Specification Tests for Time-Varying Coefficient Panel Models*

Alev Atak[a], Yonghui Zhang[b*], Qiankun Zhou[c]

[a]*Department of Economics, City, University of London*

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

[c]*Department of Economics, Louisiana State University*

February 4, 2019

Abstract

This paper provides a nonparametric test for the commonly-used structure, the homogeneity and stability, on the parameters in panels. We first get the augmented residuals by estimating the model under the null hypothesis of homogeneity and stability, then run auxiliary time series regressions of residuals on the regressors with time-varying coefficients via sieve methods. The test statistic is constructed by averaging the squared fitted values, which is close to zero under the null and deviates from zero under the alternative. We show that the test statistic, after being appropriately standardized, is asymptotically normally distributed under the null and a sequence of Pitman local alternatives as both cross-sectional and time dimensions tend to infinity. A bootstrap procedure is proposed to improve the finite sample performance of our test. Monte Carlo simulations indicate that our test performs reasonably well in finite samples. We apply the test to study the Environmental Kuznets Curve in U.S. and reject the homogeneity and stablility of the coefficients for all states. In addition, we extend the procedure to test other structures such as the homogeneity of time-varying coefficients or the stability of heterogeneous coefficients.

Key words: Homogeneity, Panel data, Nonparametric test, Sieve method, Stability, Time-varying coefficient

JEL Classification: C12, C23, C33, C38, C52.

*We are grateful to Jiti Gao for helpful comments and discussions. Address correspondence to: Yonghui Zhang, School of Economics, Renmin University of China, 59 Zhongguancun Road, Beijing, 100872, China; Phone: +86-10-82500732; e-mail: yonghui.zhang@hotmail.com. Zhang gratefully acknowledges the financial support from the National Natural Science Foundation of China (Project No.71401166). All errors are the authors' sole responsibilities.

1 Introduction

The relationship between economic variables usually changes slowly over a long time span, which is possibly influenced by preference change, technological progress, or some other driving forces such as institutional transformation, economic transition, policy switch, etc., see Chen and Hong (2012). For this reason, mainly motivated by time-varying or functional coefficient models in the literature of semiparametric regression, numerous studies have been devoted to capture the important feature of time-varying coefficients (TVC) or smoothing time trends in the panel data framework. For example, Li et al. (2011) propose a local linear dummy variable approach for estimating panel models with TVC, which is an extension of Cai et al.'s (2000) and Cai's (2007) TVC time series models; Robinson (2012) studies the kernel estimation of nonparametric trending panel data models with cross-sectional dependence; Chen et al. (2012) include exogenous regressors in Robinson's (2012) nonparametric panel trending model with a partially linear structure; Atak et al. (2011) adopt a semiparametric unbalanced panel data model with smoothing time trends to study the climate change in the United Kingdom. For other related works on time-varying or functional coefficients panel data models, see Zhao et al. (2016), Gao et al. (2018), among many others.

Almost all the aforementioned papers assume that all cross-sectional units in panels share the same vector of constant coefficients, and that the heterogeneity among individual units is fully captured by the additive unobservable individuals fixed effects. Even if the homogeneity assumption greatly reduces the dimension of parameter space, and significantly simplifies the processes of estimation and inference, however, this assumption may be inappropriate in practice and the restricted estimator with homogeneity may cause a biased estimator for the cross-sectional simple "average" or "mean" of slopes, and further lead to misleading conclusions (e.g., Hsiao and Tahmiscioglu (1997) and Lee et al. (1997)). A conservative way is to allow individual-specific or group-specific slope coefficients. For example, Ma et al. (2018) consider testing empirical asset pricing models with individual-specific time-varying factor loadings and intercepts; Su et al. (2018) propose a heterogeneous time-varying panel data model with a latent group structure and apply the classified-Lasso of Su et al. (2016) to estimate the TVCs and group memberships jointly; Liu et al. (2018) study a class of time-varying panel data models with individual-specific regression coefficients in the presence of common factors, and propose a unified semiparametric profile method to estimate the TVCs and the factor loadings simultaneously.

Since the specification of stability and/or homogeneity of coefficients plays a critical role in obtaining consistent estimation and valid statistical inference for panel data models, it is necessary and prudent for researchers to carry out certain specification or diagnostic tests before embarking on the estimation with such restrictions. However, there are only several specification tests for the heterogenous time-varying panel data models. For example, Zhang et al. (2012) and Hidalgo and Lee (2014) propose nonparametric tests for the common time trends in a semiparametric panel data model with homogeneous linear slopes; Chen and Huang (2018) suggest a nonparametric Wald-type test for the stability of coefficients while assuming that all the coefficients are homogenous among individuals; Gao et al. (2018) provide a test for homogeneity of constant slopes while allowing individual-specific and nonparametric time trends; Ma et al. (2018) test whether all the individual-specific time trends are equal to zero jointly for the asset pricing model with heterogenous time-varying factor loadings.

Yet there is no available test for the joint structure of homogeneity and stability on the coefficients for panel data models. The joint structure implies that all the coefficients in panels are fixed constant along both the time series and cross-sectional dimensions, i.e., the usual homogeneous linear panel data model, which is the simplest and most widely-used specification in empirical studies. To fill the gap, in this paper, we provide a nonparametric test for the joint structure on the heterogeneous TVC panel data model. We first estimate the model under the null hypothesis and obtain the augmented residuals, which consistently estimate the sums of fixed effect and the disturbance errors if the null is true. Then we run auxiliary time series regressions of the augmented residuals on regressors and constant with TVCs via the sieve method and propose a testing statistic by averaging all the squared fitted values across individuals and time periods. By construction, the testing statistic is close to 0 under the null and deviates from 0 under the alternative. We show that the test statistic, after being appropriately standardized, is asymptotically normally distributed under both the null and a sequence of Pitman local alternatives when both cross-sectional and time dimensions tend to infinity. A bootstrap procedure is proposed to improve the finite sample performance of the test. Extensions of the proposed testing to other commonly-used specifications such as the homogeneity of TVCs and the stability of heterogenous coefficients in panels are also discussed. Monte Carlo simulations indicate that the proposed test performs reasonably well in finite samples in a variety setup of data generating processes. We apply our test to Environmental Kuznets Curve estimation and reject the assumption of homogeneous and stable coefficients in the model.

The rest of the paper is organized as follows. In Section 2, we introduce the basic framework including the model, the hypothesis of interest, and the proposed test based on the estimation under the null hypothesis. The large sample theory for the proposed test and extension of the test for models with homogeneous TVC or stable heterogeneous coefficients are provided in Section 3 and Section 4, respectively. Section 5 conducts a set of Monte Carlo simulations to investigate the finite sample performance of our test. We apply our proposed test to study the Environmental Kuznets Curve (EKC) in US in Section 6. Section 7 concludes. The proofs for main theorems and the lemmas, additional simulation results are contained in appendix.

Notation. We use $\lambda_{\min}(A)$, $\lambda_{\max}(A)$ and $\mathrm{tr}(A)$ to denote the smallest eigenvalue, largest eigenvalue and the trace of a matrix A, respectively. For any $n \times m$ matrix A, let A' be its transpose, $\|A\| \equiv \sqrt{\mathrm{tr}(A'A)}$ its Frobenius norm, $P_A = A(A'A)^{-1}A'$ and $M_A = I_m - P_A$, where I_m is an $m \times m$ identity matrix. We use p.s.d. (p.d.) for the abbreviation for "positive semi-definite (positive definite)". The symbols $\to_p$ and $\to_d$ denote convergence in probability and in distribution, respectively. $(N,T) \to \infty$ signifies that N and T tend to infinity jointly.

2 Basic Framework

In this section, we first introduce the heterogeneous TVC panel data model and the main hypothesis of interest, then discuss the motivation of our testing approach with the restricted estimation under the null hypothesis, and finally propose a testing statistic based on auxiliary time series regressions with a TVC structure.

2.1 The Model and Hypothesis

We consider the following heterogeneous TVC panel data models with fixed effects and time trend

$$Y_{it} = X_{it}'\beta_{it} + f_{it} + \alpha_i + \varepsilon_{it},\ i = 1, \ldots, N, t = 1, \ldots, T, \tag{2.1}$$

where Y_{it} is a scalar, X_{it} is a d-vector of time-varying exogenous explanatory variables which may include some common regressors such as macroeconomic variables or financial factors, α_i represents the individual-specific unobservable effect which may be correlated with the regressors X_{it}. β_{it} is a vector of deterministic time-varying coefficients and f_{it} is the time trend for the ith individual. For the idiosyncratic error ε_{it}, we follow Su et al. (2018) and assume

$$\varepsilon_{it} = \sigma_{it}\epsilon_{it} \text{ with } \sigma_{it}^2 = \sigma_i^2 (X_{it}, t/T), \tag{2.2}$$

where ϵ_{it} has zero mean and variance one conditional on X_{it}.

Following the literature of nonparametric time-varying regressions (e.g. Cai (2007), Robinson (1989, 1991, 2012), Li et al. (2011), Zhang et al. (2012), Chen et al. (2012), Chen and Huang (2018)), we assume that for each i both slope β_{it} and trend f_{it} change slowly over a long time span as follows

$$\beta_{it} = \beta_i (\tau_t) \text{ and } f_{it} = f_i (\tau_t) \text{ for } t = 1, \ldots, T, \tag{2.3}$$

where $\tau_t \equiv t/T$ is the time regressor, and $\beta_i (\cdot) : [0, 1] \to \mathbb{R}^d$ and $f_i (\cdot) : [0, 1] \to \mathbb{R}$ are all unknown smooth functions. To identify $f_i (\cdot)$ and α_i in (2.1), we impose that[1]

$$\int_0^1 f_i (\tau) \, d\tau = 0 \text{ for } i = 1, \ldots, N. \tag{2.4}$$

Denote the component in Y_{it} explained by regressors (X_{it}) and 1 with TVCs as[2]

$$g_{it} \equiv g_i (X_{it}, \tau_t) \equiv X_{it}'\beta_{it} + f_{it}. \tag{2.5}$$

The models specified in (2.1) and (2.3) are quite general and include various panel data models in the literature as special cases when different *structures* are imposed on the unknown functions $\beta_i (\cdot)$'s and $f_i (\cdot)$'s:

1. If $\beta_i (\cdot) = \beta$ and $f_i (\cdot) = 0$ for all i's, then model (2.1) reduces to the usual homogeneous linear panel data model with fixed effects in standard textbooks (see Baltagi (2012), Hsiao (2014) and Pesaran (2015)):

$$Y_{it} = X_{it}'\beta + \alpha_i + \varepsilon_{it}; \tag{2.6}$$

2. when $\beta_i (\cdot) = \beta_i$ and $f_i (\cdot) = 0$ for each i, then model (2.1) becomes the heterogeneous linear panel data model with fixed effects (see Hsiao (2014), Pesaran (2015) and Hsiao and Pesaran (2008)):

$$Y_{it} = X_{it}'\beta_i + \alpha_i + \varepsilon_{it}; \tag{2.7}$$

[1]Alternatively, we can impose that $f_i (c^*) = 0$ for $i = 1, \ldots, N$ and $c^* \in [0, 1]$.

[2]Clearly, the setup in (2.1) and (2.3) can be easily generalized to allow for a mixture structure such as

$$Y_{it} = X_{1,it}'\beta_{1,it} + X_{2,it}'\beta_{2,i} + X_{3,it}'\beta_{3,t} + X_{4,it}'\beta_4 + \alpha_i + \varepsilon_{it},$$

where the time trends (f_{it} or f_t) can be aborbed in the first or third components. To simply the illustruation, we focus on the model with a fully heterogneous TVCs.

3. when $\beta_i(\cdot) = \beta(\cdot)$ and $f_i(\cdot) = f(\cdot)$ for $i = 1, \ldots, N$, then model (2.1) is the panel data model with homogeneous TVCs studied by Chen and Huang (2018), Chen et al. (2012), Silvapulle et al. (2016), and Li et al. (2011):

$$Y_{it} = f(\tau_t) + X_{it}'\beta(\tau_t) + \alpha_i + \varepsilon_{it}; \tag{2.8}$$

4. when $\beta_i(\cdot) = \beta_i$ or β and $f_i(\cdot) \neq 0$ or $f_i(\cdot) = f(\cdot) \neq 0$, then model (2.1) becomes the following homogeneous or heterogeneous linear panel data models with homogeneous or heterogeneous nonparametric time trends:

$$Y_{it} = f(\tau_t) + X_{it}'\beta + \alpha_i + \varepsilon_{it}, \tag{2.9}$$

$$Y_{it} = f_i(\tau_t) + X_{it}'\beta + \alpha_i + \varepsilon_{it}, \tag{2.10}$$

$$Y_{it} = f(\tau_t) + X_{it}'\beta_i + \alpha_i + \varepsilon_{it}, \tag{2.11}$$

$$Y_{it} = f_i(\tau_t) + X_{it}'\beta_i + \alpha_i + \varepsilon_{it}, \tag{2.12}$$

   where models (2.9)-(2.12) have been studied by Chen et al. (2012), Zhang et al. (2012), and Atak et al. (2012), Gao et al. (2018), respectively.

5. when there is no regressors ($\beta_i(\cdot) = 0$ for all $i = 1, \ldots, N$), then model (2.1) becomes the nonparametric trending panel data models:

$$Y_{it} = f(\tau_t) + \alpha_i + \varepsilon_{it}, \tag{2.13}$$

$$Y_{it} = f_i(\tau_t) + \alpha_i + \varepsilon_{it}, \tag{2.14}$$

   where the homogeneous trending model (2.13) has been studied by Robinson (2012) and model (2.14) allows for heterogeneous trending behavior.

6. when there exists an unknown group structure for coefficients β_{it}'s (i.e., $\beta_{it} = \beta_{jt}$ when i and j lie in the same group), model (2.1) becomes the heterogeneous linear panel data model with time-invariant coefficients in Su et al. (2016) or the heterogeneous panel data model with slowly varying coefficients in Su et al. (2018).

In this paper, we are interested in the *joint* test of *homogeneity* and *stability* of parameters in model (2.1). The null hypothesis is

$$\mathbb{H}_0 : (\beta_{it}, f_{it}) = (\beta_0, 0) \text{ for some } \beta_0 \in \mathbb{R}^d \text{ and all } i\text{'s and } t\text{'s}, \tag{2.15}$$

against the alternative hypothesis

$$\mathbb{H}_1 : (\beta_{it}, f_{it}) \neq (\beta_{js}, f_{js}) \text{ for some } (i,t) \neq (j,s) . \tag{2.16}$$

When the null hypothesis holds, all the individuals share the same time-invariant slopes for regressors X_{it} and do not have time trends. Then model (2.1) under $\mathbb{H}_0$ becomes the usual homogeneous linear panel data model with fixed effects. We can estimate the model either by the usual fixed-effect (FE) estimator or first-difference (FD) estimator.[3]

For the above hypothesis testing problem, one can construct testing statistics in the spirit of LR, Wald or LM tests. In this paper, we propose a nonparametric test for the structure in (2.15) based on the estimation under the null hypothesis for several reasons: first, the restricted estimation under $\mathbb{H}_0$ is much simpler than the estimation of the model without restriction; second, models with restrictions on parameters (homogeneity across individuals and stability along time) are preferred in empirical studies and our proposed test can be seen as a diagnostic test after the simple and popular model is fitted; lastly, the testing strategy provides a unified approach to testing other structures on parameters in panel data models such as homogeneity, stability or group pattern, and so on.

2.2 Estimation under the nulls and the test statistic

Since our test is based on the estimation under the null hypothesis, we introduce the estimators first. Under $\mathbb{H}_0$, the model (2.1) reduces to

$$Y_{it} = X_{it}'\beta_0 + \alpha_i + \varepsilon_{it}. \tag{2.17}$$

We can estimate β_0 either by FE or FD estimator when X_{it} are strictly exogenous. For illustration purposes, we adopt the following FE estimator

$$\hat{\beta}_{FE} = \left(\sum_{i=1}^{N} X_i' M_{\iota_T} X_i \right)^{-1} \sum_{i=1}^{N} X_i' M_{\iota_T} Y_i, \tag{2.18}$$

where $M_{\iota_T} = I_T - \iota_T \iota_T'/T$, ι_T is a $T \times 1$ vector of ones, $X_i = (X_{i1}, \ldots, X_{iT})'$ and $Y_i = (Y_{i1}, \ldots, Y_{iT})'$. Then g_{it} in (2.5) is estimated by $\hat{g}_{it} = X_{it}'\hat{\beta}_{FE}$. Denote

$$\beta_P = \left[\sum_{i=1}^{N} E\left(X_i' M_{\iota_T} X_i \right) \right]^{-1} \sum_{i=1}^{N} E\left(X_i' M_{\iota_T} Y_i \right),$$

[3]When X_{it} include the lags of dependent variable or endogeneous variable, we can estimate the model by GMM or IV approach, the proposed test statistics to be discussed will still be valid with extra assumptions and more labrous derivation.

as the nonrandom version of $\hat{\beta}_{FE}$ and $g_{P,it} = X_{it}'\beta_P$.[4]

Let $\hat{u}_{it} \equiv Y_{it} - \hat{g}_{it}$ be the *augmented* residual and $\eta_{it} = \hat{g}_{it} - g_{P,it}$ the "estimation error" when one use $\hat{g}_{it}$ to estimate $g_{P,it}$. Then we can decompose $\hat{u}_{it}$ as follows

$$\hat{u}_{it} = Y_{it} - \hat{g}_{it} = (g_{it} - g_{P,it}) + (g_{P,it} - \hat{g}_{it}) + (\alpha_i + \varepsilon_{it}) \equiv g_{it}^{\dagger} - \eta_{it} + u_{it}, \text{ say,} \tag{2.19}$$

where $u_{it} = \alpha_i + \varepsilon_{it}$ is the generalized error.

For (2.19), we note that, first, the second component η_{it} $(= \hat{g}_{it} - g_{P,it})$ is asymptotically negligible either under the null or alternative hypotheses. Second, the first component $g_{it}^{\dagger}$ $(= g_{it} - g_{P,it})$ can be rewritten as

$$g_{it}^{\dagger} = f_i(\tau_t) + X_{it}'[\beta_i(\tau_t) - \beta_P] \equiv f_i^{\dagger}(\tau_t) + X_{it}'\beta_i^{\dagger}(\tau_t).$$

Clearly, $\beta_i(\cdot) = \beta_0 = \beta_P$ and $f_i^{\dagger}(\cdot) = 0$ for all i's under $\mathbb{H}_0$, and then we have $g_{it}^{\dagger} = 0$ for all (i,t)'s. However, β_{it} and f_{it} have variation either across i or over t under $\mathbb{H}_1$, and then we in general have $\beta_i^{\dagger}(\tau_t) \neq 0$ and $f_i^{\dagger}(\tau_t) \neq 0$. It follows that $g_{it}^{\dagger}$'s are generally away from 0 when $\mathbb{H}_1$ holds.

The opposite behavior of $g_{it}^{\dagger}$ under $\mathbb{H}_0$ and $\mathbb{H}_1$ motivates us to consider the following test statistic based on the weighted sum of squared $g_{it}^{\dagger}$:

$$\Gamma_{NT}^0 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(g_{it}^{\dagger}\right)^2 w_{it}, \tag{2.20}$$

where $w_{it} \equiv w_i(\tau_t)$ and $w_i(\cdot)$'s are some user-specified non-negative weighting functions. By construction, $\Gamma_{NT}^0 \geq 0$. Clearly, Γ_{NT}^0 equals 0 under $\mathbb{H}_0$ but is greater than 0 under $\mathbb{H}_1$. However, in practice, Γ_{NT}^0 is infeasible because $\{g_{it}^{\dagger}, i = 1, \ldots, T, i = 1 \ldots, N\}$ are unknown to the researchers. In the following section, we propose the sieve estimation of $g_{it}^{\dagger}$.

2.3 Auxiliary time series regressions with TVCs

As mentioned above, to obtain a feasible testing statistic, we need to estimate $g_{it}^{\dagger}$. Noting that $\hat{u}_{it}$ is a consistent estimator for the composite error u_{it} under $\mathbb{H}_0$ and for $g_{it}^{\dagger} + u_{it}$ under $\mathbb{H}_1$,

[4]When the FD estimator is used, we have $\hat{\beta}_{FD} = \left(\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta X_{it}\Delta X_{it}'\right)^{-1}\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta X_{it}\Delta Y_{it}$, and $\beta_P = \left(\sum_{i=1}^{N}\sum_{t=2}^{T}E(\Delta X_{it}\Delta X_{it}')\right)^{-1}\sum_{i=1}^{N}\sum_{t=2}^{T}E(\Delta X_{it}\Delta Y_{it})$, where $\Delta X_{it} = X_{it} - X_{i,t-1}$ and $\Delta Y_{it} = Y_{it} - Y_{i,t-1}$.

we can estimate $\left\{g_{it}^{\dagger}\right\}_{t=1}^{T}$ based on $\{\hat{u}_{it}\}_{t=1}^{T}$ by the auxiliary time series regression of $\hat{u}_{it}$ on X_{it} and 1 with TVCs. For each i, we run an auxiliary time series regression with TVCs:[5]

$$\hat{u}_{it} = f_i^{\dagger}(\tau_t) + X_{it}'\beta_i^{\dagger}(\tau_t) + \alpha_i + \varepsilon_{it}^{\dagger},\ t = 1, \ldots, T, \tag{2.21}$$

where $\varepsilon_{it}^{\dagger} \equiv \varepsilon_{it} - \eta_{it}$. Noting that $f_i^{\dagger}(\cdot) : [0,1] \to \mathbb{R}$ and $\beta_i^{\dagger}(\cdot) : [0,1] \to \mathbb{R}^d$ are all unknown functions, which can be estimated either by the kernel method (e.g., Li et al. (2011), Chen and Huang (2018)) or the sieve method (e.g., Dong and Linton (2018), Su and Zhang (2016), Zhang and Zhou (2018)). In this paper, we focus on the sieve estimation of the unknown functions in (2.21).

Let $L^2[0,1] = \left\{u(\tau) : \int_0^1 u^2(\tau)\, d\tau < \infty\right\}$, in which $\langle u_1, u_2\rangle = \int_0^1 u_1(\tau)\, u_2(\tau)\, d\tau$ is the inner product and the induced norm is $\|u\| = \langle u, u\rangle^{1/2}$. Following Dong and Linton (2018), we choose cosine functions as basis functions.[6] Let $B_0(\tau) = 1$, and $B_j(\tau) = \sqrt{2}\cos(j\pi\tau)$ for $j \geq 1$. Then $\{B_j(\tau)\}_{j=1}^{\infty}$ forms an orthonormal basis in the Hilbert space $L^2[0,1]$ such that $\langle B_i, B_j\rangle = \delta_{ij}$, where δ_{ij} is the Kronecker delta. For any unknown continuous function $u(\tau) \in L^2[0,1]$, we obtain

$$u(\tau) = \sum_{j=0}^{\infty} \pi_{u,j} B_j(\tau),\ \text{where } \pi_{u,j} \equiv \langle u, B_j\rangle.$$

Suppose that for each i, $\beta_{il}^{\dagger}(\cdot) \in L^2[0,1]$ for $l = 1, \ldots, d$ and $f_i^{\dagger}(\cdot) \in L^2[0,1]$. Let $B^K(\cdot) = (B_0(\cdot), B_1(\cdot), \ldots, B_{K-1}(\cdot))'$ and $B_{-1}^K(\cdot) = (B_1(\cdot), \ldots, B_{K-1}(\cdot))'$ be the sequences of basis functions to approximate unknown functions $\beta_{il}^{\dagger}(\cdot)$ $(l = 1, \ldots, d)$ and $f_i^{\dagger}(\cdot)$, respectively.[7] Then for each i, we obtain[8]

$$\beta_{il}^{\dagger}(\cdot) = \sum_{j=0}^{\infty} \vartheta_{\beta,il,j} B_j(\cdot) = \vartheta_{\beta,il}' B^K(\cdot) + r_{\beta_{il}^{\dagger}}^{(K)}(\cdot),\ l = 1, \ldots, d \tag{2.22}$$

$$f_i^{\dagger}(\cdot) = \sum_{j=1}^{\infty} \vartheta_{f,i,j} B_j(\cdot) = \vartheta_{f,i}' B_{-1}^K(\cdot) + r_{f_i^{\dagger}}^{(K)}(\cdot), \tag{2.23}$$

[5] In testing the stability of homogeneous time-varying coefficients, the pooled estimation is more efficient since $f_i^{\dagger} = f_j^{\dagger}$ and $\beta_i^{\dagger} = \beta_j^{\dagger}$ for all $i \neq j$.

[6] As mentioned in Dong and Linton (2018), the cosine basis functions can be replaced by any other orthonormal basis in Hilbert space. However, the use of specific basis other than some general ones simplifies the assumptions on basis functions and leads to simpler calculation.

[7] Noting that the constant term is left out in the approximation of $f(\cdot)$ to impose the identification restriction $\int_0^1 f(\tau)\, d\tau = 0$ automatically.

[8] We can let the number of basis functions vary across different functions $f_i^*(\cdot)$ and $\beta_{il}^*(\cdot)$, $i = 1, ..., N$ and $l = 1, ..., d$. For simplicity, we adopt the same number of basis functions K in the sieve approximation of different unknown functions.

where $\vartheta_{\beta,il,j} = \left\langle \beta_{il}^{\dagger}, B_j \right\rangle$ for any integer $j \geq 0$, and $\vartheta_{f,i,j} = \left\langle f_i^{\dagger}, B_j \right\rangle$ for any integer $j \geq 1$, $\vartheta_{\beta,il} = (\vartheta_{\beta,il,0}, \ldots, \vartheta_{\beta,il,K-1})'$ and $\vartheta_{f,i} = (\vartheta_{\beta,il,1}, \ldots, \vartheta_{\beta,il,K-1})'$, $r_{\beta_{il}^{\dagger}}^{(K)}(\cdot) = \sum_{j=K}^{\infty} \vartheta_{\beta,il,j} B_j(\cdot)$ and $r_{f_i^{\dagger}}^{(K)}(\cdot) = \sum_{j=K}^{\infty} \vartheta_{f,i,j} B_j(\cdot)$. By Assumption 3 in Newey (1997), $\sup_{\tau\in[0,1]} \left| r_{\beta_{il}^{\dagger}}^{(K)}(\tau) \right| = O(K^{-\kappa})$ and $\sup_{\tau\in[0,1]} \left| r_{f_i^{\dagger}}^{(K)}(\cdot) \right| = O(K^{-\kappa})$ as $K \to \infty$ when $\beta_{il}^{\dagger}(\cdot)$ and $f_i^{\dagger}(\cdot)$ have κth continuous derivatives. Then we approximate $\beta_{il}^{\dagger}(\cdot)$ by $\vartheta_{\beta,il}' B^K(\cdot)$, and $f_i^{\dagger}(\cdot)$ by $\vartheta_{f,i}' B_{-1}^K(\cdot)$. Let $B_t \equiv B^K(\tau_t)$ and $B_{-1,t} \equiv B_{-1}^K(\tau_t)$, where the dependence on K is suppressed to simplify the notation. Using the approximations in (2.22)-(2.23) yields

$$g_{it}^{\dagger} = X_{it}'\beta_{it}^{\dagger} + f_{it}^{\dagger} \approx \sum_{l=1}^{d} X_{it,l} B_t' \vartheta_{\beta,il} + B_{-1,t}'\vartheta_{f,i} = Z_{it}'\vartheta_i,$$

where $\vartheta_i \equiv (\vartheta_{f,i}', \text{vec}(\vartheta_{\beta,i})')'$, $\vartheta_{\beta,i} = (\vartheta_{\beta,i1}, \ldots, \vartheta_{\beta,id})$ and $Z_{it} \equiv (B_{-1,t}, (X_{it} \otimes B_t)')'$ with $\otimes$ being the Kronecker product. As a result, the linearized time series regression model with sieve approximation is given by

$$\hat{u}_{it} = Z_{it}'\vartheta_i + \alpha_i + v_{it},\ t = 1, \ldots, T, \tag{2.24}$$

where $v_{it} = \varepsilon_{it} - \eta_{it} + r_{it}^{\dagger}$, and $r_{it}^{\dagger} \equiv g_{it}^{\dagger} - Z_{it}'\vartheta_i = \sum_{l=1}^{d} r_{\beta_{il}^{\dagger}}^{(K)}(\tau_t) X_{it,l} + r_{f_i^{\dagger}}^{(K)}(\tau_t)$ is the sieve approximation error of $g_{it}^{\dagger}$. Rewrite the model (2.24) in vector form

$$\hat{u}_i = Z_i\vartheta_i + \iota_T\alpha_i + v_i, \tag{2.25}$$

where $\hat{u}_i = (\hat{u}_{i1}, \ldots, \hat{u}_{iT})'$, $Z_i = (Z_{i1}', \ldots, Z_{iT}')'$, and $v_i = (v_{i1}, \ldots, v_{iT})'$. The usual OLS estimator for ϑ_i and the corresponding estimator for $g_{it}^{\dagger}$ are respectively given by

$$\hat{\vartheta}_i = \left(Z_i'M_{\iota_T}Z_i\right)^{-1} Z_i'M_{\iota_T}\hat{u}_i \text{ and } \hat{g}_{it}^{\dagger} = Z_{it}'\hat{\vartheta}_i. \tag{2.26}$$

Based on the sieve estimators $\hat{g}_{it}^{\dagger}$, we can construct a feasible version of Γ_{NT}^0 as follows

$$\Gamma_{NT} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\hat{g}_{it}^{\dagger}\right)^2 w_{it}. \tag{2.27}$$

Under certain regular conditions, we show later that after being appropriately centered and scaled, Γ_{NT} follows a standard normal distribution asymptotically under the null hypothesis.

3 Asymptotic theory

In this section, we study the large sample properties for the above test statistics.

3.1 Assumptions

In order to study the asymptotic properties for Γ_{NT} under the null hypothesis, we make the following assumptions.

Assumption 1. (i) ϵ_{it} in (2.2) is independent of X_{js} for any (i,t) and (j,s), $E(\epsilon_{it})=0$ and $\mathrm{Var}(\epsilon_{it})=1$;

(ii) $\{(X_i,\epsilon_i)\}_{i=1}^N$ are independent across i, where $X_i=(X_{i1},\ldots,X_{iT})'$ and $\epsilon_i=(\epsilon_{i1},\ldots,\epsilon_{iT})'$;

(iii) For each i, $\{(X_{it},\epsilon_{it})\}_{t=1}^T$ is strong mixing with mixing coefficients $\alpha_i(l)$ satisfying $\alpha(l)=\max_{1\leq i\leq N}\{\alpha_i(l)\}\leq C_\alpha\rho^l$ for some $C_\alpha<\infty$ and $\rho\in[0,1)$;

(iv) $(\epsilon_{it},\mathcal{F}_t)$ is a martingale difference sequence (MDS) such that $E(\epsilon_{it}|\mathcal{F}_{t-1})=0$, where $\mathcal{F}_{t-1}$ is the σ-field generated by $\{\epsilon_{js}, j=1,\ldots,N, s=1,\ldots,t-1\}$;

(v) $\max_{i,t}E|\epsilon_{it}|^{8+8\eta}<\infty$, $\max_{i,t}E\left\|X_{it}\right\|^{8+8\eta}<\infty$, and $\max_{i,t}E\sigma_{it}^4<\infty$ for some $\eta>0$, where $\max_{i,t}$ denotes $\max_{1\leq i\leq N}\max_{1\leq t\leq T}$;

(vi) $\mathrm{Var}(X_{it})=\Omega_i(t/T)$, where $\Omega_i(\cdot)$ is a $d\times d$ matrix of bounded functions defined on $[0,1]$. There exist some positive constants $\underline{c}_{xx}$ and $\bar{c}_{xx}$ such that

$$0<\underline{c}_{xx}\leq\min_{1\leq i\leq N}\inf_{\tau\in[0,1]}\left[\lambda_{\min}\left(\Omega_i(\tau)\right)\right]\leq\max_{1\leq i\leq N}\sup_{\tau\in[0,1]}\left[\lambda_{\max}\left(\Omega_i(\tau)\right)\right]\leq\bar{c}_{xx}<\infty;$$

(vii) Let $\tilde{X}_{it}^{(\sigma)}\equiv(1,X_{it}')'\sigma_{it}$ where $\sigma_{it}^2=\sigma_i^2(X_{it},t/T)$ and $\mathrm{Var}(\tilde{X}_{it}^{(\sigma)})=\Omega_i^{(\sigma)}(t/T)$, where $\Omega_i^{(\sigma)}(\cdot)$ is a $(d+1)\times(d+1)$ matrix of bounded functions defined on $[0,1]$. There exist some positive constants $\underline{c}_{xx}^{(\sigma)}$ and $\bar{c}_{xx}^{(\sigma)}$ such that

$$0<\underline{c}_{xx}^{(\sigma)}\leq\min_{1\leq i\leq N}\inf_{\tau\in[0,1]}[\lambda_{\min}(\Omega_i^{(\sigma)}(\tau))]\leq\max_{1\leq i\leq N}\sup_{\tau\in[0,1]}[\lambda_{\max}(\Omega_i^{(\sigma)}(\tau))]\leq\bar{c}_{xx}^{(\sigma)}<\infty.$$

Assumption 2. As $(N,T)\to\infty$, $K\to\infty$, $K^2/T\to0$, $NK/T^2\to0$, and $N^2T^{-3-4\eta}\ln(N)^{(4+4\eta)\nu_0}\to0$ for some $\eta>0$ and $\nu_0>1$.

Several remarks can be made for the above assumptions. For Assumption 1, 1(i) requires the independence of regressors $\{X_{it}\}$ and $\{\epsilon_{it}\}$, which is also used in Robinson (2015) and Su et al. (2018); 1(ii) imposes cross-sectional independence in the regressors and errors, which can be relaxed to allow for weak dependence as Chen et al. (2012) or Robinson (2015) with much complicated arguments in the proof; 1(iii) assumes that $\{(X_{it},\epsilon_{it}), t=1,\ldots,T\}$ are strong mixing with a geometric decay rate, which can be satisfied by many well-known linear processes such as ARMA processes and nonlinear processes; 1(iv) imposes a martingale difference structure on ϵ_{it} with filtrations $\{\mathcal{F}_t\}_{t=1}^T$, which is also used in Chen and Huang (2018); some moments conditions on ϵ_{it}, X_{it} and σ_{it} are given in 1(v); We assume the variance of X_{it} and $\tilde{X}_{it}^{(\sigma)}$ are both time-varying in 1(vi)-(vii), and their eigenvalues are both bounded and bounded away from 0. Assumption 2 provides the rate conditions on sample size (N,T) and the number of sieve basis terms K, and it can be easily satisfied if T/N converges to a nonzero constant as $(N,T)\to\infty$.

3.2 Asymptotic Distribution

We first introduce some notations. Let $Q_{\dot{z},i}=T^{-1}Z_i'M_{\iota_T}Z_i=\dot{Z}_i'\dot{Z}_i/T$ with $\dot{Z}_i=M_{\iota_T}Z_i$ and $Q_{w,i}=T^{-1}Z_i'W_iZ_i$ with $W_i=\mathtt{diag}(w_{i1},\ldots,w_{iT})$. We define a $T\times T$ matrix

$$\mathcal{K}_i\equiv M_{\iota_T}Z_iQ_{\dot{z},i}^{-1}Q_{w,i}Q_{\dot{z},i}^{-1}Z_i'M_{\iota_T}=\dot{Z}_iQ_{\dot{z},i}^{-1}Q_{w,i}Q_{\dot{z},i}^{-1}\dot{Z}_i',$$

and let $\mathcal{K}_{i,ts}$ denote its (t,s)-th element. Then denote the asymptotic bias and variance terms of Γ_{NT} as

$$\mathbb{B}_{NT}=\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathcal{K}_{i,tt}\sigma_{it}^2 \text{ and } \mathbb{V}_{NT}=\frac{2}{NT^2}\sum_{i=1}^{N}\sum_{1\le t\neq s\le T}\mathcal{K}_{i,ts}^2\sigma_{it}^2\sigma_{is}^2, \tag{3.1}$$

respectively. The standardized testing statistic is given by

$$J_{NT}=\frac{N^{1/2}T\Gamma_{NT}-\mathbb{B}_{NT}}{\sqrt{\mathbb{V}_{NT}}}. \tag{3.2}$$

Under certain regularity conditions, we can show that J_{NT} follows a standard normal distribution asymptotically under $\mathbb{H}_0$. However, the testing statistic J_{NT} is infeasible because $\mathbb{B}_{NT}$ and $\mathbb{V}_{NT}$ are both unknown. We can estimate $\mathbb{B}_{NT}$ and $\mathbb{V}_{NT}$ using their corresponding sample analogs

$$\hat{\mathbb{B}}_{NT}=\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathcal{K}_{i,tt}\hat{\varepsilon}_{r,it}^2 \text{ and } \hat{\mathbb{V}}_{NT}=\frac{2}{NT^2}\sum_{i=1}^{N}\sum_{1\le s\neq t\le T}\mathcal{K}_{i,ts}^2\hat{\varepsilon}_{r,it}^2\hat{\varepsilon}_{r,is}^2, \tag{3.3}$$

respectively, where $\hat{\varepsilon}_{r,it}=\hat{u}_{it}-\overline{\hat{u}}_i$ and $\overline{\hat{u}}_i=T^{-1}\sum_{t=1}^{T}\hat{u}_{it}$.[9] Consequently, a feasible testing statistic for J_{NT} is

$$\hat{J}_{NT}=\frac{N^{1/2}T\Gamma_{NT}-\hat{\mathbb{B}}_{NT}}{\sqrt{\hat{\mathbb{V}}_{NT}}}. \tag{3.4}$$

The following theorem gives the asymptotic distribution of $\hat{J}_{NT}$ under the null hypothesis.

[9] Alternatively, we can choose $\hat{\varepsilon}_{r,it}=\hat{u}_{it}-\hat{g}_{it}^{\dagger}-(\overline{\hat{u}}_i-\overline{\hat{g}}_i^{\dagger})$, where $\overline{\hat{g}}_i^{\dagger}=T^{-1}\sum_{t=1}^{T}\hat{g}_{it}^{\dagger}$.

Theorem 3.1 *Under Assumptions 1-2, we have* $\hat{J}_{NT} \xrightarrow{d} N(0,1)$ *under* $\mathbb{H}_0$ *as* $(N,T) \to \infty$.

Remark 1. The proof is complicated and relegated to Appendix A. The above theorem indicates that our test statistic $\hat{J}_{NT}$ is asymptotically pivotal under $\mathbb{H}_0$. In principle, we can compare $\hat{J}_{NT}$ with the one-sided critical value z_α, i.e., the upper αth percentile from the standard normal distribution, and reject the null when $\hat{J}_{NT} > z_\alpha$ at the α significance level. In practice, in order to improve the finite sample performance of the test statistic, we suggest the use of bootstrap p-values and provide a procedure to obtain them, see Section 3.4 for the details.

3.3 Asymptotic distribution under local alternatives

To study the local power property of the proposed test, we consider the following Pitman local alternatives:

$$\mathbb{H}_{1,\gamma_{NT}} : \beta_{it} = \beta_0 + \gamma_{NT}\Delta_{\beta,it} \text{ and } f_{it} = \gamma_{NT}\Delta_{f,it} \tag{3.5}$$

where $\gamma_{NT} \to 0$ as $(N,T) \to \infty$, $\Delta_{\beta,it} = \Delta_{\beta,i}(\tau_t)$, $\Delta_{f,it} = \Delta_{f,i}(\tau_t)$, $\Delta_{\beta,i}(\cdot)$: $[0,1] \to \mathbb{R}^d$ and $\Delta_{f,i}(\cdot)$: $[0,1] \to \mathbb{R}$ are all nonzero and continuous functions. Clearly, γ_{NT} controls the speed at which the local alternatives converge to the null hypothesis. Let $g_{\Delta,it} \equiv X_{it}'\Delta_{\beta,it} + \Delta_{f,it}$, $g_{\Delta,i} = (g_{\Delta,i1}, \ldots, g_{\Delta,iT})'$ and $\bar{g}_{\Delta,it} = X_{it}'\bar{\Delta}_\beta$, where $\bar{\Delta}_\beta = [\sum_{i=1}^N E(X_i' M_{\iota_T} X_i)]^{-1} \sum_{i=1}^N E(X_i' M_{\iota_T} g_{\Delta,i})$. Then we define

$$\begin{aligned} \breve{g}_{\Delta,it} &= g_{\Delta,it} - \bar{g}_{\Delta,it} = X_{it}'\left(\Delta_{\beta,it} - \bar{\Delta}_\beta\right) + \Delta_{f,it} \text{ and} \\ \Phi_{\Delta,NT} &\equiv \frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T \breve{g}_{\Delta,it}^2 w_{it}. \end{aligned}$$

To study the limiting behavior of $\hat{J}_{NT}$ under the local alternative $\mathbb{H}_{1,\gamma_{NT}}$, we need some additional assumptions on the functions $\Delta_{\beta,i}(\cdot)$ and $\Delta_{f,i}(\cdot)$.

Assumption 3. For each i, $\Delta_{\beta,il}(\cdot)$ for $l = 1, \ldots, d$, and $\Delta_{f,i}(\cdot)$ are all continuously differentiable up to κ-th order for some $\kappa \geq 2$;

Assumption 4. As $(N,T) \to \infty$, $\lim_{(N,T)\to\infty} \bar{\Delta}_\beta$ exists and $\Phi_\Delta = \text{plim}_{(N,T)\to\infty} \Phi_{\Delta,NT} > 0$.

The following theorem gives the asymptotic distribution of $\hat{J}_{NT}$ under $\mathbb{H}_{1,\gamma_{NT}}$.

Theorem 3.2 *Suppose that Assumptions 1-4 hold. As* $(N,T) \to \infty$, $\hat{J}_{NT} \xrightarrow{d} N(\Phi_\Delta, 1)$ *under* $\mathbb{H}_{1,\gamma_{NT}}$ *with* $\gamma_{NT} = N^{-1/4}T^{-1/2}\mathbb{V}_{NT}^{1/4}$.

Remark 2. (i) Theorem 3.2 implies that our test has non-trivial asymptotic power against alternatives that diverge from the null at the rate $O\left(N^{-1/4}T^{-1/2}K^{1/4}\right)$ by noting that $\mathbb{V}_{NT} = O_p\left(K\right)$ (see Lemma A.5 in appendix). The power increases with the magnitude of Φ_Δ. Clearly, as either N or T increases, the power of our test will increase but it increases faster as $T \to \infty$ than as $N \to \infty$. Similar patterns have been found in the testing literature of panel data models such as Su et al. (2018). (ii) The local alternative $\mathbb{H}_{1,\gamma_{NT}}$ includes the deviations from $\mathbb{H}_0$ only along time or across individuals, which means that our proposed test can detect the instability of homogeneous coefficients or the heterogeneity of TVCs.

To study the global consistency of $\hat{J}_{NT}$ under $\mathbb{H}_1$, let $\gamma_{NT} = 1$ in (3.5). Under Assumptions 1-4, we can show that $\text{plim}_{(N,T)\to\infty}\Gamma_{NT} = \Phi_\Delta$, $\hat{\mathbb{B}}_{NT} = O_p\left(N^{1/2}K\right)$ and $\hat{\mathbb{V}}_{NT} = O_p\left(K\right)$ under $\mathbb{H}_1$. The following corollary gives the global consistency of $\hat{J}_{NT}$ under $\mathbb{H}_1$.

Corollary 3.3 *Suppose that Assumptions 1-4 hold and* $N^{1/2}TK^{-(1/2+2\kappa)} \to 0$. *Then under* $\mathbb{H}_1$, $N^{-1/2}T^{-1}\hat{\mathbb{V}}_{NT}^{1/2}\hat{J}_{NT} \xrightarrow{p} \Phi_\Delta$ *as* $(N,T) \to \infty$ *and* .

Remark 3. Corollary 3.3 establishes that $\hat{J}_{NT}$ diverges to ∞ at rate $O_p\left(N^{1/2}T/K^{1/2}\right)$ under $\mathbb{H}_1$, which means that $P(\hat{J}_{NT} > d_{NT}) \to 1$ as $(N,T) \to \infty$ for any sequence $d_{NT} = o(N^{1/2}T/K^{1/2})$ provided $\Phi_\Delta > 0$.

Remark 4. The choice of optimal number of sieve terms is important in practice. However, it is still an open question in the literature of nonparametric testing for panel data models. One possible solution is to maximize the power when the size is controlled by following the optimal choice of bandwidth in kernel testing such as Horowize and Spokoiny (2003) and Gao and Gijbels (2008). We leave it as a future research topic. In simulation and application, we adopt a sequence of numbers of sieve terms and find them work reasonable well in finite samples.

3.4 Bootstrap version of the test

Even if $\hat{J}_{NT}$ follows $N(0,1)$ asymptotically under the null $\mathbb{H}_0$, due to the nature of nonparametric estimation in the test statistics, it is well known in the literature that tests based on nonparametric estimation usually suffer severe size distortion in finite samples if the standard normal critical values is used (see Li and Wang (1998) and Su and Hoshino (2016)). As a result, in order to improve the finite sample performance of our test, we follow Hansen (2000) and propose a fixed-regressor bootstrap procedure to obtain the bootstrap p-values. The procedure goes as follows:

1. Obtain $\hat{\beta}_{FE}$ and $\hat{u}_{it}$ under $\mathbb{H}_0$. For each i, run auxiliary time series regression of $\hat{u}_{it}$ on X_{it} and constant with TVCs to get the fitted value $\hat{g}_{it}^{\dagger}$, residual $\hat{\varepsilon}_{r,it}$, and then calculate $\hat{J}_{NT}$;

2. For each i, obtain the wild bootstrap errors $\left\{\varepsilon_{r,it}^*\right\}$: $\varepsilon_{r,it}^* = \hat{\varepsilon}_{r,it}\varrho_{it}$ where ϱ_{it}'s are IID $N(0,1)$. Then generate the bootstrap analogue Y_{it}^* of Y_{it} by holding the regressors X_{it} as fixed: $Y_{it}^* = X_{it}'\hat{\beta}_{FE} + \hat{\alpha}_i + \varepsilon_{r,it}^*$, where $\hat{\alpha}_i = T^{-1}\sum_{t=1}^{T}\left(\hat{u}_{it} - \hat{g}_{it}^{\dagger}\right)$.

3. Given the bootstrap resample $\{Y_{it}^*, X_{it}\}$, estimate the linear homogenous panel data model and run N auxiliary time series regressions as Step 1. For each i and t, denote the fitted value and residual as $\hat{g}_{it}^*$ and $\hat{\varepsilon}_{r,it}^*$, respectively. Calculate the bootstrap test statistic $\hat{J}_{NT}^*$ based on $\left\{\hat{g}_{it}^*, \hat{\varepsilon}_{r,it}^*\right\}$.

4. Repeat Steps 2-3 for B times and index the bootstrap statistics as $\{\hat{J}_{NT,b}^*\}_{b=1}^{B}$. Calculate the bootstrap p-value: $p^* = B^{-1}\sum_{b=1}^{B}\mathbf{1}(\hat{J}_{NT,b}^* \geq \hat{J}_{NT})$.

It is straightforward to implement the above bootstrap procedure. Note that we impose the null hypothesis of linear and homogeneity in Step 2. Let $\mathcal{W}_{NT} \equiv \{(X_{it}, Y_{it}) : i = 1, \ldots, N, t = 1, \ldots, T\}$ be the observed sample. Denote $Q_{\dot{z},i}^{(\hat{\varepsilon})} = T^{-1}\sum_{t=1}^{T}\dot{Z}_{it}\dot{Z}_{it}'\hat{\varepsilon}_{it}^2$. The next theorem implies the asymptotic validity of the above bootstrap procedure.

Theorem 3.4 *Suppose that Assumptions 1-2 hold. Assume that* $0 < \min_i \lambda_{\min}\left(Q_{\dot{z},i}^{(\hat{\varepsilon})}\right) \leq \max_i \lambda_{\max}\left(Q_{\dot{z},i}^{(\hat{\varepsilon})}\right) < \infty$. *Then as* $(N,T) \to \infty$, $\hat{J}_{NT}^* \xrightarrow{d^*} N(0,1)$ *in probability, where* d^* *denotes weak convergence under the bootstrap probability measure conditional on* $\mathcal{W}_{NT}$.

4 Extensions to Stability Test or Homogeneity Test

When the null hypothesis $\mathbb{H}_0$ in (2.15) is rejected, one may have interest in estimating the models with heterogenous time-invariant coefficients or homogeneous TVCs. Then it is natural to test the structures imposed by these models. In this section, we briefly discuss how to extend our proposed test to these two cases.

4.1 Test for the stability of heterogeneous coefficients

When $\mathbb{H}_0$ in (2.15) is rejected, a natural choice is to estimate a panel data model with heterogeneous slope coefficients without time variation (e.g., Hsiao and Pesaran, 2008). Then the null hypothesis now is given by

$$\mathbb{H}_{s0}:(\beta_i(\cdot),f_i(\cdot))=(\beta_i,0) \text{ for some vector } \beta_i\in\mathbb{R}^d \text{ and all } i\text{'s}, \tag{4.1}$$

against the alterative hypothesis $\mathbb{H}_{s1}:(\beta_i(\cdot),f_i(\cdot))\neq(\beta_i,0)$ for some i's. To study the local power property of the proposed test, we consider the following local Pitman alternatives

$$\mathbb{H}_{s1,\gamma_{NT}}:\beta_{it}=\beta_{0i}+\gamma_{NT}\Delta_{\beta,it} \text{ and } f_{it}=\gamma_{NT}\Delta_{f,it},$$

where $\gamma_{NT}\to0$ as $(N,T)\to\infty$, $\Delta_{\beta,it}=\Delta_{\beta,i}(\tau_t)$, $\Delta_{f,it}=\Delta_{f,i}(\tau_t)$, and $\Delta_{\beta,i}(\cdot)$ and $\Delta_{f,i}(\cdot)$ are nonzero continuous functions of time regressors for some is.

Under $\mathbb{H}_{s0}$, the model (2.1) becomes the usual heterogeneous linear panel data model

$$Y_{it}=X_{it}'\beta_i+\alpha_i+\varepsilon_{it}. \tag{4.2}$$

One can estimate the individual-specific coefficients β_i by the linear regression of Y_{it} on 1 and X_{it}. With the simple OLS estimator, we can estimate β_i and g_{it} by

$$\hat{\beta}_i=\left(X_i'M_{\iota_T}X_i\right)^{-1}X_i'M_{\iota_T}Y_i \text{ and } \hat{g}_{it}=X_{it}'\hat{\beta}_i, \tag{4.3}$$

respectively. The augmented residuals are given by $\hat{u}_{it}=Y_{it}-\hat{g}_{it}$. As Section 3.2, we can run N auxiliary time-series regressions and construct the test statistic Γ_{NT} as (2.27).

Define $Q_{\dot{X},i}=X_iM_{\iota_T}X_i/T=\dot{X}_i'\dot{X}_i/T$ and $Q_{\dot{Z}\dot{X},i}=Z_i'M_{\iota_T}X_i/T=\dot{Z}_i\dot{X}_i/T$. Also define a $T\times T$ matrix $\mathcal{K}_i^{\dagger}=\dot{Z}_i^{\dagger}Q_{\dot{z},i}^{-1}Q_{w,i}Q_{\dot{z},i}^{-1}\dot{Z}_i^{\dagger\prime}$ and denote its (t,s)th element as $\mathcal{K}_{i,ts}^{\dagger}$, where $\dot{Z}_i^{\dagger}=\dot{Z}_i-\dot{X}_iQ_{\dot{X},i}^{-1}Q_{\dot{Z}\dot{X},i}'$. Define the asymptotic bias and variance terms $\mathbb{B}_{NT}^{\dagger}=\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\mathcal{K}_{i,tt}^{\dagger}\sigma_{it}^2$ and $\mathbb{V}_{NT}^{\dagger}=\frac{2}{NT^2}\sum_{i=1}^{N}\sum_{1\le t\neq s\le T}\mathcal{K}_{i,ts}^{\dagger2}\sigma_{it}^2\sigma_{is}^2$, respectively. Then the normalized test statistic is $J_{NT}^{\dagger}=\left(N^{1/2}T\Gamma_{NT}-\mathbb{B}_{NT}^{\dagger}\right)/\sqrt{\mathbb{V}_{NT}^{\dagger}}$. However, $J_{NT}^{\dagger}$ is infeasible since $\mathbb{B}_{NT}^{\dagger}$ and $\mathbb{V}_{NT}^{\dagger}$ are not observable. Let $\hat{\varepsilon}_{r,it}=\hat{u}_{it}-\overline{\hat{u}}_i$ and $\overline{\hat{u}}_i=T^{-1}\sum_{t=1}^{T}\hat{u}_{it}$. Then we can calculate the estimators for bias and variance terms respectively by

$$\hat{\mathbb{B}}_{NT}^{\dagger}=\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\mathcal{K}_{i,tt}^{\dagger}\hat{\varepsilon}_{r,it}^2, \text{and } \hat{\mathbb{V}}_{NT}^{\dagger}=\frac{2}{NT^2}\sum_{i=1}^{N}\sum_{1\le s\neq t\le T}\mathcal{K}_{i,ts}^{\dagger2}\hat{\varepsilon}_{r,it}^2\hat{\varepsilon}_{r,is}^2.$$

The feasible testing statistic is given by

$$\hat{J}_{NT}^{\dagger}=\left(N^{1/2}T\Gamma_{NT}-\hat{\mathbb{B}}_{NT}^{\dagger}\right)/\sqrt{\hat{\mathbb{V}}_{NT}^{\dagger}}.$$

Let $g_{\Delta,it}\equiv X_{it}'\Delta_{\beta,it}+\Delta_{f,it}$ and $g_{\Delta,i}=(g_{\Delta,i1},\ldots,g_{\Delta,iT})'$. Let $\bar{\beta}_{\Delta i}=[E(X_i'M_{\iota_T}X_i)]^{-1}\times E(X_i'M_{\iota_T}g_{\Delta,i})$ and $\bar{g}_{\Delta,it}=X_{it}'\bar{\beta}_{\Delta i}$ under $\mathbb{H}_{s1,\gamma_{NT}}$. Then we can define $\breve{g}_{\Delta,it}=g_{\Delta,it}-\bar{g}_{\Delta,it}$ and $\Phi_{\Delta,NT}\equiv\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\breve{g}_{\Delta,it}^2w_{it}$.

Assumption 4*. As $(N,T)\to\infty$, $\lim_{(N,T)\to\infty}\bar{\beta}_{\Delta i}$ exists and $\Phi_\Delta=\text{plim}_{(N,T)\to\infty}\Phi_{\Delta,NT}>0$.

The following theorem gives the asymptotic distributions of $\hat{J}_{NT}^{\dagger}$ under $\mathbb{H}_{s0}$ and $\mathbb{H}_{s1,\gamma_{NT}}$.

Theorem 4.1 *(i) Under Assumptions 1-2,* $\hat{J}_{NT}^{\dagger}\xrightarrow{d}N(0,1)$ *as* $(N,T)\to\infty$ *under* $\mathbb{H}_{s0}$*;*

(ii) Suppose that Assumptions 1-3, and 4 hold. As* $(N,T)\to\infty$, $\hat{J}_{NT}^{\dagger}\xrightarrow{d}N(\Phi_\Delta,1)$ *under* $\mathbb{H}_{s1,\gamma_{NT}}$ *with* $\gamma_{NT}=O_p\left(N^{-1/4}T^{-1/2}\mathbb{V}_{NT}^{1/4}\right)$.

To study the consistency of $\hat{J}_{NT}^{\dagger}$ under $\mathbb{H}_{1s}$, let $\gamma_{NT}=1$. We need to study the asymptotic properties of $\hat{\mathbb{B}}_{NT}^{\dagger}$ and $\hat{\mathbb{V}}_{NT}^{\dagger}$. The following corollary gives the global consistency of $\hat{J}_{NT}^{\dagger}$ under $\mathbb{H}_{1s}$.

Corollary 4.2 *Suppose Assumptions 1-3, and 4* hold.* $\hat{\mathbb{V}}_{NT}^{\dagger 1/2}N^{-1/2}T^{-1}\hat{J}_{NT}^{\dagger}\xrightarrow{p}\Phi_\Delta$ *as* $(N,T)\to\infty$ *under* $\mathbb{H}_{1s}$.

4.2 Test for the homogeneity of time-varying coefficients

When $\mathbb{H}_0$ is rejected, another natural choice is to fit a panel data model with homogeneous TVCs, where the parameters are common across individuals (e.g., Chen and Huang (2018) and Li et al. (2011)). Then one may be interested in testing for the *homogeneity* of TVCs. To be specific, the null hypothesis under investigation now becomes

$$\mathbb{H}_{h0}:(\beta_i(\cdot),f_i(\cdot))=(\beta_0(\cdot),f_0(\cdot))\ \text{ for some }\ (\beta_0(\cdot),f_0(\cdot))\ \text{ and all } i\text{'s},\tag{4.4}$$

against the alternative hypothesis $\mathbb{H}_{h1}:(\beta_i(\cdot),f_i(\cdot))\neq(\beta_j(\cdot),f_j(\cdot))$ for some $i\neq j$. To facilitate the study of the local power property, we consider the following Pitman local alternatives

$$\mathbb{H}_{h1,\gamma_{NT}}:\beta_{it}=\beta_0(\tau_t)+\gamma_{NT}\Delta_{\beta,it},\text{ and } f_{it}=f_0(\tau_t)+\gamma_{NT}\Delta_{f,it},$$

where $\gamma_{NT}\to 0$ as $(N,T)\to\infty$, $\Delta_{\beta,it}=\Delta_{\beta,i}(\tau_t)$, $\Delta_{f,it}=\Delta_{f,i}(\tau_t)$, and $\left(\Delta'_{\beta,i}(\cdot),\Delta_{f,i}(\cdot)\right)\neq\left(\Delta'_{\beta,j}(\cdot),\Delta_{f,j}(\cdot)\right)$ for some $i\neq j$, $\Delta_{\beta,i}(\cdot)$ and $\Delta_{f,i}(\cdot)$ are all nonzero continuous functions of time regressors.

When $\mathbb{H}_{h0}$ holds, the model reduces to

$$Y_{it}=X'_{it}\beta(\tau_t)+f(\tau_t)+\alpha_i+\varepsilon_{it}.\tag{4.5}$$

Noting that $\beta(\cdot)$ and $f(\cdot)$ are all unknown, as before, we consider the sieve estimation of the above model (4.5). Let $B_t^L\equiv B^L(\tau_t)$, $B_{-1,t}^L\equiv B_{-1}^L(\tau_t)$, and $Z_{it}^L\equiv(B_{-1,t}^L,\left(X_{it}\otimes B_t^L\right)')'$. Let

$\Pi_f = (\Pi_{f,1}, \ldots, \Pi_{f,L-1}) \in \mathbb{R}^{L-1}$ with $\Pi_{f,k} = \langle f(\cdot), B_k(\cdot)\rangle$ and $\Pi_{\beta,l} = (\Pi_{\beta,l0}, \ldots, \Pi_{\beta,l,L-1})'$ with $\Pi_{\beta,lk} = \langle \beta_l(\cdot), B_k(\cdot)\rangle$ for $k = 1, \ldots, L-1$ such that

$$f(\cdot) \approx B_{-1}^L(\cdot)' \Pi_f \text{ and } \beta_l(\cdot) \approx \Pi_{\beta,l} B^L(\cdot) \text{ for } l = 1, \ldots, d. \tag{4.6}$$

Denote $\Pi \equiv (\Pi_f', \text{vec}(\Pi_\beta)')'$, where $\Pi_\beta \equiv (\Pi_{\beta,1}, \ldots, \Pi_{\beta,d}) \in \mathbb{R}^{L\times d}$.[10] Using the approximations in (4.6), we have $g_{it} = X_{it}'\beta_t + f_t \approx Z_{it}^{L\prime}\Pi$ and the induced linearized panel data model is given by

$$Y_{it} = Z_{it}^{L\prime}\Pi + \alpha_i + \varepsilon_{r,it}^{\dagger}, \tag{4.7}$$

where $\varepsilon_{r,it}^{\dagger} = \varepsilon_{it} + r_{g,it}$, and $r_{g,it} = g_{it} - Z_{it}^{L\prime}\Pi$ is the sieve approximation error of g_{it}. The usual FE estimator for Π is

$$\hat{\Pi}_{FE} = \left(\sum_{i=1}^{N} Z_i^{L\prime} M_{\iota_T} Z_i^L\right)^{-1} \sum_{i=1}^{N} Z_i^{L\prime} M_{\iota_T} Y_i. \tag{4.8}$$

Based on (4.8), the sieve estimators for Π_f and Π_β are denoted by $\hat{\Pi}_f$ and $\hat{\Pi}_\beta$, respectively. Then $f(\cdot)$, $\beta(\cdot)$ and g_{it} are estimated by

$$\hat{f}(\cdot) = B_{-1}^L(\cdot)' \hat{\Pi}_f,\ \hat{\beta}(\cdot) = \hat{\Pi}_\beta B^L(\cdot),\ \text{and } \hat{g}_{it} = Z_{it}^{L\prime}\hat{\Pi}_{FE}. \tag{4.9}$$

The augmented residuals are given by $\hat{u}_{it} = Y_{it} - \hat{g}_{it}$. As Section 3.2, we can run the auxiliary time-series regressions and construct the test statistic Γ_{NT} as (2.27). Based on $\hat{\varepsilon}_{r,it} = \hat{u}_{it} - \overline{\hat{u}}_i$ where $\overline{\hat{u}}_i = T^{-1}\sum_{t=1}^{T}\hat{u}_{it}$, we calculate $\hat{\mathbb{B}}_{NT}^{\ddagger}$ and $\hat{\mathbb{V}}_{NT}^{\ddagger}$ as (3.3). Then the feasible test statistic is given by

$$\hat{J}_{NT}^{\ddagger} = \left(N^{1/2}T\Gamma_{NT} - \hat{\mathbb{B}}_{NT}^{\ddagger}\right)/\sqrt{\hat{\mathbb{V}}_{NT}^{\ddagger}}.$$

Let $g_{\Delta,it} \equiv X_{it}'\Delta_{\beta,it} + \Delta_{f,it}$ and $g_{\Delta,i} = (g_{\Delta,i1}, \ldots, g_{\Delta,iT})'$. Let $\bar{g}_{\Delta,it} = Z_{it}^{L\prime}\bar{\Pi}_\Delta$, where $\bar{\Pi}_\Delta = [\sum_{i=1}^{N} E(\dot{Z}_i^{L\prime}\dot{Z}_i^{L})]^{-1}\sum_{i=1}^{N} E\left(\dot{Z}_i^{L\prime}g_{\Delta,i}\right)$. Then we define $\breve{g}_{\Delta,it} = g_{\Delta,it} - \bar{g}_{\Delta,it}$ and $\Phi_{\Delta,NT} \equiv \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\breve{g}_{\Delta,it}^2 w_{it}$. We establish the limiting distribution of the test statistic $\hat{J}_{NT}^{\ddagger}$ in the following theorem.

Theorem 4.3 *(i) Suppose that Assumptions 1-2 and Assumptions 3* and 5 in Appendix B hold. Then* $\hat{J}_{NT}^{\ddagger} \xrightarrow{d} N(0,1)$ *under* $\mathbb{H}_{h0}$ *as* $(N,T) \to \infty$.

(ii) Suppose that Assumptions 1-2 and Assumptions 3, 4** and 5 in Appendix B hold. As* $(N,T) \to \infty$, $\hat{J}_{NT}^{\ddagger} \xrightarrow{d} N(\Phi_\Delta, 1)$ *under* $\mathbb{H}_{h1,\gamma_{NT}}$ *with* $\gamma_{NT} = N^{-1/4}T^{-1/2}\mathbb{V}_{NT}^{\ddagger 1/4}$.

[10] Noting that the constant term is left out in the approximation of $f(\cdot)$ to impose the identification restriction $\int_0^1 f(\tau)\,d\tau = 0$ automatically.

To study the consistency of $\hat{J}_{NT}^{\ddagger}$ under $\mathbb{H}_{1h}$, let $\gamma_{NT} = 1$. The following corollary gives the global consistency of $\hat{J}_{NT}^{\ddagger}$ under $\mathbb{H}_{h1}$.

Corollary 4.4 *Suppose that Assumptions 1-2, 4-5 and Assumptions 3* in Appendix B hold. Then under* $\mathbb{H}_{h1}$, $\mathbb{V}_{NT}^{\ddagger 1/2} N^{-1/2} T^{-1} \hat{J}_{NT}^{\ddagger} \xrightarrow{p} \Phi_{\Delta}$ *as* $(N, T) \to \infty$.

The above result establishes that $\hat{J}_{NT}^{\ddagger}$ diverges to infinity at rate $O_p\left(N^{1/2}T/K^{1/2}\right)$ under $\mathbb{H}_{h1}$, which means that $P(\hat{J}_{NT}^{\ddagger} > d_{NT}) \to 1$ as $(N, T) \to \infty$ for any sequence $d_{NT} = o(N^{1/2}T/K^{1/2})$ provided $\Phi_{\Delta} > 0$.

5 Monte Carlo Simulations

In this section, we conduct a set of Monte Carlo simulations to evaluate the finite samples performance of our proposed joint test for homogeneity and stability of coefficients. We consider the following seven data generating processes (DGPs):

DGP 1. Homogeneous constant coefficient: $Y_{it} = 2X_{it} + \alpha_i + \varepsilon_{it}$;

DGP 2. Homogeneous TVC: $Y_{it} = f_0(\tau_t) + \beta_0(\tau_t) X_{it} + \alpha_i + \varepsilon_{it}$;

DGP 3. Heterogeneous constant coefficient: $Y_{it} = \beta_i X_{it} + \alpha_i + \varepsilon_{it}$, where $\beta_i \sim$IID $U[0, 2]$;

DGP 4. Fully heterogeneous TVC: $Y_{it} = \delta_{1i} f_0(\tau_t) + \delta_{2i}\beta_0(\tau_t) X_{it} + \alpha_i + \varepsilon_{it}$, where $\delta_{1i} \sim$IID $U[0.5, 1.5]$ and $\delta_{2i} \sim$IID $U[-0.5, 0.5]$;

DGP 5. Grouped heterogeneous TVCs:

$$Y_{it} = \begin{cases} 0.25 f_0(\tau_t) + 0.25\beta_0(\tau_t) X_{it} + \alpha_i + \varepsilon_{it}, & i = 1, \ldots, \lceil N/3 \rceil, \\ 0.5 f_0(\tau_t) + 0.5\beta_0(\tau_t) X_{it} + \alpha_i + \varepsilon_{it}, & i = \lceil N/3 \rceil + 1, \ldots, \lceil 2N/3 \rceil, \\ f_0(\tau_t) + \beta_0(\tau_t) X_{it} + \alpha_i + \varepsilon_{it}, & i = \lceil 2N/3 \rceil + 1, \ldots, N; \end{cases}$$

DGP 6. Homogeneous constant coefficient with an abrupt structural break:

$$Y_{it} = \begin{cases} 2X_{it} + \alpha_i + \varepsilon_{it}, & t < T/2, \\ -2X_{it} + \alpha_i + \varepsilon_{it}, & t \geq T/2; \end{cases}$$

DGP 7. Homogeneous TVCs with an abrupt structural break:

$$Y_{it} = \begin{cases} f_0(\tau_t) + \beta_0(\tau_t) X_{it} + \alpha_i + \varepsilon_{it}, & t < T/2, \\ 0.5 f_0(\tau_t) + 1.5\beta_0(\tau_t) X_{it} + \alpha_i + \varepsilon_{it}, & t \geq T/2. \end{cases}$$

Among all DGPs, the fixed effects α_i's follow IID $N(0,1)$, the regressors X_{it}'s are generated according to

$$X_{it} = 0.5\alpha_i + \frac{2\exp\left[(\tau_t - \mu_i)/0.1\right]}{1+\exp\left[(\tau_t - \mu_i)/0.1\right]} + \varepsilon_{x,it}$$

with $\varepsilon_{x,it}$ ∼IID $N(0,1)$ and μ_i ∼IID $U[0.05, 0.1]$, and the error ε_{it}'s are conditional heteroskedastic as $\varepsilon_{it} = \sqrt{0.05X_{it}^2 + 0.5}\epsilon_{it}$ with ϵ_{it} ∼IID $N(0,1)$.[11] In DGPs 2, 4, 5, and 7, we set

$$f_0(\upsilon) = 2\upsilon^2 - \upsilon + 1/6 \text{ and } \beta_0(\upsilon) = \frac{\exp\left[(\upsilon - 0.5)/0.1\right]}{1+\exp\left[(\upsilon - 0.5)/0.1\right]},$$

which are used to generate the smooth trend functions and time-varying coefficient functions. Similar function form for $\beta_0(\cdot)$ is adopted in Su et al. (2018).

DGP 1 is for size study and the other 6 DGPs are for power study for the joint test of homogeneity and stability. In the implementation of the specification test, we use the cosine functions as our basis functions in the sieve approximation of unknown functions. To investigate the sensitivity of our test to different choices of number of basis functions, we both consider a sequence of numbers $K_c = \lfloor cT^{1/6} \rfloor$ with $c = 1, 2, 3$ and the number K_{cv} chosen by the leave-one-out cross-validation (LOOCV) method[12]. Different combinations of sample sizes are used: $T = 25, 50, 100$ and $N = 25, 50$. For each combination of sample sizes, the number of replications is 500 times. In bootstrap, we consider 400 resamples for size studies and 300 resamples for power comparisons.

The simulation results for the joint test of homogeneity and stability in DGPs 1-7 are summarized in Table 1.[13] First, for DGP 1, the empirical sizes of our test statistic are very close to their corresponding nominal values (1%, 5% and 10%) either when we use a sequence of numbers for the sieve terms or the LOOCV to choose the number of sieve terms during the estimation. Second, the proposed test has good power for DGPs 2-7: (i) for all 6 DGPs, the empirical power tends to 1 as either N or T increases, and has a larger speed when T increases than N increases, which confirms that $\hat{J}_{NT}$ diverges to infinity faster as T increases than N

[11]To save space, we only report the results for conditional heteoskedastic errors. The results for homoskedastic errors are also availabe upon request.

[12]$K_{cv} = \operatorname{argmin}_{K\in[1,K_{\max}]} \sum_{i=1}^{N}\sum_{t=1}^{T}(\hat{u}_{it} - \hat{g}_{i(-t)}^{\dagger}(K) - \hat{\alpha}_{i,-t}(K))^2$ where $\hat{g}_{i(-t)}^{\dagger}(K)$ and $\hat{\alpha}_{i(-t)}(K)$ come from the ith auxillary regression of u_{it} on $(X_{it}', 1)'$ with TVCs without using the tth observation and K or $K-1$ basis functions are adopted in the sieve approximations. The theoretical verification of LOOCV is beyond this paper.

[13]We also report the additional simulation results for the test of homogeneity for TVCs ($\mathbb{H}_{h0}$ vs $\mathbb{H}_{h1}$) and the test of stability for heterogeneous coefficients ($\mathbb{H}_{s0}$ vs $\mathbb{H}_{s1}$) in Appendix B.

increases under $\mathbb{H}_1$ as shown in Corollary 3.3; (ii) the power increases much faster in DGPs 4-5 (variation of parameter both along time and across individuals) than in DGP 2 (variation of parameters along time) and DGP 3 (variation of parameters across individuals), which comes from the fact that Φ_Δ takes larger values in the DGPs 4-5; (iii) the empirical powers for DGPs 6-7 are close to 1 for all different scenarios, where the parameters are homogenous but have jumps along time, even Corollary 3.3 does not cover the case with jump in parameters along time. Overall, we can observe that our proposed test statistic performs very well in all scenarios in simulations.

6 Empirical Application to Environmental Kuznets Curve

In this section, we apply our proposed test to study the Environmental Kuznets Curve (EKC) of U.S. We are mainly interested in testing the validity of homogeneous linearity and stability restrictions in model, which is widely used in the EKC estimation.

The EKC hypothesis is initiated by the seminal works of Grossman and Krueger (1993, 1995) and becomes popular in the World Bank. Both theoretical and empirical literature on the topic is voluminous and continues to grow, and so do the controversial findings. Many empirical works seek to establish an inverted U-shaped nexus between income per capita and environmental degradation, which implies that the level of pollution increases until some level of prosperity is obtained. However, the inverted U-shaped relationship is questioned by Millimet et al. (2003), where a semiparametric partially linear model is used to fit the model and the parametric specification is rejected. Recently, Li et al. (2016) detect multiple structural breaks in EKC. These findings show that the regression relationship between income per capita and environmental degradation may be misspecified and vary along time. Different from previous studies, we reinvestigate the parametric specification of EKC using our proposed test.

We consider the following regression model

$$\ln Pol_{it} = \beta_{1,it} \ln Inc_{it} + \beta_{2,it} (\ln Inc_{it})^2 + f_{it} + \alpha_i + \varepsilon_{it} \tag{6.1}$$

where $i = 1, .., N$, $t = 1, \ldots, T$, $\ln Pol_{it}$ is the pollutant emission of sulfur dioxide (SO_2) measured by metric tones per capita, $\ln Inc_{it}$ represents the income for state i at time t, α_i is the unobserved state-specific fixed effect; $\beta_{1,it}$ and $\beta_{2,it}$ are time-varying slope coefficients

Table 1: Simulation results for joint test for DGP 1-7

			K_1			K_2			K_3			K_{cv}		
DGP	T	N	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
1	25	25	0.020	0.066	0.126	0.010	0.044	0.090	0.010	0.050	0.094	0.022	0.066	0.126
		50	0.012	0.056	0.110	0.014	0.042	0.084	0.008	0.046	0.106	0.012	0.056	0.112
	50	25	0.012	0.040	0.078	0.010	0.034	0.084	0.008	0.046	0.106	0.012	0.040	0.078
		50	0.008	0.042	0.094	0.006	0.054	0.114	0.004	0.052	0.124	0.008	0.042	0.094
	100	25	0.008	0.056	0.114	0.010	0.046	0.092	0.014	0.054	0.106	0.008	0.046	0.098
		50	0.008	0.060	0.106	0.008	0.058	0.126	0.008	0.062	0.110	0.012	0.060	0.110
2	25	25	0.144	0.404	0.568	0.036	0.192	0.316	0.000	0.068	0.128	0.148	0.408	0.568
		50	0.288	0.568	0.752	0.104	0.244	0.444	0.032	0.108	0.216	0.288	0.568	0.752
	50	25	0.832	0.972	0.992	0.664	0.900	0.968	0.452	0.736	0.868	0.832	0.972	0.992
		50	0.988	1.000	1.000	0.932	0.996	1.000	0.752	0.952	0.980	0.988	1.000	1.000
	100	25	1.000	1.000	1.000	1.000	1.000	1.000	0.996	1.000	1.000	1.000	1.000	1.000
		50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	25	25	0.128	0.332	0.488	0.084	0.216	0.344	0.024	0.112	0.232	0.128	0.332	0.488
		50	0.160	0.420	0.612	0.084	0.236	0.400	0.044	0.160	0.292	0.160	0.420	0.612
	50	25	0.426	0.724	0.840	0.320	0.596	0.736	0.244	0.500	0.632	0.480	0.724	0.840
		50	0.744	0.936	0.964	0.604	0.844	0.928	0.464	0.740	0.868	0.744	0.936	0.964
	100	25	0.872	0.956	0.988	0.820	0.948	0.976	0.752	0.920	0.968	0.892	0.976	0.988
		50	1.000	1.000	1.000	0.980	1.000	1.000	0.976	1.000	1.000	1.000	1.000	1.000
4	25	25	0.612	0.832	0.936	0.284	0.572	0.728	0.088	0.248	0.420	0.616	0.832	0.940
		50	0.900	0.980	0.992	0.676	0.860	0.924	0.196	0.472	0.644	0.900	0.980	0.992
	50	25	1.000	1.000	1.000	0.996	1.000	1.000	0.944	0.996	1.000	1.000	1.000	1.000
		50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	100	25	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
5	25	25	1.000	1.000	1.000	0.976	1.000	1.000	0.800	0.932	1.000	1.000	1.000	1.000
		50	1.000	1.000	1.000	0.996	1.000	1.000	0.964	0.992	1.000	1.000	1.000	1.000
	50	25	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	100	25	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
6	25	25	0.884	0.976	0.992	0.716	0.924	0.968	0.116	0.304	0.472	0.844	0.952	0.980
		50	0.988	0.996	1.000	0.920	0.984	0.996	0.152	0.444	0.644	0.968	0.992	0.996
	50	25	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	100	25	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
7	25	25	0.936	0.992	1.000	0.708	0.892	0.968	0.076	0.248	0.392	0.940	0.988	1.000
		50	0.992	0.996	1.000	0.936	0.988	0.996	0.124	0.388	0.616	0.992	0.996	1.000
	50	25	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	100	25	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		50	1.000	1.000	1.000	0.980	1.000	1.000	0.976	1.000	1.000	1.000	1.000	1.000

Note: 1. $K_C = \left\lfloor CT^{1/6} \right\rfloor$, $C = 1, 2, 3$, K_{cv} refers to the number of sieve terms by LOOCV;

2. DGP 1 is for size study and DGPs 2-7 are for power comparison.

for the ith individual, and f_{it} is the heterogeneous time trend. Presumably, the time trend f_{it} is related with pollution emission across countries. We apply our test the homogeneity and stability of $(\beta_{1,it}, \beta_{2,it}, f_{it})$ jointly. The data used in our paper is from Millimet et al. (2003)[14], which includes 48 states ($N = 48$) and ranges from year 1929 to year 1994 ($T = 66$). We transform the metric tone measurement for SO_2 emission into kilogram to achieve variables of comparable magnitude as the per capita income series.

To apply the joint test of homogeneous and stable coefficients along both time and individual dimensions, we first estimate the model under the null hypothesis, which is

$$\ln Pol_{it} = \beta_1 \ln Inc_{it} + \beta_2 (\ln Inc_{it})^2 + \alpha_i + \varepsilon_{it}. \tag{6.2}$$

The estimation and testing procedure follow similarly as discussed in Section 2. The FE estimation of model (6.2) gives us that

$$\hat{\beta}_1 = 9.5706^{***}(0.4358) \text{ and } \hat{\beta}_2 = -0.5608^{***}(0.0247),$$

where the standard error is reported in parentheses. The estimators for β_1 and β_2 are both significant at 1% significant level, and we get an inverted U-shaped EKC. In the testing, we run N auxiliary regressions of augmented residuals on $\ln Inc_{it}$ and $(\ln Inc_{it})^2$ with time-varying coefficients and trends. For the sieve approximation of unknown functions, we adopt the cosine functions as basis and consider a sequence of numbers for different functions. We consider $K_1 = 4, 5, 6, 7$ in the approximation of the coefficient $\beta_1(\cdot)$ for $\ln Inc_{it}$, $K_2 = 4, 5, 6$ in the approximation of the coefficient $\beta_2(\cdot)$ for $(\ln Inc_{it})^2$, and $K_3 = 3, 4, 5$ in the approximations of time trend $f_i(\cdot)$.[15] We report the p-values with 2000 bootstrap resamples.

The results for testing homogeneity and stability are reported in Table 2. We can find that almost all the p-values are smaller than 0.01, which suggest the strong evidence of rejecting homogeneity and stability restriction on parameters in model (6.1) even at 1% significant level.

7 Conclusion

In this paper, we provide a nonparametric test for the homogeneity and stability of parameters in panel data models. After fitting the model under the null hypothesis of homogeneity and

[14] We would like to thank Daniel Millimet for sharing their data set.

[15] We don't report the result for the LOOCV K because the LOOCV procedure always reachs the upper bound $K_{1,\max}$ or $K_{2,\max}$ when we use different $K_{1,\max}$ and $K_{2,\max}$ for the used data set.

Table 2: Bootstrap p-values for the joint test of homogeneity and stability (SO2)

(K_1,K_2,K_3)	p-value	(K_1,K_2,K_3)	p-value	(K_1,K_2,K_3)	p-value	(K_1,K_2,K_3)	p-value
4,4,3	0.0415	5,4,3	0.0020	6,4,3	0.0035	7,4,3	0.0130
4,4,4	0.0090	5,4,4	0.0030	6,4,4	0.0010	7,4,4	0.0300
4,4,5	0.0025	5,4,5	0.0005	6,4,5	0.0040	7,4,5	0.0295
4,5,3	0.0030	5,5,3	0.0035	6,5,3	0.0195	7,5,3	0.0020
4,5,4	0.0030	5,5,4	0.0005	6,5,4	0.0085	7,5,4	0.0305
4,5,5	0.0005	5,5,5	0.0015	6,5,5	0.0380	7,5,5	0.0040
4,6,3	0.0035	5,6,3	0.0010	6,6,3	0.0025	7,6,3	0.0005
4,6,4	0.0005	5,6,4	0.0135	6,6,4	0.0380	7,6,4	0.0025
4,6,5	0.0090	5,6,5	0.0340	6,6,5	0.0060	7,6,5	0.0005

stability, we obtain the augmented residuals. Then we run auxiliary time series regressions of augmented residuals on regressors with time-varying coefficients via the sieve method. Our testing statistic is constructed by averaging all the squared fitted values, which is close to zero under the null and deviates from zero under the alternative. We show that the testing statistic, after being appropriately standardized, is asymptotically normally distributed under the null and a sequence of Pitman local alternatives as both cross-sectional and time dimensions tend to infinity. A bootstrap procedure is proposed to improve the finite sample performance of the test. Monte Carlo simulations indicate that the proposed test performs reasonably well in finite samples. We apply our test the pollution emission data set, and we reject the assumption of homogeneous and stable coefficients. In addition, we extend the testing approach to test other structures on parameters such as the homogeneity of time-varying coefficients or the stability of heterogeneous coefficients.

Reference:

Atak, A., O. Linton, and Z. Xiao (2011) A semiparametric panel data model for unbalanced data with application to climate change in the United Kingdom. *Journal of Econometrics* 164, 92-115.

Bai, J. (2010) Common breaks in means and variances for panel data. *Journal of Econometrics* 157, 78-92.

Baltagi, B.H. (2012) *Econometric Analysis of Panel Data*, Wiley.

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

Baltagi, B.H., J.M. Griffin, and W. Xiong (2000) To pool or not to pool: homogeneous versus heterogeneous estimators applied to cigarette demand. *Review of Economics and Statistics* 82, 117-126.

Cai, Z. (2007) Trending time-varying coefficients time series models with serially correlated errors. *Journal of Econometrics* 136, 163-88.

Cai, Z. and Q. Li (2008) Nonparametric estimation of varying coefficient dynamic panel data models. *Econometric Theory* 24, 1321-1342.

Cai, Z., L. Chen, and Y. Fang (2016) Semiparametric estimation of partially varying-coefficient dynamic panel data models. *Econometric Reviews* 34, 695-719.

Chen, B. and L. Huang (2018) Nonparametric testing for smooth structural changes in panel data models. *Journal of Econometrics* 202, 245-267.

Chen, J., J. Gao, and D. Li (2012) Semiparametric trending panel data models with cross-sectional dependence. *Journal of Econometrics* 171, 71-85.

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

De Wachter, S. and E. Tzavalis (2012) Detection of structural breaks in linear dynamic panel data models. *Computational Statistics and Data Analysis* 56, 3020-3234.

Dong, C., J. Gao, and B. Peng (2015) Semiparametric single-index panel data models with cross-sectional dependence. Journal of Econometrics 188, 301-312.

Dong, C., and O. Linton (2018) Additive nonparametric models with time variable and both stationary and nonstationary regressors, *Journal of Econometrics* 207, 212-236.

Feng, G., J. Gao, B. Peng, and X. Zhang (2016) A varying-coefficient panel data model with fixed effects: theory and an application to U.S. commercial banks. *Journal of Econometrics* 196, 68-82.

Feng, Q., C. Kao, and S. Lazarova (2009) Estimation and identification of change points in panel models. *Working paper*, Center for Policy Research, Syracuse University, Mimeo.

Galvao, A.F., T. Juhl, G. Montes-Rojas, and J. Olmo (2017) Testing slope homogeneity in quantile regression panel data with an application to the cross-section of stock returns. *Journal of Financial Econometrics* 16, 211-243.

Gao, J., and I. Gijbels (2008) Bandwidth selection in nonparametric kernel testing. *Journal of the American Statistical Association* 103(484), 1584-1594.

Gonzalez, A., T. Terasvirta, and D. van Dijk (2005) Panel smooth transition regression models. *Working Paper*, Quantitative Finance Research Centre, University of Technology Sydney.

Grossmann, G.M. and A.B. Krueger (1993) Environmental impacts of a North American Free Trade Agreement, 13-56. In Garber, P. (Ed.) *The Mexico-US Free Trade Agreement*, MIT Press. Cambridge.

Horowitz, J.L., and V.G., Spokoiny (2003) An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica* 69, 599-631.

Grossmann, G.M. and A.B. Krueger (1995) Economic growth and the environment. *Quarterly Journal of Economics* 110, 353-377.

Hsiao, C., (2014) *Analysis of Panel Data.* Cambridge University Press.

Hsiao, C., Pesaran, M. H. (2008) Random coefficient models. In: Matyas, L., Sevestre, P., eds. The Econometrics of Panel Data. Vol. 46. Berlin: Springer Berlin Heidelberg.

Hsiao, C. and Tahmiscioglu, A. K. (1997) A panel analysis of liquidity constraints and firm investment. *Journal of the American Statistical Association* 92(438), 455-465.

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

Kim, D. (2011) Estimating a common deterministic time trend break in large panels with cross sectional dependence. *Journal of Econometrics* 164, 310-330.

Lee, K., Pesaran, M., and Smith, R. (1997) Growth and convergence in a multi-country empirical stochastic growth model. *Journal of Applied Econometrics* 12, 357-392.

Li, D., J. Chen, and J. Gao (2011) Non-parametric time-varying coefficient panel data models with fixed effects. *Econometrics Journal* 14, 387-408.

Li, D., L. Su, and J. Qian (2016) Panel data models with interactive fixed effects and multiple structural breaks. *Journal of the American Statistical Association* 111, 1804-1819.

Li, Q. and S. Wang (1998) A Simple bootstrap test for a parametric regression function. *Journal of Econometrics* 87, 145-165.

Liu, F., J. Gao, and Y. Yang (2018) Nonlinear time-varying panel data models: theory and applications. *Working Paper*, Monash University.

Ma, S., W. Lan, L. Su, and C-L Tsai, 2018, Testing Alphas in Conditional Time-Varying Factor Models with High Dimensional Assets, Journal of Business & Economic Statistics, forthcoming.

Millimet, D.L., J.A. List, and T. Stengos (2003) The environmental Kuznets curve: real progress or misspecified models? *Review of Economics and Statistics* 85, 1038-1047.

Newey, W K., (1997) Convergence rates and asymptotic normality for series estimators, *Journal of Econometrics* 79, 147-168.

Pauwels, L., F. Chan, and T. Griffoli (2012) Testing for structural change in heterogeneous panels with an application to the Euro's trade effect. *Journal of Time Series Econometrics* 4 (2).

Pesaran, M.H., (2015) *Time Series and Panel Data Econometrics*, Oxford University Press.

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

Su, L. and T. Hoshino (2016) Sieve instrumental variable quantile regression estimation of functional coefficient models. *Journal of Econometrics* 191, 231-254.

Su, L., X. Wang, and S. Jin, (2018) Sieve estimation of time-varying panel data models with latent structures. *Journal of Business and Economic Statistics*, forthcoming.

Su, L. and Y. Zhang (2016) Semiparametric estimation of partially linear dynamic panel data models with fixed effects. *Advances in Econometrics* in: Essays in Honor of Aman Ullah, volume 36, pp.137-204, Emerald Publishing Ltd.

Su, L., Z. Shi, and P.C.B. Phillips (2016) Identifying latent structures in panel data. *Econometrica* 84, 2215-2264.

Sun, Y. (2006) A consistent nonparametric equality test of conditional quantile functions. *Econometric Theory* 22, 614-632.

Sun, Y., R.J. Carroll, and D. Li (2009) Semiparametric estimation of fixed-effects panel data varying coefficient models. *Advances in Econometrics* in: Nonparametric Econometric Methods, volume 25, pp.101-129, Emerald Publishing Ltd.

White, H. (2000) *Asymptotic Theory for Econometricians (Revised Edition)*, Emerald Group Publishing Limited

Zhang, Y., L. Su, and P.C.B. Phillips (2012) Testing for common trends in semiparametric panel data models with fixed effects. *Econometrics Journal* 15, 56-100.

Zhang, Y., and Q. Zhou (2018) Partially linear functional-coefficient dynamic panel data models: sieve estimation and specification testing. *Working paper*, Louisiana State University.

Appendix A

The appendix provides some facts, lemmas and the proofs of main results in Section 3.

Notation. Given sequences $\{a_n\}$ and $\{b_n\}$, let $a_n \lesssim b_n$ ($a_n \gtrsim b_n$) denote that b_n/a_n (a_n/b_n) is bounded, and $a_n \asymp b_n$ denote that both a_n/b_n and b_n/a_n are bounded. When $\{a_n\}$ and $\{b_n\}$ are stochastic sequences, $a_n \lesssim b_n$ ($a_n \gtrsim b_n$) denote that b_n/a_n (a_n/b_n) is stochastically bounded, and $a_n \asymp b_n$ mean that both a_n/b_n and b_n/a_n are stochastically bounded. For a random variable X, let $\|X\|_p = E\left(|X|^p\right)^{1/p}$ for $p \geq 1$.

A Some facts and lemmas

We first state some facts and technical lemmas that are used in the proof of the main results in Section 3. The proofs for these lemmas are given in Appendix B.

Note that we use the cosine functions basis $B_{-1}^K(\tau) = (2^{1/2}\cos(\pi\tau), \ldots, 2^{1/2}\cos((K-1)\pi\tau))'$ and $B^K(\tau) = (1, 2^{1/2}\cos(\pi\tau), \ldots, 2^{1/2}\cos((K-1)\pi\tau))'$ to approximate $f_i^\dagger(\cdot)$ and $\beta_i^\dagger(\cdot)$ in the auxiliary regressions, respectively. Recall that $B_t = B^K(\tau_t)$, $B_{-1,t} = B_{-1}^K(\tau_t)$, $Z_{it} = (B_{-1,t}', X_{it}' \otimes B_t')'$, $\dot{Z}_{it} = Z_{it} - \bar{Z}_i$, and $\mathcal{K}_i = \dot{Z}_i' Q_{\dot{z},i}^{-1} Q_{w,i} Q_{\dot{z},i}^{-1} \dot{Z}_i$. We give some facts and bounds on them:

(i) $||T^{-1}\sum_{t=1}^T B_t B_t' - I_K||^2 = O\left(K^2/T^2\right)$ (see Lemma C.4 in Dong and Linton (2018));

(ii) $\sup_{\tau\in[0,1]}\left\|B^K(\tau)\right\|^2 = 2K-1$ and $\sup_{\tau\in[0,1]}\left\|B_{-1}^K(\tau)\right\|^2 = 2K-2$;

(iii) $\|Z_{it}\|^2 = \|B_{-1,t}\|^2 + \|X_{it}\|^2\|B_t\|^2 \leq \sup_{\tau\in[0,1]}\left\|B^K(\tau)\right\|^2(1+\|X_{it}\|^2) = 2K\left\|\tilde{X}_{it}\right\|^2$, where $\tilde{X}_{it} = (1, X_{it}')'$;

(iv) $\left\|\dot{Z}_{it}\right\|^2 \leq 2(\|Z_{it}\|^2 + \left\|\bar{Z}_i\right\|^2) \leq 2(\|Z_{it}\|^2 + T^{-1}\sum_{s=1}^T\|Z_{is}\|^2) \leq 4KA_{it}$, where $A_{it} = \left\|\tilde{X}_{it}\right\|^2 + T^{-1}\sum_{s=1}^T\left\|\tilde{X}_{is}\right\|^2$;

(v) $\mathcal{K}_{i,tt} = \dot{Z}_{it}' Q_{\dot{z},i}^{-1} Q_{w,i} Q_{\dot{z},i}^{-1}\dot{Z}_{it} \leq \lambda_{\max}(Q_{w,i})\,\lambda_{\max}\left(Q_{\dot{z},i}^{-2}\right)\left\|\dot{Z}_{it}\right\|^2 \leq \lambda_{\max}(Q_{w,i})\,\lambda_{\max}\left(Q_{\dot{z},i}^{-2}\right)4KA_{it}$;

(vi) $|\mathcal{K}_{i,ts}| \leq \mathcal{K}_{i,tt}^{1/2}\mathcal{K}_{i,ss}^{1/2} = \lambda_{\max}(Q_{w,i})\,\lambda_{\max}\left(Q_{\dot{z},i}^{-2}\right)4KA_{it}^{1/2}A_{is}^{1/2}$.

Next, we give some lemmas and the first two are similar to Lemmas A1-A2 in Su, et al. (2018) where spline functions are adopted as basis functions.

Lemma A.1 *Suppose that Assumption 1 holds. Let* $\mathbf{g} = (g_0, g_1, \ldots, g_d)'$, *where* $g_l = \theta_l' B^K(\cdot) \in \mathcal{G} \equiv \{g(\cdot) = \theta' B^K(\cdot) : \theta \in \mathbb{R}^K\}$ *for* $l = 1, \ldots, d$, *and* $g_0 = \theta_l' B_{-1}^K(\cdot) \in \mathcal{G}_{-1} \equiv \{g(\cdot) = \theta' B_{-1}^K(\cdot) : \theta \in \mathbb{R}^{K-1}\}$ *Then* $\|\mathbf{g}\|_i^2 = \sum_{l=0}^d \|g_l\|_2^2 \asymp \|\theta\|^2$ *where* $\|\mathbf{g}\|_i^2 \equiv E\{T^{-1}\sum_{t=1}^T[\mathbf{g}(\tau_t)'\tilde{X}_{it}][\tilde{X}_{it}'\mathbf{g}(\tau_t)]\}$ *with* $\tilde{X}_{it} = (1, X_{it}')'$ *and* $\theta = (\theta_0', \theta_1', \ldots, \theta_d')'$.

Lemma A.2 *Suppose that Assumption 1 holds. Let* $\mathcal{G} \equiv \{g(\cdot) = \theta' B^K(\cdot) : \theta \in \mathbb{R}^K\}$. *Let* $\mathcal{G}^{\otimes d}$ *denote the collection of vector of functions* $\mathbf{g} = (g_0, g_1, \ldots, g_d)'$ *with* $g_l \in \mathcal{G}$ *for* $l = 1, \ldots, d$ *and* $g_0 \in \mathcal{G}_{-1}$. *Then for any* $\epsilon > 0$, *(i)* $P\left(\max_i \sup_{\mathbf{g}\in\mathcal{G}_{-1}\times\mathcal{G}^{\otimes d}}\left|\frac{T^{-1}\sum_{t=1}^{T}\left[\mathbf{g}(\tau_t)'\tilde{X}_{it}\right]^s}{T^{-1}\sum_{t=1}^{T}E\left[\mathbf{g}(\tau_t)'\tilde{X}_{it}\right]^s}-1\right|>\epsilon\right)=o\left(N^{-1}\right)$ *for* $s=1,2$;

(ii) $P\left(\sup_{\mathbf{g}\in\mathcal{G}_{-1}\times\mathcal{G}^{\otimes d}}\left|\frac{(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\mathbf{g}(\tau_t)'\tilde{X}_{it}\right]^2}{(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}E\left[\mathbf{g}(\tau_t)'\tilde{X}_{it}\right]^2}-1\right|>\epsilon\right)=o\left(N^{-1}\right).$

Lemma A.3 *Let* $Q_{\dot{z},i}^{(\sigma)}\equiv T^{-1}\sum_{s=1}^{T}\dot{Z}_{is}\dot{Z}_{is}'\sigma_{is}^2$ *and* $Q_{\dot{z},i}^{(\varepsilon)}\equiv T^{-1}\sum_{s=1}^{T}\dot{Z}_{is}\dot{Z}_{is}'\varepsilon_{is}^2$. *Suppose that Assumption 1 holds. Then*

(i) $P\left(\underline{c}_{\dot{z}}\leq\min_i[\lambda_{\min}\left(Q_{\dot{z},i}\right)]\leq\max_i[\lambda_{\max}\left(Q_{\dot{z},i}\right)]\leq\bar{c}_{\dot{z}}\right)=1-o\left(N^{-1}\right)$;

(ii) $P\left(\underline{c}_{w}\leq\min_i[\lambda_{\min}\left(Q_{w,i}\right)]\leq\max_i[\lambda_{\max}\left(Q_{w,i}\right)]\leq\bar{c}_{w}\right)=1-o\left(N^{-1}\right)$;

(iii) $P(\underline{c}_{\dot{z},\sigma}\leq\min_i[\lambda_{\min}(Q_{\dot{z},i}^{(\sigma)})]\leq\max_i[\lambda_{\max}(Q_{\dot{z},i}^{(\sigma)})]\leq\bar{c}_{\dot{z},\sigma})=1-o\left(N^{-1}\right)$,

(iv) $P\left(\max_i[\lambda_{\max}\left(Q_{i,\varepsilon}\right)]\leq\bar{c}_{\dot{z},\sigma}\right)=1-o\left(N^{-1}\right)$;

where $\underline{c}_{\dot{z}}$, $\bar{c}_{\dot{z}}$, $\underline{c}_{w}$, $\bar{c}_{w}$, $\underline{c}_{\dot{z},\sigma}$ *and* $\bar{c}_{\dot{z},\sigma}$ *are some finite positive constants.*

Lemma A.4 *Suppose that Assumptions 1-3 hold. Then we have*

(i) $\frac{1}{NT}\sum_{i=1}^{N}\|r_{\Delta,i}\|^2=O\left(K^{-2\kappa}\right)$; *and (ii)* $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}r_{\Delta,it}^2w_{it}=O\left(K^{-2\kappa}\right).$

Lemma A.5 *Suppose that Assumptions 1-3 hold. Then we have*

(i) $\mathbb{V}_{NT}=O_p\left(K\right)$; *and (ii)* $\mathbb{B}_{NT}=O_p\left(N^{1/2}K^{1/2}\right)$.

B Proofs of main results in Section 3

In this section, we provide the proofs for the theorems in Section 3.

Proof of Theorem 3.1. Note that the limiting distribution of $\hat{J}_{NT}$ under $\mathbb{H}_0$ is a special case of Theorem 3.2 with $\Delta_{\beta,i}\left(\cdot\right)=0$ and $\Delta_{f,i}\left(\cdot\right)=0$ for all i's, or $\gamma_{NT}=0$. See the proof of Theorem 3.2.■

Proof of Theorem 3.2. We first investigate the behavior of augmented residuals $\hat{u}_{it}$ under $\mathbb{H}_{1,\gamma_{NT}}$. Recall that $\bar{\Delta}_\beta=[\sum_{i=1}^{N}E\left(X_i'M_{\iota_T}X_i\right)]^{-1}\sum_{i=1}^{N}E\left(X_i'M_{\iota_T}g_{\Delta,i}\right)$. Let $\nu_{\Delta,NT}\equiv[\sum_{i=1}^{N}X_i'M_{\iota_T}X_i]^{-1}\sum_{i=1}^{N}X_i'M_{\iota_T}g_{\Delta,i}-\bar{\Delta}_\beta$ and $\nu_{NT}\equiv[\sum_{i=1}^{N}X_i'M_{\iota_T}X_i]^{-1}\sum_{i=1}^{N}X_i'M_{\iota_T}\varepsilon_i$. By the definition of β_P, we have $\beta_P=\beta_0+\gamma_{NT}\bar{\Delta}_\beta$. Then $\hat{\beta}_{FE}-\beta_P=\gamma_{NT}\nu_{\Delta,NT}+\nu_{NT}\equiv\breve{\nu}_{NT}$ and $\beta_{it}-\beta_P=\gamma_{NT}\Delta_{\beta,it}^c$, where $\Delta_{\beta,it}^c\equiv\Delta_{\beta,it}-\bar{\Delta}_\beta$. It follows that $g_{\Delta,it}-\bar{g}_{\Delta,it}=X_{it}'\left(\beta_{it}-\beta_P\right)+\gamma_{NT}\Delta_{f,it}=\gamma_{NT}\left(X_{it}'\Delta_{\beta,it}^c+\Delta_{f,it}\right)=\gamma_{NT}\breve{g}_{\Delta,it}$, where $\breve{g}_{\Delta,it}=X_{it}'\Delta_{\beta,it}^c+\Delta_{f,it}$. Then

$$\hat{u}_{it}=\gamma_{NT}\breve{g}_{\Delta,it}-X_{it}'\breve{\nu}_{NT}+\alpha_i+\varepsilon_{it}\text{ and }\hat{u}_i=\gamma_{NT}\breve{g}_{\Delta,i}-X_i\breve{\nu}_{NT}+\iota_T\alpha_i+\varepsilon_i. \tag{A.1}$$

Using (A.1) and $\Gamma_{NT}=\frac{1}{NT^2}\sum_{i=1}^{N}\hat{u}_i'\mathcal{K}_i\hat{u}_i$, we have

$$\Gamma_{NT}=\frac{1}{NT^2}\sum_{i=1}^{N}\left(\varepsilon_i+\gamma_{NT}\breve{g}_{\Delta,i}-X_i\breve{\nu}_{NT}\right)'\mathcal{K}_i\left(\varepsilon_i+\gamma_{NT}\breve{g}_{\Delta,i}-X_i\breve{\nu}_{NT}\right)\equiv\sum_{s=1}^{6}\Gamma_{NT}^{(s)}, \tag{A.2}$$

where

$$\Gamma_{NT}^{(1)}\equiv\frac{1}{NT^2}\sum_{i=1}^{N}\varepsilon_i'\mathcal{K}_i\varepsilon_i,\quad \Gamma_{NT}^{(2)}\equiv\frac{\gamma_{NT}^2}{NT^2}\sum_{i=1}^{N}\breve{g}_{\Delta,i}'\mathcal{K}_i\breve{g}_{\Delta,i},\quad \Gamma_{NT}^{(3)}\equiv\frac{1}{NT^2}\sum_{i=1}^{N}\breve{\nu}_{NT}'X_i'\mathcal{K}_iX_i\breve{\nu}_{NT},$$
$$\Gamma_{NT}^{(4)}\equiv\frac{2\gamma_{NT}}{NT^2}\sum_{i=1}^{N}\varepsilon_i'\mathcal{K}_i\breve{g}_{\Delta,i},\quad \Gamma_{NT}^{(5)}\equiv\frac{-2}{NT^2}\sum_{i=1}^{N}\varepsilon_i'\mathcal{K}_iX_i\breve{\nu}_{NT},\quad \Gamma_{NT}^{(6)}\equiv\frac{-2\gamma_{NT}}{NT^2}\sum_{i=1}^{N}\breve{g}_{\Delta,i}'\mathcal{K}_iX_i\breve{\nu}_{NT}.$$

Using (A.2), $\hat{J}_{NT}$ can be decomposed as follows

$$\hat{J}_{NT} = \frac{N^{1/2}T\Gamma_{NT} - \hat{\mathbb{B}}_{NT}}{\hat{\mathbb{V}}_{NT}^{1/2}} = \left(J_{NT} + \sum_{s=2}^{6} \frac{N^{1/2}T\Gamma_{NT}^{(s)}}{\mathbb{V}_{NT}^{1/2}} + \frac{\mathbb{B}_{NT} - \hat{\mathbb{B}}_{NT}}{\mathbb{V}_{NT}^{1/2}} \right) \frac{\mathbb{V}_{NT}^{1/2}}{\hat{\mathbb{V}}_{NT}^{1/2}}.$$

We complete the proof by showing that, as $(N,T) \to \infty$: (i) $J_{NT} = (N^{1/2}T\Gamma_{NT}^{(1)} - \mathbb{B}_{NT})/\mathbb{V}_{NT}^{1/2} \xrightarrow{d} N(0,1)$; (ii) $J_{NT}^{(2)} \equiv N^{1/2}T\Gamma_{NT}^{(2)}/\mathbb{V}_{NT}^{1/2} = \Phi_\Delta + o_p(1)$; (iii) $J_{NT}^{(s)} \equiv N^{1/2}T\Gamma_{NT}^{(s)}/\mathbb{V}_{NT}^{1/2} = o_p(1)$ for $s = 3,4,5,6$; (iv) $\hat{\mathbb{B}}_{NT} - \mathbb{B}_{NT} = o_p(K^{1/2})$; (v) $\hat{\mathbb{V}}_{NT}/\mathbb{V}_{NT} = 1 + o_p(1)$. Note that the proofs for (iv) and (v) are given in Propositions B.2 and B.3, respectively. We are left to show (i)-(iii).

Proof of (i). Write $\Gamma_{NT}^{(1)} = \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{1\le t\neq s\le T}\mathcal{K}_{i,ts}\varepsilon_{is}\varepsilon_{it} + \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathcal{K}_{i,tt}\varepsilon_{it}^2 \equiv \Gamma_{NT}^{(1a)} + \Gamma_{NT}^{(1b)}$, say. Then J_{NT} can be further decomposed as follows

$$J_{NT} = \frac{N^{1/2}T\Gamma_{NT}^{(1a)}}{\sqrt{\mathbb{V}_{NT}}} + \frac{N^{1/2}T\Gamma_{NT}^{(1b)} - \mathbb{B}_{NT}}{\sqrt{\mathbb{V}_{NT}}} \equiv J_{NT}^{(a)} + J_{NT}^{(b)}, \text{ say.}$$

We complete the proof by showing that (ia) $J_{NT}^{(a)} \to_d N(0,1)$ and (ib) $J_{NT}^{(b)} = o_p(1)$. The justification of (ia) is given in Proposition B.1 below. We are left to show (ib).

To show (ib), write $J_{NT}^{(b)} = \tilde{J}_{NT}^{(b)}/\mathbb{V}_{NT}^{1/2}$ where $\tilde{J}_{NT}^{(b)} \equiv (N^{1/2}T\Gamma_{NT}^{(1b)} - \mathbb{B}_{NT})$. Noting that $\mathbb{V}_{NT} = O_p(K)$ by Lemma A.5(i), we want to verify that $\tilde{J}_{NT}^{(b)} = o_p\left(K^{1/2}\right)$. By the definition of $\mathbb{B}_{NT}$ in (3.1) and using $\varepsilon_{it}^2 = \sigma_{it}^2\epsilon_{it}^2$, we write $\tilde{J}_{NT}^{(b)} = N^{-1/2}T^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathcal{K}_{i,tt}\sigma_{it}^2\left(\epsilon_{it}^2 - 1\right)$. Let $\mathbf{X} \equiv (X_1,\ldots,X_N)$. Clearly, $E(\tilde{J}_{NT}^{(b)}|\mathbf{X}) = 0$ by Assumption 1(i) and

$$\begin{aligned}\text{Var}\left(\tilde{J}_{NT}^{(b)}|\mathbf{X}\right) &= \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathcal{K}_{i,tt}^2\sigma_{it}^4\text{Var}\left(\epsilon_{it}^2\right) + \frac{2}{NT^2}\sum_{i=1}^{N}\sum_{1\le t<s\le T}\mathcal{K}_{i,tt}\mathcal{K}_{i,ss}\sigma_{it}^2\sigma_{is}^2\text{Cov}\left(\epsilon_{it}^2,\epsilon_{is}^2\right) \\ &\equiv VJ_1 + VJ_2, \text{ say.}\end{aligned}$$

By the fact (v) and Lemma A.3(i)-(ii), we have $\mathcal{K}_{i,tt} \le \lambda_{\max}\left(Q_{w,i}\right)\lambda_{\min}^{-2}\left(Q_{\dot{z},i}\right)4KA_{it} \le C_*KA_{it}$ uniformly with $C_* \equiv 4\bar{c}_w\underline{c}_{\dot{z}}^{-2}$, we have $VJ_1 \le \max_{i,t}E\left(\epsilon_{it}^4\right)\frac{C_*^2}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}K^2A_{it}^2\sigma_{it}^4 \lesssim \frac{K^2}{T} \times \left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}A_{it}^2\sigma_{it}^4\right) = O_p\left(K^2/T\right) = o_p(K)$ by the fact that the term in the previous parentheses is $O_p(1)$ by Markov inequality and moment conditions on X_{it} and σ_{it} in Assumption 1(iv). By Assumption 1(iii), $\left\{\epsilon_{it}^2\right\}_{t=1}^{T}$ are strong mixing. Then we have $|\text{Cov}\left(\epsilon_{it}^2,\epsilon_{is}^2\right)| \le 8\alpha^{\eta/(1+\eta)}(s-t)\left\|\epsilon_{it}^2\right\|_{2+2\eta}\left\|\epsilon_{is}^2\right\|_{2+2\eta}$ by Davydov inequality (Bosq, 1998). Then for VJ_2,

$$
\begin{aligned}
|VJ_2| &\leq \frac{16}{NT^2} \sum_{i=1}^{N} \sum_{1 \leq t < s \leq T} \mathcal{K}_{i,tt} \mathcal{K}_{i,ss} \sigma_{it}^2 \sigma_{is}^2 \alpha^{\frac{\eta}{1+\eta}} (s-t) \left\| \epsilon_{it}^2 \right\|_{2+2\eta} \left\| \epsilon_{is}^2 \right\|_{2+2\eta} \\
&\leq \frac{16 C_*^2 K^2}{NT^2} \sum_{i=1}^{N} \sum_{1 \leq t < s \leq T} A_{it} A_{is} \sigma_{it}^2 \sigma_{is}^2 \alpha^{\frac{\eta}{1+\eta}} (s-t) \left\| \epsilon_{it}^2 \right\|_{2+2\eta} \left\| \epsilon_{is}^2 \right\|_{2+2\eta} \\
&\leq \left[\max_{i,t} (\left\| \epsilon_{it}^2 \right\|_{2+2\eta}) \right]^2 \frac{16 C_*^2 K^2}{T} \left[\frac{1}{NT} \sum_{i=1}^{N} \sum_{1 \leq t < s \leq T} A_{it} A_{is} \sigma_{it}^2 \sigma_{is}^2 \alpha^{\frac{\eta}{1+\eta}} (s-t) \right] \\
&\lesssim \frac{K^2}{T} \times \overline{VJ}_2
\end{aligned}
$$

where $\overline{VJ}_2 \equiv \frac{1}{NT} \sum_{i=1}^{N} \sum_{1 \leq t < s \leq T} A_{it} A_{is} \sigma_{it}^2 \sigma_{is}^2 \alpha^{\eta/(1+\eta)} (s-t)$. Noting that $\overline{VJ}_2 \geq 0$ and $E\left(\overline{VJ}_2\right) \leq \max_{i,t} E(A_{it}^2 \sigma_{it}^4) \frac{1}{T} \sum_{1 \leq t < s \leq T} \alpha^{\eta/(1+\eta)} (s-t) < \infty$ by Assumptions 1(iii)-(iv). Then we have $\overline{VJ}_2 = O_p(1)$ by the Markov inequality. It follows that $VJ_2 = O_p\left(K^2/T\right)$ and $\text{Var}(\tilde{J}_{NT}^{(b)}|\mathbf{X}) = O_p\left(K^2/T\right)$. By the Chebyshev inequality, $\tilde{J}_{NT}^{(b)} = O_p\left(K/T^{1/2}\right) = o_p\left(K^{1/2}\right)$ by Assumption 2**.**

Proof of (ii). By Assumption 3, for given $B^K(\cdot)$, there exist $\Pi_{\Delta,i}^{(\beta)} \in \mathbb{R}^{Kd}$ and $\Pi_{\Delta,i}^{(f)} \in \mathbb{R}^{K-1}$ such that

$$
\breve{g}_{\Delta,it} = X_{it}' \left(\Delta_{\beta,it} - \bar{\Delta}_{\beta,NT} \right) + \Delta_{f,it} = Z_{it}' \Pi_{\Delta,i} + r_{\Delta,it}, \tag{A.3}
$$

using the decomposition of $\Delta_{\beta,i}(\cdot) - \bar{\Delta}_{\beta,NT}$ and $\Delta_{f,i}(\cdot)$ similar to (2.22)-(2.23), where $\Pi_{\Delta,i} \equiv (\Pi_{\Delta,i}^{(f)\prime}, \text{vec}(\Pi_{\Delta,i}^{(\beta)\prime}))'$ and $r_{\Delta,it}$ is the sieve approximation error. We have

$$
J_{NT}^{(2)} \equiv \frac{1}{NT^2} \sum_{i=1}^{N} \left(\Pi_{\Delta,i}' Z_i' \mathcal{K}_i Z_i \Pi_{\Delta,i} + r_{\Delta,i}' \mathcal{K}_i r_{\Delta,i} + 2 r_{\Delta,i}' \mathcal{K}_i Z_i \Pi_{\Delta,i} \right) \equiv \breve{J}_{NT}^{(2a)} + \breve{J}_{NT}^{(2b)} + \breve{J}_{NT}^{(2c)}, \text{ say,}
$$

where $r_{\Delta,i} = (r_{\Delta,i1}, \ldots, r_{\Delta,iT})'$. First, noting that $Z_i' \mathcal{K}_i Z_i / T = Z_i' W_i Z_i$ and using (A.3), we have $\breve{J}_{NT}^{(2a)} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \breve{g}_{\Delta,it}^2 w_{it} + \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} r_{\Delta,it}^2 w_{it} - \frac{2}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \breve{g}_{\Delta,it} r_{\Delta,it} w_{it} \equiv \breve{J}_{NT1}^{(2a)} + \breve{J}_{NT2}^{(2a)} - 2\breve{J}_{NT3}^{(2a)}$, say. Clealry, $\breve{J}_{NT1}^{(2a)} = \Phi_\Delta + o_p(1)$. By Lemma A.4(ii), $\breve{J}_{NT2}^{(2a)} = O_p\left(K^{-2\kappa}\right)$, and further $\breve{J}_{NT3}^{(2a)} = O_p\left(K^{-\kappa}\right)$ by Cauchy-Schwarz inequality. It follows that $\breve{J}_{NT}^{(2a)} = \Phi_\Delta + o_p(1)$. Second, we have $\breve{J}_{NT}^{(2b)} = \frac{1}{NT^2} \sum_{i=1}^{N} r_{\Delta,i}' M_{\iota_T} Z_i Q_{\dot{z},i}^{-1} Q_{w,i} Q_{\dot{z},i}^{-1} Z_i' M_{\iota_T} r_{\Delta,i} \leq \max_i \lambda_{\max}(Q_{w,i}) \max_i \lambda_{\max}\left(Q_{\dot{z},i}^{-1}\right) \frac{1}{NT^2} \sum_{i=1}^{N} r_{\Delta,i}' \dot{Z}_i Q_{\dot{z},i}^{-1} \dot{Z}_i' r_{\Delta,i} \leq \bar{c}_w \underline{c}_{\dot{z}}^{-1} \max_i \lambda_{\max}\left(T^{-1} \dot{Z}_i Q_{\dot{z},i}^{-1} \dot{Z}_i'\right) \times \frac{1}{NT} \sum_{i=1}^{N} \left\| r_{\Delta,i} \right\|^2 = O_p\left(K^{-2\kappa}\right)$ by Lemma A.4(i) and the fact that $T^{-1} \dot{Z}_i Q_{\dot{z},i}^{-1} \dot{Z}_i'$ has the largest eigenvalue 1 because it is a projection matrix. By the Cauchy-Schwarz inequality, $\breve{J}_{NT}^{(2c)} = O_p\left(K^{-\kappa}\right) = o_p(1)$. Then we have shown that $J_{NT}^{(2)} = \Phi_\Delta + o_p(1)$.

Proof of (iii). When $l = 3$, by the repeatedly use of $x' A x \leq \lambda_{\max}(A) x' x$ for any symmetric matrix A and conformable vector x, we have

$$\begin{aligned}
\Gamma_{NT}^{(3)} &= \frac{1}{NT^2}\sum_{i=1}^{N}\breve{\nu}_{NT}'X_i'M_{\iota_T}Z_iQ_{\dot{z},i}^{-1}Q_{w,i}Q_{\dot{z},i}^{-1}Z_i'M_{\iota_T}X_i\breve{\nu}_{NT} \\
&\leq \max_i \lambda_{\max}\left(Q_{w,i}\right)\max_i \lambda_{\max}\left(Q_{\dot{z},i}^{-1}\right)\frac{1}{NT^2}\sum_{i=1}^{N}\breve{\nu}_{NT}'X_i'M_{\iota_T}Z_iQ_{\dot{z},i}^{-1}Z_i'M_{\iota_T}X_i\breve{\nu}_{NT} \\
&\leq \bar{c}_w\underline{c}_{\dot{z}}^{-1}\max_i \lambda_{\max}\left(T^{-1}\dot{Z}_iQ_{\dot{z},i}^{-1}\dot{Z}_i'\right)\left\|\breve{\nu}_{NT}\right\|^2\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\dot{X}_{it}\right\|^2 \\
&= \left[O_p((NT)^{-1}) + o_p\left(\gamma_{NT}^2\right)\right]O_p\left(1\right) = o_p\left(N^{-1/2}T^{-1}K^{1/2}\right)
\end{aligned}$$

because of $\breve{\nu}_{NT} = \gamma_{NT}\nu_{\Delta,NT} + \nu_{NT} = o_p\left(\gamma_{NT}\right) + O_p\left((NT)^{-1/2}\right)$. Noting that $\mathbb{V}_{NT}^{1/2} = O_p\left(K^{1/2}\right)$ by Lemma A.5(i), we have $J_{NT}^{(3)} = N^{1/2}T\Gamma_{NT}^{(3)}/\mathbb{V}_{NT}^{1/2} = o_p\left(1\right)$.

When $l = 4$, we write $\Gamma_{NT}^{(4)} = \frac{2\gamma_{NT}}{NT^2}\sum_{i=1}^{N}\varepsilon_i'\mathcal{K}_i\breve{g}_{\Delta,i} = \frac{2\gamma_{NT}}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{it}\dot{Z}_{it}'G_i$, where $G_i \equiv T^{-1}Q_{\dot{z},i}^{-1}Q_{w,i}Q_{\dot{z},i}^{-1}Z_i'M_{\iota_T}\breve{g}_{\Delta,i}$. Note that $E(\Gamma_{NT}^{(4)}|\mathbf{X}) = 0$ by Assumption 1(ii) and

$$\begin{aligned}
\mathrm{Var}(\Gamma_{NT}^{(4)}|\mathbf{X}) &= \frac{4\gamma_{NT}^2}{N^2T^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\dot{Z}_{it}'G_iG_i'Z_{it}\sigma_{it}^2 + \frac{8\gamma_{NT}^2}{N^2T^2}\sum_{i=1}^{N}\sum_{1\leq t<s\leq T}\dot{Z}_{it}'G_iG_i'\dot{Z}_{is}\sigma_{it}\sigma_{is}\mathrm{Cov}\left(\epsilon_{it},\epsilon_{is}\right) \\
&\equiv V\Gamma_{NT}^{(4a)} + V\Gamma_{NT}^{(4b)},\text{ say.}
\end{aligned}$$

For $V\Gamma_{NT}^{(4a)}$, we have

$$\begin{aligned}
V\Gamma_{NT}^{(4a)} &= \frac{4\gamma_{NT}^2}{N^2T^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\dot{Z}_{it}'Q_{\dot{z},i}^{-1}Q_{w,i}Q_{\dot{z},i}^{-1}Z_i'M_{\iota_T}\frac{\breve{g}_{\Delta,i}\breve{g}_{\Delta,i}'}{T}M_{\iota_T}Z_i/T\left(Q_{\dot{z},i}^{-1}Q_{w,i}Q_{\dot{z},i}^{-1}\right)\dot{Z}_{it}\sigma_{it}^2 \\
&\leq \frac{4\gamma_{NT}^2}{N^2T^2}\sum_{i=1}^{N}\lambda_{\max}\left(\frac{\breve{g}_{\Delta,i}\breve{g}_{\Delta,i}'}{T}\right)\sum_{t=1}^{T}\dot{Z}_{it}'Q_{\dot{z},i}^{-1}Q_{w,i}Q_{\dot{z},i}^{-1}Q_{w,i}Q_{\dot{z},i}^{-1}\dot{Z}_{it}\sigma_{it}^2 \\
&\leq \max_i \lambda_{\max}^2\left(Q_{\dot{z},i}^{-1}\right)\max_i \lambda_{\max}^2\left(Q_{w,i}\right)\frac{4\gamma_{NT}^2}{NT}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\left\|\breve{g}_{\Delta,i}\right\|^2}{T}\frac{1}{T}\sum_{t=1}^{T}\left\|\dot{Z}_{it}\right\|^2\sigma_{it}^2\right) \\
&\lesssim \frac{4\gamma_{NT}^2}{NT}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\left\|\breve{g}_{\Delta,i}\right\|^4}{T^2}\right)^{1/2}\left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{2K}{T}\sum_{t=1}^{T}A_{it}\sigma_{it}^2\right)^2\right]^{1/2} = O_p\left(\frac{\gamma_{NT}^2K}{NT}\right).
\end{aligned}$$

where we use $\lambda_{\max}\left(T^{-1}\breve{g}_{\Delta,i}\breve{g}_{\Delta,i}'\right) = T^{-1}\mathrm{tr}\left(\breve{g}_{\Delta,i}\breve{g}_{\Delta,i}'\right) = T^{-1}\left\|\breve{g}_{\Delta,i}\right\|^2$ in the second inequality, and in the last equation we use $N^{-1}T^{-2}\sum_{i=1}^{N}\left\|\breve{g}_{\Delta,i}\right\|^4 = O_p\left(1\right)$ and $N^{-1}\sum_{i=1}^{N}(T^{-1}\sum_{t=1}^{T}A_{it}\sigma_{it}^2)^2$ $= O_p\left(K^2\right)$ which can be easily verified by Markov inequality and moment conditions in Assumption 1(iv). For $V\Gamma_{NT}^{(4b)}$, by Davydov inequality (Bosq, 1998) again, we have

$$V\Gamma_{NT}^{(4b)} \leq \frac{8\gamma_{NT}^2}{N^2T^2}\sum_{i=1}^{N}\sum_{1\leq t<s\leq T}\left|\dot{Z}_{it}'G_iG_i'\dot{Z}_{is}\right|\sigma_{it}\sigma_{is}\left\|\epsilon_{it}\right\|_{2+2\eta}\left\|\epsilon_{is}\right\|_{2+2\eta}\alpha^{\frac{\eta}{1+\eta}}(t-s)$$
$$\lesssim \frac{\gamma_{NT}^2K}{N^2T^3}\sum_{i=1}^{N}\sum_{1\leq t<s\leq T}\left\|\breve{g}_{\Delta,i}\right\|^2A_{it}^{1/2}A_{is}^{1/2}\sigma_{it}\sigma_{is}\left\|\epsilon_{it}\right\|_{2+2\eta}\left\|\epsilon_{is}\right\|_{2+2\eta}\alpha^{\frac{\eta}{1+\eta}}(t-s) = O_p\left(\frac{\gamma_{NT}^2K}{NT}\right)$$

where we use the fact that $|\dot{Z}_{it}'G_iG_i'\dot{Z}_{is}| \leq T^{-1}\left\|\breve{g}_{\Delta,i}\right\|^2\lambda_{\max}(Q_{w,i}Q_{\dot{z},i}^{-1}Q_{w,i})\lambda_{\max}^2(Q_{\dot{z},i}^{-1})\left\|\dot{Z}_{it}\right\|\left\|\dot{Z}_{is}\right\|$ $\lesssim KA_{it}^{1/2}A_{is}^{1/2}T^{-1}\left\|\breve{g}_{\Delta,i}\right\|^2$ uniformly in i,t and s in the second inequality and the last equation can be verified as the determination of probability order of $\overline{VJ}_2$. By Chebyshev inequality, $\Gamma_{NT}^{(4)} = O_p(\gamma_{NT}\sqrt{K/(NT)}) = o_p\left(N^{-1/2}T^{-1}K^{1/2}\right)$. It follows that $J_{NT}^{(4)} = o_p(1)$.

When $l=5$, we can write $\Gamma_{NT}^{(5)} = F\breve{\nu}_{NT}$, where $F \equiv N^{-1}T^{-2}\sum_{i=1}^{N}\varepsilon_i'\mathcal{K}_iX_i$. Following the proof of $\Gamma_{NT}^{(4)}$, we can show that $F = O_p(\sqrt{K/(NT)})$. Then we have $\left|\Gamma_{NT}^{(5)}\right| \leq O_p(\sqrt{K/(NT)})[o_p(\gamma_{NT}) + O_p((NT)^{-1/2})] = o_p\left(N^{-1/2}T^{-1}K^{1/2}\right)$. It follows that $J_{NT}^{(5)} = o_p(1)$.

When $l=6$, we have $J_{NT}^{(6)} = o_p(1)$ by Cauchy-Schwarz inequality.■

Proposition B.1 *Suppose Assumptions 1-4 hold. We have* $J_{NT}^{(a)} = N^{1/2}T\Gamma_{NT}^{(1a)}/\mathbb{V}_{NT}^{1/2} \to_d N(0,1)$ *as* $(N,T)\to\infty$.

Proof. Write $J_{NT}^{(a)} = \sqrt{N}\overline{\mathcal{Z}}_N$, $\overline{\mathcal{Z}}_N = \frac{1}{N}\sum_{i=1}^{N}\mathcal{Z}_i$ with $\mathcal{Z}_i = \frac{2}{T\mathbb{V}_{NT}^{1/2}}\sum_{1\leq t<s\leq T}\tilde{\mathcal{K}}_{i,ts}\epsilon_{it}\epsilon_{is}$ and $\tilde{\mathcal{K}}_{i,ts} \equiv \mathcal{K}_{i,ts}\sigma_{it}\sigma_{is}$. Noting that $\mathcal{Z}_i$'s are independent but not identically distributed (inid) across i, we prove the proposition by the Linderberg-Feller CLT conditional on $\mathbf{X}$. We complete the proof by verifying Theorem 5.10 in White (2001). It suffices to show that (i) $\bar{\sigma}_N^2 = N\mathrm{Var}\left(\overline{\mathcal{Z}}_N|\mathbf{X}\right) = \mathrm{Var}\left(J_{NT}^{(a)}|\mathbf{X}\right) = 1 + o_p(1)$; and (ii) $E\mathcal{Z}_i^4 \leq C < \infty$ for all i.

Proof of (i). Noting that $\{\epsilon_{it}\}$ are an m.d.s., we have

$$\mathrm{Var}\left(J_{NT}^{(a)}|\mathbf{X}\right) = \frac{4}{NT^2\mathbb{V}_{NT}}\mathrm{Var}\left(\sum_{i=1}^{N}\sum_{1\leq t<s\leq T}\mathcal{K}_{i,ts}\varepsilon_{it}\varepsilon_{is}\right)$$
$$= \frac{4}{NT^2\mathbb{V}_{NT}}\sum_{i=1}^{N}\sum_{1\leq t_1<s_1\leq T}\sum_{1\leq t_2<s_2\leq T}\tilde{\mathcal{K}}_{i,t_1s_1}\tilde{\mathcal{K}}_{i,t_2s_2}E\left(\epsilon_{it_1}\epsilon_{it_2}\epsilon_{is_1}\epsilon_{is_2}\right)$$
$$= \frac{4}{NT^2\mathbb{V}_{NT}}\sum_{i=1}^{N}\sum_{1\leq t<s\leq T}\tilde{\mathcal{K}}_{i,ts}^2 + \frac{4}{NT^2\mathbb{V}_{NT}}\sum_{i=1}^{N}\sum_{1\leq t_1\neq t_2<t_3\leq T}\tilde{\mathcal{K}}_{i,t_1t_3}\tilde{\mathcal{K}}_{i,t_2t_3}E\left(\epsilon_{it_1}\epsilon_{it_2}\epsilon_{it_3}^2\right)$$
$$\equiv 1 + VJ_{NT}^{(a)}, \text{ say.}$$

We are left to show that $VJ_{NT}^{(a)} = o_p(1)$. For $VJ_{NT}^{(a)}$, we consider two cases for the time indices t_1,t_2,t_3: (a1) $|t_1-t_2| > t_3 - \max(t_1,t_2)$ and (a2) $|t_1-t_2| \leq t_3 - \max(t_1,t_2)$. Then we can write

$$VJ_{NT}^{(a)} = \frac{4}{NT^2\mathbb{V}_{NT}} \sum_{i=1}^{N} \left\{ \sum_{\text{case (a1)}} + \sum_{\text{case (a2)}} \right\} \tilde{\mathcal{K}}_{i,t_1t_3}\tilde{\mathcal{K}}_{i,t_2t_3} E\left(\epsilon_{it_1}\epsilon_{it_2}\epsilon_{it_3}^2\right) \equiv VJ_{NT}^{(a1)} + VJ_{NT}^{(a2)}, \text{ say.}$$

For $VJ_{NT}^{(a1)}$, we have $\left|E\left(\epsilon_{it_1}\epsilon_{it_2}\epsilon_{it_3}^2\right)\right| \leq 8\alpha^{\eta/(1+\eta)}\left(|t_2-t_1|\right)\|\epsilon_{it_1}\|_{2+2\eta}\left\|\epsilon_{it_2}\epsilon_{it_3}^2\right\|_{2+2\eta}$ by Davydov inequality. Then

$$\begin{aligned}
\left|VJ_{NT}^{(a1)}\right| &\leq \frac{64\max_{i,t}\|\epsilon_{it}\|_{2+2\eta}\max_{i,ts}\left(\left\|\epsilon_{it}\epsilon_{is}^2\right\|_{2+2\eta}\right)}{NT^2\mathbb{V}_{NT}} \sum_{i=1}^{N}\sum_{1\leq t_1<t_2<t_3\leq T}\left|\tilde{\mathcal{K}}_{i,t_1t_3}\right|\left|\tilde{\mathcal{K}}_{i,t_2t_3}\right|\alpha^{\frac{\eta}{1+\eta}}\left(t_2-t_1\right) \\
&\lesssim \frac{1}{NT^2\mathbb{V}_{NT}}\sum_{i=1}^{N}\sum_{1\leq t_1<t_2<t_3\leq T}\left\|\dot{Z}_{it_1}^*\right\|\left\|\dot{Z}_{it_2}^*\right\|\left\|\dot{Z}_{it_3}^*\right\|^2\alpha^{\frac{\eta}{1+\eta}}\left(t_2-t_1\right) \equiv \overline{VJ}_{NT}^{(a1)}, \text{ say,}
\end{aligned}$$

where $\dot{Z}_{it}^* = \dot{Z}_{it}\sigma_{it}$. Note that

$$\begin{aligned}
E\left(\overline{VJ}_{NT}^{(a1)}\right) &\leq C\max_{i,t_1,t_2,t_3} E\left(\left\|\dot{Z}_{it_1}^*\right\|\left\|\dot{Z}_{it_2}^*\right\|\left\|\dot{Z}_{it_3}^*\right\|^2\right)\frac{1}{T^2\mathbb{V}_{NT}}\sum_{t_1=1}^{T-2}\sum_{t_2=t_1+t_3-t_2}\sum_{t_3=t_2+1}^{T}\alpha^{\frac{\eta}{1+\eta}}\left(t_2-t_1\right) \\
&\leq \max_{i,t} E\left(\left\|\dot{Z}_{it}\sigma_{it}\right\|^4\right)\frac{C}{T^2\mathbb{V}_{NT}}\sum_{t_1=1}^{T-2}\sum_{t_2=t_1+t_3-t_2}\sum_{t_3=t_2+1}^{T}\alpha^{\frac{\eta}{1+\eta}}\left(t_2-t_1\right) \\
&\lesssim \frac{K^2}{T\mathbb{V}_{NT}}\frac{1}{T}\sum_{t_1=1}^{T-2}\sum_{l=2}^{T-1} l^2\alpha^{\frac{\eta}{1+\eta}}(l) = O\left(K/T\right) = o\left(1\right).
\end{aligned}$$

By the Markov inequality, we have $\overline{VJ}_{NT}^{(a1)} = o_p(1)$ and then $VJ_{NT}^{(a1)} = o_p(1)$. For (a2) with $t_1 < t_2 < t_3$, we have $\left|E\left(\epsilon_{it_1}\epsilon_{it_2}\epsilon_{it_3}^2\right)\right| \leq 8\alpha^{\eta/(1+\eta)}\left(\Delta t_3\right)\|\epsilon_{it_1}\epsilon_{it_2}\|_{2+2\eta}\left\|\epsilon_{it_3}^2\right\|_{2+2\eta}$, where $\Delta t_3 = t_3 - t_2$. Then

$$\begin{aligned}
E\left|VJ_{NT}^{(a2)}\right| &\leq 64\max_{i,ts}\left(\|\epsilon_{it}\epsilon_{is}\|_{2+2\eta}\right)\max_{i,t}\left(\left\|\epsilon_{it}^2\right\|_{2+2\eta}\right) \\
&\quad\times\frac{1}{NT^2\mathbb{V}_{NT}}\sum_{i=1}^{N}\sum_{1\leq t_1<t_2<t_3\leq T}E\left(\left|\tilde{\mathcal{K}}_{i,t_1t_3}\right|\left|\tilde{\mathcal{K}}_{i,t_2t_3}\right|\right)\alpha^{\frac{\eta}{1+\eta}}\left(\Delta t_3\right) \\
&\lesssim \max_{i,t} E\left(\left\|\dot{Z}_{it}\sigma_{it}\right\|^4\right)\frac{1}{NT^2\mathbb{V}_{NT}}\sum_{i=1}^{N}\sum_{1\leq t_1<t_2<t_3\leq T}\alpha^{\frac{\eta}{1+\eta}}\left(\Delta t_3\right) = O\left(K^2/T\right)
\end{aligned}$$

It follows that $VJ_{NT}^{(a2)} = O_p\left(K^2/T\right) = o_p(1)$ by the Markov inequality.

Proof of (ii) Note that

$$\begin{aligned}
E\left(\mathcal{Z}_i^4|\mathbf{X}\right) &= \frac{16}{T^4\mathbb{V}_{NT}^2}\sum_{\substack{1\leq t_1<t_2\leq T, 1\leq t_5<t_6\leq T\\ 1\leq t_3<t_4\leq T, 1\leq t_7<t_8\leq T}}\tilde{\mathcal{K}}_{i,t_1t_2}\tilde{\mathcal{K}}_{i,t_3t_4}\tilde{\mathcal{K}}_{i,t_5t_6}\tilde{\mathcal{K}}_{i,t_7t_8}E\left(\epsilon_{it_1}\epsilon_{it_2}\epsilon_{it_3}\epsilon_{it_4}\epsilon_{it_5}\epsilon_{it_6}\epsilon_{it_7}\epsilon_{it_8}\right) \\
&\equiv DJ_{i2} + \cdots + DJ_{i7}, \text{ say,}
\end{aligned}$$

where $DJ_{i2}, \ldots, DJ_{i7}$ denote the summations of terms with $2, \ldots, 7$ different time indices in the expectation, respectively. Note that the expectation for any term with 8 distinct time indices is 0 since $\{\epsilon_{it}\}_{t=1}^{T}$ is an MDS.

First, we consider the case with two different time indices (DJ_{i2}). We have

$$DJ_{i2} = \frac{16}{T^4 \mathbb{V}_{NT}^2} \sum_{1 \le t_1 < t_2 \le T} \tilde{\mathcal{K}}_{i,t_1 t_2}^4 E\left(\epsilon_{it_1}^4 \epsilon_{it_2}^4\right) \le \frac{16 C_*^2 K^4}{T^4 \mathbb{V}_{NT}^2} \sum_{1 \le t_1 < t_2 \le T} A_{it_1}^2 A_{it_2}^2 E\left(\epsilon_{it_1}^4 \epsilon_{it_2}^4\right)$$
$$= O_p\left(K^2 T^{-2}\right) = o_p(1).$$

because of $\tilde{\mathcal{K}}_{i,ts}^2 \le \tilde{\mathcal{K}}_{i,tt} \tilde{\mathcal{K}}_{i,ss} \le C_*^2 \sigma_{it}^2 \sigma_{it}^2 A_{it} A_{is}$. Similarly, we can show that $DJ_{i3} = O_p\left(K^2 T^{-1}\right)$.

Second, we consider the case with four different time indices (DJ_{i4}). As we will see from the proof of DJ_{i7} below, the leading term in DJ_{i4} is

$$DJ_{i4}^{\blacklozenge} \lesssim \frac{1}{T^4 \mathbb{V}_{NT}^2} \sum_{t \ne s \ne l \ne q} \left(\tilde{\mathcal{K}}_{i,ts}^2 \tilde{\mathcal{K}}_{i,lq}^2 + \tilde{\mathcal{K}}_{i,ts} \tilde{\mathcal{K}}_{i,tl} \tilde{\mathcal{K}}_{i,lq} \tilde{\mathcal{K}}_{i,qs}\right) E\left(\epsilon_{it}^2 \epsilon_{is}^2 \epsilon_{il}^2 \epsilon_{iq}^2\right) \equiv DJ_{i41}^{*\blacklozenge} + DJ_{i42}^{\blacklozenge}, \text{ say,}$$

where 8 time indices form 4 different pairs. Let $\breve{Q}_i \equiv Q_{\dot{z},i}^{-1} Q_{w,i} Q_{\dot{z},i}^{-1}$. For $DJ_{i41}^{\blacklozenge}$, we have

$$\begin{aligned}
DJ_{i41}^{\blacklozenge} &\le \max_{i,tsql} \left\{E\left(\epsilon_{it}^2 \epsilon_{is}^2 \epsilon_{il}^2 \epsilon_{iq}^2\right)\right\} \frac{1}{T^4 \mathbb{V}_{NT}^2} \left(\textstyle\sum_{1 \le t,s \le T} \tilde{\mathcal{K}}_{i,ts}^2\right)^2 \\
&\lesssim \frac{1}{T^4 \mathbb{V}_{NT}^2} \left(\textstyle\sum_{1 \le t,s \le T} \dot{Z}_{it} Q_{\dot{z},i}^{-1} Q_{w,i} Q_{\dot{z},i}^{-1} \dot{Z}_{is} \sigma_{it} \sigma_{is}\right)^2 \\
&= \left[\frac{\operatorname{tr}\left(Q_{\dot{z},i}^{(\sigma)} \breve{Q}_i Q_{\dot{z},i}^{(\sigma)} \breve{Q}_i\right) + o_p(1)}{N^{-1} \sum_{i=1}^{N} \operatorname{tr}\left(Q_{\dot{z},i}^{(\sigma)} \breve{Q}_i Q_{\dot{z},i}^{(\sigma)} \breve{Q}_i\right) + o_p(1)}\right]^2 \\
&\le \left[\frac{\max_i \lambda_{\max}^2 (Q_{w,i}) \max_i \lambda_{\max}^4 \left(Q_{\dot{z},i}^{-1}\right) \max_i \lambda_{\max}^2 \left(Q_{\dot{z},i}^{(\sigma)}\right)}{\min_i \lambda_{\min}^2 (Q_{w,i}) \min_i \lambda_{\min}^4 \left(Q_{\dot{z},i}^{-1}\right) \min_i \lambda_{\min}^2 \left(Q_{\dot{z},i}^{(\sigma)}\right)}\right]^2 + o_p(1) \le C < \infty.
\end{aligned}$$

For $DJ_{i42}^{*\blacklozenge}$, we have

$$\begin{aligned}
DJ_{i42}^{\blacklozenge} &\le \max_{i,tsql} \left\{E\left(\epsilon_{it}^2 \epsilon_{is}^2 \epsilon_{il}^2 \epsilon_{iq}^2\right)\right\} \frac{1}{T^4 \mathbb{V}_{NT}^2} \sum_{t \ne s \ne l \ne q} \tilde{\mathcal{K}}_{i,ts} \tilde{\mathcal{K}}_{i,tl} \tilde{\mathcal{K}}_{i,lq} \tilde{\mathcal{K}}_{i,qs} \\
&\lesssim \frac{1}{T^4 \mathbb{V}_{NT}^2} \sum_{t \ne s \ne l \ne q} \sigma_{is} \dot{Z}_{is}' \breve{Q}_i \sigma_{it}^2 \dot{Z}_{it} \dot{Z}_{it}' \breve{Q}_i \sigma_{il}^2 \dot{Z}_{il} \breve{Q}_i \sigma_{iq}^2 \dot{Z}_{iq} \breve{Q}_i \dot{Z}_{is} \sigma_{is} \\
&\lesssim \frac{\operatorname{tr}\left(Q_{\dot{z},i}^{(\sigma)} \breve{Q}_i Q_{\dot{z},i}^{(\sigma)} \breve{Q}_i Q_{\dot{z},i}^{(\sigma)} \breve{Q}_i Q_{\dot{z},i}^{(\sigma)} \breve{Q}_i\right) (1 + o_p(1))}{\left[N^{-1} \sum_{i=1}^{N} \operatorname{tr}\left(Q_{\dot{z},i}^{(\sigma)} \breve{Q}_i Q_{\dot{z},i}^{(\sigma)} \breve{Q}_i\right)\right]^2 (1 + o_p(1))} = O_p\left(K^{-1}\right) < \infty.
\end{aligned}$$

Now, we consider DJ_{i7}. Without loss of generality (WLOG), let $s_1 < \cdots < s_7$ be the rearranged time indices, and two t_l's take the same value s_7. Otherwise, the expectation should be 0 because $\{\epsilon_{it}\}_{t=1}^{T}$ is an MDS. Then following the proof of Lemma A.1 in Gao (2007), let d_1 be the first largest difference among $\{\Delta s_{j+1} = s_{j+1} - s_j\}_{j=1}^{6}$. Noting that $E(\prod_{j=1}^{j^*} \epsilon_{is_j}) = 0$, we can apply the Davydov inequality to $E\left(\epsilon_{is_1}\epsilon_{is_2}\epsilon_{is_3}\epsilon_{is_4}\epsilon_{is_5}\epsilon_{is_6}\epsilon_{is_7}^2\right)$ by separating the set of time indices into two subsets $\{s_1, \ldots, s_{j^*}\}$ and $\{s_{j^*+1}, \ldots, s_7\}$. Then we have

$$\begin{aligned}\left|E\left(\epsilon_{is_1}\epsilon_{is_2}\epsilon_{is_3}\epsilon_{is_4}\epsilon_{is_5}\epsilon_{is_6}\epsilon_{is_7}^2\right)\right| &= \left|E\left(\epsilon_{is_1}\epsilon_{is_2}\epsilon_{is_3}\epsilon_{is_4}\epsilon_{is_5}\epsilon_{is_6}\epsilon_{is_7}^2\right) - E\left(\prod_{j=1}^{j^*}\epsilon_{is_j}\right)E\left(\prod_{j=j^*+1}^{6}\epsilon_{is_j}\epsilon_{is_7}^2\right)\right| \\ &\leq 8\left\|\prod_{j=1}^{j^*}\epsilon_{is_j}\right\|_{2+2\eta}\left\|\epsilon_{is_7}^2\prod_{j=j^*+1}^{6}\epsilon_{is_j}\right\|_{2+2\eta}\alpha^{\frac{\eta}{1+\eta}}(d_1)\end{aligned}$$

and

$$|DJ_{NT,7}| \leq \frac{128C_7^2}{T^4\mathbb{V}_{NT}^2}\sum_{j^*=1}^{6}\sum_{\substack{1\leq s_1<\cdots<s_7\leq T\\ \Delta s_{j^*+1}=d_1}}\alpha^{\frac{\eta}{1+\eta}}(d_1)\left\|\dot{Z}_{is_7}^*\right\|^2\prod_{l=1}^{6}\left\|\dot{Z}_{is_l}^*\right\| \equiv \sum_{j^*=1}^{6}\overline{DJ}_{i7j^*},$$

where $C_7 \equiv \max_{i,s_1,\ldots,s_7}\max_{j^*=1,\ldots,6}\left(\left\|\prod_{j=1}^{j^*}\epsilon_{is_j}\right\|_{2+2\eta}\left\|\prod_{j=j^*+1}^{6}\epsilon_{is_j}\epsilon_{is_7}^2\right\|_{2+2\eta}\right)$ and

$$\overline{DJ}_{i7j^*} = \frac{128C_7^2}{T^4\mathbb{V}_{NT}^2}\sum_{1\leq s_1<\cdots<s_7\leq T,\Delta s_{j^*+1}=d_1}\alpha^{\frac{\eta}{1+\eta}}(d_1)\left\|\dot{Z}_{is_7}^*\right\|^2\prod_{l=1}^{6}\left\|\dot{Z}_{is_l}^*\right\|$$

for $j^* = 1, \ldots, 6$. We show that $\overline{DJ}_{i7j^*} = O_p\left(K^2T^{-3}\right)$ for all $j^* = 1, \ldots, 6$. For example, when $j^* = 2$, we have

$$\begin{aligned}E\left(\overline{DJ}_{i72}\right) &\leq \max_{i,s_1,\ldots,s_7}E\left(\left\|\dot{Z}_{is_7}^*\right\|^2\prod_{l=1}^{6}\left\|\dot{Z}_{is_l}^*\right\|\right)\frac{128C_7^2}{T^4\mathbb{V}_{NT}^2}\sum_{1\leq s_1<s_2\leq\cdots\leq s_7<T,\ \Delta s_3=d_1}\alpha^{\frac{\eta}{1+\eta}}(d_1) \\ &\lesssim \frac{K^4}{T^4\mathbb{V}_{NT}^2}\sum_{s_2=2}^{T-5}\sum_{d_1=2}^{T-4}\sum_{s_1=\max\{s_2-d_1+1,1\}}^{s_2-1}\sum_{s_4=s_2+d_1+1}^{s_2+2d_1}\sum_{s_5=s_4+1}^{s_4+d_1}\sum_{s_6=s_5+1}^{s_5+d_1}\sum_{s_7=s_6+1}^{\min\{s_6+d_1-1,T\}}\alpha^{\frac{\eta}{1+\eta}}(d_1) \\ &\lesssim \frac{K^4}{T^4\mathbb{V}_{NT}^2}\sum_{s=2}^{T}\sum_{d=1}^{T}d_1^5\alpha^{\frac{\eta}{1+\eta}}(d_1) = O\left(K^2T^{-3}\right).\end{aligned}$$

Similarly, $\overline{DJ}_{i7j^*} = O_p\left(K^2T^{-3}\right)$ for $j^* = 1, 3, \ldots, 6$. It follows that $DJ_{i7} = o_p(1)$.

For DJ_{i6}, WLOG, let $s_1 < \cdots < s_6$ be the rearranged time indices. Then we have: (a) three t_j's take the same value s_6; (b) two t_l's take the same value s_6, two t_l's take s_j for some $j < 6$, and remaining 4 t_l's take different values. Without confusion, we decompose $DJ_{i6} = DJ_{i6}^{(a)} + DJ_{i6}^{(b)}$ according to two subcases (a) and (b). For subcase (a), following the proof of DJ_{i7}, we have $DJ_{i6}^{(a)} = O_p\left(K^2T^{-3}\right)$. For subcase (b), we further decompose $DJ_{i6}^{(b)} \equiv \sum_{j=1}^{5}DJ_{i6j}^{(b)}$, where $DJ_{i6j}^{(b)}$ corresponds to the term with two t's take the same value s_j for $j = 1, \ldots, 5$. We first consider $DJ_{i65}^{(b)}$. Let d_1 be the first largest difference among $\Delta s_2, \Delta s_3, \Delta s_4$, and Δs_5. Then we have 4 subsubcases according to $d_1 = \Delta s_{j^*}$ for $j^* = 1, \ldots, 4$, respectively. By Davydov inequality, we have

$$\begin{aligned}\left|E\left(\epsilon_{is_1}\epsilon_{is_2}\epsilon_{is_3}\epsilon_{is_4}\epsilon_{is_5}^2\epsilon_{is_6}^2\right)\right| &= \left|E\left(\epsilon_{is_1}\epsilon_{is_2}\epsilon_{is_3}\epsilon_{is_4}\epsilon_{is_5}^2\epsilon_{is_6}^2\right) - E\left(\prod_{j=1}^{j^*}\epsilon_{is_j}\right)E\left(\epsilon_{is_5}^2\epsilon_{is_6}^2\prod_{j=j^*+1}^{4}\epsilon_{is_j}\right)\right| \\ &\leq 8\left\|\prod_{j=1}^{j^*}\epsilon_{is_j}\right\|_{2+2\eta}\left\|\epsilon_{is_5}^2\epsilon_{is_6}^2\prod_{j=j^*+1}^{4}\epsilon_{is_j}\right\|_{2+2\eta}\alpha^{\frac{\eta}{1+\eta}}(d_1),\end{aligned}$$

where we separate $\{s_1,\ldots,s_6\}$ into $\{s_1,\ldots,s_{j^*}\}$ and $\{s_{j^*+1},\ldots,s_6\}$. Let

$$C_{65} \equiv \max_{i,s_1,\ldots,s_6}\max_{j^*=1,\ldots,4}\left\{\left\|\prod_{j=1}^{j^*}\epsilon_{is_j}\right\|_{2+2\eta}\left\|\epsilon_{is_5}^2\epsilon_{is_6}^2\prod_{j=j^*+1}^{4}\epsilon_{is_j}\right\|_{2+2\eta}\right\}.$$

Then following the proof of DJ_{i7}, we have

$$E\left|DJ_{i65}^{(b)}\right| \leq 8C_{65}\max_{i,s_1,\ldots,s_6}E\left(\left\|\dot{Z}_{is_5}^*\right\|^2\left\|\dot{Z}_{is_6}^*\right\|^2\prod_{l=1}^{4}\left\|\dot{Z}_{is_l}^*\right\|\right)\times\sum_{j^*=1}^{4}EDJ_{i65,j^*}^{(b)}.$$

where $EDJ_{i65,j^*}^{(b)} \equiv \frac{K^4}{T^4\mathbb{V}_{NT}^2}\sum_{1\leq s_1<\cdots<s_6\leq T,\Delta s_{j^*+1}=d_1}\alpha^{\eta/(1+\eta)}(d_1)$ for $j^*=1,2,3,4$. For $EDJ_{i65,1}^{(b)}$, we have

$$\begin{aligned}EDJ_{i65,1}^{(b)} &= \frac{K^4}{T^4\mathbb{V}_{NT}^2}\sum_{s_1=1}^{T-6}\sum_{d_1=2}^{T-6}\sum_{s_3=s_1+d_1+1}^{s_1+2d_1}\sum_{s_4=s_3+1}^{s_3+d_1}\sum_{s_5=s_4+1}^{s_4+d_1}\sum_{s_6=s_5+1}^{T}\alpha^{\frac{\eta}{1+\eta}}(d_1) \\ &\leq \frac{K^4}{T^4\mathbb{V}_{NT}^2}\sum_{s_1=1}^{T}\sum_{d_1=1}^{T}\sum_{s_6=1}^{T}d_1^3\alpha^{\frac{\eta}{1+\eta}}(d_1) = O\left(K^2/T^2\right).\end{aligned}$$

Similarly, we have $EDJ_{i65,j^*}^{(b)} = O\left(K^2T^{-2}\right)$ for all $j^*=2,\ldots,4$. It follows that $E\left|DJ_{i65}^{(b)}\right| = O\left(K^2T^{-2}\right)$ and $DJ_{i65}^{(b)} = O_p\left(K^2T^{-2}\right)$ by Markov inequality. Now, we turn to the term $DJ_{i64}^{(b)}$. Let d_1 and d_2 be the first and second largest difference among $\Delta s_2,\Delta s_3,\Delta s_4,\Delta s_5$ and Δs_6. We consider two subsubcases for $DJ_{i64}^{(b)}$: (b1) $d_1\neq\Delta s_5$ and (b2) $d_1=\Delta s_5$. Let $DJ_{i641}^{(b)}$ and $DJ_{i642}^{(b)}$ be the corresponding terms for (b1) and (b2). Then we have $DJ_{i64}^{(b)} = DJ_{i641}^{(b)} + DJ_{i642}^{(b)}$. Following the proof of DJ_{i7}, we have $DJ_{i641}^{(b)} = O_p\left(K^2/T^2\right)$. For the subsubcase (b2), it must be $d_2=\Delta s_j$, where $j=2,3,4$,or 6 since $d_1=\Delta s_5$. We can decompose $DJ_{i642}^{(b)} = DJ_{i6422}^{(b)} + DJ_{i6423}^{(b)} + DJ_{i6424}^{(b)} + DJ_{i6426}^{(b)}$, where $DJ_{i642j}^{(b)}$ is the term with $d_2=\Delta s_j$. For $DJ_{i6422}^{(b)}$, we apply Davydov inequality to get that

$$\begin{aligned}\left|E\left(\epsilon_{is_1}\epsilon_{is_2}\epsilon_{is_3}\epsilon_{is_4}^2\epsilon_{is_5}\epsilon_{is_6}^2\right)\right| &= \left|E\left(\epsilon_{is_1}\epsilon_{is_2}\epsilon_{is_3}\epsilon_{is_4}^2\epsilon_{is_5}\epsilon_{is_6}^2\right) - E\left(\epsilon_{is_1}\right)E\left(\epsilon_{is_2}\epsilon_{is_3}\epsilon_{is_4}^4\epsilon_{is_5}\epsilon_{is_6}^2\right)\right| \\ &\leq 8\left\|\epsilon_{is_1}\right\|_{2+2\eta}\left\|\epsilon_{is_2}\epsilon_{is_3}\epsilon_{is_4}^2\epsilon_{is_5}\epsilon_{is_6}^2\right\|_{2+2\eta}\alpha^{\frac{\eta}{1+\eta}}(d_2)\end{aligned}$$

by separating $\{s_1,\ldots,s_6\}$ into $\{s_1\}$ and $\{s_2,\ldots,s_6\}$ according to the second largest increment.

Let $C_{64} \equiv \max_{i,s_1,\ldots,s_6}\left\{\|\epsilon_{is_1}\|_{2+2\eta}\left\|\epsilon_{is_2}\epsilon_{is_3}\epsilon_{is_4}^2\epsilon_{is_5}\epsilon_{is_6}^2\right\|_{2+2\eta}\right\}$. Then we have

$$
\begin{aligned}
E\left|DJ_{i6422}^{(b)}\right| &\leq C_{64}\max_{i,s_1,\ldots,s_6} E\left(\left\|\dot{Z}_{is_4}^*\right\|^2\left\|\dot{Z}_{is_5}^*\right\|\left\|\dot{Z}_{is_6}^*\right\|^2\prod_{l=1}^3\left\|\dot{Z}_{is_l}^*\right\|\right)\frac{1}{T^4\mathbb{V}_{NT}^2}\sum_{\substack{1\leq s_1<\cdots<s_6\leq T\\ \Delta s_2=d_2,\Delta s_5=d_1}}\alpha^{\frac{\eta}{1+\eta}}(d_2)\\
&\lesssim \frac{K^2}{T^4\mathbb{V}_{NT}^2}\sum_{s_1=1}^{T-6}\sum_{d_2=2}^{d_1}\sum_{s_3=s_1+d_2+1}^{s_1+2d_2}\sum_{s_4=s_3+1}^{s_3+d_2}\sum_{d_1=2}^{T-4}\sum_{s_6=s_4+d_1+1}^{\min\{s_4+d_1+d_2,T\}}\alpha^{\frac{\eta}{1+\eta}}(d_2)\\
&\leq \frac{TK^2}{T^4\mathbb{V}_{NT}^2}\sum_{d_1=2}^{T}\sum_{d_2=1}^{d_1}d_2^3\alpha^{\frac{\eta}{1+\eta}}(d_2)=O\left(K^2/T^2\right)
\end{aligned}
$$

and $DJ_{i6422}^{(b)}=O_p\left(K^2/T^2\right)$ by Markov inequality. Similarly, we have $DJ_{i642j}^{(b)}=O_p\left(K^2/T^2\right)$ for $j=3,4,6$. Then $DJ_{i642}^{(b)}=O_p\left(K^2/T^2\right)$. It follows that $DJ_{i64}^{(b)}=O_p\left(K^2/T^2\right)$. In the same way, we can show that $DJ_{i6j}^{(b)}=O_p\left(K^2T^{-2}\right)$ for $j=1,2,3$. Then we have 4 subsubcases according to $d_1=\Delta s_{j^*}$ for $j^*=1,\ldots,4$, respectively.

Similarly, we can show that $DJ_{i5}=O_p\left(K^2T^{-3}\right)+O_p\left(K^2T^{-2}\right)+O_p\left(K^2T^{-1}\right)=o_p(1)$. ■

Proposition B.2 *Under Assumptions 1-4,, we have* $\hat{\mathbb{B}}_{NT}-\mathbb{B}_{NT}=o_p(K^{1/2})$.

Proof. Note that $\hat{\varepsilon}_{r,it}=\hat{u}_{it}-\overline{\hat{u}}_i=\varepsilon_{it}-\bar{\varepsilon}_i+\gamma_{NT}\breve{g}_{\Delta,it}^{(c)}-\dot{X}_{it}'\breve{\nu}_{NT}$ under $\mathbb{H}_{1,\gamma_{NT}}$, where $\breve{g}_{\Delta,it}^{(c)}=\breve{g}_{\Delta,it}-\bar{\breve{g}}_{\Delta,i}$, $\dot{X}_{it}=X_{it}-\bar{X}_i$, $\bar{\varepsilon}_i,\bar{\breve{g}}_{\Delta,i}$ and $\bar{X}_i$ are time series average of ε_{it}'s, $\breve{m}_{\Delta,it}$'s and X_{it}'s for the ith individual, respectively. Then we can write

$$
\hat{\mathbb{B}}_{NT}=\frac{1}{\sqrt{N}T}\sum_{i=1}^N\sum_{t=1}^T\mathcal{K}_{i,tt}\left(\varepsilon_{it}-\bar{\varepsilon}_i+\gamma_{NT}\breve{g}_{\Delta,it}^{(c)}-\dot{X}_{it}'\breve{\nu}_{NT}\right)^2=\sum_{l=1}^{10}\hat{\mathbb{B}}_{NTl},
$$

where

$$
\begin{aligned}
&\hat{\mathbb{B}}_{NT1}\equiv\tfrac{1}{\sqrt{N}T}\sum_{i=1}^N\sum_{t=1}^T\mathcal{K}_{i,tt}\varepsilon_{it}^2, && \hat{\mathbb{B}}_{NT2}\equiv\tfrac{1}{\sqrt{N}T}\sum_{i=1}^N\sum_{t=1}^T\mathcal{K}_{i,tt}\bar{\varepsilon}_i^2,\\
&\hat{\mathbb{B}}_{NT3}\equiv\tfrac{\gamma_{NT}^2}{\sqrt{N}T}\sum_{i=1}^N\sum_{t=1}^T\mathcal{K}_{i,tt}\left(\breve{g}_{\Delta,it}^{(c)}\right)^2, && \hat{\mathbb{B}}_{NT4}\equiv\tfrac{1}{\sqrt{N}T}\sum_{i=1}^N\sum_{t=1}^T\mathcal{K}_{i,tt}\breve{\nu}_{NT}'\dot{X}_{it}\dot{X}_{it}'\breve{\nu}_{NT},\\
&\hat{\mathbb{B}}_{NT5}\equiv\tfrac{-2}{\sqrt{N}T}\sum_{i=1}^N\sum_{t=1}^T\mathcal{K}_{i,tt}\varepsilon_{it}\bar{\varepsilon}_i, && \hat{\mathbb{B}}_{NT6}\equiv\tfrac{2\gamma_{NT}}{\sqrt{N}T}\sum_{i=1}^N\sum_{t=1}^T\mathcal{K}_{i,tt}\varepsilon_{it}\breve{g}_{\Delta,it}^{(c)},\\
&\hat{\mathbb{B}}_{NT7}\equiv\tfrac{2}{\sqrt{N}T}\sum_{i=1}^N\sum_{t=1}^T\mathcal{K}_{i,tt}\varepsilon_{it}\dot{X}_{it}'\breve{\nu}_{NT}, && \hat{\mathbb{B}}_{NT8}\equiv\tfrac{-2\gamma_{NT}}{\sqrt{N}T}\sum_{i=1}^N\sum_{t=1}^T\mathcal{K}_{i,tt}\bar{\varepsilon}_i\breve{g}_{\Delta,it}^{(c)},\\
&\hat{\mathbb{B}}_{NT9}\equiv\tfrac{-2}{\sqrt{N}T}\sum_{i=1}^N\sum_{t=1}^T\mathcal{K}_{i,tt}\bar{\varepsilon}_i\dot{X}_{it}'\breve{\nu}_{NT}, && \hat{\mathbb{B}}_{NT10}\equiv\tfrac{\gamma_{NT}^2}{\sqrt{N}T}\sum_{i=1}^N\sum_{t=1}^T\mathcal{K}_{i,tt}\breve{g}_{\Delta,it}^{(c)}\dot{X}_{it}'\breve{\nu}_{NT}.
\end{aligned}
$$

We complete the proof of (iv) by showing that $\hat{\mathbb{B}}_{NT1}-\mathbb{B}_{NT}=o_p\left(K^{1/2}\right)$, and $\hat{\mathbb{B}}_{NTs}=o_p\left(K^{1/2}\right)$ for $s=2,\ldots,10$.

First, we have shown that $\tilde{J}_{NT}^{(b)}=\hat{\mathbb{B}}_{NT1}-\mathbb{B}_{NT}=o_p\left(K^{1/2}\right)$ in the proof of (i) of Theorem 3.2. Second, we have $\hat{\mathbb{B}}_{NT2}\leq\frac{1}{N^{1/2}T}\sum_{i=1}^N\bar{\varepsilon}_i^2\mathrm{tr}(\mathcal{K}_i)=\frac{1}{N^{1/2}}\sum_{i=1}^N\bar{\varepsilon}_i^2\mathrm{tr}(Q_{\dot{z},i}^{-1}Q_{w,i})\leq\underline{c}_{\dot{z}}^{-1}\bar{c}_wK\times\left(\frac{1}{N^{1/2}}\sum_{i=1}^N\bar{\varepsilon}_i^2\right)=O(N^{1/2}KT^{-1})=o_p\left(K^{1/2}\right)$. Third, $\hat{\mathbb{B}}_{NT3}\leq C_*\gamma_{NT}^2KN^{1/2}\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^TA_{it}[\breve{g}_{\Delta,it}^{(c)}]^2$

$= O_p\left(KN^{1/2}\gamma_{NT}^2\right) = o_p\left(K^{1/2}\right)$. Fourth, $\hat{\mathbb{B}}_{NT4} \leq C_* KN^{1/2}\left\|\breve{\nu}_{NT}\right\|^2 \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T A_{it}\left\|\dot{X}_{it}\right\|^2$ $= O_p(KN^{1/2}\left\|\breve{\nu}_{NT}\right\|^2) = o_p\left(K^{1/2}\right)$. By the Cauchy-Schwarz inequality, we can show that $\hat{\mathbb{B}}_{NTs} = o_p\left(K^{1/2}\right)$ for $s = 5, \ldots, 10$. ■

Proposition B.3 *Under Assumptions 1-4, we have* $\hat{\mathbb{V}}_{NT}/\mathbb{V}_{NT} = 1 + o_p(1)$.

Proof. We consider the following decomposition

$$\hat{\mathbb{V}}_{NT} - \mathbb{V}_{NT} = \frac{2}{NT^2}\sum_{i=1}^N\sum_{1\leq s\neq t\leq T}\mathcal{K}_{i,ts}^2\left(\hat{\varepsilon}_{r,it}^2\hat{\varepsilon}_{r,is}^2 - \varepsilon_{it}^2\varepsilon_{is}^2\right) + \frac{2}{NT^2}\sum_{i=1}^N\sum_{1\leq s\neq t\leq T}\mathcal{K}_{i,ts}^2\left(\varepsilon_{it}^2\varepsilon_{is}^2 - \sigma_{it}^2\sigma_{is}^2\right)$$
$$\equiv \Delta\hat{\mathbb{V}}_{NT}^{(a)} + \Delta\hat{\mathbb{V}}_{NT}^{(b)}, \text{ say.}$$

We first show that $\Delta\hat{\mathbb{V}}_{NT}^{(a)} = o_p(K)$. Let $\breve{\varepsilon}_{R,it} = \bar{\varepsilon}_i + \gamma_{NT}\breve{g}_{\Delta,it}^{(c)} - \dot{X}_{it}'\breve{\nu}_{NT}$. It is straightforward to verify that

$$\text{(i) } \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\breve{\varepsilon}_{R,it}^2 = O_p\left(T^{-1}\right) \text{ and (ii) } \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\breve{\varepsilon}_{R,it}^4 = O_p\left(T^{-2}\right). \tag{A.4}$$

We rewrite $\Delta\hat{\mathbb{V}}_{NT}^{(a)}$ as

$$\begin{aligned}
\Delta\hat{\mathbb{V}}_{NT}^{(a)} &= \frac{2}{NT^2}\sum_{i=1}^N\sum_{1\leq s\neq t\leq T}\mathcal{K}_{i,ts}^2\left(\hat{\varepsilon}_{r,it}\hat{\varepsilon}_{r,is} - \varepsilon_{it}\varepsilon_{is}\right)\left(\hat{\varepsilon}_{r,it}\hat{\varepsilon}_{r,is} + \varepsilon_{it}\varepsilon_{is}\right) \\
&= \frac{2}{NT^2}\sum_{i=1}^N\sum_{1\leq s\neq t\leq T}\mathcal{K}_{i,ts}^2\left(\breve{\varepsilon}_{R,it}\varepsilon_{is} + \breve{\varepsilon}_{R,is}\varepsilon_{it} + \breve{\varepsilon}_{R,is}\breve{\varepsilon}_{R,it}\right)\left(2\varepsilon_{it}\varepsilon_{is} + \breve{\varepsilon}_{R,it}\varepsilon_{is} + \breve{\varepsilon}_{R,is}\varepsilon_{it} + \breve{\varepsilon}_{R,is}\breve{\varepsilon}_{R,it}\right) \\
&= \frac{2}{NT^2}\sum_{i=1}^N\sum_{1\leq s\neq t\leq T}\mathcal{K}_{i,ts}^2\left(4\varepsilon_{is}^2\varepsilon_{it}\breve{\varepsilon}_{R,it} + 4\breve{\varepsilon}_{R,it}\breve{\varepsilon}_{R,is}\varepsilon_{it}\varepsilon_{is} + 4\breve{\varepsilon}_{R,is}^2\varepsilon_{it}^2 + 4\breve{\varepsilon}_{R,it}\breve{\varepsilon}_{R,is}^2\varepsilon_{it} + \breve{\varepsilon}_{R,is}^2\breve{\varepsilon}_{R,it}^2\right) \\
&\equiv \sum_{s=1}^5\Delta\hat{\mathbb{V}}_{NT,s}^{(a)}, \text{ say,}
\end{aligned}$$

by the symmetricity between time indices t and s.

First, we can decompose $\Delta\hat{\mathbb{V}}_{NT,1}^{(a)}$ as follows

$$\begin{aligned}
\Delta\hat{\mathbb{V}}_{NT,1}^{(a)} &= \frac{8}{NT^2}\sum_{i=1}^N\sum_{s=1}^T\sum_{t=1}^T\mathcal{K}_{i,ts}^2\varepsilon_{is}^2\varepsilon_{it}\bar{\varepsilon}_i + \frac{8\gamma_{NT}}{NT^2}\sum_{i=1}^N\sum_{s=1}^T\sum_{t=1}^T\mathcal{K}_{i,ts}^2\varepsilon_{is}^2\varepsilon_{it}\breve{g}_{\Delta,it}^{(c)} \\
&\quad - \frac{8}{NT^2}\sum_{i=1}^N\sum_{s=1}^T\sum_{t=1}^T\mathcal{K}_{i,ts}^2\varepsilon_{is}^2\varepsilon_{it}\dot{X}_{it}'\breve{\nu}_{NT} - \frac{8}{NT^2}\sum_{i=1}^N\sum_{t=1}^T\mathcal{K}_{i,tt}^2\varepsilon_{it}^3\breve{\varepsilon}_{R,it} \\
&= \Delta\hat{\mathbb{V}}_{NT,11}^{(a)} + \Delta\hat{\mathbb{V}}_{NT,12}^{(a)} + \Delta\hat{\mathbb{V}}_{NT,13}^{(a)} + \Delta\hat{\mathbb{V}}_{NT,14}^{(a)}, \text{ say.}
\end{aligned}$$

By Cauchy-Schwarz inequality, we have

$$\left|\Delta\hat{\mathbb{V}}_{NT,14}^{(a)}\right| \leq \frac{8}{T}\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathcal{K}_{i,tt}^{4}\varepsilon_{it}^{6}\right)^{1/2}\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\breve{\varepsilon}_{R,it}^{2}\right)^{1/2}$$
$$= T^{-1}O_p\left(K^2\right)O_p\left(T^{-1/2}\right) = O_p\left(K^2/T^{3/2}\right) = o_p\left(K\right)$$

by (i) in (A.4) and the fact that $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathcal{K}_{i,tt}^{4}\varepsilon_{it}^{6} \leq C_*^4K^4\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}A_{it}^{4}\varepsilon_{it}^{6}\right) = O_p\left(K^4\right)$, where the term in parentheses is $O_p\left(1\right)$ by the Markov inequality and moment conditions on X_{it} and ε_{it}. For $\Delta\hat{\mathbb{V}}_{NT,11}^{(a)}$, we first define $V_{\varepsilon,i} = T^{-1/2}\sum_{t=1}^{T}\dot{Z}_{it}\dot{Z}_{it}'\varepsilon_{it}$. Then we have

$$\begin{aligned}
\left|\Delta\hat{\mathbb{V}}_{NT,11}^{(a)}\right| &= \left|\frac{8}{NT^2}\sum_{i=1}^{N}\sum_{s=1}^{T}\sum_{t=1}^{T}\mathcal{K}_{i,ts}^{2}\varepsilon_{is}^{2}\varepsilon_{it}\bar{\varepsilon}_i\right| \\
&= \left|\frac{8}{NT^2}\sum_{i=1}^{N}\sum_{s=1}^{T}\sum_{t=1}^{T}\operatorname{tr}\left(\breve{Q}_i\dot{Z}_{is}\dot{Z}_{is}'\varepsilon_{is}^{2}\breve{Q}_i\dot{Z}_{it}\dot{Z}_{it}'\varepsilon_{it}\right)\bar{\varepsilon}_i\right| \\
&= \frac{8}{T}\left|\frac{1}{N}\sum_{i=1}^{N}\operatorname{tr}\left(\breve{Q}_iQ_{\varepsilon,i}\breve{Q}_iV_{\varepsilon,i}\right)T^{1/2}\bar{\varepsilon}_i\right| \\
&\leq \frac{8}{TN}\sum_{i=1}^{N}\left\|\breve{Q}_iQ_{i,\varepsilon}\breve{Q}_i\right\|\left\|V_{i,\varepsilon}\right\|\left|T^{1/2}\bar{\varepsilon}_i\right| \\
&\leq \frac{8}{T}\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\breve{Q}_iQ_{\dot{z},i}^{(\varepsilon)}\breve{Q}_i\right\|^2\left|T^{1/2}\bar{\varepsilon}_i\right|^2\right)^{1/2}\left(\frac{1}{N}\sum_{i=1}^{N}\left\|V_{i,\varepsilon}\right\|^2\right)^{1/2} \\
&\equiv 8T^{-1}\left(\Delta\hat{\mathbb{V}}_{NT,111}^{(a)}\right)^{1/2}\left(\Delta\hat{\mathbb{V}}_{NT,112}^{(a)}\right)^{1/2}, \text{ say,}
\end{aligned}$$

where $Q_{\varepsilon,i} = T^{-1}\sum_{s=1}^{T}\dot{Z}_{is}\dot{Z}_{is}'\varepsilon_{is}^2$ and $\breve{Q}_i = Q_{\dot{z},i}^{-1}Q_{w,i}Q_{\dot{z},i}^{-1}$. Note that by Lemma A.3

$$
\begin{aligned}
|\Delta\hat{\mathbb{V}}_{NT,111}^{(a)}| &= \frac{1}{N}\sum_{i=1}^{N}\operatorname{tr}\left(\breve{Q}_i Q_{\dot{z},i}^{(\varepsilon)}\breve{Q}_i^2 Q_{\dot{z},i}^{(\varepsilon)}\breve{Q}_i\right)\left(T^{1/2}\bar{\varepsilon}_i\right)^2 \\
&\leq \max_i \lambda_{\max}^8\left(Q_{\dot{z},i}^{-1}\right)\max_i \lambda_{\max}^4\left(Q_{w,i}\right)\frac{1}{N}\sum_{i=1}^{N}\operatorname{tr}\left(Q_{\dot{z},i}^{(\varepsilon)}Q_{\dot{z},i}^{(\varepsilon)}\right)\left(T^{1/2}\bar{\varepsilon}_i\right)^2 \\
&\leq \max_i \lambda_{\max}^8\left(Q_{\dot{z},i}^{-1}\right)\max_i \lambda_{\max}^4\left(Q_{w,i}\right)\max_i \lambda_{\max}\left(Q_{\dot{z},i}^{(\varepsilon)}\right)\frac{1}{N}\sum_{i=1}^{N}\left(T^{1/2}\bar{\varepsilon}_i\right)^2\frac{1}{T}\sum_{s=1}^{T}\operatorname{tr}\left(\dot{Z}_{is}\dot{Z}_{is}'\varepsilon_{is}^2\right) \\
&\lesssim \frac{1}{N}\sum_{i=1}^{N}\left(T^{1/2}\bar{\varepsilon}_i\right)^2\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\dot{Z}_{is}\right\|^4\sigma_{is}^4\frac{1}{T}\sum_{s=1}^{T}\epsilon_{is}^4\right)^{1/2} \\
&\leq \max_i\left(\frac{1}{T}\sum_{s=1}^{T}\epsilon_{is}^4\right)^{1/2}\max_i\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\dot{Z}_{is}\right\|^4\sigma_{is}^4\right)^{1/2}\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\bar{\varepsilon}_i}{\sqrt{T}}\right)^2 \\
&= O_p(1)\,O_p(K)\,O_p(1) = O_p(K)
\end{aligned}
$$

because of $\max_i\left(T^{-1}\sum_{s=1}^{T}\epsilon_{is}^4\right) = \max_i\left(T^{-1}\sum_{s=1}^{T}E\epsilon_{is}^4\right) + o_p(1)$ and $\max_i\left(T^{-1}\sum_{s=1}^{T}\left\|\dot{Z}_{is}\right\|^4\sigma_{is}^4\right)$ $\leq 4K^2\max_i\left(T^{-1}\sum_{s=1}^{T}A_{is}^2\sigma_{is}^4\right) = 4K^2\max_i\left(T^{-1}\sum_{s=1}^{T}E\left(A_{is}^2\sigma_{is}^4\right)\right)^{1/2} + o_p(1)$, which can be shown as the proof of Lemma A.5 in the online supplementary material to Su, Wang and Jin (2018). Second, $\Delta\hat{\mathbb{V}}_{NT,112}^{(a)} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}\sum_{t=1}^{T}\operatorname{tr}\left(\dot{Z}_{is}'\dot{Z}_{it}\dot{Z}_{it}'\dot{Z}_{is}\sigma_{is}\sigma_{it}\epsilon_{it}\epsilon_{is}\right) = O_p\left(K^2\right)$ by the conditional Markov inequality with the fact

$$
\begin{aligned}
E\left(\Delta\hat{\mathbb{V}}_{NT,112}^{(a)}|\mathbf{X}\right) &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}\sum_{t=1}^{T}\operatorname{tr}\left[\dot{Z}_{is}\dot{Z}_{is}'\dot{Z}_{it}\dot{Z}_{it}'\sigma_{is}\sigma_{it}E\left(\epsilon_{it}\epsilon_{is}\right)\right] \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}\left\|\dot{Z}_{it}\right\|^4\sigma_{it}^2E\left(\epsilon_{it}^2\right) = O_p\left(K^2\right).
\end{aligned}
$$

It follows that $\Delta\hat{\mathbb{V}}_{NT,11}^{(a)} = O_p\left(K^2/T\right) = o_p(K)$. For $\Delta\hat{\mathbb{V}}_{NT,12}^{(a)}$, we have

$$\begin{aligned}
\left|\Delta\hat{\mathbb{V}}_{NT,12}^{(a)}\right| &= \frac{\gamma_{NT}}{NT^2}\sum_{i=1}^{N}\sum_{s=1}^{T}\sum_{t=1}^{T}\mathcal{K}_{i,ts}^2\varepsilon_{is}^2\varepsilon_{it}\breve{g}_{\Delta,it}^{(c)} \\
&= 8\gamma_{NT}\left|\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{s=1}^{T}\sum_{t=1}^{T}\operatorname{tr}\left(\breve{Q}_i\dot{Z}_{is}\dot{Z}_{is}'\varepsilon_{is}^2\breve{Q}_i\dot{Z}_{it}\dot{Z}_{it}'\breve{g}_{\Delta,it}^{(c)}\varepsilon_{it}\right)\right| \\
&\leq 8\gamma_{NT}T^{-1/2}\frac{1}{N}\sum_{i=1}^{N}\left\|\breve{Q}_iQ_{i,\varepsilon}\breve{Q}_i\right\|\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\dot{Z}_{it}\dot{Z}_{it}'\breve{g}_{\Delta,it}^{(c)}\varepsilon_{it}\right\| \\
&\leq 8\gamma_{NT}T^{-1/2}\left[\frac{1}{N}\sum_{i=1}^{N}\operatorname{tr}\left(\breve{Q}_iQ_{\dot{z},i}^{(\varepsilon)}\breve{Q}_i\breve{Q}_iQ_{\dot{z},i}^{(\varepsilon)}\breve{Q}_i\right)\right]^{1/2}\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\dot{Z}_{it}\dot{Z}_{it}'\breve{g}_{\Delta,it}^{(c)}\varepsilon_{it}\right\|^2\right)^{1/2} \\
&= 8\gamma_{NT}T^{-1/2}O_p\left(K^{1/2}\right)O_p(K) = o_p(K).
\end{aligned}$$

Similarly, we can show that $\Delta\hat{\mathbb{V}}_{NT,13}^{(a)} = O_p\left(K^2T^{-1/2}\|\nu_{NT}\|\right) = o_p(K)$. It follows that $\Delta\hat{\mathbb{V}}_{NT,1}^{(a)} = o_p(K)$.

Second, for $\Delta\hat{\mathbb{V}}_{NT,2}^{(a)}$, by Cauchy-Schwarz inequality, we have

$$\begin{aligned}
\Delta\hat{\mathbb{V}}_{NT,2}^{(a)} &= \frac{8}{NT^2}\sum_{i=1}^{N}\sum_{1\leq s\neq t\leq T}\mathcal{K}_{i,ts}^2\breve{\varepsilon}_{R,it}\breve{\varepsilon}_{R,is}\varepsilon_{it}\varepsilon_{is} \\
&\leq \left(\frac{8}{NT^2}\sum_{i=1}^{N}\sum_{1\leq s\neq t\leq T}\mathcal{K}_{i,ts}^4\varepsilon_{it}^2\varepsilon_{is}^2\right)^{1/2}\left(\frac{8}{NT^2}\sum_{i=1}^{N}\sum_{1\leq s\neq t\leq T}\breve{\varepsilon}_{R,it}^2\breve{\varepsilon}_{R,is}^2\right)^{1/2} \\
&\leq \left(\frac{8C_*^4K^4}{NT^2}\sum_{i=1}^{N}\sum_{1\leq s\neq t\leq T}A_{it}^2A_{is}^2\varepsilon_{it}^2\varepsilon_{is}^2\right)^{1/2}\left[\frac{8}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}\breve{\varepsilon}_{R,it}^2\right)^2\right]^{1/2} \\
&= O_p\left(K^2\right)O_p\left(T^{-1}\right) = o_p(K).
\end{aligned}$$

Similarly, we can show $\Delta\hat{\mathbb{V}}_{NT,s}^{(a)} = o_p(K)$ for $s = 3,4,5$ by Cauchy-Schwarz inequality. Hence, $\Delta\hat{\mathbb{V}}_{NT}^{(a)} = o_p(K)$.

For $\Delta\hat{\mathbb{V}}_{NT}^{(b)}$, let $\dot{\epsilon}_{2,it} \equiv \epsilon_{it}^2 - 1$. Then we can write

$$\Delta\hat{\mathbb{V}}_{NT}^{(b)} = \frac{2}{NT^2}\sum_{i=1}^{N}\sum_{1\leq s\neq t\leq T}\mathcal{K}_{i,ts}^2\sigma_{it}^2\sigma_{is}^2\left(2\dot{\epsilon}_{2,it} + \dot{\epsilon}_{2,it}\dot{\epsilon}_{2,is}\right) \equiv \Delta\hat{\mathbb{V}}_{NT}^{(b1)} + \Delta\hat{\mathbb{V}}_{NT}^{(b2)}, \text{ say.}$$

For $\Delta\hat{\mathbb{V}}_{NT}^{(b1)}$, we have $\mathrm{E}(\Delta\hat{\mathbb{V}}_{NT}^{(b1)}|\mathbf{X}) = 0$ and

$$
\begin{aligned}
\mathrm{Var}\left(\Delta\hat{\mathbb{V}}_{NT1}^{(b1)}|\mathbf{X}\right) &= \frac{16}{N^2T^4}\sum_{i=1}^{N}\sum_{1\leq s_1\neq t_1\leq T}\sum_{1\leq s_2\neq t_2\leq T}\mathcal{K}_{i,t_1s_1}^2\sigma_{it_1}^2\sigma_{is_1}^2\mathcal{K}_{i,t_2s_2}^2\sigma_{it_2}^2\sigma_{is_2}^2\mathrm{Cov}\left(\dot{\epsilon}_{2,it_1},\dot{\epsilon}_{2,it_2}\right)\\
&\leq \frac{C}{N^2T^2}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\dot{Z}_{is}\right\|^2\sigma_{is}^2\right)^2\sum_{t=1}^{T}\left\|\dot{Z}_{it}\right\|^4\sigma_{it}^4\mathrm{Var}\left(\dot{\epsilon}_{2,it}\right)\\
&+\frac{8C}{N^2T^4}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\dot{Z}_{is}\right\|^2\sigma_{is}^2\right)^2\sum_{1\leq t<s\leq T}\left\|\dot{Z}_{it}\right\|^2\left\|\dot{Z}_{is}\right\|^2\sigma_{it}^2\sigma_{is}^2\mathrm{Cov}\left(\dot{\epsilon}_{2,it},\dot{\epsilon}_{2,is}\right)\\
&= O_p\left(\frac{K^4}{NT}\right)
\end{aligned}
$$

by following the proof of VJ_2. It follows that $\Delta\hat{\mathbb{V}}_{NT}^{(b1)} = O_p(K^2/\sqrt{NT}) = o_p(K)$. For $\Delta\hat{\mathbb{V}}_{NT}^{(b2)}$, we have

$$
\begin{aligned}
\mathrm{E}(\Delta\hat{\mathbb{V}}_{NT}^{(b2)}|\mathbf{X}) &= \frac{2}{NT^2}\sum_{i=1}^{N}\sum_{1\leq s\neq t\leq T}\tilde{\mathcal{K}}_{i,ts}^2\mathrm{Cov}\left(\epsilon_{it}^2,\epsilon_{is}^2\right) = O_p\left(K^2/T\right) \text{ and}\\
\mathrm{Var}(\Delta\hat{\mathbb{V}}_{NT}^{(b2)}|\mathbf{X}) &= \frac{4}{N^2T^4}\sum_{i=1}^{N}\sum_{1\leq s_1\neq t_1\leq T}\sum_{1\leq s_2\neq t_2\leq T}\tilde{\mathcal{K}}_{i,t_1s_1}^2\tilde{\mathcal{K}}_{i,t_2s_2}^2\mathrm{Cov}\left(\dot{\epsilon}_{2,it_1}\dot{\epsilon}_{2,is_1},\dot{\epsilon}_{2,it_2}\dot{\epsilon}_{2,is_2}\right)\\
&= O_p\left(K^4/(NT^2)\right)
\end{aligned}
$$

by following the proof of Lemma A.1 in Gao (2007, p.193). It follows that $\Delta\hat{\mathbb{V}}_{NT}^{(b2)} = O_p\left(K^2/T\right) + O_p\left(K/(N^{1/2}T)\right) = o_p(K)$. Then we show that $\Delta\hat{\mathbb{V}}_{NT}^{(b)} = o_p(K)$. ■

Proof of Corollary 3.3. Under the global alternative $\mathbb{H}_1$, we have $\breve{\nu}_{NT} = \nu_{\Delta,NT} + \nu_{NT} = o(1) + O_p((NT)^{-1/2}) = o_p(1)$. Then

$$
\begin{aligned}
\Gamma_{NT} &= \frac{1}{NT^2}\sum_{i=1}^{N}\left(\varepsilon_i + \breve{g}_{\Delta,i} - X_i\breve{\nu}_{NT}\right)'\mathcal{K}_i\left(\varepsilon_i + \breve{g}_{\Delta,i} - X_i\breve{\nu}_{NT}\right)\\
&= \frac{1}{NT^2}\sum_{i=1}^{N}\left\{\varepsilon_i'\mathcal{K}_i\varepsilon_i + \breve{g}_{\Delta,i}'\mathcal{K}_i\breve{g}_{\Delta,i} + \breve{\nu}_{NT}'X_i'\mathcal{K}_iX_i\breve{\nu}_{NT} + 2\varepsilon_i'\mathcal{K}_i\breve{g}_{\Delta,i} - 2\breve{g}_{\Delta,i}'\mathcal{K}_iX_i\breve{\nu}_{NT} - 2\varepsilon_i'\mathcal{K}_iX_i\breve{\nu}_{NT}\right\}\\
&= \sum_{l=1}^{6}\Gamma_{NT,l}, \text{ say.}
\end{aligned}
$$

Then we have (i) $\Gamma_{NT,1} = \frac{1}{NT^2}\sum_{i=1}^{N}\left\{\sum_{t=1}^{T}\sum_{s=t}^{T} + \sum_{t=1}^{T}\sum_{s=1,\neq t}^{T}\right\}\varepsilon_{is}\varepsilon_{it}\mathcal{K}_{i,ts} = O_p\left(T^{-1}K\right) + O_p\left(N^{-1/2}T^{-1}K^{1/2}\right)$; (ii) $\Gamma_{NT,1} = \Phi_\Delta + o_p(1)$; (iii) $\Gamma_{NT,3} \leq \|\breve{\nu}_{NT}\|^2 = O_p((NT)^{-1}) + o_p(1)$. Then by Cauchy-Schwarz inequality, we have $|\Gamma_{NT,l}| = o_p(1)$ for $l = 4,5,6$. It follows that $\Gamma_{NT} = \Phi_\Delta + o_p(1)$ and $P(\Gamma_{NT} \geq \Phi_\Delta/2) \to 1$. In addition, we can still show that $\hat{\mathbb{V}}_{NT} =$

$\mathbb{V}_0 + o_p(K)$ for some $\mathbb{V}_0 = O(K)$ and $\hat{\mathbb{B}}_{NT} = O_p\left(N^{1/2}K\right)$. It follows that

$$\begin{aligned}\hat{J}_{NT} &= \frac{N^{1/2}T\Gamma_{NT} - \hat{\mathbb{B}}_{NT}}{\hat{\mathbb{V}}_{NT}^{-1/2}} = \left(\frac{\hat{\mathbb{V}}_{NT}}{\mathbb{V}_0}\right)^{1/2}\left(\frac{N^{1/2}T\Gamma_{NT} - \hat{\mathbb{B}}_{NT}}{\mathbb{V}_0^{-1/2}}\right) \\ &= (1 + o_p(1))\left(\frac{N^{1/2}TO_p(1) + O_p\left(N^{1/2}K\right)}{O\left(K^{1/2}\right)}\right) = O_p\left(TN^{1/2}K^{-1/2}\right).\end{aligned}$$

Consequently, we have $P\left(\hat{J}_{NT} > d_{NT}\right) \to 1$ as $(N,T) \to \infty$ for any $d_{NT} = o\left(TN^{1/2}K^{-1/2}\right)$.■

Proof of Theorem 3.4. Let P^* denote the probability measure induced by the wild bootstrap conditional on the original sample $\mathcal{W}_{NT} \equiv \{(X_{it}, Y_{it}) : i = 1, \ldots, N,\ t = 1, \ldots, T\}$. Let E^* and Var^* denote the expectation and variance w.r.t. P^*. Let $O_{P^*}(\cdot)$ and $o_{P^*}(\cdot)$ denote the probability order under P^*; e.g., $b_{NT} = o_{P^*}(1)$ if for any $\epsilon > 0$, $P^*(\|b_{NT}\| > \epsilon) = o_P(1)$. We will use the fact that $b_{NT} = o_P(1)$ implies that $b_{NT} = o_{P^*}(1)$.

Observing that $Y_{it}^* = X_{it}'\hat{\beta}_{FE} + \hat{\alpha}_i + \varepsilon_{r,it}^*$, the null hypothesis of homogenous and time-invariant coefficients is maintained in the bootstrap world. Given $\mathcal{W}_{NT}$, $\varepsilon_{r,it}^*$ are independent across i and t, and independent of X_{js} for all i, t, j, and s, because the latter objects are fixed in the fixed-design bootstrap world. Let $\mathcal{F}_t^*$ be the σ-field generated by $\left\{\varepsilon_{r,i1}^*, \ldots, \varepsilon_{r,iT}^*\right\}_{i=1}^N$. For each i, $\left\{\varepsilon_{r,it}^*, \mathcal{F}_t^*\right\}$ is an m.d.s. such that $E^*\left(\varepsilon_{r,it}^*|\mathcal{F}_{t-1}^*\right) = \hat{\varepsilon}_{r,it}E(\varrho_{it}) = 0$ and $E^*[\left(\varepsilon_{r,it}^*\right)^2|\mathcal{F}_{t-1}^*]$ $= \hat{\varepsilon}_{r,it}{}^2E\left(\varrho_{it}^2\right) = \hat{\varepsilon}_{r,it}^2$. These observations greatly simplify the proofs in the bootstrap world. Note that $\hat{u}_{it}^* = -X_{it}'\nu_{NT}^* + \alpha_i + \varepsilon_{r,it}^*$ where $\nu_{NT}^* = [\sum_{i=1}^N X_i'M_{\iota_T}X_i]^{-1}\sum_{i=1}^N X_i'M_{\iota_T}\varepsilon_{r,i}^*$ and $\varepsilon_{r,i}^* = \left(\varepsilon_{r,i1}^*, \ldots, \varepsilon_{r,iT}^*\right)'$.

Let Γ_{NT}^*, $\mathbb{B}_{NT}^*$, $\mathbb{V}_{NT}^*$, $\hat{\mathbb{B}}_{NT}^*$, and $\hat{\mathbb{V}}_{NT}^*$ be the bootstrap analogues of Γ_{NT}, $\mathbb{B}_{NT}$, $\mathbb{V}_{NT}$, $\hat{\mathbb{B}}_{NT}$, and $\hat{\mathbb{V}}_{NT}$, respectively. Then

$$\begin{aligned}\Gamma_{NT}^* &= \frac{1}{NT^2}\sum_{i=1}^N\left(\varepsilon_{r,it}^* - X_i\nu_{NT}^*\right)'\mathcal{K}_i\left(\varepsilon_{r,it}^* - X_i\nu_{NT}^*\right) \\ &= \frac{1}{NT^2}\sum_{i=1}^N\varepsilon_{r,i}^{*\prime}\mathcal{K}_i\varepsilon_{r,i}^* - \frac{2}{NT^2}\sum_{i=1}^N\varepsilon_{r,i}^{*\prime}\mathcal{K}_iX_i\nu_{NT}^* + \frac{1}{NT^2}\sum_{i=1}^N\nu_{NT}^{*\prime}X_i'\mathcal{K}_iX_i\nu_{NT}^* \\ &\equiv \Gamma_{NT}^{(*1)} - 2\Gamma_{NT}^{(*2)} + \Gamma_{NT}^{(*3)},\ \text{say.}\end{aligned}$$

We decompose $\hat{J}_{NT}^*$ as follows

$$\hat{J}_{NT}^* = \frac{N^{1/2}T\Gamma_{NT}^* - \hat{\mathbb{B}}_{NT}^*}{\hat{\mathbb{V}}_{NT}^{*1/2}} = \left(J_{NT}^* - \frac{2N^{1/2}T\Gamma_{NT}^{(*2)}}{\mathbb{V}_{NT}^{*1/2}} + \frac{N^{1/2}T\Gamma_{NT}^{(*3)}}{\mathbb{V}_{NT}^{*1/2}} + \frac{\mathbb{B}_{NT}^* - \hat{\mathbb{B}}_{NT}^*}{\mathbb{V}_{NT}^{*1/2}}\right)\frac{\mathbb{V}_{NT}^{*1/2}}{\hat{\mathbb{V}}_{NT}^{*1/2}}$$

In particular, we can show that: (i) $J_{NT}^* = (N^{1/2}T\Gamma_{NT}^{*(1)} - \mathbb{B}_{NT}^*)/\mathbb{V}_{NT}^{*1/2} \xrightarrow{d^*} N(0,1)$, where d^*; (ii) $J_{NT}^{(2)} \equiv N^{1/2}T\Gamma_{NT}^{(*s)}/\mathbb{V}_{NT}^{*1/2} = o_{P^*}(1)$ for $s = 2, 3$; (iii) $\hat{\mathbb{B}}_{NT}^* - \mathbb{B}_{NT}^* = o_{P^*}(K^{1/2})$; (iv) $\hat{\mathbb{V}}_{NT}^*/\mathbb{V}_{NT}^* = 1 + o_{P^*}(1)$.

We only outline the proof of (i) as we can follow the proofs of Theorems 3.2 to show (ii)-(iv). Write $\Gamma_{NT}^{(*1)} = \frac{1}{NT^2}\sum_{i=1}^N \sum_{1\leq t\neq s\leq T} \mathcal{K}_{i,ts}\varepsilon_{r,is}^* \varepsilon_{r,it}^* + \frac{1}{NT^2}\sum_{i=1}^N \sum_{t=1}^T \mathcal{K}_{i,tt}\left(\varepsilon_{r,it}^*\right)^2 \equiv \Gamma_{NT}^{(*1a)} + \Gamma_{NT}^{(*1b)}$, say. Then J_{NT}^* can be further decomposed as follows

$$J_{NT}^* = \frac{N^{1/2}T\Gamma_{NT}^{(*1a)}}{\sqrt{\mathbb{V}_{NT}^*}} + \frac{N^{1/2}T\Gamma_{NT}^{(*1b)} - \mathbb{B}_{NT}^*}{\sqrt{\mathbb{V}_{NT}^*}} \equiv J_{NT}^{(*a)} + J_{NT}^{(*b)}, \text{ say.}$$

We complete the proof by showing that (ia) $J_{NT}^{(*a)} \xrightarrow{d^*} N(0,1)$ and (ib) $J_{NT}^{(*b)} = o_p(1)$. For (ia), analogously to the proof of Proposition B.1, we can show that $J_{NT}^{(*a)} = \sqrt{N}\overline{\mathcal{Z}}_N^*$, $\overline{\mathcal{Z}}_N^* = \frac{1}{N}\sum_{i=1}^N \mathcal{Z}_i^*$ with $\mathcal{Z}_i^* = \frac{2}{T\mathbb{V}_{NT}^{*1/2}}\sum_{1\leq t<s\leq T}\breve{\mathcal{K}}_{i,ts}\varrho_{it}\varrho_{is}$ and $\breve{\mathcal{K}}_{i,ts} \equiv \mathcal{K}_{i,ts}\hat{\varepsilon}_{r,it}\hat{\varepsilon}_{r,is}$. Noting that $\mathcal{Z}_i^*$'s are independent but not identically distributed (inid) across i conditional on $\mathcal{W}_{NT}$, we prove (ia) by the Linderberg-Feller CLT conditional on $\mathcal{W}_{NT}$. It suffices to show that (ia.1) $\bar{\sigma}_N^{*2} = N\text{Var}^*\left(\overline{\mathcal{Z}}_N^*\right) = \text{Var}\left(J_{NT}^{(*a)}|\mathcal{W}_{NT}\right) = 1$; and (ia.2) $E^*\left(\mathcal{Z}_i^4\right) \leq C < \infty$ for all i. For (ia.1), noting that $\{\varrho_{it}\}$ are iid across i and along t, we have

$$\begin{aligned}
\text{Var}^*\left(J_{NT}^{(*a)}\right) &= \frac{4}{NT^2\mathbb{V}_{NT}^*}\text{Var}^*\left(\sum_{i=1}^N \sum_{1\leq t<s\leq T}\mathcal{K}_{i,ts}\hat{\varepsilon}_{r,it}\hat{\varepsilon}_{r,is}\varrho_{it}\varrho_{is}\right) \\
&= \frac{4}{NT^2\mathbb{V}_{NT}^*}\sum_{i=1}^N \sum_{1\leq t_1<s_1\leq T}\sum_{1\leq t_2<s_2\leq T}\breve{\mathcal{K}}_{i,t_1s_1}\breve{\mathcal{K}}_{i,t_2s_2}E^*\left(\varrho_{it_1}\varrho_{it_2}\varrho_{is_1}\varrho_{is_2}\right) \\
&= \frac{4}{NT^2\mathbb{V}_{NT}^*}\sum_{i=1}^N \sum_{1\leq t<s\leq T}\breve{\mathcal{K}}_{i,ts}^2 = 1
\end{aligned}$$

by noting that $\mathbb{V}_{NT}^* = \frac{2}{NT^2}\sum_{i=1}^N \sum_{1\leq t<s\leq T}\mathcal{K}_{i,ts}^2\hat{\varepsilon}_{r,it}^2\hat{\varepsilon}_{r,is}^2$. For (ia.2), note that

$$\begin{aligned}
E^*\left[(\mathcal{Z}_i^*)^4\right] &= \frac{16}{T^4\mathbb{V}_{NT}^{*2}}\sum_{\substack{1\leq t_1<t_2\leq T, 1\leq t_5<t_6\leq T \\ 1\leq t_3<t_4\leq T, 1\leq t_7<t_8\leq T}}\breve{\mathcal{K}}_{i,t_1t_2}\breve{\mathcal{K}}_{i,t_3t_4}\breve{\mathcal{K}}_{i,t_5t_6}\breve{\mathcal{K}}_{i,t_7t_8}E^*\left(\varrho_{it_1}\varrho_{it_2}\varrho_{it_3}\varrho_{it_4}\varrho_{it_5}\varrho_{it_6}\varrho_{it_7}\varrho_{it_8}\right) \\
&\equiv DJ_{i2}^* + DJ_{i3}^* + DJ_{i4}^*, \text{ say,}
\end{aligned}$$

where $DJ_{i2}^*, DJ_{i3}^*, DJ_{i4}^*$ denote the summations of terms with $2, 3, 4$ different time indices in the expectation, respectively. For DJ_{i2}^*, we have $DJ_{i2}^* \asymp \frac{1}{T^4\mathbb{V}_{NT}^{*2}}\sum_{1\leq t<s\leq T}\mathcal{K}_{i,ts}^4\hat{\varepsilon}_{r,it}^4\hat{\varepsilon}_{r,is}^4 E^*\left(\varrho_{it}^4\right)E^*\left(\varrho_{is}^4\right) = O_{P^*}\left(K^2/T\right)$ by noting that $\mathbb{V}_{NT}^* = O_{P^*}(K)$; for DJ_{i4}^*, we have

$$DJ_{i4}^* \asymp \frac{1}{T^4\mathbb{V}_{NT}^2}\sum_{t\neq s\neq l\neq q}\left(\breve{\mathcal{K}}_{i,ts}^2\breve{\mathcal{K}}_{i,lq}^2 + \breve{\mathcal{K}}_{i,ts}\breve{\mathcal{K}}_{i,tl}\breve{\mathcal{K}}_{i,lq}\breve{\mathcal{K}}_{i,qs}\right) \equiv DJ_{i4a}^* + DJ_{i4b}^*, \text{say.}$$

First,

$$
\begin{aligned}
DJ_{i4a}^{*} &= \frac{1}{T^{4}\mathbb{V}_{NT}^{*2}}\left(\textstyle\sum_{1\le t,s\le T}\breve{\mathcal{K}}_{i,ts}^{2}\right)^{2} = \frac{1}{T^{4}\mathbb{V}_{NT}^{*2}}\left(\textstyle\sum_{1\le t,s\le T}\dot{Z}_{it}Q_{\dot{z},i}^{-1}Q_{w,i}Q_{\dot{z},i}^{-1}\dot{Z}_{is}\hat{\varepsilon}_{r,it}\hat{\varepsilon}_{r,is}\right)^{2} \\
&= \frac{1}{\mathbb{V}_{NT}^{*2}}\left[\operatorname{tr}\left(Q_{\dot{z},i}^{(\hat{\varepsilon})}Q_{\dot{z},i}^{-1}Q_{w,i}Q_{\dot{z},i}^{-1}Q_{\dot{z},i}^{(\hat{\varepsilon})}Q_{\dot{z},i}^{-1}Q_{w,i}Q_{\dot{z},i}^{-1}\right)\right]^{2} \\
&\le \frac{1}{\mathbb{V}_{NT}^{*2}}\left[\lambda_{\max}^{2}\left(Q_{\dot{z},i}^{-1}Q_{w,i}Q_{\dot{z},i}^{-1}\right)\lambda_{\max}\left(Q_{\dot{z},i}^{(\hat{\varepsilon})}\right)\operatorname{tr}\left(Q_{\dot{z},i}^{(\hat{\varepsilon})}\right)\right]^{2} \\
&\le \frac{1}{\mathbb{V}_{NT}^{*2}}\left[\lambda_{\max}^{2}\left(Q_{\dot{z},i}^{-1}Q_{w,i}Q_{\dot{z},i}^{-1}\right)\lambda_{\max}^{2}\left(Q_{\dot{z},i}^{(\hat{\varepsilon})}\right)K\right]^{2} \\
&= \frac{1}{O_{P^{*}}\left(K^{2}\right)}O_{p}\left(K^{2}\right) = O_{P^{*}}\left(1\right).
\end{aligned}
$$

where $Q_{\dot{z},i}^{(\hat{\varepsilon})} = T^{-1}\sum_{t=1}^{T}\dot{Z}_{it}\dot{Z}_{it}'\hat{\varepsilon}_{it}^{2}$. Second,

$$
\begin{aligned}
DJ_{i4b}^{*} &= \frac{1}{T^{4}\mathbb{V}_{NT}^{*2}}\sum_{t\ne s\ne l\ne q}\tilde{\mathcal{K}}_{i,ts}\tilde{\mathcal{K}}_{i,tl}\tilde{\mathcal{K}}_{i,lq}\tilde{\mathcal{K}}_{i,qs} \\
&\lesssim \frac{1}{T^{4}\mathbb{V}_{NT}^{*2}}\sum_{t\ne s\ne l\ne q}\hat{\varepsilon}_{r,is}\dot{Z}_{is}'\breve{Q}_{i}\hat{\varepsilon}_{r,it}^{2}\dot{Z}_{it}\dot{Z}_{it}'\breve{Q}_{i}\hat{\varepsilon}_{r,il}^{2}\dot{Z}_{il}\breve{Q}_{i}\hat{\varepsilon}_{r,iq}^{2}\dot{Z}_{iq}\breve{Q}_{i}\dot{Z}_{is}\hat{\varepsilon}_{r,is} \\
&\lesssim \frac{1}{\mathbb{V}_{NT}^{*2}}\operatorname{tr}\left(Q_{\dot{z},i}^{(\hat{\varepsilon})}\breve{Q}_{i}Q_{\dot{z},i}^{(\hat{\varepsilon})}\breve{Q}_{i}Q_{\dot{z},i}^{(\hat{\varepsilon})}\breve{Q}_{i}Q_{\dot{z},i}^{(\hat{\varepsilon})}\breve{Q}_{i}\right) \\
&\le \frac{1}{\mathbb{V}_{NT}^{*2}}\lambda_{\max}^{3}\left(Q_{\dot{z},i}^{(\hat{\varepsilon})}\right)\operatorname{tr}\left(Q_{\dot{z},i}^{(\hat{\varepsilon})}\right)\lambda_{\max}^{4}\left(Q_{\dot{z},i}^{-1}Q_{w,i}Q_{\dot{z},i}^{-1}\right) \\
&\le \frac{1}{\mathbb{V}_{NT}^{*2}}\lambda_{\max}^{3}\left(Q_{\dot{z},i}^{(\hat{\varepsilon})}\right)K\lambda_{\max}\left(Q_{\dot{z},i}^{(\hat{\varepsilon})}\right)\lambda_{\max}^{4}\left(Q_{\dot{z},i}^{-1}Q_{w,i}Q_{\dot{z},i}^{-1}\right) \\
&= O_{P^{*}}\left(K^{-1}\right) < \infty.
\end{aligned}
$$

where $\breve{Q}_{i} = Q_{\dot{z},i}^{-1}Q_{w,i}Q_{\dot{z},i}^{-1}$. It follows that $DJ_{i4}^{*} = O_{P^{*}}(1) + O_{P^{*}}\left(K^{-1}\right) = O_{P^{*}}(1)$. Similarly, we can show that $DJ_{i3}^{*} < C \le \infty$ conditional on $\mathcal{W}_{NT}$.■

Appendix B: Proofs for Lemmas and Sketch Proofs for Section 4

This appendix provides the proofs of technical lemmas which are used in the proofs of the main results in Section 3, gives assumptions and the sketch of proofs for main theorems in Section 4.

C Proofs for lemmas

Proof for Lemma A.1. Let $\mathbf{g}_{(2)} = (g_1, \ldots, g_d)'$. Then $\mathbf{g} = (g_0, \mathbf{g}_{(2)})'$ and

$$\begin{aligned}
\|\mathbf{g}\|_i^2 &= E\left[\frac{1}{T}\sum_{t=1}^{T}(\mathbf{g}(\tau_t)'\tilde{X}_{it})(\tilde{X}_{it}'\mathbf{g}(\tau_t))\right] \\
&= \frac{1}{T}\sum_{t=1}^{T}\left[\mathbf{g}_{(2)}(\tau_t)' E\left(X_{it}X_{it}'\right)\mathbf{g}_{(2)}(\tau_t) + g_0^2(\tau_t)\right] \\
&\asymp \frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{(2)}(\tau_t)'\mathbf{g}_{(2)}(\tau_t) + \frac{1}{T}\sum_{t=1}^{T} g_0^2(\tau_t) \\
&= \sum_{l=1}^{d}\theta_l'\left[\frac{1}{T}\sum_{t=1}^{T}B^K(\tau_t)B^K(\tau_t)'\right]\theta_l + \theta_0'\left[\frac{1}{T}\sum_{t=1}^{T}B_{-1}^K(\tau_t)B_{-1}^K(\tau_t)'\right]\theta_0 \\
&= \sum_{l=1}^{d}\theta_l'\theta_l + \theta_0'\theta_0 + o(1) = \|\theta\|^2 + o(1)
\end{aligned}$$

by Assumption 1(v) and the fact that $T^{-1}\sum_{t=1}^{T} B^K(\tau_t) B^K(\tau_t)' = I_K + o(K/T)$ (see Lemma C.4.(i) in Dong and Linton (2018)). ■

Proof for Lemma A.2. The proofs of (i) and (ii) are analogous to that of Lemma A.2(i)-(ii) in Su, Wang and Jin (2018). The only difference is that we use Cosine functions as basis function. One is readily to modify their proofs to obtain the above claims for our orthonormal basis functions under the conditions stated in Assumption 1. ■

Proof for Lemma A.3. We first prove (i). Recall that $Z_{it} = (B_{-1,t}, B_t \otimes X_{it})'$ and $\dot{Z}_{it} = Z_{it} - \bar{Z}_i$. Write

$$Q_{\dot{z},i} = \frac{1}{T}\sum_{t=1}^{T} Z_{it}'Z_{it} - \bar{Z}_i'\bar{Z}_i \equiv Q_{\dot{z},i}^{(1)} - Q_{\dot{z},i}^{(2)}, \text{ say.} \tag{A.1}$$

Let $\varpi = (\varpi_0', \varpi_1', \ldots, \varpi_d')' = \left(\varpi_0', \varpi^{(2)\prime}\right)'$ with $\varpi_0 \in \mathbb{R}^{K-1}$ and $\varpi_l \in \mathbb{R}^K$ for $l = 1, \ldots, d$, and $\|\varpi\| \leq C \leq \infty$. Let $g_l(\tau, \varpi_l) = \varpi_l' B^K(\tau)$ and $g_0(\tau, \varpi_0) = \varpi_0' B_{-1}^K(\tau)$. Let $\mathbf{g}_\varpi = \left(g_0(\tau, \varpi_0), \mathbf{g}_{\varpi^{(2)}}'\right)'$, where $\mathbf{g}_{\varpi^{(2)}} = (g_1(\tau, \varpi_1), \ldots, g_d(\tau, \varpi_d))'$.

First, we show that $\lambda_{\max}(Q_{\dot{z},i})$ is bounded by some positive number uniformly in i. By Lemmas A.1 and A.3, we have that uniformly in i and ϖ,

$$\varpi' Q_{\dot{z},i}^{(1)} \varpi = \frac{1}{T}\sum_{t=1}^{T}\left[\mathbf{g}_{\varpi}(\tau_t)' \tilde{X}_{it}\right]^2 = \frac{1}{T}\sum_{t=1}^{T} E\left[\mathbf{g}_{\varpi}(\tau_t)' \tilde{X}_{it}\right]^2 (1+o_p(1)) \asymp \|\varpi\|^2 .$$

Then the largest eigenvalue of $Q_{\dot{z},i}^{(1)}$ and thus $Q_{\dot{z},i}$ is bounded above by some positive number $\bar{c}_{\dot{z}}$ uniformly in i with probability $1-o(N^{-1})$.

Second, we prove that $\lambda_{\min}(Q_{\dot{z},i})$ is bounded away from zero uniformly in i. By Lemma A.2, $\varpi' Q_{\dot{z},i}^{(2)} \varpi = [\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{\varpi}(\tau_t)' \tilde{X}_{it}]^2 = [\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{\varpi}(\tau_t)' E\tilde{X}_{it}]^2 (1+o(1))$ uniformly in i and ϖ. By Cauchy-Schwarz inequality, we have $[\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{\varpi}(\tau_t)' E\tilde{X}_{it}]^2 \leq \frac{1}{T}\sum_{t=1}^{T}\left\|E\tilde{X}_{it}\right\|^2 \times \frac{1}{T}\sum_{t=1}^{T}\|\mathbf{g}_{\varpi}(\tau_t)\|^2 \leq C\|\varpi\|^2 < \infty$ uniformly in i and ϖ because of $\frac{1}{T}\sum_{t=1}^{T}\|\mathbf{g}_{\varpi}(\tau_t)\|^2 = \|\varpi\|^2(1+o(1))$ (see the proof of Lemma A.1. It follows that

$$\varpi' Q_{\dot{z},i} \varpi = \frac{1}{T}\sum_{t=1}^{T} E\left\{[\mathbf{g}_{\varpi}(\tau_t)' \tilde{X}_{it}]^2\right\} - \left[\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_{\varpi}(\tau_t)' E\tilde{X}_{it}\right]^2 + o_p(1) \equiv A_{i,\varpi} + o_p(1).$$

We want to show that $A_{i,\varpi} \geq C\|\varpi\|^2$ for some positive constant. Recall that $\mu_i(\tau_t) = EX_{it}$. For any $\tau \in [0,1]$, let $\Omega_i(\tau) \equiv \mathrm{Var}(X_{it}) = \Xi_i(\tau) - \mu_i(\tau)\mu_i(\tau)'$ and $\tilde{\mu}_i(\tau) \equiv E(\tilde{X}_{it}) = \begin{pmatrix} 1 \\ \mu_i(\tau) \end{pmatrix}$, $\tilde{\Xi}_i(\tau) \equiv E(\tilde{X}_{it}\tilde{X}_{it}') = \begin{pmatrix} 1 & \mu_i(\tau)' \\ \mu_i(\tau) & \Xi_i(\tau) \end{pmatrix}$, and $\tilde{\Omega}_i(\tau) \equiv \mathrm{Var}\left(\tilde{X}_{it}\right) = \begin{pmatrix} 0 & \mathbf{0}_{d\times 1} \\ \mathbf{0}_{d\times 1} & \Omega_i(\tau) \end{pmatrix}$. Then we have

$$\begin{aligned} A_{i,\varpi} &= \int_0^1 \mathbf{g}_{\varpi}(\tau)' \tilde{\Xi}_i(\tau)\mathbf{g}_{\varpi}(\tau)\,d\tau - \left\{\int_0^1 \mathbf{g}_{\varpi}(\tau)_i' \tilde{\mu}_i(\tau)\,d\tau\right\}^2 + o(1) \\ &= \int_0^1 \mathbf{g}_{\varpi^{(2)}}'(\tau)\Omega_i(\tau)\mathbf{g}_{\varpi^{(2)}}(\tau)\,d\tau \\ &+ \int_0^1 \left[\mathbf{g}_{\varpi}'(\tau)\tilde{\mu}_i(\tau)\right]^2 d\tau - \left(\int_0^1 \mathbf{g}_{\varpi}'(\tau)\tilde{\mu}_i(\tau)\,d\tau\right)^2 + o(1) \\ &= A_{i,\varpi}^{(1)} + A_{i,\varpi}^{(2)} + o(1) \end{aligned}$$

For the first term, we have

$$A_{i,\varpi}^{(1)} = \int_0^1 \mathbf{g}_{\varpi^{(2)}}'(\tau)\Omega_i(\tau)\mathbf{g}_{\varpi^{(2)}}(\tau)\,d\tau = \varpi' \begin{pmatrix} \mathbf{0}_{(K-1)\times(K-1)} & \mathbf{0}_{(K-1)\times dK} \\ \mathbf{0}_{dK\times(K-1)} & \int_0^1 \left(\Omega_i(\tau)\otimes B(\tau)B(\tau)'\right)d\tau \end{pmatrix} \varpi.$$

Let $\underline{\mu}_i(\tau) = (B(\tau)\otimes\mu_i(\tau))'$ and $\underline{\mu}_i^{(c)}(\tau) = \underline{\mu}_i(\tau) - \int_0^1 \underline{\mu}_i(\tau)\,d\tau$. Define

$$\mathbb{B}_i = \begin{pmatrix} \int_0^1 B_{-1}(\tau)B_{-1}(\tau)'\,d\tau & \int_0^1 B_{-1}(\tau)\underline{\mu}_i^{(c)}(\tau)'\,d\tau \\ \int_0^1 \underline{\mu}_i^{(c)}(\tau)B_{-1}(\tau)'\,d\tau & \int_0^1 \underline{\mu}_i^{(c)}(\tau)\underline{\mu}_i^{(c)}(\tau)'\,d\tau \end{pmatrix}$$

Then for the second term, we have $A_{i,\varpi}^{(2)} = \varpi'\mathbb{B}_i\varpi$. Since $\int_0^1 B_{-1}(\tau) B_{-1}(\tau)' d\tau = I_{K-1}$, it follows that

$$\begin{aligned}
&A_{i,\varpi}^{(1)} + A_{i,\varpi}^{(2)} \\
&= \varpi' \begin{pmatrix} I_{K-1} & \int_0^1 B_{-1}(\tau)\, \underline{\mu}_i^{(c)}(\tau)' d\tau \\ \int_0^1 \underline{\mu}_i^{(c)}(\tau) B_{-1}(\tau)' d\tau & \int_0^1 \underline{\mu}_i^{(c)}(\tau)\, \underline{\mu}_i^{(c)}(\tau)' d\tau + \int_0^1 \left(\Omega_i(\tau) \otimes B(\tau) B(\tau)'\right) d\tau \end{pmatrix} \varpi \\
&= \varpi' D_{1i} \begin{pmatrix} I_{K-1} & \mathbf{0}_{(K-1)\times dK} \\ \mathbf{0}_{dK\times(K-1)} & D_{0i} \end{pmatrix} D_{1i}'\varpi
\end{aligned}$$

where $D_{1i} = \begin{pmatrix} I_{K-1} & \mathbf{0} \\ -\int_0^1 \underline{\mu}_i^{(c)}(\tau) B_{-1}(\tau)' d\tau & I_{Kd} \end{pmatrix}$, $D_{0i} = \int_0^1 \left(\Omega_i(\tau) \otimes B(\tau) B(\tau)'\right) d\tau + \bar{D}_{0i}$, and

$$\bar{D}_{0i} = \int_0^1 \underline{\mu}_i^{(c)}(\tau)\, \underline{\mu}_i^{(c)}(\tau)' d\tau - \int_0^1 \underline{\mu}_i^{(c)}(\tau) B_{-1}(\tau)' d\tau \int_0^1 B_{-1}(\tau)\, \underline{\mu}_i^{(c)}(\tau)' d\tau.$$

Noting that $D_{1i}D_{1i}' = I$, we have $A_{i,\varpi}^{(1)} + A_{i,\varpi}^{(2)} \geq \lambda_{\min}(D_{0i})\, \varpi' D_{1i} D_{1i}' \varpi = \lambda_{\min}(D_{0i}) \|\varpi\|^2$

$$\begin{aligned}
A_{i,\varpi}^{(1)} + A_{i,\varpi}^{(2)} &\geq \lambda_{\min}(D_{0i})\, \varpi' D_{1i} D_{1i}' \varpi = \lambda_{\min}(D_{0i}) \|\varpi\|^2 \\
&\geq \lambda_{\min}\left(\bar{D}_{0i}\right) \|\varpi\|^2 + \lambda_{\min}\left[\int_0^1 \left(\Omega_i(\tau) \otimes B(\tau) B(\tau)'\right) d\tau\right] \|\varpi\|^2
\end{aligned}$$

by Weyl inequality. Noting that

$$\begin{aligned}
\lambda_{\min}\left[\int_0^1 \left(\Omega_i(\tau) \otimes B(\tau) B(\tau)'\right) d\tau\right] &= \inf_{\|C\|=1, C\in\mathbb{R}^{d\times K}} \int_0^1 \operatorname{vec}(C)' \left(\Omega_i(\tau) \otimes B(\tau) B(\tau)'\right) \operatorname{vec}(C)\, d\tau \\
&= \inf_{\|C\|=1} \int_0^1 B(\tau)' C' \Omega_i(\tau) C B(\tau)\, d\tau \\
&\geq \lambda_{\min}(\Omega_i(\tau)) \int_0^1 \operatorname{tr}\left[B(\tau)' C' C B(\tau)\right] d\tau \\
&= \lambda_{\min}(\Omega_i(\tau)) \operatorname{tr}\left[C'C \left(\int_0^1 B(\tau) B(\tau)' d\tau\right)\right] \\
&= \lambda_{\min}(\Omega_i(\tau)) \operatorname{tr}\left(C'C\right) \\
&= \|C\|^2 \lambda_{\min}(\Omega_i(\tau)) = \lambda_{\min}(\Omega_i(\tau)) \geq \min_i \left[\lambda_{\min}(\Omega_i(\tau))\right]
\end{aligned}$$

we are left to show that $\bar{D}_{0i}$ is semi-positive definite (s.p.d.). Define

$$\underline{\mu}_{i,P}^{(c)}(\tau) = \int_0^1 \underline{\mu}_i^{(c)}(\tau) B_{-1}(\tau)' d\tau \left\{\int_0^1 B_{-1}(\tau) B_{-1}(\tau)' d\tau\right\}^{-1} B_{-1}(\tau)$$

Clearly, by the fact that $\int_0^1 B_{-1}(\tau) B_{-1}(\tau)' d\tau = I_{K-1}$, we have

$$\underline{\mu}_{i,P}^{(c)}(\tau) = \int_0^1 \underline{\mu}_i^{(c)}(\tau) B_{-1}(\tau)' d\tau B_{-1}(\tau),$$

$$\int_0^1 \underline{\mu}_{i,P}^{(c)}(\tau) \underline{\mu}_{i,P}^{(c)}(\tau)' d\tau = \int_0^1 \underline{\mu}_i^{(c)}(\tau) B_{-1}(\tau)' d\tau \left(\int_0^1 B_{-1}(\tau) B_{-1}(\tau)' d\tau \right) \int_0^1 B_{-1}(\tau) \underline{\mu}_i^{(c)}(\tau)' d\tau$$
$$= \int_0^1 \underline{\mu}_i^{(c)}(\tau) B_{-1}(\tau)' d\tau \int_0^1 B_{-1}(\tau) \underline{\mu}_i^{(c)}(\tau)' d\tau.$$

Observing that

$$\int_0^1 \underline{\mu}_i^{(c)}(\tau) \underline{\mu}_{i,P}^{(c)}(\tau)' d\tau = \int_0^1 \underline{\mu}_i^{(c)}(\tau) B_{-1}(\tau)' d\tau \int_0^1 B_{-1}(\tau) \underline{\mu}_i^{(c)}(\tau)' d\tau = \int_0^1 \underline{\mu}_{i,P}^{(c)}(\tau) \underline{\mu}_{i,P}^{(c)}(\tau)' d\tau$$

we can write $\bar{D}_{0i}$ as

$$\bar{D}_{0i} = \int_0^1 \left[\underline{\mu}_i^{(c)}(\tau) - \underline{\mu}_{i,P}^{(c)}(\tau) \right] \left[\underline{\mu}_i^{(c)}(\tau) - \underline{\mu}_{i,P}^{(c)}(\tau) \right]' d\tau.$$

Clearly, $\bar{D}_{0i}$ is s.p.d. and $\lambda_{\min}\left(\bar{D}_{0i}\right) \geq 0$.

(ii) The proof of (ii) is much simpler than that of (i). It is omitted here.

(iii)-(iv) The proofs of (iii) and (iv) are analogous to that of (i) and thus are omitted. We can replace X_{it} by $\sigma_{it} X_{it}$, or $\varepsilon_{it} X_{it}$ and apply Assumption 1(vi) in place of Assumption (v). Noting that Var$(\varepsilon_{it} X_{it})$ =Var$(\sigma_{it} X_{it})$. Assumption 1(v) and moment conditions on $\varepsilon_{it} X_{it}$ are suffice to the proof of (v). ■

Proof for Lemma A.4. Since the proofs for (i)-(ii) are similar, we only show (i). Note that $\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T r_{g,it}^2 = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \left(r_{f,it} + X_{it}' r_{\beta,it} \right)^2 \leq \frac{2}{NT} \sum_{i=1}^N \sum_{t=1}^T X_{it}' r_{\beta,it} r_{\beta,it} X_{it} + \frac{2}{NT} \sum_{i=1}^N \sum_{t=1}^T r_{f,it}^2 \leq \sup_{\tau \in [0,1]} r_{f,i}^2(\tau) + \sup_{\tau \in [0,1]} \| r_{\beta,i}(\tau) \|^2 \frac{2}{NT} \sum_{i=1}^N \sum_{t=1}^T \| X_{it} \|^2 = O\left(K^{-2\kappa}\right)$ $+ O_p\left(K^{-2\kappa}\right) O_p(1) = O_p\left(K^{-2\kappa}\right)$ by Assumption 3 in Newey (1997). ■

Proof for Lemma A.5. (i) First, we have

$$\mathbb{V}_{NT} = \frac{2}{NT^2} \sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^T \mathcal{K}_{i,ts}^2 \sigma_{is}^2 \sigma_{it}^2 - \frac{2}{NT^2} \sum_{i=1}^N \sum_{t=1}^T \mathcal{K}_{i,tt}^2 \sigma_{it}^4 \equiv \mathbb{V}_{NT,1} - \mathbb{V}_{NT,2}, \text{ say.}$$

For $\mathbb{V}_{NT,1}$, we have

$$\mathbb{V}_{NT,1} = \frac{2}{NT^2} \sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^T \text{tr}\left(Q_{\dot{z},i}^{-1} Q_{w,i} Q_{\dot{z},i}^{-1} \dot{Z}_{is} \dot{Z}_{is}' \sigma_{is}^2 Q_{\dot{z},i}^{-1} Q_{w,i} Q_{\dot{z},i}^{-1} \dot{Z}_{it}' \dot{Z}_{it}' \sigma_{it}^2 \right)$$
$$= \frac{2}{N} \sum_{i=1}^N \text{tr}\left(Q_{w,i} Q_{\dot{z},i}^{-1} Q_{\dot{z},i}^{(\sigma)} Q_{\dot{z},i}^{-1} Q_{w,i} Q_{\dot{z},i}^{-1} Q_{\dot{z},i}^{(\sigma)} Q_{\dot{z},i}^{-1} \right)$$
$$\leq 2K \max_i \lambda_{\max}^2\left(Q_{w,i}\right) \max_i \lambda_{\max}^4\left(Q_{\dot{z},i}^{-1}\right) \max_i \lambda_{\max}^2\left(Q_{\dot{z},i}^{(\sigma)}\right)$$
$$= 2K \bar{c}_{\dot{z},\sigma}^2 \underline{c}_{\dot{z}}^{-4} \bar{c}_w = O_p(K)$$

by Lemma A.3, the repeatedly use of the rotation property of trace operator and two inequalities: (i) $\text{tr}(A) \leq n\lambda_{\max}(A)$ for any $n \times n$ symmetric positive definite matrix A and (ii) $\lambda_{\max}(BC) \leq \lambda_{\max}(B)\lambda_{\max}(C)$ for any symmetric p.s.d matrices B and C. For $\mathbb{V}_{NT,2}$, we have

$$\begin{aligned}
\mathbb{V}_{NT,2} &= \frac{2}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\text{tr}\left(Q_{\dot{z},i}^{-1}Q_{w,i}Q_{\dot{z},i}^{-1}\dot{Z}_{it}\dot{Z}_{it}'Q_{\dot{z},i}^{-1}Q_{w,i}Q_{\dot{z},i}^{-1}\dot{Z}_{it}\dot{Z}_{it}'\sigma_{it}^4\right) \\
&\leq \max_i \lambda_{\max}^4\left(Q_{\dot{z},i}^{-1}\right)\max_i \lambda_{\max}^2\left(Q_{w,i}\right)\frac{2}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\text{tr}\left(\dot{Z}_{it}\dot{Z}_{it}'\dot{Z}_{it}\dot{Z}_{it}'\sigma_{it}^4\right) \\
&\leq \frac{2\left[\underline{c}_{\dot{z}}^{-4}\bar{c}_w^2 + o_p(1)\right]}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\dot{Z}_{it}\right\|^4\sigma_{it}^4 \\
&\leq \frac{CK^2}{T}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}A_{it}^2\sigma_{it}^4 = O\left(K^2/T\right) = o_p(K).
\end{aligned}$$

where we use the fact that $\left\|\dot{Z}_{it}\right\|^2 \leq 2KA_{it}$ in the last inequality. It follows that $\mathbb{V}_{NT} = O_p(K) + O_p\left(K^2/T\right) = O_p(K)$.

(ii) Note that $\mathcal{K}_{i,tt} = \dot{Z}_{it}'Q_{\dot{z},i}^{-1}Q_{w,i}Q_{\dot{z},i}^{-1}Z_i'M_{\iota_T}\dot{Z}_{it} \leq \lambda_{\max}^2\left(Q_{\dot{z},i}^{-1}\right)\lambda_{\max}\left(Q_{w,i}\right)\left\|\dot{Z}_{it}\right\|^2 \leq \bar{c}_w\underline{c}_{\dot{z}}^{-2}\left\|\dot{Z}_{it}\right\|^2$ uniformly in i and t. Similarly, $\mathcal{K}_{i,tt} \geq \lambda_{\min}^2\left(Q_{\dot{z},i}^{-1}\right)\lambda_{\min}\left(Q_{w,i}\right)\left\|\dot{Z}_{it}\right\|^2 \geq \bar{c}_{\dot{z}}^{-2}\underline{c}_w\left\|\dot{Z}_{it}\right\|^2$ uniformly in i and t. It follows $\mathbb{B}_{NT} \asymp \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\dot{Z}_{it}\right\|^2\sigma_{it}^2 = O_p\left(KN^{1/2}\right)$. ■

D The sketch of proofs for main results in Section 4

In this section, we give some additional assumptions for the tests for stability of heterogeneous coefficients and for *homogeneity* of time-varying coefficients. Since the proofs for Theorems 4.3 and 4.1 are similar to that of Theorem 3.1, we provide the sketch of the proofs.

D.1 Test for the stability of heterogeneous coefficients

To start, we first study the behavior of $\hat{u}_{it}$ under $\mathbb{H}_{s1,\gamma_{NT}}$. By the definition of $\bar{\beta}_{P,i}$, we still have we have $\bar{\beta}_{P,i} = \beta_i$ under $\mathbb{H}_{s1,\gamma_{NT}}$ and

$$\begin{aligned}
\hat{\beta}_i - \beta_i &= \gamma_{NT}\left(X_i'M_{\iota_T}X_i\right)^{-1}X_i'M_{\iota_T}g_{\Delta,i} + \left(X_i'M_{\iota_T}X_i\right)^{-1}X_i'M_{\iota_T}\varepsilon_i \\
&= \gamma_{NT}\bar{\beta}_{\Delta i} + \gamma_{NT}\nu_{\Delta i,T} + \nu_{i,T},
\end{aligned}$$

where $\bar{\beta}_{\Delta i} = \gamma_{NT} \left(E\left(X_i' M_{\iota_T} X_i \right) \right)^{-1} E\left(X_i' M_{\iota_T} g_{\Delta,i} \right)$, $\nu_{\Delta i,T} = \hat{\beta}_{\Delta i,T} - \bar{\beta}_{\Delta i}$ with $\hat{\beta}_{\Delta i,T} = \left(X_i' M_{\iota_T} X_i \right)^{-1} \times X_i' M_{\iota_T} g_{\Delta,i}$ and $\nu_{i,T} = \left(X_i' M_{\iota_T} X_i \right)^{-1} X_i' M_{\iota_T} \varepsilon_i$. Then

$$\hat{u}_{it} = \left(\varepsilon_{it} - X_{it}' \nu_{i,T} \right) + \alpha_i + \gamma_{NT} \breve{g}_{\Delta,it} - \gamma_{NT} X_{it}' \nu_{\Delta i,T} \text{ and} \tag{A.1}$$

$$\hat{u}_i = \vec{\varepsilon}_i + \alpha_i \iota_T + \gamma_{NT} \breve{g}_{\Delta,i} - \gamma_{NT} X_i \nu_{\Delta i,T} \tag{A.2}$$

where $\vec{\varepsilon}_{it} = \varepsilon_{it} - X_{it}' \nu_{i,T}$, $\vec{\varepsilon}_i = \left(\vec{\varepsilon}_{i1}, \ldots, \vec{\varepsilon}_{iT} \right)' = \varepsilon_i - X_i \left(X_i' M_{\iota_T} X_i \right)^{-1} X_i' M_{\iota_T} \varepsilon_i$, $\breve{g}_{\Delta,it} = g_{\Delta,it} - X_{it}' \bar{\beta}_{\Delta i}$, and $\breve{g}_{\Delta,i} = \left(\breve{g}_{\Delta,i1}, \ldots, \breve{g}_{\Delta,iT} \right)'$.

Now we give the sketch of the proof of Theorem 4.1.

The Sketch of proof for Theorem 4.1. We only give the sketch proof for (ii) because (i) can be seen as a special case of (ii) with $\gamma_{NT} = 0$. Using (A.2), we can decompose Γ_{NT} as follows

$$\Gamma_{NT} = \frac{1}{NT^2} \sum_{i=1}^{N} \left(\vec{\varepsilon}_i + \gamma_{NT} \breve{g}_{\Delta,i} - \gamma_{NT} X_i \nu_{\Delta i,T} \right)' \mathcal{K}_i \left(\vec{\varepsilon}_i + \gamma_{NT} \breve{g}_{\Delta,i} - \gamma_{NT} X_i \nu_{\Delta i,T} \right) \equiv \sum_{s=1}^{6} \Gamma_{NT}^{(s)}, \text{ say}$$

where

$$\Gamma_{NT}^{(1)} = \tfrac{1}{NT^2} \textstyle\sum_{i=1}^{N} \vec{\varepsilon}_i' \mathcal{K}_i \vec{\varepsilon}_i, \quad \Gamma_{NT}^{(2)} = \tfrac{\gamma_{NT}^2}{NT^2} \textstyle\sum_{i=1}^{N} \breve{g}_{\Delta,i}' \mathcal{K}_i \breve{g}_{\Delta,i}, \quad \Gamma_{NT}^{(3)} = \tfrac{\gamma_{NT}^2}{NT^2} \textstyle\sum_{i=1}^{N} \nu_{\Delta i,T}' X_i' \mathcal{K}_i X_i \nu_{\Delta i,T},$$

$$\Gamma_{NT}^{(4)} = \tfrac{2\gamma_{NT}}{NT^2} \textstyle\sum_{i=1}^{N} \vec{\varepsilon}_i' \mathcal{K}_i \breve{g}_{\Delta,i}, \quad \Gamma_{NT}^{(5)} = \tfrac{-2\gamma_{NT}}{NT^2} \textstyle\sum_{i=1}^{N} \vec{\varepsilon}_i' \mathcal{K}_i X_i \nu_{\Delta i,T}, \quad \Gamma_{NT}^{(6)} = \tfrac{-2\gamma_{NT}^2}{NT^2} \textstyle\sum_{i=1}^{N} \breve{g}_{\Delta,i}' \mathcal{K}_i X_i \nu_{\Delta i,T}.$$

With the decomposition, we have

$$\hat{J}_{NT}^{\dagger} = \frac{N^{1/2} T \Gamma_{NT} - \hat{\mathbb{B}}_{NT}^{\dagger}}{\hat{\mathbb{V}}_{NT}^{\dagger 1/2}} = \left(J_{NT}^{\dagger} + \sum_{s=2}^{6} \frac{N^{1/2} T \Gamma_{NT}^{(s)}}{\mathbb{V}_{NT}^{\dagger 1/2}} + \frac{\mathbb{B}_{NT}^{\dagger} - \hat{\mathbb{B}}_{NT}^{\dagger}}{\mathbb{V}_{NT}^{\dagger 1/2}} \right) \frac{\mathbb{V}_{NT}^{\dagger 1/2}}{\hat{\mathbb{V}}_{NT}^{\dagger 1/2}}$$

where $J_{NT}^{\dagger} = (N^{1/2} T \Gamma_{NT}^{(1)} - \mathbb{B}_{NT}^{\dagger}) / \mathbb{V}_{NT}^{\dagger 1/2}$. We can complete the proof by showing that (i) $J_{NT}^{\dagger} \xrightarrow{d} N(0,1)$; (ii) $N^{1/2} T \Gamma_{NT}^{(2)} / \mathbb{V}_{NT}^{\dagger 1/2} = \Phi_{\Delta} + o_p(1)$, where $\Phi_{\Delta} = \text{plim}_{(N,T) \to \infty} \Phi_{\Delta,NT}$ with $\Phi_{\Delta,NT} = \frac{1}{NT^2} \sum_{i=1}^{N} \breve{g}_{\Delta,it}^2 w_{it}$; (iii) $N^{1/2} T \Gamma_{NT}^{(s)} / \mathbb{V}_{NT}^{\dagger 1/2} = o_p(1)$ for $s = 3, \ldots, 6$; (iv) $\hat{\mathbb{B}}_{NT}^{\dagger} - \mathbb{B}_{NT}^{\dagger} = o_p(K^{1/2})$; (v) $\hat{\mathbb{V}}_{NT}^{\dagger} / \mathbb{V}_{NT}^{\dagger} = 1 + o_p(1)$.

First, it is straightforward to show (i), (ii), (iv) and (v) by modifying the corresponding proofs for Theorem 3.1. For (iii), following the proof of (iii) in Theorem 3.1, we can show that $\Gamma_{NT}^{(3)} = o_p\left(\gamma_{NT}^2 \right) = o_p(K^{1/2}/(N^{1/2}T))$, $\Gamma_{NT}^{(4)} = \gamma_{NT} O_p(\sqrt{K/(NT)}) = o_p(K^{1/2}/(N^{1/2}T))$, $\Gamma_{NT}^{(5)} = \gamma_{NT} O_p(\sqrt{K/(NT)}) o_p(1) = o_p(K^{1/2}/(N^{1/2}T))$, and $\Gamma_{NT}^{(6)} = o_p\left(\gamma_{NT}^2 \right) = o_p(K^{1/2}/(N^{1/2}T))$. ■

Proof of Corollary 4.2. We can follow the proof of Theorem 3.2 to show the corollary. The details are omitted here. ■

D.2 Test the homogeneity of time-varying coefficients

We first study the behavior of $\hat{u}_{it}$ and $\hat{g}_{it}$ under the local alternative. There exist $\Pi_{\beta}^0 \in \mathbb{R}^{d \times L}$ and $\Pi_f^0 \in \mathbb{R}^{L-1}$ such that $\beta_0(\cdot) \approx \Pi_{\beta}^0 B^L(\cdot)$ and $f_0(\cdot) \approx \Pi_f^{0\prime} B_{-1}^L(\cdot)$. Let $g_{it} = g_{0,it} + \gamma_{NT} g_{\Delta,it}$, where $g_{0,it} = X_{it}'\beta_0(\tau_t) + f_0(\tau_t)$. Given $Z_{it}^L = (B_{-1t}^{L\prime}, (X_{it} \otimes B_t^L)')'$, denote $r_{g_0,it} = g_{0,it} - Z_{it}^{L\prime}\Pi^0$, where $\Pi^0 = \left(\Pi_f^{0\prime}, \text{vec}(\Pi_\beta^{0\prime})\right)'$. Let $S_{\dot{Z}\dot{Z}} = \sum_{i=1}^N \dot{Z}_i^{L\prime}\dot{Z}_i^L$, $\hat{\Pi}_{\Delta,NT} = S_{\dot{Z}\dot{Z}}^{-1}\sum_{i=1}^N \dot{Z}_i^{L\prime} g_{\Delta,i}$, $\bar{\Pi}_\Delta = \left[E\left(S_{\dot{Z}\dot{Z}}\right)\right]^{-1}\sum_{i=1}^N E\left(\dot{Z}_i^{L\prime} g_{\Delta,i}\right)$, $R_{g_0,i} \equiv (r_{g_0,i1}, \ldots, r_{g_0,iT})'$ and $g_{\Delta,i} \equiv (g_{\Delta,i1}, \ldots, g_{\Delta,iT})'$. Then we have

$$\begin{aligned}\hat{\Pi}_{FE} - \Pi^0 &= S_{\dot{Z}\dot{Z}}^{-1}\sum_{i=1}^N \dot{Z}_i^{L\prime} R_{g_0,i} + \gamma_{NT}\bar{\Pi}_\Delta + \gamma_{NT}\left[\hat{\Pi}_{\Delta,NT} - \bar{\Pi}_\Delta\right] + S_{\dot{Z}\dot{Z}}^{-1}\sum_{i=1}^N \dot{Z}_i^{L\prime}\varepsilon_i \\ &\equiv R_{g_0,NT} + \gamma_{NT}\bar{\Pi}_\Delta + \gamma_{NT}\nu_{\Pi_\Delta,NT} + \nu_{L,NT},\end{aligned}$$

where $R_{g_0,NT} = S_{\dot{Z}\dot{Z}}^{-1}\sum_{i=1}^N \dot{Z}_i^{L\prime} R_{g_0,i}$, $\nu_{\Pi_\Delta,NT} = \hat{\Pi}_{\Delta,NT} - \bar{\Pi}_\Delta$ and $\nu_{L,NT} = S_{\dot{Z}\dot{Z}}^{-1}\sum_{i=1}^N \dot{Z}_i^{L\prime}\varepsilon_i$. Let $\breve{g}_{\Delta,it} = g_{\Delta,it} - Z_{it}^{L\prime}\bar{\Pi}_\Delta$ and $\breve{\nu}_{L,NT} = \gamma_{NT}\nu_{\Pi_\Delta,NT} + \nu_{L,NT}$. We can write

$$\begin{aligned}g_{it} - \hat{g}_{it} &= (g_{0,it} + \gamma_{NT} g_{\Delta,it}) - Z_{it}^{L\prime}\left(\Pi^0 + R_{g_0,NT} + \gamma_{NT}\bar{\Pi}_\Delta + \breve{\nu}_{L,NT}\right) \\ &= \left(g_{0,it} - Z_{it}^{L\prime}\Pi^0\right) + \gamma_{NT}\left(g_{\Delta,it} - Z_{it}^{L\prime}\bar{\Pi}_\Delta\right) - Z_{it}^{L\prime} R_{g_0,NT} - Z_{it}^{L\prime}\breve{\nu}_{L,NT} \\ &= \left(r_{g_0,it} - Z_{it}^{L\prime} R_{g_0,NT}\right) + \gamma_{NT}\breve{g}_{\Delta,it} - Z_{it}^{L\prime}\breve{\nu}_{L,NT} \\ &= \breve{r}_{g_0,it} + \gamma_{NT}\breve{g}_{\Delta,it} - Z_{it}^{L\prime}\breve{\nu}_{L,NT}\end{aligned}$$

where $\breve{r}_{g_0,it} = r_{g_0,it} - Z_{it}^{L\prime} R_{g_0,NT}$. Let $\breve{R}_{g_0,i} = (\breve{r}_{g_0,i1}, \ldots, \breve{r}_{g_0,iT})'$ and $\breve{g}_{\Delta,i} = (\breve{g}_{\Delta,i1}, \ldots, \breve{g}_{\Delta,iT})'$. Then we have

$$\hat{u}_{it} = \varepsilon_{it} + \alpha_i + \breve{r}_{g_0,it} + \gamma_{NT}\breve{g}_{\Delta,it} - Z_{it}^{L\prime}\breve{\nu}_{L,NT} \text{ and} \tag{A.3}$$

$$\hat{u}_i = \varepsilon_i + \alpha_i \iota_T + \gamma_{NT}\breve{g}_{\Delta,i} + \breve{R}_{g_0,i} - Z_i^L\breve{\nu}_{L,NT}. \tag{A.4}$$

To establish the asymptotic distribution of $\hat{J}_{NT}^\ddagger$, we need the following assumptions.

Assumption 3*. (i) $f(\cdot)$ and $\beta_{0,l}(\cdot)$ for $l = 1, \ldots, d$ are all continuously differentiable up to κ-th order for some $\kappa > 0$; (ii) For each i, $\Delta_{\beta,il}(\cdot)$ for $l = 1, \ldots, d$, and $\Delta_{f,i}(\cdot)$ are all continuously differentiable up to κ-th order for some $\kappa > 0$.

Assumption 4.** As $(N,T) \to \infty$, $\Phi_\Delta = \text{plim}_{(N,T)\to\infty}\Phi_{\Delta,NT} > 0$ under $\mathbb{H}_{1h,\gamma_{NT}}$.

Assumption 5. As $(N,T) \to \infty$, $L \to \infty$, $L^2/T \to 0$, and $K/L \to 0$.

Now we give the sketch for the proof of Theorem 4.3.

Sketch of Proof for Theorem 4.3. We only give the sketch proof for (ii) since (i) can be seen as special case of (ii) with $\gamma_{NT} = 0$. Using (A.4) and $\Gamma_{NT} = \frac{1}{NT^2}\sum_{i=1}^N \hat{u}_i'\mathcal{K}_i\hat{u}_i$, we have $\Gamma_{NT} \equiv \sum_{s=1}^{10}\Gamma_{NT}^{(s)}$, where $\Gamma_{NT}^{(1)} \equiv \frac{1}{NT^2}\sum_{i=1}^N \varepsilon_i'\mathcal{K}_i\varepsilon_i$, $\Gamma_{NT}^{(2)} \equiv \frac{\gamma_{NT}^2}{NT^2}\sum_{i=1}^N \breve{g}_{\Delta,i}'\mathcal{K}_i\breve{g}_{\Delta,i}$, $\Gamma_{NT}^{(3)} \equiv \frac{1}{NT^2}\sum_{i=1}^N \breve{R}_{g_0,i}'\mathcal{K}_i\breve{R}_{g_0,i}$, $\Gamma_{NT}^{(4)} \equiv \frac{1}{NT^2}\sum_{i=1}^N \breve{\nu}_{L,NT}' Z_i^{L\prime}\mathcal{K}_i Z_i^L\breve{\nu}_{L,NT}$, $\Gamma_{NT}^{(5)} \equiv \frac{2\gamma_{NT}}{NT^2}\sum_{i=1}^N \varepsilon_i'\mathcal{K}_i\breve{g}_{\Delta,i}$, $\Gamma_{NT}^{(6)} \equiv \frac{2}{NT^2}\sum_{i=1}^N \varepsilon_i'\mathcal{K}_i\breve{R}_{g_0,i}$, $\Gamma_{NT}^{(7)} \equiv \frac{-2}{NT^2}\sum_{i=1}^N \varepsilon_i\mathcal{K}_i Z_i^L\breve{\nu}_{L,NT}$, $\Gamma_{NT}^{(8)} \equiv \frac{2\gamma_{NT}}{NT^2}\sum_{i=1}^N \breve{g}_{\Delta,i}\mathcal{K}_i\breve{R}_{g_0,i}$, $\Gamma_{NT}^{(9)} \equiv \frac{-2\gamma_{NT}}{NT^2}\sum_{i=1}^N \breve{g}_{\Delta,i}\mathcal{K}_i Z_i^L\breve{\nu}_{L,NT}$, and $\Gamma_{NT}^{(10)} \equiv \frac{-2}{NT^2}\sum_{i=1}^N \breve{R}_{g_0,i}\mathcal{K}_i Z_i^L\breve{\nu}_{L,NT}$. Then $\hat{J}_{NT}^\ddagger$ can be decomposed as follows

$$\hat{J}_{NT}^\ddagger = \frac{N^{1/2}T\Gamma_{NT} - \hat{\mathbb{B}}_{NT}^\ddagger}{\hat{\mathbb{V}}_{NT}^{\ddagger 1/2}} = \left(J_{NT}^\ddagger + \sum_{s=2}^{10}\frac{N^{1/2}T\Gamma_{NT}^{(s)}}{\mathbb{V}_{NT}^{\ddagger 1/2}} + \frac{\mathbb{B}_{NT}^\ddagger - \hat{\mathbb{B}}_{NT}^\ddagger}{\mathbb{V}_{NT}^{\ddagger 1/2}}\right)\frac{\mathbb{V}_{NT}^{\ddagger 1/2}}{\hat{\mathbb{V}}_{NT}^{\ddagger 1/2}}.$$

We complete the proof by showing that: (i) $J_{NT}^{\ddagger}=(N^{1/2}T\Gamma_{NT}^{(1)}-\mathbb{B}_{NT}^{\ddagger})/\mathbb{V}_{NT}^{\ddagger 1/2}\overset{d}{\to}N(0,1)$; (ii) $J_{NT}^{(2)}\equiv N^{1/2}T\Gamma_{NT}^{(2)}/\mathbb{V}_{NT}^{\ddagger 1/2}=\Phi_{\Delta}+o_p(1)$, where $\Phi_{\Delta}=\text{plim}_{(N,T)\to\infty}\Phi_{\Delta,NT}$ with $\Phi_{\Delta,NT}=\frac{1}{NT^2}\sum_{i=1}^{N}\breve{g}_{\Delta,it}^2 w_{it}$; (iii) $J_{NT}^{(s)}\equiv N^{1/2}T\Gamma_{NT}^{(s)}/\mathbb{V}_{NT}^{\ddagger 1/2}=o_p(1)$ for $s=3,\ldots,10$; (iv) $\hat{\mathbb{B}}_{NT}^{\ddagger}-\mathbb{B}_{NT}^{\ddagger}=o_p(K^{1/2})$; (v) $\hat{\mathbb{V}}_{NT}^{\ddagger}/\mathbb{V}_{NT}^{\ddagger}=1+o_p(1)$.

First, we can show that (i), (ii), (iv) and (v) in the proof of Theorem 3.1. Second, we can follow the proofs of (iii) for Theorem 3.1 to show that $\Gamma_{NT}^{(3)}=O_p\left(L^{-2\gamma}\right)=o_p(\mathbb{V}_{NT}^{\ddagger 1/2}/(N^{1/2}T))$, $\Gamma_{NT}^{(4)}=o_p\left(\gamma_{NT}^2\right)+O_p\left(L/(NT)\right)=o_p(\mathbb{V}_{NT}^{\ddagger 1/2}/(N^{1/2}T))$, $\Gamma_{NT}^{(5)}=O_p(\gamma_{NT}\sqrt{K/(NT)})=o_p(\mathbb{V}_{NT}^{\ddagger 1/2}/(N^{1/2}T))$, $\Gamma_{NT}^{(6)}=O_p(L^{-\gamma}\sqrt{K/(NT)})=o_p(\mathbb{V}_{NT}^{\ddagger 1/2}/(N^{1/2}T))$, $\Gamma_{NT}^{(7)}=O_p(\sqrt{K/(NT)})\left[o_p(\gamma_{NT})+O_p(\sqrt{L/(NT)})\right]=o_p(\mathbb{V}_{NT}^{\ddagger 1/2}/(N^{1/2}T))$, $\Gamma_{NT}^{(8)}=o_p\left(\gamma_{NT}L^{-\gamma}\right)=o_p(\mathbb{V}_{NT}^{\ddagger 1/2}/(N^{1/2}T))$, $\Gamma_{NT}^{(9)}=o_p\left(\gamma_{NT}^2\right)+O_p(\gamma_{NT}\sqrt{L/(NT)})=o_p(\mathbb{V}_{NT}^{\ddagger 1/2}/(N^{1/2}T))$, $\Gamma_{NT}^{(10)}=O_p\left(L^{-\gamma}\right)\left[o_p(\gamma_{NT})+O_p(\sqrt{L/(NT)})=o_p(\mathbb{V}_{NT}^{\ddagger 1/2}/(N^{1/2}T))\right.$. ■

Proof for Corollary 4.4. We can follow the proof of Theorem 3.2 to show the corollary. The details are omitted here. ■

E Additional simulation results

In this section, we present the testing results for the two tests discussed in Section 4.

First, we test the stability of heterogeneous coefficients and intercepts for DGPs 1-7. DGPs 1 and 3 are for size study, and other 5 DGPs are for power comparison. Under the null hypothesis $\mathbb{H}_{s,0}$, we use the simple OLS to estimate the heterogenous slopes and intercepts. In the construction of testing statistic, we consider the cosine functions as basis and the same numbers of sieve terms K_1, K_2, K_3 and K_{cv} as in Section 5. We also report the bootstrap p-value, where the null hypothesis of constant slopes and intercepts are imposed in the bootstrap world. Different combinations of sample sizes are used: $T=25,50,100$ and $N=25,50$. For each combination of sample sizes, the number of replications is 500 times. In bootstrap, we consider 400 resamples for size studies and 300 resamples for power comparisons. Table 3 reports the testing results for the stability test.

Second, we test the homogeneity of TVCs in DGPs 1-5. DGPs 1-2 are for size study and DGPs 3-5 are for power comparison. Although DGPs 6-7 have homogeneous coefficients, we do not report the testing results because their coefficient functions are not continuous. Under the null $\mathbb{H}_{h,0}$, we also adopt the cosine functions as basis functions in the estimation of homogeneous time-varying coefficients. The numbers of basis functions in the sieve approximation of $\beta(\cdot)$ and $f(\cdot)$ are both $L=\lfloor 2(NT)^{1/5}\rfloor$. In the construction of testing statistic, we consider the same numbers of sieve terms K_1, K_2, K_3 and K_{cv} as in Section 5. We also report the bootstrap p-value, where the null hypothesis of common TVCs are imposed in the bootstrap world. Different combinations of sample sizes are used: $T=25,50,100$ and $N=25,50$. For each combination of sample sizes, the number of replications is 500 times. In bootstrap, we consider 400 resamples for size studies and 300 resamples for power comparisons. The testing results are reported in Table 4.

Table 3: Simulation results for stability test

			K_1			K_2			K_3			K_{cv}		
DGP	T	N	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
1	25	25	0.014	0.080	0.150	0.140	0.056	0.098	0.020	0.090	0.154	0.014	0.084	0.154
		50	0.008	0.0602	0.146	0.004	0.064	0.132	0.018	0.072	0.138	0.008	0.062	0.146
	50	25	0.018	0.050	0.108	0.012	0.062	0.122	0.010	0.060	0.122	0.020	0.056	0.114
		50	0.006	0.052	0.118	0.010	0.074	0.154	0.016	0.084	0.140	0.006	0.052	0.118
	100	25	0.006	0.056	0.134	0.010	0.060	0.150	0.016	0.080	0.128	0.020	0.066	0.108
		50	0.022	0.076	0.134	0.014	0.068	0.130	0.016	0.058	0.126	0.014	0.060	0.114
2	25	25	0.884	0.972	0.988	0.308	0.588	0.748	0.064	0.196	0.348	0.884	0.972	0.988
		50	0.968	0.996	1.000	0.532	0.776	0.880	0.096	0.296	0.456	0.968	0.996	1.000
	50	25	1.000	1.000	1.000	0.992	1.000	1.000	0.932	0.984	0.996	1.000	1.000	1.000
		50	1.000	1.000	1.000	1.000	1.000	1.000	0.992	0.996	0.996	1.000	1.000	1.000
	100	25	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	25	25	0.020	0.068	0.120	0.016	0.960	0.172	0.012	0.088	0.136	0.020	0.072	0.124
		50	0.012	0.068	0.0128	0.012	0.072	0.152	0.012	0.048	0.108	0.012	0.068	0.128
	50	25	0.008	0.060	0.140	0.028	0.080	0.120	0.020	0.092	0.148	0.012	0.068	0.152
		50	0.004	0.048	0.112	0.012	0.052	0.148	0.008	0.064	0.116	0.004	0.048	0.112
	100	25	0.004	0.064	0.136	0.000	0.032	0.080	0.000	0.020	0.104	0.012	0.048	0.108
		50	0.008	0.052	0.092	0.020	0.056	0.120	0.020	0.092	0.120	0.012	0.056	0.108
4	25	25	0.876	0.956	0.988	0.496	0.716	0.840	0.104	0.252	0.404	0.884	0.964	0.992
		50	0.996	1.000	1.000	0.780	0.948	0.972	0.232	0.464	0.652	0.996	1.000	1.000
	50	25	1.000	1.000	1.000	0.992	1.000	1.000	0.984	1.000	1.000	1.000	1.000	1.000
		50	1.000	1.000	1.000	1.000	1.000	1.000	0.996	1.000	1.000	1.000	1.000	1.000
	100	25	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
5	25	25	1.000	1.000	1.000	0.860	0.944	0.968	0.232	0.488	0.632	1.000	1.000	1.000
		50	1.000	1.000	1.000	0.964	0.996	0.996	0.364	0.664	0.788	1.000	1.000	1.000
	50	25	1.000	1.000	1.000	1.000	1.000	1.000	0.996	0.996	1.000	1.000	1.000	1.000
		50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	100	25	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
6	25	25	1.000	1.000	1.000	1.000	1.000	1.000	0.804	0.932	0.960	1.000	1.000	1.000
		50	1.000	1.000	1.000	1.000	1.000	1.000	0.928	0.988	0.996	1.000	1.000	1.000
	50	25	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	100	25	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
7	25	25	1.000	1.000	1.000	1.000	1.000	1.000	0.768	0.892	0.960	1.000	1.000	1.000
		50	1.000	1.000	1.000	1.000	1.000	1.000	0.912	0.964	0.988	1.000	1.000	1.000
	50	25	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	100	25	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Note: 1. 1. $K_C = \left\lfloor CT^{1/6} \right\rfloor$, $C = 1, 2, 3$, K_{cv} refers to the number by LOOCV;

2. DGPs 1 and 3 are for size study and all the other DGPs are for power comparison.

Table 4: Simulation results for homogeneity test

			K_1			K_2			K_3			K_{cv}		
DGP	T	N	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
1	25	25	0.000	0.026	0.840	0.002	0.034	0.072	0.008	0.040	0.092	0.000	0.260	0.840
		50	0.008	0.062	0.114	0.010	0.056	0.102	0.006	0.0520	0.122	0.008	0.062	0.114
	50	25	0.006	0.042	0.112	0.008	0.060	0.132	0.012	0.042	0.102	0.008	0.042	0.112
		50	0.009	0.052	0.110	0.012	0.050	0.110	0.010	0.048	0.100	0.008	0.052	0.110
	100	25	0.012	0.048	0.118	0.010	0.048	0.136	0.012	0.058	0.116	0.010	0.058	0.124
		50	0.008	0.034	0.090	0.002	0.040	0.082	0.008	0.042	0.078	0.006	0.040	0.092
2	25	25	0.000	0.030	0.082	0.002	0.036	0.072	0.008	0.038	0.094	0.000	0.030	0.082
		50	0.010	0.062	0.106	0.010	0.054	0.098	0.006	0.050	0.126	0.010	0.062	0.106
	50	25	0.006	0.048	0.116	0.010	0.052	0.128	0.012	0.044	0.102	0.008	0.048	0.116
		50	0.008	0.052	0.112	0.010	0.048	0.110	0.010	0.054	0.096	0.008	0.052	0.112
	100	25	0.010	0.048	0.118	0.010	0.048	0.132	0.012	0.050	0.112	0.010	0.056	0.124
		50	0.060	0.036	0.092	0.004	0.042	0.084	0.006	0.040	0.078	0.006	0.038	0.090
3	25	25	0.108	0.316	0.444	0.076	0.216	0.324	0.028	0.116	0.240	0.108	0.316	0.444
		50	0.172	0.430	0.620	0.088	0.256	0.424	0.048	0.168	0.276	0.172	0.432	0.620
	50	25	0.492	0.728	0.856	0.320	0.576	0.732	0.236	0.484	0.636	0.492	0.728	0.856
		50	0.764	0.932	0.964	0.604	0.848	0.932	0.456	0.720	0.868	0.764	0.932	0.964
	100	25	0.872	0.960	0.988	0.824	0.940	0.976	0.752	0.912	0.960	0.892	0.980	0.988
		50	1.000	1.000	1.000	0.984	1.000	1.000	0.964	1.000	1.000	1.000	1.000	1.000
4	25	25	0.272	0.556	0.692	0.132	0.300	0.444	0.048	0.176	0.268	0.272	0.556	0.692
		50	0.584	0.836	0.932	0.336	0.568	0.752	0.104	0.316	0.472	0.584	0.836	0.932
	50	25	0.900	0.976	0.992	0.808	0.952	0.984	0.664	0.884	0.940	0.900	0.976	0.992
		50	0.998	1.000	1.000	0.980	1.000	1.000	0.940	0.988	1.000	0.996	1.000	1.000
	100	25	0.992	0.996	1.000	0.988	0.996	1.000	0.984	0.996	1.000	0.996	1.000	1.000
		50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
5	25	25	1.000	1.000	1.000	0.996	1.000	1.000	0.912	0.972	0.984	1.000	1.000	1.000
		50	1.000	1.000	1.000	1.000	1.000	1.000	0.988	0.996	1.000	1.000	1.000	1.000
	50	25	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	100	25	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Note: 1. $K_C = \lfloor CT^{1/6} \rfloor$, $C = 1, 2, 3$, K_{cv}refers to the number chose by LOOCV;

2. DGP 1-2 are for size study and DGPs 3-5 are for power comparison.