1-2013

Specification Test for Panel Data Models with Interactive Fixed Effects

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

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

Y. Zhang

Follow this and additional works at: [http://ink.library.smu.edu.sg/soe_research](http://ink.library.smu.edu.sg/soe_research)

Part of the [Growth and Development Commons](http://ink.library.smu.edu.sg/soe_research)

Citation

**Available at:** [http://ink.library.smu.edu.sg/soe_research/1486](http://ink.library.smu.edu.sg/soe_research/1486)

This Working Paper is brought to you for free and open access by the School of Economics at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection School Of Economics by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email libIR@smu.edu.sg.
Specification Test for Panel Data Models with Interactive Fixed Effects∗

Liangjun Su, Sainan Jin, Yonghui Zhang

School of Economics, Singapore Management University,
90 Stamford Road, Singapore 178903

September 14, 2012

Abstract

In this paper, we propose a consistent nonparametric test for linearity in a large dimensional panel data model with interactive fixed effects. Both lagged dependent variables and conditional heteroskedasticity of unknown form are allowed in the model. We estimate the model under the null hypothesis of linearity to obtain the restricted residuals which are then used to construct the test statistic. We show that after being appropriately centered and standardized, the test statistic is asymptotically normally distributed under both the null hypothesis and a sequence of Pitman local alternatives by using the concept of conditional strong mixing that was recently introduced by Prakasa Rao (2009). To improve the finite sample performance, we propose a bootstrap procedure to obtain the bootstrap p-value. A small set of Monte Carlo simulations illustrates that our test performs well in finite samples. An application to an economic growth panel dataset indicates significant nonlinear relationships between economic growth, initial income level and capital accumulation.

Key Words: Common factors; Conditional strong mixing; Cross-sectional dependence; Economic Growth; Interactive fixed effects; Linearity; Panel data models; Specification test.

JEL Classifications: C12, C14, C23

∗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. The second author gratefully acknowledges financial support from SMU with grant number C244/MSS11E005.
1 Introduction

Recently there has been a growing literature on large dimensional panel data models with interactive fixed effects. The term “interactive fixed effects” refers to the scenario where individual and time effects enter the model multiplicatively. These models can capture heterogeneity more flexibly than the traditional fixed/random effects models by the adoption of time-varying common factors that affect the cross sectional units with individual specific factor loadings. It is this flexibility that drives the models to become one of the most popular and successful tools to handle cross sectional dependence, especially when both the cross sectional dimension ($N$) and the time period ($T$) are large. For example, Pesaran (2006) considers estimation and inference in heterogeneous panel data models with a general multifactor error structure; Bai (2009) proposes a principal component analysis (PCA) estimator and establishes its asymptotic normality for homogeneous panel data models with interactive fixed effects; Zafaroni (2010) applies generalized least squares (GLS) method to estimate panel data models with common shocks; Moon and Weidner (2010; 2012) reinvestigate Bai’s (2009) PCA estimation and put it in the framework of Gaussian quasi maximum likelihood estimation (QMLE) of panel data models with interactive fixed effects; for other developments on this models, see Ahn, Lee, and Smith (2001, 2007) for GMM approach with fixed $T$ and large $N$, Kapetanos and Pesaran (2007) and Greenaway-McGrevy, Han, and Sul (2012) for factor-augmented panel regressions, Pesaran and Tosetti (2011) for estimation of panel data models with a multifactor error structure and spatial error correlation, and Su and Jin (2012) and Jin and Su (2012) for nonparametric estimation and test in panel data models with interactive fixed effects.

Panel data models with interactive fixed effects have been widely used in economics. Examples from labor economics include Carneiro, Hansen, and Heckman (2003) and Cunha, Heckman, and Navarro (2005), both of which employ a factor error structure to study individuals’ education decision. In macroeconomics, Giannone and Lenza (2005) provide an explanation for Feldstein-Horioka’s (1980) puzzle by using interactive fixed effects models. In finance, the arbitrage pricing theory of Ross (1976) is built on a factor model for assets returns. Bai and Ng (2006) develop several tests to evaluate the latent and observed factors in macroeconomics and finance. Ludvigson and Ng (2009) investigate the empirical risk-return relation by using dynamic factor analysis for large datasets to summarize a large amount of economic information by few estimated factors. Ludvigson and Ng (2011) use factor augmented regressions to analyze the relationship between bond excess returns and macroeconomic factors.

Almost all of the above papers focus on the linear specification of regression relationship in panel data models with interactive fixed effects. The only exceptions are Su and Jin (2012) and Jin and Su (2012). The former paper extends the linear model of Pesaran (2006) to a nonparametric model with interactive fixed effects and propose a sieve estimator for the unknown regression function of interest; the latter constructs a nonparametric test for poolability in nonparametric regression models with a multi-factor error structure. Despite the robustness of nonparametric estimates and tests, they are usually subject to slower convergence rates than their parametric counterparts. On the other hand, estimation and tests based on parametric (usually linear) models can be misleading if the underlying models are misspecified. For this reason, it is worthwhile to propose a test for the correct specification for the widely used linear
panel data models with interactive effects.

In this paper we are interested in testing for linearity in the following panel data model

\[ Y_{it} = m(X_{it}) + F_0^\prime \lambda_0^\prime + \varepsilon_{it}, \]  

(1.1)

where \( i = 1, \ldots, N, \ t = 1, \ldots, T, \ X_{it} \) is a \( p \times 1 \) vector of observed regressors, \( m(\cdot) \) is an unknown smooth function, \( F_0^\prime \) is an \( R \times 1 \) vector of unobserved common factors, \( \lambda_0^\prime \) is an \( R \times 1 \) vector of unobserved factor loadings, \( \varepsilon_{it} \) is an idiosyncratic error term. When \( m(X_{it}) \equiv X_{it} \beta_0^\prime \) almost surely (a.s.) for some \( \beta_0^\prime \in \mathbb{R}^p \), (1.1) becomes the most popular linear panel data model with interactive fixed effects, which is investigated by Pesaran (2006), Bai (2009) and Moon and Weidner (2010; 2012), among others. These authors consider various estimates for \( \beta \) and/or \( (\lambda_i, F_i) \) in the model. Asymptotic distributions for all estimators have been established and bias-correction is generally needed.

Although economic theory dictates that some economic variables are important for the causal effects of the others, rarely does it state exactly how the variables should enter a statistical model. Models derived from first-principles such as utility or production functions only have linear dynamics under some narrow functional form restrictions. Linear models are usually adopted for convenience. A correctly specified linear model may afford precise inference whereas a badly misspecified one may offer seriously misleading inference. When \( m(\cdot) \) is a nonlinear function, the previously reviewed methods generally cannot provide consistent estimates for the underlying regression function, and the estimated factor space would be inconsistent too. As a result, tests based on these estimates would be completely misleading.

For example, testing/determining the number of common factors is a very important problem in factor analysis (see, e.g., Bai and Ng (2002), Onatski (2009), and Lu and Su (2012)); testing for additivity versus interactivity in panel data models (see, e.g., Bai (2009)) is another one. But both are generally invalid if they are based on the estimation of a misspecified model. Therefore, to avoid the serious consequence of misspecification, it is necessary and prudent to test for linearity before we embark on statistical inference about the coefficients and factor space.

There have been many tests for linearity or more generally the correct specification of parametric models in the literature. The RESET test of Ramsey (1969) is the common used specification test for the linear regression model. But it is well known that this test is not consistent because it fails to detect some unknown nonlinear alternatives. Since Hausman (1978) a large literature on testing for the correct specification of functional forms has developed; see Bierens (1982, 1990), Wooldridge (1992), Yatchew (1992), Härdle and Mammen (1993), Hong and White (1995), Fan and Li (1996), Zheng (1996), Bierens and Ploberger (1997), Li and Wang (1998), Stinchcombe and White (1998), Chen and Gao (2007), Hsiao, Li, and Racine (2007), and Su and Ullah (2012), among others. In addition, Hjellvik and Tjøstheim (1995) and Hjellvik, Yao, and Tjøstheim (1998) derive tests for linearity specification in nonparametric regressions and Hansen (1999) reviews the problem of testing for linearity in the context of self-exciting threshold autoregressive (SETAR) models. More recently, Lee (2011) and Su and Lu (2012) consider testing for linearity in dynamic panel data models based on individual-specific generalized spectral derivative and the \( L^2 \)-distance between parametric and nonparametric estimates, respectively. Nevertheless, to the best of our knowledge, there is no available test of linearity for large dimensional panel data models with interactive fixed effects.
In this paper, we propose a nonparametric test for linearity in panel data models with interactive fixed effects. We first estimate the model under the null hypothesis of linearity and obtain the parametric residuals that are used to construct our test statistic. The parametric residual contains no useful information about the regression function when the linear model is correctly specified; it does otherwise. As a result, the projection of the parametric residual to the regressor space is expected to be zero under the null and nonzero under the alternative. This motives our residual-based test, like many other residual-based tests in the literature (e.g., Fan and Li (1996), Zheng (1996), Hsiao, Li and Racine (2007), and Su and Ullah (2012)). We show that after being appropriately centered and standardized, our test statistic is asymptotically normally distributed under the null hypothesis and a sequence of Pitman local alternatives. We also propose a bootstrap procedure to obtain the bootstrap $p$-value. Clearly, in the case of rejecting the null hypothesis, the conventional linear panel data models with interactive fixed effects cannot be used, and one has to consider nonlinear or nonparametric modelling. We apply our test to an economic growth panel dataset from the Penn World Table (PWT 7.1) and find significant nonlinear relationships across different model specifications and periods. This suggests the empirical relevance of our test and calls upon nonparametric or nonlinear modeling of panel data models with interactive fixed effects.

In comparison with the existing tests for other models in the literature, the major difficulties in analyzing our test lie in three aspects. The first one is due to the slow convergence rate of the estimates of factors and factor loadings. In the papers mentioned above, the parametric residuals converge to the true random error terms under the null at the usual parametric rate and thus the parametric estimation error does not play a role in the asymptotic distribution of the test statistic under either the null or nontrivial Pitman local alternatives. In contrast, for panel data models with interactive fixed effects, the factors and factor loadings can only be estimated at a slower rate than the slope coefficients of the observed regressors, and their estimation error plays an important role and complicates the asymptotic analysis of the local power function significantly. The second major difficulty is due to the allowance for dynamic structure in the panel data models. The test statistic (see (2.4) below) itself possesses the structure of a two-fold $V$-statistic where double summations are needed along both the individual and time dimensions. The asymptotic analysis of such a statistic becomes extremely involved with the presence of lagged dependent variables when the first-stage parameter estimation errors enter the asymptotics. The third major difficulty arises because the observations are typically not independent across the cross sectional units or strong mixing over time in dynamic panel data models with interactive fixed effects. This occurs in dynamic panel data models when we allow both the unobserved factors and factor loadings to be stochastic. Nevertheless, conditional on the unobserved factors and factor loadings, we may have independence across cross sectional units and strong mixing over time. Fortunately, the classical central limit theorem (CLT) for second-order degenerate $U$-statistics for independent but nonidentically distributed (INID) observations (see, e.g., de Jong (1987)) can be extended straightforwardly to the case of conditionally independent but nonidentically distributed (CINID) observations. The classical Davydov’s and Bernstein’s inequalities for strong mixing processes also have their analog for conditional strong mixing processes which were formerly introduced by Prakasa Rao (2009). The study of the asymptotic properties of our test statistic relies on these innovations.
The rest of the paper is organized as follows. In Section 2, we introduce the hypothesis and the test statistics. The asymptotic distributions of our test are established both under the null hypothesis and the local alternatives in Section 3. In Section 4 we conduct a small set of Monte Carlo experiments to evaluate the finite sample performance of our test and apply our test to an economic growth data set. Section 5 concludes. All proofs are relegated to the Appendices and additional proofs for the technical lemmas are provided in the supplement.

NOTATION. Throughout the paper we adopt the following notation. For an \( m \times n \) real matrix \( A \), we denote its transpose as \( A' \), its Frobenius norm as \( \| A \|_F \) \((\equiv \| \text{tr} (AA') \|^{1/2})\), its spectral norm as \( \| A \| \) \((\equiv \sqrt{\mu_1 (AA')}\)\), where \( \equiv \) means “is defined as” and \( \mu_1 (\cdot) \) denotes the largest eigenvalue of a real symmetric matrix. Let \( \mu_{\min} (\cdot) \) denote the minimum eigenvalue of a real symmetric matrix. More generally, we use \( \mu_s (\cdot) \) to denote the \( s \)th largest eigenvalue of a real symmetric matrix by counting multiple eigenvalues multiple times. Let \( P_A \equiv A (A'A)^{-1} A' \) and \( M_A \equiv I_m - P_A \) where \( I_m \) denotes an \( m \times m \) identity matrix. We use “p.d.” and “p.s.d.” to abbreviate “positive definite” and “positive semidefinite”, respectively. For symmetric matrices \( A \) and \( B \), we use \( A \succeq B \) \((A \preceq B)\) to indicate that \( A-B \) is p.d. \((\text{p.s.d.})\). The operator \( \mathcal{D} \) denotes convergence in probability, \( \mathcal{D} \) convergence in distribution, and plim probability limit. We use \((N,T) \to \infty\) to denote the joint convergence of \( N \) and \( T \) when both pass to the infinity simultaneously.

2 Basic Framework

In this section, we first formulate the hypotheses and test statistic, and then introduce the estimation of panel data models with interactive fixed effects under the null restriction.

2.1 The hypotheses and test statistic

The main objective is to construct a test for linearity in model (1.1). We are interested in testing the null hypothesis

\[
H_0 : \text{Pr} \left[ m (X_{it}) = X_{it}' \beta^0 \right] = 1 \text{ for some } \beta^0 \in \mathbb{R}^p. \tag{2.1}
\]

The alternative hypothesis is the negation of \( H_0 \):

\[
H_1 : \text{Pr} \left[ m (X_{it}) = X_{it}' \beta \right] < 1 \text{ for all } \beta \in \mathbb{R}^p. \tag{2.2}
\]

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

\[
H_1 (\gamma_{NT}) : m (X_{it}) = X_{it}' \beta^0 + \gamma_{NT} \Delta (X_{it}) \text{ a.s. for some } \beta^0 \in \mathbb{R}^p \tag{2.3}
\]

where \( \Delta (\cdot) \equiv \Delta_{NT} (\cdot) \) is a measurable nonlinear function, \( \gamma_{NT} \to 0 \) as \((N,T) \to \infty\), and the rate is specified in Theorem 3.3 below.

Let \( e_{it} \equiv Y_{it} - X_{it}' \beta^0 - F_t' \lambda_t^0 \). Let \( f_i (\cdot) \) denote the probability density function (PDF) of \( X_{it} \). In view of the fact that \( e_{it} = e_{it} \) and \( E (e_{it} | X_{it}) = 0 \) under \( H_0 \), we have

\[
J \equiv E [e_{it} E (e_{it} | X_{it}) f_i (X_{it})] = E \left\{ [E (e_{it} | X_{it})]^2 f_i (X_{it}) \right\} = 0
\]
under $\mathbb{H}_0$. Nevertheless, under $\mathbb{H}_1$ we have $e_{it} = \varepsilon_{it} + m (X_{it}) - X_{it}' \beta_0$. So $E (e_{it} | X_{it}) = m (X_{it}) - X_{it}' \beta_0$ is not equal 0 a.s., implying that $E [e_{it} E (e_{it} | X_{it}) f_i (X_{it})] > 0$ under $\mathbb{H}_1$. Below we propose a consistent test for the correct specification of the linear panel data model based on this observation.

To implement our test, we need to estimate the model under the null hypothesis (2.1). Under $\mathbb{H}_0$ under parameter, and $\mathcal{K}_{ij}$ is not equal 0 a.s., implying that $H_{0}$ under 2.2 Estimation under the null

To proceed, let $X_{it,k}$ denote the $k$th element of $X_{it}$ for $k = 1, ..., p$. Define

$$
Y_i \equiv (Y_{i1}, \ldots, Y_{iT})', \quad X_i \equiv (X_{i1}, \ldots, X_{iT})', \quad \varepsilon_i \equiv (\varepsilon_{i1}, \ldots, \varepsilon_{iT})', \quad e_i \equiv (e_{i1}, \ldots, e_{iT})', \\
F^0 \equiv (F_{10}, \ldots, F_{p0})', \quad \lambda^0 \equiv (\lambda_{10}, \ldots, \lambda_{0N})', \quad X_{i,k} \equiv (X_{i1,k}, \ldots, X_{iT,k})', \quad Y \equiv (Y_1, \ldots, Y_N)', \\
X_k \equiv (X_{1,k}, \ldots, X_{N,k})', \quad \varepsilon \equiv (\varepsilon_1, \ldots, \varepsilon_N)', \quad \text{and } e \equiv (e_1, \ldots, e_N)'.
$$

Clearly, $Y$, $X_k$, $\varepsilon$ and $e$ all denote $N \times T$ matrices.

As mentioned above, we need to estimate the model under the null hypothesis (2.1). Under $\mathbb{H}_0$, we can rewrite the model in vector and matrix notation as

$$
Y_i = X_i \beta^0 + F^0 \lambda^0 + \varepsilon_i \quad (2.5)
$$

and

$$
Y = \sum_{k=1}^{p} \beta_k X_k + \lambda F' + \varepsilon, \quad (2.6)
$$

where $\beta^0 \equiv (\beta_{10}, \ldots, \beta_{p0})'$. Following Moon and Weidner (2010, 2012), the Gaussian quasi-maximum likelihood estimator (QMLE) $(\hat{\beta}, \hat{\lambda}, \hat{F})$ of $(\beta, \lambda, F)$ can be obtained as follows

$$
\left( \hat{\beta}, \hat{\lambda}, \hat{F} \right) = \arg\min_{(\beta, \lambda, F)} \mathcal{L}_{NT} (\beta, \lambda, F) \quad (2.7)
$$

where

$$
\mathcal{L}_{NT} (\beta, \lambda, F) \equiv \frac{1}{NT} \text{tr} \left[ \left( Y - \sum_{k=1}^{p} \beta_k X_k - \lambda F' \right)' \left( Y - \sum_{k=1}^{p} \beta_k X_k - \lambda F' \right) \right], \quad (2.8)
$$

$\beta \equiv (\beta_1, \ldots, \beta_p)'$ is a $p \times 1$ vector of parameter coefficients, $F \equiv (F_1, \ldots, F_T)'$ and $\lambda \equiv (\lambda_1, \ldots, \lambda_N)'$. In particular, the main object of interest $\beta$ can be estimated by

$$
\hat{\beta} = \arg\min_{\beta} L_{NT} (\beta) \quad (2.9)
$$
where the negative profile quasi log-likelihood function $L_{NT} (\beta)$ is given by

$$L_{NT} (\beta) = \min_{\lambda, F} \mathcal{L}_{NT} (\beta, \lambda, F)$$

$$= \min_{F} \frac{1}{NT} \text{tr} \left[ \left( Y - \sum_{k=1}^{p} \beta_k X_k \right) M_F \left( Y - \sum_{k=1}^{p} \beta_k X_k \right)' \right]$$

$$= \frac{1}{NT} \sum_{t=R+1}^{T} \mu_t \left[ \left( Y - \sum_{k=1}^{p} \beta_k X_k \right)' \left( Y - \sum_{k=1}^{p} \beta_k X_k \right) \right]. \quad (2.10)$$

See Moon and Weidner (2010) for the demonstration of the equivalence of the last three expressions.

As (2.9) and (2.10) suggest, it is convenient to compute the QMLE: one only needs to calculate the eigenvalues of a $\lambda \times \lambda$ matrix at each step of the numerical optimization over $\beta$. For statistical inference, one also needs to obtain consistent estimates of $\lambda^0$ and $F^0$ under certain identification restrictions.

Following Bai (2009), we consider the following identification restrictions

$$\lambda^0 \equiv \lambda = \text{diagonal matrix}. \quad (2.11)$$

Upon obtaining $\hat{\beta}$, the QMLE $(\hat{\lambda}, \hat{F})$ of $(\lambda, F)$ are given by the solutions of the following set of nonlinear restrictions:

$$\frac{1}{NT} \sum_{i=1}^{N} (Y_i - X_i \hat{\beta})(Y_i - X_i \hat{\beta})' \hat{F} = \hat{F}V_{NT}, \quad (2.12)$$

and

$$\hat{\lambda} = (\hat{\lambda}_1, ..., \hat{\lambda}_N) = T^{-1} \left[ \hat{F}'(Y_1 - X_1 \hat{\beta}), ..., \hat{F}'(Y_N - X_N \hat{\beta}) \right], \quad (2.13)$$

where $V_{NT}$ is a diagonal matrix that consists of the $R$ largest eigenvalues of the bracketed matrix in (2.12), arranged in decreasing order.

After obtaining $(\hat{\beta}, \hat{\lambda}, \hat{F})$, we can estimate $\varepsilon_i$ by $\hat{\varepsilon}_i \equiv Y_i - X_i \hat{\beta} - \hat{F} \hat{\lambda}_i$ under the null. It is easy to verify that

$$\hat{\varepsilon}_i = M_F \varepsilon_i - M_F X_i (\hat{\beta} - \beta^0) + M_F F^0 \lambda^0 + M_F (m_i - X_i \beta^0) \quad (2.14)$$

where $m_i \equiv (m(X_{i1}), m(X_{i2}), \cdots, m(X_{iT}))'$. $\hat{\varepsilon}_i$ is then used in constructing the test statistic $J_{NT}$ defined in (2.4).

## 3 Asymptotic Distribution

In this section we first study the asymptotic behavior of $\hat{\beta}$ under $\mathbb{H}_1 (\gamma_{NT})$ and then the asymptotic distribution of our test statistic under $\mathbb{H}_1 (\gamma_{NT})$. We also propose a bootstrap method to obtain the bootstrap $p$-values for our test.

### 3.1 Asymptotic behavior of $\hat{\beta}$ under $\mathbb{H}_1 (\gamma_{NT})$

To study the asymptotic behavior of $\hat{\beta}$ under $\mathbb{H}_1 (\gamma_{NT})$, we make the following assumptions.

**Assumption A.1.** (i) $N^{-1} \lambda^0 \lambda^0 \overset{d}{\rightarrow} \Sigma_{\lambda} > 0$ for some $R \times R$ matrix $\Sigma_{\lambda}$ as $N \rightarrow \infty$. 

(ii) \( T^{-1}F_0'F_0 \overset{p}{\to} \Sigma_F > 0 \) for some \( R \times R \) matrix \( \Sigma_F \) as \( T \to \infty \).

(iii) \( \|\varepsilon\| = O_P\left(\max\left(\sqrt{N}, \sqrt{T}\right)\right) \).

(iv) \( \|X_k\| = O_P\left(\sqrt{NT}\right) \) for \( k = 1, \ldots, p \).

**Assumption A.2** (i) \((NT)^{-1/2}tr(X_k\varepsilon') = O_P(1) \) for \( k = 1, \ldots, p \).

(ii) Let \( X(\alpha) = \sum_{k=1}^{p} \alpha_k X_k \) such that \( \|\alpha\| = 1 \) where \( \alpha = (\alpha_1, \ldots, \alpha_p)' \). There exists a finite constant \( C > 0 \) such that

\[
\min_{\{\alpha \in \mathbb{R}^p : \|\alpha\| = 1\}} \sum_{t=2R+1}^{T} \mu_t \left( X(\alpha)'X(\alpha) \right) \geq C \text{ with probability approaching 1 (w.p.a.1).}
\]

**Assumption A.3** (i) As \((N, T) \to \infty, \gamma_{NT} \to 0\) and \( \delta_{NT}^{-2} / \gamma_{NT} = O(1) \), where \( \delta_{NT} = \min\left(\sqrt{N}, \sqrt{T}\right) \).

(ii) Let \( \Delta_i = (\Delta(X_{1i}), \ldots, \Delta(X_{iT}))' \) and \( \Delta = (\Delta_1, \ldots, \Delta_T)' \). \( \|\Delta\| = O_P\left(\sqrt{NT}\right) \).

Assumptions A.1-A.2 are also made in Moon and Weidner (2010). A.1(i), (ii) and (iv) can be easily satisfied and A.1(iii) can be met for various error processes. Assumption A.2(i) requires weak exogeneity of the regressors \( X_k \) and A.2(ii) imposes the usual non-collinearity condition on \( X_k \). Note that A.2(ii) rules out time-invariant regressors or cross-sectionally invariant regressors, but it can be modified as in Moon Weidner (2012) to allow for both with more complicated notation and special treatment. Assumption A.3 specifies conditions on \( \gamma_{NT} \) and \( \Delta \) relative to the sample sizes \( N \) and \( T \). A.3(i) will be automatically satisfied for the local alternative studied below.

Let \( C^{(1)}_{NT} \) and \( C^{(2)}_{NT} \) denote \( p \times 1 \) vectors whose \( k \)th elements are respectively given by

\[
C^{(1)}_{NT,k} = \frac{1}{NT\gamma_{NT}} \text{tr} \left( M^0_k X_k M_F e' \right), \quad \text{and}
\]

\[
C^{(2)}_{NT,k} = \frac{1}{NT\gamma_{NT}} \text{tr} \left( e'M_F e'M^0_k X_k \Phi_1' + e'M^0_k eM_F e X_k \Phi_1 + e'M^0_k X_k e e' \Phi_1 \right),
\]

where \( \Phi_1 = \lambda^0 (\lambda^0 \lambda^0)^{-1} (F_0 F_0)^{-1} F_0 \). Let \( D_{NT} \) denote a \( p \times p \) matrix whose \((k_1, k_2)\)th element is given by

\[
D_{NT,k_1k_2} = \frac{1}{NT} \text{tr} \left( M^0_k X_{k_1} M_F e X_{k_2}' \right).
\]

Following Moon and Weidner (2012) we refer to \( C^{(1)}_{NT} + C^{(2)}_{NT} \) and \( D_{NT} \) as the approximated score and Hessian matrix for the profile quasi-likelihood function. The following theorem states the asymptotic expansion of \( \hat{\beta} \) under \( \mathbb{H}_1(\gamma_{NT}) \).

**Theorem 3.1** Suppose Assumptions A.1-A.3 hold. Then under \( \mathbb{H}_1(\gamma_{NT}) \)

\[
\hat{\beta} - \beta^0 = \gamma_{NT} D_{NT}^{-1} \left( C^{(1)}_{NT} + C^{(2)}_{NT} \right) + O_P \left\{ \left[ \gamma_{NT}^2 (\delta_{NT}^{-1} + \gamma_{NT}) + \gamma_{NT} \delta_{NT}^{-3} \right]^{1/2} \right\}.
\]

**Remark 1.** The above result is comparable with that in Corollary 3.2 of Moon and Weidner (2010). In view of the fact that \( D_{NT} = D + o_P(1) \), \( C^{(1)}_{NT} = O_P(1) \) and \( C^{(2)}_{NT} = O_P \left( \delta_{NT}^{-2}/\gamma_{NT} \right) = O_P(1) \) where \( D = \lim_{(N,T) \to \infty} D_{NT} \) (see the remark in Appendix A), we have \( \hat{\beta} - \beta^0 = O_P(\gamma_{NT}) \). That is, the local
deviations from the null model control the convergence rate of \( \hat{\beta} \) to \( \beta^0 \). Note that we do not require \( N \) and \( T \) diverge to \( \infty \) at the same speed, nor do we require that one diverge to \( \infty \) faster than the other.

Let \( \alpha_{ik} = \xi_i^0 (\lambda_i^0 / N)^{-1} \lambda_k^0 \) and \( \hat{X}_i = M_{F_0} X_i - \frac{1}{N} \sum_{j=1}^N \alpha_{ij} M_{F_0} X_j \). It is easy to see that an alternative expression for \( D_{NT} \) is given by

\[
D_{NT} = D_{NT} (F^0) = \frac{1}{NT} \sum_{i=1}^N X_i' M_{F_0} X_i - \frac{1}{T} \left( \frac{1}{N^2} \sum_{i=1}^N \sum_{k=1}^N X_i' M_{F_0} X_k \alpha_{ik} \right) = \frac{1}{NT} \sum_{i=1}^N \hat{X}_i' \hat{X}_i,
\]

which is used by Bai (2009). But he also requires that \( D_{NT} (F) \) be asymptotically non-singular for all \( F \) such that \( F'F/T = I_R \).

### 3.2 Asymptotic distribution of the test statistic

First, we introduce the concept of conditional strong mixing.

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

\[
|P_B (A \cap B) - P_B (A) P_B (B)| \leq \alpha^B (t) \quad \text{a.s.}
\]

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

The above definition is due to Prakasa Rao (2009); see also Roussas (2008). When one takes \( \alpha^B (t) \) as the supremum of the left hand side object in (3.4) over the set \( \{ A \in \sigma (\xi_1, \ldots, \xi_k), B \in \sigma (\xi_{k+t}, \xi_{k+t+1}, \ldots) \}, k \geq 1 \), we refer it to the \( \mathcal{B} \)-strong-mixing coefficient.

Let \( q_0 \) and \( q_1 \) be as specified in Assumption A.5 below. Let \( q_2 \in (1, 4/3) \) and \( \tilde{q}_2 \equiv q_1 q_2 / (q_1 + q_2) \). Let \( \tilde{q}_3 > 0 \) be such that \( 1 - \frac{1}{\tilde{q}_3} = \frac{1}{q_2} + \frac{1}{q_3} \). Let \( h! \equiv \Pi_{k=1}^h k \). Let \( ||A||_q \equiv \{ E \|A\|^q \}^{1/q} \) for any random scalar or vector \( A \). Let \( \mathcal{D} \equiv \sigma (F^0, X^0) \), the \( \sigma \)-field generated by \( (F^0, X^0) \). To study the asymptotic distribution of the test statistic, we add the following assumptions.

**Assumption A.4** (i) For each \( i = 1, \ldots, N \), \( \{(X_{it}, \xi_{it}) : t = 1, 2, \ldots \} \) is conditionally strong mixing given \( \mathcal{D} \) (or \( \mathcal{D} \)-strong-mixing) with mixing coefficients \( \{\alpha_{NT,i}^D (t), 1 \leq t \leq T - 1 \} \). \( \alpha_D (\cdot) \equiv \alpha_{NT}^D (\cdot) \equiv \max_{1 \leq i \leq N} \alpha_{NT,i}^D (\cdot) \) satisfies \( \sum_{s=1}^\infty \alpha_D (s)^{1/q_3} \leq C_\alpha < \infty \) a.s. and \( \sum_{r=1}^\infty \alpha_D (r)^{\tilde{q}_3 / (1 + \tilde{q}_3)} \leq C_\alpha < \infty \) a.s. for some \( \tilde{q}_3 \in (0, 1/3) \). In addition, there exists \( \tau \in (1, Th!) \) such that \( Th! / \tau \gg T^\eta \) for some \( \eta > 0 \) and \( (NT)^{1 + p/q_0} (h!)^{-1} \alpha_D (\tau) = o_{a.s.} (1) \) as \( (N, T) \to \infty \).

(ii) \( \{(\xi_{it}, X_i), i = 1, \ldots, N, \} \) are mutually independent of each other conditional on \( \mathcal{D} \).

(iii) For each \( i = 1, \ldots, N \), \( E (\xi_{it} | \mathcal{F}_{NT,i-1}) = 0 \) a.s. where \( \mathcal{F}_{NT,i-1} \equiv \sigma (\{F^0, X^0, X_{it}, X_{i,t-1}, \xi_{i,t-1}, X_{i,t-2}, \xi_{i,t-2}, \ldots \}_{i=1}^N) \).

(iv) For each \( i = 1, \ldots, N \), let \( f_{i,t} (x) \) denote the marginal PDF of \( X_{it} \) given \( \mathcal{D} \), and \( f_{i,t,i} (x, \bar{x}) \) the joint PDF of \( X_{it} \) and \( X_{is} \) given \( \mathcal{D} \). \( f_{i,t} (\cdot) \) and \( f_{i,t,i} (\cdot, \cdot) \) are continuous in their arguments and uniformly bounded by \( C_f < \infty \).

**Assumption A.5** (i) \( \max_{1 \leq i \leq N} \|X_{it}\|_{q_0} \leq C_X < \infty \) for some \( q_0 \geq 4 \).
(ii) max_{1 \leq i \leq N} \|\varepsilon_{it}\|_{q_1} \leq C_\varepsilon < \infty \text{ for some } q_1 > 4.

(iii) max_{1 \leq i \leq N} \|\lambda_t^0\|_{q_1} \leq C_\lambda < \infty \text{ and max}_{1 \leq t \leq T} \|F_t^\varepsilon\|_{q_1} \leq C_F < \infty.

(iv) Either \(\Delta (\cdot)\) is uniformly bounded or there exists a function \(D_\Delta (\cdot)\) such that \(|\Delta (x + \hat{x}) - \Delta (x)| \leq D_\Delta (x) \|\hat{x}\|\) for all \(x \in \mathcal{X}\) and \(\hat{x} = o(1)\), and \((NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} E[\|F_t^\varepsilon\| + \|\lambda_t^0\|] D_\Delta (X_{it})^4 + |\Delta (X_{it})|^4] = O(1).

(v) \((NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} E[\|F_t^0\| + \|\lambda_t^0\| \varepsilon_{it}^2] = O(1)\) and \((NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} E[\|F_t^0\|^2 + \|\lambda_t^0\|^2 \varepsilon_{it}^2] = O(1)\).

**Assumption A.6** (i) The kernel function \(k(\cdot) : \mathbb{R} \rightarrow \mathbb{R}\) is a symmetric, continuous and bounded PDF. Let \(\bar{K} \equiv \sup_u [k(u)]^\rho\).

(ii) For some \(C_k < \infty\) and \(L < \infty\), either \(k(u) = 0\) for \(|u| > L\) and for all \(u\) and \(\bar{u} \in \mathbb{R}\), \(|k(u) - k(\bar{u})| \leq C_k |u - \bar{u}|\), or \(k(u)\) is differentiable, \(\sup_u |(\partial/\partial u)k(u)| \leq C_k\), \(k(u) \leq C_k |u|^{-\rho_0}\) and \(|(\partial/\partial u)k(u)| \leq C_k |u|^{-\nu}\) for \(|u| > L\) and for some \(\nu > 1\).

**Assumption A.7** (i) As \((N, T) \rightarrow \infty\), \(|h| \rightarrow 0\), \(NT\delta_N^{-4} (h!)^{1/2} \rightarrow 0\), \(Th! \rightarrow \infty\), and \(Nh! \rightarrow \infty\).

(ii) As \((N, T) \rightarrow \infty\), \(NT^{-1} (h!)^{(2(1-\rho)/q_1)} + (h!)^{-2h!(1+\rho)}\] \(\rightarrow 0\) and \(N^2T^{-2} (h!)^{(1-3\rho)/q_2} \rightarrow 0\).

A.4(i) requires that each individual time series \(\{(X_{it}, \varepsilon_{it}) : t = 1, 2, ...\}\) be \(\mathcal{D}\)-strong-mixing with an algebraic mixing rate. Prakasa Rao (2009) extends the concept of (unconditional) strong mixing to conditional strong mixing. It turns out that several well-known inequalities for strong mixing processes also have their conditional versions. See Lemmas E.1-E.3 in the supplementary appendix. As Su and Chen (2012) notice, even if \(\{(\varepsilon_{it}, F_t^\varepsilon), t \geq 1\}\) is strong mixing, the simple AR(1) panel data model with interactive fixed effects, \(Y_{it} = \rho_0 Y_{i,t-1} + \lambda_t^0 F_t^\varepsilon + \varepsilon_{it}\), does not yield a strong mixing process \(\{Y_{it}, t \geq 1\}\) unless one assumes that \(\lambda_t^0\) is nonstochastic. For this reason, Hahn and Kuersteiner (2011) assume that the individual fixed effects are nonrandom and uniformly bounded in their study of nonlinear dynamic panel data models. Here we take a different approach and adopt the notion of \(\mathcal{D}\)-strong-mixing, which in conjunction with A.4(ii)-(iii) facilitates our asymptotic analysis. The dependence of the mixing rate on \(\tilde{q}_3\) and \(\tilde{q}\) in A.4(i) reflects the trade-off between the degree of dependence and the moment bounds of the process \(\{(X_{it}, \varepsilon_{it}), t \geq 1\}\). If the process is \(\mathcal{D}\)-strong-mixing with a geometric mixing rate, the conditions on \(\sigma_T (\cdot)\) can easily be met by specifying \(\tau = [C_r \log T]\) for some sufficiently large \(C_r\), where \([a]\) denotes the integer part of \(a\).

A.4(ii) requires that \((\varepsilon_{it}, X_{it})\) be conditionally independent across \(i\) but does not rule out cross sectional dependence among them. When \(X_{it} = Y_{i,t-1}\) and \(\varepsilon_{it}\) exhibits conditional heteroskedasticity (e.g., \(\varepsilon_{it} = \sigma_0 (Y_{i,t-1}) \varepsilon_{it}\) where \(\varepsilon_{it} \sim \text{IID}(0, 1)\) and \(\sigma_0 (\cdot)\) is an unknown smooth function in the above panel AR(1) model), \((X_{it}, \varepsilon_{it})\) are not independent across \(i\) because of the presence of common factors irrespective of whether one allows \(\lambda_t^0\) to be independent across \(i\) or not. Nevertheless, conditional on \(\mathcal{D}\), it is possible that \((X_{it}, \varepsilon_{it})\) is independent across \(i\) such that A.4(ii) is still satisfied. Here the cross sectional dependence is similar to the type of cross sectional dependence generated by common shocks studied by Andrews (2005).

The difference is that Andrews (2005) assumes IID observations conditional on the \(\sigma\)-field generated by the common shocks in a cross-section framework, whereas we have conditionally independent but non-identically distributed (CINID) observations across the individual dimension in a panel framework.

A.4(iii) requires that the error terms \(\varepsilon_{it}\) be a martingale difference sequence (m.d.s.) with respect
to the filter $F_{NT,t-1}$, which allows lagged dependent variables in $X_{it}$ and conditional heteroskedasticity, skewness or kurtosis in $\varepsilon_{it}$. Of course, if one assumes that $X_{it}$ is strictly exogenous, then the proofs for the following theorems can be greatly simplified. In sharp contrast, early literature on panel data models with interactive fixed effects typically assumes that $\varepsilon_{it}$ is independent of $\lambda^0_i$ and $F^0_i$ for all $i,j,t,s$; see, e.g., Pesaran (2006), Bai (2009), Moon and Weidner (2010, 2012), and Bai and Li (2012). In particular, Moon and Weidner (2010, 2012) and Bai and Li (2012) assume that both the factors and factor loadings are fixed constants and treat them as parameters to be estimated. A.4(iii) imposes conditions on the conditional densities $f_{i,t}$ and $f_{i,ts}$. The uniform boundedness condition can be relaxed at the cost of more complicated proofs.

A.5 mainly specifies moment conditions on $\varepsilon_{it}$, $\lambda^0_i$, $F^0_i$, $X_{it}$, $\Delta(X_{it})$ and $D_\Delta(X_{it})$ as well in the case where $\Delta(\cdot)$ is not uniformly bounded. A.6 specifies conditions on the kernel function $k(\cdot)$ which, in conjunction with A.4(i) and A.5(iv) are mainly used to demonstrate that $\max_{1 \leq i,j \leq N} \|K_{ij}\| = O_P(1)$ in Lemma D.1 in Appendix D. A.7 specifies conditions on the bandwidth in relation to the sample sizes $(N,T)$. Note that $NT\delta^{-4}_N(h!)/2 \to 0$ is equivalent to $(NT^{-1}T^{-1}) (h!)^{1/2} \to 0$, which restricts the relative speed at which $N$ and $T$ diverge to $\infty$ in relation with $h!$.

Let $\Pi_{NT}$ be a $p \times 1$ vector whose $k$th element is given by\(^1\)

$$\Pi_{NT,k} = (NT)^{-1} \text{tr} \left( M_N \mathbf{X}_k M_F \Delta \right).$$

(3.5)

Then under Assumption A.3,

$$\hat{\beta} - \beta^0 = \gamma_{NT} D_{NT}^{-1} \Pi_{NT} + O_P \left( \delta^{-2}_{NT} \right).$$

(3.6)

Let

$$B_{1NT} = \left( \frac{h!}{NT} \right)^{1/2} \sum_{i=1}^N \varepsilon_i^t M_F \Delta_i M_F \varepsilon_i,$$

(3.7)

$$B_{2NT} = \left( \frac{1}{NT} \right)^2 \sum_{1 \leq i,j \leq N} \left( M_F \Delta_i - \hat{X}_i D_{NT}^{-1} \Pi_{NT} \right)^t K_{ij} \left( M_F \Delta_j - \hat{X}_j D_{NT}^{-1} \Pi_{NT} \right),$$

(3.8)

$$V_{NT} = \left( \frac{2h!}{(NT)^2} \right) \sum_{1 \leq i,j \leq N} \sum_{t=1}^T \sum_{s=1}^T E_D \left( K_{ij,ts}^2 \varepsilon_i^t \varepsilon_j^s \right).$$

(3.9)

As will be clear, $B_{1NT}$ and $V_{NT}$ stand for the asymptotic bias and variance of our test statistic, respectively; $B_{2NT}$ contributes to its asymptotic local power. The following theorem states the asymptotic distribution of the test statistic $J_{NT}$ under $\mathbb{H}_1(\gamma_{NT})$.

**Theorem 3.2** Suppose Assumptions A.1-A.7 hold. Then under $\mathbb{H}_1(\gamma_{NT})$ with $\gamma_{NT} \equiv (NT)^{-1/2} (h!)^{-1/4}$,

$$NT (h!)^{1/2} J_{NT} - B_{1NT} \overset{D}{\to} N \left( B_2, V_0 \right)$$

where $B_2 = \text{plim}_{(N,T) \to \infty} B_{2NT}$ and $V_0 = \text{plim}_{(N,T) \to \infty} V_{NT}$.

\(^1\)Using the notation $X_i$, one can also write $\Pi_{NT} = \frac{1}{NT} \sum_{i=1}^N X_i \Delta_i$. 

11
Remark 2. The proof of the above theorem is tedious and is relegated to Appendix B. The idea is simple but the details are quite involved. We can show that $NT (h!)^{1/2} J_{NT} - B_{1NT} - B_{2NT} = A_{NT} + o_P (1)$ under $H_1 (\gamma_{NT})$, where

$$A_{NT} \equiv \sum_{1 \leq i < j \leq N} W_{NT} (u_i, u_j)$$

$W_{NT} (u_i, u_j) \equiv 2 (h!)^{1/2} \sum_{1 \leq t, s \leq T} \varepsilon_i K_{ij,ts} \varepsilon_s$ and $u_i \equiv (X_i, \varepsilon_i)$. Noting that $A_{NT}$ is a degenerate second order $U$-statistic, we apply a conditional version of de Jong's (1987) central limit theorem (CLT) for independently but nonidentically distributed (INID) observations to show that $A_{NT} \overset{D}{\to} N (0, V_0)$ under Assumptions A.1-A.7.\(^2\)

In view of the fact $B_{2NT} = 0$ under $H_0$, an immediate consequence of Theorem 3.2 is

$$NT (h!)^{1/2} J_{NT} - B_{1NT} \overset{D}{\to} N (0, V_0) \text{ under } H_0.$$

To implement the test, we need to estimate the asymptotic bias $B_{1NT}$ and asymptotic variance $V_{NT}$ consistently under $H_0$. We propose to estimate $B_{1NT}$ and $V_{NT}$ respectively by

$$\hat{B}_{1NT} \equiv (h!)^{1/2} \sum_{i=1}^N \hat{\varepsilon}_i K_{ii} \hat{\varepsilon}_i \text{ and } \hat{V}_{NT} \equiv \frac{2h!}{(NT)^2} \sum_{1 \leq t, s \leq T} \sum_{1 \leq i \neq j \leq N} K_{ij,ts} \hat{\varepsilon}_i \hat{\varepsilon}_j \hat{\varepsilon}_s.$$

Then we define a feasible test statistic

$$\hat{\Gamma}_{NT} \equiv \left( NT (h!)^{1/2} J_{NT} - \hat{B}_{1NT} \right) / \sqrt{\hat{V}_{NT}}.$$

The following theorem establishes the asymptotic distribution of $\hat{\Gamma}_{NT}$ under $H_1 (\gamma_{NT})$.

**Theorem 3.3** Suppose Assumptions A.1-A.7 hold. Then under $H_1 (\gamma_{NT})$, $\hat{\Gamma}_{NT} \overset{D}{\to} N \left( B_2 / \sqrt{V_0}, 1 \right).$

Remark 3. The above theorem implies that the test has nontrivial asymptotic power against local alternatives that converge to the null at the rate $\gamma_{NT} = (NT)^{-1/2} (h!)^{-1/4}$. The local power function is given by

$$\Pr \left( \hat{\Gamma}_{NT} > z \mid H_1 (\gamma_{NT}) \right) \rightarrow 1 - \Phi \left( z - B_2 / \sqrt{V_0} \right) \text{ as } (N, T) \rightarrow \infty,$$

where $\Phi (\cdot)$ is the standard normal cumulative distribution function (CDF). We obtain this distributional result despite the fact that the unobserved factors $f_i^0$ and factor loadings $\lambda_i^0$ can only be estimated at a slower rates $(N^{-1/2}$ for the former and $T^{-1/2}$ for the latter, subject to certain matrix rotation). Even though the slow convergence rates of these factors and factor loadings estimates do not have adverse asymptotic effects on the estimation of the bias term $B_{1NT}$, the variance term $V_{NT}$, and the asymptotic distribution of $\hat{\Gamma}_{NT}$, they may play an important role in finite samples. For this reason, we will also propose a bootstrap procedure to obtain the bootstrap $p$-values for our test.

Again, under $H_0$, $B_2 = 0$, and $\hat{\Gamma}_{NT}$ is asymptotically distributed $N (0, 1)$. This is stated in the following corollary.

\(^2\)The CLT in de Jong (1987) works for second order $U$-statistics associated with INID observations. A close examination of his proof shows that it also works for conditionally independent but nonidentically distributed (CINID) observations.
Theorem 3.3 holds. Then under $H_0$, $\hat{\Gamma}_{NT} \overset{D}{\rightarrow} N(0,1)$.

In principle, one can compare $\hat{\Gamma}_{NT}$ with the one-sided critical value $z_\alpha$, the upper $\alpha$th percentile from the standard normal distribution, and reject the null hypothesis when $\hat{\Gamma}_{NT} > z_\alpha$ at $\alpha$ significance level.

**Remark 4.** Theorem 3.1 says nothing about the asymptotic property of the QMLE $\hat{\beta}$ under the global alternative $H_1$. In this case, we can define the pseudo-true parameter $\beta^*$ as the probability limit of $\hat{\beta}$. Then $\Delta(X_{it}) = m(X_{it}) - \beta^*X_{it}$ does not equal 0 a.s. Let $\tilde{\Delta}$ be analogously defined as $\Delta$ but with the local deviation $\Delta(X_{it})$ replaced by the global one $\tilde{\Delta}(X_{it})$. In this case, we can show that under the additional assumption $\|\tilde{\Delta}\| = o_P((NT)^{1/2})$,

$$\hat{\beta} - \beta^* = D_{NT}^{-1}\tilde{\Pi}_{NT} + o_P(1)$$

where $\tilde{\Pi}_{NT}$ is a $p \times 1$ vector whose $k$th element is given by $\tilde{\Pi}_{NT,k} = (NT)^{-1}\text{tr}(M_{x_0}X_kM_{x_0}D')$. In addition, following the proof of Theorem 3.2, we can show that

$$J_{NT} = \frac{1}{(NT)^{2}} \sum_{1 \leq i,j \leq N} \left( M_{x_0}\tilde{D}_{i} - \tilde{X}_i D_{NT}^{-1}\tilde{\Pi}_{NT} \right)'K_{ij} \left( M_{x_0}\tilde{D}_{j} - \tilde{X}_j D_{NT}^{-1}\tilde{\Pi}_{NT} \right) + o_P(1)$$

which has a positive probability limit. This, together with the fact that $\tilde{B}_{1NT} = O_P((hl)^{-1/2})$ and $\hat{V}_{NT}$ has a well behaved probability limit under $H_1$, implies that our test statistic $\hat{\Gamma}_{NT}$ diverges at the usual nonparametric rate $NT(hl)^{1/2}$ under $H_1$. That is

$$\Pr(\hat{\Gamma}_{NT} > b_{NT} | H_1) \rightarrow 1 \text{ as } (N,T) \rightarrow \infty$$

for any nonstochastic sequence $b_{NT} = o(NT(hl)^{1/2})$. So our test achieves consistency against any fixed global alternatives.

### 3.3 A Bootstrap version of the tests

Despite the fact that Corollary 3.4 provides the asymptotic normal null distribution for our test statistic, we cannot rely on the asymptotic normal critical values to make inference for two reasons. One is inherited from many kernel-based nonparametric tests, and the other is associated with the slow convergence rates of the factors and factors loadings estimates as mentioned above. It is well known that the asymptotic normal distribution may not serve as a good approximation for many kernel-based tests and tests based on normal critical values can be sensitive to the choice of bandwidths and suffer from substantial finite sample size distortions. The slow convergence of the estimates of factors and factor loadings may further lead to some finite sample size distortions. Therefore it is worthwhile to propose a bootstrap procedure to improve the finite sample performance of our test. Below we propose a fixed-regressor wild bootstrap method in the spirit of Hansen (2000). The procedure goes as follows:
1. Obtain the restricted residuals \( \hat{\varepsilon}_{it} = Y_{it} - X'_{it}\hat{\beta} - \hat{\Phi}_i\hat{\lambda}_i \) where \( \hat{\beta}, \hat{\Phi}_i \) and \( \hat{\lambda}_i \) are estimates under the null hypothesis of linearity. Calculate the test statistic \( \hat{\Gamma}_{NT} \) based on \( \{\hat{\varepsilon}_{it}, X_{it}\} \).

2. For \( i = 1, \ldots, N \) and \( t = 1, 2, \ldots, T \), obtain the bootstrap error \( \tilde{\varepsilon}_{it} = \varepsilon_{it}\tilde{\eta}_{it} \) where \( \tilde{\eta}_{it} \) are independently and identically distributed (IID) \( N(0, 1) \) across \( i \) and \( t \). Generate the bootstrap analogue \( \tilde{Y}_{it}^* \) of \( Y_{it} \) by holding \( (X_{it}, \tilde{\Phi}_i, \tilde{\lambda}_i) \) as fixed: \( \tilde{Y}_{it}^* = \tilde{\beta} X_{it} + \tilde{\Phi}_i + \tilde{\varepsilon}_{it} \) for \( i = 1, 2, \ldots, N \) and \( t = 1, 2, \ldots, T \).\(^3\)

3. Given the bootstrap resample \( \{\tilde{Y}_{it}^*, X_{it}\} \), obtain the QMLEs \( \tilde{\beta}^*, \tilde{\Phi}_i^* \) and \( \tilde{\lambda}_i^* \). Obtain the residuals \( \tilde{\varepsilon}_{it}^* = Y_{it}^* - X_{it}\tilde{\beta}^* - \tilde{\Phi}_i^*\tilde{\lambda}_i^* \) and calculate the bootstrap test statistic \( \tilde{\Gamma}_{NT}^* \) based on \( \{\tilde{\varepsilon}_{it}, X_{it}\} \).

4. Repeat Steps 2-3 for \( B \) times and index the bootstrap statistics as \( \{\tilde{\Gamma}_{NT,b}^*\}_{b=1}^B \). The bootstrap \( p \)-value is calculated as \( p^* \equiv B^{-1}\sum_{b=1}^B 1(\tilde{\Gamma}_{NT,b}^* \geq \hat{\Gamma}_{NT}) \), where \( 1(\cdot) \) is the usual indicator function.

It is straightforward to implement the above bootstrap procedure. Note that we impose the null hypothesis of linearity in Step 2. Following Su and Chen (2012), we can readily establish the asymptotic validity of the above bootstrap procedure. To save space, we only state the result here.

**Theorem 3.5** Suppose the conditions in Theorem 3.3 hold. Then \( \hat{\Gamma}_{NT} \to^D N(0, 1) \) conditionally on the observed sample \( \mathcal{W}_{NT} = \{(X_1, Y_1), \ldots, (X_N, Y_N)\} \).

The above result holds no matter whether the original sample satisfies the null, local alternative or global alternative hypothesis. On the one hand, if \( \mathbb{H}_0 \) holds for the original sample, \( \hat{\Gamma}_{NT} \) also converges in distribution to \( N(0, 1) \) so that a test based on the bootstrap \( p \)-value will have the right asymptotic level. On the other hand, if \( \mathbb{H}_1 \) holds for the original sample, as we argue in Remark 4, \( \hat{\Gamma}_{NT} \) diverges at rate \( NT(h!)^{1/2} \) whereas \( \hat{\Gamma}_{NT}^* \) is asymptotically \( N(0, 1) \), which implies the consistency of the bootstrap-based test.

4 **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 our test to an economic growth panel dataset.

4.1 **Monte Carlo Simulation Study**

4.1.1 **Data generating processes**

We consider the following six data generating processes (DGPs)

<table>
<thead>
<tr>
<th>DGP</th>
<th>Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( Y_{it} = \rho Y_{i,t-1} + \lambda_{it} F_i^t + \varepsilon_{it} )</td>
</tr>
<tr>
<td>2</td>
<td>( Y_{it} = \beta_1 X_{i,t-1} + \beta_2 X_{i,t-2} + \lambda_{it} F_i^t + \varepsilon_{it} )</td>
</tr>
<tr>
<td>3</td>
<td>( Y_{it} = \rho Y_{i,t-1} + \beta_1 X_{i,t-1} + \beta_2 X_{i,t-2} + \lambda_{it} F_i^t + \varepsilon_{it} )</td>
</tr>
<tr>
<td>4</td>
<td>( Y_{it} = \delta (Y_{i,t-1} - Y_{i,t-2}) + \lambda_{it} F_i^t + \varepsilon_{it} )</td>
</tr>
<tr>
<td>5</td>
<td>( Y_{it} = \beta_1 X_{i,t-1} + \beta_2 X_{i,t-2} + 2\delta X_{i,t} \Phi (X_{i,t-2}) + \lambda_{it} F_i^t + \varepsilon_{it} )</td>
</tr>
</tbody>
</table>

\( ^3 \)This is the case even if \( X_{it} \) contains lagged dependent variables, say, \( Y_{i,t-1} \) and \( Y_{i,t-2} \).
DGP 6: \( Y_{it} = \frac{1}{2} \Phi(Y_{i,t-2}) Y_{i,t-2} + \beta_0 X_{it,1} + \beta_2 X_{it,2} + \delta X_{it,1} \Phi(X_{it,2}) + \lambda_i \Phi^0 + \epsilon_{it} \),
where \( i = 1, 2, \ldots, N, t = 1, 2, \ldots, T \), \( (\rho_0, \rho_1, \rho_2) = (0.6, 1, 3) \), \( \delta = 0.25 \), \( \epsilon_{it} \sim I(0, 1) \), and \( \Phi(\cdot) \) is the standard normal CDF. Here \( \lambda_i^0 = (\lambda_{i1}, \lambda_{i2})' \), \( F_t^0 = (F_{t1}, F_{t2})' \), and the regressors are generated according to

\[
X_{it,1} = \mu_1 + c_1 \lambda_i^0 F_t^0 + \eta_{it,1},
\]

\[
X_{it,2} = \mu_2 + c_2 \lambda_i^0 F_t^0 + \eta_{it,2},
\]

where the variables \( \lambda_{ij}, F_{ij}^0, \) and \( \eta_{it,j}, j = 1, 2 \), are all IID \( N(0, 1) \), mutually independent of each other, and independent of \( \{\epsilon_{it}\} \). Clearly, the regressors \( X_{it,1} \) and \( X_{it,2} \) are correlated with \( \lambda_i^0 \) and \( F_t^0 \). We set \( \mu_1 = c_1 = 0.25 \) and \( \mu_2 = c_2 = 0.5 \). Note that DGPs 1-3 are used for the level study and DGPs 4-6 for the power study. For the dynamic models (DGPs 1, 3, 4 and 6), we discard the first 100 observations along the time dimension when generating the data.

Note that the idiosyncratic error terms in the above six DGPs are all homoskedastic both conditionally and unconditionally. To allow for conditional heteroskedasticity, which may be relevant in empirical applications, we consider another set of DGPs, namely DGPs 1h-6h which are identical to DGPs 1-6, respectively in the mean regression components but different from the latter in the generation of the idiosyncratic error terms. For DGPs 1h and 4h, the errors are generated from the process

\[
\epsilon_{it} = \sigma_{it} \epsilon_{it},
\]

\[
\sigma_{it} = (0.1 + 0.2 Y_{i,t-1}^2)^{1/2},
\]

\[
\epsilon_{it} \sim I(0, 1).
\]

For DGPs 2h-3h and 5h-6h, the errors are generated from the process

\[
\epsilon_{it} = \sigma_{it} \epsilon_{it},
\]

\[
\sigma_{it} = (0.1 + 0.1 (X_{it,1}^2 + X_{it,2}^2))^{1/2},
\]

\[
\epsilon_{it} \sim I(0, 1).
\]

### 4.1.2 Implementation

To calculate the test statistic, we need to choose both the kernel function and bandwidth parameter \( h = (h_1, \ldots, h_p) \) where \( p = 1 \) in DGPs 1, 4, 1h, and 4h; \( p = 2 \) in DGPs 2, 5, 2h, and 5h; and \( p = 3 \) in DGPs 3, 6, 3h and 6h. Let \( X_{it} \) denote the collection of the observable regressors in the above DGPs. For example, \( X_{it} = (Y_{i,t-1}, X_{it,1}, X_{it,2})' \) in DGPs 3, 6, 3h, and 6h. Throughout, we use the Gaussian kernel \( k(u) = (2\pi)^{-1/2} \exp (-u^2/2) \), and choose the bandwidth by the “rule of thumb”: \( h_l = c_0 s_l (NT)^{-1/(4+p)} \), where \( s_l \) stands for the sample standard deviation for the \( l \)th element in \( X_{it} \). We set \( c_0 = 0.5, 1, \) and \( 2 \) to examine the sensitivity of our test to the choice of bandwidth. We leave the development of a data-driven rule for the selection of an “optimal” bandwidth for future research.\(^4\)

\(^4\) Alternatively, we conjecture that one can follow Horowitz and Spokoiny (2001) and Chen and Gao (2007) and prove the rate-optimality of our test with panel data. If this is the case, then in practice one can choose the bandwidth as in these papers.
For the \((N,T)\) pair, we consider \(N,T = 20, 40,\) and \(60.\) For each scenario, we use 250 replications and 200 bootstrap resamples in each replication.

To implement the testing procedure, we need to obtain the estimators under the null hypothesis of linearity. We first obtain the initial estimators of \((\beta^0, \lambda^0, F^0)\) using Bai’s (2009) principal component approach, and then calculate the bias corrected QMLE estimator \((\hat{\beta}, \hat{\lambda}, \hat{F})\) following Moon and Weidner (2010) (see their section 3.3 in particular). We then calculate the bootstrap test statistic \(\hat{\Gamma}_{NT}^*\) based on the bias corrected QMLE estimators.

### 4.1.3 Test results

Table 1 reports the empirical rejection frequencies of our test at 1%, 5%, and 10% nominal levels when the null hypothesis holds true in DGPs 1-3 for different bandwidth choices, assuming that the idiosyncratic error terms \(\varepsilon_{it}\) are IID \(N(0, 1)\) and independent of \(\lambda^0_{ij}, F^0_{ij},\) and \(\eta_{it,j}.\) Table 1 suggests that the level of our test behaves reasonably well across all DGPs and sample sizes under investigation. More importantly, the level of our test is robust to different choices of bandwidth.

Table 2 reports the finite sample power of our test against the alternatives specified in DGPs 4-6 at 1%, 5%, and 10% nominal levels. We present some important findings from Table 2. First, as either \(N\) or \(T\) increases, the power of our test generally increases and finally reaches 1. Second, the choice of bandwidth appears to have some effect on the power of the test. A larger value of \(c\) leads to a larger testing power.

Tables 3 and 4 report the simulation results for DGPs 1h-6h when the idiosyncratic errors are conditionally heteroskedastic. To a large extent the results are similar to the homoskedastic case, although there are some slight differences. For pure dynamic models (DGP 1h), the levels of our test in the heteroskedastic case oversize in some scenarios. For example, when \((N,T) = (20, 40), (20, 60), (40, 20),\) \(c = 1\) and \(2,\) there are slightly more size distortions of our test at the 5% and 10% nominal levels in the heteroskedastic case than in the homoskedastic case; however, for DGPs 2h-3h, the levels of our test in the heteroskedastic case generally perform similarly or slightly better than the corresponding homoskedastic cases in DGPs 2-3. In addition, the power of our test continues to perform well in the case of heteroskedasticity.

### 4.2 An application to the economic growth data

In this application we consider nonparametric dynamic panel data models for the economic growth data which incorporate common shocks. We shall consider the model

\[
Y_{it} = m (Y_{i,t-1}, \ldots Y_{i,t-s}, X_{it}) + F^0_t \lambda^0_i + \varepsilon_{it},
\]

where \(Y_{it} = \log(GDP_{it}) - \log(GDP_{i,t-1})\) denotes the growth rate of GDP for country \(i\) in year \(t,\) and \(GDP_{it}\) is the real GDP per worker of country \(i\) over year \(t.\) We set \(s = 1, 2, 3\) to allow for different time lags in the regressor. \(F_t\) denotes common shocks, e.g. technological shocks and financial crises, and \(\lambda_i\) represents the heterogeneous impact of common shocks on country \(i.\) We are interested in examining the relation between a country’s economic growth and its initial economic condition as well as the relation
Table 1: Finite sample rejection frequency for DGPs 1-3 (level study, homoskedastic case)

<table>
<thead>
<tr>
<th>DGP</th>
<th>N</th>
<th>T</th>
<th>c = 0.5</th>
<th></th>
<th>c = 1</th>
<th></th>
<th>c = 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>1%</td>
<td>5%</td>
<td>10%</td>
<td>1%</td>
<td>5%</td>
<td>10%</td>
</tr>
<tr>
<td>1</td>
<td>20</td>
<td>20</td>
<td>0.008</td>
<td>0.052</td>
<td>0.124</td>
<td>0.004</td>
<td>0.056</td>
<td>0.116</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>40</td>
<td>0.024</td>
<td>0.088</td>
<td>0.124</td>
<td>0.020</td>
<td>0.076</td>
<td>0.136</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>60</td>
<td>0.000</td>
<td>0.036</td>
<td>0.092</td>
<td>0.004</td>
<td>0.036</td>
<td>0.116</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>20</td>
<td>0.020</td>
<td>0.080</td>
<td>0.132</td>
<td>0.012</td>
<td>0.060</td>
<td>0.116</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>40</td>
<td>0.032</td>
<td>0.088</td>
<td>0.144</td>
<td>0.028</td>
<td>0.084</td>
<td>0.148</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>60</td>
<td>0.000</td>
<td>0.040</td>
<td>0.100</td>
<td>0.004</td>
<td>0.032</td>
<td>0.112</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>20</td>
<td>0.008</td>
<td>0.040</td>
<td>0.096</td>
<td>0.004</td>
<td>0.048</td>
<td>0.092</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>40</td>
<td>0.024</td>
<td>0.076</td>
<td>0.116</td>
<td>0.028</td>
<td>0.088</td>
<td>0.116</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>60</td>
<td>0.000</td>
<td>0.064</td>
<td>0.096</td>
<td>0.012</td>
<td>0.036</td>
<td>0.096</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>20</td>
<td>0.016</td>
<td>0.052</td>
<td>0.112</td>
<td>0.020</td>
<td>0.060</td>
<td>0.108</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>40</td>
<td>0.008</td>
<td>0.060</td>
<td>0.108</td>
<td>0.012</td>
<td>0.044</td>
<td>0.116</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>60</td>
<td>0.008</td>
<td>0.036</td>
<td>0.072</td>
<td>0.016</td>
<td>0.040</td>
<td>0.100</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>20</td>
<td>0.012</td>
<td>0.040</td>
<td>0.076</td>
<td>0.008</td>
<td>0.028</td>
<td>0.076</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>40</td>
<td>0.008</td>
<td>0.044</td>
<td>0.092</td>
<td>0.000</td>
<td>0.052</td>
<td>0.092</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>60</td>
<td>0.012</td>
<td>0.056</td>
<td>0.088</td>
<td>0.004</td>
<td>0.052</td>
<td>0.096</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>20</td>
<td>0.012</td>
<td>0.036</td>
<td>0.088</td>
<td>0.016</td>
<td>0.056</td>
<td>0.084</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>40</td>
<td>0.016</td>
<td>0.040</td>
<td>0.076</td>
<td>0.008</td>
<td>0.048</td>
<td>0.084</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>60</td>
<td>0.016</td>
<td>0.068</td>
<td>0.104</td>
<td>0.020</td>
<td>0.068</td>
<td>0.108</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>20</td>
<td>0.016</td>
<td>0.080</td>
<td>0.116</td>
<td>0.020</td>
<td>0.064</td>
<td>0.096</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>40</td>
<td>0.028</td>
<td>0.064</td>
<td>0.100</td>
<td>0.012</td>
<td>0.064</td>
<td>0.124</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>60</td>
<td>0.020</td>
<td>0.092</td>
<td>0.120</td>
<td>0.024</td>
<td>0.064</td>
<td>0.116</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>20</td>
<td>0.012</td>
<td>0.064</td>
<td>0.104</td>
<td>0.020</td>
<td>0.052</td>
<td>0.140</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>40</td>
<td>0.004</td>
<td>0.052</td>
<td>0.112</td>
<td>0.008</td>
<td>0.056</td>
<td>0.116</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>60</td>
<td>0.020</td>
<td>0.056</td>
<td>0.096</td>
<td>0.016</td>
<td>0.040</td>
<td>0.068</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>20</td>
<td>0.020</td>
<td>0.044</td>
<td>0.116</td>
<td>0.008</td>
<td>0.052</td>
<td>0.108</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>40</td>
<td>0.008</td>
<td>0.088</td>
<td>0.136</td>
<td>0.012</td>
<td>0.040</td>
<td>0.088</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>60</td>
<td>0.012</td>
<td>0.064</td>
<td>0.128</td>
<td>0.028</td>
<td>0.060</td>
<td>0.096</td>
</tr>
</tbody>
</table>

Note: The bandwidth is chosen as $h = cS_X (NT)^{-1/(4+p)}$ where $S_X$ is the sample standard deviation of $\{X_{it}, i = 1, ..., N, t = 1, ..., T\}$. 

17
Table 2: Finite sample rejection frequency for DGPs 4-6 (power study, homoskedastic case)

<table>
<thead>
<tr>
<th>DGP</th>
<th>N</th>
<th>T</th>
<th>c = 0.5</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>1%</td>
<td>5%</td>
<td>10%</td>
<td>1%</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
<td>20</td>
<td>0.076</td>
<td>0.260</td>
<td>0.384</td>
<td>0.144</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>40</td>
<td>0.296</td>
<td>0.568</td>
<td>0.704</td>
<td>0.424</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>60</td>
<td>0.528</td>
<td>0.748</td>
<td>0.844</td>
<td>0.668</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>20</td>
<td>0.300</td>
<td>0.544</td>
<td>0.684</td>
<td>0.384</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>40</td>
<td>0.756</td>
<td>0.904</td>
<td>0.944</td>
<td>0.864</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>60</td>
<td>0.928</td>
<td>0.976</td>
<td>0.996</td>
<td>0.956</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>20</td>
<td>0.476</td>
<td>0.772</td>
<td>0.888</td>
<td>0.636</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>40</td>
<td>0.956</td>
<td>0.996</td>
<td>1.000</td>
<td>0.988</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>60</td>
<td>0.996</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>20</td>
<td>0.028</td>
<td>0.116</td>
<td>0.180</td>
<td>0.056</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>40</td>
<td>0.056</td>
<td>0.140</td>
<td>0.256</td>
<td>0.112</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>60</td>
<td>0.060</td>
<td>0.200</td>
<td>0.316</td>
<td>0.192</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>20</td>
<td>0.020</td>
<td>0.120</td>
<td>0.212</td>
<td>0.052</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>40</td>
<td>0.100</td>
<td>0.268</td>
<td>0.380</td>
<td>0.284</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>60</td>
<td>0.220</td>
<td>0.432</td>
<td>0.576</td>
<td>0.544</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>20</td>
<td>0.040</td>
<td>0.204</td>
<td>0.308</td>
<td>0.144</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>40</td>
<td>0.180</td>
<td>0.468</td>
<td>0.612</td>
<td>0.540</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>60</td>
<td>0.500</td>
<td>0.724</td>
<td>0.828</td>
<td>0.876</td>
</tr>
<tr>
<td>6</td>
<td>20</td>
<td>20</td>
<td>0.016</td>
<td>0.104</td>
<td>0.192</td>
<td>0.088</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>40</td>
<td>0.036</td>
<td>0.140</td>
<td>0.220</td>
<td>0.200</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>60</td>
<td>0.092</td>
<td>0.244</td>
<td>0.376</td>
<td>0.380</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>20</td>
<td>0.028</td>
<td>0.112</td>
<td>0.220</td>
<td>0.120</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>40</td>
<td>0.100</td>
<td>0.316</td>
<td>0.460</td>
<td>0.568</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>60</td>
<td>0.308</td>
<td>0.556</td>
<td>0.684</td>
<td>0.824</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>20</td>
<td>0.072</td>
<td>0.252</td>
<td>0.400</td>
<td>0.316</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>40</td>
<td>0.284</td>
<td>0.540</td>
<td>0.660</td>
<td>0.832</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>60</td>
<td>0.568</td>
<td>0.776</td>
<td>0.868</td>
<td>0.952</td>
</tr>
</tbody>
</table>
Table 3: Finite sample rejection frequency for DGPs 1h-3h (level study, heteroskedastic case)

<table>
<thead>
<tr>
<th>DGP</th>
<th>N</th>
<th>T</th>
<th>c = 0.5</th>
<th></th>
<th>c = 1</th>
<th></th>
<th>c = 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>1%</td>
<td>5%</td>
<td>10%</td>
<td>1%</td>
<td>5%</td>
<td>10%</td>
</tr>
<tr>
<td>1h</td>
<td>20</td>
<td>20</td>
<td>0.016</td>
<td>0.068</td>
<td>0.124</td>
<td>0.016</td>
<td>0.056</td>
<td>0.140</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>40</td>
<td>0.040</td>
<td>0.084</td>
<td>0.144</td>
<td>0.036</td>
<td>0.084</td>
<td>0.148</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>60</td>
<td>0.016</td>
<td>0.068</td>
<td>0.148</td>
<td>0.024</td>
<td>0.084</td>
<td>0.156</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>20</td>
<td>0.028</td>
<td>0.056</td>
<td>0.112</td>
<td>0.024</td>
<td>0.076</td>
<td>0.128</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>40</td>
<td>0.032</td>
<td>0.084</td>
<td>0.144</td>
<td>0.028</td>
<td>0.084</td>
<td>0.134</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>60</td>
<td>0.008</td>
<td>0.044</td>
<td>0.104</td>
<td>0.008</td>
<td>0.060</td>
<td>0.112</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>20</td>
<td>0.004</td>
<td>0.052</td>
<td>0.100</td>
<td>0.012</td>
<td>0.048</td>
<td>0.104</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>40</td>
<td>0.012</td>
<td>0.048</td>
<td>0.108</td>
<td>0.016</td>
<td>0.052</td>
<td>0.096</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>60</td>
<td>0.013</td>
<td>0.038</td>
<td>0.113</td>
<td>0.013</td>
<td>0.044</td>
<td>0.106</td>
</tr>
<tr>
<td>2h</td>
<td>20</td>
<td>20</td>
<td>0.008</td>
<td>0.056</td>
<td>0.096</td>
<td>0.016</td>
<td>0.068</td>
<td>0.112</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>40</td>
<td>0.008</td>
<td>0.060</td>
<td>0.108</td>
<td>0.004</td>
<td>0.064</td>
<td>0.100</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>60</td>
<td>0.004</td>
<td>0.028</td>
<td>0.076</td>
<td>0.000</td>
<td>0.048</td>
<td>0.084</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>20</td>
<td>0.008</td>
<td>0.052</td>
<td>0.116</td>
<td>0.004</td>
<td>0.028</td>
<td>0.100</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>40</td>
<td>0.008</td>
<td>0.032</td>
<td>0.076</td>
<td>0.008</td>
<td>0.044</td>
<td>0.084</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>60</td>
<td>0.012</td>
<td>0.064</td>
<td>0.108</td>
<td>0.012</td>
<td>0.056</td>
<td>0.088</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>20</td>
<td>0.016</td>
<td>0.048</td>
<td>0.100</td>
<td>0.020</td>
<td>0.048</td>
<td>0.092</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>40</td>
<td>0.020</td>
<td>0.052</td>
<td>0.092</td>
<td>0.020</td>
<td>0.060</td>
<td>0.084</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>60</td>
<td>0.012</td>
<td>0.064</td>
<td>0.112</td>
<td>0.016</td>
<td>0.052</td>
<td>0.096</td>
</tr>
<tr>
<td>3h</td>
<td>20</td>
<td>20</td>
<td>0.012</td>
<td>0.052</td>
<td>0.080</td>
<td>0.016</td>
<td>0.044</td>
<td>0.088</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>40</td>
<td>0.020</td>
<td>0.072</td>
<td>0.116</td>
<td>0.024</td>
<td>0.076</td>
<td>0.136</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>60</td>
<td>0.008</td>
<td>0.064</td>
<td>0.100</td>
<td>0.020</td>
<td>0.076</td>
<td>0.116</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>20</td>
<td>0.008</td>
<td>0.060</td>
<td>0.096</td>
<td>0.012</td>
<td>0.088</td>
<td>0.136</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>40</td>
<td>0.004</td>
<td>0.036</td>
<td>0.100</td>
<td>0.004</td>
<td>0.036</td>
<td>0.116</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>60</td>
<td>0.016</td>
<td>0.060</td>
<td>0.108</td>
<td>0.020</td>
<td>0.052</td>
<td>0.092</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>20</td>
<td>0.008</td>
<td>0.048</td>
<td>0.088</td>
<td>0.008</td>
<td>0.048</td>
<td>0.116</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>40</td>
<td>0.016</td>
<td>0.040</td>
<td>0.104</td>
<td>0.012</td>
<td>0.036</td>
<td>0.076</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>60</td>
<td>0.004</td>
<td>0.044</td>
<td>0.104</td>
<td>0.008</td>
<td>0.060</td>
<td>0.124</td>
</tr>
</tbody>
</table>
Table 4: Finite sample rejection frequency for DGPs 4h-6h (power study, heteroskedastic case)

<table>
<thead>
<tr>
<th>DGP</th>
<th>N</th>
<th>T</th>
<th>c = 0.5</th>
<th></th>
<th></th>
<th></th>
<th>c = 1</th>
<th></th>
<th></th>
<th></th>
<th>c = 2</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>1%</td>
<td>5%</td>
<td>10%</td>
<td>1%</td>
<td>5%</td>
<td>10%</td>
<td>1%</td>
<td>5%</td>
<td>10%</td>
<td>1%</td>
<td>5%</td>
</tr>
<tr>
<td>4h</td>
<td>20</td>
<td>20</td>
<td>0.152</td>
<td>0.352</td>
<td>0.460</td>
<td>0.228</td>
<td>0.448</td>
<td>0.564</td>
<td>0.296</td>
<td>0.528</td>
<td>0.656</td>
<td>0.296</td>
<td>0.528</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>40</td>
<td>0.500</td>
<td>0.704</td>
<td>0.796</td>
<td>0.608</td>
<td>0.832</td>
<td>0.900</td>
<td>0.708</td>
<td>0.900</td>
<td>0.936</td>
<td>0.708</td>
<td>0.900</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>60</td>
<td>0.660</td>
<td>0.848</td>
<td>0.928</td>
<td>0.784</td>
<td>0.932</td>
<td>0.964</td>
<td>0.852</td>
<td>0.964</td>
<td>0.976</td>
<td>0.852</td>
<td>0.964</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>20</td>
<td>0.404</td>
<td>0.636</td>
<td>0.752</td>
<td>0.520</td>
<td>0.736</td>
<td>0.828</td>
<td>0.632</td>
<td>0.816</td>
<td>0.896</td>
<td>0.632</td>
<td>0.816</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>40</td>
<td>0.856</td>
<td>0.972</td>
<td>0.984</td>
<td>0.920</td>
<td>0.980</td>
<td>0.992</td>
<td>0.936</td>
<td>0.988</td>
<td>0.992</td>
<td>0.936</td>
<td>0.988</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>60</td>
<td>0.976</td>
<td>0.996</td>
<td>1.000</td>
<td>0.984</td>
<td>0.996</td>
<td>1.000</td>
<td>0.992</td>
<td>1.000</td>
<td>1.000</td>
<td>0.992</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>20</td>
<td>0.648</td>
<td>0.892</td>
<td>0.948</td>
<td>0.752</td>
<td>0.948</td>
<td>0.980</td>
<td>0.820</td>
<td>0.972</td>
<td>0.984</td>
<td>0.820</td>
<td>0.972</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>40</td>
<td>0.948</td>
<td>1.000</td>
<td>1.000</td>
<td>0.956</td>
<td>1.000</td>
<td>1.000</td>
<td>0.964</td>
<td>0.996</td>
<td>1.000</td>
<td>0.964</td>
<td>0.996</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>60</td>
<td>0.996</td>
<td>1.000</td>
<td>1.000</td>
<td>0.996</td>
<td>1.000</td>
<td>1.000</td>
<td>0.996</td>
<td>1.000</td>
<td>1.000</td>
<td>0.996</td>
<td>1.000</td>
</tr>
<tr>
<td>5h</td>
<td>20</td>
<td>20</td>
<td>0.096</td>
<td>0.232</td>
<td>0.344</td>
<td>0.192</td>
<td>0.420</td>
<td>0.612</td>
<td>0.220</td>
<td>0.536</td>
<td>0.716</td>
<td>0.220</td>
<td>0.536</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>40</td>
<td>0.360</td>
<td>0.552</td>
<td>0.684</td>
<td>0.644</td>
<td>0.820</td>
<td>0.896</td>
<td>0.756</td>
<td>0.920</td>
<td>0.980</td>
<td>0.756</td>
<td>0.920</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>60</td>
<td>0.532</td>
<td>0.776</td>
<td>0.832</td>
<td>0.852</td>
<td>0.956</td>
<td>0.976</td>
<td>0.952</td>
<td>0.996</td>
<td>0.996</td>
<td>0.952</td>
<td>0.996</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>20</td>
<td>0.304</td>
<td>0.572</td>
<td>0.716</td>
<td>0.616</td>
<td>0.856</td>
<td>0.904</td>
<td>0.740</td>
<td>0.940</td>
<td>0.980</td>
<td>0.740</td>
<td>0.940</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>40</td>
<td>0.772</td>
<td>0.928</td>
<td>0.964</td>
<td>0.960</td>
<td>0.992</td>
<td>1.000</td>
<td>0.984</td>
<td>1.000</td>
<td>1.000</td>
<td>0.984</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>60</td>
<td>0.960</td>
<td>0.988</td>
<td>0.992</td>
<td>0.996</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>20</td>
<td>0.516</td>
<td>0.772</td>
<td>0.864</td>
<td>0.828</td>
<td>0.960</td>
<td>0.988</td>
<td>0.904</td>
<td>0.992</td>
<td>0.992</td>
<td>0.904</td>
<td>0.992</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>40</td>
<td>0.964</td>
<td>0.996</td>
<td>1.000</td>
<td>0.992</td>
<td>1.000</td>
<td>1.000</td>
<td>0.992</td>
<td>1.000</td>
<td>1.000</td>
<td>0.992</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>60</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>6h</td>
<td>20</td>
<td>20</td>
<td>0.124</td>
<td>0.292</td>
<td>0.388</td>
<td>0.352</td>
<td>0.664</td>
<td>0.812</td>
<td>0.664</td>
<td>0.900</td>
<td>0.948</td>
<td>0.664</td>
<td>0.900</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>40</td>
<td>0.320</td>
<td>0.596</td>
<td>0.732</td>
<td>0.828</td>
<td>0.980</td>
<td>0.988</td>
<td>0.948</td>
<td>1.000</td>
<td>1.000</td>
<td>0.948</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>60</td>
<td>0.516</td>
<td>0.772</td>
<td>0.832</td>
<td>0.980</td>
<td>1.000</td>
<td>1.000</td>
<td>0.988</td>
<td>1.000</td>
<td>1.000</td>
<td>0.988</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>20</td>
<td>0.312</td>
<td>0.600</td>
<td>0.720</td>
<td>0.752</td>
<td>0.960</td>
<td>0.976</td>
<td>0.856</td>
<td>0.996</td>
<td>1.000</td>
<td>0.856</td>
<td>0.996</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>40</td>
<td>0.788</td>
<td>0.952</td>
<td>0.976</td>
<td>0.960</td>
<td>1.000</td>
<td>1.000</td>
<td>0.964</td>
<td>1.000</td>
<td>1.000</td>
<td>0.964</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>60</td>
<td>0.944</td>
<td>0.988</td>
<td>1.000</td>
<td>0.984</td>
<td>1.000</td>
<td>1.000</td>
<td>0.984</td>
<td>1.000</td>
<td>1.000</td>
<td>0.984</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>20</td>
<td>0.576</td>
<td>0.796</td>
<td>0.900</td>
<td>0.840</td>
<td>0.980</td>
<td>1.000</td>
<td>0.840</td>
<td>0.980</td>
<td>1.000</td>
<td>0.840</td>
<td>0.980</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>40</td>
<td>0.908</td>
<td>1.000</td>
<td>1.000</td>
<td>0.948</td>
<td>1.000</td>
<td>1.000</td>
<td>0.944</td>
<td>1.000</td>
<td>1.000</td>
<td>0.944</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>60</td>
<td>0.968</td>
<td>1.000</td>
<td>1.000</td>
<td>0.968</td>
<td>1.000</td>
<td>1.000</td>
<td>0.968</td>
<td>1.000</td>
<td>1.000</td>
<td>0.968</td>
<td>1.000</td>
</tr>
</tbody>
</table>
between a country’s economic growth and its capital accumulation. \( X_{it} \) thus includes two variables, a country’s initial economic condition \( (X_{i,1}) \), which is defined as the logarithm of country i’s real GDP per worker in the initial year, and its investment share \( (X_{it,2}) \), which is defined as the logarithm of the average share of physical investment of country i over its GDP in the \( t \)th year.

Different economic models predict different relations between economic growth and its initial condition. For example, Solow (1956) finds a negative relation between the two and Barro (1991) reinforces Solow’s prediction using a cross country data in the period of 1960 to 1985. On the other hand, the endogenous growth model (see Romer (1986) and Lucas (1988) for references) predicts that the initial economic conditions do not affect the long run economic growth. The relation between a country’s economic growth and its capital accumulation is not conclusive either. Solow (1956) argues there is no association between the two and Jones (1995) concludes this point empirically. The endogenous growth model predicts a positive relation and the argument is reinforced by Bond, Leblebicioglu and Schiantarelli (2010)’s empirical findings. Most of the empirical studies above use linear models despite the fact that there are no economic theories suggesting the two relations are linear. In view of this, Su and Lu (2012) apply a new nonparametric dynamic panel data model and find nonlinear relations between economic growth and its lagged value and initial condition.

The models we use are clearly different from Su and Lu (2012). Su and Lu (2012) use a short panel data with \( N = 71 \) and \( T = 4 \) which only allows for additive fixed effects as in traditional panel data models. Our model incorporates cross sectional dependence and allows for interactive fixed effects using a large dimensional panel dataset. We use data from the Penn World Table (PWT 7.1). The panel data covers 104 countries over 50 years (1960-2009). Following Bond, Leblebicioglu and Schiantarelli (2010), we exclude oil production countries and Botswana, because of the dominant role of mining. We also drop Nicaragua and Chad for negative record of gross investment in some years. China has two versions of variable values and we choose to use version one. The results are similar if we use version two instead.

We try different model specifications: pure dynamic models with \( s = 1, 2, \) and 3 respectively in (4.1), and dynamic models with \( 1 - 3 \) lags, and \( X_{i,1}, X_{it,2} \), or both as exogenous regressors in (4.1). Therefore we have the following twelve models in total.

Model 1: \( Y_{it} = m (Y_{i,t-1}) + F_t^0 \lambda_i^0 + \varepsilon_{it} \),
Model 2: \( Y_{it} = m (Y_{i,t-1}, X_{i,1}) + F_t^0 \lambda_i^0 + \varepsilon_{it} \),
Model 3: \( Y_{it} = m (Y_{i,t-1}, X_{it,2}) + F_t^0 \lambda_i^0 + \varepsilon_{it} \),
Model 4: \( Y_{it} = m (Y_{i,t-1}, X_{i,1}, X_{it,2}) + F_t^0 \lambda_i^0 + \varepsilon_{it} \),
Model 5: \( Y_{it} = m (Y_{i,t-1}, Y_{i,t-2}) + F_t^0 \lambda_i^0 + \varepsilon_{it} \),
Model 6: \( Y_{it} = m (Y_{i,t-1}, Y_{i,t-2}, X_{i,1}) + F_t^0 \lambda_i^0 + \varepsilon_{it} \),
Model 7: \( Y_{it} = m (Y_{i,t-1}, Y_{i,t-2}, X_{it,2}) + F_t^0 \lambda_i^0 + \varepsilon_{it} \),
Model 8: \( Y_{it} = m (Y_{i,t-1}, Y_{i,t-2}, X_{i,1}, X_{it,2}) + F_t^0 \lambda_i^0 + \varepsilon_{it} \),
Model 9: \( Y_{it} = m (Y_{i,t-1}, Y_{i,t-2}, Y_{i,t-3}) + F_t^0 \lambda_i^0 + \varepsilon_{it} \),
Model 10: \( Y_{it} = m (Y_{i,t-1}, Y_{i,t-2}, Y_{i,t-3}, X_{i,1}) + F_t^0 \lambda_i^0 + \varepsilon_{it} \),
Model 11: \( Y_{it} = m (Y_{i,t-1}, Y_{i,t-2}, Y_{i,t-3}, X_{it,2}) + F_t^0 \lambda_i^0 + \varepsilon_{it} \),
Model 12: \( Y_{it} = m (Y_{i,t-1}, Y_{i,t-2}, Y_{i,t-3}, X_{i,1}, X_{it,2}) + F_t^0 \lambda_i^0 + \varepsilon_{it} \).

In all these models, the number of factors has to be determined although it is assumed to be known.
in the theoretical development. Following Bai and Ng (2002), we use the following recommended criteria to choose the number of factors:

\[
P_{C_1} (R) = V \left( R, \hat{F}^R \right) + R \hat{\sigma}^2 \left( \frac{N + T}{NT} \right) \ln \left( \frac{NT}{N + T} \right),
\]

\[
P_{C_2} (R) = V \left( R, \hat{F}^R \right) + R \hat{\sigma}^2 \left( \frac{N + T}{NT} \right) \ln C_{NT}^2,
\]

\[
I_{C_1} (R) = \ln \left( V \left( R, \hat{F}^R \right) \right) + R \left( \frac{N + T}{NT} \right) \ln C_{NT}^2,
\]

\[
I_{C_2} (R) = \ln \left( V \left( R, \hat{F}^R \right) \right) + R \left( \frac{N + T}{NT} \right) \ln C_{NT}^2,
\]

where \( C_{NT}^2 = \min \{ N, T \} \), \( V \left( R, \hat{F}^R \right) = \frac{1}{2NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left( \hat{\epsilon}_it \right)^2 \), \( \hat{\epsilon}_it = Y_{it} - X_{it}' \hat{\beta} - F_{it}' \hat{\lambda}_i \), \( \hat{\lambda}_i \), \( \hat{\beta} \), \( F_{it}' \) and \( \hat{\lambda}_i \) are estimates under the null hypothesis of linearity when \( R \) factors are used, and \( \hat{\sigma}^2 \) is a consistent estimate of \( \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbb{E} \left( \hat{\epsilon}_it^2 \right) \) and can be replaced by \( V \left( R_{\text{max}}, \hat{F}^{R_{\text{max}}} \right) \) in applications. Following Bai and Ng (2002) we set \( R_{\text{max}} \) to be 8, 10 and 15, and recognize explicitly that \( P_{C_1} (R) \) and \( P_{C_2} (R) \) depend on the choice of \( R_{\text{max}} \) through \( \hat{\sigma}^2 \) and that different criteria may yield different choices of optimal number of factors \( R^* \). Therefore we choose the number of factors that have the majority recommendations from these four criteria and three choices of \( R_{\text{max}} \). Where there is a tie, we use the larger number of factors. For example, in Model 4 the optimal number of factors is 1 for all four criteria when \( R_{\text{max}} = 8 \), both \( P_{C_1} \) and \( P_{C_2} \) suggest \( R^* \) to be 7 and both \( I_{C_1} \) and \( I_{C_2} \) suggest 1 when \( R_{\text{max}} = 10 \), \( P_{C_1} \) and \( P_{C_2} \) suggest 5, and \( I_{C_1} \) and \( I_{C_2} \) suggest 1 when \( R_{\text{max}} = 15 \). So our choice of \( R^* \) will be 1 for Model 4.

Table 5 presents the number of factors determined for each model by using the above procedure and the bootstrap p-values for our linearity test for different values of \( c \). We use 1000 bootstrap resamples. The number of chosen factors is either 1 or 2 and the bootstrap p-values are very small in almost all cases. The latter suggests that the relation between a country’s economic growth rate and its lagged values is nonlinear, and that the relation between a country’s economic growth rate and its initial economic condition as well as its investment share may be nonlinear too.

As a robustness check, we conduct the same analysis using different sample periods. Table 6 presents the bootstrap p-values for our linearity test for different values of \( c \) for the sample period from 1950 to 2009 with \( N = 52 \), Table 7 presents the results for the period 1970-2009 with \( N = 147 \), and Table 8 for the period 1980-2009 with \( N = 148 \). In each scenario, the number of bootstrap resamples is 1000. The bootstrap p-values are very small in most cases in Tables 6-8, except Models 4, 9, 10 and 12 in Table 6, and Models 7 and 8 in Table 7. In these cases, we are not able to reject the null of linearity at the 5% level for all three choices of bandwidth. In addition, when \( N = 52 \) is small in Table 6, using Bai and Ng’s method tends to choose a larger number of factors than when \( N \) is large. For Table 8 in all cases, the bootstrap p-value is 0.000, indicating strong rejection of the null of linearity.

---

5 Note that Bai and Ng (2002) study the determination of number of factors in purely approximating factor models. Following Moon and Weidner (2010) their method can be extended to linear dynamic panel data models with interactive fixed effects. Such an extension is also possible under the local alternative considered in this paper. To conserve space we do not report the details.
Table 5: Bootstrap p-values for the application to economic growth data (1960-2009, N=104)

<table>
<thead>
<tr>
<th>Number of factors</th>
<th>$c = 0.5$</th>
<th>$c = 1$</th>
<th>$c = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>2</td>
<td>0.004</td>
<td>0.004</td>
</tr>
<tr>
<td>Model 2</td>
<td>1</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Model 3</td>
<td>1</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Model 4</td>
<td>1</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Model 5</td>
<td>2</td>
<td>0.023</td>
<td>0.024</td>
</tr>
<tr>
<td>Model 6</td>
<td>1</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Model 7</td>
<td>1</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Model 8</td>
<td>1</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Model 9</td>
<td>2</td>
<td>0.050</td>
<td>0.047</td>
</tr>
<tr>
<td>Model 10</td>
<td>1</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Model 11</td>
<td>1</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Model 12</td>
<td>1</td>
<td>0.001</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Note: The numbers in the main entries are the $p$-values based on 1000 bootstrap resamples.

Table 6: Bootstrap p-values for the application to economic growth data (1950-2009, N=52)

<table>
<thead>
<tr>
<th>Number of factors</th>
<th>$c = 0.5$</th>
<th>$c = 1$</th>
<th>$c = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>3</td>
<td>0.022</td>
<td>0.016</td>
</tr>
<tr>
<td>Model 2</td>
<td>3</td>
<td>0.030</td>
<td>0.015</td>
</tr>
<tr>
<td>Model 3</td>
<td>3</td>
<td>0.035</td>
<td>0.037</td>
</tr>
<tr>
<td>Model 4</td>
<td>3</td>
<td>0.187</td>
<td>0.106</td>
</tr>
<tr>
<td>Model 5</td>
<td>2</td>
<td>0.025</td>
<td>0.024</td>
</tr>
<tr>
<td>Model 6</td>
<td>1</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Model 7</td>
<td>3</td>
<td>0.006</td>
<td>0.016</td>
</tr>
<tr>
<td>Model 8</td>
<td>1</td>
<td>0.027</td>
<td>0.001</td>
</tr>
<tr>
<td>Model 9</td>
<td>4</td>
<td>0.019</td>
<td>0.052</td>
</tr>
<tr>
<td>Model 10</td>
<td>3</td>
<td>0.207</td>
<td>0.015</td>
</tr>
<tr>
<td>Model 11</td>
<td>3</td>
<td>0.066</td>
<td>0.027</td>
</tr>
<tr>
<td>Model 12</td>
<td>3</td>
<td>0.138</td>
<td>0.055</td>
</tr>
</tbody>
</table>

Table 7: Bootstrap p-values for the application to economic growth data (1970-2009, N=147)

<table>
<thead>
<tr>
<th>Number of factors</th>
<th>$c = 0.5$</th>
<th>$c = 1$</th>
<th>$c = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>1</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Model 2</td>
<td>2</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Model 3</td>
<td>2</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Model 4</td>
<td>2</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Model 5</td>
<td>2</td>
<td>0.024</td>
<td>0.025</td>
</tr>
<tr>
<td>Model 6</td>
<td>2</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Model 7</td>
<td>2</td>
<td>0.197</td>
<td>0.202</td>
</tr>
<tr>
<td>Model 8</td>
<td>2</td>
<td>0.208</td>
<td>0.181</td>
</tr>
<tr>
<td>Model 9</td>
<td>2</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Model 10</td>
<td>2</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Model 11</td>
<td>2</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Model 12</td>
<td>2</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>
Table 8: Bootstrap p-values for the application to economic growth data (1980-2009, N=148)

<table>
<thead>
<tr>
<th>Model</th>
<th>Number of factors</th>
<th>$c = 0.5$</th>
<th>$c = 1$</th>
<th>$c = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>1</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Model 2</td>
<td>1</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Model 3</td>
<td>1</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Model 4</td>
<td>1</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Model 5</td>
<td>1</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Model 6</td>
<td>1</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Model 7</td>
<td>1</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Model 8</td>
<td>1</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Model 9</td>
<td>1</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Model 10</td>
<td>1</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Model 11</td>
<td>1</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Model 12</td>
<td>1</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

5 Concluding remarks

In this paper we propose a nonparametric consistent test for the correct specification of linear panel data models with interactive fixed effects. After we estimate the model under the null hypothesis of linearity, we obtain the residuals which are then used to construct our test statistic. We show that our test is asymptotically normally distributed under the null hypothesis and a sequence of Pitman local alternatives and propose a bootstrap procedure to obtain the bootstrap $p$-value. Simulations suggest that our bootstrap-based test works well in finite samples. We illustrate our method by applying it to an economic growth dataset. We find significant nonlinear relationship in the dataset.
APPENDIX

Let $C$ signify a generic constant whose exact value may vary from case to case. Let $|a|$ denote the integer part of real number $a$. Let $\delta_{NT} \equiv \min(\sqrt{N}, \sqrt{T})$. Let $E_D(\cdot)$ and $\text{Var}_D(\cdot)$ denote the conditional expectation and variance given $D = \{F_0, \lambda^0\}$, respectively. Let $\alpha_{k} \equiv \lambda_{k}^0 (\lambda^0 \lambda_k^0/N)^{-1} X_{k}^0$ and $\eta_{k} \equiv F_k^0 (F_0^0 F_0^0/T)^{-1} F_k^0$. Let $\Phi_1 \equiv \lambda^0 (\lambda^0 \lambda^0)^{-1} (F_0^0 F_0^0)^{-1} F_0^0$, $\Phi_2 \equiv F_0^0 (F_0^0 F_0^0)^{-1} (\lambda^0 \lambda^0)^{-1} (F_0^0 F_0^0)^{-1} F_0^0$, and $\Phi_3 \equiv \lambda^0 (\lambda^0 \lambda^0)^{-1} (F_0^0 F_0^0)^{-1} (\lambda^0 \lambda^0)^{-1} F_0^0$.

A Proof of Theorem 3.1

The proof follows closely from the proofs of Theorems 2.1 and 3.1 in Moon and Weidner (2010, MW hereafter). So we only outline the difference. By allowing local deviations from the linear panel data models, the consistency of $\hat{\beta}$ can be demonstrated as in MW. Let $X_0 \equiv (\sqrt{NT}/\|e\|) e$, $c_0 \equiv \|e\|/\sqrt{NT}$, and $\epsilon_k \equiv \beta^0_k - \beta_k$ for $k = 1, \ldots, p$. Note that under $H_1(\gamma_{NT})$, conditions (A.6) and (A.7) in MW continue to hold for sufficiently large $(N, T)$ as

$$v_{1NT} \equiv \frac{\sum_{k=1}^{p} (\beta^0_k - \beta_k) \|X_k\| + \|e\|}{\sqrt{v_{1NT}}} = o_P(1) + O_P(\delta_{NT}^{-1} + \gamma_{NT}) = o_P(1),$$

under Assumptions A.1(iii) and (iv) provided $\|\beta^0 - \beta\| = o(1)$. This enables us to apply Lemma A.1(iii) of MW to obtain

$$L_{NT}(\beta) = \frac{1}{N T} \sum_{k_1=0}^{p} \sum_{k_2=0}^{p} \epsilon_{k_1} \epsilon_{k_2} L^{(2)}(\lambda^0, F_0^0, X_{k_1}, X_{k_2})$$

$$+ \frac{1}{N T} \sum_{k_1=0}^{p} \sum_{k_2=0}^{p} \sum_{k_3=0}^{p} \epsilon_{k_1} \epsilon_{k_2} \epsilon_{k_3} L^{(3)}(\lambda^0, F_0^0, X_{k_1}, X_{k_2}, X_{k_3}) + O_P((v_{1NT})^4),$$

where for any integer $g \geq 1$,

$$L^{(g)}(\lambda^0, F_0^0, X_{k_1}, \ldots, X_{k_g}) = \frac{1}{g!} \sum_{\text{all } \text{permutations of } (k_1, \ldots, k_g)} \bar{L}^{(g)}(\lambda^0, F_0^0, X_{k_1}, \ldots, X_{k_g}),$$

$$\bar{L}^{(g)}(\lambda^0, F_0^0, X_{k_1}, \ldots, X_{k_g}) = \sum_{l=1}^{g} (-1)^{l+1} \sum_{v_1 + \cdots + v_{g} = g} \sum_{m_1 + \cdots + m_{l-1} = l-1} \sum_{2 \leq v_{l} > 1, m_{j} \geq 0} \text{tr}\left\{S^{(m_{1})} T_{k_1}^{(v_{1})} S^{(m_{2})} \cdots S^{(m_{l})} T_{k_g}^{(v_{l})} S^{(m_{l+1})}\right\},$$

$$S^{(0)} = -M_{\lambda^0}, S^{(m)} = \Phi_{3}^{m}, T_{k}^{(1)} = \lambda^0 F_0^0 X_{k} + X_{k} F_0^0 \lambda^0$$

for $k = 0, 1, \ldots, p$, and $T_{k_1 k_2}^{(2)} = X_{k_1} X_{k_2}'$ for $k_1, k_2 = 0, 1, \ldots, p$. By straightforward calculations, one verifies that

$$\bar{L}^{(2)}(\lambda^0, F_0^0, X_{k_1}, X_{k_2})$$

$$= \text{tr}\left\{S^{(0)} T_{k_1}^{(2)} T_{k_2}^{(0)} - \left[S^{(1)} T_{k_1}^{(1)} T_{k_2}^{(1)} S^{(0)} + S^{(0)} T_{k_1}^{(1)} T_{k_2}^{(1)} S^{(1)} + S^{(0)} T_{k_1}^{(1)} S^{(0)} T_{k_2}^{(1)} S^{(1)}\right]\right\}$$

$$= \text{tr}\left(M_{\lambda^0} X_{k_1} X_{k_2}' M_{\lambda^0} - M_{\lambda^0} X_{k_1} F_0^0 \lambda^0 \Phi_{3} \lambda^0 F_0^0 X_{k_2}' M_{\lambda^0}\right)$$

$$= \text{tr}\left(M_{\lambda^0} X_{k_1} M_{F_0^0} X_{k_2}' M_{\lambda^0}\right).$$

The subscript indices in $T_{k_1}^{(v_{1})}$ or $T_{k_2}^{(v_{2})}$ may contain either one (e.g., $k_1$ or $k_2$) or two elements (e.g., $(k_1, k_2)$ or $(k_{g-1}, k_g)$) depending on whether $v_{1}$ or $v_{2}$ takes value 1 or 2.
where we use the fact that $S^{(0)}T^{(1)}S^{(0)} = 0$ and $F^{0} \lambda^{0} \Phi^{3} \lambda^{0} F^{0} = P_{F^{0}}$. Similarly,

$$
\bar{L}^{(3)}(\lambda^{0}, F^{0}, X_{k_{1}}, X_{k_{2}}, X_{k_{3}}) = \text{tr}\{ -\{S^{(1)}T^{(1)}S^{(0)}T^{(2)}S^{(0)} + S^{(0)}T^{(1)}S^{(1)}T^{(2)}S^{(0)} + S^{(0)}T^{(2)}T^{(2)}S^{(1)}T^{(1)}S^{(0)} + S^{(0)}T^{(2)}S^{(0)}T^{(1)}S^{(1)}T^{(1)}S^{(0)} + S^{(0)}T^{(1)}T^{(1)}T^{(1)}S^{(0)}S^{(0)}T^{(1)}S^{(1)} \}
+ \{S^{(1)}T^{(1)}S^{(0)}T^{(2)}S^{(3)}T^{(1)}S^{(0)} + S^{(0)}T^{(1)}S^{(1)}T^{(2)}S^{(1)}T^{(1)}S^{(1)}T^{(1)}S^{(1)}T^{(1)}S^{(1)} \}\}
$$

$$
= -\text{tr}\{ \Phi^{3} M_{\lambda^{0}} X_{k_{1}} \Phi^{1} X_{k_{2}} X_{k_{3}} M_{F^{0}} X_{k_{3}'} \} + M_{\lambda^{0}} X_{k_{1}} M_{F^{0}} X_{k_{2}} \Phi_{1} X_{k_{3}} \Phi_{1} X_{k_{3}} \}.
$$

where we use the additional fact that $\Phi^{3} T^{(1)} M_{\lambda^{0}} = \Phi_{1} X_{k} M_{\lambda^{0}}, M_{\lambda^{0}} T^{(1)} \Phi_{3} = M_{\lambda^{0}} X_{k} \Phi^{1}, M_{\lambda^{0}} T^{(1)} \Phi_{1} = M_{\lambda^{0}} X_{k} P_{F^{0}}, \Phi_{1} T^{(1)} M_{\lambda^{0}} = P_{F^{0}} X_{k} M_{\lambda^{0}}, \Phi^{1} T^{(1)} \Phi_{1} = P_{F^{0}} X_{k} \Phi_{1} + \Phi_{1} X_{k} P_{F^{0}},$ and $M_{\lambda^{0}} \Phi_{1} = 0$. It follows that

$$
L^{(2)}(\lambda^{0}, F^{0}, X_{k_{1}}, X_{k_{2}}, X_{k_{3}}) = \text{tr}\{ \Phi^{3} M_{\lambda^{0}} X_{k_{1}} M_{F^{0}} X_{k_{2}} M_{\lambda^{0}} \} = \text{tr}\{ \Phi^{1} X_{k} X_{k_{2}} M_{\lambda^{0}} \}, \quad \text{and}
$$

$$
L^{(3)}(\lambda^{0}, F^{0}, X_{k_{1}}, X_{k_{2}}, X_{k_{3}}) = -\frac{1}{3} \sum_{\text{all 6 permutations of } (k_{1}, k_{2}, k_{3})} \text{tr}\{ \Phi^{1} X_{k} X_{k_{2}} M_{F^{0}} X_{k_{3}} \}.
$$

Furthermore, we have

$$
\mathcal{L}_{NT}(\beta) = \mathcal{L}_{NT}(\beta^{0}) + L_{1NT}(\beta) + L_{2NT}(\beta) + R_{NT} + O_{P}(v_{1NT}^{2}) - O_{P}(c_{0}^{2})
$$

where

$$
L_{1NT}(\beta) = \frac{2}{NT} \sum_{k=1}^{p} \epsilon_{k} \epsilon_{0} L^{(2)}(\lambda^{0}, F^{0}, X_{k}, X_{0}) + \frac{3}{NT} \sum_{k=1}^{p} \epsilon_{k} \epsilon_{0} L^{(3)}(\lambda^{0}, F^{0}, X_{k}, X_{0}, X_{0}),
$$

$$
L_{2NT}(\beta) = \frac{1}{NT} \sum_{k_{1}=1}^{p} \sum_{k_{2}=1}^{p} \epsilon_{k_{1}} \epsilon_{k_{2}} L^{(2)}(\lambda^{0}, F^{0}, X_{k_{1}}, X_{k_{2}}),
$$

$$
R_{NT}(\beta) = \frac{3}{NT} \sum_{k_{1}=1}^{p} \sum_{k_{2}=1}^{p} \sum_{k_{3}=1}^{p} \epsilon_{k_{1}} \epsilon_{k_{2}} \epsilon_{0} L^{(3)}(\lambda^{0}, F^{0}, X_{k_{1}}, X_{k_{2}}, X_{0})
+ \frac{1}{NT} \sum_{k_{1}=1}^{p} \sum_{k_{2}=1}^{p} \sum_{k_{3}=1}^{p} \sum_{k_{4}=1}^{p} \epsilon_{k_{1}} \epsilon_{k_{2}} \epsilon_{k_{3}} \epsilon_{k_{4}} L^{(3)}(\lambda^{0}, F^{0}, X_{k_{1}}, X_{k_{2}}, X_{k_{3}}).\quad \text{N}
$$

Clearly, $L_{1NT}$ and $L_{2NT}$ are linear and quadratic in $\epsilon_{k} = \beta^{0} - \beta_{k}, k = 1, ..., p$, respectively, and $R_{NT}$ reflects the terms in the third order likelihood expansion that are asymptotically negligible (argued below). Noting that $L^{(g)}(\lambda^{0}, F^{0}, X_{k_{1}}, X_{k_{2}}, ..., X_{k_{g}})$ is linear in the last $g$ elements and $\epsilon_{0} X_{0} = e$, we have

$$
L_{1NT}(\beta) = \frac{2}{NT} \sum_{k=1}^{p} \epsilon_{k} \left[ L^{(2)}(\lambda^{0}, F^{0}, X_{k}, e) + \frac{3}{2} L^{(3)}(\lambda^{0}, F^{0}, X_{k}, e, e) \right]
$$

$$
= \frac{2}{NT} \sum_{k=1}^{p} \epsilon_{k} \left[ \text{tr}\{ \Phi^{1} X_{k} M_{F^{0}} e' \} - \frac{1}{2} \sum_{\text{all 6 permutations of } (X_{k}, e, e)} \text{tr}\{ \Phi^{1} X_{k} M_{F^{0}} e' \Phi_{1} e' \} \right]
$$

$$
= -2\gamma_{NT}(\beta - \beta^{0})'(C_{NT}^{(1)} + C_{NT}^{(2)})
$$

where the $p \times 1$ vectors $C_{NT}^{(1)}$ and $C_{NT}^{(2)}$ are defined in (3.1) and (3.2), respectively. Next,

$$
L_{2NT}(\beta) = \frac{1}{NT} \sum_{k_{1}=1}^{p} \sum_{k_{2}=1}^{p} \epsilon_{k_{1}} \epsilon_{k_{2}} \text{tr}\{ \Phi^{1} X_{k_{1}} M_{F^{0}} X_{k_{2}}' \} = (\beta - \beta^{0})' D_{NT} (\beta - \beta^{0}),
$$

26
where $D_{NT}$ is defined in (3.3). As in MW, noticing that
\[
\frac{1}{NT} (\epsilon_0)^{g-r} L^{(g)} (\lambda^0, F^0, X_k, ..., X_k, X_0) = O_P \left( \left( \|e\| / \sqrt{NT} \right)^{g-r} \right) = O_P \left( (\delta^{-1}_{NT} + \gamma_{NT})^{g-r} \right),
\]
we can readily determine the probability order of $R_{NT}$ as $R_{NT} = O_P \left( \|\beta - \beta^0\|^2 (\delta^{-1}_{NT} + \gamma_{NT}) + \|\beta - \beta^0\|^3 \right)$. It follows that
\[
L_{NT} (\beta) = L_{NT} (\beta^0) - 2 \gamma_{NT} (\beta - \beta^0)' (C^{(1)}_{NT} + C^{(2)}_{NT}) + (\beta - \beta^0)' D_{NT} (\beta - \beta^0) + \tilde{R}_{NT} (\beta),
\]
where
\[
\tilde{R}_{NT} (\beta) = O_P \left\{ \|\beta - \beta^0\|^2 (\delta^{-1}_{NT} + \gamma_{NT}) + \|\beta - \beta^0\|^3 + \|\beta - \beta^0\| (\delta^{-3}_{NT} + \gamma_{NT}) \right\}.
\]
Under Assumptions A.1-A.3, we can readily show that $D_{NT} = D + o_P (1)$, $C^{(1)}_{NT} = O_P (1)$, and $C^{(2)}_{NT} = O_P (\delta^{-2}_{NT}/\gamma_{NT})$. Let $\vartheta_{NT} \equiv \gamma_{NT} D_{NT}^{-1} (C^{(1)}_{NT} + C^{(2)}_{NT})$. In view of the fact that $L_{NT}(\hat{\beta}) \leq L_{NT}(\beta^0 + \vartheta_{NT})$, we apply (A.1) to the objects on both sides to obtain
\[
(\beta - \beta^0 - \vartheta_{NT})' D_{NT} (\beta - \beta^0 - \vartheta_{NT}) \leq \tilde{R}_{NT} (\beta + \vartheta_{NT}) - \tilde{R}_{NT} (\hat{\beta}) = O_P \left[ \gamma_{NT}^2 (\delta^{-1}_{NT} + \gamma_{NT}) + \gamma_{NT} \delta^{-3}_{NT} \right] - \tilde{R}_{NT} (\hat{\beta})
\]
where the last line follows from (A.2) and Assumption A.3. Consequently,
\[
\beta - \beta^0 = \vartheta_{NT} + O_P \left\{ \left[ \gamma_{NT} (\delta^{-1}_{NT} + \gamma_{NT}) + \gamma_{NT} \delta^{-3}_{NT} \right]^{1/2} \right\},
\]
and the result follows.

**Remark.** Noting that under $H_1 (\gamma_{NT})$,
\[
C^{(1)}_{NT,k} = \frac{1}{NT} \text{tr} \left( M_{\lambda} X_k M_{F^0} \Delta' \right) + \frac{1}{NT \gamma_{NT}} \text{tr} \left( M_{\lambda} X_k M_{F^0} e' \right)
\]
\[
= (NT)^{-1} \text{tr} \left( M_{\lambda} X_k M_{F^0} \Delta' \right) + O_P \left( \delta^{-2}_{NT} \gamma^{-1}_{NT} \right)
\]
\[
= (NT)^{-1} \text{tr} \left( M_{\lambda} X_k M_{F^0} \Delta' \right) + O_P (1) = O_P (1),
\]
and similarly $C^{(2)}_{NT,k} = O_P (\delta^{-2}_{NT} / \gamma_{NT} + \gamma_{NT}) = O_P (\delta^{-2}_{NT} / \gamma_{NT}) = O_P (1)$, we have
\[
\hat{\beta} - \beta^0 = \gamma_{NT} D_{NT}^{-1} P_{NT} + O_P (\delta^{-2}_{NT})
\]
(A.3)
where $P_{NT}$ is defined in (3.5). If $\gamma_{NT} = N^{-1/2} T^{-1/2} (h!)^{-1/4}$, then $\delta^{-2}_{NT} = o(\gamma_{NT})$ under Assumption A.7(i). This means that $C^{(2)}_{NT}$ and the second term in $C^{(1)}_{NT}$ are asymptotically smaller than the first term in $C^{(1)}_{NT}$.

**B** Proof of Theorem 3.2

Following MW, we can readily show that
\[
M^\varphi = M_{F^0} + \sum_{k=1}^{P} (\beta_k^0 - \hat{\beta}_k) M_k^{(0)} + M^{(1)} + M^{(2)} + M^{(rem)},
\]
(B.1)
where
\[ M_k^{(0)} = -MF^0 X_k \Phi_1 - \Phi_1' X_k M_{F^0} \text{ for } k = 1, \ldots, p, \]
\[ M^{(1)} = -MF^0 \varepsilon \Phi_1 - \Phi_1' \varepsilon M_{F^0}, \]
\[ M^{(2)} = MF^0 \varepsilon \Phi_1 + \Phi_1' \varepsilon \Phi_1 M_{F^0} - MF^0 \varepsilon \Phi_2 - \Phi_2 \varepsilon \Phi_1 M_{F^0} - MF^0 \varepsilon \Phi_3 + \Phi_1' \varepsilon M_{F^0} \varepsilon \Phi_1, \]
and the remainder \( M^{(rem)} \) satisfies
\[
\begin{align*}
\left\| M^{(rem)} \right\|_F &= OP \left( \left\| \delta_{NT}^{-1} + \gamma_{NT} + \left\| \hat{\beta} - \beta^0 \right\| \right\| \left\| \hat{\beta} - \beta^0 \right\| + (NT)^{-3/2} \max \left( \sqrt{N}, \sqrt{T} \right)^3 + \gamma_{NT}^3 \right) \\
&= OP \left( \delta_{NT}^{-1} \gamma_{NT} + \delta_{NT}^{-3} \right) = OP \left( \delta_{NT}^{-2} \gamma_{NT} \right) \text{ under Assumption A.7 (i)}. \quad (B.2)
\end{align*}
\]
It is straightforward to show that
\[
\begin{align*}
\left\| M_k^{(0)} \right\|_F &= OP \left( 1 \right) \text{ for } k = 1, \ldots, p, \quad \left\| M^{(1)} \right\|_F = OP \left( N^{-1/2} \right), \quad \text{and} \quad \left\| M^{(2)} \right\|_F = OP \left( \delta_{NT}^{-2} \right). \quad (B.3)
\end{align*}
\]
Combining (B.1) with (2.14) yields
\[
\begin{align*}
\hat{\varepsilon}_i &= MF^0 (\varepsilon_i + c_i) + \sum_{k=1}^{p} \left( \beta_k^0 - \hat{\beta}_k \right) M_k^{(0)} (\varepsilon_i + F^0 \lambda_i^0 + c_i) + M^{(1)} (\varepsilon_i + F^0 \lambda_i^0 + c_i) \\
&+ M^{(2)} + M^{(rem)} (\varepsilon_i + F^0 \lambda_i^0 + c_i) \\
&= d_{1i} + d_{2i} + d_{3i} + d_{4i}, \text{ say,} \quad (B.4)
\end{align*}
\]
where \( c_i \equiv X_i (\beta^0 - \hat{\beta}) + (m_i - X_i \beta^0) \). It follows that
\[
\begin{align*}
NT \left( \Delta \right)^{1/2} \hat{\Delta}_{NT} &= a_{NT} \sum_{1 \leq i, j \leq N} \left( d_{1i} + d_{2i} + d_{3i} + d_{4i} \right) \hat{K}_{ij} (d_{1j} + d_{2j} + d_{3j} + d_{4j}) \\
&= a_{NT} \sum_{1 \leq i, j \leq N} \left\{ d_{1i} \hat{K}_{ij} d_{1j} + d_{2i} \hat{K}_{ij} d_{2j} + d_{3i} \hat{K}_{ij} d_{3j} + d_{4i} \hat{K}_{ij} d_{4j} + 2d_{1i} \hat{K}_{ij} d_{2j} \\
&+ 2d_{1i} \hat{K}_{ij} d_{3j} + 2d_{1i} \hat{K}_{ij} d_{4j} + 2d_{2i} \hat{K}_{ij} d_{3j} + 2d_{2i} \hat{K}_{ij} d_{4j} + 2d_{3i} \hat{K}_{ij} d_{4j} \right\} \\
&= A_1 + A_2 + A_3 + A_4 + A_5 + A_6 + A_7 + 2A_8 + 2A_9 + 2A_{10}, \text{ say,}
\end{align*}
\]
where \( a_{NT} \equiv \left( \Delta \right)^{1/2} / (NT) \). We complete the proof by showing that under \( \mathbb{P}_1 (\gamma_{NT}) \), (i) \( A_1 \equiv A_1 - B_{1,NT} - B_{2,NT} - \frac{D}{N} (0, V_0), \) (ii) \( A_2 = B_{2,2NT} + o_P \left( 1 \right) \), (iii) \( A_5 = B_{2,3NT} + o_P \left( 1 \right) \), and (iv) \( A_s = o_P \left( 1 \right) \) for \( s = 3, 4, 6, \ldots, 10 \), where \( B_{1,NT} \) is defined in (3.7),
\[
\begin{align*}
B_{1,NT} &= (NT)^{-2} \sum_{1 \leq i, j \leq N} \left( \Delta_i - X_i D_{NT}^{-1} \Pi_{NT} \right) \hat{K}_{ij} M_{F^0} \hat{K}_{ij} M_{F^0} \left( \Delta_j - X_j D_{NT}^{-1} \Pi_{NT} \right), \\
B_{2,2,NT} &= (NT)^{-2} \sum_{1 \leq i, j \leq N} \left( D_{NT}^{-1} \Pi_{NT} \right) \hat{X}_{ij} \hat{X}_{ij} \left( D_{NT}^{-1} \Pi_{NT} \right) + o_P \left( 1 \right), \\
B_{2,3,NT} &= (NT)^{-2} \sum_{1 \leq i, j \leq N} \left( \Delta_i - X_i D_{NT}^{-1} \Pi_{NT} \right) \hat{K}_{ij} \hat{X}_{ij} \left( D_{NT}^{-1} \Pi_{NT} \right),
\end{align*}
\]
and \( \hat{X}_i \equiv N^{-1} \sum_{l=1}^{N} \alpha_i l M_{F^0} X_l \). This is true because in view of the fact that
\[
\begin{align*}
M_{F^0} \Delta_i - M_{F^0} X_i D_{NT}^{-1} \Pi_{NT} + \hat{X}_i D_{NT}^{-1} \Pi_{NT} &= M_{F^0} \Delta_i - \left( M_{F^0} X_i - N^{-1} \sum_{l=1}^{N} \alpha_i l M_{F^0} X_l \right) D_{NT}^{-1} \Pi_{NT} \\
&= M_{F^0} \Delta_i - \hat{X}_i D_{NT}^{-1} \Pi_{NT},
\end{align*}
\]
we have
\[ B_{2,1NT} + B_{2,2NT} + 2B_{2,3NT} = \frac{1}{(NT)^2} \sum_{1 \leq i,j \leq N} \left( M_{F_0} \Delta_i - M_{F_0} X_i D_{NT}^{-1} \Pi_{NT} + \tilde{X}_i D_{NT}^{-1} \Pi_{NT} \right) K_{ij} \]
\[ \times \left( M_{F_0} \Delta_j - M_{F_0} X_j D_{NT}^{-1} \Pi_{NT} + \tilde{X}_j D_{NT}^{-1} \Pi_{NT} \right) \]
\[ = \frac{1}{(NT)^2} \sum_{1 \leq i,j \leq N} \left( M_{F_0} \Delta_i - \tilde{X}_i D_{NT}^{-1} \Pi_{NT} \right) K_{ij} \left( M_{F_0} \Delta_j - \tilde{X}_j D_{NT}^{-1} \Pi_{NT} \right) \]
\[ = B_{2NT}. \]

We prove (i), (ii), and (iii) in Propositions B.1, B.2, and B.5, respectively. (iv) is proved in Propositions B.3, B.4, and B.6-B.10 below.

**Proposition B.1** \( \tilde{A}_1 \overset{D}{\rightarrow} N(0, V_0) \) under \( \mathbb{H}_1(\gamma_{NT}) \).

**Proof.** Noting that \( B_{1NT} = a_{NT} \sum_{i=1}^{N} \varepsilon_i M_{F_0} K_{ii} M_{F_0} \varepsilon_i \), we have
\[ \tilde{A}_1 = a_{NT} \sum_{1 \leq i,j \leq N} (\varepsilon_i + c_i) \varepsilon_j M_{F_0} K_{ij} M_{F_0} (\varepsilon_j + c_j) - B_{1NT} - B_{2,1NT} \]
\[ = a_{NT} \sum_{1 \leq i \neq j \leq N} \varepsilon_i \varepsilon_j M_{F_0} K_{ij} M_{F_0} \varepsilon_j + a_{NT} \sum_{1 \leq i,j \leq N} \varepsilon_i M_{F_0} K_{ij} M_{F_0} c_j - B_{2,1NT} \]
\[ + 2a_{NT} \sum_{1 \leq i,j \leq N} \varepsilon_i M_{F_0} K_{ij} M_{F_0} c_j \]
\[ \equiv A_{1,1} + A_{1,2} + 2A_{1,3}, \text{ say.} \]

We complete the proof by showing that: (i) \( A_{1,1} \overset{D}{\rightarrow} N(0, V_0) \), (ii) \( A_{1,2} = o_P(1) \), and (iii) \( A_{1,3} = o_P(1) \).

First, we show (i). Using \( M_{F_0} = I_T - P_{F_0} \) and the fact that \( K_{ji} = K_{ij} \) we can decompose \( A_{1,1} \) as follows
\[ A_{1,1} = a_{NT} \sum_{1 \leq i \neq j \leq N} \varepsilon_i K_{ij} \varepsilon_j - 2a_{NT} \sum_{1 \leq i \neq j \leq N} \varepsilon_i P_{F_0} K_{ij} \varepsilon_j + a_{NT} \sum_{1 \leq i \neq j \leq N} \varepsilon_i P_{F_0} K_{ij} P_{F_0} \varepsilon_j \]
\[ \equiv A_{1,11} - 2A_{1,12} + A_{1,13}. \]

By Lemmas D.3(i) and (ii), \( A_{1,12} = o_P(1) \) and \( A_{1,13} = o_P(1) \). So we can prove (i) by showing that \( A_{1,11} \overset{D}{\rightarrow} N(0, V_0) \). To achieve this goal, we rewrite \( A_{1,11} \) as follows
\[ A_{1,11} = \frac{(ht)^{1/2}}{NT} \sum_{1 \leq i \neq j \leq N} \varepsilon_i K_{ij} \varepsilon_j = \sum_{1 \leq i \leq j \leq N} W_{ij} \]
where \( W_{ij} \equiv W_{NT}(u_i, u_j) \equiv 2(ht)^{1/2} (NT)^{-1} \sum_{1 \leq t \leq T} K_{ij,t} \varepsilon_i \varepsilon_t \) and \( u_i \equiv (X_i, \varepsilon_i) \). Noting that \( A_{1,11} \) is a second order degenerate U-statistic that is "clean" \( E_{D}[W_{NT}(u, u)] = E_{D}[W_{NT}(u, u)] = 0 \) a.s. for each nonrandom \( u \), we apply Proposition 3.2 in de Jong (1987) to prove the CLT for \( A_{1,11} \) by showing that (i) \( \text{Var}_{D}(A_{1,11}) = V_0 + o_P(1) \), (ii) \( G_{I} \equiv \sum_{1 \leq i < j \leq N} E_{D}(W_{ij}^2) = O_{P}(1) \), and (iii) \( G_{II} \equiv \sum_{1 \leq i < j < k \leq N} E_{D}(W_{ij} W_{ik} W_{jk} + W_{ki} W_{jk}^2 + W_{kj} W_{ji}^2) = O_{P}(1) \).
For (i1), noting that $E_D(A_{1,11}) = 0$ by Assumptions A.4(ii)-(iii), by the same assumptions, we have

\[
\text{Var}_D(A_{1,11}) = \frac{4h!}{(NT)^2} \sum_{1 \leq i < j \leq N} \sum_{t=1}^T \sum_{t'=1}^T \sum_{s=1}^T E_D(K_{ij,t,s_1}K_{ij,t,s_2}\varepsilon_{it}\varepsilon_{jt}\varepsilon_{js_1}\varepsilon_{js_2})
\]

\[
= \frac{4h!}{(NT)^2} \sum_{1 \leq i < j \leq N} \sum_{t=1}^T \sum_{s=1}^T E_D(K_{ij,t,s}^2\varepsilon_{it}^2\varepsilon_{js}^2) = V_0 + o_P(1) .
\]

(i2) follows from the Markov inequality and the fact that

\[
E(G_1) = \frac{16(h!)^2}{(NT)^4} \sum_{1 \leq i < j < k \leq N} \sum_{1 \leq t_1, \ldots, t_8 \leq T} E[D(e_{it_1}e_{jt_2}e_{kt_3}e_{kt_4}e_{kt_5}e_{kt_6}K_{ik,t_1,t_3}K_{ik,t_1,t_4}K_{ik,t_1,t_5}K_{ik,t_1,t_6}K_{ik,t_2,t_6}K_{ik,t_2,t_6})]
\]

\[
= \frac{(h!)^2}{(NT)^4} O \left[N^2 \left(T^6 + T^4(h!)^{-2}\right)\right] = O \left[N^{-2}T^2(h!)^2 + N^{-2}\right] = o(1),
\]

where we use the fact that the term inside the last summation takes value 0 if either $\{t_1, t_3, t_5, t_7\} = 4$ or $\{t_2, t_4, t_6, t_8\} = 4$ by Assumptions A.4(ii)-(iii). For (i3), we write $G_{11} = \sum_{1 \leq i < j < k \leq N} E_D(W_{ik}W_{jk}^2 + W_{ik}^2W_{jk} + W_{ik}W_{jk}^2) \equiv G_{11,1} + G_{11,2} + G_{11,3}$. Then

\[
E(G_{11,1}) = \frac{16(h!)^2}{(NT)^4} \sum_{1 \leq i < j < k \leq N} \sum_{1 \leq t_1, \ldots, t_8 \leq T} E[D(e_{it_1}e_{jt_2}e_{kt_3}e_{kt_4}e_{kt_5}e_{kt_6}K_{ik,t_1,t_3}K_{ik,t_1,t_4}K_{ik,t_1,t_5}K_{ik,t_1,t_6}K_{ik,t_2,t_6}K_{ik,t_2,t_6})]
\]

\[
= \frac{(h!)^2}{(NT)^4} O \left[N^3 \left(T^6(h!)^{-1} + T^4(h!)^{-2}\right)\right] = O \left(TN^{-1}h! + N^{-1}\right) = o(1),
\]

where we use the fact that the term inside the last summation takes value 0 if either $\{t_1, t_3, t_5, t_7\} = 4$ by Assumptions A.4(ii)-(iii). It follows that $G_{11,1} = o_P(1)$ by the Markov inequality. Similarly, $G_{11,s} = o_P(1)$ for $s = 2, 3$. Thus we have $G_{11} = o_P(1)$. For (iv), we write $G_{11} = \sum_{1 \leq i < j < k \leq N} E_D(W_{ij}W_{ik}W_{ij}W_{ik}) + E_D(W_{ij}W_{ik}W_{kj}W_{ji}) \equiv G_{11,1} + G_{11,2} + G_{11,3}$. Then by Assumptions A.4(ii)-(iii)

\[
E(G_{11,1}) = \frac{16(h!)^2}{(NT)^4} \sum_{1 \leq i < j < k \leq N} \sum_{1 \leq t_1, \ldots, t_8 \leq T} E[D(e_{it_1}e_{jt_2}e_{kt_3}e_{kt_4}e_{kt_5}e_{kt_6}K_{ik,t_1,t_3}K_{ik,t_1,t_4}K_{ik,t_1,t_5}K_{ik,t_1,t_6}K_{ik,t_2,t_6}K_{ik,t_2,t_6})]
\]

\[
= \frac{(h!)^2}{(NT)^4} O \left[N^4T^4\right] = O \left((h!)^2\right) = o(1).
\]

So $G_{11,1} = o_P(1)$. By the same token, $G_{11,s} = o_P(1)$ for $s = 2, 3$. It follows that $G_{11} = o_P(1)$.

Next we show (ii). Let $\hat{c}_i = \gamma_{NT} (\Delta_i - X_i D_N^{-1} \Pi_{NT} )$. Then by (A.3)

\[
c_i = \gamma_{NT} (\Delta_i - X_i D_N^{-1} \Pi_{NT}) + O_P(\delta_{NT}^2)X_i = \hat{c}_i + O_P(\delta_{NT}^2)X_i.
\]

Noting that $a_{NT} \sum_{1 \leq i,j \leq N} \varepsilon_i^2 M_{ij} \hat{K}_{ij} M_{ij} \hat{c}_j = (NT)^{-2} \sum_{1 \leq i,j \leq N} (\Delta_i - X_i D_N^{-1} \Pi_{NT})' M_{ij} \hat{K}_{ij} M_{ij} \hat{c}_j = (\Delta_i - X_i D_N^{-1} \Pi_{NT})' B_{2,1} \text{NT},$ we have

\[
A_{1,2} = a_{NT} \sum_{1 \leq i,j \leq N} (c_i - \hat{c}_i)' M_{ij} \hat{K}_{ij} M_{ij} (c_j - \hat{c}_j) + 2a_{NT} \sum_{1 \leq i,j \leq N} \hat{c}_i M_{ij} \hat{K}_{ij} M_{ij} (c_j - \hat{c}_j)
\]

\[
= A_{1,21} + 2A_{1,22}, \text{ say.}
\]
Let $c_K \equiv \max_{1 \leq i,j \leq N} \|K_{ij}\|$. Then $c_K = O_P(T)$ by Lemma D.1. By (B.5), the fact that $\sum_{i=1}^N \|X_i\| = O_P(NT^{1/2})$ by the Markov inequality, and the fact that $\|M_{F_0}\| = 1$,

$$|A_{1,21}| \leq a_{NT} \sum_{1 \leq i,j \leq N} \|M_{F_0}\|^2 \|K_{ij}\| \|c_i - \bar{c}_i\| \|c_j - \bar{c}_j\| = a_{NT} c_K O_P(\delta_{NT}^{-4/n}) \sum_{1 \leq i,j \leq N} \|X_i\| \|X_j\| = O_P(a_{NT} \delta_{NT}^{-4/n} T) O_P \left( N^2 T \right) = O_P \left( NT \delta_{NT}^{-4/n} (h!)^{1/2} \right) = o_P(1).$$

Similarly, we can show that $A_{1,22} = o_P(1)$. This completes the proof of (ii).

Now we show (iii). We decompose $A_{1,3}$ as follows

$$A_{1,3} = \gamma_{NT} a_{NT} \sum_{1 \leq i,j \leq N} \varepsilon_i' M_{F_0} K_{ij} M_{F_0} \Delta_j + a_{NT} \sum_{1 \leq i,j \leq N} \varepsilon_i' M_{F_0} K_{ij} M_{F_0} X_j (\beta^0 - \hat{\beta})$$

$$= \gamma_{NT} A_{1,31} + A_{1,32} (\beta^0 - \hat{\beta}), \text{ say.}$$

In view of the fact that $\|\beta^0 - \hat{\beta}\| = O_P(\gamma_{NT})$, we can prove $A_{1,3} = o_P(1)$ by showing that (iii1) $\gamma_{NT} A_{1,31} = o_P(1)$ and (iii2) $\gamma_{NT} A_{1,32} = o_P(1)$. The last two claims are proved in Lemma D.2(i) and (ii), respectively. This completes the proof.

**Proposition B.2** $A_2 = B_{2,2NT} + o_P(1)$ under $H_1(\gamma_{NT})$, where $B_{2,2NT} = (NT)^{-2} \sum_{1 \leq i,j \leq N} (D_{NT}^{-1} \Pi_{NT})' \hat{X}' K_{ij} \hat{X} (D_{NT}^{-1} \Pi_{NT})$.

**Proof.** First, we decompose $A_2$ as follows

$$A_2 = \sum_{k=1}^P \left( \beta_k^0 - \hat{\beta}_k \right) \sum_{l=1}^P \left( \beta_l^0 - \hat{\beta}_l \right) a_{NT} \sum_{1 \leq i,j \leq N} \left( \varepsilon_i' + \lambda_i^0 F^0 \varepsilon_i + c_i \right) M_k^{(0)} K_{ij} M_l^{(0)} (\varepsilon_j + F^0 \lambda_j^0 + c_j)$$

$$= \sum_{k=1}^P \left( \beta_k^0 - \hat{\beta}_k \right) \sum_{l=1}^P \left( \beta_l^0 - \hat{\beta}_l \right) a_{NT} \sum_{1 \leq i,j \leq N} \lambda_i^0 F^0 M_k^{(0)} K_{ij} M_l^{(0)} F^0 \lambda_j^0$$

$$+ \sum_{k=1}^P \left( \beta_k^0 - \hat{\beta}_k \right) \sum_{l=1}^P \left( \beta_l^0 - \hat{\beta}_l \right) a_{NT} \sum_{1 \leq i,j \leq N} \left\{ \varepsilon_i' M_k^{(0)} K_{ij} M_l^{(0)} \varepsilon_j + c_i M_k^{(0)} K_{ij} M_l^{(0)} c_j + 2 \varepsilon_i' M_k^{(0)} K_{ij} M_l^{(0)} F^0 \lambda_j^0 + 2 \varepsilon_i' M_k^{(0)} K_{ij} M_l^{(0)} c_j + 2 \lambda_i^0 F^0 M_k^{(0)} K_{ij} M_l^{(0)} c_j \right\}$$

$$= A_{2,1} + A_{2,2}, \text{ say.}$$

We prove the proposition by showing that (i) $A_{2,1} = B_{2,2NT} + o_P(1)$ and (ii) $A_{2,2} = o_P(1)$. (i) follows
because

\[
A_{2,1} = a_{NT} \sum_{k=1}^{p} \left( \beta_k^{0} - \hat{\beta}_k \right) \sum_{l=1}^{p} \left( \beta_l^{0} - \hat{\beta}_l \right) \sum_{1 \leq i,j \leq N} \lambda_{ij}^{0} F^{0} \Phi_i^0 X_k M F^0 \Phi_j^0 \lambda_j^{0} \\
= a_{NT} \sum_{k=1}^{p} \left( \beta_k^{0} - \hat{\beta}_k \right) \sum_{l=1}^{p} \left( \beta_l^{0} - \hat{\beta}_l \right) \sum_{1 \leq i,j \leq N} \lambda_{ij}^{0} (\lambda^{0} \lambda^{0})^{-1} \lambda_{ij}^{0} X_k M F^0 \Phi_j^0 \lambda_j^{0} (\lambda^{0} \lambda^{0})^{-1} \lambda_j^{0} \\
= \frac{1}{(NT)^2} \sum_{k=1}^{p} \rho_k D_{NT}^{-1} \sum_{l=1}^{p} \rho_l D_{NT}^{-1} \sum_{1 \leq i,j \leq N} \lambda_{ij}^{0} (\lambda^{0} \lambda^{0})^{-1} \lambda_{ij}^{0} X_k M F^0 \Phi_j^0 \lambda_j^{0} (\lambda^{0} \lambda^{0})^{-1} \lambda_j^{0} \\
+ o_P (1) \\
= \frac{1}{(NT)^2} \sum_{1 \leq i,j \leq N} \sum_{k=1}^{p} \rho_k D_{NT}^{-1} \Pi_{NT} X_{k,i} \Phi_j^0 \sum_{l=1}^{p} \rho_l D_{NT}^{-1} \Pi_{NT} \tilde{X}_{l,j} + o_P (1) \\
= \frac{1}{(NT)^2} \sum_{1 \leq i,j \leq N} \left( D_{NT}^{-1} \Pi_{NT} \right)^T \tilde{X}_{i} \Phi_j^0 \tilde{X}_{j} (D_{NT}^{-1} \Pi_{NT}) + o_P (1) = B_{2,2NT} + o_P (1) ,
\]

where \( \rho_k \) is a \( p \times 1 \) vector with 1 in the \( k \)th place and zeros elsewhere, and \( \tilde{X}_i = N^{-1} \sum_{l=1}^{N} \alpha_{il} M F^0 X_l \) is a \( T \times p \) matrix whose \( k \)th column is given by \( \tilde{X}_{k,k} = (\lambda^{0} \lambda^{0})^{-1} \lambda_{i}^{0} X_{k} M F^0 \).

To show (ii), we assume that \( p = 1 \) for notational simplicity. In this case, we can write \( X_k \) and \( \sum_{k=1}^{p} (\beta_k^{0} - \hat{\beta}_k) M (0)^0 \) simply as \( X \) and \( (\beta^0 - \hat{\beta}) M (0)^0 \), respectively, where \( M (0)^0 = -M F^0 X \Phi_1 - \Phi_j^0 X M F^0 \).

Then

\[
A_{2,2} = \left( \beta^0 - \hat{\beta} \right)^2 a_{NT} \sum_{1 \leq i,j \leq N} \left\{ \varepsilon_i M (0)^0 K_{ij} M (0)^0 \varepsilon_j + \varepsilon_i M (0)^0 K_{ij} M (0)^0 \varepsilon_j + 2 \varepsilon_i M (0)^0 K_{ij} M (0)^0 \right\} \\
= \left( \beta^0 - \hat{\beta} \right)^2 \left\{ A_{2,21} + A_{2,22} + 2A_{2,23} + 2A_{2,24} + 2A_{2,25} \right\} ,
\]

noting that \( \| \beta^0 - \hat{\beta} \| = O_P (\gamma_{NT}) \), we prove (ii) by showing that \( \bar{A}_{2,2s} = \gamma_{NT} A_{2,2s} = o_P (1) \) for \( s = 1, 2, \ldots, 5 \).

Noting that \( \| M (0)^0 \| = O_P (1) \) by (B.3) and

\[
\| M (0)^0 \varepsilon_i \| = \|(M F^0 X \Phi_1 + \Phi_j^0 X M F^0) \varepsilon_i \| = O_P \left( T^{-1/2} \right) \| F^0 \varepsilon_i \| + O_P \left( N^{-1/2} T^{-1/2} \right) \| X^T \varepsilon_i \| , \quad (B.6)
\]

we have by Assumptions A.4(iii) and A.5 and Lemma D.1,

\[
| \bar{A}_{2,21} | \leq (NT)^{-2} \sum_{1 \leq i,j \leq N} \| K_{ij} \| \| M (0)^0 \varepsilon_i \| \| M (0)^0 \varepsilon_j \| \\
\leq c_{\bar{K}} O_P \left( N^{-2} T^{-2} \right) \sum_{1 \leq i,j \leq N} \left[ O_P \left( T^{-1/2} \right) \| F^0 \varepsilon_i \| + O_P \left( N^{-1/2} T^{-1/2} \right) \| X^T \varepsilon_i \| \right] \\
\times \left[ O_P \left( T^{-1/2} \right) \| F^0 \varepsilon_j \| + O_P \left( N^{-1/2} T^{-1/2} \right) \| X^T \varepsilon_j \| \right] \\
= TO_P \left( N^{-2} T^{-2} \right) O_P \left( N^2 \right) = O_P \left( T^{-1} \right) = o_P (1) ,
\]

32
\[ |\tilde{A}_{2.23}^t| \leq (NT)^{-2} \sum_{1 \leq i,j \leq N} \| K_{ij} \| \| M^{(0)} \| \| M^{(0)} F^0 \lambda_i^0 \| \]
\[ \leq c \kappa (NT)^{-2} \| M^{(0)} F^0 \| \sum_{1 \leq i,j \leq N} \left[ O_P \left( T^{-1/2} \right) \| F^0 \| + O_P \left( N^{-1/2} T^{-1/2} \right) \| X_{\epsilon_i} \| \right] \| \lambda_i^0 \| \]
\[ = T (NT)^{-2} O_P \left( T^{1/2} \right) O_P \left( N^2 \right) = O_P \left( T^{-1/2} \right) = o_P \left( 1 \right), \]
and
\[ |\tilde{A}_{2.24}^t| \leq (NT)^{-2} \sum_{1 \leq i,j \leq N} \| K_{ij} \| \| M^{(0)} \| \| M^{(0)} c_j \| \]
\[ \leq c \kappa \gamma_{NT} (NT)^{-2} \| M^{(0)} \| \sum_{1 \leq i,j \leq N} \left[ O_P \left( T^{-1/2} \right) \| F^0 \| + O_P \left( N^{-1/2} T^{-1/2} \right) \| X_{\epsilon_i} \| \right] \times \left( \| X_j \| + \| \Delta_j \| \right) \]
\[ = T \gamma_{NT} (NT)^{-2} O_P \left( N^2 T^{-1/2} \right) = O_P \left( T^{-1/2} \gamma_{NT} \right) = o_P \left( 1 \right). \]

In addition, by (B.3) and (B.5)
\[ |\tilde{A}_{2.22}^t| \leq c \kappa O_P (\gamma_{NT}^2) (NT)^{-2} \| M^{(0)} \| \sum_{1 \leq i,j \leq N} \left( \| X_i \| + \| \Delta_i \| \right) \left( \| X_j \| + \| \Delta_j \| \right) \]
\[ = T O_P (\gamma_{NT}^2) (NT)^{-2} O_P \left( 1 \right) O_P \left( N^2 \right) = O_P \left( \gamma_{NT}^2 \right) = o_P \left( 1 \right) \]
and
\[ |\tilde{A}_{2.25}^t| \leq c \kappa \gamma_{NT} (NT)^{-2} \| M^{(0)} \| \sum_{1 \leq i,j \leq N} \| F^0 \lambda_i^0 \| \left( \| X_j \| + \| \Delta_j \| \right) \]
\[ = T \gamma_{NT} (NT)^{-2} O_P \left( 1 \right) O_P \left( N^2 \right) = O_P \left( \gamma_{NT} \right) = o_P \left( 1 \right). \]

**Proposition B.3** \( A_3 = o_P \left( 1 \right) \) under \( H_1 \left( \gamma_{NT} \right) \).

**Proof.** Recall \( M^{(1)} = -M_{F^0 e^t} \Phi_1^t e^t M_{F^0} \) and \( \Phi_1 = \lambda^0 \left( \lambda^0 \lambda^0 \right)^{-1} \left( F_0 F_0 \right)^{-1} F^0 \). Noting that \( \Phi_1 M_{F^0} = 0 \) and \( \mu_1 \left( M_{F^0} \right) = 1 \), we have

\[ \| M^{(1)} \|_{F^t}^2 = \text{tr} \left[ e^t \left( M_{F^0 e^t} \Phi_1 + \Phi_1^t e^{t} M_{F^0} \right) M_{F^0 e^t} e^t \Phi_1 e^{t} M_{F^0} e^t \right] \]
\[ = 2 \text{tr} \left[ e^t \left( \Phi_1^t e^t M_{F^0} e^t \Phi_1 \right) \right] \leq 2 \text{tr} \left[ e^t \left( \Phi_1^t e^t e^t \Phi_1 \right) \right] \]
\[ = 2 \text{tr} \left[ e^t F^0 \left( F^0 F^0 \right)^{-1} \left( \lambda^0 \lambda^0 \right)^{-1} \lambda^0 e^t e^t \lambda^0 \left( \lambda^0 \lambda^0 \right)^{-1} \left( F^0 F^0 \right)^{-1} F^0 e^t \right] \]
\[ = 2 \text{tr} \left[ \left( F^0 F^0 \right)^{-1} \left( \lambda^0 \lambda^0 \right)^{-1} \lambda^0 e^t e^t \lambda^0 \left( \lambda^0 \lambda^0 \right)^{-1} \left( F^0 F^0 \right)^{-1} F^0 e^t \right] \]
\[ \leq 2 \text{tr} \left[ \left( F^0 F^0 \right)^{-1} \left( \lambda^0 \lambda^0 \right)^{-1} \lambda^0 e^t e^t \lambda^0 \left( \lambda^0 \lambda^0 \right)^{-1} \left( F^0 F^0 \right)^{-1} \right] \text{tr} \left( F^0 e^t e^t F^0 \right) \]
\[ \leq 2 \text{tr} \left[ \left( \lambda^0 \lambda^0 \right)^{-1} \left( F^0 F^0 \right)^{-1} \left( F^0 F^0 \right)^{-1} \left( \lambda^0 \lambda^0 \right)^{-1} \left( \lambda^0 \lambda^0 \right)^{-1} \right] \text{tr} \left( \lambda^0 e^t e^t \lambda^0 \right) \text{tr} \left( F^0 e^t e^t F^0 \right) \]
\[ = O_P \left( (NT)^{-2} \right) O_P \left( NT \right) \text{tr} \left( F^0 e^t e^t F^0 \right) = O_P \left( (NT)^{-1} \right) \| F^0 e^t \|_F^2, \]

33
where we have repeatedly used the rotational property of the trace operator, the fact that
\[
\text{tr}(AB) \leq \mu_1(A) \text{tr}(B)
\] (B.7)
for any symmetric matrix A and p.s.d. matrix B (see, e.g., Bernstein, 2005, Proposition 8.4.13), and the fact that
\[
\text{tr}(AB) \leq \text{tr}(A) \text{tr}(B)
\] (B.8)
for any two p.s.d. matrices A and B (see, e.g., Bernstein, 2005, Fact 8.10.7). It follows that
\[
\left\| M^{(1)} \varepsilon_i \right\| = O_P \left( (NT)^{-1/2} \right) \left\| F^{0} \varepsilon_i \right\| \quad (B.9)
\]
By the fact that \( \| M^{(1)} \| = O_P \left( N^{-1/2} \right) \) (see (B.3)) and (A.3),
\[
\left\| M^{(1)} C_i \right\| \leq \left\| M^{(1)} \right\| \left\| C_i \right\| = O_P \left( N^{-1/2} \gamma_{NT} \right) \left( \| X_i \| + \| \Delta_i \| \right).
\] (B.10)
Combining (B.9) and (B.10) yields
\[
\left\| M^{(1)} (\varepsilon_i + c_i) \right\| = O_P \left( (NT)^{-1/2} \right) \left\| F^{0} \varepsilon_i \right\| + O_P \left( N^{-1/2} \gamma_{NT} \right) \left( \| X_i \| + \| \Delta_i \| \right).
\] (B.11)
We will use these results frequently.

Now, we decompose \( A_3 \) as follows.
\[
A_3 = a_{NT} \sum_{1 \leq i, j \leq N} \left( \varepsilon_i + \lambda_i^0 F^{0} + c_i \right) M^{(1)} K_{ij} M^{(1)} \left( \varepsilon_j + F^{0} \lambda_j^0 + c_j \right)
= a_{NT} \sum_{1 \leq i, j \leq N} \left\{ \varepsilon_i^j M^{(1)} K_{ij} M^{(1)} \varepsilon_j + \lambda_i^0 F^{0} M^{(1)} K_{ij} M^{(1)} F^{0} \lambda_j^0 + c_i^j M^{(1)} K_{ij} M^{(1)} c_j + 2 \varepsilon_i^j M^{(1)} F^{0} \lambda_j^0 + 2 \varepsilon_i^j M^{(1)} K_{ij} M^{(1)} c_j + 2 \lambda_i^0 F^{0} M^{(1)} K_{ij} M^{(1)} c_j \right\}
= A_{3,1} + A_{3,2} + A_{3,3} + 2A_{3,4} + 2A_{3,5} + 2A_{3,6}, \text{ say.}
\]
We prove the proposition by demonstrating that \( A_{3,s} = o_P(1) \) for \( s = 1, 2, \ldots, 6 \). By (B.9)-(B.11), (B.3), Assumptions A.4(iii) and A.5, and Lemma D.1, we have
\[
|A_{3,1}| \leq a_{NT} \sum_{1 \leq i, j \leq N} \left\| \kappa_{ij} \right\| \left\| M^{(1)} \varepsilon_i \right\| \left\| M^{(1)} \varepsilon_j \right\|
\leq c_{\kappa} a_{NT} O_P \left( (NT)^{-1} \right) \sum_{1 \leq i, j \leq N} \left\| F^{0} \varepsilon_i \right\| \left\| F^{0} \varepsilon_j \right\|
= T O_P \left( a_{NT} (NT)^{-1} \right) O_P (N^2 T) = O_P \left( (h!)^{1/2} \right) = o_P (1),
\]
\[
|A_{3,2}| \leq a_{NT} \sum_{1 \leq i, j \leq N} \left\| \kappa_{ij} \right\| \left\| M^{(1)} C_i \right\| \left\| M^{(1)} C_j \right\|
\leq c_{\kappa} a_{NT} O_P (N^{-1} \gamma_{NT}^2) \sum_{1 \leq i, j \leq N} \left( \| X_i \| + \| \Delta_i \| \right) \left( \| X_j \| + \| \Delta_j \| \right)
= T O_P (a_{NT} N^{-1} \gamma_{NT}^2) O_P (N^2 T) = O_P (N^{-1}) = o_P (1),
\]
\[
34
\]
\[ |A_{3,5}| \leq a_{NT} \sum_{1 \leq i,j \leq N} \|K_{ij}\| \left\| M^{(1)} e_i \right\| \left\| M^{(1)} c_j \right\| \]
\[ \leq c_K a_{NT} O_P \left( N^{-1/2} \gamma_{NT} (NT)^{-1/2} \right) \sum_{1 \leq i,j \leq N} \|F^0 e_i\| (\|X_j\| + \|\Delta_j\|) \]
\[ = TO_P \left( a_{NT} \gamma_{NT} N^{-1} T^{-1/2} \right) O_P \left( N^2 T \right) = O_P \left( N^{-1/2} (ht)^{1/4} \right) = o_P \left( 1 \right), \]
and
\[ |A_{3,6}| \leq a_{NT} \sum_{1 \leq i,j \leq N} \|K_{ij}\| \left\| M^{(1)} F^0 \lambda_i^0 \right\| \left\| M^{(1)} c_j \right\| \]
\[ \leq c_K a_{NT} O_P \left( N^{-1/2} \gamma_{NT} \right) O_P \left( T^{1/2} N^{-1/2} \right) \sum_{1 \leq i,j \leq N} \|\lambda_i^0\| (\|X_j\| + \|\Delta_j\|) \]
\[ = TO_P \left( a_{NT} \gamma_{NT} T^{1/2} N^{-1} \right) O_P \left( N^2 T^{1/2} \right) = O_P \left( T^{1/2} N^{-1/2} (ht)^{1/4} \right) = o_P \left( 1 \right). \]

By Lemmas D.3(iii)-(iv) and the fact that \( M_{F^0} = I_T - P_{F^0}, \ A_{3,2} = a_{NT} \sum_{1 \leq i,j \leq N} \lambda_i^0 (\lambda^0 \lambda^0)^{-1} \lambda_j^0 \varepsilon M_{F^0} K_{ij} M_{F^0} \varepsilon' (\lambda^0 \lambda^0)^{-1} \lambda_j^0 = o_P \left( 1 \right). \) By Lemma D.4(i), \( A_{3,4} = a_{NT} \sum_{1 \leq i,j \leq N} \varepsilon_i' M^{(1)} K_{ij} M^{(1)} F^0 \lambda_j^0 \)
\[ = o_P \left( 1 \right). \] This completes the proof. \[ \blacksquare \]

**Proposition B.4** \( A_4 = o_P \left( 1 \right) \) under \( \mathbb{H}_1 \left( \gamma_{NT} \right). \)

**Proof.** Noting that \( \left\| M^{(2)} + M^{(rem)} \right\| = O_P \left( \delta_{NT}^{-2} \right) \) by (B.2) and (B.3), we have by Assumption A.5
\[ |A_4| \leq a_{NT} \sum_{1 \leq i,j \leq N} \left( (e_i^0 + \lambda_i^0 F^0 + c_i^0) (M^{(2)} + M^{(rem)})' K_{ij} (M^{(2)} + M^{(rem)}) (e_j + F^0 \lambda_j^0 + c_j) \right) \]
\[ \leq c_K a_{NT} \left\| M^{(2)} + M^{(rem)} \right\|^2 \sum_{1 \leq i,j \leq N} \|e_i + F^0 \lambda_i^0 + c_i\| \|e_j + F^0 \lambda_j^0 + c_j\| \]
\[ = TA_{NT} O_P \left( \delta_{NT}^{-1} \right) O_P \left( N^{2} T \right) = O_P \left( N T \delta_{NT}^{-1} (ht)^{1/2} \right) = o_P \left( 1 \right). \]
\[ \blacksquare \]

**Proposition B.5** \( A_5 = B_{2,3NT} + o_P \left( 1 \right) \) under \( \mathbb{H}_1 \left( \gamma_{NT} \right), \) where \( B_{2,3NT} = (NT)^{-2} \sum_{1 \leq i,j \leq N} (\Delta_i - X_i D_{NT}^{-1} \Pi_{NT})' M_{F^0} K_{ij} X_j (D_{NT}^{-1} \Pi_{NT}) \).

**Proof.** First, we decompose \( A_5 \) as follows
\[ A_5 = \sum_{k=1}^{p} \left( \beta_k^0 - \beta_k \right) a_{NT} \sum_{1 \leq i,j \leq N} (e_i^0 + c_i^0) M_{F^0} K_{ij} M_k^{(0)} (e_j + F^0 \lambda_j^0 + c_j) \]
\[ = \sum_{k=1}^{p} \left( \beta_k^0 - \beta_k \right) a_{NT} \sum_{1 \leq i,j \leq N} c_i^0 M_{F^0} K_{ij} M_k^{(0)} F^0 \lambda_j^0 \]
\[ + \sum_{k=1}^{p} \left( \beta_k^0 - \beta_k \right) a_{NT} \sum_{1 \leq i,j \leq N} \{ (e_i^0 + c_i^0) M_{F^0} K_{ij} M_k^{(0)} (e_j + c_j) + e_i^0 M_{F^0} K_{ij} M_k^{(0)} F^0 \lambda_j^0 \} \]
\[ = A_{5,1} + A_{5,2}, \text{ say.} \]
We prove the proposition by showing that (i) $A_{5.1} = B_{2,3NT} + o_P(1)$, and (ii) $A_{5.2} = o_P(1)$. (i) follows because by (A.3)

\[
A_{5.1} = -a_{NT} \sum_{k=1}^{p} \sum_{1 \leq i, j \leq N} c_k' M_{Fi} K_{ij} M_{Fi} \Phi_i F^0 \lambda_j^0
\]

\[
= -a_{NT} \sum_{k=1}^{p} \sum_{1 \leq i, j \leq N} c_k' M_{Fi} K_{ij} M_{Fi} \Phi_i (\lambda^0 \lambda^0)^{-1} \lambda_j^0
\]

\[
= \frac{1}{(NT)^2} \sum_{k=1}^{p} \left( \sum_{1 \leq i, j \leq N} (\Delta_i - X_i D_{i,j}^{-1} \Pi_{NT})' M_{Fi} K_{ij} M_{Fi} \Phi_i (\lambda^0 \lambda^0)^{-1} \lambda_j^0 + o_P(1) \right)
\]

To show (ii), again we assume that $p = 1$ for notational simplicity. As before, we now write $X_k$ and $\sum_{k=1}^{p} (\hat{\beta}_k^0 - \hat{\beta}_k) M_k^{(0)}$ simply as $X$ and $(\hat{\beta}^0 - \hat{\beta}) M^{(0)}$, respectively. Then

\[
A_{5.2} = \left( \hat{\beta}^0 - \hat{\beta} \right) a_{NT} \sum_{1 \leq i, j \leq N} \{ c_i' M_{Fi} K_{ij} M^{(0)} \epsilon_j + \hat{c}_i' M_{Fi} K_{ij} M^{(0)} \hat{c}_j + c_i' M_{Fi} K_{ij} M^{(0)} \epsilon_j \}
\]

\[
= \left( \hat{\beta}^0 - \hat{\beta} \right) (A_{5.21} + A_{5.22} + A_{5.23} + A_{5.24} + A_{5.25}).
\]

We prove the proposition by showing that $A_{5.2s} = \gamma_{NT} A_{5.2s} = o_P(1)$ for $s = 1, 2, \ldots, 5$. By Lemma D.3(iv), $A_{5.21} = o_P(1)$. By Lemma D.2(iii), $A_{5.25} = o_P(1)$. So we are left to show that $A_{5.2s} = o_P(1)$ for $s = 2, 3, 4$.

For $A_{5.22}$, we have

\[
A_{5.22} = \gamma_{NT} a_{NT} \sum_{1 \leq i, j \leq N} \hat{c}_i' K_{ij} M^{(0)} \hat{c}_j - \gamma_{NT} a_{NT} \sum_{1 \leq i, j \leq N} \hat{c}_i' P_{Fi} K_{ij} M^{(0)} \hat{c}_j \equiv A_{5.22a} - A_{5.22b}.
\]

Using (B.3) and (B.5), Lemma D.1, Assumptions A.4(iii) and A.5, and the fact that $\| P_{Fi} \epsilon_i \| = O_P \left( T^{-1/2} \right) \| F^0 \epsilon_i \|$, we can bound $A_{5.22b}$ directly:

\[
| A_{5.22b} | \leq O_P \left( \gamma_{NT} T^{-1/2} \right) c_k \gamma_{NT} a_{NT} \sum_{1 \leq i, j \leq N} \| F^0 \epsilon_i \| (\| X_j \| + \| \Delta_j \|) = O_P \left( T^{-1/2} \right) = o_P(1).
\]

For $A_{5.22a}$, we can easily show that $A_{5.22a} = A_{5.22a} + o_P(1)$ where $A_{5.22a} = \gamma_{NT} a_{NT} \sum_{1 \leq i \neq j \leq N} \hat{c}_i' K_{ij} M^{(0)} \hat{c}_j$. Noting that

\[
E \left( \sum_{1 \leq i \neq j \leq N} \Delta_j \hat{c}_i' K_{ij} \right)^2 = \sum_{1 \leq i \neq j \leq N} \sum_{1 \leq i \neq j \neq k \leq N} \sum_{1 \leq t, \ldots, t_4 \leq T} E (\epsilon_{1, t, K_{ij, t_1, t_2} K_{ij, t_3, t_4} \hat{c}_j \Delta_j \hat{c}_j \Delta_j \hat{c}_j) = O \left( N^3 T^4 \right),
\]
we have \( \left\| \sum_{1 \leq i \neq j \leq N} \Delta_j \tilde{c}_j \tilde{K}_{ij} \right\|_F = O_p \left( N^{3/2} T^{3/2} \right) \). Similar result holds when \( \Delta_j \) is replaced by \( X_j D_{NT}^{-1} \Pi_{NT} \). Then by Cauchy-Schwarz’s and Minkowski’s inequalities

\[
\begin{align*}
|A_{5,22a}| &= \gamma_{NT} a_{NT} \left\| \text{tr} \left( M^{(0)} \sum_{1 \leq i \neq j \leq N} \Delta_j \tilde{c}_j \tilde{K}_{ij} \right) \right\|_F \\
&\leq \gamma_{NT}^2 a_{NT} \left\| M^{(0)} \right\|_F \left\| \sum_{1 \leq i \neq j \leq N} \left( \Delta_j - X_j D_{NT}^{-1} \Pi_{NT} \right) \tilde{c}_j \tilde{K}_{ij} \right\|_F \\
&= O_p \left( N^{-2} T^{-2} \right) O_p \left( N^{3/2} T^{3/2} \right) = O_p \left( N^{-1/2} T^{-1/2} \right) = o_p \left( 1 \right).
\end{align*}
\]

It follows that \( \tilde{A}_{5,22} = o_p \left( 1 \right) \).

By (B.5), (B.6) and (B.3),

\[
|\tilde{A}_{5,23}| \leq c_N \gamma_{NT}^2 a_{NT} \sum_{1 \leq i, j \leq N} \left( \|X_i\| + \|X_j\| \right) \left[ O_p \left( T^{-1/2} \right) \left\| F^{0p} \| \right\| + O_p \left( N^{-1/2} T^{-1/2} \right) \|X||X_j\| \right]
\]

\[
= T \gamma_{NT}^2 a_{NT} O_p \left( N^{2} \right) = O_p \left( T^{-1/2} \right) = o_p \left( 1 \right),
\]

and

\[
|\tilde{A}_{5,24}| \leq c_N \gamma_{NT}^3 a_{NT} \sum_{1 \leq i, j \leq N} \left( \|X_i\| + \|X_j\| \right) \left( \|X_i\| + \|X_j\| \right)
\]

\[
= T \gamma_{NT}^3 a_{NT} O_p \left( N^{2} \right) = O_p \left( 1 \right).
\]

This completes the proof. \( \blacksquare \)

**Proposition B.6** \( A_6 = o_p \left( 1 \right) \) under \( \mathbb{H}_1 \left( \gamma_{NT} \right) \).

**Proof.** First, we decompose \( A_6 \) as follows

\[
A_6 = a_{NT} \sum_{1 \leq i, j \leq N} \left( \varepsilon_j + c_j \right) M_{F \circ K_{ij}} M^{(1)} \left( \varepsilon_j + F^{0p} \lambda_j + c_j \right)
\]

\[
= a_{NT} \sum_{1 \leq i, j \leq N} \left\{ \varepsilon_j M_{F \circ K_{ij}} M^{(1)} \varepsilon_j + \varepsilon_j M_{F \circ K_{ij}} M^{(1)} F^{0p} \lambda_j + \varepsilon_j M_{F \circ K_{ij}} M^{(1)} c_j \right. \right.

\[
\left. + c_j M_{F \circ K_{ij}} M^{(1)} \varepsilon_j + c_j M_{F \circ K_{ij}} M^{(1)} F^{0p} \lambda_j + c_j M_{F \circ K_{ij}} M^{(1)} c_j \right\}
\]

\[
= A_{6,1} + A_{6,2} + A_{6,3} + A_{6,4} + A_{6,5} + A_{6,6}.
\]

By Lemma D.4(ii), \( A_{6,1} = o_p \left( 1 \right) \). By Lemmas D.3(vi)-(viii), \( A_{6,2} = o_p \left( 1 \right) \) and \( A_{6,3} = o_p \left( 1 \right) \). By Lemma D.2(iv), \( A_{6,5} = o_p \left( 1 \right) \). We finish the proof of the proposition by showing that \( A_{6,s} = o_p \left( 1 \right) \) for \( s = 4, 6 \).

By (B.9)-(B.10), and Lemma D.1

\[
|A_{6,4}| \leq a_{NT} \sum_{1 \leq i, j \leq N} \left\| K_{ij} \right\| \left\| M_{F \circ c_j} \right\| \left\| M^{(1)} \varepsilon_j \right\|
\]

\[
\leq c_k O_p \left( \gamma_{NT} (NT)^{-1/2} \right) a_{NT} \sum_{1 \leq i, j \leq N} \left\| X_i \right\| \left\| \Delta_i \right\| \left\| F^{0p} \varepsilon_j \right\|
\]

\[
= T O_p \left( \gamma_{NT} (NT)^{-1/2} \right) O_p \left( N^2 T \right) = o_p \left( \left( NT \right)^{1/4} \right) = o_p \left( 1 \right).
\]
This completes the proof of the proposition. ■

**Proposition B.7** $A_7 = o_P (1)$ under $H_1 (\gamma_{NT})$.

**Proof.** First we decompose $A_7$ as follows

$$A_7 = a_{NT} \sum_{1 \leq i,j \leq N} (\varepsilon_i^j + c_i^j) M_{F^o}K_{ij} \left( M^{(2)} + M^{(rem)} \right) (\varepsilon_j^0 + F^0 \lambda_j^0 + c_j)$$

$$= a_{NT} \sum_{1 \leq i,j \leq N} \left\{ \varepsilon_i^j h_{F^o}K_{ij} M^{(2)} + M^{(rem)} \right\} (\varepsilon_j^0 + F^0 \lambda_j^0 + c_j) + \varepsilon_i^j M_{F^o}K_{ij} M^{(2)} \varepsilon_j + \varepsilon_i^j M_{F^o}K_{ij} M^{(2)} F^0 \lambda_j^0 + \varepsilon_i^j M_{F^o}K_{ij} M^{(rem)} c_j \right\}$$

$$= A_{7,1} + A_{7,2} + A_{7,3} + A_{7,4} + A_{7,5} + A_{7,6} + A_{7,7}.$$ 

By Lemma D.5(i), $A_{7,2} = o_P (1)$. By Lemma D.4(iii), $A_{7,3} = o_P (1)$. We complete the proof of the proposition by showing that $A_{7,s} = o_P (1)$ for $s = 1, 4, 5, 6, 7$.

By (B.5), (B.2), (B.3), and Lemma D.1

$$|A_{7,1}| \leq c_k O_P \left( \gamma_{NT} a_{NT} \right) \left\| M^{(2)} + M^{(rem)} \right\| \sum_{1 \leq i,j \leq N} \left( \| X_i \| + \| \Delta_i \| \right) \left( \| \varepsilon_j \| + \left\| F^0 \lambda_j^0 \right\| + \| c_j \| \right)$$

$$= T O_P \left( \gamma_{NT} a_{NT} \delta_{NT}^2 \right) O_P \left( N^2 T \right) = O_P \left( \sqrt{NT} \delta_{NT}^2 (h!)^1/2 \right) = o_P (1),$$

$$|A_{7,4}| \leq c_k O_P \left( \gamma_{NT} a_{NT} \right) \sum_{1 \leq i,j \leq N} \left( \| X_i \| + \| \Delta_j \| \right)$$

$$= T O_P \left( \gamma_{NT} a_{NT} \delta_{NT}^2 \right) O_P \left( N^2 T \right) = O_P \left( \sqrt{NT} \delta_{NT}^2 (h!)^1/2 \right) = o_P (1),$$

and

$$|A_{7,7}| \leq c_k a_{NT} O_P \left( \gamma_{NT} \right) \left\| M^{(rem)} \right\| \sum_{1 \leq i,j \leq N} \left( \| X_i \| + \| \Delta_j \| \right)$$

$$= T O_P \left( a_{NT} \delta_{NT}^1 \gamma_{NT} \right) O_P \left( N^2 T \right) = O_P \left( \delta_{NT}^1 \right) = o_P (1).$$

For $A_{7,5}$, we decompose it as follows

$$A_{7,5} = a_{NT} \sum_{1 \leq i,j \leq N} \varepsilon_i^j h_{F^o}K_{ij} M^{(rem)} \varepsilon_j - a_{NT} \sum_{1 \leq i,j \leq N} \varepsilon_i^j h_{F^o}K_{ij} M^{(rem)} \varepsilon_j \equiv A_{7,51} - A_{7,52}.$$
Noting that \( \|P_0 \xi_i\| = O_p(T^{-1/2}) \| F^0 \xi_i\| \), we have by (B.2),
\[
|A_{7,52}| \leq c_k a_{NT} O_p(T^{-1/2}) O_p(\delta_{NT}^{-1}) \sum_{1 \leq i,j \leq N} \| F^0 \xi_i\| \| \varepsilon_j\|
= O_p(a_{NT} T^{1/2} \delta_{NT}^{-1}) O_p(N^2 T) = O_p(N^{1/2} \delta_{NT}^{-1}(h!)^{1/4}) = o_p(1).
\]
For \( A_{7,51} \), we have by (B.2),
\[
|A_{7,51}| = a_{NT} \left\| \text{tr} \left( M^{(\text{rem})} \sum_{1 \leq i,j \leq N} \varepsilon_j \varepsilon_i^i K_{ij} \right) \right\| \leq a_{NT} \left\| M^{(\text{rem})} \right\| \left\| \sum_{1 \leq i,j \leq N} \varepsilon_j \varepsilon_i^i K_{ij} \right\|_F
= a_{NT} O_p(\delta_{NT}^{-1}) O_p(N^{3/2} T^{3/2}) = o_p(1),
\]
where we use the fact that
\[
E \left\| \sum_{1 \leq i,j \leq N} \varepsilon_j \varepsilon_i^i K_{ij} \right\|_F^2 = \sum_{1 \leq i_1, \ldots, i_4 \leq N} \sum_{1 \leq i_1, \ldots, i_4 \leq N} E (\varepsilon_{i_1 i_2 i_3 i_4} \varepsilon_i \varepsilon_{i_1 i_2 i_3 i_4} \varepsilon_{i_1 i_2 i_3 i_4}) = O(N^3 T^3).
\]
It follows that \( A_{7,5} = o_p(1) \).

Now, we decompose \( A_{7,6} \) as follows
\[
A_{7,6} = a_{NT} \sum_{1 \leq i,j \leq N} \varepsilon_i^i K_{ij} M^{(\text{rem})} F^0 \lambda_j^0 - a_{NT} \sum_{1 \leq i,j \leq N} \varepsilon_i^i P_{\alpha i} K_{ij} M^{(\text{rem})} F^0 \lambda_j^0 \equiv A_{7,61} - A_{7,62}.
\]
As in the study of \( A_{7,52} \), we can bound \( A_{7,62} \) by \( o_p(1) \). Similarly, as in the study of \( A_{7,51} \), we have by (B.2) and Chebyshev’s inequality
\[
|A_{7,61}| = a_{NT} \left\| \text{tr} \left( M^{(\text{rem})} F^0 \sum_{1 \leq i,j \leq N} \lambda_j^0 \varepsilon_i^i K_{ij} \right) \right\| \leq a_{NT} \left\| M^{(\text{rem})} F^0 \right\| \left\| \sum_{1 \leq i,j \leq N} \lambda_j^0 \varepsilon_i^i K_{ij} \right\|_F
= a_{NT} O_p(\delta_{NT}^{-1}) O_p(N^{3/2} T) = o_p(1).
\]
It follows that \( A_{7,6} = o_p(1) \).

**Proposition B.8** \( A_8 = o_p(1) \) under \( H_1(\gamma_{NT}) \).

**Proof.** Again, assuming \( p = 1 \), we can decompose \( A_8 \) as follows
\[
A_8 = (\beta^0 - \hat{\beta}) \sum_{1 \leq i,j \leq N} \{(\varepsilon_i^0 + c_i^0) M^{(1)} K_{ij} M^{(0)} (\varepsilon_j^0 + F^0 \lambda_j^0 + c_j)
+ \lambda_j^0 F^0 M^{(1)} K_{ij} M^{(0)} \varepsilon_i^0 + \lambda_i^0 F^0 M^{(1)} K_{ij} M^{(0)} F^0 \lambda_j^0 + \lambda_i^0 F^0 M^{(1)} K_{ij} M^{(0)} c_j \}
= (\beta^0 - \hat{\beta}) (A_{8,1} + A_{8,2} + A_{8,3} + A_{8,4}).
\]
We prove the claim by showing that \( A_{8,s} \equiv \gamma_{NT} A_{8,s} = o_p(1) \) for \( s = 1, 2, 3, 4 \). By Lemma D.3(viii), \( A_{8,2} = o_p(1) \). By Lemma D.2(v), \( A_{8,3} = o_p(1) \).
By (B.11), (B.5), and Lemma D.1, we can readily show that

\[ |\tilde{A}_{8,1}| \leq \gamma_{NT} a_{NT} \left\| M^{(0)} \right\| \sum_{1 \leq i,j \leq N} \left\| \mathcal{K}_{ij} \right\| \left\| M^{(1)} (\varepsilon_i + c_i) \right\| \left\| F^0 \lambda_j^0 + c_j \right\| \]

\leq c_\mathcal{K} \gamma_{NT} a_{NT} \left\| M^{(0)} \right\| \sum_{1 \leq i,j \leq N} \left\| \varepsilon_j + F^0 \lambda_j^0 + c_j \right\|

\times \left\{ O_P \left( (NT)^{-1/2} \right) \left\| F^0 \varepsilon_i \right\| + O_P( N^{-1/2} \gamma_{NT} ) \left( \|X_i\| + \|\Delta_i\| \right) \right\}

= T \gamma_{NT} a_{NT} O_P \left( N^{3/2} T^{1/2} \right) = O_P \left( (h!)^{1/4} \right) = o_P (1),

and

\[ |\tilde{A}_{8,4}| \leq c_\mathcal{K} O_P (\gamma_{NT}^2 a_{NT}) \left\| M^{(1)} F^0 \right\| \sum_{1 \leq i,j \leq N} \left\| \lambda_j^0 \right\| (\|X_i\| + \|\Delta_j\|)

= N^{-2} T^{-1} O_P \left( N^{-1/2} T^{-1/2} \right) O_P \left( N^2 T^{1/2} \right) = O_P \left( N^{-1/2} \right) = o_P (1). \]

This completes the proof.

**Proposition B.9** \( A_0 = o_P (1) \) under \( \mathbb{H}_1 (\gamma_{NT}) \).

**Proof.** Again, we assume that \( p = 1 \). By the fact that \( \|\beta^0 - \hat{\beta}\| = O_P (\gamma_{NT}) \), (B.2)-(B.3), and Lemma D.1, we have

\[ |A_0| \leq c_\mathcal{K} a_{NT} \left\| \beta^0 - \hat{\beta} \right\| \left\| M^{(0)} \right\| \left\| M^{(2)} + M^{(rem)} \right\|

\times \sum_{1 \leq i,j \leq N} (\|\varepsilon_i\| + \left\| F^0 \lambda_i^0 \right\| + \|c_i\|) (\|\varepsilon_j\| + \left\| F^0 \lambda_j^0 \right\| + \|c_j\|)

= T O_P \left( a_{NT} \gamma_{NT} \delta_{NT}^{-2} \right) O_P \left( N^2 T \right) = O_P \left( \sqrt{NT} \delta_{NT}^{-2} (h!)^{1/4} \right) = o_P (1). \]

**Proposition B.10** \( A_{10} = o_P (1) \) under \( \mathbb{H}_1 (\gamma_{NT}) \).

**Proof.** First we decompose \( A_{10} \) as follows

\[ A_{10} = a_{NT} \sum_{1 \leq i,j \leq N} \left\{ (\varepsilon_i' + c_i') M^{(1)} \mathcal{K}_{ij} (M^{(2)} + M^{(rem)}) (\varepsilon_j + F^0 \lambda_j^0 + c_j)

+ \lambda_i^0 F^0 M^{(1)} \mathcal{K}_{ij} M^{(rem)} (\varepsilon_j + F^0 \lambda_j^0 + c_j)

+ \lambda_i^0 F^0 M^{(1)} \mathcal{K}_{ij} M^{(2)} F^0 \lambda_j^0 + \lambda_i^0 F^0 M^{(1)} \mathcal{K}_{ij} M^{(2)} c_j \right\}

= A_{10,1} + A_{10,2} + A_{10,3} + A_{10,4} + A_{10,5}. \]

By Lemma D.5(ii), \( A_{10,3} = o_P (1) \). By Lemma D.4(iii), \( A_{10,4} = o_P (1) \). We complete the proof of the proposition by showing that \( A_{10,s} = o_P (1) \) for \( s = 1, 2, 5 \).
By (B.2), (B.3), (B.11), and Lemma D.1

\[ |A_{10,1}| \leq a_{NT} \left\| M^{(2)} + M^{(rem)} \right\| \sum_{1 \leq i, j \leq N} \| K_{ij} \| \left\| M^{(1)} (\varepsilon_i + c_i) \right\| \left( \| \varepsilon_j \| + \| F^0 \lambda_j^0 \| + \| c_j \| \right) \]

\[ \leq c_k O \left( a_{NT} \delta_{NT}^2 \right) \left\{ O \left( (NT)^{-1/2} \right) \sum_{1 \leq i, j \leq N} \| F^0 \varepsilon_i \| \left( \| \varepsilon_j \| + \| F^0 \lambda_j^0 \| + \| c_j \| \right) \right. \]

\[ + O \left( N^{-1/2} \gamma_{NT} \right) \sum_{1 \leq i, j \leq N} \left( \| X_i \| + \| \Delta_i \| \right) \left( \| \varepsilon_j \| + \| F^0 \lambda_j^0 \| + \| c_j \| \right) \}

\[ = TO \left( a_{NT} \delta_{NT}^2 \right) O \left( N^{3/2} T^{1/2} + N^{-1/2} \gamma_{NT} N^2 T \right) = o_p \left( 1 \right), \]

\[ |A_{10,2}| \leq c_k a_{NT} \left\| M^{(1)} F^0 \right\| \left\| M^{(rem)} \right\| \sum_{1 \leq i, j \leq N} \| \lambda_i^0 \| \left( \| \varepsilon_j \| + \| F^0 \lambda_j^0 \| + \| c_j \| \right) \]

\[ \leq T a_{NT} O \left( N^{-1/2} T^{1/2} \right) O \left( \delta_{NT}^{-1} \gamma_{NT} \right) O \left( N^2 T^{1/2} \right) = O \left( T^{1/2} \delta_{NT}^{-1} (h!)^{1/4} \right) = o_p \left( 1 \right), \]

and

\[ |A_{10,5}| \leq c_k O \left( \gamma_{NT} a_{NT} \right) \left\| M^{(1)} F^0 \right\| \left\| M^{(2)} \right\| \sum_{1 \leq i, j \leq N} \| X_j \| \left( \| \Delta_j \| \right) \]

\[ = TO \left( \gamma_{NT} a_{NT} \right) O \left( N^{-1/2} T^{1/2} \right) O \left( \delta_{NT}^2 \right) O \left( N^2 T^{1/2} \right) = O \left( T^{1/2} \delta_{NT}^2 (h!)^{1/4} \right) = o_p \left( 1 \right). \]

This completes the proof of the proposition. ■

### C Proof of Theorem 3.3

By Theorem 2.2, it suffices to prove the theorem by showing that (i) \( \hat{\mathcal{B}}_{1NT} = B_{1NT} + o_p \left( 1 \right) \) and (ii) \( \hat{\mathcal{V}}_{NT} = \mathcal{V}_{NT} + o_p \left( 1 \right) \) under \( \mathbb{H}_1 \left( \gamma_{NT} \right) \). For (i), we apply (B.4) to obtain

\[ \hat{\mathcal{B}}_{1NT} = a_{NT} \sum_{i=1}^{N} \left( d_{i1} + d_{i2} + d_{i3} + d_{i4} \right) K_{ii} \left( d_{i1} + d_{i2} + d_{i3} + d_{i4} \right) \]

\[ = a_{NT} \sum_{i=1}^{N} \left( d'_{i1} K_{ii} d_{i1} + d'_{i2} K_{ii} d_{i2} + d'_{i3} K_{ii} d_{i3} + d'_{i4} K_{ii} d_{i4} + 2d'_{i1} K_{ii} d_{i2} + 2d'_{i2} K_{ii} d_{i3} \right. \]

\[ + 2d'_{i3} K_{ii} d_{i4} + 2d'_{i4} K_{ii} d_{i1} + 2d'_{i2} K_{ii} d_{i3} + 2d'_{i3} K_{ii} d_{i4} \}

\[ \equiv \hat{A}_1 + \hat{A}_2 + \hat{A}_3 + \hat{A}_4 + 2\hat{A}_5 + 2\hat{A}_6 + 2\hat{A}_7 + 2\hat{A}_8 + 2\hat{A}_9 + 2\hat{A}_{10}, \text{ say,} \]

where \( a_{NT} \equiv (h!)^{1/2} / (NT) \). Following the proof of Theorem 2.2, it is straightforward to show that under \( \mathbb{H}_1 \left( \gamma_{NT} \right), \hat{A}_1 = B_{1NT} + o_p \left( 1 \right) \) and \( \hat{A}_s = 0 \) for \( s = 2, 3, \ldots, 10 \). For example, for \( \hat{A}_1 \) we have

\[ \hat{A}_1 = a_{NT} \sum_{i=1}^{N} \left( \varepsilon_i + c_i \right) M_{F^0} K_{ii} M_{F^0} \left( \varepsilon_i + c_i \right) \]

\[ = a_{NT} \sum_{i=1}^{N} \varepsilon_i M_{F^0} K_{ii} M_{F^0} + a_{NT} \sum_{i=1}^{N} c_i M_{F^0} K_{ii} M_{F^0} + 2a_{NT} \sum_{i=1}^{N} \varepsilon_i M_{F^0} K_{ii} M_{F^0} c_i \]

\[ \equiv \hat{A}_{1,1} + \hat{A}_{1,2} + 2\hat{A}_{1,3}, \text{ say.} \]
The first term is $B_{1NT}$. By (B.5), Lemma D.1, and Assumptions A.5 and A.7(i), the second and third terms are respectively bounded above by

$$a_{NT} \sum_{i=1}^{N} \| \mathcal{K}_{ii} \| \| M_{F0} c_{i} \|^2 \leq O_P(\gamma_{NT}^2) a_{NT} \sum_{i=1}^{N} \| \mathcal{K}_{ii} \| (\| X_i \| + \| \Delta_i \|)^2$$

$$= O_P(\gamma_{NT}^2) O_P(NT^2) = O_P(N^{-1}) = o_P(1)$$

and

$$a_{NT} \sum_{i=1}^{N} \| \mathcal{K}_{ii} \| \| M_{F0} \varepsilon_i \| \| M_{F0} c_{i} \| \leq O_P(\gamma_{NT} a_{NT} \sum_{i=1}^{N} \| \mathcal{K}_{ii} \| \| \varepsilon_i \| (\| X_i \| + \| \Delta_i \|))$$

$$= O_P(\gamma_{NT} a_{NT} O_P(NT^2) = O_P \left( T^{1/2} N^{-1/2} (h!)^{1/4} \right) = o_P(1).$$

It follows that $\hat{A}_1 = B_{1NT} + o_P(1)$.

To show (ii), we decompose $\hat{V}_{NT} - V_{NT}$ as follows

$$\hat{V}_{NT} - V_{NT} = 2h!(NT)^{-2} \sum_{1 \leq t, s \leq T} \sum_{1 \leq i \neq j \leq N} \left[ \mathcal{K}_{ij,ts}^2 \varepsilon_{it}^2 \varepsilon_{js}^2 - E_D \left( \mathcal{K}_{ij,ts}^2 \varepsilon_{it}^2 \varepsilon_{js}^2 \right) \right]$$

$$+ 2h!(NT)^{-2} \sum_{1 \leq t, s \leq T} \sum_{1 \leq i \neq j \leq N} \mathcal{K}_{ij,ts}^2 \left( \varepsilon_{it}^2 \varepsilon_{js}^2 - \varepsilon_{it}^2 \varepsilon_{js}^2 \right) = V_{1NT} + V_{2NT}, \text{ say.}$$

Noting that $E_D(V_{1NT}) = 0$ and $E_D(V_{2NT}) = O_P(N^{-1})$ by the independence of $(\varepsilon_{it}, X_{it})$ across $i$ given $\mathcal{D}$ under Assumption A.4(ii), we have $V_{1NT} = o_P(1)$ by the Chebyshev inequality. For $V_{2NT}$, we further decompose it as follows

$$V_{2NT} = 2h!(NT)^{-2} \sum_{1 \leq t, s \leq T} \sum_{1 \leq i \neq j \leq N} \mathcal{K}_{ij,ts}^2 \left( \varepsilon_{it}^2 - \varepsilon_{it}^2 \right) \left( \varepsilon_{js}^2 - \varepsilon_{js}^2 \right)$$

$$+ 2h!(NT)^{-2} \sum_{1 \leq t, s \leq T} \sum_{1 \leq i \neq j \leq N} \mathcal{K}_{ij,ts}^2 \left( \varepsilon_{it}^2 - \varepsilon_{it}^2 \right) \varepsilon_{js}^2$$

$$+ 2h!(NT)^{-2} \sum_{1 \leq t, s \leq T} \sum_{1 \leq i \neq j \leq N} \mathcal{K}_{ij,ts}^2 \left( \varepsilon_{js}^2 - \varepsilon_{js}^2 \right)$$

$$= 2V_{2NT,1} + 2V_{2NT,2} + 2V_{2NT,3}.$$
In view of the fact that $\sum_{i,t} \tilde{\epsilon}_{it}^2 \leq \sum_{i,t} \tilde{\epsilon}_{it}^2$ and using Markov’s inequality, we have

$$(NT)^{-1} \sum_{i,t} (\tilde{\epsilon}_{it} + \epsilon_{it})^2 \leq 2 (NT)^{-1} \sum_{i,t} (\tilde{\epsilon}_{it}^2 + \epsilon_{it}^2) \leq 4 (NT)^{-1} \sum_{i,t} \epsilon_{it}^2 = o_P(1).$$

So we can prove $V_{2NT,1} = o_P(1)$ by showing that $V_{2NT,11} \equiv (h!)^{-1} (NT)^{-1} \sum_{i,t} (\tilde{\epsilon}_{it} - \epsilon_{it})^2 = o_P(1)$. By (B.4), $\hat{\epsilon}_i - \epsilon_i = \hat{d}_{1i} + d_{2i} + d_{3i} + d_{4i}$ where $\hat{d}_{1i} \equiv d_{1i} - \epsilon_i$. It follows that

$$V_{2NT,11} = (h!)^{-1} (NT)^{-1} \sum_{i=1}^N \left\| \hat{d}_{1i} + d_{2i} + d_{3i} + d_{4i} \right\|_F^2 \leq 4 (h!)^{-1} (NT)^{-1} \sum_{i=1}^N \left\{ \left\| \hat{d}_{1i} \right\|_F^2 + \left\| d_{2i} \right\|_F^2 + \left\| d_{3i} \right\|_F^2 + \left\| d_{4i} \right\|_F^2 \right\} = 4V_{2NT,11a} + V_{2NT,11b} + V_{2NT,11c} + V_{2NT,11d},$$

Noting that $\|P_{F^0} \epsilon_i\|_F^2 = O_P(T^{-1}) \|F^0 \epsilon_i\|_F^2$ and $\|M_{F^0} c_i\|_F^2 \leq \|c_i\|_F^2 = O (\gamma_{NT}^2) (\|X_i\|_F^2 + \|\Delta_i\|_F^2)$, we have by Assumptions A.4(iii) and A.5 and Markov’s inequality

$$V_{2NT,11a} = (h!)^{-1} (NT)^{-1} \sum_{i=1}^N \|P_{F^0} \epsilon_i + M_{F^0} c_i\|_F^2 \leq 2 (h!)^{-1} (NT)^{-1} \sum_{i=1}^N \left( \|P_{F^0} \epsilon_i\|_F^2 + \|M_{F^0} c_i\|_F^2 \right) = 2 (h!)^{-1} (NT)^{-1} \left\{ O_P(T^{-1}) \sum_{i=1}^N \|F^0 \epsilon_i\|_F^2 + O_P(\gamma_{NT}^2) \sum_{i=1}^N \left( \|X_i\|_F^2 + \|\Delta_i\|_F^2 \right) \right\} \leq O_P \left( (T^{-1} + \gamma_{NT}^2) (h!)^{-1} \right).$$

Similarly,

$$V_{2NT,11b} \leq \left\| \beta^0 - \beta \right\|_F^2 \sum_{k=1}^p \left\| M_k^{(0)} \right\|_F (h!)^{-1} (NT)^{-1} \sum_{i=1}^N \left\| \epsilon_i + F^0 \lambda_i^0 + c_i \right\|_F^2 = O_P \left( \gamma_{NT}^2 (h!)^{-1} \right),$$

$$V_{2NT,11c} \leq \left\| M^{(1)} \right\|_F (h!)^{-1} (NT)^{-1} \sum_{i=1}^N \left\| \epsilon_i + F^0 \lambda_i^0 + c_i \right\|_F^2 = O_P \left( N^{-1} (h!)^{-1} \right),$$

$$V_{2NT,11d} = \left\| M^{(2)} + M^{(rem)} \right\|_F (h!)^{-1} (NT)^{-1} \sum_{i=1}^N \left\| \epsilon_i + F^0 \lambda_i^0 + c_i \right\|_F^2 = O_P \left( \delta_{NT}^4 (h!)^{-1} \right).$$

It follows that $V_{2NT,11} = O_P((T^{-1} + N^{-1} + \gamma_{NT}^2) (h!)^{-1}) = o_P(1)$ by Assumption A.7(i) and thus $V_{2NT,1} = o_P(1)$.

For (ii2), we use the fact when $K$ is a symmetric PDF under Assumption A.6(i), there exists another symmetric PDF $K^0$ such that $K$ can be written as a two-fold convolution of $K^0 : K(u) = \int K^0(v) K^0(u - v) dv$. Define $K^0_h$ analogously as $K_h$. By Minkowski’s inequality, the fact that $K_{ij,ts} = \int$
It follows that
proof of
Lemma D.1
Suppose Assumptions A.4-A.7 hold. Then
which is not intended for publication but will be made available online.
prove the first lemma, and the proofs of the other lemmas are provided in the supplementary appendix,
Thus we have shown that
and
Again, by the relationship between \( K \) and \( K^0 \), the study of \( V_{2NT,1} \), Markov’s inequality, and Assumption A.7(i), we have
It follows that \( V_{2NT,2} = o_P(1) \). Thus we have shown that \( V_{2NT} = o_P(1) \). This completes the proof of (ii) and the theorem.

D Some Technical Lemmas

In this appendix we provide some technical lemmas that are used in the proof of Theorem 3.1. We only prove the first lemma, and the proofs of the other lemmas are provided in the supplementary appendix, which is not intended for publication but will be made available online.

Lemma D.1 Suppose Assumptions A.4-A.7 hold. Then \( c_K \equiv \max_{1 \leq i,j \leq N} \| K_{ij} \| = O_P(T) \).

Proof. Noting that \( \| K_{ij} \|^2 \leq \| K_{ij} \|_1 \| K_{ij} \|_\infty \) where \( \| K_{ij} \|_1 = \max_{1 \leq t \leq T} \sum_{s=1}^T |K_h(X_{it} - X_{js})| \) and \( \| K_{ij} \|_\infty = \max_{1 \leq i,j \leq N} \| K_h(X_{it} - X_{js}) \| \), it suffices to prove the lemma by showing that (i) \( \max_{1 \leq i,j \leq N} T^{-1} \| K_{ij} \|_1 = O_P(1) \) and (ii) \( \max_{1 \leq i,j \leq N} T^{-1} \| K_{ij} \|_\infty = O_P(1) \). We only prove (i) as the proof of (ii) is almost identical.
Let \( c_{NT} \equiv (NT)^{1/q_0} \), \( \eta_{iT,js} \equiv T^{-1} \sum_{t=1}^{T} K_h (X_{it} - X_{js}) \), and \( \bar{\eta}_{iT,js} \equiv T^{-1} \sum_{t=1}^{T} K_h (X_{it} - X_{js}) \times 1 \{ ||X_{js}|| \leq c_{NT} \} \). Then by Markov’s inequality and Assumption A.5(i), for any \( \epsilon^* > 0 \)

\[
\Pr\left( \max_{1 \leq i,j \leq N} \max_{1 \leq s \leq T} |\eta_{iT,js} - \bar{\eta}_{iT,js}| \geq \epsilon^* \right)
\]

\[
= \Pr\left( \max_{1 \leq i,j \leq N} \max_{1 \leq s \leq T} T^{-1} \sum_{t=1}^{T} K_h (X_{it} - X_{js}) 1 \{ ||X_{js}|| > c_{NT} \} \geq \epsilon^* \right)
\]

\[
\leq \Pr\left( \max_{1 \leq i,j \leq N} \max_{1 \leq s \leq T} ||X_{js}|| > c_{NT} \right) \leq \sum_{j=1}^{N} \sum_{s=1}^{T} \Pr ( ||X_{js}||^{q_0} > c_{NT}^{q_0} )
\]

\[
\leq 1 \sum_{j=1}^{N} \sum_{s=1}^{T} E ( ||X_{js}||^{q_0} 1 ( ||X_{js}||^{q_0} > c_{NT}^{q_0} ) ) = o (1).
\]

It follows that we can prove (i) by showing that \( L_{NT} \equiv \max_{1 \leq i \leq N} L_{iNT} = O_P (1) \), where

\[
L_{iNT} \equiv \max_{||x|| \leq c_{NT}} T^{-1} \sum_{t=1}^{T} K_h (X_{it} - x).
\]

By the Minkowski inequality

\[
L_{iNT} \leq \max_{||x|| \leq c_{NT}} T^{-1} \sum_{t=1}^{T} K_h (X_{it} - x) - E_D [K_h (X_{it} - x)] + \max_{||x|| \leq c_{NT}} T^{-1} \sum_{t=1}^{T} E_D [K_h (X_{it} - x)]
\]

\[
\equiv L_{iNT,1} + L_{iNT,2}, \text{ say.}
\]

By the change of variables and Assumptions A.4(iv), A.6(i) and A.7

\[
\max_{1 \leq i \leq N} L_{iNT,2} = \max_{1 \leq i \leq N} \max_{||x|| \leq c_{NT}} T^{-1} \sum_{t=1}^{T} \int f_{i,t} (x + h \odot u) K(u) du \leq C_f,
\]

where \( \odot \) denotes the Hadamard product. If \( \{ X_{it}, t = 1, 2, \ldots \} \) is strictly stationary and strong mixing, then one could replace \( E_D [K_h (X_{it} - x)] \) by its unconditional version and then apply Theorems 2 and 4 in Hansen (2008) so show that \( L_{iNT,1} = o_P (1) \) for each \( i \). Here, \( \{ X_{it}, t = 1, 2, \ldots \} \) is conditionally strong mixing given \( D \) so Hansen’s (2008) results are not applicable. We complete the proof of (i) by showing that

\[
\max_{1 \leq i \leq N} L_{iNT,1} = o_P (1).
\]

Take any small \( \epsilon > 0 \) and cover the compact set \( \{ ||x|| \leq c_{NT} \} \) with \( Q = c_{NT}^{q_0} (h!)^{-1} \epsilon^{-p} \) balls of the form \( A_l = \{ s : ||x - x_l|| \leq \epsilon (h!)^{1/p} \} \). The main step in the proof of (D.3) is to show that for any finite \( C_1 > 0 \)

\[
P_D \left( \max_{1 \leq i \leq N} \sup_{1 \leq s \leq Q} |\varphi_{iT} (x_l)| \geq cC_1 \right) = o (1),
\]

where \( P_D (\cdot) = P (|D|, \varphi_{iT} (x) = (Th!)^{-1} \sum_{t=1}^{T} Z_{i,t} (x) \text{ and } Z_{i,t} (x) = h! [K_h (X_{it} - x) - E_D [K_h (X_{it} - x)]] \).

Noting that \( \max_{1 \leq i \leq N} \max_{1 \leq s \leq T} \sup_x |Z_{i,t} (x)| \leq K \) and \( \max_{1 \leq i \leq N} \max_{1 \leq s \leq T} \sup_x E_D [Z_{i,t} (x)]^2 \leq C_2 h! \)

for some \( C_2 < \infty \), we can apply Boole’s inequality and the exponential inequality for conditional strongly
mixing processes (see Lemma E.2 in the supplemental appendix) to bound the left hand side of (D.4) from above by

\[
\sum_{i=1}^{N} \sum_{l=1}^{Q} P_D(|\varphi_{lT}(x_i)| \geq \varepsilon C_1) \leq NQ \sup_{1 \leq i \leq N} \sup_{1 \leq l \leq Q} P_D \left( \left\| T^{-1} \sum_{t=1}^{T} Z_{it}(x_i) \right\| \geq C_1 \varepsilon h! \right) \\
\leq 2NQ \left[ \tau \exp \left( -\frac{TC_1^2 \varepsilon^2 \langle h l \rangle}{4T} + \frac{2K C_1 \varepsilon h!}{\tau^3} \right) + T \alpha_D(\tau) \right] \\
\leq 2NQ \left[ \tau \exp \left( -\frac{C_1^2 \varepsilon^2 Th!}{4C_2^2 + 2C_1 K \varepsilon h^3 / 3} \right) + T \alpha_D(\tau) \right] \\
\to 0 \text{ as } (N,T) \to \infty,
\]

provided that \( \tau \in (1, Th!) \) such that \( Th! / \tau \gg T^\eta \) for some \( \eta > 0 \) and \( (NT)^{(1+p/\theta_0)} (h!)^{-1} \alpha_D(\tau) = o_{a.s.}(1) \). Assumption A.4(i) ensures the existence of such a \( \tau \). As a result, (D.4) holds and one can complete the rest of the proof for (D.3) following similar arguments as used in Hansen (2008). Combining (D.1), (D.2), and (D.3) yields \( L_{NT} = O_P(1) \). This completes the proof. \( \blacksquare \)

**Lemma D.2** Suppose the conditions in Theorem 3.2 hold. Then

(i) \( D_{1.1} \equiv \gamma_{NT} a_{NT} \sum_{1 \leq i, j \leq N} \varepsilon_i M_{i0} K_{ij} M_{j0} \Delta_j = o_P(1) \);

(ii) \( D_{1.2} \equiv \gamma_{NT} a_{NT} \sum_{1 \leq i, j \leq N} \varepsilon_i M_{i0} K_{ij} M_{j0} X_j = o_P(1) \);

(iii) \( D_{1.3} \equiv \gamma_{NT} a_{NT} \sum_{1 \leq i, j \leq N} \varepsilon_i M_{i0} K_{ij} M^{(0)} F^0 \lambda_j^0 = o_P(1) \);

(iv) \( D_{1.4} \equiv \gamma_{NT} a_{NT} \sum_{1 \leq i, j \leq N} \varepsilon_i M_{i0} K_{ij} M^{(1)} F^0 \lambda_j^0 = o_P(1) \);

(v) \( D_{1.5} \equiv \gamma_{NT} a_{NT} \sum_{1 \leq i, j \leq N} \lambda_i^0 F^0 M^{(1)} K_{ij} M^{(0)} \lambda_j^0 = o_P(1) \).

**Lemma D.3** Suppose the conditions in Theorem 3.2 hold. Then

(i) \( D_{2.1} \equiv a_{NT} \sum_{1 \leq i, j \leq N} \varepsilon_i^r P_{i0} K_{ij} \lambda_j^0 = o_P(1) \);

(ii) \( D_{2.2} \equiv a_{NT} \sum_{1 \leq i, j \leq N} \varepsilon_i^r P_{i0} K_{ij} \lambda_j^0 = o_P(1) \);

(iii) \( D_{2.3} \equiv a_{NT} \sum_{1 \leq i, j \leq N} \varepsilon_i^r K_{ij} M^{(0)} \lambda_j^0 = o_P(1) \);

(iv) \( D_{2.4} \equiv a_{NT} \sum_{1 \leq i, j \leq N} \varepsilon_i^r K_{ij} M^{(1)} \lambda_j^0 = o_P(1) \);

(v) \( D_{2.5} \equiv a_{NT} \sum_{1 \leq i, j \leq N} \varepsilon_i^r M_{i0} K_{ij} M^{(1)} \lambda_j^0 = o_P(1) \);

(vi) \( D_{2.6} \equiv a_{NT} \sum_{1 \leq i, j \leq N} \varepsilon_i^r M_{i0} K_{ij} M^{(1)} \lambda_j^0 = o_P(1) \);

(vii) \( D_{2.7} \equiv a_{NT} \sum_{1 \leq i, j \leq N} \varepsilon_i^r M_{i0} K_{ij} M^{(1)} \lambda_j^0 = o_P(1) \);

(viii) \( D_{2.8} \equiv a_{NT} \sum_{1 \leq i, j \leq N} \varepsilon_i^r M_{i0} K_{ij} M^{(1)} \lambda_j^0 = o_P(1) \).

**Lemma D.4** Suppose the conditions in Theorem 3.2 hold. Then

(i) \( D_{3.1} \equiv a_{NT} \sum_{1 \leq i, j \leq N} \varepsilon_i^r M^{(1)} K_{ij} M^{(1)} \lambda_j^0 = o_P(1) \);

(ii) \( D_{3.2} \equiv a_{NT} \sum_{1 \leq i, j \leq N} \varepsilon_i^r M_{i0} K_{ij} M^{(1)} = o_P(1) \);

(iii) \( D_{3.3} \equiv a_{NT} \sum_{1 \leq i, j \leq N} \varepsilon_i^r M_{i0} K_{ij} M^{(2)} \lambda_j^0 = o_P(1) \);

(iv) \( D_{3.4} \equiv a_{NT} \sum_{1 \leq i, j \leq N} \lambda_i^0 F^0 M^{(1)} K_{ij} M^{(2)} \lambda_j^0 = o_P(1) \).

**Lemma D.5** Suppose the conditions in Theorem 3.2 hold. Then

(i) \( D_{4.1} \equiv a_{NT} \sum_{1 \leq i, j \leq N} \varepsilon_i^r M_{i0} K_{ij} M^{(2)} \lambda_j^0 = o_P(1) \);

(ii) \( D_{4.2} \equiv a_{NT} \sum_{1 \leq i, j \leq N} \lambda_i^0 F^0 M^{(1)} K_{ij} M^{(2)} \lambda_j^0 = o_P(1) \).
References


Moon, H. & M. Weidner (2012) Linear regression for panel with unknown number of factors as interactive fixed effect. Working paper, USC.


