Journal of Econometrics 227 (2022) 114–133

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Testing capital asset pricing models using functional-coefficient panel data models with cross-sectional dependence

Zongwu Cai [a], Ying Fang [b], Qiuhua Xu [c,*]

[a] Department of Economics, University of Kansas, Lawrence, KS 66045, USA
[b] Wang Yanan Institute for Studies in Economics, Ministry of Education Key Laboratory of Econometrics and Fujian Key Laboratory of Statistical Sciences, Xiamen University, Xiamen, Fujian 361005, China
[c] School of Finance, Southwestern University of Finance and Economics, Chengdu, Sichuan 611130, China

A R T I C L E I N F O

Article history:
Received 2 November 2018
Received in revised form 2 December 2019
Accepted 20 July 2020
Available online 7 August 2020

JEL classification:
C12
C13
C14
C23

Keywords:
Cross-sectional dependence
Functional coefficients
Local linear estimation
Nonlinear panel data models
Nonparametric test

A B S T R A C T

This paper proposes a functional-coefficient panel data model with cross-sectional dependence motivated by re-examining the empirical performance of conditional capital asset pricing model. In order to characterize the time-varying property of assets' betas and alpha, our proposed model allows the betas to be unknown functions of some macroeconomic and financial instruments. Moreover, a common factor structure is introduced to characterize cross-sectional dependence which is an attractive feature under a panel data regression setting as different assets or portfolios may be affected by same unobserved shocks. Compared to the existing studies, such as the classic Fama–MacBeth two-step procedure, our model can achieve substantial efficiency gains for inference by adopting a one-step procedure using the entire sample rather than a single cross-sectional regression at each time point. We propose a local linear common correlated effects estimator for estimating time-varying betas by pooling the data. The consistency and asymptotic normality of the proposed estimators are established. Another methodological and empirical challenge in asset pricing is how to test the constancy of conditional betas and the significance of pricing errors, we echo this challenge by constructing an L_2-norm statistic for functional-coefficient panel data models allowing for cross-sectional dependence. We show that the new test statistic has a limiting standard normal distribution under the null hypothesis. Finally, the method is applied to test the model in Fama and French (1993) using Fama–French 25 and 100 portfolios, sorted by size and book-to-market ratio, respectively, dated from July 1963 to July 2018.

1. Introduction

A central issue in theoretical and empirical finance literature is how to explain the cross-sectional variation in assets' expected returns. The standard method, for example, introduced by Fama and MacBeth (1973), consists of a two-step procedure. In the first step, a cross-sectional regression is run by excess returns on multiple risk factors at each period with an intercept, and in the second step, the time series averages of these cross-sectional regression coefficients are

* Corresponding author.
E-mail address: xuqh@swufe.edu.cn (Q. Xu).

https://doi.org/10.1016/j.jeconom.2020.07.018
0304-4076/

computed and tests are conducted to examine whether these time series averages significantly differ from zero. The advantage of the Fama and MacBeth procedure is that it allows us to easily include many risk factors in the model. However, the standard procedure restricts the relation between expected returns and risk factors to be linear. In other words, the Fama and MacBeth procedure assumes all regression coefficients to be constant.

Indeed, it has been well recognized by many studies that the betas of risk factors in the capital asset pricing models depend on the state of economy; see, for example, Fama and French (1997), Ferson and Harvey (1999), Lettau and Ludvigson (2001), Zhang (2005), Lewellen and Nagel (2006), Cai et al. (2015b), Guo et al. (2017), and among others. An influential paper by Ferson and Harvey (1999) argued to adopt a set of lagged economy-wide instruments to capture common dynamic patterns in returns. Both pricing errors (the alphas) and betas were set as a linear function of these lagged instruments. Furthermore, some recent studies focused on estimating alphas and betas semi-parametrically or nonparametrically. For example, Ferreira et al. (2011) and Ang and Kristensen (2012) considered the estimation of time-varying betas, where betas are regarded as an unknown smooth function of time, and other studies treated betas as an unknown function of state variables. To avoid the curse of dimensionality problem in a nonparametric setting, Connor et al. (2012) considered an additive nonparametric regression model, while Cai et al. (2015b) and Guo et al. (2017) used single index functional-coefficient models, which include the model in Ferson and Harvey (1999) as a special case. But, both Cai et al. (2015b) and Guo et al. (2017) considered only time series model with ignoring the dependence among individuals.

Another deficiency of most existing studies is to estimate a time series of betas only using single cross-sectional data at each time period and then computing a time series average using a same weight for all cross sections. Such a procedure causes two problems. Obviously, using a single cross section and then taking time series averages with same weights might lose estimation efficiency. On the other hand, running cross section regressions separately is hard to control possible cross-sectional relations among error terms, which may immensely distort estimation and testing results.

We propose a new model to deal with the aforementioned issues: the possible misspecification of betas and alpha and the ignorance of potential cross-sectional dependence, in a simultaneous way. To this end, a functional-coefficient panel data model with cross-sectional dependence is considered so that our model allows all coefficients, both the pricing error alpha and the beta coefficients of risk factors, to depend on macroeconomic or financial instruments and be estimated using the pool data rather than a single cross section, which should reach more powerful results for estimation and testing. In addition, the model allows the error term to include an unobserved factor structure which is used to capture cross-sectional dependence. Ferreira et al. (2011) considered to estimate betas using a system of equations allowing for cross-sectional relations. Our model differs from theirs in not only capturing cross-sectional dependence but also allowing for possible correlations between risk factors and other unobserved factors. Finally, we propose a local linear common correlated effect estimator for both conditional alphas and betas and construct a nonparametric L_2 type statistic for testing conditional alphas and betas in a panel data model framework, together with establishing its asymptotic properties.

Our paper is closely related to the econometric literature on nonparametric and semiparametric panel data models which impose relatively little restriction on model's structures; see, for example, to name just a few, Cai and Li (2008), Sun et al. (2009), Cai et al. (2015a), Sun et al. (2015), and Cai et al. (2018). However, the aforementioned literature focused on nonparametric or semiparametric inference by assuming cross-sectional independence. Recently, studies on panel data models with cross-sectional dependence have received increased attentions; for example, see the papers by Pesaran (2006), Bai (2009), and Moon and Weidner (2015), and among others. While there is relatively rich literature on linear panel data models with cross-sectional dependence, to the best of our knowledge, little work has been done in estimating nonparametric or semiparametric panel data models with cross-sectional dependence. For example, Su and Jin (2012) extended the work of Pesaran (2006) to a nonparametric heterogeneous panel data model with factor structure and proposed a sieve estimation for their model and Huang (2013) considered a similar nonparametric panel data model with fixed effect and factor structure. Note that a sieve method is a global parametric approximation, as far as testing for local curvature of a smooth function, which is the focus of many empirical studies, it is natural to use a local estimator (such as that of functional-coefficient models) to conduct a statistical testing. For details on the nonparametric or semiparametric panel data models with cross-sectional dependence, readers are referred to the recent survey paper by Xu et al. (2016). Our model can be regarded as an extension of Pesaran (2006) which contributes to the literature by proposing a kernel based estimator for functional coefficients and a nonparametric test in a panel data model with cross-sectional dependence. In particular, our model can be applied to making inferences on capital asset pricing models incorporating the dependence among individuals.

The rest of the paper is organized as follows. Section 2 introduces the model and discusses the estimation method. Asymptotic properties of proposed estimators are also established. In Section 3, a nonparametric test for testing constancy in functional coefficients is proposed, together with providing not only the asymptotic distributions of the proposed statistic but also a simply-implemented Bootstrap procedure to improve finite sample performance. Monte Carlo simulation results are presented in Section 4. Section 5 revisits the empirical performance of the Fama–French model using Fama–French 25 and 100 portfolios formed on size and book-to-market ratio dated from July 1963 to July 2018. Finally, a concluding remark is given in Section 6. All proofs for the theorems are provided in the supplement (see Appendix).

2. Estimation methods and asymptotic theories

In this section, the model is first introduced, and then the idea of local linear common correlated effect (CCE) estimation is presented for both heterogeneous and homogeneous functional coefficients. Finally, some theoretical assumptions are given and the asymptotic properties of proposed estimators are stated.

2.1. Model setup

To be specific, let $\{Y_{it}, X_{it}, U_{it}\}$ be the observed data on the ith cross-sectional unit at time t for $1 \le i \le N$ and $1 \le t \le T$. The following general functional-coefficient panel data model is considered

$$Y_{it} = \beta_i^\top(U_{it})X_{it} + \gamma_{1i}^\top f_{1t} + e_{it}, \tag{2.1}$$

where $A^\top$ denotes the transpose of A, $X_{it} \in \mathbb{R}^p$ is a vector of individual-specific explanatory variables on the ith cross-sectional unit at time t, $\beta_i(\cdot)$ is a vector of smooth functions defined on $\mathbb{R}^{du}$ which has a continuous second derivative and may take different functional forms for different individuals, $U_{it} \in \mathbb{R}^{du}$ is a smooth variable,[1] f_{1t} is an $m_1 \times 1$ vector of observed common factors, and $\{\gamma_{1i},\ i = 1, 2, \ldots, N\}$ are factor loadings. The intercept term can be included in X_{it} or f_{1t}. However, they cannot both contain an intercept for the purpose of identification. To characterize cross-sectional dependence, it is conventional to assume that e_{it} follows the following multi-factor structure

$$e_{it} = \gamma_{2i}^\top f_{2t} + \varepsilon_{it}, \tag{2.2}$$

where f_{2t} is an $m_2 \times 1$ vector of unobserved common factors, $\{\gamma_{2i},\ i = 1, 2, \ldots, N\}$ are factor loadings, and ε_{it} is the idiosyncratic error of Y_{it}. In general, the unobserved factors f_{2t} could be correlated with the observed data (X_{it}, U_{it}, f_{1t}). To allow for such a possibility, similar to Pesaran (2006), the following fairly general model is adopted,

$$\boldsymbol{\omega}_{it} \equiv \begin{pmatrix} X_{it} \\ U_{it} \\ Z_{it} \end{pmatrix} = \Gamma_{1i}^\top f_{1t} + \Gamma_{2i}^\top f_{2t} + v_{it}, \tag{2.3}$$

where Z_{it} is $p_z \times 1$ vector of covariates specific to unit i, Γ_{1i} and Γ_{2i} are $m_1 \times (p + p_z + 1)$ and $m_2 \times (p + p_z + 1)$ factor loading matrices, and $\{v_{it}\}$ are the error for $\boldsymbol{\omega}_{it}$. Of interest for model (2.1) is to estimate nonparametric functionals $\beta_i(\cdot)$ and γ_{1i} is treated as a nuisance parameter.

In some applications, one might be interesting in considering a restricted model of (2.1) as follows:

$$Y_{it} = \beta^\top(U_{it})X_{it} + \gamma_{1i}^\top f_{1t} + e_{it}. \tag{2.4}$$

That is, $\beta_i(\cdot) = \beta(\cdot)$, where $\beta(\cdot)$ is assumed to have a continuous second derivative, for all i in model (2.1). Following Pesaran (2006), the functional-coefficient functions are called to be homogeneous when $\beta_i(\cdot) = \beta(\cdot)$ for all i and heterogeneous otherwise.

A multi-factor asset pricing model introduced in the introduction can be regarded as a special case of Eq. (2.1). For example, Y_{it} can be an asset or portfolio excess return, X_{it} denotes a vector of risk factors, and U_{it} can be a vector of economy-wide state variables. Particularly, in the Fama–French three-factor model, X_{it} includes market risk, size risk and value risk factors. Note that the model allows for cross-sectional dependence among the errors e_{it}, and furthermore, the model also allows for possible correlations between risk factors and unobserved factors f_{2t}. Details on how to apply model (2.1) to making inferences on capital asset pricing models to analyze real examples are presented in Section 5.

To estimate both the heterogeneous functional coefficients $\beta_i(\cdot)$ in (2.1) and the homogeneous functional coefficient $\beta(\cdot)$ in (2.4), instead of using any global parametric approximation approaches, we propose a locally (kernel method) common correlated effect estimator (LCCE) for the heterogeneous case by combining the CCE approach with the local linear estimation method. For homogeneous case, we propose a locally common correlated effect pooled estimator (LCCEP) to improve estimation efficiency by pooling the data. It will be shown that as both N and T go to infinity, the LCCE and LCCEP estimators are not only consistent but also asymptotically normally distributed under certain mild conditions. Moreover, a nonparametric goodness-of-fit statistic is proposed for testing the constancy of functional coefficients. This is of interests in practice since the smoothing variables are often selected based on certain economic hypothesis. A constancy test here amounts to testing economic theories.

2.2. Local linear estimation of $\beta_i(\cdot)$ and $\beta(\cdot)$

Following the idea of CCE estimation proposed by Pesaran (2006), first, unobserved factor f_{2t} is approximated by the cross-sectional averages of observable variables in Eq. (2.3). To be specific, let $\overline{\boldsymbol{\omega}}_t \equiv N^{-1}\sum_{i=1}^N \boldsymbol{\omega}_{it}$. By the same token, one can define $\overline{\Gamma}_1, \overline{\Gamma}_2$ and $\overline{v}_t$ as cross-section averages of Γ_{1i}, Γ_{2i} and v_{it}, respectively. Then, (2.3) implies that

$$\overline{\boldsymbol{\omega}}_t = \overline{\Gamma}_1^\top f_{1t} + \overline{\Gamma}_2^\top f_{2t} + \overline{v}_t. \tag{2.5}$$

Premultiplying both sides of (2.5) by $\overline{\Gamma}_2$ and solving for f_{2t} implies that

$$f_{2t} = \left(\overline{\Gamma}_2\overline{\Gamma}_2^\top\right)^{-1}\overline{\Gamma}_2\left(\overline{\boldsymbol{\omega}}_t - \overline{\Gamma}_1^\top f_{1t} - \overline{v}_t\right),$$

[1] For simplicity, we only consider the case $d_u = 1$ in (2.1). Extension to the case $d_u > 1$ involves no fundamentally new ideas. Also, note that models with large d_u are not practically useful due to the so-called "curse of dimensionality". Usually, $d_u \le 3$ in real application.

provided that $rank(\overline{\Gamma}_2) = m_2 \leq p + p_z + 1$ for sufficiently large N, which is imposed by Pesaran (2006). For each t, as $N \to \infty$, $\overline{v}_t \xrightarrow{p} 0$ under some weak conditions. Then, it follows that

$$f_{2t} - \left(\overline{\Gamma}_2\overline{\Gamma}_2^\top\right)^{-1}\overline{\Gamma}_2\left(\overline{\boldsymbol{\omega}}_t - \overline{\Gamma}_1^\top f_{1t}\right) \xrightarrow{p} 0$$

as $N \to \infty$, which suggests using $q_t = \left(f_{1t}^\top, \overline{\boldsymbol{\omega}}_t^\top\right)^\top$ as a proxy for f_{2t}. Note that it is not necessary to use all variables in $\boldsymbol{\omega}_t$ to filter unobserved common factors. For example, Hjalmarsson (2010) only adopted the included explanatory variables.

To estimate the heterogeneous coefficients $\beta_i(\cdot)$, the following augmented regression is considered

$$Y_{it} = \beta_i^\top(U_{it})X_{it} + \vartheta_i^\top q_t + \varepsilon_{it}^* \tag{2.6}$$

where ε_{it}^* is the new error term which includes ε_{it} and the approximation error coming from replacing f_{2t} by q_t. Indeed, (2.6) is a partially varying-coefficient model, which is usually estimated by a profile likelihood method or other two-step semiparametric methods in order to achieve parametric convergence rate for the parametric part; see, for example, Fan and Huang (2005). However, in model (2.6), ϑ_i is a nuisance parameter which is lack of interest, so that it simply projects the term $\vartheta_i^\top q_t$ out by using the idea of partitioned regression. Let $Y_i = (Y_{i1}, \ldots, Y_{iT})^\top$, $Q = (q_1, \ldots, q_T)^\top$ and $\varepsilon_i^* = \left(\varepsilon_{i1}^*, \ldots, \varepsilon_{iT}^*\right)^\top$. Rewriting (2.6) in a vector form, one has

$$Y_i = \begin{pmatrix} \beta_i^\top(U_{i1})X_{i1} \\ \vdots \\ \beta_i^\top(U_{iT})X_{iT} \end{pmatrix} + Q\vartheta_i + \varepsilon_i^*. \tag{2.7}$$

To estimate the unknown function $\beta_i(\cdot)$, the local linear estimation method is employed. Given $u_0 \in \mathbb{R}$, if $|U_{it} - u_0| \leq h$, $t = 1, 2, \ldots, T$, using first order Taylor's expansion, one obtains

$$\begin{pmatrix} \beta_i^\top(U_{i1})X_{i1} \\ \vdots \\ \beta_i^\top(U_{iT})X_{iT} \end{pmatrix} = \tilde{X}_i\beta_i^*(u_0) + O_p(h^2), \tag{2.8}$$

where $\tilde{X}_i$ is the $T \times 2p$ matrix whose tth row is $(X_{it}^\top, X_{it}^\top(U_{it} - u_0/h))$, and $\beta_i^*(u_0) = \left(\beta_i^\top(u_0), h\beta_i'^\top(u_0)\right)^\top$ with $\beta_i'(\cdot)$ denoting the first order derivative of $\beta_i(\cdot)$. By substituting (2.8) into (2.7) and ignoring the higher order term $O_p(h^2)$, it is easy to see that the estimator of $\beta_i^*(u_0)$, denoted by $\hat{\beta}_i^*(u_0)$, can be obtained by minimizing the following sum of locally weighted squares,

$$\hat{\beta}_i^*(u_0) = \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} = \arg\min_{a,b} \sum_{t=1}^{T} [Y_{it} - X_{it}^\top(a + b(U_{it} - u_0)/h) - \vartheta_i^\top q_t]^2 k_h(U_{it} - u_0),$$

where $k(\cdot)$ is a kernel function and $k_h(\cdot) = k(\cdot/h)/h$. Note that the above local linear estimator can be viewed as a weighted OLS estimator of a working linear model

$$w_{it}Y_{it} = w_{it}(X_{it}^\top, X_{it}^\top(\frac{U_{it} - u_0}{h}))\beta_i^*(u_0) + w_{it}q_t^\top\vartheta_i + w_{it}\varepsilon_{it}^*, \tag{2.9}$$

where $w_{it} = \sqrt{k_h(U_{it} - u_0)}$. Clearly, (2.9) can be re-written in a vector form as

$$\mathbf{W}_{i,h}^{1/2}(u_0)Y_i = \mathbf{W}_{i,h}^{1/2}(u_0)\widetilde{X}_i\beta_i^*(u_0) + \mathbf{W}_{i,h}^{1/2}(u_0)Q\vartheta_i + \mathbf{W}_{i,h}^{1/2}(u_0)\varepsilon_i^*$$

which is a linear regression model, where $\mathbf{W}_{i,h}(u_0) = \mathrm{diag}(w_{i1}^2, \ldots, w_{iT}^2)$. By Frisch–Waugh–Lovell Theorem in a linear regression, one can obtain the following locally common correlated effect (LCCE) estimator,

$$\hat{\beta}_i^*(u_0) = [\widetilde{X}_i^\top\mathbf{W}_{i,h}^{1/2}(u_0)M_i(u_0)\mathbf{W}_{i,h}^{1/2}(u_0)\widetilde{X}_i]^{-1}\widetilde{X}_i^\top\mathbf{W}_{i,h}^{1/2}(u_0)M_i(u_0)\mathbf{W}_{i,h}^{1/2}(u_0)Y_i, \tag{2.10}$$

where $M_i(u_0) = I_T - \mathbf{W}_{i,h}^{1/2}(u_0)Q[Q^\top\mathbf{W}_{i,h}(u_0)Q]^{-1}Q^\top\mathbf{W}_{i,h}^{1/2}(u_0)$ and I_K is a $K \times K$ identity matrix.

In sum, Eq. (2.6) is a transformation of Eq. (2.1) by approximating the unobservable factor f_{2t} with cross-sectional averages of observables. Estimating the workable model (2.6) instead of the original model (2.1) incurs some approximation error. However, as one can see in Lemma A.1.2 in the appendix, the error is of order $O_p(1/N)$ which is negligible when $N \to \infty$.

Now, it turns to the estimation of the homogeneous coefficients $\beta(\cdot)$. In order to improve estimation efficiency for the homogeneous coefficients, $\beta(\cdot)$ is estimated by pooling all data together. Consider the following augmented regression in a vector form

$$Y_i = \begin{pmatrix} \beta^\top(U_{i1})X_{i1} \\ \vdots \\ \beta^\top(U_{iT})X_{iT} \end{pmatrix} + Q\vartheta_i + \varepsilon_i^*.$$

For ease of notations, ε_i^* is also used to denote the new error term as in (2.7). Using the first order Taylor's expansion, Y_i can be approximated locally by

$$Y_i \approx \tilde{X}_i\beta^*(u_0) + Q\vartheta_i + \varepsilon_i^*,$$

where $\beta^*(u_0) = \left(\beta^\top(u_0), h\beta'^\top(u_0)\right)^\top$. Let $Y = \left(Y_1^\top, \dots, Y_N^\top\right)^\top$, $\tilde{X} = \left(\tilde{X}_1^\top, \dots, \tilde{X}_N^\top\right)^\top$, $\mathcal{Q} = I_N \otimes Q$, $\vartheta = \left(\vartheta_1^\top, \dots, \vartheta_N^\top\right)^\top$, and $\varepsilon^* = \left(\varepsilon_1^{*\top}, \dots, \varepsilon_N^{*\top}\right)^\top$, where $\otimes$ denotes the Kronecker product. By pooling all data together, one has

$$Y \approx \tilde{X}\beta^*(u_0) + \mathcal{Q}\vartheta + \varepsilon^*. \tag{2.11}$$

Using the same argument as in deriving (2.10), it is easy to show that with $\mathbf{W}_h(u_0) = \text{diag}(\mathbf{W}_{1,h}(u_0), \dots, \mathbf{W}_{N,h}(u_0))$,

$$\begin{aligned}\hat{\beta}^*(u_0) &= \left[\widetilde{X}^\top\mathbf{W}_h^{1/2}(u_0)M(u_0)\mathbf{W}_h^{1/2}(u_0)\widetilde{X}\right]^{-1}\widetilde{X}^\top\mathbf{W}_h^{1/2}(u_0)M(u_0)\mathbf{W}_h^{1/2}(u_0)Y \\ &= \left[\sum_{i=1}^N \widetilde{X}_i^\top\mathbf{W}_{i,h}^{1/2}(u_0)M_i(u_0)\mathbf{W}_{i,h}^{1/2}(u_0)\widetilde{X}_i\right]^{-1}\left[\sum_{i=1}^N \widetilde{X}_i^\top\mathbf{W}_{i,h}^{1/2}(u_0)M_i(u_0)\mathbf{W}_{i,h}^{1/2}(u_0)Y_i\right],\end{aligned} \tag{2.12}$$

where $M(u_0) = I_{NT} - \mathbf{W}_h^{1/2}(u_0)\mathcal{Q}[\mathcal{Q}^\top\mathbf{W}_h(u_0)\mathcal{Q}]^{-1}\mathcal{Q}^\top\mathbf{W}_h^{1/2}(u_0)$. This is termed as the locally common correlated effect pooled (LCCEP) estimator for homogeneous functional coefficients.

2.3. Large sample theories

2.3.1. Notations and assumptions

To simplify notations used in the analysis, $f_i(U_{it})$ is used to denote the marginal density of U_{it}, $f_{i1t}(U_{i1}, U_{it})$ to be the joint density of U_{i1} and U_{it}, and $f_i(U_{it}, \xi_{it})$ to be the joint density of U_{it} and ξ_{it}, where $\xi_{it} \in \{X_{it}, v_{it}\}$. Let $\mu_j = \int_{-\infty}^{\infty} u^j k(u)du$, $\lambda_j = \int_{-\infty}^{\infty} u^j k^2(u)du$, $\eta_t = \left(f_{1t}^\top, f_{2t}^\top\right)^\top$, $\eta = (\eta_1, \dots, \eta_T)^\top$, $\Gamma = \begin{pmatrix} I_{m_1} & \overline{\Gamma}_1 \\ 0 & \overline{\Gamma}_2 \end{pmatrix}$, $b_t = \Gamma^\top\eta_t$ and $B = \eta\Gamma = (b_1, \dots, b_T)^\top$. Now, let $f_i(U_{it}, \varsigma_t)$ to be the joint density of U_{it} and ς_t, where $\varsigma_t \in \{q_t, b_t, f_{2t}\}$, and $f_i(U_{it}, X_{it}, q_t)$ to be the joint density of U_{it}, X_{it} and q_t. Further, introduce the following notations: $\sigma_{i1t} \equiv E(\varepsilon_{i1}\varepsilon_{it})$, $\sigma_i^2 \equiv E(\varepsilon_{i1}^2)$, $\Omega_{ixx}(U_{it}) = E(X_{it}X_{it}^\top|U_{it})$, $\Omega_{iqq}(U_{it}) = E(q_tq_t^\top|U_{it})$, $\Omega_{ixq}(U_{it}) = E(X_{it}q_t^\top|U_{it})$, $\Pi_i(u_0) = \Omega_{ixq}(u_0)\Omega_{iqq}^{-1}(u_0)$, $M_{it}(u_0) = X_{it} - \Pi_i(u_0)q_t$, $\Omega_i(U_{it}) = E\left\{M_{i1}(u_0)M_{i1}^\top(u_0)|U_{it}\right\}$, $\Omega_{it}(U_{i1}, U_{it}) = E\left\{M_{i1}(u_0)M_{it}^\top(u_0)|U_{i1}, U_{it}\right\}$, $\Omega(u_0) \equiv p\lim_{N\to\infty} N^{-1}\sum_{i=1}^N f_i(u_0)\Omega_i(u_0)$, and $\Omega^*(u_0) \equiv p\lim_{N\to\infty} N^{-1}\sum_{i=1}^N \sigma_i^2 f_i(u_0)\Omega_i(u_0)$.

Below are a set of assumptions needed for deriving asymptotic results. Note that these may not be necessary to be the weakest conditions.

A1. The kernel $k(\cdot)$ is a symmetric, nonnegative and bounded probability density function *(PDF)* with a compact support.

A2. (i) There exists some $\delta > 0$ such that $\max_{1\le i\le N} E|\zeta_i|^{2(1+\delta)} < \infty$ for $\zeta_i = \varepsilon_{i1}, f_{11}, f_{21}$ and v_{i1}. (*ii*) For each fixed i, $\{(\varepsilon_{it}, v_{it}) : t \ge 1\}$ is strictly stationary and α-mixing with the mixing coefficient satisfying $\alpha_i(\tau) = O(\tau^{-\theta})$, where $\theta = (2+\delta)(1+\delta)/\delta$. (*iii*) The individual-specific errors ε_{it} and v_{js} are distributed independently for all i, j, t and s. Let $\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{iT})^\top$ and $v_i = (v_{i1}, \dots, v_{iT})^\top$. ε_i and v_i are independently distributed across i with zero means. (*iv*) The common factors $\{(f_{1t}, f_{2t}) : t \ge 1\}$ are strictly stationary and α-mixing with the mixing coefficient satisfying $\alpha_f(\tau) = O(\tau^{-\theta})$. (*v*) (f_{1t}, f_{2t}) are distributed independently of ε_{is} and v_{is} for all i, t and s. (*vi*) There exists $m \in (1, Th)$ such that $Th/m \gg T^{2\bar{\eta}}$ for some $\bar{\eta} > 0$ and $NT\alpha(m) = o_p(1)$ as $(N, T) \to \infty$, where $\alpha(\tau) = \max\{\alpha_1(\tau), \dots, \alpha_N(\tau), \alpha_f(\tau)\}$, which also satisfies $\alpha(\tau) = O(\tau^{-\theta})$.

A3. (i) The factor loadings Γ_{1i}, γ_{2i} and Γ_{2i} are independently and identically distributed with finite $2(1+\delta)$th moments. (*ii*) γ_{2i} and Γ_{2i} are independent of v_{jt}, ε_{jt} and (f_{1t}, f_{2t}) for all j and t. (*iii*) Γ_{1i} are independent of Γ_{2j}, v_{jt}, ε_{jt} and (f_{1t}, f_{2t}) for all j and t. (*iv*) $Rank(\Gamma_2) = m_2 \le p + p_z + 1$, where $\Gamma_2 = E(\Gamma_{2i})$.

A4. (i) For any $\tau \ge 1$, $f_i(U_{i0}, U_{i\tau}|(X_{i0}, q_0), (X_{i\tau}, q_\tau))$, $f_i(U_{i0}, U_{i\tau}|X_{i0}, X_{i\tau})$ and $f_i(U_{i0}, U_{i\tau}|q_0, q_\tau)$ are bounded conditional density of $(U_{i0}, U_{i\tau})$ given $((X_{i0}, q_0), (X_{i\tau}, q_\tau))$, $(X_{i0}, X_{i\tau})$ and (q_0, q_τ). (*ii*) $f_i(U_{it}|\xi_{it})$, $f_i(U_{it}|\varsigma_t)$ and $f_i(U_{it}|X_{it}, q_t)$ are uniformly bounded conditional density of U_{it} given ξ_{it}, ς_t and (X_{it}, q_t), where the definitions of ξ_{it} and ς_t are given earlier. (*iii*) For any $1 \le t \le T$, $f_{i1t}(U_{i1}, U_{it})$ and $\Omega_{it}(U_{i1}, U_{it})$ are continuous at $(U_{i1} = u_0, U_{it} = u_0)$. (*iv*) $\bar{\Omega}_i(U_{it})$ and $f_i(U_{it})$ are continuous at u_0, where $\bar{\Omega}_i(U_{it}) \in \{\Omega_{ixx}(U_{it}),\ \Omega_{ixq}(U_{it}),\ \Omega_{iqq}(U_{it}),\ \Omega_i(U_{it})\}$.

A5. (i) $h \to 0$, $Th \to \infty$ as $T \to \infty$ and $N \to \infty$. (*ii*) $\max_{1\le i\le N} E\left\|X_{it}X_{it}^\top\right\|^{2(1+\delta)} < \infty$, $\max_{1\le i\le N} E\left\|X_{it}q_t^\top\right\|^{2(1+\delta)} < \infty$ and $E\left\|q_tq_t^\top\right\|^{2(1+\delta)} < \infty$. (*iii*) For $j = 0, 1, 2, 3$, $\int |v|^j|k(v)|dv = \kappa_j < \infty$ and $|v^jk(v)| \le \bar{K}$. (*iv*) $\max_{1\le i\le N}\sup_u E[\Xi_{it}|U = u]f_i(u) \le \mathbb{B}_1 < \infty$, where $\Xi_{it} \in \{\|X_{it}X_{it}^\top\|^{2(1+\delta)}, \|X_{it}q_t^\top\|^{2(1+\delta)}, \|q_tq_t^\top\|^{2(1+\delta)}\}$. (*v*) $\max_{1\le i\le N}|\Omega_i'(\cdot)f_i'(\cdot)|$ is bounded. Also, $\max_{1\le i\le N}|\tilde{\Omega}_i'(\cdot)f_i(\cdot)|$ and $\max_{1\le i\le N}|\tilde{\Omega}_i(\cdot)f_i'(\cdot)|$ are bounded, where $\tilde{\Omega}_i(\cdot) \in \{\Omega_{ixx}(\cdot),\ \Omega_{ixq}(\cdot),\ \Omega_{iqq}(\cdot)\}$. (*vi*) $T^{1-1/\theta}h^{(2+\delta)/(1+\delta)} \to \infty$.

Remark 2.1 (*Discussions of Conditions*). A1 is a commonly used condition for kernel functions, which can be satisfied by many widely used kernel functions including the Epanechnikov kernel. A2 imposes strict stationarity, and moments and mixing conditions on common factors and error terms as well. Furthermore, A2(iii) and (v) also impose some independence assumptions which are commonly used in the literature; see Pesaran (2006) and Su and Jin (2012). A3(iv) is the rank condition required by the CCE approach, which is similar to Pesaran (2006). A4 requires some uniform boundedness of conditional densities and smoothness conditions on involved functionals, which are quite mild and standard. A5(i) is standard for nonparametric kernel estimation. Conditions A5(ii) and (vi) are similar to those widely used for nonlinear time series models (e.g., Cai et al. (2000)). Note that A5(ii) imposes moment conditions directly on X_{it} and q_t. Instead, we can also make similar restrictions on factors, loadings, and error terms at the cost of more tedious presentation. A5(iii)–(v) are required for establishing the uniform convergence results in order to prove Theorem 2.3. Similar conditions are also imposed by Hansen (2008).

2.3.2. Asymptotic properties of $\hat{\beta}_i(u_0)$

An application of Taylor's expansion of $\beta_i(U_{it})$ around u_0 leads to

$$\beta_i(U_{it}) = \beta_i(u_0) + h\beta_i'(u_0)(\frac{U_{it}-u_0}{h}) + \frac{h^2}{2}\beta_i''(u_0)(\frac{U_{it}-u_0}{h})^2 + o_p(h^2). \quad (2.13)$$

Inserting (2.13) into (2.1) and rewriting all the observations of i in vector forms implies that

$$Y_i = \widetilde{X}_i\beta_i^*(u_0) + \frac{h^2}{2}A_i(u_0)\beta_i''(u_0) + f_1\gamma_{1i} + f_2\gamma_{2i} + \varepsilon_i + o_p(h^2), \quad (2.14)$$

where $f_1 = (f_{11}, \ldots, f_{1T})^\top, f_2 = (f_{21}, \ldots, f_{2T})^\top$, $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iT})^\top$, and

$$A_i(u_0) = \begin{pmatrix} X_{i1}^\top((U_{i1}-u_0)/h)^2 \\ \vdots \\ X_{iT}^\top((U_{iT}-u_0)/h)^2 \end{pmatrix}.$$

Combining (2.14) and (2.10), then, one has

$$\begin{aligned}
&\hat{\beta}_i^*(u_0) - \beta_i^*(u_0) \\
=\ & \frac{h^2}{2}[\widetilde{X}_i^\top \mathbf{W}_{i,h}^{1/2}(u_0)M_i(u_0)\mathbf{W}_{i,h}^{1/2}(u_0)\widetilde{X}_i]^{-1}\widetilde{X}_i^\top \mathbf{W}_{i,h}^{1/2}(u_0)M_i(u_0)\mathbf{W}_{i,h}^{1/2}(u_0)A_i(u_0)\beta_i''(u_0) \\
&+ [\widetilde{X}_i^\top \mathbf{W}_{i,h}^{1/2}(u_0)M_i(u_0)\mathbf{W}_{i,h}^{1/2}(u_0)\widetilde{X}_i]^{-1}\widetilde{X}_i^\top \mathbf{W}_{i,h}^{1/2}(u_0)M_i(u_0)\mathbf{W}_{i,h}^{1/2}(u_0)f_2\gamma_{2i} \\
&+ [\widetilde{X}_i^\top \mathbf{W}_{i,h}^{1/2}(u_0)M_i(u_0)\mathbf{W}_{i,h}^{1/2}(u_0)\widetilde{X}_i]^{-1}\widetilde{X}_i^\top \mathbf{W}_{i,h}^{1/2}(u_0)M_i(u_0)\mathbf{W}_{i,h}^{1/2}(u_0)\varepsilon_i \\
&+ o_p(h^2), \quad (2.15)
\end{aligned}$$

which expresses the difference between $\hat{\beta}_i^*(u_0)$ and its true value as the sum of three terms plus a higher order term. Clearly, each term has own mean. For example, the first term is a standard bias term of local linear estimators, which contributes to the asymptotic bias, the second one is due to the replacement of f_{2t} by a linear combination of $(f_{1t}, \overline{\boldsymbol{\omega}}_t)$ so that it gives an asymptotic bias too, the third contains the idiosyncratic errors ε_i, which determines the variance, and the last term is a higher order reminder of Taylor expansion. Now, the asymptotic properties of $\hat{\beta}_i(u_0)$ are stated in the following theorems.

Theorem 2.1. *Suppose Conditions A1, A2(i)–(v), A3, A4, and A5(i)–(ii) hold, then, for any given i, we have*

$$\hat{\beta}_i(u_0) - \beta_i(u_0) - \frac{h^2}{2}\mu_2\beta_i''(u_0) = o_p(h^2) + O_p(\frac{1}{\sqrt{Th}}) + M_N, \quad (2.16)$$

where $M_N = O_p(1/\sqrt{N})$.

Theorem 2.2. *Under Conditions A1, A2(i)–(v), A3, A4, and A5(i)(ii)(vi), for any given i, we have*

$$\sqrt{Th}\left(\hat{\beta}_i(u_0) - \beta_i(u_0) - \frac{h^2}{2}\mu_2\beta_i''(u_0) + M_N\right) \xrightarrow{d} N(0, \Theta_i^2(u_0)), \quad (2.17)$$

where $h^2\mu_2\beta_i''(u_0)/2$ is the asymptotic bias term and $\Theta_i^2(u_0) = \lambda_0\sigma_i^2\Omega_i^{-1}(u_0)/f_i(u_0)$ is the asymptotic variance term with $\Omega_i(\cdot)$ given in Section 2.3.1.

Remark 2.2. Theorem 2.1 gives the convergence rate of the LCCE estimator. Under A5(i), this result shows that the LCCE estimator is consistent. Theorem 2.2 reveals that the LCCE estimator is asymptotically normally distributed. The asymptotic distribution in (2.17) is a standard result for local linear estimation, except for the term M_N, the probability order corresponding to the second summand in (2.15), which is the approximation error due to the replacement of f_{2t} by the linear combination of $(f_{1t}, \overline{\omega}_t)$. Note that, as aforementioned, when estimating the heterogeneous model for fixed $i \in N$, one actually considers a time series with length T. Therefore, the convergence rate for $\hat{\beta}_i(u_0)$ is $\sqrt{Th}$, depending on the time dimension T and bandwidth h.

2.3.3. Asymptotic properties of $\hat{\beta}(u_0)$

Similarly, by Taylor's expansion of the homogeneous functional coefficients $\beta(U_{it})$ around u_0, one has

$$\beta(U_{it}) = \beta(u_0) + h\beta'(u_0)(\frac{U_{it} - u_0}{h}) + \frac{h^2}{2}\beta''(u_0)(\frac{U_{it} - u_0}{h})^2 + o_p(h^2). \tag{2.18}$$

By substituting (2.18) into (2.4) and rewriting all the observations of i in vector forms, one obtains

$$Y_i = \widetilde{X}_i\beta^*(u_0) + \frac{h^2}{2}A_i(u_0)\beta''(u_0) + f_1\gamma_{1i} + f_2\gamma_{2i} + \varepsilon_i + o_p(h^2).$$

Pooling all data together gives that

$$Y = \widetilde{X}\beta^*(u_0) + \frac{h^2}{2}A(u_0)\beta''(u_0) + F_1\gamma_{\mathbf{1}} + F_2\gamma_{\mathbf{2}} + \varepsilon + o_p(h^2), \tag{2.19}$$

where $A(u_0) = \left(A_1^\top(u_0), \ldots, A_N^\top(u_0)\right)^\top$, $\gamma_{\mathbf{1}} = \left(\gamma_{11}^\top, \ldots, \gamma_{1N}^\top\right)^\top$, $\gamma_{\mathbf{2}} = \left(\gamma_{21}^\top, \ldots, \gamma_{2N}^\top\right)^\top$, $F_1 = I_N \otimes f_1$, $F_2 = I_N \otimes f_2$ and $\varepsilon = \left(\varepsilon_1^\top, \ldots, \varepsilon_N^\top\right)^\top$. By combining (2.19) and (2.12), similar to (2.15), we express the difference between $\hat{\beta}^*(u_0)$ and its true value as a summation of four terms as follows:

$$\begin{aligned}
&\hat{\beta}^*(u_0) - \beta^*(u_0) \\
=\ &\frac{h^2}{2}[\widetilde{X}^\top\mathbf{W}_h^{1/2}(u_0)M(u_0)\mathbf{W}_h^{1/2}(u_0)\widetilde{X}]^{-1}\widetilde{X}^\top\mathbf{W}_h^{1/2}(u_0)M(u_0)\mathbf{W}_h^{1/2}(u_0)A(u_0)\beta''(u_0) \\
&+ [\widetilde{X}^\top\mathbf{W}_h^{1/2}(u_0)M(u_0)\mathbf{W}_h^{1/2}(u_0)\widetilde{X}]^{-1}\widetilde{X}^\top\mathbf{W}_h^{1/2}(u_0)M(u_0)\mathbf{W}_h^{1/2}(u_0)F_2\gamma_{\mathbf{2}} \\
&+ [\widetilde{X}^\top\mathbf{W}_h^{1/2}(u_0)M(u_0)\mathbf{W}_h^{1/2}(u_0)\widetilde{X}]^{-1}\widetilde{X}^\top\mathbf{W}_h^{1/2}(u_0)M(u_0)\mathbf{W}_h^{1/2}(u_0)\varepsilon \\
&+ o_p(h^2).
\end{aligned} \tag{2.20}$$

Then, the asymptotic distribution of the LCCEP estimator can be established by studying each of the above four terms. Now, the asymptotic properties of $\hat{\beta}(u_0)$ are stated in the following two theorems. Note that $\lambda_{max}(A)$ denotes the maximum eigenvalue of a real symmetric matrix A.

Theorem 2.3. *Suppose Conditions A1–A3, A4(i)–(iii), and A5(i)–(v) hold, and $\lambda_{max}(E(\gamma_{\mathbf{2}}\gamma_{\mathbf{2}}^\top)) = O(r_N)$, then*

$$\hat{\beta}(u_0) - \beta(u_0) - \frac{h^2}{2}\mu_2\beta''(u_0) = o_p(h^2) + O_p(\frac{1}{\sqrt{NTh}}) + M_N^*, \tag{2.21}$$

where $M_N^* = O_p((r_N^{1/2} + 1)/N)$.

Theorem 2.4. *Under Conditions A1–A5, we have*

$$\sqrt{NTh}\left(\hat{\beta}(u_0) - \beta(u_0) - \frac{h^2}{2}\mu_2\beta''(u_0) + M_N^*\right) \xrightarrow{d} N(0, \lambda_0\Omega^{-1}(u_0)\Omega^*(u_0)\Omega^{-1}(u_0)),$$

where $\Omega(u_0)$ and $\Omega^*(u_0)$ are defined in Section 2.3.1.

Remark 2.3. Theorem 2.3 gives the convergence rate of the LCCEP estimator. Compared with Theorem 2.1, there are two obvious changes. First, the term $O_p(1/\sqrt{NTh})$ has a faster convergence rate than the corresponding term $O_p(1/\sqrt{Th})$ in (2.16) as both N and T go to infinity simultaneously in Theorem 2.3. This reflects the benefit of using pooled data to estimate the homogeneous model (2.4). Second, the probability order of the approximate error due to the replacement of f_{2t} by observable variables has changed from M_N to M_N^*. Note that M_N^* is not only dependent on N but also the order of the maximum eigenvalue of $E(\gamma_{\mathbf{2}}\gamma_{\mathbf{2}}^\top)$, which measures the strength of cross-sectional dependence (see Chudik et al. (2011)). Theorem 2.4 shows that the LCCEP estimator is asymptotically normally distributed with convergence rate $\sqrt{NTh}$. This convergence rate is standard in the nonparametric/semiparametric panel literature (see, for example, Cai and Li (2008); Chen et al. (2012)).

3. Hypothesis testing

In this section, a nonparametric test is proposed for testing constancy on functional coefficients and two kinds of constancy tests are considered. The first is for testing heterogeneous functional coefficients, i.e., testing whether $\beta_i(\cdot) = \beta_i$, where β_i is a constant which may assume different values for different i's. The second is for testing homogeneous functional coefficients; that is to test whether $\beta(\cdot) = \beta_0$ for some constant β_0. The test statistic of the former is easy to be constructed and the asymptotic distribution under null hypothesis is standard by employing the results in Hjellvik et al. (1998) and Fan and Li (1999), because given i, it is actually a time series model. The latter is more involved since the test statistic (see (3.3) later) itself possesses the structure of a two-fold V-statistic where double summations are needed along both the individual and time dimensions. Therefore, only focus here is on testing $\beta(\cdot) = \beta_0$ in the following study.

3.1. Test statistic

Consider the following null and alternative hypotheses:

$$H_0 : \beta(u) = \beta_0 \quad \text{versus} \quad H_a : \beta(u) \neq \beta_0 \quad \text{for some} \quad \beta_0 \in \mathbb{R}^p.$$

The test statistic is constructed based on the weighted integrated squared difference between the constant and varying coefficients; that is, $L \equiv \int [\beta(u) - \beta_0]^\top \mathbb{W}(u)[\beta(u) - \beta_0]du$, where $\mathbb{W}(u)$ is a weight matrix. In Section 2, $\beta(u)$ is estimated based on the local linear method. This method is one of the best approaches for boundary correction since its bias does not depend on the density function of u. However, when constructing a test statistic, the local linear estimator complicates the asymptotic analysis of the test statistic. Furthermore, as shown in Lin et al. (2014), it suffices to use the local constant estimator of $\beta(u)$ to construct the test statistic. The local constant estimator of $\beta(u)$ in (2.4), denoted by $\hat{\beta}_{lc}(u)$, is given by

$$\begin{aligned} \hat{\beta}_{lc}(u) &= [(NT)^{-1}X^\top \mathbf{W}_h^{1/2}(u)M(u)\mathbf{W}_h^{1/2}(u)X]^{-1}(NT)^{-1}X^\top \mathbf{W}_h^{1/2}(u)M(u)\mathbf{W}_h^{1/2}(u)Y \\ &\equiv D^{-1}(u)W(u)Y, \end{aligned} \tag{3.1}$$

where the definition of $D(u)$ and $W(u)$ should be apparent from the above equation. The asymptotic distribution of $\hat{\beta}_{lc}(u)$ can be derived following similar steps as in the proof of Theorem 2.4. It is well known that the local constant estimator and the local linear estimator share identical asymptotic variance. To get a feasible test statistic, one can replace $\beta(u)$ in L with $\hat{\beta}_{lc}(u)$ and replace β_0 with the CCEP estimator $\hat{\beta}_0$ proposed by Pesaran (2006) which has the following closed-form expression

$$\hat{\beta}_0 = \left[\sum_{i=1}^N X_i^\top M_w X_i\right]^{-1} \sum_{i=1}^N X_i^\top M_w Y_i,$$

where $M_w = I_T - H_w[H_w^\top H_w]^{-1}H_w^\top$, $H_w = (h_{w1}, \ldots, h_{wT})^\top$ and $h_{wt} = (f_{1t}^\top, \overline{X}_t^\top, \overline{Y}_t)^\top$ for $1 \le t \le T$. Since the random denominator $D(u)$ in (3.1) is not bounded away from 0, a test statistic L_{NT} is proposed based on a weighted integrated squared difference with $\mathbb{W}(u) = D^\top(u)D(u)$ as the weight matrix; that is

$$\begin{aligned} L_{NT} &\equiv \int \{D(u)[\hat{\beta}_{lc}(u) - \hat{\beta}_0]\}^\top \{D(u)[\hat{\beta}_{lc}(u) - \hat{\beta}_0]\}du \\ &= (Y - X\hat{\beta}_0)^\top \left[\int W^\top(u)W(u)du\right](Y - X\hat{\beta}_0). \end{aligned}$$

The test statistic L_{NT} contains the integration $\int W^\top(u)W(u)du$ which is not easy to compute in practice. The main difficulty in getting rid of the integration in L_{NT} is that the multiplicand $M(u)$ in $W(u)$ is a function of u. Note that the purpose of employing $M(u)\mathbf{W}_h^{1/2}(u)$ is to remove $\mathcal{Q}\vartheta$ in (2.11) in the limit. The matrix $M_{\mathcal{Q}} \equiv I_N \otimes M_Q$, which does not depend on u, can achieve the same goal, where $M_Q = I_T - Q(Q^\top Q)^{-1}Q^\top$. Thus, $W^*(u) \equiv (NT)^{-1}X^\top \mathbf{W}_h^{1/2}(u)M_{\mathcal{Q}}$ is used instead of using $W(u)$. In addition, the matrix $\mathbf{W}_h^{1/2}(u)$ in $W^*(u)$ can be replaced by $\mathbf{W}_h(u)$. The previous analysis results in the following test statistic

$$\begin{aligned} L_{NT}^* &= \frac{1}{N^2T^2}(Y - X\hat{\beta}_0)^\top M_{\mathcal{Q}} \left[\int \mathbf{W}_h(u)XX^\top \mathbf{W}_h(u)du\right] M_{\mathcal{Q}}(Y - X\hat{\beta}_0) \\ &= \frac{1}{N^2T^2h^2}\sum_{i=1}^N\sum_{j=1}^N (Y_i - X_i\hat{\beta}_0)^\top M_Q \left[\int \mathbf{W}_i(u)X_iX_j^\top \mathbf{W}_j(u)du\right] M_Q(Y_j - X_j\hat{\beta}_0) \\ &= \frac{1}{N^2T^2h^2}\sum_{i=1}^N\sum_{j=1}^N \tilde{\nu}_i^\top \left[\int \mathbf{W}_i(u)X_iX_j^\top \mathbf{W}_j(u)du\right]\tilde{\nu}_j, \end{aligned}$$

where $\tilde{\nu}_i = (\tilde{\nu}_{i1}, \ldots, \tilde{\nu}_{iT})^\top = M_Q(Y_i - X_i\hat{\beta}_0)$ and $\mathbf{W}_i(\cdot)$ is obtained by replacing each appearance of $k_h(\cdot)$ in $\mathbf{W}_{i,h}(\cdot)$ by $k(\cdot/h)$ for $1 \le i \le N$. The (t, s)th element of $\int \mathbf{W}_i(u)X_iX_j^\top \mathbf{W}_j(u)du$ is $X_{it}^\top X_{js}\int k((U_{it} - u)/h)k((U_{js} - u)/h)du = hX_{it}^\top X_{js}\bar{k}((U_{it} - U_{js})/h)$, where $\bar{k}(v) = \int k(u)k(v - u)du$ is the twofold convolution kernel derived from $k(\cdot)$. As pointed out by Li et al. (2002) and Lin et al. (2014), one does not even have to use the convolution kernels. Simply replacing $\bar{k}((U_{it} - U_{js})/h)$ by $k((U_{it} - U_{js})/h)$

results in the following test statistic

$$L_{NT}^{**} = \frac{1}{N^2T^2h}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\tilde{v}_{it}\tilde{v}_{js}X_{it}^{\top}X_{js}k(\frac{U_{it}-U_{js}}{h})$$
$$= \frac{1}{N^2T^2h}\sum_{i=1}^{N}\sum_{j=1}^{N}\tilde{v}_i^{\top}\mathbf{W}(\boldsymbol{\omega}_i,\boldsymbol{\omega}_j)\tilde{v}_j, \tag{3.2}$$

where $\mathbf{W}(\boldsymbol{\omega}_i,\boldsymbol{\omega}_j)$ is the $T\times T$ matrix whose (t,s)th element is $X_{it}^{\top}X_{js}k((U_{it}-U_{js})/h)$. Finally, the $i=j$ term in (3.2) is dropped to remove a nonzero center term from L_{NT}^{**} under H_0. Therefore, the final form of our test statistic is given by

$$\hat{L}_{NT} = \frac{1}{N^2T^2h}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\tilde{v}_i^{\top}\mathbf{W}(\boldsymbol{\omega}_i,\boldsymbol{\omega}_j)\tilde{v}_j. \tag{3.3}$$

It is obvious that H_0 is rejected if and only if $|\hat{L}_{NT}|$ is over a certain critical value.

3.2. Asymptotic theory of test statistic

In order to derive the asymptotic distribution of the test statistic, the following assumptions are necessary:

T1. (i) Let $\varepsilon_i=(\varepsilon_{i1},\ldots,\varepsilon_{iT})^{\top}$, $v_i=(v_{i1},\ldots,v_{iT})^{\top}$, where $\{\varepsilon_i\}_{i=1}^N$ and $\{v_i\}_{i=1}^N$ are independently and identically distributed across i. (ii) Let $\|\varepsilon_{it}\|_{\delta_1}=(E|\varepsilon_{it}|^{\delta_1})^{1/\delta_1}$, $\|\varepsilon_{it}\|_{\delta_1}<\infty$, where $\delta_1>4$. Given i, let $\alpha_\varepsilon(\tau)$ be the α-mixing coefficient of $\{\varepsilon_{it}:t\geq 1\}$ such that $\sum_{\tau=1}^{\infty}\tau\alpha_\varepsilon(\tau)^{(\delta_1-4)/\delta_1}<\infty$.

T2. (i) Let $f_{U_{it_1}}(U_{j_1t_2})$ denote the conditional density of $U_{j_1t_2}$ given U_{it_1}, and $f_{U_{it_1},U_{it_2}}(U_{j_1t_3},U_{j_2t_4})$ be the conditional density of $(U_{j_1t_3},U_{j_2t_4})$ given (U_{it_1},U_{it_2}). $f_{U_{it_1}}(U_{j_1t_2})$ and $f_{U_{it_1},U_{it_2}}(U_{j_1t_3},U_{j_2t_4})$ are uniformly bounded. (ii) Let $\mathscr{D}=\{U_{it}:i=1,\ldots,N;t=1,\ldots,T\}$, $X_{it,l}$ the lth element of X_{it}, and $q_{t,k}$ the kth element of q_t, respectively. For all $1\leq i_1,i_2\leq N$, $1\leq t_1,t_2\leq T$, $1\leq l,m\leq p$, and $1\leq k_1,k_2\leq d$ with $d=m_1+p+p_z+1$, $E_{\mathscr{D}}|X_{i_1t_1}^{\top}X_{i_2t_2}|^8<\infty$, $E_{\mathscr{D}}|X_{i_1t_1,l}X_{i_2t_2,m}|^8<\infty$ and $E_{\mathscr{D}}|q_{t_1,k_1}q_{t_2,k_2}|^8<\infty$, where $E_{\mathscr{D}}(\cdot)$ is the conditional expectation conditional on $\mathscr{D}$.

T3. $h\to 0,\ Th\to\infty,\ T/N\to 0$ as $(N,T)\to\infty$.

Remark 3.1. Note that the proposed test statistic is of the form of a U-statistic. It intends to employ the theories developed by Hall (1984) and Powell et al. (1989) for a U-statistic to derive the asymptotic distributions of $\hat{L}_{NT}$. To reach this goal, in T1(i), ε_i and v_i are assumed to be independently and identically distributed across the cross-sectional dimension. T1(ii) is required by the use of the Davydov's inequality to obtain some upper bound for the cross-moment for mixing random variables and T2 includes some boundedness conditions on functionals involved. Finally, in T3, it needs an additional requirement that $T/N\to 0$ as $(N,T)\to\infty$, which is also used in Theorem 4 of Pesaran (2006).

The asymptotic distributions of our proposed test statistic $\hat{L}_{NT}$ under the null and alternative hypotheses are given in the following two theorems.

Theorem 3.1. *If Assumptions A1, A2(i)–(v), A3 and Assumptions T1–T3 hold, we have, under the null hypothesis H_0,*

$$\mathbb{J}_{NT} = NTh^{1/2}\hat{L}_{NT}/\sqrt{\hat{V}} \xrightarrow{d} N(0,1), \tag{3.4}$$

where

$$\hat{V} = \frac{2}{N^2T^2h}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{1\leq t,s\leq T}\tilde{v}_{it}^2\tilde{v}_{js}^2(X_{it}^{\top}X_{js})^2k^2(\frac{U_{it}-U_{js}}{h}), \tag{3.5}$$

is a consistent estimator of the asymptotic variance of $NTh^{1/2}\hat{L}_{NT}$, i.e.,

$$V = plim_{T\to\infty}\frac{2}{T^2}\sum_{1\leq t,s\leq T}E_{\mathscr{D}_f}\left[\varepsilon_{1t}^2\varepsilon_{2s}^2(X_{1t}^{\top}X_{2s})^2\frac{1}{h}k^2(\frac{U_{1t}-U_{2s}}{h})\right],$$

where $E_{\mathscr{D}_f}(\cdot)$ denotes the conditional expectation given $\mathscr{D}_f=\{(f_{1t},f_{2t}):t=1,\ldots,T\}$.

Theorem 3.2. *If Assumptions A1, A2(i)–(v), A3 and Assumptions T1–T3 hold, we have, under the alternative H_a,*

$$\Pr\{\mathbb{J}_{NT}\geq M_{NT}\} \to 1 \quad as\ (N,T)\to\infty,$$

where M_{NT} is any non-stochastic, positive sequence such that $M_{NT}=o(NTh^{1/2})$.

Remark 3.2. Theorem 3.1 shows that, under H_0, the standardized statistic $\mathbb{J}_{NT}$ approaches toward the asymptotic standard normal distribution at the rate $NTh^{1/2}$. In addition, estimation of the asymptotic variance V involves the use of the transformed residuals $\tilde{\nu}_i = M_Q(Y_i - X_i\hat{\beta}_0)$, which does not need to know the information of the unobserved factors. Therefore, our test statistic is user-friendly and easily computed. Theorem 3.2 shows that, under the alternative H_a, the probability that the proposed test rejects the null hypothesis approaches 1 as both N and T go to infinity simultaneously, which implies that the proposed test is consistent.

3.3. A bootstrap procedure

It is well-known that kernel-based nonparametric tests converge to the asymptotic distributions very slowly and often suffer from finite sample size distortions. Therefore, in addition to Theorem 3.1, where the asymptotic normal distribution is derived under H_0 for the proposed test statistic, a bootstrap approach is proposed to improve its finite sample performance.

For sake of exposition, here we only consider the case that $f_{1t} = 1$ for all t. The general case can be dealt with in a similar way. Therefore, under H_0, the following model is considered:

$$Y_{it} = X_{it}^\top \beta_0 + \gamma_{1i} + \gamma_{2i}^\top f_{2t} + \varepsilon_{it}.$$

Our bootstrap procedure is based on the "fixed regressor bootstrap" given by Hansen (2000) which generates a bootstrapping sample using the observable covariates X_{it} and the estimated residuals $\hat{\varepsilon}_{it}$ which is estimated by following the common practice in panel data literature (e.g. Jin and Su (2013)) to estimate unobservable factors f_{2t} and factor loadings γ_{2i} with the principal component analysis (PCA) method. Specifically, the bootstrapping procedure is given as follows:

(i) Estimate the linear panel data model under H_0 using the CCEP method proposed by Pesaran (2006) and obtain $\hat{\nu}_{it} = Y_{it} - X_{it}^\top \hat{\beta}_0$ and $\tilde{\nu}_i = M_Q(Y_i - X_i\hat{\beta}_0)$ for all i and t. Calculate $\mathbb{J}_{NT}$ using (3.3)–(3.5).

(ii) Estimate the unobservable common factor f_{2t} and factor loadings γ_{2i} by the PCA method. The estimates are denoted by $\hat{f}_{2t}$ and $\hat{\gamma}_{2i}$, respectively. Estimate γ_{1i} using $\hat{\gamma}_{1i} = T^{-1}\sum_{t=1}^{T}[Y_{it} - X_{it}^\top\hat{\beta}_0 - \hat{\gamma}_{2i}^\top\hat{f}_{2t}]$ for all i. Obtain the residuals $\hat{\varepsilon}_{it} = \hat{\nu}_{it} - \hat{\gamma}_{1i} - \hat{\gamma}_{2i}^\top\hat{f}_{2t}$.

(iii) Compute the wild bootstrap errors from $\{\hat{\varepsilon}_{it}\}_{i=1,\ldots,N;t=1,\ldots,T}$ by $\hat{\varepsilon}_{it}^* = \hat{\varepsilon}_{it}\eta_{it}^*$ where η_{it}^* is generated from IIDN(0,1) for all i and t. Generate Y_{it}^* via $Y_{it}^* = X_{it}^\top\hat{\beta}_0 + \hat{\gamma}_{1i} + \hat{\gamma}_{2i}^\top\hat{f}_{2t} + \hat{\varepsilon}_{it}^*$ for $i = 1,\ldots,N$ and $t = 1,\ldots,T$. Then regard $\{(X_{it}, U_{it}, Y_{it}^*)\}_{i=1,\ldots,N;t=1,\ldots,T}$ as the bootstrapping sample.

(iv) Using the bootstrapping sample, one can recalculate the CCEP estimator for β_0, denoted by $\hat{\beta}_0^*$, and then obtain the bootstrapping transformed residuals via $\tilde{\nu}_i^* = (\tilde{\nu}_{i1}^*, \ldots, \tilde{\nu}_{iT}^*)^\top = M_Q(Y_i^* - X_i\hat{\beta}_0^*)$ for all i, where Y_i^* is defined in the same way as Y_i except that Y_{it} is replaced by Y_{it}^* wherever it occurs.

(v) Compute the bootstrap test statistic $\mathbb{J}_{NT}^*$ in the same way as $\mathbb{J}_{NT}$ by using $\tilde{\nu}_i^*$ obtained from step (iv) instead. That is,

$$\mathbb{J}_{NT}^* = NTh^{1/2}\hat{L}_{NT}^*/\sqrt{\hat{V}^*},$$

where

$$\hat{L}_{NT}^* = \frac{1}{N^2T^2h}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\tilde{\nu}_i^{*\top}\mathbf{W}(\boldsymbol{\omega}_i, \boldsymbol{\omega}_j)\tilde{\nu}_j^*$$

and

$$\hat{V}^* = \frac{2}{N^2T^2h}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{1\le t,s\le T}\tilde{\nu}_{it}^{*2}\tilde{\nu}_{js}^{*2}(X_{it}^\top X_{js})^2k^2(\frac{U_{it}-U_{js}}{h}).$$

(vi) Repeat steps (iii)–(v) B^* times to get B^* bootstrapping test statistics $\{\mathbb{J}_{NT,b}^*\}_{b=1}^{B^*}$. Calculate the bootstrapping p-value p^* via $p^* = B^{*-1}\sum_{b=1}^{B^*}\mathbf{1}(\mathbb{J}_{NT,b}^* \ge \mathbb{J}_{NT})$. Then, reject the null hypothesis of constancy in coefficients if p^* is smaller than the prescribed level of significance.

4. Monte Carlo simulations

In this section, Monte Carlo simulations are conducted to evaluate the finite sample performance of the proposed estimators and test statistic. Section 4.1 investigates the finite sample properties of LCCE and LCCEP estimators defined by (2.10) and (2.12) and Section 4.2 assesses the size and power performance of our proposed test statistic for a given sample size.

Table 1
Median of RMSE_1 of LCCE estimator.

N/T	50		100		200	
	LCCE_{fr}	LCCE_{rd}	LCCE_{fr}	LCCE_{rd}	LCCE_{fr}	LCCE_{rd}
50	0.5738	0.9161	0.3236	0.5304	0.2201	0.3562
100	0.7252	1.2353	0.3923	0.7134	0.2532	0.4494
200	0.9116	1.7876	0.4684	0.9838	0.3106	0.6251

4.1. Finite sample performance of estimators

Consider the following data generating process (DGP):

$$Y_{it} = \beta_i(U_{it})X_{it} + \gamma_{1i} + e_{it} \quad \text{and} \quad e_{it} = \gamma_{2i,1}f_{2t,1} + \gamma_{2i,2}f_{2t,2} + \varepsilon_{it},$$

where $\beta_i(U_{it}) = \exp(U_{it})/(\exp(U_{it}) + 1) + \delta_i(0.5U_{it} - 0.25U_{it}^2)$, $X_{it} = \Gamma_{1i,x} + \Gamma_{2i,x1}f_{2t,1} + \Gamma_{2i,x2}f_{2t,2} + v_{it,x}$, and $U_{it} = \Gamma_{1i,u} + \Gamma_{2i,u1}f_{2t,1} + \Gamma_{2i,u2}f_{2t,2} + v_{it,u}$. The above DGP follows closely to those of Pesaran (2006) and Su and Jin (2012). In this DGP, there are one observable common factor ($f_{1t} = 1$), and two unobservable common factors ($f_{2t,1}, f_{2t,2}$). The individual effect γ_{1i} is generated as $\gamma_{1i} = 0.5\bar{X}_i$, where $\bar{X}_i = T^{-1}\Sigma_{t=1}^{T}X_{it}$, so that γ_{1i} and X_{it} are correlated for $1 \le t \le T$. Whether or not the functional coefficients are heterogeneous depends on δ_i. The δ_i is generated from IIDU(0,1) for heterogeneous functional coefficients, while δ_i is set to be 0.5 for all $1 \le i \le N$ in the homogeneous case.

Other components in the DGP are generated as follows:

(1) The idiosyncratic errors ε_{it} are generated independently of each other as stationary AR(1) processes: $\varepsilon_{it} = \rho_{i\varepsilon}\varepsilon_{i,t-1} + \sigma_i\sqrt{(1-\rho_{i\varepsilon}^2)}\zeta_{it}$, $t = -49, \ldots, 1, \ldots, T$, where $\zeta_{it} \sim \text{IIDN}(0, 1)$ across i and t, $\rho_{i\varepsilon} \sim \text{IIDU}(0.05, 0.95)$ across i, and $\sigma_i^2 \sim \text{IIDU}(0.5, 1.5)$ across i.

(2) The individual-specific errors $v_{it,x}$ of X_{it} are generated independently of each other as stationary AR(1) processes: $v_{it,x} = \rho_{vix}v_{i,t-1,x} + w_{it,x}$, $t = -49, \ldots, 1, \ldots, T$, where $w_{it,x} \sim \text{IIDN}(0, 1-\rho_{vix}^2)$ across i and t, $\rho_{vix} \sim \text{IIDU}(0.05, 0.95)$ across i, and $w_{i,-50,x} = 0$ for all i. The individual-specific errors $v_{it,u}$ of U_{it} are generated in the same way as $v_{it,x}$ are generated. For each i, $\{\varepsilon_{it}\}_{t=1}^{T}$, $\{v_{it,x}\}_{t=1}^{T}$ and $\{v_{it,u}\}_{t=1}^{T}$ are generated independently of each other.

(3) The unobservable common factors $f_{2t,j}(j = 1, 2)$ are generated as independently stationary AR(1) processes: $f_{2t,j} = \rho_{fj}f_{2,t-1,j} + v_{ft,j}$, $t = -49, \ldots, 1, \ldots, T$, where $v_{ft,j} \sim \text{IIDN}(0, 1-\rho_{fj}^2)$ across t, $\rho_{fj} = 0.5$, and $f_{2,-50,j} = 0$.

(4) The factor loadings $\Gamma_{1i} = (\Gamma_{1i,x}, \Gamma_{1i,u})^{\top}$ are generated as follows

$$\Gamma_{1i} = (\Gamma_{1i,x}, \Gamma_{1i,u})^{\top} \sim \text{IIDN}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & 0.5\\0.5 & 1\end{pmatrix}\right).$$

The factor loadings $\gamma_{2i} = (\gamma_{2i,1}, \gamma_{2i,2})^{\top}$ of the unobservable common factors in e_{it} are generated in the same way as Γ_{1i} are generated.

(5) For the factor loadings Γ_{2i} of the unobservable common factors in X_{it} and U_{it}, two cases, denoted by A and B, are considered, respectively. Γ_{2i} are generated such that $\text{vec}(\Gamma_{2i}) = (\Gamma_{2i,x1}, \Gamma_{2i,x2}, \Gamma_{2i,u1}, \Gamma_{2i,u2})^{\top} \sim \text{IIDN}(\Gamma_{2,\tau}, I_4)$, $\tau = A, B$, where $\Gamma_{2,A} = (1, 0, 0, 1)^{\top}$ and $\Gamma_{2,B} = (1, 1, 0, 0)^{\top}$. In view of A3(iv), Case A satisfies the full rank condition. However, in Case B, the full rank condition is violated.

(N, T) are taken as pairs with $(N, T = 50, 100, 200)$. The number of Monte Carlo replications is 1000. The Epanechnikov kernel $k(u) = 0.75(1-u^2)\mathbf{1}\{|u| \le 1\}$ is used to compute the local linear estimators. Similar to (5.9) in Fan and Yao (2003), it is suggested to choose $h = c_0\hat{\sigma}_u T^{-1/5}$ for heterogeneous model and $h = c_0\hat{\sigma}_u(NT)^{-1/5}$ for homogeneous model, where $\hat{\sigma}_u$ is the sample standard deviation of smoothing variable $\{U_{it}\}_{i=1,\ldots,N;t=1,\ldots,T}$ and $c_0 = 2.34$.

Table 1 reports the medians of the root mean squared error (RMSE) from 1000 replications for the LCCE estimators. The RMSE for heterogeneous model is defined by

$$\text{RMSE}_1^2 = \frac{1}{ND_{grid}}\Sigma_{i=1}^{N}\Sigma_{k=1}^{D_{grid}}\left[\hat{\beta}_i(u_k) - \beta_i(u_k)\right]^2,$$

where $\{u_k : k = 1, \ldots, D_{grid}\}$ are grid points. The simulation results are summarized in the case that the full rank condition is satisfied in the columns below LCCE_{fr} in Table 1. Firstly, one can observe that the increase of N does not help the estimation of the unknown heterogeneous coefficient functions, which is as expected since larger N implies more heterogeneous relationships and the convergence rate of LCCE estimators only depends on the time dimension T. However, as T increases, RMSEs decrease for the estimators. The columns below LCCE_{rd} of Table 1 report the results for the case when the rank condition is not satisfied. Clearly, in the rank deficient case, the LCCE estimators loss certain extent of efficiency compared to the full rank case.

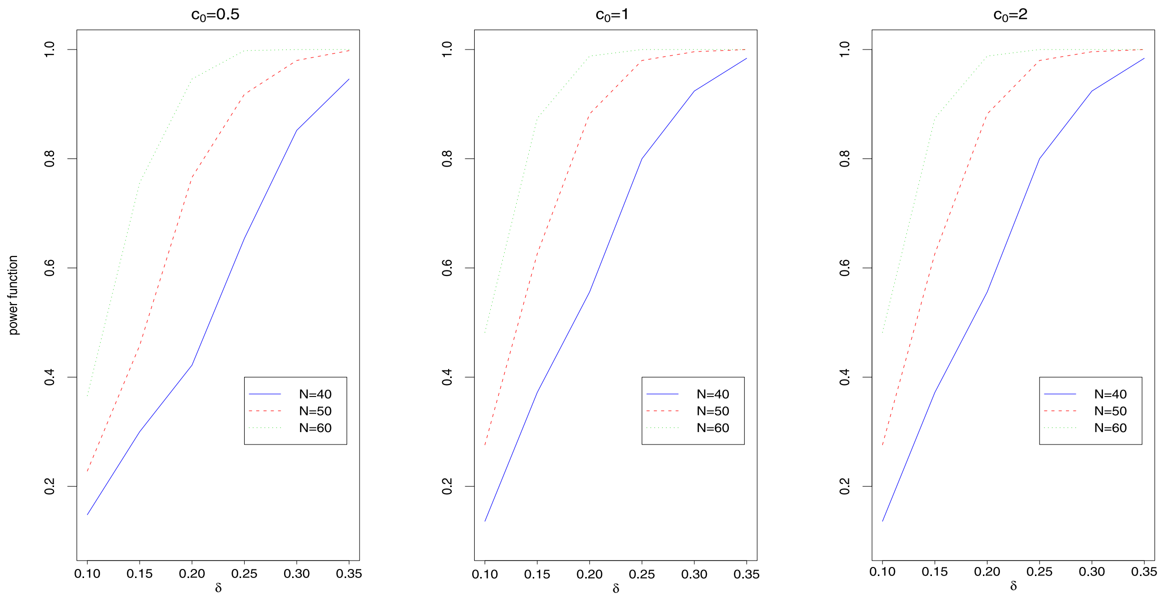

Fig. 1. From left to right: plots of power curves against δ for $c_0 = 0.5$, 1.0, and 2.0 at 5% significance level. For each c_0, power curves for $N = 40$ (blue solid line), $N = 50$ (red dashed line), and $N = 60$ (green dotted line) are depicted, respectively. For each N, T is set to be $\lfloor N/4 \rfloor$.

Table 2
Median of RMSE_2 of LCCEP estimator.

N/T	50		100		200	
	LCCEP_{fr}	LCCEP_{rd}	LCCEP_{fr}	LCCEP_{rd}	LCCEP_{fr}	LCCEP_{rd}
50	0.0604	0.0720	0.0443	0.0568	0.0328	0.0455
100	0.0471	0.0542	0.0337	0.0414	0.0245	0.0337
200	0.0361	0.0416	0.0250	0.0321	0.0186	0.0255

To evaluate the estimation accuracy of the LCCEP estimators, the medians of RMSE from 1000 replications are computed too and the RMSE for homogeneous model is defined by

$$\text{RMSE}_2^2 = \frac{1}{D_{grid}} \Sigma_{k=1}^{D_{grid}} \left[\hat{\beta}(u_k) - \beta(u_k) \right]^2 .$$

The simulation results are presented in Table 2, which imply that, as either N or T increases, the RMSEs of the proposed LCCEP estimators decrease significantly as expected for both the full rank case and rank deficiency case. Moreover, although there still exists efficiency loss, the difference between two cases in terms of RMSEs becomes smaller than those in the LCCE estimators.

4.2. Finite sample performance of test

To assess the size and power performance of the proposed test statistic, one can consider the null hypothesis $H_0 : \beta_j(U_{it}) = \theta_j$ for $j = 0, 1$ versus the alternative $H_a : \beta_j(U_{it}) \neq \theta_j$ for at least one j. The power is evaluated under a series of alternative models indexed by δ, such that $H_a : \beta_j(U_{it}) = \theta_j + \delta(\beta_j^*(U_{it}) - \theta_j)$ for $j = 0, 1$ and $0 \leq \delta \leq 1$, where $\beta_0^*(U_{it}) = \exp(U_{it})/(\exp(U_{it}) + 1)$ and $\beta_1^*(U_{it}) = 2\sin(U_{it})$. θ_j is chosen to be the average height of $\beta_j^*(U_{it})$ for $j = 0, 1$ ($\theta_0 = 0.48$ and $\theta_1 = 0.01$).

The following DGPs are considered in the evaluation of the finite sample size and power performance of our proposed test statistic:

$$\begin{aligned} \text{DGP}_0 : Y_{it} &= \theta_0 X_{it,1} + \theta_1 X_{it,2} + \gamma_{1i} + e_{it}, \\ \text{DGP}_\delta : Y_{it} &= \beta_0(U_{it}) X_{it,1} + \beta_1(U_{it}) X_{it,2} + \gamma_{1i} + e_{it}, \end{aligned}$$

where $e_{it} = \gamma_{2i,1} f_{2t,1} + \gamma_{2i,2} f_{2t,2} + \varepsilon_{it}$, $X_{it,1} = \Gamma_{1i,x1} + \Gamma_{2i,x1} f_{2t,1} + \Gamma_{2i,x2} f_{2t,2} + v_{it,x1}$, $X_{it,2} = \Gamma_{1i,x2} + \Gamma_{2i,x1}^* f_{2t,1} + \Gamma_{2i,x2}^* f_{2t,2} + v_{it,x2}$, $U_{it} = \Gamma_{1i,u} + \Gamma_{2i,u1} f_{2t,1} + \Gamma_{2i,u2} f_{2t,2} + v_{it,u}$, and DGP_δ indicates that $\beta_0(\cdot)$ and $\beta_1(\cdot)$ depend on δ. In order to demonstrate how powers change over different values of δ, δ is set to be 0.10, 0.15, 0.20, 0.25, 0.30, and 0.35. $v_{it,xj}$, $v_{it,u}$, $f_{2t,j}$, $\Gamma_{1i,xj}$, $\Gamma_{1i,u}$

Table 3
Proportion of rejections under H_0.

DGP	N	T	$c_0 = 0.5$			$c_0 = 1$			$c_0 = 2$		
			1%	5%	10%	1%	5%	10%	1%	5%	10%
Tests based on standard normal critical values											
0	50	$\lfloor N/8 \rfloor$	0.006	0.044	0.096	0.008	0.042	0.066	0.006	0.014	0.032
		$\lfloor N/4 \rfloor$	0.016	0.054	0.088	0.014	0.046	0.078	0.004	0.024	0.046
		$\lfloor N/2 \rfloor$	0.008	0.050	0.116	0.010	0.038	0.054	0.008	0.034	0.058
	60	$\lfloor N/8 \rfloor$	0.010	0.054	0.086	0.018	0.048	0.078	0.006	0.030	0.054
		$\lfloor N/4 \rfloor$	0.018	0.052	0.080	0.006	0.058	0.096	0.014	0.030	0.056
		$\lfloor N/2 \rfloor$	0.016	0.056	0.098	0.010	0.058	0.088	0.006	0.024	0.058
	70	$\lfloor N/8 \rfloor$	0.014	0.050	0.082	0.010	0.048	0.076	0.008	0.028	0.054
		$\lfloor N/4 \rfloor$	0.016	0.052	0.082	0.012	0.032	0.068	0.008	0.050	0.076
		$\lfloor N/2 \rfloor$	0.016	0.050	0.082	0.012	0.046	0.078	0.010	0.040	0.058
Tests based on Bootstrap p-values											
0	50	$\lfloor N/8 \rfloor$	0.010	0.052	0.114	0.010	0.048	0.106	0.006	0.052	0.110
		$\lfloor N/4 \rfloor$	0.010	0.050	0.102	0.012	0.050	0.088	0.008	0.040	0.082
		$\lfloor N/2 \rfloor$	0.016	0.042	0.086	0.014	0.052	0.108	0.014	0.046	0.098
	60	$\lfloor N/8 \rfloor$	0.010	0.060	0.106	0.010	0.054	0.116	0.016	0.054	0.096
		$\lfloor N/4 \rfloor$	0.008	0.044	0.102	0.018	0.070	0.106	0.018	0.056	0.094
		$\lfloor N/2 \rfloor$	0.010	0.036	0.086	0.022	0.052	0.120	0.008	0.048	0.096
	70	$\lfloor N/8 \rfloor$	0.014	0.042	0.098	0.010	0.068	0.096	0.012	0.062	0.106
		$\lfloor N/4 \rfloor$	0.018	0.046	0.090	0.024	0.054	0.110	0.012	0.062	0.116
		$\lfloor N/2 \rfloor$	0.012	0.046	0.096	0.010	0.046	0.100	0.006	0.034	0.094

and ε_{it} are generated independently from IIDN(0,1) for $j = 1$ and 2. The factor loadings Γ_{2i} of the unobservable common factors in $X_{it,j}$ $(j = 1, 2)$ and U_{it} are generated from $\text{vec}(\Gamma_{2i}) = (\Gamma_{2i,x1}, \Gamma_{2i,x2}, \Gamma^*_{2i,x1}, \Gamma^*_{2i,x2}, \Gamma_{2i,u1}, \Gamma_{2i,u2})^\top \sim \text{IIDN}(\Gamma_2, I_6)$, where $\Gamma_2 = (1, 1.5, 1, 0, 0, 1)^\top$. The factor loadings $\gamma_{2i} = (\gamma_{2i,1}, \gamma_{2i,2})^\top$ of the unobservable common factors in e_{it} are generated in the same way as step (4) in Section 4.1 and $\gamma_{1i} = 0.5\bar{X}_{i1} + 0.3\bar{X}_{i2}$, where $\bar{X}_{ij} = T^{-1}\Sigma_{t=1}^T X_{it,j}$ for $j = 1$ and 2. For evaluating the size performance, N is chosen to 50, 60, and 70. Furthermore, in view of Assumption T3, we set $T = \lfloor N/8 \rfloor, \lfloor N/4 \rfloor, \lfloor N/2 \rfloor$ for each N, where $\lfloor r \rfloor$ denotes the integral part of r. Table 3 reports the actual proportion of rejections under H_0 at nominal levels 1%, 5%, and 10%, respectively. The bandwidth h is taken as $c_0\hat{\sigma}_u(NT)^{-1/5}$ with $c_0 = 0.5$, 1.0, and 2.0 to examine the sensitivity of the proposed test to the bandwidth selection. The upper panel of Table 3 reports the test results using critical values taken from the standard normal distribution. It is obvious that the results based on asymptotic distribution are sensitive to the choice of bandwidth. It tends to under-rejection when the value of c_0 is large at 5% and 10% nominal levels. Therefore, it turns to using the proposed Bootstrapping method. The number of replications is 500, and within each replication, 300 Bootstrapping samples are used to calculate the Bootstrapping p-values. The test results based on Bootstrapping p-values are reported in the lower panel of Table 3. It is obvious to observe that the proposed Bootstrapping method is much less sensitive to the choice of bandwidth than the asymptotic method, and the former also improves the finite sample size performance immensely. The next is to demonstrate the power performance of our test statistic. N is taken as 40, 50, and 60. Meanwhile, T and c_0 are chosen the same as those in the size evaluation. Tables 4 and 5 display the actual proportion of rejections under the alternative H_a for different values of δ. One can see that, for a fixed δ, the rejection rate increases as either N or T increases. In addition, when δ gets larger, which indicates a larger deviation from the null model, the rejection rate increases for all (N, T) combinations. To better demonstrate the power performance, the power functions are plotted against δ for $c_0 = 0.5$, 1.0, and 2.0 at 5% nominal level in Fig. 1. For each c_0, three power curves are computed corresponding to $N = 40$, 50, and 60 and $T = \lfloor N/4 \rfloor$. From these figures, one can find that the power curves are similar for different choices of c_0. The power increases rapidly to 1 as δ increases, which shows a good power performance in finite samples.

5. Empirical analysis

In this section, the empirical performance of the Fama and French (1993) (FF hereafter) three-factor model[2] is re-examined using the proposed functional-coefficient panel data model with cross-sectional dependence. There has been numerous empirical evidence of showing that betas in capital asset pricing models are in general time varying. For example, based on 25 portfolios formed on size and book-to-market ratio, Ferson and Harvey (1999) found a strong evidence that betas in the FF model vary with lagged macroeconomic and financial instruments. Wang (2003) also provided a clear empirical evidence to argue that the market betas vary with conditioning variables such as the dividend price ratio and the one-month Treasury bill rate. To characterize the dynamic pattern of conditional betas, beta coefficients

[2] A comprehensive discussion about the issue of choosing risk factors in a linear asset pricing model can be found in Fama and French (2018).

Table 4
Proportion of rejections under H_a.

DGP	N	T	$c_0 = 0.5$			$c_0 = 1$			$c_0 = 2$		
			1%	5%	10%	1%	5%	10%	1%	5%	10%
δ	Tests based on Bootstrap critical values										
0.100	40	$\lfloor N/8 \rfloor$	0.010	0.052	0.102	0.032	0.090	0.142	0.014	0.048	0.094
		$\lfloor N/4 \rfloor$	0.032	0.148	0.206	0.054	0.136	0.250	0.056	0.156	0.250
		$\lfloor N/2 \rfloor$	0.160	0.328	0.472	0.286	0.452	0.580	0.230	0.448	0.604
	50	$\lfloor N/8 \rfloor$	0.006	0.096	0.176	0.038	0.094	0.180	0.028	0.084	0.166
		$\lfloor N/4 \rfloor$	0.058	0.228	0.352	0.118	0.276	0.378	0.098	0.286	0.442
		$\lfloor N/2 \rfloor$	0.246	0.482	0.626	0.498	0.732	0.836	0.486	0.750	0.844
	60	$\lfloor N/8 \rfloor$	0.058	0.112	0.178	0.048	0.120	0.200	0.036	0.116	0.198
		$\lfloor N/4 \rfloor$	0.148	0.366	0.508	0.286	0.482	0.640	0.234	0.468	0.624
		$\lfloor N/2 \rfloor$	0.576	0.760	0.878	0.706	0.880	0.948	0.794	0.936	0.962
0.150	40	$\lfloor N/8 \rfloor$	0.014	0.060	0.128	0.042	0.098	0.184	0.018	0.074	0.142
		$\lfloor N/4 \rfloor$	0.110	0.300	0.446	0.154	0.372	0.498	0.134	0.342	0.470
		$\lfloor N/2 \rfloor$	0.564	0.740	0.842	0.726	0.856	0.900	0.690	0.898	0.942
	50	$\lfloor N/8 \rfloor$	0.030	0.134	0.212	0.062	0.152	0.226	0.036	0.122	0.212
		$\lfloor N/4 \rfloor$	0.240	0.458	0.580	0.332	0.626	0.724	0.352	0.574	0.758
		$\lfloor N/2 \rfloor$	0.806	0.948	0.964	0.948	0.990	0.990	0.954	0.996	0.996
	60	$\lfloor N/8 \rfloor$	0.060	0.190	0.328	0.144	0.270	0.384	0.140	0.282	0.374
		$\lfloor N/4 \rfloor$	0.508	0.756	0.840	0.732	0.874	0.924	0.756	0.896	0.944
		$\lfloor N/2 \rfloor$	0.984	1.000	1.000	0.992	1.000	1.000	0.998	1.000	1.000
0.200	40	$\lfloor N/8 \rfloor$	0.024	0.104	0.178	0.044	0.114	0.192	0.018	0.108	0.180
		$\lfloor N/4 \rfloor$	0.246	0.422	0.574	0.328	0.556	0.674	0.280	0.634	0.744
		$\lfloor N/2 \rfloor$	0.824	0.932	0.962	0.964	0.994	0.996	0.970	0.990	0.998
	50	$\lfloor N/8 \rfloor$	0.088	0.226	0.342	0.086	0.246	0.338	0.076	0.230	0.356
		$\lfloor N/4 \rfloor$	0.544	0.766	0.846	0.698	0.882	0.920	0.698	0.888	0.938
		$\lfloor N/2 \rfloor$	0.992	0.998	0.998	1.000	1.000	1.000	1.000	1.000	1.000
	60	$\lfloor N/8 \rfloor$	0.128	0.326	0.488	0.254	0.440	0.590	0.266	0.488	0.588
		$\lfloor N/4 \rfloor$	0.860	0.946	0.968	0.950	0.988	0.992	0.954	0.992	0.996
		$\lfloor N/2 \rfloor$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

are treated as unknown functions of macroeconomic and financial variables used in previous studies. Moreover, in contrast with existing studies, to the best of our knowledge, this is the first attempt to estimate conditional betas using the entire panel data allowing for possible cross-sectional dependence among portfolio returns. Finally, the nonparametric L_2-norm statistic is applied to testing the constancy of betas and the significance of pricing errors.

We collect monthly returns of the Fama–French 25 and 100 portfolios which are sorted by size ("S") and book-to-market ratio ("B") for the period from July 1963 to July 2018. The data of portfolio returns, the monthly risk-free rate, and the FF three factors are all downloaded from the Kenneth French Data Library.[3] Following Ferson and Harvey (1999) and Cai et al. (2015b), various conditional variables are considered, including the one-month Treasury bill yield (RF), the spread between the returns of the three-month and the one-month Treasury bill (r3m1), the spread between Moody's Baa and Aaa corporate bond yield (BmA), and the spread between a ten-year and one-year Treasury bond yields (r10m1).

The following heterogeneous vary-coefficient panel data model is adopted for estimating the conditional FF three-factor model:

$$R_{p,t+1} = \alpha_p(U_{p,t}) + \beta_{p,1}(U_{p,t})\mathrm{MKT}_{t+1} + \beta_{p,2}(U_{p,t})\mathrm{HML}_{t+1} + \beta_{p,3}(U_{p,t})\mathrm{SMB}_{t+1} + e_{p,t+1} \quad (5.1)$$

with $e_{p,t} = \gamma_{2p}^{\top} f_{2t} + \varepsilon_{p,t}$, where $R_{p,t+1}$ is the excess return of portfolio p at time $t+1$, MKT_{t+1}, HML_{t+1} and SMB_{t+1} are the Fama–French three factors at time $t+1$, and $U_{p,t}$ is one of the four aforementioned lagged instruments. The f_{2t} denotes a vector of unobservable factors. In order to implement our proposed estimation procedure, additional covariates Z_{pt} are needed to filter the unobserved common factors. Particularly, Z_{pt} are chosen to be the volatility, coskewness of Harvey and Siddique (2000), and kurtosis of portfolio returns.

First, testing the constancy of alphas and betas using single time series data of each portfolio is considered for the FF 25 portfolios. The null hypothesis is

$$H_0 : \alpha_p(u) = \alpha_p, \quad \beta_{p,j}(u) = \beta_{p,j} \quad \text{for } j = 1, 2, 3,$$

and the alternative hypothesis is

$$H_a : \text{either } \alpha_p(u) = f_{p,0}(u) \text{ or } \beta_{p,j}(u) = f_{p,j}(u) \text{ for some } j \in \{1, 2, 3\},$$

3 http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

Table 5
Proportion of rejections under H_a.

DGP	N	T	$c_0 = 0.5$			$c_0 = 1$			$c_0 = 2$		
			1%	5%	10%	1%	5%	10%	1%	5%	10%
δ	Tests based on Bootstrap critical values										
0.250	40	$\lfloor N/8 \rfloor$	0.052	0.124	0.212	0.064	0.162	0.248	0.068	0.150	0.240
		$\lfloor N/4 \rfloor$	0.416	0.654	0.784	0.542	0.800	0.872	0.590	0.838	0.900
		$\lfloor N/2 \rfloor$	0.988	1.000	1.000	0.996	0.998	1.000	0.998	1.000	1.000
	50	$\lfloor N/8 \rfloor$	0.112	0.314	0.432	0.254	0.454	0.560	0.150	0.338	0.468
		$\lfloor N/4 \rfloor$	0.758	0.918	0.952	0.920	0.980	0.990	0.918	0.986	0.992
		$\lfloor N/2 \rfloor$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	60	$\lfloor N/8 \rfloor$	0.232	0.560	0.676	0.342	0.650	0.738	0.414	0.598	0.720
		$\lfloor N/4 \rfloor$	0.986	0.998	0.998	0.994	1.000	1.000	0.998	1.000	1.000
		$\lfloor N/2 \rfloor$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.300	40	$\lfloor N/8 \rfloor$	0.064	0.158	0.266	0.072	0.182	0.312	0.070	0.154	0.292
		$\lfloor N/4 \rfloor$	0.710	0.852	0.928	0.814	0.924	0.964	0.742	0.926	0.962
		$\lfloor N/2 \rfloor$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	50	$\lfloor N/8 \rfloor$	0.196	0.396	0.518	0.326	0.486	0.624	0.250	0.458	0.590
		$\lfloor N/4 \rfloor$	0.920	0.980	0.992	0.986	0.996	0.998	0.982	0.998	0.998
		$\lfloor N/2 \rfloor$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	60	$\lfloor N/8 \rfloor$	0.530	0.708	0.800	0.610	0.820	0.890	0.614	0.842	0.912
		$\lfloor N/4 \rfloor$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$\lfloor N/2 \rfloor$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.350	40	$\lfloor N/8 \rfloor$	0.068	0.186	0.308	0.082	0.218	0.392	0.070	0.224	0.350
		$\lfloor N/4 \rfloor$	0.878	0.964	0.982	0.950	0.984	0.994	0.942	0.992	0.998
		$\lfloor N/2 \rfloor$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	50	$\lfloor N/8 \rfloor$	0.254	0.478	0.628	0.372	0.574	0.708	0.324	0.608	0.758
		$\lfloor N/4 \rfloor$	0.992	0.998	1.000	1.000	1.000	1.000	0.996	1.000	1.000
		$\lfloor N/2 \rfloor$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	60	$\lfloor N/8 \rfloor$	0.618	0.858	0.908	0.844	0.926	0.964	0.782	0.918	0.954
		$\lfloor N/4 \rfloor$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$\lfloor N/2 \rfloor$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 6
Bootstrap p-values for constancy test based on time series data.

	B1	B2	B3	B4	B5
Panel A: p-values using RF as conditioning variable					
S1	0.0180	0.0821	0.0841	0.0000	0.0150
S2	0.2062	0.0120	0.0220	0.0070	0.3784
S3	0.6176	0.0340	0.0020	0.0000	0.0200
S4	0.1181	0.0080	0.0000	0.0010	0.2943
S5	0.1502	0.0320	0.0400	0.0010	0.1041
Panel B: p-values using r3m1 as conditioning variable					
S1	0.0050	0.0350	0.1251	0.0000	0.0090
S2	0.1962	0.0000	0.0050	0.0060	0.3754
S3	0.8669	0.0160	0.0000	0.0000	0.0050
S4	0.4114	0.0000	0.0000	0.0000	0.0881
S5	0.0781	0.0000	0.0210	0.0000	0.0470
Panel C: p-values using BmA as conditioning variable					
S1	0.1912	0.2523	0.0000	0.6056	0.0130
S2	0.1351	0.0010	0.0460	0.0521	0.0330
S3	0.0571	0.1051	0.0010	0.0000	0.1622
S4	0.1772	0.0010	0.0000	0.0541	0.3133
S5	0.0190	0.0000	0.0030	0.0280	0.1652
Panel D: p-values using r10m1 as conditioning variable					
S1	0.0160	0.1852	0.0090	0.7427	0.0170
S2	0.1512	0.0000	0.0190	0.0280	0.0000
S3	0.9640	0.0030	0.0050	0.0000	0.0140
S4	0.7548	0.0000	0.0000	0.0060	0.0561
S5	0.6246	0.0000	0.0180	0.0250	0.0420

where $f_{p,j}(\cdot)$ for $j = 0, 1, 2, 3$ are unknown functions and at least one of them is varying with conditioning variable. Following Wang (2003), the test is conducted for different choice of conditioning variables from RF, r3m1, BmA and r10m1, respectively. The nonparametric test of Fan and Li (1999) is employed for testing the constancy of betas and

Table 7
Bootstrap *p*-values for testing alpha based on time series data.

	B1	B2	B3	B4	B5
Panel A: *p*-values using rf as conditioning variable					
S1	0.2120	0.0470**	0.8660	0.2800	0.3170
S2	0.1520	0.7740	0.9430	0.4190	0.7380
S3	0.9920	0.5070	0.8550	0.6970	0.9900
S4	0.3150	0.5710	0.4530	0.4990	0.1400
S5	0.1310	0.0880*	0.7900	0.2140	0.2240
Panel B: *p*-values using r3m1 as conditioning variable					
S1	0.1880	0.1850	0.1520	0.7900	0.6110
S2	0.5710	0.3110	0.0550*	0.1560	0.8900
S3	0.3310	0.0630*	0.2550	0.9670	0.6790
S4	0.3160	0.9090	0.6400	0.8230	0.4570
S5	0.1530	0.5210	0.5870	0.4500	0.0680*
Panel C: *p*-values using BmA as conditioning variable					
S1	0.5190	0.6280	0.4850	0.2930	0.6310
S2	0.1000	0.3990	0.8890	0.2220	0.3570
S3	0.7000	0.5890	0.8600	0.3230	0.7850
S4	0.2150	0.8470	0.1540	0.1770	0.8250
S5	0.0810*	0.7600	0.5450	0.7910	0.6870
Panel D: *p*-values using r10m1 as conditioning variable					
S1	0.2340	0.9690	0.1690	0.8350	0.1700
S2	0.8030	0.8340	0.1570	0.8580	0.8010
S3	0.1840	0.2020	0.1790	0.0070***	0.3990
S4	0.4950	0.9900	0.0800*	0.7730	0.0700*
S5	0.2550	0.6110	0.5170	0.4600	0.6370

*Rejections at 10% appear in the table.
**Rejections at 5% appear in the table.
***Rejections at 1% appear in the table.

Table 8
Bootstrap *p*-values for constancy test based on panel data.

	B1	B2	B3	B4	B5
Panel A: *p*-values using RF as conditioning variable					
S1	0.4565	0.0521	0.0000	0.9540	0.9980
S2	0.0100	0.0000	0.0000	0.0000	0.0000
S3	0.0000	0.0000	0.0000	0.0000	0.0000
S4	0.0000	0.0000	0.0000	0.0000	0.0140
S5	1.0000	0.0000	0.0250	0.0020	0.0240
Panel B: *p*-values using r3m1 as conditioning variable					
S1	0.4645	0.0581	0.0000	0.9520	0.9880
S2	0.3734	0.0000	0.0000	0.0000	0.0000
S3	0.0000	0.0000	0.0010	0.0000	0.0000
S4	0.0000	0.0000	0.0010	0.0000	0.0030
S5	1.0000	0.0000	0.0370	0.0040	0.0200
Panel C: *p*-values using BmA as conditioning variable					
S1	0.4525	0.0671	0.0000	1.0000	1.0000
S2	0.0320	0.0000	0.0000	0.0000	0.0000
S3	0.0000	0.0000	0.0000	0.0000	0.0000
S4	0.0000	0.0000	0.0000	0.0000	0.0010
S5	1.0000	0.0000	0.0480	0.0030	0.0190
Panel D: *p*-values using r10m1 as conditioning variable					
S1	0.4424	0.0621	0.0000	0.9910	0.9990
S2	0.0220	0.0000	0.0000	0.0000	0.0000
S3	0.0000	0.0000	0.0000	0.0000	0.0000
S4	0.0000	0.0000	0.0000	0.0010	0.0340
S5	1.0000	0.0000	0.0200	0.0010	0.0180

alphas. Table 6 reports the Bootstrapping *p*-values for testing H_0 based on 1000 Bootstrapping samples. The notations S1 through S5 and B1 through B5 stand for the FF quantiles on size and book-to-market ratio. The numbers in the row of S1 and the column of B5, for example, are the *p*-values for testing the constancy for the portfolio of stocks in the smallest size quantile and the highest book-to-market quantile. For each conditioning variable U_t, the bandwidth $h = 2.34\hat{\sigma}_u T^{-0.2}$ is

Table 9
Bootstrap p-values for testing alpha based on panel data.

	B1	B2	B3	B4	B5
Panel A: p-values using RF as conditioning variable					
S1	1.0000	0.3640	0.4170	0.4530	0.6060
S2	0.1520	1.0000	0.5660	1.0000	1.0000
S3	0.9840	0.5200	0.4820	0.9850	0.4140
S4	0.1340	0.9360	1.0000	0.1410	0.3210
S5	0.9310	1.0000	0.5910	0.1480	0.9980
Panel B: p-values using r3m1 as conditioning variable					
S1	0.9070	0.7680	0.5580	0.5470	0.6970
S2	0.3840	0.0940*	0.2870	0.1500	0.1890
S3	0.1060	0.1160	0.4150	0.3290	0.5290
S4	0.2020	0.3340	0.2840	0.3110	0.0930*
S5	0.0320**	0.1430	0.7370	0.1750	0.9960
Panel C: p-values using BmA as conditioning variable					
S1	0.5910	0.8180	0.6950	0.8180	0.2650
S2	0.8870	0.7870	0.5380	0.8850	0.2460
S3	0.6140	0.4840	0.6580	0.4310	0.5030
S4	0.9310	0.6590	0.2710	0.6890	0.9260
S5	0.5530	0.2860	0.5260	0.8150	0.9570
Panel D: p-values using r10m1 as conditioning variable					
S1	0.9370	0.9460	0.5500	0.7490	0.7730
S2	0.3060	0.1050	0.8410	0.7370	0.0940*
S3	0.4010	0.1600	0.4190	0.5940	0.7480
S4	0.4540	0.4180	0.1420	0.6720	0.4790
S5	0.6680	0.2680	0.6880	0.2920	0.9680

*Rejections at 10% appear in the table.
**Rejections at 5% appear in the table.

chosen, where $\hat{\sigma}_u$ is the sample standard deviation of U_t and $T = 660$. From Table 6, one can find that the rejection ratios at 5% nominal level for RF, r3m1, BmA and r10m1 are 0.64 (16/25, which means that 16 out of 25 portfolios reject the null and the following notations are defined in the same fashion), 0.72 (18/25), 0.52 (13/25) and 0.72 (18/25), respectively. These results are in line with the testing results (the majority of the 25 portfolios (17 out of 25) are in favor of time-varying betas) in Cai et al. (2015b) using the Fama–French 25 portfolios from July 1963 to December 2009.

The next is to test H_0 using FF 100 portfolios which are also formed on size and book-to-market ratio. Due to the issue of missing data, the actual number of portfolios under study is 96. The four portfolios excluded from our analysis are S7/B10, S10/B8, S10/B9, and S10/B10. The Bootstrapping p-values show that the rejection ratios at 5% nominal level for RF, r3m1, BmA and r10m1 are now 0.50 (48/96), 0.55 (53/96), 0.42 (40/96) and 0.49 (47/96), respectively. These rejection rates are significantly lower than those for 25 portfolios.

Now, it turns to testing the significance of the pricing error alpha. In view of the theory of conditional asset pricing model, alpha represents the abnormal returns of risky assets. If the functional-coefficient FF three-factor model is correct, one would expect that the estimated alphas are insignificant. Therefore, the following test is considered:

$$H_{\alpha,0} : \alpha_p = \alpha_p(u) = 0 \quad \text{versus} \quad H_{\alpha,1} : \alpha_p \neq 0.$$

Here, restrictions are only on the alphas, but the betas are allowed to vary with the conditioning variables. Table 7 reports the Bootstrapping p-values for testing $H_{\alpha,0}$ using 1000 Bootstrapping samples, which gives that the rejection rates for conditional variables r3m1 and BmA are 0 at 5% nominal level, but for the cases of RF and r10m1, the rejection rate is 0.04 (1/25) at 5% nominal level. The same procedure is then applied to the FF 100 portfolios. The rejection ratios at 5% nominal level for RF, r3m1, BmA and r10m1 are 0.021 (2/96), 0.063 (6/96), 0.031 (3/96) and 0.042 (4/96), respectively.

Now, the proposed test is applied to test the constancy of alphas and betas using the entire panel data to examine whether the power performance can be improved by using the pooled data. First, H_0 and $H_{\alpha,0}$ are re-tested using the Bootstrapping procedure described in Section 3.3 for the FF 25 portfolios. The Bootstrapping p-values for testing H_0 are reported in Table 8, from which one can observe clearly that the rejection ratio at 5% nominal level for r3m1 is 0.76 (19/25) and 0.80 (20/25) for other conditioning variables. The Bootstrapping p-values for testing $H_{\alpha,0}$ are reported in Table 9, from which one can see that only $H_{\alpha,0}$ for the portfolio S5/B1 is rejected using r3m1 as the conditioning variable at 5% nominal level.

Next, H_0 and $H_{\alpha,0}$ are also re-tested using the entire panel data for the FF 100 portfolios. The p-values of both tests are demonstrated in Figs. 2 and 3 via heat maps, respectively.[4] From Fig. 2, one can observe that the majority of the

[4] Note that, in each heat map given in Figs. 2 or 3, the upper right corner has 4 blank grids which indicate the four portfolios whose data is missing.

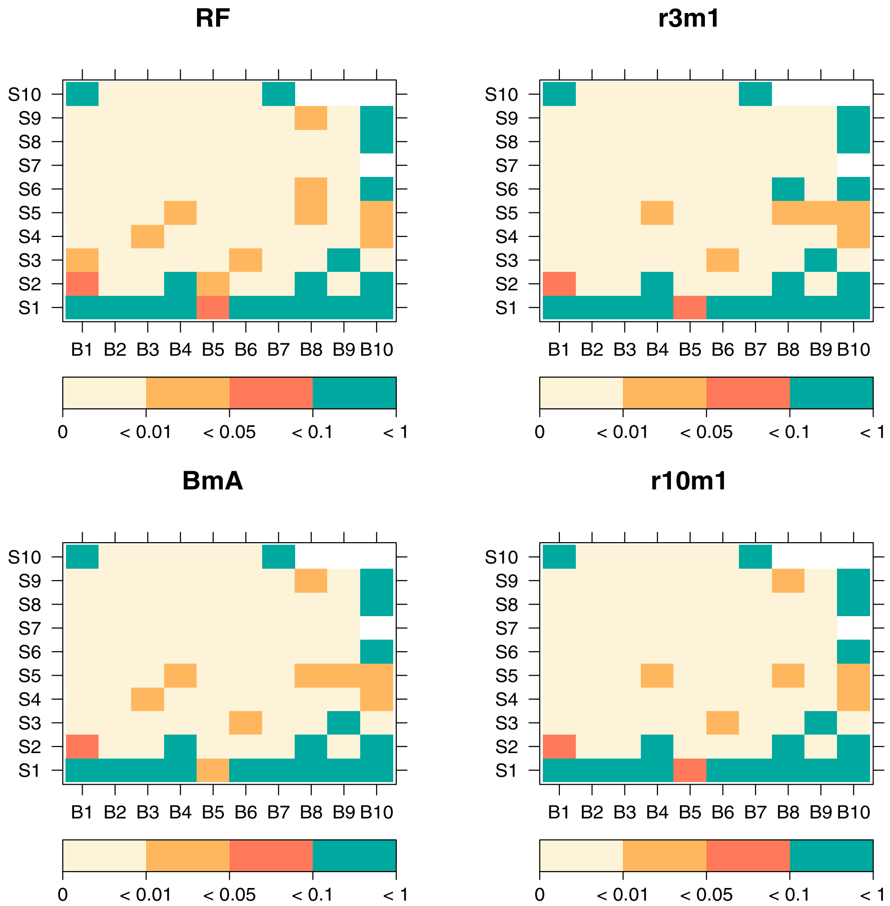

Fig. 2. Heat maps of Bootstrap *p*-values for constancy test based on FF 100 portfolios panel data.

portfolios have *p*-values smaller than 0.05 in the constancy test no matter which conditioning variable is used. In fact, the rejection ratios at 5% nominal level for RF, r3m1, BmA and r10m1 are 0.79 (76/96), 0.78 (75/96), 0.80 (77/96) and 0.79 (76/96), respectively. Therefore, the constancy test using the entire panel data does increase the test powers in general for all conditioning variables. Moreover, contrary to the testing results based on time series data, our method gives a much stronger support to the conditional asset pricing models with three factors.

For the pricing error testing results given in Fig. 3, it is obvious to see again that our method offers more powerful support to the conditional FF three-factor model than the method using time series data. For example, using r3m1 as the conditioning variable, the number of portfolios rejects by the time series data is 6. However, from the upper right heat map in Fig. 3, one can easily recognize that only 3 portfolios (S3/B1, S5/B5, S8/B9) are rejected by our method.

Finally, to assess the robustness of these results, we conduct our previous tests using two different samples: the period from July 1963 to December 1994, which is the sample used by Ferson and Harvey (1999), and the period from July 1963 to December 2009, which is analyzed by Cai et al. (2015b). The results are qualitatively similar to those reported above. For brevity, these results are not shown here, but are available in the online supplementary material.[5]

6. Conclusion

Inferences on capital asset pricing model have gained a lot of attentions in the literature in the recent years. In this paper, a novel functional-coefficient panel data model with cross-sectional dependence is proposed to revisit this issue. In our model, the time-varying property of betas is characterized through unknown functions of certain macroeconomic and financial instruments. Moreover, the model allows for possible cross-sectional dependence via a multi-factor structure. A local common correlated effect estimation method is proposed to estimate the proposed model. The consistency and asymptotic normality of the proposed estimators have been derived when both the cross-sectional dimension N and the

[5] The web site is http://www.people.ku.edu/~z397c158/CFX-Supplement.pdf.

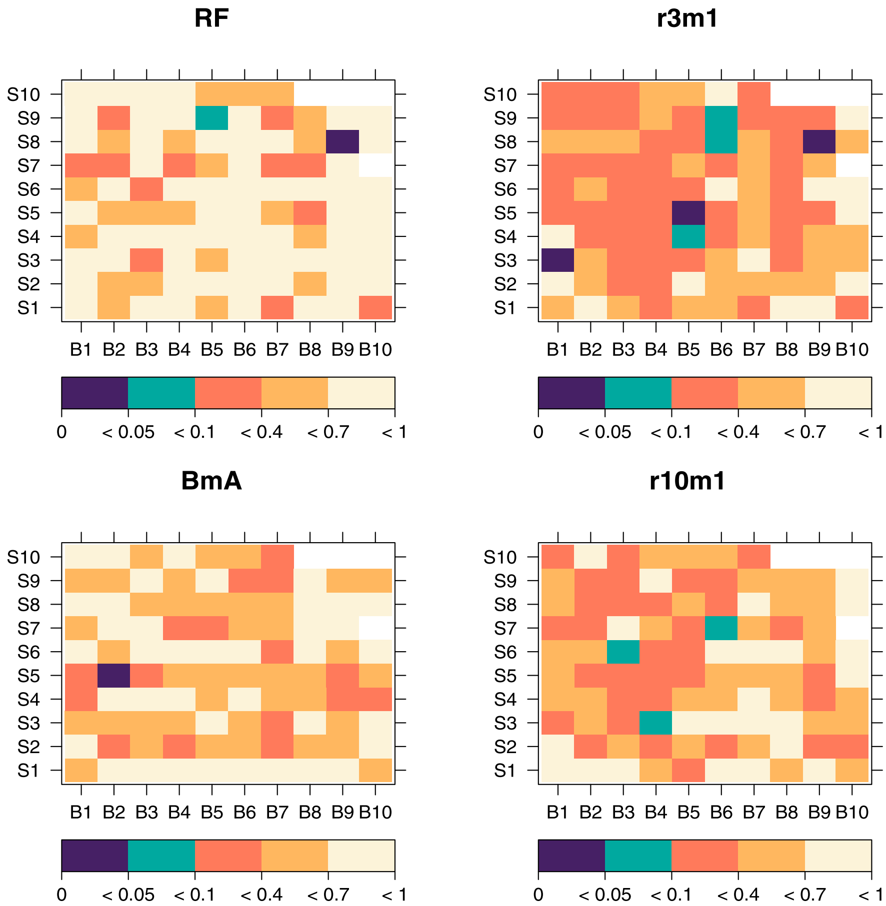

Fig. 3. Heat maps of Bootstrap *p*-values for testing alpha based on FF 100 portfolios panel data.

time series dimension T tend to infinity simultaneously. We have also constructed a simple goodness-of-fit test for testing the stability of the model coefficients based on L_2-norm and showed that the new test statistic has an asymptotically standard normal distribution. The proposed test statistic is applied to test the constancy of conditional betas and the significance of alphas in the Fama–French three-factor model. The empirical results of our study advocate the finding that betas of risk factors depend on certain state variables. Besides, it is discovered that testing on conditional betas and pricing errors can achieve substantial efficiency gains by utilizing panel data. Our empirical findings suggest that applications of well-known asset pricing models should control for time-varying betas and take dependence among different assets into account.

Future studies can be conducted on more general inference problems on varying and/or partially varying coefficients in panel data models with cross-sectional dependence and it might be of interest in using the interactive fixed effect estimation approach proposed in Bai (2009) to consider our model in (2.1). Such new models and their modeling approaches can be applied to making inference on the conditional capital asset pricing models.

Acknowledgments

The authors thank the Editor (Professor Torben Andersen) and the Guest Editor (Professor Michael McAleer) and two anonymous reviewers for their helpful and constructive comments to improve significantly the quality of the paper. This research was partially supported from the National Natural Science Foundation of China (NSFC) for Distinguished Young Scholars (71625001), the NSFC key projects with Grant Numbers 71631004 and 71131008, and the Basic Science Center Program of NSFC (71988101).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2020.07.018. All theoretical proofs are presented in the form of supplementary material, which can be found online at http://www.people.ku.edu/~z397c158/CFX-Supplement.pdf.

References

Ang, A., Kristensen, D., 2012. Testing conditional factor models. J. Financ. Econom. 106 (1), 132–156.
Bai, J., 2009. Panel data models with interactive fixed effects. Econometrica 77 (4), 1229–1279.
Cai, Z., Chen, L., Fang, Y., 2015a. Semiparametric estimation of partially varying-coefficient dynamic panel data models. Econometric Rev. 34 (6–10), 695–719.
Cai, Z., Chen, L., Fang, Y., 2018. A semiparametric quantile panel data model with an application to estimating the growth effect of FDI. J. Econometrics 206 (2), 531–553.
Cai, Z., Fan, J., Yao, Q., 2000. Functional-coefficient regression models for nonlinear time series. J. Amer. Statist. Assoc. 95 (451), 941–956.
Cai, Z., Li, Q., 2008. Nonparametric estimation of varying coefficient dynamic panel data models. Econometric Theory 24 (05), 1321–1342.
Cai, Z., Ren, Y., Yang, B., 2015b. A semiparametric conditional capital asset pricing model. J. Bank. Financ. 61, 117–126.
Chen, J., Gao, J., Li, D., 2012. Semiparametric trending panel data models with cross-sectional dependence. J. Econometrics 171 (1), 71–85.
Chudik, A., Pesaran, M., Tosetti, E., 2011. Weak and strong cross-section dependence and estimation of large panels. Econom. J. 14 (1), C45–C90.
Connor, G., Hagmann, M., Linton, O., 2012. Efficient semiparametric estimation of the Fama–French model and extensions. Econometrica 80 (2), 713–754.
Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. J. Financ. Econom. 33 (1), 3–56.
Fama, E.F., French, K.R., 1997. Industry costs of equity. J. Financ. Econom. 43 (2), 153–193.
Fama, E.F., French, K.R., 2018. Choosing factors. J. Financ. Econom. 128 (2), 234–252.
Fama, E.F., MacBeth, J.D., 1973. Risk, return, and equilibrium: Empirical tests. J. Political Econ. 81 (3), 607–636.
Fan, J., Huang, T., 2005. Profile likelihood inferences on semiparametric varying-coefficient partially linear models. Bernoulli 11 (6), 1031–1057.
Fan, J., Yao, Q., 2003. Nonlinear Time Series: Nonparametric and Parametric Methods. Springer Science & Business Media.
Fan, Y., Li, Q., 1999. Central limit theorem for degenerate U-statistics of absolutely regular processes with applications to model specification testing. J. Nonparametr. Stat. 10 (3), 245–271.
Ferreira, E., Gil-Bazo, J., Orbe, S., 2011. Conditional beta pricing models: A nonparametric approach. J. Bank. Financ. 35 (12), 3362–3382.
Ferson, W.E., Harvey, C.R., 1999. Conditioning variables and the cross section of stock returns. J. Finance 54 (4), 1325–1360.
Guo, H., Wu, C., Yu, Y., 2017. Time-varying beta and the value premium. J. Financ. Quant. Anal. 52 (4), 1551–1576.
Hall, P., 1984. Central limit theorem for integrated square error of multivariate nonparametric density estimators. J. Multivariate Anal. 14 (1), 1–16.
Hansen, B.E., 2000. Testing for structural change in conditional models. J. Econometrics 97 (1), 93–115.
Hansen, B.E., 2008. Uniform convergence rates for kernel estimation with dependent data. Econometric Theory 24 (3), 726–748.
Harvey, C.R., Siddique, A., 2000. Conditional skewness in asset pricing tests. J. Finance 55 (3), 1263–1295.
Hjalmarsson, E., 2010. Predicting global stock returns. J. Financ. Quant. Anal. 45 (1), 49–80.
Hjellvik, V., Yao, Q., Tjøstheim, D., 1998. Linearity testing using local polynomial approximation. J. Statist. Plann. Inference 68 (2), 295–321.
Huang, X., 2013. Nonparametric estimation in large panels with cross-sectional dependence. Econometric Rev. 32 (5–6), 754–777.
Jin, S., Su, L., 2013. A nonparametric poolability test for panel data models with cross section dependence. Econometric Rev. 32 (4), 469–512.
Lettau, M., Ludvigson, S., 2001. Resurrecting the (C) CAPM: A cross-sectional test when risk premia are time-varying. J. Political Econ. 109 (6), 1238–1287.
Lewellen, J., Nagel, S., 2006. The conditional CAPM does not explain asset-pricing anomalies. J. Financ. Econom. 82 (2), 289–314.
Li, Q., Huang, C.J., Li, D., Fu, T.T., 2002. Semiparametric smooth coefficient models. J. Bus. Econom. Statist. 20 (3), 412–422.
Lin, Z., Li, Q., Sun, Y., 2014. A consistent nonparametric test of parametric regression functional form in fixed effects panel data models. J. Econometrics 178 (1), 167–179.
Moon, H.R., Weidner, M., 2015. Linear regression for panel with unknown number of factors as interactive fixed effects. Econometrica 83 (4), 1543–1579.
Pesaran, M.H., 2006. Estimation and inference in large heterogeneous panels with a multifactor error structure. Econometrica 74 (4), 967–1012.
Powell, J.L., Stock, J.H., Stoker, T.M., 1989. Semiparametric estimation of index coefficients. Econometrica 57 (6), 1403–1430.
Su, L., Jin, S., 2012. Sieve estimation of panel data models with cross section dependence. J. Econometrics 169 (1), 34–47.
Sun, Y., Carroll, R., Li, D., 2009. Semiparametric estimation of fixed-effects panel data varying coefficient models. Adv. Econom. 25, 101–129.
Sun, Y., Zhang, Y., Li, Q., 2015. Nonparametric panel data regression models. In: Baltagi, B. (Ed.), The Oxford Handbook of Panel Data. Oxford University Press, New York.
Wang, K.Q., 2003. Asset pricing with conditioning information: A new test. J. Finance 58 (1), 161–196.
Xu, Q., Cai, Z., Fang, Y., 2016. Panel data models with cross-sectional dependence: a selective review. Appl. Math.-J. Chin. Univ. 31 (2), 127–147.
Zhang, L., 2005. The value premium. J. Finance 60 (1), 67–103.