Journal of Econometrics 178 (2014) 146–166

Contents lists available at ScienceDirect

Journal of Econometrics

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

Testing a linear dynamic panel data model against nonlinear alternatives

Yoon-Jin Lee

Department of Economics, Indiana University, Wylie Hall 201, 100 South Woodlawn, Bloomington, IN 47405, USA

ARTICLE INFO

Article history:
Available online 16 August 2013

JEL classification:
C4
C3

Keywords:
Conditional heteroskedasticity,
Degenerate U-statistics
Dynamic panel data model
Generalized spectral derivative
Joint limit asymptotics
Linearity
Martingale
Specification testing

ABSTRACT

The most popular econometric models in the panel data literature are the class of linear panel data models with unobserved individual- and/or time-specific effects. The consistency of parameter estimators and the validity of their economic interpretations as marginal effects depend crucially on the correct functional form specification of the linear panel data model. In this paper, a new class of residual-based tests is proposed for checking the validity of dynamic panel data models with both large cross-sectional units and time series dimensions. The individual and time effects can be fixed or random, and panel data can be balanced or unbalanced. The tests can detect a wide range of model misspecifications in the conditional mean of a dynamic panel data model, including functional form and lag misspecification. They check a large number of lags so that they can capture misspecification at any lag order asymptotically. No common alternative is assumed, thus allowing for heterogeneity in the degrees and directions of functional form misspecification across individuals. Thanks to the use of panel data with large N and T, the proposed nonparametric tests have an asymptotic normal distribution under the null hypothesis without requiring the smoothing parameters to grow with the sample sizes. This suggests better nonparametric asymptotic approximation for the panel data than for time series or cross sectional data. This is confirmed in a simulation study. We apply the new tests to test linear specification of cross-country growth equations and found significant nonlinearities in mean for OECD countries' growth equation for annual and quintannual panel data.

1. Introduction

Structural relationships are needed for testing economic theories and evaluating policy. Many economic relationships are dynamic in nature and there has been an increase in interest in dynamic panel data models in recent literature (e.g., Arellano and Bond (1991), Kiviet (1995), Hahn and Kuersteiner (2002), Alvarez and Arellano (2003)). In practice, a linear functional form for dynamic structural economic relationships is often assumed, and there has been a relatively well-established econometric theory for linear panel data models (Baltagi, 2001; Wooldridge, 2002; Arellano, 2003; Hsiao, 2003). However, linearity for panel data models is assumed for convenience. Economic theories usually do not suggest a linear functional form and, in fact, many theories indicate a nonlinear relationship.

While there are a large number of papers on the estimation of panel data models, not much effort has been devoted to functional form specification and evaluation of dynamic panel data models in the literature (Granger, 1996). To fill this gap, the present paper develops a new class of specification tests for the linear functional form in dynamic panel data models with conditional heteroskedasticity of unknown form. By construction, the proposed tests are also applicable to static panel data models, but the motivation and emphasis are tailored toward dynamic panel data models. Various specification tests for panel data models have been proposed in the literature (e.g., Hausman (1978), Hausman and Taylor (1981), Arellano (1990), Arellano and Bond (1991), Liu and Stegnos (1999), Baltagi (1997, 1999), Baltagi (2001, ch. 4), and Metcalf (1996)). Nevertheless, none of these tests are designed to check the functional form specification of panel data models. The most popular specification tests by Hausman (1978) and Hausman and Taylor (1981), for example, check whether unobserved individual effects are correlated with observed explanatory variables, under the maintained assumption of a linear functional form. It would be also desirable if a test for the linear functional form specification precedes Hausman test. Recently, Henderson et al. (2008) proposed bootstrap procedures to determine whether a parametric, semiparametric or nonparametric model is appropriate.

This paper will propose a class of specification tests for linearity in dynamic panel data models, with no prior knowledge of possible alternatives (including functional forms, lag structures and heterogeneity across individuals). The main methodology used characterizes the correct specification of the panel data models by

E-mail address: lee243@indiana.edu.

0304-4076/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jeconom.2013.08.013

the martingale (difference) property of model residuals. The procedure proposed in the paper tests the linear functional form specification of a dynamic panel data model with both a large cross section dimension N and a large time series dimension T. Testing the functional form of panel data models is not trivial. Besides the impact of parameter estimation uncertainty, care is needed to control for unobserved heterogeneity and the difficulty of controlling N and T jointly when allowing them to grow to infinity simultaneously (see, Phillipse and Moon, 1999, for more discussion on the joint limit theory). Panel data with large N and T are particularly informative about the dynamic nature of a panel process, offering additional insights into the data-generating process (see, e.g., Bai et al., 2009; Chen et al., 2012; Li et al., 2011, for more discussion on the motivation of studying the large panel data). The asymptotic analysis and limiting results for such panel data substantially differ from those for time series or cross-sectional data. This paper imposes relatively mild conditions on the relative rates of growth between N and T, thus the proposed tests are applicable to panel data with various size combinations of N and T.

The asymptotic results here are in contrast with those of Hong and Lee (2005), who, in a univariate time series context, derive the asymptotic normality of nonparametric kernel-based tests by requiring the smoothing parameter to grow to infinity as T increases. As shown in this paper, this condition is not required for the asymptotic normality of the nonparametric test statistics in panel contexts, thanks to the benefits of using large N and T panels. Also, in a time series or cross-sectional analysis, it is well known that the asymptotic theory for nonparametric statistics usually provides poor finite sample approximation (e.g., Skaug and Tjøstheim (1993a,b)), because the asymptotically negligible higher order terms depend on the smoothing parameter and are very close in regard to the order of magnitude to the dominant term that determines the asymptotic distribution. Accordingly, the asymptotic approximation of test statistics delicately depends upon the smoothing parameter. The fact that the asymptotic distribution of the proposed nonparametric test statistics in panel data contexts does not depend on the smoothing parameters indicates that the asymptotic theory may provide a reasonable approximation in a panel data context even for moderately small samples. This is an advantage of using panel data. In this sense, the results of the present paper provide a new insight into the nonparametric applications for the panel data analysis. Indeed, simulation study shows that the asymptotic theory works reasonably well for (N, T) as small as $(25, 25)$ and for different combinations of large N and small T, or small N and large T.

Because there exist infinite nonlinear alternatives with potential heterogeneity across individuals, we do not assume a specific common alternative model for all individuals. This flexibility is attributed to the use of individual-specific generalized spectral derivative functions, which can capture both linear and nonlinear serial dependence and heterogeneous dynamics in the conditional mean for each individual. In addition, the individual-specific generalized spectrum enjoys the nice features of spectral analysis. For example, they incorporate information about serial dependence from all lags and can capture cyclical dynamics caused by linear or nonlinear serial dependence for each individual. Thus, this approach can detect a wide variety of heterogeneous misspecifications in both the functional form and lag structure. The proposed tests are particularly useful when the information set contains a large time dimension.

Dynamic economic theory, while having implications for the conditional mean dynamics, is usually silent in regard to the higher order conditional moment dynamics. It is important to develop tests of the functional form for the conditional mean of panel data models that are robust to conditional heteroskedasticity and other higher order time-varying moments of unknown form (e.g., Meghir and Windmeijer, 1999; Cermeño and Grier, 2001). Failure to accommodate conditional heteroskedasticity will cause improper sizes for the tests. As an important feature, the proposed tests are robust to conditional heteroskedasticity and higher order conditional moments of unknown form. The associated asymptotic analysis is non-trivial because we can no longer exploit the implications of serial independence, particularly in a panel data context where the control of unobserved heterogeneity is needed.

Section 2 describes the panel data framework, states the hypotheses of interest, and proposes the individual-specific generalized spectral tests. The asymptotic distribution of the test statistics is derived in Section 3, while Section 4 studies asymptotic power of the test statistics under a general class of alternatives. Section 5 discusses a data-driven lag order selection, and Section 6 contains simulation study. Section 7 applies the new tests to growth empirics and Section 8 contains conclusions. All proofs are given in the mathematical appendix and Supplementary Material. Throughout the paper, the following notations will be used: C denotes a generic bounded constant, $\|\cdot\|$ denotes the Euclidean norm and A^* denotes the complex conjugate of A.

2. Hypotheses of interest and approach

2.1. Hypotheses of interest

The most popular econometric models in panel data literature are the class of linear panel data models with additive individual and time effects (e.g., Baltagi, 2001; Wooldridge, 2002; Arellano, 2003; Hsiao, 2003)

$$Y_{it} = X_{it}'\beta + \alpha_i + \lambda_t + \varepsilon_{it}, \quad i = 1, \dots, N, \ t = 1, \dots, T_i, \tag{2.1}$$

where α_i is a time-invariant individual-specific effect; λ_t is a time-specific effect that is common to all individuals; Y_{it} is a scalar dependent variable; X_{it} is a $k \times 1$ vector of explanatory variables that may contain lagged dependent variables, and current and lagged exogenous variables; and β is a $k \times 1$ parameter vector that is common to all individuals and does not change over time. We observe Y_{it} and X_{it}, but not α_i and λ_t. Both α_i and λ_t can be fixed or random, in the sense that they can be correlated or uncorrelated with the observed explanatory variables X_{it}. The time-specific effect λ_t is used to capture the intercorrelation (i.e., cross-sectional dependence) across i. We allow for unbalanced panel data, which is very common in practice. Eq. (2.1) covers most popular econometric linear dynamic panel data models in the literature (e.g., Holtz-Eakin et al. (1988), Arellano and Bond (1991), Arellano (2003), Hsiao (2003), Hjellvik and Tjøstheim (1999)).

In practice, the specification (2.1) is usually an economic structural model that is derived from an economic theory such as the Rational Expectations. The models or applications that fall within this class of specification have been proposed in the literature and have been widely used in practice. For example, Arellano and Bond (1991) consider the Euler equation model for firm investment and from that derive the dynamic employment equation. MaCurdy (1981, 1985) developed the life-cycle labor supply model under uncertainty, which suggests a rational-expectations solution to the consumer's problem. Altonji (1986) formed the basis of the empirical study of the life-cycle labor supply model and Ziliak (1997) employed this model, within which the empirical performance of instrumental variables estimators with predetermined instruments was examined.

The focus of the panel data literature has been on the inference of the true parameter value β_0, the partial effect of X_{it} on Y_{it}, after controlling for the impact of unobserved heterogeneity. The consistency of $\hat{\beta}$ for β_0 and the validity of the economic interpretation of β_0 as the marginal effects of X_{it} on Y_{it} critically depend upon the linearity with suitable control of the unobserved heterogeneity. In this paper, we are interested in whether the linear functional form in a dynamic panel data model is correctly specified after controlling for the unobserved heterogeneity across