ARTICLE IN PRESS

Journal of Econometrics ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Journal of Econometrics

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

Specification test for panel data models with interactive fixed effects☆

Liangjun Su[a,*], Sainan Jin[a], Yonghui Zhang[b]

[a] School of Economics, Singapore Management University, Singapore

[b] School of Economics, Renmin University of China, China

ARTICLE INFO

Article history:
Received 17 September 2012
Received in revised form
6 January 2014
Accepted 27 June 2014
Available online xxxx

JEL classification:
C12
C14
C23

Keywords:
Common factors
Conditional strong mixing
Cross-sectional dependence
Economic growth
Interactive fixed effects
Linearity
Panel data models
Specification test

ABSTRACT

In this paper, we propose a consistent nonparametric test for linearity in a large dimensional panel data model with interactive fixed effects. Both lagged dependent variables and conditional heteroskedasticity of unknown form are allowed in the model. We estimate the model under the null hypothesis of linearity to obtain the restricted residuals which are then used to construct the test statistic. We show that after being appropriately centered and standardized, the test statistic is asymptotically normally distributed under both the null hypothesis and a sequence of Pitman local alternatives by using the concept of conditional strong mixing that was recently introduced by Prakasa Rao (2009). To improve the finite sample performance, we propose a bootstrap procedure to obtain the bootstrap p-value. A small set of Monte Carlo simulations illustrates that our test performs well in finite samples. An application to an economic growth panel dataset indicates significant nonlinear relationships between economic growth, initial income level and capital accumulation.

1. Introduction

Recently there has been a growing literature on large dimensional panel data models with interactive fixed effects (IFE hereafter). These models can capture heterogeneity more flexibly than the traditional fixed/random effects models by the adoption of time-varying common factors that affect the cross sectional units with individual-specific factor loadings. It is this flexibility that drives the models to become one of the most popular and successful tools to handle cross sectional dependence, especially when both the cross sectional dimension (N) and the time period (T) are large. For example, Pesaran (2006) proposes common correlated effect (CCE) estimation of panel data models with IFE; Bai (2009) proposes principal component analysis (PCA) estimation; Moon and Weidner (2010, 2013) reinvestigate Bai's (2009) PCA estimation and put it in the framework of Gaussian quasi maximum likelihood estimation (QMLE) framework; Su and Chen (2013) consider testing for slope homogeneity in panel data models with IFE. For other developments on this type of models, see Ahn et al. (2001, 2013) for GMM approach with large N and fixed T, Kapetanios and Pesaran (2007) and Greenaway-McGrevy et al. (2012) for factor-augmented panel regressions, Pesaran and Tosetti (2011) for estimation of panel data models with a multifactor error structure and spatial error correlation, Avarucci and Zafaroni (2012) for generalized least squares (GLS) estimation, to name just a few.

Panel data models with IFE have been widely used in economics. Examples from labor economics include Carneiro et al. (2003) and

☆ The authors gratefully thank the co-editor Cheng Hsiao, the associate editor, and two anonymous referees for their many constructive comments on the previous version of the paper. They express their sincere appreciation to Jushan Bai, James Heckman, Hidehiko Ichimura, and Lung-fei Lee for discussions on the subject matter and valuable comments on the paper. They also thank the conference and seminar participants at the 8th Symposium on Econometric Theory and Application (SETA 2012), the 2013 China Meeting of the Econometric Society, the 2013 Tsinghua International Conference in Econometrics, the 2013 Asian Meeting of the Econometric Society (AMES 2013), and Fudan University for their valuable comments. Su and Jin gratefully acknowledge the Singapore Ministry of Education for Academic Research Fund under grant number MOE2012-T2-2-021. Jin also acknowledges the financial support from SMU with grant number C244/MSS11E005. All errors are the authors' sole responsibilities.

* Correspondence to: School of Economics, Singapore Management University, 90 Stamford Road, Singapore, 178903, Singapore. Tel.: +65 6828 0386; fax: +65 6828 0833.

E-mail address: ljsu@smu.edu.sg (L. Su).

http://dx.doi.org/10.1016/j.jeconom.2014.06.018
0304-4076/© 2014 Elsevier B.V. All rights reserved.

Cunha et al. (2005), both of which employ a factor error structure to study individuals' education decision. In macroeconomics, Giannone and Lenza (2010) provide an explanation for Feldstein and Horioka's (1980) puzzle by using IFE models. In finance, the arbitrage pricing theory of Ross (1976) is built on a factor model for assets returns. Bai and Ng (2006) develop several tests to evaluate the latent and observed factors in macroeconomics and finance. Ludvigson and Ng (2009) investigate the empirical risk-return relation by using dynamic factor analysis for large datasets to summarize a large amount of economic information by few estimated factors. Ludvigson and Ng (2011) use factor augmented regressions to analyze the relationship between bond excess returns and macroeconomic factors.

All of the aforementioned papers focus on the linear specification of regression relationship in panel data models with IFE. Recently nonparametric panel data models with IFE have started to receive attention; see Freyberger (2012), Su and Jin (2012), Jin and Su (2013), and Su and Zhang (2013). Freyberger (2012) considers identification and sieve estimation of nonparametric panel data models with IFE when N is large and T is fixed. Su and Jin (2012) extend the CCE estimation of Pesaran (2006) from the static linear model to a static nonparametric model via the method of sieves. Jin and Su (2013) construct a nonparametric test for poolability in nonparametric regression models with IFE. Su and Zhang (2013) extend the PCA estimation of Bai (2009) to nonparametric dynamic panel data models with IFE. Despite the robustness of nonparametric estimates and tests, they are usually subject to slower convergence rates than their parametric counterparts. On the other hand, estimation and tests based on parametric (usually linear) models can be misleading if the underlying models are misspecified. For this reason, it is worthwhile to propose a test for the correct specification of the widely used linear panel data models with interactive effects.

In this paper we are interested in testing for linearity in the following panel data model

$$Y_{it} = m(X_{it}) + F_t^{0\prime}\lambda_i^0 + \varepsilon_{it}, \tag{1.1}$$

where $i = 1, \ldots, N$, $t = 1, \ldots, T$, X_{it} is a $p \times 1$ vector of observed regressors that may contain lagged dependent variables, $m(\cdot)$ is an unknown smooth function, F_t^0 is an $R \times 1$ vector of unobserved common factors, λ_i^0 is an $R \times 1$ vector of unobserved factor loadings, ε_{it} is an idiosyncratic error term. When $m(X_{it}) = X_{it}'\beta^0$ almost surely (a.s.) for some $\beta^0 \in \mathbb{R}^p$, (1.1) becomes the most popular linear panel data model with IFE, which is investigated by Pesaran (2006), Bai (2009), and Moon and Weidner (2010, 2013), among others. These authors consider various estimates for β and (λ_i, F_t) in the model. Asymptotic distributions for all estimators have been established and bias-correction is generally needed.

To motivate our test and study of the nonparametric model in (1.1), we take the economic growth model as an example. Prior to the middle 1990s, almost all empirical cross-country growth studies were based on the assumption that all countries obey a parametric (commonly linear) specification as required by the Solow model or its variants. Several studies conducted in the mid to late 1990s question the assumption of linearity and propose nonlinear alternatives for growth model. For example, in a cross sectional study Liu and Stegnos (1999) employ a partially linear model to uncover the nonlinear pattern that initial income and schooling levels affect growth rates. Recently Su and Lu (2013) and Lee (2014) study economic growth via a dynamic panel data model and find significant nonlinear patterns. The former paper considers the traditional panel data model with only individual fixed effects when N is large and T is fixed; the latter considers large dimensional panel with both individual and time effects when both N and T are large. Given the fact that the linear dynamic panel data model is rejected in either paper, we can consider the following nonparametric panel data model

$$Y_{it} = m(X_{it}) + \alpha_i + f_t + \varepsilon_{it}, \tag{1.2}$$

where α_i and f_t are the usual individual and time fixed effects, Y_{it} is the growth rate of GDP per capita in country i at time period t, X_{it} is a vector that may include the last period economic growth rate ($Y_{i,t-1}$) as well as some economic growth determinants such as initial income level, human capital, and investment as a share of GDP. Obviously, employing the panel data model in (1.2) to growth allows us to control not only the country-specific effects but also the time-specific effects, but its limitation is also apparent. Loosely speaking, (1.2) assumes that the common shocks such as technology shocks, oil price shocks, and financial crises enter the equation through the time-specific effects f_t and have the same effects on all individual countries. This is certainly not the case in reality as a small economy tends to be more vulnerable to such shocks than a large economy. This motivates the use of nonparametric panel data models with IFE in (1.1) in the growth literature. We shall examine whether we can continue to find evidence of nonlinear patterns when the usual additive fixed effects is replaced by the IFE.

More generally, although economic theory dictates that some economic variables are important for the causal effects of the others, rarely does it state exactly how the variables should enter a statistical model. Models derived from first-principles such as utility or production functions only have linear dynamics under some narrow functional form restrictions. Linear models are usually adopted for convenience. A correctly specified linear model may afford precise inference whereas a badly misspecified one may offer seriously misleading inference. When $m(\cdot)$ is a nonlinear function, the previously reviewed parametric methods generally cannot provide consistent estimates for the underlying regression function, and the estimated factor space would be inconsistent too. As a result, tests based on these estimates would be completely misleading. For example, it is very important to determine the number of common factors in factor analysis (e.g., Bai and Ng (2002), Onatski (2009), and Lu and Su (2013)) and to test for additivity versus interactivity in panel data models (e.g., Bai (2009)). But both are generally invalid if they are based on the estimation of a misspecified model. Therefore, to avoid the serious consequence of misspecification, it is necessary and prudent to test for linearity before we embark on statistical inference about the coefficients and factor space.

There have been many tests for linearity or more generally the correct specification of parametric models in the literature. The RESET test of Ramsey (1969) is the commonly used specification test for the linear regression model but it is not consistent. Since Hausman (1978) a large literature on testing for the correct specification of functional forms has developed; see Bierens (1982, 1990), Wooldridge (1992), Yatchew (1992), Härdle and Mammen (1993), Hong and White (1995), Fan and Li (1996), Zheng (1996), Li and Wang (1998), Stinchcombe and White (1998), Chen and Gao (2007), Hsiao et al. (2007), and Su and Ullah (2013), to name just a few. In addition, Hjellvik and Tjøstheim (1995) and Hjellvik et al. (1998) derive tests for linearity specification in nonparametric regressions and Hansen (1999) reviews the problem of testing for linearity in the context of self-exciting threshold autoregressive (SETAR) models. More recently, Su and Lu (2013) and Lee (2014) consider testing for linearity in *dynamic* panel data models based on the weighted square distance between parametric and nonparametric estimates and individual-specific generalized spectral derivative, respectively; Lin et al. (2014) propose a consistent test for a linear functional form in a *static* panel data model with fixed effects. Nevertheless, to the best of our knowledge, there is no available test of linearity for panel data models with IFE.

Please cite this article in press as: Su, L., et al., Specification test for panel data models with interactive fixed effects. Journal of Econometrics (2014), http://dx.doi.org/10.1016/j.jeconom.2014.06.018