Contents lists available at ScienceDirect

Journal of Econometrics

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

Testing error serial correlation in fixed effects nonparametric panel data models[☆]

Carl Green[a,b], Wei Long[c,*], Cheng Hsiao[d]

[a] ISEM, Capital University of Economics and Business, Beijing, 100070, PR China
[b] Department of Economics, Texas A&M University, College Station, TX 77843-4228, USA
[c] Department of Economics, Tulane University, New Orleans, LA 70118, USA
[d] Department of Economics, University of Southern California, Los Angeles, CA 90089-0253, USA

ARTICLE INFO

Article history:
Available online xxxx

JEL classification:
C12
C14

Keywords:
Panel data model
Nonparametric
Test serial correlation
Fixed effects

ABSTRACT

In this paper we consider the problem of testing serial correlation in fixed effects panel data model in a nonparametric framework. Using asymptotic results developed in Su and Lu (2013), we show that our test statistic has a standard normal distribution under the null hypothesis of zero serial correlation. The test statistic diverges to infinity at the rate of $\sqrt{N}$ under the alternative hypothesis that error is serially correlated, where N is the cross sectional sample size. Simulations show that the proposed test works well in finite sample applications.

1. Introduction

Nonparametric and semiparametric methods allow for the estimation of panel data models that impose relatively few assumptions. This flexibility has made these methods increasingly popular among applied researchers. An early paper by Li and Stengos (1996) proposes a method for estimating a fixed effects panel data model that uses standard methods for estimating nonparametric additive models such as the marginal integration method of Linton and Nielsen (1995) or a backfitting method such as in Opsomer and Ruppert (1997) or Mammen et al. (1999). However, this method does not take full advantage of the structure of the model, and several more recent papers introduce methods that use more of this structure. Baltagi and Li (2002) propose a method that uses a series approximation to estimate the regression function. Henderson et al. (2008) introduce an iterative nonparametric kernel estimator and conjecture its asymptotic distribution. This conjecture is confirmed in Li and Liang (2015).

At the same time, parametric dynamic panel models, which allow for the inclusion of lagged dependent variables as regressors, are also becoming more popular. Dynamic panel models are useful not only in applications in which the relationship between the dependent variable and its lagged values is of direct interest, but also in applications in which the lagged dependent variable is an important control variable. For an overview of dynamic panel models, see Baltagi (2008). While parametric dynamic panel models are increasingly popular, until very recently few, if any, estimators for dynamic panel models allowed the lagged dependent variable to enter the regression function nonparametrically. A recent paper by Su and Lu (2013) addresses this gap in the literature. The authors introduce a recursive local polynomial estimation method for fixed effects dynamic panel models. They use methods developed in Mammen et al. (2009) to derive the uniform consistency and asymptotic normality of the estimators under the assumption of zero serial correlation in the idiosyncratic errors.

We propose a test for the null hypothesis zero serial correlation. As argued in Li and Hsiao (1998), testing for serial correlation has long been a standard practice in applied econometric analysis because if the errors are serially correlated, not only an estimator ignoring serial correlation is generally inefficient, it can be inconsistent if the regressors contain lagged dependent variables. Moreover, strong serial correlation is often an indication of omitting important explanatory variables. Hence, testing autocorrelation is important because the choice of an appropriate estimation procedure for a given panel data model crucially depends on the error structure assumed by the model. Often the estimation methods could be considerably simplified if the errors are not autocorrelated. In this paper, we will generalize Li and Hsiao's test for zero

☆ We would like to thank two referees for their helpful comments.
* Corresponding author.
E-mail address: wei.long.cu@gmail.com (W. Long).

http://dx.doi.org/10.1016/j.jeconom.2015.03.011
0304-4076/© 2015 Elsevier B.V. All rights reserved.

error serial correlation in a nonparametric model to a fixed effects nonparametric model.

The remainder of the paper is organized as follows: Section 2 introduces the test statistic for a nonparametric model fixed effects model and derives its asymptotic distribution. Section 3 proposes using a bootstrap method to better approximate the null distribution of the test statistics. Section 4 reports Monte Carlo simulation results to examine the finite sample performance of the proposed test. The proofs of the main results are given in the two Appendices.

2. The nonparametric fixed effects panel data model

We consider the following fixed effects nonparametric panel data model:

$$y_{it} = g(x_{it}) + \mu_i + \nu_{it}, \quad i = 1, \ldots, N; \; t = 1, \ldots, T, \tag{2.1}$$

where $x_{it} = (y_{i,t-1}, \tilde{x}_{it}')'$, $\tilde{x}_{it}$ is of dimension $(d-1) \times 1$ $(d \geq 2)$ vector of explanatory variable that does not contain any lagged value of the dependent variable, μ_i is the fixed effect term.

We are interested in testing the null hypothesis that there is zero first order serial correlation in ν_{it}. That is, we test

$$H_0: \quad E(\nu_{it}\nu_{i,t-1}) = 0.$$

We would like to test H_0 against an alternative that $E(\nu_{it}\nu_{i,t-1}) \neq 0$. However, since we have to first remove the fixed effects μ_i by first difference, the first difference error $\epsilon_{it} \equiv \nu_{it} - \nu_{i,t-1}$ at an MA(1) error structure when ν_{it} is serially correlated, our test statistic will be based on the sample analogue of $E(\epsilon_{it}\epsilon_{i,t-1})$ which equals to zero under H_0. If H_0 is false, ν_{it} is serially correlated, then $E(\epsilon_{it}\epsilon_{i,t-1}) = E[(\nu_{it} - \nu_{i,t-1})(\nu_{i,t-1} - \nu_{i,t-3})] = 2\gamma_2 - \gamma_1 - \gamma_3$, where $\gamma_j = E(\nu_{i,t-j}\nu_{it})$. Thus, our test will have power against the alternative hypothesis that $2\gamma_2 - \gamma_1 - \gamma_3 \neq 0$.

Because ν_{it} is not observable, we need to first estimate the $g(\cdot)$ function in order to estimate ν_{it}. Also, since the fixed effects can be arbitrarily correlated with the regressor x_{it} and there are no instrumental variables available that can take care of the correlation between x_{it} and μ_t, following Henderson et al. (2008) and Su and Lu (2013) we take a first difference to remove the fixed effects:

$$y_{it} - y_{i,t-1} = g(x_{it}) - g(x_{i,t-1}) + \nu_{it} - \nu_{i,t-1}. \tag{2.2}$$

Model (2.2) is an additive model with the restriction that, except for the negative sign in front of the second function, the two additive functions have identical functional forms. Henderson et al. (2008) proposed using a profile likelihood back-fitting method to estimate model (2.2) under the assumptions that x_{it} and ν_{js} are independent with each other for all it and js. Su and Lu (2013) consider a similar dynamic panel data model in which x_{it} contains one lagged dependent variable, $y_{i,t-1}$, and propose to use a local polynomial method to estimate the $g(\cdot)$ function using a back-fitting method. In this paper we will adopt the estimation method proposed by Su and Lu (2013).

Note that x_{it} contains $y_{i,t-1}$ which is correlated with $\nu_{i,t-1}$. However, given that ν_{it} is a serially uncorrelated process, $x_{i,t-1} = (y_{i,t-2}, \tilde{x}_{i,t-1}')'$ is uncorrelated with $\nu_{it} - \nu_{i,t-1}$. Hence, taking the conditional expectation of (2.2) conditional on $x_{i,t-1} = x$, we obtain

$$E(\Delta y_{it}|x_{i,t-1} = x) = E[g(x_{it})|x_{i,t-1} = x] - g(x). \tag{2.3}$$

Let $f_{t,t-1}(z|x)$ denote the conditional density function of x_{it} at $x_{it} = z$ conditional on $x_{i,t-1} = x$ and define $r(x) = -E(\Delta y_{it}|x_{i,t-1} = x)$. Then we can re-write (2.2) as

$$r(x) = g(x) - \int f_{t,t-1}(z|x)g(z)dz \equiv g(x) - (Ag)(x), \tag{2.4}$$

where $(Ag)(x) = \int f_{t,t-1}(z|x)g(z)dz$.

Note that A is a linear operator. Eqs. (2.3) or (2.4) suggest a recursive (back-fitting) method to estimate $g(x)$. For expositional simplicity we will discuss a local constant recursive estimator below; see Su and Lu (2013) for a general local polynomial estimator. Let $\hat{g}_{[l-1]}(x)$ denote the $l-1$ step estimate of $g(x)$. Then the next step estimator is given by

$$\hat{g}_{[l]}(x) = \hat{r}(x) + \hat{E}[g_{[l-1]}(x_{it})|x_{i,t-1} = x], \tag{2.5}$$

where

$$\hat{r}(x) = -\frac{\frac{1}{NT_3}\sum_{j=1}^{N}\sum_{s=4}^{T} \Delta y_{js} K_{j,s-1,x}}{\hat{f}(x)}, \tag{2.6}$$

$$\hat{E}[g_{[l-1]}(x_{it})|x_{i,t-1} = x] = \frac{\frac{1}{NT_3}\sum_{j=1}^{N}\sum_{s=4}^{T} \hat{g}_{[l-1]}(x_{js}) K_{j,s-1,x}}{\hat{f}(x)}, \tag{2.7}$$

$$\hat{f}(x) = \frac{1}{NT_3}\sum_{j=1}^{N}\sum_{s=4}^{T} K_{js,x}, \tag{2.8}$$

where $T_j = T - j$ and $K_{js,x} = K((x_{js} - x)/h) = \prod_{m=1}^{d} k((x_{js,m} - x_m)/h_m)$ is the product kernel function.

The above estimation procedure requires one to use an initial estimator to start the iterative procedure. Following Henderson et al. (2008) and Su and Lu (2013) we use a nonparametric series estimator as an initial estimator. Letting $p(x)$ be a $L \times 1$ vector of series base functions, we use the linear combination of them: $p(x)'\beta$ to approximate $g(x)$, so that the initial estimator of $g(x)$ is given by

$$\hat{g}_{[0]}(x) = p(x)'\hat{\beta} = p(x)'(\tilde{P}'\tilde{P})^{-1}\tilde{P}\Delta Y,$$

where $\tilde{P}$ is a $(nT_3) \times L$ matrix with a typical row given by $p(x_{it})' - p(x_{i,t-1})'$ and ΔY is $(nT_3) \times 1$ with a typical element given by $y_{it} - y_{i,t-1}$.

We define ϵ_{it} and $\hat{\epsilon}_{it}$ as follows:

$$\epsilon_{it} = \nu_{it} - \nu_{i,t-1},$$
$$\hat{\epsilon}_{it} = y_{it} - y_{i,t-1} - (\hat{g}_{it} - \hat{g}_{i,t-1}),$$

where g_{it} denotes $g(x_{it})$.

Then our test statistic I_N is based on the sample analogue of $E(\epsilon_{it}\epsilon_{i,t-2})$ defined as follows:

$$I_N \equiv \frac{1}{NT_3}\sum_{i=1}^{N}\sum_{t=4}^{T} \hat{\epsilon}_{it}\hat{\epsilon}_{i,t-2}. \tag{2.9}$$

We derive the asymptotic distribution of I_N under zero serial correlation in ν_{it} under the following assumptions which are similar to the ones imposed in Su and Lu (2013):

Assumption A1. (i) The random variables (y_i, x_i, μ_i, ν_i), $i = 1, \ldots, N$ are independent and identically distributed across the i index, where $y_i = (y_{i1}, \ldots, y_{iT})$, $x_i = (x_{i1}, \ldots, x_{iT})$, $\nu_i = (\nu_{i1}, \ldots, \nu_{iT})$.
(ii) $(y_{it}, x_{it}, \nu_{it})$ is strictly stationary in t.
(iii) $E[\epsilon_{it}^2|x_{it}] = \sigma_\epsilon^2$.
(iv) Let $f_t(\cdot)$ denote the PDF of x_{it}, and let $\mathcal{D}$ denote its support. We assume that $\mathcal{D}$ is a compact set.
(v) The PDF $f_t(\cdot)$ is uniformly bounded and is bounded below from 0 on its support.
(vi) $E(\nu_{it}|x_{it}, x_{i,t-1}, \ldots, x_{i1}) = 0$ a.s. under H_0.
(vii) $\|g\|_2 < C$ for some $C < \infty$, where $\|g\|_2 \equiv \left(\int g(x)^2 f(x)dx\right)^{1/2}$.
(viii) $\int\int [g(z) - g(x)]^2 f_t(x) f_{t|t-1}(z|x)dxdz > 0$ for $t = 2, \ldots, T$.

Please cite this article in press as: Green, C., et al., Testing error serial correlation in fixed effects nonparametric panel data models. Journal of Econometrics (2015), http://dx.doi.org/10.1016/j.jeconom.2015.03.011