $\hat{F}\hat{\lambda}_i = P_{\hat{F}}(Y_i - X_i\hat{\beta})$ according to (2.7).

2.3 A R^2-based test for slope homogeneity

We consider the time series linear regression model

$$\hat{\varepsilon}_{it} = \delta_i + \phi_i' X_{it} + \eta_{it},\ t = 1, ..., T, \tag{2.9}$$

for each cross sectional unit $i = 1, \ldots, N$, where η_{it} is the error term. Under the null hypothesis of homogeneous slopes, we expect $\phi_i = 0$ for all i, because

$$\hat{\varepsilon}_{it} = (\beta - \hat{\beta})' X_{it} + \lambda_i' F_t - \hat{\lambda}_i' \hat{F}_t + \varepsilon_{it}$$

where $\hat{\beta} - \beta \xrightarrow{p} 0$ and $\lambda_i' F_t - \hat{\lambda}_i' \hat{F}_t \xrightarrow{p} 0$ under $\mathbb{H}_0$. Thus the goodness-of-fit measure R_i^2 for the above regression should be close to 0. Under the alternative hypothesis,

$$\hat{\varepsilon}_{it} = (\beta_i - \hat{\beta})' X_{it} + \lambda_i' F_t - \hat{\lambda}_i' \hat{F}_t + \varepsilon_{it}.$$

In general, $\beta_i - \hat{\beta}$ does not converge to 0 in probability and hence R_i^2 should deviate from 0. This enlightens us to propose a test based on an average of cross sectional R_i^2.

Under the null, we first estimate the restricted panel data model with interactive fixed effects

$$Y_i = X_i\beta + F\lambda_i + \boldsymbol{\varepsilon}_i,\ i = 1, ..., N.$$

Then we run the individual time series regression of $\hat{\varepsilon}_{it}$ on $Z_{it}' = (1, X_{it}')$ for $i = 1, \ldots, N$, i.e.,

$$\hat{\boldsymbol{\varepsilon}}_i = Z_i\gamma_i + \boldsymbol{\eta}_i,$$

where $\gamma_i \equiv (\delta_i, \phi_i')'$, $Z_i \equiv (\mathbf{i}_T, X_i)$, and $\boldsymbol{\eta}_i \equiv (\eta_{i1}, \eta_{i2}, \ldots, \eta_{iT})'$. For each cross sectional unit i, we calculate

$$R_i^2 \equiv \frac{ESS_i}{TSS_i} \equiv \frac{\sum_{t=1}^{T}(Z_{it}'\hat{\gamma}_i - \bar{\hat{\varepsilon}}_i)^2}{\sum_{t=1}^{T}(\hat{\varepsilon}_{it} - \bar{\hat{\varepsilon}}_i)^2} = \frac{\hat{\boldsymbol{\varepsilon}}_i'(P_{Z_i} - L)\hat{\boldsymbol{\varepsilon}}_i}{\hat{\boldsymbol{\varepsilon}}_i' M_0 \hat{\boldsymbol{\varepsilon}}_i} \tag{2.10}$$

where $ESS_i \equiv \sum_{t=1}^{T}(Z_{it}'\hat{\gamma}_i - \bar{\hat{\varepsilon}}_i)^2$, $TSS_i \equiv \sum_{t=1}^{T}(\hat{\varepsilon}_{it} - \bar{\hat{\varepsilon}}_i)^2$, $\bar{\hat{\varepsilon}}_i \equiv T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_{it}$, and $M_0 \equiv I_T - L$. We define the average goodness-of-fit for all individual time series regressions as

$$\bar{R}_{NT}^2 = \frac{1}{N}\sum_{i=1}^{N} R_i^2 \tag{2.11}$$

which is used to test for slope homogeneity in the panel data model with interactive fixed effects. Clearly, $0 \le \bar{R}^2_{NT} \le 1$ by construction. It is close to 0 under $\mathbb{H}_0$ because $\{\hat{\varepsilon}_{it}\}$ contains no useful information about $\{X_{it}\}$ and deviates from 0 otherwise. We will show that after being appropriately centered and scaled, $\bar{R}^2_{NT}$ is asymptotically normally distributed under the null and a sequence of Pitman local alternatives.

2.4 Alternative approaches

Alternatively, we can consider estimating the model (2.1) under the null and alternative hypotheses respectively, and comparing the restricted and unrestricted estimators of β_i in the spirit of Hausman test. Nevertheless, Bai's (2009) iterative PCA method is not applicable to heterogenous panel data models and we have to resort to Pesaran's (2006) CCE method to obtain the unrestricted estimators of β_i, $i = 1, ..., N$. The latter method would require that certain rank conditions must be satisfied, which are not needed in this paper.

PY (2008) propose a test of slope homogeneity for large panel data models with fixed effects. Specifically, they consider testing the null that $\beta_i = \beta$ for all i in the following conventional fixed effects panel data model:

$$Y_{it} = \alpha_i + \beta_i' X_{it} + \varepsilon_{it},\ i = 1, ..., N,\ t = 1, \ldots, T. \tag{2.12}$$

To construct their test statistic, one needs to run both restricted and unrestricted regressions. Let

$$\begin{aligned} \hat{\beta}_i &\equiv (X_i' M_0 X_i)^{-1} X_i' M_0 Y_i, \\ \hat{\beta}_{FE} &\equiv \left(\sum_{i=1}^N X_i' M_0 X_i\right)^{-1} \sum_{i=1}^N X_i' M_0 Y_i, \\ \tilde{\beta}_{WFE} &\equiv \left(\sum_{i=1}^N \tilde{\sigma}_{i,PY}^{-2} X_i' M_0 X_i\right)^{-1} \sum_{i=1}^N \tilde{\sigma}_{i,PY}^{-2} X_i' M_0 Y_i, \end{aligned}$$

where $\tilde{\sigma}^2_{i,PY} = (T-1)^{-1} (Y_i - X_i'\hat{\beta}_{FE})' M_0 (Y_i - X_i'\hat{\beta}_{FE})$. PY's standardized Swamy test statistic is

$$\tilde{\Delta}^{PY}_{adj} \equiv \sqrt{\frac{N(T+1)}{T-K-1}} \left(\frac{N^{-1}\tilde{S}^{PY} - K}{\sqrt{2K}}\right), \tag{2.13}$$

where

$$\tilde{S}^{PY} \equiv \sum_{i=1}^N \left(\hat{\beta}_i - \tilde{\beta}_{WFE}\right)' \frac{X_i' M_0 X_i}{\tilde{\sigma}^2_{i,PY}} \left(\hat{\beta}_i - \tilde{\beta}_{WFE}\right). \tag{2.14}$$

PY (2008) prove that $\tilde{\Delta}^{PY}_{adj} \xrightarrow{d} N(0,1)$ under certain regularity conditions.

Here, we can also apply PY's (2008) method to test $\phi_i = \phi = 0$ for all i in (2.9). In this case, we only need to obtain the unrestricted estimate of ϕ_i by $\hat{\phi}_i = (X_i' M_0 X_i)^{-1} X_i' M_0 \hat{\varepsilon}_i$ because the analogue of either $\hat{\beta}_{FE}$ or $\tilde{\beta}_{WFE}$ is given by 0. Let $\tilde{\sigma}^2_i \equiv (T-1)^{-1} \hat{\varepsilon}_i' M_0 \hat{\varepsilon}_i$. Then we can consider the following analogue of $\tilde{S}^{PY}$:

$$\tilde{S} = \sum_{i=1}^N \left(\hat{\phi}_i - 0\right)' \frac{X_i' M_0 X_i}{\tilde{\sigma}^2_i} \left(\hat{\phi}_i - 0\right) = \sum_{i=1}^N \hat{\varepsilon}_i' M_0 X_i \left(X_i' M_0 X_i\right)^{-1} X_i' M_0 \hat{\varepsilon}_i / \tilde{\sigma}^2_i.$$

In view of the fact that $M_0 X_i \left(X_i' M_0 X_i\right)^{-1} X_i' M_0 = P_{Z_i} - L$,[1] it is interesting to see that

$$\tilde{S} = (T-1) \sum_{i=1}^{N} \frac{\hat{\boldsymbol{\varepsilon}}_i' (P_{Z_i} - L) \hat{\boldsymbol{\varepsilon}}_i}{\hat{\boldsymbol{\varepsilon}}_i' M_0 \hat{\boldsymbol{\varepsilon}}_i} = N (T-1) \frac{1}{N} \sum_{i=1}^{N} R_i^2 = N (T-1) \bar{R}_{NT}^2.$$

That is, PY's non-standardized Swamy test statistic $\tilde{S}$ is proportional to our average R^2-test statistic $\bar{R}_{NT}^2$. Unfortunately, PY's standardization method in (2.13) does not work in our setup due to the slow convergence rates of the estimators of the factors and factor loadings.

3 Asymptotic Distributions

In this section we first present a set of assumptions that are necessary for asymptotic analyses, and then study the asymptotic distributions of $\bar{R}_{NT}^2$ under the null hypothesis and a sequence of Pitman local alternatives. We also propose a bootstrap procedure to obtain the bootstrap p-values for our test.

3.1 Assumptions

Let $\mathcal{F} \equiv \{F : F'F/T = I_r\}$. Let $\mathcal{F}_t(\boldsymbol{\varepsilon}_i)$ denote the σ-field generated by $\{\varepsilon_{it}, ..., \varepsilon_{i1}\}$. Let M denote a generic positive constant whose value may change across lines. We make the following assumptions.

Assumption A1. (i) $E||X_{it}||^4 \le M$ and $\inf_{F \in \mathcal{F}} D(F) > 0$, where

$$D(F) = \frac{1}{NT} \sum_{i=1}^{N} X_i' M_F X_i - \frac{1}{T} \left[\frac{1}{N^2} \sum_{i=1}^{N} \sum_{k=1}^{N} X_i' M_F X_k a_{ik} \right] \tag{3.1}$$

and $a_{ik} = \lambda_i' (\Lambda'\Lambda/N)^{-1} \lambda_k$.

(ii) $E||F_t||^4 \le M$ and $T^{-1} \sum_{t=1}^{T} F_t F_t' \xrightarrow{p} \Sigma_F > 0$ for some $r \times r$ matrix Σ_F as $T \to \infty$.

(iii) $E||\lambda_i||^4 \le M$ and $\Lambda'\Lambda/N \xrightarrow{p} \Sigma_\Lambda > 0$ for some $r \times r$ matrix Σ_Λ as $N \to \infty$.

(iv) Let $Z_i \equiv (Z_{i1}, ..., Z_{iT})'$ where $Z_{it} \equiv (1, X_{it}')'$. $T^{-1} Z_i' Z_i \xrightarrow{p} \Sigma_{Z_i} > 0$ for some $(K+1) \times (K+1)$ matrix Σ_{Z_i} as $T \to \infty$. $\min_{1 \le i \le N} \lambda_{\min}(\Sigma_{Z_i}) \ge \underline{c}_Z$ for some $\underline{c}_Z > 0$.

(v) Let $\varsigma_{it} \equiv ||X_{it}||^2 - E||X_{it}||^2$. $\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} E(\varsigma_{it} \varsigma_{jt}) \le M$ and $\frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} E(\varsigma_{it} \varsigma_{is}) \le M$.

(vi) Let $\zeta_i \equiv ||\lambda_i||^2 - E||\lambda_i||^2$. There exists an even number $\vartheta \ge 2$ such that $E||\zeta_i||^\vartheta \le M$, and $\frac{1}{N^{\vartheta/2}} \sum_{1 \le i_1, i_2, ..., i_\vartheta \le N} E\left(\zeta_{i_1} \zeta_{i_2} ... \zeta_{i_\vartheta}\right) \le M$.

Assumption A2. (i) ε_{it} is independent of X_{js}, λ_j, and F_s for all i, t, j and s. $E(\varepsilon_{it}^8) \le M$.

(ii) $\boldsymbol{\varepsilon}_i$, $i = 1, ..., N$, are mutually independent of each other.

Assumption A3. (i) For each i, $\{\varepsilon_{it}, \mathcal{F}_t(\boldsymbol{\varepsilon}_i)\}$ is a martingale difference sequence (m.d.s.) such that $E[\varepsilon_{it} | \mathcal{F}_{t-1}(\boldsymbol{\varepsilon}_i)] = 0$ a.s.

[1] This can be proven by noticing that $Z_i Z_i' = \begin{pmatrix} \mathbf{i}_T' \mathbf{i}_T & \mathbf{i}_T' X_i \\ X_i' \mathbf{i}_T & X_i X_i' \end{pmatrix}$ and using the inverse formula for partitioned matrix. Intuitively, noting that $M_0 X_i \left(X_i' M_0 X_i\right)^{-1} X_i' M_0$ and L are projection matrices associated with the space spanned by the column vectors of $M_0 X_i$ and $\mathbf{i}_T$, respectively, P_{Z_i} is a projection matrix associated with the space spanned by the column vectors of Z_i, or equivalently $[\mathbf{i}_T,\ M_0 X_i,]$, we must have $M_0 X_i \left(X_i' M_0 X_i\right)^{-1} X_i' M_0 + L = P_{Z_i}$ as $(M_0 X_i)' \mathbf{i}_T = 0$.

(ii) $E\left[\varepsilon_{it}^2|\mathcal{F}_{t-1}(\boldsymbol{\varepsilon}_i)\right] = \sigma_i^2$ a.s. such that $\underline{c}_\sigma \leq \min_{1\leq i\leq N}\sigma_i^2 \leq \max_{1\leq i\leq N}\sigma_i^2 \leq \overline{c}_\sigma$ for some $\underline{c}_\sigma, \overline{c}_\sigma \in (0,\infty)$.

(iii) Let $\xi_{it} \equiv \varepsilon_{it}^2 - \sigma_i^2$. $E|\xi_{it}|^\vartheta \leq M$ and $\frac{1}{NT^{\vartheta/2}}\sum_{i=1}^N \sum_{1\leq t_1,t_2,\ldots,t_\vartheta\leq T} E\left(\xi_{it_1}\xi_{it_2}...\xi_{it_\vartheta}\right) \leq M$, where ϑ is given in A1(vi).

Assumption A4. (i) As $(N,T)\to\infty$, $N/T^2 \to 0$, $T/N^{3/2} \to 0$.

(ii) As $(N,T)\to\infty$, $N^{1/2+2/\vartheta}/T \to 0$.

A1(i), A1(ii)-(iii), and A2(i) are identical to Assumptions A, B, and D in Bai (2009a), respectively. They are required for identification and consistent estimation of the parameters in the model. As Moon and Meidner (2010b) show, we can relax A2(i) to allow lagged dependent variables as regressors, but the proof strategy will be totally different from that in Bai (2009a). A1(iv)-(vi) are new and needed to establish the asymptotic distribution of our test statistic. A1(iv) implies that the minimum eigenvalue of $T^{-1}Z_i'Z_i$ is also uniformly bounded below from 0 with probability approaching 1 as $(N,T)\to\infty$. A1(v) and (vi) impose restrictions on the dependence among $\{X_{it}\}$ and $\{\lambda_i\}$. In addition, A1(vi) strengthens the moment conditions of λ_i in A1(iii).

A2(ii) rules out cross sectional dependence between $\boldsymbol{\varepsilon}_i$'s and A3(i) rules out serial dependence among $\{\varepsilon_{it}, t\geq 1\}$. It is worth mentioning that either assumption can be relaxed and neither one causes much further technical difficulty in establishing the asymptotic normal distribution of our test statistic under the null provided the other one is assumed. Nevertheless, given the complicated form of the dominant term in our test statistic, it seems extremely difficult to relax both assumptions simultaneously unless one wants to impose that certain version of central limit theorem (CLT) holds for a complicated U-statistic as in Bai (2009a).[2] In this paper, we impose both assumptions for the simplicity of estimating the asymptotic variance of our test statistic. See Remark 2 after Theorem 3.1.[3]

A3(ii) imposes conditional homoskedasticity along the time dimension but not along the cross sectional dimension. It can be relaxed at the cost of further complication in the derivation of our distributional theory. A3(iii) imposes some moment conditions that are needed to verify the martingale CLT and to prove the consistency of the estimates for the asymptotic bias and variance for our test statistic.

A4 imposes some conditions on the rate at which N and T pass to the infinity, and the interaction of (N,T) with ϑ. A4(i) is also assumed in previous literature on panel data models with interactive fixed effects [e.g., Bai (2009a), Theorem 4]. A4(ii) is new and will be needed to establish the consistency of the estimate of the asymptotic variance of our test statistic. Under A4(ii), $N/T^2 \to 0$ in A4(i) becomes redundant.

[2] Without using either A2(ii) or A3(i), Bai (2009a) assumes instead that a high level CLT holds for some U-statistic under some moment conditions. Nevertheless, the U-statistic that drives the asymptotic normal distribution of our test statistic is much more complicated than that studied by Bai (2009a). Moreover, even though we can assume ad hoc that some CLT holds for our U-statistic, we find that it seems impossible to estimate the asymptotic variance consistently under the null without either A2(ii) or A3(i).

[3] Assumptions A2(i)-(ii) are also implied by Assumptions 1-3 in Pesaran (2006).

3.2 Asymptotic null distribution

Let $h_{i,ts}$ denote the (t,s)th element of $H_i \equiv M_F(P_{Z_i} - L)M_F$. Define

$$\begin{aligned} B_{NT} &\equiv \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{\varepsilon_{it}^2 h_{i,tt}}{T^{-1}TSS_i}, \\ V_{NT} &\equiv \frac{2}{N} \sum_{i=1}^{N} \sum_{1 \le t \ne s \le T} h_{i,ts}^2, \\ J_{NT} &\equiv \sqrt{N} T \bar{R}_{NT}^2 - B_{NT}. \end{aligned}$$

The following theorem states the asymptotic null distribution of the infeasible statistic J_{NT}.

Theorem 3.1 *Suppose Assumptions A1-A4 hold. Then under* $\mathbb{H}_0$,

$$J_{NT} \xrightarrow{d} N(0, V_0)$$

where $V_0 \equiv \lim_{(N,T)\to\infty} V_{NT}$.

Remark 1. The proof of the above theorem is tedious and relegated to the appendix. The key step in the proof is to show that under $\mathbb{H}_0$, $J_{NT} = R_{1NT,1} + o_P(1)$, where

$$R_{1NT,1} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sigma_i^{-2} \sum_{1 \le s \ne t \le T} \varepsilon_{it}\varepsilon_{is} h_{i,ts}.$$

With this, we can apply the martingale central limit theorem (CLT) to show that $R_{1NT,1} \xrightarrow{d} N(0, V_0)$ under Assumptions A1-A4. [4] Note that V_{NT} would be observed if the factor F were observable. In this sense, we can say that J_{NT} is almost asymptotically pivotal. This is one of the advantages to base a test on the measure of goodness-of-fit.

Remark 2. As mentioned above, either Assumption A2(ii) or A3(i) can be relaxed. If we relax A2(ii) to allow for cross sectional dependence among $\boldsymbol{\varepsilon}_i$, we can modify the proof of (B.1) in Appendix B and show that $R_{1NT,1}$ continues to satisfy the martingale CLT under some auxiliary conditions. Specifically, we can replace A2(ii) by the following assumption:

Assumption A5. Let $\boldsymbol{\varepsilon}_{\cdot t} \equiv (\varepsilon_{1t}, ..., \varepsilon_{Nt})'$. Let $\mathcal{F}_{N,t}(\boldsymbol{\varepsilon}_i, \boldsymbol{\varepsilon}_j)$ and $\mathcal{F}_{N,t}(\boldsymbol{\varepsilon})$ denote the σ-field generated by $\{(\varepsilon_{it}, \varepsilon_{jt}), (\varepsilon_{i,t-1}, \varepsilon_{j,t-1}), ..., (\varepsilon_{i1}, \varepsilon_{j1})\}$ and $\{\boldsymbol{\varepsilon}_{\cdot t}, \boldsymbol{\varepsilon}_{\cdot t-1}, \ldots, \boldsymbol{\varepsilon}_{\cdot 1}\}$, respectively.

(i) For all (i,j), $E[\varepsilon_{it}|\mathcal{F}_{N,t-1}(\boldsymbol{\varepsilon}_i, \boldsymbol{\varepsilon}_j)] = 0$ a.s.

(ii) For all (i,j,k,l), $E[\varepsilon_{it}\varepsilon_{jt}|\mathcal{F}_{N,t-1}(\boldsymbol{\varepsilon})] = \omega_{ij}$ a.s. and $E[\varepsilon_{it}\varepsilon_{jt}\varepsilon_{kt}\varepsilon_{lt}|\mathcal{F}_{N,t-1}(\boldsymbol{\varepsilon})] = \kappa_{ijkl}$ a.s. Moreover

$$\frac{1}{N} \sum_{1 \le i,j \le N} |\omega_{ij}| \le M, \ \frac{1}{N^2} \sum_{1 \le i,j,k,l \le N} |\kappa_{ijkl}| \le M, \text{ and } \max_{1 \le i,j \le N} \kappa_{iijj} \le M.$$

[4] Note that $h_{i,ts}$ depends on F and X_i. Without Assumptions A2-A3, we cannot establish this asymptotic normality result.

Then under Assumptions A1, A2(i), A3-A5, we can show that $R_{1NT,1} \xrightarrow{d} N(0, V_1)$, where $V_1 = \lim_{(N,T)\to\infty} V_{1NT}$, and

$$V_{1NT} = \frac{4}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=2}^{T} \sum_{s=1}^{t-1} \sigma_i^{-2} \sigma_j^{-2} \omega_{ij}^2 h_{i,ts} h_{j,ts}.$$

Obviously $V_{1NT} = V_{NT}$ under Assumption A2(ii).

Similarly, if A3(i) is relaxed to allow for some sort of weak dependence among $\{\varepsilon_{it}, t \geq 1\}$ for each i, then under A2(ii), we can apply the CLT for independent heterogenous processes to derive the asymptotic null distribution of our test statistic. In this case, we can replace A3(i) by the following high level assumption:

Assumption A6. Let $\varsigma_{i,ts} \equiv \varepsilon_{it}\varepsilon_{is} - E(\varepsilon_{it}\varepsilon_{is})$ and $\sigma_i^2 \equiv E(\varepsilon_{it}^2)$. σ_i^2 is uniformly bounded and uniformly bounded below from 0. $Q_{NT} \equiv N^{-1} \sum_{i=1}^{N} \sigma_i^{-8} \sum_{1\leq t_1 \neq t_2 \leq T} \sum_{1\leq t_3 \neq t_4 \leq T} \sum_{1\leq t_5 \neq t_6 \leq T} \sum_{1\leq t_7 \neq t_8 \leq T}$ $E(\varsigma_{i,t_1t_2}\varsigma_{i,t_3t_4}\varsigma_{i,t_5t_6}\varsigma_{i,t_7t_8})\, h_{i,t_1t_2} h_{i,t_3t_4} h_{i,t_5t_6} h_{i,t_7t_8} = O_P(1)$.

Noting that each element of $h_{i,ts}$ is $O_P(T^{-1})$, A6 can be satisfied by assuming some strong mixing conditions on the process $\{\varepsilon_{it}, t \geq 1\}$ and it ensures that the Lindeberg condition holds. Under Assumptions A1, A2, A3(iii), A4 and A6, we can show that $R_{1NT,1} - B_{2NT} \xrightarrow{d} N(0, V_2)$, where $B_{2NT} = N^{-1/2} \sum_{i=1}^{N} (T^{-1}TSS_i)^{-1} \sum_{1\leq s\neq t\leq T} E(\varepsilon_{it}\varepsilon_{is})\, h_{i,ts}$ is the bias term to be corrected, $V_2 = \lim_{(N,T)\to\infty} V_{2NT}$, and

$$V_{2NT} = \frac{1}{N} \sum_{i=1}^{N} \sum_{1\leq s\neq t\leq T} \sum_{1\leq r\neq q\leq T} \sigma_i^{-4} \mathrm{Cov}(\varepsilon_{it}\varepsilon_{is}, \varepsilon_{ir}\varepsilon_{iq})\, h_{i,ts} h_{i,rq}.$$

Under Assumptions A3(i)-(ii), straightforward calculations show that $V_{2NT} = V_{NT}$ and $B_{2NT} = 0$.

To implement the test, we need to estimate the asymptotic variance V_{1NT} in the presence of cross sectional dependence, and the asymptotic variance V_{2NT} and the asymptotic bias B_{2NT} in the presence of serial dependence. Unfortunately, the estimation of such objects seems to be a daunting task which is beyond the scope of this paper.[5] For this reason, we restrict our attention to the case where the idiosyncratic error terms ε_{it} exhibit neither cross sectional dependence nor serial dependence. Then we only need consistent estimates of both B_{NT} and V_{NT} defined above.

We propose to estimate B_{NT} by

$$\hat{B}_{NT} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{h}_{i,tt} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \mathrm{tr}(\hat{H}_i)$$

and V_{NT} by

$$\hat{V}_{NT} = \frac{2}{N} \sum_{i=1}^{N} \sum_{1\leq t,s\leq T} \hat{h}_{i,ts}^2 = \frac{2}{N} \sum_{i=1}^{N} \mathrm{tr}(\hat{H}_i^2)$$

[5] In a related framework, Kim (2010) considers the estimation of the asymptotic variance-covariance (VC) of the coefficient estimators in linear panel data models that is robust to both spatial and serial dependence. Nevertheless, the structure of his VC matrix is much simpler than that of our asymptotic variance here.

where $\hat{h}_{i,ts}$ is the (t,s)th element of $\hat{H}_i \equiv M_{\hat{F}}(P_{Z_i} - L)M_{\hat{F}}$. Then we can define a feasible test statistic:

$$\bar{J}_{NT} \equiv \left(\sqrt{N}T\bar{R}_{NT}^2 - \hat{B}_{NT}\right)/\sqrt{\hat{V}_{NT}}.$$

The following corollary establishes the consistency of $\hat{B}_{NT}$ and $\hat{V}_{NT}$ and the asymptotic distribution of $\bar{J}_{NT}$ under $\mathbb{H}_0$.

Corollary 3.2 *Suppose Assumptions A1-A4 hold. Then under* $\mathbb{H}_0$, $\hat{B}_{NT} = B_{NT} + o_P(1)$, $\hat{V}_{NT} = V_{NT} + o_P(1)$, *and* $\bar{J}_{NT} \xrightarrow{d} N(0,1)$.

Remark 3. Corollary 3.2 implies that the test statistic $\bar{J}_{NT}$ is asymptotically pivotal. We can compare $\bar{J}_{NT}$ with the one-sided critical value z_α, i.e., the upper αth percentile from the standard normal distribution, and reject the null when $\bar{J}_{NT} > z_\alpha$ at the asymptotic α significance level.

Remark 4. We obtain the above distributional results despite the fact that the unobserved factors and factor loadings can only be estimated at slower rates ($N^{-1/2}$ for the former and $T^{-1/2}$ for the latter) than that at which the homogeneous slope parameter β can be estimated under the null. The slow convergence rates of these factor and factor loadings estimates do not have adverse asymptotic effects on the estimation of the bias term B_{NT}, the variance term V_{NT}, and the asymptotic distribution of $\bar{J}_{NT}$. Nevertheless, they can play an important role in finite samples. For this reason, we will also propose a bootstrap procedure to obtain the bootstrap p-values for the $\bar{J}_{NT}$ test.

3.3 Asymptotic local power property

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

$$\mathbb{H}_{1,NT} : \beta_i = \beta + N^{-1/4}T^{-1/2}\delta_i \text{ for } i = 1, 2, \ldots, N, \tag{3.2}$$

where the δ_i's are $K \times 1$ vectors of fixed constants such that $\|\delta_i\| < M$ for all i and $\delta_i \neq \delta_j$ for some pair $i \neq j$. Let

$$\begin{aligned} \Theta_0 \quad \equiv \quad & \operatorname*{plim}_{(N,T)\to\infty} \frac{1}{NT}\sum_{i=1}^{N} \sigma_i^{-2} \left\{ X_i\delta_i - X_iD(F)^{-1}\frac{1}{NT}\sum_{k=1}^{N}\Pi_k' X_k\delta_k - \frac{1}{N}\sum_{k=1}^{N} a_{ik}X_k\delta_k \right\}' H_i \\ & \times \left\{ X_i\delta_i - X_iD(F)^{-1}\frac{1}{NT}\sum_{k=1}^{N}\Pi_k' X_k\delta_k - \frac{1}{N}\sum_{k=1}^{N} a_{ik}X_k\delta_k \right\}, \end{aligned} \tag{3.3}$$

where $D(F)^{-1}\frac{1}{NT}\sum_{k=1}^{N}\Pi_k' X_k\delta_k$ can be viewed as a weighted average of δ_k's, and $\frac{1}{N}\sum_{k=1}^{N} a_{ik}X_k\delta_k$ is a weighted average of $X_k\delta_k$, and $\Pi_i \equiv M_F X_i - \frac{1}{N}\sum_{k=1}^{N} a_{ik}M_F X_k$. Clearly, $\Theta_0 = 0$ under the null and is no less than 0 otherwise. See the discussion in Remark 6 below.

In the appendix we show that we can extend the asymptotic analysis of Bai's (2009a) PCA estimator to allow for non-homogeneous slopes, but the extension works only when we consider local deviations from the null hypothesis of homogenous slopes. In particular, we demonstrate that under $\mathbb{H}_{1,NT}$ in (3.2),

$\hat{\beta} - \beta = O_P(N^{-1/4}T^{-1/2})$ and the convergence rates of the estimates of F_t and λ_i (after certain matrix rotation) are the same as those under $\mathbb{H}_0$. With this and some tedious calculations, we can establish the local power property of our test.

Theorem 3.3 *Suppose Assumptions A1-A4 hold. Then the local power of our test satisfies*

$$P\left(\bar{J}_{NT} > z_\alpha | \mathbb{H}_{1,NT}\right) \to 1 - \Phi\left(z_\alpha - \Theta_0/\sqrt{V_0}\right)$$

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

Remark 5. Theorem 3.3 implies that our test has nontrivial asymptotic power against the sequence of local alternatives that deviate from the null at the rate $N^{-1/4}T^{-1/2}$ provided $\Theta_0 > 0$, and the asymptotic local power increases with the magnitude of Θ_0. In this case, as either N or T increases, the power of our test will increase but it is expected to increase faster as $T \to \infty$ than as $N \to \infty$. The rate $N^{-1/4}T^{-1/2}$ is the same as that obtained by PY (2008), indicating that the estimation of factors and factor loadings does not affect the rate at which our test can detect the local alternatives.

Remark 6. The requirement $\Theta_0 > 0$ imposes some restrictions on the degree of slope heterogeneity under the local alternatives, and on the interaction between the heterogeneity parameters δ_i, the observed regressors X_i, and the unobserved factors F. In terms of degree of slope heterogeneity, it requires that β_i and β_j differ from each other for a "large" number of pairs (i,j) with $i \neq j$. In particular, it rules out the case where only a finite fixed number of slope parameters are distinct from a finite number of others (e.g., only β_1 is different from a finite number of others), or the case where the cardinality of the set $\{\beta_1, \beta_2, ..., \beta_N\}$ is diverging to infinity as $N \to \infty$ but at a rate slower than N. In terms of interaction between δ_i, X_i, and F, the expression of Θ_0 in (3.3) is too complicated to analyze. Clearly, the complicated form of Θ_0 arises mainly due to the presence of the unobservable factors (or factor loadings). If F were observable, as in Bai (2009a), the second term in the expression of $D(F)$ in (3.1) and the terms associated with a_{ik} in (3.3) and the definition of Π_i would drop. In this case, Θ_0 reduces to the probability limit of

$$\frac{1}{NT}\sum_{i=1}^{N}\sigma_i^{-2}\left\{X_i\delta_i - X_i D_F^{-1}\frac{1}{NT}\sum_{k=1}^{N}X_k' M_F X_k \delta_k\right\}' H_i \left\{X_i\delta_i - X_i D_F^{-1}\frac{1}{NT}\sum_{k=1}^{N}X_k' M_F X_k \delta_k\right\},$$

where $D_F \equiv \frac{1}{NT}\sum_{i=1}^{N} X_i' M_F X_i$ and $X_i\delta_i - D_F^{-1}\frac{1}{NT}\sum_{k=1}^{N} X_k' M_F X_k \delta_k$ denotes the residual from the $\mathcal{L}_2$ projection of $X_i\delta_i$ on the space spanned by the columns of $M_F X_i$. In the special case where F is absent in the panel data model, then Θ_0 further reduces to the probability limit of

$$\frac{1}{NT}\sum_{i=1}^{N}\sigma_i^{-2}\delta_i' X_i'(P_{Z_i} - L)X_i\delta_i.$$

Noting that $P_{Z_i} - L \geq 0$ as it is a projection matrix, now it becomes transparent that the requirement that the probability limit of the above object is strictly positive does not seem stringent at all.

Remark 7. Under the global alternative $\mathbb{H}_1$, we cannot study the asymptotic properties of Bai's (2009a) PCA estimator because the latter imposes homogeneity on the slope parameters. For this reason, we cannot study the consistency of our test against global alternatives. Even so, we conjecture that the $\bar{J}_{NT}$ test statistic diverges to infinity for fixed alternatives at the rate $N^{1/2}T$ as $(N,T) \to \infty$ provided $\Theta_0 > 0$ in (3.3) where δ_i is now redefined as $\beta_i - \beta$.

3.4 A bootstrap version of the test

As mentioned above, because of the slow convergence rates of the factors and factor loadings estimates, the asymptotic normal null distribution of our test statistic may not approximate its finite sample distribution well in practice. Therefore it is worthwhile to propose a bootstrap procedure to improve the finite sample performance of our test. Below we propose a fixed-regressor bootstrap method to obtain the bootstrap p-values for out test. The procedure goes as follows:

1. Estimate the restricted model in (2.4) and obtain the residuals $\hat{\varepsilon}_{it} = Y_{it} - \hat{\beta}' X_{it} - \hat{\lambda}_i' \hat{F}_t$, where $\hat{\beta}$, $\hat{\lambda}_i$ and $\hat{F}_t$ are obtained by Bai's (2009a) PCA method. Calculate the test statistic $\bar{J}_{NT}$ based on $\{\hat{\varepsilon}_{it}, X_{it}\}$.

2. For $i = 1, ..., N$, obtain the bootstrap error ε_{it}^* randomly from $\left\{\hat{\varepsilon}_{i1} - \bar{\hat{\varepsilon}}_i, \hat{\varepsilon}_{i2} - \bar{\hat{\varepsilon}}_i, \ldots, \hat{\varepsilon}_{iT} - \bar{\hat{\varepsilon}}_i\right\}$. Generate the bootstrap analogue Y_{it}^* of Y_{it} by holding $(X_{it}, \hat{F}_t)$ as fixed: $Y_{it}^* = \hat{\beta}' X_{it} + \hat{\lambda}_i' \hat{F}_t + \varepsilon_{it}^*$ for $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$.

3. Given the bootstrap resample $\{Y_{it}^*, X_{it}\}$, run the restricted model estimation and obtain the bootstrap residuals $\hat{\varepsilon}_{it}^* = Y_{it}^* - \hat{\beta}^{*\prime} X_{it} - \hat{\lambda}_i^{*\prime} \hat{F}_t^*$, where $\hat{\beta}^*$, $\hat{\lambda}_i^*$ and $\hat{F}_t^*$ are obtained by Bai's (2009a) PCA method. Calculate the test statistic $\bar{J}_{NT}^*$ based on $\{\hat{\varepsilon}_{it}^*, X_{it}\}$.

4. Repeat steps 2 and 3 for B times and index the bootstrap test statistics as $\{\bar{J}_{NT,l}^*\}_{l=1}^B$. The bootstrap p-value is calculated by $p^* \equiv B^{-1} \sum_{l=1}^B 1\{\bar{J}_{NT,l}^* > \bar{J}_{NT}\}$, where $1\{\cdot\}$ is the usual indicator function.

Remark 8. It is straightforward to implement the above bootstrap procedure. Note that we impose the null hypothesis of homogeneous slopes in Step 2. In the proof of Theorem 3.4 below, we also verify that Assumptions A2 and A3 are satisfied in the bootstrap world. This is important and will greatly facilitate the proof of the bootstrap's validity.[6]

Remark 9. Even though the asymptotic analysis in Bai (2009a) does not allow predetermined regressor in the model, simulations there indicate that the slope estimators continue to work well with the inclusion of a lagged dependent variable as a regressor. In fact, Moon and Weidner (2010b) allow for lagged dependent variables and demonstrate that in this case the QMLE estimator of the slope coefficient continues to be $\sqrt{NT}$-consistent under some conditions despite some difference in the bias formula. Below we also consider a dynamic panel data model in our simulations, where Y_{it} is generated according to: $Y_{it} = \rho Y_{i,t-1} + \beta' X_{it} + \lambda_i' F_t + \varepsilon_{it}$. Despite the presence of $Y_{i,t-1}$ on the right hand side of the

[6]Kapetanios (2008) considers various resampling scheme for panel data models to account for either cross-sectional dependence or serial dependence. Neither of them is needed here under our assumptions on the idiosyncratic error terms.

last equation, it is well known that we can treat it, like X_{it}, as a fixed regressor in the bootstrap world. In this case, we generate the bootstrap analogue of Y_{it} as follows: $Y_{it}^* = \hat{\rho} Y_{i,t-1} + \hat{\beta}' X_{it} + \hat{\lambda}_i' \hat{F}_t + \varepsilon_{it}^*$.

The following theorem states the main result in this subsection.

Theorem 3.4 *Suppose that Assumptions A1-A4 hold. Then* $\bar{J}_{NT}^* \xrightarrow{d} N(0,1)$ *conditionally on the observed sample* $\mathcal{W}_{NT} \equiv \{(X_1, Y_1), ..., (X_N, Y_N)\}$.

The above theorem shows that the bootstrap provides an asymptotic valid approximation to the limit null distribution of $\bar{J}_{NT}$. This holds as long as we generate the bootstrap data by imposing the null hypothesis. If the null hypothesis does not hold in the observed sample, then we expect $\bar{J}_{NT}$ to explode at the rate $N^{1/4}T^{1/2}$, which delivers the consistency of the bootstrap-based test $\bar{J}_{NT}^*$.

3.5 Discussions and extensions

The focus of this paper is to design a test for the slope homogeneity in large dimensional panel data models with interactive fixed effects. It turns out that our test statistic $\bar{J}_{NT}$ can be used for other testing purposes after minor modifications.

3.5.1 Test of model (2.1) against a pure factor model

First, we can test the specification of the model (2.1) against a pure factor model. Specifically, we can test the null hypothesis

$$\mathbb{H}_0^* : \beta_i = \mathbf{0}_{K\times 1} \text{ for all } i = 1, ..., N,$$

against the alternative hypothesis

$$\mathbb{H}_1^* : \beta_i \neq \mathbf{0}_{K\times 1} \text{ for some } i = 1, ..., N,$$

where $\mathbf{0}_{K\times 1}$ is a $K \times 1$ vector of zeros. Under $\mathbb{H}_0^*$, β_i is a constant that does not vary across i and it is identically equal to 0, implying that the regressor X_{it} has no explanatory power for Y_{it}. Under $\mathbb{H}_1^*$, we may have either heterogeneous slopes or homogeneous non-zero slopes.

There are various locations where such a test is applicable. Here we focus on a potential application to the asset returns in finance. With the advance of the capital asset pricing model (CAPM) and the arbitrary pricing theory (APT), factor models have become one of the most important tools in modern finance. The traditional factor model specifies the excess returns of asset i at time t as

$$R_{it} = \lambda_i' F_t + \eta_{it}$$

where λ_i is a $r \times 1$ vector of factor loadings and F_t is a $r \times 1$ vector of latent factors, and η_{it} is the usual idiosyncratic error term. Even though the development of the asset pricing theory can proceed without a complete specification of how many and what factors are required, empirical testing does not have this luxury. For this reason, some authors [e.g., Lehmann and Modest (1988), Connor and Korajzcyk (1998)] use estimated factors to test the asset pricing theory despite the drawback that the statistically estimated factors do not have immediate economic interpretation. A more popular approach is to rely on economic intuition and theory as a guideline to come up with a list of observed variables/factors G_t to serve as proxies of the unobservable factors F_t. The most eminent example is the three observable risk factors discussed in Fama and French (1993, FF hereafter): the market excess return, small minus big factor, and high minus low factor. Then an appealing question is whether these observable factors are, in fact, the underlying latent factors. In their seminal paper Bai and Ng (2006) considered statistics to determine if the observed and latent factors are exactly the same and applied their tests to assess how well the FF factors and several business cycle indicators can approximate the latent factors in portfolio and stock returns.

Here we offer an alternative approach by considering the following model

$$R_{it} = \beta_i' G_t + \lambda_i' F_t + \varepsilon_{it} \tag{3.4}$$

where G_t denotes a $K \times 1$ vector of observable factors and plays the role of X_{it} in (2.1). Clearly, we cannot estimate the above model by using either Bai's (2009a) PCA method or Pesaran's (2006) CCE method. Nevertheless, as Bai (2009b) demonstrates, the above model is identified under the null

$$\mathbb{H}_{01} : \beta_i = \beta \text{ for all } i = 1, ..., N \tag{3.5}$$

provided $\frac{1}{T} G' M_F G > 0$ where $G \equiv (G_1, G_2, ..., G_T)'$, i.e., there is no multicollinearity between G and $F \equiv (F_1, F_2, ..., F_T)'$. Let $G_{t,k}$ denote the kth element of G_t, $k = 1, ..., K$. If there exists a $r \times 1$ vector α_k such that $G_{t,k} = \alpha_k' F_t$ for all t, we can say that $G_{t,k}$ is an exact factor. If the kth column of G lies in the space spanned by the column vectors of F, which is the case when $G_{t,k}$ is an exact factor, then we cannot estimate the restricted model under $\mathbb{H}_{01}$. This motivates us to consider the following null instead

$$\mathbb{H}_{02} : \beta_i = \mathbf{0}_{K \times 1} \text{ for all } i = 1, ..., N. \tag{3.6}$$

Intuitively speaking, $\mathbb{H}_{02}$ says that given the r latent factors in F_t, the K observable risk factors in G_t are redundant in explaining the asset returns in (3.4). In the case when we reject $\mathbb{H}_{02}$, it means that the r latent factors in F_t cannot span the space of the K observable factors. Various reasons can cause the latter to occur. One reason is that the K observable factors are all relevant but $r < K$. If this is the case, we should observe the change from rejecting $\mathbb{H}_{02}$ to failing to reject $\mathbb{H}_{02}$ as we increase r. Another reason is that the observable factors in G_t are bad proxies for the latent factors. This suggests the importance of testing $\mathbb{H}_{02}$ against its alternative

$$\mathbb{H}_{12} : \beta_i \neq \mathbf{0}_{K \times 1} \text{ for some } i = 1, ..., N.$$

Note that we allow heterogenous factor loadings for the observable factors under $\mathbb{H}_{12}$.

Our $\bar{J}_{NT}$ test can be used to test $\mathbb{H}_{02}$ against $\mathbb{H}_{12}$ with minor modifications. Under $\mathbb{H}_{02}$, we have a pure factor model so that both the latent factors F_t and the factor loadings λ_i can be estimated, say, by $\hat{F}_t$ and $\hat{\lambda}_i$, respectively, via the PCA method. Let $\hat{\varepsilon}_{it} = R_{it} - \hat{\lambda}_i' \hat{F}_t$. Then we can base our $\bar{J}_{NT}$ test on the averaging of the cross sectional R^2's by running the time series least squares linear regression of $\hat{\varepsilon}_{it}$ on $(1, G_t)$. It is easy to see that the asymptotic distribution theory in the above analysis continues to hold in this case.

3.5.2 Test of the linear functional form in (2.1)

We can also test the correct specification of the functional form in (2.1) by considering a nonparametric heterogeneous panel data model with interactive fixed effects

$$Y_{it} = m_i\left(X_{it}\right) + \lambda_i^{'} F_t + \varepsilon_{it},\ i = 1, ..., N,\ t = 1, \ldots, T, \tag{3.7}$$

where $m_i\left(\cdot\right), i = 1, ..., N$, are unknown but smooth functions. The null hypothesis is

$$\mathbb{H}_0^{(1)} : m_i\left(x\right) = \beta_i' x \text{ for all } i = 1, ..., N.$$

Under $\mathbb{H}_0^{(1)}$ and certain rank conditions, we can estimate the heterogeneous linear panel in (2.1) by Pesaran's (2006) CCE method, obtain the residuals and run the time series regression of these residuals on X_{it} nonparametrically to construct a test statistic similar to ours based on the nonparametric goodness-of-fit measure.

Alternatively, we can consider Bai's canonical model

$$Y_{it} = \beta' X_{it} + \lambda_i^{'} F_t + \varepsilon_{it},\ i = 1, ..., N,\ t = 1, \ldots, T, \tag{3.8}$$

and test whether the above linear model is correctly specified. The model under the alternative is obtained by replacing $\beta' X_{it}$ in the above model by $m(X_{it})$, where $m\left(\cdot\right)$ is an unknown but smooth function. In this case, we can obtain the residuals $\hat{\varepsilon}_{it}$ from the model (3.8) and run the panel nonparametric regression of $\hat{\varepsilon}_{it}$ on X_{it} to obtain the nonparametric measure of goodness-of-fit and propose a test based on such a measure. We leave the details for the future research.

4 Monte Carlo Simulations and Applications

In this section, we first conduct a small set of Monte Carlo simulations to evaluate the finite sample performance of our test and then apply the test to the OECD real GDP growth data and asset returns data.

4.1 Simulations

4.1.1 Data generating processes (DGP)

Following Bai (2009a), we use the following two DGPs for level study.

DGP 1:

$$Y_{it} = X_{it,1}\beta_1 + X_{it,2}\beta_2 + \lambda_i' F_t + \varepsilon_{it},\ i = 1, 2, \ldots, N, t = 1, 2, \ldots, T,$$

where $(\beta_1, \beta_2) = (1, 3)$, $\lambda_i = (\lambda_{i1}, \lambda_{i2})'$, $F = (F_{t1}, F_{t2})'$, and the regressors are generated according to

$$\begin{aligned} X_{it,1} &= \mu_1 + c_1 \lambda_i' F_t + \iota' \lambda_i + \iota' F_t + \eta_{it,1}, \\ X_{it,2} &= \mu_2 + c_2 \lambda_i' F_t + \iota' \lambda_i + \iota' F_t + \eta_{it,2}, \end{aligned}$$

with $\iota' = (1, 1)$. Clearly, the regressors are correlated with λ_i and F_t. The variables λ_{ij}, F_{tj}, and $\eta_{it,j}$ are all independently and identically distributed (i.i.d.) $N(0, 4)$, and mutually independent of each other, and the regression errors ε_{it} are i.i.d. $N(0,1)$ and independent of λ_{ij}, F_{tj}, and $\eta_{it,j}$. We set $\mu_1 = \mu_2 = c_1 = c_2 = 1$.

DGP 2:

$$Y_{it} = \rho Y_{i,t-1} + X_{it,1}\beta_1 + X_{it,2}\beta_2 + \lambda_i' F_t + \varepsilon_{it},\ i = 1,2,\ldots,N,\ t = 1,2,\ldots,T,$$

where $(\rho, \beta_1, \beta_2) = (0.75, 1, 3)$, $Y_{i,0} \sim N(0,1)$, $X_{it,1}, X_{it,2}, \lambda_i$, F_t and ε_{it} are generated as in DGP 1.

To evaluate the power performance of our test, we consider the following two DGPs.

DGP 3:

$$Y_{it} = X_{it,1}\beta_{i,1} + X_{it,2}\beta_{i,2} + \lambda_i' F_t + \varepsilon_{it}, i = 1,2,\ldots,N,\ t = 1,2,\ldots,T,$$

where $\beta_{i,1}$ are i.i.d. $N(1, 0.2^2)$, $\beta_{i,2}$ are i.i.d. $N(3, 0.2^2)$ and independent of $\beta_{i,1}$. The generation of other variables in this DGP is the same as in DGP 1.

DGP 4:

$$Y_{it} = \rho_i Y_{i,t-1} + X_{it,1}\beta_{i,1} + X_{it,2}\beta_{i,2} + \lambda_i' F_t + \varepsilon_{it}, i = 1,2,\ldots,N,\ t = 1,2,\ldots,T,$$

where $\beta_{i,1}$, $\beta_{i,2}$, $X_{it,1}$, $X_{it,2}$, λ_i, F_t, and ε_{it} are generated as in DGP 3, and ρ_i are i.i.d. $U(0.70, 0.75)$ and independent of all other parameters or variables on the right hand side of the above the equation.

4.1.2 Test results

We consider two tests of slope homogeneity. The first one is our $\bar{J}_{NT}$ test. The second one is the PY's test. We are interested in seeing how PY's test statistic $\tilde{\Delta}_{adj}^{PY}$ in (2.13) behaves in the panel data models with interactive fixed effects. For comparison purpose, we also consider Blomquist's (2010) bootstrap version of $\tilde{\Delta}_{adj}^{PY}$, which is claimed to be robust to general forms of both cross-sectional dependence and serial correlation. His bootstrap procedure works as follows:

1. Estimate model (2.12) by OLS applied to each i and obtain the residuals $\hat{\varepsilon}_{it}$. For each i, calculate the Bartlett-kernel-based estimator of the autocorrelation-consistent variance, say,

$$\hat{\omega}_i^2 = \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^2 + \frac{2}{T}\sum_{j=1}^{k_i-1}\left(1 - \frac{j}{k_i}\right)\sum_{t=j+1}^{T}\hat{\varepsilon}_{it}\hat{\varepsilon}_{it-j},$$

   where k_i is the bandwidth.

2. Compute the $\tilde{\Delta}_{adj}^{\omega}$ test statistic as in (2.13) by replacing $\tilde{\sigma}_{i,PY}^2$ with $\hat{\omega}_i^2$. At the same time, obtain the residuals in the fixed effects regression for the restricted model of (2.12) under $\beta_i = \beta$. That is, $\tilde{\varepsilon}_{it} = Y_{it} - \hat{\beta}_{FE}' X_{it} - \hat{\alpha}_i$, where $\hat{\alpha}_i = T^{-1}\sum_{t=1}^{T}(Y_{it} - \hat{\beta}_{FE}' X_{it})$. Format $\tilde{\varepsilon}_{it}$ in a $T \times N$ matrix $\tilde{\boldsymbol{\varepsilon}}$.

3. In order to obtain a pseudo panel of errors $\boldsymbol{\varepsilon}^*$, we apply the stationary bootstrap to $\tilde{\boldsymbol{\varepsilon}}$. For $t = 1,2,\ldots,T$, let $\tilde{\boldsymbol{\varepsilon}}_{\cdot t} \equiv (\tilde{\varepsilon}_{1t}, \tilde{\varepsilon}_{2t}, \ldots, \tilde{\varepsilon}_{Nt})'$, and let $B_{tl} = (\tilde{\boldsymbol{\varepsilon}}_{\cdot t}, \tilde{\boldsymbol{\varepsilon}}_{\cdot t+1}, \cdots, \tilde{\boldsymbol{\varepsilon}}_{\cdot t+l-1})'$ be the block of l consecutive estimated errors starting at date t. Sample a sequence of block lengths (say, $l_1, l_2, \ldots$) randomly from a geometric distribution with mean $\bar{l}$ and a sequence of i.i.d. random integers (say, $I_1, I_2, \ldots$) from a discrete uniform distribution on $\{1, 2, \ldots, T\}$. Thus the first l_1 rows of $\boldsymbol{\varepsilon}^*$ are

generated as $B_{I_1 l_1}$ and the next l_2 rows of $\boldsymbol{\varepsilon}^*$ is given by $B_{I_1 l_1}$. The procedure goes on until T rows of $\boldsymbol{\varepsilon}^*$ have been obtained. [7]

4. Generate the bootstrap analog of Y_{it} by holding X_{it} as fixed: $Y_{it}^* = \hat{\beta}_{FE}' X_{it} + \hat{\alpha}_i + \varepsilon_{it}^*$ for $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$, where ε_{it}^* is the (t,i)th element in the matrix $\boldsymbol{\varepsilon}^*$.

5. Given the bootstrap resample $\{Y_{it}^*, X_{it}\}$, compute the bootstrap analogue $\tilde{\Delta}_{adj}^{\omega *}$ of $\tilde{\Delta}_{adj}^{\omega}$ as in step 2.

6. Repeat steps 3-5 for B times and index the bootstrap statistics as $\{\tilde{\Delta}_{adj,l}^{\omega *}\}_{l=1}^{B}$. Calculate The bootstrap p-value as $p^* \equiv B^{-1} \sum_{l=1}^{B} 1\{\tilde{\Delta}_{adj,l}^{\omega *} > \tilde{\Delta}_{adj}^{\omega}\}$.

We consider N, $T = 25$, 50, 100. For each combination of N and T, we consider 1000 simulations for the non-bootstrap version of the test. For the bootstrap version of the test, we use 500 replications for each scenario and $B = 400$ bootstrap resamples for each replication.

Table 1 reports the finite sample rejection frequency of our test and PY's test when the nominal levels are 5% and 10%. We first focus on the non-bootstrap version of the two tests. Table 1 indicates that the levels of both tests are highly distorted, and the distortion tends to increase as N increases for fixed T. For the PY test, the distortion also increases as T increases. For our test, nevertheless, when T is large, the size distortion becomes mild. Now we investigate the bootstrap version of the two tests. It is clear from Table 1 that the use of Blomquist's (2010) bootstrap does not help improve the level behavior of the PY's test at all,[8] and we should not apply the PY test if one doubts the presence of interactive fixed effects. In contrast, the proposed bootstrap can annihilate the oversize issue of the non-bootstrap version of our test.

Table 2 reports the finite sample power performance of our test. We summarize some important findings from Table 2. First, the non-bootstrap version of our test tends to have higher finite sample rejection frequency than the bootstrap version. This is due to the over-sized issue of the former test. Second, for large T, the bootstrap version of our test tends to catch up with the non-bootstrap version of our test in terms of rejection frequency. Third, the finite sample power behavior of the bootstrap version of our test is quite satisfactory in both DGPs 3 and 4. As either N or T increases, the power of our test increases quickly, and it increases faster as T doubles for fixed N than as N doubles for fixed T.

4.2 Two applications

4.2.1 An application to the OECD economic growth data

In this subsection we apply our test to the OECD economic growth data which were analyzed in Zhang, Su and Phillips (2011) for different modelling strategy. The data set consists of four economic variables

[7] We choose $k_i = \lfloor c_1 T^{1/3} \rfloor$ and $\bar{l} = \lfloor c_2 T^{1/3} \rfloor$ in the simulations for different choices of c_1 and c_2, where $\lfloor A \rfloor$ denotes the integer part of A. We find that the results are qualitatively similar for $c_1, c_2 = 1$, 1.5, and 2. To conserve space, we only report the results for $(c_1, c_2) = (1, 1)$ in Table 2 below.

[8] Our simulation result is quite different from Blomquist's (2010). Note that the latter did not demonstrate the asymptotic validity of their bootstrap procedure and his simulation did not allow the correlation between X_{it} and the non-spherical errors which may present either cross-sectional or serial dependence.

Table 1: Finite sample rejection frequency under the null

			Non-bootstrap				Bootstrap			
			Our test		PY's test		Our test		PY's test	
DGP	T	N	5%	10%	5%	10%	5%	10%	5%	10%
1	25	25	16.5	24.5	68.7	77.9	5.0	10.2	49.8	69.0
		50	18.5	28.9	91.9	95.4	5.0	9.0	65.4	79.8
		100	30.3	41.9	99.3	99.6	7.4	10.8	83.4	93.2
	50	25	8.9	14.5	98.1	99.2	4.6	9.4	91.8	95.2
		50	10.6	16.8	100	100	6.4	11.4	98.6	99.4
		100	11.5	18.8	100	100	4.0	7.2	100	100
	100	25	7.8	13.4	100	100	5.6	10.6	100	100
		50	7.4	13.6	100	100	3.8	10.6	100	100
		100	7.0	14.3	100	100	4.0	9.0	100	100
2	25	25	19.1	30.0	52.9	64.9	3.8	10.8	31.0	48.8
		50	26.9	41.3	78.7	87.9	6.2	10.0	38.8	59.4
		100	45.5	58.2	95.8	98.1	6.6	13.8	51.8	76.0
	50	25	11.3	17.7	94.4	96.5	6.0	12.0	79.6	89.4
		50	13.9	21.1	99.5	99.9	8.0	12.0	94.2	97.8
		100	16.6	26.1	100	100	5.2	10.6	99.8	100
	100	25	8.4	13.6	99.9	100	7.4	13.0	99.8	100
		50	6.7	12.6	100	100	3.6	8.4	100	100
		100	8.9	15.9	100	100	6.0	10.4	100	100

Note: PY refers to Pesaran and Yamagata. The bootstrap version of PY's test was studied in Blomquist's (2010).

Table 2: Finite sample rejection frequency under the alternative

		Non-bootstrap				Bootstrap			
		DGP 3		DGP 4		DGP 3		DGP 4	
T	N	5%	10%	5%	10%	5%	10%	5%	10%
25	25	40.7	53.4	54.1	66.6	21.8	32.8	31.0	42.0
	50	64.3	76.3	80.4	88.0	33.6	46.6	53.2	66.2
	100	87.5	92.3	97.7	98.5	54.2	67.6	78.0	87.0
50	25	60.0	69.2	78.6	85.8	49.2	63.0	69.8	80.6
	50	86.0	91.4	96.8	98.6	76.4	86.4	92.8	96.8
	100	98.5	99.7	100	100	95.2	98.2	99.0	99.6
100	25	89.4	93.5	98.6	99.0	86.0	92.8	97.6	98.6
	50	99.1	99.5	99.9	99.9	98.6	99.0	100	100
	100	100	100	100	100	100	100	100	100

for $N = 16$ OECD countries, which are Gross domestic product (GDP), Capital stock (K), Labor input (L), and Human capital (H). The first three are seasonally adjusted quarterly data from 1975Q4 to 2010Q3 ($T = 140$), while we use linear interpolation to obtain the quarterly observations for Human capital as there are only 5-year census data available.

We consider the following two economic growth models:

Model 1:

$$\Delta \ln GDP_{it} = \beta_{i,1}\Delta \ln K_{it} + \beta_{i,2}\Delta \ln L_{it} + \beta_{i,3}\Delta \ln H_{it} + \lambda_i' F_t + \varepsilon_{it},$$

Model 2:

$$\Delta \ln GDP_{it} = \rho_i \Delta \ln GDP_{it-1} + \beta_{i,1}\Delta \ln K_{it} + \beta_{i,2}\Delta \ln L_{it} + \beta_{i,3}\Delta \ln H_{it} + \lambda_i' F_t + \varepsilon_{it},$$

where F_t is a $r \times 1$ vector that represents common shocks such as technological shocks and financial crises, λ_i represents the heterogeneous impact of common shocks on country i, and $\Delta \ln Z_{it} = \ln Z_{it} - \ln Z_{it-1}$ for $Z = GDP$, K, L and H. $\beta_{i,1}, \beta_{i,2}$ and $\beta_{i,3}$ are coefficients of growth rate of K, L, and H respectively. In Model 2, ρ_i represents the impact of previous quarter GDP growth rate on the current one in country i. We are interested in testing for homogeneous coefficients for the 16 OECD countries.

We consider $r = 1, 2, \ldots, 8$ to capture the interactive fixed effects in the growth model.[9] Table 3 reports the test statistics and the bootstrap p-values for our test of slope homogeneity. From the table, we see that the bootstrap p-values for all numbers of factors under investigation are uniformly much smaller than 0.01. So we can reject the null hypothesis of homogeneous slopes at the 1% level for both models. The results imply that the slope homogeneity assumption may not be plausible at all despite the fact it is commonly assumed in the literature.

Table 3: Test statistics and bootstrap p-values for the application to the OECD real GDP growth rate data

Model \ r	1	2	3	4	5	6	7	8
Model 1	25.01	9.22	7.42	8.16	7.74	6.21	5.62	5.05
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0002)	(0.0000)	(0.0005)
Model 2	34.55	21.32	17.33	15.54	14.33	12.38	12.59	10.57
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)

Note: The numbers in braces are bootstrap p-values where the bootstrap number B is 10000.

4.2.2 An application to asset returns

In this application, we test the ability of the FF factors in explaining the excess asset returns in the financial market. FF (1993) proposed three observable risk factors to reflect the excess returns of asset, which are $R_{mt} - R_{ft}$ (the excess return of market portfolio), SMB_t (small market capitalization minus big) and HML_t (high book-to-market ratio minus low). Various empirical studies suggest that these

[9] Alternatively, one can use the information criteria proposed by Bai and Ng (2002) to determine the number r of factors. But it is well known that their criteria tend to fail when the cross sectional unit N is small, which is the case here.

three factors are good proxies for the latent factors in accounting for the excess asset returns. Bai and Ng (2006) developed several tests that can serve as guides as to which observable variables are close to the latent factors in asset returns and concluded that the FF factors can approximate the factors in portfolios and individual stock returns much better than any single macroeconomic variable even though no decisive conclusion is reached.

Here we aim to test the effect of the FF factors on the excess returns of asset when the unobserved factors are added in the model. We consider the following model

$$R_{it} - R_{ft} = \beta_{i,1}(R_{mt} - R_{ft}) + \beta_{i,2} SMB_t + \beta_{i,3} HML_t + \lambda_i' F_t + \varepsilon_{it},\ i = 1, 2, \ldots, N,\ t = 1, 2, \ldots, T, \quad (4.1)$$

where R_{it} is the return of asset i at time t, R_{ft} is the risk-free return rate at time t, F_t is a $r \times 1$ vector of unobservable factor returns and λ_i represents the factor loadings. As explained in Section 3.5.1, we are interested in testing the null hypothesis

$$\mathbb{H}_0 : (\beta_{i,1}, \beta_{i,2}, \beta_{i,3}) = (0, 0, 0). \quad (4.2)$$

If the FF factors are the dominant factors in explaining the excess returns, we expect to reject the null as long as $r \leq 2$ because the FF factors cannot be spanned by the column vectors of $F = (F_1, ..., F_T)'$ in this case. As r increases, however, we should observe the change from rejecting $\mathbb{H}_0$ to failing to reject $\mathbb{H}_0$. On the other hand, if we continuously reject $\mathbb{H}_0$ for sufficiently large r, it means that the FF factors do not lie on the space spanned by the (estimated) large number of latent factors, and they cannot be the dominant factors despite the fact that they have certain power in explaining the excess returns.

We collect monthly data on the excess returns for 100 portfolios and the three FF factors for the period from 1973m1 to 2008m12 from Professor Kenneth French's web site. A total of 97 portfolios are available for two subsamples. To remove the outliers of the return data, we truncate the data using 95% percentile of original data as upper bound and 5% as lower bound. Like Bai and Ng (2006), we standardize the data on the observable factors before the implementation of the test. To minimize the risk of structural change, we consider testing (4.2) for the model (4.1) for four subsamples listed in Table 4, the first three of which are studied in Bai and Ng (2006). We consider the number of unobserved factors $r = 1, 2, \ldots, 10$ in the model and construct the test statistic as detailed at the end of Section 3.5.1. Table 4 reports the test statistics and the corresponding bootstrap p-values. Two features are noteworthy. First, for both the 1988-1996 and 1997-2008 subsamples, our tests suggest that we always reject the null in (4.2) at the 1% nominal level. This questions the use of the three FF factors to approximate the latent factors for these two subsamples. Secondly, for the 1973-1987 subsample we do observe the phenomenon of change of rejection conclusions: for small values of r (≤ 4), we reject the null at the 1% nominal level, which means the three FF factors do not lie in the space spanned by the first four estimated latent factors; but as long as $r > 5$, we fail to reject the null at the 10% nominal level so that the three FF factors do lie in the space spanned by the first six or more estimated factors. In sum, we conclude that the FF factors surely have certain explanatory power in explaining the excess returns, and they do so very well for some subsamples, but may not do so well for other subsamples.

Table 4: Test statistics and bootstrap p-values for the application to asset return data

Subsample \ r	1	2	3	4	5	6	7	8	9	10
1983 − 1996 (T=168, N=100)	195.80 (0.000)	124.01 (0.000)	40.98 (0.000)	15.47 (0.000)	11.98 (0.000)	10.12 (0.000)	5.28 (0.000)	4.57 (0.000)	3.20 (0.040)	2.76 (0.085)
1973 − 1987 (T=180, N=97)	167.29 (0.000)	110.69 (0.000)	8.22 (0.000)	8.22 (0.000)	2.41 (0.049)	2.17 (0.110)	1.64 (0.174)	1.78 (0.170)	1.79 (0.194)	0.43 (0.605)
1988 − 1996 (T=108, N=100)	132.99 (0.000)	82.63 (0.000)	38.74 (0.003)	16.93 (0.000)	14.60 (0.000)	14.34 (0.000)	8.66 (0.000)	8.44 (0.000)	6.90 (0.000)	7.00 (0.000)
1997 − 2008 (T=144, N=97)	207.72 (0.000)	120.50 (0.000)	37.53 (0.000)	16.04 (0.000)	15.10 (0.000)	14.73 (0.000)	7.71 (0.000)	5.57 (0.001)	6.05 (0.000)	4.66 (0.003)

Note: The numbers in braces are bootstrap p-values where the bootstrap number B is 1000.

5 Concluding Remarks

In this paper we propose a R^2-based test for slope heterogeneity in large dimensional panel data models with interactive fixed effects. We first estimate the restricted model to obtain the residuals and run the linear regression of the residuals on a constant and the observable regressors for each cross sectional unit to obtain N measures of R^2. We construct our test statistic by averaging these individual R^2's, and demonstrate that after being appropriately normalized, it is asymptotically normally distributed under the null hypothesis of homogeneous slopes. We show that our test has power to detect Pitman local alternatives at the rate of $T^{-1/2}N^{-1/4}$ and propose a bootstrap procedure to obtain the bootstrap p-values. Simulations demonstrate that the bootstrap version of our test behaves reasonably well in finite samples. The application to the OECD economic growth data indicates that the commonly imposed slope homogeneity assumption is rather fragile. The application to the FF three factor model suggests some other potential applications of our test.

When the null hypothesis of homogeneous slopes is rejected, we may consider applying Pesaran's (2006) CCE method to obtain consistent estimates of both individual slopes and their cross-sectional average under certain rank conditions. If some prior information is available, one can divide the cross sectional units into several groups, test the slope homogeneity within each group, and estimate the homogenous slopes with each individual group in the case of failure of rejection. Alternatively, a panel structure model in the spirit of Sun (2005) may be considered.

APPENDIX

In this appendix we first prove some technical lemmas and then prove the main results in Section 3.

A Some Technical Lemmas

Let $\bar{P}_{Z_i} \equiv P_{Z_i} - L$, $\delta_{NT} \equiv \min[\sqrt{N}, \sqrt{T}]$, and $\gamma_{NT} \equiv N^{-1/4}T^{-1/2}$. By Assumptions A1(i)-(iii) and the Chebyshev inequality, $||X_i|| = O_P(T^{1/2})$ for all i, $(NT)^{-1}\sum_{i=1}^N ||X_i||^2 = O_P(1)$, and $||F|| = O_P(T^{1/2})$. Note that $||\hat{F}|| = T^{1/2}\sqrt{r}$. Let $\mathcal{D} \equiv \{X_1, ..., X_N, F, \Lambda\}$. We use $E_{\mathcal{D}}$ and $\text{Var}_{\mathcal{D}}$ to denote the expectation and variance conditional on $\mathcal{D}$. In addition, we will frequently use the following decomposition:

$$\begin{aligned} M_F - M_{\hat{F}} &= P_{\hat{F}} - P_F \\ &= T^{-1}(\hat{F} - FH)H'F' + T^{-1}(\hat{F} - FH)(\hat{F} - FH)' \\ &\quad + T^{-1}FH(\hat{F} - FH)' + T^{-1}F[HH' - \left(T^{-1}F'F\right)^{-1}]F' \\ &\equiv a_1 + a_2 + a_3 + a_4, \text{ say.} \end{aligned} \tag{A.1}$$

Lemma A.1 *Suppose Assumptions A1-A3 and A4(i) hold. Then under $\mathbb{H}_{1,NT}$ we have*

(i) $T^{-1/2}||\hat{F} - FH|| = O_P(||\hat{\beta} - \beta||) + O_P(\delta_{NT}^{-1}) + O_P(\gamma_{NT})$,

(ii) $T^{-1}\boldsymbol{\varepsilon}_i'(\hat{F} - FH) = T^{-1/2}O_P(||\hat{\beta} - \beta||) + O_P(\delta_{NT}^{-2}) + O_P(T^{-1/2}\gamma_{NT})$ *for all i,*

(iii) $T^{-1}F'(\hat{F} - FH) = O_P(||\hat{\beta} - \beta||) + O_P(\delta_{NT}^{-2}) + O_P(\gamma_{NT})$,

(iv) $T^{-1}Z_i'M_{\hat{F}}(F - \hat{F}H^{-1}) = O_P(||\hat{\beta} - \beta||) + O_P(\delta_{NT}^{-2}) - Z_i'M_{\hat{F}}\frac{\gamma_{NT}}{NT}\sum_{k=1}^N X_k\delta_k\lambda_k\,(\Lambda'\Lambda/N)^{-1}$ *for all i,*

(v) $T^{-1}\mathbf{i}_T'M_{\hat{F}}(F - \hat{F}H^{-1}) = O_P(||\hat{\beta} - \beta||) + O_P(\delta_{NT}^{-2}) - \mathbf{i}_T'M_{\hat{F}}\frac{\gamma_{NT}}{NT}\sum_{k=1}^N X_k\delta_k\lambda_k\,(\Lambda'\Lambda/N)^{-1}$,

(vi) $HH' - (T^{-1}F'F)^{-1} = O_P(||\hat{\beta} - \beta||) + O_P(\delta_{NT}^{-2}) + O_P(\gamma_{NT})$,

(vii) $\left\|P_F - P_{\hat{F}}\right\|^2 = O_P(||\hat{\beta} - \beta||) + O_P(\delta_{NT}^{-2}) + O_P(\gamma_{NT}^2)$,

(viii) $N^{-1}\sum_{i=1}^N \left\|T^{-1}\boldsymbol{\varepsilon}_i'(\hat{F} - FH)\right\|^2 = T^{-1}O_P(||\hat{\beta} - \beta||^2) + O_P(\delta_{NT}^{-4}) + O_P(T^{-1}\gamma_{NT}^2)$,

(ix) $N^{-1}\sum_{i=1}^N T^{-1}Z_i'M_{\hat{F}}(F - \hat{F}H^{-1}) = O_P(||\hat{\beta} - \beta||) + O_P(\delta_{NT}^{-2}) + O_P\left(\gamma_{NT}\right)$,

where $H \equiv (\Lambda'\Lambda/N)(F'\hat{F}/T)V_{NT}^{-1}$ and V_{NT} is defined after (2.7).

Proof. (i) Substituting $Y_i - X_i\hat{\beta} = X_i(\beta - \hat{\beta}) + F\lambda_i + \boldsymbol{\varepsilon}_i + \gamma_{NT}X_i\delta_i$ into (2.6) yields

$$\begin{aligned} &\hat{F}V_{NT} - F\left(\Lambda'\Lambda/N\right)\left(F'\hat{F}/T\right) \\ &\quad = \frac{1}{NT}\sum_{i=1}^N X_i(\beta - \hat{\beta})(\beta - \hat{\beta})'X_i'\hat{F} + \frac{1}{NT}\sum_{i=1}^N X_i(\beta - \hat{\beta})\lambda_i'F'\hat{F} + \frac{1}{NT}\sum_{i=1}^N X_i(\beta - \hat{\beta})\boldsymbol{\varepsilon}_i'\hat{F} \\ &\quad\quad + \frac{1}{NT}\sum_{i=1}^N F\lambda_i(\beta - \hat{\beta})'X_i'\hat{F} + \frac{1}{NT}\sum_{i=1}^N \boldsymbol{\varepsilon}_i(\beta - \hat{\beta})'X_i'\hat{F} + \frac{1}{NT}\sum_{i=1}^N F\lambda_i\boldsymbol{\varepsilon}_i'\hat{F} \\ &\quad\quad + \frac{1}{NT}\sum_{i=1}^N \boldsymbol{\varepsilon}_i\lambda_i'F'\hat{F} + \frac{1}{NT}\sum_{i=1}^N \boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\hat{F} + \frac{\gamma_{NT}}{NT}\sum_{i=1}^N X_i(\beta - \hat{\beta})\delta_i'X_i'\hat{F} \end{aligned}$$

$$
\begin{aligned}
&+\frac{\gamma_{NT}}{NT}\sum_{i=1}^{N}X_i\delta_i(\beta-\hat{\beta})'X_i'\hat{F}+\frac{\gamma_{NT}}{NT}\sum_{i=1}^{N}F\lambda_i\delta_i'X_i'\hat{F}\ +\frac{\gamma_{NT}}{NT}\sum_{i=1}^{N}X_i\delta_i\lambda_i'F'\hat{F}\\
&+\frac{\gamma_{NT}}{NT}\sum_{i=1}^{N}\varepsilon_i\delta_i'X_i'\hat{F}\ +\frac{\gamma_{NT}}{NT}\sum_{i=1}^{N}X_i\delta_i\varepsilon_i'\hat{F}\ +\frac{\gamma_{NT}^2}{NT}\sum_{i=1}^{N}X_i\delta_i\delta_i'X_i'\hat{F}\\
\equiv\ & B_1+B_2+\cdots+B_{15},\ \text{say.} \qquad\qquad (A.2)
\end{aligned}
$$

The first eight terms also appear under $\mathbb{H}_0$ and can be analyzed as in the proof of Proposition A.1 in Bai (2009a). In particular, $T^{-1/2}\|B_l\|=O_P(||\beta-\hat{\beta}||)$ for $l=1,2,\ldots,5$, and $T^{-1/2}\|B_l\|=O_P(\delta_{NT}^{-1})$ for $l=6,7$ and 8. For B_9, using $||\hat{F}||=\sqrt{Tr}$ we have

$$
T^{-1/2}\|B_9\|\leq\frac{\gamma_{NT}}{NT}\sum_{i=1}^{N}||X_i||^2||\beta-\hat{\beta}||\sqrt{r}\,\|\delta_i\|=O_P(\gamma_{NT}||\beta-\hat{\beta}||)=o_P(||\beta-\hat{\beta}||).
$$

Similarly, $T^{-1/2}\|B_{10}\|=o_P(||\beta-\hat{\beta}||)$. For B_{11}, we have

$$
T^{-1/2}\|B_{11}\|\leq\frac{\gamma_{NT}}{N}\sum_{i=1}^{N}\frac{\|F\|}{\sqrt{T}}\|\lambda_i\|\,\|\delta_i\|\,\frac{\|X_i\|}{\sqrt{T}}\sqrt{r}=O_P(\gamma_{NT}).
$$

Similarly, $T^{-1/2}\|B_{12}\|=O_P(\gamma_{NT})$. For B_{13}, we have

$$
T^{-1/2}\|B_{13}\|\leq\gamma_{NT}\left\|\frac{1}{NT}\sum_{i=1}^{N}\varepsilon_i\delta_i'X_i'\right\|\sqrt{r}=O_P(N^{-1/2}\gamma_{NT})
$$

because

$$
\begin{aligned}
E\left\|\frac{1}{NT}\sum_{i=1}^{N}\varepsilon_i\delta_i'X_i'\right\|^2 &= \frac{1}{N^2T^2}\operatorname{tr}\left[\sum_{i=1}^{N}\sum_{j=1}^{N}E\left(\delta_i'X_i'X_j\delta_j\right)E\left(\varepsilon_j'\varepsilon_i\right)\right]\\
&= \frac{1}{N^2}\sum_{i=1}^{N}\frac{E\left\|X_i\delta_i\right\|^2}{T}\sigma_i^2=O\left(N^{-1}\right). \qquad (A.3)
\end{aligned}
$$

Similarly, $T^{-1/2}\|B_{14}\|=O_P(N^{-1/2}\gamma_{NT})$. For B_{15}, we have

$$
T^{-1/2}\|B_{15}\|\leq\frac{\gamma_{NT}^2}{N}\sum_{i=1}^{N}\frac{\|X_i\|^2}{T}\|\delta_i\|^2\sqrt{r}=O_P(\gamma_{NT}^2).
$$

Following the same arguments as used in the proof of Proposition A.1 in Bai (2009a), we obtain

$$
\begin{aligned}
T^{-1/2}||\hat{F}-FH|| &= O_P(||\beta-\hat{\beta}||)+T^{-1/2}\left(B_6+B_7+B_8\right)V_{NT}^{-1}+O_P(\gamma_{NT})\\
&= O_P(||\beta-\hat{\beta}||)+O_P(\delta_{NT}^{-1})+O_P(\gamma_{NT}).
\end{aligned}
$$

(ii) By (A.2), we have the following decomposition:

$$
T^{-1}\varepsilon_k'(\hat{F}-FH)=T^{-1}\varepsilon_k'\left(B_1+B_2+\cdots+B_{15}\right)V_{NT}^{-1}. \qquad (A.4)
$$

The first eight terms can be analyzed as in the proof of Lemma A.4(i) in Bai (2009b) to obtain

$$
T^{-1}\varepsilon_k'\left(B_1+B_2+\cdots+B_8\right)V_{NT}^{-1}=T^{-1/2}O_P(||\hat{\beta}-\beta||)+O_P(\delta_{NT}^{-2}).
$$

For the other terms in (A.4), by the proof of (i) we only need to prove that the dominant terms $T^{-1}\varepsilon_k' B_{11} V_{NT}^{-1}$ and $T^{-1}\varepsilon_k' B_{12} V_{NT}^{-1}$ are $O_P(T^{-1/2}\gamma_{NT})$. For $T^{-1}\varepsilon_k' B_{11} V_{NT}^{-1}$, we have

$$\left\|T^{-1}\varepsilon_k' B_{11} V_{NT}^{-1}\right\| \leq \frac{\gamma_{NT}}{\sqrt{T}} \frac{\|\varepsilon_k' F\|}{\sqrt{T}} \left\| \frac{1}{NT} \sum_{i=1}^{N} \lambda_i \delta_i' X_i' \hat{F} \right\| \left\|V_{NT}^{-1}\right\| = O_P(T^{-1/2}\gamma_{NT})$$

as one can readily show that $\frac{1}{\sqrt{T}} \|\varepsilon_k' F\| = O_P(1)$ and $\frac{1}{NT}\sum_{i=1}^{N} \lambda_i \delta_i' X_i' \hat{F} = O_P(1)$. Similarly, $||T^{-1}\varepsilon_k' \times B_{12} V_{NT}^{-1}|| = O_P(T^{-1/2}\gamma_{NT})$. Thus the result in (ii) follows.

(iii) By (A.2), we have the decomposition

$$T^{-1}F'(\hat{F} - FH) = T^{-1}F'\left(B_1 + B_2 + \cdots + B_{15}\right) V_{NT}^{-1}. \tag{A.5}$$

The first eight terms can be analyzed as in the proof of Lemma A.3(i) in Bai (2009b) to obtain

$$T^{-1}F'\left(B_1 + B_2 + \cdots + B_8\right) V_{NT}^{-1} = O_P(||\hat{\beta} - \beta||) + O_P(\delta_{NT}^{-2}).$$

By the proof of the corresponding terms in (i), we can readily show that $T^{-1}F'B_9V_{NT}^{-1}$ and $T^{-1}F'B_{10}V_{NT}^{-1}$ are $o_P(||\beta - \hat{\beta}||)$, and that $T^{-1}F'B_{11}V_{NT}^{-1}$ and $T^{-1}F'B_{12}V_{NT}^{-1}$ are $O_P(\gamma_{NT})$. For $T^{-1}F'B_{13}V_{NT}^{-1}$, by (A.3) we have

$$\left\|T^{-1}F'B_{13}V_{NT}^{-1}\right\| \leq \gamma_{NT} \frac{\|F\|}{\sqrt{T}} \left\| \frac{1}{NT} \sum_{i=1}^{N} \varepsilon_i \delta_i' X_i' \right\| \sqrt{r} \left\|V_{NT}^{-1}\right\| = O_P(N^{-1/2}\gamma_{NT}).$$

Similarly, $T^{-1}F'B_{13}V_{NT}^{-1} = O_P(N^{-1/2}\gamma_{NT})$. Finally, $\left\|T^{-1}F'B_{15}V_{NT}^{-1}\right\| = T^{-1/2}\|F\| T^{-1/2}\|B_{15}\| \left\|V_{NT}^{-1}\right\| = O_P(\gamma_{NT}^2)$. Thus the result in (iii) follows.

(iv) The proof of (iv) is similar to that of (iii) by using the decomposition in (A.2) to write

$$T^{-1}Z_i' M_{\hat{F}}(F - \hat{F}H^{-1}) = -T^{-1}Z_i' M_{\hat{F}}(B_1 + B_2 + \cdots + B_{15})G, \tag{A.6}$$

where $G \equiv (F'\hat{F}/T)^{-1}(\Lambda'\Lambda/N)^{-1}$. We can readily show that $T^{-1}Z_i' M_{\hat{F}}(B_1 + B_2 + \cdots + B_8)G = O_P(||\hat{\beta} - \beta||) + O_P(\delta_{NT}^{-2})$. For the other terms in (A.6), we only study the two dominant terms that are associated with B_{11} and B_{12}. By the repeated use of the fact that

$$|\mathrm{tr}\,(AB)| \leq \lambda_{\max}(A)\,\mathrm{tr}\,(B) \tag{A.7}$$

for any conformable symmetric matrix A and p.s.d. matrix B and the fact that $\lambda_{\max}\left(M_{\hat{F}}\right) = 1$ (see, e.g., Bernstein, 2005, p. 309), we can show that $||T^{-1}Z_i' M_{\hat{F}}(F - \hat{F}H^{-1})|| \leq T^{-1}||Z_i'(F - \hat{F}H^{-1})|| \leq T^{-1}||Z_i|| \, ||F - \hat{F}H^{-1}||$. Using this and (i) we can show that

$$T^{-1}Z_i' M_{\hat{F}} B_{11} G = \gamma_{NT}\left[T^{-1}Z_i' M_{\hat{F}}\left(F - \hat{F}H^{-1}\right)\right]\left[\frac{1}{NT}\sum_{k=1}^{N} \lambda_k \delta_k' X_k' \hat{F}\right] G = O_P(T^{-1/2}\gamma_{NT}).$$

In addition, $T^{-1}Z_i' M_{\hat{F}} B_{12} G = Z_i' M_{\hat{F}} \frac{\gamma_{NT}}{NT}\sum_{k=1}^{N} X_k \delta_k \lambda_k' (\Lambda'\Lambda/N)^{-1} = O_P(\gamma_{NT})$.

The proof of (v) follows from the proof of (iv) by replacing Z_i with $\mathbf{i}_T$. (vi) and (vii) can be proved by following the proof of Lemmas A.7(i)-(ii) in Bai (2009b). The proofs of (viii) and (ix) follow closely from those of (ii) and (iv), respectively. ■

Lemma A.2 *Suppose Assumptions A1-A3 and A4(i) hold. Then under* $\mathbb{H}_{1,NT}$ *we have* $\hat{\beta} - \beta = D(F)^{-1}\frac{\gamma_{NT}}{NT}\sum_{i=1}^{N}\Pi_i' X_i \delta_i + o_P(\gamma_{NT})$, *where* $\Pi_i = M_F X_i - \frac{1}{N}\sum_{k=1}^{N} a_{ik} M_F X_k$.

Proof. By (2.5) and using $Y_i = X_i\beta + F\lambda_i + \boldsymbol{\varepsilon}_i + \gamma_{NT} X_i \delta_i$ under $\mathbb{H}_{1,NT}$, we have

$$\left(\frac{1}{NT}\sum_{i=1}^{N} X_i' M_{\hat{F}} X_i\right)(\hat{\beta} - \beta) = \frac{1}{NT}\sum_{i=1}^{N} X_i' M_{\hat{F}} F\lambda_i + \frac{1}{NT}\sum_{i=1}^{N} X_i' M_{\hat{F}} \boldsymbol{\varepsilon}_i + \frac{\gamma_{NT}}{NT}\sum_{i=1}^{N} X_i' M_{\hat{F}} X_i \delta_i. \quad \text{(A.8)}$$

First, by (A.2) the first term on the right hand side of (A.8) can be decomposed as follows:

$$\begin{aligned}\frac{1}{NT}\sum_{i=1}^{N} X_i' M_{\hat{F}} F\lambda_i &= \frac{1}{NT}\sum_{i=1}^{N} X_i' M_{\hat{F}}(F - \hat{F}H^{-1})\lambda_i \\ &= -\frac{1}{NT}\sum_{i=1}^{N} X_i' M_{\hat{F}}\left(B_1 + B_2 + \cdots + B_{15}\right) G\lambda_i \\ &\equiv C_1 + C_2 + \cdots + C_{15}, \text{ say,}\end{aligned}$$

where recall $G \equiv (F'\hat{F}/T)^{-1}(\Lambda'\Lambda/N)^{-1}$. The first eight terms can be analyzed as in the proof of Proposition A.2 of Bai (2009a). In particular, $C_l = o_P(||\hat{\beta} - \beta||)$ for $l = 1, 3, 4, 5$,

$$\begin{aligned} C_2 &= \frac{1}{T}\left[\frac{1}{N}\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{N} X_i' M_{\hat{F}} X_k a_{ik}\right](\hat{\beta} - \beta), \\ C_6 &= o_P(||\hat{\beta} - \beta||) + o_P(1/\sqrt{NT}) + O_P(N^{-1}\delta_{NT}^{-2}) + N^{-1/2} O_P(\delta_{NT}^{-4}), \\ C_7 &= -\frac{1}{N^2 T}\sum_{i=1}^{N}\sum_{k=1}^{N} a_{ik} X_i' M_{\hat{F}} \varepsilon_k, \text{ and} \\ C_8 &= A_{NT} + o_P(1/\sqrt{NT}) + o_P(\hat{\beta} - \beta) + N^{-1/2} O_P(\delta_{NT}^{-2}),\end{aligned}$$

where $A_{NT} \equiv -\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\sigma_k^2 X_i' M_{\hat{F}}\hat{F}\ (F'\hat{F}/T)^{-1}(\Lambda'\Lambda/N)^{-1}\lambda_i$. C_9 and C_{10} are bounded in the Euclidean norm by $o_P(1)||\hat{\beta} - \beta||$. For C_{11}, as in the proof of Lemma A.1(iv) we have

$$C_{11} = -\frac{\gamma_{NT}}{N^2T^2}\sum_{i=1}^{N} X_i' M_{\hat{F}}\left(F - \hat{F}H^{-1}\right)\sum_{j=1}^{N}\lambda_j \delta_j' X_j' \hat{F} G \lambda_i = O_P(T^{-1/2}\gamma_{NT}).$$

For C_{12}, we have

$$C_{12} = -\frac{\gamma_{NT}}{N^2 T}\sum_{i=1}^{N}\sum_{j=1}^{N} X_i' M_{\hat{F}} X_j \delta_j \lambda_j' \left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\lambda_i = -\frac{\gamma_{NT}}{N^2 T}\sum_{i=1}^{N}\sum_{k=1}^{N} X_i' M_{\hat{F}} X_k \delta_k a_{ik} = O_P(\gamma_{NT}).$$

Moreover, we can show that C_{13} and C_{14} are $O_P(N^{-1/2}\gamma_{NT})$ and C_{15} is $O_P(\gamma_{NT}^2)$. Thus, we obtain

$$\frac{1}{NT}\sum_{i=1}^{N} X_i' M_{\hat{F}} F\lambda_i = C_2 + C_7 + A_{NT} + o_P(||\hat{\beta} - \beta||) + C_{12} + O_P(\delta_{NT}^{-1}\gamma_{NT}). \quad \text{(A.9)}$$

The last term in (A.8) is $O_P(\gamma_{NT})$. Thus, combining (A.8) and (A.9) yields

$$\begin{aligned}\left(\frac{1}{NT}\sum_{i=1}^{N} X_i' M_{\hat{F}} X_i + o_P(1)\right)(\hat{\beta} - \beta) - C_2 &= \frac{1}{NT}\sum_{i=1}^{N} X_i' M_{\hat{F}}\boldsymbol{\varepsilon}_i + C_7 + A_{NT} + o_P(||\hat{\beta} - \beta||) \\ &\quad + C_{12} + O_P(\delta_{NT}^{-1}\gamma_{NT}) + \frac{\gamma_{NT}}{NT}\sum_{i=1}^{N} X_i' M_{\hat{F}} X_i \delta_i.\end{aligned}$$

Observing that $\gamma_{NT}^{-1}A_{NT} = o_P(1)$ and $\gamma_{NT}^{-1}(N^{-1}T^{-1}\sum_{i=1}^{N} X_i' M_{\hat{F}}\varepsilon_i + C_7) = o_P(1)$, multiplying both sides of the above equation by γ_{NT}^{-1} yields

$$[D(\hat{F}) + o_P(1)]\gamma_{NT}^{-1}(\hat{\beta} - \beta) = \frac{1}{NT}\sum_{i=1}^{N}\left[X_i' M_{\hat{F}} - \frac{1}{N}\sum_{k=1}^{N} a_{ik} X_k' M_{\hat{F}}\right] X_i\delta_i + o_P(1).$$

It can be shown that $D(\hat{F}) = D(F) + o_P(1)$ and

$$\frac{1}{NT}\sum_{i=1}^{N}\left[X_i' M_{\hat{F}} - \frac{1}{N}\sum_{k=1}^{N} a_{ik} X_k' M_{\hat{F}}\right] X_i\delta_i = \frac{1}{NT}\sum_{i=1}^{N}\left[X_i' M_F - \frac{1}{N}\sum_{k=1}^{N} a_{ik} X_k' M_F\right] X_i\delta_i + o_P(1).$$

Thus we have

$$\gamma_{NT}^{-1}(\hat{\beta} - \beta) = D(F)^{-1}\frac{1}{NT}\sum_{i=1}^{N}\left[X_i' M_F - \frac{1}{N}\sum_{k=1}^{N} a_{ik} X_k' M_F\right] X_i\delta_i + o_P(1).$$

That is, $\hat{\beta} - \beta = D(F)^{-1}\frac{\gamma_{NT}}{NT}\sum_{i=1}^{N}\Pi_i' X_i\delta_i + o_P(\gamma_{NT})$, where $\Pi_i = M_F X_i - \frac{1}{N}\sum_{k=1}^{N} a_{ik} M_F X_k$ and $D(F) = \frac{1}{NT}\sum_{i=1}^{N}\Pi_i'\Pi_i$. ■

Lemma A.3 *Suppose Assumptions A1-A3 and A4(i) hold. Then under $\mathbb{H}_{1,NT}$ we have*

(i) $\Gamma_{1NT} \equiv N^{-1/2}\sum_{i=1}^{N}\sigma_i^{-2}\varepsilon_i'(M_{\hat{F}} - M_F)\bar{P}_{Z_i}(M_{\hat{F}} - M_F)\varepsilon_i = o_P(1)$,

(ii) $\Gamma_{2NT} \equiv \left\|N^{-1/2}\sum_{i=1}^{N}\sigma_i^{-2}\lambda_i\varepsilon_i' M_F\bar{P}_{Z_i}\right\| = O_P(1)$,

(iii) $\Gamma_{3NT} \equiv \left\|N^{-1/2}\sum_{i=1}^{N}\sigma_i^{-2}\varepsilon_i\varepsilon_i' M_F\bar{P}_{Z_i}\right\| = O_P(\sqrt{N+T})$,

(iv) $\Gamma_{4NT} \equiv \left\|N^{-1/2}\sum_{i=1}^{N}\sigma_i^{-2}F'\varepsilon_i\varepsilon_i' M_F\bar{P}_{Z_i}\right\| = O_P(\sqrt{N+T})$,

(v) $\Gamma_{5NT} \equiv N^{-1/2}\sum_{i=1}^{N}\sigma_i^{-2}\varepsilon_i' M_F\bar{P}_{Z_i} M_F(F - \hat{F}H^{-1})\lambda_i = o_P(1)$,

(vi) $\Gamma_{6NT} \equiv N^{-1/2}\sum_{j=1}^{N}\sigma_i^{-2}\varepsilon_i' M_F\bar{P}_{Z_i} T^{-1} FH(\hat{F} - FH)'\varepsilon_i = o_P(1)$.

Proof. (i) Noting that $\Gamma_{1NT} \le \underline{c}_\sigma^{-1}\bar{\Gamma}_{1NT}$ where $\bar{\Gamma}_{1NT} = N^{-1/2}\sum_{i=1}^{N}\varepsilon_i'(M_{\hat{F}} - M_F)\bar{P}_{Z_i}(M_{\hat{F}} - M_F)\varepsilon_i$, we prove (i) by showing that $\bar{\Gamma}_{1NT} = o_P(1)$. Using (A.1), we can decompose $\bar{\Gamma}_{1NT}$ as follows

$$\begin{aligned}\bar{\Gamma}_{1NT} &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\varepsilon_i'(a_1 + a_2 + a_3 + a_4)'\bar{P}_{Z_i}(a_1 + a_2 + a_3 + a_4)\varepsilon_i \\ &= \bar{\Gamma}_{1NT,1} + \bar{\Gamma}_{1NT,2} + \cdots + \bar{\Gamma}_{1NT,10},\end{aligned}$$

where

$$\begin{aligned}
&\bar{\Gamma}_{1NT,1} \equiv \tfrac{1}{\sqrt{N}}\sum_{i=1}^{N}\varepsilon_i' a_1'\bar{P}_{Z_i} a_1\varepsilon_i, && \bar{\Gamma}_{1NT,6} \equiv \tfrac{2}{\sqrt{N}}\sum_{i=1}^{N}\varepsilon_i' a_1'\bar{P}_{Z_i} a_3\varepsilon_i,\\
&\bar{\Gamma}_{1NT,2} \equiv \tfrac{1}{\sqrt{N}}\sum_{i=1}^{N}\varepsilon_i' a_2'\bar{P}_{Z_i} a_2\varepsilon_i, && \bar{\Gamma}_{1NT,7} \equiv \tfrac{2}{\sqrt{N}}\sum_{i=1}^{N}\varepsilon_i' a_1'\bar{P}_{Z_i} a_4\varepsilon_i,\\
&\bar{\Gamma}_{1NT,3} \equiv \tfrac{1}{\sqrt{N}}\sum_{i=1}^{N}\varepsilon_i' a_3'\bar{P}_{Z_i} a_3\varepsilon_i, && \bar{\Gamma}_{1NT,8} \equiv \tfrac{2}{\sqrt{N}}\sum_{i=1}^{N}\varepsilon_i' a_2'\bar{P}_{Z_i} a_3\varepsilon_i,\\
&\bar{\Gamma}_{1NT,4} \equiv \tfrac{1}{\sqrt{N}}\sum_{i=1}^{N}\varepsilon_i' a_4'\bar{P}_{Z_i} a_4\varepsilon_i, && \bar{\Gamma}_{1NT,9} \equiv \tfrac{2}{\sqrt{N}}\sum_{i=1}^{N}\varepsilon_i' a_2'\bar{P}_{Z_i} a_4\varepsilon_i,\\
&\bar{\Gamma}_{1NT,5} \equiv \tfrac{2}{\sqrt{N}}\sum_{i=1}^{N}\varepsilon_i' a_1'\bar{P}_{Z_i} a_2\varepsilon_i, && \bar{\Gamma}_{1NT,10} \equiv \tfrac{2}{\sqrt{N}}\sum_{i=1}^{N}\varepsilon_i' a_3'\bar{P}_{Z_i} a_4\varepsilon_i.
\end{aligned}$$

We first consider $\bar{\Gamma}_{1NT,1}$. Noting that $\bar{P}_{Z_i} = P_{Z_i} - L$ is a projection matrix, $\lambda_{\max}(\bar{P}_{Z_i}) = 1$ and

$$\begin{aligned}\bar{\Gamma}_{1NT,1} &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(T^{-1}(\hat{F}-FH)H'F'\boldsymbol{\varepsilon}_i\right)'\bar{P}_{Z_i}T^{-1}(\hat{F}-FH)H'F'\boldsymbol{\varepsilon}_i \\ &\leq \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_{\max}(\bar{P}_{Z_i})||T^{-1}(\hat{F}-FH)H'F'\boldsymbol{\varepsilon}_i||^2 \\ &\leq ||\hat{F}-FH||^2||H||^2\frac{1}{T^2\sqrt{N}}\sum_{i=1}^{N}||F'\boldsymbol{\varepsilon}_i||^2 \\ &= O_P(\max[T/N,1])O_P(N^{1/2}T^{-1}) = o_P(1),\end{aligned}$$

by Lemma A.1 (i), Assumption A4(i), and the fact that $\sum_{i=1}^{N}E||F'\boldsymbol{\varepsilon}_i||^2 = O(NT)$ under Assumptions A1(ii) and A2. For $\bar{\Gamma}_{1NT,2}$, using Lemmas A.1(i) and (viii), Lemma A.2 and Assumption A4(i) we have

$$\begin{aligned}\bar{\Gamma}_{1NT,2} &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[T^{-1}(\hat{F}-FH)(\hat{F}-FH)'\boldsymbol{\varepsilon}_i\right]'\bar{P}_{Z_i}T^{-1}(\hat{F}-FH)(\hat{F}-FH)'\boldsymbol{\varepsilon}_i \\ &\leq \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_{\max}(\bar{P}_{Z_i})||T^{-1}(\hat{F}-FH)(\hat{F}-FH)'\boldsymbol{\varepsilon}_i||^2 \\ &\leq \sqrt{N}||\hat{F}-FH||^2\left[\frac{1}{N}\sum_{i=1}^{N}||T^{-1}(\hat{F}-FH)'\boldsymbol{\varepsilon}_i||^2\right] \\ &= \sqrt{N}O_P(\max[T/N,1])O_P(\max[N^{-2},T^{-2}]) = o_P(1).\end{aligned}$$

Similarly,

$$\begin{aligned}\bar{\Gamma}_{1NT,3} &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[T^{-1}FH(\hat{F}-FH)'\boldsymbol{\varepsilon}_i\right]'\bar{P}_{Z_i}T^{-1}FH(\hat{F}-FH)'\boldsymbol{\varepsilon}_i \\ &\leq \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_{\max}(\bar{P}_{Z_i})||T^{-1}FH(\hat{F}-FH)'\boldsymbol{\varepsilon}_i||^2 \\ &\leq \sqrt{N}||FH||^2\left[\frac{1}{N}\sum_{i=1}^{N}||T^{-1}(\hat{F}-FH)'\boldsymbol{\varepsilon}_i||^2\right] \\ &= \sqrt{N}O_P(T)O_P(\max[N^{-2},T^{-2}]) = o_P(1).\end{aligned}$$

For $\bar{\Gamma}_{1NT,4}$, using Lemma A.1 (vi) and A.2 and Assumption A4(i) we have

$$\begin{aligned}\bar{\Gamma}_{1NT,4} &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{T^{-1}F\left[HH'-\left(T^{-1}F'F\right)^{-1}\right]F'\boldsymbol{\varepsilon}_i\right\}'\bar{P}_{Z_i}\left\{T^{-1}F\left[HH'-\left(T^{-1}F'F\right)^{-1}\right]F'\boldsymbol{\varepsilon}_i\right\} \\ &\leq \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_{\max}(\bar{P}_{Z_i})\left\|T^{-1}F\left[HH'-\left(T^{-1}F'F\right)^{-1}\right]F'\boldsymbol{\varepsilon}_i\right\|^2 \\ &\leq N^{1/2}\left(\frac{1}{T}||F||^2\right)\left[\left\|HH'-\left(T^{-1}F'F\right)^{-1}\right\|^2\right]\left(\frac{1}{NT}\sum_{i=1}^{N}||F'\boldsymbol{\varepsilon}_i||^2\right) \\ &= N^{1/2}O_P(1)O_P(N^{-1/2}T^{-1})O_P(1) = o_P(1).\end{aligned}$$

By the above results and the Cauchy-Schwarz inequality, $\bar{\Gamma}_{1NT,l} = o_P(1)$ for $l = 5, 6, ..., 10$. It follows that $\bar{\Gamma}_{1NT} = o_P(1)$.

(ii) Observe that $E_{\mathcal{D}}(N^{-1/2}\sum_{i=1}^N \sigma_i^{-2}\lambda_i\varepsilon_i' M_F\bar{P}_{Z_i}) = 0$, where recall that $E_{\mathcal{D}}$ denotes expectation conditional on $\mathcal{D} \equiv \{X_1, ..., X_N, F, \Lambda\}$. By the repeated use of the fact in (A.7) and the fact that $\lambda_{\max}(M_F) = 1$, we have

$$\begin{aligned} E_{\mathcal{D}}\left(\Gamma_{2NT}^2\right) &= \frac{1}{N}\mathrm{tr}\left[\sum_{i=1}^N\sum_{j=1}^N \sigma_i^{-2}\sigma_j^{-2}\lambda_i'\lambda_j M_F\bar{P}_{Z_j}\bar{P}_{Z_i}M_F E\left(\varepsilon_i\varepsilon_j'\right)\right] \\ &= \frac{1}{N}\mathrm{tr}\left(\sum_{i=1}^N \sigma_i^{-2}\lambda_i'\lambda_i M_F\bar{P}_{Z_i}M_F\right) \leq (K+1)\underline{c}_\sigma^{-1}\frac{1}{N}\sum_{i=1}^N\|\lambda_i\|^2 = O_P(1). \end{aligned}$$

Therefore $\Gamma_{2NT} = O_P(1)$ by the conditional Chebyshev inequality.

(iii) Noting that $E(\varepsilon_i\varepsilon_i'\varepsilon_j\varepsilon_j') = \varpi_{ij}I_T$ where $\varpi_{ij} = (T-1)\,\omega_{ij}^2 + E(\varepsilon_{it}^2\varepsilon_{jt}^2)$ and $\omega_{ij} = E(\varepsilon_{it}\varepsilon_{jt}) = \sigma_i^2 1\{i=j\}$, we have by arguments analogous to those used in the study of Γ_{2NT},

$$\begin{aligned} E_{\mathcal{D}}\left(\Gamma_{3NT}^2\right) &= \mathrm{tr}\left[\frac{1}{N}\sum_{j=1}^N\sum_{i=1}^N \sigma_i^{-2}\sigma_j^{-2}M_F\bar{P}_{Z_j}\bar{P}_{Z_i}M_F E\left(\varepsilon_i\varepsilon_i'\varepsilon_j\varepsilon_j'\right)\right] \\ &= \mathrm{tr}\left(M_F\frac{1}{N}\sum_{j=1}^N\sum_{i=1}^N \sigma_i^{-2}\sigma_j^{-2}\varpi_{ij}\bar{P}_{Z_j}\bar{P}_{Z_i}\right) \leq \frac{1}{N}\underline{c}_\sigma^{-2}\sum_{j=1}^N\sum_{i=1}^N\varpi_{ij}\mathrm{tr}\left(\bar{P}_{Z_j}\bar{P}_{Z_i}\right) \\ &\leq (K+1)\underline{c}_\sigma^{-2}\frac{1}{N}\sum_{j=1}^N\sum_{i=1}^N\varpi_{ij} = O(N+T). \end{aligned}$$

Thus $\Gamma_{3NT} = O_P(\sqrt{N+T})$.

(iv) Analogously to the proof of (iii), we can show that $\Gamma_{4NT} = O_P(\sqrt{N+T})$.

(v) By the proof of Lemma A.1(i), we have

$$\begin{aligned} F - \hat{F}H^{-1} &= -\left[B_1 + B_2 + \cdots + B_{15}\right]G \\ &= T^{1/2}O_P(\beta - \hat{\beta}) + T^{1/2}O_P(\gamma_{NT}) - \left[B_6 + B_7 + B_8\right]G \\ &\equiv G_1 + G_2 - G_3, \text{ say,} \end{aligned} \tag{A.10}$$

where recall $G \equiv (F'\hat{F}/T)^{-1}(\Lambda'\Lambda/N)^{-1}$. Thus we have $\Gamma_{5NT} = \Gamma_{5NT,1} + \Gamma_{5NT,2} - \Gamma_{5NT,3}$, where $\Gamma_{5NT,l} = N^{-1/2}\sum_{i=1}^N \sigma_i^{-2}\varepsilon_i' M_F\bar{P}_{Z_i}M_F G_l\lambda_i$, $l = 1, 2, 3$. For $\Gamma_{5NT,1}$, by (ii) we have

$$\begin{aligned} \Gamma_{5NT,1} &= \mathrm{tr}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^N \sigma_i^{-2}\lambda_i\varepsilon_i' M_F\bar{P}_{Z_i}M_F G_1\right) \\ &\leq \left\|\frac{1}{\sqrt{N}}\sum_{i=1}^N \sigma_i^{-2}\lambda_i\varepsilon_i' M_F\bar{P}_{Z_i}M_F\right\| T^{1/2}O_P(||\beta - \hat{\beta}||) \\ &= O_P(1)O_P(T^{1/2}||\beta - \hat{\beta}||) = o_P(1). \end{aligned}$$

By the same token, $\Gamma_{5NT,2}=O_P(T^{1/2}\gamma_{nT})=o_P(1)$. Now we decompose $\Gamma_{5NT,3}$ as follows

$$\begin{aligned}\Gamma_{5NT,3} &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sigma_i^{-2}\varepsilon_i' M_F\bar{P}_{Z_i}M_F\left[B_6+B_7+B_8\right]G\lambda_i \\ &\equiv \Gamma_{5NT,31}+\Gamma_{5NT,32}+\Gamma_{5NT,33},\ \text{say.}\end{aligned}$$

Obviously, $\Gamma_{5NT,31}=0$ as $M_F F=0$. For $\Gamma_{5NT,32}$, we have

$$\begin{aligned}\Gamma_{5NT,32} &= \frac{1}{N^{3/2}}\sum_{i=1}^{N}\sum_{j=1}^{N}\sigma_i^{-2}\varepsilon_i' M_F\bar{P}_{Z_i}M_F\varepsilon_j a_{ij}. \\ &= \frac{1}{N^{3/2}}\sum_{i=1}^{N}\sigma_i^{-2}\varepsilon_i' M_F\bar{P}_{Z_i}M_F\varepsilon_i a_{ii}+\frac{1}{N^{3/2}}\sum_{1\le i\neq j\le N}\sigma_i^{-2}\varepsilon_i' M_F\bar{P}_{Z_i}M_F\varepsilon_j a_{ij} \\ &\equiv \Gamma_{5NT,321}+\Gamma_{5NT,322},\ \text{say.}\end{aligned}$$

By the repeated use of (A.7),

$$\begin{aligned}E_{\mathcal{D}}\left|\Gamma_{5NT,321}\right| &= \frac{1}{N^{3/2}}\sum_{i=1}^{N}\sigma_i^{-2}\text{tr}\left[E_{\mathcal{D}}\left(\varepsilon_i\varepsilon_i'\right)M_F\bar{P}_{Z_i}M_F\right]a_{ii} \\ &\le \frac{K+1}{N^{3/2}}\sum_{i=1}^{N}\lambda_i'(\Lambda'\Lambda/N)^{-1}\lambda_i=\frac{K+1}{N^{3/2}}\text{tr}\left(\Lambda(\Lambda'\Lambda/N)^{-1}\Lambda'\right) \\ &= r\left(K+1\right)N^{-1/2},\end{aligned}$$

it follows that $\Gamma_{5NT,321}=O_P(N^{-1/2})$ by the conditional Markov inequality. Noting that $E_{\mathcal{D}}\left(\Gamma_{5NT,322}\right)=0$ and

$$\begin{aligned}E_{\mathcal{D}}\left(\Gamma_{5NT,322}^2\right) &= \frac{1}{N^3}\sum_{1\le i\neq j\le N}\sum_{1\le k\neq l\le N}\sigma_i^{-2}\sigma_k^{-2}E_{\mathcal{D}}\left(\varepsilon_i' M_F\bar{P}_{Z_i}M_F\varepsilon_j\varepsilon_k' M_F\bar{P}_{Z_k}M_F\varepsilon_k a_{ij}a_{kl}\right) \\ &= \frac{1}{N^3}\sum_{1\le i\neq j\le N}a_{ij}^2\sigma_i^{-4}\text{tr}\left[M_F\bar{P}_{Z_i}M_F E_{\mathcal{D}}\left(\varepsilon_j\varepsilon_j'\right)M_F\bar{P}_{Z_i}M_F E_{\mathcal{D}}\left(\varepsilon_i\varepsilon_i'\right)\right] \\ &\quad +\frac{1}{N^3}\sum_{1\le i\neq j\le N}a_{ij}^2\sigma_i^{-2}\sigma_j^{-2}\text{tr}\left[M_F\bar{P}_{Z_i}M_F E_{\mathcal{D}}\left(\varepsilon_j\varepsilon_j'\right)M_F\bar{P}_{Z_j}M_F E_{\mathcal{D}}\left(\varepsilon_i\varepsilon_i'\right)\right] \\ &= \frac{1}{N^3}\sum_{1\le i\neq j\le N}a_{ij}^2\sigma_i^{-2}\sigma_j^{2}\text{tr}\left(M_F\bar{P}_{Z_i}M_F\bar{P}_{Z_i}\right)+\frac{1}{N^3}\sum_{1\le i\neq j\le N}a_{ij}^2\text{tr}\left(M_F\bar{P}_{Z_i}M_F\bar{P}_{Z_j}\right) \\ &\le \frac{\bar{c}_\sigma\underline{c}_\sigma^{-1}\left(K+1\right)}{N^3}\sum_{1\le i\neq j\le N}a_{ij}^2+\frac{K+1}{N^3}\sum_{1\le i\neq j\le N}a_{ij}^2 \\ &\le \frac{\left(\bar{c}_\sigma\underline{c}_\sigma^{-1}+1\right)\left(K+1\right)}{N^3}\left\|\Lambda(\Lambda'\Lambda/N)^{-1}\Lambda'\right\|^2=O_P(N^{-1}),\end{aligned}$$

it follows that $\Gamma_{5NT,322}=O_P(N^{-1/2})$ by the conditional Chebyshev inequality. Hence $\Gamma_{5NT,32}=O_P(N^{-1/2})=o_P(1)$. For $\Gamma_{5NT,33}$, it can be shown that $\Gamma_{5NT,33}=\bar{\Gamma}_{5NT,33}+o_P(1)$, where

$$\bar{\Gamma}_{5NT,33}\equiv\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sigma_i^{-2}\varepsilon_i' M_F\bar{P}_{Z_i}M_F\frac{1}{NT}\sum_{j=1}^{N}\varepsilon_j\varepsilon_j' FG\lambda_i.$$

Noting that

$$E_{\mathcal{D}}\left\|\frac{1}{NT}\sum_{j=1}^{N}\varepsilon_j\varepsilon_j'F\right\|^2 = \operatorname{tr}\left[\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}FF'E_{\mathcal{D}}\left(\varepsilon_i\varepsilon_i'\varepsilon_j\varepsilon_j'\right)\right]$$
$$= \frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\varpi_{ij}\left\|F'F/T\right\| = O_P(N^{-1}),$$

by (ii) we have

$$\begin{aligned}\bar{\Gamma}_{5NT,33} &= \operatorname{tr}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sigma_i^{-2}\lambda_i\varepsilon_i'M_F\bar{P}_{Z_i}M_F\frac{1}{NT}\sum_{j=1}^{N}\varepsilon_j\varepsilon_j'FG\right)\\ &\leq \left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sigma_i^{-2}\varepsilon_i'M_F\bar{P}_{Z_i}M_F\right\|\left\|\frac{1}{NT}\sum_{j=1}^{N}\varepsilon_j\varepsilon_j'F\right\|\left\|G\right\|\\ &= O_P(1)\,O_P(N^{-1/2})O_P(1) = o_P(1).\end{aligned}$$

It follows that $\Gamma_{5NT,33} = o_P(1)$. Hence $\Gamma_{5NT,3} = o_P(1)$ and $\Gamma_{5NT} = o_P(1)$.

(vi) By the same arguments as used in (v), it suffices to show that

$$\Gamma_{6NT,1} \equiv \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sigma_i^{-2}\varepsilon_i'M_F\bar{P}_{Z_i}T^{-1}FH\left(V_{NT}^{-1}\right)'(B_6+B_7+B_8)'\varepsilon_i = o_P(1).$$

Let $\Gamma_{6NT,1l} \equiv \frac{1}{\sqrt{N}}\sum\limits_{i=1}^{N}\sigma_i^{-2}\varepsilon_i'M_F\bar{P}_{Z_i}T^{-1}FH\left(V_{NT}^{-1}\right)'B_{5+l}'\varepsilon_i$ for $l = 1, 2, 3$. For $\Gamma_{6NT,11}$, by (iv) we have

$$\begin{aligned}\Gamma_{6NT,11} &= \frac{1}{N^{3/2}T^2}\sum_{i=1}^{N}\sigma_i^{-2}\operatorname{tr}\left[\varepsilon_i'M_F\bar{P}_{Z_i}FH\left(V_{NT}^{-1}\right)'\sum_{j=1}^{N}\hat{F}'\varepsilon_j\lambda_j'F'\varepsilon_i\right]\\ &= \operatorname{tr}\left[\frac{1}{N^{3/2}T^2}\sum_{i=1}^{N}\sigma_i^{-2}F'\varepsilon_i\varepsilon_i'M_F\bar{P}_{Z_i}FH\left(V_{NT}^{-1}\right)'\sum_{j=1}^{N}\hat{F}'\varepsilon_j\lambda_j'\right]\\ &\leq \left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sigma_i^{-2}F'\varepsilon_i\varepsilon_i'M_F\bar{P}_{Z_i}\right\|\frac{1}{\sqrt{T}}\left\|F\right\|\left\|H\right\|\left\|V_{NT}^{-1}\right\|\left\|\frac{1}{NT^{3/2}}\sum_{j=1}^{N}\hat{F}'\varepsilon_j\lambda_j'\right\|\\ &= O_P(\sqrt{N+T})O_P(1)O_P(N^{-1/2}T^{-1/2}) = o_P(1).\end{aligned}$$

By the same token, $\Gamma_{6NT,12}$ and $\Gamma_{6NT,13}$ are $o_P(1)$. Therefore $\Gamma_{6NT,1} = o_P(1)$ and $\Gamma_{6NT} = o_P(1)$. ■

Lemma A.4 *Suppose Assumptions A1-A3 and A4(i) hold. Then*

(i) $\max_{1\leq t\leq T}N^{-1}\sum_{i=1}^{N}\left\|Z_{it}\right\|^2 = O_P(\sqrt{T/N}+1)$,

(ii) $\max_{1\leq i\leq N}T^{-1}\sum_{t=1}^{T}\left\|Z_{it}\right\|^2 = O_P(\sqrt{N/T}+1)$,

(iii) $N^{-1}\sum_{i=1}^{N}\sum_{s=1}^{T}h_{i,ts}^2 = O_P\left(T^{-1}\right)$ *for each* t,

(iv) $N^{-1}\sum_{i=1}^{N}\sum_{s=1}^{T}h_{i,ts}^2 \leq (\alpha_{1NT}+\alpha_{2NT})\left\|F_t\right\|^2 + o_P(T^{-1/2})$ *uniformly in* t,

(v) $N^{-2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\sum_{r=1}^{t-1}h_{i,ts}^2h_{j,tr}^2 = o_P(1)$,

where $\alpha_{1NT} = c_F^2c_ZN^{-1}T^{-3}\left\|F\right\|^2\sum_{i=1}^{N}\sum_{s=1}^{T}\left\|Z_{is}\right\|^2$, $\alpha_{2NT} = c_F^4c_Z^2N^{-1}T^{-6}\left\|F\right\|^6\sum_{i=1}^{N}(\sum_{r=1}^{T}||Z_{ir}||^2)^2$, $c_Z \equiv \{\min_{1\leq i\leq N}\left[\lambda_{\min}\left(T^{-1}Z_i'Z_i\right)\right]\}^{-1}$ *and* $c_F \equiv \left[\lambda_{\min}\left(T^{-1}F'F\right)\right]^{-1}$.

Proof. (i) Let $\varsigma_{it} \equiv \|Z_{it}\|^2 - E\|Z_{it}\|^2$. Write $N^{-1}\sum_{i=1}^N \|Z_{it}\|^2 = N^{-1}\sum_{i=1}^N \varsigma_{it} + N^{-1}\sum_{i=1}^N E\|Z_{it}\|^2$. The second term is $O_P(1)$ by Assumption A1(i). By Assumption A1(v), for any $\epsilon > 0$ we have

$$P\left(\max_{1\leq t\leq T} N^{-1}\sum_{i=1}^N \varsigma_{it} \geq \epsilon\sqrt{T/N}\right) \leq \sum_{t=1}^T P\left(\sum_{i=1}^N \varsigma_{it} \geq \epsilon\sqrt{NT}\right) \leq \epsilon^{-2}(NT)^{-1}\sum_{t=1}^T E\left(\sum_{i=1}^N \varsigma_{it}\right)^2 = O(1).$$

It follows that $\max_{1\leq t\leq T}|N^{-1}\sum_{i=1}^N \varsigma_{it}| = O_P(\sqrt{T/N})$ and $\max_{1\leq t\leq T} N^{-1}\sum_{i=1}^N \|Z_{it}\|^2 = O_P(\sqrt{T/N}) + O_P(1) = O_P(\sqrt{T/N}+1)$.

(ii) The proof is analogous to that of (i) and thus omitted.

(iii) Note that $H_i = M_F(P_{Z_i} - L)M_F$. Let $m_{F,ts}$ and $p_{Z_i,ts}$ denote the (t,s)th element of M_F and P_{Z_i} respectively, i.e., $m_{F,ts} = 1_{ts} - p_{F,ts}$ and $p_{Z_i,ts} = Z_{it}'(Z_i'Z_i)^{-1}Z_{is}$, where $1_{ts} \equiv 1\{t=s\}$, Z_{it} denotes the tth column of Z_i', and $p_{F,ts} \equiv F_t'(F'F)^{-1}F_s$. Then $h_{i,ts} = \sum_{r=1}^T\sum_{q=1}^T m_{F,tr}\left(p_{Z_i,rq} - T^{-1}\right)m_{F,qs}$. Observe that

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^N\sum_{s=1}^T h_{i,ts}^2 &\leq \frac{2}{N}\sum_{i=1}^N\sum_{s=1}^T\left[\sum_{r=1}^T\sum_{q=1}^T m_{F,tr}p_{Z_i,rq}m_{F,qs}\right]^2 + \frac{2}{NT^2}\sum_{i=1}^N\sum_{s=1}^T\left[\sum_{r=1}^T\sum_{q=1}^T m_{F,tr}m_{F,qs}\right]^2 \\ &\equiv 2J_{t1} + 2J_{t2},\ \text{say,}\end{aligned}$$

and

$$\begin{aligned}J_{t1} &\leq \frac{4}{N}\sum_{i=1}^N\sum_{s=1}^T\left[\sum_{r=1}^T\sum_{q=1}^T 1_{tr}p_{Z_i,rq}1_{qs}\right]^2 + \frac{4}{N}\sum_{i=1}^N\sum_{s=1}^T\left[\sum_{r=1}^T\sum_{q=1}^T 1_{tr}p_{Z_i,rq}p_{F,qs}\right]^2 \\ &\quad + \frac{4}{N}\sum_{i=1}^N\sum_{s=1}^T\left[\sum_{r=1}^T\sum_{q=1}^T p_{F,tr}p_{Z_i,rq}1_{qs}\right]^2 + \frac{4}{N}\sum_{i=1}^N\sum_{s=1}^T\left[\sum_{r=1}^T\sum_{q=1}^T p_{F,tr}p_{Z_i,rq}p_{F,qs}\right]^2 \\ &\equiv 4J_{t11} + 4J_{t12} + 4J_{t13} + 4J_{t14},\ \text{say.}\end{aligned}$$

Noting that $|F_t'(T^{-1}F'F)^{-1}F_s| \leq \{F_t'(T^{-1}F'F)^{-1}F_t\}^{1/2}\{F_s'(T^{-1}F'F)^{-1}F_s\}^{1/2} \leq c_F\|F_t\|\|F_s\|$, by (A.7) and the Cauchy-Schwarz inequality we have

$$\begin{aligned}J_{t11} &= \frac{1}{N}\sum_{i=1}^N\sum_{s=1}^T (p_{Z_i,ts})^2 = \frac{1}{NT}\sum_{i=1}^N \mathrm{tr}\left(\left(T^{-1}Z_i'Z_i\right)^{-1}Z_{it}Z_{it}'\right) \leq \frac{c_Z}{NT}\sum_{i=1}^N\|Z_{it}\|^2, \\ J_{t12} &= \frac{1}{NT}\sum_{i=1}^N\sum_{s=1}^T\sum_{r=1}^T p_{Z_i,ts}p_{Z_i,tr}F_r'\left(T^{-1}F'F\right)^{-1}F_s \leq c_F\frac{1}{NT}\sum_{i=1}^N\left\{\sum_{s=1}^T|p_{Z_i,ts}|\|F_s\|\right\}^2 \\ &\leq \frac{c_Fc_Z^2}{NT}\sum_{i=1}^N\|Z_{it}\|^2\left\{\frac{1}{T}\sum_{s=1}^T\|Z_{is}\|\|F_s\|\right\}^2 \leq \frac{c_Fc_Z^2\|F\|^2}{NT^3}\sum_{i=1}^N\|Z_{it}\|^2\sum_{s=1}^T\|Z_{is}\|^2,\end{aligned}$$

$$
\begin{aligned}
J_{t13} &= \frac{1}{N}\sum_{i=1}^{N}\sum_{s=1}^{T}\sum_{r=1}^{T}\sum_{q=1}^{T} F_t'\left(F'F\right)^{-1}F_r\ F_t'\left(F'F\right)^{-1}F_q\ Z_{ir}'\left(Z_i'Z_i\right)^{-1}Z_{is}\ Z_{is}'\left(Z_i'Z_i\right)^{-1}Z_{iq} \\
&\leq \frac{c_F^2\left\|F_t\right\|^2}{NT^2}\sum_{i=1}^{N}\sum_{r=1}^{T}\sum_{q=1}^{T}\left\|F_r\right\|\left\|F_q\right\|\left|Z_{ir}'\left(Z_i'Z_i\right)^{-1}Z_{iq}\right| \\
&\leq \frac{c_F^2 c_Z\left\|F_t\right\|^2}{NT^3}\sum_{i=1}^{N}\sum_{r=1}^{T}\sum_{q=1}^{T}\left\|F_r\right\|\left\|F_q\right\|\left\|Z_{ir}'\right\|\left\|Z_{iq}\right\| \\
&= \frac{c_F^2 c_Z\left\|F_t\right\|^2}{NT}\sum_{i=1}^{N}\left\{\frac{1}{T}\sum_{r=1}^{T}\left\|F_r\right\|\left\|Z_{ir}\right\|\right\}^2 \leq \frac{c_F^2 c_Z\left\|F_t\right\|^2\left\|F\right\|^2}{NT^3}\sum_{i=1}^{N}\sum_{s=1}^{T}\left\|Z_{is}\right\|^2,
\end{aligned}
$$

and similarly,

$$
\begin{aligned}
J_{t14} &= \frac{1}{N}\sum_{i=1}^{N}\sum_{s=1}^{T}\left[\sum_{r=1}^{T}\sum_{q=1}^{T}F_t'\left(F'F\right)^{-1}F_rZ_{ir}'\left(Z_i'Z_i\right)^{-1}Z_{iq}F_q'\left(F'F\right)^{-1}F_s\right]^2 \\
&\leq \frac{c_F^4 c_Z^2\left\|F_t\right\|^2}{NT^2}\sum_{i=1}^{N}\sum_{s=1}^{T}\left\|F_s\right\|^2\left(\frac{1}{T}\sum_{r=1}^{T}\left\|F_r\right\|\left\|Z_{ir}\right\|\right)^4 \\
&\leq \frac{c_F^4 c_Z^2\left\|F_t\right\|^2\left\|F\right\|^2}{NT^2}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{r=1}^{T}\left\|F_r\right\|\left\|Z_{ir}\right\|\right)^4 \leq \frac{c_F^4 c_Z^2\left\|F_t\right\|^2\left\|F\right\|^6}{NT^6}\sum_{i=1}^{N}\left(\sum_{r=1}^{T}\left\|Z_{ir}\right\|^2\right)^2.
\end{aligned}
$$

We can readily show by the Markov inequality that $J_{t1l} = O_P\left(T^{-1}\right)$ for $l = 1, 2, 3, 4$, implying that $J_{t1} = O_P\left(T^{-1}\right)$. By the same token, $J_{t2} = O_P\left(T^{-1}\right)$. Thus $\frac{1}{N}\sum_{i=1}^{N}\sum_{s=1}^{T}h_{i,ts}^2 = O_P\left(T^{-1}\right)$.

(iv) By (i)-(ii) and the proof of (iii), we have

$$
\begin{aligned}
\max_{1\leq t\leq T} J_{t11} &= T^{-1}O_P\left(\sqrt{T/N}+1\right) = O_P\left((NT)^{-1/2}+T^{-1}\right) = o_P\left(T^{-1/2}\right), \\
\max_{1\leq t\leq T} J_{t12} &\leq \frac{c_F c_Z^2\left\|F\right\|^2}{T^2}\max_{1\leq i\leq N}\left(\frac{1}{T}\sum_{s=1}^{T}\left\|Z_{is}\right\|^2\right)\max_{1\leq t\leq T}\left(\frac{1}{N}\sum_{i=1}^{N}\left\|Z_{it}\right\|^2\right) \\
&= O_P\left(T^{-1}\right)O_P(\sqrt{N/T}+1)O_P(\sqrt{T/N}+1) = o_P(T^{-1/2}), \\
J_{t13} &\leq \alpha_{1NT}\left\|F_t\right\|^2, \text{ and } J_{t14} \leq \alpha_{2NT}\left\|F_t\right\|^2.
\end{aligned}
$$

Thus (iv) follows.

(v) By Markov inequality, we can show that $\alpha_{1NT} = O_P\left(T^{-1}\right)$, and $\alpha_{2NT} = O_P(T^{-1})$. It follows that

$$
\begin{aligned}
\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\sum_{r=1}^{t-1}h_{i,ts}^2h_{j,tr}^2 &= \sum_{t=2}^{T}\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{s=1}^{t-1}h_{i,ts}^2\right)^2 \leq \sum_{t=2}^{T}\left[\left(\alpha_{1NT}+\alpha_{2NT}\right)\left\|F_t\right\|^2+o_P(T^{-1/2})\right]^2 \\
&= \left(\alpha_{1NT}+\alpha_{2NT}\right)^2\sum_{t=2}^{T}\left\|F_t\right\|^4+o_P(1) = O_P(T^{-1})+o_P(1) = o_P(1).
\end{aligned}
$$

■

Lemma A.5 *Let $\hat{\sigma}_i^2 \equiv TSS_i/T$ and $\upsilon_{NT} \equiv N^{1/\vartheta}T^{-1/2}$. Suppose Assumptions A1-A3 hold. Then under* $\mathbb{H}_{1,NT}$, $\max_{1\leq i\leq N}\left|\hat{\sigma}_i^2-\sigma_i^2\right| = O_P(\upsilon_{NT}+N^{1/2}T^{-1}+\gamma_{NT}^{1/2}+\delta_{NT}^{-1}N^{1/(2\vartheta)})$.

Proof. Noting that $\hat{\boldsymbol{\varepsilon}}_i = M_{\hat{F}}\boldsymbol{\varepsilon}_i + M_{\hat{F}}X_i(\beta-\hat{\beta}) + M_{\hat{F}}F\lambda_i + M_{\hat{F}}X_i(\beta_i-\beta)$, we have

$$\hat{\sigma}_i^2 = \hat{\boldsymbol{\varepsilon}}_i' M_0 \hat{\boldsymbol{\varepsilon}}_i / T = \sum_{l=1}^{10} TSS_{il}/T, \tag{A.11}$$

where

$$\begin{aligned}
&TSS_{i1} \equiv \boldsymbol{\varepsilon}_i' M_{\hat{F}} M_0 M_{\hat{F}} \boldsymbol{\varepsilon}_i, && TSS_{i6} \equiv 2\boldsymbol{\varepsilon}_i' M_{\hat{F}} M_0 M_{\hat{F}} F\lambda_i,\\
&TSS_{i2} \equiv (\beta-\hat{\beta})' X_i' M_{\hat{F}} M_0 M_{\hat{F}} X_i(\beta-\hat{\beta}), && TSS_{i7} \equiv 2\boldsymbol{\varepsilon}_i' M_{\hat{F}} M_0 M_{\hat{F}} X_i(\beta_i-\beta),\\
&TSS_{i3} \equiv \lambda_i' F' M_{\hat{F}} M_0 M_{\hat{F}} F\lambda_i, && TSS_{i8} \equiv 2(\beta-\hat{\beta})' X_i' M_{\hat{F}} M_0 M_{\hat{F}} F\lambda_i,\\
&TSS_{i4} \equiv (\beta_i-\beta)' X_i' M_{\hat{F}} M_0 M_{\hat{F}} X_i(\beta_i-\beta), && TSS_{i9} \equiv 2(\beta-\hat{\beta})' X_i' M_{\hat{F}} M_0 M_{\hat{F}} X_i(\beta_i-\beta),\\
&TSS_{i5} \equiv 2\boldsymbol{\varepsilon}_i' M_{\hat{F}} M_0 M_{\hat{F}} X_i(\beta-\hat{\beta}), && TSS_{i10} \equiv 2\lambda_i' F' M_{\hat{F}} M_0 M_{\hat{F}} X_i(\beta_i-\beta).
\end{aligned}$$

We prove the lemma by showing that

$$\max_{1\le i\le N} \left| T^{-1} TSS_{i1} - \sigma_i^2 \right| = O_P(\upsilon_{NT} + N^{1/2}T^{-1} + \delta_{NT}^{-1} + \gamma_{NT}), \tag{A.12}$$

and

$$\sum_{l=2}^{10} \max_{1\le i\le N} \left| T^{-1} TSS_{il} \right| = O_P(\upsilon_{NT} + \delta_{NT}^{-1} N^{1/(2\vartheta)}). \tag{A.13}$$

First, we prove (A.12). Observe that

$$T^{-1}TSS_{i1} - \sigma_i^2 = \left(T^{-1}\boldsymbol{\varepsilon}_i' M_0 \boldsymbol{\varepsilon}_i - \sigma_i^2\right) + T^{-1}\boldsymbol{\varepsilon}_i' P_{\hat{F}} M_0 P_{\hat{F}} \boldsymbol{\varepsilon}_i - 2T^{-1}\boldsymbol{\varepsilon}_i' M_0 P_{\hat{F}} \boldsymbol{\varepsilon}_i. \tag{A.14}$$

For the first term in (A.14), write $T^{-1}\boldsymbol{\varepsilon}_i' M_0 \boldsymbol{\varepsilon}_i - \sigma_i^2 = T^{-1}\sum_{t=1}^T (\varepsilon_{it} - \bar{\varepsilon}_i)^2 - \sigma_i^2 = T^{-1}\sum_{t=1}^T \xi_{it} - \bar{\varepsilon}_i^2$, where $\xi_{it} \equiv \varepsilon_{it}^2 - \sigma_i^2$. Then by Assumption A3(iii), for any $\epsilon > 0$ we have

$$\begin{aligned}
P\left(\max_{1\le i\le N} \frac{1}{T}\sum_{t=1}^T \xi_{it} \ge \epsilon\upsilon_{NT}\right) &\le \sum_{i=1}^N P\left(\frac{1}{T}\sum_{t=1}^T \xi_{it} \ge \epsilon\upsilon_{NT}\right) \le \epsilon^{-\vartheta}\upsilon_{NT}^{-\vartheta} \sum_{i=1}^N E\left|\frac{1}{T}\sum_{t=1}^T \xi_{it}\right|^{\vartheta}\\
&= \frac{\epsilon^{-\vartheta}\upsilon_{NT}^{-\vartheta}}{T^{\vartheta}} \sum_{i=1}^N \sum_{1\le t_1,\ldots,t_\vartheta \le T} E\left(\xi_{it_1}\xi_{it_2}\ldots\xi_{it_\vartheta}\right)\\
&= O(NT^{-\vartheta/2}\upsilon_{NT}^{-\vartheta}) = O(1).
\end{aligned} \tag{A.15}$$

It follows that $\max_{1\le i\le N} |T^{-1}\sum_{t=1}^T \varepsilon_{it}^2 - \sigma_i^2| = O_P(\upsilon_{NT})$. Similarly, $\max_{1\le i\le N} |\bar{\varepsilon}_i| = O_P(\upsilon_{NT}^2) = o_P(\upsilon_{NT})$. It follows that

$$\max_{1\le i\le N} \left| T^{-1}\boldsymbol{\varepsilon}_i' M_0 \boldsymbol{\varepsilon}_i - \sigma_i^2 \right| = O_P(\upsilon_{NT}). \tag{A.16}$$

For the second term in (A.14), observe that $T^{-1}\boldsymbol{\varepsilon}_i' P_{\hat{F}} M_0 P_{\hat{F}} \boldsymbol{\varepsilon}_i \le T^{-1}\boldsymbol{\varepsilon}_i' P_{\hat{F}} \boldsymbol{\varepsilon}_i$ as $\lambda_{\max}(M_0) = 1$. Further, $T^{-1}\boldsymbol{\varepsilon}_i' P_{\hat{F}} \boldsymbol{\varepsilon}_i = T^{-1}\boldsymbol{\varepsilon}_i' P_F \boldsymbol{\varepsilon}_i + T^{-1}\boldsymbol{\varepsilon}_i' \left(P_{\hat{F}} - P_F\right)\boldsymbol{\varepsilon}_i$. First, $\max_{1\le i\le N} T^{-1}\boldsymbol{\varepsilon}_i' P_F \boldsymbol{\varepsilon}_i \le c_F \max_{1\le i\le N} T^{-2}\boldsymbol{\varepsilon}_i' FF' \boldsymbol{\varepsilon}_i$, where $c_F \equiv \left[\lambda_{\min}\left(T^{-1}F'F\right)\right]^{-1}$. Writing $T^{-2}\boldsymbol{\varepsilon}_i' FF'\boldsymbol{\varepsilon}_i = T^{-2}\left[\boldsymbol{\varepsilon}_i' FF'\boldsymbol{\varepsilon}_i - E\left(\boldsymbol{\varepsilon}_i' FF'\boldsymbol{\varepsilon}_i\right)\right] + T^{-2}E\left(\boldsymbol{\varepsilon}_i' FF'\boldsymbol{\varepsilon}_i\right)$, as in (A.15) we can show that the first term is $O_P\left(N^{1/2}T^{-1}\right)$ and the second term is $O\left(T^{-1}\right)$, both uniformly in i. It follows that $\max_{1\le i\le N} T^{-2}\boldsymbol{\varepsilon}_i' FF'\boldsymbol{\varepsilon}_i = O_P\left(N^{1/2}T^{-1}\right)$, implying that $\max_{1\le i\le N} T^{-1}\boldsymbol{\varepsilon}_i' P_F \boldsymbol{\varepsilon}_i = O_P\left(N^{1/2}T^{-1}\right)$. Now, by Lemmas A.1(vii) and A.2, $\max_{1\le i\le N} \left|T^{-1}\boldsymbol{\varepsilon}_i' \left(P_{\hat{F}} - P_F\right)\boldsymbol{\varepsilon}_i\right| \le ||P_{\hat{F}} - P_F|| \times \max_{1\le i\le N}(T^{-1}\left\|\boldsymbol{\varepsilon}_i\right\|^2) = O_P(\delta_{NT}^{-1} + \gamma_{NT})$. Consequently,

$$\max_{1\le i\le N} T^{-1}\boldsymbol{\varepsilon}_i' P_{\hat{F}} M_0 P_{\hat{F}} \boldsymbol{\varepsilon}_i = O_P(N^{1/2}T^{-1} + \delta_{NT}^{-1} + \gamma_{NT}). \tag{A.17}$$

For the third term in (A.14), write $T^{-1}\boldsymbol{\varepsilon}_i' M_0 P_{\hat{F}} \boldsymbol{\varepsilon}_i = T^{-1}\boldsymbol{\varepsilon}_i' P_{\hat{F}} \boldsymbol{\varepsilon}_i - T^{-1}\boldsymbol{\varepsilon}_i' L P_{\hat{F}} \boldsymbol{\varepsilon}_i$. The uniform bound for the first term was obtained above, and we can show that the second term is also bounded by $O_P(N^{1/2}T^{-1} + \delta_{NT}^{-1} + \gamma_{NT})$. Hence we have

$$\max_{1\le i\le N} T^{-1} \left|\boldsymbol{\varepsilon}_i' M_0 P_{\hat{F}} \boldsymbol{\varepsilon}_i\right| = O_P(N^{1/2}T^{-1} + \delta_{NT}^{-1} + \gamma_{NT}). \tag{A.18}$$

Combining (A.16)-(A.18) delivers (A.12).

Now, we prove (A.13). For TSS_{i2}, by Lemmas A.2 and A.4(i) we have

$$\max_{1\le i\le N} T^{-1} TSS_{i2} \le ||\hat{\beta} - \beta||^2 \max_{1\le i\le N} (T^{-1} \left\|X_i\right\|^2) = O_P(\gamma_{NT}^2) O_P(\sqrt{N/T} + 1).$$

For TSS_{i3}, by the repeated use of (A.7), Lemmas A.1(i) and A.2, and Assumption A1(vi), we have

$$\begin{aligned}
\max_{1\le i\le N} \{T^{-1} TSS_{i3}\} &= \max_{1\le i\le N} T^{-1} \lambda_i' \left(F - \hat{F}H\right)' M_{\hat{F}} M_0 M_{\hat{F}} \left(F - \hat{F}H\right) \lambda_i \\
&\le \max_{1\le i\le N} T^{-1} \lambda_i' \left(F - \hat{F}H\right)' \left(F - \hat{F}H\right) \lambda_i \le T^{-1} \left\|F - \hat{F}H\right\|^2 \max_{1\le i\le N} \left\|\lambda_i\right\|^2 \\
&= O_P(\delta_{NT}^{-2} + \gamma_{NT}^2) O_P(N^{1/\vartheta}) = O_P\left(\delta_{NT}^{-2} N^{1/\vartheta}\right),
\end{aligned}$$

where we used the fact that $\max_{1\le i\le N} \left\|\lambda_i\right\|^2 = O_P(N^{1/\vartheta})$ by Assumption A1(vi) and the Markov inequality.

For TSS_{i4}, by (3.2) we can obtain

$$\max_{1\le i\le N} \left\{T^{-1} TSS_{i4}\right\} \le M^2 \gamma_{NT}^2 \max_{1\le i\le N} T^{-1} \left\|X_i\right\|^2 = O_P(\gamma_{NT}^2) O_P(\sqrt{N/T} + 1).$$

By the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
&\max_{1\le i\le N} T^{-1} \left|TSS_{i5}\right| = O_P(\gamma_{NT}((N/T)^{1/4} + 1)), \quad \max_{1\le i\le N} T^{-1} \left|TSS_{i6}\right| = O_P(\delta_{NT}^{-1} N^{1/(2\vartheta)}), \\
&\max_{1\le i\le N} T^{-1} \left|TSS_{i7}\right| = O_P(\gamma_{NT}((N/T)^{1/4} + 1)), \quad \max_{1\le i\le N} T^{-1} \left|TSS_{i8}\right| = O_P(\gamma_{NT}((N/T)^{1/4} + 1)\delta_{NT}^{-1} N^{1/(2\vartheta)}), \\
&\max_{1\le i\le N} T^{-1} \left|TSS_{i9}\right| = O_P(\gamma_{NT}^2((N/T)^{1/2} + 1)), \quad \max_{1\le i\le N} T^{-1} \left|TSS_{i10}\right| = O_P(\gamma_{NT}((N/T)^{1/4} + 1)\delta_{NT}^{-1} N^{1/(2\vartheta)}).
\end{aligned}$$

Noting that $\gamma_{NT}((N/T)^{1/4} + 1) = o\left(\delta_{NT}^{-1}\right)$, (A.13) follows. ■

B Proof of the Results in Section 3

Proof of Theorem 3.1.

The proof is a special case of that of Theorem 3.2 and thus omitted. ■

Proof of Corollary 3.2.

By Theorem 3.1 and the Slutsky lemma, it suffices to prove the first two parts of the corollary. In fact, we prove a slightly stronger result, i.e., under $\mathbb{H}_{1,NT}$ in (3.2), $\hat{B}_{NT} = B_{NT} + o_P(1)$ and $\hat{V}_{NT} = V_{NT} + o_P(1)$. This stronger result will be needed in the proof of Theorem 3.3 below.

(i) We prove $\hat{B}_{NT} = B_{NT} + o_P(1)$ under $H_{1,NT}$. Recall $\hat{\sigma}_i^2 \equiv T^{-1}TSS_i$. We can decompose $\hat{B}_{NT} - B_{NT}$ as follows:

$$\begin{aligned}
\hat{B}_{NT} - B_{NT} &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\operatorname{tr}\left(\hat{H}_i\right) - \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\hat{\sigma}_i^{-2}\sum_{t=1}^{T}\varepsilon_{it}^2 h_{i,tt} \\
&= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\operatorname{tr}\left(\hat{H}_i - H_i\right) + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sigma_i^{-2}\sum_{t=1}^{T}\left(\sigma_i^2 - \varepsilon_{it}^2\right)h_{i,tt} + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\sigma_i^{-2} - \hat{\sigma}_i^{-2}\right)\sum_{t=1}^{T}h_{i,tt}\varepsilon_{it}^2 \\
&\equiv \hat{B}_{NT,1} + \hat{B}_{NT,2} + \hat{B}_{NT,3},\ \text{say.}
\end{aligned}$$

Noting that $\hat{H}_i - H_i = (M_{\hat{F}} - M_F)\bar{P}_{Z_i}(M_{\hat{F}} - M_F) + M_F\bar{P}_{Z_i}(M_{\hat{F}} - M_F) + (M_{\hat{F}} - M_F)\bar{P}_{Z_i}M_F$, we have

$$\begin{aligned}
\hat{B}_{NT,1} &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\operatorname{tr}\left[(M_{\hat{F}} - M_F)\bar{P}_{Z_i}(M_{\hat{F}} - M_F)\right] + \frac{2}{\sqrt{N}}\sum_{i=1}^{N}\operatorname{tr}\left[(M_{\hat{F}} - M_F)\bar{P}_{Z_i}M_F\right] \\
&\equiv \hat{B}_{NT,11} + 2\hat{B}_{NT,12},\ \text{say.}
\end{aligned}$$

By Lemmas A.1(vii) and A.2, (A.7) and Assumption A4(i), we have $\hat{B}_{NT,11} \le \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\|M_{\hat{F}} - M_F\right\|^2$ $= O_P(N^{1/2}\delta_{NT}^{-2}) = o_P(1)$. For $\hat{B}_{NT,12}$, using (A.1) we have

$$\hat{B}_{NT,12} = -\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\operatorname{tr}\left[(a_1 + a_2 + a_3 + a_4)\bar{P}_{Z_i}M_F\right] \equiv -\hat{B}_{NT,121} - \hat{B}_{NT,122} - \hat{B}_{NT,123} - \hat{B}_{NT,124},\ \text{say.}$$

For $\hat{B}_{NT,121}$, we have

$$\begin{aligned}
\hat{B}_{NT,121} &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\operatorname{tr}\left[T^{-1}M_F(\hat{F} - FH)H'F'P_{Z_i}\right] - \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\operatorname{tr}\left[T^{-1}M_F(\hat{F} - FH)H'F'\frac{1}{T}\mathbf{i}_T\mathbf{i}_T'\right] \\
&\equiv \hat{B}_{NT,1211} + \hat{B}_{NT,1212},\ \text{say.}
\end{aligned}$$

We further decompose $\hat{B}_{NT,1211}$ as follows

$$\begin{aligned}
\hat{B}_{NT,1211} &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\operatorname{tr}\left[T^{-1}M_{\hat{F}}(\hat{F} - FH)H'F'P_{Z_i}\right] + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\operatorname{tr}\left[T^{-1}\left(M_F - M_{\hat{F}}\right)(\hat{F} - FH)H'F'P_{Z_i}\right] \\
&\equiv \hat{B}_{NT,1211a} + \hat{B}_{NT,1211b}.
\end{aligned}$$

By the repeated use of the matrix version of Cauchy-Schwarz inequality, Assumptions A1(iv) and A4(i), Lemmas A.1(ix) and A.2,

$$\begin{aligned}
\left|\hat{B}_{NT,1211a}\right| &= \left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\operatorname{tr}\left[T^{-1}Z_i'M_{\hat{F}}(\hat{F} - FH)H'F'Z_i\left(Z_i'Z_i\right)^{-1}\right]\right| \\
&\le \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\|T^{-1}Z_i'M_{\hat{F}}(\hat{F} - FH)\right\|\left\|H'F'Z_i\left(Z_i'Z_i\right)^{-1}\right\| \\
&\le \sqrt{\frac{1}{TN}}\left\|FH\right\|\left\{\min_{1\le i\le N}\lambda_{\min}\left(T^{-1}Z_i'Z_i\right)\right\}^{-1/2}\sum_{i=1}^{N}\left\|T^{-1}Z_i'M_{\hat{F}}(\hat{F} - FH)\right\| \\
&= \sqrt{\frac{1}{TN}}O_P(\sqrt{T})O_P(1)O_P\left(N\delta_{NT}^{-2} + N\gamma_{NT}\right) = O_P\left(N^{1/2}(\delta_{NT}^{-2} + \gamma_{NT})\right) = o_P\left(1\right).
\end{aligned}$$

By the repeated use of the matrix version of Cauchy-Schwarz inequality, (A.7), Lemmas A.1(i) and (vii) and Lemma A.2, and Assumption A4(i),

$$\begin{aligned}
\left|\hat{B}_{NT,1211b}\right| &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\text{tr}\left[T^{-1}\left(M_F - M_{\hat{F}}\right)(\hat{F}-FH)H'F'P_{Z_i}\right] \\
&\leq \left\|M_F - M_{\hat{F}}\right\|\left\{T^{-1/2}\left\|\hat{F}-FH\right\|\right\}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left\{\text{tr}\left(H'F'P_{Z_i}FH\right)\right\}^{1/2} \\
&= O_P\left(\delta_{NT}^{-1}\right)O_P\left(\delta_{NT}^{-1}\right)\sqrt{\frac{N}{T}}\left\|FH\right\| = O_P\left(N^{1/2}\delta_{NT}^{-2}\right) = o_P(1).
\end{aligned}$$

It follows that $\hat{B}_{NT,1211} = o_P(1)$. By the same token, $\hat{B}_{NT,1212} = o_P(1)$. Thus $\hat{B}_{NT,121} = o_P(1)$. Analogously, we can show that $\hat{B}_{NT,12l} = o_P(1)$ for $l = 2,3,4$. It follows that $\hat{B}_{NT,12} = o_P(1)$ and $\hat{B}_{NT,1} = o_P(1)$.

For $\hat{B}_{NT,2}$, noting that $E_{\mathcal{D}}(\hat{B}_{NT,2}) = 0$, and by Assumption A2(v) and Lemma A.4(iii)

$$\begin{aligned}
E_{\mathcal{D}}\left(\hat{B}_{NT,2}^2\right) &= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\sigma_i^{-2}\sigma_j^{-2}E\left[\left(\sigma_i^2-\varepsilon_{it}^2\right)\left(\sigma_j^2-\varepsilon_{js}^2\right)\right]h_{i,tt}h_{j,ss} \\
&= \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\sigma_i^{-4}h_{i,tt}^2E\left[\left(\varepsilon_{it}^2-\sigma_i^2\right)^2\right] \leq \frac{\underline{c}_\sigma^{-2}M}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}h_{i,tt}^2 = o_P(1).
\end{aligned}$$

It follows that $\hat{B}_{NT,2} = o_P(1)$.

By a geometric expansion, $1/\hat{\sigma}_i^2 - 1/\sigma_i^2 = -\left(\hat{\sigma}_i^2-\sigma_i^2\right)/\sigma_i^4 + (\hat{\sigma}_i^2-\sigma_i^2)^2/(\sigma_i^4\hat{\sigma}_i^2)$. It follows that

$$\begin{aligned}
\hat{B}_{NT,3} &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\hat{\sigma}_i^2-\sigma_i^2}{\sigma_i^4}\sum_{t=1}^{T}h_{i,tt}\varepsilon_{it}^2 - \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\left(\hat{\sigma}_i^2-\sigma_i^2\right)^2}{\sigma_i^4\hat{\sigma}_i^2}\sum_{t=1}^{T}h_{i,tt}\varepsilon_{it}^2 \\
&\equiv \hat{B}_{NT,31} - \hat{B}_{NT,32}, \text{ say.}
\end{aligned}$$

By Lemma A.5, we can readily show that $\hat{B}_{NT,32} = \sqrt{N}O_P\left(\upsilon_{NT}^2 + NT^{-2} + \gamma_{NT}^2 + \delta_{NT}^{-2}N^{1/\vartheta}\right) = o_P(1)$. Using the decomposition of $\hat{\sigma}_i^2$ in (A.11) and following the arguments analogous to those used in the proof of Lemma A.5, we can show $\hat{B}_{NT,31} = o_P(1)$. Thus $\hat{B}_{NT,3} = o_P(1)$, and $\hat{B}_{NT} - B_{NT} = o_P(1)$.

(ii) We prove $\hat{V}_{NT} = V_{NT} + o_P(1)$ under $\mathbb{H}_{1,NT}$. Note that

$$\hat{V}_{NT} - V_{NT} = \frac{2}{N}\sum_{i=1}^{N}\text{tr}\left(\hat{H}_i^2 - H_i^2\right) + \frac{2}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}h_{i,tt}^2 \equiv 2\hat{V}_{NT,1} + 2\hat{V}_{NT,2}, \text{ say.}$$

We further decompose $\hat{V}_{NT,1}$ as follows:

$$\hat{V}_{NT,1} = \frac{1}{N}\sum_{i=1}^{N}\text{tr}\left[\hat{H}_i\left(\hat{H}_i - H_i\right)\right] + \frac{1}{N}\sum_{i=1}^{N}\text{tr}\left[\left(\hat{H}_i - H_i\right)H_i\right] \equiv \hat{V}_{NT,11} + \hat{V}_{NT,12}, \text{ say.}$$

Noting that

$$\begin{aligned}
\hat{H}_i - H_i &= M_{\hat{F}}\bar{P}_{Z_i}M_{\hat{F}} - M_F\bar{P}_{Z_i}M_F \\
&= (M_{\hat{F}} - M_F)\bar{P}_{Z_i}(M_{\hat{F}} - M_F) + M_F\bar{P}_{Z_i}(M_{\hat{F}} - M_F) + (M_{\hat{F}} - M_F)\bar{P}_{Z_i}M_F,
\end{aligned}$$

we have

$$
\begin{aligned}
\hat{V}_{NT,11} &= \frac{1}{N}\sum_{i=1}^{N}\operatorname{tr}\left(\hat{H}_i(M_{\hat{F}}-M_F)\bar{P}_{Z_i}(M_{\hat{F}}-M_F)\right)+\frac{1}{N}\sum_{i=1}^{N}\operatorname{tr}\left(\hat{H}_iM_F\bar{P}_{Z_i}(M_{\hat{F}}-M_F)\right) \\
&\quad +\frac{1}{N}\sum_{i=1}^{N}\operatorname{tr}\left(\hat{H}_i(M_{\hat{F}}-M_F)\bar{P}_{Z_i}M_F\right) \\
&\equiv \hat{V}_{NT,11a}+\hat{V}_{NT,11b}+\hat{V}_{NT,11c},\ \text{say.}
\end{aligned}
$$

By the repeated use of (A.7), the fact that $\lambda_{\max}(M_F)=\lambda_{\max}\left(\bar{P}_{Z_i}\right)=1$, and Lemma A.1(vii) and A.2, we have

$$
\begin{aligned}
\left|\hat{V}_{NT,11a}\right| &= \frac{1}{N}\sum_{i=1}^{N}\operatorname{tr}\left(\bar{P}_{Z_i}M_{\hat{F}}(M_{\hat{F}}-M_F)\bar{P}_{Z_i}(M_{\hat{F}}-M_F)M_{\hat{F}}\right) \\
&\leq \frac{1}{N}\sum_{i=1}^{N}\operatorname{tr}\left(M_{\hat{F}}(M_{\hat{F}}-M_F)\bar{P}_{Z_i}(M_{\hat{F}}-M_F)\right)\leq\frac{1}{N}\sum_{i=1}^{N}\operatorname{tr}\left(\bar{P}_{Z_i}(M_{\hat{F}}-M_F)(M_{\hat{F}}-M_F)\right) \\
&\leq \left\|M_{\hat{F}}-M_F\right\|^2=O_P\left(\delta_{NT}^{-2}+\gamma_{NT}^2\right)=o_P(1).
\end{aligned}
$$

Similarly, $\hat{V}_{NT,11b}=O_P(\delta_{NT}^{-1}+\gamma_{NT})=o_P(1)$ and $\hat{V}_{NT,11c}=O_P(\delta_{NT}^{-1}+\gamma_{NT})=o_P(1)$. It follows that $\hat{V}_{NT,11}=o_P(1)$. By the same token $\hat{V}_{NT,12}=o_P(1)$ and hence we have $\hat{V}_{NT,1}=o_P(1)$. By Lemma A.4(iii), $\hat{V}_{NT,2}=O_P\left(T^{-1}\right)=o_P(1)$. Hence $\hat{V}_{NT}-V_{NT}=o_P(1)$. ■

Proof of Theorem 3.3.

By (2.10) and (2.11), we have

$$
\begin{aligned}
J_{NT} &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\hat{\boldsymbol{\varepsilon}}_i'\bar{P}_{Z_i}\hat{\boldsymbol{\varepsilon}}_i}{T^{-1}TSS_i}-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\varepsilon_{it}^2h_{i,tt}}{T^{-1}TSS_i} \\
&= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sigma_i^{-2}\left[\hat{\boldsymbol{\varepsilon}}_i'\bar{P}_{Z_i}\hat{\boldsymbol{\varepsilon}}_i-\boldsymbol{\varepsilon}_i'Q_i\boldsymbol{\varepsilon}_i\right]+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[\hat{\boldsymbol{\varepsilon}}_i'\bar{P}_{Z_i}\hat{\boldsymbol{\varepsilon}}_i-\boldsymbol{\varepsilon}_i'Q_i\boldsymbol{\varepsilon}_i\right]\left[(T^{-1}TSS_i)^{-1}-\sigma_i^{-2}\right] \\
&\equiv J_{NT,1}+J_{NT,2},\ \text{say,}
\end{aligned}
$$

where $Q_i\equiv\operatorname{diag}(h_{i,11},h_{i,22},\ldots,h_{i,TT})$. We prove the theorem by showing that: (i) $J_{NT,1}\xrightarrow{d}N(\Theta_0,V_0)$, (ii) $J_{NT,2}=o_P(1)$, (iii) $\hat{B}_{NT}=B_{NT}+o_P(1)$, and (iv) $\hat{V}_{NT}=V_{NT}+o_P(1)$. (iii) and (iv) are proved in the proof of Corollary 3.2. So we complete the proof of the theorem by showing (i) and (ii) respectively in Propositions B.1 and B.2 below.

Proposition B.1 *Under the conditions of Theorem 3.3,* $J_{NT,1}\xrightarrow{d}N(\Theta_0,V_0)$.

Proof. By (2.8), we have

$$
J_{NT,1}=R_{1NT}+R_{2NT}+R_{3NT}+R_{4NT}+2R_{5NT}+2R_{6NT}+2R_{7NT}+2R_{8NT}+2R_{9NT}+2R_{10NT},
$$

where

$$R_{1NT} \equiv \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sigma_i^{-2} \left(\varepsilon_i' M_{\hat{F}} \bar{P}_{Z_i} M_{\hat{F}} \varepsilon_i - \varepsilon_i' Q_i \varepsilon_i \right), \qquad R_{6NT} \equiv \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sigma_i^{-2} \varepsilon_i' M_{\hat{F}} \bar{P}_{Z_i} M_{\hat{F}} F \lambda_i,$$

$$R_{2NT} \equiv \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sigma_i^{-2} (\beta - \hat{\beta})' X_i' M_{\hat{F}} \bar{P}_{Z_i} M_{\hat{F}} X_i (\beta - \hat{\beta}), \quad R_{7NT} \equiv \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sigma_i^{-2} \varepsilon_i' M_{\hat{F}} \bar{P}_{Z_i} M_{\hat{F}} X_i (\beta_i - \beta),$$

$$R_{3NT} \equiv \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sigma_i^{-2} \lambda_i' F' M_{\hat{F}} \bar{P}_{Z_i} M_{\hat{F}} F \lambda_i, \qquad R_{8NT} \equiv \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sigma^{-2} (\beta - \hat{\beta})' X_i' M_{\hat{F}} \bar{P}_{Z_i} M_{\hat{F}} F \lambda_i,$$

$$R_{4NT} \equiv \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sigma_i^{-2} (\beta_i\text{-}\beta)' X_i' M_{\hat{F}} \bar{P}_{Z_i} M_{\hat{F}} X_i (\beta_i\text{-}\beta), \qquad R_{9NT} \equiv \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sigma_i^{-2} (\beta\text{-}\hat{\beta})' X_i' M_{\hat{F}} \bar{P}_{Z_i} M_{\hat{F}} X_i (\beta_i\text{-}\beta),$$

$$R_{5NT} \equiv \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sigma_i^{-2} \varepsilon_i' M_{\hat{F}} \bar{P}_{Z_i} M_{\hat{F}} X_i (\beta - \hat{\beta}), \qquad R_{10NT} \equiv \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sigma_i^{-2} \lambda_i' F' M_{\hat{F}} \bar{P}_{Z_i} M_{\hat{F}} X_i (\beta_i - \beta).$$

We prove the proposition by showing that

$$R_{1NT} \xrightarrow{d} N(0, V_0), \tag{B.1}$$

$$R_{2NT} + R_{3NT} + R_{4NT} + 2R_{8NT} + 2R_{9NT} + 2R_{10NT} = \Theta_0 + o_P(1), \tag{B.2}$$

$$R_{sNT} = o_P(1), \;\; s = 5, 6, 7. \tag{B.3}$$

(i) First, we prove (B.1). We decompose R_{1NT} as follows:

$$\begin{aligned} R_{1NT} &= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sigma_i^{-2} \varepsilon_i' \left(M_F \bar{P}_{Z_i} M_F - Q_i \right) \varepsilon_i + \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sigma_i^{-2} \varepsilon_i' (M_{\hat{F}} - M_F) \bar{P}_{Z_i} (M_{\hat{F}} - M_F) \varepsilon_i. \\ &\quad + \frac{2}{\sqrt{N}} \sum_{i=1}^{N} \sigma_i^{-2} \varepsilon_i' M_F \bar{P}_{Z_i} (M_{\hat{F}} - M_F) \varepsilon_i \\ &\equiv R_{1NT,1} + R_{1NT,2} + 2R_{1NT,3}. \end{aligned}$$

It suffices to show that $R_{1NT,2}$ and $R_{1NT,3}$ are $o_P(1)$, and $R_{1NT,1} \xrightarrow{d} N(0, V_0)$.

By Lemma A.3(i), $R_{1NT,2} = \Gamma_{1NT} = o_P(1)$. For $R_{1NT,3}$, using (A.1) we have

$$\begin{aligned} R_{1NT,3} &= -N^{-1/2} \sum_{i=1}^{N} \sigma_i^{-2} \varepsilon_i' M_F \bar{P}_{Z_i} (a_1 + a_2 + a_3 + a_4) \varepsilon_i \\ &\equiv -R_{1NT,31} - R_{1NT,32} - R_{1NT,33} - R_{1NT,34}, \text{ say,} \end{aligned}$$

where, e.g., $R_{1NT,31} = N^{-1/2} \sum_{i=1}^{N} \sigma_i^{-2} \varepsilon_i' M_F \bar{P}_{Z_i} a_1 \varepsilon_i$. By Lemmas A.3(iii) and A.1(i) and A.2, we have

$$\begin{aligned} R_{1NT,31} &= N^{-1/2} \sum_{i=1}^{N} \sigma_i^{-2} \text{tr} \left[\varepsilon_i' M_F \bar{P}_{Z_i} T^{-1} (\hat{F} - FH) H' F' \varepsilon_i \right] \\ &= N^{-1/2} \text{tr} \left[\sum_{i=1}^{N} \sigma_i^{-2} F' \varepsilon_i \varepsilon_i' M_F \bar{P}_{Z_i} T^{-1} (\hat{F} - FH) H' \right] \\ &\leq \left\| N^{-1/2} \sum_{i=1}^{N} \sigma_i^{-2} F' \varepsilon_i \varepsilon_i' M_F \bar{P}_{Z_i} \right\| T^{-1} \left\| \hat{F} - FH \right\| \|H\| \\ &= O_P(\sqrt{N+T}) O_P(T^{-1/2} (\delta_{NT}^{-1} + \gamma_{NT})) O_P(1) = o_P(1). \end{aligned}$$

By the Cauchy-Schwarz inequality, Lemmas A.1(i) and A.3(iii), we have

$$\begin{aligned} R_{1NT,32} &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sigma_i^{-2}\text{tr}\left[\boldsymbol{\varepsilon}_i' M_F \bar{P}_{Z_i} T^{-1}(\hat{F}-FH)(\hat{F}-FH)'\boldsymbol{\varepsilon}_i\right] \\ &= \text{tr}\left[T^{-1}(\hat{F}-FH)(\hat{F}-FH)' N^{-1/2}\sum_{i=1}^{N}\sigma_i^{-2}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i' M_F\bar{P}_{Z_i}\right] \\ &\leq T^{-1}\left\|\hat{F}-FH\right\|^2\left\|N^{-1/2}\sum_{i=1}^{N}\sigma_i^{-2}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i' M_F\bar{P}_{Z_i}\right\| \\ &= O_P\left(\delta_{NT}^{-2}+\gamma_{NT}^2\right)O_P\left(\sqrt{N+T}\right)=o_P(1). \end{aligned}$$

By the Lemma A.3(iv), $R_{1NT,33}=\Gamma_{6NT}=o_P(1)$. Analogously to the case of $R_{1NT,31}$, using Lemmas A.1(vi) and A.2 we can show $R_{1NT,34}=o_P(1)$. It follows that $R_{1NT,3}=o_P(1)$.

Now we prove $R_{1NT,1}\xrightarrow{d} N(0,V_0)$. Noting that $M_F\bar{P}_{Z_i}M_F=H_i$ and $Q_i\equiv\text{diag}(h_{i,11},h_{i,22},\ldots,h_{i,TT})$ we have

$$R_{1NT,1}=\frac{2}{\sqrt{N}}\sum_{i=1}^{N}\sigma_i^{-2}\sum_{1\leq s<t\leq T}\varepsilon_{it}\varepsilon_{is}h_{i,ts}\equiv\sum_{t=2}^{T}Z_{NT,t},$$

where $Z_{NT,t}\equiv 2N^{-1/2}\sum_{i=1}^{N}\sigma_i^{-2}\sum_{s=1}^{t-1}\varepsilon_{it}\varepsilon_{is}h_{i,ts}$. Let $\mathcal{F}_{NT,t}$ denote the σ-field generated by $\{X_1,X_2,\ldots,X_N,F,\Lambda,\boldsymbol{\varepsilon}_{\cdot t},\ldots,\boldsymbol{\varepsilon}_{\cdot 1}\}$ where recall $\boldsymbol{\varepsilon}_{\cdot t}\equiv(\varepsilon_{1t},\ldots,\varepsilon_{Nt})'$. By Assumptions A2(i) and A3(i), $\{Z_{NT,t},\ \mathcal{F}_{NT,t}\}$ is an m.d.s. because

$$E\left(Z_{NT,t}|\mathcal{F}_{NT,t-1}\right)\equiv 2N^{-1/2}\sum_{i=1}^{N}\sigma_i^{-2}\sum_{s=1}^{t-1}\varepsilon_{is}h_{i,ts}E(\varepsilon_{it}|\mathcal{F}_{NT,t-1})=0.$$

By the martingale CLT [e.g., Pollard (1984, p. 171)], it suffices to show that:

$$\mathcal{Z}\equiv\sum_{t=2}^{T}E_{\mathcal{F}_{NT,t-1}}\left|Z_{NT,t}\right|^4=o_P(1),\ \text{and}\ \sum_{t=2}^{T}Z_{NT,t}^2-V_{NT}=o_P(1). \tag{B.4}$$

where $E_{\mathcal{F}_{NT,t-1}}$ denotes expectation conditional on $\mathcal{F}_{NT,t-1}$. Using Assumptions A2 and A3(i)-(ii) we have

$$\begin{aligned} \mathcal{Z} &= \frac{16}{N^2}\sum_{t=2}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{l=1}^{N}\sum_{1\leq r,s,q,v\leq t-1}\sigma_i^{-2}\sigma_j^{-2}\sigma_k^{-2}\sigma_l^{-2}h_{i,ts}h_{j,tr}h_{k,tq}h_{l,tv}\varepsilon_{is}\varepsilon_{jr}\varepsilon_{kq}\varepsilon_{lv}E\left(\varepsilon_{it}\varepsilon_{jt}\varepsilon_{kt}\varepsilon_{lt}\right) \\ &= \frac{48}{N^2}\sum_{t=2}^{T}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\sum_{1\leq r,s,q,v\leq t-1}\sigma_i^{-2}\sigma_j^{-2}h_{i,ts}h_{i,tr}h_{j,tq}h_{j,tv}\varepsilon_{is}\varepsilon_{ir}\varepsilon_{jq}\varepsilon_{jv} \\ &\quad +\frac{16}{N^2}\sum_{t=2}^{T}\sum_{i=1}^{N}\sum_{1\leq r,s,q,v\leq t-1}\sigma_i^{-8}h_{i,ts}h_{i,tr}h_{i,tq}h_{i,tv}\varepsilon_{is}\varepsilon_{ir}\varepsilon_{iq}\varepsilon_{iv}E\left(\varepsilon_{it}^4\right) \\ &\equiv 48\mathcal{Z}_1+16\mathcal{Z}_2,\ \text{say}. \end{aligned}$$

Noting that $\mathcal{Z}\geq 0$, it suffices to show $\mathcal{Z}=o_P(1)$ by showing that $E_{\mathcal{D}}(\mathcal{Z})=48E_{\mathcal{D}}(\mathcal{Z}_1)+16E_{\mathcal{D}}(\mathcal{Z}_2)=$

$o_P(1)$ by the conditional Markov inequality. By straightforward calculations and Lemma A.4(v),

$$\begin{aligned} E_{\mathcal{D}}(\mathcal{Z}_1) &= \frac{1}{N^2}\sum_{t=2}^{T}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\sum_{1\leq r,s,q,v\leq t-1}\sigma_i^{-2}\sigma_j^{-2}h_{i,ts}h_{i,tr}h_{j,tq}h_{j,tv}E(\varepsilon_{is}\varepsilon_{ir})E(\varepsilon_{jq}\varepsilon_{jv}) \\ &= \frac{1}{N^2}\sum_{t=2}^{T}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\sum_{s=1}^{t-1}\sum_{r=1}^{t-1}h_{i,ts}^2h_{j,tr}^2 = o_P(1), \end{aligned}$$

and

$$\begin{aligned} E_{\mathcal{D}}(\mathcal{Z}_2) &= 3\frac{1}{N^2}\sum_{t=2}^{T}\sum_{i=1}^{N}\sum_{s=1}^{t-1}\sum_{r=1,r\neq s}^{t-1}\sigma_i^{-4}h_{i,ts}^2h_{i,tr}^2E(\varepsilon_{it}^4)+\frac{1}{N^2}\sum_{t=2}^{T}\sum_{i=1}^{N}\sum_{s=1}^{t-1}\sigma_i^{-8}h_{i,ts}^4E(\varepsilon_{is}^4)E(\varepsilon_{it}^4) \\ &\leq \frac{3\underline{c}_\sigma^{-2}M^{1/2}}{N^2}\sum_{t=2}^{T}\sum_{i=1}^{N}\sum_{s=1}^{t-1}\sum_{r=1,r\neq s}^{t-1}h_{i,ts}^2h_{i,tr}^2+\frac{\underline{c}_\sigma^{-4}M}{N^2}\sum_{t=2}^{T}\sum_{i=1}^{N}\sum_{s=1}^{t-1}h_{i,ts}^4 = o_P(1), \end{aligned}$$

where we used the fact that $E(\varepsilon_{it}^4)\leq[E(\varepsilon_{it}^8)]^{1/2}\leq M^{1/2}$. It follows that $E_{\mathcal{D}}(\mathcal{Z})=o_P(1)$ and thus $\mathcal{Z}=o_P(1)$. Consequently the first part of (B.4) follows.

For the second part of (B.4), we have

$$\begin{aligned} \sum_{t=2}^{T}E_{\mathcal{D}}(Z_{NT,t}^2) &= 4N^{-1}\sum_{t=2}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s=1}^{t-1}\sum_{r=1}^{t-1}\sigma_i^{-2}\sigma_j^{-2}h_{i,ts}h_{j,tr}E(\varepsilon_{it}\varepsilon_{is}\varepsilon_{jt}\varepsilon_{jr}) \\ &= 4N^{-1}\sum_{t=2}^{T}\sum_{i=1}^{N}\sum_{s=1}^{t-1}\sigma_i^{-4}h_{i,ts}^2E(\varepsilon_{it}^2\varepsilon_{is}^2) = 4N^{-1}\sum_{t=2}^{T}\sum_{i=1}^{N}\sum_{s=1}^{t-1}h_{i,ts}^2 = V_{NT}, \end{aligned}$$

where the second and third equalities follow from Assumptions A2(ii) and A3(i) and Assumption A3(ii), respectively. In addition, we can show by straightforward moment calculations that $E_{\mathcal{D}}(\sum_{t=2}^{T}Z_{NT,t}^2)^2=V_{NT}^2+o_P(1)$. Thus $\mathrm{Var}_{\mathcal{D}}(\sum_{t=2}^{T}Z_{NT,t}^2)=o_P(1)$ and the second part of (B.4) follows.

(ii) Next, we prove (B.2). We first consider R_{2NT}. By Lemma A.2, we have the following decomposition:

$$\begin{aligned} R_{2NT} &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sigma_i^{-2}b_1'X_i'M_{\hat{F}}\bar{P}_{Z_i}M_{\hat{F}}X_ib_1+\frac{2}{\sqrt{N}}\sum_{i=1}^{N}\sigma_i^{-2}b_1'X_i'M_{\hat{F}}\bar{P}_{Z_i}M_{\hat{F}}X_ib_2 \\ &\quad +\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sigma_i^{-2}b_2'X_i'M_{\hat{F}}\bar{P}_{Z_i}M_{\hat{F}}X_ib_2 \\ &\equiv R_{2NT,1}+R_{2NT,2}+R_{2NT,3},\ \text{say.} \end{aligned}$$

where $b_1=o_P(\gamma_{NT})$ and $b_2=D(F)^{-1}\frac{\gamma_{NT}}{NT}\sum_{i=1}^{N}\Pi_i'X_i\delta_i$. Noting that $\lambda_{\max}(\bar{P}_{Z_i})=1$ and $\lambda_{\max}(M_{\hat{F}})=1$ and using (A.7) repeatedly, we have

$$\begin{aligned} R_{2NT,1} &\leq \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sigma_i^{-2}b_1'X_i'X_ib_1\leq\underline{c}_\sigma^{-1}\|b_1\|^2\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\|X_i\|^2\right) \\ &= o_P(T^{-1}N^{-1/2})O_P(N^{1/2}T)=o_P(1). \end{aligned}$$

Using Lemma A.1(vii), we can easily show that

$$R_{2NT,3} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sigma_i^{-2} b_2' X_i' M_F \bar{P}_{Z_i} M_F X_i b_2 + o_P(1) = O_P(1).$$

Then $R_{2NT,3} = o_P(1)$ by the Cauchy-Schwarz inequality and we have

$$R_{2NT} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sigma_i^{-2} b_2' X_i' M_F \bar{P}_{Z_i} M_F X_i b_2 + o_P(1). \tag{B.5}$$

For R_{3NT}, we have

$$\begin{aligned} R_{3NT} &= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sigma_i^{-2} \lambda_i' (F - \hat{F} H^{-1})' M_{\hat{F}} \bar{P}_{Z_i} M_{\hat{F}} (F - \hat{F} H^{-1}) \lambda_i \\ &= \frac{1}{T\sqrt{N}} \sum_{i=1}^{N} \sigma_i^{-2} \lambda_i' (F - \hat{F} H^{-1})' M_{\hat{F}} Z_i \left(T^{-1} Z_i' Z_i\right)^{-1} Z_i' M_{\hat{F}} (F - \hat{F} H^{-1}) \lambda_i \\ &\quad - \frac{1}{T\sqrt{N}} \sum_{i=1}^{N} \sigma_i^{-2} \lambda_i' (F - \hat{F} H^{-1})' M_{\hat{F}} \mathbf{i}_T \mathbf{i}_T' M_{\hat{F}} (F - \hat{F} H^{-1}) \lambda_i \\ &\equiv R_{3NT,1} - R_{3NT,2}, \text{ say.} \end{aligned}$$

Using Lemma A.1(iv) yields

$$\begin{aligned} R_{3NT,1} &= \frac{T}{\sqrt{N}} \sum_{i=1}^{N} \sigma_i^{-2} \lambda_i' c_1' \left(T^{-1} Z_i' Z_i\right)^{-1} c_1 \lambda_i + \frac{2T}{\sqrt{N}} \sum_{i=1}^{N} \sigma_i^{-2} \lambda_i' c_1' \left(T^{-1} Z_i' Z_i\right)^{-1} c_{i2} \lambda_i \\ &\quad + \frac{T}{\sqrt{N}} \sum_{i=1}^{N} \sigma_i^{-2} \lambda_i' c_{i2}' \left(T^{-1} Z_i' Z_i\right)^{-1} c_{i2} \lambda_i \\ &\equiv R_{3NT,11} + R_{3NT,12} + R_{3NT,13}, \text{ say,} \end{aligned}$$

where $c_1 = O_P(||\hat{\beta} - \beta||) + O_P(\delta_{NT}^{-2})$ and $c_{i2} = -Z_i' M_{\hat{F}} \frac{\gamma_{NT}}{NT} \sum_{k=1}^{N} X_k \delta_k \lambda_k \left(\Lambda' \Lambda / N\right)^{-1}$. Using Lemma A.2, we can readily show that $R_{3NT,11} = O_P(T N^{1/2} \delta_{NT}^{-4}) = o_P(1)$. Using Lemma A.1(vii), we can show that

$$\begin{aligned} R_{3NT,13} &= \frac{1}{NT} \sum_{i=1}^{N} \sigma_i^{-2} \left(\frac{1}{N} \sum_{k=1}^{N} a_{ik} X_k \delta_k \right)' M_{\hat{F}} P_{Z_i} M_{\hat{F}} \left(\frac{1}{N} \sum_{k=1}^{N} a_{ik} X_k \delta_k \right) \\ &= \frac{1}{NT} \sum_{i=1}^{N} \sigma_i^{-2} \left(\frac{1}{N} \sum_{k=1}^{N} a_{ik} X_k \delta_k \right)' M_F P_{Z_i} M_F \left(\frac{1}{N} \sum_{k=1}^{N} a_{ik} X_k \delta_k \right) + o_P(1) \\ &= O_P(1). \end{aligned}$$

Then $R_{3NT,12} = o_P(1)$ by the Cauchy-Schwarz inequality and

$$R_{3NT,1} = \frac{1}{NT} \sum_{i=1}^{N} \sigma_i^{-2} \left(\frac{1}{N} \sum_{k=1}^{N} a_{ik} X_k \delta_k \right)' M_F P_{Z_i} M_F \left(\frac{1}{N} \sum_{k=1}^{N} a_{ik} X_k \delta_k \right) + o_P(1).$$

By the same token, we can obtain

$$R_{3NT,2} = \frac{1}{NT}\sum_{i=1}^{N}\sigma_i^{-2}\left(\frac{1}{N}\sum_{k=1}^{N}a_{ik}X_k\delta_k\right)' M_F L M_F \left(\frac{1}{N}\sum_{k=1}^{N}a_{ik}X_k\delta_k\right) + o_P(1).$$

It follows that

$$R_{3NT} = \frac{1}{NT}\sum_{i=1}^{N}\sigma_i^{-2}\left(\frac{1}{N}\sum_{k=1}^{N}a_{ik}X_k\delta_k\right)' M_F \bar{P}_{Z_i} M_F \left(\frac{1}{N}\sum_{k=1}^{N}a_{ik}X_k\delta_k\right) + o_P(1). \quad \text{(B.6)}$$

For R_{4NT}, by Lemma A.1(vii) we have

$$\begin{aligned} R_{4NT} &= \frac{1}{NT}\sum_{i=1}^{N}\sigma_i^{-2}\delta_i' X_i' M_{\hat{F}} \bar{P}_{Z_i} M_{\hat{F}} X_i \delta_i \\ &= \frac{1}{NT}\sum_{i=1}^{N}\sigma_i^{-2}\delta_i' X_i' M_F \bar{P}_{Z_i} M_F X_i \delta_i + o_P(1) = O_P(1). \end{aligned} \quad \text{(B.7)}$$

For R_{8NT}, R_{9NT} and R_{10NT}, by the Cauchy-Schwarz inequality and analogous arguments as used above we obtain

$$R_{8NT} = \frac{1}{NT}\sum_{i=1}^{N}\sigma_i^{-2}\left(D(F)^{-1}\frac{1}{NT}\sum_{k=1}^{N}\Pi_k' X_k\delta_k\right)' X_i' M_F \bar{P}_{Z_i} M_F \left(\frac{1}{N}\sum_{k=1}^{N}a_{ik}X_k\delta_k\right) + o_P(1), \quad \text{(B.8)}$$

$$R_{9NT} = -\frac{1}{NT}\sum_{i=1}^{N}\sigma_i^{-2}\left(D(F)^{-1}\frac{1}{NT}\sum_{k=1}^{N}\Pi_k' X_k\delta_k\right)' X_i' M_F \bar{P}_{Z_i} M_F X_i\delta_i + o_P(1), \text{ and} \quad \text{(B.9)}$$

$$R_{10NT} = -\frac{1}{NT}\sum_{i=1}^{N}\sigma_i^{-2}\left(\frac{1}{N}\sum_{k=1}^{N}a_{ik}X_k\delta_k\right)' X_i' M_F \bar{P}_{Z_i} M_F X_i\delta_i + o_P(1). \quad \text{(B.10)}$$

Combining (B.5)-(B.10) yields (B.2).

(iii) Now, we prove (B.3). For R_{5NT}, we have

$$R_{5NT} = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sigma_i^{-2}\boldsymbol{\varepsilon}_i'(M_{\hat{F}} - M_F)\bar{P}_{Z_i} M_{\hat{F}} X_i(\hat{\beta}-\beta) + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sigma_i^{-2}\boldsymbol{\varepsilon}_i' M_F \bar{P}_{Z_i} M_{\hat{F}} X_i(\hat{\beta}-\beta).$$

The first term in absolute value is bounded by $\{\Gamma_{1NT}\}^{1/2} \times \{R_{2NT}\}^{1/2} = o_P(1)$ by the Cauchy-Schwarz inequality, Lemma A.3(i), and (B.5). To show that the second term is $o_P(1)$, it suffices to demonstrate that $R_{5NT,1} \equiv N^{-1/2}\sum_{i=1}^{N}\sigma_i^{-2}\boldsymbol{\varepsilon}_i' M_F \bar{P}_{Z_i} M_{\hat{F}} X_i$ is $o_P\left(\gamma_{NT}^{-1}\right)$ by Lemma A.2. We further decompose $R_{5NT,1}$ as follows

$$\begin{aligned} R_{5NT,1} &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sigma_i^{-2}\boldsymbol{\varepsilon}_i' M_F \bar{P}_{Z_i} M_F X_i + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sigma_i^{-2}\boldsymbol{\varepsilon}_i' M_F \bar{P}_{Z_i}(M_{\hat{F}} - M_F)X_i \\ &\equiv R_{5NT,11} + R_{5NT,12}, \text{ say.} \end{aligned}$$

Observe that $E(R_{5NT,11}) = 0$ and by the repeated use of (A.7),

$$\begin{aligned} E\left\|R_{5NT,11}\right\|^2 &= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\sigma_i^{-2}\sigma_j^{-2}E\left[\operatorname{tr}\left(M_F\bar{P}_{Z_j}M_FX_jX_i'M_F\bar{P}_{Z_i}M_FE\left(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_j'\right)\right)\right] \\ &= \frac{1}{N}\sum_{i=1}^{N}\sigma_i^{-2}E\left[\operatorname{tr}\left(X_i'M_F\bar{P}_{Z_i}M_F\bar{P}_{Z_i}M_FX_i\right)\right] \\ &\le \frac{1}{N}\sum_{i=1}^{N}\sigma_i^{-2}E\left[\operatorname{tr}\left(X_i'X_i\right)\right] = \frac{\underline{c}_\sigma^{-1}}{N}\sum_{i=1}^{N}E\left\|X_i\right\|^2 = O\left(T\right). \end{aligned}$$

It follows that $R_{5NT,11} = O_P(\sqrt{T}) = o_P\left(\gamma_{NT}^{-1}\right)$. By the Cauchy-Schwarz inequality and Lemmas A.1(vii) and A.2,

$$\begin{aligned} \left\|R_{5NT,12}\right\| &\le \frac{\underline{c}_\sigma^{-1}}{\sqrt{N}}\left\|M_{\hat{F}}-M_F\right\|\sum_{i=1}^{N}\left\|\boldsymbol{\varepsilon}_i'M_F\bar{P}_{Z_i}\right\|\left\|X_i\right\| \\ &\le \underline{c}_\sigma^{-1}\left\|M_{\hat{F}}-M_F\right\|\left(\sum_{i=1}^{N}\left\|\boldsymbol{\varepsilon}_i'M_F\bar{P}_{Z_i}\right\|^2\right)^{1/2}\left(\frac{1}{N}\sum_{i=1}^{N}\left\|X_i\right\|^2\right)^{1/2} \\ &= O_P(T^{-1/2}+N^{-1/2})O_P(N^{1/2})O_P(T^{1/2}) = o_P\left(\gamma_{NT}^{-1}\right). \end{aligned}$$

It follows that $R_{5NT,1} = o_P\left(\gamma_{NT}^{-1}\right)$ and $R_{5NT} = o_P\left(1\right)$.

For R_{6NT}, we write

$$R_{6NT} = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sigma_i^{-2}\boldsymbol{\varepsilon}_i'(M_{\hat{F}}-M_F)\bar{P}_{Z_i}M_{\hat{F}}F\lambda_i + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sigma_i^{-2}\boldsymbol{\varepsilon}_i'M_F\bar{P}_{Z_i}M_{\hat{F}}\left(F-\hat{F}H^{-1}\right)\lambda_i.$$

By the Cauchy-Schwarz inequality, the first term in absolute value is bounded by $\{\Gamma_{1NT}\}^{1/2}\times\{R_{3NT}\}^{1/2} = o_P(1)$ by Lemmas A.3(i) and (B.6). Denoting the second term as $R_{6NT,1}$, we decompose it as follows

$$\begin{aligned} R_{6NT,1} &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sigma_i^{-2}\boldsymbol{\varepsilon}_i'M_F\bar{P}_{Z_i}\left(M_{\hat{F}}-M_F\right)\left(F-\hat{F}H^{-1}\right)\lambda_i \\ &\quad -\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sigma_i^{-2}\boldsymbol{\varepsilon}_i'M_F\bar{P}_{Z_i}M_F\left(F-\hat{F}H^{-1}\right)\lambda_i \\ &\equiv R_{6NT,11} - R_{6NT,12},\ \text{say.} \end{aligned}$$

By Lemmas A.1(i) and (vii) and Lemma A.2,

$$\begin{aligned} \left\|R_{6NT,11}\right\| &= \operatorname{tr}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sigma_i^{-2}\lambda_i\boldsymbol{\varepsilon}_i'M_F\bar{P}_{Z_i}\left(M_{\hat{F}}-M_F\right)\left(F-\hat{F}H^{-1}\right)\right] \\ &\le \left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sigma_i^{-2}\lambda_i\boldsymbol{\varepsilon}_i'M_F\bar{P}_{Z_i}\right\|\left\|M_{\hat{F}}-M_F\right\|\left\|F-\hat{F}H^{-1}\right\| \\ &= O_P(1)O_P(\delta_{NT}^{-1})O_P(T^{1/2}\delta_{NT}^{-1}) = o_P(1). \end{aligned}$$

By Lemma A.3(v), $R_{6NT,12} = \Gamma_{5NT} = o_P(1)$. It follows that $R_{6NT,1} = o_P\left(1\right)$ and $R_{6NT} = o_P\left(1\right)$. Analogously to the analysis of R_{5NT}, we can show that $R_{7NT} = o_P\left(1\right)$. This completes the proof of (B.3). ■

Proposition B.2 *Under the conditions of Theorem 3.3, $J_{NT,2} = o_P(1)$.*

Proof. By a geometric expansion, $1/\hat{\sigma}_i^2 - 1/\sigma_i^2 = -(\hat{\sigma}_i^2 - \sigma_i^2)/\sigma_i^4 + (\hat{\sigma}_i^2 - \sigma_i^2)^2/(\sigma_i^4\hat{\sigma}_i^2)$. It follows that

$$\begin{aligned} J_{NT,2} &= -\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\hat{\boldsymbol{\varepsilon}}_i' \bar{P}_{Z_i} \hat{\boldsymbol{\varepsilon}}_i - \boldsymbol{\varepsilon}_i' Q_i \boldsymbol{\varepsilon}_i\right) \frac{\hat{\sigma}_i^2 - \sigma_i^2}{\sigma_i^4} + \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\hat{\boldsymbol{\varepsilon}}_i' \bar{P}_{Z_i} \hat{\boldsymbol{\varepsilon}}_i - \boldsymbol{\varepsilon}_i' Q_i \boldsymbol{\varepsilon}_i\right) \frac{\left(\hat{\sigma}_i^2 - \sigma_i^2\right)^2}{\sigma_i^4 \hat{\sigma}_i^2} \\ &\equiv -J_{NT,21} + J_{NT,22}, \text{ say.} \end{aligned}$$

Using $\hat{\boldsymbol{\varepsilon}}_i = M_{\hat{F}} \boldsymbol{\varepsilon}_i + M_{\hat{F}} X_i(\beta - \hat{\beta}) + M_{\hat{F}} F \lambda_i + M_{\hat{F}} X_i(\beta_i - \beta)$, Lemmas A.1(vii) and A.2, and the conditional Markov inequality, we can show that under (3.2),

$$\begin{aligned} N^{-1} \sum_{i=1}^{N} \left(\hat{\boldsymbol{\varepsilon}}_i' \bar{P}_{Z_i} \hat{\boldsymbol{\varepsilon}}_i - \boldsymbol{\varepsilon}_i' Q_i \boldsymbol{\varepsilon}_i\right)^2 &= N^{-1} \sum_{i=1}^{N} \left(\boldsymbol{\varepsilon}_i' M_F \bar{P}_{Z_i} M_F \boldsymbol{\varepsilon}_i - \boldsymbol{\varepsilon}_i' Q_i \boldsymbol{\varepsilon}_i\right)^2 + o_P(1) \\ &= 4N^{-1} \sum_{i=1}^{N} \left(\sum_{1 \le t < s \le T} \varepsilon_{it} \varepsilon_{is} h_{i,ts} \right)^2 + o_P(1) = O_P(1). \quad \text{(B.11)} \end{aligned}$$

Then by Lemma A.5, (B.11), and Assumption A4, we have

$$\begin{aligned} J_{NT,22} &\le \frac{\max_{1 \le i \le N} \left|\hat{\sigma}_i^2 - \sigma_i^2\right|^2}{\min_{1 \le i \le N} \sigma_i^4 \hat{\sigma}_i^2} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left|\hat{\boldsymbol{\varepsilon}}_i' \bar{P}_{Z_i} \hat{\boldsymbol{\varepsilon}}_i - \boldsymbol{\varepsilon}_i' Q_i \boldsymbol{\varepsilon}_i\right| \\ &\le \frac{\sqrt{N} \max_{1 \le i \le N} \left|\hat{\sigma}_i^2 - \sigma_i^2\right|^2}{\min_{1 \le i \le N} \sigma_i^4 \hat{\sigma}_i^2} \left[\frac{1}{N} \sum_{i=1}^{N} \left(\hat{\boldsymbol{\varepsilon}}_i' \bar{P}_{Z_i} \hat{\boldsymbol{\varepsilon}}_i - \boldsymbol{\varepsilon}_i' Q_i \boldsymbol{\varepsilon}_i\right)^2 \right]^{1/2} \\ &= \sqrt{N} O_P(\upsilon_{NT}^2 + NT^{-2} + \gamma_{NT}^2 + \delta_{NT}^{-2} N^{1/\vartheta}) O_P(1) = o_P(1). \end{aligned}$$

For $J_{NT,21}$, we have $J_{NT,21} = \sum_{l=1}^{10} J_{NT,21l}$, where

$$\begin{aligned} J_{NT,211} &= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sigma_i^{-4} (\hat{\boldsymbol{\varepsilon}}_i' \bar{P}_{Z_i} \hat{\boldsymbol{\varepsilon}}_i - \boldsymbol{\varepsilon}_i' Q_i \boldsymbol{\varepsilon}_i)(T^{-1} TSS_{i1} - \sigma_i^2), \text{ and} \\ J_{NT,21l} &= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sigma_i^{-4} (\hat{\boldsymbol{\varepsilon}}_i' \bar{P}_{Z_i} \hat{\boldsymbol{\varepsilon}}_i - \boldsymbol{\varepsilon}_i' Q_i \boldsymbol{\varepsilon}_i)(T^{-1} TSS_{il}) \text{ for } l = 2, 3, \ldots 10, \end{aligned}$$

where TSS_{il}, $l = 1, ..., 10$, are defined after (A.11). Following the same steps in the proof of Proposition B.1 and the analysis for TSS_{il} in the proof of Lemma A.5, we can show that $J_{NT,21l} = o_P(1)$ for all $l = 1, ..., 10$. ■

Proof of Theorem 3.4.

Let P^* denote the probability conditional on the original sample $\mathcal{W}_{NT} \equiv \{(Y_i, X_i), i = 1, ..., N\}$ and E^* and Var* denote the expectation and variance with respect to P^*. Let $O_{P^*}(\cdot)$ and $o_{P^*}(\cdot)$ denote the probability order under P^*, for example, $a_{NT} = o_{P^*}(1)$ if for any $\epsilon > 0$, $P^*(|a_{NT}| > \epsilon) = o_P(1)$. Note that $a_{NT} = o_P(1)$ implies that $a_{NT} = o_{P^*}(1)$.

Observing that $Y_{it}^* = \hat{\beta}' X_{it} + \hat{\lambda}_i' \hat{F}_t + \varepsilon_{it}^*$, the null hypothesis is maintained in the bootstrap world. Given $\mathcal{W}_{NT}$, ε_{it}^* are independent across i, and are independent of X_{js}, $\hat{\lambda}_j$, and $\hat{F}_s$ for all i, t, j, s, because the latter objects are fixed in the bootstrap world. Let $\boldsymbol{\varepsilon}_i^* \equiv (\varepsilon_{i1}^*, ..., \varepsilon_{iT}^*)'$. Let $\mathcal{F}_t(\boldsymbol{\varepsilon}_i^*)$ denote the σ-field generated by $\{\varepsilon_{it}^*, ..., \varepsilon_{i1}^*\}$. For each i, $\{\varepsilon_{it}^*, \mathcal{F}_t(\boldsymbol{\varepsilon}_i^*)\}$ is also an m.d.s. such that $E^*(\varepsilon_{it}^* | \mathcal{F}_{t-1}(\boldsymbol{\varepsilon}_i^*)) = E^*(\varepsilon_{it}^*) = T^{-1}\sum_{t=1}^T (\hat{\varepsilon}_{it} - \bar{\hat{\varepsilon}}_i) = 0$, and $E^*[(\varepsilon_{it}^*)^2 | \mathcal{F}_{t-1}(\boldsymbol{\varepsilon}_i^*)] = E^*[(\varepsilon_{it}^*)^2] = T^{-1}\sum_{t=1}^T (\hat{\varepsilon}_{it} - \bar{\hat{\varepsilon}}_i)^2 = \hat{\sigma}_i^2$. Under either $\mathbb{H}_0$ or $\mathbb{H}_{1,NT}$, Lemma A.5 indicates that $\hat{\sigma}_i^2$ is uniformly bounded and bounded away from 0 with probability approaching 1 as $(N,T) \to \infty$. In addition, letting $\xi_{it}^* \equiv \varepsilon_{it}^* - \hat{\sigma}_i^2$, we can verify that $E^* |\xi_{it}^*|^\vartheta$ exists provided $E|\xi_{it}|^\vartheta$ exists and $\frac{1}{NT^{\vartheta/2}}\sum_{i=1}^N \sum_{1\le t_1,t_2,...,t_\vartheta \le T} E^*(\xi_{it_1}^* \xi_{it_2}^* ... \xi_{it_\vartheta}^*) = O_{P^*}(1)$ by the serial independence of $\{\varepsilon_{it}^*, t \ge 1\}$ and thus $\{\xi_{it}^*, t \ge 1\}$. Thus we have verified that Assumptions A2(i)-(ii) and A3(i)-(iii) are satisfied in the bootstrap world.

Note that the bootstrap analogue of $\{X_{it}, \lambda_i, F_t\}$ is $\{X_{it}, \hat{\lambda}_i, \hat{F}_t\}$ which is known given $\mathcal{W}_{NT}$. The conditions on $\{X_{it}\}$ alone in Assumptions A1(i), (iv) and (v) remain satisfied in the bootstrap world. Under either $\mathbb{H}_0$ or $\mathbb{H}_{1,NT}$, using Lemmas A.1 and A.2 we can show that $T^{-1}\sum_{t=1}^T \hat{F}_t \hat{F}_t' = \Sigma_F + o_P(1)$, $N^{-1}\hat{\Lambda}'\hat{\Lambda} = \Sigma_\Lambda + o_P(1)$, $\frac{1}{N^{\vartheta/2}}\sum_{1\le i_1,i_2,...,i_\vartheta \le N} \hat{\zeta}_{i_1}...\hat{\zeta}_{i_\vartheta} = O_P(1)$, and $D(\hat{F}) = D(F) + o_P(1)$. This indicates that the other conditions in Assumption A1 are also met in the bootstrap world. Note that Assumption A2(ii) is mainly needed to simplify the calculation of the asymptotic variance of J_{NT} in the proof of Proposition B.1.

By the above discussions, we can verify that Lemmas A.1, A.3, and A.4 remain valid in the bootstrap world by replacing $\{F, \hat{F}, H, H_i, \boldsymbol{\varepsilon}_i, \lambda_i, \beta, \hat{\beta}, \sigma_i^2, \delta_i, \gamma_{NT}\}$, $O_P(\cdot)$ and $o_P(\cdot)$ by $\{\hat{F}, \hat{F}^*, H^*, H_i^*, \boldsymbol{\varepsilon}_i^*, \hat{\lambda}_i, \hat{\beta}, \hat{\beta}^*, \hat{\sigma}_i^2, 0, 0\}$, $O_{P^*}(\cdot)$ and $o_{P^*}(\cdot)$, respectively, where $\hat{F}^* \equiv (\hat{F}_1^*, ..., \hat{F}_T^*)'$, $H^* \equiv (\hat{\Lambda}'\hat{\Lambda}/N)(\hat{F}'\hat{F}^*/T)V_{NT}^{*-1}$, V_{NT}^* satisfies $[\frac{1}{NT}\sum_{i=1}^N (Y_i^* - X_i\hat{\beta}^*)(Y_i^* - X_i\hat{\beta}^*)']\hat{F}^* = \hat{F}^* V_{NT}^*$, $Y_i^* \equiv (Y_{i1}^*, ..., Y_{iT}^*)'$, and $H_i^* \equiv M_{\hat{F}}(P_{Z_i} - L)M_{\hat{F}}$. The results in Lemmas A.2 and A.5 now become $\hat{\beta}^* - \hat{\beta} = O_{P^*}(N^{-1/2}T^{-1/2})$, and $\max_{1\le i\le N} |\hat{\sigma}_i^{*2} - \hat{\sigma}_i^2| = O_{P^*}(\upsilon_{NT} + N^{1/2}T^{-1} + \delta_{NT}^{-1}N^{1/(2\vartheta)})$, where $\hat{\sigma}_i^{*2} \equiv T^{-1}TSS_i^* \equiv T^{-1}\sum_{t=1}^T (\hat{\varepsilon}_{it}^* - \bar{\hat{\varepsilon}}_i^*)^2$, $\hat{\varepsilon}_{it}^* \equiv Y_{it}^* - \hat{\beta}^{*\prime} X_{it} - \hat{\lambda}_i^{*\prime}\hat{F}_t^*$, and $\bar{\hat{\varepsilon}}_i^* \equiv T^{-1}\sum_{t=1}^T \hat{\varepsilon}_{it}^*$.

Let $\bar{R}_{NT}^{*2}$, J_{NT}^*, B_{NT}^*, V_{NT}^*, $\hat{B}_{NT}^*$, and $\hat{V}_{NT}^*$ denote the bootstrap analogue of $\bar{R}_{NT}^2$, J_{NT}, B_{NT}, V_{NT}, $\hat{B}_{NT}$, and $\hat{V}_{NT}$, respectively. Then $J_{NT}^* \equiv (\sqrt{N}T\bar{R}_{NT}^{*2} - B_{NT}^*)/\sqrt{V_{NT}^*}$ and $\hat{J}_{NT}^* \equiv (\sqrt{N}T\bar{R}_{NT}^{*2} - \hat{B}_{NT}^*)/\sqrt{\hat{V}_{NT}^*}$. As in the proof of Theorem 3.3, we have

$$\begin{aligned} J_{NT}^* &= \frac{1}{\sqrt{N}}\sum_{i=1}^N \frac{\hat{\boldsymbol{\varepsilon}}_i^{*\prime}\bar{P}_{Z_i}\hat{\boldsymbol{\varepsilon}}_i^*}{\hat{\sigma}_i^{*2}} - \frac{1}{\sqrt{N}}\sum_{i=1}^N\sum_{t=1}^T \frac{\varepsilon_{it}^{*2}h_{i,tt}^*}{\hat{\sigma}_i^{*2}} \\ &= \frac{1}{\sqrt{N}}\sum_{i=1}^N \hat{\sigma}_i^{-2}(\hat{\boldsymbol{\varepsilon}}_i^{*\prime}\bar{P}_{Z_i}\hat{\boldsymbol{\varepsilon}}_i^* - \boldsymbol{\varepsilon}_i^{*\prime}Q_i^*\boldsymbol{\varepsilon}_i^*) + \frac{1}{\sqrt{N}}\sum_{i=1}^N (\hat{\boldsymbol{\varepsilon}}_i^{*\prime}\bar{P}_{Z_i}\hat{\boldsymbol{\varepsilon}}_i^* - \boldsymbol{\varepsilon}_i^{*\prime}Q_i^*\boldsymbol{\varepsilon}_i^*)(\hat{\sigma}_i^{*-2} - \hat{\sigma}_i^{-2}) \\ &\equiv J_{NT,1}^* + J_{NT,2}^*, \text{ say,} \end{aligned}$$

where $Q_i^* \equiv \text{diag}(h_{i,11}^*, h_{i,22}^*, \ldots, h_{i,TT}^*)$, $h_{i,ts}^*$ is the (t,s) element of H_i^*. We prove the theorem by showing that: (i) $J_{NT,1}^* \xrightarrow{d} N(0, V_0)$ conditional on $\mathcal{W}_{NT}$, (ii) $J_{NT,2}^* = o_{P^*}(1)$, (iii) $\hat{B}_{NT}^* = B_{NT}^* + o_{P^*}(1)$, and (iv) $\hat{V}_{NT}^* = V_{NT}^* + o_{P^*}(1)$.

We only outline the proof of (i) as the proofs of other parts are analogous to those in the proof of Theorem 3.3. By (2.8), we have

$$J_{NT,1}^* = R_{1NT}^* + R_{2NT}^* + R_{3NT}^* + 2R_{5NT}^* + 2R_{6NT}^* + 2R_{8NT}^*,$$

where R_{lNT}^* is the bootstrap analogue of R_{lNT} for $l = 1,2,3,5,6,8$, e.g., $R_{1NT}^* \equiv N^{-1/2}\sum_{i=1}^N \hat{\sigma}_i^{-2}(\boldsymbol{\varepsilon}_i^{*\prime}M_{\hat{F}^*}\bar{P}_{Z_i}M_{\hat{F}^*}\boldsymbol{\varepsilon}_i^* - \boldsymbol{\varepsilon}_i^{*\prime}Q_i^*\boldsymbol{\varepsilon}_i^*)$. Analogously to the proof of Proposition B.1, we can show that $R_{lNT}^* = o_{P^*}(1)$ for

$l=2,3,5,6,8,$ and $R_{1NT}^{*}=\sum_{t=2}^{T}Z_{NT,t}^{*}+o_{P^{*}}(1)$, where $Z_{NT,t}^{*}\equiv 2N^{-1/2}\sum_{i=1}^{N}\hat{\sigma}_{i}^{-2}\sum_{s=1}^{t-1}\varepsilon_{it}^{*}\varepsilon_{is}^{*}h_{i,ts}^{*}$. Let $\mathcal{F}_{NT,t}^{*}$ denote the σ-field generated by $\{\mathcal{W}_{NT},\boldsymbol{\varepsilon}_{\cdot t}^{*},...,\boldsymbol{\varepsilon}_{\cdot 1}^{*}\}$ where recall $\boldsymbol{\varepsilon}_{\cdot t}^{*}\equiv(\varepsilon_{1t}^{*},...,\varepsilon_{Nt}^{*})'$. Then $\{Z_{NT,t}^{*},\mathcal{F}_{NT,t}^{*}\}$ is an m.d.s. because $E\left(Z_{NT,t}^{*}|\mathcal{F}_{NT,t-1}^{*}\right)\equiv 2N^{-1/2}\sum_{i=1}^{N}\hat{\sigma}_{i}^{-2}\sum_{s=1}^{t-1}\varepsilon_{is}^{*}h_{i,ts}^{*}E(\varepsilon_{it}^{*}|\mathcal{F}_{NT,t-1}^{*})=0$. So we can continue to apply the martingale CLT in Pollard (1984, p. 171) by showing that

$$\mathcal{Z}^{*}\equiv\sum_{t=2}^{T}E_{\mathcal{F}_{NT,t-1}^{*}}\left|Z_{NT,t}^{*}\right|^{4}=o_{P^{*}}(1),\text{ and }\sum_{t=2}^{T}Z_{NT,t}^{*2}-V_{NT}^{*}=o_{P^{*}}(1). \tag{B.12}$$

where $E_{\mathcal{F}_{NT,t-1}^{*}}$ denotes expectation conditional on $\mathcal{F}_{NT,t-1}^{*}$. By direct calculations and the bootstrap version of Lemma A.4,

$$\begin{aligned}E^{*}\left(\mathcal{Z}^{*}\right) &= \frac{48}{N^{2}}\sum_{t=2}^{T}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\sum_{s=1}^{t-1}\sum_{r=1}^{t-1}h_{i,ts}^{*2}h_{j,tr}^{*2}+\frac{48}{N^{2}}\sum_{t=2}^{T}\sum_{i=1}^{N}\sum_{s=1}^{t-1}\sum_{r=1,r\neq s}^{t-1}\hat{\sigma}_{i}^{-4}h_{i,ts}^{*2}h_{i,tr}^{*2}E^{*}\left(\varepsilon_{it}^{*4}\right)\\ &\quad+\frac{16}{N^{2}}\sum_{t=2}^{T}\sum_{i=1}^{N}\sum_{s=1}^{t-1}\hat{\sigma}_{i}^{-8}h_{i,ts}^{*4}E^{*}\left(\varepsilon_{is}^{*4}\right)E^{*}\left(\varepsilon_{it}^{*4}\right)\\ &= o_{P}(1)\end{aligned}$$

where we use the fact that $E^{*}\left(\varepsilon_{it}^{*4}\right)=T^{-1}\sum_{t=1}^{T}\left(\hat{\varepsilon}_{it}-\bar{\hat{\varepsilon}}_{i}\right)^{4}=T^{-1}\sum_{t=1}^{T}\varepsilon_{it}^{4}+o_{P}(1)$ uniformly in i and $\hat{\sigma}_{i}^{-2}\leq 2\sigma_{i}^{-2}$ with probability arbitrarily close to 1 as $(N,T)\to\infty$ by Lemma A.5. It follows that $\mathcal{Z}^{*}=o_{P^{*}}(1)$. Now, $\sum_{t=1}^{T}E^{*}(Z_{NT,t}^{*2})=4N^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{s=1}^{t-1}h_{i,ts}^{*2}=V_{NT}^{*}$. In addition, we can show by straightforward moment calculations that $E^{*}(\sum_{t=2}^{T}Z_{NT,t}^{*2})^{2}=V_{NT}^{*2}+o_{P}(1)$. Thus $\text{Var}^{*}(\sum_{t=2}^{T}Z_{NT,t}^{*2})=o_{P}(1)$ and $\sum_{t=2}^{T}Z_{NT,t}^{*2}-V_{NT}^{*}=o_{P^{*}}(1)$. ■

References

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

[2] Bai, J., 2009b. Supplement to "Panel data models with interactive fixed effects": technical details and proofs, Econometrica Supplementary Material.

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

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

[5] Baltagi, B. H., G. Bresson, and A. Pirotte, 2008. To pool or not to pool? In L. Mátyás and P. Sevestre (eds.), The Econometrics of Panel Data, pp. 517-546, Springer-Verlag, Berlin.

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

[7] Blomquist, J., 2010. A panel bootstrap test for slope homogeneity. Working paper, Lund University.

[8] Connor, G. and R. Korajzcyk, 1998. Risk and return in an equilibrium APT application of a new test methodology. Journal of Financial Economics 21, 225-289.

[9] Fama, E., K. and K. French, 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 18, 61-90.

[10] Greenaway-McGrevy, R., C. Han, D. Sul, 2010. Asymptotic distribution of factor augmented estimators for panel regression. Forthcoming in Journal of Econometrics.

[11] Hsiao, C., 2003. Analysis of Panel Data. Cambridge University Press, Cambridge.

[12] Jin, S. and L. Su, 2011. A nonparametric poolability test for panel data models with cross section dependence. Working paper, Singapore Management University.

[13] Kapetanios, G., 2008. A bootstrap procedure for panel data sets with many cross-sectional units. Econometrics Journal 11, 377-395.

[14] Kapetanios, G. and M. H. Pesaran, 2007. Alternative approaches to estimation and inference in large multifactor panels: small sample results with an application to modelling of asset return, in G. Phillips and E. Tzavalis (eds.), The Refinement of Econometric Estimation and Testing Procedures: Finite Sample and Asymptotic Analysis, Cambridge University Press, Cambridge.

[15] Kim, M., 2010. Heteroskedasticity and spatiotemporal dependence robust inference for linear panel models with fixed effects. Working paper, UCSD.

[16] Lehmann, B. and D. Modest, 1988. The empirical foundations of the arbitrage pricing theory. Journal of Financial Economics 21, 213-254.

[17] Lin, C-C., 2010. Testing for slope homogeneity in a linear panel model with fixed effects and conditional heteroskedasticity. Working paper, Institute of Economics, Academia Sinica.

[18] Moon, H. and M. Weidner, 2010a. Linear regression for panel with unknown number of factors as interactive fixed effects. Unpublished manuscript, University of Southern California.

[19] Moon, H. and M. Weidner, 2010b. Dynamic linear panel regression models with interactive fixed effects. Manuscript. Unpublished manuscript, University of Southern California.

[20] Pesaran, M. H., 2006. Estimation and inference in large heterogeneous panels with a multifactor error structure. Econometrica 74, 967-1012.

[21] Pesaran, H., R. Smith, and K. S. Im, 1996. Dynamic linear models for heterogeneous panels. In: L. Mátyás and P. Sevestre (Eds.), The Econometrics of Panel Data: A Handbook of the Theory with Applications, second revised edition, pp.145–195, Kluwer Academic Publishers, Dordrecht.

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

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

[24] Phillips, P. C. B. and D. Sul, 2003. Dynamic panel estimation and homogeneity testing under cross section dependence. Econometrics Journal 6, 217–259.

[25] Pollard, D., 1984. Convergence of Stochastic Processes. Springer-Verlag, New York.

[26] Su, L. and S. Jin, 2010. Sieve estimation of panel data models with cross section dependence. Forthcoming in Journal of Econometrics.

[27] Sun, Y., 2005. Estimation and inference in panel structure models. Working paper, Department of Economics, UCSD.

[28] Zhang, Y., L. Su, and P. C. B. Phillips, 2011. Testing for common trends in semiparametric panel data models with fixed effects. Working paper, Singapore Management University.