

Econ 7218 Problem Set 3

Due by Monday, June 5, 2023

Consider a spatial autoregressive (SAR) model for studying social interactions,

$$y_{i,g} = \lambda \sum_{j=1}^{n_g} W_{ij,g} y_{j,g} + x_{i,g} \beta_1 + \sum_{j=1}^{n_g} W_{ij,g} x_{j,g} \beta_2 + \alpha_g + \epsilon_{i,g} \quad (1)$$

where $g = 1, \dots, G$ denote the groups (networks). $W_{ij,g} = 1$ if individual i links to individual j and zero otherwise. The vector form of this SAR model is

$$y_g = \lambda W_g y_g + x_g \beta_1 + W_g x_g \beta_2 + \alpha_g \ell_g + \epsilon_g, \quad (2)$$

where ℓ_g is a $n_g \times 1$ vector of ones. Assume $\epsilon_g | W_g, x_g \sim N(0, \sigma^2 I_{n_g})$, denote the common model parameters by $\theta = (\lambda, \beta', \sigma^2)'$ and group fixed effects by $\alpha = \{\alpha_1, \dots, \alpha_G\}$, the log-likelihood of the SAR model is given by

$$\begin{aligned} \mathcal{L}(\theta; \alpha) = & - \sum_g \frac{n_g}{2} \ln(2\pi) - \sum_g \frac{n_g}{2} \ln \sigma^2 + \sum_g \ln (\det(I_{n_g} - \lambda W_g)) \\ & - \frac{1}{2\sigma^2} \sum_g \epsilon_g(\lambda, \beta, \alpha)' \epsilon_g(\lambda, \beta, \alpha), \end{aligned} \quad (3)$$

where $\det(A)$ is the determinant of matrix A and $\epsilon_g(\lambda, \beta, \alpha) = (I_{n_g} - \lambda W_g)y_g - x_g \beta_1 - W_g x_g \beta_2 - \alpha_g \ell_g$. Notice that for the SAR model to be stationary, we need the coefficient λ to be properly bounded. One sufficient bound is $|\lambda| \leq 1/\|W_g\|_\infty$, for $g = 1, \dots, G$, where $\|W_g\| = \max_i \sum_{j=1}^{n_g} W_{ij,g}$ is the maximum row sum of W_g . You are given a sample of student networks and school outcome (GPA). The sample consists of 76 networks. Individual characteristics and GPA are available in the file *group*. The list of variables in each column is summarized in Table 1. The network adjacency matrix of each group is provided in the file *network*.

Questions:

1. Define the dependent variable y as student's GPA and the independent variables x using age, male, black, Asian, hisp, race other, less hs, more hs, momedu miss, welfare, momjob miss, prof, and job other. Estimate the common parameters (θ) of the SAR model in Equation (1) by the maximum likelihood method to study the peer effect and determination of student's GPA using the sample of 76 student networks. Report your parameter estimates and standard errors.

Hint: In this data set the maximum number of friends for each individual is 10, so you can set the constraint of λ to be $|\lambda| \leq 1/10$.

2. Consider the Bayesian estimation of the SAR model in Equation (1):

2-1 $\beta = (\beta'_1, \beta'_2)'$ is a $2k \times 1$ parameter vector, where k denotes the dimension of x . Assume the prior distribution of β , $\beta \sim \mathcal{N}(\beta_0, B_0)$, where β_0 is a $2k \times 1$ vector and B_0 is a $2k \times 2k$ matrix denoting fixed hyperparameters. Derive the conditional posterior distribution of β , i.e., $p(\beta|\{y_g\}, \{W_g\}, \{x_g\}, \lambda, \{\alpha_g\})$.

2-2 Assume the prior distribution of α_g , $\alpha_g \sim \mathcal{N}(\alpha_0, A_0)$, where α_0 and A_0 are fixed hyperparameters specified by a user. Derive the conditional posterior distribution of α_g , i.e., $p(\alpha_g|y_g, W_g, x_g, \lambda, \beta)$.

2-3 Using the mixed MCMC strategy (M-H within Gibbs) to estimate the common parameters (θ) of the SAR model in Equation (1) to study the peer effect and determination of student's GPA. Report the posterior mean and posterior standard deviation for each parameter as a point estimate.

Turn in your code and a short write-up which includes the results and algorithms on NTU COOL.

Table 1: Variable definitions

Variable	Definition
age	student's age
male	male dummy
black	black dummy
Asian	asian dummy
hisp	hispanic dummy
race other	other race dummy
both par	living with both parents
less hs	mom's education less than high school
more hs	mom's education higher than high school
momedu miss	mom's education is missing
welfare	mom participate in welfare programs
momjob miss	mom's job is missing
prof	mom's job is professional
job other	mom's job is other types
sport	participate in sport clubs
white	white dummy
yr school	number of years stayed in this school
gpa	GPA
overage	dummy of older than the average age in the grade