Economics Letters 147 (2016) 27–31

Contents lists available at ScienceDirect

Economics Letters

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

# 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, Singapore
[b] School of Economics, Renmin University of China, China
[c] School of Economics, Huazhong University of Science and Technology, China

## HIGHLIGHTS

- We provide a practical test for strict exogeneity in linear panel data models with fixed effects when $N$ is large and $T$ is fixed.
- We establish the asymptotic theory of the test, propose a bootstrap procedure for the test and justify its validity.
- Simulations are conducted for our proposed test in comparison with Wooldridge's Wald test.

ARTICLE INFO

*Article history:*
Received 11 June 2016
Received in revised form
8 August 2016
Accepted 10 August 2016
Available online 16 August 2016

*JEL classification:*
C12
C23

*Keywords:*
Bootstrap
Fixed effects
Panel data
Strict exogeneity
Wald test

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.


## 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):

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

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, $\alpha_i$ represents an unobserved individual effect, and $u_{it}$ is the idiosyncratic error. Depending on whether $\alpha_i$ 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 $\alpha_i$ 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 $\alpha_i$, 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

* Correspondence to: School of Economics, Huazhong University of Science and Technology, 1037 Luoyu Road, Wuhan, Hubei, 430074, China.
*E-mail address:* jiewei.econ@gmail.com (J. Wei).

http://dx.doi.org/10.1016/j.econlet.2016.08.012
0165-1765/

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 concludes. Proof of the theorems are relegated to the Appendix A.

Notation. Let $\iota_a$ 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

Let $x_i = (x_{i1}, \ldots, x_{iT})$. The strict exogeneity used in the linear panel data model with fixed effects can be stated as

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

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:

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

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

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

where $\mathcal{S}_T \equiv \{-T_2, -T_3 \ldots, -1, 1, \ldots, T_3, T_2\}$,[1] and $T_a = T - a$ for any positive integer $a$ such that $a \leq T - 1$. When $s > 0$, all observations with $t = 1, \ldots, T - s$ are used in the estimation of $\beta$ and $\delta_s$ in (2.3); similarly, when $s < 0$, all observations with $t = 1 - s, \ldots, T$ 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

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

against the alternative

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

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 $\delta_s$'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 $\xi_{i,t}^{t+d} = (\xi_{it}, \ldots, \xi_{i,t+d})'$ for $d > 0$, where $\xi = y, x$, or $u$. Define a series of $T_{|s|} \times 1$ vectors or $T_{|s|} \times k$ matrices as follows,

$$Y_i^{(s)} = \begin{cases} y_{i,1}^{T-s}, & s > 0 \\ y_{i,1-s}^{T}, & s < 0 \end{cases} \qquad X_i^{(s)} = \begin{cases} x_{i,1}^{T-s}, & s > 0 \\ x_{i,1-s}^{T}, & s < 0 \end{cases}$$

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

Then the model in (2.3) can be rewritten as

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

or in a vector form

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

[1] First, if we are certain about that $\mathrm{E}\,(u_{it}x_{is}) = 0$ 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 $\mathcal{S}_T = \{-T_2, -T_3, \ldots, -1\}$, we test the sequential exogeneity of $x_{it}$ given $x_{it}$ being weak exogenous. In general, $\mathcal{S}_T$ can be any subset of $\{-T_2, \ldots, -1, 1, \ldots, T_2\}$. Second, as in Wooldridge (2010), we can also replace $x_{i,t+s}$ by a subset $w_{i,t+s}$.