Econometric Reviews

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/lecr20

Semiparametric inferences for panel data models with fixed effects via nearest neighbor difference transformation

Qiuhua Xu, Zongwu Cai & Ying Fang

To cite this article: Qiuhua Xu, Zongwu Cai & Ying Fang (2021) Semiparametric inferences for panel data models with fixed effects via nearest neighbor difference transformation, Econometric Reviews, 40:10, 919-943, DOI: 10.1080/07474938.2021.1889197

To link to this article: https://doi.org/10.1080/07474938.2021.1889197

Published online: 05 Aug 2021.

ECONOMETRIC REVIEWS
2021, VOL. 40, NO. 10, 919–943
https://doi.org/10.1080/07474938.2021.1889197

Semiparametric inferences for panel data models with fixed effects via nearest neighbor difference transformation

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

[a]School of Finance, Southwestern University of Finance and Economics, Chengdu, Sichuan 611130, China; [b]Department of Economics, University of Kansas, Lawrence, KS, 66045, USA; [c]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

ABSTRACT
In this paper, we propose a simple method to estimate a partially varying-coefficient panel data model with fixed effects. By taking difference upon the nearest neighbor of the smoothing variables to remove the fixed effects, we employ the profile least squares method and local linear fitting to estimate the parametric and nonparametric parts, respectively. Moreover, a functional form specification test and a nonparametric Hausman type test are constructed and their asymptotic properties are derived. Monte Carlo simulations are conducted to examine the finite sample performance of our estimators and test statistics.

KEYWORDS
Fixed effects; Hausman-type test; panel data; specification test; varying coefficients

JEL CLASSIFICATION
C12; C13; C14; C23

1. Introduction

Panel data models have been of great interest in the literature of theoretical econometrics and applied economics since the seminal work of Balestra and Nerlove (1966). Arellano (2003), Baltagi (2008), and Hsiao (2014) provided excellent and detailed surveys of the panel data literature. Panel data can provide a sample of individual observations over time. Compared to cross-sectional and time-series data, one appealing feature of a panel data model is to control unobserved heterogeneity in a regression model by employing individual effects.[1] There are two common methods to define the individual effects. The fixed effects model allows individual effects to have an unspecified dependence upon regressors, while the random effects model assumes zero correlation between individual effects and regressors. The random effects model can achieve efficient estimation, but the lack of correlation between random effects and regressors is usually hard to verify in practice. The fixed effects model can obtain consistent estimation even with nonzero correlation between individual effects and regressors. As for how to choose them in real applications, Hausman and Taylor (1981) proposed a Hausman-type test to select a model specification between the fixed and random effects models.

Nonparametric and semiparametric panel data models have become increasingly important because they can provide flexible model specifications under relatively weak conditions. However, the literature on nonparametric or semiparametric inference on fixed effects panel data models is not well developed. In linear panel data models, the fixed effects can be removed by a first-order differencing transformation, but in nonparametric or semiparametric panel data models, a

CONTACT Ying Fang yifst1@xmu.edu.cn 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.
[1]Panel data have many advantages over cross-sectional or time series data. More details can be found in Hsiao (2014).

transformation based on a first-order difference leads to a difference of the same unknown function evaluated at different values, which may cause difficulty in estimation.

One approach to this problem is to treat the fixed effects as dummy variables and then avoid the first-order differencing transformation. Indeed, Su and Ullah (2006) considered a partially linear panel data model with fixed effects and proposed dummy variable profile likelihood estimation for both the parametric and nonparametric parts, Zhang et al. (2011) proposed an empirical likelihood estimation for the same model as the one in Su and Ullah (2006), and Mammen et al. (2009) considered a general additive panel data model with fixed effects and proposed a smoothed backfitting approach to estimate the nonparametric part. Some studies focused on estimating the difference of the same unknown function caused from the first order differencing transformation. For example, Henderson et al. (2008) proposed an iterative procedure to estimate the nonparametric part based on maximizing a local weighted profile likelihood, but without theory, while Qian and Wang (2012) employed marginal integration to recover the nonparametric part. Recent contributions to the nonparametric or semiparametric estimation of panel data models include, but not limited to, papers by Cai and Li (2008), Huang (2013), Hoover et al. (1998), Li and Stengos (1996), Li et al. (2011), Su and Ullah (2007), Su and Jin (2012), and Xue and Zhu (2007), and the references therein.

In this paper, we consider the estimation and testing of this partially varying-coefficient panel data model with fixed effects:

$$Y_{it} = X_{it}^{\top}\beta(u_{it}) + Z_{it}^{\top}\gamma + \alpha_i + \epsilon_{it}. \tag{1.1}$$

When regressors X_{it} only contain a constant term, the above model is reduced to a partially linear panel data model with fixed effects; see, for example, Henderson et al. (2008), Qian and Wang (2012), Su and Ullah (2006) and the references therein. By ignoring the fixed effects, model (1.1) becomes to the one studied in Li and Stengos (1996) and a special case of model (1.1) was considered by Hoover et al. (1998) without including the parametric part and the fixed effects.

Our estimation of model (1.1) is based on a differencing transformation. However, local polynomial estimation cannot be directly applied to the first-order differencing transformed model because the Taylor series expansion is not guaranteed to well approximate both $\beta(u_{it})$ and $\beta(u_{i,t-1})$ simultaneously, and furthermore, we do not know the order of the difference between $\beta(u_{it})$ and $\beta(u_{i,t-1})$. Instead, we solve the above problem by taking the difference to the nearest neighbor, denoted by u^*, of u_{it} among the sample observations across t. Therefore, the maximal distance $\|\beta(u_{it}) - \beta(u^*)\|$ is bounded by $O_p(\log T/T)$ based on the result in Janson (1987), and then profile least square and local polynomial estimation can be applied to estimate the constant and varying coefficients, respectively. Compared to existing estimation methods, our approach based on nearest neighbor difference transformation is simple and can share most merits of local polynomial estimation.

In partially varying-coefficient panel data models, it is interesting to test a specified parametric model versus a nonparametric model. We construct a test statistic based on the weighted integrated squared difference between the specified parametric function and the nonparametric estimator, similar to Li et al. (2002). Another important specification issue in the panel data model is to test a fixed effects model versus a random effects model. We construct a nonparametric Hausman-type test similar to the one studied in Henderson et al. (2008). We derive the limiting distributions of both test statistics and conduct simulations to investigate their finite sample performance. However, note that the test statistics proposed in Henderson et al. (2008) are based on an iteratively derived estimator, so that Henderson et al. (2008) did not provide their asymptotic properties.

In comparison with the existing literature on estimating and testing semiparametric fixed effects panel data models, the main difficulty in deriving asymptotic properties is due to the change of the variance–covariance matrix of the error terms after the nearest neighbor differencing transformation. Although the original observations are independently and identically distributed across individuals, the transformed data may alter the distribution of data for a given individual and lead to independently but nonidentically distributed observations across the cross-sectional units. Therefore, the asymptotic analysis of the proposed estimators and statistics are nontrivial, and it is much more involved to establish the asymptotical normality of parametric functional form test statistics.

The rest of the paper is organized as follows. Section 2 introduces the model and estimation methods and Section 3 details the construction of the functional form specification test and the nonparametric Hausman-type test and presents their asymptotic distributions. We provide asymptotic results of our proposed estimators and test statistics in Section 4. Monte Carlo simulations are conducted in Section 5 to investigate the finite sample performance of the proposed estimators and test statistics. Section 6 summarizes our conclusions. All proofs are given in the appendices.

2 Model and estimation procedure

This paper considers the following partially varying-coefficient fixed effects panel data model:

$$Y_{it} = X_{it}^{\top}\beta(u_{it}) + Z_{it}^{\top}\gamma + \alpha_i + \epsilon_{it}, \quad 1 \le i \le N, 1 \le t \le T, \tag{2.1}$$

where Y_{it} is a scalar dependent variable; X_{it} and Z_{it} are $p \times 1$ and $q \times 1$ vectors, respectively; γ denotes a $q \times 1$ constant coefficient, and $\beta(u_{it})$ denotes a $p \times 1$ vector of unknown smooth functions defined on $\mathbb{R}^d$, which has a continuous second derivative.[2] Here, α_i represents the fixed effects and model (2.1) allows the fixed effects to be correlated with the regressors in an arbitrary way. Finally, ϵ_{it} is independently identically distributed (i.i.d.) random error. Note that $A^{\top}$ denotes the transpose of A. In practice, X_{it} is the vector of main variables that one cares about for their time-varying partial effects on Y_{it} given U_{it}; that is, the conditional correlation between Y_{it} and X_{it} given U_{it} is nonlinear in U_{it}, whereas Z_{it} is the vector of other covariates that should be controlled for. The choice of U_{it} is usually based on certain economic theory. Of course, it can be chosen by using a data-driven method as suggested by Cai et al. (2000). Also, it is possible that U_{it} is a proper subset of X_{it} or Z_{it}.

To remove the fixed effects, we first take the difference of model (2.1) upon the nearest neighbor of u_{it} among all within-group observations. Denote by t^* the position of the nearest neighbor of t such that $u_{it^*} = \text{argmin}_{\{u_{is}:1 \le s \le T, s \ne t\}} \|u_{is} - u_{it}\|$. Then the transformed model is defined as

$$\Delta^* Y_{it} = X_{it}^{\top}\beta(u_{it}) - X_{it^*}^{\top}\beta(u_{it^*}) + \Delta^* Z_{it}^{\top}\gamma + \Delta^*\epsilon_{it}, \tag{2.2}$$

where $\Delta^* Y_{it} = Y_{it} - Y_{it^*}$, and $\Delta^* X_{it}, \Delta^* Z_{it}$, and $\Delta^*\epsilon_{it}$ are defined in the same fashion. Under some regular conditions, $\|\beta(u_{it}) - \beta(u_{it^*})\| = O_p(\log T/T)$, based on Janson (1987), where $\|A\|$ denotes the Frobenius norm of A, and then (2.2) can be further simplified to

$$\Delta^* Y_{it} = \Delta^* X_{it}^{\top}\beta(u_{it}) + \Delta^* Z_{it}^{\top}\gamma + P_{it^*,T} + \Delta^*\epsilon_{it},$$

where $P_{it^*,T} \equiv X_{it^*}^{\top} O_p(\log T/T)$. However, the variance–covariance matrix of $\Delta^*\epsilon_{it}$ is usually singular due to the nearest neighbor difference. For example, if u_{it} and u_{is} are nearest neighbors, then the variance–covariance matrix must contain two linear dependent columns. To avoid this problem, we can achieve a nonsingular variance–covariance matrix by simply dropping some sample observations within each group $i \in N$ according to the following rules: (1) When there is more than one nearest neighbor, we choose the one with the smallest time index. For example, if u_{i2} and u_{i3} are nearest to u_{i1}, then we choose u_{i2}. (2) If, in the variance–covariance matrix, one

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

row can be expressed by another row, or more than one row, then we delete that row. For example, if the nearest neighbor of u_{i1} is u_{i2}, and the nearest neighbor of u_{i2} is u_{i1}, then u_{i2} is deleted. Rearranging all the remaining observations, the transformed model is now given by

$$\Delta^* y_{it} = \Delta^* x_{it}^\top \beta(u_{it}) + \Delta^* z_{it}^\top \gamma + P_{it^*,T} + \Delta^* \epsilon_{it}, \quad 1 \le i \le N, 1 \le t \le T_i, \tag{2.3}$$

where $\Delta^* y_{it}, \Delta^* x_{it}$, and $\Delta^* z_{it}$ represent transformed and rearranged sample observations, and T_i denotes the remaining number of observations in each group i. To simplify the notation, we consider a balanced panel by assuming T_i is the same for all i, denoted by $T_i = T$ for $i = 1, ..., N$.

Indeed, (2.3) is a partially varying-coefficient model, which is usually estimated by a profile likelihood method; see, for example, Chen et al. (2012) and Fan and Huang (2005). Define $Y = (\Delta^* y_{11}, ..., \Delta^* y_{1T}, ..., \Delta^* y_{N1}, ..., \Delta^* y_{NT})^\top$, and define X, Z, and ϵ in the same fashion. Then model (2.3) can be expressed in matrix form as

$$Y = M + Z\gamma + P_T + \epsilon, \tag{2.4}$$

where

$$M = \begin{pmatrix} \Delta^* x_{11}^\top \beta(u_{11}) \\ \vdots \\ \Delta^* x_{1T}^\top \beta(u_{1T}) \\ \vdots \\ \Delta^* x_{N1}^\top \beta(u_{N1}) \\ \vdots \\ \Delta^* x_{NT}^\top \beta(u_{NT}) \end{pmatrix}, \quad P_T = \begin{pmatrix} P_{11^*,T} \\ \vdots \\ P_{1T^*,T} \\ \vdots \\ P_{N1^*,T} \\ \vdots \\ P_{NT^*,T} \end{pmatrix} = \begin{pmatrix} X_{11^*}^\top \\ \vdots \\ X_{1T^*}^\top \\ \vdots \\ X_{N1^*}^\top \\ \vdots \\ X_{NT^*}^\top \end{pmatrix} O_p(\log T/T), \tag{2.5}$$

which implies that $\tilde{Y} \equiv Y - Z\gamma - P_T = M + \epsilon$. Assuming that γ is known, the local linear estimator of $\beta(u)$ and its first-order derivative are given by

$$\begin{pmatrix} \hat{\beta}_\gamma(u) \\ h\hat{\beta}_\gamma'(u) \end{pmatrix} = \{D_u^\top H_u D_u\}^{-1} D_u^\top H_u (Y - Z\gamma), \tag{2.6}$$

where $\beta_\gamma'(u)$ denotes the first-order derivative of $\beta_\gamma(\cdot)$ at u, and the subscript "γ" signifies that the estimator depends on γ. In addition, h is the bandwidth, $H_u = \text{diag}\{K_h(u_{11} - u), ..., K_h(u_{NT} - u)\}$ with $K_h(\cdot) = K(\cdot/h)/h, K(\cdot)$ is the kernel function, and

$$D_u = \begin{pmatrix} \Delta^* x_{11}^\top & \Delta^* x_{11}^\top \frac{u_{11}-u}{h} \\ \vdots & \vdots \\ \Delta^* x_{1T}^\top & \Delta^* x_{1T}^\top \frac{u_{1T}-u}{h} \\ \vdots & \vdots \\ \Delta^* x_{N1}^\top & \Delta^* x_{N1}^\top \frac{u_{N1}-u}{h} \\ \vdots & \vdots \\ \Delta^* x_{NT}^\top & \Delta^* x_{NT}^\top \frac{u_{NT}-u}{h} \end{pmatrix}. \tag{2.7}$$

Therefore, the local linear estimator of M is given by

$$\hat{M} = \begin{pmatrix} (\Delta^* x_{11}^\top, \mathbf{0})\{D_{u_{11}}^\top H_{u_{11}} D_{u_{11}}\}^{-1} D_{u_{11}}^\top H_{u_{11}} \\ \vdots \\ (\Delta^* x_{NT}^\top, \mathbf{0})\{D_{u_{NT}}^\top H_{u_{NT}} D_{u_{NT}}\}^{-1} D_{u_{NT}}^\top H_{u_{NT}} \end{pmatrix} (Y - Z\gamma) \equiv \hat{S}(Y - Z\gamma), \tag{2.8}$$

where the definition of $\hat{S}$ should be apparent from the above equation. Next, replacing M in (2.4) by (2.8), we obtain

$$(I - \hat{S})Y = (I - \hat{S})Z\gamma + P_T + \epsilon. \tag{2.9}$$

Simple calculation leads to the profile least square estimator of γ:

$$\hat{\gamma} = \left\{Z^\top (I - \hat{S})^\top (I - \hat{S})Z\right\}^{-1} Z^\top (I - \hat{S})^\top (I - \hat{S})Y. \tag{2.10}$$

Finally, substituting (2.10) into (2.6), the local linear estimator of $\beta(u)$ and its first-order derivative are given by

$$\begin{pmatrix} \hat{\beta}(u) \\ h\hat{\beta}'(u) \end{pmatrix} = \{D_u^\top H_u D_u\}^{-1} D_u^\top H_u (Y - Z\hat{\gamma}). \tag{2.11}$$

3. Model specification tests

In this section, we develop two types of model specification tests in the partially varying-coefficient panel data model: the first one is to test whether the varying coefficients $\beta(u)$ can take a specified parametric functional form, and the second is to test the fixed effects model versus the random effects model.

3.1. Testing functional forms of varying coefficients

Consider the case that the varying coefficients $\beta(u_{it})$ in model (2.1) can take a specified functional form due to some prior knowledge. Although the nonparametric treatment of varying coefficients accommodates more model flexibility, a parametric estimate can achieve efficiency gains if one believes that the parametric functional form is adequate. Therefore, it is of interest to test the null hypothesis $H_{10} : \beta(u) = \beta_0(u; \theta)$ for all u versus the alternative $H_{11} : \beta(u) \neq \beta_0(u; \theta)$, where $\beta_0(u; \theta)$ denotes a specified parametric function with unknown parameter θ. Note that the model under the null hypothesis has several interesting cases. For example, the test is reduced to a significance test of X_{it} when $\beta_0(u; \theta) = 0$, and it further reduces to a constancy test when $\beta_0(u; \theta) = \theta$.

Our consistent test statistic is constructed based on the weighted integrated squared difference between the estimators of $\beta(u)$ and $\beta_0(u; \theta)$. In Section 2, $\beta(u)$ is estimated based on the local linear method. Although it can take care of boundary problems, the local linear estimator complicates the asymptotic analysis of the test statistic if it is used to construct the test statistic. Fortunately, as shown in Li et al. (2002), it suffices to use the local constant estimator of $\beta(u)$ to construct the test statistic. The local constant estimator of $\beta(u)$ in (2.1), denoted by $\hat{\beta}_{lc}(u)$, is given by

$$\hat{\beta}_{lc}(u) = B_u^{-1} \left[\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \Delta^* x_{it} (\Delta^* y_{it} - \Delta^* z_{it}^\top \hat{\gamma}) K_h(u_{it} - u) \right],$$

where $B_u = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Delta^* x_{it}\Delta^* x_{it}^\top K_h(u_{it}-u)$. Let $\hat{\beta}_0(u) \equiv \beta_0(u;\hat{\theta})$, where $\hat{\theta}$ is a root-NT consistent estimator of θ. Our test statistic is given by

$$W_{NT} = \int \left[B_u(\hat{\beta}_{lc}(u) - \hat{\beta}_0(u))\right]^\top \left[B_u(\hat{\beta}_{lc}(u) - \hat{\beta}_0(u))\right] du. \tag{3.1}$$

By substituting the expression of B_u into the above equation, one obtains

$$\begin{aligned} W_{NT} = \int &\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Delta^* x_{it}(\Delta^* y_{it} - \Delta^* z_{it}^\top \hat{\gamma} - \Delta^* x_{it}^\top \hat{\beta}_0(u))K_h(u_{it}-u)\right]^\top \\ \times &\left[\frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\Delta^* x_{js}(\Delta^* y_{js} - \Delta^* z_{js}^\top \hat{\gamma} - \Delta^* x_{js}^\top \hat{\beta}_0(u))K_h(u_{js}-u)\right] du. \end{aligned} \tag{3.2}$$

Following Li et al. (2002), we delete the $i=j$ terms in (3.2) to remove the nonzero center term of W_{NT} under H_{10}. Furthermore, $\hat{\beta}_0(u)$ in the first and second brackets is replaced by $\hat{\beta}_0(u_{it})$ and $\hat{\beta}_0(u_{js})$, respectively. These procedures lead to the following test statistic:

$$\begin{aligned} W_{NT}^* = \frac{1}{N^2T^2h}\sum_{i\neq j}\sum_{t=1}^{T}\sum_{s=1}^{T} &\left[\Delta^* x_{it}(\Delta^* y_{it} - \Delta^* z_{it}^\top \hat{\gamma} - \Delta^* x_{it}^\top \hat{\beta}_0(u_{it}))\right]^\top \\ \times &\left[\Delta^* x_{js}(\Delta^* y_{js} - \Delta^* z_{js}^\top \hat{\gamma} - \Delta^* x_{js}\hat{\beta}_0(u_{js}))\right]\bar{K}\left(\frac{u_{it}-u_{js}}{h}\right), \end{aligned} \tag{3.3}$$

where $\sum_{i\neq j}$ is the abbreviation for $\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}$, and $\bar{K}(v) = \int K(u)K(v-u)du$ is the twofold convolution kernel derived from $K(\cdot)$. As suggested by Li et al. (2002) and Lin et al. (2014), one need not even use the convolution kernels. Simply replacing $\bar{K}((u_{it}-u_{js})/h)$ by $K((u_{it}-u_{js})/h)$ results in the following simple test statistic:

$$\begin{aligned} \tilde{W}_{NT} = \frac{1}{N^2T^2h}\sum_{i\neq j}\sum_{t=1}^{T}\sum_{s=1}^{T} &\left[\Delta^* x_{it}(\Delta^* y_{it} - \Delta^* z_{it}^\top \hat{\gamma} - \Delta^* x_{it}^\top \hat{\beta}_0(u_{it}))\right]^\top \\ \times &\left[\Delta^* x_{js}(\Delta^* y_{js} - \Delta^* z_{js}^\top \hat{\gamma} - \Delta^* x_{js}\hat{\beta}_0(u_{js}))\right]K\left(\frac{u_{it}-u_{js}}{h}\right). \end{aligned} \tag{3.4}$$

3.2. Testing fixed versus random effects

The selection between the fixed effects model and random effects model is an important issue in the panel data literature. The fixed effects model can achieve consistent estimation under very mild conditions, while the random effects model can achieve efficient estimation when the individual effects are uncorrelated with the regressors. In other words, the fixed effects estimators can guarantee consistent estimation, but the random effects estimators may achieve efficient estimation bounds. Under the null hypothesis of the random effects, both estimators should be asymptotically close. Hausman and Taylor (1981) used this simple fact to develop a Hausman-type test to compare both estimators. Henderson et al. (2008) proposed a nonparametric test of fixed effects versus random effects. Instead of comparing fixed effects and random effects estimators, they directly tested whether the individual effects are mean independence conditional on all regressors. Since the test is based on conditional moments, it becomes a special case of Zheng (1996). However, Henderson et al. (2008) did not provide the asymptotic property of their proposed nonparametric Hausman-type test because their test is based on an iteratively derived estimator. This paper adopts a similar nonparametric test but constructs the test statistic using the local linear estimator proposed in Section 2 to estimate $\beta(u_{it})$.

Consider the original model (2.1). The null hypothesis of interest is $H_{20} : E(\alpha_i|\mathscr{R}_i) = 0$ almost everywhere, where $\mathscr{R}_i$ denotes the set of all the regressors for unit i, and $1 \leq t \leq T$, i.e., $\mathscr{R}_i = (R_{i1}^\top, ..., R_{iT}^\top)^\top$ with $R_{it} = (u_{it}, X_{it}^\top, Z_{it}^\top)^\top$. The alternative hypothesis is $H_{21} : E(\alpha_i|\mathscr{R}_i) \neq 0$ on a set with positive measure. Let $v_{it} = \alpha_i + \epsilon_{it}$. Throughout this paper, we assume that ϵ_{it} is mean independent of $\mathscr{R}_i$ under either H_{20} or H_{21}. Therefore, the null and alternative hypotheses are equivalent to

$$H_{20} : E(v_{it}|\mathscr{R}_i) = 0 \quad \text{almost everywhere,}$$

and

$$H_{21} : E(v_{it}|\mathscr{R}_i) \neq 0 \quad \text{on a set with positive measure.}$$

Let $f(R_{it})$ be the joint probability density function of R_{it}. Following Zheng (1996) and Henderson et al. (2008), the proposed test statistic is based on $E\{v_{it}E(v_{it}|R_{it})f(R_{it})\}$ with the sample analogue given by

$$\begin{aligned}\tilde{H}_{NT} &= (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{v}_{it}\hat{E}_{-it}(\hat{v}_{it}|R_{it})\hat{f}_{-it}(R_{it}) \\ &= (NT(NT-1))^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{s=1,\,\{j,s\}\neq\{i,t\}}^{T}\hat{v}_{it}\hat{v}_{js}K^*_{h,\,it,js},\end{aligned} \tag{3.5}$$

where $\hat{v}_{it} = Y_{it} - X_{it}^\top\hat{\beta}(u_{it}) - Z_{it}^\top\hat{\gamma}$ with $\hat{\beta}(u_{it})$ being the local linear estimator, as in Section 2, $K^*_{h,\,it,js} = K^*_h(R_{it} - R_{js}), K^*_h(w) = h^{-m}\prod_{l=1}^{m} K(w_l/h)$ with $m = 1 + p + q$,[3] and

$$\hat{E}_{-it}(\hat{v}_{it}|R_{it}) = (N(T-1))^{-1}\sum_{j=1}^{N}\sum_{s=1,\,\{j,s\}\neq\{i,t\}}^{T}\hat{v}_{js}K^*_{h,\,it,js}/\hat{f}_{-it}(R_{it})$$

and

$$\hat{f}_{-it}(R_{it}) = (N(T-1))^{-1}\sum_{j=1}^{N}\sum_{s=1,\,\{j,s\}\neq\{i,t\}}^{T}K^*_{h,\,it,js}$$

are the leave-one-out estimators of $E(v_{it}|R_{it})$ and $f(R_{it})$, respectively.

4. Asymptotic theories

In this section, we derive the asymptotic properties of the proposed estimators and test statistics with their detailed proofs relegated to the appendices. To establish the asymptotic results, the following assumptions are needed, although they might not be the weakest ones.

4.1. Assumptions

A1. *The observed data* $\{(Y_{it}, X_{it}, Z_{it}, u_{it}) : 1 \leq i \leq N, 1 \leq t \leq T\}$ *are independently and identically distributed.*
A2. *The kernel* $K(\cdot)$ *is a symmetric and bounded probability density function with compact support. Furthermore, it satisfies the Lipschitz condition.*

[3]For simplicity, we use the same bandwidth h for different covariates.

A3. *The density function of* u_{it} *is Lipschitz continuous and bounded away from zero on its support* $\boldsymbol{\Omega}$.
A4. *The* $p\times p$ *matrix* $E(\Delta^* x_{it}\Delta^* x_{it}^\top|u)$ *is nonsingular for each* $u\in\boldsymbol{\Omega}$. $E(\Delta^* x_{it}\Delta^* x_{it}^\top|u)$ *and* $E(\Delta^* x_{it}\Delta^* z_{it}^\top|u)$ *are all Lipschitz continuous.*
A5. $h\to 0$ *and* $NTh\to\infty$ *as* $(N,T)\to\infty$, *and* $\left(\sqrt{NT}\log T\right)/T\to 0$.
A6. $E|X_{it}|^4<\infty, E|Z_{it}|^4<\infty$.

4.2. Asymptotic properties

In order to present the main results, we first introduce some notation. Let $c_{NT}=\left\{\log(1/h)/NTh\right\}^{1/2}+h^2$, and define

$$\Phi(u)=f(u)\underset{(N,T)\to\infty}{\text{plim}}\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T E(\Delta^* x_{it}\Delta^* x_{it}^\top|u_{it}=u),$$

$$\Omega^*(u)=f(u)\nu_0\underset{(N,T)\to\infty}{\text{plim}}\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T E\left\{\Delta^* x_{it}\Delta^* x_{it}^\top E\left[\Delta^*\epsilon_{it}^2|u_{it},\Delta^* x_{it}\right]|u_{it}=u\right\},$$

where $\nu_0=\int K^2(u)du$.

Now, we state the asymptotic properties of the proposed estimators and test statistics in the following theorems.

Theorem 4.1. *Suppose that Assumptions A1–A6 hold. Then we have*

$$\sqrt{NT}(\hat\gamma-\gamma)\xrightarrow{d}N(\mathbf{0},\Sigma^{-1}\Sigma^*\Sigma^{-1})\quad\text{as }(N,T)\to\infty,$$

where $\Sigma=\text{plim}_{(N,T)\to\infty}\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\left\{E(\Delta^* z_{it}\Delta^* z_{it}^\top)-E\left[f^{-1}(u_{it})\Psi^\top(u_{it})\Phi^{-1}(u_{it})\Psi(u_{it})\right]\right\}$, *and* $\Sigma^*=\text{plim}_{(N,T)\to\infty}\frac{1}{NT}E(Z^\top(I-\hat S)^\top\epsilon\epsilon^\top(I-\hat S)Z)$.

Theorem 4.2. *Suppose that Assumptions A1–A6 hold. Then the following result holds:*

$$\sqrt{NTh}\left[\hat\beta(u)-\beta(u)-\frac{h^2}{2}\mu_2\beta''(u)\left\{1+O_p(c_{NT})\right\}\right]\xrightarrow{d}N(\mathbf{0},\boldsymbol{\Theta}(u))$$

as $(N,T)\to\infty$, *where* $\boldsymbol{\Theta}(u)=\Phi^{-1}(u)\Omega^*(u)\Phi^{-1}(u)$ *and* $\mu_2=\int u^2K(u)du$.

Theorem 4.3. *Under Assumptions A1–A6 and* H_{10}, *we have*

$$\mathbb{W}_{NT}=NTh^{1/2}\tilde W_{NT}/\sqrt{\hat V_W}\xrightarrow{d}N(0,1),\tag{4.1}$$

where

$$\hat V_W=\frac{2}{N^2T^2h}\sum_{i\neq j}^N\sum_{t=1}^T\sum_{s=1}^T\Delta^*\hat\epsilon_{it}^2\Delta^*\hat\epsilon_{js}^2(\Delta^* x_{it}^\top\Delta^* x_{js})^2K^2\left(\frac{u_{it}-u_{js}}{h}\right),$$

with $\Delta^*\hat\epsilon_{it}$ *defined as* $\Delta^* y_{it}-\Delta^* z_{it}^\top\hat\gamma-\Delta^* x_{it}^\top\hat\beta_0(u_{it})$.

Theorem 4.4. *Under Assumptions A1–A6 and* H_{20}, *we have*

$$\mathbb{H}_{NT}=NTh^{m/2}\tilde H_{NT}/\sqrt{\hat V_H}\xrightarrow{d}N(0,1)\tag{4.2}$$

as $h\to 0$ *and* $NTh^m\to\infty$, *where*

$$\hat{V}_H = \frac{2}{N^2T^2h^m}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{s=1,\,\{j,s\}\neq\{i,t\}}^{T}\hat{v}_{it}^2\hat{v}_{js}^2K_{h,it,js}^{*2}.$$

Remark 1. Theorems 4.1 and 4.2 indicate that both estimators $\hat{\gamma}$ and $\hat{\beta}(u)$ are asymptotically normal. Notice that $\hat{\gamma}$ achieves the parametric convergence rate $\sqrt{NT}$, while the convergence rate of $\hat{\beta}(u)$ is $\sqrt{NTh}$, which is standard for nonparametric estimators using kernel methods. Theorems 4.3 and 4.4 show that, under H_{10} and H_{20}, the standardized statistics $\mathbb{W}_{NT}$ and $\mathbb{H}_{NT}$ approach the asymptotic standard normal distribution.

Theorem 4.5. *Given Assumptions A1–A6,*

(a) *under H_{11}, we have*

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

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

(b) *if $h \to 0$ and $NTh^m \to \infty$, then under H_{21},*

$$\Pr\left\{\mathbb{H}_{NT} \geq M_{NT}^*\right\} \to 1 \quad as \quad (N,T) \to \infty,$$

where M_{NT}^* is any nonstochastic, positive sequence such that $M_{NT}^* = o(NTh^{m/2})$.

Remark 2. Theorem 4.5 (a) shows that, under the alternative H_{11}, the probability that the proposed functional form test rejects the null hypothesis approaches 1 as both N and T go to infinity simultaneously, which implies that the proposed functional form test is consistent. Similarly, Theorem 4.5 (b) shows that the proposed nonparametric Hausman test is consistent.

5. Monte Carlo simulations

In this section, we conduct Monte Carlo simulations to evaluate the finite sample performances of the proposed estimators and test statistics. Section 5.1 investigates the finite sample properties of estimators defined by (2.10) and (2.11), and Sections 5.2 and 5.3 assess the size and power performance of the functional form test and nonparametric Hausman-type test, respectively.

5.1. Finite sample performance of estimators

We consider the following data generating process (DGP):

$$Y_{it} = \beta(u_{it})X_{it} + \gamma Z_{it} + \alpha_i + \epsilon_{it}, \quad 1 \leq i \leq N, \quad 1 \leq t \leq T, \tag{5.1}$$

where $\beta(u_{it}) = \sin(2u_{it})$, $\gamma = 1$, the smooth variable u_{it} is generated independently from a uniform distribution $U(0, 1)$, Z_{it} and X_{it} are generated independently from normal distribution $N(0, 2^2)$ and $N(1, 3^2)$, respectively, the individual effect α_i is generated as $\alpha_i = \kappa\frac{1}{T}\sum_{t=1}^{T}X_{it} + \eta_i, i = 1, ..., N$ with $\kappa = 1/2$, and η_i is generated independently from normal distributions $N(0, 0.1^2)$. Finally, the idiosyncratic error ϵ_{it} is generated from a standard normal distribution.

The number of Monte Carlo replications is 1000 for each of (N, T) pairs with $N, T = 20$, 30, 40. The Epanechnikov kernel $K(u) = \frac{3}{4}(1 - u^2)1\{|u| \leq 1\}$ is employed to compute the semiparametric estimators given in (2.10) and (2.11). When estimating γ using (2.10), a relatively small bandwidth $h_1 = (NT)^{-2/5}$ is used to reduce the influence of the estimation bias. The bandwidth $h_2 = c_0\hat{\sigma}_u n^{-1/5}$ is used for the estimation of $\beta(\cdot)$ with $c_0 = 2.0,\ 2.5,\ 3.0$, where $\hat{\sigma}_u$ is the sample

Table 1. Means and SDs of ADEs of estimator for $\gamma = 1$.

	$c_0 = 2.0$			$c_0 = 2.5$			$c_0 = 3.0$		
$N\backslash T$	20	30	40	20	30	40	20	30	40
20	0.0275 (0.0207)	0.0233 (0.0171)	0.0195 (0.0147)	0.0275 (0.0212)	0.0240 (0.0179)	0.0195 (0.0147)	0.0281 (0.0216)	0.0239 (0.0179)	0.0199 (0.0151)
30	0.0233 (0.0174)	0.0182 (0.0135)	0.0157 (0.0119)	0.0220 (0.0169)	0.0185 (0.0137)	0.0159 (0.0122)	0.0223 (0.0172)	0.0182 (0.0139)	0.0167 (0.0119)
40	0.0207 (0.0157)	0.0153 (0.0115)	0.0144 (0.0110)	0.0195 (0.0150)	0.0156 (0.0118)	0.0140 (0.0099)	0.0194 (0.0155)	0.0154 (0.0122)	0.0142 (0.0110)

standard deviation of all the remaining smooth variables after data deletion,[4] and n is the number of all remaining observations; i.e., $n = \sum_{i=1}^{N} T_i$.

To evaluate the estimation accuracy of the proposed parametric estimator of γ, the absolute deviation error (ADE) is used, and is defined by

$$\text{ADE} = |\hat{\gamma} - \gamma|.$$

To evaluate the performance of the proposed nonparametric estimator of $\beta(\cdot)$, the root mean squared error (RMSE) is used, given by

$$\text{RMSE} = \sqrt{\frac{1}{D}\sum_{k=1}^{D}\left[\hat{\beta}(u_k) - \beta(u_k)\right]^2},$$

where $\{u_k : k = 1, ..., D\}$ are grid points. Table 1 reports the means and standard deviations (SDs) (in parentheses) of the ADEs from 1000 replications for the estimation of $\gamma = 1$. One can find that the estimation accuracy in terms of ADE improves as either N or T increases. One can also see that the simulation results are largely unaffected by the different values taken by c_0, which is as expected, since the choice of h_2 has little or no impact on the estimation of γ. Table 2 reports the means and SDs (in parentheses) of the RMSEs from 1000 replications for the estimation of $\beta(u) = \sin(2u)$. This table shows that RMSEs decrease as either N or T increases. Moreover, it seems that the RMSEs of the estimator for $\beta(u)$ perform best for $c_0 = 2.5$ in h_2.

5.2. Finite sample performance of functional form tests

Next, we study the finite sample performance of the functional form test. We consider the null hypothesis $H_0 : \beta(u_{it}) = \beta_0$ versus the alternative $H_1 : \beta(u_{it}) \neq \beta_0$. The power is evaluated under a series of alternative models indexed by δ, such that $H_1 : \beta(u_{it}) = \beta_0 + \delta(\beta^*(u_{it}) - \beta_0)$ for $0 \leq \delta < 1$, where $\beta^*(u_{it}) = \sin(2u_{it})$. We choose β_0 to be the average height of $\sin(2u_{it})$ with u_{it} generated from $U(0, 1)$ ($\beta_0 = 0.71$). The following DGP is considered in the evaluation of the size performance of the proposed test statistic:

$$Y_{it} = \beta_0 X_{it} + \gamma Z_{it} + \alpha_i + \epsilon_{it}, \quad 1 \leq i \leq N, 1 \leq t \leq T,$$

where X_{it}, Z_{it}, α_i, and ϵ_{it} are generated in the same way as in (5.1). To investigate the size of the test, the simulation is conducted with (N, T) pairs, where N, T = 30, 40, and 50. The kernel function and bandwidth are chosen to be the Epanechnikov kernel and $h = c_0\hat{\sigma}_u n^{-1/5}$, respectively. To investigate whether the size performance is sensitive to the choice of bandwidth, c_0 is chosen to be 0.5, 1.0, 2.0. The critical values for the test are taken from the standard normal distribution. Table 3 reports the empirical rejection frequencies under H_0 based on 2000 replications of our proposed functional form test at 1%, 5%, and 10% nominal levels. As seen in Table 3, the sizes are close to the nominal sizes for most (N, T) pairs. In addition, the size performance seems not to be sensitive to the choice of bandwidth.

[4]Recall that data deletion is required to obtain a nonsingular variance–covariance matrix of errors in (2.3).

Table 2. Means and SDs of RMSEs of estimator for $\beta(u) = \sin(2u)$.

	$c_0 = 2.0$			$c_0 = 2.5$			$c_0 = 3.0$		
$N\backslash T$	20	30	40	20	30	40	20	30	40
20	0.0403 (0.0155)	0.0348 (0.0126)	0.0300 (0.0109)	0.0382 (0.0153)	0.0322 (0.0132)	0.0286 (0.0113)	0.0388 (0.0164)	0.0332 (0.0138)	0.0299 (0.0127)
30	0.0341 (0.0128)	0.0297 (0.0106)	0.0263 (0.0090)	0.0329 (0.0138)	0.0274 (0.0105)	0.0245 (0.0091)	0.0334 (0.0139)	0.0278 (0.0116)	0.0255 (0.0107)
40	0.0299 (0.0106)	0.0260 (0.0088)	0.0230 (0.0078)	0.0290 (0.0115)	0.0244 (0.0086)	0.0220 (0.0076)	0.0299 (0.0110)	0.0248 (0.0090)	0.0223 (0.0078)

Table 3. Size performance of functional form test.

		$c_0 = 0.5$			$c_0 = 1$			$c_0 = 2$		
N	T	1%	5%	10%	1%	5%	10%	1%	5%	10%
30	30	0.0155	0.0500	0.0940	0.0125	0.0450	0.0880	0.0220	0.0565	0.0945
	40	0.0130	0.0520	0.0810	0.0190	0.0575	0.0935	0.0175	0.0550	0.0870
	50	0.0130	0.0610	0.0965	0.0240	0.0610	0.1000	0.0235	0.0555	0.0900
40	30	0.0145	0.0520	0.0880	0.0190	0.0540	0.0975	0.0175	0.0515	0.0815
	40	0.0130	0.0435	0.0810	0.0190	0.0550	0.0920	0.0150	0.0600	0.1110
	50	0.0175	0.0520	0.0980	0.0175	0.0510	0.0940	0.0190	0.0570	0.0925
50	30	0.0155	0.0485	0.0945	0.0185	0.0545	0.0950	0.0175	0.0465	0.0815
	40	0.0105	0.0500	0.0935	0.0150	0.0515	0.0910	0.0230	0.0570	0.0980
	50	0.0150	0.0560	0.1000	0.0210	0.0580	0.0970	0.0195	0.0545	0.0905

We now turn to the evaluation of the power performance. The empirical rejection frequencies are computed under H_1 based on 2000 replications. To better visualize the change of powers as the sample sizes and δ grow, the power functions are plotted against δ for $c_0 = 0.5,\ 1.0$, and 2.0 in Figures 1 and 2 at the 5% nominal level. Figure 1 presents the power curves when T is fixed at 30 and $N = 30$, 40, and 50, while Figure 2 is for cases when N is fixed at 30 and $T = 30$, 40, and 50. As shown in these figures, for fixed δ, the power of our test increases as either N or T increases before it reaches 1. When $\delta = 0$, the specified alternative hypothesis reduces to the null hypothesis. Therefore, the power curves approach 0.05 (the horizontal dotted lines), the nominal significance level. However, the power rapidly tends toward 1 as δ increases. These power functions show that our proposed test indeed has good power performance.

5.3. Finite sample performance of nonparametric Hausman-type tests

Finally, we evaluate the finite sample performance of the nonparametric Hausman-type test for differentiating between the fixed effects model and random effects model. The DGP is the same as in Section 5.1 except that the individual effect α_i is generated with $\kappa = 0$, 0.1, 0.2, and 0.3. Note that κ characterizes the extent that (5.1) deviates from random effects specification. In particular, when $\kappa = 0$, (5.1) collapses to a random effects model. The power behavior of our nonparametric Hausman-type test is assessed by letting κ gradually deviate from zero. To compute $\tilde{H}_{NT}$ in (3.5), we employ the univariate Gaussian kernel function for each variable in $R_{it} = (u_{it}, X_{it}, Z_{it})^\top$. The bandwidth for u_{it} is chosen to be $h_u = \hat{\sigma}_u n^{-1/5}$, where the definitions of $\hat{\sigma}_u$ and n are the same as in Section 5.1. The bandwidths for X_{it} and Z_{it} are defined similarly. Tables 4 and 5 report the simulation results based on 2000 Monte Carlo replications. The critical values for the test are still taken from the standard normal distribution. From these tables, one can observe that the estimated sizes of our nonparametric Hausman-type test are close to the asymptotic sizes when $\kappa = 0$. As κ gradually departs from zero, the power of the test quickly converges to 1 for any (N, T) combination. Furthermore, for fixed $\kappa > 0$, the power increases rapidly as either N or T increases. These results show that the proposed nonparametric Hausman-type test has a good finite sample size and power performance.

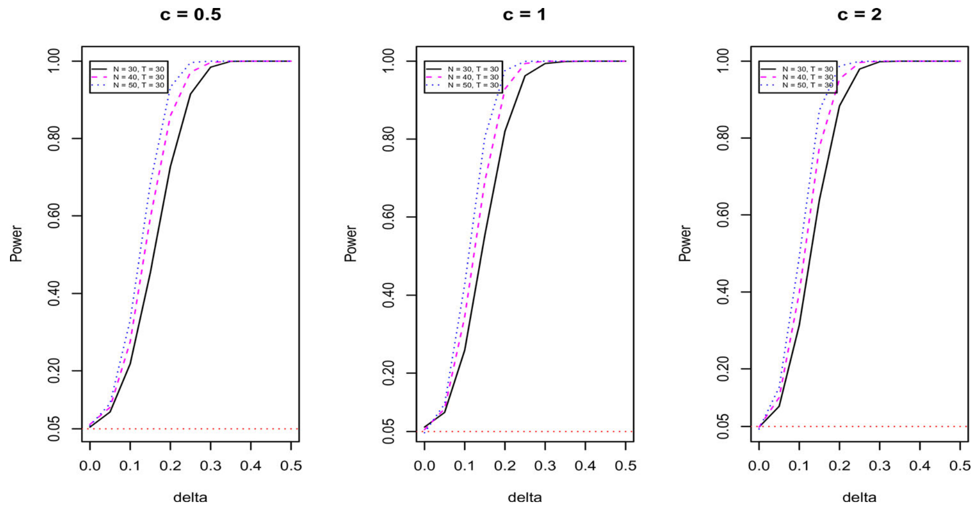

Figure 1. Plot of power curves against δ for fixed T.

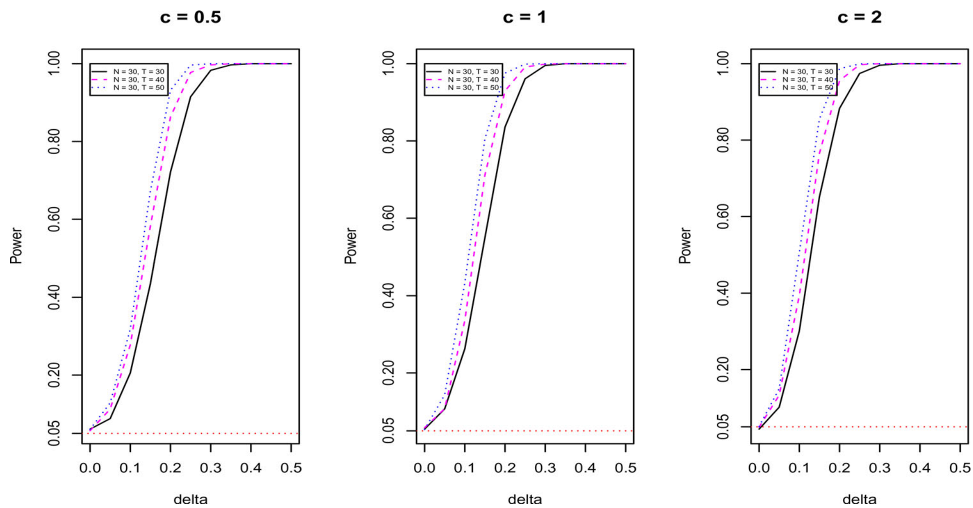

Figure 2. Plots of power curves against δ for fixed N.

Table 4. Size and power performance of nonparametric Hausman-type test ($T = 40$).

	$N = 30$			$N = 40$			$N = 50$		
κ	1%	5%	10%	1%	5%	10%	1%	5%	10%
0	0.0205	0.0675	0.1125	0.0165	0.0605	0.1135	0.0160	0.0635	0.1115
0.1	0.2700	0.4180	0.5210	0.3420	0.5125	0.6090	0.4305	0.5820	0.6785
0.2	0.9165	0.9605	0.9750	0.9750	0.9880	0.9925	0.9940	0.9970	0.9990
0.3	0.9995	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

6. Conclusion

In this paper, we propose a novel way to estimate a partially varying-coefficient panel data model with fixed effects which removes the individual effects by taking the difference of the nearest neighbor of the smoothing variables. The profile least squares method based on first-stage local linear fitting is developed to estimate both the parametric and nonparametric parts of the difference-transformed model. Moreover, a functional form specification test and a nonparametric

Table 5. Size and power performance of nonparametric Hausman-type test ($N = 40$).

	$T = 30$			$T = 40$			$T = 50$		
κ	1%	5%	10%	1%	5%	10%	1%	5%	10%
0	0.0165	0.0640	0.1135	0.0160	0.0580	0.1170	0.0200	0.0630	0.1135
0.1	0.2785	0.4300	0.5210	0.3530	0.5059	0.6020	0.4210	0.5725	0.6785
0.2	0.9170	0.9665	0.9785	0.9665	0.9875	0.9940	0.9885	0.9970	0.9980
0.3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Hausman-type test are constructed. The asymptotic properties of the proposed estimator and test statistics are derived as well. Monte Carlo simulations are conducted to illustrate good finite sample performance for the proposed estimator and the test statistics.

There are some limitations in the current study. This paper assumes that the observed data is independently and identically distributed. However, this assumption may be relaxed to stationary time series on the time dimension and cross-sectional dependence on the cross-sectional dimension. We leave it as a future research topic.

Appendix

The following notations and definitions will be used in the proof of lemmas and theorems. Define $\mu_k = \int u^k K(u)du, \nu_k = \int u^k K^2(u)du$, and

$$\Psi(u) = f(u) \operatorname*{plim}_{(N,T)\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} E(\Delta^* x_{it} \Delta^* z_{it}^\top | u_{it} = u).$$

Further let $\tilde{Z} = (I - \hat{S})Z$ and $\tilde{Z}_{it}$ denote the $[(i-1)T + t]$th column of $\tilde{Z}^\top$.

Appendix A: Useful lemmas

Lemma A1. *Let $\{W_{it}, u_{it}\}$ be independently and identically distributed bivariate random vectors, Δ^* denote the nearest neighbor difference defined in Section 2, and $K(\cdot)$ be a bounded positive function with a bounded support, satisfying a Lipschitz condition. Further assume that $E|W|^s < \infty$ and $\sup_u \int |W|^s f(W,u)dW < \infty$, where f denotes the joint density of (W, u). Given that $(NT)^{2\delta-1}h \to \infty$ for some $\delta < 1 - s^{-1}$, then*

$$\sup_u \left| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} [\Delta^* W_{it} K_h(u_{it} - u) - E(\Delta^* W_{it} K_h(u_{it} - u))] \right| = O_p\left(\left\{ \frac{\log(1/h)}{NTh} \right\}^{1/2} \right).$$

Proof of Lemma A1. Note that $\Delta^* W_{it} = W_{it} - W_{it^*}$ where W_{it^*} is the nearest neighbor of W_{it}. Therefore, we have

$$\begin{aligned} &\sup_u \left| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} [\Delta^* W_{it} K_h(u_{it} - u) - E(\Delta^* W_{it} K_h(u_{it} - u))] \right| \\ &\le \sup_u \left| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} [W_{it} K_h(u_{it} - u) - E(W_{it} K_h(u_{it} - u))] \right| \\ &\quad + \sup_u \left| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} [W_{it^*} K_h(u_{it} - u) - E(W_{it^*} K_h(u_{it} - u))] \right| \\ &\equiv \mathbb{I}_1 + \mathbb{I}_2. \end{aligned}$$

By Proposition 4 in Mack and Silverman (1982), it can be shown that

$$\mathbb{I}_1 = O_p\left(\left\{ \frac{\log(1/h)}{NTh} \right\}^{1/2} \right), \quad \mathbb{I}_2 = O_p\left(\left\{ \frac{\log(1/h)}{NTh} \right\}^{1/2} \right).$$

This completes the proof of the lemma.

Lemma A2. *Suppose that Assumptions A1–A5 hold, we have*

$$\frac{1}{NT} Z^\top (I-\hat{S})^\top (I-\hat{S}) Z \underset{p}{\to} \Sigma,$$

where $\Sigma = \text{plim}_{(N,T)\to\infty} \frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T \left\{E(\Delta^* z_{it} \Delta^* z_{it}^\top) - E\left[f^{-1}(u_{it})\Psi^\top(u_{it})\Phi^{-1}(u_{it})\Psi(u_{it})\right]\right\}$.

Proof of Lemma A2. Firstly, we observe that

$$\hat{S}Z = \begin{pmatrix} (\Delta^* x_{11}^\top, \mathbf{0})\left\{D_{u_{11}}^\top H_{u_{11}} D_{u_{11}}\right\}^{-1} D_{u_{11}}^\top H_{u_{11}} Z \\ \vdots \\ (\Delta^* x_{NT}^\top, \mathbf{0})\left\{D_{u_{NT}}^\top H_{u_{NT}} D_{u_{NT}}\right\}^{-1} D_{u_{NT}}^\top H_{u_{NT}} Z \end{pmatrix}.$$

Also note that $D_u^\top H_u D_u$ is given by

$$\begin{pmatrix} \sum_{i=1}^N \sum_{t=1}^T \Delta^* x_{it} \Delta^* x_{it}^\top K_h(u_{it}-u) & \sum_{i=1}^N \sum_{t=1}^T \Delta^* x_{it} \Delta^* x_{it}^\top \frac{u_{it}-u}{h} K_h(u_{it}-u) \\ \sum_{i=1}^N \sum_{t=1}^T \Delta^* x_{it} \Delta^* x_{it}^\top \frac{u_{it}-u}{h} K_h(u_{it}-u) & \sum_{i=1}^N \sum_{t=1}^T \Delta^* x_{it} \Delta^* x_{it}^\top \left(\frac{u_{it}-u}{h}\right)^2 K_h(u_{it}-u) \end{pmatrix}.$$

By Lemma A1, for the upper-left entry in the above matrix, it is easy to show that

$$\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T \Delta^* x_{it} \Delta^* x_{it}^\top K_h(u_{it}-u) \\
&= \frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T E(\Delta^* x_{it} \Delta^* x_{it}^\top K_h(u_{it}-u))\left\{1 + O_p\left(\left\{\frac{\log(1/h)}{NTh}\right\}^{1/2}\right)\right\} \\
&= \frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T \int \Delta^* x_{it} \Delta^* x_{it}^\top K_h(u_{it}-u) f(\Delta^* x_{it}|u_{it}) f(u_{it}) d\Delta^* x_{it} du_{it} \\
&\qquad \times \left\{1 + O_p\left(\left\{\frac{\log(1/h)}{NTh}\right\}^{1/2}\right)\right\} \\
&= f(u)\left\{\frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T E(\Delta^* x_{it} \Delta^* x_{it}^\top | u_{it} = u)\right\}\{1 + O_p(c_{NT})\} \\
&= \Phi(u)\{1 + O_p(c_{NT})\}.
\end{aligned}$$

By employing the same arguments, we can derive

$$\frac{1}{NT} D_u^\top H_u D_u = \begin{pmatrix} \Phi(u) & \mathbf{0} \\ \mathbf{0} & \Phi(u)\mu_2 \end{pmatrix}\{1 + O_p(c_{NT})\}, \tag{A.1}$$

and, therefore,

$$\left(\frac{1}{NT} D_u^\top H_u D_u\right)^{-1} = \begin{pmatrix} \Phi^{-1}(u) & \mathbf{0} \\ \mathbf{0} & \Phi^{-1}(u)\mu_2^{-1} \end{pmatrix}\{1 + O_p(c_{NT})\}.$$

Similarly, a simple manipulation leads to

$$\frac{1}{NT} D_u^\top H_u Z = \begin{pmatrix} \Psi(u) \\ \mathbf{0} \end{pmatrix}\{1 + O_p(c_{NT})\}. \tag{A.2}$$

Combining the previous results gives

$$\hat{S}Z = \begin{pmatrix} \Delta^* x_{11}^\top \Phi^{-1}(u_{11})\Psi(u_{11}) \\ \vdots \\ \Delta^* x_{NT}^\top \Phi^{-1}(u_{NT})\Psi(u_{NT}) \end{pmatrix}\{1 + O_p(c_{NT})\}. \tag{A.3}$$

Finally, we observe that

$$
\begin{aligned}
&\frac{1}{NT}Z^\top(I-\hat{S})^\top(I-\hat{S})Z\\
&=\frac{1}{NT}(Z-\hat{S}Z)^\top(Z-\hat{S}Z)\\
&=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\Delta^* z_{it}-\Psi^\top(u_{it})\Phi^{-1}(u_{it})\Delta^* x_{it}\right]\left[\Delta^* z_{it}-\Psi^\top(u_{it})\Phi^{-1}(u_{it})\Delta^* x_{it}\right]^\top\left\{1+O_p(c_{NT})\right\}\\
&\equiv \mathbb{I}_3\left\{1+O_p(c_{NT})\right\}.
\end{aligned}
$$

Note that the nearest neighbor difference causes some dependent structure over t. However, for a given $i\in N$, there are at most two elements among $\left\{(\Delta^* z_{is},\Delta^* x_{is}):s\neq t\right\}$ that are correlated to $(\Delta^* z_{it},\Delta^* x_{it})$. Therefore, it is easy to show that $Var(\mathbb{I}_3)\to 0$. Using Chebyshev inequality, we have

$$
\frac{1}{NT}Z^\top(I-\hat{S})^\top(I-\hat{S})Z\xrightarrow[p]{}\Sigma.
$$

The lemma is established.

Lemma A3. *Under Assumptions A1–A5, we have*

$$
\frac{1}{NT}Z^\top(I-\hat{S})^\top(I-\hat{S})M=O_p(c_{NT}^2).
$$

Proof of Lemma A3. Similarly to (A.3), we can show that

$$
\hat{S}M=\begin{pmatrix}\Delta^* x_{11}^\top\beta(u_{11})\\ \vdots\\ \Delta^* x_{NT}^\top\beta(u_{NT})\end{pmatrix}\left\{1+O_p(c_{NT})\right\}.
$$

Then, simple calculation gives

$$
\begin{aligned}
&\frac{1}{NT}Z^\top(I-\hat{S})^\top(I-\hat{S})M\\
&=\frac{1}{NT}(Z-\hat{S}Z)^\top(M-\hat{S}M)\\
&=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\Delta^* z_{it}-\Psi^\top(u_{it})\Phi^{-1}(u_{it})\Delta^* x_{it}\right]\Delta^* x_{it}^\top\beta(u_{it})\left\{1+O_p(c_{NT})\right\}O_p(c_{NT})\\
&=O_p(c_{NT}^2).
\end{aligned}
$$

This completes the proof of the lemma.

Appendix B:Proofs of theorems

Proof of Theorem 4.1. From Equation (2.10) and Lemma A2, we can derive that

$$
\begin{aligned}
\sqrt{NT}(\hat{\gamma}-\gamma)&=\left\{\frac{1}{NT}Z^\top(I-\hat{S})^\top(I-\hat{S})Z\right\}^{-1}\left\{\frac{1}{\sqrt{NT}}Z^\top(I-\hat{S})^\top(I-\hat{S})(M+P_T+\epsilon)\right\}\\
&=\Sigma^{-1}\left\{\frac{1}{\sqrt{NT}}Z^\top(I-\hat{S})^\top(I-\hat{S})(M+P_T+\epsilon)\right\}\left\{1+o_p(1)\right\}.
\end{aligned}
$$

Next, we consider

$$
\begin{aligned}
&\frac{1}{\sqrt{NT}}Z^\top(I-\hat{S})^\top(I-\hat{S})(M+P_T+\epsilon)\\
&=\frac{1}{\sqrt{NT}}Z^\top(I-\hat{S})^\top(I-\hat{S})M+\frac{1}{\sqrt{NT}}Z^\top(I-\hat{S})^\top(I-\hat{S})P_T\\
&\quad+\frac{1}{\sqrt{NT}}Z^\top(I-\hat{S})^\top(I-\hat{S})\epsilon\\
&\equiv\mathbb{I}_4+\mathbb{I}_5+\mathbb{I}_6.
\end{aligned}
$$

From Lemma A3, we have $\mathbb{I}_4 = O_p\left(\sqrt{NT}c_{NT}^2\right)$. According to the definition of P_T in (2.5), we have

$$\begin{aligned}\mathbb{I}_5 &= \frac{1}{\sqrt{NT}} Z^\top (I-\hat{S})^\top (I-\hat{S}) P_T \\ &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T [\Delta^* z_{it} - \Psi^\top(u_{it})\Phi^{-1}(u_{it})\Delta^* x_{it}] \left\{ X_{it^*}^\top - [\Delta^* x_{it}^\top \Phi^{-1}(u_{it}), \mathbf{0}] \frac{D_{u_{it}}^\top H_{u_{it}} X_{it^*}^\top}{NT} \right\} \\ &\quad \times \{1 + O_p(c_{NT})\} O_p(\log T/T) \\ &= O_p\left(\frac{\sqrt{NT}\log T}{T}\right)\{1 + O_p(c_{NT})\}.\end{aligned}$$

Finally, we turn to the analysis of $\mathbb{I}_6$. Note that

$$\mathbb{I}_6 = \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \tilde{Z}_{it} \left\{ \Delta^* \epsilon_{it} - (\Delta^* x_{it}^\top, \mathbf{0}) \left\{ D_{u_{it}}^\top H_{u_{it}} D_{u_{it}} \right\}^{-1} D_{u_{it}}^\top H_{u_{it}} \epsilon \right\}.$$

By using the same argument as in the proof of Lemma A2, we have

$$(\Delta^* x_{it}^\top, \mathbf{0}) \left\{ D_{u_{it}}^\top H_{u_{it}} D_{u_{it}} \right\}^{-1} D_{u_{it}}^\top H_{u_{it}} \epsilon = \Delta^* x_{it}^\top \Phi^{-1}(u_{it}) f(u_{it}) E(\Delta^* x_{it} | u_{it}) O_p(c_{NT}).$$

Then we can show that

$$\begin{aligned}\mathbb{I}_6 &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \{\Delta^* z_{it} - \Psi^\top(u_{it})\Phi^{-1}(u_{it})\Delta^* x_{it}\} \Delta^* \epsilon_{it} \{1 + o_p(1)\} \\ &\equiv \frac{1}{\sqrt{N}} \sum_{i=1}^N Q_i \{1 + o_p(1)\},\end{aligned}$$

where $Q_i = (1/\sqrt{T}) \sum_{t=1}^T \{\Delta^* z_{it} - \Psi^\top(u_{it})\Phi^{-1}(u_{it})\Delta^* x_{it}\}\Delta^* \epsilon_{it}$. We now show that $\frac{1}{\sqrt{N}}\sum_{i=1}^N Q_i$ is normally distributed by employing the Cramér–Wold theorem and the Lindeberg–Feller central limit theorem.

For any unit vector $\boldsymbol{d} \in \mathbb{R}^q$, let $\omega_i = \boldsymbol{d}^\top Q_i$. Then,

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \boldsymbol{d}^\top Q_i = \frac{1}{\sqrt{N}} \sum_{i=1}^N \omega_i.$$

Note that, for a given $i \in N$, there are at most two elements among $\{(\Delta^* z_{is}, \Delta^* x_{is}) : s \neq t\}$ that are correlated to $(\Delta^* z_{it}, \Delta^* x_{it})$. By Assumption 6, it is easy to show that

$$\frac{1}{\rho_0^4 N^2} \sum_{i=1}^N E|\omega_i|^4 \to 0,$$

as $(N,T) \to \infty$, where $\rho_0^2 = \text{plim}_{N\to\infty}(1/N)\sum_{i=1}^N \boldsymbol{d}^\top E(Q_i Q_i^\top)\boldsymbol{d}$. This completes the proof of the theorem.

Proof of Theorem 4.2. From (2.11), we can rewrite the local linear estimator of $\beta(u)$ as follows:

$$\begin{aligned}\hat{\beta}(u) &= [I_p, \mathbf{0}_p]\{D_u^\top H_u D_u\}^{-1} D_u^\top H_u (Y - Z\hat{\gamma}) \\ &= [I_p, \mathbf{0}_p]\{D_u^\top H_u D_u\}^{-1} D_u^\top H_u [M + Z(\gamma - \hat{\gamma}) + P_T + \epsilon],\end{aligned}$$

where I_p is a $p \times p$ identity matrix, and $\mathbf{0}_p$ is a $p \times p$ zero matrix. Therefore, it is simple to show that

$$\hat{\beta}(u) - \beta(u) = \mathbb{I}_7 + \mathbb{I}_8 + \mathbb{I}_9 + \mathbb{I}_{10},$$

where

$$\begin{aligned}\mathbb{I}_7 &= [I_p, \mathbf{0}_p]\{D_u^\top H_u D_u\}^{-1} D_u^\top H_u \left[M - D_u \begin{pmatrix} \beta(u) \\ h\beta'(u) \end{pmatrix} \right], \\ \mathbb{I}_8 &= [I_p, \mathbf{0}_p]\{D_u^\top H_u D_u\}^{-1} D_u^\top H_u Z(\gamma - \hat{\gamma}), \\ \mathbb{I}_9 &= [I_p, \mathbf{0}_p]\{D_u^\top H_u D_u\}^{-1} D_u^\top H_u P_T, \\ \mathbb{I}_{10} &= [I_p, \mathbf{0}_p]\{D_u^\top H_u D_u\}^{-1} D_u^\top H_u \epsilon.\end{aligned}$$

We first consider $\mathbb{I}_7$. Note that

$$\mathbb{I}_7 = [I_p, \mathbf{0}_p]\{D_u^\top H_u D_u\}^{-1} D_u^\top H_u \left[M - D_u \begin{pmatrix} \beta(u) \\ h\beta'(u) \end{pmatrix} \right]$$
$$= [I_p, \mathbf{0}_p]\{D_u^\top H_u D_u\}^{-1} D_u^\top H_u \begin{pmatrix} \Delta^* x_{11}^\top \left(\frac{u_{11}-u}{h}\right)^2 \\ \vdots \\ \Delta^* x_{NT}^\top \left(\frac{u_{NT}-u}{h}\right)^2 \end{pmatrix} \frac{h^2}{2}\beta''(u)\{1 + o_p(h^2)\}.$$

Similar to the derivation of (A.1), it is easy to show that

$$\frac{1}{NT} D_u^\top H_u \begin{pmatrix} \Delta^* x_{11}^\top \left(\frac{u_{11}-u}{h}\right)^2 \\ \vdots \\ \Delta^* x_{NT}^\top \left(\frac{u_{NT}-u}{h}\right)^2 \end{pmatrix} = \begin{pmatrix} \Phi(u)\mu_2 \\ \mathbf{0} \end{pmatrix} \{1 + O_p(c_{NT})\}$$

Therefore, we have

$$\mathbb{I}_7 = \frac{h^2}{2}\mu_2 \beta''(u)\{1 + O_p(c_{NT})\},$$

and

$$\sqrt{NTh}\left[\hat{\beta}(u) - \beta(u) - \frac{h^2}{2}\mu_2\beta''(u)\{1 + O_p(c_{NT})\}\right] = \sqrt{NTh}(\mathbb{I}_8 + \mathbb{I}_9 + \mathbb{I}_{10}).$$

Next, we turn to the analysis of $\sqrt{NTh}\mathbb{I}_8$. By (A.1), (A.2), and Theorem 4.1, we can derive that

$$\sqrt{NTh}\mathbb{I}_8 = \sqrt{h}[I_p, \mathbf{0}_p]\{D_u^\top H_u D_u\}^{-1} D_u^\top H_u Z \sqrt{NT}(\gamma - \hat{\gamma})$$
$$= \sqrt{h}\Phi^{-1}(u)\Psi(u)\{1 + O_p(c_{NT})\}O_p(1).$$

As for $\sqrt{NTh}\mathbb{I}_9$, it can be shown that

$$\sqrt{NTh}\mathbb{I}_9 = \sqrt{NTh}[I_p, \mathbf{0}_p]\left\{\frac{D_u^\top H_u D_u}{NT}\right\}^{-1} \frac{D_u^\top H_u P_T}{NT}$$
$$= O_p\left(\frac{\sqrt{NTh}\log T}{T}\right)\{1 + O_p(c_{NT})\}.$$

Finally, we consider $\sqrt{NTh}\mathbb{I}_{10}$. Note that

$$\sqrt{NTh}\mathbb{I}_{10} = \sqrt{NTh}[I_p, \mathbf{0}_p]\{D_u^\top H_u D_u\}^{-1} D_u^\top H_u \epsilon$$
$$= \Phi^{-1}(u)\sqrt{NTh}\left\{\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \Delta^* x_{it}\Delta^*\epsilon_{it}K_h(u_{it} - u)\right\}\{1 + O_p(c_{NT})\}.$$

The variance of $\sqrt{NTh}\left\{\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \Delta^* x_{it}\Delta^*\epsilon_{it}K_h(u_{it} - u)\right\}$ can be derived as follows:

$$NTh \cdot Var\left\{\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \Delta^* x_{it}\Delta^*\epsilon_{it}K_h(u_{it} - u)\right\}$$
$$= \frac{h}{NT}\sum_{i=1}^N\sum_{t=1}^T Var\big(\Delta^* x_{it}\Delta^*\epsilon_{it}K_h(u_{it} - u)\big)$$
$$+ \frac{h}{NT}\sum_{i=1}^N\sum_{t=1}^T\sum_{s\neq t} E\big(\Delta^* x_{it}\Delta^*\epsilon_{it}\Delta^*\epsilon_{is}\Delta^* x_{is}^\top K_h(u_{it} - u)K_h(u_{is} - u)\big)$$
$$= f(u)\nu_0\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T E\{\Delta^* x_{it}\Delta^* x_{it}^\top E[\Delta^*\epsilon_{it}^2 | u_{it}, \Delta^* x_{it}] | u_{it} = u\}\{1 + O(h)\}$$
$$= \mathbf{\Omega}^*(u)\{1 + O(h)\}.$$

Now, note that

$$\sqrt{NTh}\left\{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Delta^* x_{it}\Delta^*\epsilon_{it}K_h(u_{it}-u)\right\} = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}Q_i^*,$$

where $Q_i^* = (1/\sqrt{T})\sum_{t=1}^{T}\sqrt{h}\Delta^* x_{it}\Delta^*\epsilon_{it}K_h(u_{it}-u)$. For any unit vector $\boldsymbol{d} \in \mathbb{R}^q$, let $\omega_i^* = \boldsymbol{d}^\top Q_i^*$. It easy to show that

$$\frac{1}{\rho_1^4 N^2}\sum_{i=1}^{N}E\left|\omega_i^*\right|^4 \to 0,$$

as $(N, T) \to \infty$, where $\rho_1^2 = \boldsymbol{d}^\top\Omega^*(u)\boldsymbol{d}$. By the Cramér–Wold theorem and the Lindeberg–Feller central limit theorem, the asymptotic normality is established.

Proof of Theorem 4.3. Under H_{10}, Equation (2.3) becomes $\Delta^* y_{it} = \Delta^* x_{it}^\top\beta_0(u_{it}) + \Delta^* z_{it}^\top\gamma + P_{it^*,T} + \Delta^*\epsilon_{it}$. Let $\Delta^*\varpi_{it}$ be $(\Delta^* x_{it}^\top, \Delta^* z_{it}^\top)^\top$. Further, define $\Gamma_0(u_{it})$ and $\hat{\Gamma}_0(u_{it})$ as $(\beta_0^\top(u_{it}), \gamma^\top)^\top$ and $(\hat{\beta}_0^\top(u_{it}), \hat{\gamma}^\top)^\top$, respectively. Then, $\tilde{W}_{NT}$ can be decomposed as follows:

$$\begin{aligned}
\tilde{W}_{NT} &= \frac{1}{N^2T^2h}\sum_{i\neq j}\sum_{t=1}^{T}\sum_{s=1}^{T}\left[\Delta^* x_{it}(\Delta^* y_{it} - \Delta^* z_{it}^\top\hat{\gamma} - \Delta^* x_{it}^\top\hat{\beta}_0(u_{it}))\right]^\top \\
&\quad\times\left[\Delta^* x_{js}(\Delta^* y_{js} - \Delta^* z_{js}^\top\hat{\gamma} - \Delta^* x_{js}\hat{\beta}_0(u_{js}))\right]K\left(\frac{u_{it}-u_{js}}{h}\right) \\
&= \frac{1}{N^2T^2h}\sum_{i\neq j}\sum_{t=1}^{T}\sum_{s=1}^{T}\left\{\Delta^* x_{it}\left[\Delta^*\epsilon_{it} + \Delta^*\varpi_{it}^\top(\Gamma_0(u_{it}) - \hat{\Gamma}_0(u_{it})) + P_{it^*,T}\right]\right\}^\top \\
&\quad\times\left\{\Delta^* x_{js}\left[\Delta^*\epsilon_{js} + \Delta^*\varpi_{js}^\top\left(\Gamma_0(u_{js}) - \hat{\Gamma}_0(u_{js})\right) + P_{js^*,T}\right]\right\}K\left(\frac{u_{it}-u_{js}}{h}\right) \\
&= \mathbb{A}_{1NT} + 2\mathbb{A}_{2NT} + 2\mathbb{A}_{3NT} + \mathbb{A}_{4NT} + 2\mathbb{A}_{5NT} + \mathbb{A}_{6NT},
\end{aligned}$$

where

$$\begin{aligned}
\mathbb{A}_{1NT} &= \frac{1}{N^2T^2h}\sum_{i\neq j}\sum_{t=1}^{T}\sum_{s=1}^{T}\Delta^*\epsilon_{it}\Delta^*\epsilon_{js}\Delta^* x_{it}^\top\Delta^* x_{js}K\left(\frac{u_{it}-u_{js}}{h}\right), \\
\mathbb{A}_{2NT} &= \frac{1}{N^2T^2h}\sum_{i\neq j}\sum_{t=1}^{T}\sum_{s=1}^{T}\Delta^*\epsilon_{it}\Delta^* x_{it}^\top\Delta^* x_{js}\Delta^*\varpi_{js}^\top\left(\Gamma_0(u_{js}) - \hat{\Gamma}_0(u_{js})\right)K\left(\frac{u_{it}-u_{js}}{h}\right), \\
\mathbb{A}_{3NT} &= \frac{1}{N^2T^2h}\sum_{i\neq j}\sum_{t=1}^{T}\sum_{s=1}^{T}\Delta^*\epsilon_{it}\Delta^* x_{it}^\top\Delta^* x_{js}P_{js^*,T}K\left(\frac{u_{it}-u_{js}}{h}\right), \\
\mathbb{A}_{4NT} &= \frac{1}{N^2T^2h}\sum_{i\neq j}\sum_{t=1}^{T}\sum_{s=1}^{T}(\Gamma_0(u_{it}) - \hat{\Gamma}_0(u_{it}))^\top\Delta^*\varpi_{it}\Delta^* x_{it}^\top\Delta^* x_{js} \\
&\qquad\times\ \Delta^*\varpi_{js}^\top\left(\Gamma_0(u_{js}) - \hat{\Gamma}_0(u_{js})\right)K\left(\frac{u_{it}-u_{js}}{h}\right), \\
\mathbb{A}_{5NT} &= \frac{1}{N^2T^2h}\sum_{i\neq j}\sum_{t=1}^{T}\sum_{s=1}^{T}(\Gamma_0(u_{it}) - \hat{\Gamma}_0(u_{it}))^\top\Delta^*\varpi_{it}\Delta^* x_{it}^\top\Delta^* x_{js}P_{js^*,T}K\left(\frac{u_{it}-u_{js}}{h}\right), \\
\mathbb{A}_{6NT} &= \frac{1}{N^2T^2h}\sum_{i\neq j}\sum_{t=1}^{T}\sum_{s=1}^{T}P_{it^*,T}\Delta^* x_{it}^\top\Delta^* x_{js}P_{js^*,T}K\left(\frac{u_{it}-u_{js}}{h}\right).
\end{aligned}$$

We complete the proof by showing that under H_{10}, (i) $NTh^{1/2}\mathbb{A}_{1NT} \xrightarrow{d} N(0, V_W)$, where $V_W = \operatorname{plim}_{(N,T)\to\infty}\frac{4}{N^2T^2h}\sum_{1\leq i<j\leq N}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left[\Delta^*\epsilon_{it}^2\Delta^*\epsilon_{js}^2(\Delta^* x_{it}^\top\Delta^* x_{js})^2K^2\left(\frac{u_{it}-u_{js}}{h}\right)\right]$, (ii) $NTh^{1/2}\mathbb{A}_{rNT} = o_p(1)$ for $r = 2, 3, \ldots, 6$, and (iii) $\hat{V}_W = V_W + o_p(1)$.

Proposition 1. *Under H_{10}, $NTh^{1/2}\mathbb{A}_{1NT} \xrightarrow{d} N(0, V_W)$.*

Proof. Rewrite $NTh^{1/2}\mathbb{A}_{1NT}$ as

$$NTh^{1/2}\mathbb{A}_{1NT} = \sum_{1\le i<j\le N} W_{ij},$$

where $W_{ij} = 2(NTh^{1/2})^{-1}\sum_{1\le t,s\le T}\Delta^*\epsilon_{it}\Delta^*\epsilon_{js}\Delta^*x_{it}^\top\Delta^*x_{js}K\left(\frac{u_{it}-u_{js}}{h}\right)$. Note that $\sum_{1\le i<j\le N}W_{ij}$ is a degenerate second order U statistic. The asymptotically normality can be established through the application of Proposition 3.2 in de Jong (1987) for independently but nonidentically distributed (INID) observations. To achieve this goal, we need to verify the following conditions: (C1) $G_I \equiv \sum_{1\le i<j\le N}E(W_{ij}^4) = o(1)$, (C2) $G_{II} \equiv \sum_{1\le i<j<k\le N}E(W_{ij}^2W_{ik}^2 + W_{ji}^2W_{jk}^2 + W_{ki}^2W_{kj}^2) = o(1)$, and (C3) $G_{III} \equiv \sum_{1\le i<j<k<l\le N}E(W_{ij}W_{ik}W_{lj}W_{lk} + W_{ij}W_{il}W_{kj}W_{kl} + W_{ik}W_{il}W_{jk}W_{jl}) = o(1)$.

For (C1), noting that

$$\begin{aligned}\sum_{1\le i<j\le N}E(W_{ij}^4) = \frac{16}{N^4T^4h^2}\sum_{1\le i<j\le N}\sum_{1\le t_1,\ldots,t_8\le T}&E(\Delta^*\epsilon_{it_1}\Delta^*\epsilon_{it_3}\Delta^*\epsilon_{it_5}\Delta^*\epsilon_{it_7}\Delta^*\epsilon_{jt_2}\Delta^*\epsilon_{jt_4}\Delta^*\epsilon_{jt_6}\\ &\times\Delta^*\epsilon_{jt_8}\Delta^*x_{it_1}^\top\Delta^*x_{jt_2}\Delta^*x_{it_3}^\top\Delta^*x_{jt_4}\Delta^*x_{it_5}^\top\Delta^*x_{jt_6}\Delta^*x_{it_7}^\top\Delta^*x_{jt_8}\\ &\times K_{ij,t_1t_2}K_{ij,t_3t_4}K_{ij,t_5t_6}K_{ij,t_7t_8}),\end{aligned}\tag{A.4}$$

where $K_{ij,ts} = K\left(\frac{u_{it}-u_{js}}{h}\right)$. Let $\mathscr{I}$ be any index set, $\#\mathscr{I}$ denote the number of distinct elements in $\mathscr{I}$. Further, let $\mathscr{S}$ be any set of random variables, $\#_{ind}\mathscr{S}$ denote the number of elements in the maximal mutually independent subset of $\mathscr{S}$. If $\#\{t_1,\ldots,t_8\} = 8$, we consider four cases:

(Ia) $\#_{ind}\{\Delta^*\epsilon_{it_1},\Delta^*\epsilon_{it_3},\Delta^*\epsilon_{it_5},\Delta^*\epsilon_{it_7}\} = 3$ or 4, and $\#_{ind}\{\Delta^*\epsilon_{jt_2},\Delta^*\epsilon_{jt_4},\Delta^*\epsilon_{jt_6},\Delta^*\epsilon_{jt_8}\} = 3$ or 4,

(Ib) $\#_{ind}\{\Delta^*\epsilon_{it_1},\Delta^*\epsilon_{it_3},\Delta^*\epsilon_{it_5},\Delta^*\epsilon_{it_7}\} = 3$ or 4, and $\#_{ind}\{\Delta^*\epsilon_{jt_2},\Delta^*\epsilon_{jt_4},\Delta^*\epsilon_{jt_6},\Delta^*\epsilon_{jt_8}\} = 2$,

(Ic) $\#_{ind}\{\Delta^*\epsilon_{it_1},\Delta^*\epsilon_{it_3},\Delta^*\epsilon_{it_5},\Delta^*\epsilon_{it_7}\} = 2$, and $\#_{ind}\{\Delta^*\epsilon_{jt_2},\Delta^*\epsilon_{jt_4},\Delta^*\epsilon_{jt_6},\Delta^*\epsilon_{jt_8}\} = 3$ or 4,

(Id) $\#_{ind}\{\Delta^*\epsilon_{it_1},\Delta^*\epsilon_{it_3},\Delta^*\epsilon_{it_5},\Delta^*\epsilon_{it_7}\} = \#_{ind}\{\Delta^*\epsilon_{jt_2},\Delta^*\epsilon_{jt_4},\Delta^*\epsilon_{jt_6},\Delta^*\epsilon_{jt_8}\} = 2$.

According to these cases, (A.4) can be written as

$$\sum_{1\le i<j\le N}E(W_{ij}^4) = \mathbb{S}_{Ia} + \mathbb{S}_{Ib} + \mathbb{S}_{Ic} + \mathbb{S}_{Id} + \mathbb{S}_{Ir},$$

where $\mathbb{S}_{Il}$ denotes the corresponding summation for case (Il), $l\in\{a,b,c\}$, and $\mathbb{S}_{Ir}$ corresponds to the summation for $\#\{\{t_1,\ldots,t_8\} < 8$. One can easily prove that $\mathbb{S}_{Ia} = \mathbb{S}_{Ib} = \mathbb{S}_{Ic} = 0$. As for $\mathbb{S}_{Id}$, recall that, for any t, $\Delta^*\epsilon_{it}$ are correlated with at most two elements among $\{\Delta^*\epsilon_{is} : s\neq t\}$. Thus, it is easy to show that $\mathbb{S}_{Id} = O(h^2/N^2) = o(1)$. Using similar argument, we can prove that $\mathbb{S}_{Ir} = o(1), G_{II} = o(1)$, and $G_{III} = o(1)$.

Finally, we have to show that $Var(NTh^{1/2}\mathbb{A}_{1NT}) = V_W + o(1)$. By Assumption A1, it is easy to see that $E(\mathbb{A}_{1NT}) = 0$. Therefore,

$$\begin{aligned}Var(NTh^{1/2}\mathbb{A}_{1NT}) &= E(N^2T^2h\cdot\mathbb{A}_{1NT}^2)\\ &= \frac{4}{N^2T^2h}\sum_{1\le i<j\le N}\sum_{t_1=1}^T\sum_{t_2=1}^T\sum_{s_1=1}^T\sum_{s_2=1}^T E(\Delta^*\epsilon_{it_1}\Delta^*\epsilon_{it_2}\Delta^*\epsilon_{js_1}\Delta^*\epsilon_{js_2}\\ &\times\ \Delta^*x_{it_1}^\top\Delta^*x_{js_1}\Delta^*x_{it_2}^\top\Delta^*x_{js_2}K_{ij,t_1s_1}K_{ij,t_2s_2})\\ &= \frac{4}{N^2T^2h}\sum_{1\le i<j\le N}\sum_{t=1}^T\sum_{s=1}^T E\left[\Delta^*\epsilon_{it}^2\Delta^*\epsilon_{js}^2(\Delta^*x_{it}^\top\Delta^*x_{js})^2K_{ij,ts}^2\right] + o(1)\\ &= V_W + o(1).\end{aligned}$$

Therefore, the asymptotic normality of $NTh^{1/2}\mathbb{A}_{1NT}$ has been established.

Proposition 2. *Under H_{10}, $NTh^{1/2}\mathbb{A}_{rNT} = o_p(1)$ for $r = 2,3,\ldots,6$.*

Proof. We only prove $NTh^{1/2}\mathbb{A}_{4NT} = o_p(1)$, and the proofs of other results follow in a similar manner. With the help of the Mean Value Theorem, we can rewrite $\mathbb{A}_{4NT}$ as follows

$$\mathbb{A}_{4NT} = \begin{pmatrix} \hat{\theta}-\theta \\ \hat{\gamma}-\gamma \end{pmatrix}^{\top} \left\{ \frac{1}{N^2T^2h} \sum_{i\neq j} \sum_{1\leq t,s\leq T} \Xi(u_{it},\bar{\theta}_1)^{\top} \Delta^* \varpi_{it} \Delta^* x_{it}^{\top} \Delta^* x_{js} \Delta^* \varpi_{js}^{\top} \Xi(u_{js},\bar{\theta}_2) K_{ij,ts} \right\} \times \begin{pmatrix} \hat{\theta}-\theta \\ \hat{\gamma}-\gamma \end{pmatrix}$$
$$\equiv \begin{pmatrix} \hat{\theta}-\theta \\ \hat{\gamma}-\gamma \end{pmatrix}^{\top} \mathbb{S}_{4NT} \begin{pmatrix} \hat{\theta}-\theta \\ \hat{\gamma}-\gamma \end{pmatrix},$$

where $\bar{\theta}_1$, depending on u_{it}, lies between θ and $\hat{\theta}$, and $\bar{\theta}_2$, which is dependent on u_{js}, also lies between θ and $\hat{\theta}$, and

$$\Xi(u_{it},\bar{\theta}_1) = \begin{pmatrix} \frac{\partial\beta^{\top}(u_{it};\bar{\theta}_1)}{\partial\theta} & \mathbf{0}_{p\times q} \\ \mathbf{0}_{q\times d_\theta} & I_{q\times q} \end{pmatrix}, \quad \Xi(u_{js},\bar{\theta}_2) = \begin{pmatrix} \frac{\partial\beta^{\top}(u_{js};\bar{\theta}_2)}{\partial\theta} & \mathbf{0}_{p\times q} \\ \mathbf{0}_{q\times d_\theta} & I_{q\times q} \end{pmatrix},$$

where d_θ is the dimension of θ.

From Theorem 4.1 and the fact that $\hat{\theta}$ is a $\sqrt{NT}$ consistent estimator of θ, we have $\begin{pmatrix} \hat{\theta}-\theta \\ \hat{\gamma}-\gamma \end{pmatrix} = O_p(1/\sqrt{NT})$. In addition, we can easily show that $\mathbb{S}_{4NT} = O_p(1)$. Thus, we have

$$NTh^{1/2}\mathbb{A}_{4NT} = NTh^{1/2} \cdot O_p\left(1/\sqrt{NT}\right) \cdot O_p(1) \cdot O_p\left(1/\sqrt{NT}\right) = O_p(h^{1/2}) = o_p(1).$$

This proof is completed.

Proposition 3. *Under H_{10}, $\hat{V}_W = V_W + o_p(1)$.*

Proof. It suffices to show that $\hat{V}_W = \tilde{V}_W + o_p(1)$, where

$$\tilde{V}_W = \frac{2}{N^2T^2h} \sum_{i\neq j}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} E\left[\Delta^*\epsilon_{it}^2 \Delta^*\epsilon_{js}^2 (\Delta^* x_{it}^{\top} \Delta^* x_{js})^2 K^2\left(\frac{u_{it}-u_{js}}{h}\right)\right].$$

We first decompose $\hat{V}_W - \tilde{V}_W$ as follows

$$\begin{aligned} \hat{V}_W - \tilde{V}_W = &\frac{2}{N^2T^2h} \sum_{i\neq j}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} \Big\{ \Delta^*\epsilon_{it}^2 \Delta^*\epsilon_{js}^2 (\Delta^* x_{it}^{\top}\Delta^* x_{js})^2 K_{ij,ts}^2 \\ &- E\Big[\Delta^*\epsilon_{it}^2 \Delta^*\epsilon_{js}^2 (\Delta^* x_{it}^{\top}\Delta^* x_{js})^2 K_{ij,ts}^2\Big]\Big\} \\ &+ \frac{2}{N^2T^2h} \sum_{i\neq j}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} \Big\{ \Delta^*\hat{\epsilon}_{it}^2 \Delta^*\hat{\epsilon}_{js}^2 (\Delta^* x_{it}^{\top}\Delta^* x_{js})^2 K_{ij,ts}^2 \\ &- \Delta^*\epsilon_{it}^2 \Delta^*\epsilon_{js}^2 (\Delta^* x_{it}^{\top}\Delta^* x_{js})^2 K_{ij,ts}^2 \Big\} \\ \equiv &\ \mathbb{V}_{1NT} + \mathbb{V}_{2NT}. \end{aligned}$$

Noting that $E(\mathbb{V}_{1NT}) = 0$, and $E(\mathbb{V}_{1NT}^2) = o(1)$ by Assumption A1. Applying the Chebyshev inequality, we obtain $\mathbb{V}_{1NT} = o_p(1)$.

For $\mathbb{V}_{2NT}$, we can write

$$\begin{aligned} \mathbb{V}_{2NT} = &\frac{2}{N^2T^2h} \sum_{i\neq j}^{N} \sum_{1\leq t,s\leq T} (\Delta^* x_{it}^{\top}\Delta^* x_{js})^2 K_{ij,ts}^2 (\Delta^*\hat{\epsilon}_{it}^2 - \Delta^*\epsilon_{it}^2)(\Delta^*\hat{\epsilon}_{js}^2 - \Delta^*\epsilon_{js}^2) \\ &+ \frac{2}{N^2T^2h} \sum_{i\neq j}^{N} \sum_{1\leq t,s\leq T} (\Delta^* x_{it}^{\top}\Delta^* x_{js})^2 K_{ij,ts}^2 (\Delta^*\hat{\epsilon}_{it}^2 - \Delta^*\epsilon_{it}^2)\Delta^*\epsilon_{js}^2 \\ &+ \frac{2}{N^2T^2h} \sum_{i\neq j}^{N} \sum_{1\leq t,s\leq T} (\Delta^* x_{it}^{\top}\Delta^* x_{js})^2 K_{ij,ts}^2 \Delta^*\epsilon_{it}^2(\Delta^*\hat{\epsilon}_{js}^2 - \Delta^*\epsilon_{js}^2) \\ \equiv &\ \mathbb{V}_{2NT,1} + \mathbb{V}_{2NT,2} + \mathbb{V}_{2NT,3} = \mathbb{V}_{2NT,1} + 2\mathbb{V}_{2NT,2}. \end{aligned}$$

We proceed to prove $\mathbb{V}_{2NT} = o_p(1)$ by showing that (i1) $\mathbb{V}_{2NT,1} = o_p(1)$, and (i2) $\mathbb{V}_{2NT,2} = o_p(1)$.

For $\mathbb{V}_{2NT,1}$, we have

$$
\begin{aligned}
|\mathbb{V}_{2NT,1}| &\le \frac{C}{N^2T^2h}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{1\le t,s\le T}\left|\left(\Delta^*\hat{\epsilon}_{it}^2-\Delta^*\epsilon_{it}^2\right)\left(\Delta^*\hat{\epsilon}_{js}^2-\Delta^*\epsilon_{js}^2\right)\right|\cdot\left|\left(\Delta^*x_{it}^\top\Delta^*x_{js}\right)^2\right| \\
&= \frac{C}{N^2T^2h}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{1\le t,s\le T}\left|\left(\Delta^*\hat{\epsilon}_{it}^2-\Delta^*\epsilon_{it}^2\right)\left(\Delta^*\hat{\epsilon}_{js}^2-\Delta^*\epsilon_{js}^2\right)\right| \\
&\qquad\times\left|\sum_{l=1}^{p}\sum_{m=1}^{p}\Delta^*x_{it,l}\Delta^*x_{js,l}\Delta^*x_{it,m}\Delta^*x_{js,m}\right| \\
&\le \frac{C}{N^2T^2h}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{1\le t,s\le T}\left|\left(\Delta^*\hat{\epsilon}_{it}^2-\Delta^*\epsilon_{it}^2\right)\left(\Delta^*\hat{\epsilon}_{js}^2-\Delta^*\epsilon_{js}^2\right)\right|\left(\sum_{l=1}^{p}\sum_{m=1}^{p}|\Delta^*x_{it,l}\Delta^*x_{it,m}|\right) \\
&\qquad\times\left(\sum_{l'=1}^{p}\sum_{m'=1}^{p}|\Delta^*x_{js,l'}\Delta^*x_{js,m'}|\right) \\
&= \frac{C}{h}\left\{(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{l=1}^{p}\sum_{m=1}^{p}|\Delta^*\hat{\epsilon}_{it}-\Delta^*\epsilon_{it}|\cdot|\Delta^*\hat{\epsilon}_{it}+\Delta^*\epsilon_{it}|\cdot|\Delta^*x_{it,l}\Delta^*x_{it,m}|\right\}^2 \\
&\le C\left\{(NTh)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}(\Delta^*\hat{\epsilon}_{it}-\Delta^*\epsilon_{it})^2\right\} \\
&\times\left\{(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{l=1}^{p}\sum_{m=1}^{p}(\Delta^*\epsilon_{it}+\Delta^*\epsilon_{it})^2(\Delta^*x_{it,l}\Delta^*x_{it,m})^2\right\},
\end{aligned}
$$

where $\Delta^*x_{it,l}$ denotes the lth element of Δ^*x_{it}, and the last inequality follows from the Cauchy–Schwarz inequality. Note that $\Delta^*\hat{\epsilon}_{it}=\Delta^*z_{it}^\top(\gamma-\hat{\gamma})+\Delta^*x_{it}^\top(\beta_0(u_{it})-\hat{\beta}_0(u_{it}))+P_{it^*,T}+\Delta^*\epsilon_{it}$. Under H_{10}, the following relation holds

$$
\sum_{i=1}^{N}\sum_{t=1}^{T}\Delta^*\hat{\epsilon}_{it}^{*2}\le 4\sum_{i=1}^{N}\sum_{t=1}^{T}\Delta^*\epsilon_{it}^2,
$$

which implies that

$$
\begin{aligned}
&(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{l=1}^{p}\sum_{m=1}^{p}(\Delta^*\hat{\epsilon}_{it}+\Delta^*\epsilon_{it})^2(\Delta^*x_{it,l}\Delta^*x_{it,m})^2 \\
&\le 2(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{l=1}^{p}\sum_{m=1}^{p}(\Delta^*\hat{\epsilon}_{it}^2+\Delta^*\epsilon_{it}^2)(\Delta^*x_{it,l}\Delta^*x_{it,m})^2 \\
&\le C(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{l=1}^{p}\sum_{m=1}^{p}\Delta^*\epsilon_{it}^2(\Delta^*x_{it,l}\Delta^*x_{it,m})^2 \\
&= O_p(1).
\end{aligned}
$$

Therefore, $\mathbb{V}_{2NT,1}=o_p(1)$ can be established by showing that $(NTh)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}(\Delta^*\hat{\epsilon}_{it}-\Delta^*\epsilon_{it})^2=o_p(1)$. In fact, since $\Delta^*\hat{\epsilon}_{it}-\Delta^*\epsilon_{it}=\Delta^*z_{it}^\top(\gamma-\hat{\gamma})+\Delta^*x_{it}^\top(\beta_0(u_{it})-\hat{\beta}_0(u_{it}))+P_{it^*,T}$,

$$
\begin{aligned}
&(NTh)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}(\Delta^*\hat{\epsilon}_{it}-\Delta^*\epsilon_{it})^2 \\
&= (NTh)^{-1}\sum_{i=1}^{N}||\Delta^*\hat{\epsilon}_i-\Delta^*\epsilon_i||^2 \\
&= (NTh)^{-1}\sum_{i=1}^{N}||\Delta^*\Xi_i(\theta-\hat{\theta})+\Delta^*z_i(\gamma-\hat{\gamma})+P_{i,T}||^2 \\
&\le 2(NTh)^{-1}\sum_{i=1}^{N}\left(||\Delta^*\Xi_i(\theta-\hat{\theta})||^2+||\Delta^*z_i(\gamma-\hat{\gamma})||^2+||P_{i,T}||^2\right) \\
&= o_p(1),
\end{aligned}
$$

where $\Delta^*\hat{\epsilon}_i=(\Delta^*\hat{\epsilon}_{i1},...,\Delta^*\hat{\epsilon}_{iT})^\top,\Delta^*\epsilon_i=(\Delta^*\epsilon_{i1},...,\Delta^*\epsilon_{iT})^\top,\Delta^*z_i=(\Delta^*z_{i1},...,\Delta^*z_{iT})^\top,P_{i,T}=(P_{i1^*,T},...,P_{iT^*,T})$, and

$$\Delta^*\Xi_i = \begin{pmatrix} \Delta^* x_{i1}^\top \frac{\partial \beta^\top(u_{i1};\bar{\theta}_1)}{\partial\theta} \\ \vdots \\ \Delta^* x_{iT}^\top \frac{\partial \beta^\top(u_{iT};\bar{\theta}_T)}{\partial\theta} \end{pmatrix}, \quad \bar{\theta}_t \text{ lies between } \theta \text{ and } \hat{\theta}, t = 1, ..., T.$$

Now, we consider $\mathbb{V}_{2NT,2}$. Following the arguments in the proof of Lemma A.2.3 in Cai, Fang and Xu (2020), we rewrite $K(\cdot)$ as a twofold convolution of another symmetric PDF $\bar{K}(\cdot) : K(v) = \int \bar{K}(u)\bar{K}(v-u)du$. Thus,

$$\begin{aligned}
|\mathbb{V}_{2NT,2}| &\le h(NT)^{-2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{1\le t,s\le T}|\Delta^*\hat{\epsilon}_{it}^2 - \Delta^*\epsilon_{it}^2|\Delta^*\epsilon_{js}^2\left(\Delta^* x_{it}^\top \Delta^* x_{js}\right)^2 K_h^2(u_{it}-u_{js}) \\
&\le h(NT)^{-2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{1\le t,s\le T}|\Delta^*\hat{\epsilon}_{it}^2 - \Delta^*\epsilon_{it}^2|\Delta^*\epsilon_{js}^2\left(\sum_{l=1}^{p}\sum_{m=1}^{p}|\Delta^* x_{it,l}\Delta^* x_{it,m}|\right) \\
&\quad \times \left(\sum_{l'=1}^{p}\sum_{m'=1}^{p}|\Delta^* x_{js,l'}\Delta^* x_{js,m'}|\right)K_h^2(u_{it}-u_{js}) \\
&= h(NT)^{-2}\int\int\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{l=1}^{p}\sum_{m=1}^{p}|\Delta^*\hat{\epsilon}_{it}^2 - \Delta^*\epsilon_{it}^2||\Delta^* x_{js,l'}\Delta^* x_{js,m'}|\bar{K}_h(u_{it}-u)\bar{K}_h(u_{it}-u^*) \\
&\quad \times \sum_{j=1}^{N}\sum_{s=1}^{T}\sum_{l'=1}^{p}\sum_{m'=1}^{p}\Delta^*\epsilon_{js}^2|\Delta^* x_{js,l'}\Delta^* x_{js,m'}|\bar{K}_h(u_{js}-u)\bar{K}_h\left(u_{js}-u^*\right)dudu^* \\
&\le \mathbb{V}_{2NT,2,1}^{1/2}\mathbb{V}_{2NT,2,2}^{1/2},
\end{aligned}$$

where

$$\begin{aligned}
\mathbb{V}_{2NT,2,2} &= h(NT)^{-2}\int\int\left[\sum_{j=1}^{N}\sum_{s=1}^{T}\sum_{l'=1}^{p}\sum_{m'=1}^{p}\Delta^*\epsilon_{js}^2|\Delta^* x_{js,l'}\Delta^* x_{js,m'}|\right. \\
&\qquad \left. \times\ \bar{K}_h(u_{js}-u)\bar{K}_h(u_{js}-u^*)\right]^2 dudu^* \\
&= h(NT)^{-2}\sum_{1\le i,j\le N}\sum_{1\le t,s\le T}\sum_{1\le l',m'\le p}\sum_{1\le l,m\le p}\Delta^*\epsilon_{it}^2\Delta^*\epsilon_{js}^2|\Delta^* x_{js,l'}\Delta^* x_{js,m'}||\Delta^* x_{it,l}\Delta^* x_{it,m}| \\
&\quad \times\ K_h^2(u_{it}-u_{js}) \\
&= O_p(1),
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{V}_{2NT,2,1} &= h(NT)^{-2}\int\int\left[\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{l=1}^{p}\sum_{m=1}^{p}|\Delta^*\hat{\epsilon}_{it}^2 - \Delta^*\epsilon_{it}^2||\Delta^* x_{it,l}\Delta^* x_{it,m}|\right. \\
&\qquad \left. \times\ \bar{K}_h(u_{it}-u)\bar{K}_h(u_{it}-u^*)\right]^2 dudu^* \\
&= h(NT)^{-2}\sum_{1\le i,j\le N}\sum_{1\le t,s\le T}\sum_{1\le l',m'\le p}\sum_{1\le l,m\le p}|\Delta^*\hat{\epsilon}_{it}^2 - \Delta^*\epsilon_{it}^2||\Delta^*\hat{\epsilon}_{js}^2 - \Delta^*\epsilon_{js}^2| \\
&\quad \times\ |\Delta^* x_{it,l}\Delta^* x_{it,m}||\Delta^* x_{js,l'}\Delta^* x_{js,m'}|K_h^2(u_{it}-u_{js}) \\
&= o_p(1),
\end{aligned}$$

where the last equality is obtained by employing the same arguments as we have shown in the analysis of $|\mathbb{V}_{2NT,1}|$. Therefore, $\mathbb{V}_{2NT} = o_p(1)$, which completes the proof.

To sum up, we have shown that

$$NTh^{1/2}\tilde{W}_{NT} \xrightarrow{d} N(0, V_W),$$

and V_W can be consistently estimated by $\hat{V}_W$. Therefore, $\mathbb{W}_{NT} \xrightarrow{d} N(0,1)$.

Proof of Theorem 4.4. First, noting that $\tilde{H}_{NT}$ can be decomposed as follows:

$$\begin{aligned}\tilde{H}_{NT} &= (NT(NT-1))^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{s=1,\,\{j,s\}\neq\{i,t\}}^{T}\hat{v}_{it}\hat{v}_{js}K^{*}_{h,\,it,\,js}\\ &= (NT(NT-1))^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\sum_{j\neq i}^{N}\sum_{s=1}^{T}+\sum_{j=i}^{N}\sum_{s\neq t}^{T}\right)\hat{v}_{it}\hat{v}_{js}K^{*}_{h,\,it,\,js}\\ &\equiv \tilde{H}_{NT,1}+\tilde{H}_{NT,2}.\end{aligned}$$

Comparing with $\tilde{H}_{NT,1}$, it is not difficult to prove that $\tilde{H}_{NT,2}$ is negligible. Thus, we focus on $\tilde{H}_{NT,1}$. By the definition of $\hat{v}_{it}$, we have

$$\hat{v}_{it} = v_{it} + X_{it}^{\top}(\beta(u_{it})-\hat{\beta}(u_{it})) + Z_{it}^{\top}(\gamma-\hat{\gamma}).$$

Therefore,

$$\begin{aligned}\tilde{H}_{NT,1} &= (NT(NT-1))^{-1}\sum_{i\neq j}^{N}\sum_{1\leq t,s\leq T}\hat{v}_{it}\hat{v}_{js}K^{*}_{h,\,it,\,js}\\ &= (NT(NT-1))^{-1}\sum_{i\neq j}^{N}\sum_{1\leq t,s\leq T}\left[v_{it}+X_{it}^{\top}(\beta(u_{it})-\hat{\beta}(u_{it}))+Z_{it}^{\top}(\gamma-\hat{\gamma})\right]\\ &\quad\times\left[v_{js}+X_{js}^{\top}(\beta(u_{js})-\hat{\beta}(u_{js}))+Z_{js}^{\top}(\gamma-\hat{\gamma})\right]K^{*}_{h,\,it,\,js}\\ &= (NT(NT-1))^{-1}\sum_{i\neq j}^{N}\sum_{1\leq t,s\leq T}v_{it}v_{js}K^{*}_{h,\,it,\,js}\\ &\quad+2(NT(NT-1))^{-1}\sum_{i\neq j}^{N}\sum_{1\leq t,s\leq T}v_{it}X_{js}^{\top}(\beta(u_{js})-\hat{\beta}(u_{js}))K^{*}_{h,\,it,\,js}\\ &\quad+2(NT(NT-1))^{-1}\sum_{i\neq j}^{N}\sum_{1\leq t,s\leq T}v_{it}Z_{js}^{\top}(\gamma-\hat{\gamma})K^{*}_{h,\,it,\,js}\\ &\quad+(NT(NT-1))^{-1}\sum_{i\neq j}^{N}\sum_{1\leq t,s\leq T}X_{it}^{\top}(\beta(u_{it})-\hat{\beta}(u_{it}))X_{js}^{\top}(\beta(u_{js})-\hat{\beta}(u_{js}))K^{*}_{h,\,it,\,js}\\ &\quad+2(NT(NT-1))^{-1}\sum_{i\neq j}^{N}\sum_{1\leq t,s\leq T}X_{it}^{\top}(\beta(u_{it})-\hat{\beta}(u_{it}))Z_{js}^{\top}(\gamma-\hat{\gamma})K^{*}_{h,\,it,\,js}\\ &\quad+(NT(NT-1))^{-1}\sum_{i\neq j}^{N}\sum_{1\leq t,s\leq T}Z_{it}^{\top}(\gamma-\hat{\gamma})Z_{js}^{\top}(\gamma-\hat{\gamma})K^{*}_{h,\,it,\,js}\\ &= \mathbb{B}_{1NT}+2\mathbb{B}_{2NT}+2\mathbb{B}_{3NT}+\mathbb{B}_{4NT}+2\mathbb{B}_{5NT}+\mathbb{B}_{6NT}.\end{aligned}$$

An argument similar to the one used in the proof of Theorem 4.3 shows that $NTh^{m/2}\mathbb{B}_{rNT}=o_p(1)$ for $r=2,3,\ldots,6$.

For $\mathbb{B}_{1NT}$, under H_{20}, it is a degenerate second-order U statistic with independent and identically distributed observations. Using Theorem 1 of Hall (1984), one can show that $NTh^{m/2}\mathbb{B}_{1NT}$ is asymptotically normal distributed. The detailed proof is a modification to the proofs of Lemma 3.3 in Zheng (1996) and thus is omitted. The fact that the asymptotical variance of $NTh^{m/2}\mathbb{B}_{1NT}$ can be consistently estimated by $\hat{V}_H$ can be proved by employing Lemma 3.1 of Powell et al. (1989). The detailed proof follows closely from the proof of Lemma 3.3e in Zheng (1996) and is also omitted.

Proof of Theorem 4.5. The proof of Theorem 4.5 is similar to the proofs of Theorems 4.3 and 4.4, and thus we only sketch it.

(a) Under H_{11}, Assumptions A1–A6 ensure that $\hat{\gamma}=\bar{\gamma}+O_p((NT)^{-1/2})$ and $\hat{\theta}=\bar{\theta}+O_p((NT)^{-1/2})$, where $\bar{\gamma}$ and $\bar{\theta}$ are the probability limits of $\hat{\gamma}$ and $\hat{\theta}$, respectively. Then, we have

$$
\begin{aligned}
\tilde{W}_{NT} &= \frac{1}{N^2T^2h}\sum_{i\neq j}\sum_{t=1}^{T}\sum_{s=1}^{T}\left[\Delta^* x_{it}(\Delta^* y_{it} - \Delta^* z_{it}^\top \bar{\gamma} - \Delta^* x_{it}^\top \bar{\beta}_0(u_{it}))\right]^\top \\
&\quad \times \left[\Delta^* x_{js}(\Delta^* y_{js} - \Delta^* z_{js}^\top \bar{\gamma} - \Delta^* x_{js}\bar{\beta}_0(u_{js}))\right] K\left(\frac{u_{it}-u_{js}}{h}\right)\left[1+o_p(1)\right] \\
&= \frac{1}{N^2T^2h}\sum_{i\neq j}\sum_{t=1}^{T}\sum_{s=1}^{T}\left\{\Delta^* x_{it}\left[\Delta^*\epsilon_{it} + \Delta^*\varpi_{it}^\top\left(\Gamma(u_{it}) - \bar{\Gamma}_0(u_{it})\right) + P_{it^*,T}\right]\right\}^\top \\
&\quad \times \left\{\Delta^* x_{js}\left[\Delta^*\epsilon_{js} + \Delta^*\varpi_{js}^\top\left(\Gamma(u_{js}) - \bar{\Gamma}_0(u_{js})\right) + P_{js^*,T}\right]\right\} K\left(\frac{u_{it}-u_{js}}{h}\right) \\
&\quad \times \left[1+o_p(1)\right] \\
&= \left\{\mathbb{A}_{1NT} + 2\bar{\mathbb{A}}_{2NT} + 2\mathbb{A}_{3NT} + \bar{\mathbb{A}}_{4NT} + 2\bar{\mathbb{A}}_{5NT} + \mathbb{A}_{6NT}\right\}\left[1+o_p(1)\right],
\end{aligned}
$$

where $\bar{\beta}_0(u) \equiv \beta_0(u;\bar{\theta}), \Gamma(u_{it}) \equiv (\beta^\top(u_{it}), \gamma^\top)^\top, \bar{\Gamma}_0(u_{it}) \equiv (\bar{\beta}_0^\top(u_{it}), \bar{\gamma}^\top)^\top$, and

$$
\bar{\mathbb{A}}_{2NT} = \frac{1}{N^2T^2h}\sum_{i\neq j}\sum_{t=1}^{T}\sum_{s=1}^{T}\Delta^*\epsilon_{it}\Delta^* x_{it}^\top \Delta^* x_{js}\Delta^*\varpi_{js}^\top\left(\Gamma(u_{js}) - \bar{\Gamma}_0(u_{js})\right)K\left(\frac{u_{it}-u_{js}}{h}\right),
$$

$$
\begin{aligned}
\bar{\mathbb{A}}_{4NT} &= \frac{1}{N^2T^2h}\sum_{i\neq j}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\Gamma(u_{it}) - \bar{\Gamma}_0(u_{it})\right)^\top \Delta^*\varpi_{it}\Delta^* x_{it}^\top \Delta^* x_{js} \\
&\qquad \times \Delta^*\varpi_{js}^\top\left(\Gamma(u_{js}) - \bar{\Gamma}_0(u_{js})\right)K\left(\frac{u_{it}-u_{js}}{h}\right),
\end{aligned}
$$

$$
\bar{\mathbb{A}}_{5NT} = \frac{1}{N^2T^2h}\sum_{i\neq j}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\Gamma(u_{it}) - \bar{\Gamma}_0(u_{it})\right)^\top \Delta^*\varpi_{it}\Delta^* x_{it}^\top \Delta^* x_{js}P_{js^*,T}K\left(\frac{u_{it}-u_{js}}{h}\right).
$$

By utilizing the similar arguments in the proof of Theorem 3.2 in Lin et al. (2014), we can show that $\mathbb{A}_{1NT} \xrightarrow{p} C_1$, where C_1 is a positive constant. Furthermore, following similar steps as in the proof of Proposition 2, we have $NTh^{1/2}\bar{\mathbb{A}}_{rNT} = o_p(1)$ for $r=2$, 4, 5, and $NTh^{1/2}\mathbb{A}_{rNT} = o_p(1)$ for $r=3$, 6. A modification of the proof of Proposition 3 shows that $\hat{V}_W \xrightarrow{p} C_2$, where C_2 is also a positive constant. Therefore,

$$
\mathbb{W}_{NT} = NTh^{1/2}\tilde{W}_{NT}/\sqrt{\hat{V}_W} = \left\{NTh^{1/2}C_1/\sqrt{C_2}\right\}\left[1+o_p(1)\right],
$$

which leads to (a).

(b) The detailed proof is a modification of the proof of Theorem 2 in Zheng (1996) and thus is omitted.

Funding

This research was partially supported from the National Science Fund of China (NSFC) for Distinguished Scholars (71625001), the NSFC key projects with grant numbers 71631004 and 71131008, and the Basic Scientific Center Project of NSFC (71988101).

References

Arellano, M. (2003). *Panel Data Econometrics*, Oxford: Oxford University Press.

Balestra, P., Nerlove, M. (1966). Pooling cross section and time series data in the estimation of a dynamic model: the demand for natural gas. *Econometrica* 34(3):585–612. doi:10.2307/1909771

Baltagi, B. (2008). *Econometric Analysis of Panel Data*, New York: John Wiley & Sons.

Cai, Z., Fan, J., Yao, Q. (2000). Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association* 95(451):941–956. doi:10.1080/01621459.2000.10474284

Cai, Z., Fang, Y., Xu, Q. (2020). Testing capital asset pricing models using functional-coefficient panel data models with cross-sectional dependence. *Journal of Econometrics*. Advance online publication. doi:10.1016/j.jeconom.2020.07.018

Cai, Z., Li, Q. (2008). Nonparametric estimation of varying coefficient dynamic panel data models. *Econometric Theory* 24(5):1321–1342. doi:10.1017/S0266466608080523

Chen, J., Gao, J., Li, D. (2012). Semiparametric trending panel data models with cross-sectional dependence. *Journal of Econometrics* 171(1):71–85. doi:10.1016/j.jeconom.2012.07.001

de Jong, P. (1987). A Central limit theorem for generalized quadratic forms. *Probability Theory and Related Fields* 75(2):261–277. doi:10.1007/BF00354037

Fan, J., Huang, T. (2005). Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli* 11(6):1031–1057. doi:10.3150/bj/1137421639

Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *Journal of Multivariate Analysis* 14(1):1–16. doi:10.1016/0047-259X(84)90044-7

Hausman, J. A., Taylor, W. E. (1981). Panel data and unobservable individual effects. *Econometrica* 49(6): 1377–1398. doi:10.2307/1911406

Henderson, D. J., Carroll, R. J., Li, Q. (2008). Nonparametric estimation and testing of fixed effects panel data models. *Journal of Econometrics* 144(1):257–275. doi:10.1016/j.jeconom.2008.01.005

Hoover, D. R., Rice, J. A., Wu, C. O., Yang, L.-P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* 85(4):809–822. doi:10.1093/biomet/85.4.809

Hsiao, C. (2014). *Analysis of Panel Data, 3rd edition*, Cambridge: Cambridge University Press.

Huang, X. (2013). Nonparametric estimation in large panels with cross-sectional dependence. *Econometric Reviews* 32(5–6):754–777. doi:10.1080/07474938.2013.740998

Janson, S. (1987). Maximal spacings in several dimensions. *The Annals of Probability* 15(1):274–280. doi:10.1214/aop/1176992269

Li, D., Chen, J., Gao, J. (2011). Non-parametric time-varying coefficient panel data models with fixed effects. *The Econometrics Journal* 14(3):387–408. doi:10.1111/j.1368-423X.2011.00350.x

Li, Q., Huang, C. J., Li, D., Fu, T.-T. (2002). Semiparametric smooth coefficient models. *Journal of Business and Economic Statistics* 20(3):412–422. doi:10.1198/073500102288618531

Li, Q., Stengos, T. (1996). Semiparametric estimation of partially linear panel data models. *Journal of Econometrics* 71(1–2):389–397. doi:10.1016/0304-4076(94)01711-5

Lin, Z., Li, Q., Sun, Y. (2014). A consistent nonparametric test of parametric regression functional form in fixed effects panel data models. *Journal of Econometrics* 178(1):167–179. doi:10.1016/j.jeconom.2013.08.014

Mack, Y.-P., Silverman, B. W. (1982). Weak and strong uniform consistency of kernel regression estimates. *Zeitschrift für Wahrscheinlichkeitstheorie Und Verwandte Gebiete* 61(3):405–415. doi:10.1007/BF00539840

Mammen, E., Støve, B., Tjøstheim, D. (2009). Nonparametric additive models for panels of time series. *Econometric Theory* 25(2):442–481. doi:10.1017/S0266466608090142

Powell, J. L., Stock, J. H., Stoker, T. M. (1989). Semiparametric estimation of index coefficients. *Econometrica* 57(6):1403–1430. doi:10.2307/1913713

Qian, J., Wang, L. (2012). Estimating semiparametric panel data models by marginal integration. *Journal of Econometrics* 167(2):483–493. doi:10.1016/j.jeconom.2011.09.030

Su, L., Jin, S. (2012). Sieve estimation of panel data models with cross section dependence. *Journal of Econometrics* 169(1):34–47. doi:10.1016/j.jeconom.2012.01.006

Su, L., Ullah, A. (2006). Profile likelihood estimation of partially linear panel data models with fixed effects. *Economics Letters* 92(1):75–81. doi:10.1016/j.econlet.2006.01.019

Su, L., Ullah, A. (2007). More efficient estimation of nonparametric panel data models with random effects. *Economics Letters* 96(3):375–380. doi:10.1016/j.econlet.2007.02.018

Xue, L., Zhu, L. (2007). Empirical likelihood for a varying coefficient model with longitudinal data. *Journal of the American Statistical Association* 102(478):642–654. doi:10.1198/016214507000000293

Zhang, J., Feng, S., Li, G., Lian, H. (2011). Empirical likelihood inference for partially linear panel data models with fixed effects. *Economics Letters* 113(2):165–167. doi:10.1016/j.econlet.2011.07.014

Zheng, J. X. (1996). A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics* 75(2):263–289.