

# A Social Interactions Model with Endogenous Friendship Formation and Selectivity

Chih-Sheng Hsieh and Lung-fei Lee

Department of Economics  
Ohio State University  
Columbus, Ohio

## Abstract

This paper considers the problem caused by endogenous association of members within a group when applying the spatial autoregressive (SAR) model to study outcomes of social interactions. When the spatial weights matrix in the SAR model is designed to represent the network structure within a group, it could be correlated with the disturbance term of the SAR model and then cause an endogeneity problem. We try to correct this selection bias problem with a new modeling approach. In this approach, a statistical network model is adopted to explain the endogenous network formation process. By specifying an unobserved component in both the network model and the SAR model, we can capture the correlation between the network formation process and the outcome process and then restore a proper estimation procedure for the SAR model. This study provides a Monte Carlo experiment and an empirical example using the Add Health data to show the usefulness of this new modeling approach.

**JEL classification:** C21, C25, I21, J13

**Keywords:** spatial autoregressive, network, social interaction, Bayesian estimation

# 1 Introduction

Since Manski (1993), there has been a wide range of discussion on the potential problems inherited in the empirical analysis of social interactions. Typically researchers worry about two problems. The first is the difficulty to distinguish between endogenous and contextual effects – the ‘reflection’ problem. It is realized that we can not avoid the reflection problem in Manski’s linear-in-means model unless strong exclusion assumptions are imposed. However, if we consider the binary choice model setting in Brock and Durlauf (2001, 2007) or the Spatial Autoregressive (SAR) model setting in Lee (2007) which involves many different-sized groups, the reflection problem can be resolved. The second problem is the influences of unobservables in a group, also known as the ‘correlated effects’ in Manski’s (1993) terminology. Moffitt (2001) argues that correlated unobservables in a group may contribute to correlations of outcomes and cause an identification problem by confounding with endogenous effects. There are some existing solutions for this problem, including instrumental variable strategy (e.g. Evans *et al.* 1992; Rivkin 2001), family fixed effect strategy (e.g. Aaronson 1998; Plotnick *et al.* 1999) and experiment-type data strategy (e.g. Sacerdote 2001; Zimmerman 2003). Under the SAR model setting, Lee (2007) also provides a solution for this problem by incorporating correlated unobservables into the model as a fixed effect in a group.

Since the SAR model specified in Lee (2007) handles above two problems in the study of social interactions, it has been applied in a number of empirical works (e.g. Hauser *et al.* 2009; Lin 2010; Lee *et al.* 2010; Boucher *et al.* 2010), especially on studying peer influences to students. The reason why this specific social interaction receives researchers’ attentions is due to its importance to school policies. Once the magnitude of peer influence on students is realized, it becomes new evidence to support or oppose school policies such as ability tracking, affirmative action and the like. Lin (2010) is among the first to apply the SAR model in Lee (2007) to the real data set – Add Health data (Udry, 2003) and confirm the existence of endogenous and contextual effects in students’ academic achievement. During the application, she generalizes the SAR model in Lee (2007) by incorporating the more realistic asymmetric social interaction pattern within groups, which means that every student has his (her) own friendship network within a group and most interactions are happening with network members. This interaction pattern implies that the friendship network provides the main source of influences to students and influences from other individuals outside the friendship network can be neglected. This feature makes the SAR model used in Lin (2010) different from Lee (2007) in which each individual is assumed to be equally affected by all the other members in the same group. Although the asymmetric interaction pattern is more realistic in characterizing interactions among students, it requires specific information from

data on how students are interacted within a group. The Add Health data used in Lin (2010) provides information on ‘named’ friends of each student within the sample and makes this approach plausible. The information of whether link or not for each pair of students can be summarized by a spatial weights matrix with nonzero elements at some entries of each row and the resulting weights matrix will represent a network structure. To account for the network structure and group fixed effects in the SAR model, Lin (2010) estimates the model using the maximum likelihood (ML) method developed in Lee *et al.* (2010).

In this paper, we continue with the SAR model used in Lin (2010) and discuss the issue of endogenous friendship formation and its possible selection bias consequence on outcomes within groups. While it is known that endogenous association within groups will also cause biases on the estimates of interaction effects, there are only few papers formally dealing with this problem (e.g., Weinberg 2007; Fletcher and Tienda 2008; Mihaly 2009). Weinberg (2007) develops a theoretical model of social interactions which allows endogenous association within groups and generates a nonlinear relationship between the group composition and behaviors. This nonlinearity could provide a potential solution to Manski’s reflection problem. Also, his model shows that sorting would be greater in large groups than in small groups. This finding has an important policy implication, that is, any relocation experiments will be more effective when experiment participants are relocated to small groups. Fletcher and Tienda (2008) recognize that one particular network for college freshmen is consisted of members who attended the same high school. Therefore, they use the data from the University of Texas-Austin to examine whether high school peer networks influence college freshmen’s academic achievement. They find that college freshmen with larger high school peer networks outperform their counterparts with smaller networks. Mihaly (2009) examines the relationship between students’ popularity and academic achievement. She controls the endogenous friendship formation by instrumenting the popularity variable with the interaction of individual demographic characteristics and the mean of that characteristic in the grade by gender. After controlling for endogenous friendship formation, she finds that the original positive effect of popularity on academic achievement becomes negative and much larger. She concludes that the original positive effect of popularity is biased due to endogeneity. Under the SAR model, endogenous association of members within a group will result in an endogenous spatial weights matrix. If some variables which affect both network structures within a group and individuals’ outcomes were not properly specified in the model, the spatial weights matrix would be correlated with the disturbance term in the SAR model. The consequence is that standard estimation methods for the SAR model such as the 2SLS and GMM methods which construct instrument variables or

moment conditions from the spatial weights matrix would not be valid. Even the ML method provided in Lee *et al.* (2010) would not be appropriate and need to be reformulated. To correct this problem, we propose a new modeling approach which takes account of the possible endogenous spatial weights matrix. In this modeling approach, entries of the spatial weights matrix are viewed as endogenous and will be captured by a network formation process. By properly specifying an unobserved component in both the network formation process and the outcome process to make them connected, we can capture the problem caused by endogenous association of members within a group and restore a selection bias corrected estimation procedure for the SAR model. The model specifications in this paper are parametric in nature.<sup>1</sup>

The remainder of this paper is organized as follows. Section 2 presents a new modeling approach which should be applied to the original SAR model when there is a concern of endogenous association within groups. A Bayesian estimation method for the proposed models is provided in Section 3. Section 4 contains a simulation study to investigate the sampling property of our Bayesian estimation method. Section 5 includes an empirical study. We conclude this paper in Section 6.

## 2 The SAR model with an endogenous weights matrix

The SAR model used in Lin (2010) is given by

$$Y_g = \lambda W_g Y_g + l_{m_g} \beta_1 + X_g \beta_2 + W_g X_g \beta_3 + l_{m_g} \alpha_g + \epsilon_g, \quad g = 1, \dots, G. \quad (1)$$

$Y_g = (y_{1,g}, \dots, y_{m_g,g})'$  is an  $m_g$ -dimensional vector of  $y_{i,g}$ , where  $y_{i,g}$  is the observed outcome of the  $i$ th member in the group  $g$  and  $m_g$  is the number of members in the  $g$ th group.  $X_g$  is a  $m_g \times k$  matrix of exogenous variables.  $l_{m_g}$  is the  $m_g$ -dimensional vector of ones and  $\alpha_g$  represents unobserved group-specific effects.  $W_g$  is an  $m_g \times m_g$  spatial weights matrix which summarizes the network structure of the group  $g$ . In this paper, we start from the SAR model in equation (1) and try to deal with the problem caused by an endogenous weights matrix. In the usual specification of  $W_g$  from current social network study (see Wasserman and Faust 1994), each entry of  $W_g$  will take the value of 1 if there is a link and 0 otherwise. In general, elements of  $W_g$  can be functions of those 0 and 1 choices for each individual. The row normalization on  $W_g$  such that the sum of each row is the unity, i.e.,  $W_g l_{m_g} = l_{m_g}$  when all the outdegrees are not zero, is one specific case. With the row normalization on  $W_g$ , each element of  $W_g Y_g$  summarizes the weighted average influence of

---

<sup>1</sup>It remains to be shown whether any nonparametric approach can be feasible.

the connected peers. When  $W_g$  is not row-normalized and each entry is either 0 or 1, the coefficient  $\lambda$  should be interpreted as the average influence from one single connection. We then consider the specification of an endogenous network formation process for  $W_g$ , and the estimation of such a model. For notational simplicity, we adopt the 0-1 entries case for  $W_g$  in subsequent analysis.<sup>2</sup>

## 2.1 The Network model

To model the endogenous weights matrix,  $W_g$ , we adopt one statistical network model – latent position model from Hoff *et al.* (2002). The main assumption of this latent position model is that, in each group  $g$ , each individual  $i$  has an unknown position  $z_{i,g}$  in a social space. Therefore, the latent position model uses a function of these positions such as the distance between two individuals in the social space as well as other observed characteristics as independent variables to explain individuals' links. We use  $C_g^r$ ,  $r = 2, \dots, q$ , each of which is a  $m_g \times m_g$  matrix, to summarize observed characteristics, such as age, sex or race, shared by each pair of individuals in the group  $g$ . Each element of  $C_g^r$ ,  $c_{ij,g}^r$  could be either a dummy variable indicates whether individuals  $i$  and  $j$  share the same observed characteristics or a continuous variable which measures the distance between  $i$ 's and  $j$ 's observed characteristics. Let  $C_g = \{C_g^2, \dots, C_g^q\}$  be a collection of such matrices. We also let  $Z_g = (z_{1,g}, \dots, z_{m_g,g})$ , be the column vector of latent positions for all individuals in the group  $g$ . The latent position model is parameterized as a logit model where the probability of each link  $w_{ij,g}$  in the group  $g$  can be written as

$$P(w_{ij,g} | c_{ij,g}, z_{i,g}, z_{j,g}, \gamma) = \left( \frac{e^{\psi_{ij,g}}}{1 + e^{\psi_{ij,g}}} \right)^{w_{ij,g}} \left( \frac{1}{1 + e^{\psi_{ij,g}}} \right)^{1-w_{ij,g}}$$

$$\psi_{ij,g} = \gamma_1 + \sum_{r=2}^q \gamma_r c_{ij,g}^r + \gamma_{q+1} \cdot |z_{i,g} - z_{j,g}|. \quad (2)$$

Following Hoff *et al.* (2002), we call this model the distance model. Given these latent position variables  $Z_g$ , each link in the network  $g$  is assumed to be conditionally independent of other links. Therefore, the probability of observed links of the network  $g$  can be calculated as

$$P(W_g | C_g, Z_g, \gamma) = \prod_{i \neq j}^{m_g} P(w_{ij,g} | c_{ij,g}^2, \dots, c_{ij,g}^q, z_{i,g}, z_{j,g}, \gamma), \quad (3)$$

where  $\gamma = (\gamma_1, \dots, \gamma_{q+1})$ . In our application, the role of  $Z_g$  does not necessarily represent the social position, it could be any variable unobserved to researchers, for instance, taste or interest. Here

---

<sup>2</sup>Our Bayesian approach can be adopted to the general case that  $W_g$  is a function of those 0-1 variables. This is so, because, in the Bayesian approach, conditional analysis on  $W_g$  can be replaced by conditioning on those 0-1 decisions of the members in the  $g$  group.

we simply assume  $z_{i,g}$  to be one-dimensioned. In the study of social network, we usually find some common features from data: homophily on attitudes, transitivity of relations and clustering (see review in Wasserman and Faust 1994). The homophily means individuals are more likely to be connected with others who share the same characteristics. Transitivity of relations means if there are two individuals both connected with the third individual, then these two individuals are more likely to be connected. Clustering is interpreted as the result of transitivity but it could also be generated by homophily on unobserved attributes. In the distance model specified in equation (2), parameters  $\gamma_2$  to  $\gamma_q$  can capture the homophily on some observed characteristics and  $\gamma_{q+1}$  can capture the inherited transitivity of relations. The idea of using  $\gamma_{q+1}$  to capture the transitivity of relations can be illustrated with the triangle inequality,  $|z_{i,g} - z_{j,g}| \leq |z_{i,g} - z_{k,g}| + |z_{k,g} - z_{j,g}|$ . When observing links between  $\{i, k\}$  and  $\{k, j\}$  it suggests that  $|z_{i,g} - z_{k,g}|$  and  $|z_{k,g} - z_{j,g}|$  are small. Therefore,  $|z_{i,g} - z_{j,g}|$  cannot not be large and a link between  $\{i, j\}$  is expected.

## 2.2 The SAR model

In the distance model, we introduce a vector of unobserved variables  $Z_g$ . These  $Z_g$  may be useful to characterize the problem of association within groups – by allowing  $Z_g$  be correlated with the error term  $\epsilon_g$  in the SAR model. Based on the idea that  $Z_g$  may capture the unobservable in association with bias of outcomes, we propose two new specifications of the SAR model by taking into account the possible correlation between  $Z_g$  and  $\epsilon_g$ . In the first specification we assume

$$(z_{i,g}, \epsilon_{i,g}) \sim i.i.d. \mathcal{N}_2 \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_z^2 & \sigma_{z\epsilon} \\ \sigma_{\epsilon z} & \sigma_\epsilon^2 \end{pmatrix} \right). \quad (4)$$

With this assumption, the SAR model in equation (1) can be rewritten as

$$Y_g = \lambda W_g Y_g + l_{m_g} \beta_1 + X_g \beta_2 + W_g X_g \beta_3 + \sigma_{\epsilon z} \sigma_z^{-2} Z_g + l_{m_g} \alpha_g + u_g, \quad (5)$$

where  $u_g \sim \mathcal{N}_{m_g}(0, (\sigma_\epsilon^2 - \sigma_{\epsilon z}^2 \sigma_z^{-2}) I_{m_g})$ . From this specification  $Z_g$  appears in the SAR model as one of the determinants for  $Y_g$ . If the correlation between  $Z_g$  and  $\epsilon_g$  specified in equation (4) is true, using the original SAR model in equation (1) which neglects the potential effect of  $Z_g$  will be problematic because we will face the problem of endogeneity on  $W_g$ , which is exactly the problem of association of members within a group discussed so far. Therefore, one may replace the original SAR model with the new SAR model in equation (5) and joint it with the distance model as the generalized model to study the interaction effect. By doing that, we may capture possible association within a group. One thing should be clear is that if  $\sigma_{\epsilon z}$  is equal to zero, which means

$Z_g$  and  $\epsilon_g$  are uncorrelated, the concern of endogeneity on  $W_g$  is eliminated and then the distance model and the SAR model can be estimated separately.

For the second specification, we directly specify the linear relationship between  $\epsilon_g$  and  $Z_g$ , more explicitly, we let  $\epsilon_g = Z_g\delta_1 + W_gZ_g\delta_2 + u_g$ , where  $Z_g \sim \mathcal{N}_{m_g}(\mu_z, \sigma_z^2 I_{m_g})$  and  $u_g \sim \mathcal{N}_{m_g}(0, \sigma_u^2 I_{m_g})$ . By bring this assumption into the original SAR model in equation (1) we get

$$Y_g = \lambda W_g Y_g + l_{m_g} \beta_1 + X_g \beta_2 + W_g X_g \beta_3 + Z_g \delta_1 + W_g Z_g \delta_2 + l_{m_g} \alpha_g + u_g, \quad (6)$$

where  $u_g \sim \mathcal{N}_{m_g}(0, \sigma_u^2 I_{m_g})$ . In this specification  $Z_g$  plays the role as an unobserved counterpart of  $X_g$ . Therefore, the parameter  $\delta_1$  and  $\delta_2$  will represent the own and contextual effect of  $Z_g$  separately. By the inspection of equation (5) and equation (6), we can see that the second specification is more general than the first one by considering more on the contextual effect of  $Z_g$ . We joint the SAR model in equation (6) with the distance model to be the second generalized model. Here after we call the first generalized model the Type-I model and the second generalized model the Type-II model.

The preceding models consider the case where the outcome variable  $Y_g$  in the SAR model is continuous. However, in some situations we only observe a binary indicator of the outcome, namely,  $y_{i,g} = 1$  if  $y_{i,g}^* \geq 0$  and  $y_{i,g} = 0$  if  $y_{i,g}^* < 0$ , where  $y_{i,g}^*$  is a latent outcome which is not directly observable. Models with interaction on observed binary choices have been considered in Brock and Durlauf (2001, 2007), Krauth (2006) and Lee *et al.* (2007). Brock and Durlauf (2001, 2007) model binary choices with social interactions under homogeneous rational expectations and Lee *et al.* (2007) extend it to heterogeneous rational expectations. The model in Krauth (2006) is a simultaneous discrete choice model in which the equilibrium is not based on rational expectations but game theoretical solutions. The game theoretical model will face the potential problem of multiple equilibria on outcomes. The same problem may also occur in the rational expectations models when interactions are strong. Except these models, Davezies *et al.* (2009) consider the case where interactions happen on latent variables  $Y_g^* = (y_{1,g}^*, \dots, y_{m_g,g}^*)$  which determine observed outcome variables  $Y_g$ . They argue that the resulting latent variables model “*is not a discrete choice model but rather a continuous choice model with imperfect observations of the choices.*” In this paper we follow Davezies *et al.* (2009) to consider the latent variables model. We assume

$$Y_g^* = \lambda W_g Y_g^* + l_{m_g} \beta_1 + X_g \beta_2 + W_g X_g \beta_3 + l_{m_g} \alpha_g + \epsilon_g, \quad (7)$$

and

$$\begin{aligned}
y_{i,g} &= \mathbf{1}\{y_{i,g}^* \geq 0\} \\
&= \mathbf{1}\{(I_g - \lambda W_g)_i^{-1}(l_{m_g}\beta_1 + X_g\beta_2 + W_gX_g\beta_3 + l_{m_g}\alpha_g + \epsilon_g) \geq 0\} \\
&= \mathbf{1}\{S_{g,i}^{-1}(\mathbf{X}_g\beta + l_{m_g}\alpha_g + \epsilon_g) \geq 0\}
\end{aligned} \tag{8}$$

where  $S_{g,i} = (I_g - \lambda W_g)_i$  is the  $i$ th row of  $(I_g - \lambda W_g)$ ,  $\mathbf{X}_g = (l_{m_g}, X_g, W_gX_g)$ ,  $\beta = (\beta_1, \beta_2', \beta_3')'$  and  $\mathbf{1}\{\cdot\}$  represents an indicator function. This specification is meaningful in the sense that the latent variable  $Y_g^*$  can be treated as unobserved utilities, motivations or intensions from individuals. Interactions between individuals may be generated via the inter-dependence of these latent variables and then the interactions are reflecting on the observed binary outcomes of  $Y_g$ . When the observed binary outcome  $Y_g$  is augmented with the latent outcome  $Y_g^*$  modeled in equation (7), it is also possible to apply the Type-I and Type-II specifications on  $Y_g^*$  in order to capture the effect of endogenous association within groups.

### 3 Estimation

In this study we apply the Bayesian method to estimate the Type-I and Type-II models. Different from the classical estimation approach which employs the concept of extremum estimation with an objective function, the Bayesian estimation relies purely on probability concepts and statistical inferences are based on the posterior distributions of parameters. One may utilize the Markov Chain Monte Carlo (MCMC) method to draw samples from posterior distributions and then calculate sample posterior means as parameter estimates. The superiority of dealing with complex models especially with the presence of latent variable  $Z_g$  in both the distance model and the SAR model is the main reason why adopt the Bayesian estimation instead of the classical estimation. The Bayesian estimation has been applied to the distance model in Handcock *et al.* (2007), and the SAR model in LeSage (1997, 2000) and Smith and LeSage (2002). In this study, we generalize their estimating procedures taking into account the endogenous link between these two models. To better organize the presentation without unnecessary repetitions, we will present the general estimating procedure using the Type-I model as an example in the first subsection. For the Type-II model only required modifications from this general procedure will be discussed. The extension to the case of binary dependent variable will be discussed in another subsection.



### 3.1 The case of continuous dependent variable

Before estimating the Type-I model, we have to discuss two identification problems surrounding the variable  $Z_g$ . First, the parameter  $\sigma_z^2$  is not identified from either the distance model or the SAR model. Therefore, as  $Z_g$  are normally distributed, the covariance matrix is assumed to be equal to  $I_{m_g}$  during the estimation. The consequence of restricting  $Z_g$  with an identity covariance matrix is that the parameter  $\gamma_{q+1}$  in equation (2) and  $\sigma_{\epsilon z}\sigma_z^{-2}$  in equation (5) will both absorb  $\sigma_z$  and become  $\gamma_{q+1}\sigma_z$  and  $\sigma_{\epsilon z}\sigma_z^{-1}$ . There is no way to further separate  $\sigma_z$  from  $\gamma_{q+1}$  and  $\sigma_{\epsilon z}$ , hence, they should be treated as confounded effects from  $Z_g$ . Second, the sign of  $\sigma_{\epsilon z}$  is arbitrary. Since the distance model is used as our network model, its identification problem is inherited, that is,  $|z_{i,g} - z_{j,g}|$  is invariant under rotation, reflection, and translation of  $z_{i,g}$  and  $z_{j,g}$ . Although including  $Z_g$  into equation (5) creates new identification source, we still face the problem that  $|z_{i,g} - z_{j,g}|$  is invariant under the sign translation of  $z_{i,g}$  and  $z_{j,g}$ . Since  $Z_g$  is unknown to researchers and its sign cannot be identified from the distance model, it results in an arbitrary sign on the parameter  $\sigma_{\epsilon z}$ . In the concern of this, we normalize  $\sigma_{\epsilon z}$  to be positive during the estimation and this can be achieved easily in the Bayesian method by assuming a truncated prior distribution for  $\sigma_{\epsilon z}$ .

To prepare for the posterior analysis, we have to define prior distributions for all unobservables in the model, including parameters  $\theta_1 = (\gamma, \lambda, \beta, \sigma_\alpha^2, \sigma_\epsilon^2, \sigma_{\epsilon z})$ , latent variables  $\{Z_g\}$  and group effects  $\{\alpha_g\}$ . In this study, we treat  $\alpha_g$  as a random effect<sup>3</sup> and assume  $\alpha_g \sim \mathcal{N}(0, \sigma_\alpha^2)$ . By properly blocking parameters in  $\theta_1$  into subgroups and setting up a hierarchical prior for  $\alpha_g$ , we define prior distributions for all unobservables in the Type-I model as follows:

$$Z_g \sim \mathcal{N}_{m_g}(0, I_{m_g}), \quad g = 1, \dots, G, \quad (9)$$

$$\gamma \sim \mathcal{N}_{q+2}(\gamma_0, \Gamma_0), \quad (10)$$

$$\lambda \sim U[-1/\tau_G, 1/\tau_G], \quad (11)$$

$$\beta \sim \mathcal{N}_{2k+1}(\beta_0, B_0), \quad (12)$$

$$\sigma = (\sigma_\epsilon^2, \sigma_{\epsilon z}) \sim \mathcal{TN}_2(\sigma_0, \Sigma_0), \quad (13)$$

$$\alpha_g | \sigma_\alpha^2 \sim \mathcal{N}(0, \sigma_\alpha^2), \quad g = 1, \dots, G, \quad (14)$$

$$\sigma_\alpha^2 \sim \mathcal{IG}(\frac{\rho_0}{2}, \frac{\eta_0}{2}), \quad (15)$$

---

<sup>3</sup>The Bayesian estimation is still a likelihood-based approach. If we assume  $\alpha_g$  is a group fixed effect, in order to derive the likelihood function of the SAR model, we must get rid of  $\alpha_g$  by a transformation on the SAR model. See Lee *et al.* (2008) for details of this transformation. However, this transformation requires that sum of each rows in  $W_g$  equal to one. In this study we do not want to impose this restriction, therefore, we use the random instead of the fixed effect.

where  $\mathcal{TN}_2(\sigma_0, \Sigma_0)$  is a truncated bivariate-normal distribution and  $\mathcal{IG}(a, b)$  is the inverse-gamma distribution with a shape parameter  $a$  and a scale parameter  $b$ . These prior distributions except for  $\lambda$  are conjugate priors commonly used in the Bayesian literature. For  $\lambda$  we employ a uniform prior suggested in Smith and LeSage (2002). We restrict the valid value of  $\lambda$  between  $-1/\tau_G$  to  $1/\tau_G$ , where  $\tau_G = \max\{\tau_1^*, \dots, \tau_G^*\}$  and  $\tau_g^* = \min\{\max_{1 \leq i \leq m_g} \sum_{j=1}^{m_g} |w_{ij,g}|, \max_{1 \leq j \leq m_g} \sum_{i=1}^{m_g} |w_{ij,g}|\}^4$ . We especially put  $\sigma_\epsilon^2$  and  $\sigma_{\epsilon z}$  into a group and call it  $\sigma$ . By designing a truncated distribution for  $\sigma$  to a area  $C$ , where  $C$  is a convex area that leads to a proper correlation matrix with  $\sigma_{\epsilon z} \geq 0$ , we can make sure that  $\sigma_\epsilon^2$  and  $\sigma_{\epsilon z}$  form a proper correlation matrix and  $\sigma_{\epsilon z}$  is positive.

The joint posterior distribution of all unobservables in the Type-I model can be constructed using Baye's theorem  $P(\theta|Y) \propto \pi(\theta)P(Y|\theta)$ :

$$P(\{Z_g\}, \{\alpha_g\}, \theta_1 | \{Y_g\}, \{W_g\}) \propto \pi(\{Z_g\}, \{\alpha_g\}, \theta_1) \cdot P(\{Y_g\}, \{W_g\} | \{Z_g\}, \{\alpha_g\}, \theta_1), \quad (16)$$

where  $\pi(\cdot)$  represents the density function of the prior distribution and independent variables  $\{X_g\}$  and  $\{C_g\}$  are suppressed from the above expression for simplicity. Here we assume independence between prior distributions of all unobservables and hence,  $\pi(\{Z_g\}, \{\alpha_g\}, \theta_1) = \pi_1(\{Z_g\})\pi_2(\{\alpha_g\})\pi_3(\theta_1)$ . Drawing samples from this joint posterior distribution can rely on the Gibbs sampling. The idea of Gibbs sampling is that we first divide all parameters in the model into blocks such that drawing from their conditional posterior distributions is feasible. Then we make draws sequentially from these conditional posterior distributions by cycling and in the limit these draws can be treated as draws from the joint posterior distribution. One strategy of the Gibbs sampling is to sample each single parameter from the full conditional posterior distribution (the conditional distribution of each parameter given the data and the remaining parameters). However, Gelfand *et al.* (1995) point out that this strategy might suffer from slow convergence due to high correlations between, or weak identifiability of, certain model parameters and this is referred to the problem with mixing. A case of this problem can be found in the linear hierarchical model studied in Chib and Carlin (1999) where the random effects are highly negatively correlated with covariate parameters  $\beta$  and the variances of the random effects and error terms are only identified corporately, not independently from the data. To deal with this problem, Chib and Carlin (1999) apply the grouped Gibbs sampler introduced by Liu (1994) to sample the highly-correlated parameters in the same block. In our model we use the similar approach to deal with high correlations between the random effects  $\{\alpha_g\}$  and other parameters in the SAR model. We first put  $\gamma, \lambda, \beta, \sigma, \{Z_g\}$  with  $\{\alpha_g\}$  in the same block and leave  $\sigma_\alpha^2$  in another block. Therefore, the grouped Gibbs sampler requires

---

<sup>4</sup>This interval is suggested by Kelejian and Prucha (2010) in which  $I_{m_g} - \lambda W_g$  is nonsingular for all values of  $\lambda$  in this interval.

drawing:

- (i)  $(\{Z_g\}, \{\alpha_g\}, \gamma, \lambda, \beta, \sigma) \sim P(\{Z_g\}, \{\alpha_g\}, \gamma, \lambda, \beta, \sigma | \{Y_g\}, \{W_g\}, \sigma_\alpha^2);$
- (ii)  $\sigma_\alpha^2 \sim P(\sigma_\alpha^2 | \{Y_g\}, \{W_g\}, \{Z_g\}, \{\alpha_g\}, \gamma, \lambda, \beta, \sigma).$

We can use the method of composition to rewrite step (i) as

$$(i') (\{Z_g\}, \{\alpha_g\}, \gamma, \lambda, \beta, \sigma) \sim P(\{\alpha_g\} | \{Y_g\}, \{W_g\}, \{Z_g\}, \gamma, \lambda, \beta, \sigma, \sigma_\alpha^2) \cdot P(\{Z_g\}, \gamma, \lambda, \beta, \sigma | \{Y_g\}, \{W_g\}, \sigma_\alpha^2).$$

In step (i') we first draw  $(\{Z_g\}, \gamma, \lambda, \beta, \sigma)$  marginalizing over  $\{\alpha_g\}$  and then draw  $\{\alpha_g\}$  conditioning on all other parameters. The fact that the conditional distribution of the outcomes  $Y_g$  marginalized over  $\alpha_g$  is still normal makes this composition method beneficial. However, it is still difficult to draw  $(\{Z_g\}, \gamma, \lambda, \beta, \sigma)$  jointly in step (i'). Therefore, we use the following modified version of the grouped Gibbs sampler:

- (i'a)  $\{Z_g\} \sim P(\{Z_g\} | \{Y_g\}, \{W_g\}, \gamma, \lambda, \beta, \sigma, \sigma_\alpha^2);$
- (i'b)  $\gamma \sim P(\gamma | \{Y_g\}, \{W_g\}, \{Z_g\}, \lambda, \beta, \sigma, \sigma_\alpha^2);$
- (i'c)  $\lambda \sim P(\lambda | \{Y_g\}, \{W_g\}, \{Z_g\}, \gamma, \beta, \sigma, \sigma_\alpha^2);$
- (i'd)  $\beta \sim P(\beta | \{Y_g\}, \{W_g\}, \{Z_g\}, \gamma, \lambda, \sigma, \sigma_\alpha^2);$
- (i'e)  $\sigma \sim P(\sigma | \{Y_g\}, \{W_g\}, \{Z_g\}, \gamma, \lambda, \beta, \sigma_\alpha^2);$
- (i'f)  $\{\alpha_g\} \sim P(\{\alpha_g\} | \{Y_g\}, \{W_g\}, \{Z_g\}, \gamma, \lambda, \beta, \sigma, \sigma_\alpha^2);$
- (ii)  $\sigma_\alpha^2 \sim P(\sigma_\alpha^2 | \{Y_g\}, \{W_g\}, \{Z_g\}, \{\alpha_g\}, \gamma, \lambda, \beta, \sigma).$

It can be shown that the draws from this modified grouped Gibbs sampler still converge to our objective joint posterior distribution in equation (16)<sup>5</sup>. Now we list the set of conditional posterior distributions required in this modified grouped Gibbs sampler:

- (i'a) Using Baye's theorem it shows that  $P(Z_g | Y_g, W_g, \theta_1) \propto \pi(Z_g) \cdot P(Y_g, W_g | Z_g, \theta_1)$ . Therefore, we can get

$$P(Z_g | Y_g, W_g, \theta_1) \propto \phi_{m_g}(Z_g; 0, I_{m_g}) \cdot P(Y_g, W_g | Z_g, \theta_1), \quad g = 1, \dots, G. \quad (17)$$

- (i'b) First we can simplify the conditional posterior distribution of  $\gamma$  to  $P(\gamma | \{W_g\}, \{Z_g\})$ . By Baye's theorem we get

$$P(\gamma | \{W_g\}, \{Z_g\}) \propto \mathcal{N}_{q+2}(\gamma; \gamma_0, \Gamma_0) \cdot \prod_{g=1}^G P(W_g | Z_g, \gamma). \quad (18)$$

---

<sup>5</sup>See discussions of convergence for modified Gibbs sampler in Chen *et al.*(2000).

(i'c) First remove irrelevant argument  $\gamma$  from the conditioning set. By applying Baye's theorem we get

$$P(\lambda|\{Y_g\}, \{W_g\}, \{Z_g\}, \beta, \sigma, \sigma_\alpha^2) \propto \prod_{g=1}^G P(Y_g|W_g, Z_g, \lambda, \beta, \sigma, \sigma_\alpha^2), \quad \lambda \in [-1/\tau_G, 1/\tau_G]. \quad (19)$$

(i'd) First remove irrelevant argument  $\gamma$  from the conditioning set. By applying Baye's theorem we get

$$P(\beta|\{Y_g\}, \{W_g\}, \{Z_g\}, \lambda, \sigma, \sigma_\alpha^2) \propto \pi(\beta) \cdot \prod_{g=1}^G P(Y_g|W_g, Z_g, \lambda, \beta, \sigma, \sigma_\alpha^2).$$

Since both  $\pi(\beta)$  and  $P(Y_g|W_g, Z_g, \lambda, \beta, \sigma, \sigma_\alpha^2)$  are in terms of normal density, we obtain the standard linear model result

$$P(\beta|\{Y_g\}, \{W_g\}, \{Z_g\}, \lambda, \sigma, \sigma_\alpha^2) \propto \mathcal{N}_{2k+1}(\hat{\beta}, \mathbf{B}), \quad (20)$$

where  $\hat{\beta} = \mathbf{B} \left( B_0^{-1} \beta_0 + \sum_{g=1}^G \mathbf{X}_g' V_g^{-1} (S_g Y_g - \sigma_{\epsilon z} \sigma_z^{-1} Z_g) \right)$  and  $\mathbf{B} = \left( B_0^{-1} + \sum_{g=1}^G \mathbf{X}_g' V_g^{-1} \mathbf{X}_g \right)^{-1}$ ;  
 $V_g = (\sigma_\epsilon^2 - \sigma_{\epsilon z}^2 \sigma_z^{-2}) I_{m_g} + \sigma_\alpha^2 l_{m_g} l_{m_g}'$ .

(i'e) First remove irrelevant argument  $\gamma$  from the conditioning set. By applying Baye's theorem we get

$$P(\sigma|\{Y_g\}, \{W_g\}, \{Z_g\}, \lambda, \beta, \sigma_\alpha^2) \propto \mathcal{T} \mathcal{N}_2(\sigma; \sigma_0, \Sigma_0) \prod_{g=1}^G P(Y_g|W_g, X_g, Z_g, \lambda, \beta, \sigma, \sigma_\alpha^2), \quad \sigma \in C. \quad (21)$$

(i'f) First remove irrelevant argument  $\gamma$  from the conditioning set. According to Chib (2008), we have

$$P(\alpha_g|Y_g, W_g, Z_g, \lambda, \beta, \sigma, \sigma_\alpha^2) \propto \mathcal{N}(\hat{\alpha}_g, D_g), \quad g = 1, \dots, G, \quad (22)$$

where  $\hat{\alpha}_g = (\sigma_\epsilon^2 - \sigma_{\epsilon z}^2 \sigma_z^{-2})^{-1} D_g l_{m_g}' (S_g Y_g - \mathbf{X}_g \beta - \sigma_{\epsilon z} \sigma_z^{-1} Z_g)$  and  $D_g = (\sigma_\alpha^{-2} + (\sigma_\epsilon^2 - \sigma_{\epsilon z}^2 \sigma_z^{-2})^{-1} l_{m_g}' l_{m_g})^{-1}$ .

(ii) First remove irrelevant argument  $\gamma$  from the conditioning set. According to Chib (2008), we have

$$P(\sigma_\alpha^2|\{Y_g\}, \{W_g\}, \{Z_g\}, \{\alpha_g\}, \lambda, \beta, \sigma) \propto \mathcal{JG}\left(\frac{\rho_0 + G}{2}, \frac{\eta_0 + \sum_{g=1}^G \alpha_g^2}{2}\right), \quad (23)$$

Except  $\beta$ ,  $\{\alpha_g\}$  and  $\sigma_\alpha^2$ , other conditional posterior distributions are not available in a closed form. Therefore, we use Metropolis-Hastings (M-H) algorithm to draw from these full conditional distributions. It has been shown in Tierney (1994), Chib and Greenberg (1996) that the combination of Markov chains (Metropolis-within-Gibbs) is still a Markov chain with the invariant distribution equal to the correct objective distribution. The procedure of MCMC sampling is starting with arbitrary initial values of parameters  $\theta_1^{(0)} = (\gamma^{(0)}, \lambda^{(0)}, \beta^{(0)}, \sigma_\alpha^{2(0)}, \sigma^{(0)})$ ,  $\{Z_g\}^{(0)}$  and  $\{\alpha_g\}^{(0)}$  and then sampling sequentially from the above set of conditional posterior distributions.

### Algorithm 1 – the Type-I model

Step 1. Sample  $Z_g$  from  $P(Z_g|Y_g, W_g, \theta_1^{(0)})$  specified in equation (17) using M-H algorithm for  $g = 1, \dots, G$ .

(a) Propose  $z'_{i,g} \sim \mathcal{N}(z_{i,g}^{(0)}, \xi_z^2)$ ,  $\xi_z^2$  is chosen by users.<sup>6</sup> Let  $Z'_g = (z_{1,g}^{(0)}, \dots, z_{i-1,g}^{(0)}, z'_{i,g}, z_{i+1,g}^{(0)}, \dots, z_{m_g,g}^{(0)})$ .

(b) With probability equal to

$$\alpha(Z_g^{(0)}, Z'_g) = \min \left\{ \frac{P(Y_g|W_g, Z'_g; \lambda^{(0)}, \beta^{(0)}, \sigma^{(0)}, \sigma_\alpha^{2(0)}) \prod_{i \neq j}^{m_g} P(w_{ij,g}|z'_{i,g}, z'_{j,g}, \gamma^{(0)}) \phi_{m_g}(Z'_g|0, I_{m_g})}{P(Y_g|W_g, Z_g^{(0)}; \lambda^{(0)}, \beta^{(0)}, \sigma^{(0)}, \sigma_\alpha^{2(0)}) \prod_{i \neq j}^{m_g} P(w_{ij,g}|z_{i,g}^{(0)}, z_{j,g}^{(0)}, \gamma^{(0)}) \phi_{m_g}(Z_g^{(0)}|0, I_{m_g})}, 1 \right\}$$

set  $z_{i,g}^{(1)}$  equal to  $z'_{i,g}$ . Otherwise set it to  $z_{i,g}^{(0)}$ . Repeat these two steps for  $i = 1, \dots, m_g$ .

Step 2. Sample  $\gamma$  from  $P(\gamma|\{W_g\}, \{Z_g\}^{(1)})$  specified in equation (18) using M-H algorithm.

(a) Propose  $\gamma' \sim \mathcal{N}_{q+2}(\gamma^{(0)}, \xi_\gamma^2 I_{q+2})$ ,  $\xi_\gamma^2$  is chosen by users.

(b) With probability equal to

$$\alpha(\gamma^{(0)}, \gamma') = \min \left\{ \prod_{g=1}^G \left( \frac{\prod_{i \neq j}^{m_g} P(w_{ij,g}|z_{i,g}^{(1)}, z_{j,g}^{(1)}, \gamma')}{\prod_{i \neq j}^{m_g} P(w_{ij,g}|z_{i,g}^{(1)}, z_{j,g}^{(1)}, \gamma^{(0)})} \right) \cdot \frac{\phi_{q+2}(\gamma'|\gamma_0, \Gamma_0)}{\phi_{q+2}(\gamma^{(0)}|\gamma_0, \Gamma_0)}, 1 \right\}$$

set  $\gamma^{(1)}$  equal to  $\gamma'$ . Otherwise set it to  $\gamma^{(0)}$ .

Step 3. Sample  $\lambda$  from  $P(\lambda|\{Y_g\}, \{W_g\}, \{Z_g\}^{(1)}, \beta^{(0)}, \sigma^{(0)}, \sigma_\alpha^{2(0)})$ ,  $\lambda \in [-1/\tau_G, 1/\tau_G]$  specified in equation (19) using M-H algorithm.

(a) Propose  $\lambda' \sim \mathcal{N}(\lambda^{(0)}, \xi_\lambda^2)$ ,  $\xi_\lambda^2$  is chosen by users.

---

<sup>6</sup>We choose  $\xi_z^2$  depending on the performance of the algorithm. Users can make  $\xi_z^2$  floatable during the loop. If the acceptance rate of  $Z'_g$  is too low, then lower  $\xi_z^2$ . Similarly, if the acceptance rate of  $Z'_g$  is too high, then raise  $\xi_z^2$ . This strategy can be applied to other steps which use M-H algorithm.

(b) With probability equal to

$$\alpha(\lambda^{(0)}, \lambda') = \min \left\{ \prod_{g=1}^G \left( \frac{p(Y_g|W_g, Z_g^{(1)}; \lambda', \beta^{(0)}, \sigma^{(0)}, \sigma_\alpha^{2(0)})}{p(Y_g|W_g, Z_g; \lambda^{(0)}, \beta^{(0)}, \sigma^{(0)}, \sigma_\alpha^{2(0)})} \right) \cdot \frac{I(\lambda' \in A)}{I(\lambda^{(0)} \in A)}, 1 \right\}, \quad A = [-1/\tau_G, 1/\tau_G].$$

set  $\lambda^{(1)}$  equal to  $\lambda'$ . Otherwise set it to  $\lambda^{(0)}$ .

Step 4. Sample  $\beta$  from  $P(\beta|\{Y_g\}, \{W_g\}, \{Z_g\}^{(1)}, \lambda^{(1)}, \sigma^{(0)}, \sigma_\alpha^{2(0)})$  specified in equation (20).

Step 5. Sample  $\sigma$  from  $P(\sigma|\{Y_g\}, \{W_g\}, \{Z_g\}^{(1)}, \lambda^{(1)}, \beta^{(1)}, \sigma_\alpha^{2(0)})$ ,  $\sigma \in C$  specified in equation (21) using M-H algorithm.

(a) Propose  $\sigma' \sim \mathcal{N}_2(\sigma^{(0)}, \xi_\sigma^2 I_2)$ ,  $\xi_\sigma^2$  is chosen by users.

(b) with the probability

$$\alpha(\sigma, \sigma') = \min \left\{ \prod_{g=1}^G \left( \frac{p(Y_g|W_g, Z_g^{(1)}; \lambda^{(1)}, \beta^{(1)}, \sigma', \sigma_\alpha^{2(0)})}{p(Y_g|W_g, Z_g^{(1)}; \lambda^{(1)}, \beta^{(1)}, \sigma^{(0)}, \sigma_\alpha^{2(0)})} \right) \frac{\phi_2(\sigma'|\sigma_0, \Sigma_0)I(\sigma' \in C)}{\phi_2(\sigma^{(0)}|\sigma_0, \Sigma_0)I(\sigma^{(0)} \in C)}, 1 \right\}$$

set  $\sigma^{(1)}$  equal to  $\sigma'$ . Otherwise set it to  $\sigma^{(0)}$ .

Step 6. Sample  $\alpha_g$  from  $P(\alpha_g|Y_g, W_g, Z_g^{(1)}, \lambda^{(1)}, \beta^{(1)}, \sigma^{(1)}, \sigma_\alpha^{2(0)})$  specified in equation (22) for  $g = 1, \dots, G$ .

Step 7. Sample  $\sigma_\alpha^2$  from  $P(\sigma_\alpha^2|\{Y_g\}, \{W_g\}, \{Z_g\}^{(1)}, \{\alpha_g\}^{(1)}, \lambda^{(1)}, \beta^{(1)}, \sigma^{(1)})$  specified in equation (23).

Step 8. Repeat Steps 1-7 using the most current values for the conditioning variables.

We collect the draws from the above procedures for a long run and used them to construct the posterior distributions of the parameters.

In the Type-II model there are also some identification problems surrounding the variable  $Z_g - \sigma_z^2$  is not identified and the signs of  $\delta_1$  and  $\delta_2$  are arbitrary. Again, we deal with them by normalizing the covariance matrix of  $Z_g$  to  $I_{m_g}$  and restricting the parameters  $\delta_1$  and  $\delta_2$  to be positive. The consequence of the first action is that  $\delta_1$  and  $\delta_2$  will absorb  $\sigma_z$  and become  $\delta_1\sigma_z$  and  $\delta_2\sigma_z$ . For estimating the Type-II model, we still employ the modified grouped Gibbs sampler to simulate draws for the parameter vector  $\theta_2 = (\gamma, \lambda, \beta, \delta, \mu_z, \sigma_u^2, \sigma_\alpha^2)$ , where  $\delta = (\delta_1, \delta_2)$  and two unknowns  $\{Z_g\}$  and  $\{\alpha_g\}$  – put  $\{Z_g\}, \gamma, \lambda, \beta, \delta, \mu_z, \sigma_u^2$  and  $\{\alpha_g\}$  into one block and  $\sigma_\alpha^2$  into another block. The complete estimating procedure for the Type-II model can be modified from

the Algorithm 1 by excluding  $\sigma$  and including  $\delta$ ,  $\mu_z$  and  $\sigma_u^2$  into the MCMC steps. To emphasize the identification of the parameter  $\mu_z$ , we can rewrite equation (6) as

$$Y_g = \lambda W_g Y_g + l_{m_g}(\beta_0 + \mu_z \delta_1) + X_g \beta_1 + W_g X_g \beta_2 + Z_g^* \delta_1 + W_g Z_g^* \delta_2 + W_g l_{m_g} \mu_z \delta_2 + l_{m_g} \alpha_g + u_g, \quad (24)$$

where  $Z_g^* = Z_g - l_{m_g} \mu_z \sim \mathcal{N}_{m_g}(0, I_{m_g})$ . From equation (24) we can observe that  $\mu_z$  is identifiable as long as  $W_g l_{m_g} \neq l_{m_g}$ . We define the prior distributions of  $\delta$ ,  $\mu_z$  and  $\sigma_u^2$  as follows:

$$\delta \sim \mathcal{TN}_2(\delta_0, D_0), \quad (25)$$

$$\mu_z \sim \mathcal{N}(0, M_0), \quad (26)$$

$$\sigma_u^2 \sim \mathcal{IG}\left(\frac{\nu_0}{2}, \frac{\omega_0}{2}\right), \quad (27)$$

where  $\delta$  is truncated to the area with both  $\delta_1$  and  $\delta_2$  positive. Joint with the Type-II model specified in equation (2) and equation (6) we can derive the conditional posterior distributions of  $\delta$ ,  $\mu_z$  and  $\sigma_u^2$  marginalized over  $\{\alpha_g\}$  as:

$$P(\delta | \{Y_g\}, \{W_g\}, \{Z_g\}, \lambda, \beta, \mu_z, \sigma_u^2, \sigma_\alpha^2) \propto \mathcal{TN}_2(\delta_0, D_0) \cdot \prod_{g=1}^G P(Y_g | W_g, Z_g, \lambda, \beta, \delta, \mu_z, \sigma_u^2, \sigma_\alpha^2). \quad (28)$$

$$P(\mu_z | \{Y_g\}, \{W_g\}, \{Z_g\}, \lambda, \beta, \delta, \sigma_u^2, \sigma_\alpha^2) \propto \mathcal{N}\left(\frac{\sum_{g=1}^G \sum_{i=1}^{m_g} z_{i,g}}{\sum_{g=1}^G m_g + 1/M_0}, \frac{1}{\sum_{g=1}^G m_g + 1/M_0}\right). \quad (29)$$

$$P(\sigma_u^2 | \{Y_g\}, \{W_g\}, \{Z_g\}, \lambda, \beta, \delta, \mu_z, \sigma_\alpha^2) \propto \mathcal{IG}\left(\frac{\nu_0 + \sum_{g=1}^G m_g}{2}, \frac{\omega_0 + \sum_{g=1}^G u'_g u_g}{2}\right). \quad (30)$$

We can now write down the complete MCMC procedure for the Type-II model:

### Algorithm 2 – the Type-II model

Step 1. Sample  $Z_g$  from  $P(Z_g | Y_g, W_g, \theta_2^{(0)})$ ,  $g = 1, \dots, G$  using M-H algorithm.

Step 2. Sample  $\gamma$  from  $P(\gamma | \{W_g\}, \{Z_g\}^{(1)})$  using M-H algorithm.

Step 3. Sample  $\lambda$  from  $P(\lambda | \{Y_g\}, \{W_g\}, \{Z_g\}^{(1)}, \beta^{(0)}, \delta^{(0)}, \mu_z^{(0)}, \sigma_u^{2(0)}, \sigma_\alpha^{2(0)})$ ,  $\lambda \in [-1/\tau_G, 1/\tau_G]$  using M-H algorithm.

Step 4. Sample  $\beta$  from  $P(\beta | \{Y_g\}, \{W_g\}, \{Z_g\}^{(1)}, \lambda^{(1)}, \delta^{(0)}, \mu_z^{(0)}, \sigma_u^{2(0)}, \sigma_\alpha^{2(0)})$ .

Step 5. Sample  $\delta$  from  $P(\delta|\{Y_g\}, \{W_g\}, \{Z_g\}^{(1)}, \lambda^{(1)}, \beta^{(1)}, \mu_z^{(0)}, \sigma_u^{2(0)}, \sigma_\alpha^{2(0)})$  using M-H algorithm.

Step 6. Sample  $\mu_z$  from  $P(\mu_z|\{Y_g\}, \{W_g\}, \{Z_g\}^{(1)}, \lambda^{(1)}, \beta^{(1)}, \delta^{(1)}, \sigma_u^{2(0)}, \sigma_\alpha^{2(0)})$ .

Step 7. Sample  $\sigma_u^2$  from  $P(\sigma_u^2|\{Y_g\}, \{W_g\}, \{Z_g\}^{(1)}, \lambda^{(1)}, \beta^{(1)}, \delta^{(1)}, \mu_z^{(1)}, \sigma_\alpha^{2(0)})$ .

Step 8. Sample  $\alpha_g$  from  $P(\alpha_g|Y_g, W_g, X_g, Z_g^{(1)}, \lambda^{(1)}, \beta^{(1)}, \delta^{(1)}, \mu_z^{(1)}, \sigma_u^{2(1)}, \sigma_\alpha^{2(0)}), g = 1, \dots, G$ .

Step 9. Sample  $\sigma_\alpha^2$  from  $P(\sigma_\alpha^2|\{Y_g\}, \{W_g\}, \{Z_g\}^{(1)}, \{\alpha_g\}^{(1)}, \lambda^{(1)}, \beta^{(1)}, \delta^{(1)}, \mu_z^{(1)}, \sigma_u^{2(1)})$ .

Step 10. Repeat Steps 1-9 using the most current values for the conditioning variables.

### 3.2 The case of binary dependent variable

For the case of binary dependent variable, we follow the approach of Albert and Chib (1993) to include the sampling of latent variables  $\{Y_g^*\}$  in the MCMC procedure along with other unobservables. To give more details of this approach, we assume  $\{Y_g^*\}$  follow the Type-I model in equation (5),

$$Y_g^* = \lambda W_g Y_g^* + l_{m_g} \beta_1 + X_g \beta_2 + W_g X_g \beta_3 + \sigma_{\epsilon z} \sigma_z^{-2} Z_g + l_{m_g} \alpha_g + u_g, \quad g = 1 \dots, G, \quad (31)$$

and  $u_g \sim \mathcal{N}_{m_g}(0, I_{m_g})$  by the usual normalization on the variance of  $u_g$ . From equation (31) we know  $Y_g^* \sim \mathcal{N}_{m_g}(S_g^{-1}(\mathbf{X}_g \beta + \sigma_{\epsilon z} \sigma_z^{-2} Z_g), \Sigma_g)$ , where  $\Sigma_g := \text{var}(Y_g^*) = S_g^{-1}(\sigma_\alpha^2 l_{m_g} l_{m_g}' + I_{m_g}) S_g^{-1'}$ . Using Baye's theorem we can write down the joint posterior distribution of  $\{Y_g^*\}$  and other unobservables as

$$\begin{aligned} & P(\theta_1, \{Z_g\}, \{\alpha_g\}, \{Y_g^*\} | \{Y_g\}, \{W_g\}) \\ & \propto \pi(\theta_1, \{Z_g\}, \{\alpha_g\}) \cdot \prod_{g=1}^G f(Y_g^* | \theta_1, Z_g, \alpha_g) \cdot P(Y_g, W_g | \theta_1, Z_g, \alpha_g, Y_g^*) \\ & \propto \pi(\theta_1, \{Z_g\}, \{\alpha_g\}) \cdot \prod_{g=1}^G \left\{ \prod_{i=1}^{m_g} \{ \mathbf{1}(y_{i,g}^* \geq 0) \mathbf{1}(y_{i,g} = 1) + \mathbf{1}(y_{i,g}^* < 0) \mathbf{1}(y_{i,g} = 0) \} P(Y_g^*, W_g | \theta_1, Z_g, \alpha_g) \right\}. \end{aligned} \quad (32)$$

The last line of equation (32) is derived from equation (8) and it much simplifies this joint posterior distribution. The Gibbs sampler corresponding to this model can be constructed based on the existing Gibbs steps for the Type-I model with one more step: drawing  $\{Y_g^*\}$  from the conditional posterior distribution of  $\{Y_g^*\}$  given other unobservables. Also note that we should remove  $\sigma_\epsilon^2$  from  $\theta_1$  because the variance of  $u_g$ ,  $\sigma_\epsilon^2 - \sigma_{\epsilon z}^2 \sigma_z^{-2}$  is normalized to one and hence  $\sigma_\epsilon^2$  will be no



longer identifiable. Let  $\theta'_1$  denote this new parameter set without  $\sigma_\epsilon^2$ . The conditional posterior distribution of  $Y_g^*$  given  $\theta'_1$ ,  $Z_g$  and data (marginalize over  $\alpha_g$ ) can be written as

$$P(Y_g^* | \theta_1, Z_g, Y_g, W_g) \propto \phi_{m_g}(Y_g^* | S_g^{-1}(\mathbf{X}_g \beta + \sigma_{\epsilon z} \sigma_z^{-2} Z_g), \Sigma_g) \prod_{i=1}^{m_g} \{\mathbf{1}(y_{i,g}^* \geq 0) \mathbf{1}(y_{i,g} = 1) + \mathbf{1}(y_{i,g}^* < 0) \mathbf{1}(y_{i,g} = 0)\}. \quad (33)$$

Equation (33) shows that the conditional posterior distribution of  $Y_g^*$  is in fact a truncated multivariate normal distribution. In order to sampling  $Y_g^*$  from this distribution we use the method in Geweke (1991) which will be explained below.

Consider generating  $Y_g^*$  from a multivariate normal distribution subject to linear inequality restrictions,

$$Y_g^* \sim \mathcal{N}_{m_g}(S_g^{-1}(\mathbf{X}_g \beta + \sigma_{\epsilon z} \sigma_z^{-2} Z_g), \Sigma_g), \text{ with } a_{i,g} \leq y_{i,g}^* \leq b_{i,g}, \quad i = 1, \dots, m_g, \quad (34)$$

it is equivalent to generating  $Y_g^{**}$  from

$$Y_g^{**} \sim \mathcal{N}_{m_g}(0, \Sigma_g), \text{ with } \alpha_{i,g} \leq y_{i,g}^{**} \leq \beta_{i,g}, \quad i = 1, \dots, m_g, \quad (35)$$

where  $\alpha_{i,g} = a_{i,g} - (S_g^{-1}(\mathbf{X}_g \beta + \sigma_{\epsilon z} \sigma_z^{-2} Z_g))_i$ ;  $\beta_{i,g} = b_{i,g} - (S_g^{-1}(\mathbf{X}_g \beta + \sigma_{\epsilon z} \sigma_z^{-2} Z_g))_i$  and then calculating  $Y_g^* = Y_g^{**} + S_g^{-1}(\mathbf{X}_g \beta + \sigma_{\epsilon z} \sigma_z^{-2} Z_g)$ . Utilizing the fact that the distribution of each element of  $Y_g^{**}$  conditional on all other elements of  $Y_g^{**}$  is truncated normal and known, we can employ the Gibbs sampler to draw  $y_{i,g}^{**}$  from  $m_g$  conditional distributions and these draws will converge to the distribution of  $Y_g^{**}$ . Suppose in the non-truncated distribution  $\mathcal{N}_{m_g}(0, \Sigma_g)$ ,

$$\mathbb{E}[y_{i,g}^{**} | y_{1,g}^{**}, \dots, y_{i-1,g}^{**}, y_{i+1,g}^{**}, \dots, y_{m_g,g}^{**}] = \sum_{j \neq i} c_g^{ij} y_{j,g}^{**}, \quad (36)$$

then in the truncated distribution of equation (35),  $y_{i,g}^{**}$  conditional on  $(y_{1,g}^{**}, \dots, y_{i-1,g}^{**}, y_{i+1,g}^{**}, \dots, y_{m_g,g}^{**})$  can be constructed as

$$y_{i,g}^{**} = \sum_{j \neq i} c_g^{ij} y_{j,g}^{**} + h_{i,g} \epsilon_{i,g}^{**}, \quad \epsilon_{i,g}^{**} \sim \mathcal{TN}(0, 1). \quad (37)$$

where  $\epsilon_{i,g}^{**}$  is truncated to the interval  $\left( \frac{\alpha_{i,g} - \sum_{j \neq i} c_g^{ij} y_{j,g}^{**}}{h_{i,g}}, \frac{\beta_{i,g} - \sum_{j \neq i} c_g^{ij} y_{j,g}^{**}}{h_{i,g}} \right)$ .

Denote  $c_g^i = (c_g^{i1}, \dots, c_g^{i,i-1}, c_g^{i,i+1}, \dots, c_g^{i,m_g})'$ , the coefficients  $c_g^{ij}$  and  $h_{i,g}$  in equation (36) and equation (37) are

$$c_g^i = -(\Sigma_g^{ii})^{-1} \Sigma_g^{i,<i}, \quad h_{i,g}^2 = (\Sigma_g^{ii})^{-1},$$

where  $\Sigma_g^{ii}$  is the element in row  $i$  and column  $i$  of  $\Sigma_g^{-1}$  and  $\Sigma_g^{i,<i}$  is row  $i$  of  $\Sigma_g^{-1}$  with  $\Sigma_g^{ii}$  deleted. In the case where the observation  $y_{i,g} = 1$ , we know  $y_{i,g}^* \geq 0$ , which is equivalent to  $y_{i,g}^{**} \geq -(S_g^{-1}(\mathbf{X}_g\beta + \sigma_{\epsilon z}\sigma_z^{-2}Z_g))_i$ . From equation (37) we get  $\epsilon_{i,g}^{**} \geq -(S_g^{-1}(\mathbf{X}_g\beta + \sigma_{\epsilon z}\sigma_z^{-2}Z_g))_i - \sum_{j \neq i} c_g^{ij} y_{j,g}^{**} / h_{i,g}$ . Therefore, we can simulate  $y_{i,g}^{**}$  by using the inverse-cdf method.

$$y_{i,g}^{**} = \sum_{j \neq i} c_g^{ij} y_{j,g}^{**} + h_{i,g} \Phi^{-1} [\Phi(\varpi_{i,g}) + U(1 - \Phi(\varpi_{i,g}))], \quad (38)$$

where  $\varpi_{i,g} = -(S_g^{-1}(\mathbf{X}_g\beta + \sigma_{\epsilon z}\sigma_z^{-2}Z_g))_i - \sum_{j \neq i} c_g^{ij} y_{j,g}^{**} / h_{i,g}$  and  $U$  is a uniform random variable on  $(0, 1)$ . Similarly, in the case where the observation  $y_{i,g} = 0$ ,  $y_{i,g}^{**}$  can be simulated by

$$y_{i,g}^{**} = \sum_{j \neq i} c_g^{ij} y_{j,g}^{**} + h_{i,g} \Phi^{-1} [U\Phi(\varpi_{i,g})]. \quad (39)$$

After each draw of  $y_{i,g}^{**}$ , we can recover  $y_{i,g}^* = y_{i,g}^{**} + (S_g^{-1}(\mathbf{X}_g\beta + \sigma_{\epsilon z}\sigma_z^{-2}Z_g))_i$ . The procedure of MCMC sampling for the type-I model with binary dependent variables can now be summarized as follows:

### Algorithm 3 – the Binary Type-I model

- Step 1. Sample  $\{Y_g^*\}$  from  $P(Y_g^* | \theta_1^{(0)}, Z_g^{(0)}, Y_g, W_g)$  with Gibbs sampler where each of the full conditional truncated normal variable can be drawn by the inverse-cdf method as specified in equation (38) and equation (39).
- Step 2. Sample  $\{Z_g\}$  from  $P(Z_g | Y_g^{*(1)}, W_g, \theta_1^{(0)})$ ,  $g = 1, \dots, G$  as specified in equation (17) using M-H algorithm.
- Step 3. Sample  $\gamma$  from  $P(\gamma | \{W_g\}, \{Z_g\}^{(1)})$  as specified in equation (18) using M-H algorithm.
- Step 4. Sample  $\lambda$  from  $P(\lambda | \{Y_g^{*(1)}\}, \{W_g\}, \{Z_g\}^{(1)}, \beta^{(0)}, \sigma_{\epsilon z}^{(0)}, \sigma_{\alpha}^{2(0)})$ ,  $\lambda \in [-1/\tau_G, 1/\tau_G]$  as specified in equation (19) using M-H algorithm.
- Step 5. Sample  $\beta$  from  $P(\beta | \{Y_g^{*(1)}\}, \{W_g\}, \{Z_g\}^{(1)}, \lambda^{(1)}, \sigma_{\epsilon z}^{(0)}, \sigma_{\alpha}^{2(0)})$  as specified in equation (20).
- Step 6. Sample  $\sigma_{\epsilon z}$  from  $P(\sigma_{\epsilon z} | \{Y_g^{*(1)}\}, \{W_g\}, \{Z_g\}^{(1)}, \lambda^{(1)}, \beta^{(1)}, \sigma_{\alpha}^{2(0)})$ ,  $\sigma_{\epsilon z} \geq 0$  using M-H algorithm.
- Step 7. Sample  $\{\alpha_g\}$  from  $P(\alpha_g | Y_g^{*(1)}, W_g, Z_g^{(1)}, \lambda^{(1)}, \beta^{(1)}, \sigma_{\epsilon z}^{(1)}, \sigma_{\alpha}^{2(0)})$ ,  $g = 1, \dots, G$  as specified in equation (22).

Step 8. Sample  $\sigma_\alpha^2$  from  $P(\sigma_\alpha^2 | \{Y_g^{*(1)}\}, \{W_g\}, \{Z_g\}^{(1)}, \{\alpha_g\}^{(1)}, \lambda^{(1)}, \beta^{(1)}, \sigma_{\epsilon z}^{(1)})$  as specified in equation (23).

Step 9. Repeat Steps 1-8 using the most current values for the conditioning variables.

## 4 Sampling performance of Bayesian estimates

In this section we examine the performance of the Bayesian estimation for the Type-I and the Type-II models with a Monte Carlo experiment. We consider both continuous dependent variable and binary dependent variable cases. In this experiment all groups are assumed to have the same size  $n$  and the number of groups is equal to  $G$ . We set  $(n, G) = (30, 50)$  for the case of continuous dependent variable and  $(n, G) = (30, 100)$  for the case of binary dependent variable. The experiment data are generated from the data generating process (DGP) specified as follows:

- DGP I<sup>7</sup>

- Distance model:  $\gamma_1 = -1.5$ ;  $\gamma_2 = 0.5$ ;  $\gamma_3 = -1$ .
- Type-I:  $\lambda = 0.05$ ;  $\beta_1 = 0.5$ ;  $\beta_2 = 0.5$ ;  $\beta_3 = 0.5$ ;  $\sigma_\alpha^2 = 0.5$ ;  $\sigma_z^2 = 1$ ;  $\sigma_{\epsilon z} = 0.5$ ;  $\sigma_\epsilon^2 = 1.25$ .
- Type-II:  $\lambda = 0.05$ ;  $\beta_1 = 0.5$ ;  $\beta_2 = 0.5$ ;  $\beta_3 = 0.5$ ;  $\delta_1 = 0.5$ ;  $\delta_2 = 0.5$ ;  $\sigma_\alpha^2 = 0.5$ ;  $\mu_z = 0.5$ ;  $\sigma_z^2 = 1$ ;  $\sigma_u^2 = 1$ .

We generate the independent variable  $C_g^1$  in the distance model by first drawing two  $n \times 1$  vectors of  $U(0, 1)$  random variables,  $U_{1,g}$  and  $U_{2,g}$ . If the  $i$ th element of  $U_{1,g}$  and the  $j$ th element of  $U_{2,g}$  are both larger or less than 0.5, we set  $C_{ij,g}^1$  equal to one. Otherwise, we set it to zero. The independent variable  $X_g$  is generated from  $\mathcal{N}_n(0, I_n)$ . The continuous dependent variable  $Y_g^*$  can be directly generated based on DGP I. We then generate the binary dependent variables by  $y_{i,g} = \mathbf{1}\{y_{i,g}^* \geq 0\}$  for  $i = 1, \dots, n$  and  $g = 1, \dots, G$ . In order to better understand the identification problems discussed for the Type-I and the Type-II models, we also consider DGP II which changes  $\sigma_z^2$  from 1 to 2 in DGP I. This change would be expected to affect the estimates of  $\gamma_2$  and  $\sigma_{\epsilon z}$  ( $\delta_1, \delta_2$ ) in the Type-I (Type-II) model.

- DGP II

---

<sup>7</sup>Given the generated  $W_g$ ,  $\lambda$  is set to 0.05 in order to satisfy the nonsingularity requirement on  $I_n - \lambda W_g$  as mentioned in footnote 3. The  $\sigma_\epsilon^2$  in the Type-I model is set to 1.25 such that under the binary dependent variable case when normalizing the variance of  $u_g$ ,  $(\sigma_\epsilon^2 - \sigma_{\epsilon z}^2 \sigma_z^{-2})I_{m_g}$  to  $I_{m_g}$  during the estimation will not cause any effects to the estimates of other coefficients.

Table 1: The average, minimum and maximum outdegree of generated networks

	DGP I				DGP II			
	Continuous		Binary		Continuous		Binary	
	Type-I	Type-II	Type-I	Type-II	Type-I	Type-II	Type-I	Type-II
Average	3.112	3.105	3.107	3.106	2.472	2.462	2.468	2.469
Minimum	0	0	0	0	0	0	0	0
Maximum	13	13	13	14	12	12	11	12

Note: The above results are calculated from 1,500 samples in the continuous case and 3,000 samples in the binary case across 50 repetitions.

- Distance model: same as DGP I.
- Type-I: same as DGP I except  $\sigma_z^2 = 2$  and  $\sigma_\epsilon^2 = 1.125$ .
- Type-II: same as DGP I except  $\sigma_z^2 = 2$ .

The MCMC sampling is implemented in a single long-run with the first 500 samples discarded for convergence. In order to produce approximate independence between each draws, we only collect every 10th point along the sample path and total 500 draws are collected. Finally, the number of Monte Carlo repetitions is set to 50. Given the networks generated by these two DGPs, we report the average, minimum and maximum outdegrees calculated from  $n \times G$  samples across 50 repetitions in Table 1. The parameters in prior distributions are specified as follows:  $\gamma_0 = (0, 0, 0)$ ;  $\Gamma_0 = 10I_3$ ;  $\beta_0 = (0, 0, 0)$ ;  $B_0 = 10I_3$ ;  $\sigma_0 = (0, 0)$ ;  $\Sigma_0 = I_2$ ;  $\rho_0 = 5$ ;  $\eta_0 = 1$ ;  $\delta_0 = (0, 0)$ ;  $D_0 = 10I_2$ ;  $M_0 = 2$ ;  $\nu_0 = 5$ ;  $\omega_0 = 1$ . These parameters are designed to allow relative flat prior densities over the range of the data.

Table 2 shows the experiment results for both cases of continuous and binary dependent variables using DGP I. The values shown for each parameter are the average and the standard deviation of the posterior means across 50 repetitions. The MCMC sampler is coded in MATLAB and implemented in a 2.60 GHz desktop computer. For the case of continuous dependent variable, it takes the Type-I (Type-II) model 3,800 (6,600) seconds to finish one repetition. The case of binary dependent variable takes longer, which is 9,700 and 16,000 seconds, respectively, for the Type-I and Type-II model. From the left panel of Table 2 we can observe that in the case of continuous dependent variable our Bayesian estimates for both the Type-I and the Type-II model are close

to the true values and have small standard deviations. In the case of binary dependent variable, the estimates for the Type-I model in the right panel of Table 2 are close to the values we expect. However, the estimates for the Type-II model exhibit some degrees of biases. The average of posterior means for  $\lambda$  is biased downward and the average of posterior means for  $\beta$ ,  $\delta$ ,  $\sigma_\alpha^2$  are biased upward. The standard deviations of posterior means for these parameters are also higher. Since the Type-II model with binary dependent variables is the most complicated model considered in our study, we may attribute this unexpected result to insufficient samples needed for obtaining accurate estimates.

Table 3 presents the experiment results using DGP II. When we change  $\sigma_z^2$  from 1 to 2, the coefficients  $\gamma_3$  and  $\sigma_{\epsilon z}$  ( $\delta$ ) in the Type-I (Type-II) model will absorb  $\sigma_z$  and change their magnitudes with  $\sqrt{2}$  (see reasons in Section 3.1). This change is clearly reflected in our estimates. Under the case of continuous dependent variable, we have the average of posterior means for  $\gamma_3$  equal to -1.419 (-1.420) in the Type-I (Type-II) model, the average of posterior means for  $\sigma_{\epsilon z}$  equal to 0.325 and the average of posterior means for  $\delta_1$  ( $\delta_2$ ) equal to 0.698 (0.703). Under the case of binary dependent variable, Most of our estimates for both Type-I and Type-II models are also close to true values except that The estimate of  $\sigma_{\epsilon z}$  in the Type-I model and the estimates of  $\beta$  and  $\delta$  in the Type-II model have some degrees of biases. Again these biases could be attributed to insufficient samples needed for obtaining accurate estimates.

## 5 Empirical Example

In this empirical example, we follow Lin (2010) and others applying our proposed models to the Add Health data which contains information on how each student interacts with others in the same group, i.e., the network structure of each group as well as outcome variables for students. The Add Health data is a nationally representative study of adolescents' health-related behaviors which covers students in grade 7 through 12 from 132 schools. With the original purpose on understanding how to protect young people from risky behaviors, such as using illicit drugs, smoking, drinking or having early and unprotected sex, the Add Health data covers detailed information about respondents' demographic backgrounds, academic outcomes, health related behaviors and most uniquely, their friendship networks. Four waves of surveys were conducted from 1994 to 2008, where Wave I in-school survey consists of a sample of 90,182 students and the following waves are in-home surveys and only cover subsets of the sample. Our study is based on Wave I in-school survey where each respondents were asked to nominate up to five male and five female friends

Table 2: Sampling performance of Bayesian estimators with DGP I

Variable	Continuous dependent variable				Binary dependent variable			
	Type-I		Type-II		Type-I		Type-II	
	Average	S.D.	Average	S.D.	Average	S.D.	Average	S.D.
$\gamma_1$	-1.504	0.039	-1.495	0.043	-1.507	0.029	-1.513	0.036
$\gamma_2$	0.504	0.039	0.499	0.039	0.501	0.024	0.507	0.026
$\gamma_3$	-0.997	0.048	-1.007	0.048	-0.991	0.041	-0.990	0.039
$\lambda$	0.049	0.009	0.054	0.011	0.053	0.014	0.044	0.017
$\beta_1$	0.500	0.097	0.487	0.112	0.493	0.079	0.547	0.140
$\beta_2$	0.499	0.029	0.497	0.028	0.495	0.029	0.559	0.086
$\beta_3$	0.499	0.017	0.495	0.023	0.500	0.030	0.553	0.063
$\delta_1$	-	-	0.490	0.056	-	-	0.548	0.140
$\delta_2$	-	-	0.495	0.032	-	-	0.623	0.093
$\sigma_\alpha^2$	0.515	0.101	0.465	0.113	0.476	0.105	0.556	0.208
$\sigma_\epsilon^2$	1.258	0.045	-	-	-	-	-	-
$\sigma_{\epsilon z}$	0.500	0.044	-	-	0.460	0.081	-	-
$\sigma_u^2$	-	-	1.074	0.043	-	-	-	-
$\mu_z$	-	-	0.502	0.079	-	-	0.489	0.060

1) Under the continuous dependent variable, the true values for  $(\gamma_1, \gamma_2, \gamma_3, \lambda, \beta_1, \beta_2, \beta_3, \delta_1, \delta_2, \sigma_\alpha^2, \sigma_\epsilon^2, \sigma_{\epsilon z}, \sigma_u^2, \mu_z)$ :

1a) (-1.5, 0.5, -1, 0.05, 0.5, 0.5, 0.5, -, -, 0.5, 1.25, 0.5, -, -) in the Type-I model;

1b) (-1.5, 0.5, -1, 0.05, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, -, -, 1, 0.5) in the Type-II model.

2) Under the binary dependent variable, the true values for  $(\gamma_1, \gamma_2, \gamma_3, \lambda, \beta_1, \beta_2, \beta_3, \delta_1, \delta_2, \sigma_\alpha^2, \sigma_\epsilon^2, \sigma_{\epsilon z}, \sigma_u^2, \mu_z)$ :

2a) (-1.5, 0.5, -1, 0.05, 0.5, 0.5, 0.5, -, -, 0.5, -, 0.5, -, -) in the Type-I model;

2b) (-1.5, 0.5, -1, 0.05, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, -, -, -, 0.5) in the Type-II model.

Table 3: Sampling performance of Bayesian estimators with DGP II

Variable	Continuous dependent variable				Binary dependent variable			
	Type-I		Type-II		Type-I		Type-II	
	Average	S.D.	Average	S.D.	Average	S.D.	Average	S.D.
$\gamma_1$	-1.501	0.046	-1.501	0.039	-1.508	0.033	-1.514	0.029
$\gamma_2$	0.507	0.039	0.503	0.041	0.504	0.027	0.503	0.023
$\gamma_3$	-1.419	0.071	-1.420	0.056	-1.406	0.052	-1.396	0.046
$\lambda$	0.051	0.010	0.055	0.013	0.056	0.017	0.045	0.018
$\beta_1$	0.501	0.098	0.503	0.104	0.483	0.067	0.530	0.129
$\beta_2$	0.499	0.029	0.498	0.029	0.494	0.031	0.546	0.067
$\beta_3$	0.500	0.021	0.495	0.029	0.492	0.029	0.554	0.059
$\delta_1$	-	-	0.698	0.048	-	-	0.771	0.106
$\delta_2$	-	-	0.703	0.039	-	-	0.846	0.143
$\sigma_\alpha^2$	0.506	0.098	0.472	0.121	0.454	0.097	0.509	0.159
$\sigma_\epsilon^2$	1.135	0.043	-	-	-	-	-	-
$\sigma_{\epsilon z}$	0.325	0.077	-	-	0.262	0.105	-	-
$\sigma_u^2$	-	-	1.073	0.052	-	-	-	-
$\mu_z$	-	-	0.349	0.055	-	-	0.358	0.048

1) Under the continuous dependent variable, the true values for  $(\gamma_1, \gamma_2, \gamma_3, \lambda, \beta_1, \beta_2, \beta_3, \delta_1, \delta_2, \sigma_\alpha^2, \sigma_\epsilon^2, \sigma_{\epsilon z}, \sigma_u^2, \mu_z)$ :

1a) (-1.5, 0.5, -1.414, 0.05, 0.5, 0.5, 0.5, -, -, 0.5, 1.125, 0.354, -, -) in the Type-I model;

1b) (-1.5, 0.5, -1.414, 0.05, 0.5, 0.5, 0.5, 0.707, 0.707, 0.5, -, -, 1, 0.354) in the Type-II model.

2) Under the binary dependent variable, the true values for  $(\gamma_1, \gamma_2, \gamma_3, \lambda, \beta_1, \beta_2, \beta_3, \delta_1, \delta_2, \sigma_\alpha^2, \sigma_\epsilon^2, \sigma_{\epsilon z}, \sigma_u^2, \mu_z)$ :

2a) (-1.5, 0.5, -1.414, 0.05, 0.5, 0.5, 0.5, -, -, 0.5, -, -, 0.354, -, -) in the Type-I model;

2b) (-1.5, 0.5, -1.414, 0.05, 0.5, 0.5, 0.5, 0.707, 0.707, 0.5, -, -, -, 0.354) in the Type-II model.

and this provides information about students’ friendship networks. Given the rich information provided in Add Health data, we would like to study peer effects on two important subjects. One is student’s academic achievement, i.e., GPA and another is a health related issue – smoking behavior. To study these two subjects it should apply our approaches for the case of continuous dependent variable and the case of binary dependent variable separately.

## 5.1 Data Summary

Before estimating our Type-I and Type-II model using the Add Health data, we first define the dependent and independent variables. The dependent variable GPA is calculated from the average of respondent’s grades in several subjects, including language, social science, mathematics and science. The highest GPA is equal to 4 and the lowest is equal to 1. The dependent variable Smoking takes the value 1 or 0 which indicates the respondent is a smoker or a non-smoker.<sup>8</sup> We choose and define independent variables in the SAR model following Lin (2010) and Lee *et al.* (2007, 2010) and the complete variable list is shown in Table 4. The sample consists of 70,639 observations available for the analysis of GPA and 75,586 observations available for the analysis of Smoking<sup>9</sup>. Then we assign those respondents into groups, which is defined by respondents’ school-grade. For the case of GPA (Smoking), we have total 574 (584) school-grade groups. We do not plan to use all groups for the analysis because that will take a tremendous time on computation. Instead, we only consider the groups which consist of the number of members between 10 to 100. The lower bound of the group size is chosen at 10 because we do not want the restriction on the group size to prevent respondents to name their 10 friends in the same group. We choose the upper bound at 100 because that still allows us to keep reasonably large amount of observations for analysis but largely decrease the time required for computation. After defining our sample as (intermediate) groups with the size between 10 to 100, we have 9,891 (8,233) observations left for the analysis of GPA (Smoking) and the average group size is equal to 68.5 (68). The corresponding group numbers is 181 (157) and the average number of nominated friends is equal to 2.74 (3.03). Excluding groups with members more than 100 from our sample does not cause a systematic downward bias on the average number of friends. When those large sized groups are added back to our sample and the

---

<sup>8</sup>The respondent is coded as non-smoker if she never smoked or just smoked once or twice during the last 12 months before being surveyed. Otherwise she is coded as smoker.

<sup>9</sup>We currently delete observations without complete information on variables listed in Table 4. This might not be an appropriate method to deal with network data since each individual plays a role in the network. An alternative method is to impute missing values using the Bayesian approach. This is a critical revision for this empirical study we are currently processing.



average number of nominated friends is calculated again, we get 2.80 (2.96) and that is similar to what we got from our sample.

In the distance model, we capture observed characteristics shared by each pair of respondents using four dummy variables – age, sex, race and whether both attend sport clubs or not. Table 4 presents the descriptive statistics of all variables under the intermediate groups and all groups. First we can observe that in the intermediate groups there is no significant difference between the sample used for GPA and the sample used for Smoking. Further comparing the descriptive statistics in the intermediate groups with those in all groups, we find that there are less Asian and Hispanic students in the intermediate groups. Also, students of intermediate groups tend to stay in the same school longer and have higher percentage in joining sport clubs. Otherwise, students of the intermediate groups show similar characteristics as those of all groups.

## 5.2 Estimation Results

Table 5 presents the estimation results for GPA. The estimate for each coefficient is the posterior mean constructed from MCMC draws and the value between the parenthesis is the standard deviation of the draws. In the MCMC sampling, the parameters in prior distributions and the length of sampling are same as used in our Monte Carlo experiment. From the left panel of Table 5, we can observe that the endogenous effect coefficient in the Type-I model is equal to 0.064, which means that on average if one of your friends raises (or lowers) his (her) GPA by one unit, the interaction with this friend will raise (or lower) your GPA by 0.064 unit. When we multiply this endogenous effect coefficient with the average number of nominated friends, which is 3.5 shown in Lin (2010), we get 0.221 and this is close to the endogenous effect coefficient 0.259 shown in Table 5 of Lin (2010). For own effects, our estimates show that students who have certain characteristics tend to have lower GPA, including older; male; black, hispanic or other races; mom’s education is less than high school or missing; mom’s job is missing or belonging to other categories, i.e., neither professional nor staying home. Most of our estimates for own effects have same signs as those shown in Lin (2010), except for the coefficient of whether mom received welfare or not, which is anyway not significant. Half of our estimates for contextual variables are imprecise given that their standard deviations are close to or larger than the corresponding estimates and hence we will not emphasize their impacts on students’ GPA. For remaining contextual variables, our estimates show that age, year, black, other race, less HS, edu missing, job missing and sport have negative effects on students’ GPA. The above results show that after controlling for endogenous association of members within a group, we still find evidences of endogenous effect and contextual

Table 4: Summary Statistics

variable	min	max	Intermediate groups only				All groups			
			GPA		Smoking		GPA		Smoking	
			mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
GPA	1	4	2.913	0.770	-	-	2.809	0.806	-	-
Smoking	0	1	-	-	0.348	0.476	-	-	0.357	0.479
Age	10	19	14.91	1.855	14.96	1.839	15.07	1.855	15.05	1.690
Years in school (Year)	1	6	3.513	1.808	3.688	1.834	2.500	1.417	2.511	1.424
Male	0	1	0.481	0.500	0.469	0.499	0.487	0.500	0.487	0.500
<i>Female</i>	0	1	0.519	0.500	0.531	0.499	0.512	0.500	0.512	0.500
<i>White</i>	0	1	0.595	0.491	0.618	0.486	0.562	0.496	0.558	0.497
Black	0	1	0.202	0.401	0.185	0.388	0.178	0.382	0.182	0.386
Asian	0	1	0.053	0.223	0.049	0.217	0.066	0.249	0.066	0.248
Hispanic	0	1	0.094	0.291	0.088	0.284	0.138	0.345	0.138	0.345
Other race	0	1	0.057	0.232	0.059	0.235	0.056	0.229	0.056	0.230
Participate in sport club (Sport)	0	1	0.611	0.488	0.607	0.489	0.524	0.499	0.526	0.499
<i>No sport club</i>	0	1	0.389	0.488	0.393	0.489	0.476	0.499	0.474	0.499
Live with both Parents (Both parents)	0	1	0.733	0.442	0.748	0.434	0.736	0.441	0.731	0.443
Mother education less than HS (Less HS)	0	1	0.098	0.297	0.100	0.300	0.100	0.300	0.100	0.300
<i>Mother education HS</i>	0	1	0.317	0.465	0.312	0.463	0.304	0.460	0.303	0.460
Mother education more than HS (More HS)	0	1	0.418	0.493	0.420	0.494	0.419	0.493	0.412	0.492
Mother education missing (Edu missing)	0	1	0.093	0.292	0.097	0.296	0.102	0.302	0.107	0.309
Mother job professional (Professional)	0	1	0.263	0.440	0.266	0.442	0.264	0.441	0.262	0.440
<i>Mother staying home</i>	0	1	0.210	0.407	0.219	0.414	0.205	0.403	0.204	0.403
Mother other jobs (Other Jobs)	0	1	0.359	0.480	0.351	0.477	0.362	0.481	0.358	0.479
Mother on welfare (Welfare)	0	1	0.009	0.097	0.008	0.091	0.008	0.087	0.009	0.092
Mother job missing (Job missing)	0	1	0.085	0.279	0.082	0.275	0.086	0.280	0.090	0.286

Note: Intermediate groups means groups with size between 10 to 100.

professional jobs include scientist, teacher, executive, director and the like.

other jobs refers to mother's occupation is not among "professional" or "staying home".

The variables in italics are omitted categories in the estimation.

effects in student academic achievement as shown in Lin (2010). In the distance model, all four dummy variables used to capture same observed characteristics shared by two individuals have positive effects on friendship formation, where the race shows the largest impact and then followed by the sex. Also, the estimate of  $\gamma_5$  is negative, which shows that the difference of unobserved characteristics between two students has a negative effect on friendship formation.

From the right panel of Table 5, we observe that the estimates for the SAR model and the distance model in the Type-II model are consistent with those in the Type-I model. The main differences between the Type-I model and the Type-II model are on the parameters designed to connect the distance model and the SAR model. In the Type-I model we have the estimate for  $\sigma_{\epsilon z}$  equal to 0.056 and in the Type-II model we have the estimates for  $\delta_1$  and  $\delta_2$  equal to 0.104 and 0.009, which are all significant. Given these estimates, we do not find strong correlations (even significant) between student's academic achievement and choices of friends in terms of unobserved variables  $Z_g$ .

The estimation result for Smoking is in Table 6. Similar to the case of GPA, The estimates for the SAR model and the distance model in both the Type-I and the Type-II models are highly consistent. Therefore, we will just look at the Type-I model for details of estimates. In the Type-I model, The endogenous effect coefficient is equal to 0.096. From this coefficient we can infer that on average if one of your friends raise (or lower) his (her) utility of smoking with one unit, by interacting with this friend you will raise (or lower) your utility of smoking by 0.096 unit. This interaction effect is larger than what we found in student's academic achievement.<sup>10</sup> From estimates of own effects, we can first observe that older students are more likely to smoke. Between different races, black, asian, and hispanic students are less likely to smoke compared to their white counterparts. Other factors such as living with both parents, mother's education is missing, and participating in sport clubs will also decrease the probability of smoking. For contextual effects, most of our estimates are imprecise and small. However, from the precise estimates we can still observe that the contextual variables such as Less HS, Professional and Other Jobs will contribute to students' smoking behaviors. The estimation result for the distance model in the study of Smoking is similar to what we got from the study of GPA. For the parameters connecting the distance model and the SAR model, we get the estimate of  $\sigma_{\epsilon z}$  in the Type-I model equal to 0.064 and  $\delta_1$ ,  $\delta_2$  in the Type-II model equal to 0.054 and 0.017. Given these results we find small correlations between students' smoking behaviors and choices of friends in terms of unobserved

---

<sup>10</sup>The direct interpretation based on the utility level does not tell us how likely students will change their smoking behaviors from interaction effects. Therefore, we would like to follow Lee *et al.* (2007) to calculate marginal effects based on the estimates. This job is now under processing.

variables  $Z_g$ .

## 6 Conclusion

This paper extends the SAR model used in Lin (2010) by considering endogenous associations of members within a group. When the spatial weights matrix is designed to represent the network structure within a group, it becomes endogenous and could be correlated with the disturbance term in the SAR model. When this correlation exists but is not considered in the model, we will have an endogeneity problem on the spatial weights matrix and standard estimation methods for the SAR model such as 2SLS and GMM methods which construct instruments or moment conditions from the spatial weights matrix would not be valid. We try to address this potential endogeneity problem with a new modeling approach. In this new approach, a statistical network model is adopted to explain the endogenous network formation process. Then we specify an unobserved component in both the network model and the SAR model. With the correlation between the network formation process and the outcome process built by this unobserved component, we may capture the selection problem caused by the endogenous association of members within a group.

The generalized models implied by our modeling approach are estimated with the Bayesian method. We consider both the case of continuous dependent variable and the case of binary dependent variable. The sampling performances of the Bayesian estimation method are examined with a Monte Carlo experiment. The estimates under continuous dependent variables are very close to true values, while the estimates under binary dependent variables display some degree of biases. These biases might be attributed to insufficient samples required to achieve accurate estimates. In the empirical example, we apply our generalized models to study students' academic achievement and smoking behaviors using the Add Health data. Even after controlling for endogenous association of members within a group, we still find evidences of endogenous effects and contextual effects in student academic achievement as shown in Lin (2010). From both the results of GPA and smoking behaviors, we only find some weak correlations between students' outcomes and the choices of friends in terms of unobserved characteristics.

For the future study, we will put more emphasis on modeling the network formation process. Our current strategy of modeling each link of the network with a pairwise regression is relying on the independence assumption between different links given latent position variables. However, as pointed out by Bramoullé and Fortin (2009), there are two problems accompanying with this strategy. First, it requires that the latent utility function behind the binary variable regression

Table 5: Peer Effect Estimation Results for GPA

SAR model	Type-I				Type-II			
	$\lambda$				$\lambda$			
	0.064 (0.004)				0.060 (0.004)			
	Own		Contextual		Own		Contextual	
Constant	3.542	(0.136 )	-		3.548	(0.137 )	-	
Age	-0.059	(0.009 )	-0.009	(0.001 )	-0.059	(0.009 )	-0.008	(0.001 )
Year	0.019	(0.005 )	-0.004	(0.002 )	0.018	(0.005 )	-0.005	(0.002 )
Male	-0.144	(0.015 )	0.005	(0.010 )	-0.146	(0.015 )	0.003	(0.009 )
Black	-0.143	(0.025 )	-0.019	(0.009 )	-0.144	(0.024 )	-0.023	(0.010 )
Asian	0.158	(0.036 )	-0.011	(0.012 )	0.141	(0.036 )	-0.022	(0.014 )
Hispanic	-0.140	(0.030 )	0.001	(0.015 )	-0.137	(0.029 )	0.002	(0.015 )
Other race	-0.058	(0.030 )	-0.028	(0.019 )	-0.063	(0.031 )	-0.031	(0.018 )
Both Parents	0.122	(0.017 )	0.012	(0.010 )	0.124	(0.017 )	0.017	(0.011 )
Less HS	-0.087	(0.025 )	-0.061	(0.017 )	-0.080	(0.025 )	-0.052	(0.016 )
More HS	0.163	(0.018 )	0.003	(0.009 )	0.160	(0.018 )	0.004	(0.010 )
Edu missing	-0.024	(0.025 )	-0.032	(0.018 )	-0.019	(0.026 )	-0.029	(0.018 )
Welfare	0.050	(0.072 )	-0.026	(0.059 )	0.027	(0.072 )	-0.045	(0.060 )
Job missing	-0.060	(0.029 )	-0.036	(0.018 )	-0.057	(0.027 )	-0.029	(0.018 )
Professional	0.035	(0.020 )	-0.010	(0.011 )	0.039	(0.020 )	-0.003	(0.012 )
Other Jobs	-0.025	(0.018 )	-0.007	(0.010 )	-0.022	(0.017 )	-0.003	(0.011 )
Sport	0.117	(0.015 )	-0.022	(0.008 )	0.115	(0.015 )	-0.022	(0.007 )
<b>Distance model</b>								
$\gamma_0$	-2.014	(0.068)			-2.005	( 0.043 )		
Age ( $\gamma_1$ )	0.141	(0.017)			0.141	( 0.016 )		
Sex ( $\gamma_2$ )	0.592	(0.017)			0.587	( 0.015 )		
Race ( $\gamma_3$ )	0.660	(0.024)			0.653	( 0.022 )		
Sport ( $\gamma_4$ )	0.340	(0.018)			0.344	( 0.018 )		
$\gamma_5$	-4.759	(0.296)			-4.779	( 0.197 )		
$\sigma_\alpha^2$	0.048	(0.007)			0.047	(0.007)		
$\sigma_\epsilon^2$	0.462	(0.007)			0.450	(0.007)		
$\sigma_{z\epsilon}$	0.056	(0.013)			-			
$\delta_1$	-				0.104	(0.013)		
$\delta_2$	-				0.009	(0.005)		
$\mu_z$	-				-0.013	(0.020)		

The number of groups is 181. The size of group is between 10 to 100.  $T = 500$ ;  $M = 500$ ;  $\ell = 10$ .

The standard deviation of MCMC draws is in the parenthesis.

Table 6: Peer Effect Estimation Results for Smoking

	Type-I				Type-II			
SAR model								
$\lambda$	0.096 (0.004)				0.095 (0.004)			
	Own		Contextual		Own		Contextual	
Constant	-2.224	(0.243 )	-		-2.238	(0.264)	-	
Age	0.127	(0.015 )	0.000	( 0.002)	0.127	(0.017)	0.000	(0.002)
Year	0.001	(0.011 )	0.002	( 0.005)	0.002	(0.011)	0.003	(0.005)
Male	0.001	(0.039 )	0.007	( 0.022)	0.004	(0.039)	0.006	(0.023)
Black	-0.607	(0.065 )	0.032	( 0.023)	-0.608	(0.063)	0.033	(0.023)
Asian	-0.282	(0.089 )	0.054	( 0.032)	-0.275	(0.101)	0.050	(0.033)
Hispanic	-0.141	(0.072 )	0.011	( 0.035)	-0.135	(0.069)	0.013	(0.034)
Other race	0.038	(0.072 )	0.021	( 0.040)	0.044	(0.069)	0.025	(0.040)
Both Parents	-0.169	(0.041 )	-0.078	( 0.027)	-0.167	(0.040)	-0.080	(0.023)
Less HS	0.089	(0.058 )	0.082	( 0.037)	0.087	(0.058)	0.076	(0.036)
More HS	-0.033	(0.044 )	0.007	( 0.020)	-0.033	(0.041)	0.006	(0.021)
Edu missing	-0.166	(0.070 )	0.065	( 0.044)	-0.167	(0.066)	0.058	(0.040)
Welfare	0.323	(0.161 )	0.012	( 0.134)	0.339	(0.178)	0.038	(0.134)
Job missing	-0.023	(0.069 )	0.064	( 0.046)	-0.027	(0.069)	0.066	(0.045)
Professional	-0.041	(0.051 )	0.065	( 0.029)	-0.036	(0.051)	0.068	(0.029)
Other Jobs	0.032	(0.041 )	0.046	( 0.025)	0.032	(0.042)	0.042	(0.025)
Sport	-0.125	(0.036 )	0.005	( 0.019)	-0.121	(0.038)	0.003	(0.018)
Distance model								
$\gamma_0$	-1.914 (0.069)				-1.917 (0.050)			
Age ( $\gamma_1$ )	0.129 (0.020)				0.128 (0.015)			
Sex ( $\gamma_2$ )	0.583 (0.016)				0.586 (0.015)			
Race ( $\gamma_3$ )	0.660 (0.027)				0.652 (0.018)			
Sport ( $\gamma_4$ )	0.337 (0.018)				0.353 (0.020)			
$\gamma_5$	-4.588 (0.416)				-4.649 (0.190)			
$\sigma_\alpha^2$	0.069 (0.014)				0.071 (0.015)			
$\sigma_{z\epsilon}^2$	0.064 (0.019)				-			
$\delta_1$	-				0.054 (0.023)			
$\delta_2$	-				0.017 (0.009)			
$\mu_z$	-				-0.031 (0.025)			

The number of groups is 157. The size of group is between 10 to 100.  $T = 500$ ;  $M = 500$ ;  $\ell = 10$ .

The standard deviation of MCMC draws is in the parenthesis.

is separable<sup>11</sup> in order to support the independence assumption between different links. But this assumption on the utility function is too strong to be held. Second, it assumes that every individuals in a group are potential partners and hence any pair of individuals, connected or not, is an observation to analyze. However, this assumption is problematic because usually time and social constraint will restrict potential partners into a subset of the group (Weinberg 2007). Our next goal will be constructing a new network formation process which can adress these two problems.

---

<sup>11</sup>This means that the utility derived from the network is equal to the sum of utilities from each link and not affected by the structure of the network.

## References

- [1] Aaronson, D. (1998). Using sibling data to estimate the impact of neighborhoods on children's educational outcomes. *Journal of Human Resources* 33, 915-46.
- [2] Boucher, V., Y. Bramoullé, H. Djebbari and B. Fortin (2010). Do peers affect student achievement? Evidence from Canada using group size variation. IZA Discussion Papers 4723, Institute for the Study of Labor (IZA).
- [3] Bramoullé, Y. and B. Fortin (2009). The econometrics of social networks. Cahiers de recherche 0913, CIRPEE.
- [4] Brock, W. A. and S. M. Durlauf (2001). Discrete choice with social interactions. *Review of Economic Studies* 68, 235-60.
- [5] Brock, W. A. and S. M. Durlauf (2007). Identification of binary choice models with social interactions. *Journal of Econometrics* 140, 52-75.
- [6] Chen, M.-H., Shao, Q.-M., and Ibrahim, J. G. (2000). *Monte Carlo Methods in Bayesian Computation*. New York: Springer-Verlag.
- [7] Chib, S. (2008). Panel data modeling and inference: A Bayesian primer. In Mátyás, L., Sevestre, P. (Eds.), *The Econometrics of Panel Data*, 3rd ed. Springer-Verlag, Berlin, Heidelberg, 479-515.
- [8] Chib, S. and E. Greenberg (1996). Markov chain monte carlo simulation methods in Econometrics. *Econometric Theory* 12, 409-31.
- [9] Chib, S. and B. P. Carlin (1999). On MCMC sampling in hierarchical longitudinal models. *Statistics and Computing* 9, 17-26.
- [10] Evans, W. N., W. E. Oates and R. M. Schwab (1992). Measuring peer group effects: A study of teenage behavior. *Journal of Political Economy* 100, 966-91.
- [11] Fletcher, J. M. and Marta T. (2008) High school peer networks and college success: Lessons from Texas. Unpublished working paper.
- [12] Gelfand, A. E., S. K. Sahu and B. P. Carlin (1995). Efficient parameterizations for normal linear mixed models. *Biometrika* 82, 479-88.



- [13] Geweke, J. (1991). Efficient simulation from the multivariate normal and student- $t$  distributions subject to linear constraints. In Keramidas, E., kaufman, S. (Eds.) *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*. Fairfax Station, VA: Interface Foundation of North America, 571-8.
- [14] Handcock, M. S., A. E. Raftery, and J. M. Tantrum (2007). Model-based clustering for social networks. *Journal of the Royal Statistical Society* 170, 301-54.
- [15] Hauser, C., M. Pfaffermayr, and G. Tappeiner (2009). Social capital formation and intra familial correlation: A social panel perspective. *The Singapore Economic Review* 54, 473-88.
- [16] Hoff, P. D., A. E. Raftery, and M. S. Handcock (2002). Latent Space Approaches to Social Network Analysis. *Journal of the American Statistical Association*, 97, 1090-98.
- [17] Kelejian, H. H. and I. R. Prucha (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics* 157, 53-67.
- [18] Krauth, B. V. (2006). Simulation-based estimation of peer effects. *Journal of Econometrics* 133, 243-71.
- [19] Lee, L. F. (2007), Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. *Journal of Econometrics*, 140, 333-74.
- [20] Lee, L. F., Li, J., and X. Lin (2007). Binary choice models with social interactions under heterogeneous rational expectations. Unpublished working paper.
- [21] Lee, L. F., Liu, X., and X. Lin (2010). Specification and estimation of social interaction models with network structures. Forthcoming in *The Econometrics Journal*.
- [22] LeSage, J. P. (1997). Bayesian estimation of spatial autoregressive models. *International Regional Science Review* 20, 113-29.
- [23] LeSage, J. P. (2000). Bayesian estimation of limited dependent variable spatial autoregressive models. *Geographical Analysis* 32, 19-35.
- [24] Lin, Xu (2010). Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables. Forthcoming in *Journal of Labor Economics*.
- [25] Liu, J. S. (1994). The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem. *Journal of the American Statistical Association* 89, 958-66.

- [26] Manski, C. F. (1993). Identification of endogenous social effects: the reflection problem. *Review of Economic Studies*, **60**, 531-42.
- [27] Mihaly, K. (2009). Do more friends mean better grades? Student popularity and academic achievement. RAND Working Paper Series WR-678.
- [28] Moffitt, R. A. (2001). Policy interventions, low-level equilibria, and social interactions. In Durlauf, S., Young, H. P. (Eds.), *Social Dynamics*. MIT Press, Cambridge, MA, 45-82.
- [29] Plotnick, R. D. and S. D. Huffman (1999). The effect of neighborhood characteristics on young adult outcomes: Alternative estimates. *Social Science Quarterly* 80, 1-18.
- [30] Rivkin, S. G. (2001). Tiebout sorting, aggregation and the estimation of peer group effects. *Economics of Education Review* 20, 201-09.
- [31] Sacerdote, B. (2001). Peer effects with random assignment: Results for Dartmouth roommates. *Quarterly Journal of Economics* 116, 681-704.
- [32] Smith, T. E. and J. P. LeSage (2004). A Bayesian probit model with spatial dependencies. In J. P. LeSage, Pace, R. K. (Eds.), *Spatial and Spatiotemporal Econometrics. Advances in Econometrics* Vol. 18. Elsevier Science, Oxford, UK, 127V160.
- [33] Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *Annals of Statistics* 22, 1701-62.
- [34] Udry, J. R. (2003). The national longitudinal study of adolescent health (add health), waves 1.
- [35] Wasserman, S. S. and K. Faust (1994). *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.
- [36] Weinberg, B. A. (2007), Social Interactions with Endogenous Associations. NBER Working Paper Series, Vol. w13038.
- [37] Zimmerman, D. J. (2003). Peer effects in academic outcomes: Evidence from a natural experiment. *The Review of Economics and Statistics* 85, 9-23.