Journal of Econometrics 159 (2010) 99–115

Contents lists available at ScienceDirect

Journal of Econometrics

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

GMM estimation of social interaction models with centrality[☆]

Xiaodong Liu[a], Lung-fei Lee[b,*]

[a] University of Colorado, Boulder, CO 80309, United States

[b] Ohio State University, Columbus, OH 43210, United States

ARTICLE INFO

Article history:
Received 28 January 2009
Received in revised form
13 April 2010
Accepted 27 April 2010
Available online 11 May 2010

JEL classification:
C13
C21

Keywords:
Social network
Centrality
Spatial autoregressive model
GMM
Efficiency

ABSTRACT

This paper considers the specification and estimation of social interaction models with network structures and the presence of endogenous, contextual, correlated, and group fixed effects. When the network structure in a group is captured by a graph in which the degrees of nodes are not all equal, the different positions of group members as measured by the Bonacich (1987) centrality provide additional information for identification and estimation. In this case, the Bonacich centrality measure for each group can be used as an instrument for the endogenous social effect, but the number of such instruments grows with the number of groups. We consider the 2SLS and GMM estimation for the model. The proposed estimators are asymptotically efficient, respectively, within the class of IV estimators and the class of GMM estimators based on linear and quadratic moments, when the sample size grows fast enough relative to the number of instruments.

1. Introduction

This paper studies social interaction models with network structures. The model considered has the specification of a spatial autoregressive (SAR) model but has features and implications directly relevant to social interaction issues. With such a specification, the information on network structures is usually summarized in the spatial weights matrix, also known as the sociomatrix (or adjacency matrix), in social interaction models.

A general social interaction model not only allows possible endogenous interactions, but also exogenous interactions, unobserved group effects, and correlation of unobservables. Identification of the endogenous interaction effect from the other effects is a major interest in social interaction models (see, e.g., Manski, 1993; Moffitt, 2001). Linear regression models with endogenous interaction based on rational expectations of the group behavior would suffer from the 'reflection problem' of Manski (1993), and the various interaction effects cannot be separately identified. Lee (2007b) considers a group setting where an individual is equally influenced by all the other members in the group and the average outcome of peers represents the source of the endogenous effect. Lee's (2007b) social interaction model is identifiable only if there is variation in group sizes in the sample. The reason for the possible identification is that individuals in a small group will have stronger endogenous interactions than those in a larger group. The identification, however, can be weak if all of the groups have large sizes, even if there is group size variation. The sociomatrix in Lee's (2007b) model has zero diagonal and all of its off-diagonal entries take the value of $\frac{1}{m-1}$, where m is the group size. Such a sociomatrix represents a rather restrictive network structure, but may be practical when there is no information on how individuals interact with each other.

In some data sets, one may have information on network structures. Based on a specific network structure, the (i, j) entry of the sociomatrix is one if i is influenced by j, and zero otherwise. The corresponding sociomatrix represents a directed graph with a directed edge leading from j to i if j affects i.[1] Such a directed-graph sociomatrix has been considered in Lee et al. (2010), where it is row-normalized such that each row sums to unity. For SAR models in empirical studies, row-normalized spatial weights matrices are

☆ We thank John Chao and Ingmar Prucha for helpful discussions. We are also grateful to the editor, the associate editor, three referees, and seminar participants at the II World Conference of the Spatial Econometrics Association in New York, the Far East and South Asia Meeting of the Econometric Society in Tokyo, Yale University, and University of Maryland for valuable comments and suggestions.

* Corresponding author. Tel.: +1 614 292 5508; fax: +1 614 292 4192.
E-mail address: lflee@econ.ohio-state.edu (L.-f. Lee).

[1] Note that the influence may or may not be reciprocal, so the sociomatrix could be asymmetric.

0304-4076/$ – see front matter
doi:10.1016/j.jeconom.2010.04.009

typical with a few exceptions,[2] because the spatial effect can be interpreted as a (weighted) average of neighborhood effects. The social interaction models based on expected group means in Brock and Durlauf (2001) and Manski (1993), and the one in Lee (2007b), all have the endogenous effect being an average of peers' outcomes.

The row-normalized sociomatrix in Lee et al. (2010) has some limitations. First, it implicitly rules out the possibility that an individual's outcome might affect peers' outcomes but he/she might not be affected by peers.[3] In addition, for social interaction studies, one may be interested in the aggregate influence of an individual's peers instead of the average influence. One may also be interested in how an individual's position in a network would influence peers' behavior. Notions such as prestige and centrality have received attention in network studies (Wasserman and Faust, 1994). When the social interaction is specified as a SAR model, the measure of centrality in Bonacich (1987) comes out naturally in the reduced form equation. If the sociomatrix represents the directed graph mentioned above, the sum of the ith row is the indegrees (the number of inward directed edges) of the node i in the graph. All group members (nodes) would have the same level of centrality by the Bonacich measure if and only if the indegrees of all nodes are equal. Thus, if the indegrees have a non-zero variation, so does the Bonacich centrality measure for the group members. The variation in the Bonacich centrality measure helps to identify the various interaction effects. Yet, row-normalization would eliminate the variation in the Bonacich centrality measure. So, for social network studies, sometimes a sociomatrix without row-normalization would be appropriate. In this paper, we study the identification and estimation of network effects without requiring row-normalization of the sociomatrix.

Similar to the model in Lee et al. (2010), the social interaction model in this paper has the specification of a SAR model and incorporates endogenous, exogenous, correlated, and unobserved group effects. The unobserved group effect is captured by a group dummy variable, which is allowed to have a conditional mean that depends on the exogenous variables and/or the sociomatrices (due to self-selection), and so it is treated as a fixed effect. With many groups in the sample, the group dummies may induce the incidental parameter problem as in Neyman and Scott (1948). Based on a transformed model that has the group dummies eliminated, Lee et al. (2010) has generalized the ML estimation approach in Lee (2007b) to the network model with a sociomatrix having constant row sums (including the special case of a row-normalized sociomatrix).[4] However, when the sociomatrix does not have constant row sums, the likelihood function for the transformed model could not be derived, and alternative estimation approaches need to be considered.

This paper considers the 2SLS and generalized method of moments (GMM) estimation approaches. The 2SLS approach has been proposed for the estimation of SAR models in Kelejian and Prucha (1998). The GMM method has been considered for the estimation of a spatial process in Kelejian and Prucha (1999), and SAR models in Lee (2007c) and Lee and Liu (2010). The 2SLS and GMM approaches can be generalized for the estimation of social network models. When the sociomatrix is not row-normalized and the indegrees of its nodes are not all equal, the Bonacich centrality measure for each group can be used as an additional IV to improve estimation efficiency. The number of such instruments depends on the number of groups. If the number of groups grows with the sample size, so does the number of IVs. We show that the proposed 2SLS and GMM estimators can be consistent and asymptotically normal, and they can be efficient when the sample size grows fast enough relative to the number of instruments. We also suggest bias-correction procedures for both estimators based on the estimated leading order many-instrument biases.

Since Bekker's (1994) seminal work, the study of many-instrument asymptotics, where the number of instruments increases with the sample size, has attracted a lot of attention in the IV estimation literature. Some recent developments in this area include Anderson et al. (2007), Donald and Newey (2001), Hansen et al. (2008) and van Hasselt (2010), to name a few. In particular, Bekker and van der Ploeg (2005) has considered IV estimation of a model where group indicators are used as (dummy) instruments and the number of groups goes to infinity. In this paper, we also consider many-group asymptotics, where the number of instruments depends on the number of groups. However, the instruments based centrality measures are not dummy variables. Our model also relaxes the i.i.d. assumption for observations within a group in Bekker and van der Ploeg (2005) by allowing for possible spatial (or social) correlation among group members. Similar to Donald and Newey (2001), we focus on the case where the number of instruments grows with, but at a slower rate than, the sample size.[5] Another important direction of research in the IV estimation literature is on weak instruments or weak identification (see, e.g., Chao and Swanson, 2005, 2007). In this paper, we assume the concentration parameter grows at the same rate as the sample size.[6] Hence, we restrict our attention to scenarios where instruments are stronger than assumed in the weak-instrument literature.

The rest of the paper is organized as follows. Section 2 presents the network model and suggests a transformation of the model to eliminate group fixed effects. Sections 3 and 4 propose the 2SLS and GMM approaches for the estimation of the model. We prove consistency of the proposed estimators, derive the asymptotic distributions, and suggest bias correction procedures for the many-instrument bias. The detailed proofs are given in Appendix C. Monte Carlo evidence on the small sample performance of the proposed estimators is given in Section 5. Section 6 briefly concludes.

2. The network model with group fixed effects

The model considered has the specification

$$Y_r = \lambda_0 W_r Y_r + X_{1r}\beta_{01} + W_r X_{2r}\beta_{02} + l_{m_r}\alpha_{0r} + u_r, \tag{1}$$

and $u_r = \rho_0 M_r u_r + \epsilon_r$, for $r = 1, \ldots, \bar{r}$, where $\bar{r}$ is the total number of groups in the sample, m_r is the number of individuals

2 An exception is argued in Bell and Bockstael (2000) for real estate problems with micro-level data. Kelejian and Robinson (1995) has argued that the parameter space should be free except some singularity points for the spatial matrix—and discussed not-row-normalized spatial matrices. More recently, Kelejian and Prucha (2007) consider implications on the parameter space of the SAR model when the spatial matrix is not row-normalized.

3 In this case, the corresponding row in the sociomatrix will have all zeros and cannot be normalized to sum to unity.

4 The resulting likelihood function can be shown to be a partial likelihood function under normal disturbances. The notion of partial likelihood is introduced in Cox (1975); see also Lancaster (2000).

5 Under the asymptotic sequence that the number of instruments increases at the same rate as the same size, the asymptotic distribution of IV-based estimators has been established by Bekker (1994), Bekker and van der Ploeg (2005), Hansen et al. (2008) and van Hasselt (2010). However, their CLTs assume independent observations and might not be easy to modify for the case with (spatially) correlated observations without imposing strong regularity conditions.

6 This condition on the concentration parameter is implied by Assumption 4 in Section 3. The assumption of independent observations is omnipresent in the literature of weak instruments. This model allows the observations within a group to be correlated. The analysis of asymptotic properties of IV estimators in the presence of weak instruments in a model with correlated observations is a difficult problem, which is beyond the scope of this paper.