Journal of Multivariate Analysis 130 (2014) 64–89

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Panel data partially linear model with fixed effects, spatial autoregressive error components and unspecified intertemporal correlation

Jianhua Hu [a,b], Fuxiang Liu [a,c,*], Jinhong You [a,b]

[a] School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai, 200433, China
[b] Key Laboratory of Mathematical Economics (SUFE), Ministry of Education, Shanghai, 200433, China
[c] Science College, China Three Gorges University, Yichang, Hubei, 443002, China

ARTICLE INFO

Article history:
Received 12 June 2012
Available online 17 May 2014

AMS 2000 subject classifications:
primary 62H12
secondary 62A10

Keywords:
Panel data
Partially linear model
Fixed effect
Spatial correlation
Temporal correlation
Generalized method of moments

ABSTRACT

This paper considers the estimating problem of a panel data partially linear model with spatial autoregressive errors and fixed effects. In addition, we allow the idiosyncratic errors to be intertemporally correlated. By combining the polynomial spline series approximation, the semiparametric least squares method and the difference based technique, a new generalized moment estimator for the autoregressive parameter of the spatial model is constructed. Its consistency and asymptotic normality are established. In order to avoid the incidental parameter problem, a difference based intertemporal covariance matrix estimator is proposed. Based on the estimated spatially and time-wise correlated error structure, we further construct a weighted difference based semiparametric least squares estimator (WDSLSE) and a weighted difference based polynomial spline series estimator (WDPSSE) for the parametric and nonparametric components of the mean model, respectively. We develop an asymptotic theory for these two estimators, including the asymptotic normality, asymptotic efficiency and convergence rate. In particular, we show that the parametric component estimator has the same asymptotic distribution as that based on completely known spatial autoregressive parameter and intertemporal covariance matrix. Simulation studies demonstrate that our asymptotic theory is applicable for finite samples, and the analysis of a real data set illustrates the usefulness of our developed methodology.

1. Introduction

The pooling of observations on a cross-section of countries, firms, households, patients etc. over several time periods will result in the panel data. Therefore, observations in panel data involve at least two dimensions: a cross-sectional dimension and a time series dimension. Such two-dimensional information set makes the researchers be able to construct complex models and conduct efficient statistical inferences which may not be possible using pure cross-section data or time-series data. With the increased availability of panel data, both theoretical and applied work in panel data analysis have become more popular in the last decade. Based on whether the individual effects being correlated with the observed explanatory variables panel data model can be divided into two classes. One is the random effects panel data model in which the individual effects are random and uncorrelated with the explanatory variables, and the other is the fixed effects panel

* Corresponding author at: Science College, China Three Gorges University, Yichang, Hubei, 443002, China.
E-mail address: liufuxiangst@gmail.com (F. Liu).

http://dx.doi.org/10.1016/j.jmva.2014.05.002
0047-259X/

data model in which the individual effects are not random or the individual effects are random but are correlated with the explanatory variables. If the individual effects are fixed, using random effects panel data modeling method will result in an inconsistent estimator. If the individual effects are random, using the fixed effects panel data modeling method will still result in a consistent estimator although it is not efficient. Therefore, the fixed effects panel data modeling method is more robust than the random effects panel data modeling. Fixed effects panel data modeling has been a hot topic in econometrics since it was proposed. We refer to the book by Baltagi [2] for an overview of the fixed effects panel data modeling and corresponding statistical inference. Recently, due to their flexibility fixed effects panel data semiparametric and structural nonparametric modeling have attracted great interest. For the fixed effects panel data partially linear model, Baltagi and Li [3] proposed difference-based series estimators for the parametric component and the nonparametric components, established the asymptotic normality of the former and derived the convergence rate of the latter. By applying the back fitting method, Fan, Peng and Huang [9] further proposed a profile least squares estimator and a local linear estimator for the parametric and nonparametric components, respectively. They established the asymptotic normality for the former and the mean square error upper bound for the latter. Li, Chen and Gao [25] used the local linear dummy variable approach to remove the fixed effects for the estimating procedure of non-parametric time-varying coefficient models with fixed components. And Henderson, Carroll and Li [10] and Sun, Carroll and Li [33] considered a functional-coefficient version of it.

All those modeling techniques and corresponding statistical inference methods for fixed effects panel data need the assumption that there are no correlation among the individuals. If the individuals are randomly drawn from the population, the cross-sectional independence may not be worried. However, the competition between cross-sectional units, copy-cat policies, network issues, spill-overs, externalities, regional issues, etc. are very common in regional science and urban economics (see Kapoor, Kelejian and Prucha [13]). These issues will result in the sham of the cross-sectional independence. An attractive way of allowing for interdependence between cross-sectional units in empirical models is by means of the so-called spatial methods. In spatial methods, interactions between cross sectional units are typically modeled in terms of some measure of distance between them. The most widely used spatial methods are variants of the ones considered in Cliff and Ord [5]. The book by Anselin and Florax [1] is a good reference for an overview of the estimation and testing of spatial cross-sectional data models. With the increasing availability of spatial panel data it is natural to take the spatial correlation into account for panel data modeling. Recently, several authors have investigated the statistical inference problems of fixed effects panel data models with spatially correlated error components structure. Debarsy and Ertur [6] derived several Lagrange multiplier statistics and the corresponding likelihood ratio statistics to test for spatial autocorrelation in a fixed effects panel data linear model. These tests allow discriminating between the two main types of spatial autocorrelation, namely endogenous spatial lag versus spatially autocorrelated errors. Moscone and Tosetti [27] investigated the estimation of a panel data regression model with spatial autoregressive disturbances, fixed effects and unknown heteroskedasticity. Following the work by Kelejian and Prucha [15], Lee and Liu [22] and others, they adopted the generalized method of moments (GMM) and considered as moments a set of linear quadratic conditions in the disturbances. Kim and Sun [19] studied the robust inference for fixed effects panel data linear models in the presence of heteroskedasticity and unknown spatiotemporal dependence and proposed a general bivariate kernel covariance estimator. Lee and Yu [24] presented a comprehensive summary for some recent developments in fixed effects spatial panel data models and corresponding statistical inference methods.

Although Su [30] and Su and Jin [31] have investigated the estimating problem of spatial cross-sectional data semiparametric (partially linear) regression models, all the work about the fixed effects spatial panel data modeling mentioned above focuses on the parametric models, especially the linear models. The parametric models are easy to make calculation, interpretation and extrapolation. However, many data that arise in variety of disciplines, such as, economics, political science, geography, and epidemiology, require more flexible models and these models are usually hard to be specified in advance. For these scenarios, semiparametric and structural nonparametric regression models are good alternatives which embody a compromise between a general nonparametric framework and a fully parametric specification. In this paper, motivated by an application, we propose a semiparametric fixed effects spatial panel data model, namely the fixed effects panel data partially linear model with a spatially correlated error structure. In addition, we also allow the idiosyncratic errors are intertemporally correlated and the correlation is unspecified in advance. The partially linear model is a realistic and parsimonious candidate if one believes that the relationship between the response and some of the explanatory variables has a parametric form, while the relationship between the response and the remaining explanatory variables may not be known.

A fixed effects panel data partially linear model can be written as

$$\mathbf{Y}_N(t) = \boldsymbol{\mu}_N + \mathbf{X}_N(t)\boldsymbol{\beta} + \mathbf{M}_N(\mathbf{U}_N(t)) + \boldsymbol{\varepsilon}_N(t), \quad t = 1, \dots, T, \tag{1.1}$$

where $\mathbf{Y}_N(t) = (Y_{1t,N}, \dots, Y_{Nt,N})^\tau$ are the responses, $\mathbf{X}_N(t) = (\mathbf{X}_{1t,N}, \dots, \mathbf{X}_{Nt,N})^\tau$ and $\mathbf{U}_N(t) = (U_{1t,N}, \dots, U_{Nt,N})^\tau$ are explanatory variables, $\boldsymbol{\beta}$ is an unknown $p \times 1$ parametric vector, $\mathbf{M}_N(\mathbf{U}_N(t)) = (m(U_{1t,N}), \dots, m(U_{Nt,N}))^\tau$ with $m(\cdot)$ being an unknown function, $\boldsymbol{\mu}_N = (\mu_{1,N}, \dots, \mu_{N,N})^\tau$ are the fixed individual effects, $\boldsymbol{\varepsilon}_N(t) = (\varepsilon_{1t,N}, \dots, \varepsilon_{Nt,N})^\tau$ are the random errors and "τ" denotes a transpose of a matrix or vector. Typically, we choose $\sum_{i=1}^N \mu_{i,N} = 0$ as our identification condition.

We further assume that the random errors $\boldsymbol{\varepsilon}_N(t)$ in model (1.1) follow a spatial and time-wise correlated structure:

$$\boldsymbol{\varepsilon}_N(t) = \lambda \mathbf{W}_N \boldsymbol{\varepsilon}_N(t) + \boldsymbol{\nu}_N(t), \quad t = 1, \dots, T, \tag{1.2}$$

with λ being a scalar cross-sectional autoregressive parameter, $\mathbf{W}_N$ being an $N \times N$ weighting matrix of known constants which does not involve t, $\nu_{it,N}$ is the idiosyncratic error with $\mathrm{Cov}(\boldsymbol{\nu}_{i,N}) = \boldsymbol{\Sigma}$ and $E(\boldsymbol{\nu}_{i_1,N}\boldsymbol{\nu}_{i_2,N}^\tau) = \mathbf{0}$ for $i_1 \neq i_2$ where