10-2016

A practical test for strict exogeneity in linear panel data models with fixed effects

Liangjun SU
Singapore Management University, ljsu@smu.edu.sg

Yonghui ZHANG
Renmin University of China

Jie WEI
Huazhong University of Science and Technology

Published in Economics Letters, October 2016, Volume 147, Pages 27-31.
http://doi.org/10.1016/j.econlet.2016.08.012

A Practical Test for Strict Exogeneity in Linear Panel Data Models with Fixed Effects

Liangjun Su[a], Yonghui Zhang[b], Jie Wei[c*]
[a]School of Economics, Singapore Management University
[b]School of Economics, Renmin University of China
[c]School of Economics, Huazhong University of Science and Technology

July 15, 2016

Abstract

This paper provides a practical test for strict exogeneity in linear panel data models with fixed effects when the number of individuals N goes to infinity while the number of time periods T is fixed. The test is based on the supremum of a sequence of Wald test statistics. Under suitable conditions, we establish the asymptotic distribution of the test statistic and consistency of the test. A bootstrap procedure is proposed to improve the finite sample performance and the validity of the procedure is justified. We investigate the finite sample performance of the test via a small set of Monte Carlo simulations. The application to a panel data set of agricultural production rejects the strict exogeneity assumption.

JEL Classification: C12, C23

Key Words: Bootstrap, Fixed effects, Panel data, Strict exogeneity, Wald test

*Su gratefully acknowledges the Singapore Ministry of Education for Tier-2 Academic Research Fund under grant number MOE2012-T2-2-021 and the funding support provided by the Lee Kong Chian Fund for Excellence. Zhang gratefully acknowledges the financial support from the National Natural Science Foundation of China (Project No.71401166). Wei gratefully acknowledges the the financial support from the Fundamental Research Funds for the Central Universities, HUST (310-0118310052). All errors are the authors' sole responsibilities.
Address correspondence to: Jie Wei, School of Economics, Huazhong University of Science and Technology, 1037 Luoyu Road, Wuhan, Hubei, 430074, China; Phone: +86 13429820860; e-mail: jiewei.econ@gmail.com.

1 Introduction

With the availability of a wealth of sources of panel data, researchers usually estimate a textbook panel data model in Baltagi (2013), Hsiao (2014) or Wooldridge (2010):

$$y_{it} = x_{it}'\beta + \alpha_i + u_{it},\ i = 1, 2, \ldots, N, t = 1, 2, \ldots, T, \tag{1.1}$$

where y_{it} is the dependent variable for individual i at time period t, x_{it} is a $k \times 1$ vector of explanatory variables, α_i represents an unobserved individual effect, and u_{it} is the idiosyncratic error. Depending on whether α_i is correlated with x_{it} or not, it is referred to as either fixed effects or random effects. Due to the prevailingness of correlation between x_{it} and α_i in empirical works, the fixed effects approach has received more attention than the random effects approach.

For a fixed effect model, both the fixed effects (FE) estimator and the first-difference (FD) estimator, which adopt within-group and first-difference transformation to eliminate α_i, respectively, have been employed in empirical works. One standard assumption to ensure the $\sqrt{N}$-consistency of these estimators in the large N and fixed T framework is the strict exogeneity of x_{it}. For some recent empirical applications to panel data sets adopting the strict exogeneity assumption, see Boumparis et al. (2015), Earnhart (2004), and Papageorgiadis and Sharma (2016), among others.

However, strict exogeneity of x_{it} may not hold in many applications due to the possible existence of feedback effects or economic periodicity. Wooldridge (2010, p.324) has showed that the FE estimator is generally biased and inconsistent, and its' probability limit is different from the FD estimator when x_{it} is not strictly exogenous. Due to the important role of this assumption, it is necessary to develop a formal test to detect its violation. The only available test in the literature of linear panel data models is constructed by Wooldridge (2010), who introduces a simple test based on an augmented regression, where a subset ($w_{i,t+1}$) of the first order leading term $x_{i,t+1}$ is included in the level equation as additional regressors. Under the null hypothesis of strict exogeneity, the coefficient of $w_{i,t+1}$ should be equal to zero. Then one can construct a Wald test that is robust to arbitrary serial correlation and heteroskedasticity of unknown form. Nevertheless, the test only includes a subset of the first leading explanatory variables, which implies that the test may have power only when $x_{i,t+1}$ is correlated with u_{it}. Clearly, it may not detect a potentially more general structure of intertemporal correlation between $\{u_{it}\}$ and $\{x_{it}\}$. To fill the gap, we propose a practical test for strict exogeneity of regressors in this paper, which generalizes the test by Wooldridge to detect all orders of intertemporal correlation between $\{u_{it}\}$ and $\{x_{it}\}$. Because the limiting distribution of our test statistic is nonstandard, we will propose a bootstrap method to obtain the p-values and justify its asymptotic validity.

The rest of this paper is organized as follows. We formalize the hypotheses in Section 2.

We introduce the test statistic in Section 3, and study its asymptotic properties in Section 4. We evaluate the finite sample performance of our proposed test in Section 5. Section 6 gives an empirical application of the test. Section 7 concludes.

Notation. Let ι_a be an $a \times 1$ vector of ones, $0_{a\times b}$ be an $a \times b$ matrix of zeros and I_a be an $a \times a$ identity matrix. We use $\|A\| = [\mathtt{tr}\,(A'A)]^{1/2}$ to denote the Euclidean norm of matrix A. Denote $\Delta c_{it} = c_{it} - c_{i,t-1}$ and $\dot{c}_{it} = c_{it} - \bar{c}_i$, where $\bar{c}_i = T^{-1}\sum_{s=1}^{T} c_{is}$. The symbols $\to_p$ and $\to_d$ denote convergence in probability and in distribution, respectively.

2 The hypothesis

The strict exogeneity used in the linear panel data model with fixed effects can be stated as

$$\mathtt{E}\,(u_{it}|x_i, \alpha_i) = 0; \tag{2.1}$$

see (10.14) in Wooldridge (2010, p.288). A direct implication of this assumption is that the explanatory variables at a given time period are uncorrelated with the idiosyncratic errors at any given time period:

$$\mathtt{E}\,(u_{it}x_{is}) = 0 \text{ for all } t \text{ and } s. \tag{2.2}$$

Then we have $\mathtt{E}\,(\Delta u_{it}\Delta x_{it}) = 0$ and $\mathtt{E}\,(\dot{u}_{it}\dot{x}_{it}) = 0$ for all t, ensuring the consistency of the FD and FE estimators for β, respectively.

Since the conditions in (2.2) are essential for consistency and the fixed effects are wiped out through transformation, we can consider a test for (2.1) based on the implication of (2.1). Wooldridge (2010) proposes a simple test for strict exogeneity by testing whether $\gamma = 0$ in the following augmented regression:

$$y_{it} = x_{it}'\beta + w_{i,t+1}'\gamma + \alpha_i + u_{it},$$

where $w_{i,t+1}$ is a subset of $x_{i,t+1}$. Clearly, under the null hypothesis of strict exogeneity, $\gamma = 0$ and we can carry out the test using FE estimation. However, since Wooldridge's test only includes a subset of $x_{i,t+1}$, the test may not be able to detect general intertemporal correlation between u_{it} and x_{is} when $|t-s| \geq 2$. To improve the power, we propose to check all possible intertemporal correlations between u_{it} and x_{is} for $|t-s| \geq 1$.

Following the idea of Wooldridge (2010), we consider a sequence of augmented linear panel regressions

$$y_{it} = x_{it}'\beta + x_{i,t+s}'\delta_s + \alpha_i + u_{it},\ s \in \mathcal{S}_T \tag{2.3}$$

where $\mathcal{S}_T \equiv \{-T_2, -T_3 \ldots, -1, 1, \ldots, T_3, T_2\}$.[1] Here $T_a = T - a$ for any positive integer a such that $a \leq T-1$. When $s > 0$, all observations with $t = 1, ..., T-s$ are used in the estimation of β and δ_s in (2.3); similarly, when $s < 0$, all observations with $t = 1-s, ..., T$ are used in the estimation. Under the assumption of strict exogeneity, $\delta_s = 0$ for all $s \in \mathcal{S}_T$. Consequently, we can test the strict exogeneity assumption by testing the null hypothesis

$$\mathbb{H}_0 : \delta_s = 0 \text{ for all } s \in \mathcal{S}_T$$

against the alternative

$$\mathbb{H}_1 : \delta_s \neq 0 \text{ for some } s \in \mathcal{S}_T.$$

Under $\mathbb{H}_0$, $\delta_s = 0$ implies that the idiosyncratic error u_{it} does not include any further information about $x_{i,t+s}$, and thus there is no need to include $x_{i,t+s}$ as regressors in model (1.1).

3 The test statistic

We construct our test statistic based on a sequence of estimators $\hat{\delta}_s$, $s \in \mathcal{S}_T$. One way to check whether all δ_s's being equal to zero simultaneously or not is to consider the following sup-Wald test statistic

$$supW_N = \sup_{s \in \mathcal{S}_T} \left\{ N \hat{\delta}_s' \hat{V}_s^{-1} \hat{\delta}_s \right\}$$

where $\hat{V}_s$ is a data-dependent normalizing matrix, often taken as the estimator of the asymptotic variance of $\sqrt{N}\hat{\delta}_s$, i.e., $\hat{V}_s = \widehat{\mathtt{Avar}}(\sqrt{N}\hat{\delta}_s)$. Under some assumptions to be specified in the next section, we can establish the consistency and asymptotic distribution for $supW_N$.

To state how to obtain $\hat{\delta}_s$ and $\hat{V}_s$, we define $a_{i,t}^{t+d} = (a_{it}, ..., a_{i,t+d})'$ for $d > 0$, where $a = y$, x, or u. Define a series of $T_{|s|} \times 1$ vectors or $T_{|s|} \times k$ matrices as follows,

$$Y_i^{(s)} = \begin{cases} y_{i,1}^{T-s}, s > 0 \\ y_{i,1-s}^{T}, s < 0 \end{cases} \quad X_i^{(s)} = \begin{cases} x_{i,1}^{T-s}, s > 0 \\ x_{i,1-s}^{T}, s < 0 \end{cases} \quad X_{a,i}^{(s)} = \begin{cases} x_{i,1+s}^{T}, s > 0 \\ x_{i,1}^{T+s}, s < 0 \end{cases} \quad \text{and } u_i^{(s)} = \begin{cases} u_{i,1}^{T-s}, s > 0 \\ u_{i,1-s}^{T}, s < 0. \end{cases}$$

Then the model can be rewritten as

$$Y_i^{(s)} = X_i^{(s)}\beta + X_{a,i}^{(s)}\delta_s + \alpha_i \iota_{T_{|s|}} + u_i^{(s)}, \quad s \in \mathcal{S}_T \tag{3.1}$$

[1] First, if we are certain about that $\mathtt{E}(u_{it}x_{is}) = 0$ for $s < t$, then we can set $\mathcal{S}_T = \{1, \ldots, T_3, T_2\}$. This is relevant when we believe that u_{it} affects x_{is} in the future but not in the past, i.e., x_{is} is sequentially exogenous. When $S_T = \{-T_2, -T_3, ..., -1\}$, we test the sequential exogeneity of x_{it} given x_{it} being weak exogenous. In general, S_T can be any subset of $\{-T_2, ..., -1, 1, ..., T_2\}$. Second, as in Wooldridge (2010), we can also replace $x_{i,t+s}$ by a subset $w_{i,t+s}$.

or in a vector form

$$\begin{pmatrix} Y_1^{(s)} \\ Y_2^{(s)} \\ \vdots \\ Y_N^{(s)} \end{pmatrix} = \begin{pmatrix} X_1^{(s)} \\ X_2^{(s)} \\ \vdots \\ X_N^{(s)} \end{pmatrix} \beta + \begin{pmatrix} X_{a,1}^{(s)} \\ X_{a,2}^{(s)} \\ \vdots \\ X_{a,N}^{(s)} \end{pmatrix} \delta_s + \begin{pmatrix} \iota_{T_{|s|}} & 0_{T_{|s|}\times 1} & \cdots & 0_{T_{|s|}\times 1} \\ 0_{T_{|s|}\times 1} & \iota_{T_{|s|}} & \cdots & 0_{T_{|s|}\times 1} \\ \vdots & \vdots & \ddots & \vdots \\ 0_{T_{|s|}\times 1} & 0_{T_{|s|}\times 1} & \cdots & \iota_{T_{|s|}} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{pmatrix} + \begin{pmatrix} u_1^{(s)} \\ u_2^{(s)} \\ \vdots \\ u_N^{(s)} \end{pmatrix}. \tag{3.2}$$

Let D_s denote a $T_{|s|}\times T_{|s|}$ or $T_{|s|+1}\times T_{|s|}$ transformation matrix such that $D_s \iota_{T_{|s|}} = 0$ depending on whether the within-group transformation ($D_s = I_{T_{|s|}} - \iota_{T_{|s|}} \iota'_{T_{|s|}}/T_{|s|}$) or the first difference transformation ($D_s = [0_{T_{|s|+1}\times 1}, I_{T_{|s|+1}}] - [I_{T_{|s|+1}}, 0_{T_{|s|+1}\times 1}]$) is used. Denote $M_{D_s} = \texttt{Diag}\,(D_s, ..., D_s)$, an $NT_{|s|} \times NT_{|s|}$ or $NT_{|s|+1} \times NT_{|s|}$ block diagonal matrix. Pre-multiplying M_{D_s} on both sides of (3.2) leads to

$$M_{D_s} Y^{(s)} = M_{D_s} X^{(s)} \beta + M_{D_s} X_a^{(s)} \delta_s + M_{D_s} u^{(s)},$$

where $Y^{(s)} = (Y_1^{(s)\prime}, ..., Y_N^{(s)\prime})'$, $X^{(s)} = (X_1^{s\prime}, ..., X_N^{s\prime})'$, $X_a^{(s)}$ and $u^{(s)}$ are defined similarly. Denote the OLS estimators for β and δ_s as $\hat{\beta}_{(s)}$ and $\hat{\delta}_s$, respectively. By the formula of partitioned regression, we have

$$\hat{\delta}_s = \left(X_a^{(s)\prime} M^{(s)} X_a^{(s)} \right)^{-1} X_a^{(s)\prime} M^{(s)} Y^{(s)} \text{ for } s \in \mathcal{S}_T,$$

where $M^{(s)} = M'_{D_s} M_s M_{D_s}$, $M_s = I_{n^*} - M_{D_s} X^{(s)} \left(X^{(s)\prime} M'_{D_s} M_{D_s} X^{(s)} \right)^{-1} X^{(s)\prime} M'_{D_s}$, and n^* is the row number of M_{D_s}. Let $\hat{u}^{(s)} = Y^{(s)} - X^{(s)} \hat{\beta}_{(s)} - X_a^{(s)} \hat{\delta}_s - (I_N \otimes \iota_{T_{|s|}})\hat{\alpha}$, where $\otimes$ is the Kronecker product, $\hat{\alpha} = (\hat{\alpha}_1, ..., \hat{\alpha}_N)$ and $\hat{\alpha}_i = \bar{y}_i - \bar{x}'_i \hat{\beta}_{(s)}$. Then we can estimate the asymptotic variance of $\sqrt{N}\hat{\delta}_s$ by $\hat{V}_s = N(X_a^{(s)\prime} M^{(s)} X_a^{(s)})^{-1} X_a^{(s)\prime} M^{(s)} \hat{u}^{(s)} \hat{u}^{(s)\prime} M^{(s)} X_a^{(s)} (X_a^{(s)\prime} M^{(s)} X_a^{(s)})^{-1}$. Note that this estimate allows both conditional (or unconditional) serial correlation and heteroskedasticity of unknown form.

4 The asymptotic properties of the test statistic

To state the main assumptions, we introduce some notation. Denote

$$\Omega = \begin{pmatrix} \Omega_{-T_2,-T_2} & \cdots & \Omega_{-T_2,-1} & \Omega_{-T_2,1} & \cdots & \Omega_{-T_2,T_2} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \Omega_{-1,-T_2} & \cdots & \Omega_{-1,-1} & \Omega_{-1,1} & \cdots & \Omega_{-1,T_2} \\ \Omega_{1,-T_2} & \cdots & \Omega_{1,-1} & \Omega_{1,1} & \cdots & \Omega_{1,T_2} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \Omega_{T_2,-T_2} & \cdots & \Omega_{T_2,-1} & \Omega_{T_2,1} & \cdots & \Omega_{T_2,T_2} \end{pmatrix},$$

where $\Omega_{s,r} = \mathtt{E}[(X_{a,i}^{(s)} - X_i^{(s)}\rho_s)' D_s' D_s u_i^{(s)} u_i^{(r)\prime} D_r' D_r (X_{a,i}^{(r)} - X_i^{(r)}\rho_r)]$ where $\rho_s = [\mathtt{E}(X_i^{(s)\prime} D_s' D_s X_i^{(s)})]^{-1} \times \mathtt{E}(X_i^{(s)\prime} D_s' D_s X_{i,a}^{(s)})$. Define $\Sigma_s = \mathtt{E}[(X_{a,i}^{(s)} - X_i^{(s)}\rho_s)' D_s' D_s (X_{a,i}^{(s)} - X_i^{(s)}\rho_s)]$.

To establish the asymptotic properties of our test statistic, we need the following assumptions.

Assumption 1. (i) (X_i, α_i, u_i) are independent and identically distributed (IID) across i;

(ii) $\mathtt{E}(u_{it} x_{it}) = 0$ for all t.

Assumption 2. (i) $\mathtt{E}(X_i^{(s)\prime} D_s' D_s X_i^{(s)})$ is nonsingular for any $s \in \mathcal{S}_T$;

(ii) Σ_s is positive definite (p.d.) for any $s \in \mathcal{S}_T$.

(iii) Ω is p.d..

Remark 1. Assumption 1(i) is commonly used in the literature on panel data models with fixed T, which rules out the cross-sectional dependence; Assumption 1(ii) assumes contemporary uncorrelation between u_{it} and x_{it}, which means that we test intertemporal correlation between $\{u_{it}\}$ and $\{x_{it}\}$ given x_{it} being contemporarily exogenous. Testing of contemporaneous exogeneity requires existence of instrumental variables for x_{it}, and is beyond the scope of this paper. Assumption 2 imposes moment conditions on x_{it} and u_{it}. Assumption 2(i) ensures that we can project $M_{D_s} X_a^{(s)}$ on $M_{D_s} X^{(s)}$ and then ρ_s is well-defined; Assumption 2(ii) implies that the matrix of augmented regressors, $D_s X_{a,i}^{(s)}$, after removing its linear projection on $D_s X_i^{(s)}$, is still of full column rank; Assumption 2(iii) is a regular second moment condition which is used in the central limit theorem for the IID sequence.

Denote $\hat{\delta} = (\hat{\delta}_{-T_2}', \hat{\delta}_{-T_3}', ..., \hat{\delta}_{-1}', \hat{\delta}_1', ..., \hat{\delta}_{T_3}', \hat{\delta}_{T_2}')'$. Let ξ_s be a $k \times 1$ random vector for $s \in \mathcal{S}_T$ and $\xi = \left(\xi_{-T_2}', \xi_{-T_3}', ..., \xi_{-1}', \xi_1', ..., \xi_{T_3}', \xi_{T_2}'\right)' \sim N(0, \mathbb{V})$, where $\mathbb{V} = \vec{\Sigma}' \Omega \vec{\Sigma}$, $\vec{\Sigma} = \mathtt{Diag}(\Sigma_{-T_2}^{-1}, \Sigma_{-T_3}^{-1}, ..., \Sigma_{-1}^{-1}, \Sigma_1^{-1}, ..., \Sigma_{T_3}^{-1}, \Sigma_{T_2}^{-1})'$. Clearly, we have $\mathtt{Cov}(\xi_s, \xi_r) = \mathbb{V}_{s,r} = \Sigma_s^{-1\prime} \Omega_{s,r} \Sigma_r^{-1}$ for $s, r \in \mathcal{S}_T$.

Theorem 4.1 *Suppose Assumptions 1-2 hold. Under $\mathbb{H}_0$, we have $\sqrt{N}\hat{\delta} \xrightarrow{d} N(0, \mathbb{V})$ and $supW_N \xrightarrow{d} \sup_{s \in \mathcal{S}_T} \left\{\xi_s' \mathbb{V}_{s,s}^{-1} \xi_s\right\}$ as $N \to \infty$.*

Remark 2. The proof is straightforward and relegated to the Appendix. In principle, we reject the null hypothesis if $supW_N > C_\alpha$, where C_α is the α-level critical value from the asymptotic distribution of $supW_N$ under $\mathbb{H}_0$. Note that the covariance matrices $\mathbb{V}_{s,r}$ of ξ_s and ξ_r are generally not equal to zero. So the critical values C_α's depend on the joint distribution of ξ_s's and we cannot tabulate them. Below we will propose a bootstrap method to obtain the p-values instead.

Remark 3. When the null hypothesis is rejected, the test suggests the source for the breakdown of strict exogeneity. Let $k_0 = \text{argmax}_{s \in \mathcal{S}_T} \{N\hat{\delta}_s' V_s^{-1} \hat{\delta}_s\}$. Intuitively, k_0 indicates that the strongest correlation exists between $x_{i,t+k_0}$ and u_{it}, after controlling for x_{it}. Further, we can execute the same testing approach over the set $\mathcal{S}_T \backslash \{k_0\}$ and check whether the strict exogeneity may be violated for other leading or lagged regressors. Continuing in this fashion, we can in principle detect all violations of strict exogeneity by implementing a sequential testing procedure. At the end, we may sort out the set of valid moment conditions which offers basis for consistent GMM estimation of β. We leave the topic for future research.

Remark 4. It is possible to include any subset of $\{x_{i1}, ..., x_{i,t-1}, x_{i,t+1}, ..., x_{iT}\}$ in the augmented regression in (2.3) to improve the power performance. However, as the dimension increases, there will be size distortion especially when T is mildly large in finite samples.

Remark 5. Alternatively, we can construct the test based on either $\sum_{s\in\mathcal{S}_T} N\hat{\delta}_s' V_s^{-1}\hat{\delta}_s$ or $N\hat{\delta}' V^{-1}\hat{\delta}$, where V is a consistent estimator for $\mathbb{V}$. These tests have an advantage that they are less sensitive to outliers. However, the tests cumulate too much estimation error and will have good power only if many $\delta_s \neq 0$ simultaneously. In the case of large T, we can modify these tests to improve the power performance via the technique developed in Fan, Liao and Yao (2015).

Theorem 4.2 *Suppose Assumptions 1-2 hold. Under $\mathbb{H}_1$, we have*

$$\Pr\left(supW_N > C_\alpha | \mathbb{H}_1\right) \to 1 \ \text{as}\ N \to \infty,$$

where C_α is the critical value from the asymptotic distribution of $supW_N$ under $\mathbb{H}_0$.

To implement our test, we propose a bootstrap method to obtain the bootstrap p-value. Following Su and Chen (2013) and Su et al. (2015), we adopted the fixed-regressor wild bootstrap:

1. Under $\mathbb{H}_0$, obtain $\hat{\beta}$ by FE or FD estimation, and then obtain $\hat{\alpha}_i = \bar{y}_i - \bar{x}_i'\hat{\beta}$ and $\hat{u}_{it} = y_{it} - x_{it}'\hat{\beta} - \hat{\alpha}_i$.

2. Let $u_{it}^* = \hat{u}_{it}e_i^*$, where e_i^*'s are IID $N(0,1)$. Generate $y_{it}^* = \hat{\alpha}_i + x_{it}'\hat{\beta} + u_{it}^*$ and obtain the bootstrap sample $\{(y_{it}^*, x_{it})\}$.

3. Obtain the estimator $\hat{\delta}_s^*$ and $\widehat{\texttt{Avar}}(\sqrt{N}\hat{\delta}_s^*)$ based on the bootstrap sample $\{(y_{it}^*, x_{it})\}$. Then calculate $supW_N^* = \sup_{s\in\mathcal{S}_T} N\hat{\delta}_s^{*\prime}[\widehat{\texttt{Avar}}(\sqrt{N}\hat{\delta}_s^*)]^{-1}\hat{\delta}_s^*$.

4. Repeat Steps 2-3 B times to obtain $\{supW_{N,b}^*\}_{b=1}^B$, and obtain the bootstrap p-value: $p^* = B^{-1}\sum_{b=1}^B \mathbf{1}(supW_{N,b}^* > supW_N)$, where $\mathbf{1}(\cdot)$ is the usual indicator function. Reject $\mathbb{H}_0$ if p^* is less than some prescribed significance level.

It is straightforward to implement the above bootstrap procedure. Note that we impose the null hypothesis $\mathbb{H}_0$ implicitly in Step 2. Due to the fact that the observations in bootstrap world are independent but not identically distributed (INID), we add the following assumption:

Assumption 3. (i) For some $r > 0$, $\mathtt{E}\left\|(D_s X_{a,i}^{(s)} - D_s X_i^{(s)} \rho_s)' D_s u_i^{(s)}\right\|^{2+r} < \infty$ for any $s \in \mathcal{S}_T$; (ii) $\mathtt{E}\left\|(D_s X_{a,i}^{(s)} - D_s X_i^{(s)} \rho_s)' D_s X_i^{(s)}\right\|^{2+r} < \infty$ for any $s \in \mathcal{S}_T$ and the same r given in (i).

Remark 6. Assumptions 3(i)-(ii) are two moment conditions imposed on the transformed regressors and errors, which are used to verify the Lyapunov condition in deriving the asymptotic distribution of $supW_N^*$. Note that we impose the conditions on the transformed regressors and errors directly.

The next theorem implies the asymptotic validity of the above bootstrap procedure.

Theorem 4.3 *Suppose Assumptions 1-3 hold. Then* $supW_N^* \xrightarrow{d^*} \sup_{s \in \mathcal{S}_T} \left\{\xi_s' \mathbb{V}_{s,s}^{-1} \xi_s\right\}$ *in probability, where* $\xrightarrow{d^*}$ *denotes weak convergence under the bootstrap probability measure conditional on the observed data set* $\{(y_{it}, x_{it}), i = 1, ..., N, t = 1, ..., T\}$.

5 Monte Carlo simulations

In this section, we carry out a small set of Monte Carlo simulations to examine the finite sample performance of our test.

5.1 Data generating processes

We design the following data generating processes (DGPs) in our experiments:

$$\begin{aligned} y_{it} &= x_{it}\beta + x_{i,t+s}\delta_s + \alpha_i + u_{it}, \\ x_{it} &= -1 + 0.5x_{i,t-1} + \alpha_i + \varepsilon_{it}, \\ u_{it} &= \sqrt{0.1 + 0.25x_{it}^2}(e_{it} + 0.3e_{i,t-1}), \end{aligned}$$

where α_i are IID $N(1, 0.25)$, ε_{it} are IID $N(0,1)$, e_{it} are IID $N(0,1)$, and they are mutually independent of each other; we set the initial value of $\{x_{it}\}$ as $x_{i,1-T} = 0.5$. We fix $\beta = 1$ in both size and power studies and let $\delta_s \in \{0, 0.1, 0.2\}$, where $\delta_s = 0$ and $\neq 0$ are used in the size and power studies, respectively. To avoid confusion, we simply use "$\mathbb{H}_0$" to represent the DGPs in size studies while "$\mathbb{H}_{1,s}(\delta_s)$" (where $\delta_s \neq 0$) in power studies. Note that we allow both conditional heteroskedasticity and serial correlation in $\{u_{it}\}$.

We consider different sample size combinations: $T \in \{5, 8\}$ and $N \in \{100, 200\}$. We compare the performance of our sup-Wald test ($supW$) with that of Wooldridge's Wald test (W). The test statistics are calculated based on FE and FD transformations. We use W_{FE} and $supW_{FE}$ to indicate the tests based on FE estimation, and W_{FD} and $supW_{FD}$ to represent the tests based on FD estimation. In all scenarios we consider 500 replications and 400 bootstrap resamples.

Table 1. Finite sample rejection frequency (%)

	W_{FE}	$supW_{FE}$	W_{FD}	$supW_{FD}$	W_{FE}	$supW_{FE}$	W_{FD}	$supW_{FD}$
	(N,T)=(100,5)				(N,T)=(200,5)			
$\mathbb{H}_0$	6.4	5.4	4.6	5.6	3.4	5.2	6.4	5.4
$\mathbb{H}_{1,-3}(0.1)$	5.0	17.8	6.0	16.6	4.0	26.0	5.4	26.0
$\mathbb{H}_{1,-2}(0.1)$	6.8	25.2	7.2	25.4	6.4	43.2	6.2	47.8
$\mathbb{H}_{1,-1}(0.1)$	9.4	35.2	8.4	36.6	8.6	63.0	9.0	65.2
$\mathbb{H}_{1,1}(0.1)$	51.8	34.8	55.6	35.6	82.8	63.0	82.6	67.2
$\mathbb{H}_{1,2}(0.1)$	6.8	21.8	10.4	23.6	3.6	39.8	13.0	43.0
$\mathbb{H}_{1,3}(0.1)$	9.4	14.6	6.6	16.6	8.6	26.6	6.8	27.6
$\mathbb{H}_{1,-3}(0.2)$	7.6	54.6	7.2	57.4	9.0	84.0	6.2	83.6
$\mathbb{H}_{1,-2}(0.2)$	10.6	71.0	8.4	78.0	11.8	95.6	8.6	98.0
$\mathbb{H}_{1,-1}(0.2)$	16.4	89.4	15.4	88.4	20.4	99.4	18.4	100
$\mathbb{H}_{1,1}(0.2)$	96.2	91.2	97.8	91.8	100	100	100	100
$\mathbb{H}_{1,2}(0.2)$	5.8	77.0	25.2	81.2	6.4	97.2	34.8	98.6
$\mathbb{H}_{1,3}(0.2)$	15.0	58.0	13.0	57.6	21.8	84.6	12.8	83.8
	(N,T)=(100,8)				(N,T)=(200,8)			
$\mathbb{H}_0$	4.6	6.2	4.6	5.8	6.2	6.4	4.2	6.2
$\mathbb{H}_{1,-5}(0.1)$	8.2	21.6	6.6	22.0	7.0	42.4	3.4	39.2
$\mathbb{H}_{1,-3}(0.1)$	10.2	35.0	7.2	37.6	10.2	65.8	4.6	72.8
$\mathbb{H}_{1,-1}(0.1)$	9.8	63.4	10.6	56.2	10.4	93.4	13.4	88.2
$\mathbb{H}_{1,1}(0.1)$	85.6	61.6	83.4	57.2	98.6	91.8	98.2	89.4
$\mathbb{H}_{1,3}(0.1)$	7.4	40.6	8.6	40.4	7.8	67.6	8.4	69.8
$\mathbb{H}_{1,5}(0.1)$	10.8	22.2	6.4	17.6	9.8	40.8	4.6	39.8
$\mathbb{H}_{1,-5}(0.2)$	5.8	70.8	4.6	75.8	6.6	95.8	7.0	96.4
$\mathbb{H}_{1,-3}(0.2)$	14.6	96.0	6.0	98.2	22.6	100	8.6	100
$\mathbb{H}_{1,-1}(0.2)$	16.6	100	25.4	99.6	25.4	100	36.8	100
$\mathbb{H}_{1,1}(0.2)$	100	100	100	100	100	100	100	100
$\mathbb{H}_{1,3}(0.2)$	6.0	94.4	15.6	94.8	7.2	100	24.0	100
$\mathbb{H}_{1,5}(0.2)$	16.6	76.2	5.0	77.8	23.8	95.6	6.2	97.2

5.2 Simulation results

Table 1 presents the finite sample rejection frequency at the 5% nominal level. We summarize the findings as follows. (i) As expected, Wooldridge's test has good power against $\mathbb{H}_{1,1}(0.1)$ and $\mathbb{H}_{1,1}(0.2)$ and it has little or no power against the other alternatives. (ii) The sup-Wald test behaves reasonably well. When the null of strict exogeneity holds, the empirical rejection frequencies are close to the nominal level 5%; when the null is violated, the empirical rejection frequencies are larger than the nominal level and increase rapidly either as δ_s changes from 0.1 to 0.2 or as the sample size N increases. (iii) Both tests $supW_{FE}$ and $supW_{FD}$ have similar performance in size and power studies.

6 An application to agricultural production

In this section, we apply our test to a cross-country panel data set of agricultural production. We collect all the variables from the website of Food and Agriculture Organization of the United Nations (http://faostat3.fao.org/home/E). The data set covers 109 countries (regions) over the years from 2002 to 2007. So $N = 109$ and $T = 6$.

We model the production function in a traditional Cobb-Douglas form as in Mundlak (1961) with panel data. To be specific, we consider the following panel data model with fixed effects

$$\begin{aligned} \log output_{it} &= \beta_1 \log area_{it} + \beta_2 \log pop_{it} + \beta_3 \log govexp_{it} + \alpha_i + u_{it}, \\ i &= 1, ..., 109, t = 2002, ..., 2007, \end{aligned}$$

where the dependent variable $\log output_{it}$ is the logarithm of output for the i-th country or region in the t-th year and is measured by *net production value*; $\log area_{it}$ and $\log pop_{it}$ are two explanatory variables representing inputs *agricultural area* and totally economically *active population in agriculture* in logarithm, respectively; $\log govexp_{it}$ is the *government expenditure in agriculture* in logarithm which is used as a proxy for the investment in agricultural industry, due to the lack of data for capital stock and machinery and the fact that agricultural industry is usually subsidized by a country's government.

We first report the conventional estimation results in Table 2. From this table, we can see that all the inputs are significant, at least at the 10% level. However, the estimates based on different methods are different from each other. First, the estimation results in the first three columns show an inconsistency among the pooled OLS, random effects (RE) and fixed effects (FE) estimators. The differences between the RE and FE estimators indicate the random effects assumption is fragile. The Hausman test statistic (Hausman, 1978) for the random effects assumption is given by 120.59 with a p-value smaller than 0.001. More importantly, we can see a clear difference between the FE and FD estimators. The FD estimator for $\log area_{it}$ is 0.504, which is almost half of that for the FE estimator. This indicates the possible breakdown of strict exogeneity of x_{it} because FE and FD estimators, when strict exogeneity fails, can be both inconsistent and have different probability limits (Wooldridge, 2010).

Next, we conduct both Wooldridge's test and our proposed test in this paper. The results are given in Table 3. From this table, we can see that both our test and Wooldridge's test based on FE estimation reject the hypothesis of strict exogeneity at the 5% level and our test has smaller p-values than Wooldridge's test. Wooldridge's test suggests that the idiosyncratic

Table 2. Estimation results for agricultural production function

	POLS	RE	FE	FD
$\log area_{it}$	$0.267^{***}(0.054)$	$0.533^{***}(0.121)$	$0.953^{**}(0.461)$	$0.504^{*}(0.269)$
$\log pop_{it}$	$0.359^{***}(0.050)$	$0.195^{*}(0.108)$	$0.307^{**}(0.125)$	$0.369^{***}(0.114)$
$\log govexp_{it}$	$0.457^{***}(0.029)$	$0.176^{***}(0.023)$	$0.136^{***}(0.016)$	$0.090^{***}(0.013)$

Note: 1.The numbers in parentheses are cluster-robust standard errors.
2. *,** and *** indicate significance at 10%, 5% and 1% level, respectively.

errors may have a feedback effect on agricultural inputs in the following period. For instance, a detrimental natural disaster in a country may cause the government to subsidize its agricultural infrastructure more in the following year. Our FE-estimation-based sup-Wald test can pick up such an effect successfully.

In contrast, when the FD estimation is in place, Wooldridge's test fails to reject the null of strict exogeneity at the 5% level while our test continues to reject it. We conjecture that this is due to the fact that Wooldridge's test has less power than ours in detecting intertemporal correlation between the regressors and error terms.

Table 3. Bootstrap p-values for testing strict exogeneity

Testing statistic	W_{FE}	$supW_{FE}$	W_{FD}	$supW_{FD}$
p-value	0.045	0.033	0.098	0.023

Note: The number of bootstrap resamples is 4000.

7 Conclusions

In this paper, we provide a practical test for strict exogeneity in linear panel data models, which is easy to implement and robust to unknown serial correlation and heteroskedasticity. The asymptotic properties of the test have been established. A small set of simulations shows that the finite sample performance of our test behaves reasonably well. An application to agricultural production function illustrates its usefulness.

Our test can be used as a diagnostic tool to justify the use of FE or FD estimators in traditional panel data models. Besides, when the null hypothesis is rejected, our test may reveal the sources responsible for the violation of strict exogeneity. This may guide us towards a consistent estimator, for example, by GMM. We leave this for a future research.

References

Baltagi, B.H., 2013. *Econometric Analysis of Panel Data, 5th Edition.* John Wiley & Sons, Ltd.

Boumparis, P., Milas, C., and Panagiotidis, T., 2015. Has the crisis affected the behavior of the rating agencies? Panel evidence from the Eurozone. *Economics Letters* 136, 118-124.

Earnhart, D., 2004. Panel data analysis of regulatory factors shaping environmental performance. *The Review of Economics and Statistics* 86 (1), 391-401.

Fan, J., Liao, Y., and Yao, J., 2015. Power enhancement in high-dimensional cross-sectional tests. *Econometrica* 83 (4), 1497-1541.

Hausman, J. A., 1978. Specification test in econometrics. *Econometrica* 46 (6), 1251-71.

Hsiao, C., 2014. *Analysis of Panel Data, 5th Edition.* Cambridge University Press, Cambridge.

Mundlak, Y., 1961. Empirical production function free of management bias. *Journal of Farm Economics* 43 (1), 44-56.

Papageorgiadis, N., and Sharma, A., 2016. Intellectual property rights and innovation: A panel analysis. *Economics Letters* 141, 70-72.

Su, L., and Chen, Q., 2013. Testing homogeneity in panel data models with interactive fixed effects. *Econometric Theory* 29 (6), 1079-1135.

Su, L., Jin, S., and Zhang, Y., 2015. Specification test for panel data models with interactive fixed effects. *Journal of Econometrics* 186 (1), 222-244.

Wooldridge, J. M., 2010. *Econometric Analysis of Cross Section and Panel Data, 2nd Edition.* The MIT Press, Cambridge, Massachusetts.

APPENDIX

In this appendix we prove the main results in the paper.

Proof of Theorem 4.1. Let $A_{s,N} \equiv X_a^{(s)\prime} M^{(s)} X_a^{(s)}/N$ and $B_{s,N} \equiv X_a^{(s)\prime} M^{(s)} u^{(s)}/\sqrt{N}$. We have $\sqrt{N}\hat{\delta}_s = A_{s,N}^{-1} B_{s,N}$ for $s \in \mathcal{S}_T$ because $\delta_s = 0$ under $\mathbb{H}_0$. By Assumptions 1(i) and 2(i) , we can readily show that $A_{s,N} = \Sigma_s + o_P(1)$ for all $s \in \mathcal{S}_T$. For the term $B_{s,N}$, we have $\hat{\rho}_s = (\sum_{i=1}^N X_i^{(s)\prime} D_s' D_s X_i^{(s)})^{-1} \times \sum_{i=1}^N X_i^{(s)\prime} D_s' D_s X_{a,i}^{(s)} = \rho_s + o_P(1)$ under Assumptions 1(i) and 2(i) by the weak law of large numbers (WLLN). Then

$$\begin{aligned} B_{s,N} &\equiv \frac{1}{\sqrt{N}} X_a^{(s)\prime} M^{(s)} u^{(s)} = \frac{1}{\sqrt{N}} \sum_{i=1}^N \left(D_s X_{a,i}^{(s)} - D_s X_i^{(s)} \hat{\rho}_s \right)' D_s u_i^{(s)} \\ &= \frac{1}{\sqrt{N}} \sum_{i=1}^N \left(D_s X_{a,i}^{(s)} - D_s X_i^{(s)} \rho_s \right)' D_s u_i^{(s)} + (\rho_s - \hat{\rho}_s)' \cdot \frac{1}{\sqrt{N}} \sum_{i=1}^N X_i^{(s)\prime} D_s' D_s u_i^{(s)} \\ &\equiv B_{s,N}^0 + o_P(1). \end{aligned}$$

By the Cramér-Wold device and the central limit theorem (CLT) for IID random sequences, we can show that under Assumptions 1-2

$$B_N^0 = \left(B_{-T_2,N}^{0\prime}, B_{-T_3,N}^{0\prime} \ldots, B_{-1,N}^{0\prime}, B_{1,N}^{0\prime}, \ldots, B_{T_3,N}^{0\prime}, B_{T_2,N}^{0\prime} \right)' \xrightarrow{d} N(0, \Omega).$$

It follows that $B_N = (B_{-T_2,N}', B_{-T_3,N}' \ldots, B_{-1,N}', B_{1,N}', \ldots, B_{T_3,N}', B_{T_2,N}')' \to_d N(0,\Omega)$. In addition, $A_N = (A_{-T_2,N}, A_{-T_3,N}, \ \ldots, A_{-1,N}, A_{1,N}, \ \ldots, A_{T_3,N}, A_{T_2,N}) \to_p (\Sigma_{-T_2}, \Sigma_{-T_3}, \ldots, \Sigma_{-1}, \Sigma_1, \ldots,$

$\Sigma_{T_3}, \Sigma_{T_2})$. By Slutsky's lemma, we have $\sqrt{N}\hat{\delta} \rightarrow_d N(0, \mathbb{V})$. It is straightforward to show that $\hat{V}_s = \mathbb{V}_{s,s} + o_P(1)$. It follows that $supW_N \rightarrow_d \sup_{s \in \mathcal{S}_T} \left\{\xi_s' \mathbb{V}_{s,s}^{-1} \xi_s\right\}$ by the continuous mapping theorem. ■

Proof of Theorem 4.2. Let $\mathcal{S}_T^\dagger = \{s : \delta_s \neq 0 \text{ and } s \in \mathcal{S}_T\}$. Under $\mathbb{H}_1$, $\mathcal{S}_T^\dagger \neq \varnothing$. We have $\hat{\delta}_s = \delta_s + A_{s,N}^{-1} B_{s,N}$ for $s \in \mathcal{S}_T^\dagger$ and $\sqrt{N}\hat{\delta}_s = A_{s,N}^{-1} B_{s,N}$ for $s \in \mathcal{S}_T \backslash \mathcal{S}_T^\dagger$. Then following the proof of Theorem 4.1, we can show that $\sqrt{N}(\hat{\delta} - \delta) \rightarrow_d N(0, \mathbb{V})$. Write

$$supW_N = \max\left\{W_N^0, W_N^\dagger\right\}$$

where $W_N^0 = \sup_{s \in \mathcal{S}_T \backslash \mathcal{S}_T^\dagger}\{N\hat{\delta}_s' \hat{V}_s^{-1} \hat{\delta}_s\}$ and $W_N^\dagger = \sup_{s \in \mathcal{S}_T^\dagger}\{N\hat{\delta}_s' \hat{V}_s^{-1} \hat{\delta}_s\}$. Clearly, $W_N^0 = O_P(1)$. For $W_N^\dagger$, we have

$$\begin{aligned} N\hat{\delta}_s' \hat{V}_s^{-1} \hat{\delta}_s &= \sqrt{N}\left(\hat{\delta}_s - \delta_s + \delta_s\right)' \hat{V}_s^{-1} \sqrt{N}\left(\hat{\delta}_s - \delta_s + \delta_s\right) \\ &= \sqrt{N}\left(\hat{\delta}_s - \delta_s\right)' \hat{V}_s^{-1} \sqrt{N}\left(\hat{\delta}_s - \delta_s\right) + 2\sqrt{N}\delta_s' \hat{V}_s^{-1} \sqrt{N}\left(\hat{\delta}_s - \delta_s\right) + N\delta_s' \hat{V}_s^{-1} \delta_s \\ &= O_P(1) + O_P(\sqrt{N}) + O_P(N) = O_P(N) \end{aligned}$$

for $s \in \mathcal{S}_T^\dagger$. Then we have $W_N^\dagger = O_P(N)$ as $N \rightarrow \infty$ under $\mathbb{H}_1$. It follows that $\Pr(supW_N > C_\alpha | \mathbb{H}_1) \rightarrow 1$ as $N \rightarrow \infty$. ■

Proof of Theorem 4.3. Let $\Pr^*$ denote the probability measure induced by the wild bootstrap conditional on the original sample $\{(x_{it}, y_{it}) : i = 1, ..., N; t = 1, ..., T\}$. Let $\mathtt{E}^*$ and $\mathtt{Var}^*$ denote the expectation and variance with respect to $\Pr^*$. Let $O_{P^*}(1)$ and $o_{P^*}(1)$ denote the probability order under $\Pr^*$; e.g., $b_{NT} = o_{P^*}(1)$ if for any $c > 0$, $\Pr^*(\|b_{NT}\| > c) = o_P(1)$. We will use the fact that $b_{NT} = o_P(1)$ implies that $b_{NT} = o_{P^*}(1)$.

Observing that $y_{it}^* = \hat{\beta}' x_{it} + \hat{\alpha}_i + u_{it}^*$ in the second step, the null hypothesis is maintained in the bootstrap world since $\mathtt{E}^*(x_{it} u_{is}^*) = \mathtt{E}^*(x_{it} \hat{u}_{is} e_i^*) = x_{it} \hat{u}_{is} \mathtt{E}^*(e_i^*) = 0$ for any t and s, and $\delta_s = 0$ for all $s \in S_T$. Then we can rewrite the estimator in the bootstrap world as follows

$$\begin{aligned} \sqrt{N}\hat{\delta}_s^* &= \left(\frac{1}{N} X_a^{(s)\prime} M^{(s)} X_a^{(s)}\right)^{-1} \frac{1}{\sqrt{N}} X_a^{(s)\prime} M^{(s)} Y^{(s)*} \\ &= \left(\frac{1}{N} X_a^{(s)\prime} M^{(s)} X_a^{(s)}\right)^{-1} \frac{1}{\sqrt{N}} X_a^{(s)\prime} M^{(s)} u^{(s)*} = A_{s,N}^{-1} B_{s,N}^*, \end{aligned}$$

where $Y^{(s)*}$, $u^{(s)*}$ and $B_{s,N}^*$ are the bootstrap version of $Y^{(s)}$, $u^{(s)}$ and $B_{s,N}$, respectively. First, we still have $A_{s,N} = \Sigma_s + o_P(1)$, implying that $A_N = (\Sigma_{-T_2}, \Sigma_{-T_3}, ..., \Sigma_{-1}, \Sigma_1, ..., \Sigma_{T_3}, \Sigma_{T_2}) + o_{P^*}(1)$. Second, we make the following decomposition for $B_{s,N}^*$:

$$\begin{aligned} B_{s,N}^* &= \frac{1}{\sqrt{N}} \sum_{i=1}^N \left(D_s X_{a,i}^{(s)} - D_s X_i^{(s)} \rho_s\right)' D_s \hat{u}_i^{(s)} e_i^* + (\rho_s - \hat{\rho}_s)' \cdot \frac{1}{\sqrt{N}} \sum_{i=1}^N X_i^{(s)\prime} D_s' D_s \hat{u}_i^{(s)} e_i^* \\ &= B_{s,N}^{0*} + o_P(1) \cdot B_{s,N}^{1*}. \end{aligned}$$

We complete the proof by showing that (i) $B_N^{0*} \rightarrow_{d^*} N(0, \Omega)$ in probability, where B_N^{0*} is the bootstrap version of B_N^0; and (ii) $B_{s,N}^{1*} = O_{P^*}(1)$. By straightforward moment calculations and Chebyshev inequality, we can readily show $B_{s,N}^{1*} = O_{P^*}(1)$ under Assumption 3.

We are left to show (i). We use the Cramér-Wold device and the Lyapunov CLT for the INID random sequence. Let $\hat{Z}_i^{(s)} = (D_s X_{a,i}^{(s)} - D_s X_i^{(s)} \rho_s)' D_s \hat{u}_i^{(s)}$. Let $a = \left(a'_{-T_2}, ...a'_{-1}, a'_1, ...a'_{T_2}\right)'$ be any $2T_2 k \times 1$ real non-random vector and $\|a\| = 1$, where a_l is $k \times 1$ for $l \in S_T$. Denote $a' B_N^{0*} = \sum_{i=1}^N \sum_{s \in \mathcal{S}_T} N^{-1/2} a'_s \hat{Z}_i^{(s)} e_i^* = \sum_{i=1}^N N^{-1/2} \hat{G}_i e_i^*$, where $\hat{G}_i = \sum_{s \in \mathcal{S}_T} a'_s \hat{Z}_i^{(s)}$. Note that $\mathtt{E}^*(\hat{G}_i e_i^*) = \hat{G}_i \mathtt{E}^*(e_i^*) = 0$ and $\mathtt{Var}^*(\hat{G}_i e_i^*) = \hat{G}_i^2$. We can establish the CLT for $a' B_N^{0*}$ by verifying the Lyapunov condition. Let $\bar{\sigma}_N^2 = N^{-1} \sum_{i=1}^N \hat{G}_i^2$. We have

$$\frac{1}{\bar{\sigma}_N^{2+r}} \sum_{i=1}^N \mathtt{E}^* \left| \frac{\hat{G}_i e_i^*}{\sqrt{N}} \right|^{2+r} = \frac{1}{N^{r/2} \bar{\sigma}_N^{2+r}} \frac{1}{N} \sum_{i=1}^N \left| \hat{G}_i \right|^{2+r} \cdot \mathtt{E}^* \left| e_i^* \right|^{2+r} = N^{-r/2} O_P(1) = o_P(1)$$

provided that $\frac{1}{N} \sum_{i=1}^N \left| \hat{G}_i \right|^{2+r} = O_P(1)$ and $1/\bar{\sigma}_N^{2+r} = O_P(1)$. In fact,

$$\begin{aligned}
\frac{1}{N} \sum_{i=1}^N \left| \hat{G}_i \right|^{2+r} &= \frac{1}{N} \sum_{i=1}^N \left| \sum_{s \in \mathcal{S}_T} a'_s \hat{Z}_i^{(s)} \right|^{2+r} \\
&\leq \frac{1}{N} \sum_{i=1}^N \left(\sum_{s \in \mathcal{S}_T} \left\| \hat{Z}_i^{(s)} \right\|^2 \right)^{1+r/2} \\
&\leq (2T_2)^{r/2} \sum_{s \in \mathcal{S}_T} \frac{1}{N} \sum_{i=1}^N \left\| \hat{Z}_i^{(s)} \right\|^{2+r} \\
&\leq 2^{1+r} (2T_2)^{r/2} \sum_{s \in \mathcal{S}_T} \frac{1}{N} \sum_{i=1}^N \left\| (D_s X_{a,i}^{(s)} - D_s X_i^{(s)} \rho_s)' D_s u_i^{(s)} \right\|^{2+r} \\
&\quad + 2^{1+r} (2T_2)^{r/2} \sum_{s \in \mathcal{S}_T} \frac{1}{N} \sum_{i=1}^N \left\| (D_s X_{a,i}^{(s)} - D_s X_i^{(s)} \rho_s)' D_s X_i^{(s)} \right\|^{2+r} \cdot \left\| \beta - \hat{\beta} \right\|^{2+r} \\
&= O_P(1) + O_P(1) \cdot O_P(1) = O_P(1),
\end{aligned}$$

where the first inequality follows from the Cauchy-Schwarz inequality:

$$\left(\sum_{s \in \mathcal{S}_T} a'_s \hat{Z}_i^{(s)} \right)^2 \leq \sum_{s \in \mathcal{S}_T} \|a_s\|^2 \sum_{s \in \mathcal{S}_T} \left\| \hat{Z}_i^{(s)} \right\|^2 = \sum_{s \in \mathcal{S}_T} \left\| \hat{Z}_i^{(s)} \right\|^2,$$

the second inequality follows from the Jensen inequality, the third inequality follows from the C_r inequality and the fact that $\|AB\| \leq \|A\| \|B\|$ for any two conformable matrices A and B, and the last line follows from the WLLN and Assumptions 1 and 3. In addition, it is easy to show that $\bar{\sigma}_N^2$ converge to a positive number under Assumptions 1-3. Then $B_N^{0*} \to_{d^*} N(0, \Omega)$ in probability and $\sqrt{N} \hat{\delta}^* \to_{d^*} N(0, \mathbb{V})$ in probability. It is easy to show that $V_s^* = V_{s,s} + o_{P^*}(1)$. Then by the continuous mapping theorem, $supW_N^* \to_{d^*} \sup_{s \in \mathcal{S}_T} \left\{ \xi'_s \mathbb{V}_{s,s}^{-1} \xi_s \right\}$ in probability. ■