Econometric Theory, **0**, 2023, 1–48.
doi:10.1017/S026646662300018X

SPECIFICATION TESTS FOR TIME-VARYING COEFFICIENT PANEL DATA MODELS

Alev Atak
Middle East Technical University

Thomas Tao Yang
Australian National University

Yonghui Zhang
Renmin University of China

Qiankun Zhou
Louisiana State University

This paper provides nonparametric specification tests for the commonly used homogeneous and stable coefficients structures in panel data models. We first obtain the augmented residuals by estimating the model under the null hypothesis and then run auxiliary time series regressions of augmented residuals on covariates with time-varying coefficients (TVCs) via sieve methods. The test statistic is then constructed by averaging the squared fitted values, which are close to zero under the null and deviate from zero under the alternatives. We show that the test statistic, after being appropriately standardized, is asymptotically normal under the null and under a sequence of Pitman local alternatives. A bootstrap procedure is proposed to improve the finite sample performance of our test. In addition, we extend the procedure to test other structures, such as the homogeneity of TVCs or the stability of heterogeneous coefficients. The joint test is extended to panel models with two-way fixed effects. Monte Carlo simulations indicate that our tests perform reasonably well in finite samples. We apply the tests to re-examine the environmental Kuznets curve in the United States, and find that the model with homogenous TVCs is more appropriate for this application.

1. INTRODUCTION

A panel dataset follows a given sample of entities over time, and it possesses several advantages over a cross-sectional or time series dataset (see, e.g., Hsiao,

We are deeply indebted to the Editor (Peter Phillips), the Co-Editor (Guido Kuersteiner), and three referees for their many constructive comments, which improve the paper substantially. We are grateful to Jiti Gao and participants of the CES 2019 North American conference at Kansas University for helpful comments and discussions. Zhang acknowledges the financial support from the National Natural Science Foundation of China under Grant Nos. 71973141 and 71873033. All errors are the authors' sole responsibilities. The MATLAB codes for our tests are available upon request. Address correspondence to Yonghui Zhang, School of Economics, Renmin University of China, Beijing, China; e-mail: yonghui.zhang@hotmail.com.

The original version of this article had an incorrect author name. A notice has been published and the error rectified in the online PDF and HTML version.

1985, 1995). Panel data models are widely used in applied works. We refer readers to standard textbooks, such as Wooldridge (2010) and Hsiao (2014), for its history and a comprehensive review. In the last few decades, both econometricians and statisticians propose and study various forms of panel data models that are more reasonable in certain scenarios. For instance, entities observed in the panel data could respond to certain covariates differently from each other, and one may need to consider heterogeneous panel data models. In this direction, Lombardia and Sperlich (2008) estimate and conduct inference for a partial linear panel data model with some group structures. Other related and more recent studies include Su et al. (2016) and Lumsdaine et al. (2023). The heterogeneity may possess other forms. Kneip et al. (2012) study a panel data model with heterogeneous time trends. Boneva et al. (2015) propose a semiparametric model for heterogeneous panel data. Gao et al. (2020) study heterogeneous panel data models with cross-sectional dependence. Furthermore, a common source for heterogeneity in panels is that entities may exhibit cross-sectional dependence. To accommodate this feature, researchers study panels with interactive fixed effects (FEs) or latent factor structures. To name a few, Pesaran (2006) proposes a common correlated effects estimator; Bai (2009) considers an iterative principal components estimator; and Hsiao et al. (2022) propose a transformed estimator for panel interactive effects models. See also Moon and Weidner (2015), Hsiao (2018), and the references therein.

Another feature we observe in some data is that entities might respond differently across time. One possibility is that entities may experience structural breaks at certain times. We refer readers to Bada et al. (2022) and the references therein for a review in this direction. One other possibility is that entities' responses vary across time continuously. This behavior can be modeled as TVC semiparametric models. This strand of literature has been studied extensively. To name a few, Li et al. (2011) propose a local linear dummy variable approach for estimating panel models with TVCs, which is a panel data extension of Cai et al. (2000) and Cai (2007); 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; and Atak et al. (2011) adopt a semiparametric unbalanced panel data model with common smoothing time trends to study climate change in the United Kingdom. For other related works on time-varying or functional coefficients panel data models, see Chen and Hong (2012), Feng et al. (2016), and Zhao et al. (2018), among many others.

Almost all the aforementioned papers in the TVC literature assume that all cross-sectional units share the same vector of constant coefficients and that the heterogeneity among individual units is captured by additive unobservable individuals FEs. Even if the homogeneity assumption (i.e., that the slope coefficients are homogeneous) greatly reduces the dimension of the parameter space, and thus significantly simplifies the processes of estimation and inference, this assumption may be inappropriate in practice, and the constrained estimator with homogeneity may result in a biased estimator for panels with heterogeneity. This may further lead to misleading conclusions (see, e.g., Hsiao and Tahmiscioglu, 1997; Lee et al., 1997). A conservative specification is to allow individual-specific or group-specific slope coefficients. For example, Ma et al. (2020) consider testing empirical asset-pricing models with individual-specific time-varying factor loadings and intercepts; Su et al. (2019) propose a panel data model with grouped TVCs and apply the classified Lasso in Su et al. (2016) to estimate the TVCs and group memberships jointly; and Liu et al. (2020) study a class of panel data models with individual-specific TVCs in the presence of common factors.

Since the specification of stability and/or homogeneity of coefficients plays a critical role in obtaining a consistent estimation and a valid statistical inference for panel data models, it is necessary and prudent for researchers to carry out certain specification or diagnostic tests for the structure of parameters. However, there are only a few tests for the homogeneity of parameters either along time or across individuals in the literature of panel data models. For example, Pesaran and Yamagata (2008) consider testing slope homogeneity in large linear panels; 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; Bartolucci et al. (2015) study the test for time-invariant (against time-variant) unobserved heterogeneity in generalized linear models for panel data; Jin and Su (2013) provide a nonparametric poolability test for panel data models with cross-sectional dependence; Chen and Huang (2018) suggest a nonparametric Wald-type test for the stability of coefficients while assuming that all the coefficients are common among individuals; Gao et al. (2020) provide a test for homogeneity of constant slopes while allowing for individual-specific and nonparametric time trends; and Ma et al. (2020) test whether all the individual-specific time trends are equal to zero jointly for the asset-pricing model with heterogeneous time-varying factor loadings.

Yet there is no test available in the literature 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 constants along both the time series and the cross-sectional dimensions, that is, the usual homogeneous linear panel data model. To fill this gap, in this paper, we provide a nonparametric test for the joint structure on the heterogeneous TVC panel data model. 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.

More importantly, we provide a *unified approach* to examine the commonly used panel data model specifications. The basic idea underlying our approach is to estimate the model under the null hypothesis, and then to explore the information in the generalized residuals using further parametric or nonparametric regressions. Consequently, the approach of testing variation of parameters along time or among individuals is quite flexible and not restricted to the joint test. Indeed, as explored

https://doi.org/10.1017/S026646662300018X Published online by Cambridge University Press

in Section 4, this idea can be extended to constructing tests of the homogeneity of coefficients among individuals or the stability of coefficients along time. This has been illustrated in our application to environmental Kuznets curve (EKC) estimation. The EKC is a hypothesis that suggests that environmental degradation initially increases with economic growth, but eventually decreases as income levels rise. Based on our proposed tests, both the structure of joint homogeneity and stability and the structure of stability are rejected. We reach the conclusion that the model with homogeneous TVCs is more appropriate for the application.

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 is provided in Section 3. In Section 4, we consider the extensions of the test to panel models with homogeneous TVCs, stable heterogeneous coefficients, and two-way FEs. In Section 5, a set of Monte Carlo simulations is conducted to investigate the finite sample performance of our test. We apply our proposed test to study the EKC in the United States in Section 6. Section 7 concludes. The proofs of the main theorems and used lemmas are collected in the Appendix. The proofs of lemmas and some additional theoretical and simulation results are relegated to the Supplementary Material.

Notation. We use $\lambda_{\min}(A)$, $\lambda_{\max}(A)$, and $\mathrm{tr}(A)$ to denote the smallest eigenvalue, the largest eigenvalue, and the trace of a matrix A, respectively. For any $n \times m$ matrix A, let A' be its transpose, and let $\|A\| = [\mathrm{tr}(A'A)]^{1/2}$ be its Frobenius norm. We use p.s.d. (p.d.) for the abbreviation for "positive semidefinite (positive definite)." The symbols $\xrightarrow{p}$ and $\xrightarrow{d}$ denote convergence in probability and 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 constrained estimation under the null hypothesis, and finally propose a feasible test 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 model with FEs and time trends:

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

where Y_{it} is a scalar dependent variable, X_{it} is a d-vector of time-varying exogenous explanatory variables, α_i represents the individual-specific unobservable effect that may be arbitrarily correlated with the regressors X_{it}, and ε_{it} is the idiosyncratic error. The parameters of interest are the unknown vector of TVCs β_{it} and the time trends f_{it}.[1]

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

$$\beta_{it} = \beta_i(\tau_t) \text{ and } f_{it} = f_i(\tau_t) \text{ for } t = 1, \ldots, T, \tag{2.2}$$

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. We note that the value of τ_t depends on T. This fact is important for deriving the asymptotics. We keep using τ_t for convenience. To identify $f_i(\cdot)$ and α_i in (2.1), we impose that $\int_0^1 f_i(\tau)\,d\tau = 0$, for $i = 1, \ldots, N$.[2] Denote the component in Y_{it} explained by regressors (X_{it} and 1) with TVCs as

$$g_{it} \equiv g_i(X_{it}, \tau_t) \equiv X_{it}'\beta_{it} + f_{it}. \tag{2.3}$$

The models specified in (2.1) and (2.2) are quite general and include various existing panel data models 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 standard homogeneous linear panel data model with FEs found in many textbooks (see Baltagi, 2012; Hsiao, 2014; Pesaran, 2015): $Y_{it} = X_{it}'\beta + \alpha_i + \varepsilon_{it}$.
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 FEs (see Hsiao, 2014; Pesaran, 2015; Hsiao and Pesaran, 2008): $Y_{it} = X_{it}'\beta_i + \alpha_i + \varepsilon_{it}$.
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 Li et al. (2011), Chen et al. (2012), Silvapulle et al. (2016), and Chen and Huang (2018): $Y_{it} = f(\tau_t) + X_{it}'\beta(\tau_t) + \alpha_i + \varepsilon_{it}$.
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:

$$Y_{it} = f(\tau_t) + X_{it}'\beta + \alpha_i + \varepsilon_{it}, \tag{2.4}$$

$$Y_{it} = f_i(\tau_t) + X_{it}'\beta + \alpha_i + \varepsilon_{it}, \tag{2.5}$$

$$Y_{it} = f(\tau_t) + X_{it}'\beta_i + \alpha_i + \varepsilon_{it}, \tag{2.6}$$

$$Y_{it} = f_i(\tau_t) + X_{it}'\beta_i + \alpha_i + \varepsilon_{it}, \tag{2.7}$$

[1] The setup in (2.1) can be easily generalized to allow for a mixture structure such as $Y_{it} = X_{1,it}'\beta_{1,it} + X_{2,it}'\beta_{2,i} + X_{3,it}'\beta_{3,t} + X_{4,it}'\beta_4 + \alpha_i + \varepsilon_{it}$ with time trends (f_{it} or f_t) being absorbed in the first or third component, respectively.

[2] Another identification restrictions can be $f_i(c^*) = 0$ for some $c^* \in [0,1]$, $i = 1, \ldots, N$.

https://doi.org/10.1017/S026646662300018X Published online by Cambridge University Press

where models (2.4)–(2.7) have been studied by Chen et al. (2012), Zhang et al. (2012), Atak et al. (2011), and Gao et al. (2020), respectively.

5. When there are no regressors ($\beta_i(\cdot) = 0$ for all i's), then model (2.1) becomes the nonparametric trending panel data models:

$$Y_{it} = f(\tau_t) + \alpha_i + \varepsilon_{it} \text{ or } Y_{it} = f_i(\tau_t) + \alpha_i + \varepsilon_{it},$$

where the homogeneous trending model has been studied by Robinson (2012) and the later model allows for individual-specific 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. (2019).

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

$$\mathbb{H}_0 : (\beta_{it}, f_{it}) = (\beta_0, 0) \text{ for some } \beta_0 \in \mathbb{R}^d \text{ and all } (i,t)\text{'s} \tag{2.8}$$

against the alternative hypothesis

$$\mathbb{H}_1 : (\beta_{it}, f_{it}) \neq \left(\beta_{js}, f_{js}\right) \text{ for some } (i,t) \neq (j,s). \tag{2.9}$$

When the null hypothesis holds, all the cross-sectional units share the same time-invariant slopes for regressors X_{it} and do not have time trends. Then model (2.1) becomes the usual homogeneous linear panel data model with FEs, which is the most widely used setup in empirical applications. We can estimate the model either by the usual FE estimator or the first-difference (FD) estimator. When X_{it} include the lags of the dependent variable or endogenous variables, we can estimate the model by the generalized method of moments or instrumental variables approach, and the proposed test statistics to be discussed are still valid with extra assumptions and more laborious derivation.

For the above hypothesis testing problem, the test statistic can be constructed in the spirit of Wald, Lagrange multiplier, or likelihood ratio tests. In this paper, we propose a nonparametric test for the structure in (2.8) based on the residuals from estimation under the null hypothesis for several reasons: first, constrained estimation under $\mathbb{H}_0$ usually estimates fewer parameters and is much simpler than estimation without restriction; second, parsimonious models with restrictions on parameters (homogeneity across individuals and stability along time) are usually the starting point of many empirical studies; third, our proposed test provides a diagnostic check when a simple and popular model is fitted by exploring the information underlying the residuals; and finally, the testing strategy provides a unified approach to testing other commonly used structures such as homogeneity, stability, or group pattern on parameters in panel data models.

2.2. The Test Statistic

We first consider the estimation of the model under the null hypothesis $\mathbb{H}_0$. The model (2.1) reduces to

$$Y_{it} = X_{it}'\beta_0 + \alpha_i + \varepsilon_{it}, \tag{2.10}$$

then we can estimate β_0 either by the FE or the FD estimator. For illustrative purposes, we adopt the FE estimator

$$\hat{\beta}_{FE} = \left(\sum_{i=1}^{N} X_i' M_T X_i\right)^{-1} \sum_{i=1}^{N} X_i' M_T Y_i, \tag{2.11}$$

where $M_T \equiv I_T - \iota_T \iota_T'/T$, 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 (2.3) is estimated by $\hat{g}_{it} = X_{it}'\hat{\beta}_{FE}$.

Denote $g_{P,it} = X_{it}'\beta_P$ with $\beta_P = \left[\sum_{i=1}^{N} E\left(X_i' M_T X_i\right)\right]^{-1} \sum_{i=1}^{N} E\left(X_i' M_T Y_i\right)$. Let $\hat{u}_{it} = Y_{it} - \hat{g}_{it}$ be the augmented residual and $\eta_{it} = \hat{g}_{it} - g_{P,it}$ the "estimation error" when using $\hat{g}_{it}$ to estimate $g_{P,it}$. We decompose $\hat{u}_{it}$ as follows:

$$\hat{u}_{it} = \left(g_{it} - g_{P,it}\right) + \left(g_{P,it} - \hat{g}_{it}\right) + \alpha_i + \varepsilon_{it} \equiv g_{it}^{\dagger} - \eta_{it} + u_{it}, \tag{2.12}$$

where $u_{it} = \alpha_i + \varepsilon_{it}$ is the generalized error. Note that η_{it} is asymptotically negligible under either the null or the alternative hypotheses,[3] and $g_{it}^{\dagger} (\equiv g_{it} - g_{P,it})$ can be rewritten as

$$g_{it}^{\dagger} = f_i(\tau_t) + X_{it}'[\beta_i(\tau_t) - \beta_P] \equiv f_i(\tau_t) + X_{it}'\beta_i^{\dagger}(\tau_t). \tag{2.13}$$

Clearly, under $\mathbb{H}_0$, we have $g_{it}^{\dagger} = 0$ for all i and t because of $\beta_i(\cdot) = \beta_0 = \beta_P$ and $f_i(\cdot) = 0$ for all i's; however, β_{it} and f_{it} have variation either across i or over t under $\mathbb{H}_1$, and then in general $\beta_i^{\dagger}(\tau_t) \neq 0$ or $f_i(\tau_t) \neq 0$. 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 average of squared $g_{it}^{\dagger}$'s:[4]

$$\Gamma_{NT}^0 = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} g_{it}^{\dagger 2}.$$

Clearly, by construction, Γ_{NT}^0 equals 0 under $\mathbb{H}_0$ but is greater than 0 under $\mathbb{H}_1$. However, Γ_{NT}^0 is infeasible because $g_{it}^{\dagger}$'s are unknown to the researchers. Therefore, we need a consistent estimation of $g_{it}^{\dagger}$.

[3] The statement holds under Assumptions 1 and 2 in Section 3.1.

[4] Alternatively, we can consider a weighted version $\Gamma_{NT}^0 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} g_{it}^{\dagger 2} w_{it}$, where $w_{it} \equiv w_i(\tau_t)$ and $w_i(\cdot)$'s are some user-specified nonnegative weighting functions. In practice, we can use $w_i(\tau) = 1\,(c \leq \tau \leq 1-c)$ with a small $c > 0$ to remove the boundary observations to improve the finite sample performance.

https://doi.org/10.1017/S026646662300018X Published online by Cambridge University Press

Noting that $\hat{u}_{it}$ is a consistent estimator for u_{it} under $\mathbb{H}_0$ and for $g_{it}^{\dagger}+u_{it}$ under $\mathbb{H}_1$, we can estimate $\{g_{it}^{\dagger}\}_{t=1}^{T}$ from $\left\{\hat{u}_{it}\right\}_{t=1}^{T}$ by the regression of $\hat{u}_{it}$ on X_{it} and 1 with TVCs. To be specific, for each $i=1,\ldots,N$, we run the following auxiliary time series regression with TVCs:

$$\hat{u}_{it}=f_i(\tau_t)+X_{it}'\beta_i^{\dagger}(\tau_t)+\alpha_i+\varepsilon_{it}^{\dagger},\ t=1,\ldots,T, \tag{2.14}$$

where $\varepsilon_{it}^{\dagger}\equiv\varepsilon_{it}-\eta_{it}$. Noting that $f_i(\cdot)$ and $\beta_i^{\dagger}(\cdot)$ are all unknown functions defined on [0,1], which can be estimated either by the kernel method (e.g., Li et al., 2011; Chen and Huang, 2018) or by the sieve method (e.g., Su and Zhang, 2016; Dong and Linton, 2018; Zhang and Zhou, 2021). In this paper, we consider the sieve estimation of the unknown functions in (2.14).

Let $L^2[0,1]=\{m(\tau):\int_0^1 m^2(\tau)\,d\tau<\infty\}$, in which $\langle m_1,m_2\rangle=\int_0^1 m_1(\tau)m_2(\tau)\,d\tau$ is the inner product and the induced norm is $\|m\|=\langle m,m\rangle^{1/2}$. Following Dong and Linton (2018), we choose cosine functions as basis functions.[5] Let $B_0(\tau)=1$, and $B_j(\tau)=\sqrt{2}\cos(j\pi\tau)$, for $j\geq 1$. Then $\left\{B_j(\tau)\right\}_{j=1}^{\infty}$ forms an orthonormal basis for the Hilbert space $L^2[0,1]$ such that $\left\langle B_i,B_j\right\rangle=\delta_{ij}$, where δ_{ij} is the Kronecker delta. For any unknown continuous function $m\in L^2[0,1]$, it can be written as

$$m(\tau)=\sum_{j=0}^{\infty}\pi_jB_j(\tau),\text{ where }\pi_j\equiv\left\langle m,B_j\right\rangle.$$

For model (2.14), we further assume $\beta_{il}^{\dagger}(\cdot)\in L^2[0,1]$, for $l=1,\ldots,d$, and $f_i(\cdot)\in L^2[0,1]$, for each i. 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 two sequences of basis functions to approximate unknown functions $\beta_{il}^{\dagger}(\cdot)$ $(l=1,\ldots,d)$ and $f_i(\cdot)$, respectively. The constant term is excluded from $B_{-1}^K(\cdot)$ because of the identification restriction on $f_i(\cdot)$'s. Then, for each i, we obtain[6]

$$\beta_{il}^{\dagger}(\cdot)=\sum_{j=0}^{\infty}\vartheta_{\beta,il,j}B_j(\cdot)=\vartheta_{\beta,il}'B^K(\cdot)+r_{\beta_{il}^{\dagger}}^{(K)}(\cdot),\ l=1,\ldots,d, \tag{2.15}$$

$$f_i(\cdot)=\sum_{j=1}^{\infty}\vartheta_{f,i,j}B_j(\cdot)=\vartheta_{f,i}'B_{-1}^K(\cdot)+r_{f_i^{\dagger}}^{(K)}(\cdot), \tag{2.16}$$

where $\vartheta_{\beta,il,j}=\left\langle\beta_{il}^{\dagger},B_j\right\rangle$ and $\vartheta_{f,i,j}=\left\langle f_i,B_j\right\rangle$ for integer $j\geq 1$, $\vartheta_{\beta,il}=\left(\vartheta_{\beta,il,0},\ldots,\vartheta_{\beta,il,K-1}\right)'$, $\vartheta_{f,i}=\left(\vartheta_{f,i,1},\ldots,\vartheta_{f,i,K-1}\right)'$, $r_{\beta_{il}^{\dagger}}^{(K)}(\cdot)=\sum_{j=K}^{\infty}\vartheta_{\beta,il,j}B_j(\cdot)$, and $r_{f_i}^{(K)}(\cdot)=\sum_{j=K}^{\infty}\vartheta_{f,i,j}B_j(\cdot)$. By Assumption 3 in Newey (1997), $\sup_{\tau\in[0,1]}\left|r_{\beta_{il}^{\dagger}}^{(K)}(\tau)\right|=O\left(K^{-\kappa}\right)$

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

[6] Different numbers of basis functions can be adopted in estimating different functions. For simplicity, we use the same number K in the sieve approximation of different functions.

https://doi.org/10.1017/S026646662300018X Published online by Cambridge University Press

and $\sup_{\tau\in[0,1]}\left|r_{f_i}^{(K)}(\cdot)\right| = O\left(K^{-\kappa}\right)$, for $i = 1,\ldots,N$, as $K \to \infty$ when $\beta_{il}^{\dagger}(\cdot)$ and $f_i(\cdot)$ have κth-order continuous derivatives. Then we can approximate $\beta_{il}^{\dagger}(\cdot)$ by $\vartheta_{\beta,il}' B^K(\cdot)$ and $f_i(\cdot)$ by $\vartheta_{f,i}' B_{-1}^K(\cdot)$, respectively.

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 (2.15) and (2.16) yields

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

where $\vartheta_i \equiv (\vartheta_{f,i}', \text{vec}(\vartheta_{\beta,i})')'$, $\vartheta_{\beta,i} = \left(\vartheta_{\beta,i1},\ldots,\vartheta_{\beta,id}\right)$, 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 based on the sieve approximation is given by

$$\hat{u}_{it} = Z_{it}'\vartheta_i + \alpha_i + v_{it},\ t = 1,\ldots,T, \tag{2.17}$$

where $v_{it} = \varepsilon_{it} - \eta_{it} + r_{it}^{\dagger}$, and $r_{it}^{\dagger} \equiv g_{it}^{\dagger} - Z_{it}'\vartheta_i = \sum_{l=1}^{d} r_{\beta_{il}^{\dagger}}^{(K)}(\tau_t) X_{it,l} + r_{f_i}^{(K)}(\tau_t)$ comes from the sieve approximation error of $g_{it}^{\dagger}$. Rewrite the model (2.17) in vector form

$$\hat{u}_i = Z_i\vartheta_i + \iota_T\alpha_i + v_i, \tag{2.18}$$

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

$$\hat{\vartheta}_i = \left(Z_i'M_TZ_i\right)^{-1} Z_i'M_T\hat{u}_i \text{ and } \hat{g}_{it}^{\dagger} = Z_{it}'\hat{\vartheta}_i. \tag{2.19}$$

On the basis of the sieve estimator $\hat{g}_{it}^{\dagger}$ for $g_{it}^{\dagger}$ in (2.19), we estimate Γ_{NT}^0 by[7]

$$\Gamma_{NT} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{g}_{it}^{\dagger 2} = \frac{1}{NT}\sum_{i=1}^{N}\hat{\vartheta}_i'Z_i'Z_i\hat{\vartheta}_i \tag{2.20}$$

$$= \frac{1}{NT}\sum_{i=1}^{N}\hat{u}_i'M_TZ_i\left(Z_i'M_TZ_i\right)^{-1}Z_i'Z_i\left(Z_i'M_TZ_i\right)^{-1}Z_i'M_T\hat{u}_i. \tag{2.21}$$

Note that $\hat{u}_{it} = u_{it} + X_{it}'(\beta_0 - \hat{\beta}_{FE}) \approx u_i$ under $\mathbb{H}_0$. Then $\Gamma_{NT} \approx \Gamma_{NT}^* \equiv \frac{1}{NT}\sum_{i=1}^{N}\varepsilon_i'\mathcal{A}_i\varepsilon_i$ with $\mathcal{A}_i \equiv M_TZ_i\left(Z_i'M_TZ_i\right)^{-1}Z_i'Z_i\left(Z_i'M_TZ_i\right)^{-1}Z_i'M_T$. Note that Γ_{NT}^* is a cross-sectional average of quadratic form of ε_i. When X_{it} are strictly exogenous, we can easily derive $E\left(\Gamma_{NT}^*\right)$ and $\text{Var}\left(\Gamma_{NT}^*\right)$ as the asymptotic bias and variance, respectively.

[7]Instead of taking the average in (2.20), one may adopt statistics taking the maximum of the absolute deviation from $\mathbb{H}_0$ over individual $i = 1,2,\ldots,N$. The techniques of this kind of test statistics are more involved, and the limiting distribution and the associated inference procedure can be nonstandard (see, for example, Chernozhukov and Fernández-Val, 2011). We leave this interesting direction to future research.

https://doi.org/10.1017/S026646662300018X Published online by Cambridge University Press

In the next section, we show that after being appropriately centered and scaled, Γ_{NT} follows a standard normal distribution asymptotically under $\mathbb{H}_0$ and a set of regular conditions.

3. ASYMPTOTIC THEORY

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

3.1. Assumptions

Let $\max_{i,t}$ and $\min_{i,t}$ denote $\max_{1\le i\le N}\max_{1\le t\le T}$ and $\min_{1\le i\le N}\min_{1\le t\le T}$, respectively.

To study the asymptotic properties for Γ_{NT} under the null hypothesis, we make the following assumptions:

Assumption 1. (i) $\{(X_i,\varepsilon_i)\}_{i=1}^N$ are independent across i, where $X_i=(X_{i1},\ldots,X_{iT})'$ and $\varepsilon_i=(\varepsilon_{i1},\ldots,\varepsilon_{iT})'$.

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

(iii) $\{\varepsilon_{it}\}_{t=1}^T$ is a martingale difference sequence (MDS) with respect to (w.r.t.) filtrations $\{\mathcal{F}_t\}_{t=1}^T$ such that $E(\varepsilon_{it}|\mathcal{F}_t)=0$, where $\mathcal{F}_t$ is the σ-field generated by $\left\{\left(X_{it},X_{i,t-1},\varepsilon_{i,t-1},\ldots,X_{i1},\varepsilon_{i1}\right)\right\}_{i=1}^N$.

(iv) $\max_{i,t}E|\varepsilon_{it}|^{16(1+\eta)}<\infty$ and $\max_{i,t}E\|X_{it}\|^{16(1+\eta)}<\infty$ for some constant $\eta>0$.

(v) $\mathrm{Var}(X_{it})=\Omega_i^{(x)}(\tau_t)$, where $\Omega_i^{(x)}(\cdot)$ is a $d\times d$ matrix of piecewise continuous functions on $[0,1]$ with countable discontinuities. There exist some positive constants $\underline{c}_{xx}$ and $\bar{c}_{xx}$ such that

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

(vi) Let $\tilde{X}_{it}^{(\varepsilon)}\equiv\left(1,X_{it}'\right)'\varepsilon_{it}$ and $\mathrm{Var}(\tilde{X}_{it}^{(\varepsilon)})=\Omega_i^{(\varepsilon)}(\tau_t)$, where $\Omega_i^{(\varepsilon)}(\cdot)$ is a $(d+1)\times(d+1)$ matrix of piecewise continuous functions on $[0,1]$ with countable discontinuities. There exist some positive constants $\underline{c}_{xx}^{(\varepsilon)}$ and $\bar{c}_{xx}^{(\varepsilon)}$ such that

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

Assumption 2. As $(N,T)\to\infty$, $K\to\infty$, $K^2/N\to0$, $NK/T^2\to0$, and $K^2(\log T)^4/T\to0$.

Several remarks can be made about the above assumptions. Assumption 1(i) requires the cross-sectional independence of $\{(X_i,\varepsilon_i)\}_{i=1}^N$, which is also used in Lee and Robinson (2016) and Su et al. (2019); the assumption can be relaxed to allow for weak cross-sectional dependence among error terms as in Bai (2009), Chen et al. (2012), and Lee and Robinson (2016) with more complicated arguments in the proofs; and for models with two-way FEs, it can be replaced by the

https://doi.org/10.1017/S026646662300018X Published online by Cambridge University Press

cross-sectional independence of (X_i, ε_i) conditional on the σ-field generated by all FEs. Assumption 1(ii) assumes that $\{(X_{it}, \varepsilon_{it}), t = 1, \ldots, T\}$ is strong mixing with a geometric decay rate, which can be easily satisfied by many well-known linear processes such as autoregressive–moving-average processes and nonlinear processes. Note that the mixing coefficient $\alpha_i(l)$ is defined as

$$\alpha_i(l) = \sup_{k\in\mathbb{Z}} \alpha_i\left(\mathcal{F}^i_{-\infty,k}, \mathcal{F}^i_{k+l,\infty}\right) = \sup_{k\in\mathbb{Z}} \sup_{A\in\mathcal{F}^i_{-\infty,k}, B\in\mathcal{F}^i_{k+l,\infty}} 2\left|P(A\cap B) - P(A)P(B)\right|, \tag{3.1}$$

where $\mathcal{F}^i_{n,m}$ is the σ-field generated by $\{(X_{it}, \varepsilon_{it}), n < t \leq m\}$ (Bosq, 1998, p. 19). Assumption 1(iv) imposes a martingale difference structure on $\{\varepsilon_{it}\}$ with filtrations $\{\mathcal{F}_t\}_{t=1}^T$, which is also adopted in Chen and Huang (2018). The MDS assumption is suitable when lagged dependent variables are included in X_{it}, and can be relaxed with more complicated proofs. Assumption 1(iv) specifies some high-order moment conditions on ε_{it} and X_{it} in testing. They are required to verify a key condition involving fourth-order moments of a quadratic form of $Z_{it}\varepsilon_{it}$ (or $X_{it}\varepsilon_{it}$). There is no need to impose such high-order moments in estimation. Assumption 1(v) and (vi) allows the variance of X_{it} and $\tilde{X}_{it}^{(\varepsilon)}$ to be time-varying and requires that their eigenvalues be bounded and bounded away from 0. It is possible to weaken the current mixing condition to $\alpha(l) \leq Cl^{-\theta}$ for some positive θ. With the weaker mixing condition, we need to face some trade-off of θ and some moment conditions, similar to those in Lemma A.8 in the Appendix.

Assumption 2 provides conditions on the relative rate of the sample size (N, T) and the number of sieve basis terms K. Note that $K^2/N \to 0$ is used to show the consistency of the asymptotic variance term estimator ($\hat{\mathbb{V}}_{NT}$ in Section 3.2); $NK/T \to 0$ and $K^2(\log T)^4/T \to 0$ are used in the establishment of central limit theorem (CLT) in the proofs of Proposition A.10. These requirements are quite mild and include various combinations of (N, T, K). It allows for $T/N \to c \in [0, \infty]$. To see that, if the $K \propto T^{1/5}$ (as chosen in the simulations) is used, the condition on the relationship between T and N is reduced to $N/T^{1.8} \to 0$ and $T^{0.4}/N \to 0$. Clearly, it permits $N = T^b$ with $b \in (0.4, 1.8)$. Note that we can let K be fixed under $\mathbb{H}_0$ because there is no sieve approximation error under $\mathbb{H}_0$. It implies that both Theorems 3.1 and 4.1(i) (to be introduced) hold when K is fixed.

3.2. Asymptotic Distribution under the Null

We first introduce some notations. Let $\hat{Q}_{i,\dot{z}\dot{z}} = T^{-1}\sum_{t=1}^T \dot{Z}_{it}\dot{Z}_{it}'$ and $\hat{Q}_{i,zz} = T^{-1}\sum_{t=1}^T Z_{it}Z_{it}'$, where $\dot{Z}_{it} = Z_{it} - \bar{Z}_i$ and $\bar{Z}_i = T^{-1}\sum_{t=1}^T Z_{it}$. Denote $Q_{i,\dot{z}\dot{z}} \equiv T^{-1}\sum_{t=1}^T E\left(\dot{Z}_{it}\dot{Z}_{it}'\right)$, $Q_{i,zz} \equiv T^{-1}\sum_{t=1}^T E\left(Z_{it}Z_{it}'\right)$, and $\Omega_i \equiv T^{-1}\sum_{t=1}^T E(\mathring{Z}_{it}\mathring{Z}_{it}'\varepsilon_{it}^2)$, where $\mathring{Z}_{it} \equiv Z_{it} - E\left(\bar{Z}_i\right)$. Further, let $\hat{\mathbb{Q}}_i = \hat{Q}_{i,\dot{z}\dot{z}}^{-1}\hat{Q}_{i,zz}\hat{Q}_{i,\dot{z}\dot{z}}^{-1}$ and $\mathbb{Q}_i = Q_{i,\dot{z}\dot{z}}^{-1}Q_{i,zz}Q_{i,\dot{z}\dot{z}}^{-1}$. To obtain the asymptotic distribution of Γ_{NT}, we define

$$\mathbb{B}_{NT} = \frac{1}{\sqrt{N}}\sum_{i=1}^N \operatorname{tr}(\mathbb{Q}_i\Omega_i) \text{ and } \mathbb{V}_{NT} = \frac{2}{N}\sum_{i=1}^N \operatorname{tr}(\mathbb{Q}_i\Omega_i\mathbb{Q}_i\Omega_i). \tag{3.2}$$

https://doi.org/10.1017/S026646662300018X Published online by Cambridge University Press

Note that Q_i and Ω_i are both well-defined matrices. It is easy to show that $\mathbb{B}_{NT} = O_p\left(N^{1/2}K\right)$ and $\mathbb{V}_{NT} = O_p(K)$, and they can be estimated by sample analogs:

$$\hat{\mathbb{B}}_{NT} = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\operatorname{tr}\left(\hat{\mathbb{Q}}_i\hat{\Omega}_i\right) \text{ and } \hat{\mathbb{V}}_{NT} = \frac{2}{N}\sum_{i=1}^{N}\operatorname{tr}\left(\hat{\mathbb{Q}}_i\hat{\Omega}_i\hat{\mathbb{Q}}_i\hat{\Omega}_i\right), \tag{3.3}$$

where $\hat{\Omega}_i = T^{-1}\sum_{t=1}^{T}\dot{Z}_{it}\dot{Z}_{it}'\hat{\varepsilon}_{r,it}^2$, $\hat{\varepsilon}_{r,it} = \hat{u}_{it} - \overline{\hat{u}}_i$ and $\overline{\hat{u}}_i = T^{-1}\sum_{t=1}^{T}\hat{u}_{it}$.[8] Finally, we propose the normalized test statistic:

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

The following theorem provides the asymptotic distribution of $\hat{J}_{NT}$ under the null hypothesis $\mathbb{H}_0$.

THEOREM 3.1. *Suppose that Assumptions 1 and 2 hold. Under* $\mathbb{H}_0$, $\hat{J}_{NT} \xrightarrow{d} N(0,1)$ *as* $(N,T) \to \infty$.

Remark 1. (i) The proof is complicated and relegated to the Appendix. 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, that is, the upper αth percentile from the standard normal distribution, and reject the null when $\hat{J}_{NT} > z_\alpha$ at the α significance level. (ii) Note that under the null $\mathbb{H}_0$, $f_i(\cdot) = 0$ and $\beta_i^\dagger(\cdot) = 0$, which can be exactly expressed as a linear combination of any K basis functions. The limiting distribution in Theorem 3.1 still holds when K is fixed, but in practice a diverging number of basis functions is used to increase the power of the test. (iii) We can modify the test statistics to test the structure of a subvector of coefficients or only time trends.

3.3. Asymptotic Distribution under Local Alternatives

To study the local power properties 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 $g_{\Delta,it} \equiv X_{it}'\Delta_{\beta,it} + \Delta_{f,it}$, $g_{\Delta,i} = \left(g_{\Delta,i1},\ldots,g_{\Delta,iT}\right)'$ and $\bar{g}_{\Delta,it} = X_{it}'\bar{\Delta}_\beta$, where $\bar{\Delta}_\beta = \left[\sum_{i=1}^{N}E\left(X_i'M_TX_i\right)\right]^{-1}\sum_{i=1}^{N}E\left(X_i'M_Tg_{\Delta,i}\right)$. Then we define

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

https://doi.org/10.1017/S026646662300018X Published online by Cambridge University Press

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

Before we establish 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 the κth 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 $\gamma_{NT} \equiv N^{-1/4}T^{-1/2}\mathbb{V}_{NT}^{1/4}$, where Φ_{Δ} is defined in Assumption 4.

Remark 2. (i) Theorem 3.2 implies that our test has nontrivial asymptotic power against alternatives that diverge from the null at rate $O\left(N^{-1/4}T^{-1/2}K^{1/4}\right)$ by noting that $\mathbb{V}_{NT} = O_p(K)$ (see Lemma A.6 in the Appendix). The power increases with the magnitude of Φ_{Δ}. Clearly, as either N or T increases, the power increases and 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. (2019). (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. (iii) Our test may have low or no power against the "sparse" alternatives with the majority of $\breve{g}_{\Delta,it}$ being close to 0 such that the probability limit of $\Phi_{\Delta,NT}$ is 0 or is close to 0. We do need a nonnegligible proportion of individuals that deviate from the null significantly so that our test can detect this deviation. We note that the low power problem happens for almost all average-type tests in large panel data models. The power enhancement device in Fan et al. (2015) can be adopted to boost the power when the alternative has a sparse structure.

To study the global consistency of $\hat{J}_{NT}$ under $\mathbb{H}_1$, let $\gamma_{NT} = 1$ in (3.5). Note that we impose Assumption 3 for $\gamma_{NT} = 1$ in the corollary followed. That is equivalent to assuming that $\beta_i(\cdot)$ and $f_i(\cdot)$ are κth-order continuously differentiable under $\mathbb{H}_1$. 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 $\hat{\mathbb{V}}_{NT} = O_p(K)$ 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. Then, under $\mathbb{H}_1$, $N^{-1/2}T^{-1}\hat{\mathbb{V}}_{NT}^{1/2}\hat{J}_{NT} \xrightarrow{p} \Phi_{\Delta}$ as $(N,T) \to \infty$.*

https://doi.org/10.1017/S026646662300018X Published online by Cambridge University Press

Remark 3. Corollary 3.3 establishes that $\hat{J}_{NT}$ diverges to ∞ at rate $O_p(N^{1/2}T/K^{1/2})$ 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 $\Phi_\Delta > 0$. Note under $\mathbb{H}_1 : \beta_{it} = \beta_0 + \Delta_{\beta,it}$ and $f_{it} = \Delta_{f,it}$, the power of our test comes from the magnitudes of $\|\Delta_{\beta,it}\|$ and $\|\Delta_{f,it}\|$. The bias of estimating $\Delta_{\beta,it}$ is a small order term of $\|\Delta_{\beta,it}\|$, and thus the bias term is asymptotically negligible compared to $\|\Delta_{\beta,it}\|$. This also applies to the estimation of $\Delta_{f,it}$. The consistency of the estimates is enough to guarantee the power. Therefore, we do not need the bias to be smaller than the standard deviation (mathematically $N^{1/2}TK^{-(1/2+2\kappa)} \to 0$ in this case) that is typically needed in nonparametric estimations for inferences.

Remark 4. Choosing the optimal number of sieve terms is important in practice. To the best of our knowledge, there is no existing work on the optimal K in nonparametric testing using the sieve regression method. 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 Horowitz and Spokoiny (2003), Gao and King (2004), or Gao and Gijbels (2008). We do not pursue the optimal choice of K theoretically and leave it to future research; instead, we propose to choose K based on some unique features of our test. Theorem 3.1 implies that the results under $\mathbb{H}_0$ hold as long as K does not diverge too fast (hold even under fixed K). This suggests that the size of the test should not be very sensitive to the choice of K. In terms of power, intuition suggests that the test should be more powerful when $\beta_i(\cdot)$ and $f_i(\cdot)$ are estimated more precisely. Theoretically, we should set $K \propto T^{1/(2\kappa+1)}$ to minimize the root-mean-square error (RMSE). In practice, we may choose K by the leave-one-out cross-validation (LOOCV) method (see Section 5 for details). The simulation results in Section 5 basically confirm these findings; the size of the test is not very sensitive to the choice of K, and the test is relatively more powerful using the K chosen by the LOOCV method.

Remark 5.[9] The test statistic in one single step may be constructed as follows. We directly regress the demeaned Y_{it} on the demeaned Z_{it} for each i and obtain the fitted coefficients $\tilde{\vartheta}_i = \left(Z_i'M_TZ_i\right)^{-1} Z_i'M_TY_i$. Note that X_{it} is a subvector of $Z_{it} \equiv (B_{-1,t}', X_{it}', \left(X_{it} \otimes B_{-1,t}\right)')'$. Then, under $\mathbb{H}_0$, the coefficients of X_{it} should be constants across i and the coefficients for the rest component in Z_{it} are 0. Using this observation, we can proceed to study the statistic

$$\tilde{\Gamma}_{NT} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[X_{it}'\left(\tilde{\vartheta}_{X_i} - \frac{1}{N}\sum_{i=1}^{N}\tilde{\vartheta}_{X_i}\right) + Z_{-X,it}'\tilde{\vartheta}_{-X_i}\right]^2,$$

where $\tilde{\vartheta}_{X_i}$ denotes the estimated coefficients of X_{it} in $\tilde{\vartheta}_i$, $Z_{-X,it}$ denotes the vector Z_{it} after removing X_{it}, and $\tilde{\vartheta}_{-X_i}$ picks the corresponding coefficients for $Z_{-X,it}$ from $\tilde{\vartheta}_i$. We can show that this one-step procedure is equivalent to ours

[9] We thank a referee for pointing this out.

https://doi.org/10.1017/S026646662300018X Published online by Cambridge University Press

as $(N,T) \to \infty$ (for details, see Appendix E of the Supplementary Material). This paper focuses on the two-step procedure to provide a unified approach for testing the structure in panel data models.

Remark 6. Our proof can be applied to unbalanced panels when only a fixed number of observations for each i are missing. Suppose, for each i, that there are c_i ($c_i \geq 0$) missing observations. Specifically, we need $\max_{1\leq i\leq N} c_i \leq C < \infty$ uniformly for all t for our proof to go through. For each i, only a fixed number of observations are missing. In this case, missing observations are asymptotically negligible for the estimation for each i, as $T \to \infty$. As a result, all the proofs are expected to go through. For other cases when the number of missing observations diverges for some i, a careful and tedious treatment is needed. We leave this for future work.

3.4. Bootstrap Procedure

It is well known that tests based on nonparametric estimation usually suffer severe size distortion in finite samples if standard normal critical values are used (see Li and Wang, 1998; Su and Hoshino, 2016). The empirical size of these tests can be quite sensitive to the choice of basis number or highly distorted in finite samples. Therefore, we suggest using a bootstrap method to obtain bootstrap p-values. We follow Hansen (2000) and propose a fixed-regressor bootstrap procedure to obtain bootstrap p-values.

The bootstrap procedure is as follows:

1. Obtain $\hat{\beta}_{FE}$ and $\hat{u}_{it}$ under $\mathbb{H}_0$. For each i, run auxiliary time series regressions of $\hat{u}_{it}$ on X_{it} and 1 with TVCs to obtain fitted values $\hat{g}_{it}^{\dagger}$, residuals $\hat{\varepsilon}_{r,it}$, and then calculate $\hat{J}_{NT}$.
2. For each i, obtain the wild bootstrap errors $\{\varepsilon^*_{r,it}\} : \varepsilon^*_{r,it} = \hat{\varepsilon}_{r,it}\varrho_{it}$, where ϱ_{it}'s are i.i.d. $N(0,1)$ across i and t. Then generate the bootstrap analog 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 $\hat{\alpha}_i = T^{-1}\sum_{t=1}^{T}(\hat{u}_{it} - \hat{g}_{it}^{\dagger})$.
3. Given the bootstrap resample $\{Y^*_{it}, X_{it}\}$, estimate the linear homogeneous panel data model using all data and run N auxiliary time series regressions as Step 1. Obtain the fitted value $\hat{g}^*_{it}$ and residual $\hat{\varepsilon}^*_{r,it}$. Calculate the bootstrap test statistic $\hat{J}^*_{NT}$ based on $\{\hat{g}^*_{it}, \hat{\varepsilon}^*_{r,it}\}$.
4. Repeat Steps 2 and 3 B times and index the bootstrap statistics as $\{\hat{J}^*_{NT,b}\}_{b=1}^{B}$. Calculate the bootstrap p-value by $p^* = B^{-1}\sum_{b=1}^{B}\mathbf{1}(\hat{J}^*_{NT,b} \geq \hat{J}_{NT})$.

It is straightforward to implement the above bootstrap procedure. Note that for the bootstrap resample, we impose the null hypothesis of linearity and homogeneity on parameters in Step 2.

Let $\mathcal{W}_{NT} \equiv \{(X_{it}, Y_{it}) : i = 1,\ldots,N, t = 1,\ldots,T\}$ be the observed sample. Recall that $\hat{\Omega}_i = T^{-1}\sum_{t=1}^{T}\dot{Z}_{it}\dot{Z}'_{it}\hat{\varepsilon}^2_{r,it}$. The next theorem establishes the asymptotic validity of the above bootstrap procedure.

THEOREM 3.4. *Suppose that the assumptions in Theorem 3.1 or 3.2 hold. Assume that* $0 < \min_{1\leq i\leq N}\lambda_{\min}(\hat{\Omega}_i) \leq \max_{1\leq i\leq N}\lambda_{\max}(\hat{\Omega}_i) < \infty$. *Then, as* $(N,T)\to\infty$,

$$\hat{J}^*_{NT} \xrightarrow{d^*} N(0,1)$$

in probability, where d^* *denotes the weak convergence under the bootstrap probability measure conditional on* $\mathcal{W}_{NT}$.

Remark 7. (i) The proof of Theorem 3.4 is given in the Appendix. Note that ϱ_{it}'s are i.i.d. $N(0,1)$ across i and t, and the proof is much simpler than that of Theorem 3.1. (ii) We provide a set of low-level conditions for $0 < \min_{1\leq i\leq N}\lambda_{\min}(\hat{\Omega}_i) \leq \max_{1\leq i\leq N}\lambda_{\max}(\hat{\Omega}_i) < \infty$ to hold uniformly after some large T almost surely, and show its validity in Lemma A.9.

4. SEVERAL EXTENSIONS

In this section, we consider several extensions. We provide a test for the stability of heterogeneous coefficients (Section 4.1) and a test for homogeneity of TVCs (Section 4.2). These two specifications of parameters are also commonly used in empirical studies. After that, we consider the panel data models with both the individual and the time FEs in Section 4.3. In the following, we provide an overview of the results in this section for an easier reference for applied researchers.

In Section 4.1, we consider the test of

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

against the alternative hypothesis $\mathbb{H}_{s1}: (\beta_i(\cdot), f_i(\cdot)) \neq (\beta_i, 0)$ for some i's. The model under $\mathbb{H}_{s0}$ is linear with heterogeneous coefficients:

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

In Section 4.2, we consider the test of

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

against the alternative hypothesis $\mathbb{H}_{h1}: (\beta_i(\cdot), f_i(\cdot)) \neq \left(\beta_j(\cdot), f_j(\cdot)\right)$ for some $i \neq j$. Under the null, the model is a panel data model with homogeneous TVCs:

$$Y_{it} = X_{it}'\beta_0(\tau_t) + f_0(\tau_t) + \alpha_i + \varepsilon_{it},\ i = 1,\ldots,N,\ t = 1,\ldots,T.$$

For both cases, we propose test statistics, establish their theoretical properties, and investigate their finite sample properties via simulations.

In Section 4.3, we extend our results to the panel data model with two-way FEs. The model under consideration becomes

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

https://doi.org/10.1017/S026646662300018X Published online by Cambridge University Press

with an additional time FE λ_t added to model (2.1). The null hypothesis is still the *joint* test of *homogeneity* and *stability* of parameters in model:

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

against the alternative hypothesis $\mathbb{H}_1 : (\beta_{it}, f_{it}) \neq \left(\beta_{js}, 0\right)$ for some $(i,t) \neq (j,s)$. We propose a test statistic, discuss the examination of its limiting distributions, and demonstrate its finite sample properties via simulations.

4.1. Testing for the Stability of Heterogeneous Coefficients

In addition to the homogenous linear panel data model in $\mathbb{H}_0$, estimating a panel data model with heterogeneous constant slope coefficients (e.g., Hsiao and Pesaran, 2008) may be of interest. For this model, the hypothesis testing problem is

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

against the alternative hypothesis $\mathbb{H}_{s1} : (\beta_i(\cdot), f_i(\cdot)) \neq (\beta_i, 0)$ for some i's and all $\beta_i \in \mathbb{R}^d$. To examine the local power property of the proposed test, we consider the following local Pitman alternatives:

$$\mathbb{H}_{s1,\gamma_{NT}} : \beta_{it} = \beta_{0i} + \gamma_{NT}\Delta_{\beta,it} \text{ and } f_{it} = \gamma_{NT}\Delta_{f,it}, \tag{4.2}$$

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 $\Delta_{\beta,i}(\cdot)$ and $\Delta_{f,i}(\cdot)$ are nonzero continuous functions of the time regressor for some i's.

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

$$Y_{it} = X_{it}'\beta_i + \alpha_i + \varepsilon_{it}. \tag{4.3}$$

We note that β_i in (4.3) can be estimated by linear regression of Y_{it} on 1 and X_{it}, and the resulting estimators of β_i and g_{it} are, respectively, given by

$$\hat{\beta}_i = \left(X_i'M_TX_i\right)^{-1} X_i'M_TY_i \text{ and } \hat{g}_{it} = X_{it}'\hat{\beta}_i. \tag{4.4}$$

The augmented residuals are $\hat{u}_{it} = Y_{it} - \hat{g}_{it}$, for $t = 1, \ldots, T$. As in Section 2.2, we can run N auxiliary time series regressions and construct Γ_{NT} as (2.20). Define $\hat{\mathbb{B}}_{NT} = N^{-1/2}\sum_{i=1}^N \text{tr}(\hat{\mathbb{Q}}_i\hat{\Omega}_i)$ and $\hat{\mathbb{V}}_{NT} = 2N^{-1}\sum_{i=1}^N \text{tr}(\hat{\mathbb{Q}}_i\hat{\Omega}_i\hat{\mathbb{Q}}_i\hat{\Omega}_i)$. The test statistic for $\mathbb{H}_{s0}$ versus $\mathbb{H}_{s1}$ is given by

$$\hat{J}_{NT} = \left(N^{1/2}T\Gamma_{NT} - \hat{\mathbb{B}}_{NT}\right)/\sqrt{\hat{\mathbb{V}}_{NT}}. \tag{4.5}$$

Let $g_{\Delta,it} \equiv X_{it}'\Delta_{\beta,it} + \Delta_{f,it}$ and $g_{\Delta,i} = \left(g_{\Delta,i1}, \ldots, g_{\Delta,iT}\right)'$. Let $\bar{\beta}_{\Delta i} = \left[E\left(X_i'M_TX_i\right)\right]^{-1} \times E\left(X_i'M_Tg_{\Delta,i}\right)$ and $\bar{g}_{\Delta,it} = X_{it}'\bar{\beta}_{\Delta i}$. Then we define $\breve{g}_{\Delta,it} = g_{\Delta,it} - \bar{g}_{\Delta,it}$ and $\Phi_{\Delta,NT} \equiv \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \breve{g}_{\Delta,it}^2$.

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

https://doi.org/10.1017/S026646662300018X Published online by Cambridge University Press

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

THEOREM 4.1. (i) *Suppose that Assumptions 1 and 2 hold. As* $(N,T)\to\infty$, *under* $\mathbb{H}_{s0}$,

$$\hat{J}_{NT}\xrightarrow{d}N(0,1).$$

(ii) *Suppose that Assumptions 1–3 and 4* hold. As* $(N,T)\to\infty$,

$$\hat{J}_{NT}\xrightarrow{d}N(\Phi_{\Delta},1),$$

under $\mathbb{H}_{s1,\gamma_{NT}}$ *with* $\gamma_{NT}=O\left(N^{-1/4}T^{-1/2}\mathbb{V}_{NT}^{1/4}\right)$.

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

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

4.2. Testing for the Homogeneity of TVCs

Another natural specification for the panel data model assumes homogeneous TVCs (see, for example, Li et al., 2011; Chen and Huang, 2018). Then testing for the *homogeneity* of TVCs may be of interest. To be specific, we now consider testing the null hypothesis

$$\mathbb{H}_{h0}:(\beta_i(\cdot),f_i(\cdot))=(\beta_0(\cdot),f_0(\cdot))\ \text{for some}\ (\beta_0(\cdot),f_0(\cdot))\ \text{and all}\ i\text{'s} \tag{4.6}$$

against the alternative hypothesis $\mathbb{H}_{h1}:(\beta_i(\cdot),f_i(\cdot))\neq\left(\beta_j(\cdot),f_j(\cdot)\right)$ for some $i\neq j$. To conduct the local power analysis, we also consider the Pitman local alternatives

$$\mathbb{H}_{h1,\gamma_{NT}}:\beta_{it}=\beta_0(\tau_t)+\gamma_{NT}\Delta_{\beta,it},\ \text{and}\ f_{it}=f_0(\tau_t)+\gamma_{NT}\Delta_{f,it}, \tag{4.7}$$

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 $(\Delta'_{\beta,i}(\cdot),\Delta_{f,i}(\cdot))\neq(\Delta'_{\beta,j}(\cdot),\Delta_{f,j}(\cdot))$ for some $i\neq j$, $\Delta_{\beta,i}(\cdot)$ and $\Delta_{f,i}(\cdot)$ are all nonzero continuous functions of the time regressor.

When $\mathbb{H}_{h0}$ holds, model (2.1) reduces to

$$Y_{it}=X'_{it}\beta(\tau_t)+f(\tau_t)+\alpha_i+\varepsilon_{it}. \tag{4.8}$$

Noting that $\beta(\cdot)$ and $f(\cdot)$ are all unknown, as before, we consider the sieve estimation of the above model (4.8). Let $B_t^L\equiv B^L(\tau_t)$, $B_{-1,t}^L\equiv B_{-1}^L(\tau_t)$, and $Z_{it}^L\equiv(B_{-1,t}^{L\prime},\left(X_{it}\otimes B_t^L\right)')'$. Also, let $\Pi_f=\left(\Pi_{f,1},\ldots,\Pi_{f,L-1}\right)\in\mathbb{R}^{L-1}$ with $\Pi_{f,k}=$

https://doi.org/10.1017/S026646662300018X Published online by Cambridge University Press

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

$$f(\cdot) \approx B_{-1}^{L}(\cdot)'\,\Pi_f \text{ and } \beta_l(\cdot) \approx \Pi_{\beta,l} B^L(\cdot) \text{ for } l = 1, \ldots, d. \tag{4.9}$$

Denote $\Pi \equiv (\Pi_f', \text{vec}(\Pi_\beta)')'$, where $\Pi_\beta \equiv (\Pi_{\beta,1}, \ldots, \Pi_{\beta,d}) \in \mathbb{R}^{L\times d}$. Using the approximations in (4.9), 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

$$Y_{it} = Z_{it}^{L\prime}\Pi + \alpha_i + \varepsilon_{r,it}^{\dagger}, \tag{4.10}$$

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 sieve estimator for Π is

$$\hat{\Pi}_{FE} = \left(\sum_{i=1}^{N} Z_i^{L\prime} M_T Z_i^L\right)^{-1} \sum_{i=1}^{N} Z_i^{L\prime} M_T Y_i. \tag{4.11}$$

On the basis of (4.11), 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

$$\hat{f}(\cdot) = B_{-1}^{L}(\cdot)'\,\hat{\Pi}_f,\ \hat{\beta}(\cdot) = \hat{\Pi}_\beta B^L(\cdot), \text{ and } \hat{g}_{it} = Z_{it}^{L\prime}\hat{\Pi}_{FE}. \tag{4.12}$$

The augmented residuals are now given by $\hat{u}_{it} = Y_{it} - \hat{g}_{it}$. As in Section 2.2, we can run the auxiliary time series regressions and construct the test statistic Γ_{NT} as (2.20). On the basis of $\hat{\varepsilon}_{r,it} = \hat{u}_{it} - \bar{\hat{u}}_i$, we calculate $\hat{\mathbb{B}}_{NT}$ and $\hat{\mathbb{V}}_{NT}$ as (3.3). Then the feasible test statistic for (4.6) is given by

$$\hat{J}_{NT} = \left(N^{1/2} T \Gamma_{NT} - \hat{\mathbb{B}}_{NT}\right) / \sqrt{\hat{\mathbb{V}}_{NT}}. \tag{4.13}$$

Let $g_{\Delta,it} \equiv X_{it}'\Delta_{\beta,it} + \Delta_{f,it}$ and $g_{\Delta,i} = \left(g_{\Delta,i1}, \ldots, g_{\Delta,iT}\right)'$. Also, let $\bar{g}_{\Delta,it} = Z_{it}^{L\prime}\bar{\Delta}_\beta$, where $\bar{\Pi}_\Delta = \left[\sum_{i=1}^N E\left(\dot{Z}_i^{L\prime}\dot{Z}_i^L\right)\right]^{-1} \sum_{i=1}^N E\left(\dot{Z}_i^{L\prime} g_{\Delta,i}\right)$. We further define

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

Then, for $\hat{J}_{NT}$ in (4.13), we have the following theorem.

THEOREM 4.3. (i) *Suppose that Assumptions 1 and 2 and Assumptions 3* and 5 in Appendix D of the Supplementary Material hold. Then, under $\mathbb{H}_{h0}$, as $(N,T) \to \infty$,*

$$\hat{J}_{NT} \xrightarrow{d} N(0,1).$$

https://doi.org/10.1017/S026646662300018X Published online by Cambridge University Press

(ii) *Suppose that Assumptions 1 and 2 and Assumptions 3*, 4**, and 5 in Appendix D of the Supplementary Material hold. As* $(N,T) \to \infty$,

$$\hat{J}_{NT} \xrightarrow{d} N(\Phi_\Delta, 1),$$

under $\mathbb{H}_{h1,\gamma_{NT}}$ *with* $\gamma_{NT} = N^{-1/4}T^{-1/2}\mathbb{V}_{NT}^{1/4}$.

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

COROLLARY 4.4. *Suppose that Assumptions 1, 2, 4, and 5 and Assumption 3* in Appendix D of the Supplementary Material hold. As* $(N,T) \to \infty$, *under* $\mathbb{H}_{h1}$,

$$\mathbb{V}_{NT}^{1/2}N^{-1/2}T^{-1}\hat{J}_{NT} \xrightarrow{p} \Phi_\Delta.$$

The above result establishes that $\hat{J}_{NT}$ diverges to ∞ at rate $O_p\left(N^{1/2}T/K^{1/2}\right)$ under $\mathbb{H}_{h1}$, 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 $\Phi_\Delta > 0$.

4.3. Test Panel Data Models with Two-Way Fixed Effects

In empirical studies, two-way FEs are often used to capture the individual-specific and period-specific heterogeneity. Our method can be used to test the structure underlying the parameters in panel data models with two-way FEs. Now we consider the following model:

$$Y_{it} = X_{it}'\beta_{it} + \alpha_i + \lambda_t + \varepsilon_{it},\ i = 1,\dots,N, t = 1,\dots,T, \tag{4.14}$$

where λ_t is time effect. In addition, it is possible to incorporate heterogeneous smoothing time trends f_{it} with some additional identification restrictions on trending functions and time FEs.

The null hypothesis is the same *joint* test of *homogeneity* and *stability* of parameters as in model (4.14):

$$\mathbb{H}_{2W,0} : \beta_{it} = \beta_0 \text{ for some } \beta_0 \in \mathbb{R}^d \text{ and all } (i,t)\text{'s} \tag{4.15}$$

against the alternative hypothesis

$$\mathbb{H}_{2W,1} : \beta_{it} \neq \beta_{js} \text{ for some } (i,t) \neq (j,s). \tag{4.16}$$

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

$$\mathbb{H}_{2W,1,\gamma_{NT}} : \beta_{it} = \beta_0 + \gamma_{NT}\Delta_{\beta,it},$$

where $\gamma_{NT} \to 0$ as $(N,T) \to \infty$, $\Delta_{\beta,it} = \Delta_{\beta,i}(\tau_t)$, and $\Delta_{\beta,i}(\cdot) \neq \Delta_{\beta,j}(\cdot)$ for some $i \neq j$ with $\Delta_{\beta,i}(\cdot)$ being nonzero continuous functions of the time regressor.

Let $\ddot{a}_{it} = a_{it} - \bar{a}_i - \bar{a}_{\cdot t} + \bar{a}$, where $\bar{a}_{\cdot t} = N^{-1}\sum_{j=1}^{N} a_{jt}$ and $\bar{a} = (NT)^{-1}\sum_{j=1}^{N}\sum_{s=1}^{T} a_{js}$ for $a = X$ or Y. The two-way FE estimator for β_0 is given by

$$\hat{\beta}_{2WFE} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{X}_{it}\ddot{X}_{it}'\right)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{X}_{it}\ddot{Y}_{it}.$$

Similarly, we define the augmented residuals

$$\hat{u}_{it} = Y_{it} - X_{it}'\hat{\beta}_{2WFE} = X_{it}'\beta_{it}^{\dagger} + \eta_{it} + \alpha_i + \lambda_t + \varepsilon_{it},$$

where $\beta_{it}^{\dagger} = \beta_{it} - \beta$, $\beta_P = \left[\sum_{i=1}^{N}\sum_{t=1}^{T} E(\ddot{X}_{it}\ddot{X}_{it}')\right]^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} E\left(\ddot{X}_{it}\ddot{Y}_{it}\right)$ and $\eta_{it} = X_{it}'(\beta_P - \hat{\beta}_{2WFE})$. Under $\mathbb{H}_{2W,0}$, $\beta_{it} = \beta_P$ and $\hat{\beta}_{2WFE} = \beta_P + O_p[(NT)^{-1/2}]$, and these imply that $\hat{u}_{it} \approx \alpha_i + \lambda_t + \varepsilon_{it}$; under $\mathbb{H}_{2W,1}$, $\beta_{it}^{\dagger} = \beta_{it} - \beta_P$ is a nonzero function of the time regressor, and in general $g_{it}^{\dagger} = X_{it}'\beta_{it}^{\dagger} \neq 0$. Then we can construct a test statistic by

$$\Gamma_{NT}^{0} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} g_{it}^{\dagger 2}.$$

To obtain the quantity Γ_{NT}^{0}, we need the estimation of $g_{it}^{\dagger}$. Note that the augmented residual has two-way FEs:[10]

$$\hat{u}_{it} \approx X_{it}'\beta_{it}^{\dagger} + \alpha_i + \lambda_t + \varepsilon_{it}.$$

We cannot estimate $g_{it}^{\dagger}$ by N auxiliary time series regressions without modification because of the presence of time FEs. If λ_t can be treated as a function of time trends, we can still estimate $\beta_{it}^{\dagger}$ and λ_t using TVC time series regression as before. Without a smoothing structure on λ_t's, we can assume that $X_{it} = \mu_i^{(x)}(\tau_t) + V_{it}$ and $E(V_{it}) = 0$. Denote $\mu_{it}^{(x)} = \mu_i^{(x)}(\tau_t)$. Then the cross-sectional mean of X_{it} and g_{it} can be, respectively, written as

$$\bar{X}_{\cdot t} = \frac{1}{N}\sum_{i=1}^{N}\mu_{it}^{(x)} + \frac{1}{N}\sum_{i=1}^{N} V_{it} = \bar{\mu}^{(x)}(\tau_t) + O_p\left(N^{-1/2}\right) \text{ and}$$

$$\bar{g}_{\cdot t} = \frac{1}{N}\sum_{i=1}^{N}\mu_{it}^{(x)\prime}\beta_{it}^{\dagger} + \frac{1}{N}\sum_{i=1}^{N} V_{it}'\beta_{it}^{\dagger} = G(\tau_t) + O_p\left(N^{-1/2}\right),$$

where $\bar{\mu}^{(x)}(\tau_t) = \lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}\mu_{it}^{(x)}$ and $G(\tau_t) = \lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}\mu_{it}^{(x)\prime}\beta_{it}^{\dagger}$. One method is to use the cross-sectional demeaning to remove λ_t and obtain

[10] We can also consider the joint estimation of $\{g_{it}^{\dagger}\}$, $\{\alpha_i\}$, and $\{\lambda_t\}$ by minimizing the least-squares objective function under some identification restrictions (see Lu and Su, 2022). To unify our testing approach in this paper, we adopt the demeaned method to remove time FEs and then can run N time series regressions to estimate $\{g_{it}^{\dagger}\}$.

https://doi.org/10.1017/S026646662300018X Published online by Cambridge University Press

$$\begin{aligned}\hat{u}_{it}^{c} &= X_{it}^{c\prime}\beta_{it}^{\dagger} + \bar{X}_{\cdot t}^{\prime}\beta_{it}^{\dagger} - \bar{g}_{\cdot t} + \alpha_{i}^{c} + \varepsilon_{it}^{c} + \eta_{it}^{c} \\ &\approx X_{it}^{c\prime}\beta_{it}^{\dagger} + [\bar{\mu}^{(x)}(\tau_t)\beta_{i}^{\dagger}(\tau_t) - G(\tau_t)] + \alpha_{i}^{c} + \varepsilon_{it}^{c} \\ &\equiv X_{it}^{c\prime}\beta_{it}^{\dagger} + f_{it} + \alpha_{i}^{c} + \varepsilon_{it}^{c}, \text{ say,}\end{aligned}$$

where $a_{it}^{c} = a_{it} - N^{-1}\sum_{j=1}^{N} a_{jt}$ for $A = \hat{u}$, X, ε, and η, $\alpha_{i}^{c} = \alpha_{i} - N^{-1}\sum_{j=1}^{N}\alpha_{j}$. Clearly, as model (2.14), the above model has TVCs ($\beta_{it}^{\dagger}$) and smoothing time trends (f_{it}). Then we can still estimate $\beta_{i}^{\dagger}(\cdot)$ and f_{it} by running nonparametric regressions of $\hat{u}_{it}^{c}$ on X_{it}^{c} and 1 with TVCs as before. Let $Z_{it}^{c} \equiv X_{it}^{c} \otimes B_{t}$, $Z_{i}^{c} = \left(Z_{i1}^{c\prime}, \cdots, Z_{iT}^{c\prime}\right)^{\prime}$, and $\hat{u}_{i}^{c} = \left(\hat{u}_{i1}^{c}, \cdots, \hat{u}_{iT}^{c}\right)^{\prime}$. The linearized model is given by

$$\hat{u}_{it}^{c} = \text{vec}(\vartheta_{\beta,i})^{\prime} Z_{it}^{c} + B_{-1,t}^{\prime}\vartheta_{f,i} + \alpha_{i}^{c} + \varepsilon_{it}^{*},$$

where ε_{it}^{*} includes ε_{it}^{c} and the sieve approximation errors. The usual OLS estimator for sieve coefficients $\vartheta_{\beta,i}$ is given by $\text{vec}(\hat{\vartheta}_{\beta,i}) = \left(Z_{i}^{c\prime}M_{B}Z_{i}^{c}\right)^{-1} Z_{i}^{c\prime}M_{B}\hat{u}_{i}^{c}$, where $M_{B} = I_{T} - B\left(B^{\prime}B\right)^{-1}B^{\prime}$ with $B = (B_{1},\ldots,B_{T})^{\prime}$. Then $\hat{g}_{it}^{\dagger} = Z_{it}^{c\prime}\text{vec}(\hat{\vartheta}_{\beta,i})$ and we can estimate Γ_{NT}^{0} by

$$\Gamma_{NT} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{g}_{it}^{\dagger 2}.$$

Let $\hat{Q}_{i,1}^{c} = Z_{i}^{c\prime}M_{B}Z_{i}^{c}/T$, $\hat{Q}_{i,0}^{c} = Z_{i}^{c\prime}Z_{i}^{c}/T$, $\hat{\mathbb{Q}}_{i}^{c} = (\hat{Q}_{i,1}^{c})^{-1}\hat{Q}_{i,0}^{c}(\hat{Q}_{i,1}^{c})^{-1}$, and $\hat{\Omega}_{i}^{c} = T^{-1}\sum_{t=1}^{T}\breve{Z}_{it}^{c}\breve{Z}_{it}^{c\prime}\hat{\varepsilon}_{r,it}^{2}$, where $\hat{\varepsilon}_{r,it} = \hat{u}_{it} - \bar{\hat{u}}_{i} - \bar{\hat{u}}_{\cdot t} + \bar{\hat{u}}$ and $\breve{Z}_{it}^{c\prime}$ is the tth row of $M_{B}Z_{i}^{c}$. To obtain the asymptotic distribution of Γ_{NT}, we define

$$\hat{\mathbb{B}}_{NT}^{c} = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\text{tr}\left(\hat{\mathbb{Q}}_{i}^{c}\hat{\Omega}_{i}^{c}\right) \text{ and } \hat{\mathbb{V}}_{NT}^{c} = \frac{2}{N}\sum_{i=1}^{N}\text{tr}\left(\hat{\mathbb{Q}}_{i}^{c}\hat{\Omega}_{i}^{c}\hat{\mathbb{Q}}_{i}^{c}\hat{\Omega}_{i}^{c}\right).$$

Then a feasible test statistic for $\mathbb{H}_{2W,0}$ versus $\mathbb{H}_{2W,1}$ is given by

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

Let $g_{\Delta,it} \equiv X_{it}^{\prime}\Delta_{\beta,it}$ and $g_{\Delta,i} = \left(g_{\Delta,i1},\ldots,g_{\Delta,iT}\right)^{\prime}$. Also, let $\bar{g}_{\Delta,it} = X_{it}^{\prime}\bar{\Delta}_{\beta}$, where $\bar{\Delta}_{\beta} = \left[\sum_{i=1}^{N}\sum_{t=1}^{T}E(\ddot{X}_{it}\ddot{X}_{it}^{\prime})\right]^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}E(\ddot{X}_{it}\ddot{g}_{\Delta,it})$. We further define $\breve{g}_{\Delta,it} = g_{\Delta,it} - \bar{g}_{\Delta,it}$ and $\Phi_{\Delta,NT} \equiv \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\breve{g}_{\Delta,it}^{2}$. We can follow Section 3 to establish the large sample properties for the above test statistic under some suitable conditions. Here, we only provide the main results and leave the rigorous justification for future work. As $(N,T)\to\infty$, we have

(i) $\hat{J}_{NT} \xrightarrow{d} N(0,1)$ under $\mathbb{H}_{2W,0}$;

(ii) $\hat{J}_{NT} \xrightarrow{d} N(\Phi_{\Delta},1)$ under $\mathbb{H}_{2W,1,\gamma_{NT}}$ with $\gamma_{NT} = N^{-1/4}T^{-1/2}\left(\mathbb{V}_{NT}^{c}\right)^{1/4}$, where $\mathbb{V}_{NT}^{c}$ is the population version of $\hat{\mathbb{V}}_{NT}^{c}$, and $\Phi_{\Delta} = \text{plim}_{(N,T)\to\infty}\Phi_{\Delta,NT}$;

(iii) $\left(\mathbb{V}_{NT}^{c}\right)^{1/2}N^{-1/2}T^{-1}\hat{J}_{NT} \xrightarrow{p} \Phi_{\Delta}$ under $\mathbb{H}_{2W,1}$, where $\gamma_{NT} = 1$.

https://doi.org/10.1017/S026646662300018X Published online by Cambridge University Press

5. MONTE CARLO SIMULATIONS

5.1. Simulations for Testing $\mathbb{H}_0$ versus $\mathbb{H}_1$

In this section, we conduct a set of Monte Carlo simulations to evaluate the finite sample performance of our proposed joint test for homogeneity and stability of coefficients. We consider the following six data generating processes (DGPs):
DGP 1. Homogeneous constant coefficients: $Y_{it} = 2X_{it} + \alpha_i + \varepsilon_{it}$.
DGP 2. Homogeneous TVCs: $Y_{it} = f_0(\tau_t) + \beta_0(\tau_t) X_{it} + \alpha_i + \varepsilon_{it}$.
DGP 3. Heterogeneous constant coefficients: $Y_{it} = \beta_i X_{it} + \alpha_i + \varepsilon_{it}$, where $\beta_i \sim$i.i.d. $U[0.7, 1.3]$.
DGP 4. Fully heterogeneous TVCs: $Y_{it} = \delta_{1i} f_0(\tau_t) + \delta_{2i}\beta_0(\tau_t) X_{it} + \alpha_i + \varepsilon_{it}$, where $\delta_{1i} \sim$i.i.d. $U[0.5, 1.5]$ and $\delta_{2i} \sim$i.i.d. $U[-0.5, 0.5]$.
DGP 5. Grouped heterogeneous TVCs:

$$Y_{it} = \begin{cases} 0.5 f_0(\tau_t) + 0.5\beta_0(\tau_t) X_{it} + \alpha_i + \varepsilon_{it}, & i = 1, \ldots, \lceil N/3 \rceil, \\ 0.75 f_0(\tau_t) + 0.75\beta_0(\tau_t) X_{it} + \alpha_i + \varepsilon_{it}, & i = \lceil N/3 \rceil + 1, \ldots, \lceil 2N/3 \rceil, \\ f_0(\tau_t) + \beta_0(\tau_t) X_{it} + \alpha_i + \varepsilon_{it}, & i = \lceil 2N/3 \rceil + 1, \ldots, N. \end{cases}$$

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

$$Y_{it} = \begin{cases} 0.25 X_{it} + \alpha_i + \varepsilon_{it}, & t < T/2, \\ -0.25 X_{it} + \alpha_i + \varepsilon_{it}, & t \geq T/2. \end{cases}$$

Among all DGPs, the FEs α_i's follow i.i.d. $N(0,1)$, the regressors X_{it}'s are generated according to

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

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

$$f_0(\upsilon) = \upsilon^2 - \upsilon + 1/6 \text{ and } \beta_0(\upsilon) = \frac{\exp[(\upsilon - 0.5)/0.4]}{1 + \exp[(\upsilon - 0.5)/0.4]},$$

which are used to generate the smooth trend functions and TVC functions. A similar function form for $\beta_0(\cdot)$ is adopted in Su et al. (2019).

DGP 1 is for size study, and the other five 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. We choose K following the discussion in Remark 4. Note that $d = 1$ (we estimate $\beta_i(\cdot)$ and $f_i(\cdot)$ additively), and we assume that $\kappa \geq 2$. In theory, we should set $K \propto T^{1/5}$ taking care of the least smooth functions we assumed (i.e., $\kappa = 2$) to minimize the RMSE. For practice, we propose to adopt the data-

https://doi.org/10.1017/S026646662300018X Published online by Cambridge University Press

Table 1. Simulation results for DGPs 1–6 using K from LOOCV

		DGP 1			DGP 2			DGP 3		
T	N	10%	5%	1%	10%	5%	1%	10%	5%	1%
25	25	0.104	0.057	0.012	0.533	0.370	0.143	0.721	0.602	0.369
	50	0.108	0.060	0.017	0.724	0.556	0.267	0.859	0.775	0.540
50	25	0.099	0.052	0.011	0.978	0.945	0.766	0.985	0.966	0.866
	50	0.116	0.061	0.014	1.000	0.998	0.976	1.000	0.998	0.988
		DGP 4			DGP 5			DGP 6		
		10%	5%	1%	10%	5%	1%	10%	5%	1%
25	25	0.658	0.537	0.306	0.639	0.481	0.232	0.668	0.496	0.246
	50	0.819	0.710	0.474	0.808	0.680	0.421	0.840	0.715	0.412
50	25	0.978	0.941	0.813	0.988	0.957	0.835	0.997	0.991	0.927
	50	0.999	0.998	0.975	1.000	0.998	0.988	1.000	1.000	0.996

Note: DGP 1 is for the size study, and DGPs 2–6 are for the power comparison.

driven K_{cv} chosen by the LOOCV method.[11] This choice of K_{cv} works well for all DGPs and is recommended for practice. To investigate the sensitivity of our test to different choices of number of basis functions, we consider a sequence of numbers $K_c = \lceil CT^{1/5} \rceil$ with $C = 1, 1.5, 2$.

Different combinations of sample sizes are used: $T = 25, 50$ and $N = 25, 50$. For each combination of sample sizes, the number of replications is 1,000 times. For the bootstrap, we consider 299 resamples for size studies and power comparisons.

The simulation results for the joint test of homogeneity and stability in DGPs 1–6 with the proposed K_{cv} are summarized in Table 1. The results using $K_c = \lceil CT^{1/5} \rceil$ for the size study (DGP 1) and the power study (DGPs 2–6) are reported in Table 2 and Table 5 in Appendix F of the Supplementary Material, respectively. We summarize the results as follows. 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. This fits our intuition because our size results hold even under fixed K, as discussed in Remark 4. Second, the proposed test has good power for DGPs 2–6: (i) in general, the test is more powerful or stable when using K_{cv} (as discussed in Remark 4), but the power can be sensitive to K_c (see Table 5 in Appendix F of the Supplementary Material); (ii) for all six DGPs, the empirical power tends to 1 as either N or T increases, and has a larger speed when T increases than when N increases, which confirms that $\hat{J}_{NT}$ diverges to infinity faster as T increases than N increases under $\mathbb{H}_1$ as shown in Corollary 3.3; (iii) the empirical powers for DGP 6 are close to 1

[11] $K_{cv} = \operatorname{argmin}_{K \in [1, K_{\max}]} \sum_{i=1}^{N} \sum_{t=1}^{T} (\hat{u}_{it} - \hat{g}^{\dagger}_{i(-t)}(K) - \hat{\alpha}_{i(-t)}(K))^2$ where $\hat{g}^{\dagger}_{i(-t)}(K)$ and $\hat{\alpha}_{i(-t)}(K)$ come from the ith auxiliary regression of $\hat{u}_{it}$ on X_{it} with TVCs and trends without using the tth observation and K or $K-1$ basis functions are adopted in the sieve approximations.

https://doi.org/10.1017/S026646662300018X Published online by Cambridge University Press

TABLE 2. Size sensitivity studies: DGP 1 using different K

		K_1			K_2			K_3		
T	N	10%	5%	1%	10%	5%	1%	10%	5%	1%
25	25	0.099	0.057	0.019	0.092	0.048	0.018	0.114	0.060	0.010
	50	0.108	0.052	0.023	0.109	0.062	0.020	0.090	0.044	0.012
50	25	0.125	0.055	0.009	0.107	0.051	0.007	0.099	0.043	0.012
	50	0.102	0.034	0.007	0.097	0.053	0.008	0.108	0.053	0.011

Note: $K_1 = \lceil T^{1/5} \rceil$, $K_2 = \lceil 1.5T^{1/5} \rceil$, and $K_3 = \lceil 2T^{1/5} \rceil$.

when $T = 50$, where the parameters are homogeneous but have jumps along time, although Corollary 3.3 does not cover the case with jump in parameters along time. A final remark is that the size results using normal critical values are worse than they are using critical values based on the bootstrap (for details, see Table 4 in Appendix F of the Supplementary Material). Overall, we can observe that our proposed test statistic performs very well in all scenarios in simulations.

5.2. Simulations for Extensions in Section 4

We consider the same DGPs as those in testing $\mathbb{H}_0$ for the tests $\mathbb{H}_{s0}$ versus $\mathbb{H}_{s1}$ and $\mathbb{H}_{h0}$ versus $\mathbb{H}_{h1}$ in Section 4. For testing the stability of heterogeneous coefficients, DGPs 1 and 3 are for size study and DGPs 2 and 4–6 are for power analysis. For testing the homogeneity of time-varying coefficients, DGPs 1, 2, and 6 are for size study and DGPs 3–5 are for power comparison. Note that DGP 6 satisfies $\mathbb{H}_{h0}$, but they come with nonsmooth $\beta_0(\cdot)$ and $f_0(\cdot)$. So we expect some size distortion for DGP 6 in testing $\mathbb{H}_{h0}$ against $\mathbb{H}_{h1}$ because the sieve approximation errors are not asymptotically negligible.

For the joint test in the presence of both the individual and the time FEs in Section 4.3, we add time FEs λ_t and remove time trending functions in all DGPs. For example, DGPs 1 and 2 become

$$Y_{it} = 2X_{it} + \alpha_i + \lambda_t + \varepsilon_{it} \text{ and } Y_{it} = \beta_0(\tau_t) X_{it} + \alpha_i + \varepsilon_{it},$$

respectively. Other DGPs are modified similarly. The distribution of individual FEs α_i's is changed to i.i.d. $N(0, 1/\sqrt{2})$. The distribution of time FEs λ_t's is i.i.d. $N(0, 1/\sqrt{2})$. The regressors X_{it}'s are changed to

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

with $\varepsilon_{x,it}$ ~i.i.d. $N(0,1)$ and μ_i ~i.i.d. $U[0.05, 0.1]$, and the error ε_{it}'s are conditional heteroskedastic as $\varepsilon_{it} = \sqrt{0.05X_{it}^2 + 0.5}\epsilon_{it}$ with ϵ_{it} ~i.i.d. $N(0,1)$. $\alpha_i, \lambda_t, \varepsilon_{x,it}, \mu_i$, and ϵ_{it} are independent of each other. As before, DGP 1 is for size study, and DGPs 2–6 are for power study.

https://doi.org/10.1017/S026646662300018X Published online by Cambridge University Press

We report the results in Tables 6–8 in Appendix F of the Supplementary Material to save space. The tuning parameter is set to K_{cv}, chosen by the LOOCV. We also study the sensitivity of the tuning parameters by setting $K_c = \lceil CT^{1/5} \rceil$ with $C = 1, 1.5, 2$. The number of replications and the number of resamples are the same as in the last section. From Tables 5–7 in Appendix F of the Supplementary Material, we can see that our tests perform rather well in terms of size results, and are insensitive to the choice of K. Of course, we see some size distortions for DGP 6 when testing $\mathbb{H}_{h0}$ versus $\mathbb{H}_{h1}$. The reason is that β and f are nonsmooth in this DGP and do not satisfy the smoothness conditions required for our test. Our tests also perform very well in terms of power, especially when using K_{cv}, but power can be sensitive to the choice of K. To summarize, our extended tests perform reasonably well in small samples.

6. APPLICATION TO ENVIRONMENTAL KUZNETS CURVE

In this section, we apply our proposed test to study the EKC for the data of emissions published in the U.S. Environmental Protection Agency's *National Air Pollutant Emission Trends, 1900–1994*. We are mainly interested in testing the validity of homogeneity and stability restrictions in the panel data model, which is widely used for EKC estimation.

The EKC hypothesis dates back to the seminal work of Grossmann and Krueger (1993, 1995) and becomes popular at the World Bank. Both the theoretical literature and the empirical literature on the topic is voluminous and continues to grow, and so do the controversial findings. Many empirical studies 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), who use a semiparametric partially linear panel model to fit the data, and reject the parametric specification. 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. Note that the test in Li et al. (2016) assumes homogeneity, and it might suffer from the possibility that the coefficients are heterogeneous across individuals. To alleviate this problem, we reinvestigate the parametric specification of EKC using our proposed test.

We consider the following regression model:

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

where $i = 1, \ldots, N$, $t = 1, \ldots, T$, $\ln Pol_{it}$ is the logarithm of pollutant emission of sulfur dioxide (SO_2) measured by metric tons per capita, $\ln Inc_{it}$ represents the logarithm of income for state i at time t, α_i is the unobserved state-specific FE; $\beta_{1,it}$ and $\beta_{2,it}$ are time-varying slope coefficients for the ith individual, and f_{it} is the heterogeneous time trend. Presumably, the time trend f_{it} is related to pollution

TABLE 3. Bootstrap p-values for three tests

Tests	$C=1$	$C=1.5$	$C=2$	K_{cv}
$\mathbb{H}_0$ vs. $\mathbb{H}_1$	0.000	0.027	0.048	0.009
$\mathbb{H}_{s0}$ vs. $\mathbb{H}_{s1}$	0.000	0.000	0.000	0.001
$\mathbb{H}_{h0}$ vs. $\mathbb{H}_{h1}$	0.093	0.175	0.212	0.258

emission across countries. We test the homogeneity and stability of $(\beta_{1,it}, \beta_{2,it}, f_{it})$ jointly. The data used in our paper are from Millimet et al. (2003), which includes 48 states ($N=48$) and ranges from year 1929 to year 1994 ($T=66$). We transform the metric ton measurement for SO_2 emission into kilograms to achieve variables of magnitude comparable to those of 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

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

The estimation and testing procedures are similar to those discussed in Section 2. The FE estimation of model (6.2) gives us the following:

$$\hat{\beta}_1 = 9.5706^{***} (0.4358) \text{ and } \hat{\beta}_2 = -0.5608^{***} (0.0247), \qquad \textbf{(6.3)}$$

where the standard error is reported in parentheses. The estimators for β_1 and β_2 are both significant at the 1% significance level, and we obtain 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 TVCs and smoothing time trends. For the sieve approximation of unknown functions, we adopt the cosine functions as the bases and consider a sequence of numbers for different functions. Consistently with the setting in the simulations, we set $K_1 = K_2 = \lfloor C \cdot 66^{1/5} \rfloor$ and $K_3 = \max\{2, K_1 - 1\}$, with $C = 1, 1.5, 2$, where K_1, K_2, and K_3 are the number of sieve terms for $\beta_{1i}(\cdot)$, $\beta_{2i}(\cdot)$, and $f_i(\cdot)$, respectively. We also consider K_{cv} by the LOOCV. In search of the K_{cv}, we set the minimum of K_{cv} as 2 and maximum of K_{cv} as 6, considering that T is only 66. We report the bootstrap p-values with 2,000 bootstrap resamples.

The results for testing homogeneity and stability are reported in the second row of Table 3. We can find that almost all the p-values are smaller than 0.05, which provides strong evidence for rejecting homogeneity and stability restrictions on parameters in model (6.1). We also apply our tests for $\mathbb{H}_{s0}$ and $\mathbb{H}_{h0}$ developed in Section 4, and we report the results in the third and fourth rows of Table 3. $\mathbb{H}_{s0}$ is rejected at the 5% level for all K's, whereas $\mathbb{H}_{h0}$ is not rejected at the 5% level for all K's. With the evidence reported in Table 3, we conclude that the model with homogeneous but TVCs are more appropriate for this application.

https://doi.org/10.1017/S026646662300018X Published online by Cambridge University Press

Based on the test results, we conduct the estimation under $\mathbb{H}_{h0}$, where

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

We report the estimates with 95% confidence intervals (CIs) in Figure 1 in Appendix F of the Supplementary Material. The 95% CIs are obtained using the fixed-regressor bootstrap procedure. We only report the results using L_{cv} obtained by LOOCV, and the results using other numbers of sieve terms are very similar. We can see that the averages of estimated $\hat{\beta}_1(\cdot)$ and $\hat{\beta}_2(\cdot)$ over t are indeed consistent with the FE estimates in equation (6.3), but the estimates differ over time. Noticeably, one of the estimates of β's at $T = 28$ exhibits different signs. We run the regression of the demeaned Y on the demeaned X for $T = 28$, and find that $\hat{\beta}_1$ and $\hat{\beta}_2$ are of the opposite signs of the FE estimates and are significant.[12]

7. CONCLUSION

This paper provides a nonparametric test of homogeneity and coefficient stability in panel data models. We establish the theoretical properties for the test and generalize it to test homogeneity across individuals and stability over time in large panel data models with one-way or two-way FEs. We suggest using bootstrap p-values for better finite sample performance. Through simulation, we have demonstrated that the proposed tests have excellent finite sample properties in various designs. In addition, we have illustrated the usefulness of the tests in analyzing the EKC. It should be noted that we impose cross-sectional independence of errors only for convenience. Our approach could accommodate cross-sectional dependence straightforwardly with more laborious derivations and some extra conditions. Furthermore, our tests may have low power under some sparse alternative hypotheses, as they are based on the average of squared fitted values. Finally, we draw readers' attention to the fact that while the number of sieve terms by LOOCV works well in simulations and applications, it lacks theoretical justification and is not generally optimal for hypothesis testing. We leave these topics—constructing more powerful tests for detecting sparse alternatives and the optimal number of sieve terms in test—as future research.

APPENDIX

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

Notation. Given sequences $\{a_n\}$ and $\{b_n\}$, let $a_n \lesssim b_n$ denote that 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, $a_n \lesssim b_n$ denotes that a_n/b_n is stochastically bounded, and $a_n \asymp b_n$ means 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 $p \geq 1$. To simplify the notation, we use $\max_i$, $\max_t$, and $\max_{i,t}$ to

[12]The cross-sectional estimates for $T = 28$ are $\hat{\beta}_1 = -39.264$ (13.905) and $\hat{\beta}_2 = 2.281$ (0.764).

https://doi.org/10.1017/S026646662300018X Published online by Cambridge University Press

denote $\max_{1\le i\le N}$, $\max_{1\le t\le T}$, and $\max_{1\le i\le N,\,1\le t\le T}$, respectively; $\min_i$, $\min_t$, and $\min_{i,t}$ are defined similarly.

A.1. Some Facts and Lemmas

We first state some facts related to basis functions and several technical lemmas that are used in the proofs of the main results in Section 3. The proofs of these lemmas are given in the Supplementary Material.

Note that we use the cosine 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 $f_i(\cdot)$ and $\beta_i^{\dagger}(\cdot)$ in the auxiliary regressions, respectively. Recall that $B_t=B^K(\tau_t)$, $B_{-1,t}=B_{-1}^K(\tau_t)$, $Z_{it}=(B_{-1,t}',X_{it}'\otimes B_t')'$, $Z_i=\left(Z_{i1},\ldots,Z_{i,T}\right)$, $\hat{Q}_{i,zz}=Z_iZ_i'/T$, $Q_{i,zz}=E(\hat{Q}_{i,zz})$, $\dot{Z}_{it}=Z_{it}-\bar{Z}_i$, $\dot{Z}_i=M_TZ_i$, $\hat{Q}_{i,\dot{z}\dot{z}}=\dot{Z}_i\dot{Z}_i'/T$, $\mathring{Z}_{it}=Z_{it}-E\left(\bar{Z}_i\right)$,$\mathring{Z}_i=(\mathring{Z}_{i1},\ldots,\mathring{Z}_{iT})'$, and $Q_{i,\dot{z}\dot{z}}=(E\hat{Q}_{i,\dot{z}\dot{z}})$. Denote $\mathcal{K}_i\equiv\dot{Z}_i\hat{Q}_{i,\dot{z}\dot{z}}^{-1}\hat{Q}_{i,zz}\hat{Q}_{i,\dot{z}\dot{z}}^{-1}\dot{Z}_i'=\dot{Z}_i\hat{\mathbb{Q}}_i\dot{Z}_i'$ and $\mathring{\mathcal{K}}_i\equiv\mathring{Z}_iQ_{i,\dot{z}\dot{z}}^{-1}Q_{i,zz}Q_{i,\dot{z}\dot{z}}^{-1}\mathring{Z}_i'=\mathring{Z}_i\mathbb{Q}_i\mathring{Z}_i'$. We have the following facts.

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

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

(iii) $\|Z_{it}\|^2=\left\|B_{-1,t}\right\|^2+\|X_{it}\|^2\|B_t\|^2\le C_{B,K}||\tilde{X}_{it}||^2$ for all i,t, where $\tilde{X}_{it}\equiv\left(1,X_{it}'\right)'$;

(iv) $\|E(Z_{it})\|^2=\left\|B_{-1,t}\right\|^2+\|E(X_{it})\|^2\|B_t\|^2\le C_{B,K}(1+\|E(X_{it})\|^2)\le C_XC_{B,K}$ for all i,t, where $C_X\equiv 1+\max_{i,t}(\|E(X_{it})\|^2)$;

(v) $||\dot{Z}_{it}||^2\le 2(\|Z_{it}\|^2+\left\|\bar{Z}_i\right\|^2)\le 2(\|Z_{it}\|^2+T^{-1}\sum_{s=1}^{T}\|Z_{is}\|^2)\le 2C_{B,K}A_{it}$ for all i,t, where $A_{it}\equiv||\tilde{X}_{it}||^2+T^{-1}\sum_{s=1}^{T}||\tilde{X}_{is}||^2$;

(vi) $||\mathring{Z}_{it}||^2\le 2(\|Z_{it}\|^2+\left\|E\left(\bar{Z}_i\right)\right\|^2)\le 2C_{B,K}(||\tilde{X}_{it}||^2+C_X)\le 2C_{B,K}\mathring{A}_{it}$ for all i,t, where $\mathring{A}_{it}\equiv||\tilde{X}_{it}||^2+C_X$;

(vii) $|\mathcal{K}_{i,ts}|=|\dot{Z}_{it}'\hat{\mathbb{Q}}_i\dot{Z}_{is}|\le\lambda_{\max}(\hat{\mathbb{Q}}_i)||\dot{Z}_{it}||||\dot{Z}_{is}||\le\lambda_{\max}(\hat{\mathbb{Q}}_i)2C_{B,K}A_{it}^{1/2}A_{is}^{1/2}$ for all i,t,s.

Next, we give some lemmas. The first two lemmas are for cosine basis functions, and their proofs are similar to Lemmas A1 and A2 in Su et al. (2019), where splines are used as basis functions. Lemmas A.1.3–A.1.7 are some intermediate results needed in the proofs of our main theorems. Lemma A.8 is Theorem 4.1 in Shao and Yu (1996), which is the main tool for showing Lemma A.9. We refer readers to the original paper for the detailed proof. Note that we apply results in the second part of Lemma A.8 with a stronger mixing condition in Assumption 1. Lemma A.9 gives a set of low-level sufficient conditions to ensure that $0<\min_i\lambda_{\min}\left(\hat{\Omega}_i\right)\le\max_i\lambda_{\max}\left(\hat{\Omega}_i\right)<\infty$ holds uniformly after some large T almost surely. This condition is used in the proof of asymptotic validity of bootstrap procedure of our test.

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* $\|\mathbf{g}\|_i^2\equiv E\{T^{-1}\sum_{t=1}^{T}[\mathbf{g}(\tau_t)'\tilde{X}_{it}][\tilde{X}_{it}'\mathbf{g}(\tau_t)]\}=\sum_{l=0}^{d}\|g_l\|_2^2\asymp\|\theta\|^2$, *where* $\tilde{X}_{it}=\left(1,X_{it}'\right)'$ *and* $\theta=\left(\theta_0',\theta_1',\ldots,\theta_d'\right)'$.

LEMMA A.2. *Suppose that Assumptions 1 and 2 hold. Let* $\mathcal{G}\equiv\{g(\cdot)=\theta'B^K(\cdot):\theta\in\mathbb{R}^K\}$. *Let* $\mathcal{G}_{-1}\times\mathcal{G}^{\otimes d}$ *denote the function space of* $\mathbf{g}=(g_0,g_1,\ldots,g_d)'$ *with* $g_0\in\mathcal{G}_{-1}$ *and* $g_l\in\mathcal{G}$ *for* $l=1,\ldots,d$. *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. *Suppose that Assumption 1 holds. Then:*

(i) $\max_i ||\hat{Q}_{i,\dot{z}\dot{z}}-Q_{i,\dot{z}\dot{z}}||=O_p[K(T/\ln N)^{-1/2}]$;
(ii) $\max_i ||\hat{Q}_{i,zz}-Q_{i,zz}||=O_p[K(T/\ln N)^{-1/2}]$;
(iii) $\max_i ||\hat{\mathbb{Q}}_i-\mathbb{Q}_i||=O_p[K(T/\ln N)^{-1/2}]$;
(iv) $\max_i ||\hat{Q}_{i,\dot{z}\varepsilon}||=O_p[K^{1/2}(T/\ln N)^{-1/2}]$;
(v) $\max_i ||\mathring{\Omega}_i-\Omega_i||=O_p[K(T/\ln N)^{-1/2}]$;
(vi) $\max_i ||\dot{\Omega}_i-\Omega_i||=O_p[K(T/\ln N)^{-1/2}]$,

where $\hat{Q}_{i,\dot{z}\varepsilon}=T^{-1}Z_i'M_T\varepsilon_i$, $\mathring{\Omega}_i=T^{-1}\sum_{t=1}^{T}\mathring{Z}_{it}\mathring{Z}_{it}'\varepsilon_{it}^2$, *and* $\dot{\Omega}_i=T^{-1}\sum_{t=1}^{T}\dot{Z}_{it}\dot{Z}_{it}'\varepsilon_{it}^2$.

LEMMA A.4. *Suppose that Assumptions 1 and 2 hold. Then:*

(i) $P\left(\underline{c}\le\min_i\lambda_{\min}(\hat{Q}_{i,\dot{z}\dot{z}})\le\max_i\lambda_{\max}(\hat{Q}_{i,\dot{z}\dot{z}})\le\bar{c}\right)=1-o\left(N^{-1}\right)$;
(ii) $P\left(\underline{c}\le\min_i\lambda_{\min}(\hat{Q}_{i,zz})\le\max_i\lambda_{\max}(\hat{Q}_{i,zz})\le\bar{c}\right)=1-o\left(N^{-1}\right)$;
(iii) $P\left(\underline{c}\le\min_i\lambda_{\min}(\hat{\mathbb{Q}}_i)\le\max_i\lambda_{\max}(\hat{\mathbb{Q}}_i)\le\bar{c}\right)=1-o\left(N^{-1}\right)$;
(iv) $P(\underline{c}\le\min_i\lambda_{\min}(\mathring{\Omega}_i)\le\max_i\lambda_{\max}(\mathring{\Omega}_i)\le\bar{c})=1-o\left(N^{-1}\right)$,

where $\underline{c}$ *and* $\bar{c}$ *are some finite positive constants.*

LEMMA A.5. *Suppose Assumptions 1–3 hold. Then* $(NT)^{-1}\sum_{i=1}^{N}\left\|r_{\Delta,i}\right\|^2=O\left(K^{-2\kappa}\right)$.

LEMMA A.6. *Suppose Assumptions 1–3 hold. Then (i)* $\mathbb{V}_{NT}\asymp K$ *and (ii)* $\mathbb{B}_{NT}\asymp N^{1/2}K$.

LEMMA A.7. *Suppose Assumptions 1–3 hold. Then (i)* $\frac{1}{N^{1/2}T}\sum_{i=1}^{N}\varepsilon_i'\left(\mathcal{K}_i-\mathring{\mathcal{K}}_i\right)\varepsilon_i=o_p\left(K^{1/2}\right)$ *and (ii)* $\frac{1}{N^{1/2}T}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\mathcal{K}_{i,tt}-\mathring{\mathcal{K}}_{i,tt}\right)\varepsilon_{it}^2=o_p\left(K^{1/2}\right)$.

LEMMA A.8. *Let* $r>2$, $\delta>0$, $2<v\le r+\delta$, *and* $\{X_t,t\ge 1\}$ *be an* α*-mixing sequence of random variables with* $E(X_t)=0$ *and* $\|X_t\|_{r+\delta}<\infty$. *Assume that* $\alpha(l)\le Cl^{-\theta}$, *for some* $C>0$ *and* $\theta>0$. *Then, for any* $\epsilon>0$, *there exists* $C^*<\infty$ *such that*

$$E\left|\sum_{t=1}^{T}X_t\right|^r\le C^*\left[(TC_T)^{r/2}\max_{1\le t\le T}\left(\|X_t\|_v\right)^r+T^{(r-\delta\theta/(r+\delta))\vee(1+\epsilon)}\max_{1\le t\le T}\left(\|X_t\|_{r+\delta}\right)^r\right],$$

where $C_T=\left[\sum_{t=1}^{T}(t+1)^{2/(v-2)}\alpha(t)\right]^{(v-2)/v}$. *In particular, for any* $\epsilon>0$,

$$E\left|\sum_{t=1}^{T}X_t\right|^r\le C^*\left[T^{r/2}\max_{1\le t\le T}\left(\|X_t\|_v\right)^r+T^{1+\epsilon}\max_{1\le t\le T}\left(\|X_t\|_{r+\delta}\right)^r\right]$$

if $\theta > v/(v-2)$ *and* $\theta \geq (r-1)(r+\delta)/\delta$, *and*

$$E\left|\sum_{t=1}^{T} X_t\right|^r \leq C^* T^{r/2} \max_{1\leq t\leq T} \left(\|X_t\|_{r+\delta}\right)^r,$$

if $\theta \geq r(r+\delta)/(2\delta)$.

LEMMA A.9. *Suppose Assumptions 1 and 2 hold. Further, assume that* $NK^{6+9\eta/40}(\ln T)^2/T^{1+9\eta/80} = O(1)$, *and* $\beta_i(\cdot)$ *and* $f_i(\cdot)$ *are uniformly bounded on* $[0,1]$ *for all i's. Then*

$$0 < \min_i \lambda_{\min}(\hat{\Omega}_i) \leq \max_i \lambda_{\max}(\hat{\Omega}_i) < \infty$$

holds after some large T almost surely, where $\hat{\Omega}_i = T^{-1}\sum_{t=1}^{T} \dot{Z}_{it}\dot{Z}_{it}'\hat{\varepsilon}_{r,it}^2$.

A.2. Proofs of the Main Results in Section 3

Proof of Theorem 3.1. The limiting distribution of $\hat{J}_{NT}$ under $\mathbb{H}_0$ is a special case of Theorem 3.2 with $\Delta_{\beta,i}(\cdot) = 0$ and $\Delta_{f,i}(\cdot) = 0$ 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 residual $\hat{u}_{it}$ under $\mathbb{H}_{1,\gamma_{NT}}$. Recall that $\bar{\Delta}_\beta = \left[\sum_{i=1}^{N} E\left(X_i'M_TX_i\right)\right]^{-1}\sum_{i=1}^{N} E\left(X_i'M_Tg_{\Delta,i}\right)$. Let $\nu_{\Delta,NT} \equiv \left(\sum_{i=1}^{N} X_i'M_TX_i\right)^{-1}\sum_{i=1}^{N} X_i'M_Tg_{\Delta,i} - \bar{\Delta}_\beta$ and $\nu_{NT} \equiv \left(\sum_{i=1}^{N} X_i'M_TX_i\right)^{-1}\sum_{i=1}^{N} X_i'M_T\varepsilon_i$. Under $\mathbb{H}_{1,\gamma_{NT}}$, we can rewrite $\beta_P = \beta_0 + \gamma_{NT}\bar{\Delta}_\beta$ by the definition of β_P (see Section 2.2). 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 $g_{\Delta,it} - \bar{g}_{\Delta,it} = X_{it}'(\beta_{it} - \beta_P) + \gamma_{NT}\Delta_{f,it} = \gamma_{NT}\breve{g}_{\Delta,it}$, where $\breve{g}_{\Delta,it} \equiv X_{it}'\Delta_{\beta,it}^c + \Delta_{f,it}$. Then $\hat{u}_{it} = \gamma_{NT}\breve{g}_{\Delta,it} - X_{it}'\breve{\nu}_{NT} + \alpha_i + \varepsilon_{it}$ 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), we have

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

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:

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

https://doi.org/10.1017/S026646662300018X Published online by Cambridge University Press

where $J_{NT} \equiv (N^{1/2}T\Gamma_{NT}^{(1)} - \mathbb{B}_{NT})/\mathbb{V}_{NT}^{1/2}$ and $J_{NT}^{(l)} \equiv N^{1/2}T\Gamma_{NT}^{(l)}/\mathbb{V}_{NT}^{1/2}$, for $l = 2,\ldots,6$. We complete the proof by showing that, as $(N,T) \to \infty$: (i) $J_{NT} \overset{d}{\to} N(0,1)$; (ii) $J_{NT}^{(2)} = \Phi_\Delta + o_p(1)$; (iii) $J_{NT}^{(l)} \equiv N^{1/2}T\Gamma_{NT}^{(l)}/\mathbb{V}_{NT}^{1/2} = o_p(1)$, for $l = 3,4,5,6$; (iv) $\hat{\mathbb{B}}_{NT} - \mathbb{B}_{NT} = o_p(K^{1/2})$; and (v) $\hat{\mathbb{V}}_{NT}/\mathbb{V}_{NT} = 1 + o_p(1)$. The proofs of (iv) and (v) are given in Propositions A.11 and A.12, respectively. We are left to show (i)–(iii). □

Proof of (i). Write $\Gamma_{NT}^{(1)} = \frac{1}{NT^2}\sum_{i=1}^N \sum_{1\le t\neq s\le T} \mathring{\mathcal{K}}_{i,ts}\varepsilon_{is}\varepsilon_{it} + \frac{1}{NT^2}\sum_{i=1}^N\sum_{t=1}^T \mathring{\mathcal{K}}_{i,tt}\varepsilon_{it}^2 + \frac{1}{NT^2}\sum_{i=1}^N \varepsilon_i'(\mathcal{K}_i - \mathring{\mathcal{K}}_i)\varepsilon_i \equiv \Gamma_{NT}^{(1a)} + \Gamma_{NT}^{(1b)} + \Gamma_{NT}^{(1c)}$, say. We can decompose J_{NT} as follows:

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

Then we want to show that (ia) $J_{NT,a} \to_d N(0,1)$; (ib) $J_{NT,b} = o_p(1)$; and (ic) $J_{NT,c} = o_p(1)$. The proof of (ia) is given in Proposition A.10. (ic) holds by the fact that $\mathbb{V}_{NT} = O(K)$ by Lemma A.6(i) and $\Gamma_{NT}^{(1c)} = o_p\left(N^{-1/2}T^{-1}K^{1/2}\right)$ by Lemma A.7(i). We are only left to show (ib). Noting that $\mathbb{V}_{NT} = O(K)$ by Lemma A.6(i) and the definition of $\mathbb{B}_{NT}$ in (3.2), we want to verify that $\mathring{J}_{NT}^{(b)} \equiv N^{1/2}T(\Gamma_{NT}^{(1b)} - \mathbb{B}_{NT}) = N^{-1/2}T^{-1}\sum_{i=1}^N\sum_{t=1}^T[\mathring{\mathcal{K}}_{i,tt}\varepsilon_{it}^2 - E(\mathring{\mathcal{K}}_{i,tt}\varepsilon_{it}^2)] = o_p\left(K^{1/2}\right)$. Clearly, $E(\mathring{J}_{NT}^{(b)}) = 0$ and

$$\mathrm{Var}\left(\mathring{J}_{NT}^{(b)}\right) = \frac{1}{NT^2}\sum_{i=1}^N\sum_{t=1}^T \mathrm{Var}\left(\mathring{\mathcal{K}}_{i,tt}\varepsilon_{it}^2\right) + \frac{2}{NT^2}\sum_{i=1}^N\sum_{1\le t<s\le T}\mathrm{Cov}\left(\mathring{\mathcal{K}}_{i,tt}\varepsilon_{it}^2, \mathring{\mathcal{K}}_{i,ss}\varepsilon_{is}^2\right) \equiv VJ_1 + VJ_2.$$

By the fact (vi), we have $\mathrm{Var}(\mathring{\mathcal{K}}_{i,tt}\varepsilon_{it}^2) \le E(\mathring{\mathcal{K}}_{i,tt}^2\varepsilon_{it}^4) \le \lambda_{\max}^2(\mathbb{Q}_i)E(||\mathring{Z}_{it}||^4\varepsilon_{it}^4) \lesssim K^2E(\mathring{A}_{it}^2\varepsilon_{it}^4)$. Then $VJ_1 \le O_p\left(K^2/T\right)$. For VJ_2, noting that $\{\mathring{\mathcal{K}}_{i,tt}\varepsilon_{it}^2\}_{t=1}^T$ are strong mixing by Assumption 1(ii), using the Davydov inequality (Bosq 1998), we have $|\mathrm{Cov}(\mathring{\mathcal{K}}_{i,tt}\varepsilon_{it}^2, \mathring{\mathcal{K}}_{i,ss}\varepsilon_{is}^2)| \le 8\alpha^{\eta/(4+\eta)}(s-t) \times ||\mathring{\mathcal{K}}_{i,tt}\varepsilon_{it}^2||_{2+\eta/2}||\mathring{\mathcal{K}}_{i,ss}\varepsilon_{is}^2||_{2+\eta/2}$, where $||\mathring{\mathcal{K}}_{i,tt}\varepsilon_{it}^2||_{2+\eta/2} \le ||2C_B(K)\lambda_{\max}(\mathbb{Q}_i)\varepsilon_{it}^2\mathring{A}_{it}||_{(4+\eta)/2} \le CK||\varepsilon_{it}^2\mathring{A}_{it}||_{(4+\eta)/2} \lesssim CK$ by Assumption 1(iv). Then, for VJ_2, we have

$$\begin{aligned}|VJ_2| &\le \frac{16}{NT^2}\sum_{i=1}^N\sum_{1\le t<s\le T}\alpha^{\frac{\eta}{4+\eta}}(s-t)\left\|\mathring{\mathcal{K}}_{i,tt}\varepsilon_{it}^2\right\|_{2+\eta/2}\left\|\mathring{\mathcal{K}}_{i,ss}\varepsilon_{is}^2\right\|_{2+\eta/2}\\ &\le C^2K^2\frac{1}{T^2}\sum_{1\le t<s\le T}\alpha^{\frac{\eta}{4+\eta}}(s-t) \lesssim \frac{CK^2}{T}\end{aligned}$$

by Assumption 1(ii). Then we have $VJ_2 = O\left(K^2/T\right)$. It follows that $\mathrm{Var}(\mathring{J}_{NT}^{(b)}) = O\left(K^2/T\right)$. By the Chebyshev inequality and Assumption 2, we have $\mathring{J}_{NT}^{(b)} = O_p\left(K/T^{1/2}\right) = o_p\left(K^{1/2}\right)$. □

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

https://doi.org/10.1017/S026646662300018X Published online by Cambridge University Press

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

using the decomposition of $\Delta_{\beta,i}(\cdot) - \bar{\Delta}_{\beta,NT}$ and $\Delta_{f,i}(\cdot)$ similar to (2.15) and (2.16), where $\Pi_{\Delta,i} \equiv (\Pi_{\Delta,i}^{(f)\prime}, \text{vec}(\Pi_{\Delta,i}^{(\beta)\prime}))'$ and $r_{\Delta,it}$ is the error coming from the sieve approximation. We have $J_{NT}^{(2)} \equiv \frac{1}{NT^2}\sum_{i=1}^N \left(\Pi_{\Delta,i}'Z_i'\mathcal{K}_iZ_i\Pi_{\Delta,i} + r_{\Delta,i}'\mathcal{K}_ir_{\Delta,i} + 2r_{\Delta,i}'\mathcal{K}_iZ_i\Pi_{\Delta,i}\right) \equiv J_{NT}^{(2a)} + J_{NT}^{(2b)} + J_{NT}^{(2c)}$, say, where $r_{\Delta,i} = \left(r_{\Delta,i1},\ldots,r_{\Delta,iT}\right)'$. First, noting that $Z_i'\mathcal{K}_iZ_i/T^2 = Z_i'Z_i/T$ and using (A.3), we have $J_{NT}^{(2a)} = \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \breve{g}_{\Delta,it}^2 + \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T r_{\Delta,it}^2 - \frac{2}{NT}\sum_{i=1}^N\sum_{t=1}^T \breve{g}_{\Delta,it}r_{\Delta,it} \equiv J_{NT1}^{(2a)} + J_{NT2}^{(2a)} - 2J_{NT3}^{(2a)}$, say. Clearly, $J_{NT1}^{(2a)} = \Phi_\Delta + o_P(1)$. By Lemma A.5, $J_{NT2}^{(2a)} = O_P\left(K^{-2\kappa}\right)$, and further $J_{NT3}^{(2a)} = O_P\left(K^{-\kappa}\right)$ by the Cauchy–Schwarz inequality. It follows that $J_{NT}^{(2a)} = \Phi_\Delta + o_P(1)$. Second, by the definition of $\mathcal{K}_i$ and the repeated use of $x'Ax \le \lambda_{\max}(A)x'x$ for any symmetric matrix A and conformable vector x, we have $J_{NT}^{(2b)} = \frac{1}{NT^2}\sum_{i=1}^N r_{\Delta,i}'M_TZ_i\hat{\mathbb{Q}}_iZ_i'M_Tr_{\Delta,i} \le \max_i\lambda_{\max}(\hat{Q}_{i,zz})$ $\max_i\lambda_{\max}(\hat{Q}_{i,\dot{z}\dot{z}}^{-1})\frac{1}{NT^2}\sum_{i=1}^N r_{\Delta,i}'\dot{Z}_i\hat{Q}_{i,\dot{z}\dot{z}}^{-1}\dot{Z}_i'r_{\Delta,i} \le \bar{c}_W\underline{c}_{\dot{z}}^{-1}\max_i\lambda_{\max}(T^{-1}\dot{Z}_i\hat{Q}_{i,\dot{z}\dot{z}}^{-1}\dot{Z}_i')\frac{1}{NT}\sum_{i=1}^N$ $\left\|r_{\Delta,i}\right\|^2 = O_P\left(K^{-2\kappa}\right)$ by Lemma A.5 and the fact that $T^{-1}\dot{Z}_i\hat{Q}_{i,\dot{z}\dot{z}}^{-1}\dot{Z}_i'$ has the largest eigenvalue 1 because it is a projection matrix. By the Cauchy–Schwarz inequality, we have $J_{NT}^{(2c)} = O_P\left(K^{-\kappa}\right) = o_P(1)$. It follows that $J_{NT}^{(2)} = \Phi_\Delta + o_P(1)$. □

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

$$\begin{aligned}
\Gamma_{NT}^{(3)} &= \frac{1}{NT^2}\sum_{i=1}^N \breve{\nu}_{NT}'X_i'M_TZ_i\hat{Q}_{i,\dot{z}\dot{z}}^{-1}\hat{Q}_{i,zz}\hat{Q}_{i,\dot{z}\dot{z}}^{-1}Z_i'M_TX_i\breve{\nu}_{NT}\\
&\le \max_i\lambda_{\max}\left(\hat{Q}_{i,zz}\right)\max_i\lambda_{\max}\left(\hat{Q}_{i,\dot{z}\dot{z}}^{-1}\right)\frac{1}{NT^2}\sum_{i=1}^N \breve{\nu}_{NT}'X_i'M_TZ_i\hat{Q}_{i,\dot{z}\dot{z}}^{-1}Z_i'M_TX_i\breve{\nu}_{NT}\\
&\le \bar{c}_W\underline{c}_{\dot{z}}^{-1}\max_i\lambda_{\max}\left(T^{-1}\dot{Z}_i\hat{Q}_{i,\dot{z}\dot{z}}^{-1}\dot{Z}_i'\right)\|\breve{\nu}_{NT}\|^2\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\left\|\dot{X}_{it}\right\|^2\\
&= \left[O_P((NT)^{-1}) + o_P\left(\gamma_{NT}^2\right)\right]O_P(1) = o_P\left(N^{-1/2}T^{-1}K^{1/2}\right)
\end{aligned}$$

because of $\breve{\nu}_{NT} = \gamma_{NT}\nu_{\Delta,NT} + \nu_{NT} = o_P(\gamma_{NT}) + O_P[(NT)^{-1/2}]$. Noting that $\mathbb{V}_{NT}^{1/2} \asymp K^{1/2}$ by Lemma A.6(i), we have $J_{NT}^{(3)} = N^{1/2}T\Gamma_{NT}^{(3)}/\mathbb{V}_{NT}^{1/2} = o_P(1)$.

When $l = 4$, we can follow the proof of Lemma A.7 to show that $\Gamma_{NT}^{(4)} = \frac{2\gamma_{NT}}{NT}\sum_{i=1}^N\sum_{t=1}^T \varepsilon_{it}\dot{Z}_{it}'\hat{\mathbb{Q}}_i \times (Z_i'M_T\breve{g}_{\Delta,i}/T) = \tilde{\Gamma}_{NT}^{(4)} + o_P\left(N^{-1/2}T^{-1}K^{1/2}\right)$, where $\tilde{\Gamma}_{NT}^{(4)} \equiv \frac{2\gamma_{NT}}{NT}\sum_{i=1}^N\sum_{t=1}^T \varepsilon_{it}\mathring{Z}_{it}'\mathbb{Q}_iG_i$ and $G_i \equiv T^{-1}E[Z_i'M_T\breve{g}_{\Delta,i}]$. Note that $E(\tilde{\Gamma}_{NT}^{(4)}) = 0$ by Assumption 1(ii) and

$$\begin{aligned}
\text{Var}(\tilde{\Gamma}_{NT}^{(4)}) &= \frac{4\gamma_{NT}^2}{N^2T^2}\sum_{i=1}^N\sum_{t=1}^T E\left(\mathring{Z}_{it}'\mathbb{Q}_iG_iG_i'\mathbb{Q}_i\mathring{Z}_{it}\varepsilon_{it}^2\right) + \frac{8\gamma_{NT}^2}{N^2T^2}\sum_{i=1}^N\sum_{1\le t<s\le T} E\left(\mathring{Z}_{it}'\mathbb{Q}_iG_iG_i'\mathbb{Q}_i\mathring{Z}_{is}\varepsilon_{it}\varepsilon_{is}\right)\\
&\equiv V\Gamma_{NT}^{(4a)} + V\Gamma_{NT}^{(4b)}, \text{ say.}
\end{aligned}$$

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

$$V\Gamma_{NT}^{(4a)} = \frac{4\gamma_{NT}^2}{N^2T^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\text{tr}\left\{\mathbb{Q}_iG_iG_i'\mathbb{Q}_iE(\mathring{Z}_{it}\mathring{Z}_{it}'\varepsilon_{it}^2)\right\}$$

$$\leq \max_i \lambda_{\max}\left(\mathbb{Q}_iG_iG_i'\mathbb{Q}_i\right)\frac{4\gamma_{NT}^2}{NT}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}E\left(\left\|\mathring{Z}_{it}\right\|^2\varepsilon_{it}^2\right) = O_p\left(\frac{\gamma_{NT}^2K^2}{NT}\right),$$

where we use $\lambda_{\max}\{\mathbb{Q}_iG_iG_i'\mathbb{Q}_i\} \leq \lambda_{\max}^2(\mathbb{Q}_i)\max_i\|G_i\|^2 \leq CK^2$ in the last equation. For $V\Gamma_{NT}^{(4b)}$, by Assumption 1(ii) and the Davydov inequality (Bosq 1998) again, we have

$$V\Gamma_{NT}^{(4b)} \leq \frac{8\gamma_{NT}^2}{N^2T^2}\sum_{i=1}^{N}\sum_{1\leq t<s\leq T}E\left(\mathring{Z}_{it}'G_iG_i'\mathring{Z}_{is}\varepsilon_{it}\varepsilon_{is}\right)$$

$$\lesssim \frac{\gamma_{NT}^2K^2}{N^2T^3}\sum_{i=1}^{N}\sum_{1\leq t<s\leq T}\alpha^{\frac{\eta}{1+\eta}}(t-s) = O_p\left(\frac{\gamma_{NT}^2K^2}{NT}\right).$$

By the Chebyshev inequality, $\tilde{\Gamma}_{NT}^{(4)} = O_p[\gamma_{NT}K/(NT)^{1/2}] = o_p\left(N^{-1/2}T^{-1}K^{1/2}\right)$. 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 $F = O_p[K^{1/2}/(NT)^{1/2}]$. Then we have $|\Gamma_{NT}^{(5)}| \leq O_p[K^{1/2}/(NT)^{1/2}][o_p(\gamma_{NT}) + O_p((NT)^{-1/2})] = o_p\left(N^{-1/2}T^{-1}K^{1/2}\right)$. It follows that $J_{NT}^{(5)} = o_p(1)$. When $l=6$, we have $J_{NT}^{(6)} = o_p(1)$ by the Cauchy–Schwarz inequality. □

PROPOSITION A.10. *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 $J_{NT,a} = \frac{1}{\sqrt{T}}\sum_{t=2}^{T}\mathcal{Z}_{NT,t}$, where $\mathcal{Z}_{NT,t} \equiv \frac{2}{\sqrt{NT\mathbb{V}_{NT}}}\sum_{s=1}^{t-1}\sum_{i=1}^{N}\mathring{\mathcal{K}}_{i,ts}\varepsilon_{it}\varepsilon_{is}$. Noting that $\{\mathcal{Z}_{NT,t}\}_{t=1}^{T}$ is an MDS w.r.t. $\mathcal{F}_t \equiv \sigma\{(X_{it},\ldots,X_{i1},\varepsilon_{i,t-1},\ldots,\varepsilon_{i1}),\ i=1,\ldots,N\}$, we prove the proposition by applying the martingale CLT. By Corollary 5.26 of White (2000), it suffices to show that: (i) $E[\mathcal{Z}_{NT,t}^4] < C < \infty$ for some constant C and all t and (ii) $T^{-1}\sum_{t=2}^{T}\mathcal{Z}_{NT,t}^2 - 1 = o_P(1)$. We first show (i). For $2\leq t\leq T$, decompose

$$\mathcal{Z}_{NT,t}^2 = \frac{4}{NT\mathbb{V}_{NT}}\sum_{s=1}^{t-1}\sum_{s_2=1}^{t-1}\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\mathring{\mathcal{K}}_{i_1,ts_1}\varepsilon_{i_1t}\varepsilon_{i_1s_1}\mathring{\mathcal{K}}_{i_2,ts_2}\varepsilon_{i_2s_2}\varepsilon_{i_2t}$$

$$= \frac{4}{NT\mathbb{V}_{NT}}\sum_{s=1}^{t-1}\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\chi_{i_1,ts}\chi_{i_2,ts} + \frac{8}{NT\mathbb{V}_{NT}}\sum_{1\leq s_1<s_2\leq t-1}\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\chi_{i_1,ts_1}\chi_{i_2,ts_2}$$

$$= \mathcal{Z}_{1t} + \mathcal{Z}_{2t},\ \text{say,}$$

https://doi.org/10.1017/S026646662300018X Published online by Cambridge University Press

where $\chi_{i,ts} \equiv \mathring{\mathcal{K}}_{i,ts}\varepsilon_{it}\varepsilon_{is}$. Then $E(\mathcal{Z}_{NT,t}^4) = E(\mathcal{Z}_{1t}+\mathcal{Z}_{2t})^2 \le 2\left[E\left(\mathcal{Z}_{1t}^2\right)+E\left(\mathcal{Z}_{2t}^2\right)\right] \equiv 2(\mathbb{Z}_{1t}+\mathbb{Z}_{2t})$, say. We show that $\mathbb{Z}_{lt} < C < \infty$, for $l = 1,2$ and all t's. For $\mathbb{Z}_{1t}$, we have

$$\begin{aligned}
\mathbb{Z}_{1t} &= \frac{16}{N^2T^2\mathbb{V}_{NT}^2}\sum_{s=1}^{t-1}\sum_{i=1}^{N}E\left(\chi_{i,ts}^4\right) + \frac{48}{N^2T^2\mathbb{V}_{NT}^2}\sum_{s=1}^{t-1}\sum_{1\le i\ne j\le N}E\left(\chi_{i,ts}^2\right)E(\chi_{j,ts}^2)\\
&+ \frac{32}{N^2T^2\mathbb{V}_{NT}^2}\sum_{1\le s_1<s_2\le t-1}\sum_{i=1}^{N}E(\chi_{i,ts_1}^2\chi_{i,ts_2}^2) + \frac{32}{N^2T^2\mathbb{V}_{NT}^2}\sum_{1\le s_1<s_2\le t-1}\sum_{1\le i\ne j\le N}E(\chi_{i,ts_1}^2)E(\chi_{j,ts_2}^2)\\
&+ \frac{64}{N^2T^2\mathbb{V}_{NT}^2}\sum_{1\le s_1<s_2\le t-1}\sum_{1\le i\ne j\le N}E(\chi_{i,ts_1}\chi_{i,ts_2})E(\chi_{j,ts_1}\chi_{j,ts_2})\\
&= \mathbb{Z}_{1t,a}+\mathbb{Z}_{1t,b}+\mathbb{Z}_{1t,c}+\mathbb{Z}_{1t,d}+\mathbb{Z}_{1t,e}, \text{ say.}
\end{aligned}$$

Note that $E(\chi_{i,ts}^4) = E[(\mathring{Z}_{it}'\mathbb{Q}_i\mathring{Z}_{is})^4\varepsilon_{it}^4\varepsilon_{is}^4] \le \lambda_{\max}^4(\mathbb{Q}_i)E(||\mathring{Z}_{it}||^4||\mathring{Z}_{is}||^4\varepsilon_{it}^4\varepsilon_{is}^4) \le K^4C^*$, where $C^* \equiv 256\lambda_{\max}^4(\mathbb{Q}_i)[\max_{i,t}E(\mathring{A}_{it}^8)]^{1/2}\left[\max_{i,t}E\left(\varepsilon_{it}^{16}\right)\right]^{1/2} < \infty$ by Assumption 1(iv). Noting that $\mathbb{V}_{NT} \asymp K$ by Lemma A.6(i), we have $\mathbb{Z}_{1t,a} \lesssim \frac{16}{N^2T^2\mathbb{V}_{NT}^2}\sum_{s=1}^{t-1}\sum_{i=1}^{N}K^4C^* = O(K^2/(NT))$. Similarly, $\mathbb{Z}_{1t,b} \lesssim O\left(K^2/T\right)$ and $\mathbb{Z}_{1t,c} \lesssim O\left(K^2/N\right)$. For $\mathbb{Z}_{1t,d}$, noting that $E(\chi_{i,ts}^2) = E[(\mathring{Z}_{it}'\mathbb{Q}_i\mathring{Z}_{is})^2\varepsilon_{it}^2\varepsilon_{is}^2] \asymp K^2$, we have

$$\mathbb{Z}_{1t,d} \lesssim \left[\frac{\sum_{s=1}^{T}\sum_{i=1}^{N}E(\chi_{i,ts}^2)}{T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}E(\chi_{i,ts}^2)}\right]^2 < C$$

using $\mathbb{V}_{NT} = \frac{2}{NT^2}\sum_{1\le s,t\le T}\sum_{i=1}^{N}E(\chi_{i,ts}^2)$. Let $m \equiv m_T = \lfloor C\ln T\rfloor$, where $\lfloor\cdot\rfloor$ is the integer part of "·"For $\mathbb{Z}_{1t,e}$, we consider two cases for time indices s_1,s_2: (a1) $d_1 \equiv \max\{s_2-s_1, t-s_2\} \ge m$ and (a2) $d_1 < m$. For Case (a1), when $d_1 = s_2-s_1$, by Lemma 2.1 of Sun and Chiang (1997), we have

$$\left|E\left(\chi_{i,ts_1}\chi_{i,ts_2}\right)\right| = \left|E\left(\chi_{i,ts_1}\chi_{i,ts_2}\right) - E_{s_2t}E_{s_1}\left(\chi_{i,ts_1}\chi_{i,ts_2}\right)\right| \le 4M_{3a}^{1/(1+\eta)}\alpha^{\eta/(1+\eta)}(m),$$

where $E_{s_2t}E_{s_1}\left(\chi_{i,ts_1}\chi_{i,ts_2}\right) \equiv \int \chi_{i,ts_1}\chi_{i,ts_2}dF_{i,s_1}^{(1)}dF_{i,s_2t}^{(2)} = 0$, $F_{i,s_1}^{(1)}$ and $F_{i,s_2t}^{(2)}$ denote the marginal CDF of $\xi_{is_1} \equiv \left(X_{is_1},\varepsilon_{is_1}\right)$ and the joint CDF of ξ_{is_2} and ξ_{it}, respectively, and $M_{3a} \equiv \max_i\max_{1\le s_1<s_2<t\le T}\int\left|\chi_{i,ts_1}\chi_{i,ts_2}\right|^{1+\eta}dF_{i,s_1}^{(1)}dF_{i,s_2t}^{(2)}$; when $d_1 = t-s_2$, we have

$$\left|E\left(\chi_{i,ts_1}\chi_{i,ts_2}\right)\right| = \left|E\left(\chi_{i,ts_1}\chi_{i,ts_2}\right) - E_{s_1s_2}E_t\left(\chi_{i,ts_1}\chi_{i,ts_2}\right)\right| \le 4M_{3b}^{1/(1+\eta)}\alpha^{\eta/(1+\eta)}(m),$$

where $E_{s_1s_2}E_t\left(\chi_{i,ts_1}\chi_{i,ts_2}\right) = 0$ and $M_{3b} \equiv \max_i\max_{1\le s_1<s_2<t\le T}\int\left|\chi_{i,ts_1}\chi_{i,ts_2}\right|^{1+\eta}dF_{i,s_1s_2}^{(2)}dF_{i,t}^{(1)}$. Clearly, $M_{3a} \lesssim K^{2(1+\eta)}\max_i\lambda_{\max}^{2+2\eta}(\mathbb{Q}_i)\max_{i,s_1,s_2,t}\int|\mathring{A}_{is_1}\varepsilon_{is_1}\mathring{A}_{is_2}\varepsilon_{is_2}\mathring{A}_{it}^2\varepsilon_{it}^2|^{1+\eta}dF_{i,s_1}^{(1)}\ dF_{i,s_2t}^{(2)} \lesssim CK^{2(1+\eta)}$. Similarly, we have $M_{3b} \lesssim CK^{2(1+\eta)}$. It follows that $M_3 \equiv \max\{M_{3a},M_{3b}\} \lesssim CK^{2(1+\eta)}$ and $\left|E\left(\chi_{i,ts_1}\chi_{i,ts_2}\right)\right| \lesssim K^2\alpha^{\eta/(1+\eta)}(m)$ for Case (a1). For Case (a2), for each expectation in summation, we have $\left|E\left(\chi_{i,ts_1}\chi_{i,ts_2}\right)\right| \lesssim K^2\lambda_{\max}^2(\mathbb{Q}_i)\max_{i,s_1,s_2,t}E\left|\mathring{A}_{is_1}\varepsilon_{is_1}\mathring{A}_{is_2}\varepsilon_{is_2}\mathring{A}_{it}^2\varepsilon_{it}^2\right|^{1+\eta} = O\left(K^2\right)$. Then

https://doi.org/10.1017/S026646662300018X Published online by Cambridge University Press

it follows

$$
\begin{aligned}
|\mathbb{Z}_{1t,e}| &\leq \frac{64}{N^2T^2\mathbb{V}_{NT}^2}\sum_{1\leq i\neq j\leq N}\left\{\sum\nolimits_{(\mathrm{a1})}+\sum\nolimits_{(\mathrm{a2})}\right\}\left|E\left(\chi_{i,ts_1}\chi_{i,ts_2}\right)\right|\left|E\left(\chi_{j,ts_1}\chi_{j,ts_2}\right)\right| \\
&\lesssim \frac{64}{N^2T^2\mathbb{V}_{NT}^2}\left(N^2t^2K^4\alpha^{\frac{2\eta}{1+\eta}}(m)+N^2m^2K^4\right)=O\left(K^2\alpha^{\frac{2\eta}{1+\eta}}(m)+m^2K^4T^{-2}\right).
\end{aligned}
$$

Under Assumption 1(ii), $K^2\alpha^{2\eta/(1+\eta)}(m)\to 0$ holds for a choice of m with a large constant C. It follows that $|\mathbb{Z}_{1t}|\leq C<\infty$. Now we consider $\mathbb{Z}_{2t}$. Write

$$
\begin{aligned}
\mathbb{Z}_{2t} &= \frac{64}{N^2T^2\mathbb{V}_{NT}^2}\sum_{1\leq s_1<s_2\leq t-1}\sum_{1\leq s_3<s_4\leq t-1}\sum_{i=1}^{N}E\left(\chi_{i,ts_1}\chi_{i,ts_2}\chi_{i,ts_3}\chi_{i,ts_4}\right) \\
&\quad+\frac{64}{N^2T^2\mathbb{V}_{NT}^2}\sum_{1\leq s_1<s_2\leq t-1}\sum_{1\leq s_3<s_4\leq t-1}\sum_{1\leq i\neq j\leq N}E\left(\chi_{i,ts_1}\chi_{i,ts_2}\right)E\left(\chi_{j,ts_3}\chi_{j,ts_4}\right) \\
&\quad+\frac{64}{N^2T^2\mathbb{V}_{NT}^2}\sum_{1\leq s_1<s_2\leq t-1}\sum_{1\leq s_3<s_4\leq t-1}\sum_{1\leq i\neq j\leq N}E\left(\chi_{i,ts_1}\chi_{i,ts_3}\right)E\left(\chi_{j,ts_2}\chi_{j,ts_4}\right) \\
&\quad+\frac{64}{N^2T^2\mathbb{V}_{NT}^2}\sum_{1\leq s_1<s_2\leq t-1}\sum_{1\leq s_3<s_4\leq t-1}\sum_{1\leq i\neq j\leq N}E\left(\chi_{i,ts_1}\chi_{i,ts_4}\right)E\left(\chi_{j,ts_2}\chi_{j,ts_3}\right) \\
&\equiv \mathbb{Z}_{2t,a}+\mathbb{Z}_{2t,b}+\mathbb{Z}_{2t,c}+\mathbb{Z}_{2t,d},\text{ say.}
\end{aligned}
$$

First, rewrite $\mathbb{Z}_{2t,a}$ as

$$
\begin{aligned}
\mathbb{Z}_{2t,a} &= \frac{16}{N^2T^2\mathbb{V}_{NT}^2}\sum_{1\leq s_1<s_2\leq t-1}\sum_{i=1}^{N}E(\chi_{i,ts_1}^2\chi_{i,ts_2}^2) \\
&\quad+\frac{16}{N^2T^2\mathbb{V}_{NT}^2}\sum_{1\leq s_1<s_2<s_4\leq t-1}\sum_{i=1}^{N}E(\chi_{i,ts_1}^2\chi_{i,ts_2}\chi_{i,ts_4}) \\
&\quad+\frac{16}{N^2T^2\mathbb{V}_{NT}^2}\sum_{1\leq s_1<s_2<s_4\leq t-1}\sum_{i=1}^{N}E(\chi_{i,ts_1}\chi_{i,ts_2}^2\chi_{i,ts_4}) \\
&\quad+\frac{16}{N^2T^2\mathbb{V}_{NT}^2}\sum_{1\leq s_1\neq s_2\neq s_3\neq s_4\leq t-1}\sum_{i=1}^{N}E(\chi_{i,ts_1}\chi_{i,ts_2}\chi_{i,ts_3}\chi_{i,ts_4}) \\
&= \mathbb{Z}_{2t,a1}+\mathbb{Z}_{2t,a2}+\mathbb{Z}_{2t,d3}+\mathbb{Z}_{2t,a4},\text{ say.}
\end{aligned}
$$

Similar to the proof of $\mathbb{Z}_{1t,a}$, we can show $\mathbb{Z}_{2t,a1}\leq O\left(K^2/N\right)$. For $\mathbb{Z}_{2t,a2}$, let $d_1\geq d_2\geq d_3$ be the decreasing ranked increments among four different time indices: s_1,s_2,s_4, and t. We consider two cases: (a1) $d_2\geq m$ and (a2) $d_2<m$. For Case (a1), it must be $d_2=s_2-s_1$, s_4-s_2, or $t-s_4$. By Lemma 2.1 in Sun and Chiang (1997) again, we have $|E(\chi_{i,ts_1}^2\chi_{i,ts_2}\chi_{i,ts_4})|=|E(\chi_{i,ts_1}^2\chi_{i,ts_2}\chi_{i,ts_4})-E_{s_2s_4t}E_{s_1}(\chi_{i,ts_1}^2\chi_{i,ts_2}\chi_{i,ts_4})|\leq M_4^{1/(1+\eta)}\alpha^{\eta/(1+\eta)}(d_2)$ when $d_2=s_2-s_1$; similarly, $|E(\chi_{i,ts_1}^2\chi_{i,ts_2}\chi_{i,ts_4})|\leq M_4^{1/(1+\eta)}$

https://doi.org/10.1017/S026646662300018X Published online by Cambridge University Press

$\alpha^{\eta/(1+\eta)}(d_2)$ when $d_2 = s_4 - s_2$ or $t - s_4$, where $M_4 \equiv \max\{M_{4a}, M_{4b}, M_{4c}\}$ with

$$M_{4a} \equiv \max_{1\le i\le N} \max_{1\le s_1<s_2<s_4<t\le T} \int \left|\chi_{i,ts_1}^2 \chi_{i,ts_2}\chi_{i,ts_4}\right|^{1+\eta} dF_{i,s_1}^{(1)} dF_{i,s_2s_4t}^{(3)},$$

$$M_{4b} \equiv \max_{1\le i\le N} \max_{1\le s_1<s_2<s_4<t\le T} \int \left|\chi_{i,ts_1}^2 \chi_{i,ts_2}\chi_{i,ts_4}\right|^{1+\eta} dF_{i,s_1s_2}^{(2)} dF_{i,s_4t}^{(2)},$$

$$M_{4c} \equiv \max_{1\le i\le N} \max_{1\le s_1<s_2<s_4<t\le T} \int \left|\chi_{i,ts_1}^2 \chi_{i,ts_2}\chi_{i,ts_4}\right|^{1+\eta} dF_{i,s_1s_2s_4}^{(3)} dF_{i,t}^{(1)}.$$

It is easy to show that $M_4 \lesssim CK^{2(1+\eta)}$ and $|E(\chi_{i,ts_1}^2 \chi_{i,ts_2}\chi_{i,ts_4})| \lesssim K^2\alpha^{\eta/(1+\eta)}(m)$. Note the total number of these terms in Case (a1) is bounded by t^3; for Case (a2), $|E(\chi_{i,ts_1}^2 \chi_{i,ts_2}\chi_{i,ts_4})| \le CK^4$ and the total number of these terms is bounded by tm^2. Then we have

$$\left|\mathbb{Z}_{2t,a2}\right| \lesssim \frac{16}{N^2T^2\mathbb{V}_{NT}^2}(K^4Nt^3\alpha^{\frac{\eta}{1+\eta}}(m) + tm^2NK^4) = O\left(\frac{K^2T}{N}\alpha^{\frac{\eta}{1+\eta}}(m) + \frac{m^2K^2}{NT}\right).$$

Similarly, $\left|\mathbb{Z}_{2t,al}\right| \le O(N^{-1}K^2\alpha^{\eta/(1+\eta)}(m) + \frac{m^2K^2}{NT})$, for $l = 3,4$. It follows that $\left|\mathbb{Z}_{2t,a}\right| \le C < \infty$. For $\mathbb{Z}_{2t,b}$, we have

$$\left|\mathbb{Z}_{2t,b}\right| \le \frac{16}{N^2T^2\mathbb{V}_{NT}^2}\left\{\sum_{1\le s_1<s_2\le t-1}\sum_{i=1}^{N}\left|E\left(\chi_{i,ts_1}\chi_{i,ts_2}\right)\right|\right\}^2 \lesssim O\left(T^2K^2\alpha^{\frac{2\eta}{1+\eta}}(m) + m^4K^2T^{-2}\right).$$

Similarly, we can write $\mathbb{Z}_{2t,c}$ as follows:

$$\begin{aligned}
\mathbb{Z}_{2t,c} =& \frac{64}{N^2T^2\mathbb{V}_{NT}^2}\sum_{1\le s_1<s_2\le t-1}\sum_{1\le i\ne j\le N} E(\chi_{i,ts_1}^2)E(\chi_{j,ts_2}^2)\\
&+ \frac{64}{N^2T^2\mathbb{V}_{NT}^2}\sum_{1\le s_1<s_2\ne s_4\le t-1}\sum_{1\le i\ne j\le N} E(\chi_{i,ts_1}^2)E\left(\chi_{j,ts_2}\chi_{j,ts_4}\right)\\
&+ \frac{64}{N^2T^2\mathbb{V}_{NT}^2}\sum_{1\le s_1\ne s_3<s_2\le t-1}\sum_{1\le i\ne j\le N} E\left(\chi_{i,ts_1}\chi_{i,ts_3}\right)E(\chi_{j,ts_2}^2)\\
&+ \frac{64}{N^2T^2\mathbb{V}_{NT}^2}\sum_{1\le s_3<s_1<s_2\le t-1}\sum_{1\le i\ne j\le N} E\left(\chi_{i,ts_1}\chi_{i,ts_3}\right)E\left(\chi_{j,ts_2}\chi_{j,ts_1}\right)\\
&+ \frac{64}{N^2T^2\mathbb{V}_{NT}^2}\sum_{1\le s_1<s_2<s_4\le t-1}\sum_{1\le i\ne j\le N} E\left(\chi_{i,ts_1}\chi_{i,ts_2}\right)E\left(\chi_{j,ts_2}\chi_{j,ts_4}\right)\\
&+ \frac{64}{N^2T^2\mathbb{V}_{NT}^2}\sum_{\substack{1\le s_1<s_2\le t-1,\,1\le s_3<s_4\le t-1\\ s_1\ne s_2\ne s_3\ne s_4}}\sum_{1\le i\ne j\le N} E\left(\chi_{i,ts_1}\chi_{i,ts_3}\right)E\left(\chi_{j,ts_2}\chi_{j,ts_4}\right)\\
=& \mathbb{Z}_{2t,c1} + \mathbb{Z}_{2t,c2} + \mathbb{Z}_{2t,c3} + \mathbb{Z}_{2t,c4} + \mathbb{Z}_{2t,5} + \mathbb{Z}_{2t,6}, \text{ say.}
\end{aligned}$$

We have $\left|\mathbb{Z}_{2t,c1}\right| \le C < \infty$ as the determination of the upper bound for $\mathbb{Z}_{1t,d}$. Similarly, we can show that $\left|\mathbb{Z}_{2t,cl}\right| \lesssim O[K^2T\alpha^{\frac{\eta}{1+\eta}}(m) + K^2m^2/T]$, for $l = 2,3$, $\left|\mathbb{Z}_{2t,cl}\right| \lesssim O[K^2T\alpha^{\frac{2\eta}{1+\eta}}(m) + K^2m\alpha^{\frac{\eta}{1+\eta}}(m) + K^2m^3/T]$, for $l = 4,5$, and

https://doi.org/10.1017/S026646662300018X Published online by Cambridge University Press

$\left|\mathbb{Z}_{2t,c6}\right| \lesssim O[K^2T^2\alpha^{\frac{2\eta}{1+\eta}}(m)+K^2m^2\alpha^{\frac{\eta}{1+\eta}}(m)+K^2m^4/T]$. It follows that $\left|\mathbb{Z}_{2t,c}\right| \leq C < \infty$. Similarly, we have $\left|\mathbb{Z}_{2t,d}\right| \leq C < \infty$. Thus, $|\mathbb{Z}_{2t}| \leq C < \infty$ for all t's.

Proof of (ii). First, note that $T^{-1}\sum_{t=2}^{T}E\left[\mathcal{Z}_{NT,t}^2\right] = \frac{1}{NT^2\mathbb{V}_{NT}}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N} E\left(\chi_{i,ts}^2\right)+O\left(K^2/T\right) = 1+o(1)$. Then we decompose

$$E\left[\left(\frac{1}{T}\sum_{t=2}^{T}\mathcal{Z}_{NT,t}^2\right)^2\right]$$

$$= \frac{1}{T^2}\sum_{t=2}^{T}E\left(\mathcal{Z}_{NT,t}^4\right)+\frac{2}{T^2}\sum_{2\leq t<s\leq T}E\left(\mathcal{Z}_{NT,t}^2\mathcal{Z}_{NT,s}^2\right) \equiv \mathfrak{Z}_{1NT}+\mathfrak{Z}_{2NT}, \text{ say.}$$

Clearly, $\mathfrak{Z}_{1NT} = O(1/T) = o(1)$; for $\mathfrak{Z}_{2NT}$, we have

$$\mathfrak{Z}_{2NT} = \frac{2}{T^2}\sum_{2\leq t<s\leq T}E\left(\mathcal{Z}_{1t}\mathcal{Z}_{1s}+\mathcal{Z}_{2t}\mathcal{Z}_{2s}+\mathcal{Z}_{1t}\mathcal{Z}_{2s}+\mathcal{Z}_{2t}\mathcal{Z}_{1s}\right) \equiv \sum_{l=1}^{4}\mathfrak{Z}_{2NT,l}, \text{ say.}$$

Write $\mathfrak{Z}_{2NT,1}$ as follows:

$$\mathfrak{Z}_{2NT,1} = \frac{32}{N^2T^4\mathbb{V}_{NT}^2}\sum_{2\leq t_1<t_2\leq T}\sum_{s_1=1}^{t_1-1}\sum_{s_2=1}^{t_2-1}\left\{\sum_{i=1}^{N}E(\chi_{i,t_1s_1}^2\chi_{i,t_2s_2}^2)+\sum_{1\leq i\neq j\leq N}E(\chi_{i,t_1s_1}^2)E(\chi_{j,t_2s_2}^2)\right\}$$

$$\lesssim \left[\frac{1}{NT^2\mathbb{V}_{NT}}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}E\left(\chi_{i,ts}^2\right)\right]^2+O\left(K^2/N\right).$$

For $\mathfrak{Z}_{2NT,2}$, we have

$$\mathfrak{Z}_{2NT,2} = \frac{2}{N^2T^4\mathbb{V}_{NT}^2}\sum_{2\leq t_1<t_2\leq T}\sum_{1\leq s_1<s_2\leq t_1-1}\sum_{1\leq s_3<s_4\leq t_2-1}\sum_{i=1}^{N}E(\chi_{i,t_1s_1}\chi_{i,t_1s_2}\chi_{i,t_2s_3}\chi_{i,t_2s_4})$$
$$+\frac{2}{N^2T^4\mathbb{V}_{NT}^2}\sum_{2\leq t_1<t_2\leq T}\sum_{1\leq s_1<s_2\leq t_1-1}\sum_{1\leq s_3<s_4\leq t_2-1}\sum_{1\leq i_1\neq i_3\leq N}E(\chi_{i_1,t_1s_1}\chi_{i_1,t_1s_2})E(\chi_{i_3,t_2s_3}\chi_{i_3,t_2s_4})$$
$$= \mathfrak{Z}_{2NT,2a}+\mathfrak{Z}_{2NT,2b}, \text{ say.}$$

For $\mathfrak{Z}_{2NT,2a}$, we consider three cases with two, three, and four different elements in $\{s_1,s_2,s_3,s_4\}$ and accordingly decompose $\mathfrak{Z}_{2NT,2a}$ into $\mathfrak{Z}_{2NT,2a} = \mathfrak{Z}_{2NT,2a2}+\mathfrak{Z}_{2NT,2a3}+\mathfrak{Z}_{2NT,2a4}$. For $\mathfrak{Z}_{2NT,2a2}$, each expectation $E\left(\chi_{i,t_1s_1}\chi_{i,t_1s_2}\chi_{i,t_2s_1}\chi_{i,t_2s_2}\right)$ is of order K^2 and we have $\mathfrak{Z}_{2NT,2a2} = O\left(K^2/N\right)$. For $\mathfrak{Z}_{2NT,2a3}$, we consider two subcases with $d_3 \geq m$ or $d_3 < m$, where d_3 is the third largest increment among five different time indices. When $d_3 \geq m$, each expectation is bounded by $CK^4\alpha^{\eta/(1+\eta)}(m)$ and the total number of such terms is of order $O\left(T^5\right)$; when $d_3 \leq m$, each expectation is of order K^4 and the total number is of order $O(m^2T^3)$. Then $\left|\mathfrak{Z}_{2NT,2a3}\right| = O[K^2T\alpha^{\eta/(1+\eta)}(m)/N+K^2m^2/(NT)] = o(1)$. Similarly, $\left|\mathfrak{Z}_{2NT,2a4}\right| = O[K^2T^2\alpha^{\eta/(1+\eta)}(m)/N+K^2m^3/(NT)]$. Lastly, we can show that $\left|\mathfrak{Z}_{2NT,2b}\right| = O[T^2K^2\alpha^{2\eta/(1+\eta)}(m)+m^4K^2/T^2]$ analogously. It follows that $\mathfrak{Z}_{2NT,2} = o(1)$. Finally, following the proof for $\mathfrak{Z}_{2NT,2}$, we can show that

https://doi.org/10.1017/S026646662300018X Published online by Cambridge University Press

$\mathfrak{z}_{2NT,l} = O[TK^2\alpha^{\eta/(1+\eta)}(m)/N + K^2m^2/(NT) + TK^2\alpha^{\eta/(1+\eta)}(m) + m^2K^2/T] = o(1)$ for $l = 3,4$. It follows that

$$E\left[\left(\frac{1}{T}\sum_{t=2}^{T}\mathcal{Z}_{NT,t}^2\right)^2\right] = \left[\frac{1}{NT^2\mathbb{V}_{NT}}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}E\left(\chi_{i,ts}^2\right)\right]^2 + o(1) \text{ and}$$

$$\mathrm{Var}\left(\frac{1}{T}\sum_{t=2}^{T}\mathcal{Z}_{NT,t}^2\right) = E\left[\left(\frac{1}{T}\sum_{t=2}^{T}\mathcal{Z}_{NT,t}^2\right)^2\right] - \left[E\left(\frac{1}{T}\sum_{t=2}^{T}\mathcal{Z}_{NT,t}^2\right)\right]^2 = o(1).$$

Consequently, (ii) holds by the Chebyshev inequality. □

PROPOSITION A.11. *Under Assumptions 1–4,* $\hat{\mathbb{B}}_{NT} - \mathbb{B}_{NT} = o_p(K^{1/2})$.

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

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

where

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

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

First, write

$$\hat{\mathbb{B}}_{NT1} - \mathbb{B}_{NT} = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\mathring{\mathcal{K}}_{i,tt}\varepsilon_{it}^2 - E\left(\mathring{\mathcal{K}}_{i,tt}\varepsilon_{it}^2\right)\right] + \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\mathcal{K}_{i,tt} - \mathring{\mathcal{K}}_{i,tt}\right)\varepsilon_{it}^2$$
$$\equiv \tilde{J}_{NT}^{(b)} + \tilde{J}_{NT}^{(\triangle b)}, \text{ say.}$$

We can show that $E[\tilde{J}_{NT}^{(b)}] = 0$ and $\mathrm{Var}(\tilde{J}_{NT}^{(b)}) = O\left(K^2/T\right) = o(1)$. By the Chebyshev inequality, $\tilde{J}_{NT}^{(b)} = O_p\left(K/T^{1/2}\right) = o_p\left(K^{1/2}\right)$. Following the proof of Lemma A.7, we can show that $\tilde{J}_{NT}^{(\triangle b)} = o_p\left(K^{1/2}\right)$. Second, $\hat{\mathbb{B}}_{NT2} \leq \frac{1}{N^{1/2}T}\sum_{i=1}^N\bar{\varepsilon}_i^2\mathrm{tr}(\mathcal{K}_i) \leq \frac{C}{N^{1/2}T}\sum_{i=1}^N\bar{\varepsilon}_i^2\|Z_i\|^2 = O(N^{1/2}K/T) = o_p\left(K^{1/2}\right)$ by the Markov inequality. Third,

$\hat{\mathbb{B}}_{NT3} \leq \frac{C\gamma_{NT}^2 K}{N^{1/2}T}\sum_{i=1}^N\sum_{t=1}^T A_{it}[\breve{g}_{\Delta,it}^{(c)}]^2 = O_p\left(KN^{1/2}\gamma_{NT}^2\right) = o_p\left(K^{1/2}\right)$. Fourth, $\hat{\mathbb{B}}_{NT4} \leq \frac{C_* K\|\breve{\nu}_{NT}\|^2}{N^{1/2}T}\sum_{i=1}^N\sum_{t=1}^T A_{it}||\dot{X}_{it}||^2 = O_p(N^{1/2}K\|\breve{\nu}_{NT}\|) = o_p\left(K^{1/2}\right)$. Lastly, by the Cauchy–Schwarz inequality, we can show that $\hat{\mathbb{B}}_{NTl} = o_p\left(K^{1/2}\right)$, for $l = 5,\ldots,10$. □

Proposition A.2. *Under Assumptions 1–4, we have* $\hat{\mathbb{V}}_{NT}/\mathbb{V}_{NT} = 1 + o_p(1)$.

Proof. We consider the following decomposition:

$$\begin{aligned}
\hat{\mathbb{V}}_{NT} - \mathbb{V}_{NT} &= \frac{2}{NT^2}\sum_{i=1}^N\sum_{1\leq s,t\leq T}\mathcal{K}_{i,ts}^2\left(\hat{\varepsilon}_{r,it}^2\hat{\varepsilon}_{r,is}^2 - \varepsilon_{it}^2\varepsilon_{is}^2\right)\\
&+ \frac{2}{NT^2}\sum_{i=1}^N\sum_{1\leq s,t\leq T}\left[\mathring{\mathcal{K}}_{i,ts}^2\varepsilon_{it}^2\varepsilon_{is}^2 - E\left(\mathring{\mathcal{K}}_{i,ts}^2\varepsilon_{it}^2\varepsilon_{is}^2\right)\right]\\
&+ \frac{2}{NT^2}\sum_{i=1}^N\sum_{1\leq s,t\leq T}\left(\mathcal{K}_{i,ts}^2 - \mathring{\mathcal{K}}_{i,ts}^2\right)\varepsilon_{it}^2\varepsilon_{is}^2 + \frac{2}{NT^2}\sum_{i=1}^N\sum_{1\leq s,t\leq T}E\left[\left(\mathring{\mathcal{K}}_{i,ts}^2 - \mathcal{K}_{i,ts}^2\right)\varepsilon_{it}^2\varepsilon_{is}^2\right]\\
&\equiv \Delta\hat{\mathbb{V}}_{NT}^{(a)} + \Delta\hat{\mathbb{V}}_{NT}^{(b)} + \Delta\hat{\mathbb{V}}_{NT}^{(c)} + \Delta\hat{\mathbb{V}}_{NT}^{(d)}, \text{ say.}
\end{aligned}$$

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

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

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

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

https://doi.org/10.1017/S026646662300018X Published online by Cambridge University Press

by the symmetry between time indices t and s. First, decompose $\Delta\hat{\mathbb{V}}_{NT,1}^{(a)}$ as follows:

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

Define $\dot{\Omega}_i = T^{-1}\sum_{s=1}^{T}\dot{Z}_{is}\dot{Z}_{is}'\varepsilon_{is}^2$ and $\dot{V}_i = T^{-1/2}\sum_{t=1}^{T}\dot{Z}_{it}\dot{Z}_{it}'\varepsilon_{it}$. For $\Delta\hat{\mathbb{V}}_{NT,11}^{(a)}$, then we have

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

By Lemma A.3(iii) and (v) and Lemma A.4(iii)–(iv), we have $|\Delta\hat{\mathbb{V}}_{NT,111}^{(a)}| = \frac{1}{NT}\sum_{i=1}^{N}\operatorname{tr}\left(\hat{\mathbb{Q}}_i\hat{\Omega}_i\hat{\mathbb{Q}}_i\hat{\mathbb{Q}}_i\hat{\Omega}_i\hat{\mathbb{Q}}_i\right)\bar{\varepsilon}_i^2 \le K\max_i\lambda_{\max}^4\left(\hat{\mathbb{Q}}_i\right)\max_i\lambda_{\max}^2\left(\hat{\Omega}_i\right)\frac{1}{NT}\sum_{i=1}^{N}\bar{\varepsilon}_i^2 = O_p(K)$. Second, $E[\Delta\hat{\mathbb{V}}_{NT,112}^{(a)}] = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}E(||\mathring{Z}_{it}||^4\varepsilon_{it}^2) = O\left(K^2\right)$ implies that $\Delta\hat{\mathbb{V}}_{NT,112}^{(a)} = O_p\left(K^2\right)$ by the Markov inequality. It follows that $\Delta\hat{\mathbb{V}}_{NT,11}^{(a)} = O_p\left(K^{3/2}/T\right) = o_p(K)$. For $\Delta\hat{\mathbb{V}}_{NT,12}^{(a)}$, we have

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

https://doi.org/10.1017/S026646662300018X Published online by Cambridge University Press

$$\leq 8\gamma_{NT}T^{-1/2}\frac{1}{N}\sum_{i=1}^{N}\left\|\hat{\mathbb{Q}}_i\dot{\Omega}_i\hat{\mathbb{Q}}_i\right\|\left\|T^{-1/2}\sum_{t=1}^{T}\dot{Z}_{it}\dot{Z}_{it}'\breve{g}_{\Delta,it}^{(c)}\varepsilon_{it}\right\|$$

$$\leq 8\gamma_{NT}T^{-1/2}\max_i\left\|\hat{\mathbb{Q}}_i\dot{\Omega}_i\hat{\mathbb{Q}}_i\right\|\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T^{1/2}}\sum_{t=1}^{T}\dot{Z}_{it}\dot{Z}_{it}'\breve{g}_{\Delta,it}^{(c)}\varepsilon_{it}\right\|^2$$

$$= 8\gamma_{NT}T^{-1/2}O_p(K)\,O_p\left(K^2\right) = o_p(K).$$

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 the Cauchy–Schwarz inequality, we have

$$\Delta\hat{\mathbb{V}}_{NT,2}^{(a)} = \frac{8}{NT^2}\sum_{i=1}^{N}\sum_{1\leq s\neq t\leq T}\mathcal{K}_{i,ts}^2\breve{\varepsilon}_{R,it}\breve{\varepsilon}_{R,is}\varepsilon_{it}\varepsilon_{is}$$

$$\leq\left(\frac{8}{NT^2}\sum_{i=1}^{N}\sum_{1\leq s\neq t\leq T}\mathcal{K}_{i,ts}^4\varepsilon_{it}^2\varepsilon_{is}^2\right)^{1/2}\left(\frac{8}{NT^2}\sum_{i=1}^{N}\sum_{1\leq s\neq t\leq T}\breve{\varepsilon}_{R,it}^2\breve{\varepsilon}_{R,is}^2\right)^{1/2}$$

$$\leq\left(\frac{CK^4}{NT^2}\sum_{i=1}^{N}\sum_{1\leq s\neq t\leq T}A_{it}^2A_{is}^2\varepsilon_{it}^2\varepsilon_{is}^2\right)^{1/2}\left[\frac{C}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}\breve{\varepsilon}_{R,it}^2\right)^2\right]^{1/2}$$

$$= O_p\left(K^2\right)O_p\left(T^{-1}\right) = o_p(K).$$

Similarly, we can show $\Delta\hat{\mathbb{V}}_{NT,s}^{(a)} = o_p(K)$, for $s = 3,4,5$, by the Cauchy–Schwarz inequality. Hence, $\Delta\hat{\mathbb{V}}_{NT}^{(a)} = o_p(K)$. For $\Delta\hat{\mathbb{V}}_{NT}^{(b)}$, we have $E(\Delta\hat{\mathbb{V}}_{NT}^{(b)}) = 0$ and $\mathrm{Var}(\Delta\hat{\mathbb{V}}_{NT1}^{(b)}) = \frac{16}{N^2T^4}\sum_{i=1}^{N}\sum_{1\leq s_1\neq t_1\leq T}\sum_{1\leq s_2\neq t_2\leq T}\mathrm{Cov}(\mathring{\mathcal{K}}_{i,t_1s_1}^2\varepsilon_{it_1}^2\varepsilon_{is_2}^2,\mathring{\mathcal{K}}_{i,t_2s_2}^2\varepsilon_{it_2}^2\varepsilon_{is_2}^2) = O\left(K^4/N\right)$ by Assumption 1(ii). It follows that $\Delta\hat{\mathbb{V}}_{NT}^{(b)} = O_p(K^2/\sqrt{N}) = o_p(K)$. For $\Delta\hat{\mathbb{V}}_{NT}^{(c)}$ and $\Delta\hat{\mathbb{V}}_{NT}^{(d)}$, we can follow the proof of Lemma A.7 to show that they are both of order $o_p(K)$. □

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

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

$$= \frac{1}{NT^2}\sum_{i=1}^{N}\varepsilon_i'\mathcal{K}_i\varepsilon_i + \frac{1}{NT^2}\sum_{i=1}^{N}\breve{g}_{\Delta,i}'\mathcal{K}_i\breve{g}_{\Delta,i} + \frac{1}{NT^2}\sum_{i=1}^{N}\breve{\nu}_{NT}'X_i'\mathcal{K}_iX_i\breve{\nu}_{NT}$$

$$+ \frac{1}{NT^2}\sum_{i=1}^{N}2\varepsilon_i'\mathcal{K}_i\breve{g}_{\Delta,i} - \frac{1}{NT^2}\sum_{i=1}^{N}2\breve{g}_{\Delta,i}'\mathcal{K}_iX_i\breve{\nu}_{NT} - \frac{1}{NT^2}\sum_{i=1}^{N}2\varepsilon_i'\mathcal{K}_iX_i\breve{\nu}_{NT}$$

$$\equiv \sum_{l=1}^{6}\Gamma_{NT,l},\ \text{say.}$$

https://doi.org/10.1017/S026646662300018X Published online by Cambridge University Press

Clearly, we have (i) $\Gamma_{NT,1} = \frac{1}{NT^2}\sum_{i=1}^{N}\left\{\sum_{1\le s=t\le T}+\sum_{t=1}^{T}\sum_{s=1,\neq t}^{T}\right\}\varepsilon_{is}\varepsilon_{it}\mathcal{K}_{i,ts} = O_p\left(T^{-1}K\right)+O_p\left(N^{-1/2}T^{-1}K^{1/2}\right)$; (ii) $\Gamma_{NT,2} = \Phi_\Delta + o_p(1)$; and (iii) $\Gamma_{NT,3} \le \|\breve{\nu}_{NT}\|^2 = O_p((NT)^{-1}) + o_p(1)$. Then, by the Cauchy–Schwarz inequality, we have $\left|\Gamma_{NT,l}\right| = o_p(1)$, for $l = 4,5,6$. It follows that $\Gamma_{NT} = \Phi_\Delta + o_p(1)$ and $P(\Gamma_{NT} \ge \Phi_\Delta/2) \to 1$. 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 $\hat{\mathbb{B}}_{NT} = O_p\left(N^{1/2}K\right)$ under $\mathbb{H}_1$. It follows that

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

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

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 $\mathcal{W}_{NT}$, $\varepsilon_{r,it}^* = \hat{\varepsilon}_{r,it}\varrho_{it}$ 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. Note that in the bootstrap world, we have $E^*\left(\varepsilon_{r,it}^*\right) = \hat{\varepsilon}_{r,it}E(\varrho_{it}) = 0$, $E^*[\left(\varepsilon_{r,it}^*\right)^2] = \hat{\varepsilon}_{r,it}^2E\left(\varrho_{it}^2\right) = \hat{\varepsilon}_{r,it}^2$, $\hat{u}_{it}^* = -X_{it}'\nu_{NT}^*+\alpha_i+\varepsilon_{r,it}^*$ where $\nu_{NT}^* = [\sum_{i=1}^N X_i'M_TX_i]^{-1}\sum_{i=1}^N X_i'M_T\varepsilon_{r,i}^*$ and $\varepsilon_{r,i}^* = \left(\varepsilon_{r,i1}^*,\ldots,\varepsilon_{r,iT}^*\right)'$. These observations greatly simplify the proofs for the test in bootstrap world.

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

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

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^* denotes the weak convergence under bootstrap probability measure conditional on $\mathcal{W}_{NT}$;

https://doi.org/10.1017/S026646662300018X Published online by Cambridge University Press

(ii) $N^{1/2}T\Gamma_{NT}^{(*s)}/\mathbb{V}_{NT}^{*1/2} = o_{P^*}(1)$, for $s = 2,3$; (iii) $\hat{\mathbb{B}}_{NT}^* - \mathbb{B}_{NT}^* = o_{P^*}(K^{1/2})$; and (iv) $\hat{\mathbb{V}}_{NT}^*/\mathbb{V}_{NT}^* = 1 + o_{P^*}(1)$.

We only outline the proof for (i) as we can follow the proof of Theorem 3.2 to show (ii)–(iv). Write $\Gamma_{NT}^{(*1)} = \frac{1}{NT^2}\sum_{i=1}^N \sum_{1\le t\ne s\le T} \mathcal{K}_{i,ts}\varepsilon_{r,is}^*\varepsilon_{r,it}^* + \frac{1}{NT^2}\sum_{i=1}^N\sum_{t=1}^T \mathcal{K}_{i,tt}\left(\varepsilon_{r,it}^*\right)^2 \equiv \Gamma_{NT}^{(*1a)} + \Gamma_{NT}^{(*1b)}$, 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), we write $J_{NT}^{(*a)} = \frac{1}{\sqrt{N}}\sum_{i=1}^N \mathcal{Z}_i^*$ with $\mathcal{Z}_i^* = \frac{2}{T\mathbb{V}_{NT}^{*1/2}}\sum_{1\le t<s\le T}\breve{\mathcal{K}}_{i,ts}\varrho_{it}\varrho_{is}$ 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 (i.n.i.d.) across i conditional on $\mathcal{W}_{NT}$, we prove (ia) by the Linderberg–Feller CLT conditional on $\mathcal{W}_{NT}$. The proof of (ib) is simple and is omitted here. To show (ia), it suffices to show that (ia.1) $\bar{\sigma}_N^{*2} = N\text{Var}^*(N^{-1}\sum_{i=1}^N \mathcal{Z}_i^*) = \text{Var}(J_{NT}^{(*a)}|\mathcal{W}_{NT}) = 1$; and (ia.2) $E^*\left(\mathcal{Z}_i^4\right) \le C < \infty$ for all i. For (ia.1), noting that ϱ_{it}'s are i.i.d. across i and along t, we have

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

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

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

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

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

https://doi.org/10.1017/S026646662300018X Published online by Cambridge University Press

First, DJ^*_{i4a} can be written as

$$DJ^*_{i4a} = \frac{1}{T^4\mathbb{V}^{*2}_{NT}}\left(\sum_{1\le t,s\le T}\breve{\mathcal{K}}^2_{i,ts}\right)^2 = \frac{1}{T^4\mathbb{V}^{*2}_{NT}}\left(\sum_{1\le t,s\le T}\dot{Z}_{it}\hat{Q}^{-1}_{i,\dot{z}\dot{z}}\hat{Q}_{i,zz}\hat{Q}^{-1}_{i,\dot{z}\dot{z}}\dot{Z}_{is}\hat{\varepsilon}_{r,it}\hat{\varepsilon}_{r,is}\right)^2$$

$$= \frac{1}{\mathbb{V}^{*2}_{NT}}\left[\operatorname{tr}\left(\hat{\Omega}_i\hat{Q}^{-1}_{i,\dot{z}\dot{z}}\hat{Q}_{i,zz}\hat{Q}^{-1}_{i,\dot{z}\dot{z}}\hat{\Omega}_i\hat{Q}^{-1}_{i,\dot{z}\dot{z}}\hat{Q}_{i,zz}\hat{Q}^{-1}_{i,\dot{z}\dot{z}}\right)\right]^2$$

$$\le \frac{1}{\mathbb{V}^{*2}_{NT}}\left[\lambda^2_{\max}\left(\hat{Q}^{-1}_{i,\dot{z}\dot{z}}\hat{Q}_{i,zz}\hat{Q}^{-1}_{i,\dot{z}\dot{z}}\right)\lambda_{\max}\left(\hat{\Omega}_i\right)\operatorname{tr}\left(\hat{\Omega}_i\right)\right]^2$$

$$\lesssim \frac{1}{K^2\lambda^4_{\min}\left(\hat{\Omega}_i\right)}\left\{\left[\lambda^2_{\max}\left(Q^{-1}_{i,\dot{z}\dot{z}}Q_{i,zz}Q^{-1}_{i,\dot{z}\dot{z}}\right)+o_p(1)\right]\lambda^2_{\max}\left(\hat{\Omega}_i\right)K\right\}^2$$

$$\lesssim O_{P^*}\left(K^{-2}\right)O_p\left(K^2\right) = O_{P^*}(1).$$

Second, we have

$$DJ^*_{i4b} = \frac{1}{T^4\mathbb{V}^{*2}_{NT}}\sum_{t\ne s\ne l\ne q}\tilde{\mathcal{K}}_{i,ts}\tilde{\mathcal{K}}_{i,tl}\tilde{\mathcal{K}}_{i,lq}\tilde{\mathcal{K}}_{i,qs}$$

$$\lesssim \frac{1}{T^4\mathbb{V}^{*2}_{NT}}\sum_{t\ne s\ne l\ne q}\hat{\varepsilon}_{r,is}\dot{Z}'_{is}\hat{\mathbb{Q}}_i\hat{\varepsilon}^2_{r,it}\dot{Z}_{it}\dot{Z}'_{it}\hat{\mathbb{Q}}_i\hat{\varepsilon}^2_{r,il}\dot{Z}_{il}\hat{\mathbb{Q}}_i\hat{\varepsilon}^2_{r,iq}\dot{Z}_{iq}\hat{\mathbb{Q}}_i\dot{Z}_{is}\hat{\varepsilon}_{r,is}$$

$$\lesssim \frac{1}{\mathbb{V}^{*2}_{NT}}\operatorname{tr}\left(\hat{\Omega}_i\hat{\mathbb{Q}}_i\hat{\Omega}_i\hat{\mathbb{Q}}_i\hat{\Omega}_i\hat{\mathbb{Q}}_i\hat{\Omega}_i\hat{\mathbb{Q}}_i\right) \le \frac{1}{\mathbb{V}^{*2}_{NT}}\lambda^3_{\max}\left(\hat{\Omega}_i\right)\lambda^4_{\max}\left(\hat{\mathbb{Q}}_i\right)\operatorname{tr}\left(\hat{\Omega}_i\right)$$

$$\le \frac{1}{K^2\lambda^4_{\min}\left(\hat{\Omega}_i\right)}\lambda^3_{\max}\left(\hat{\Omega}_i\right)K\lambda_{\max}\left(\hat{\Omega}_i\right)\lambda^4_{\max}\left(\hat{\mathbb{Q}}_i\right) = O_{P^*}\left(K^{-1}\right) < \infty.$$

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 $\mathcal{W}_{NT}$. It follows that (ia.2) holds. Then we have shown (ia).

SUPPLEMENTARY MATERIAL

Alev Atak, Thomas Tao Yang, Yonghui Zhang, Qiankun Zhou (March 2023): Online Appendix to "SPECIFICATION TESTS FOR TIME-VARYING COEFFICIENT PANEL DATA MODELS,"Econometric Theory Supplementary Material. To view, please visit: https://doi.org/10.1017/S026646662300018X

REFERENCES

Atak, A., O. Linton, & 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.

Bada, O., A. Kneip, D. Liebl, T. Mensinger, J. Gualtieri, & R.C. Sickles (2022) A wavelet method for panel models with jump discontinuities in the parameters. *Journal of Econometrics* 226, 399–422.

Bai, J. (2009) Panel data models with interactive fixed effects. *Econometrica* 74, 1229–1279.

https://doi.org/10.1017/S026646662300018X Published online by Cambridge University Press

Baltagi, B.H. (2012) *Econometric Analysis of Panel Data*. Wiley.
Bartolucci, F., F. Belotti, & F. Peracchi (2015) Testing for time-invariant unobserved heterogeneity in generalized linear models for panel data. *Journal of Econometrics* 184, 111–123.
Boneva, L., O. Linton, & M. Vogt (2015) A semiparametric model for heterogeneous panel data with fixed effects. *Journal of Econometrics* 188, 327–345.
Bosq, D. (1998) *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction*, *2nd Edition.* Springer.
Cai, Z. (2007) Trending time-varying coefficients time series models with serially correlated errors. *Journal of Econometrics* 136, 163–188.
Cai, Z., J. Fan, & Q. Yao (2000) Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association* 95, 941–956.
Chen, B. & Y. Hong (2012) Testing for smooth structural changes in time series models via nonparametric regression. *Econometrica* 80, 1157–1183.
Chen, B. & L. Huang (2018) Nonparametric testing for smooth structural changes in panel data models. *Journal of Econometrics* 202, 245–267.
Chen, J., J. Gao, & D. Li (2012) Semiparametric trending panel data models with cross-sectional dependence. *Journal of Econometrics* 171, 71–85.
Chernozhukov, V. & I. Fernández-Val (2011) Inference for extremal conditional quantile models, with an application to market and birthweight risks. *The Review of Economic Studies* 78, 559–589.
Dong, C. & O. Linton (2018) Additive nonparametric models with time variable and both stationary and nonstationary regressors. *Journal of Econometrics* 207, 212–236.
Fan, J., Y. Liao, & J. Yao (2015) Power enhancement in high-dimensional cross-sectional tests. *Econometrica* 83, 1497–1541.
Feng, G., J. Gao, B. Peng, & 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.
Gao, J. & I. Gijbels (2008) Bandwidth selection in nonparametric kernel testing. *Journal of the American Statistical Association* 103, 1584–1594.
Gao, J. & M. King (2004) Adaptive testing in continuous-time diffusion models. *Econometric Theory* 20, 844–882.
Gao, J., K. Xia, & H. Zhu (2020) Heterogeneous panel data models with cross-sectional dependence. *Journal of Econometrics* 219(2), 329–353.
Grossmann, G.M. & A.B. Krueger (1993) Environmental impacts of a North American free trade agreement. In P. Garber (ed.), *The Mexico–US Free Trade Agreement*, pp. 13–56. MIT Press.
Grossmann, G.M. & A.B. Krueger (1995) Economic growth and the environment. *Quarterly Journal of Economics* 110, 353–377.
Hansen, B. (2000) Testing for structural change in conditional models. *Journal of Econometrics* 97(1), 93–115.
Hidalgo, J. & J. Lee (2014) A CUSUM test for common trends in large heterogeneous panels. In *Essays in Honor of Peter C.B. Phillips*. Advances in Econometrics, vol. 33, pp. 303–345. Emerald Publishing Limited.
Horowitz, J.L. & V.G. Spokoiny (2003) An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica* 69, 599–631.
Hsiao, C. (1985) Benefits and limitations of panel data. *Econometric Reviews* 4, 121–174.
Hsiao, C. (1995) Panel analysis for metric data. In G. Arminger, C. C. Clogg, & M. E. Sobel (eds.), *Handbook of Statistical Modelling in the Social and Behavioral Sciences*. Springer.
Hsiao, C. (2014) *Analysis of Panel Data*, *3rd Edition*. Cambridge University Press.
Hsiao, C. (2018) Panel models with interactive effects. *Journal of Econometrics* 206, 645–673.
Hsiao, C. & M.H. Pesaran (2008) Random coefficient models. In L. Matyas & P. Sevestre (eds.), *The Econometrics of Panel Data*, pp. 185–213. Springer.
Hsiao, C., Z. Shi, & Q. Zhou (2022) Transformed estimator for panel interactive effects models. *Journal of Business & Economic Statistics* 40, 1831–1848.
Hsiao, C. & A.K. Tahmiscioglu (1997) A panel analysis of liquidity constraints and firm investment. *Journal of the American Statistical Association* 92, 455–465.

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

Kneip, A., R.C. Sickles, & W. Song (2012) A new panel data treatment for heterogeneity in time trends. *Econometric Theory* 28(3), 590–628.

Lee, J. & P.M. Robinson (2016) Series estimation under cross-sectional dependence. *Journal of Econometrics* 188, 346–362.

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

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

Li, D., L. Su, & 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. & S. Wang (1998) A simple bootstrap test for a parametric regression function. *Journal of Econometrics* 87, 145–165.

Liu, F., J. Gao, & Y. Yang (2020) Time-Varying Panel Data Models with an Additive Factor Structure. Working Paper 42/20, Monash University.

Lombardia, M.J. & S. Sperlich (2008) Semiparametric inference in generalized mixed effects models. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* 70(5), 913–930.

Lu, X. & L. Su (2022) Uniform inference in linear panel data models with two-dimensional heterogeneity. *Journal of Econometrics*, forthcoming.

Lumsdaine, R.L., R. Okui, & W. Wang (2023) Estimation of panel group structure models with structural breaks in group memberships and coefficients. *Journal of Econometrics* 233(1), 45–65.

Ma, S., W. Lan, L. Su, & C.-L. Tsai (2020) Testing alphas in conditional time-varying factor models with high dimensional assets. *Journal of Business & Economic Statistics* 38(1), 214–227.

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

Moon, H.R. & M. Weidner (2015) Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica* 83, 1543–1579.

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

Pesaran, M.H. (2006) Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica* 74(4), 967–1012.

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

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

Robinson, P.M. (2012) Nonparametric trending regression with cross-sectional dependence. *Journal of Econometrics* 169, 4–14.

Shao, Q.M. & H. Yu (1996) Weak convergence for weighted empirical processes of dependent sequences. *The Annals of Probability* 24, 2098–2127.

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

Su, L., Z. Shi, & P.C.B. Phillips (2016) Identifying latent structures in panel data. *Econometrica* 84(6), 2215–2264.

Su, L., X. Wang, & S. Jin (2019) Sieve estimation of time-varying panel data models with latent structures. *Journal of Business & Economic Statistics* 37, 334–349.

Su, L. & Y. Zhang (2016) Semiparametric estimation of partially linear dynamic panel data models with fixed effects. In *Essays in Honor of Aman Ullah*. Advances in Econometrics, vol. 36, pp. 137–204. Emerald Publishing Limited.

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

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

Wooldridge, J.M. (2010) *Econometric Analysis of Cross Section and Panel Data*, *2nd Edition*. MIT Press Academic.
Zhang, Y., L. Su, & P.C.B. Phillips (2012) Testing for common trends in semiparametric panel data models with fixed effects. *The Econometrics Journal* 15, 56–100.
Zhang, Y. & Q. Zhou (2021) Partially linear functional-coefficient dynamic panel data models: Sieve estimation and specification testing. *Econometric Reviews* 40, 983–1006.
Zhao, H., W. Tu, & Z. Yu (2018) A nonparametric time-varying model for panel count data. *Journal of Nonparametric Statistics* 30(3), 640–661.