

A note on estimating network dependence in a discrete choice model

JING ZHOU^{*}, DA HUANG[†], AND HANSHENG WANG[‡]

Discrete choice model is probably one of the most popularly used statistical methods in practice. The common feature of this model is that it considers the behavioral factors of a person and the assumption of independent individuals. However, this widely accepted assumption seems problematic because human beings do not live in isolation. They interact with each other and form complex networks. Then the application of discrete choice model to network data will allow for network dependence in a general framework. In this paper, we focus on a discrete choice model with probit error which is specified as a latent spatial autoregressive model (SAR). This model could be viewed as a natural extension of the classical SAR model. The key difference is that the network dependence is latent and unobservable. Instead, it could be measured by a binary response variable. Parameter estimation then becomes a challenging task due to the complicated objective function. Following the idea of composite likelihood, an approximated paired maximum likelihood estimator (APMLE) is developed. Numerical studies are carried out to assess the finite sample performance of the proposed estimator. Finally a real dataset of Sina Weibo is analyzed for illustration purpose.

AMS 2000 SUBJECT CLASSIFICATIONS: Primary 91B72.

KEYWORDS AND PHRASES: Approximated paired maximum likelihood estimation, Discrete choice model, Social network, Spatial autocorrelation.

1. INTRODUCTION

Discrete choice model has been extensively studied in the past few decades [10, 13, 17, 6, 14, 9]. In general, choice

model is derived in a random utility framework in which decision makers are assumed to be utility maximizers. The utility is usually decomposed into a deterministic part which is observable and a stochastic component that cannot be observed by analysts [10]. Then the specification of a choice model depends on how the distribution is defined about the stochastic term. For example, we have the popularly used logit and probit models. The common feature of a discrete choice model is that it considers the behavioral factors of a person and the assumption of independent human beings. However, this widely accepted assumption seems problematic because human beings do not live in isolation. People in the society interact with each other and form complex networks. Their behaviors are easily influenced by the connected neighbors. Then the application of discrete choice models to network data should allow for network dependence.

By network dependence, we refer to the fact that individuals living in the same social network should be correlated with each other. One typical example about network dependence could be peer-effects or social interaction. There has been a large body of recent researches considering the role of social interaction in economic behavior. For example, [5] generalize the logistic models by incorporating deterministic social utility in the utility function, which reflects the desire of individuals to conform to the behavior of others. [16] empirically test the discrete choice model with social interaction, using U.S. presidential election data to analyze voters' voting behavior. [12] use physician's self-reported relational data to investigate the impact of social network on the prescription of a new drug. They find that adoption is affected by peers' usage volume. [19] first explore the role of social interaction in customer defection behavior with a cellular company. The results show that a customer will have a higher defection rate if he is exposed to a majority of defected neighbors. A recent paper by [25] claim that network-based credit scoring primarily depends on first-order relationship information. Although these researches did not directly model network dependence, they state that individual's decision is socially affected by others' actions.

Modeling the network dependence for socially connected continuous variables has been considered in existing literatures. Two classical models have been popularly used. One is the spatial autoregressive model (SAR) and the other one is spatial error model (SEM). More sophisticated extensions do exist, such as spatial autoregressive moving average

^{*}Supported in part by National Natural Science Foundation of China (Grant No. 71702185), the fund for building world-class universities (disciplines) of Renmin University of China.

[†]Corresponding author. Supported in part by China Statistical Research (CSR, 2015LY77), the Youth Innovation Team Project for Arts and Social Science of Fudan University, and National Natural Science Foundation of China (NSFC, 71531006, 11571080, 11571081, 71672042).

[‡]Supported in part by National Natural Science Foundation of China (Grant No. 71532001, 11525101), China's National Key Research Special Program (Grant No. 2016YFC0207700) and Center for Statistical Science at Peking University.

model, matrix exponential spatial specification model, spatial Durbin Model and others [15]. Excellent overviews can be found in the handbook edited by [1], [15] and [8]. The focus of these models is to estimate the so-called spatial autocorrelation coefficient, which in our case quantifies the network dependence between connected individuals. To apply these models, we need to have the continuous responses to be directly observed. However, in a discrete choice model, the network dependence is modeled by a latent variable, which is not directly observed by the researcher [22]. Instead, the researcher observes a binary response of individual choice. This specification in the discrete choice model brings computational difficulty for estimation.

To overcome the computation challenge in spatial probit models, many methods have been proposed. Among them, the composite marginal likelihood (CML) is a popular technique. It was first introduced by [11] in the analysis of binary data in the field of geostatistics. This method is based on pairwise likelihood and can be applied to large datasets. Theoretical properties of the estimators are also investigated by the authors. [3] then proposes a maximum approximate CML in the estimation of multinomial probit (MNP) choice models. His method sheds light on the estimation for a variety of MNP models. An alternative equivalent approach to CML is the partial maximum likelihood method proposed by [24]. The CML method is effective in solving the high-dimensional numerical integration problem. However, it is infeasible in large datasets due to the inverting of high-dimensional precision matrix. The partial maximum likelihood method assumes that observations can be divided by pairwise groups and within each group, a bivariate normal distribution is specified. More methods about computational issues can be found in recent reviews of [4] and [18].

In this work, we focus on a discrete choice model with a probit error, which is specified as a latent SAR model. By doing so, we are capable of considering network dependence between individuals under a principled binary regression framework. To estimate the unknown parameter (e.g. spatial autocorrelation), we propose a pseudo maximum likelihood approach. Specifically, for a total of n subjects, we form them into different pairs and each pair only contains two nodes denoted by i and j . This leads to a total pair of $n(n-1)/2$. For a specific pair $\{i, j\}$, we can obtain an approximated log likelihood function by Taylor's expansion. By treating different pairs as independent samples, we can sum together the paired objective functions for selected pairs. However, there exists two comparatively natural choices of pairs. One way is to consider all pairs, which is computationally too expensive and cannot be used in practice. The other one is to consider connected pairs, for they are most likely to be correlated and thus provide best information. Therefore, we sum together all connected paired objective functions according to the idea of composite likelihood [21, 24] and then maximize the summation with respect to spatial autocorrelation. The resulting estimator is then referred to as an approximated paired maximum like-

lihood estimator (APMLE). It should be noted that instead of the SEM specification, which was employed in [24], our specification follows an SAR framework. We show the consistency of APMLE using extensive numeric studies and a real dataset analysis about Sina Weibo is also presented for illustration purpose.

The rest of the article is organized as follows. Section 2 presents the model setup and the difficulties associated with true likelihood function. This motivates us to develop the APMLE. To demonstrate its finite sample performance, extensive numerical studies based on various network topologies are conducted in Section 3. We then use a real dataset to examine the spatial autocorrelation in Section 4. Lastly, the article is concluded with a brief discussion and suggestions for further studies in Section 5.

2. THE METHODOLOGY

2.1 Model setup

We consider a network with n nodes (i.e., consumers) indexed by $1 \leq i \leq n$. Its network structure is captured by an adjacency matrix $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ where $a_{ij} = 1$ if node i follows node j and $a_{ij} = 0$ otherwise. For an undirected network we have $a_{ij} = a_{ji}$. However, throughout the rest of this article, we consider a directed network structure, which allows $a_{ij} \neq a_{ji}$. Because self loop is not allowed for most social network applications, we require $a_{ii} = 0$ for every $1 \leq i \leq n$. Let $Y_i \in \{0, 1\}$ be the binary response variable collected from the i th ($1 \leq i \leq n$) subject. Then to establish a linear regression model to the binary response, we usually suppose there is a latent continuous variable Z_i which satisfies

$$(1) \quad Y_i = \begin{cases} 1 & \text{if the latent } Z_i \geq \mu, \\ 0 & \text{if the latent } Z_i < \mu, \end{cases}$$

where μ is an unknown threshold that needs to be estimated. Next, we assume Z_i follows a SAR process, which could be modeled as

$$(2) \quad \mathbb{Z} = \rho W \mathbb{Z} + \varepsilon,$$

where $\mathbb{Z} = (Z_1, \dots, Z_n)^\top \in \mathbb{R}^n$ is the latent continuous response vector. It should be noted that various models may be applied for probit error specification. The adoption of a specific setting depends on the nature of problem. In this paper, we focus on spatial autoregressive framework. Without any doubt, other specifications are also extremely useful and should be considered for future studies. $W = (w_{ij}) \in \mathbb{R}^{n \times n}$ with $w_{ij} = a_{ij}/d_i$, where d_i is the out-degree for each node and $d_i = \sum_{j=1}^n a_{ij}$. So W is the row normalized adjacency matrix. Finally $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^\top \in \mathbb{R}^n$ is the residual vector with mean 0 and covariance $I \in \mathbb{R}^{n \times n}$. Here I stands for a $n \times n$ identity matrix. The parameter $\rho \in \mathbb{R}^1$ captures the strength of spatial autocorrelation. By (2), we know that $\mathbb{Z} = (I - \rho W)^{-1} \varepsilon$. This implies \mathbb{Z} follows a normal distribution with mean 0 and covariance

$$(3) \quad \Sigma = (I - \rho W)^{-1}(I - \rho W^\top)^{-1},$$

subsequently, we need to estimate the unknown parameter μ and ρ based on the observed Y_i . Rigorous mathematical calculation implies the joint probability distribution of Y_i is

$$\int_{\mathbb{R}^n} \frac{1}{\sqrt{2\pi^n} |\Sigma|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2} \mathbf{z}^\top \Sigma^{-1} \mathbf{z}\right\} \prod_{i \leq n} I(\text{sgn}(Y_i - 0.5) = \text{sgn}(z_i) - \mu) dz_1 dz_2 \cdots dz_n,$$

where $\text{sgn}(x)$ denotes the sign of x and $I(a = b) = 1$ (0) iff $a = b$ ($a \neq b$). It is remarkable that maximizing this objective function with respect to μ and ρ will take a lot of efforts due to the integration. The form is complicated and corresponding computation is infeasible in practice. An alternative estimator, which is more computationally feasible, is desired. This leads to the idea of paired maximum likelihood estimator in the following subsection.

2.2 Approximated paired maximum likelihood

To propose the new method, we consider the likelihood of two arbitrary selected nodes, denoted by i and j . The joint probability of $\{Y_i, Y_j\}$ can only have the following four cases:

$$\begin{aligned} P(Y_i = 0, Y_j = 0) &= P(Z_i < \mu, Z_j < \mu) \\ &= \int_{-\infty}^{\mu} \int_{-\infty}^{\mu} f(z_i, z_j) dz_i dz_j, \\ P(Y_i = 0, Y_j = 1) &= P(Z_i < \mu, Z_j \geq \mu) \\ &= \int_{-\infty}^{\mu} \int_{\mu}^{\infty} f(z_i, z_j) dz_i dz_j, \\ P(Y_i = 1, Y_j = 0) &= P(Z_i \geq \mu, Z_j < \mu) \\ &= \int_{\mu}^{\infty} \int_{-\infty}^{\mu} f(z_i, z_j) dz_i dz_j, \\ P(Y_i = 1, Y_j = 1) &= P(Z_i \geq \mu, Z_j \geq \mu) \\ &= \int_{\mu}^{\infty} \int_{\mu}^{\infty} f(z_i, z_j) dz_i dz_j, \end{aligned}$$

where $f_{ij}(z_i, z_j)$ is the joint probability density function (p.d.f) of (Z_i, Z_j) . Recall that $\mathbf{Z} = (Z_1, \dots, Z_n)^\top \in \mathbb{R}^n$ follows a normal distribution with zero mean and covariance matrix $\Sigma = (I - \rho W)^{-1}(I - \rho W^\top)^{-1}$. Write $\Sigma = (\sigma_{ij})$, we then define $\Sigma_{ij} = (\sigma_{ii}, \sigma_{ij}; \sigma_{ji}, \sigma_{jj}) \in \mathbb{R}^{2 \times 2}$. By definition, $\Sigma_{ij} = \text{cov}(Z_{ij})$, where $Z_{ij} = (Z_i, Z_j)^\top \in \mathbb{R}^2$. Next, by Taylor's expansion, we have $\Sigma = (\sum_{k=0}^{\infty} \rho^k W^k) \{ \sum_{k=0}^{\infty} \rho^k (W^\top)^k \}$. This suggests that we can approximate Σ by $\Sigma \approx \Sigma^{(K)} = (\sigma_{ij}^{(K)}) = (\sum_{k=0}^K \rho^k W^k) (\sum_{k=0}^K \rho^k W^k)^\top$, where K is some pre-specified approximation order. Accordingly, Σ_{ij} can be approximated by $\Sigma_{ij}^{(K)} = (\sigma_{ii}^{(K)}, \sigma_{ij}^{(K)}; \sigma_{ji}^{(K)}, \sigma_{jj}^{(K)}) \in \mathbb{R}^{2 \times 2}$. This leads to an approximated paired likelihood function as

$$(4) \quad f_{ij}^{(K)}(z_i, z_j) = \frac{1}{2\pi} |\Sigma_{ij}^{(K)}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} \mathbf{Z}_{ij}^\top (\Sigma_{ij}^{(K)})^{-1} \mathbf{Z}_{ij}\right\},$$

then the paired likelihood for each case can be approximated as

$$\begin{aligned} P(Y_i = 0, Y_j = 0) &\approx \int_{-\infty}^{\mu} \int_{-\infty}^{\mu} f_{ij}^{(K)}(z_i, z_j) dz_i dz_j \\ &\equiv \pi_{00}^{(K)}(\theta), \\ P(Y_i = 0, Y_j = 1) &\approx \int_{-\infty}^{\mu} \int_{\mu}^{\infty} f_{ij}^{(K)}(z_i, z_j) dz_i dz_j \\ &\equiv \pi_{01}^{(K)}(\theta), \\ P(Y_i = 1, Y_j = 0) &\approx \int_{\mu}^{\infty} \int_{-\infty}^{\mu} f_{ij}^{(K)}(z_i, z_j) dz_i dz_j \\ &\equiv \pi_{10}^{(K)}(\theta), \\ P(Y_i = 1, Y_j = 1) &\approx \int_{\mu}^{\infty} \int_{\mu}^{\infty} f_{ij}^{(K)}(z_i, z_j) dz_i dz_j \\ &\equiv \pi_{11}^{(K)}(\theta), \end{aligned}$$

where $\theta = (\mu, \rho)^\top \in \mathbb{R}^2$ is the unknown parameter vector. As noted by subscripts, the definition of $\pi_{00}, \pi_{01}, \pi_{10}$ and π_{11} should depend on the subscript (i, j) . However, for notation simplicity, they are omitted. Then following the idea of composite likelihood [21, 24], we can sum up all the paired objective functions. This leads to the K -th order approximated log-likelihood function

$$\begin{aligned} &\sum_{ij} \left\{ I(Y_i = 0, Y_j = 0) \log \pi_{00}^{(K)}(\theta) \right. \\ &\quad + I(Y_i = 0, Y_j = 1) \log \pi_{01}^{(K)}(\theta) \\ &\quad + I(Y_i = 1, Y_j = 0) \log \pi_{10}^{(K)}(\theta) \\ &\quad \left. + I(Y_i = 1, Y_j = 1) \log \pi_{11}^{(K)}(\theta) \right\}, \end{aligned}$$

however, considering all the pairs may take a tremendous amount of computation resources and cannot be used in practice. A compromised solution is to consider those connected pairs, for they are most likely to be correlated and thus provide relevant information. This leads to our final objective function as

$$\begin{aligned} \ell^{(K)}(\theta) &= \sum_{(i,j) \in \mathcal{D}} \left\{ I(Y_i = 0, Y_j = 0) \log \pi_{00}^{(K)}(\theta) \right. \\ &\quad + I(Y_i = 0, Y_j = 1) \log \pi_{01}^{(K)}(\theta) \\ &\quad + I(Y_i = 1, Y_j = 0) \log \pi_{10}^{(K)}(\theta) \\ &\quad \left. + I(Y_i = 1, Y_j = 1) \log \pi_{11}^{(K)}(\theta) \right\}, \end{aligned}$$

where $\mathcal{D} = \{(i, j) : a_{ij} + a_{ji} > 0\}$ collects all the connected pairs. The corresponding estimator is given by $\hat{\theta} = \text{argmax}_{\theta} \ell^{(K)}(\theta)$. Because $\hat{\theta}$ is the estimator obtained by optimizing the approximated paired likelihood function, we refer to it as approximated paired maximum likelihood estimators, APMLE. As one can see, larger K leads to bet-

ter approximation. However, this also calls for substantial computation efforts. We need to take a balance between the accuracy and cost. According to our experience, $K = 1$ and 2 are sufficient for most scenarios.

3. NUMERIC STUDIES

3.1 Simulation models

To demonstrate the finite sample performance of APMLE, we present in this subsection three simulation studies. Each simulation setting corresponds to a typical network topology, which is represented in the generation of the adjacency matrix A as well as the specification of $\theta = (\mu, \rho)^\top \in \mathbb{R}^2$. Once A is simulated, W can be computed by normalizing each row of A . Then the latent variable \mathbb{Z} is generated according to $\mathbb{Z} = (I - \rho W)^{-1} \varepsilon$, where $\varepsilon \in \mathbb{R}^n$ is simulated from a n -dimensional standard normal random vector. The binary response variable Y_i is simulated according to model (1). We consider here three ubiquitous network topologies and detail generation process is presented as follows.

EXAMPLE 1. (Random Distributed Model) We first present here a simple network structure called random distributed model for which the in-degree (i.e., $q_i = \sum_{j=1}^N a_{ji}$) follows a random distribution. In this way, no influential nodes (i.e., with a relatively large in-degree) are allowed in the network. To construct this network structure, we start with n independent and identically distributed random variables which are generated according to a uniform distribution between 0 and 5. Denote these variables by U_i with $1 \leq i \leq n$. For each node i , we randomly select a sample size of $[U_i]$ from $\mathcal{S}_F = \{1, 2, \dots, n\}$ without replacement, where $[U_i]$ stands for the smallest integer no less than U_i . Denote the sample by \mathcal{S}_i . Define $a_{ij} = 1$ if $j \in \mathcal{S}_i$ and $a_{ij} = 0$ otherwise. This leads to the adjacency matrix A . Note $[U_i]$ here is actually the out-degree of each node. Lastly, set $\theta = (0.1, 0.1)^\top$. To give a more realistic description about this network structure, we visualize a network sample with $n = 50$ nodes in one replication in Figure 1¹. From this figure we can see that distribution of in-degree is almost random and there are no influential nodes.

EXAMPLE 2. (Power-Law Distributed Model) The power-law distributed network structure [2, 7] is another popularly studied network topology in literature. A very important feature of this network structure is that the majority of nodes have very few connections while a small amount have a gigantic number of connections. To simulate this network model, we follow [7] and generate A as follows. First, we generate for each node its out-degree in the same way as in example 1. Next, we generate another n independent and identically distributed random variables (e.g., denoted as $r_i, i = 1, \dots, n$) according to the discrete power-law dis-

¹For all the figures in this article, the dot stands for node and the line stands for edge. The deeper the color and the larger the dot means a larger in-degree.

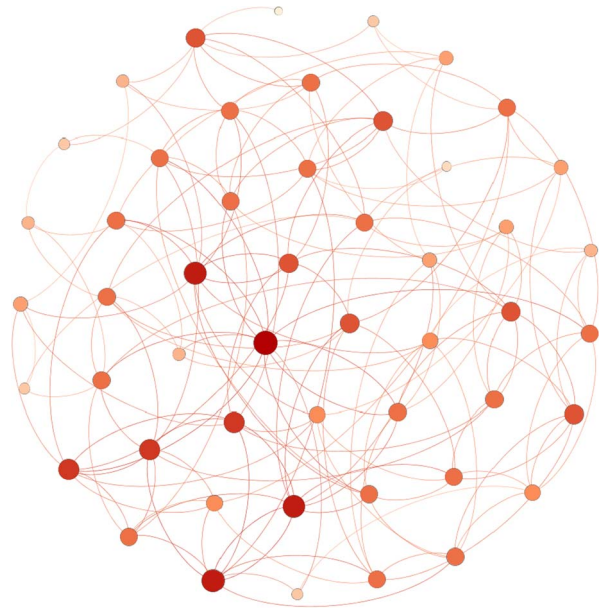


Figure 1. Visualization of the random distributed model.

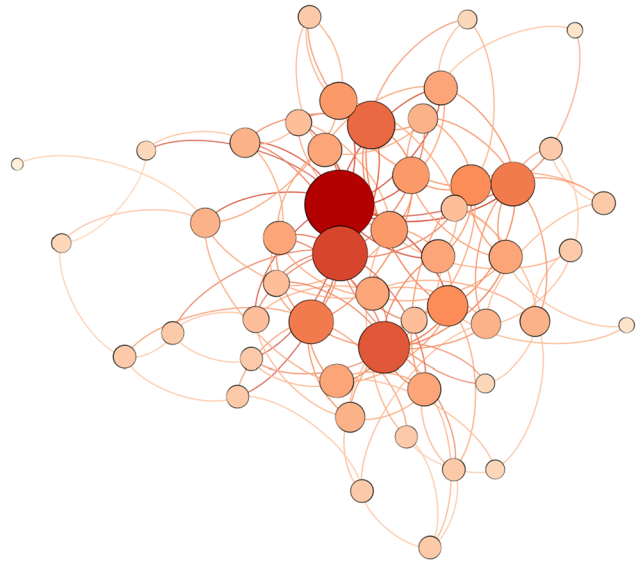


Figure 2. Visualization of the power-law distributed model.

tribution, i.e., $P(r_i) = ck^{-\alpha}$ for a normalizing constant c and the exponent parameter $\alpha = 2.5$. A smaller α value implies a heavier distribution tail. We then normalize each r_i to its corresponding probability $p_i = r_i / \sum_{i=1}^N r_i$. Recall that the out-degree for each node is represented by $[U_i]$ in example 1. Thereafter, for each node i , we select a sample size of $[U_i]$ according to the probability of p_i from $\mathcal{S}_F = \{1, 2, \dots, n\}$ without replacement. Denote the sample by \mathcal{S}_i . Define $a_{ij} = 1$ if $j \in \mathcal{S}_i$ and $a_{ij} = 0$ otherwise. Lastly, fix $\theta = (0.5, 0.2)^\top$. Visualization of this network model with a sample of $n = 50$ nodes in one replication is given in Figure 2. This network structure is very different from its counter-

Table 1. Simulation results for example 1 with $\mu = 0.1$ and $\rho = 0.1$. The RMSE and SD values for μ and ρ are reported

N	$K = 1$				$K = 2$			
	RMSE $_{\mu}$	RMSE $_{\rho}$	SD $_{\mu}$	SD $_{\rho}$	RMSE $_{\mu}$	RMSE $_{\rho}$	SD $_{\mu}$	SD $_{\rho}$
1000	0.0403	0.0744	0.0401	0.0747	0.0402	0.0756	0.0398	0.0760
2000	0.0296	0.0564	0.0282	0.0565	0.0292	0.0574	0.0276	0.0574
4000	0.0226	0.0352	0.0226	0.0353	0.0218	0.0343	0.0219	0.0345

Table 2. Simulation results for example 2 with $\mu = 0.5$ and $\rho = 0.2$. The RMSE and SD values for μ and ρ are reported

N	$K = 1$				$K = 2$			
	RMSE $_{\mu}$	RMSE $_{\rho}$	SD $_{\mu}$	SD $_{\rho}$	RMSE $_{\mu}$	RMSE $_{\rho}$	SD $_{\mu}$	SD $_{\rho}$
1000	0.0669	0.0887	0.0582	0.0867	0.0651	0.0849	0.0574	0.0818
2000	0.0430	0.0620	0.0361	0.0615	0.0421	0.0622	0.0359	0.0603
4000	0.0356	0.0482	0.0333	0.0468	0.0358	0.0484	0.0334	0.0470

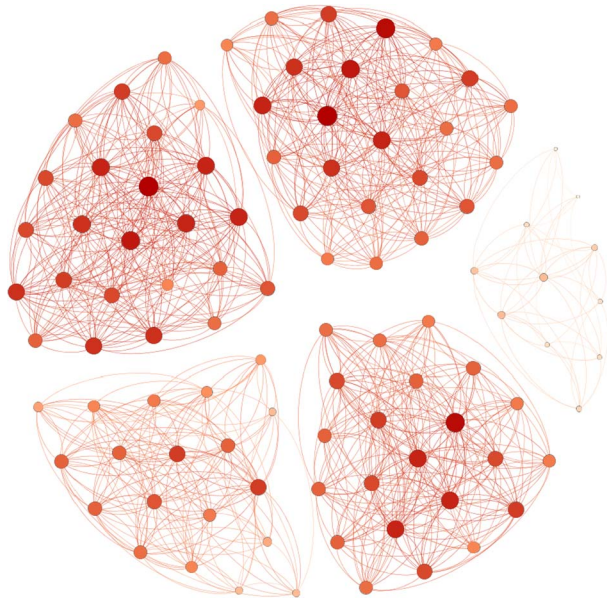


Figure 3. Visualization of the stochastic block model.

part in example 1. We can see that there is a node with very large degree, which could be an influential node.

EXAMPLE 3. (Stochastic Block Model) Finally, we consider a stochastic block model [23, 20], which is also a widely studied network topology in previous literature. For example, it is of particular interest for community detection [26]. Following [20], we randomly assign a block label ($k = 1, 2, \dots, K$) to each node with equal probability, where $K = n/20$ is the total number of blocks. Next, set $P(a_{ij} = 1) = 0.5$ if i and j belong to the same block and $P(a_{ij} = 1) = 0.001/n$ otherwise. This promises that nodes from different blocks are less likely to connect with each other compared with those in the same blocks. Lastly, assign $\theta = (0, -0.1)^\top$. Visualization of this network structure with $n = 100$ and $K = 5$ is presented in Figure 3. From Figure 3 we can clearly see there are 5 blocks in this sim-

ulated network structure. Nodes are connected with each other within a block otherwise not.

3.2 Simulation results

For each simulation example, different network sizes are considered (e.g. $n = 1000, 2000, 4000$) and the experiment is randomly replicated $M = 100$ times. For illustration purpose, we only examine the first order and second order (e.g., $K = 1$ and $K = 2$) approximated paired likelihood estimators. Let $\hat{\theta}^{(m)} = (\hat{\mu}^{(m)}, \hat{\rho}^{(m)})^\top$ be the estimator obtained in the m th replication ($1 \leq m \leq M$). We then consider the following measures to gauge their performances. First, for a given parameter θ_j with $1 \leq j \leq 2$, the root mean square error is evaluated by $\text{RMSE}_j = \{M^{-1} \sum_{m=1}^M (\hat{\theta}_j^{(m)} - \theta_j)^2\}^{1/2}$. Next, for each $1 \leq j \leq 2$, the standard deviation of the corresponding estimator is constructed as $\text{SD}_j = \{M^{-1} \sum_{m=1}^M (\hat{\theta}_j^{(m)} - \bar{\theta}_j)^2\}^{1/2}$, where $\bar{\theta}_j = M^{-1} \sum_{m=1}^M \hat{\theta}_j^{(m)}$. We then use these measures to evaluate the finite sample performance of this proposed method. Detailed results are summarized in Tables 1-3. For the first example in Table 1, we find that the estimators are consistent, with both RMSE and SD values decrease towards 0 as $n \rightarrow \infty$. Since higher order approximation is used in this method, as a side effect, our method is computationally feasible but not superior. Furthermore, we find the RMSE and SD values are very similar to each other for both $K = 1$ and $K = 2$. This means in practice, we can use the first order approximation to save computational resources. Quantitatively similar results are obtained for example 2 in Table 2 and example 3 in Table 3. All these findings confirm a fact that the proposed estimator $\hat{\theta}$ is indeed consistent.

4. REAL DATA ANALYSIS

We consider in this section a real network example about Sina Weibo (www.weibo.com) which can be viewed as a Twitter-type social media in China. Our goal is to investigate how the users of Sina Weibo interact with each other in terms of their posting behavior. For illustration purpose,

Table 3. Simulation results for example 3 with $\mu = 0$ and $\rho = -0.1$. The RMSE and SD values for μ and ρ are reported

N	$K = 1$				$K = 2$			
	RMSE $_{\mu}$	RMSE $_{\rho}$	SD $_{\mu}$	SD $_{\rho}$	RMSE $_{\mu}$	RMSE $_{\rho}$	SD $_{\mu}$	SD $_{\rho}$
1000	0.0375	0.0825	0.0376	0.0663	0.0371	0.0834	0.0370	0.0666
2000	0.0236	0.0607	0.0228	0.0484	0.0246	0.0612	0.0235	0.0489
4000	0.0187	0.0442	0.0188	0.0349	0.0199	0.0431	0.0200	0.0341

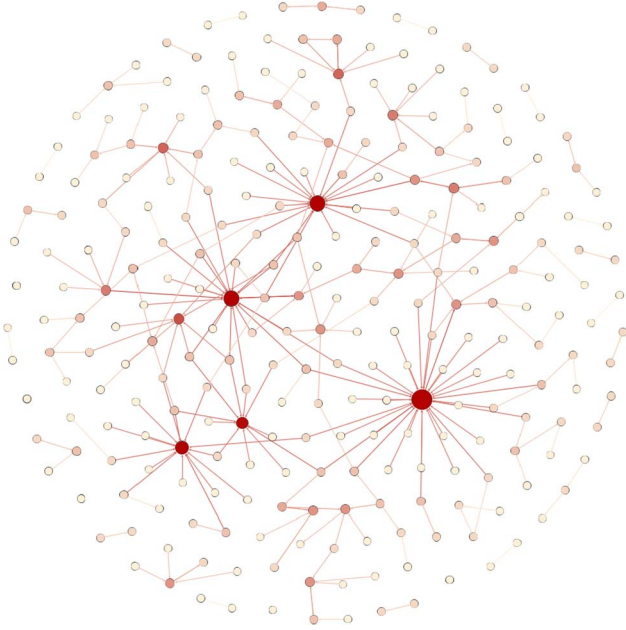


Figure 4. Visualization of the sampled Sina Weibo network.

we start with an official Weibo account and randomly select 100 nodes from their followers. Subsequently, this node's followers are also collected. Because our model is better satisfied with a sparse network structure, this motivates us to keep those users who are with relatively small degree number. This leads to a final dataset of $n = 673$ nodes and network density of 0.067%. Their follower-followee relationships (i.e., A) are also recorded. We picture this sampled real network structure in Figure 4. From this figure, we can see that the network is very sparse, which is very similar to the true world. For each node, we define the binary response as whether it posts a tweet on a specific day. In our dataset, 58.25% of the users have posted a tweet on the observed day. Then the interest is to examine whether the posting will be influenced by the connected friends. Applying the proposed method, we can calculate the spatial autocorrelation $\hat{\rho} = 0.314$ and the threshold $\mu = 0$. This means Sina Weibo users' posting activity does correlated with each other in a nontrivial way.

5. CONCLUSION

We investigate here the spatial autocorrelation estimation problem in a discrete choice model. Specifically, we

consider a discrete choice model with the probit error term specifying in a latent SAR framework. In this model, the network dependence is no longer observable, instead it is treated as a latent variable and measured by a binary response variable. Estimating spatial autocorrelation in this new model is a challenging work because of the complicate integration in the objective function. We then propose a novel approximation method to solve the estimation problem, leading to the method of APMLE. Our findings are confirmed by both numerical studies and a real dataset from Sina Weibo.

To conclude with this article, we discuss here a few interesting topics for further study. First, the spatial discrete choice model only considers those directly connected friends in the network. However, many empirical studies show that higher order relationships (e.g., indirectly connected nodes) can also have impact on consumers' behavior. Then this could be an extension of our proposed method and needs a further study. Second, we assume the spatial autocorrelation is same across different users. While this widely accepted assumption will be violated if we take consumer heterogeneity into consideration. This would be another useful extension of the proposed model which deserves a separate study. Lastly, to further understand individual choice, a more complicated model which incorporates covariates should be considered. However, due to the accessibility of real data, we leave this work for a further investigation.

ACKNOWLEDGEMENTS

The authors thank the editor, associate editor, and two referees for their insightful comments that have led to significant improvement of this article.

Received 21 December 2016

REFERENCES

- [1] ANSELIN, L. (1988). *Spatial Econometrics: Methods and Models*, Springer.
- [2] BARABASI, A. L. and ALBERT, R. (1999). Emergence of scaling in random networks. *Science*, **286** 509–512. [MR2091634](#)
- [3] BHAT, C. R. (2011). The Maximum Approximate Composite Marginal Likelihood (MACML) Estimation of Multinomial Probit-Based Unordered Response Choice Models. *Transportation Research Part B*. **45** 923–939.
- [4] BILLÉ, A. G. (2014). Computational Issues in the Estimation of the Spatial Probit Model: A Comparison of Various Estimators. *The Review of Regional Studies*. **43** 131–154.

- [5] BROCK, W. and DURLAUF, S. (2001). Discrete Choice with Social Interaction. *Review of Economic Studies*. **68** 235–260. [MR1834607](#)
- [6] CHUNG, J. and VITHALA, R.(2003). A general choptice model for bundles with multiple-category products: Application to market segmentation and optimal pricing for bundles. *Journal of Marketing Research*. **40(2)** 115–130.
- [7] CLAUSER, A., SHALIZI, C. R. and NEWMAN, M. E. J. (2009). Power-Law Distributions in Empirical Data. *Siam Review*, **51(4)** 661–703. [MR2563829](#)
- [8] ELHORST, J. P. (2014) *Spatial Econometrics: From Cross-sectional Data to Spatial Panels*. Springer. [MR3114802](#)
- [9] GREENE, W. H. and DAVID, A. H. (2013). Revealing additional dimensions of preference heterogeneity in a latent class mixed multinomial logit model. *Applied Economics*. **45(14)** 1897–1902.
- [10] GUADAGNI, P. M. and JOHN, D. C. L. (1983). A logit model of brand choice calibrated on scanner data *Marketing Science*. **2(3)** 203–238.
- [11] HEAGERTY, P. J. and LELE, S. R. (1998). A composite likelihood approach to binary spatial data. *Journal of the American Statistical Association*. **93** 1099–1111. [MR1649204](#)
- [12] IYENGAR, R., CHRISTOPHE, VAN DEN B. and THOMAS, W. V. (2011). Opinion leadership and social contagion in new product diffusion. *Marketing Science*. **30(2)** 195–212.
- [13] KAMAKURA, WAGNER A. and GARY, R. (1989). A probabilistic choice model for market segmentation and elasticity structure. *Journal of Marketing Research*. **26** 379–390.
- [14] KOPALLE, P. K., KANNAN, P. K., BOLDT, L. B., and ARORA, N. (2012). The impact of household level heterogeneity in reference price effects on optimal retailer pricing policies. *Journal of Retailing*. **88(1)** 102–114.
- [15] LESAGE, J. P. and PACE, R. K. (2009). *Introduction to Spatial Econometrics*, CRC Press, Taylor and Francis Group. [MR2485048](#)
- [16] LI, J. and LEE, L. F.(2009). Binary choice under social interactions: an empirical study with and without subjective data on expectations. *Journal of Applied Econometrics*. **24** 257–281. [MR2662226](#)
- [17] MELA, CARL F., SUNIL G., and DONALD, R. L. (1997). The long-term impact of promotion and advertising on consumer brand choice. *Journal of Marketing Research* 248–261.
- [18] MOZHAROVSKIY, P. and VOGLER, J. (2016). Composite marginal likelihood estimation of spatial autoregressive probit models feasible in very large samples. *Economic Letters*. **148** 87–90. [MR3566060](#)
- [19] NITZAN, I., and BARAK, L. (2011). Social effects on customer retention. *Journal of Marketing*. **75(6)** 24–38.
- [20] NOWICKI, K. and SNIJDERS, T. A. B. (2011). Estimation and prediction for stochastic block structures. *Journal of the American Statistical Association*. **96(455)** 1077–1087. [MR1947255](#)
- [21] SHAO, J. (2003). *Mathematical Statistics*. New York: John Wiley. [MR2002723](#)
- [22] SMIRNOV, O. A. (2010). Modeling spatial discrete choice. *Regional Science and Urban Economics*. **40(5)** 292–298.
- [23] WANG, Y. J. and WONG, G. Y. (2010). Stochastic block models for directed graphs. *Journal of the American Statistical Association*. **82(82)** 8–19. [MR0883333](#)
- [24] WANG, H., IGLESIAS, E. M. and WOOLDRIDGE, J. M. (2013). Partial maximum Likelihood Estimation of a Spatial Probit Model. *Journal of Econometrics*. **172** 77–89. [MR2997131](#)
- [25] WEI, Y., YILDIRIM, P., VAN DEN BULTE, C. and DELLAROCAS, C. (2015). Credit scoring with social network data. *Marketing Science*, To appear.
- [26] ZHAO Y., LEVINA, E. and ZHU, J. (2012). Consistency of community detection in networks under degree-corrected stochastic block models. *Annals of Statistics*. **40(40)** 2266–2292. [MR3059083](#)

Jing Zhou
 School of Statistics
 Renmin University of China
 Beijing
 China
 E-mail address: zhoujing_89@126.com

Da Huang
 School of Management
 Fudan University
 Shanghai
 China
 E-mail address: dahuang@fudan.edu.cn

Hansheng Wang
 Guanghua School of Management
 Peking University
 Beijing
 China
 E-mail address: hansheng@gsm.pku.edu.cn