Journal of Econometrics 172 (2013) 77–89

Contents lists available at SciVerse ScienceDirect

Journal of Econometrics

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

Partial maximum likelihood estimation of spatial probit models[☆]

Honglin Wang [a], Emma M. Iglesias [b,*], Jeffrey M. Wooldridge [c]

[a] Hong Kong Institute for Monetary Research, 55/F, Two International Finance Centre, 8 Finance Street, Central, Hong Kong

[b] Department of Applied Economics II. Facultad de Economía y Empresa. University of A Coruña, Campus de Elviña, 15071. A Coruña, Spain

[c] Department of Economics, Michigan State University, 101 Marshall-Adams Hall, East Lansing, MI 48824-1038, USA

ARTICLE INFO

Article history:
Received 31 October 2009
Received in revised form
17 February 2012
Accepted 13 August 2012
Available online 21 August 2012

JEL classification:
C12
C13
C21
C24
C25

Keywords:
Spatial statistics
Maximum likelihood
Probit model

ABSTRACT

This paper analyzes spatial Probit models for cross sectional dependent data in a binary choice context. Observations are divided by pairwise groups and bivariate normal distributions are specified within each group. Partial maximum likelihood estimators are introduced and they are shown to be consistent and asymptotically normal under some regularity conditions. Consistent covariance matrix estimators are also provided. Estimates of average partial effects can also be obtained once we characterize the conditional distribution of the latent error. Finally, a simulation study shows the advantages of our new estimation procedure in this setting. Our proposed partial maximum likelihood estimators are shown to be more efficient than the generalized method of moments counterparts.

1. Introduction

Most econometric techniques using cross-sectional data are based on the assumption of independence of the observations. When the data are outcomes measured at different geographical locations the assumption of independence is tenuous, especially as economic activities have become more and more correlated over space with the advent of modern communication and transportation improvements. Technological advances in the geographic information system (GIS) make collecting spatial data easier than ever before. Consequently, the possibility of spatial correlation among observations has received more and more attention in a wide range of fields, including regional, real estate, agricultural, environmental, and industrial organization economics (Lee, 2004).

Econometricians have begun to pay more attention to spatial dependence problems in the last two decades, and there have been important advances both theoretical and empirical.[1] The analysis of spatial data starts with an underlying spatial structure generating observed spatial correlations (Anselin and Florax, 1995). There are two popular ways of capturing spatial dependence. The first is in the domain of geostatistics, where the spatial index is continuous (Conley, 1999). The second is to assume that spatial sites form a countable lattice (Lee, 2004). Among lattice models, there are also two types of spatial dependence models that have received the bulk of the attention: the spatial autoregressive dependent variable model (SAR) and the spatial autoregressive error model (SAE). In most applications of spatial models, the dependent variables are continuous, work that has been added by important theoretical results in Conley (1999), Lee (2004), and Kelejian and Prucha (1999, 2001). Nevertheless, there are a handful of applications that address spatial dependence with discrete choice dependent variables

☆ We are grateful to two referees, the Associate Editor, the Co-Editor and participants at the 2009 Econometric Society European Meeting, the 2009 Simposio de Análisis Económico, the 2010 "Brunel Macroeconomic Research Centre"-QASS Conference on Macro and Financial Economics and at seminars at London City University, London School of Economics, Tinbergen Institute, UCL, University Carlos III, University of Essex and University of Exeter for very useful comments. Any remaining errors are our own.

* Corresponding author.
E-mail addresses: hwang@hkma.gov.hk (H. Wang), emma.iglesias@udc.es, iglesia5@msu.edu (E.M. Iglesias), wooldri1@msu.edu (J.M. Wooldridge).

1 Anselin et al. (2004) wrote a comprehensive review about econometrics for spatial models.

0304-4076/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2012.08.005

(Case, 1991; McMillen, 1995; Pinkse and Slade, 1998; Lesage, 2000; Beron and Vijverberg, 2003; Pinkse et al., 2006). The purpose of this paper is to advance the available estimation methods for spatially correlated binary outcomes.

While the analysis can be made more general, we focus on the probit model with spatially correlated data. As is now well known, if we ignore the spatial correlation and construct a pseudo-likelihood function as if we had independent draws, the resulting pooled maximum likelihood estimator (MLE) is, under fairly general conditions, consistent and asymptotically normal, provided the marginal model is correctly specified. Poirier and Ruud (1988) established this result for time series data, and it is pretty clear that it holds, under certain assumptions that restrict the amount of dependence for spatial data. The main drawback to applying the pooled MLE when the observations are dependent is a loss of efficiency. Some authors, for example Robinson (1982), explicitly consider joint maximum likelihood estimation of a nonlinear model with time series data. Unfortunately, in the context of spatially correlated data obtaining maximum likelihood estimators that account for the joint dependence in the data is computationally very demanding.

Rather than taking either extreme — ignoring the dependence in the data or trying to model full joint dependence — middle-ground approaches are possible. For example, Poirier and Ruud (1988) show how to estimate the probit model with dependence in time-series data using generalized conditional moment (GCM) estimators. These estimators are computationally attractive and relatively more efficient than ignoring serial dependence. Generally, nonlinear models with a time series dimension can be estimated by generalized method of moments (GMM). The GMM approach is (asymptotically) more efficient than just using a pooled MLE procedure. However, because time series dependence is ignored in forming moment conditions, GMM estimators still can be considerably less efficient than the joint MLE.

Similar considerations hold for spatially correlated data. Methods that only use information on the marginal distributions — such as Pinkse and Slade's (1998) GMM estimator of the SAE probit model based on the pooled MLE first order conditions — potentially give up much in terms of efficiency compared with a full MLE approach. The motivation for the current paper is that joint MLE is often prohibitively difficult while recognizing that methods based only on marginal distributions will often be too imprecise. Therefore, we propose a middle ground between a pooled probit approach and full maximum likelihood. In particular, we choose to capture spatial dependence by assuming that sites form a countable lattice. Then, we divide the lattice into many small groups (clusters), where the clusters are formed from adjacent observations. The resulting structure is a large number of small clusters. If we can obtain the joint density of the responses within cluster, we can improve upon methods that completely ignore the spatial dependence while arriving at estimation methods much less computationally demanding than joint MLE. We refer to our proposed method as "partial MLE" because we are only using partial joint distributions, not the entire joint distribution.

Because we model spatial correlation only within a cluster, we still need to account for spatial correlation across clusters. This feature is what distinguishes the current setting from a standard panel data setting, where independence across clusters are assumed. To obtain valid inference, we appeal to Conley (1999), who extends Newey and West (1987) to allow for data generated by a countable lattice. Conley (1999) uses metrics of economic distance to characterize dependence among agents, and shows that the GMM estimator is consistent and asymptotically normal under some assumptions similar to time-series data.

The rest of the paper is organized as follows. Section 2 provides a brief overview of popular spatial models with a binary response. Section 3 presents the bivariate spatial probit model. In Section 4, we prove consistency and asymptotic normality of the PML estimator (PMLE) under regularity assumptions, and discuss how to get consistent covariance matrix estimators. Section 5 presents a simulation study showing the advantages of our new estimation procedure in this setting. Finally, Section 6 concludes. The proofs are collected in Appendix A, while the results for the simulation study are provided in Appendix B.

2. Discrete choice models with spatial dependence

It is useful to begin with a brief discussion of general binary response models with spatial dependence. For a draw i, let Y_i be a binary outcome and X_i a $1 \times K$ vector of covariates. Assume that Y_i is generated as

$$Y_i = 1[X_i\beta + \varepsilon_i > 0], \tag{1}$$

where ε_i is an unobserved error and β is a $K \times 1$ parameter vector to be estimated. Regardless of any dependence in the data across i, if ε_i is independent of X_i, then the response probability $P(Y_i = 1|X_i)$ can be obtained if the distribution of ε_i is known. In the case where $\varepsilon_i \sim Normal(0, 1)$, it is well known that $P(Y_i = 1|X_i) = \Phi(X_i\beta)$, where Φ denotes the standard normal cumulative distribution function (cdf). The "marginal probability" can be used, under general assumptions, to consistently estimate β using a pooled MLE procedure — even though the data may not be independent. This is effectively the insight of the Poirier and Ruud (1988) results for time series data.

Allowing explicitly for spatial correlation of the kind that is popular for linear models raises a couple of important issues, as recognized in Pinkse and Slade (1998). First, the variance of the error in such models typically depends on the distances among pairs of observations in the lattice — via the matrix that is used in a weighted least squares analysis. Let W denote and $N \times N$ matrix of weights that are exogenous in the sense that

$$\varepsilon_i|X, W \sim Normal(0, h_i(W, \lambda)), \tag{2}$$

where $(h_i(W, \lambda)) > 0$ is a variance function that depends on λ. The form of $h_i(\cdot)$ differs across spatial models and is not yet important. The exogeneity assumption is embodied in the requirement $E(\varepsilon_i|X, W) = 0$, which also imposes a strict exogeneity assumption on the covariates X.

If we maintain (2) along with (1) then $D(Y_i|X, W)$ follows a so-called heteroskedastic probit model with

$$P(Y_i = 1|X, W) = \Phi\left[X_i\beta/\sqrt{h_i(W, \lambda)}\right]. \tag{3}$$

Under sufficient regularity conditions — mainly restricting the amount of spatial dependence — β and λ can be consistently and $\sqrt{n}$-asymptotically normally estimated by using a pooled heteroskedastic probit approach. These moment conditions are used in the Pinkse and Slade (1998) GMM estimator.

Before we proceed further, the presence of W in (3) raises a question about how we should summarize the partial effects of the elements of X_i on the response probability. The notion of the average structural function (ASF), proposed by Blundell and Powell (2004) in a different context, seems useful. In the present application, the ASF is defined as

$$ASF(x) = E_W\left\{\Phi\left[x\beta/\sqrt{h_i(W, \lambda)}\right]\right\}. \tag{4}$$

The average partial effects are obtained by taking changes or partial derivatives of $ASF(x)$. Given consistent estimators $\hat{\beta}$ and $\hat{\lambda}$, $ASF(x)$ can be (under regularity conditions) consistently estimated by

$$n^{-1}\sum_{i=1}^{n}\Phi\left[x\hat{\beta}/\sqrt{h_i(W, \hat{\lambda})}\right]. \tag{5}$$