8-2014

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

Yonghui ZHANG
Singapore Management University, yizhang@smu.edu.sg

SMU ECONOMICS & STATISTICS
WORKING PAPER SERIES

Specification Test for Panel Data Models with Interactive Fixed Effects

Liangjun Su, Sainan Jin and Yonghui Zhang

August 2014

Paper No. 08-2014

Specification Test for Panel Data Models with Interactive Fixed Effects*

Liangjun Su[a], Sainan Jin[a], Yonghui Zhang[b]

[a] *School of Economics, Singapore Management University*

[b] *School of Economics, Renmin University of China*

August 12, 2014

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

*The authors gratefully thank the Co-editor Cheng Hsiao, the associate editor, and two anonymous referees for their many constructive comments on the previous version of the paper. They express their sincere appreciation to Jushan Bai, James Heckman, Hidehiko Ichimura, and Lung-fei Lee for discussions on the subject matter and valuable comments on the paper. They also thank the conference and seminar participants at the 8th Symposium on Econometric Theory and Application (SETA 2012), the 2013 China Meeting of the Econometric Society, the 2013 Tsinghua International Conference in Econometrics, the 2013 Asian Meeting of the Econometric Society (AMES 2013), and Fudan University for their valuable comments. Su and Jin gratefully acknowledge the Singapore Ministry of Education for Academic Research Fund under grant number MOE2012-T2-2-021. Jin also acknowledges the financial support from SMU with grant number C244/MSS11E005. All errors are the authors' sole responsibilities. Address Correspondence to: Liangjun Su, School of Economics, Singapore Management University, 90 Stamford Road, Singapore, 178903; E-mail: ljsu@smu.edu.sg, Phone: (+65) 6828 0386.

1 Introduction

Recently there has been a growing literature on large dimensional panel data models with interactive fixed effects (IFE hereafter). 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) proposes common correlated effect (CCE) estimation of panel data models with IFE; Bai (2009) proposes principal component analysis (PCA) estimation; Moon and Weidner (2010, 2013) reinvestigate Bai's (2009) PCA estimation and put it in the framework of Gaussian quasi maximum likelihood estimation (QMLE) framework; Su and Chen (2013) consider test for slope homogeneity in panel data models with IFE. For other developments on this type of models, see Ahn et al. (2001, 2013) for GMM approach with large N and fixed T, Kapetanios and Pesaran (2007) and Greenaway-McGrevy et al. (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, Avarucci and Zafaroni (2012) for generalized least squares (GLS) estimation, to name just a few.

Panel data models with IFE have been widely used in economics. Examples from labor economics include Carneiro et al. (2003) and Cunha et al. (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 IFE 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.

All of the aforementioned papers focus on the linear specification of regression relationship in panel data models with IFE. Recently nonparametric panel data models with IFE have started to receive attention; see Freyberger (2012), Su and Jin (2012), Jin and Su (2013), and Su and Zhang (2013). Freyberger (2012) considers identification and sieve estimation of nonparametric panel data models with IFE when N is large and T is fixed. Su and Jin (2012) extend the CCE estimation of Pesaran (2006) from the static linear model to a static nonparametric model via the method of sieves. Jin and Su (2013) construct a nonparametric test for poolability in nonparametric regression models with IFE. Su and Zhang (2013) extend the PCA estimation of Bai (2009) to nonparametric dynamic panel data models with IFE. 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 of 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

$$Y_{it} = m(X_{it}) + F_t^{0\prime}\lambda_i^0 + \varepsilon_{it}, \tag{1.1}$$

where $i = 1, \ldots, N$, $t = 1, \ldots, T$, X_{it} is a $p \times 1$ vector of observed regressors that may contain lagged dependent variables, $m(\cdot)$ is an unknown smooth function, F_t^0 is an $R \times 1$ vector of unobserved common factors, λ_i^0 is an $R \times 1$ vector of unobserved factor loadings, ε_{it} is an idiosyncratic error term. When $m(X_{it}) = X_{it}'\beta^0$ almost surely (a.s.) for some $\beta^0 \in \mathbb{R}^p$, (1.1) becomes the most popular linear panel data model with IFE, which is investigated by Pesaran (2006), Bai (2009), and Moon and Weidner (2010, 2013), among others. These authors consider various estimates for β and (λ_i, F_t) in the model. Asymptotic distributions for all estimators have been established and bias-correction is generally needed.

To motivate our test and study of the nonparametric model in (1.1), we take the economic growth model as an example. Prior to the middle 1990s, almost all empirical cross-country growth studies were based on the assumption that all countries obey a parametric (commonly linear) specification as required by the Solow model or its variants. Several studies conducted in the mid to late 1990s question the assumption of linearity and propose nonlinear alternatives for growth model. For example, in a cross sectional study Liu and Stengos (1999) employ a partially linear model to uncover the nonlinear pattern that initial income and schooling levels affect growth rates. Recently Su and Lu (2013) and Lee (2014) study economic growth via a dynamic panel data model and find significant nonlinear patterns. The former paper considers the traditional panel data model with only individual fixed effects when N is large and T is fixed; the latter considers large dimensional panel with both individual and time effects when both N and T are large. Given the fact that the linear dynamic panel data model is rejected in either paper, we can consider the following nonparametric panel data model

$$Y_{it} = m(X_{it}) + \alpha_i + f_t + \varepsilon_{it}, \tag{1.2}$$

where α_i and f_t are the usual individual and time fixed effects, Y_{it} is the growth rate of GDP per capita in country i at time period t, X_{it} is a vector that may include the last period economic growth rate $(Y_{i,t-1})$ as well as some economic growth determinants such as initial income level, human capital, and investment as a share of GDP. Obviously, employing the panel data model in (1.2) to growth allows us to control not only the country-specific effects but also the time-specific effects, but its limitation is also apparent. Loosely speaking, (1.2) assumes that the common shocks such as technology shocks, oil price shocks, and financial crises enter the equation through the time-specific effects f_t and have the same effects on all individual countries. This is certainly not the case in reality as a small economy tends to be more vulnerable to such shocks than a large economy. This motivates the use of nonparametric panel data models with IFE in (1.1) in the growth literature. We shall examine whether we can continue to find evidence of nonlinear patterns when the usual additive fixed effects is replaced by the IFE.

More generally, 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 parametric 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, it is very important to determine the number of common factors in factor analysis (e.g., Bai and Ng (2002), Onatski (2009), and Lu and Su (2013)) and to test for additivity versus interactivity in panel data models (e.g., Bai (2009)). 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 not consistent. 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), Li and Wang (1998), Stinchcombe and White (1998), Chen and Gao (2007), Hsiao et al. (2007), and Su and Ullah (2013), to name just a few. In addition, Hjellvik and Tjøstheim (1995) and Hjellvik et al. (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, Su and Lu (2013) and Lee (2014) consider testing for linearity in *dynamic* panel data models based on the weighted square distance between parametric and nonparametric estimates and individual-specific generalized spectral derivative, respectively; Lin et al. (2014) propose a consistent test for a linear functional form in a *static* panel data model with fixed effects. Nevertheless, to the best of our knowledge, there is no available test of linearity for panel data models with IFE.

In this paper, we propose a nonparametric test for linearity in panel data models with IFE. 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 motivates our residual-based test, like many other residual-based tests in the literature (e.g., Fan and Li (1996), Zheng (1996), Hsiao et al. (2007), and Su and Ullah (2013)). 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 linear panel data models with IFE 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 IFE.

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 IFE, the factors and factor loadings can only be estimated at a slower rate than the slope coefficients, and their estimation error plays an important role and complicates the asymptotic analysis of the local power function substantially. 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 IFE. 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 Appendixes 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 $\|A\|_F$ ($\equiv [\mathrm{tr}(AA')]^{1/2}$), its spectral norm as $\|A\|$ ($\equiv \sqrt{\mu_1(A'A)}$), 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 sth 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 $A > B$ ($A \geq B$) to indicate that $A - B$ is p.d. (p.s.d.). The operator $\xrightarrow{P}$ denotes convergence in probability, $\xrightarrow{D}$ 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 IFE 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

$$\mathbb{H}_0 : \Pr\left[m\left(X_{it}\right) = X_{it}'\beta^0\right] = 1 \text{ for some } \beta^0 \in \mathbb{R}^p. \tag{2.1}$$

The alternative hypothesis is the negation of $\mathbb{H}_0$:

$$\mathbb{H}_1 : \Pr\left[m\left(X_{it}\right) = X_{it}'\beta\right] < 1 \text{ for all } \beta \in \mathbb{R}^p. \tag{2.2}$$

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

$$\mathbb{H}_1\left(\gamma_{NT}\right) : m\left(X_{it}\right) = X_{it}'\beta^0 + \gamma_{NT}\Delta\left(X_{it}\right) \text{ a.s. for some } \beta^0 \in \mathbb{R}^p \tag{2.3}$$

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 $e_{it} \equiv Y_{it} - X_{it}'\beta^0 - F_t^{0\prime}\lambda_i^0$. Let $f_i(\cdot)$ denote the probability density function (PDF) of X_{it}. In view of the fact that $e_{it} = \varepsilon_{it}$ and $E\left(e_{it}|X_{it}\right) = 0$ under $\mathbb{H}_0$, we have

$$J \equiv E\left[e_{it}E\left(e_{it}|X_{it}\right)f_i\left(X_{it}\right)\right] = E\left\{\left[E\left(e_{it}|X_{it}\right)\right]^2 f_i\left(X_{it}\right)\right\} = 0$$

under $\mathbb{H}_0$. Nevertheless, under $\mathbb{H}_1$ we have $e_{it} = \varepsilon_{it} + m\left(X_{it}\right) - X_{it}'\beta^0$. So $E\left(e_{it}|X_{it}\right) = m\left(X_{it}\right) - X_{it}'\beta^0$ is not equal 0 a.s., implying that $E\left[e_{it}E\left(e_{it}|X_{it}\right)f_i\left(X_{it}\right)\right] > 0$ 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 $\mathbb{H}_0$ and obtain the restricted residuals $\hat{\varepsilon}_i = \left(\hat{\varepsilon}_{i1}, ..., \hat{\varepsilon}_{iT}\right)'$ for $i = 1, ..., N$. Then one can obtain the following sample analogue of J

$$J_{NT} = \frac{1}{(NT)^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\hat{\varepsilon}_{it}\hat{\varepsilon}_{js}K_h\left(X_{it} - X_{js}\right) = \frac{1}{(NT)^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\varepsilon}_i'\mathcal{K}_{ij}\hat{\varepsilon}_j \tag{2.4}$$

where $K_h(x) = \Pi_{l=1}^p h_l^{-1}k\left(x_l/h_l\right),\ k(\cdot)$ is a univariate kernel function, $h = \left(h_1, ..., h_p\right)$ is a bandwidth parameter, and $\mathcal{K}_{ij}$ is a $T \times T$ matrix whose (t,s)th element is given by $\mathcal{K}_{ij,ts} \equiv K_h\left(X_{it} - X_{js}\right).$

2.2 Estimation under the null

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

$$\begin{aligned}
Y_i &\equiv \left(Y_{i1}, \cdots, Y_{iT}\right)',\ X_i \equiv \left(X_{i1}, \cdots, X_{iT}\right)',\ \varepsilon_i \equiv \left(\varepsilon_{i1}, \cdots, \varepsilon_{iT}\right)',\ e_i \equiv \left(e_{i1}, \cdots, e_{iT}\right)', \\
F^0 &\equiv \left(F_1^0, \cdots, F_T^0\right)',\ \lambda^0 \equiv \left(\lambda_1^0, ..., \lambda_N^0\right)',\ X_{i,\cdot k} \equiv \left(X_{i1,k}, ..., X_{iT,k}\right)',\ \mathbf{Y} \equiv \left(Y_1, ..., Y_N\right)', \\
\mathbf{X}_k &\equiv \left(X_{1,\cdot k}, ..., X_{N,\cdot k}\right)',\ \boldsymbol{\varepsilon} \equiv \left(\varepsilon_1, ..., \varepsilon_N\right)',\ \text{and } \mathbf{e} \equiv \left(e_1, ..., e_N\right)'.
\end{aligned}$$

Clearly, $\mathbf{Y}$, $\mathbf{X}_k$, $\boldsymbol{\varepsilon}$ and $\mathbf{e}$ 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

$$Y_i = X_i\beta^0 + F^0\lambda_i^0 + \varepsilon_i \tag{2.5}$$

and

$$\mathbf{Y} = \sum_{k=1}^{p} \beta_k^0 \mathbf{X}_k + \lambda^0 F^{0\prime} + \boldsymbol{\varepsilon}, \tag{2.6}$$

where $\beta^0 \equiv \left(\beta_1^0, ..., \beta_p^0\right)'$.

Following Moon and Weidner (2010, 2013), the Gaussian quasi-maximum likelihood estimator (QMLE) $(\hat{\beta}, \hat{\lambda}, \hat{F})$ of (β, λ, F) can be obtained as follows

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

where

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

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

$$\hat{\beta} = \underset{\beta}{\arg\min} L_{NT}(\beta) \tag{2.9}$$

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

$$\begin{aligned} L_{NT}(\beta) &= \min_{\lambda,F} \mathcal{L}_{NT}(\beta, \lambda, F) \\ &= \min_{F} \frac{1}{NT}\text{tr}\left[\left(\mathbf{Y} - \sum_{k=1}^{p}\beta_k\mathbf{X}_k\right) M_F \left(\mathbf{Y} - \sum_{k=1}^{p}\beta_k\mathbf{X}_k\right)'\right] \\ &= \frac{1}{NT}\sum_{t=R+1}^{T}\mu_t\left[\left(\mathbf{Y} - \sum_{k=1}^{p}\beta_k\mathbf{X}_k\right)'\left(\mathbf{Y} - \sum_{k=1}^{p}\beta_k\mathbf{X}_k\right)\right]. \end{aligned} \tag{2.10}$$

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 $T \times T$ matrix at each step of the numerical optimization over β. For statistical inference, one also needs to obtain consistent estimates of λ^0 and F^0 under certain identification restrictions.

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

$$F'F/T = I_R \text{ and } \lambda'\lambda = \text{diagonal matrix.} \tag{2.11}$$

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

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

and

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

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 ε_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

$$\hat{\varepsilon}_i = M_{\hat{F}}\varepsilon_i - M_{\hat{F}}X_i(\hat{\beta} - \beta^0) + M_{\hat{F}}F^0\lambda_i^0 + M_{\hat{F}}\left(m_i - X_i\beta^0\right) \tag{2.14}$$

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 $\mathbb{H}_1(\gamma_{NT})$

Let $C_{NT}^{(1)}$ and $C_{NT}^{(2)}$ denote $p \times 1$ vectors whose kth elements are respectively given by

$$C_{NT,k}^{(1)} = \frac{1}{NT}\mathrm{tr}\left(M_{\lambda^0}\mathbf{X}_k M_{F^0}\mathbf{e}'\right) \text{ and} \tag{3.1}$$

$$C_{NT,k}^{(2)} = -\frac{1}{NT}\left[\mathrm{tr}\left(\mathbf{e}M_{F^0}\mathbf{e}'M_{\lambda^0}\mathbf{X}_k\Phi_1'\right) + \mathrm{tr}\left(\mathbf{e}'M_{\lambda^0}\mathbf{e}M_{F^0}\mathbf{X}_k'\Phi_1\right) + \mathrm{tr}\left(\mathbf{e}'M_{\lambda^0}\mathbf{X}_k M_{F^0}\mathbf{e}'\Phi_1\right)\right], \tag{3.2}$$

where $\Phi_1 \equiv \lambda^0\left(\lambda^{0\prime}\lambda^0\right)^{-1}\left(F^{0\prime}F^0\right)^{-1}F^{0\prime}$. Let D_{NT} denote a $p \times p$ matrix whose (k_1, k_2)th element is given by

$$D_{NT,k_1k_2} = \frac{1}{NT}\mathrm{tr}\left(M_{\lambda^0}\mathbf{X}_{k_1}M_{F^0}\mathbf{X}_{k_2}'\right). \tag{3.3}$$

Let $\alpha_{ik} \equiv \lambda_i^{0\prime}\left(\lambda^{0\prime}\lambda^0/N\right)^{-1}\lambda_k^0$ and $\tilde{X}_i \equiv M_{F^0}X_i - \frac{1}{N}\sum_{j=1}^N \alpha_{ij}M_{F^0}X_j$. It is easy to see that an alternative expression for D_{NT} is given by

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

which is used by Bai (2009). Following Moon and Weidner (2013) we refer to $C_{NT}^{(1)} + C_{NT}^{(2)}$ and D_{NT} as the approximated score and Hessian matrix for the profile quasi-likelihood function. Let $\delta_{NT} \equiv \min(\sqrt{N}, \sqrt{T})$. Let $\Delta_i \equiv (\Delta(X_{i1}), ..., \Delta(X_{iT}))'$ and $\mathbf{\Delta} \equiv (\Delta_1, ..., \Delta_N)'$.

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

Assumption A.1. (i) $N^{-1}\lambda^{0\prime}\lambda^0 \xrightarrow{P} \Sigma_\lambda > 0$ for some $R \times R$ matrix Σ_λ as $N \to \infty$.

(ii) $T^{-1}F^{0\prime}F^0 \xrightarrow{P} \Sigma_F > 0$ for some $R \times R$ matrix Σ_F as $T \to \infty$.

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

(iv) $\|\mathbf{X}_k\| = O_P\left(\sqrt{NT}\right)$ for $k = 1, ..., p$.

(v) $D_{NT} \xrightarrow{P} D > 0$ for some $p \times p$ matrix D as $(N, T) \to \infty$.

Assumption A.2. *(i)* $(NT)^{-1/2}\text{tr}(\mathbf{X}_k\boldsymbol{\varepsilon}') = O_P(1)$ for $k = 1, ..., p$.

(ii) Let $\mathbf{X}_{(\alpha)} = \sum_{k=1}^{p} \alpha_k \mathbf{X}_k$ such that $\|\alpha\| = 1$ where $\alpha = (\alpha_1, ..., \alpha_p)'$. There exists a finite constant $C > 0$ such that $\min_{\{\alpha\in\mathbb{R}^p:\|\alpha\|=1\}} \sum_{t=2R+1}^{T} \mu_t \left(\mathbf{X}'_{(\alpha)}\mathbf{X}_{(\alpha)}\right) \geq C$ with probability approaching 1 (w.p.a.1).

Assumption A.3. (i) $\|\boldsymbol{\Delta}\| = O_P(\sqrt{NT})$.

(ii) As $(N, T) \to \infty$, $\gamma_{NT} \to 0$.

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. A.1(v) requires that D_{NT} be asymptotically positive definite. A.2(i) requires weak exogeneity of the regressors $\mathbf{X}_k$ and A.2(ii) imposes the usual non-collinearity condition on $\mathbf{X}_k$. Note that A.2(ii) rules out time-invariant regressors or cross-sectionally invariant regressors, but it can be modified as in Moon Weidner (2013) to allow for both with more complicated notation and special treatment. A.3(i)-(ii) specify conditions on γ_{NT} and $\boldsymbol{\Delta}$. Note that we only require that the deviation from the null hypothesis is of local nature.

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* $\mathbb{H}_1(\gamma_{NT})$

$$\hat{\beta} - \beta^0 = D_{NT}^{-1}\left(C_{NT}^{(1)} + C_{NT}^{(2)}\right) + O_P\left\{\left[\gamma_{NT}^2\left(\delta_{NT}^{-1} + \gamma_{NT}\right) + \gamma_{NT}\delta_{NT}^{-3} + \delta_{NT}^{-5}\right]^{1/2}\right\}.$$

Remark 1. The result in Theorem 3.1 is comparable with that in Corollary 3.2 of Moon and Weidner (2010). Let $\bar{C}_{NT}^{(1)}$ and $\bar{C}_{NT}^{(2)}$ denote $p \times 1$ vectors with kth elements respectively given by

$$\bar{C}_{NT,k}^{(1)} = \frac{1}{NT}\text{tr}\left(M_{\lambda^0}\mathbf{X}_k M_{F^0}\boldsymbol{\varepsilon}'\right) \text{ and} \tag{3.4}$$

$$\bar{C}_{NT,k}^{(2)} = -\frac{1}{NT}\left[\text{tr}\left(\boldsymbol{\varepsilon}M_{F^0}\boldsymbol{\varepsilon}'M_{\lambda^0}\mathbf{X}_k\Phi_1'\right) + \text{tr}\left(\boldsymbol{\varepsilon}'M_{\lambda^0}\boldsymbol{\varepsilon}M_{F^0}\mathbf{X}_k'\Phi_1\right) + \text{tr}\left(\boldsymbol{\varepsilon}'M_{\lambda^0}\mathbf{X}_k M_{F^0}\boldsymbol{\varepsilon}'\Phi_1\right)\right]. \tag{3.5}$$

Following the proof of the above theorem, we can show that under $\mathbb{H}_0$, the asymptotic representation for $\hat{\beta} - \beta^0$ is given by

$$\hat{\beta} - \beta^0 = D_{NT}^{-1}\left(\bar{C}_{NT}^{(1)} + \bar{C}_{NT}^{(2)}\right) + O_P\left(\delta_{NT}^{-5/2}\right).$$

Let $\kappa_{NT} \equiv \left[\gamma_{NT}^2\left(\delta_{NT}^{-1} + \gamma_{NT}\right) + \gamma_{NT}\delta_{NT}^{-3} + \delta_{NT}^{-5}\right]^{1/2}$. To see the effect of the local deviation from the null hypothesis on the asymptotic expansion of $\hat{\beta} - \beta^0$, we apply the fact that $\mathbf{e} = \boldsymbol{\varepsilon} + \gamma_{NT}\boldsymbol{\Delta}$ under $\mathbb{H}_1(\gamma_{NT})$ and make the following decomposition under $\mathbb{H}_1(\gamma_{NT})$:

$$\begin{aligned}\hat{\beta} - \beta^0 &= D_{NT}^{-1}\left(\bar{C}_{NT}^{(1)} + \bar{C}_{NT}^{(2)}\right) + D_{NT}^{-1}\left(C_{NT}^{(1)} - \bar{C}_{NT}^{(1)}\right) + D_{NT}^{-1}\left(C_{NT}^{(2)} - \bar{C}_{NT}^{(2)}\right) + O_P(\kappa_{NT}) \\ &= A_{1NT} + A_{2NT} + A_{3NT} + O_P(\kappa_{NT}),\end{aligned}$$

where

$$\begin{aligned}
A_{1NT} &\equiv D_{NT}^{-1}\left(\bar{C}_{NT}^{(1)}+\bar{C}_{NT}^{(2)}\right),\\
A_{2NT} &\equiv D_{NT}^{-1}\left(C_{NT}^{(1)}-\bar{C}_{NT}^{(1)}\right)=D_{NT}^{-1}\frac{\gamma_{NT}}{NT}\text{tr}\left(M_{\lambda^0}\mathbf{X}_k M_{F^0}\mathbf{\Delta}'\right),\\
A_{3NT} &\equiv D_{NT}^{-1}\left(C_{NT}^{(2)}-\bar{C}_{NT}^{(2)}\right)\\
&= -D_{NT}^{-1}\frac{\gamma_{NT}}{NT}\left\{\left[\text{tr}\left(\boldsymbol{\varepsilon}M_{F^0}\mathbf{\Delta}'M_{\lambda^0}\mathbf{X}_k\Phi_1'\right)+\text{tr}\left(\mathbf{\Delta}M_{F^0}\boldsymbol{\varepsilon}'M_{\lambda^0}\mathbf{X}_k\Phi_1'\right)+\gamma_{NT}\text{tr}\left(\mathbf{\Delta}M_{F^0}\mathbf{\Delta}'M_{\lambda^0}\mathbf{X}_k\Phi_1'\right)\right]\right.\\
&\quad+\left[\text{tr}\left(\boldsymbol{\varepsilon}'M_{\lambda^0}\mathbf{\Delta}M_{F^0}\mathbf{X}_k'\Phi_1\right)+\text{tr}\left(\mathbf{\Delta}'M_{\lambda^0}\boldsymbol{\varepsilon}M_{F^0}\mathbf{X}_k'\Phi_1\right)+\gamma_{NT}\text{tr}\left(\mathbf{\Delta}'M_{\lambda^0}\mathbf{\Delta}M_{F^0}\mathbf{X}_k'\Phi_1\right)\right]\\
&\quad\left.+\left[\text{tr}\left(\boldsymbol{\varepsilon}'M_{\lambda^0}\mathbf{X}_k M_{F^0}\mathbf{\Delta}'\Phi_1\right)+\text{tr}\left(\mathbf{\Delta}'M_{\lambda^0}\mathbf{X}_k M_{F^0}\boldsymbol{\varepsilon}'\Phi_1\right)+\gamma_{NT}\text{tr}\left(\mathbf{\Delta}'M_{\lambda^0}\mathbf{X}_k M_{F^0}\mathbf{\Delta}'\Phi_1\right)\right]\right\}.
\end{aligned}$$

Apparently, A_{1NT} denotes the dominant term in the expression of $\hat{\beta}-\beta^0$ under $\mathbb{H}_0$, A_{2NT} and A_{3NT} signify the effect of the local deviation from the null hypothesis on the asymptotic expansion.
Remark 2. In view of the fact that under Assumptions A.1-A.3(i)

$$C_{NT,k}^{(1)}=\frac{\gamma_{NT}}{NT}\text{tr}\left(M_{\lambda^0}\mathbf{X}_k M_{F^0}\mathbf{\Delta}'\right)+\frac{1}{NT}\text{tr}\left(M_{\lambda^0}\mathbf{X}_k M_{F^0}\boldsymbol{\varepsilon}'\right)=O_P\left(\gamma_{NT}\right)+O_P\left(\delta_{NT}^{-2}\right)\tag{3.6}$$

and that $C_{NT}^{(2)}=O_P\left(\delta_{NT}^{-2}+\gamma_{NT}^2\right)$, we have under $\mathbb{H}_1\left(\gamma_{NT}\right)$

$$\begin{aligned}
\hat{\beta}-\beta^0 &= D_{NT}^{-1}\left(C_{NT}^{(1)}+C_{NT}^{(2)}\right)+O_P\{\left[\gamma_{NT}^2\left(\delta_{NT}^{-1}+\gamma_{NT}\right)+\gamma_{NT}\delta_{NT}^{-3}+\delta_{NT}^{-5}\right]^{1/2}\}\\
&= O_P(\gamma_{NT}+\delta_{NT}^{-2}).
\end{aligned}$$

As expected, the convergence rate of $\hat{\beta}$ to β^0 depends on γ_{NT} and δ_{NT}^{-2} jointly under $\mathbb{H}_1\left(\gamma_{NT}\right)$. If $\delta_{NT}^{-2}=O\left(\gamma_{NT}\right)$, then the local deviation from the null model controls the convergence rate of $\hat{\beta}$ to β^0. In the following study, we consider $\mathbb{H}_1\left(\gamma_{NT}\right)$ with $\gamma_{NT}=N^{-1/2}T^{-1/2}(h!)^{-1/4}$ and restrict $\delta_{NT}^{-2}=o(\gamma_{NT})$ (see Assumption A.7(i) below). The latter condition implies that $C_{NT}^{(2)}$ and the second term in $C_{NT}^{(1)}$ in (3.6) are asymptotically smaller than the first term in $C_{NT}^{(1)}$. Then we have

$$\hat{\beta}-\beta^0=\gamma_{NT}D_{NT}^{-1}\Pi_{NT}+O_P\left(\delta_{NT}^{-2}+\gamma_{NT}^2\right)=\gamma_{NT}D_{NT}^{-1}\Pi_{NT}+o_P\left(\gamma_{NT}\right),\tag{3.7}$$

where Π_{NT} is a $p\times 1$ vector whose kth element is given by[1]

$$\Pi_{NT,k}=\left(NT\right)^{-1}\text{tr}\left(M_{\lambda^0}\mathbf{X}_k M_{F^0}\mathbf{\Delta}'\right)=O_P\left(1\right).\tag{3.8}$$

Note that we do not require N and T diverge to ∞ at the same speed, nor do we require that one diverge to ∞ faster than the other.

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-*σ*-algebra of* $\mathcal{A}$. *Let* $P_{\mathcal{B}}\left(\cdot\right)\equiv P\left(\cdot|\mathcal{B}\right)$. *Let* $\{\xi_t,t\geq 1\}$ *be a sequence of random variables defined on* $(\Omega,\mathcal{A},P)$. *The sequence* $\{\xi_t,t\geq 1\}$

[1] Using the notation $\tilde{X}_i$, one can also write $\Pi_{NT}=\frac{1}{NT}\sum_{i=1}^N\tilde{X}_i'\Delta_i$.

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 $\alpha^{\mathcal{B}}(t)$ converging to 0 a.s. as $t \to \infty$ such that

$$|P_{\mathcal{B}}(A \cap B) - P_{\mathcal{B}}(A) P_{\mathcal{B}}(B)| \leq \alpha^{\mathcal{B}}(t) \text{ a.s.} \tag{3.9}$$

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

The above definition is due to Prakasa Rao (2009); see also Roussas (2008). When one takes $\alpha^{\mathcal{B}}(t)$ as the supremum of the left hand side object in (3.9) over the set $\{A \in \sigma(\xi_1, ..., \xi_k), B \in \sigma(\xi_{k+t}, \xi_{k+t+1}, ...),$ $k \geq 1\}$, we refer it to the $\mathcal{B}$-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 $1 - \frac{1}{\tilde{q}_3} = \frac{1}{q_1} + \frac{1}{\tilde{q}_2}$. Let $h! \equiv \Pi_{k=1}^{p} h_k$. Let $\|A\|_q \equiv \{E\|A\|^q\}^{1/q}$ for any random scalar or vector A. Let $\mathcal{D} \equiv \sigma(F^0, \lambda^0)$, the σ-field generated by (F^0, λ^0). To study the asymptotic distribution of the test statistic, we add the following assumptions.

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

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

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

(iv) For each $i = 1, ..., N$, let $f_{i,t}(x)$ denote the marginal PDF of X_{it} given $\mathcal{D}$, and $f_{i,ts}(x, \bar{x})$ the joint PDF of X_{it} and X_{is} given $\mathcal{D}$. $f_{i,t}(\cdot)$ and $f_{i,ts}(\cdot, \cdot)$ 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) $\max_{1 \leq i \leq N} \|\varepsilon_{it}\|_{q_1} \leq C_\varepsilon < \infty$ for some $q_1 > 4$.

(iii) $\max_{1 \leq i \leq N} \|\lambda_i^0\|_4 \leq C_\lambda < \infty$ and $\max_{1 \leq t \leq T} \|F_t^0\|_4 \leq C_F < \infty$.

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

(v) $(NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} E\{[\|F_t^0\| + \|\lambda_i^0\|]\varepsilon_{it}^4\} = O(1)$ and $(NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} E[\|F_t^0\|^2 \|\lambda_i^0\|^2 \varepsilon_{it}^2] = O(1)$.

Assumption A.6. (i) The kernel function $k(\cdot) : \mathbb{R} \to \mathbb{R}$ is a symmetric, continuous and bounded PDF.

(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, $\sup_u |(\partial/\partial u) k(u)| \leq C_k$, $k(u) \leq C_k |u|^{-q_0}$ and $|(\partial/\partial u) k(u)| \leq C_k |u|^{-\nu}$ for $|u| > L$ and for some $\nu > 1$.

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

(ii) As $(N, T) \to \infty$, $NT^{-1}(h!)[(h!)^{2(1-q_2)/q_2} + (h!)^{-2\tilde{\eta}/(1+\tilde{\eta})}]^2 \to 0$ and $N^2 T^{-2} (h!)^{(4-3q_2)/q_2} \to 0$.

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 (α-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 (2013) notice, even if $\left\{\left(\varepsilon_{it}, F_t^0\right), t \geq 1\right\}$ is strong mixing, the simple panel AR(1) model with IFE, $Y_{it} = \rho_0 Y_{i,t-1} + \lambda_i^{0\prime} F_t^0 + \varepsilon_{it}$, usually does not yield a strong mixing process $\{Y_{it}, t \geq 1\}$ unless one assumes that λ_i^0's are 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. They also suggest that when the fixed effects are random, one should adopt the concept of conditional strong mixing where the mixing coefficient is defined by conditioning on the individual fixed effects. Lee (2013) follows this suggestion and demonstrates that under suitable conditions a nonlinear panel AR(1) process with random fixed effects is β-mixing and thus α-mixing by conditioning on the individual fixed effects. Gagliardini and Gouriéroux (2012) assume conditional β-mixing by conditioning on the factor path in a nonlinear dynamic panel data model with common unobserved factors. Here we define the conditional strong mixing processes by conditioning on the sigma-field $\mathcal{D}$. For the above panel AR(1) process, through the conditioning, we can treat $\lambda_i^{0\prime} F_t^0$ as an intercept term, so that the $\mathcal{D}-$strong mixing property simply follows from that of the usual AR(1) process which essentially requires that $|\rho_0| < 1$ and that ε_{it} have nontrivial absolutely continuous component in additional to some moment condition on ε_{it}.

A.4(i), in conjunction with A.4(ii)-(iii), facilitates our asymptotic analysis. We assume that $\alpha_{\mathcal{D}}(\cdot) \equiv \alpha_{NT}^{\mathcal{D}}(\cdot) \equiv \max_{1 \leq i \leq N} \alpha_{NT,i}^{\mathcal{D}}(\cdot)$ satisfies some summability condition. With more lengthy argument, it is possible to relax this condition, say, by assuming that $\frac{1}{N}\sum_{i=1}^{N}\sum_{s=1}^{\infty} \alpha_{NT,i}^{\mathcal{D}}(s)^{1/\tilde{q}_3} \leq C_\alpha < \infty$ a.s. The dependence of the mixing rate on $\tilde{q}_3$ and $\tilde{\eta}$ 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 $\alpha_{\mathcal{D}}(\cdot)$ can easily be met by specifying $\tau = \lfloor C_\tau \log T \rfloor$ for some sufficiently large C_τ, where $\lfloor a \rfloor$ denotes the integer part of a. A.4(ii) requires that (ε_i, X_i) be conditionally independent across i but does not rule out cross sectional dependence among them. When $X_{it} = Y_{i,t-1}$ and ε_{it} exhibits conditional heteroskedasticity (e.g., $\varepsilon_{it} = \sigma_0(Y_{i,t-1})\epsilon_{it}$ where $\epsilon_{it} \backsim$ 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 λ_i^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 σ-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 ε_{it} be a martingale difference sequence (m.d.s.) with respect to the filter $\mathcal{F}_{NT,t-1}$, which allows lagged dependent variables in X_{it} and conditional heteroskedasticity, skewness or kurtosis in ε_{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 IFE typically assumes that ε_{it} is independent of λ_j^0 and F_s^0 for all i, j, t, s; see, e.g., Pesaran (2006), Bai (2009), Moon and Weidner (2010, 2013), and Bai and Li (2012). In particular, Moon and Weidner (2010, 2013) 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(iv) 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 ε_{it}, λ_i^0, F_t^0, 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(i) are mainly used to demonstrate that $\max_{1\leq i,j\leq N} T^{-1} \|\mathcal{K}_{ij}\| = O_P(1)$ in Lemma D.1 in Appendix D. A.7 specifies conditions on the bandwidth in relation to the sample sizes (N, T). Note that $NT\delta_{NT}^{-4}(h!)^{1/2} \to 0$ is equivalent to $\left(NT^{-1} + N^{-1}T\right)(h!)^{1/2} \to 0$, which restricts the relative speed at which N and T diverge to ∞ in relation with $h!$.

Let

$$B_{1NT} \equiv \frac{(h!)^{1/2}}{NT} \sum_{i=1}^{N} \varepsilon_i' M_{F^0} \mathcal{K}_{ii} M_{F^0} \varepsilon_i, \tag{3.10}$$

$$B_{2NT} \equiv \frac{1}{(NT)^2} \sum_{1\leq i,j\leq N} \left(M_{F^0}\Delta_i - \tilde{X}_i D_{NT}^{-1}\Pi_{NT}\right)' \mathcal{K}_{ij}(M_{F^0}\Delta_j - \tilde{X}_j D_{NT}^{-1}\Pi_{NT}), \tag{3.11}$$

$$V_{NT} \equiv \frac{2h!}{(NT)^2} \sum_{1\leq i\neq j\leq N} \sum_{t=1}^{T}\sum_{s=1}^{T} E_{\mathcal{D}}\left(\mathcal{K}_{ij,ts}^2 \varepsilon_{it}^2 \varepsilon_{js}^2\right), \tag{3.12}$$

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* $\gamma_{NT} \equiv (NT)^{-1/2}(h!)^{-1/4}$,

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

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

Remark 3. 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 $\mathbb{H}_1(\gamma_{NT})$, where $A_{NT} \equiv \sum_{1\leq i<j\leq N} W_{NT}(u_i, u_j)$, $W_{NT}(u_i, u_j) \equiv 2(h!)^{1/2} \sum_{1\leq t,s\leq T} \varepsilon_{it}\mathcal{K}_{ij,ts}\varepsilon_{js}$ 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 $A_{NT} \xrightarrow{D} N(0, V_0)$ under Assumptions A.1-A.7.[2]

In view of the fact $B_{2NT} = 0$ under $\mathbb{H}_0$, an immediate consequence of Theorem 3.2 is

$$NT(h!)^{1/2} J_{NT} - B_{1NT} \xrightarrow{D} N(0, V_0) \text{ under } \mathbb{H}_0.$$

[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.

To implement the test, we need to estimate the asymptotic bias B_{1NT} and asymptotic variance V_{NT} consistently under $\mathbb{H}_0$. We propose to estimate B_{1NT} and V_{NT} respectively by

$$\hat{B}_{1NT} \equiv \frac{(h!)^{1/2}}{NT} \sum_{i=1}^{N} \hat{\varepsilon}_i' \mathcal{K}_{ii} \hat{\varepsilon}_i \text{ and } \hat{V}_{NT} \equiv \frac{2h!}{(NT)^2} \sum_{1 \leq t,s \leq T} \sum_{1 \leq i \neq j \leq N} \mathcal{K}_{ij,ts}^2 \hat{\varepsilon}_{it}^2 \hat{\varepsilon}_{js}^2.$$

Then we define a feasible test statistic

$$\hat{\Gamma}_{NT} \equiv \left(NT (h!)^{1/2} J_{NT} - \hat{B}_{1NT} \right) / \sqrt{\hat{V}_{NT}}. \tag{3.13}$$

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

Theorem 3.3 *Suppose Assumptions A.1-A.7 hold. Then under* $\mathbb{H}_1(\gamma_{NT})$, $\hat{\Gamma}_{NT} \xrightarrow{D} N\left(B_2/\sqrt{V_0}, 1\right)$.

Remark 4. 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

$$\Pr\left(\hat{\Gamma}_{NT} > z \mid \mathbb{H}_1(\gamma_{NT})\right) \to 1 - \Phi\left(z - B_2/\sqrt{V_0}\right) \text{ as } (N,T) \to \infty,$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function (CDF). We obtain this distributional result despite the fact that the unobserved factors F_t^0 and factor loadings λ_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 $\mathbb{H}_0$, $B_2 = 0$, and $\hat{\Gamma}_{NT}$ is asymptotically distributed $N(0,1)$. This is stated in the following corollary.

Corollary 3.4 *Suppose the conditions in Theorem 3.3 hold. Then under* $\mathbb{H}_0$, $\hat{\Gamma}_{NT} \xrightarrow{D} N(0,1)$.

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

Remark 5. Theorem 3.1 says nothing about the asymptotic property of the QMLE $\hat{\beta}$ under the global alternative $\mathbb{H}_1$. In this case, we can define the pseudo-true parameter β^* as the probability limit of $\hat{\beta}$. Then

$$\bar{\Delta}(X_{it}) \equiv m(X_{it}) - \beta^{*\prime} X_{it}$$

does not equal 0 a.s. Let $\bar{\mathbf{\Delta}}$ be analogously defined as $\mathbf{\Delta}$ but with the local deviation $\Delta(X_{it})$ replaced by the global one $\bar{\Delta}(X_{it})$. In this case, we can show that under the additional assumption $\|\bar{\mathbf{\Delta}}\| = o_P((NT)^{1/2})$,

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

where $\bar{\Pi}_{NT}$ is a $p \times 1$ vector whose kth element is given by $\bar{\Pi}_{NT,k} = (NT)^{-1} \text{tr}\big(M_{\lambda^0} \mathbf{X}_k M_{F^0} \bar{\mathbf{\Delta}}'\big)$. In addition, following the proof of Theorem 3.2, we can show that

$$\begin{aligned} J_{NT} &= \frac{1}{(NT)^2} \sum_{1 \le i,j \le N} \left(M_{F^0} \bar{\Delta}_i - \tilde{X}_i D_{NT}^{-1} \bar{\Pi}_{NT} \right)' \mathcal{K}_{ij} \left(M_{F^0} \bar{\Delta}_j - \tilde{X}_j D_{NT}^{-1} \bar{\Pi}_{NT} \right) + o_P(1) \\ &= \bar{B}_{2NT} + o_P(1), \end{aligned}$$

which has a positive probability limit. This, together with the fact that $\hat{B}_{1NT} = O_P\left((h!)^{-1/2}\right)$ and $\hat{V}_{NT}$ has a well behaved probability limit under $\mathbb{H}_1$, implies that our test statistic $\hat{\Gamma}_{NT}$ diverges at the usual nonparametric rate $NT(h!)^{1/2}$ under $\mathbb{H}_1$. That is

$$\Pr\left(\hat{\Gamma}_{NT} > b_{NT} \mid \mathbb{H}_1\right) \to 1 \text{ as } (N,T) \to \infty$$

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

Remark 6 (Test under strict exogeneity). Up to now we assume the existence of lagged dependent variables in the panel regression and rely on the notion of conditional strong mixing to study the asymptotic properties of our test statistic. To avoid dynamic misspecification of the model and to facilitate the asymptotic analysis of our test statistic, we assume certain m.d.s. condition in Assumption A.4(iii) which unfortunately rules our serial correlation among the idiosyncratic error terms. If the panel data model is *static* and the regressors are *strictly exogenous* as in Pesaran (2006) and Bai (2009), we can rely on the usual notion of strong mixing and allow serial correlation in the error terms. In this case, we can replace Assumption A.4 by Assumption A.4* in the supplementary appendix and demonstrate that Theorems 3.2 and 3.3 continue to hold under some modifications on Assumption A.5. To save the space, we relegate the discussions to the supplementary Appendix F.

3.3 A Bootstrap version of the test

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 factor 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 plays an important role in the determination of the asymptotic null distribution of our test statistic and may further lead to some finite sample size distortions; this also occurs in Su and Chen's (2013) LM-test for the slope homogeneity in a linear dynamic panel data models with IFE. 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 $\hat{\varepsilon}_{it} = Y_{it} - X_{it}'\hat{\beta} - \hat{F}_t'\hat{\lambda}_i$ where $\hat{\beta}$, $\hat{F}_t$ 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 $i = 1, ..., N$ and $t = 1, 2, \ldots, T$, obtain the bootstrap error $\varepsilon_{it}^* = \hat{\varepsilon}_{it}\eta_{it}$ where η_{it} are independently and identically distributed (IID) $N(0,1)$ across i and t. Generate the bootstrap analogue Y_{it}^* of Y_{it} by holding $(X_{it}, \hat{F}_t, \hat{\lambda}_i)$ as fixed: $Y_{it}^* = \hat{\beta}' X_{it} + \hat{\lambda}_i' \hat{F}_t + \varepsilon_{it}^*$ for $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$.[3]

3. Given the bootstrap resample $\{Y_{it}^*, X_{it}\}$, obtain the QMLEs $\hat{\beta}^*, \hat{F}_t^*$ and $\hat{\lambda}_i^*$. Obtain the residuals $\hat{\varepsilon}_{it}^* = Y_{it}^* - X_{it}\hat{\beta}^* - \hat{F}_t^{*\prime}\hat{\lambda}_i^*$ and calculate the bootstrap test statistic $\hat{\Gamma}_{NT}^*$ based on $\{\hat{\varepsilon}_{it}^*, X_{it}\}$.

4. Repeat Steps 2-3 for B times and index the bootstrap statistics as $\left\{\hat{\Gamma}_{NT,b}^*\right\}_{b=1}^B$. The bootstrap p-value is calculated as $p^* \equiv B^{-1}\sum_{b=1}^B 1(\hat{\Gamma}_{NT,b}^* \geq \hat{\Gamma}_{NT})$, 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 (2013), 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* $\hat{\Gamma}_{NT}^* \xrightarrow{D} N(0,1)$ *conditionally on the observed sample* $\mathcal{W}_{NT} \equiv \{(X_1, Y_1), ..., (X_N, Y_N)\}$.

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)
DGP 1: $Y_{it} = \rho^0 Y_{i,t-1} + \lambda_i^{0\prime} F_t^0 + \varepsilon_{it}$,
DGP 2: $Y_{it} = \beta_1^0 X_{it,1} + \beta_2^0 X_{it,2} + \lambda_i^{0\prime} F_t^0 + \varepsilon_{it}$,
DGP 3: $Y_{it} = \rho^0 Y_{i,t-1} + \beta_1^0 X_{it,1} + \beta_2^0 X_{it,2} + \lambda_i^{0\prime} F_t^0 + \varepsilon_{it}$,
DGP 4: $Y_{it} = \delta\Phi(Y_{i,t-1}) Y_{i,t-1} + \lambda_i^{0\prime} F_t^0 + \varepsilon_{it}$,
DGP 5: $Y_{it} = \beta_1^0 X_{it,1} + \beta_2^0 X_{it,2} + \delta\Phi(X_{it,1}X_{it,2}) + \lambda_i^{0\prime} F_t^0 + \varepsilon_{it}$,
DGP 6: $Y_{it} = \frac{1}{2}\delta\Phi(Y_{i,t-2}) Y_{i,t-2} + \beta_1^0 X_{it,1} + \beta_2^0 X_{it,2} + \delta X_{it,1}\Phi(X_{it,2}) + \lambda_i^{0\prime} F_t^0 + \varepsilon_{it}$,

[3] This is the case even if X_{it} contains lagged dependent variables, say, $Y_{i,t-1}$ and $Y_{i,t-2}$.

where $i = 1, 2, \ldots, N$, $t = 1, 2, \ldots, T$, $(\rho^0, \beta_1^0, \beta_2^0) = (0.6, 1, 3)$, $\delta = 0.25$, $\varepsilon_{it} \backsim$ IID $N(0, 1)$, and $\Phi(\cdot)$ is the standard normal CDF. Here $\lambda_i^0 = (\lambda_{i1}^0, \lambda_{i2}^0)'$, $F_t^0 = (F_{t1}^0, F_{t2}^0)'$, and the regressors are generated according to

$$X_{it,1} = \mu_1 + c_1 \lambda_i^{0\prime} F_t^0 + \eta_{it,1} \text{ and } X_{it,2} = \mu_2 + c_2 \lambda_i^{0\prime} F_t^0 + \eta_{it,2},$$

where the variables λ_{ij}^0, F_{tj}^0, and $\eta_{it,j}$, $j = 1, 2$, are all IID $N(0, 1)$, mutually independent of each other, and independent of $\{\varepsilon_{it}\}$. Clearly, the regressors $X_{it,1}$ and $X_{it,2}$ are correlated with λ_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 for each individual 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, we generate the error terms as follows: $\varepsilon_{it} = \sigma_{it}\epsilon_{it}$, $\sigma_{it} = \left(0.1 + 0.2Y_{i,t-1}^2\right)^{1/2}$, $\epsilon_{it} \sim$IID $N(0, 1)$. For DGPs 2h-3h and 5h-6h, the errors are generated as follows: $\varepsilon_{it} = \sigma_{it}\epsilon_{it}$, $\sigma_{it} = [0.1 + 0.1(X_{it,1}^2 + X_{it,2}^2)]^{1/2}$, $\epsilon_{it} \sim$IID $N(0, 1)$.

As a referee kindly pointed out, it is important to allow serial dependence in the error process. So we consider the following two additional error generating processes :

$$\text{MA}(1) \quad : \quad \varepsilon_{it} = 0.5\zeta_{i,t-1} + \zeta_{it} \text{ with } \zeta_{it} \sim \text{IID } N(0, 1), \tag{4.1}$$

$$\text{AR}(1) \quad : \quad \varepsilon_{it} = 0.3\varepsilon_{i,t-1} + \zeta_{it} \text{ with } \zeta_{it} \sim \text{IID } N(0, 1). \tag{4.2}$$

Then we consider another four DGPs as follows:
DGPs 7 and 8: $Y_{it} = \beta_1^0 X_{it,1} + \beta_2^0 X_{it,2} + \lambda_i^{0\prime} F_t^0 + \varepsilon_{it}$,
DGPs 9 and 10: $Y_{it} = \beta_1^0 X_{it,1} + \beta_2^0 X_{it,2} + \delta\Phi\left(X_{it,1}X_{it,2}\right) + \lambda_i^{0\prime} F_t^0 + \varepsilon_{it}$,
where ε_{it}'s are generated according to (4.1) in DGPs 7 and 9 and (4.2) in DGPs 8 and 10; $X_{it,1}$, $X_{it,2}$, λ_i^0, F_t^0 are generated as in the previous DGPs. As before, we set $(\beta_1^0, \beta_2^0) = (1, 3)$ and $\delta = 0.25$. DGPs 7-8 and 9-10 are for level and power studies, respectively. Clearly, in these DGPs we allow for exogenous regressors and weakly serially dependent errors.

4.1.2 Implementation

To calculate the test statistic, we need to choose both the kernel function and the bandwidth parameter $h = (h_1, ..., h_p)$ where $p = 1$ in DGPs 1, 4, 1h, and 4h, $= 2$ in DGPs 2, 5, 2h, 5h, and 7-9, and $= 3$ 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. It is well known that the choice of kernel function is not crucial for nonparametric kernel-based tests. So we adopt the Gaussian kernel throughout: $k(u) = (2\pi)^{-1/2} \exp\left(-u^2/2\right)$. As to the bandwidth, a common feature of the kernel-based tests is the involvement of a single bandwidth, which creates two limitations: one is that the tests can be sensitive to the single bandwidth used, and the other is that these tests are consistent against local alternatives of form (2.3) only when γ_{NT} is at the rate of $(NT)^{-1/2} (h!)^{-1/4}$ or larger. Therefore it is worthwhile to consider different choices of bandwidths. Generally speaking, there are at least four ways to choose the bandwidth for a nonparametric smooth test. One is based on Silverman's rule of thumb, which is simple but does not have any optimality property. The second is to choose the bandwidth by certain cross-validation methods (typically leave-one-out least squares cross-validation). The chosen bandwidth may be optimal for the estimation purpose but does not have any optimality property for the kernel-based test. The third one is the adaptive-rate-optimal rule proposed by Horowitz and Spokoiny (2001, HS hereafter). The fourth one is based on the idea of maximizing the local power while keeping the size well controlled; see, e.g., Gao and Gijbels (2008).

In this paper we consider two choices of bandwidth sequences, one is based on Silverman's rule of thumb (ROT), and the other on HS's adaptive test procedure. The former is used to examine the sensitivity of our test to the bandwidth and the latter is intended to improve the power performance of our test. We choose the ROT bandwidth sequences according to: $h_l = c_0 s_l \left(NT\right)^{-1/(4+p)}$, where s_l stands for the sample standard deviation for the lth element in X_{it} and $c_0 = 0.5$, 1, and 2. HS propose an adaptive test that combines a version of the Härdle-Mammen test statistics over a set of bandwidths. The test is called adaptive and rate optimal if it adapts to the unknown smoothness of the local alternative hypothesis and is able to achieve the optimal order in the minimax sense. To ensure the adaptive rate-optimality, HS have to impose some strong assumptions on the underlying DGP: the observations are IID and the regressors, random or not, are uniformly bounded with continuous distributions. Chen and Gao (2007) relax the IID assumptions and show that the results established by HS are valid for weakly dependent observations. We conjecture that these results can also be extended to our dynamic panel data models with IFE, but the formal study is beyond the scope of the current paper. Instead, we just apply their adaptive test procedure to our test and consider its finite sample performance. Following HS and Chen and Gao (2007), we use a geometric grid consisting of the points $h_{j,s} = \omega^j s_j h_{\min}$ $(s = 0, 1, ..., \mathcal{N} - 1; j = 1, ..., p)$, where $\mathcal{N}$ is number of grid points, $\omega = (h_{\max}/h_{\min})^{1/(\mathcal{N}-1)}$, $h_{\min} = 0.4\left(NT\right)^{-1/(2.1p)}$ and $h_{\max} = 3\left(NT\right)^{-1/1000}$. It is easy to verify these bandwidths also meet our theoretical requirements on the bandwidth when $N \propto T$. Like HS, we choose $\mathcal{N}$ according to the rule of thumb $\mathcal{N} = \lfloor \log(NT) \rfloor + 1$ where $\lfloor \cdot \rfloor$ denotes the integer part of $\cdot$. Let $h^{(s)} = (h_{1,s}, ..., h_{p,s})$, $s = 0, 1, ..., \mathcal{N} - 1$. For each $h^{(s)}$, we calculate the test statistic in (3.13) and denote it as $\hat{\Gamma}_{NT}\left(h^{(s)}\right)$. Define

$$\sup \hat{\Gamma}_{NT} = \max_{0 \le s \le \mathcal{N}-1} \hat{\Gamma}_{NT}\left(h^{(s)}\right).$$

Even though $\hat{\Gamma}_{NT}\left(h^{(s)}\right)$ is asymptotically distributed as $N(0,1)$ under the null for each s, the distribution of $\sup \hat{\Gamma}_{NT}$ is generally unknown. Fortunately, we can use bootstrap approximation. Based upon the same bootstrap resampling data $\{Y_{it}^*, X_{it}\}$ as in Section 3.3, we construct the bootstrap version $\sup \hat{\Gamma}_{NT}^*$. We repeat this procedure B times and obtain the sequence $\{\sup \hat{\Gamma}_{NT,b}^*\}_{b=1}^B$. We reject the null when $p^* = B^{-1} \sum_{b=1}^B 1(\sup \hat{\Gamma}_{NT,b}^* \ge \sup \hat{\Gamma}_{NT})$ is smaller than the given level of significance.

For the (N,T) pair, we consider $(N,T) = (20,20)$, $(20,40)$, $(20,60)$, $(40,20)$, $(40,40)$, $(40,40)$, $(60,20)$, and $(60,40)$. For each scenario, we use 500 and 250 replications for the size and power studies, respectively, and use 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 $\left(\beta^0, \lambda^0, F^0\right)$ 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 $\hat{\Gamma}^*_{NT}$ and $\sup \hat{\Gamma}^*_{NT}$, based on the bias corrected QMLE estimators.

4.1.3 Test results

Table 1 reports the empirical rejection frequencies of our $\hat{\Gamma}_{NT}$ test at 1%, 5%, and 10% nominal levels for different ROT bandwidth sequences and our $\sup \hat{\Gamma}_{NT}$ test for DGPs 1-6. We summarize some important findings from Table 1. First, when the null hypothesis holds true in DGPs 1-3, 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 and HS's adaptive test procedure seems to yield well-controlled size behavior too. Second, when the null hypothesis does not hold in DGPs 4-6, Table 1 suggests expected power behavior for our test: (i) as either N or T increases, the power of our test generally increases very rapidly; (ii) the choice of bandwidth appears to have some effect on the power of our test and a larger value of c_0 tends to yield a larger testing power; (iii) the power of our test based on HS's adaptive test procedure behaves quite well, is much larger than tests based on ROT bandwidth with $c_0 = 0.5$ and 1, and slightly outperforms tests based on ROT bandwidth with $c_0 = 2$.

Table 2 reports 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 panels (DGP 1h), the levels of our test in the heteroskedastic case oversize in some scenarios. For example, when $(N, T) = (20, 40)$, $(20, 60)$, $(40, 20)$, $c_0 = 1$ 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.

Tables 3 reports the simulation results for DGPs 7-10 when the idiosyncratic errors are serially correlated. For DGPs 7 and 8, the level of our test using the ROT bandwidth with $c_0 = 0.5$ and 1 works reasonably well. However, there is moderate size distortion when the ROT bandwidth with $c_0 = 2$ or the HS's adaptive test procedure is applied. The results for DGPs 9 and 10 indicate that the power of our test still works well when there is serial correlation in the errors.

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 consider the model

$$Y_{it} = m\left(Y_{i,t-1}, ...Y_{i,t-s}, X_{it}\right) + F_t^{0\prime} \lambda_i^0 + \varepsilon_{it}, \tag{4.3}$$

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 λ_i

Table 1: Finite sample rejection frequency for DGPs 1-6 (homoskedastic case: level study for DGPs 1-3 and power study for DGPs 4-6)

DGP	N	T	$c_0 = 0.5$			$c_0 = 1$			$c_0 = 2$			$\sup \hat{\Gamma}_{NT}$		
			1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
1	20	20	0.010	0.054	0.108	0.020	0.068	0.108	0.010	0.070	0.118	0.012	0.068	0.114
	20	40	0.008	0.022	0.084	0.010	0.042	0.086	0.014	0.042	0.086	0.010	0.036	0.076
	20	60	0.014	0.052	0.110	0.012	0.056	0.110	0.010	0.046	0.108	0.016	0.050	0.108
	40	20	0.020	0.064	0.104	0.024	0.072	0.108	0.026	0.064	0.114	0.022	0.062	0.098
	40	40	0.018	0.060	0.106	0.018	0.050	0.110	0.018	0.064	0.120	0.016	0.056	0.108
	40	60	0.014	0.062	0.112	0.016	0.050	0.106	0.014	0.044	0.096	0.016	0.054	0.112
	60	20	0.012	0.048	0.112	0.010	0.058	0.108	0.012	0.060	0.118	0.012	0.068	0.110
	60	40	0.012	0.056	0.100	0.008	0.050	0.088	0.004	0.040	0.092	0.008	0.036	0.082
2	20	20	0.002	0.058	0.108	0.016	0.046	0.094	0.018	0.056	0.108	0.010	0.048	0.106
	20	40	0.014	0.062	0.110	0.012	0.050	0.108	0.016	0.064	0.106	0.014	0.066	0.112
	20	60	0.020	0.046	0.096	0.020	0.042	0.088	0.020	0.042	0.090	0.020	0.046	0.078
	40	20	0.018	0.052	0.094	0.010	0.044	0.102	0.008	0.052	0.108	0.014	0.058	0.104
	40	40	0.010	0.044	0.090	0.006	0.048	0.094	0.006	0.040	0.080	0.008	0.040	0.086
	40	60	0.008	0.040	0.100	0.010	0.060	0.096	0.016	0.064	0.116	0.018	0.048	0.108
	60	20	0.020	0.050	0.106	0.020	0.064	0.102	0.020	0.052	0.122	0.014	0.054	0.098
	60	40	0.016	0.046	0.106	0.012	0.052	0.098	0.010	0.070	0.112	0.014	0.048	0.092
3	20	20	0.010	0.052	0.090	0.016	0.040	0.074	0.006	0.050	0.102	0.012	0.054	0.090
	20	40	0.006	0.046	0.084	0.018	0.058	0.098	0.012	0.058	0.110	0.008	0.054	0.098
	20	60	0.016	0.068	0.110	0.018	0.060	0.116	0.010	0.058	0.126	0.010	0.054	0.120
	40	20	0.024	0.064	0.118	0.008	0.060	0.104	0.012	0.056	0.104	0.010	0.060	0.100
	40	40	0.016	0.062	0.090	0.010	0.062	0.104	0.012	0.052	0.100	0.016	0.060	0.116
	40	60	0.014	0.082	0.138	0.022	0.056	0.112	0.010	0.054	0.112	0.014	0.068	0.126
	60	20	0.012	0.044	0.104	0.006	0.048	0.100	0.008	0.036	0.078	0.006	0.042	0.098
	60	40	0.012	0.056	0.098	0.012	0.046	0.096	0.006	0.056	0.108	0.004	0.050	0.116
4	20	20	0.112	0.268	0.400	0.168	0.372	0.484	0.172	0.420	0.580	0.288	0.484	0.600
	20	40	0.316	0.548	0.676	0.460	0.664	0.780	0.548	0.756	0.868	0.644	0.836	0.896
	20	60	0.532	0.792	0.864	0.676	0.864	0.944	0.752	0.936	0.972	0.884	0.952	0.980
	40	20	0.256	0.544	0.692	0.380	0.700	0.844	0.440	0.784	0.888	0.644	0.808	0.864
	40	40	0.792	0.944	0.972	0.876	0.992	1.000	0.960	1.000	1.000	0.972	1.000	1.000
	40	60	0.936	0.988	0.996	0.984	0.996	1.000	0.992	1.000	1.000	0.992	1.000	1.000
	60	20	0.496	0.788	0.872	0.668	0.876	0.936	0.732	0.916	0.956	0.844	0.916	0.956
	60	40	0.952	1.000	1.000	0.996	1.000	1.000	1.000	1.000	1.000	0.996	1.000	1.000
5	20	20	0.052	0.140	0.220	0.092	0.220	0.332	0.164	0.364	0.464	0.208	0.380	0.472
	20	40	0.076	0.228	0.352	0.156	0.464	0.616	0.280	0.688	0.820	0.344	0.700	0.804
	20	60	0.132	0.352	0.432	0.328	0.576	0.732	0.560	0.844	0.920	0.612	0.876	0.920
	40	20	0.052	0.220	0.316	0.168	0.424	0.576	0.388	0.692	0.816	0.496	0.748	0.824
	40	40	0.212	0.492	0.600	0.580	0.768	0.840	0.702	0.932	0.976	0.744	0.952	0.968
	40	60	0.424	0.692	0.796	0.680	0.912	0.956	0.776	0.992	1.000	0.792	0.996	1.000
	60	20	0.140	0.384	0.492	0.388	0.664	0.768	0.620	0.884	0.916	0.692	0.892	0.924
	60	40	0.372	0.656	0.812	0.760	0.956	0.984	0.904	1.000	1.000	0.932	1.000	1.000
6	20	20	0.012	0.060	0.152	0.032	0.216	0.348	0.100	0.424	0.568	0.228	0.408	0.548
	20	40	0.060	0.196	0.288	0.224	0.448	0.544	0.432	0.780	0.876	0.640	0.812	0.888
	20	60	0.116	0.240	0.352	0.360	0.644	0.760	0.812	0.952	0.972	0.908	0.972	0.984
	40	20	0.080	0.180	0.280	0.176	0.456	0.560	0.348	0.752	0.872	0.524	0.804	0.872
	40	40	0.140	0.348	0.484	0.588	0.844	0.900	0.812	0.988	0.996	0.872	1.000	1.000
	40	60	0.248	0.556	0.664	0.832	0.964	0.984	0.960	1.000	1.000	0.972	1.000	1.000
	60	20	0.092	0.220	0.384	0.312	0.632	0.756	0.572	0.928	0.988	0.708	0.952	0.980
	60	40	0.256	0.480	0.608	0.764	0.968	0.988	0.864	1.000	1.000	0.888	1.000	1.000

Note. For the first three test statistics, the bandwidth is chosen as $h = (h_1, ..., h_p)$ where $h_l = c_0 s_l (NT)^{-1/(4+p)}$ and s_l is the sample standard deviation of the lth element in X_{it}.

Table 2: Finite sample rejection frequency for DGPs 1h-6h (heteroskedastic case: level study for DGPs 1h-3h and power study for DGPs 4h-6h)

DGP	N	T	$c_0 = 0.5$			$c_0 = 1$			$c_0 = 2$			$\sup \hat{\Gamma}_{NT}$		
			1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
1h	20	20	0.018	0.084	0.134	0.018	0.078	0.154	0.028	0.072	0.140	0.018	0.070	0.134
	20	40	0.024	0.080	0.136	0.020	0.072	0.142	0.012	0.072	0.150	0.008	0.056	0.128
	20	60	0.022	0.058	0.124	0.018	0.072	0.124	0.020	0.072	0.128	0.020	0.068	0.112
	40	20	0.008	0.060	0.126	0.010	0.068	0.142	0.014	0.078	0.146	0.012	0.064	0.126
	40	40	0.010	0.074	0.146	0.018	0.092	0.154	0.028	0.088	0.144	0.024	0.082	0.124
	40	60	0.028	0.068	0.122	0.030	0.076	0.126	0.024	0.080	0.126	0.020	0.064	0.122
	60	20	0.022	0.070	0.122	0.020	0.064	0.118	0.022	0.064	0.130	0.020	0.068	0.114
	60	40	0.020	0.066	0.134	0.026	0.062	0.124	0.026	0.064	0.116	0.020	0.058	0.118
2h	20	20	0.006	0.042	0.106	0.010	0.050	0.106	0.014	0.050	0.102	0.014	0.044	0.084
	20	40	0.010	0.056	0.110	0.018	0.060	0.108	0.022	0.060	0.094	0.018	0.072	0.104
	20	60	0.016	0.048	0.102	0.014	0.046	0.088	0.012	0.038	0.096	0.016	0.038	0.090
	40	20	0.012	0.046	0.094	0.014	0.054	0.094	0.016	0.052	0.100	0.016	0.048	0.098
	40	40	0.010	0.054	0.102	0.016	0.056	0.094	0.014	0.042	0.108	0.012	0.034	0.088
	40	60	0.010	0.032	0.106	0.010	0.060	0.094	0.010	0.060	0.102	0.010	0.040	0.108
	60	20	0.014	0.064	0.110	0.018	0.066	0.108	0.020	0.058	0.128	0.010	0.052	0.098
	60	40	0.018	0.048	0.108	0.026	0.056	0.108	0.010	0.054	0.114	0.010	0.056	0.096
3h	20	20	0.008	0.048	0.104	0.016	0.052	0.116	0.004	0.072	0.118	0.014	0.056	0.094
	20	40	0.008	0.058	0.092	0.016	0.044	0.092	0.010	0.070	0.122	0.010	0.036	0.096
	20	60	0.010	0.050	0.106	0.020	0.070	0.126	0.014	0.072	0.126	0.018	0.066	0.106
	40	20	0.016	0.062	0.128	0.016	0.056	0.132	0.018	0.054	0.116	0.016	0.080	0.126
	40	40	0.018	0.046	0.080	0.022	0.058	0.104	0.022	0.056	0.122	0.020	0.056	0.120
	40	60	0.010	0.048	0.098	0.008	0.034	0.086	0.004	0.046	0.094	0.006	0.046	0.090
	60	20	0.010	0.044	0.080	0.006	0.044	0.112	0.006	0.050	0.098	0.006	0.036	0.080
	60	40	0.008	0.046	0.088	0.014	0.054	0.108	0.008	0.052	0.100	0.012	0.048	0.094
4h	20	20	0.184	0.364	0.484	0.304	0.496	0.624	0.376	0.576	0.684	0.436	0.628	0.680
	20	40	0.500	0.704	0.796	0.600	0.808	0.892	0.676	0.888	0.932	0.784	0.904	0.928
	20	60	0.760	0.900	0.920	0.848	0.928	0.960	0.880	0.956	0.976	0.912	0.968	0.980
	40	20	0.436	0.680	0.780	0.556	0.744	0.856	0.624	0.852	0.928	0.764	0.904	0.940
	40	40	0.896	0.956	0.976	0.928	0.980	0.992	0.964	0.988	0.988	0.980	0.996	1.000
	40	60	0.956	0.988	1.000	0.984	1.000	1.000	0.992	1.000	1.000	1.000	1.000	1.000
	60	20	0.712	0.888	0.940	0.788	0.956	0.988	0.848	0.972	0.992	0.912	0.980	0.992
	60	40	0.972	0.992	1.000	0.984	0.996	1.000	0.984	1.000	1.000	0.992	1.000	1.000
5h	20	20	0.216	0.428	0.568	0.484	0.712	0.820	0.696	0.868	0.920	0.688	0.852	0.904
	20	40	0.520	0.764	0.880	0.812	0.960	0.992	0.884	0.984	1.000	0.898	0.988	1.000
	20	60	0.732	0.920	0.968	0.764	0.992	0.996	0.780	1.000	1.000	0.784	1.000	1.000
	40	20	0.576	0.812	0.892	0.880	0.980	0.988	0.940	0.996	1.000	0.940	0.992	1.000
	40	40	0.924	0.996	1.000	0.972	1.000	1.000	0.972	1.000	1.000	0.976	1.000	1.000
	40	60	0.948	1.000	1.000	0.952	1.000	1.000	0.956	1.000	1.000	0.968	1.000	1.000
	60	20	0.776	0.920	0.968	0.908	0.992	1.000	0.928	0.998	1.000	0.936	0.996	1.000
	60	40	0.980	1.000	1.000	0.980	1.000	1.000	0.984	1.000	1.000	1.000	1.000	1.000
6h	20	20	0.124	0.276	0.416	0.424	0.672	0.796	0.652	0.932	0.968	0.720	0.892	0.956
	20	40	0.320	0.544	0.676	0.800	0.944	0.976	0.948	1.000	1.000	0.976	1.000	1.000
	20	60	0.544	0.740	0.840	0.956	1.000	1.000	0.972	1.000	1.000	0.976	1.000	1.000
	40	20	0.360	0.596	0.720	0.768	0.964	0.984	0.828	0.996	1.000	0.848	0.996	1.000
	40	40	0.816	0.940	0.964	0.980	1.000	1.000	0.960	1.000	1.000	0.984	1.000	1.000
	40	60	0.952	0.996	1.000	0.976	1.000	1.000	0.976	1.000	1.000	0.980	1.000	1.000
	60	20	0.596	0.864	0.924	0.832	0.992	1.000	0.848	0.992	1.000	0.872	0.992	1.000
	60	40	0.948	1.000	1.000	0.952	1.000	1.000	0.956	1.000	1.000	0.960	1.000	1.000

Table 3: Finite sample rejection frequency for DGPs 9-10 (serial correlation case: level study for DGPs 7-8 and power study for DGPs 9-10)

DGP	N	T	$c_0 = 0.5$			$c_0 = 1$			$c_0 = 2$			$\sup \hat{\Gamma}_{NT}$		
			1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
7	20	20	0.006	0.040	0.066	0.006	0.046	0.084	0.016	0.056	0.104	0.016	0.062	0.094
	20	40	0.016	0.056	0.100	0.016	0.056	0.084	0.020	0.062	0.120	0.026	0.062	0.128
	20	60	0.020	0.040	0.104	0.028	0.048	0.100	0.026	0.064	0.120	0.028	0.072	0.132
	40	20	0.012	0.060	0.112	0.018	0.072	0.112	0.020	0.074	0.122	0.014	0.056	0.110
	40	40	0.006	0.044	0.098	0.012	0.038	0.086	0.024	0.060	0.100	0.028	0.074	0.114
	40	60	0.020	0.060	0.096	0.018	0.060	0.120	0.024	0.068	0.124	0.024	0.078	0.144
	60	20	0.010	0.044	0.084	0.014	0.038	0.092	0.018	0.050	0.100	0.022	0.050	0.106
	60	40	0.010	0.032	0.102	0.024	0.062	0.110	0.030	0.076	0.128	0.034	0.072	0.106
8	20	20	0.018	0.056	0.094	0.016	0.054	0.114	0.020	0.078	0.138	0.020	0.076	0.124
	20	40	0.012	0.056	0.100	0.001	0.046	0.108	0.014	0.072	0.116	0.014	0.066	0.122
	20	60	0.016	0.038	0.100	0.010	0.038	0.088	0.016	0.058	0.100	0.016	0.048	0.104
	40	20	0.004	0.042	0.094	0.006	0.032	0.074	0.010	0.054	0.090	0.022	0.052	0.108
	40	40	0.006	0.040	0.086	0.010	0.050	0.092	0.024	0.066	0.114	0.020	0.078	0.132
	40	60	0.014	0.058	0.110	0.022	0.074	0.122	0.014	0.064	0.120	0.022	0.074	0.136
	60	20	0.002	0.036	0.082	0.006	0.038	0.086	0.010	0.054	0.104	0.014	0.054	0.094
	60	40	0.006	0.048	0.092	0.018	0.064	0.110	0.030	0.082	0.132	0.036	0.072	0.130
9	20	20	0.036	0.096	0.156	0.068	0.136	0.252	0.100	0.228	0.332	0.136	0.236	0.328
	20	40	0.076	0.192	0.264	0.144	0.328	0.444	0.228	0.464	0.556	0.256	0.452	0.556
	20	60	0.040	0.268	0.384	0.112	0.456	0.608	0.208	0.660	0.748	0.248	0.652	0.720
	40	20	0.064	0.208	0.300	0.140	0.352	0.432	0.228	0.440	0.592	0.284	0.440	0.548
	40	40	0.160	0.396	0.524	0.388	0.608	0.712	0.584	0.860	0.896	0.664	0.848	0.904
	40	60	0.272	0.512	0.624	0.536	0.812	0.848	0.652	0.920	0.958	0.712	0.932	0.952
	60	20	0.048	0.200	0.316	0.212	0.444	0.564	0.400	0.636	0.744	0.432	0.676	0.732
	60	40	0.268	0.564	0.640	0.532	0.820	0.852	0.804	0.948	0.972	0.840	0.972	0.988
10	20	20	0.056	0.112	0.184	0.096	0.196	0.272	0.136	0.296	0.396	0.172	0.300	0.396
	20	40	0.072	0.200	0.292	0.172	0.388	0.488	0.272	0.520	0.616	0.284	0.524	0.636
	20	60	0.084	0.308	0.444	0.164	0.524	0.652	0.264	0.708	0.780	0.284	0.716	0.784
	40	20	0.064	0.216	0.372	0.180	0.368	0.460	0.276	0.556	0.640	0.308	0.500	0.580
	40	40	0.208	0.428	0.540	0.424	0.664	0.776	0.660	0.888	0.928	0.704	0.892	0.924
	40	60	0.320	0.564	0.672	0.564	0.828	0.880	0.692	0.960	0.984	0.764	0.964	0.976
	60	20	0.088	0.260	0.352	0.272	0.484	0.616	0.428	0.700	0.804	0.496	0.724	0.788
	60	40	0.344	0.612	0.700	0.640	0.852	0.908	0.880	0.972	1.000	0.920	0.988	1.000

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 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 tth 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) confirms this point empirically. The endogenous growth model predicts a positive relation and the argument is reinforced by Bond et al. (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 (2013) 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 (2013) who consider a *short* panel with *additive* fixed effects. Our model incorporates cross sectional dependence and allows for IFE 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 et al. (2010), we exclude oil production countries and Botswana, because of the dominant role of mining. We also drop Nicaragua and Chad for their negative record of gross investment in some years. China has two versions of data and we choose 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 (4.3), and dynamic models with 1-3 lags, and $X_{i,1}$, $X_{it,2}$, or both as exogenous regressors in (4.3). Therefore we have the following twelve models in total.

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

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:[4]

$$\begin{aligned} PC_{p1}(R) &= V\left(R,\hat{F}^R\right)+R\hat{\sigma}^2\left(\frac{N+T}{NT}\right)\ln\left(\frac{NT}{N+T}\right), \\ PC_{p2}(R) &= V\left(R,\hat{F}^R\right)+R\hat{\sigma}^2\left(\frac{N+T}{NT}\right)\ln C_{NT}^2, \\ IC_{p1}(R) &= \ln\left(V\left(R,\hat{F}^R\right)\right)+R\left(\frac{N+T}{NT}\right)\ln\left(\frac{NT}{N+T}\right), \\ IC_{p2}(R) &= \ln\left(V\left(R,\hat{F}^R\right)\right)+R\left(\frac{N+T}{NT}\right)\ln C_{NT}^2, \end{aligned}$$

where $C_{NT}^2=\min\{N,T\}$, $V(R,\hat{F}^R)=\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T(\hat{\varepsilon}_{it}^R)^2$, $\hat{\varepsilon}_{it}^R=Y_{it}-X_{it}'\hat{\beta}^R-\hat{F}_t^{R\prime}\hat{\lambda}_i^R$, $\hat{\beta}^R$, $\hat{F}_t^R$ and $\hat{\lambda}_i^R$ are estimates under the null hypothesis of linearity when R factors are used, and $\hat{\sigma}^2$ is a consistent estimate of $\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T E(\varepsilon_{it}^2)$ and can be replaced by $V(R_{\max},\hat{F}^{R_{\max}})$ in applications. Following Bai and Ng (2002) we set $R_{\max}$ to be 8, 10 and 15, and recognize explicitly that $PC_{p1}(R)$ and $PC_{p2}(R)$ depend on the choice of $R_{\max}$ 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 $R_{\max}$. 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 $R_{\max}=8$, both PC_{p1} and PC_{p2} suggest R^* to be 7 and both IC_{p1} and IC_{p2} suggest 1 when $R_{\max}=10$, PC_{p1} and PC_{p2} suggest 5, and IC_{p1} and IC_{p2} suggest 1 when $R_{\max}=15$. So our choice of R^* will be 1 for Model 4.

Table 4 presents the number of factors determined for each model by using the above procedure and the bootstrap p-values for our linearity test based on the ROT bandwidth and HS's adaptive test procedure. For the purpose of comparison, we fix the list of 104 countries that have observations during the time period 1960-2009 and consider the test results by varying the time periods from 1960-2009 to 1970-2009 and 1980-2009, respectively. Table 4 reports the bootstrap p-values based on 1000 bootstrap resamples. For the time period 1960-2009, 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. Interestingly, for the time periods 1970-2009 and 1980-2009, Bai and Ng's (2002) information criteria tend to choose three or four factors in many scenarios; the bootstrap p-values are all very small except for Model 1 in the period 1970-2009. So in general we find strong evidence of nonlinearity in the panel data.

To conduct a robustness check, we do the same analysis using different sample periods for different sets of countries available in PWT 7.1. Table 5 presents the corresponding bootstrap p-values for our

[4] 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.

linearity test based on the ROT bandwidth and HS's adaptive test procedure. There are $N = 52$, 104, 147, and 148 countries in PWT 7.1 that have observations for the periods 1950-2009, 1960-2009, 1970-2009, and 1980-2009, respectively. The results for the period 1960-2009 were reported above. So Table 5 only reports the bootstrap p-values based on 1000 bootstrap resamples for the other three periods in conjunction with the number of factors determined by Bai and Ng's (2002) information criteria. The bootstrap p-values are very small in most cases in Table 5 except Models 4, 9, 10, and 12 for the period 1950-2009. In these cases, we are not able to reject the null of linearity at the 5% level for some choices of bandwidths. Nevertheless, if we uses the $\sup \hat{\Gamma}_{NT}$ statistic, we fail to reject the null of linearity at 5% level only for Models 4 and 12 for the period 1950-2009. In addition, when $N = 52$ is small in Table 5, Bai and Ng's method tends to yield a larger number of factors than when N is large. In sum, our results are generally in favor of strong degree of nonlinearity in the panel dataset.

5 Concluding remarks

In this paper we propose a nonparametric consistent test for the correct specification of linear panel data models with IFE. 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.

We only consider homogenous panel data models in this paper. As a referee kindly remarks, the assumption of common regression functions may be inappropriate in some applications. In this case, one can consider panel data models with heterogenous function forms $m_i(\cdot)$ and then test whether the commonly used heterogenous linear specification is correct or not; that is, the null hypothesis is

$$\mathbb{H}_0 : m_i(X_{it}) = X_{it}'\beta_i^0 \text{ a.s. for some } \beta_i^0 \in \mathbb{R}^p \text{ and for all } i = 1, ..., N.$$

Given the very recent contributions by Chudik and Pesaran (2013) and Song (2013) in linear dynamic panel data models with IFE, one can obtain estimates of the heterogenous slopes under the above null restrictions and then extend the asymptotic theory in the current paper to this framework. We leave this for future research.

Table 4: Bootstrap p-values for the application to economic growth data (1960-2009, 1970-2009, 1980-2009, N=104)

	Number of factors	$c_0 = 0.5$	$c_0 = 1$	$c_0 = 2$	$\sup \hat{\Gamma}_{NT}$
1960-2009					
Model 1	2	0.006	0.004	0.005	0.005
Model 2	1	0.000	0.000	0.000	0.000
Model 3	1	0.000	0.000	0.000	0.000
Model 4	1	0.000	0.000	0.000	0.000
Model 5	2	0.022	0.025	0.030	0.037
Model 6	1	0.000	0.000	0.000	0.000
Model 7	1	0.000	0.000	0.000	0.000
Model 8	1	0.000	0.000	0.000	0.000
Model 9	2	0.041	0.040	0.054	0.062
Model 10	1	0.000	0.000	0.000	0.000
Model 11	1	0.000	0.000	0.000	0.000
Model 12	1	0.000	0.000	0.000	0.000
1970-2009					
Model 1	1	0.224	0.207	0.218	0.268
Model 2	1	0.000	0.000	0.000	0.000
Model 3	1	0.000	0.000	0.000	0.000
Model 4	1	0.000	0.000	0.000	0.001
Model 5	4	0.008	0.008	0.008	0.009
Model 6	2	0.000	0.000	0.000	0.000
Model 7	1	0.000	0.000	0.000	0.001
Model 8	2	0.001	0.001	0.002	0.003
Model 9	4	0.011	0.017	0.027	0.026
Model 10	3	0.000	0.001	0.003	0.005
Model 11	3	0.000	0.000	0.000	0.000
Model 12	3	0.003	0.002	0.005	0.004
1980-2009					
Model 1	3	0.004	0.004	0.004	0.005
Model 2	3	0.008	0.007	0.009	0.010
Model 3	3	0.010	0.009	0.010	0.010
Model 4	3	0.010	0.010	0.011	0.011
Model 5	3	0.013	0.013	0.013	0.014
Model 6	3	0.003	0.005	0.005	0.010
Model 7	3	0.002	0.005	0.007	0.008
Model 8	3	0.008	0.008	0.009	0.009
Model 9	3	0.006	0.007	0.013	0.012
Model 10	3	0.001	0.003	0.006	0.008
Model 11	3	0.003	0.003	0.005	0.007
Model 12	4	0.087	0.066	0.086	0.086

Table 5: Bootstrap p-values for the application to economic growth data (1950-2009, N=52; 1970-2009, N=147; 1980-2009, N=148)

	Number of factors	$c_0 = 0.5$	$c_0 = 1$	$c_0 = 2$	$\sup \hat{\Gamma}_{NT}$
1950-2009 $(N = 47)$					
Model 1	2	0.030	0.029	0.178	0.057
Model 2	1	0.036	0.014	0.011	0.017
Model 3	1	0.036	0.038	0.044	0.066
Model 4	1	0.166	0.108	0.158	0.174
Model 5	2	0.030	0.025	0.050	0.041
Model 6	1	0.000	0.000	0.000	0.000
Model 7	1	0.009	0.019	0.016	0.020
Model 8	1	0.033	0.000	0.000	0.000
Model 9	2	0.014	0.042	0.130	0.039
Model 10	1	0.225	0.019	0.014	0.017
Model 11	1	0.062	0.027	0.010	0.018
Model 12	1	0.136	0.071	0.091	0.114
1970-2009 $(N = 147)$					
Model 1	1	0.000	0.000	0.000	0.000
Model 2	1	0.000	0.000	0.000	0.000
Model 3	1	0.000	0.000	0.000	0.000
Model 4	1	0.000	0.000	0.000	0.000
Model 5	4	0.000	0.000	0.000	0.000
Model 6	2	0.000	0.000	0.000	0.000
Model 7	1	0.000	0.000	0.000	0.000
Model 8	2	0.000	0.000	0.000	0.000
Model 9	4	0.000	0.000	0.000	0.000
Model 10	3	0.000	0.000	0.000	0.000
Model 11	3	0.000	0.000	0.000	0.000
Model 12	3	0.000	0.000	0.000	0.000
1980-2009 $(N = 148)$					
Model 1	3	0.000	0.000	0.000	0.000
Model 2	3	0.000	0.000	0.000	0.000
Model 3	3	0.000	0.000	0.000	0.000
Model 4	3	0.000	0.000	0.000	0.000
Model 5	3	0.000	0.000	0.000	0.000
Model 6	3	0.000	0.000	0.000	0.000
Model 7	3	0.000	0.000	0.000	0.000
Model 8	3	0.000	0.000	0.000	0.000
Model 9	3	0.000	0.000	0.000	0.000
Model 10	3	0.000	0.000	0.000	0.000
Model 11	3	0.000	0.000	0.000	0.000
Model 12	4	0.000	0.000	0.000	0.000

APPENDIX

Let C signify a generic constant whose exact value may vary from case to case. Let $\lfloor a \rfloor$ denote the integer part of real number a. Let $\delta_{NT} \equiv \min(\sqrt{N}, \sqrt{T})$. Let $E_{\mathcal{D}}(\cdot)$ and $\text{Var}_{\mathcal{D}}(\cdot)$ denote the conditional expectation and variance given $\mathcal{D} \equiv \{F^0, \lambda^0\}$, respectively. Let $\alpha_{ik} \equiv \lambda_i^{0\prime} \left(\lambda^{0\prime}\lambda^0/N\right)^{-1} \lambda_k^0$ and $\eta_{ts} \equiv F_t^{0\prime} \left(F^{0\prime}F^0/T\right)^{-1} F_s^0$. Let $\Phi_1 \equiv \lambda^0 \left(\lambda^{0\prime}\lambda^0\right)^{-1} \left(F^{0\prime}F^0\right)^{-1} F^{0\prime}$, $\Phi_2 \equiv F^0 \left(F^{0\prime}F^0\right)^{-1} \left(\lambda^{0\prime}\lambda^0\right)^{-1} \left(F^{0\prime}F^0\right)^{-1} F^{0\prime}$, and $\Phi_3 \equiv \lambda^0 \left(\lambda^{0\prime}\lambda^0\right)^{-1} \left(F^{0\prime}F^0\right)^{-1} \left(\lambda^{0\prime}\lambda^0\right)^{-1} \lambda^{0\prime}$.

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 $\mathbf{X}_0 \equiv (\sqrt{NT}/\|\mathbf{e}\|)\mathbf{e}$, $\epsilon_0 \equiv \|\mathbf{e}\|/\sqrt{NT}$, and $\epsilon_k \equiv \beta_k^0 - \beta_k$ for $k = 1, ..., p$. Note that under $\mathbb{H}_1(\gamma_{NT})$, conditions (A.6) and (A.7) in **MW** continue to hold for sufficiently large (N, T) as

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

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

$$\begin{aligned} L_{NT}(\beta) &= \frac{1}{NT} \sum_{k_1=0}^{p} \sum_{k_2=0}^{p} \epsilon_{k_1}\epsilon_{k_2} L^{(2)}\left(\lambda^0, F^0, \mathbf{X}_{k_1}, \mathbf{X}_{k_2}\right) \\ &\quad + \frac{1}{NT} \sum_{k_1=0}^{p} \sum_{k_2=0}^{p} \sum_{k_3=0}^{p} \epsilon_{k_1}\epsilon_{k_2}\epsilon_{k_3} L^{(3)}\left(\lambda^0, F^0, \mathbf{X}_{k_1}, \mathbf{X}_{k_2}, \mathbf{X}_{k_3}\right) + O_P\left(v_{1NT}^4\right), \end{aligned}$$

where for any integer $g \geq 1$,

$$\begin{aligned} L^{(g)}\left(\lambda^0, F^0, \mathbf{X}_{k_1}, ..., \mathbf{X}_{k_g}\right) &= \frac{1}{g!} \sum_{\text{all } g! \text{ permutations of } (k_1, ..., k_g)} \tilde{L}^{(g)}\left(\lambda^0, F^0, \mathbf{X}_{k_1}, ..., \mathbf{X}_{k_g}\right), \\ \tilde{L}^{(g)}\left(\lambda^0, F^0, \mathbf{X}_{k_1}, ..., \mathbf{X}_{k_g}\right) &= \sum_{l=1}^{g} (-1)^{l+1} \sum_{\substack{v_1+v_2+\cdots+v_l=g \\ m_1+\cdots+m_{l+1}=l-1 \\ 2 \geq v_j \geq 1,\ m_j \geq 0}} \text{tr}\left\{S^{(m_1)}\mathcal{T}_{k_1\cdots}^{(v_1)} S^{(m_2)} \cdots S^{(m_l)} \mathcal{T}_{\cdots k_g}^{(v_l)} S^{(m_{l+1})}\right\}, \end{aligned}$$

$S^{(0)} = -M_{\lambda^0}$, $S^{(m)} = \Phi_3^m$, $\mathcal{T}_k^{(1)} = \lambda^0 F^{0\prime}\mathbf{X}_k' + \mathbf{X}_k F^0 \lambda^{0\prime}$ for $k = 0, 1, ..., p$, and $\mathcal{T}_{k_1k_2}^{(2)} = \mathbf{X}_{k_1}\mathbf{X}_{k_2}'$ for $k_1, k_2 = 0, 1, ..., p$.[5] By straightforward calculations, one verifies that

$$\begin{aligned} &\tilde{L}^{(2)}\left(\lambda^0, F^0, \mathbf{X}_{k_1}, \mathbf{X}_{k_2}\right) \\ = \ &\text{tr}\left\{S^{(0)}\mathcal{T}_{k_1k_2}^{(2)}S^{(0)} - \left[S^{(1)}\mathcal{T}_{k_1}^{(1)}S^{(0)}\mathcal{T}_{k_2}^{(1)}S^{(0)} + S^{(0)}\mathcal{T}_{k_1}^{(1)}S^{(1)}\mathcal{T}_{k_2}^{(1)}S^{(0)} + S^{(0)}\mathcal{T}_{k_1}^{(1)}S^{(0)}\mathcal{T}_{k_2}^{(1)}S^{(1)}\right]\right\} \\ = \ &\text{tr}\left(M_{\lambda^0}\mathbf{X}_{k_1}\mathbf{X}_{k_2}'M_{\lambda^0} - M_{\lambda^0}\mathbf{X}_{k_1}F^0\lambda^{0\prime}\Phi_3\lambda^0F^{0\prime}\mathbf{X}_{k_2}'M_{\lambda^0}\right) \\ = \ &\text{tr}\left(M_{\lambda^0}\mathbf{X}_{k_1}M_{F^0}\mathbf{X}_{k_2}'M_{\lambda^0}\right) \end{aligned}$$

[5] The subscript indices in $\mathcal{T}_{k_1\cdots}^{(v_1)}$ or $\mathcal{T}_{\cdots k_g}^{(v_l)}$ may contain either one (e.g., k_1 or k_g) or two elements (e.g., (k_1, k_2) or (k_{g-1}, k_g)) depending on whether v_1 or v_l takes value 1 or 2.

where we use the fact that $S^{(0)}\mathcal{T}_k^{(1)}S^{(0)}=0$ and $F^0\lambda^{0\prime}\Phi_3\lambda^0F^{0\prime}=P_{F^0}$. Similarly,

$$\begin{aligned}
&\tilde{L}^{(3)}\left(\lambda^0,F^0,\mathbf{X}_{k_1},\mathbf{X}_{k_2},\mathbf{X}_{k_3}\right)\\
=\;&\mathrm{tr}\{-[S^{(1)}\mathcal{T}_{k_1}^{(1)}S^{(0)}\mathcal{T}_{k_2k_3}^{(2)}S^{(0)}+S^{(0)}\mathcal{T}_{k_1}^{(1)}S^{(1)}\mathcal{T}_{k_2k_3}^{(2)}S^{(0)}+S^{(0)}\mathcal{T}_{k_1k_2}^{(2)}S^{(1)}\mathcal{T}_{k_3}^{(1)}S^{(0)}+S^{(0)}\mathcal{T}_{k_1k_2}^{(2)}S^{(0)}\mathcal{T}_{k_3}^{(1)}S^{(1)}]\\
&+[S^{(1)}\mathcal{T}_{k_1}^{(1)}S^{(0)}\mathcal{T}_{k_2}^{(1)}S^{(1)}\mathcal{T}_{k_3}^{(1)}S^{(0)}+S^{(0)}\mathcal{T}_{k_1}^{(1)}S^{(1)}\mathcal{T}_{k_2}^{(1)}S^{(1)}\mathcal{T}_{k_3}^{(1)}S^{(0)}+S^{(0)}\mathcal{T}_{k_1}^{(1)}S^{(1)}\mathcal{T}_{k_2}^{(1)}S^{(0)}\mathcal{T}_{k_3}^{(1)}S^{(1)}]\}\\
=\;&\mathrm{tr}\{-[\Phi_1\mathbf{X}_{k_1}'M_{\lambda^0}\mathbf{X}_{k_2}\mathbf{X}_{k_3}'M_{\lambda^0}+M_{\lambda^0}\mathbf{X}_{k_1}\Phi_1'\mathbf{X}_{k_2}\mathbf{X}_{k_3}'+M_{\lambda^0}\mathbf{X}_{k_1}\mathbf{X}_{k_2}'\Phi_1\mathbf{X}_{k_3}'+M_{\lambda^0}\mathbf{X}_{k_1}\mathbf{X}_{k_2}'M_{\lambda^0}\mathbf{X}_{k_3}\Phi_1']\\
&+\Phi_1\mathbf{X}_{k_1}'M_{\lambda^0}\mathbf{X}_{k_2}P_{F^0}\mathbf{X}_{k_3}'M_{\lambda^0}+M_{\lambda^0}\mathbf{X}_{k_1}\left(\Phi_1'\mathbf{X}_{k_2}P_{F^0}+P_{F^0}\mathbf{X}_{k_2}'\Phi_1\right)\mathbf{X}_{k_3}'+M_{\lambda^0}\mathbf{X}_{k_1}P_{F^0}\mathbf{X}_{k_2}'M_{\lambda^0}\mathbf{X}_{k_3}\Phi_1'\}\\
=\;&-\mathrm{tr}\left(M_{\lambda^0}\mathbf{X}_{k_1}\Phi_1'\mathbf{X}_{k_2}M_{F^0}\mathbf{X}_{k_3}'+M_{\lambda^0}\mathbf{X}_{k_1}M_{F^0}\mathbf{X}_{k_2}'\Phi_1\mathbf{X}_{k_3}'\right)
\end{aligned}$$

where we use the additional fact that $\Phi_3\mathcal{T}_k^{(1)}M_{\lambda^0}=\Phi_1\mathbf{X}_k'M_{\lambda^0}$, $M_{\lambda^0}\mathcal{T}_k^{(1)}\Phi_3=M_{\lambda^0}\mathbf{X}_k\Phi_1'$, $M_{\lambda^0}\mathcal{T}_k^{(1)}\Phi_1=M_{\lambda^0}\mathbf{X}_kP_{F^0}$, $\Phi_1'\mathcal{T}_k^{(1)}M_{\lambda^0}=P_{F^0}\mathbf{X}_k'M_{\lambda^0}$, $\Phi_1'\mathcal{T}_k^{(1)}\Phi_1=P_{F^0}\mathbf{X}_k'\Phi_1+\Phi_1'\mathbf{X}_kP_{F^0}$, and that $M_{\lambda^0}\Phi_1=0$. It follows that

$$\begin{aligned}
L^{(2)}\left(\lambda^0,F^0,\mathbf{X}_{k_1},\mathbf{X}_{k_2}\right)&=\mathrm{tr}\left(M_{\lambda^0}\mathbf{X}_{k_1}M_{F^0}\mathbf{X}_{k_2}'M_{\lambda^0}\right)=\mathrm{tr}\left(M_{\lambda^0}\mathbf{X}_{k_1}M_{F^0}\mathbf{X}_{k_2}'\right),\text{ and}\\
L^{(3)}\left(\lambda^0,F^0,\mathbf{X}_{k_1},\mathbf{X}_{k_2},\mathbf{X}_{k_3}\right)&=-\frac{1}{3}\sum_{\text{all 6 permutations of }(k_1,k_2,k_3)}\mathrm{tr}\left(M_{\lambda^0}\mathbf{X}_{k_1}M_{F^0}\mathbf{X}_{k_2}'\Phi_1\mathbf{X}_{k_3}'M_{\lambda^0}\right).
\end{aligned}$$

Furthermore, we have

$$\mathcal{L}_{NT}(\beta)=\mathcal{L}_{NT}\left(\beta^0\right)+L_{1NT}(\beta)+L_{2NT}(\beta)+R_{NT}+O_P\left(\upsilon_{1NT}^4-\epsilon_0^4\right)$$

where

$$\begin{aligned}
L_{1NT}(\beta)&\equiv\frac{2}{NT}\sum_{k=1}^p\epsilon_k\epsilon_0L^{(2)}\left(\lambda^0,F^0,\mathbf{X}_k,\mathbf{X}_0\right)+\frac{3}{NT}\sum_{k=1}^p\epsilon_k\epsilon_0\epsilon_0L^{(3)}\left(\lambda^0,F^0,\mathbf{X}_k,\mathbf{X}_0,\mathbf{X}_0\right),\\
L_{2NT}(\beta)&\equiv\frac{1}{NT}\sum_{k_1=1}^p\sum_{k_2=1}^p\epsilon_{k_1}\epsilon_{k_2}L^{(2)}\left(\lambda^0,F^0,\mathbf{X}_{k_1},\mathbf{X}_{k_2}\right),\\
R_{NT}(\beta)&\equiv\frac{3}{NT}\sum_{k_1=1}^p\sum_{k_2=1}^p\epsilon_{k_1}\epsilon_{k_2}\epsilon_0L^{(3)}\left(\lambda^0,F^0,\mathbf{X}_{k_1},\mathbf{X}_{k_2},\mathbf{X}_0\right)\\
&\quad+\frac{1}{NT}\sum_{k_1=1}^p\sum_{k_2=1}^p\sum_{k_3=1}^p\epsilon_{k_1}\epsilon_{k_2}\epsilon_{k_3}L^{(3)}\left(\lambda^0,F^0,\mathbf{X}_{k_1},\mathbf{X}_{k_2},\mathbf{X}_{k_3}\right).
\end{aligned}$$

Clearly, L_{1NT} and L_{2NT} are linear and quadratic in $\epsilon_k=\beta_k^0-\beta_k$, $k=1,...,p$, respectively, and R_{NT} reflects the terms in the third order likelihood expansion that are asymptotically negligible (argued below). Noting that $L^{(g)}\left(\lambda^0,F^0,\mathbf{X}_{k_1},\mathbf{X}_{k_2},...,\mathbf{X}_{k_g}\right)$ is linear in the last g elements and $\epsilon_0\mathbf{X}_0=\mathbf{e}$, we have

$$\begin{aligned}
L_{1NT}(\beta)&=\frac{2}{NT}\sum_{k=1}^p\epsilon_k\left[L^{(2)}\left(\lambda^0,F^0,\mathbf{X}_k,\mathbf{e}\right)+\frac{3}{2}L^{(3)}\left(\lambda^0,F^0,\mathbf{X}_k,\mathbf{e},\mathbf{e}\right)\right]\\
&=\frac{2}{NT}\sum_{k=1}^p\epsilon_k\left[\mathrm{tr}\left(M_{\lambda^0}\mathbf{X}_kM_{F^0}\mathbf{e}'\right)-\frac{1}{2}\sum_{\text{all 6 permutations of }(\mathbf{X}_k,\mathbf{e},\mathbf{e})}\mathrm{tr}\left(M_{\lambda^0}\mathbf{X}_kM_{F^0}\mathbf{e}'\Phi_1\mathbf{e}'\right)\right]\\
&=-2\left(\beta-\beta^0\right)'\left(C_{NT}^{(1)}+C_{NT}^{(2)}\right)
\end{aligned}$$

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

$$L_{2NT}(\beta) = \frac{1}{NT} \sum_{k_1=1}^{p} \sum_{k_2=1}^{p} \epsilon_{k_1} \epsilon_{k_2} \operatorname{tr}\left(M_{\lambda^0} \mathbf{X}_{k_1} M_{F^0} \mathbf{X}_{k_2}'\right) = \left(\beta - \beta^0\right)' D_{NT}\left(\beta - \beta^0\right)$$

where D_{NT} is defined in (3.3). As in **MW**, noticing that

$$\frac{1}{NT}(\epsilon_0)^{g-r} L^{(g)}\left(\lambda^0, F^0, \mathbf{X}_{k_1}, ..., \mathbf{X}_{k_r}, \mathbf{X}_0 ..., \mathbf{X}_0\right) = O_P\left(\left(\|\mathbf{e}\| / \sqrt{NT}\right)^{g-r}\right) = O_P\left(\left(\delta_{NT}^{-1} + \gamma_{NT}\right)^{g-r}\right),$$

we can readily determine the probability order of R_{NT} : $R_{NT} = O_P\left(\left\|\beta - \beta^0\right\|^2\left(\delta_{NT}^{-1} + \gamma_{NT}\right) + \left\|\beta - \beta^0\right\|^3\right)$. It follows that

$$L_{NT}(\beta) = L_{NT}\left(\beta^0\right) - 2\gamma_{NT}\left(\beta - \beta^0\right)'\left(C_{NT}^{(1)} + C_{NT}^{(2)}\right) + \left(\beta - \beta^0\right)' D_{NT}\left(\beta - \beta^0\right) + \tilde{R}_{NT}(\beta) \quad \text{(A.1)}$$

where

$$\tilde{R}_{NT}(\beta) = O_P\left\{\left\|\beta - \beta^0\right\|^2\left(\delta_{NT}^{-1} + \gamma_{NT}\right) + \left\|\beta - \beta^0\right\|^3 + \left\|\beta - \beta^0\right\|\left(\delta_{NT}^{-3} + \gamma_{NT}^3\right)\right\} \quad \text{(A.2)}$$

by the fact that

$$\begin{aligned} v_{1NT}^4 - \epsilon_0^4 &= \left(\sum_{k=1}^{p} \left|\beta_k^0 - \beta_k\right| \frac{\|\mathbf{X}_k\|}{\sqrt{NT}} + \frac{\|\mathbf{e}\|}{\sqrt{NT}}\right)^4 - \left(\frac{\|\mathbf{e}\|}{\sqrt{NT}}\right)^4 \\ &= O_P\left(\left\|\beta - \beta^0\right\|^4 + \left\|\beta - \beta^0\right\|^3 \frac{\|\mathbf{e}\|}{\sqrt{NT}} + \left\|\beta - \beta^0\right\|^2 \frac{\|\mathbf{e}\|^2}{NT} + \left\|\beta - \beta^0\right\|\left(\frac{\|\mathbf{e}\|}{\sqrt{NT}}\right)^3\right) \end{aligned}$$

$\left\|\beta - \beta^0\right\| = o(1)$, and that $\|\mathbf{e}\| / \sqrt{NT} = O_P\left(\delta_{NT}^{-1} + \gamma_{NT}\right) = o_P(1)$.

Under Assumptions A.1-A.3(i), we can readily show that under $\mathbb{H}_1(\gamma_{NT})$,

$$\begin{aligned} C_{NT,k}^{(1)} &= \frac{\gamma_{NT}}{NT} \operatorname{tr}\left(M_{\lambda^0} \mathbf{X}_k M_{F^0} \boldsymbol{\Delta}'\right) + \frac{1}{NT} \operatorname{tr}\left(M_{\lambda^0} \mathbf{X}_k M_{F^0} \boldsymbol{\varepsilon}'\right) \\ &= \gamma_{NT}(NT)^{-1} \operatorname{tr}\left(M_{\lambda^0} \mathbf{X}_k M_{F^0} \boldsymbol{\Delta}'\right) + O_P\left(\delta_{NT}^{-2}\right) = O_P(\gamma_{NT}) + O_P\left(\delta_{NT}^{-2}\right), \end{aligned}$$

and similarly $C_{NT,k}^{(2)} = O_P\left(\gamma_{NT}^2 + \delta_{NT}^{-2}\right)$. Let $\vartheta_{NT} \equiv D_{NT}^{-1}(C_{NT}^{(1)} + C_{NT}^{(2)})$ where D_{NT} is asymptotically invertible by Assumption A.1(v). In view of the fact that $L_{NT}(\hat{\beta}) \leq L_{NT}\left(\beta^0 + \vartheta_{NT}\right)$, we apply (A.1) to the objects on both sides to obtain

$$\begin{aligned} & \left(\hat{\beta} - \beta^0 - \vartheta_{NT}\right)' D_{NT}\left(\hat{\beta} - \beta^0 - \vartheta_{NT}\right) \\ \leq\ & \tilde{R}_{NT}\left(\beta^0 + \vartheta_{NT}\right) - \tilde{R}_{NT}\left(\hat{\beta}\right) \\ =\ & O_P\left\{\left(\gamma_{NT}^2 + \delta_{NT}^{-4}\right)\left(\delta_{NT}^{-1} + \gamma_{NT}\right) + \left(\gamma_{NT}^3 + \delta_{NT}^{-6}\right) + \left(\gamma_{NT} + \delta_{NT}^{-2}\right)\left(\delta_{NT}^{-3} + \gamma_{NT}^3\right)\right\} - \tilde{R}_{NT}\left(\hat{\beta}\right) \\ =\ & O_P\left[\gamma_{NT}^2\left(\delta_{NT}^{-1} + \gamma_{NT}\right) + \gamma_{NT}\delta_{NT}^{-3} + \delta_{NT}^{-5}\right] - \tilde{R}_{NT}\left(\hat{\beta}\right) \end{aligned}$$

where the first equality follows from (A.2) and the fact that $\|\vartheta_{NT}\| = O_P(\gamma_{NT} + \delta_{NT}^{-2})$. Then by Assumption A.1(v) and the bound on $\tilde{R}_{NT}(\hat{\beta})$ via (A.2) we can readily show that $\hat{\beta} - \beta^0 = \vartheta_{NT} + O_P\{\left[\gamma_{NT}^2\left(\delta_{NT}^{-1} + \gamma_{NT}\right) + \gamma_{NT}\delta_{NT}^{-3} + \delta_{NT}^{-5}\right]^{1/2}\}$ by contradiction. This completes the proof of the theorem. ■

B Proof of Theorem 3.2

Following **MW**, we can readily show that

$$M_{\hat{F}} = M_{F^0} + \sum_{k=1}^{p} \left(\beta_k^0 - \hat{\beta}_k\right) M_k^{(0)} + M^{(1)} + M^{(2)} + M^{(rem)}, \tag{B.1}$$

where

$$\begin{aligned}
M_k^{(0)} &= -M_{F^0}\mathbf{X}_k'\Phi_1 - \Phi_1'\mathbf{X}_k M_{F^0} \text{ for } k = 1,...,p, \\
M^{(1)} &= -M_{F^0}\boldsymbol{\varepsilon}'\Phi_1 - \Phi_1'\boldsymbol{\varepsilon}M_{F^0}, \\
M^{(2)} &= M_{F^0}\boldsymbol{\varepsilon}'\Phi_1\boldsymbol{\varepsilon}'\Phi_1 + \Phi_1'\boldsymbol{\varepsilon}\Phi_1'\boldsymbol{\varepsilon}M_{F^0} - M_{F^0}\boldsymbol{\varepsilon}'M_{\lambda^0}\boldsymbol{\varepsilon}\Phi_2 - \Phi_2\boldsymbol{\varepsilon}'M_{\lambda^0}\boldsymbol{\varepsilon}M_{F^0} - M_{F^0}\boldsymbol{\varepsilon}'\Phi_3\boldsymbol{\varepsilon}M_{F^0} + \Phi_1'\boldsymbol{\varepsilon}M_{F^0}\boldsymbol{\varepsilon}'\Phi_1,
\end{aligned}$$

and the remainder $M^{(rem)}$ satisfies

$$\begin{aligned}
\left\|M^{(rem)}\right\|_F &= O_P\left(\left(\delta_{NT}^{-1} + \gamma_{NT} + \left\|\hat{\beta} - \beta^0\right\|\right)\left\|\hat{\beta} - \beta^0\right\| + (NT)^{-3/2}\max\left(\sqrt{N}, \sqrt{T}\right)^3 + \gamma_{NT}^3\right) \\
&= O_P\left(\delta_{NT}^{-1}\gamma_{NT} + \delta_{NT}^{-3}\right) = O_P\left(\delta_{NT}^{-1}\gamma_{NT}\right) \qquad \text{(B.2)}
\end{aligned}$$

by (3.7), (3.8), and Assumption A.7(i). It is straightforward to show that

$$\left\|M_k^{(0)}\right\|_F = O_P(1) \text{ for } k = 1,...,p, \quad \left\|M^{(1)}\right\|_F = O_P\left(N^{-1/2}\right), \text{ and } \left\|M^{(2)}\right\|_F = O_P\left(\delta_{NT}^{-2}\right). \tag{B.3}$$

Combining (B.1) with (2.14) yields

$$\begin{aligned}
\hat{\varepsilon}_i &= M_{F^0}(\varepsilon_i + c_i) + \sum_{k=1}^{p}\left(\beta_k^0 - \hat{\beta}_k\right) M_k^{(0)}\left(\varepsilon_i + F^0\lambda_i^0 + c_i\right) + M^{(1)}\left(\varepsilon_i + F^0\lambda_i^0 + c_i\right) \\
&\quad + \left(M^{(2)} + M^{(rem)}\right)\left(\varepsilon_i + F^0\lambda_i^0 + c_i\right) \\
&\equiv d_{1i} + d_{2i} + d_{3i} + d_{4i}, \text{ say,} \qquad \text{(B.4)}
\end{aligned}$$

where $c_i \equiv X_i(\beta^0 - \hat{\beta}) + \left(m_i - X_i\beta^0\right)$. It follows that

$$\begin{aligned}
NT(h!)^{1/2}\hat{J}_{NT} &= a_{NT}\sum_{1\le i,j\le N}(d_{1i} + d_{2i} + d_{3i} + d_{4i})'\mathcal{K}_{ij}(d_{1j} + d_{2j} + d_{3j} + d_{4j}) \\
&= a_{NT}\sum_{1\le i,j\le N}\{d_{1i}'\mathcal{K}_{ij}d_{1j} + d_{2i}'\mathcal{K}_{ij}d_{2j} + d_{3i}'\mathcal{K}_{ij}d_{3j} + d_{4i}'\mathcal{K}_{ij}d_{4j} + 2d_{1i}'\mathcal{K}_{ij}d_{2j} \\
&\qquad + 2d_{1i}'\mathcal{K}_{ij}d_{3j} + 2d_{1i}'\mathcal{K}_{ij}d_{4j} + 2d_{2i}'\mathcal{K}_{ij}d_{3j} + 2d_{2i}'\mathcal{K}_{ij}d_{4j} + 2d_{3i}'\mathcal{K}_{ij}d_{4j}\} \\
&\equiv A_1 + A_2 + A_3 + A_4 + 2A_5 + 2A_6 + 2A_7 + 2A_8 + 2A_9 + 2A_{10}, \text{ say,}
\end{aligned}$$

where $a_{NT} \equiv (h!)^{1/2}/(NT)$. We complete the proof by showing that under $\mathbb{H}_1(\gamma_{NT})$, (i) $\bar{A}_1 \equiv A_1 - B_{1NT} - B_{2,1NT} \xrightarrow{D} N(0, V_0)$, (ii) $A_2 = B_{2,2NT} + o_P(1)$, (iii) $A_5 = B_{2,3NT} + o_P(1)$, and (iv) $A_s = o_P(1)$ for $s = 3, 4, 6, ..., 10$, where B_{1NT} is defined in (3.10),

$$\begin{aligned}
B_{2,1NT} &\equiv (NT)^{-2}\sum_{1\le i,j\le N}\left(\Delta_i - X_iD_{NT}^{-1}\Pi_{NT}\right)'M_{F^0}\mathcal{K}_{ij}M_{F^0}(\Delta_j - X_jD_{NT}^{-1}\Pi_{NT}), \\
B_{2,2NT} &\equiv (NT)^{-2}\sum_{1\le i,j\le N}\left(D_{NT}^{-1}\Pi_{NT}\right)'\bar{X}_i'\mathcal{K}_{ij}\bar{X}_j\left(D_{NT}^{-1}\Pi_{NT}\right) + o_P(1), \\
B_{2,3NT} &\equiv (NT)^{-2}\sum_{1\le i,j\le N}\left(\Delta_i - X_iD_{NT}^{-1}\Pi_{NT}\right)'M_{F^0}\mathcal{K}_{ij}\bar{X}_j\left(D_{NT}^{-1}\Pi_{NT}\right),
\end{aligned}$$

and $\bar{X}_i \equiv N^{-1}\sum_{l=1}^{N}\alpha_{il}M_{F^0}X_l$. This is true because in view of the fact that

$$\begin{aligned} M_{F^0}\Delta_i - M_{F^0}X_iD_{NT}^{-1}\Pi_{NT} + \bar{X}_iD_{NT}^{-1}\Pi_{NT} &= M_{F^0}\Delta_i - \left(M_{F^0}X_i - N^{-1}\sum_{l=1}^{N}\alpha_{il}M_{F^0}X_l\right)D_{NT}^{-1}\Pi_{NT} \\ &= M_{F^0}\Delta_i - \tilde{X}_iD_{NT}^{-1}\Pi_{NT}, \end{aligned}$$

we have $B_{2,1NT}+B_{2,2NT}+2B_{2,3NT} = \frac{1}{(NT)^2}\sum_{1\le i,j\le N}(M_{F^0}\Delta_i - \tilde{X}_iD_{NT}^{-1}\Pi_{NT})'\mathcal{K}_{ij}(M_{F^0}\Delta_j - \tilde{X}_jD_{NT}^{-1}\Pi_{NT})$ $= B_{2NT}$, where B_{2NT} is defined in (3.11). 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 $\bar{A}_1 \xrightarrow{D} N(0, V_0)$ *under* $\mathbb{H}_1(\gamma_{NT})$.

Proof. Noting that $B_{1NT} = a_{NT}\sum_{i=1}^{N}\varepsilon_i'M_{F^0}\mathcal{K}_{ii}M_{F^0}\varepsilon_i$, we have

$$\begin{aligned} \bar{A}_1 &= a_{NT}\sum_{1\le i,j\le N}(\varepsilon_i + c_i)'M_{F^0}\mathcal{K}_{ij}M_{F^0}(\varepsilon_j + c_j) - B_{1NT} - B_{2,1NT} \\ &= a_{NT}\sum_{1\le i\ne j\le N}\varepsilon_i'M_{F^0}\mathcal{K}_{ij}M_{F^0}\varepsilon_j + \left(a_{NT}\sum_{1\le i,j\le N}c_i'M_{F^0}\mathcal{K}_{ij}M_{F^0}c_j - B_{2,1NT}\right) \\ &\quad +2a_{NT}\sum_{1\le i,j\le N}\varepsilon_i'M_{F^0}\mathcal{K}_{ij}M_{F^0}c_j \\ &\equiv A_{1,1} + A_{1,2} + 2A_{1,3}, \text{ say.} \end{aligned}$$

We complete the proof by showing that: (i) $A_{1,1} \xrightarrow{D} N(0, V_0)$, (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 $\mathcal{K}_{ji}' = \mathcal{K}_{ij}$ we can decompose $A_{1,1}$ as follows $A_{1,1} = a_{NT}\sum_{1\le i\ne j\le N}\varepsilon_i'\mathcal{K}_{ij}\varepsilon_j - 2a_{NT}\sum_{1\le i\ne j\le N}\varepsilon_i'P_{F^0}\mathcal{K}_{ij}\varepsilon_j + a_{NT}\sum_{1\le i\ne j\le N}\varepsilon_i'P_{F^0}\mathcal{K}_{ij}P_{F^0}\varepsilon_j$ $\equiv A_{1,11} - 2A_{1,12} + A_{1,13}$. 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 $A_{1,11} \xrightarrow{D} N(0, V_0)$. To achieve this goal, we rewrite $A_{1,11}$ as follows

$$A_{1,11} = \frac{(h!)^{1/2}}{NT}\sum_{1\le i\ne j\le N}\varepsilon_i'\mathcal{K}_{ij}\varepsilon_j = \sum_{1\le i<j\le N}W_{ij}$$

where $W_{ij} \equiv W_{NT}(u_i, u_j) \equiv 2(h!)^{1/2}(NT)^{-1}\sum_{1\le t,s\le T}\mathcal{K}_{ij,ts}\varepsilon_{js}\varepsilon_{it}$ and $u_i \equiv (X_i, \varepsilon_i)$. Noting that $A_{1,11}$ is a second order degenerate U-statistic that is "clean" ($E_{\mathcal{D}}[W_{NT}(u_i, u)] = E_{\mathcal{D}}[W_{NT}(u, u_j)] = 0$ 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 (i1) $\text{Var}_{\mathcal{D}}(A_{1,11}) = V_0 + o_P(1)$, (i2) $G_I \equiv \sum_{1\le i<j\le N}E_{\mathcal{D}}(W_{ij}^4) = o_P(1)$, (i3) $G_{II} \equiv \sum_{1\le i<j<k\le N}E_{\mathcal{D}}\left(W_{ik}^2W_{jk}^2 + W_{ik}^2W_{ij}^2 + W_{jk}^2W_{ji}^2\right) = o_P(1)$, and (i4) $G_{III} \equiv \sum_{1\le i<j<k<l\le N}E_{\mathcal{D}}(W_{ij}W_{ik}W_{lj}W_{lk} + W_{ij}W_{il}W_{kj}W_{kl} + W_{ik}W_{il}W_{jk}W_{jl}) = o_P(1)$.

For (i1), noting that $E_{\mathcal{D}}(A_{1,11}) = 0$ by Assumptions A.4(ii)-(iii), by the same assumptions, we have

$$\begin{aligned} \text{Var}_{\mathcal{D}}(A_{1,11}) &= \frac{4h!}{(NT)^2}\sum_{1\le i<j\le N}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\sum_{s_1=1}^{T}\sum_{s_2=1}^{T}E_{\mathcal{D}}\left(\mathcal{K}_{ij,t_1s_1}\mathcal{K}_{ij,t_2s_2}\varepsilon_{it_1}\varepsilon_{it_2}\varepsilon_{js_1}\varepsilon_{js_2}\right) \\ &= \frac{4h!}{(NT)^2}\sum_{1\le i<j\le N}\sum_{t=1}^{T}\sum_{s=1}^{T}E_{\mathcal{D}}\left(\mathcal{K}_{ij,ts}^2\varepsilon_{it}^2\varepsilon_{js}^2\right) = V_0 + o_P(1). \end{aligned}$$

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

$$\begin{aligned} E\left(G_I\right) &= \frac{16\left(h!\right)^2}{\left(NT\right)^4}\sum_{1\leq i<j\leq N}\sum_{1\leq t_1,\ldots,t_8\leq T} E\left[E_{\mathcal{D}}\left(\varepsilon_{it_1}\varepsilon_{it_3}\varepsilon_{it_5}\varepsilon_{it_7}\varepsilon_{jt_2}\varepsilon_{jt_4}\varepsilon_{jt_6}\varepsilon_{jt_8}\mathcal{K}_{ij,t_1t_2}\mathcal{K}_{ij,t_3t_4}\mathcal{K}_{ij,t_5t_6}\mathcal{K}_{ij,t_7t_8}\right)\right] \\ &= \frac{\left(h!\right)^2}{\left(NT\right)^4}O\left[N^2\left(T^6+T^4\left(h!\right)^{-2}+T^2\left(h!\right)^{-3}\right)\right] = O\left[N^{-2}T^2(h!)^2+N^{-2}+N^{-2}T^{-2}(h!)^{-1}\right] \\ &= o\left(1\right), \end{aligned}$$

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 $G_{II}=\sum_{1\leq i<j<k\leq N}E_{\mathcal{D}}(W_{ik}^2W_{jk}^2 +W_{ik}^2W_{ij}^2+W_{jk}^2W_{ji}^2)\equiv G_{II,1}+G_{II,2}+G_{II,3}$. Then

$$\begin{aligned} E\left(G_{II,1}\right) &= \frac{16\left(h!\right)^2}{\left(NT\right)^4}\sum_{1\leq i<j<k\leq N}\sum_{1\leq t_1,\ldots,t_6\leq T} E\left[E_{\mathcal{D}}\left(\varepsilon_{it_1}^2\varepsilon_{jt_2}^2\varepsilon_{kt_3}\varepsilon_{kt_4}\varepsilon_{kt_5}\varepsilon_{kt_6}\mathcal{K}_{ik,t_1t_3}\mathcal{K}_{ik,t_1t_4}\mathcal{K}_{jk,t_2t_5}\mathcal{K}_{jk,t_2t_6}\right)\right] \\ &= \frac{\left(h!\right)^2}{\left(NT\right)^4}O\left[N^3\left(T^5(h!)^{-1}+T^4\left(h!\right)^{-2}\right)\right] = O\left(TN^{-1}h!+N^{-1}\right) = o\left(1\right), \end{aligned}$$

where we use the fact that the term inside the last summation takes value 0 if $\#\{t_3,t_4,t_5,t_6\}=4$ by Assumptions A.4(ii)-(iii). It follows that $G_{II,1}=o_P\left(1\right)$ by the Markov inequality. Similarly, $G_{II,s}=o_P\left(1\right)$ for $s=2,3$. Thus we have $G_{II}=o_P\left(1\right)$. For (iv), we write $G_{III}=\sum_{1\leq i<j<k<l\leq N}[E_{\mathcal{D}}(W_{ij}W_{ik}W_{lj}W_{lk})+ E_{\mathcal{D}}(W_{ij}W_{il}W_{kj}W_{kl})+E_{\mathcal{D}}(W_{ik}W_{il}W_{jk}W_{jl})]\equiv G_{III,1}+G_{III,2}+G_{III,3}$. Then by Assumptions A.4(ii)-(iii)

$$\begin{aligned} E\left(G_{III,1}\right) &= \frac{16\left(h!\right)^2}{\left(NT\right)^4}\sum_{1\leq i<j<k<l\leq N}\sum_{1\leq t_1,\ldots,t_8\leq T} E[\varepsilon_{it_1}\varepsilon_{it_3}\varepsilon_{jt_2}\varepsilon_{jt_6}\varepsilon_{kt_4}\varepsilon_{kt_8}\varepsilon_{lt_5}\varepsilon_{lt_7}\mathcal{K}_{ij,t_1t_2}\mathcal{K}_{ik,t_3t_4}\mathcal{K}_{lj,t_5t_6}\mathcal{K}_{lk,t_7t_8}] \\ &= \frac{16\left(h!\right)^2}{\left(NT\right)^4}\sum_{1\leq i<j<k<l\leq N}\sum_{1\leq t_1,t_2,t_4,t_5\leq T} E[\varepsilon_{it_1}^2\varepsilon_{jt_2}^2\varepsilon_{kt_4}^2\varepsilon_{lt_5}^2\mathcal{K}_{ij,t_1t_2}\mathcal{K}_{ik,t_1t_4}\mathcal{K}_{lj,t_5t_2}\mathcal{K}_{lk,t_5t_4}] \\ &= \frac{\left(h!\right)^2}{\left(NT\right)^4}O\left(N^4T^4\right) = O\left(\left(h!\right)^2\right) = o\left(1\right). \end{aligned}$$

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

Next we show (ii). Let $\tilde{c}_i\equiv\gamma_{NT}\left(\Delta_i-X_iD_{NT}^{-1}\Pi_{NT}\right)$. Then by (3.7)

$$c_i=\gamma_{NT}\left(\Delta_i-X_iD_{NT}^{-1}\Pi_{NT}\right)+O_P(\delta_{NT}^{-2})X_i=\tilde{c}_i+O_P(\delta_{NT}^{-2})X_i. \tag{B.5}$$

Noting that $a_{NT}\sum_{1\leq i,j\leq N}\tilde{c}_i'M_{F^0}\mathcal{K}_{ij}M_{F^0}\tilde{c}_j=\left(NT\right)^{-2}\sum_{1\leq i,j\leq N}\left(\Delta_i-X_iD_{NT}^{-1}\Pi_{NT}\right)'M_{F^0}\mathcal{K}_{ij}M_{F^0}(\Delta_j -X_jD_{NT}^{-1}\Pi_{NT})=B_{2,1NT}$, we have $A_{1,2}=a_{NT}\sum_{1\leq i,j\leq N}\left(c_i-\tilde{c}_i\right)'M_{F^0}\mathcal{K}_{ij}M_{F^0}\left(c_j-\tilde{c}_j\right)+2a_{NT}\sum_{1\leq i,j\leq N} \tilde{c}_i'M_{F^0}\mathcal{K}_{ij}M_{F^0}\left(c_j-\tilde{c}_j\right)\equiv A_{1,21}+2A_{1,22}$, say. Let $c_{\mathcal{K}}\equiv\max_{1\leq i,j\leq N}\left\|\mathcal{K}_{ij}\right\|$. Then $c_{\mathcal{K}}=O_P\left(T\right)$ by Lemma D.1. By (B.5), the fact that $\sum_{i=1}^N\left\|X_i\right\|=O_P\left(NT^{1/2}\right)$ by the Markov inequality, and the fact that $\left\|M_{F^0}\right\|=1$,

$$\begin{aligned} \left|A_{1,21}\right| &\leq a_{NT}\sum_{1\leq i,j\leq N}\left\|M_{F^0}\right\|^2\left\|\mathcal{K}_{ij}\right\|\left\|c_i-\tilde{c}_i\right\|\left\|c_j-\tilde{c}_j\right\| = a_{NT}c_{\mathcal{K}}O_P(\delta_{NT}^{-4})\sum_{1\leq i,j\leq N}\left\|X_i\right\|\left\|X_j\right\| \\ &= O_P(a_{NT}\delta_{NT}^{-4}T)O_P\left(N^2T\right) = O_P(NT\delta_{NT}^{-4}(h!)^{1/2}) = o_P\left(1\right). \end{aligned}$$

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

Now we show (iii). Note that $A_{1,3} = \gamma_{NT} a_{NT} \sum_{1\le i,j\le N} \varepsilon_i' M_{F^0} \mathcal{K}_{ij} M_{F^0} \Delta_j + a_{NT} \sum_{1\le i,j\le N} \varepsilon_i' M_{F^0} \mathcal{K}_{ij} M_{F^0} X_j (\beta^0 - \hat{\beta}) \equiv \gamma_{NT} A_{1,31} + A_{1,32}(\beta^0 - \hat{\beta})$, 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 $\mathbb{H}_1(\gamma_{NT})$, where $B_{2,2NT} = (NT)^{-2} \sum_{1\le i,j\le N} \left(D_{NT}^{-1}\Pi_{NT}\right)' \bar{X}_i' \mathcal{K}_{ij} \bar{X}_j \left(D_{NT}^{-1}\Pi_{NT}\right)$.*

Proof. First, we decompose A_2 as follows

$$\begin{aligned}
A_2 &= \sum_{k=1}^{p}\left(\beta_k^0 - \hat{\beta}_k\right)\sum_{l=1}^{p}\left(\beta_l^0 - \hat{\beta}_l\right) a_{NT} \sum_{1\le i,j\le N} \left(\varepsilon_i' + \lambda_i^{0\prime} F^{0\prime} + c_i'\right) M_k^{(0)} \mathcal{K}_{ij} M_l^{(0)} \left(\varepsilon_j + F^0 \lambda_j^0 + c_j\right) \\
&= \sum_{k=1}^{p}\left(\beta_k^0 - \hat{\beta}_k\right)\sum_{l=1}^{p}\left(\beta_l^0 - \hat{\beta}_l\right) a_{NT} \sum_{1\le i,j\le N} \lambda_i^{0\prime} F^{0\prime} M_k^{(0)} \mathcal{K}_{ij} M_l^{(0)} F^0 \lambda_j^0 \\
&\quad + \sum_{k=1}^{p}\left(\beta_k^0 - \hat{\beta}_k\right)\sum_{l=1}^{p}\left(\beta_l^0 - \hat{\beta}_l\right) a_{NT} \sum_{1\le i,j\le N} \{\varepsilon_i' M_k^{(0)} \mathcal{K}_{ij} M_l^{(0)} \varepsilon_j + c_i' M_k^{(0)} \mathcal{K}_{ij} M_l^{(0)} c_j \\
&\qquad + 2\varepsilon_i' M_k^{(0)} \mathcal{K}_{ij} M_l^{(0)} F^0 \lambda_j^0 + 2\varepsilon_i' M_k^{(0)} \mathcal{K}_{ij} M_l^{(0)} c_j + 2\lambda_i' F^{0\prime} M_k^{(0)} \mathcal{K}_{ij} M_l^{(0)} c_j\} \\
&\equiv A_{2,1} + A_{2,2}, \text{ say.}
\end{aligned}$$

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

$$\begin{aligned}
A_{2,1} &= a_{NT} \sum_{k=1}^{p}\left(\beta_k^0 - \hat{\beta}_k\right)\sum_{l=1}^{p}\left(\beta_l^0 - \hat{\beta}_l\right) \sum_{1\le i,j\le N} \lambda_i^{0\prime} F^{0\prime} \Phi_1' \mathbf{X}_k M_{F^0} \mathcal{K}_{ij} M_{F^0} \mathbf{X}_l' \Phi_1 F^0 \lambda_j^0 \\
&= a_{NT} \sum_{k=1}^{p}\left(\beta_k^0 - \hat{\beta}_k\right)\sum_{l=1}^{p}\left(\beta_l^0 - \hat{\beta}_l\right) \sum_{1\le i,j\le N} \lambda_i^{0\prime} \left(\lambda^{0\prime}\lambda^0\right)^{-1} \lambda^{0\prime} \mathbf{X}_k M_{F^0} \mathcal{K}_{ij} M_{F^0} \mathbf{X}_l' \lambda^0 \left(\lambda^{0\prime}\lambda^0\right)^{-1} \lambda_j^0 \\
&= \frac{1}{(NT)^2} \sum_{k=1}^{p} \iota_k' D_{NT}^{-1} \Pi_{NT} \sum_{l=1}^{p} \iota_l' D_{NT}^{-1} \Pi_{NT} \sum_{1\le i,j\le N} \lambda_i^{0\prime} \left(\lambda^{0\prime}\lambda^0\right)^{-1} \lambda^{0\prime} \mathbf{X}_k M_{F^0} \mathcal{K}_{ij} M_{F^0} \mathbf{X}_l' \lambda^0 \left(\lambda^{0\prime}\lambda^0\right)^{-1} \lambda_j^0 \\
&\quad + o_P(1) \\
&= \frac{1}{(NT)^2} \sum_{1\le i,j\le N} \sum_{k=1}^{p} \iota_k' D_{NT}^{-1} \Pi_{NT} \bar{X}_{k,\cdot i}' \mathcal{K}_{ij} \sum_{l=1}^{p} \iota_l' D_{NT}^{-1} \Pi_{NT} \bar{X}_{l,\cdot j} + o_P(1) \\
&= \frac{1}{(NT)^2} \sum_{1\le i,j\le N} \left(D_{NT}^{-1}\Pi_{NT}\right)' \bar{X}_i' \mathcal{K}_{ij} \bar{X}_j \left(D_{NT}^{-1}\Pi_{NT}\right) + o_P(1) = B_{2,2NT} + o_P(1),
\end{aligned}$$

where ι_k is a $p \times 1$ vector with 1 in the kth place and zeros elsewhere, and $\bar{X}_i \equiv N^{-1} \sum_{l=1}^{N} \alpha_{il} M_{F^0} X_l$ is a $T \times p$ matrix whose kth column is given by $\bar{X}_{i,\cdot k} \equiv \left(\lambda_i^{0\prime} \left(\lambda^{0\prime}\lambda^0\right)^{-1} \lambda^{0\prime} \mathbf{X}_k M_{F^0}\right)'$.

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

Then

$$\begin{aligned} A_{2,2} &= \left(\beta^0-\hat{\beta}\right)^2 a_{NT} \sum_{1\le i,j\le N} \{\varepsilon_i' M^{(0)} \mathcal{K}_{ij} M^{(0)} \varepsilon_j + c_i' M^{(0)} \mathcal{K}_{ij} M^{(0)} c_j + 2\varepsilon_i' M^{(0)} \mathcal{K}_{ij} M^{(0)} F^0 \lambda_j^0 \\ &\quad + 2\varepsilon_i' M^{(0)} \mathcal{K}_{ij} M^{(0)} c_j + 2\lambda_i' F^{0\prime} M^{(0)} \mathcal{K}_{ij} M^{(0)} c_j\} \\ &\equiv \left(\beta^0-\hat{\beta}\right)^2 \{A_{2,21}+A_{2,22}+2A_{2,23}+2A_{2,24}+2A_{2,25}\}, \text{ say.} \end{aligned}$$

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

Noting that $\left\|M^{(0)}\right\| = O_P(1)$ by (B.3) and

$$\left\|M^{(0)}\varepsilon_i\right\| = \|(M_{F^0}\mathbf{X}'\Phi_1 + \Phi_1'\mathbf{X}M_{F^0})\varepsilon_i\| = O_P\left(T^{-1/2}\right)\left\|F^{0\prime}\varepsilon_i\right\| + O_P\left(N^{-1/2}T^{-1/2}\right)\|\mathbf{X}'\varepsilon_i\|, \quad \text{(B.6)}$$

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

$$\begin{aligned} \left|\bar{A}_{2,21}\right| &\le (NT)^{-2} \sum_{1\le i,j\le N} \|\mathcal{K}_{ij}\| \left\|M^{(0)}\varepsilon_i\right\| \left\|M^{(0)}\varepsilon_j\right\| \\ &\le c_{\mathcal{K}} O_P\left(N^{-2}T^{-2}\right) \sum_{1\le i,j\le N} \left[O_P\left(T^{-1/2}\right)\left\|F^{0\prime}\varepsilon_i\right\| + O_P\left(N^{-1/2}T^{-1/2}\right)\|\mathbf{X}'\varepsilon_i\|\right] \\ &\qquad \times \left[O_P\left(T^{-1/2}\right)\left\|F^{0\prime}\varepsilon_j\right\| + O_P\left(N^{-1/2}T^{-1/2}\right)\|\mathbf{X}'\varepsilon_j\|\right] \\ &= TO_P\left(N^{-2}T^{-2}\right)O_P\left(N^2\right) = O_P\left(T^{-1}\right) = o_P(1), \\ \left|\bar{A}_{2,23}\right| &\le c_{\mathcal{K}}(NT)^{-2}\left\|M^{(0)}F^0\right\| \sum_{1\le i,j\le N} \left[O_P\left(T^{-1/2}\right)\left\|F^{0\prime}\varepsilon_i\right\| + O_P\left(N^{-1/2}T^{-1/2}\right)\|\mathbf{X}'\varepsilon_i\|\right]\left\|\lambda_j^0\right\| \\ &= T(NT)^{-2} O_P\left(T^{1/2}\right)O_P\left(N^2\right) = O_P\left(T^{-1/2}\right) = o_P(1), \end{aligned}$$

and $\left|\bar{A}_{2,24}\right| \le c_{\mathcal{K}}\gamma_{NT}(NT)^{-2}\left\|M^{(0)}\right\| \sum_{1\le i,j\le N}\left[O_P\left(T^{-1/2}\right)\left\|F^{0\prime}\varepsilon_i\right\| + O_P\left(N^{-1/2}T^{-1/2}\right)\|\mathbf{X}'\varepsilon_i\|\right](\|X_j\| + \|\Delta_j\|) = T\gamma_{NT}(NT)^{-2}O_P\left(N^2T^{1/2}\right) = O_P\left(T^{-1/2}\gamma_{NT}\right) = o_P(1).$

In addition, by (B.3) and (B.5), $\left|\bar{A}_{2,22}\right| \le c_{\mathcal{K}} O_P\left(\gamma_{NT}^2\right)(NT)^{-2}\left\|M^{(0)}\right\|^2 \sum_{1\le i,j\le N}(\|X_i\|+\|\Delta_i\|)(\|X_j\|+\|\Delta_j\|) = TO_P\left(\gamma_{NT}^2\right)(NT)^{-2}O_P(1)O_P\left(N^2T\right) = O_P\left(\gamma_{NT}^2\right) = o_P(1)$ and $\left|\bar{A}_{2,25}\right| \le c_{\mathcal{K}}\gamma_{NT}(NT)^{-2}\left\|M^{(0)}\right\|^2 \sum_{1\le i,j\le N}\left\|F^0\lambda_i^0\right\|(\|X_j\|+\|\Delta_j\|) = T\gamma_{NT}(NT)^{-2}O_P(1)O_P\left(N^2T\right) = O_P(\gamma_{NT}) = o_P(1).$ ■

Proposition B.3 *$A_3 = o_P(1)$ under $\mathbb{H}_1(\gamma_{NT})$.*

Proof. Recall $M^{(1)} = -M_{F^0}\boldsymbol{\varepsilon}'\Phi_1 - \Phi_1'\boldsymbol{\varepsilon}M_{F^0}$ and $\Phi_1 = \lambda^0\left(\lambda^{0\prime}\lambda^0\right)^{-1}\left(F^{0\prime}F^0\right)^{-1}F^{0\prime}$. Noting that $\Phi_1 M_{F^0} = 0$ and $\mu_1(M_{F^0}) = 1$, we have

$$\begin{aligned} \left\|M^{(1)}\varepsilon_i\right\|_F^2 &= \operatorname{tr}\left[\varepsilon_i'\left(M_{F^0}\boldsymbol{\varepsilon}'\Phi_1 + \Phi_1'\boldsymbol{\varepsilon}M_{F^0}\right)\left(M_{F^0}\boldsymbol{\varepsilon}'\Phi_1 + \Phi_1'\boldsymbol{\varepsilon}M_{F^0}\right)\varepsilon_i\right] \\ &= 2\operatorname{tr}\left(\varepsilon_i'\Phi_1'\boldsymbol{\varepsilon}M_{F^0}\boldsymbol{\varepsilon}'\Phi_1\varepsilon_i\right) \le 2\operatorname{tr}\left(\varepsilon_i'\Phi_1'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\Phi_1\varepsilon_i\right) \\ &= 2\operatorname{tr}\left[\left(F^{0\prime}F^0\right)^{-1}\left(\lambda^{0\prime}\lambda^0\right)^{-1}\lambda^{0\prime}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\lambda^0\left(\lambda^{0\prime}\lambda^0\right)^{-1}\left(F^{0\prime}F^0\right)^{-1}F^{0\prime}\varepsilon_i\varepsilon_i'F^0\right] \\ &\le 2\operatorname{tr}\left[\left(\lambda^{0\prime}\lambda^0\right)^{-1}\left(F^{0\prime}F^0\right)^{-1}\left(F^{0\prime}F^0\right)^{-1}\left(\lambda^{0\prime}\lambda^0\right)^{-1}\right]\operatorname{tr}\left(\lambda^{0\prime}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\lambda^0\right)\operatorname{tr}\left(F^{0\prime}\varepsilon_i\varepsilon_i'F^0\right) \\ &= O_P\left((NT)^{-2}\right)O_P(NT)\operatorname{tr}\left(F^{0\prime}\varepsilon_i\varepsilon_i'F^0\right) = O_P\left((NT)^{-1}\right)\left\|F^{0\prime}\varepsilon_i\right\|^2, \end{aligned}$$

where we have repeatedly used the rotational property of the trace operator, the fact that

$$\operatorname{tr}(AB) \leq \mu_1(A) \operatorname{tr}(B) \tag{B.7}$$

for any symmetric matrix A and p.s.d. matrix B (see, e.g., Bernstein, 2005, Proposition 8.4.13), and the fact that

$$\operatorname{tr}(AB) \leq \operatorname{tr}(A) \operatorname{tr}(B) \tag{B.8}$$

for any two p.s.d. matrices A and B (see, e.g., Bernstein, 2005, Fact 8.10.7). It follows that

$$\left\| M^{(1)} \varepsilon_i \right\| = O_P\left((NT)^{-1/2} \right) \left\| F^{0\prime} \varepsilon_i \right\| \tag{B.9}$$

By the fact that $\left\| M^{(1)} \right\| = O_P\left(N^{-1/2}\right)$ (see (B.3)) and (3.7),

$$\left\| M^{(1)} c_i \right\| \leq \left\| M^{(1)} \right\| \|c_i\| = O_P\left(N^{-1/2} \gamma_{NT} \right) (\|X_i\| + \|\Delta_i\|). \tag{B.10}$$

Combining (B.9) and (B.10) yields

$$\left\| M^{(1)} (\varepsilon_i + c_i) \right\| = O_P\left((NT)^{-1/2} \right) \left\| F^{0\prime} \varepsilon_i \right\| + O_P\left(N^{-1/2} \gamma_{NT} \right) (\|X_i\| + \|\Delta_i\|). \tag{B.11}$$

We will use these results frequently.

Now, we decompose A_3 as follows.

$$\begin{aligned}
A_3 &= a_{NT} \sum_{1 \leq i,j \leq N} \left(\varepsilon_i' + \lambda_i^{0\prime} F^{0\prime} + c_i' \right) M^{(1)} \mathcal{K}_{ij} M^{(1)} \left(\varepsilon_j + F^0 \lambda_j^0 + c_j \right) \\
&= a_{NT} \sum_{1 \leq i,j \leq N} \{ \varepsilon_i' M^{(1)} \mathcal{K}_{ij} M^{(1)} \varepsilon_j + \lambda_i^{0\prime} F^{0\prime} M^{(1)} \mathcal{K}_{ij} M^{(1)} F^0 \lambda_j^0 + c_i' M^{(1)} \mathcal{K}_{ij} M^{(1)} c_j \\
&\quad + 2 \varepsilon_i' M^{(1)} \mathcal{K}_{ij} M^{(1)} F^0 \lambda_j^0 + 2 \varepsilon_i' M^{(1)} \mathcal{K}_{ij} M^{(1)} c_j + 2 \lambda_i' F^{0\prime} M^{(1)} \mathcal{K}_{ij} M^{(1)} c_j \} \\
&\equiv A_{3,1} + A_{3,2} + A_{3,3} + 2A_{3,4} + 2A_{3,5} + 2A_{3,6}, \text{ say.}
\end{aligned}$$

We prove the proposition by demonstrating that $A_{3,s} = o_P(1)$ for $s = 1, 2, ..., 6$. By (B.9)-(B.11), (B.3), Assumptions A.4(iii) and A.5, and Lemma D.1, we have $|A_{3,1}| \leq c_{\mathcal{K}} a_{NT} O_P\left((NT)^{-1}\right) \sum_{1 \leq i,j \leq N} \left\| F^{0\prime} \varepsilon_i \right\| \left\| F^{0\prime} \varepsilon_j \right\| = T O_P\left(a_{NT} (NT)^{-1}\right) O_P\left(N^2 T\right) = O_P\left((h!)^{1/2}\right) = o_P(1)$, $|A_{3,3}| \leq c_{\mathcal{K}} a_{NT} O_P(N^{-1} \gamma_{NT}^2) \sum_{1 \leq i,j \leq N} (\|X_i\| + \|\Delta_i\|)(\|X_j\| + \|\Delta_j\|) = T O_P(a_{NT} N^{-1} \gamma_{NT}^2) O_P\left(N^2 T\right) = O_P\left(N^{-1}\right) = o_P(1)$, $|A_{3,5}| \leq c_{\mathcal{K}} a_{NT} O_P\left(N^{-1/2} \gamma_{NT} (NT)^{-1/2}\right) \sum_{1 \leq i,j \leq N} \left\| F^{0\prime} \varepsilon_i \right\| (\|X_j\| + \|\Delta_j\|) = T O_P\left(a_{NT} \gamma_{NT} N^{-1} T^{-1/2}\right) O_P\left(N^2 T\right) = O_P(N^{-1/2} (h!)^{1/4}) = o_P(1)$, and $|A_{3,6}| \leq c_{\mathcal{K}} a_{NT} O_P\left(N^{-1/2} \gamma_{NT}\right) O_P\left(T^{1/2} N^{-1/2}\right) \sum_{1 \leq i,j \leq N} \left\| \lambda_i^0 \right\| (\|X_j\| + \|\Delta_j\|) = T O_P\left(a_{NT} \gamma_{NT} T^{1/2} N^{-1}\right) O_P\left(N^2 T^{1/2}\right) = O_P(T^{1/2} N^{-1/2} (h!)^{1/4}) = o_P(1)$.

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

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

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 $|A_4| \leq c_{\mathcal{K}} a_{NT} \left\| M^{(2)} + M^{(rem)} \right\|^2 \sum_{1 \leq i,j \leq N} \left\| \varepsilon_i + F^0 \lambda_i^0 + c_i \right\| \left\| \varepsilon_j + F^0 \lambda_j^0 + c_j \right\| = T a_{NT} O_P\left(\delta_{NT}^{-4}\right) O_P\left(N^2 T\right) = O_P\left(NT \delta_{NT}^{-4} (h!)^{1/2}\right) = o_P(1)$. ■

Proposition B.5 $A_5 = B_{2,3NT} + o_P(1)$ *under* $\mathbb{H}_1(\gamma_{NT})$, *where* $B_{2,3NT} \equiv (NT)^{-2}\sum_{1\le i,j\le N}(\Delta_i - X_i D_{NT}^{-1}\Pi_{NT})' M_{F^0}\mathcal{K}_{ij}\bar{X}_j\left(D_{NT}^{-1}\Pi_{NT}\right)$.

Proof. First, we decompose A_5 as follows

$$\begin{aligned} A_5 &= \sum_{k=1}^{p}\left(\beta_k^0 - \hat{\beta}_k\right) a_{NT} \sum_{1\le i,j\le N} \left(\varepsilon_i' + c_i'\right) M_{F^0}\mathcal{K}_{ij} M_k^{(0)}\left(\varepsilon_j + F^0\lambda_j^0 + c_j\right) \\ &= \sum_{k=1}^{p}\left(\beta_k^0 - \hat{\beta}_k\right) a_{NT} \sum_{1\le i,j\le N} c_i' M_{F^0}\mathcal{K}_{ij} M_k^{(0)} F^0\lambda_j^0 \\ &\quad + \sum_{k=1}^{p}\left(\beta_k^0 - \hat{\beta}_k\right) a_{NT} \sum_{1\le i,j\le N} \{\left(\varepsilon_i' + c_i'\right) M_{F^0}\mathcal{K}_{ij} M_k^{(0)}\left(\varepsilon_j + c_j\right) + \varepsilon_i' M_{F^0}\mathcal{K}_{ij} M_k^{(0)} F^0\lambda_j^0\} \\ &\equiv A_{5,1} + A_{5,2}, \text{ say.} \end{aligned}$$

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 (3.7)

$$\begin{aligned} A_{5,1} &= -a_{NT}\sum_{k=1}^{p}\left(\beta_k^0 - \hat{\beta}_k\right) \sum_{1\le i,j\le N} c_i' M_{F^0}\mathcal{K}_{ij} M_{F^0}\mathbf{X}_k'\Phi_1 F^0\lambda_j^0 \\ &= -a_{NT}\sum_{k=1}^{p}\left(\beta_k^0 - \hat{\beta}_k\right) \sum_{1\le i,j\le N} c_i' M_{F^0}\mathcal{K}_{ij} M_{F^0}\mathbf{X}_k'\lambda^0\left(\lambda^{0\prime}\lambda^0\right)^{-1}\lambda_j^0 \\ &= \frac{1}{(NT)^2}\sum_{k=1}^{p}\iota_k' D_{NT}^{-1}\Pi_{NT} \sum_{1\le i,j\le N}\left(\Delta_i - X_i D_{NT}^{-1}\Pi_{NT}\right)' M_{F^0}\mathcal{K}_{ij} M_{F^0}\mathbf{X}_k'\lambda^0\left(\lambda^{0\prime}\lambda^0\right)^{-1}\lambda_j^0 + o_P(1) \\ &= \frac{1}{(NT)^2}\sum_{1\le i,j\le N}\left(\Delta_i - X_i D_{NT}^{-1}\Pi_{NT}\right)' M_{F^0}\mathcal{K}_{ij}\sum_{k=1}^{p}\iota_l' D_{NT}^{-1}\Pi_{NT}\bar{X}_{k,\cdot j} + o_P(1) \\ &= \frac{1}{(NT)^2}\sum_{1\le i,j\le N}\left(\Delta_i - X_i D_{NT}^{-1}\Pi_{NT}\right)' M_{F^0}\mathcal{K}_{ij}\bar{X}_j\left(D_{NT}^{-1}\Pi_{NT}\right) + o_P(1) = B_{2,3NT} + o_P(1). \end{aligned}$$

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

$$\begin{aligned} A_{5,2} &= \left(\beta^0 - \hat{\beta}\right) a_{NT}\sum_{1\le i,j\le N}\{\varepsilon_i' M_{F^0}\mathcal{K}_{ij}M^{(0)}\varepsilon_j + \varepsilon_i' M_{F^0}\mathcal{K}_{ij}M^{(0)}c_j + c_i' M_{F^0}\mathcal{K}_{ij}M^{(0)}\varepsilon_j \\ &\qquad + c_i' M_{F^0}\mathcal{K}_{ij}M^{(0)}c_j + \varepsilon_i' M_{F^0}\mathcal{K}_{ij}M^{(0)}F^0\lambda_j^0\} \\ &\equiv \left(\beta^0 - \hat{\beta}\right)\left(A_{5,21} + A_{5,22} + A_{5,23} + A_{5,24} + A_{5,25}\right). \end{aligned}$$

We prove the proposition by showing that $\bar{A}_{5,2s} = \gamma_{NT}A_{5,2s} = o_P(1)$ for $s = 1, 2, ..., 5$. By Lemma D.3(iv), $\bar{A}_{5,21} = o_P(1)$. By Lemma D.2(iii), $\bar{A}_{5,25} = o_P(1)$. So we are left to show that $\bar{A}_{5,2s} = o_P(1)$ for $s = 2, 3, 4$.

For $\bar{A}_{5,22}$, we have $\bar{A}_{5,22} = \gamma_{NT}a_{NT}\sum_{1\le i,j\le N}\varepsilon_i'\mathcal{K}_{ij}M^{(0)}c_j - \gamma_{NT}a_{NT}\sum_{1\le i,j\le N}\varepsilon_i' P_{F^0}\mathcal{K}_{ij}M^{(0)}c_j \equiv \bar{A}_{5,22a} - \bar{A}_{5,22b}$. Using (B.3) and (B.5), Lemma D.1, Assumptions A.4(iii) and A.5, and the fact that $\|P_{F^0}\varepsilon_i\| = O_P\left(T^{-1/2}\right)\left\|F^{0\prime}\varepsilon_i\right\|$, we can bound $\bar{A}_{5,22b}$ directly: $\left|\bar{A}_{5,22b}\right| \le O_P\left(\gamma_{NT}T^{-1/2}\right)c_{\mathcal{K}}\gamma_{NT}a_{NT}$

$\sum_{1\leq i,j\leq N}\left\|F^{0\prime}\varepsilon_i\right\|(\|X_j\|+\|\Delta_j\|)=O_P\left(T^{-1/2}\right)=o_P(1)$. For $\bar{A}_{5,22a}$, we can easily show that $\bar{A}_{5,22a}=A_{5,22a}+o_P(1)$ where $A_{5,22a}\equiv\gamma_{NT}a_{NT}\sum_{1\leq i\neq j\leq N}\varepsilon_i'\mathcal{K}_{ij}M^{(0)}\tilde{c}_j$. Noting that $E\left\|\sum_{1\leq i\neq j\leq N}\Delta_j\varepsilon_i'\mathcal{K}_{ij}\right\|_F^2=\sum_{1\leq i_1\neq j_1\leq N}\sum_{1\leq i_2\neq j_2\leq N}\sum_{1\leq t_1,\ldots,t_4\leq T}E\left(\varepsilon_{i_1t_1}\mathcal{K}_{i_1j_1,t_1t_2}\mathcal{K}_{j_2i_2,t_2t_3}\varepsilon_{i_2t_3}\Delta_{j_2t_4}\Delta_{j_1t_4}\right)=O\left(N^3T^3\right)$, we have $\left\|\sum_{1\leq i\neq j\leq N}\Delta_j\varepsilon_i'\mathcal{K}_{ij}\right\|_F=O_P\left(N^{3/2}T^{3/2}\right)$. Similar result holds when Δ_j is replaced by $X_jD_{NT}^{-1}\Pi_{NT}$. Then by Cauchy-Schwarz's and Minkowski's inequalities

$$\begin{aligned}|A_{5,22a}| &= \gamma_{NT}a_{NT}\left|\operatorname{tr}\left(M^{(0)}\sum_{1\leq i\neq j\leq N}\tilde{c}_j\varepsilon_i'\mathcal{K}_{ij}\right)\right|\leq\gamma_{NT}^2a_{NT}\left\|M^{(0)}\right\|_F\left\|\sum_{1\leq i\neq j\leq N}\left(\Delta_j-X_jD_{NT}^{-1}\Pi_{NT}\right)\varepsilon_i'\mathcal{K}_{ij}\right\|_F\\ &= O_P\left(N^{-2}T^{-2}\right)O_P\left(N^{3/2}T^{3/2}\right)=O_P\left(N^{-1/2}T^{-1/2}\right)=o_P(1).\end{aligned}$$

It follows that $\bar{A}_{5,22}=o_P(1)$. By (B.5), (B.6) and (B.3), $\left|\bar{A}_{5,23}\right|\leq c_{\mathcal{K}}\gamma_{NT}^2a_{NT}\sum_{1\leq i,j\leq N}(\|X_i\|+\|\Delta_i\|)$ $\left[O_P\left(T^{-1/2}\right)\left\|F^{0\prime}\varepsilon_j\right\|+O_P\left(N^{-1/2}T^{-1/2}\right)\|\mathbf{X}'\varepsilon_j\|\right]=T\gamma_{NT}^2a_{NT}O_P\left(N^2T^{1/2}\right)=O_P\left(T^{-1/2}\right)=o_P(1)$, and $\left|\bar{A}_{5,24}\right|\leq c_{\mathcal{K}}\gamma_{NT}^3a_{NT}\left\|M^{(0)}\right\|\sum_{1\leq i,j\leq N}(\|X_i\|+\|\Delta_i\|)(\|X_j\|+\|\Delta_j\|)=T\gamma_{NT}^3a_{NT}O_P\left(N^2T\right)=O_P(\gamma_{NT})=o_P(1)$. This completes the proof. ■

Proposition B.6 $A_6=o_P(1)$ *under* $\mathbb{H}_1(\gamma_{NT})$.

Proof. First, we decompose A_6 as follows

$$\begin{aligned}A_6 &= a_{NT}\sum_{1\leq i,j\leq N}(\varepsilon_i'+c_i')M_{F^0}\mathcal{K}_{ij}M^{(1)}\left(\varepsilon_j+F^0\lambda_j^0+c_j\right)\\ &= a_{NT}\sum_{1\leq i,j\leq N}\{\varepsilon_i'M_{F^0}\mathcal{K}_{ij}M^{(1)}\varepsilon_j+\varepsilon_i'M_{F^0}\mathcal{K}_{ij}M^{(1)}F^0\lambda_j^0+\varepsilon_i'M_{F^0}\mathcal{K}_{ij}M^{(1)}c_j\\ &\qquad +c_i'M_{F^0}\mathcal{K}_{ij}M^{(1)}\varepsilon_j+c_i'M_{F^0}\mathcal{K}_{ij}M^{(1)}F^0\lambda_j^0+c_i'M_{F^0}\mathcal{K}_{ij}M^{(1)}c_j\}\\ &\equiv A_{6,1}+A_{6,2}+A_{6,3}+A_{6,4}+A_{6,5}+A_{6,6}.\end{aligned}$$

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

By (B.9)-(B.10) and Lemma D.1, $|A_{6,4}|\leq a_{NT}\sum_{1\leq i,j\leq N}\|\mathcal{K}_{ij}\|\|M_{F^0}c_i\|\left\|M^{(1)}\varepsilon_j\right\|\leq c_{\mathcal{K}}O_P\left(\gamma_{NT}(NT)^{-1/2}\right)$ $a_{NT}\sum_{1\leq i,j\leq N}(\|X_i\|+\|\Delta_i\|)\left\|F^{0\prime}\varepsilon_j\right\|=TO_P\left(\gamma_{NT}(NT)^{-1/2}\right)O_P\left(N^2T\right)=O_P((h!)^{1/4})=o_P(1)$, and $|A_{6,6}|\leq a_{NT}\sum_{1\leq i,j\leq N}\|\mathcal{K}_{ij}\|\|M_{F^0}c_i\|\left\|M^{(1)}c_j\right\|\leq c_{\mathcal{K}}O_P\left(N^{-1/2}\gamma_{NT}^2\right)a_{NT}\sum_{1\leq i,j\leq N}(\|X_i\|+\|\Delta_i\|)(\|X_j\|$ $+\|\Delta_j\|)\leq TO_P\left(N^{-1/2}\gamma_{NT}^2a_{NT}\right)O_P\left(N^2T\right)=O_P\left(N^{-1/2}\right)=o_P(1)$.This completes the proof. ■

Proposition B.7 $A_7=o_P(1)$ *under* $\mathbb{H}_1(\gamma_{NT})$.

Proof. First we decompose A_7 as follows

$$\begin{aligned}A_7 &= a_{NT}\sum_{1\leq i,j\leq N}(\varepsilon_i'+c_i')M_{F^0}\mathcal{K}_{ij}\left(M^{(2)}+M^{(rem)}\right)\left(\varepsilon_j+F^0\lambda_j^0+c_j\right)\\ &= a_{NT}\sum_{1\leq i,j\leq N}\{c_i'M_{F^0}\mathcal{K}_{ij}(M^{(2)}+M^{(rem)})\left(\varepsilon_j+F^0\lambda_j^0+c_j\right)+\varepsilon_i'M_{F^0}\mathcal{K}_{ij}M^{(2)}\varepsilon_j+\varepsilon_i'M_{F^0}\mathcal{K}_{ij}M^{(2)}F^0\lambda_j^0\\ &\qquad +\varepsilon_i'M_{F^0}\mathcal{K}_{ij}M^{(2)}c_j+\varepsilon_i'M_{F^0}\mathcal{K}_{ij}M^{(rem)}\varepsilon_j+\varepsilon_i'M_{F^0}\mathcal{K}_{ij}M^{(rem)}F^0\lambda_j^0+\varepsilon_i'M_{F^0}\mathcal{K}_{ij}M^{(rem)}c_j\}\\ &\equiv A_{7,1}+A_{7,2}+A_{7,3}+A_{7,4}+A_{7,5}+A_{7,6}+A_{7,7}.\end{aligned}$$

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, $|A_{7,1}| \leq c_{\mathcal{K}} O_P(\gamma_{NT}) a_{NT} \|M^{(2)} + M^{(rem)}\| \sum_{1\leq i,j\leq N}(\|X_i\| + \|\Delta_i\|) (\|\varepsilon_j\| + \|F^0\lambda_j^0\| + \|c_j\|) = TO_P(\gamma_{NT} a_{NT} \delta_{NT}^{-2}) O_P(N^2T) = o_P(1)$, $|A_{7,4}| \leq c_{\mathcal{K}} O_P(\gamma_{NT}) a_{NT} \|M^{(2)}\| \sum_{1\leq i,j\leq N} \|\varepsilon_i\| (\|X_j\| + \|\Delta_j\|) = TO_P(\gamma_{NT} a_{NT} \delta_{NT}^{-2}) O_P(N^2T) = o_P(1)$, and $|A_{7,7}| \leq c_{\mathcal{K}} a_{NT} O_P(\gamma_{NT}) \|M^{(rem)}\| \sum_{1\leq i,j\leq N} \|\varepsilon_i\| (\|X_j\| + \|\Delta_j\|) = TO_P(a_{NT} \delta_{NT}^{-1} \gamma_{NT}^2) O_P(N^2T) = o_P(1)$.

Next, $A_{7,5} = a_{NT} \sum_{1\leq i,j\leq N} \varepsilon_i' \mathcal{K}_{ij} M^{(rem)} \varepsilon_j - a_{NT} \sum_{1\leq i,j\leq N} \varepsilon_i' P_{F^0} \mathcal{K}_{ij} M^{(rem)} \varepsilon_j \equiv A_{7,51} - A_{7,52}$. Noting that $\|P_{F^0}\varepsilon_i\| = O_P(T^{-1/2}) \|F^{0\prime}\varepsilon_i\|$, we have by (B.2), $|A_{7,52}| \leq c_{\mathcal{K}} a_{NT} O_P(T^{-1/2}) O_P(\delta_{NT}^{-1}\gamma_{NT}) \sum_{1\leq i,j\leq N} \|F^{0\prime}\varepsilon_i\| \|\varepsilon_j\| = O_P(a_{NT} T^{1/2} \delta_{NT}^{-1} \gamma_{NT}) O_P(N^2T) = o_P(1)$. By (B.2), $|A_{7,51}| = a_{NT} |\text{tr}(M^{(rem)} \sum_{1\leq i,j\leq N} \varepsilon_j \varepsilon_i' \mathcal{K}_{ij})| \leq a_{NT} \|M^{(rem)}\|_F \|\sum_{1\leq i,j\leq N} \varepsilon_j \varepsilon_i' \mathcal{K}_{ij}\|_F = a_{NT} O_P(\delta_{NT}^{-1}\gamma_{NT}) O_P(N^{3/2}T^{3/2}) = o_P(1)$ where we use the fact that $E \|\sum_{1\leq i,j\leq N} \varepsilon_j \varepsilon_i' \mathcal{K}_{ij}\|_F^2 = O(N^3T^3)$. Thus $A_{7,5} = o_P(1)$.

Now, write $A_{7,6} = a_{NT} \sum_{1\leq i,j\leq N} \varepsilon_i' \mathcal{K}_{ij} M^{(rem)} F^0 \lambda_j^0 - a_{NT} \sum_{1\leq i,j\leq N} \varepsilon_i' P_{F^0} \mathcal{K}_{ij} M^{(rem)} F^0 \lambda_j^0 \equiv A_{7,61} - A_{7,62}$. 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 $|A_{7,61}| = a_{NT} |\text{tr}(M^{(rem)} F^0 \sum_{1\leq i,j\leq N} \lambda_j^0 \varepsilon_i' \mathcal{K}_{ij})| \leq a_{NT} \|M^{(rem)} F^0\|_F \|\sum_{1\leq i,j\leq N} \lambda_j^0 \varepsilon_i' \mathcal{K}_{ij}\|_F = a_{NT} O_P(\delta_{NT}^{-1} \gamma_{NT} \sqrt{T}) O_P(N^{3/2}T) = o_P(1)$. It follows that $A_{7,6} = o_P(1)$. ■

Proposition B.8 $A_8 = o_P(1)$ *under* $\mathbb{H}_1(\gamma_{NT})$.

Proof. Again, assuming $p = 1$, we can decompose A_8 as follows

$$\begin{aligned} A_8 &= \left(\beta^0 - \hat{\beta}\right) a_{NT} \sum_{1\leq i,j\leq N} \{(\varepsilon_i' + c_i') M^{(1)} \mathcal{K}_{ij} M^{(0)} (\varepsilon_j + F^0\lambda_j^0 + c_j) \\ &\quad + \lambda_i^{0\prime} F^{0\prime} M^{(1)} \mathcal{K}_{ij} M^{(0)} \varepsilon_j + \lambda_i^{0\prime} F^{0\prime} M^{(1)} \mathcal{K}_{ij} M^{(0)} F^0 \lambda_j^0 + \lambda_i^{0\prime} F^{0\prime} M^{(1)} \mathcal{K}_{ij} M^{(0)} c_j\} \\ &\equiv \left(\beta^0 - \hat{\beta}\right) (A_{8,1} + A_{8,2} + A_{8,3} + A_{8,4}). \end{aligned}$$

We prove the claim by showing that $\bar{A}_{8,s} \equiv \gamma_{NT} A_{8,s} = o_P(1)$ for $s = 1, 2, 3, 4$. By Lemma D.3(viii), $\bar{A}_{8,2} = o_P(1)$. By Lemma D.2(v), $\bar{A}_{8,3} = o_P(1)$. By (B.11), (B.5), and Lemma D.1, we can readily show that

$$\begin{aligned} |\bar{A}_{8,1}| &\leq \gamma_{NT} a_{NT} \|M^{(0)}\| \sum_{1\leq i,j\leq N} \|\mathcal{K}_{ij}\| \|M^{(1)} (\varepsilon_i + c_i)\| \|\varepsilon_j + F^0\lambda_j^0 + c_j\| \\ &\leq c_{\mathcal{K}} \gamma_{NT} a_{NT} \|M^{(0)}\| \sum_{1\leq i,j\leq N} \|\varepsilon_j + F^0\lambda_j^0 + c_j\| \\ &\quad \times \left\{O_P\left((NT)^{-1/2}\right) \|F^{0\prime}\varepsilon_i\| + O_P(N^{-1/2}\gamma_{NT}) (\|X_i\| + \|\Delta_i\|)\right\} \\ &= T\gamma_{NT} a_{NT} O_P\left(N^{3/2}T^{1/2}\right) = O_P\left((h!)^{1/4}\right) = o_P(1), \end{aligned}$$

and similarly, $|\bar{A}_{8,4}| \leq c_{\mathcal{K}} O_P(\gamma_{NT}^2 a_{NT}) \|M^{(1)} F^0\| \sum_{1\leq i,j\leq N} \|\lambda_i^0\| (\|X_j\| + \|\Delta_j\|) = N^{-2}T^{-1} O_P(N^{-1/2}T^{1/2}) O_P(N^2T^{1/2}) = o_P(1)$. This completes the proof. ■

Proposition B.9 $A_9 = o_P(1)$ *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 $|A_9|\le c_{\mathcal{K}}a_{NT}\left\|\beta^0-\hat{\beta}\right\|\left\|M^{(0)}\right\|\left\|M^{(2)}+M^{(rem)}\right\|\sum_{1\le i,j\le N}\left(\|\varepsilon_i\|+\left\|F^0\lambda_i^0\right\|+\|c_i\|\right)(\|\varepsilon_j\| +\left\|F^0\lambda_j^0\right\|+\|c_j\|)=TO_P\left(a_{NT}\gamma_{NT}\delta_{NT}^{-2}\right)O_P\left(N^2T\right)=o_P(1)$. ■

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

Proof. First we decompose A_{10} as follows

$$
\begin{aligned}
A_{10} &= a_{NT}\sum_{1\le i,j\le N}\{(\varepsilon_i'+c_i')M^{(1)}\mathcal{K}_{ij}(M^{(2)}+M^{(rem)})\left(\varepsilon_j+F^0\lambda_j^0+c_j\right)\\
&\quad+\lambda_i^{0\prime}F^{0\prime}M^{(1)}\mathcal{K}_{ij}M^{(rem)}\left(\varepsilon_j+F^0\lambda_j^0+c_j\right)+\lambda_i^{0\prime}F^{0\prime}M^{(1)}\mathcal{K}_{ij}M^{(2)}\varepsilon_j\\
&\quad+\lambda_i^{0\prime}F^{0\prime}M^{(1)}\mathcal{K}_{ij}M^{(2)}F^0\lambda_j^0+\lambda_i^{0\prime}F^{0\prime}M^{(1)}\mathcal{K}_{ij}M^{(2)}c_j\}\\
&\equiv A_{10,1}+A_{10,2}+A_{10,3}+A_{10,4}+A_{10,5}.
\end{aligned}
$$

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

$$
\begin{aligned}
|A_{10,1}| &\le a_{NT}\left\|M^{(2)}+M^{(rem)}\right\|\sum_{1\le i,j\le N}\|\mathcal{K}_{ij}\|\left\|M^{(1)}(\varepsilon_i+c_i)\right\|\left(\|\varepsilon_j\|+\left\|F^0\lambda_j^0\right\|+\|c_j\|\right)\\
&\le c_{\mathcal{K}}O_P\left(a_{NT}\delta_{NT}^{-2}\right)\Bigg\{O_P\left((NT)^{-1/2}\right)\sum_{1\le i,j\le N}\left\|F^{0\prime}\varepsilon_i\right\|\left(\|\varepsilon_j\|+\left\|F^0\lambda_j^0\right\|+\|c_j\|\right)\\
&\quad+O_P(N^{-1/2}\gamma_{NT})\sum_{1\le i,j\le N}(\|X_i\|+\|\Delta_i\|)\left(\|\varepsilon_j\|+\left\|F^0\lambda_j^0\right\|+\|c_j\|\right)\Bigg\}\\
&= TO_P\left(a_{NT}\delta_{NT}^{-2}\right)O_P(N^{3/2}T^{1/2}+N^{-1/2}\gamma_{NT}N^2T)=o_P(1).
\end{aligned}
$$

Similarly, $|A_{10,2}|\le c_{\mathcal{K}}a_{NT}\left\|M^{(1)}F^0\right\|\left\|M^{(rem)}\right\|\sum_{1\le i,j\le N}\left\|\lambda_i^0\right\|\left(\|\varepsilon_j\|+\left\|F^0\lambda_j^0\right\|+\|c_j\|\right)\le Ta_{NT}O_P(N^{-1/2}T^{1/2})O_P\left(\delta_{NT}^{-1}\gamma_{NT}\right)O_P(N^2T^{1/2})=o_P(1)$, and $|A_{10,5}|\le c_{\mathcal{K}}O_P(\gamma_{NT})a_{NT}\left\|M^{(1)}F^0\right\|\left\|M^{(2)}\right\|\sum_{1\le i,j\le N}\left\|\lambda_i^0\right\|(\|X_j\|+\|\Delta_j\|)=T\gamma_{NT}a_{NT}O_P\left(N^{-1/2}T^{1/2}\right)O_P\left(\delta_{NT}^{-2}\right)O_P(N^2T^{1/2})=o_P(1)$. This completes the proof. ■

C Proof of Theorem 3.3

By Theorem 3.2, it suffices to prove the theorem by showing that (i) $\hat{B}_{1NT}=B_{1NT}+o_P(1)$ and (ii) $\hat{V}_{NT}=V_{NT}+o_P(1)$ under $\mathbb{H}_1(\gamma_{NT})$. For (i), we apply (B.4) to obtain

$$
\begin{aligned}
\hat{B}_{1NT} &= a_{NT}\sum_{i=1}^N(d_{1i}+d_{2i}+d_{3i}+d_{4i})'\mathcal{K}_{ii}(d_{1i}+d_{2i}+d_{3i}+d_{4i})\\
&= a_{NT}\sum_{i=1}^N\{d_{1i}'\mathcal{K}_{ii}d_{1i}+d_{2i}'\mathcal{K}_{ii}d_{2i}+d_{3i}'\mathcal{K}_{ii}d_{3i}+d_{4i}'\mathcal{K}_{ii}d_{4i}+2d_{1i}'\mathcal{K}_{ii}d_{2i}+2d_{1i}'\mathcal{K}_{ii}d_{3i}\\
&\quad+2d_{1i}'\mathcal{K}_{ii}d_{4i}+2d_{2i}'\mathcal{K}_{ii}d_{3i}+2d_{2i}'\mathcal{K}_{ii}d_{4i}+2d_{3i}'\mathcal{K}_{ii}d_{4i}\}\\
&\equiv \tilde{A}_1+\tilde{A}_2+\tilde{A}_3+\tilde{A}_4+2\tilde{A}_5+2\tilde{A}_6+2\tilde{A}_7+2\tilde{A}_8+2\tilde{A}_9+2\tilde{A}_{10},\ \text{say,}
\end{aligned}
$$

where $a_{NT} \equiv (h!)^{1/2} / (NT)$. Following the proof of Theorem 3.2, it is straightforward to show that under $\mathbb{H}_1(\gamma_{NT})$, $\tilde{A}_1 = B_{1NT} + o_P(1)$ and $\tilde{A}_s = 0$ for $s = 2, 3, ..., 10$. For example, for $\tilde{A}_1$ we have

$$\begin{aligned}
\tilde{A}_1 &= a_{NT} \sum_{i=1}^{N} (\varepsilon_i + c_i)' M_{F^0} \mathcal{K}_{ii} M_{F^0} (\varepsilon_i + c_i) \\
&= a_{NT} \sum_{i=1}^{N} \varepsilon_i' M_{F^0} \mathcal{K}_{ii} M_{F^0} \varepsilon_i + a_{NT} \sum_{i=1}^{N} c_i' M_{F^0} \mathcal{K}_{ii} M_{F^0} c_i + 2 a_{NT} \sum_{i=1}^{N} \varepsilon_i' M_{F^0} \mathcal{K}_{ii} M_{F^0} c_i \\
&\equiv \tilde{A}_{1,1} + \tilde{A}_{1,2} + 2\tilde{A}_{1,3}, \text{ say.}
\end{aligned}$$

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 $a_{NT} \sum_{i=1}^{N} \|\mathcal{K}_{ii}\| \|M_{F^0} c_i\|^2 \leq O_P(\gamma_{NT}^2) a_{NT} \sum_{i=1}^{N} \|\mathcal{K}_{ii}\| (\|X_i\| + \|\Delta_i\|)^2 = O_P(\gamma_{NT}^2 a_{NT}) O_P(NT^2) = O_P(N^{-1}) = o_P(1)$ and $a_{NT} \sum_{i=1}^{N} \|\mathcal{K}_{ii}\| \|M_{F^0} \varepsilon_i\| \|M_{F^0} c_i\| \leq O_P(\gamma_{NT}) a_{NT} \sum_{i=1}^{N} \|\mathcal{K}_{ii}\| \|\varepsilon_i\| (\|X_i\| + \|\Delta_i\|) = O_P(\gamma_{NT} a_{NT}) O_P(NT^2) = O_P(T^{1/2} N^{-1/2} (h!)^{1/4}) = o_P(1)$. It follows that $\tilde{A}_1 = B_{1NT} + o_P(1)$.

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

$$\begin{aligned}
\hat{V}_{NT} - V_{NT} &= 2h! (NT)^{-2} \sum_{1 \leq t,s \leq T} \sum_{1 \leq i \neq j \leq N} \left[\mathcal{K}_{ij,ts}^2 \varepsilon_{it}^2 \varepsilon_{js}^2 - E_{\mathcal{D}} \left(\mathcal{K}_{ij,ts}^2 \varepsilon_{it}^2 \varepsilon_{js}^2 \right) \right] \\
&\quad + 2h! (NT)^{-2} \sum_{1 \leq t,s \leq T} \sum_{1 \leq i \neq j \leq N} \mathcal{K}_{ij,ts}^2 \left(\hat{\varepsilon}_{it}^2 \hat{\varepsilon}_{js}^2 - \varepsilon_{it}^2 \varepsilon_{js}^2 \right) \equiv V_{1NT} + V_{2NT}, \text{ say.}
\end{aligned}$$

Noting that $E_{\mathcal{D}}(V_{1NT}) = 0$ and $E_{\mathcal{D}}(V_{1NT}^2) = O_P(N^{-1})$ 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. Now, write $V_{2NT} = 2h! (NT)^{-2} \sum_{1 \leq t,s \leq T} \sum_{1 \leq i \neq j \leq N} \mathcal{K}_{ij,ts}^2 \left(\hat{\varepsilon}_{it}^2 - \varepsilon_{it}^2 \right) \left(\hat{\varepsilon}_{js}^2 - \varepsilon_{js}^2 \right) + 2h! (NT)^{-2} \sum_{1 \leq t,s \leq T} \sum_{1 \leq i \neq j \leq N} \mathcal{K}_{ij,ts}^2 (\hat{\varepsilon}_{it}^2 - \varepsilon_{it}^2) \varepsilon_{js}^2 + 2h! (NT)^{-2} \sum_{1 \leq t,s \leq T} \sum_{1 \leq i \neq j \leq N} \mathcal{K}_{ij,ts}^2 \varepsilon_{it}^2 \left(\hat{\varepsilon}_{js}^2 - \varepsilon_{js}^2 \right) \equiv 2V_{2NT,1} + 2V_{2NT,2} + 2V_{2NT,3}$. Noting that $V_{2NT,3} = V_{2NT,2}$ as $\mathcal{K}_{ij,ts} = \mathcal{K}_{ji,st}$ by the symmetry of K under Assumption A.6(i), we prove $V_{2NT} = o_P(1)$ by showing that (ii1) $V_{2NT,1} = o_P(1)$, and (ii2) $V_{2NT,2} = o_P(1)$.

To show (ii1), we use $\sum_{i,t}$ to denote $\sum_{i=1}^{N} \sum_{t=1}^{T}$. Let $\bar{K} \equiv \sup_u [k(u)]^p$. By the uniform boundedness of the kernel function K by $\bar{K}$ under Assumption A.6(i), and Cauchy-Schwarz inequality,

$$\begin{aligned}
|V_{2NT,1}| &\leq \bar{K}^2 (h!)^{-1} (NT)^{-2} \sum_{i,t} \sum_{j,s} \left| \left(\hat{\varepsilon}_{it}^2 - \varepsilon_{it}^2 \right) \left(\hat{\varepsilon}_{js}^2 - \varepsilon_{js}^2 \right) \right| \\
&= \bar{K}^2 (h!)^{-1} \left\{ (NT)^{-1} \sum_{i,t} \left| (\hat{\varepsilon}_{it} - \varepsilon_{it}) (\hat{\varepsilon}_{it} + \varepsilon_{it}) \right| \right\}^2 \\
&\leq \bar{K}^2 \left\{ (h!)^{-1} (NT)^{-1} \sum_{i,t} (\hat{\varepsilon}_{it} - \varepsilon_{it})^2 \right\} \left\{ (NT)^{-1} \sum_{i,t} (\hat{\varepsilon}_{it} + \varepsilon_{it})^2 \right\}.
\end{aligned}$$

In view of the fact that $\sum_{i,t} \hat{\varepsilon}_{it}^2 \leq \sum_{i,t} \varepsilon_{it}^2$ and by Markov inequality, $(NT)^{-1} \sum_{i,t} (\hat{\varepsilon}_{it} + \varepsilon_{it})^2 \leq 2 (NT)^{-1} \sum_{i,t} \left(\hat{\varepsilon}_{it}^2 + \varepsilon_{it}^2 \right) \leq 4 (NT)^{-1} \sum_{i,t} \varepsilon_{it}^2 = O_P(1)$. So we can prove $V_{2NT,1} = o_P(1)$ by showing that $V_{2NT,11} \equiv (h!)^{-1} (NT)^{-1} \sum_{i,t} (\hat{\varepsilon}_{it} - \varepsilon_{it})^2 = o_P(1)$. By (B.4), $\hat{\varepsilon}_i - \varepsilon_i = \tilde{d}_{1i} + d_{2i} + d_{3i} + d_{4i}$ where $\tilde{d}_{1i} \equiv d_{1i} - \varepsilon_i$. It follows that

$$
\begin{aligned}
V_{2NT,11} &= (h!)^{-1}(NT)^{-1}\sum_{i=1}^{N}\left\|\tilde{d}_{1i}+d_{2i}+d_{3i}+d_{4i}\right\|_F^2 \\
&\leq 4(h!)^{-1}(NT)^{-1}\sum_{i=1}^{N}\left\{\left\|\tilde{d}_{1i}\right\|_F^2+\|d_{2i}\|_F^2+\|d_{3i}\|_F^2+\|d_{4i}\|_F^2\right\} \\
&= 4V_{2NT,11a}+V_{2NT,11b}+V_{2NT,11c}+V_{2NT,11d},\text{ say.}
\end{aligned}
$$

Noting that $\|P_{F^0}\varepsilon_i\|_F^2 = O_P\left(T^{-1}\right)\left\|F^{0\prime}\varepsilon_i\right\|_F^2$ and $\|M_{F^0}c_i\|_F^2 \leq \|c_i\|_F^2 = O\left(\gamma_{NT}^2\right)\left(\|X_i\|_F^2+\|\Delta_i\|_F^2\right)$, we have by Assumptions A.4(iii) and A.5 and Markov inequality

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

Similarly,

$$
\begin{aligned}
V_{2NT,11b} &\leq \left\|\beta^0-\hat{\beta}\right\|^2\sum_{k=1}^{p}\left\|M_k^{(0)}\right\|(h!)^{-1}(NT)^{-1}\sum_{i=1}^{N}\left\|\varepsilon_i+F^0\lambda_i^0+c_i\right\|_F^2 = O_P\left(\gamma_{NT}^2(h!)^{-1}\right), \\
V_{2NT,11c} &\leq \left\|M^{(1)}\right\|_F^2(h!)^{-1}(NT)^{-1}\sum_{i=1}^{N}\left\|\varepsilon_i+F^0\lambda_i^0+c_i\right\|_F^2 = O_P\left(N^{-1}(h!)^{-1}\right), \\
V_{2NT,11d} &= \left\|M^{(2)}+M^{(rem)}\right\|_F^2(h!)^{-1}(NT)^{-1}\sum_{i=1}^{N}\left\|\varepsilon_i+F^0\lambda_i^0+c_i\right\|_F^2 = O_P\left(\delta_{NT}^{-4}(h!)^{-1}\right).
\end{aligned}
$$

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 : $K(u) = \int K^0(v)K^0(u-v)\,dv$. Define K_h^0 analogously as K_h. By Minkowski inequality, the fact that $\mathcal{K}_{ij,ts} = K_h(X_{it}-X_{js}) = \int K_h^0(X_{it}-x)K_h^0(X_{js}-x)\,dx$, Fubini theorem, and Cauchy-Schwarz inequality,

$$
\begin{aligned}
|V_{2NT,2}| &\leq h!(NT)^{-2}\sum_{i,t}\sum_{j,s}\mathcal{K}_{ij,ts}^2\left|\hat{\varepsilon}_{it}^2-\varepsilon_{it}^2\right|\varepsilon_{js}^2 \\
&= h!(NT)^{-2}\int\int\sum_{i,t}\left|\hat{\varepsilon}_{it}^2-\varepsilon_{it}^2\right|K_h^0(X_{it}-x)K_h^0(X_{it}-\bar{x})\sum_{j,s}\varepsilon_{js}^2K_h^0(X_{js}-x)K_h^0(X_{js}-\bar{x})\,dx\,d\bar{x} \\
&\leq \{V_{2NT,21}V_{2NT,22}\}^{1/2},
\end{aligned}
$$

where $V_{2NT,21} = h!\,(NT)^{-2} \int\int \left[\sum_{i,t} \left|\hat{\varepsilon}_{it}^2 - \varepsilon_{it}^2\right| K_h^0 (X_{it} - x)\, K_h^0 (X_{it} - \bar{x})\right]^2 dx d\bar{x}$ and $V_{2NT,22} = h!\,(NT)^{-2}$ $\int\int \left[\sum_{j,s} \varepsilon_{js}^2 K_h^0 (X_{js} - x)\, K_h^0 (X_{js} - \bar{x})\right]^2 dx d\bar{x}$. Again, by the relationship between K and K^0, the study of $V_{2NT,1}$, Markov inequality, and Assumption A.7(i), we have $V_{2NT,21} = h!\,(NT)^{-2} \sum_{i,t} \sum_{j,s} \left|\hat{\varepsilon}_{it}^2 - \varepsilon_{it}^2\right|$ $\left|\hat{\varepsilon}_{js}^2 - \varepsilon_{js}^2\right| [K_h (X_{it} - X_{js})]^2 = O_P((T^{-1}+N^{-1}+\gamma_{NT}^2)\,(h!)^{-1}) = o_P(1)$, and $V_{2NT,22} = h!\,(NT)^{-2} \sum_{i,t} \sum_{j,s}$ $\varepsilon_{js}^2 \varepsilon_{it}^2 [K_h (X_{it} - X_{js})]^2 = O_P(1)$. 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).

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* $c_{\mathcal{K}} \equiv \max_{1\le i,j\le N} \|\mathcal{K}_{ij}\| = O_P(T)$.

Proof. Noting that $\|\mathcal{K}_{ij}\|^2 \le \|\mathcal{K}_{ij}\|_1 \|\mathcal{K}_{ij}\|_\infty$ where $\|\mathcal{K}_{ij}\|_1 = \max_{1\le s\le T} \sum_{t=1}^T |K_h (X_{it} - X_{js})|$ and $\|\mathcal{K}_{ij}\|_\infty = \max_{1\le t\le T} \sum_{s=1}^T |K_h (X_{it} - X_{js})|$, it suffices to prove the lemma by showing that (i) $\max_{1\le i,j\le N} T^{-1} \|\mathcal{K}_{ij}\|_1 = O_P(1)$ and (ii) $\max_{1\le i,j\le N} T^{-1} \|\mathcal{K}_{ij}\|_\infty = O_P(1)$. 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 $\bar{\eta}_{iT,js} \equiv T^{-1} \sum_{t=1}^T K_h (X_{it} - X_{js})$ $\times 1\{\|X_{js}\| \le c_{NT}\}$. Then by Markov inequality, dominated convergence theorem, and Assumption A.5(i), for any $\epsilon^* > 0$

$$
\begin{aligned}
& \Pr\left(\max_{1\le i,j\le N} \max_{1\le s\le T} \left|\eta_{iT,js} - \bar{\eta}_{iT,js}\right| \ge \epsilon^*\right) \\
= \; & \Pr\left(\max_{1\le i,j\le N} \max_{1\le s\le T} T^{-1} \sum_{t=1}^T K_h (X_{it} - X_{js})\, 1\{\|X_{js}\| > c_{NT}\} \ge \epsilon^*\right) \\
\le \; & \Pr\left(\max_{1\le j\le N} \max_{1\le s\le T} \|X_{js}\| > c_{NT}\right) \le \sum_{j=1}^N \sum_{s=1}^T \Pr\left(\|X_{js}\|^{q_0} > c_{NT}^{q_0}\right) \\
\le \; & \frac{1}{c_{NT}^{q_0}} \sum_{j=1}^N \sum_{s=1}^T E\left[\|X_{js}\|^{q_0}\, 1\left(\|X_{js}\|^{q_0} > c_{NT}^{q_0}\right)\right] = o(1).
\end{aligned}
$$

It follows that we can prove (i) by showing that $L_{NT} \equiv \max_{1\le i\le N} L_{iNT} = O_P(1)$, where $L_{iNT} \equiv \max_{\|x\|\le c_{NT}} T^{-1} \sum_{t=1}^T K_h (X_{it} - x)$. By the Minkowski inequality

$$
\begin{aligned}
L_{iNT} \;\le\; & \max_{\|x\|\le c_{NT}} \left|T^{-1} \sum_{t=1}^T K_h (X_{it} - x) - E_{\mathcal{D}}[K_h (X_{it} - x)]\right| + \max_{\|x\|\le c_{NT}} \left|T^{-1} \sum_{t=1}^T E_{\mathcal{D}}[K_h (X_{it} - x)]\right| \\
\equiv\; & L_{iNT,1} + L_{iNT,2}, \text{ say.} \qquad \text{(D.1)}
\end{aligned}
$$

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

$$
\max_{1\le i\le N} L_{iNT,2} = \max_{1\le i\le N} \max_{\|x\|\le c_{NT}} \left|T^{-1} \sum_{t=1}^T \int f_{i,t}(x + h \odot u)\, K(u)\, du\right| \le C_f, \qquad \text{(D.2)}
$$

where $\odot$ denotes the Hadamard product. If $\{X_{it}, t = 1, 2, ...\}$ is strictly stationary and strong mixing, then one could replace $E_{\mathcal{D}}\left[K_h\left(X_{it} - x\right)\right]$ 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, $\{X_{it}, t = 1, 2, ...\}$ is conditionally strong mixing given $\mathcal{D}$ so Hansen's (2008) results are not applicable. We complete the proof of (i) by showing that

$$\max_{1 \le i \le N} L_{iNT,1} = o_P(1). \tag{D.3}$$

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

$$P_{\mathcal{D}}\left(\max_{1 \le i \le N} \sup_{1 \le l \le Q} |\varphi_{iT}(x_l)| \ge \epsilon C_1\right) = o(1), \tag{D.4}$$

where $P_{\mathcal{D}}(\cdot) = P(\cdot|\mathcal{D})$, $\varphi_{iT}(x) = (Th!)^{-1} \sum_{t=1}^T Z_{i,t}(x)$ and $Z_{i,t}(x) = h!\{K_h(X_{it} - x) - E_{\mathcal{D}}[K_h(X_{it} - x)]$. Let $\bar{K} \equiv \sup_u [k(u)]^p$. Noting that $\max_{1 \le i \le N} \max_{1 \le t \le T} \sup_x |Z_{i,t}(x)| \le \bar{K}$ and $\max_{1 \le i \le N} \max_{1 \le t \le T} \sup_x E_{\mathcal{D}}[Z_{i,t}(x)]^2 \le C_2 h!$ for some $C_2 < \infty$, we can apply Boole's inequality and the exponential inequality for conditional strong mixing processes (see Lemma E.2 in the supplemental appendix) to bound the left hand side of (D.4) from above by

$$\begin{aligned}
\sum_{i=1}^N \sum_{l=1}^Q P_{\mathcal{D}}\left(|\varphi_{iT}(x_l)| \ge \epsilon C_1\right) &\le NQ \max_{1 \le i \le N} \sup_{1 \le l \le Q} P_{\mathcal{D}}\left(\left|T^{-1} \sum_{t=1}^T Z_{i,t}(x_l)\right| \ge C_1 \epsilon h!\right) \\
&\le 2NQ\left[\tau \exp\left(-\frac{T C_1^2 \epsilon^2 (h!)^2}{4\tau C_2 h! + 2\bar{K}\ C_1 \epsilon h! \tau/3}\right) + T\alpha_{\mathcal{D}}(\tau)\right] \\
&\le 2NQ\left[\tau \exp\left(-\frac{C_1^2 \epsilon^2 T h!}{4 C_2 \tau + 2 C_1 \bar{K} \epsilon \tau/3}\right) + T\alpha_{\mathcal{D}}(\tau)\right] \\
&\to 0 \text{ as } (N, T) \to \infty,
\end{aligned}$$

provided that $\tau \in (1, Th!)$ such that $Th!/\tau \gg T^\eta$ for some $\eta > 0$ and $(NT)^{(1+p/q_0)} (h!)^{-1} \alpha_{\mathcal{D}}(\tau) = o_{a.s.}(1)$. Assumption A.4(i) ensures the existence of such a τ. 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. ■

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

(i) $D_{1,1} \equiv \gamma_{NT} a_{NT} \sum_{1 \le i,j \le N} \varepsilon_i' M_{F^0} \mathcal{K}_{ij} M_{F^0} \Delta_j = o_P(1)$;
(ii) $D_{1,2} \equiv \gamma_{NT} a_{NT} \sum_{1 \le i,j \le N} \varepsilon_i' M_{F^0} \mathcal{K}_{ij} M_{F^0} X_j = o_P(1)$;
(iii) $D_{1,3} \equiv \gamma_{NT} a_{NT} \sum_{1 \le i,j \le N} \varepsilon_i' M_{F^0} \mathcal{K}_{ij} M^{(0)} F^0 \lambda_j^0 = o_P(1)$;
(iv) $D_{1,4} \equiv a_{NT} \sum_{1 \le i,j \le N} c_i' M_{F^0} \mathcal{K}_{ij} M^{(1)} F^0 \lambda_j^0 = o_P(1)$;
(v) $D_{1,5} \equiv \gamma_{NT} a_{NT} \sum_{1 \le i,j \le N} \lambda_i^{0\prime} F^{0\prime} M^{(1)} \mathcal{K}_{ij} M^{(0)} F^0 \lambda_j^0 = o_P(1)$.

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

(i) $D_{2,1} \equiv a_{NT} \sum_{1 \le i \ne j \le N} \varepsilon_i' P_{F^0} \mathcal{K}_{ij} \varepsilon_j = o_P(1)$;
(ii) $D_{2,2} \equiv a_{NT} \sum_{1 \le i \ne j \le N} \varepsilon_i' P_{F^0} \mathcal{K}_{ij} P_{F^0} \varepsilon_j = o_P(1)$;

(iii) $D_{2,3} \equiv a_{NT} \sum_{1 \le i,j \le N} \lambda_i^{0\prime} \left(\lambda^{0\prime}\lambda^0\right)^{-1} \lambda^{0\prime} \varepsilon \mathcal{K}_{ij} \varepsilon' \lambda^0 \left(\lambda^{0\prime}\lambda^0\right)^{-1} \lambda_j^0 = o_P(1)$;
(iv) $D_{2,4} \equiv a_{NT} \sum_{1 \le i,j \le N} \lambda_i^{0\prime} \left(\lambda^{0\prime}\lambda^0\right)^{-1} \lambda^{0\prime} \varepsilon \mathcal{K}_{ij} P_{F^0} \varepsilon' \lambda^0 \left(\lambda^{0\prime}\lambda^0\right)^{-1} \lambda_j^0 = o_P(1)$;
(v) $D_{2,5} \equiv \gamma_{NT} a_{NT} \sum_{1 \le i \ne j \le N} \varepsilon_i' M_{F^0} \mathcal{K}_{ij} M^{(0)} \varepsilon_j = o_P(1)$;
(vi) $D_{2,6} \equiv a_{NT} \sum_{1 \le i,j \le N} \varepsilon_i' M_{F^0} \mathcal{K}_{ij} M^{(1)} F^0 \lambda_j^0 = o_P(1)$;
(vii) $D_{2,7} \equiv a_{NT} \sum_{1 \le i \ne j \le N} \varepsilon_i' M_{F^0} \mathcal{K}_{ij} M^{(1)} c_j = o_P(1)$;
(viii) $D_{2,8} \equiv a_{NT} \sum_{1 \le i \ne j \le N} \lambda_i^{0\prime} F^0 M^{(1)} \mathcal{K}_{ij} M^{(0)} \varepsilon_j = o_P(1)$.

Lemma D.4 *Suppose the conditions in Theorem 3.2 hold. Then*
(i) $D_{3,1} \equiv a_{NT} \sum_{1 \le i,j \le N} \varepsilon_i' M^{(1)} \mathcal{K}_{ij} M^{(1)} F^0 \lambda_j^0 = o_P(1)$;
(ii) $D_{3,2} \equiv a_{NT} \sum_{1 \le i \ne j \le N} \varepsilon_i' M_{F^0} \mathcal{K}_{ij} M^{(1)} \varepsilon_j = o_P(1)$;
(iii) $D_{3,3} \equiv a_{NT} \sum_{1 \le i,j \le N} \varepsilon_i' M_{F^0} \mathcal{K}_{ij} M^{(2)} F^0 \lambda_j^0 = o_P(1)$;
(iv) $D_{3,4} \equiv a_{NT} \sum_{1 \le i,j \le N} \lambda_i' F^0 M^{(1)} \mathcal{K}_{ij} M^{(2)} F^0 \lambda_j^0 = o_P(1)$.

Lemma D.5 *Suppose the conditions in Theorem 3.2 hold. Then*
(i) $D_{4,1} \equiv a_{NT} \sum_{1 \le i \ne j \le N} \varepsilon_i' M_{F^0} \mathcal{K}_{ij} M^{(2)} \varepsilon_j = o_P(1)$;
(ii) $D_{4,2} \equiv a_{NT} \sum_{1 \le i \ne j \le N} \lambda_i^{0\prime} F^{0\prime} M^{(1)} \mathcal{K}_{ij} M^{(2)} \varepsilon_j = o_P(1)$.

References

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

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

Andrews, D. W. K., 2005. Cross-section regression with common shocks. Econometrica 73, 1551-1585.

Avarucci, M., Zafaroni, P., 2012. Generalized least squares estimation of panel with common shocks. Working paper, Imperial College London.

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

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

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

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

Barro, R., 1991. Economic growth in a cross section of countries. Quarterly Journal of Economics 106, 407-443.

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

Bierens, H. J., 1982. Consistent model specification tests. Journal of Econometrics 20, 105-134.

Bierens, H. J., 1990. A consistent conditional moment test of functional form. Econometrica 58, 1443-1458.

Bond, S., Leblebicioglu, A., Schiantarelli, F., 2010. Capital accumulation and growth: a new look at the empirical evidence. Journal of Applied Econometrics 25, 1073-1099.

Carneiro, P., Hansen, K., Heckman, J., 2003. Estimating distributions of counterfactuals with an application to the returns to schooling and measurement of the effect of uncertainty on schooling choice. International Economic Review 44, 361-422.

Chen, S., Gao, J., 2007. An adaptive empirical likelihood test for parametric time series regression models. Journal of Econometrics 141, 950-972.

Chudik, A., Pesaran M. H., 2013. Common correlated effects estimation of heterogenous dynamic panel data models with weakly exogenous regressors. Working paper, Department of Economics, University of Southern California.

Cunha, F., Heckman, J., Navarro, S., 2005. Separating uncertainty from heterogeneity in life cycle earnings. Oxford Economic Papers 57, 191-261.

de Jong, P., 1987. A central limit theorem for generalized quadratic form. Probability Theory and Related Fields 75, 261-277.

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

Feldstein, M., Horioka, C., 1980. Domestic saving and international capital flows. The Economic Journal 90, 314-329.

Freyberger, J., 2012. Nonparametric panel data models with interactive fixed effects. Working paper, Department of Economics, Northwestern University.

Gao, J., Gijbels, I., 2008. Bandwidth selection in nonparametric kernel testing. Journal of the American Statistical Association 103, 1584-1594.

Giannone, D., Lenza, M., 2010. The Feldstein-Horioka fact, in The NBER International Seminar on Macroeconomics, Reichlin L. & K. West (eds.), University of Chicago Press, 103-117.

Gagliardini, P., Gouriéroux, C., 2012. Efficiency in large dynamic panel models with common factor. Working paper, Swiss Finance Institute.

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

Hahn, J., Kuersteiner, G., 2011. Bias reduction for dynamic nonlinear panel models with fixed effects. Econometric Theory 27, 1152-1191.

Hall, P., Heyde C. C., 1980. Martingale Limit Theory and Its Applications. Academic Press, New York.

Hansen, B. E., 1999. Testing for linearity. Journal of Economic Surveys 13, 551-576.

Hansen, B. E., 2000. Testing for structural change in conditional models. Journal of Econometrics 97, 93-115.

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

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

Hausman, J. A., 1978. Specification testing in econometrics. Econometrica 46, 1251-1271.

Hjellvik, V., Yao, Q., Tjøstheim, D., 1998. Linearity testing using local polynomial approximation. Journal of Statistical Planning and Inference 68, 295-321.

Hjellvik, V., Tjøstheim, D., 1995. Nonparametric tests of linearity for time series. Biometrika 82, 351-368.

Hong, Y., White, H., 1995. Consistent specification testing via nonparametric series regression. Econometrica 63, 1133-1159.

Horowitz, J. L., Spokoiny, V. G., 2001. An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. Econometrica 69, 599-631.

Hsiao, C., Li Q., Racine, J., 2007. A consistent model specification test with mixed discrete and continuous data. Journal of Econometrics 140, 802–826.

Jin, S., Su, L., 2013. A nonparametric poolability test for panel data models with cross section dependence. Econometric Reviews 32, 469-512.

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

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

Lee, Y., 2013. Nonparametric estimation of dynamic panel models with fixed effects. Econometric Theory, forthcoming.

Lee, Y-J., 2014. Testing a linear dynamic panel data model against nonlinear alternatives. Journal of Econometrics 178, 146-166.

Li, Q., Wang, S., 1998. A simple consistent bootstrap test for a parametric regression function. Journal of Econometrics 87, 145-165.

Lin, Z., Li, Q., Sun, Y., 2014. A consistent nonparametric test of parametric regression functional form in fixed-effects panel data models. Journal of Econometrics 178, 167-179.

Liu, Z., Stegnos, T., 1999. Non-linearities in cross-country growth regression: a semiparametric approach. Journal of Applied Econometrics 14, 527-538.

Lu, X., Su, L., 2013. Shrinkage estimation of dynamic panel data models with interactive fixed effects. Working paper, HKUST.

Lucas, R., 1988. On the mechanics of economic development. Journal of Monetary Economics 22, 3-42.

Ludvigson S. C., Ng, S., 2009. Macro factors in bond risk premia. The Review of Financial Studies 22, 5027-5067.

Ludvigson S. C., Ng, S., 2011. A factor analysis of bond risk premia. Handbook of Empirical Economics and Finance, Ullah A. & D. Giles (eds.), Chapman and Hall, pp. 313-372.

Moon, H., Weidner, M., 2010. Dynamic linear panel regression models with interactive fixed effects. Working paper, Department of Economics, USC.

Moon, H., Weidner, M., 2013. Linear regression for panel with unknown number of factors as interactive fixed effects. Working paper, Department of Economics, University College London.

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

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

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

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

Ramsey, J. B., 1969. Tests for specification errors in classical linear least squares regression analysis. Journal of the Royal Statistical Society: Series B 31, 350-371.

Romer, P., 1986. Increasing returns and long-run growth. Journal of Political Economy 94, 1002-1037.

Ross, S., 1976. The arbitrage theory of capital asset pricing. Journal of Economic Theory 13, 341-360.

Roussas, G. G., 2008. On conditional independence, mixing and association. Stochastic Analysis and Applications 26, 1274-1309.

Solow, R.,1956. A contribution to the theory of economic growth. Quarterly Journal of Economics 70, 65-94.

Song, M., 2013. Asymptotic theory for dynamic heterogeneous panels with cross-sectional dependence and its applications. Working paper, Columbia University.

Stinchcombe, M., White, H., 1998. Consistent specification testing with nuisance parameters present only under alternative. Econometric Theory 14, 295-324.

Su, L., Chen, Q., 2013. Testing homogeneity in panel data models with interactive fixed effects. Econometric Theory 29, 1079-1135.

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

Su, L., Lu, X., 2013. Nonparametric dynamic panel data models: kernel estimation and specification testing. Journal of Econometrics 176, 112-133.

Su, L., Ullah, A., 2013. A nonparametric goodness-of-fit-based test for conditional heteroskedasticity. Econometric Theory 29, 187-212.

Su, L., Zhang, Y., 2013. Nonparametric dynamic panel data models with interactive fixed effects: sieve estimation and specification testing. Working paper, Singapore Management University.

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

Wooldridge, J. M., 1992. A test for functional form against nonparametric alternatives. Econometric Theory 8, 452-475.

Yatchew, A. J., 1992. Nonparametric regression tests based on least squares. Econometric Theory 8, 435-451.

Zheng, X., 1996. Consistent test of functional form via nonparametric estimation technique. Journal of Econometrics 75, 263-289.