Journal of Statistical Planning and Inference 159 (2015) 64–80

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

Estimation of semi-parametric varying-coefficient spatial panel data models with random-effects

Yuanqing Zhang [a,b,c,*], Dianmei Shen [d]

[a] School of Finance and Business, Shanghai Normal University, No.100 Guilin Rd., Shanghai, 200234, PR China
[b] Key Laboratory of Mathematical Economics (SUFE), Ministry of Education, Shanghai 200433, PR China
[c] School of Statistics & Management, Shanghai University of Finance & Economics, Shanghai, 200433, PR China
[d] Shandong University of Science and Technology, Qingdao 266590, PR China

ARTICLE INFO

Article history:
Received 19 March 2014
Received in revised form 20 September 2014
Accepted 3 November 2014
Available online 18 November 2014

Keywords:
Spatial
Panel data
Varying-coefficient

ABSTRACT

This paper studies estimation of a semi-parameter varying-coefficient spatial panel data model with random-effects. Under the assumption of exogenous spatial weighting matrix and exogenous regressors, the unknown parameter and varying-coefficient component are estimated by applying the instrumental variable estimation. Under some sufficient conditions, the proposed estimator for the finite dimensional parameter is shown to be root-N consistent and asymptotically normally distributed. The proposed estimator for the varying-coefficient part is consistent and also asymptotically distributed. Consistent estimators for the asymptotic variance–covariance matrices of both the parametric and varying-coefficient components are provided. Simulation results are provided to illustrate the finite sample behavior of the proposed estimation methods and show that the proposed approach has some value in practice.

1. Introduction

Panel data analysis has received a lot of attention during the last two decades due to applications in many disciplines, such as Economics, Finance and Biology. Panel data could offer researchers extended modeling possibilities as compared to purely time-series or cross-sectional data. They control more information such as unobservable and time invariant individual heterogeneity, and they contain more variation and less collinearity among the variables. The popular linear panel data models examples include Ahn and Schmidt (1995), Baltagi (2005). As Arellano (2003) argues, the field of econometrics of panel data has expanded to cover almost any aspect of econometrics. Therefore, it is not surprising that it has included spatial econometrics as well. Spatial panel models can control for both heterogeneity and spatial correlation, see Anselin and Sheri (1992), the reviews in Anselin (2001), Elhorst (2001, 2003), as well as the recent papers by Baltagi et al. (2007), Baltagi et al. (2008), Baltagi et al. (2009), Kapoor et al. (2007), Lee and Yu (2010) and Baltagi (2005, pp. 197–200), among others. In the context of panel data, it implies the following specification

$$y_{it} = \lambda_0 \sum_{j=1}^{N} w_{ij} y_{jt} + x_{it}^{'} \beta_0 + \alpha_i + \varepsilon_{it}, \quad i = 1, \ldots, N, t = 1, \ldots, T, \tag{1}$$

where y_{it} denotes the dependent variable of individual i in period t, x_{it} denote explanatory variables, α_i denotes the unobserved and time invariant individual effect, w_{ij} denotes the spatial weight between individual i and j, ε_{it} denotes random

* Corresponding author at: School of Finance and Business, Shanghai Normal University No.100 Guilin Rd., Shanghai, 200234, PR China.
E-mail address: zhangyuanqing980@sina.com (Y. Zhang).

http://dx.doi.org/10.1016/j.jspi.2014.11.001

noise, and $\delta_0 = (\lambda_0, \beta_0')'$ denotes the unknown true parameter value. Applications about spatial panel data in economics include hedonic housing equations using residential sales (Bell and Bockstael (2000)), unemployment clustering with respect to different social and economic metrics (Conley and Topa (2002)), and spatial price competition in the wholesale gasoline markets (Pinkse et al. (2002)).

However, Model (1) is a linear specification, parametric statistical inference always necessitates some model assumptions, linearity being among the most convenient. Although their properties are very well established, linear models are often unrealistic in applications. Moreover, mis-specification of the data generation mechanism by a linear model could lead to excessive modeling biases and erroneous conclusions. To achieve greater realism, various attempts have been made to relax these model assumptions and hence widen their applicability, of importance is the varying coefficient models, proposed by Hastie and Tibshirani (1993), which widen the scope of applications by allowing regression coefficients to depend on certain covariates. Brumback and Rice (1998), Fan and Zhang (2000), and Hoover et al. (1998) have given details on novel applications of the varying-coefficient models to longitudinal data. Cai et al. (2000) have provided statistical inferences based on the functional coefficient autoregressive models. For applications in finance, econometrics and environmental study, we refer to the papers by Hong and Lee (2003) and Cai and Tiwari (2000).

Nonparametric modeling does not need any assumption, but it may fail to incorporate some prior information and the resulting estimator of the unknown function tends to lesser precision and incur greater variance. So an intermediate strategy is to apply a semiparametric form, among which semi-parametric varying-coefficient models are widely used. Examples include Zhang et al. (2002) and Li et al. (2002). The semi-parametric varying-coefficient models are not stimulated by the desire of purely mathematical extension, rather they come from the need in practice, for example, Fan and Huang (2005) illustrate the Profile Likelihood method by an application to the Boston housing data set. Cai et al. (2006) apply the two-step estimators to investigate the empirical relation between wages and education, using a random sample of young Australian female workers from the 1985 wave of the Australian Longitudinal Survey. However, none of the studies have been discussed the same type of models in the panel data setting. We wish to specify the following semi-parametric varying coefficient specified spatial panel regression (hereafter SVSPR):

$$y_{it} = \lambda_0 \sum_{j=1}^{N} w_{ij} y_{jt} + x_{it}^{'}\beta_0 + z_{it}^{'}\gamma_0(u_{it}) + \alpha_i + \varepsilon_{it}, \quad i = 1, \dots, N, t = 1, \dots, T, \tag{2}$$

where z_{it} and u_{it} denote explanatory variables, $\gamma_0(.) = (\gamma_{01}(.), \dots, \gamma_{0q}(.))'$ is a q-dimensional vector of unknown coefficient functions. This model generalizes the varying-coefficient model of Fan and Zhang (2000), Cai and Li (2008). This model generalizes the semi-parametric varying-coefficient model of Fan and Huang (2005), Zhang et al. (2002), Li et al. (2002), Cai et al. (2006); Cai (2007) and the spatial panel regression model of Lee and Yu (2007, 2010). This model generalizes semi-parametric spatial model of Sun et al. (2014) to panel data set. Furthermore, this model generalizes the partially specified spatial panel data linear regression model of Ai and Zhang (in press). Due to the curse of dimensionality, we assume, for simplicity, that u is univariate. Model (2) permits the interaction between u and z in such a way that a different level of u is associated with a different linear model. This allows us to examine the extent to which the effects of covariates z vary over different levels of the covariates of u. The main objective of the paper is to propose an approach to estimate δ_0 and $\gamma_0(.)$ consistently under the random-effects assumptions, to establish the asymptotic properties for the proposed estimators and to report on a simulation result.

The paper is organized as follows: Section 2 proposes an estimation of (2); Section 3 establishes the asymptotic properties of the proposed estimators; Section 4 gives some consistent covariance matrices; Section 5 reports some Monte Carlo simulation results and all technical proofs are relegated to the appendices.

2. Estimation

Throughout the paper, we shall consider the model (2) in micro panel i.e. for the case of small and fixed T but large N. There are two problems which will be solved with estimation of the model (2). First is the endogeneity between the spatial term $\sum_{j=1}^{N} w_{ij} y_{jt}$ and the error term ε_{it}. Second is the infinite dimensionality of the unknown parameter $\gamma_0(u_{it})$. We tackle the infinite dimensionality problem by using the sieve method, see Ai and Chen (2003) for details. Specifically, let $p^K(u) = (p_1^{K_1}(u)', \dots, p_q^{K_q}(u)')'$ denote a sequence of known basis functions that can approximate any measurable function arbitrarily well in the functional space endowed by a norm $\|\cdot\|_s$, where $p_l^{K_l}(u) = (p_{l1}^{K_l}(u), \dots, p_{lK_l}^{K_l}(u))'$. Examples of sieves include polynomials, Fourier series, splines, wavelets etc. For each $l = 1, \dots, q$, there exists some constant π_{0l} such that $\gamma_{0l}(u) = p_l^{K_l\prime}(u)\pi_{0l} + \vartheta_{0l}(u)$, with $\vartheta_{0l}(u)$ as the approximation error satisfying $\|\vartheta_{0l}(u)\|_s \to 0$ as $K_l \to \infty$. Denote $\vartheta_0(u) = (\vartheta_{01}(u), \dots, \vartheta_{0q}(u))'$, $\vartheta_{0it} = \vartheta_0(u_{it})$.

We define

$$P(u) = \begin{pmatrix} p_{11}^{K_1}(u) & \dots & p_{1K_1}^{K_1}(u) & 0 & \dots & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & 0 & \dots & 0 & p_{q1}^{K_q}(u) & \dots & p_{qK_q}^{K_q}(u) \end{pmatrix}. \tag{3}$$

The matrix $P(u)$ is a $q \times (K_1 + \dots + K_q)$ matrix.