1 Introduction

2 Bayesian estimation for SAR model

2.1 SAR model

$$\mathbf{y} = \rho \mathbf{W} \mathbf{y} + \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{1}$$

$$\boldsymbol{\epsilon} \sim MVN\left(\mathbf{0}, \mathbf{V} = \sigma^2 \mathbf{I}\right) \tag{2}$$

$$\mathbf{B}_0 = \left[\begin{array}{cccc} B_{1,1} & B_{1,2} & \cdots & B_{1,n} \\ B_{2,1} & B_{2,2} & \cdots & B_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ B_{n,1} & B_{n,2} & \cdots & B_{n,n} \end{array}\right] \tag{3}$$

$$\mathbf{W}_0 = \left[\begin{array}{cccc} W_{1,1} & W_{1,2} & \cdots & W_{1,n} \\ W_{2,1} & W_{2,2} & \cdots & W_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ W_{n,1} & W_{n,2} & \cdots & W_{n,n} \end{array}\right] \tag{4}$$

$$p\left(\mathbf{y} \mid \rho, \boldsymbol{\beta}, \sigma^2\right) = \frac{\det\left(\mathbf{A}\right)}{\left(2\pi\right)^{n/2} \det\left(\mathbf{V}\right)^{1/2}} \exp\left(-\frac{1}{2} \boldsymbol{\epsilon}^T \mathbf{V}^{-1} \boldsymbol{\epsilon}\right) \tag{5}$$

$$\boldsymbol{\epsilon} = \mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta} \tag{6}$$

$$\mathbf{A} = \mathbf{I} - \rho \mathbf{W} \tag{7}$$

2.2 Bayesian estimation for SAR model

$$p\left(\rho, \boldsymbol{\beta}, \sigma^2\right) = p\left(\rho\right) p\left(\boldsymbol{\beta}, \sigma^2\right) \tag{8}$$

$$p\left(\boldsymbol{\beta}, \sigma^2\right) = p\left(\boldsymbol{\beta} \mid \sigma^2\right) p\left(\sigma^2\right) \tag{9}$$

$$\boldsymbol{\beta} \mid \sigma^2 \sim MVN\left(\mathbf{c}_0, \mathbf{D}_0\right) \tag{10}$$

$$\sigma^2 \sim IG\left(\text{shape} = a_0, \text{scale} = b_0\right) \tag{11}$$

$$p\left(\sigma^2\right) = \frac{b_0^{a_0}}{\Gamma\left(a_0\right)} \frac{1}{\left(\sigma^2\right)^{a_0+1}} \exp\left(-\frac{b_0}{\sigma^2}\right) \quad \sigma^2 > 0 \tag{12}$$

$$\rho \sim U\left(l_0, u_0\right) \tag{13}$$

$$p\left(\rho\right) = \frac{1}{u_0 - l_0} \quad l_0 \leq \rho \leq u_0 \tag{14}$$

$$p\left(\rho, \boldsymbol{\beta}, \sigma^2 \mid \mathbf{y}\right) \propto p\left(\mathbf{y} \mid \rho, \boldsymbol{\beta}, \sigma^2\right) \times p\left(\rho\right) p\left(\boldsymbol{\beta} \mid \sigma^2\right) p\left(\sigma^2\right) \tag{15}$$

3 Bayesian case influence analysis for SAR model

3.1 Bayesian case influence analysis

$$p^*(\boldsymbol{\theta} \mid \mathbf{y}) = p(\mathbf{y} \mid \boldsymbol{\theta}) \times p(\boldsymbol{\theta}) \tag{16}$$

$$p^*(\boldsymbol{\theta} \mid \mathbf{y}_{-k}) = p(\mathbf{y}_{-k} \mid \boldsymbol{\theta}) \times p(\boldsymbol{\theta}) \tag{17}$$

$$w_{-k}\left(\boldsymbol{\theta}^{(r)}\right) = \frac{p^*\left(\boldsymbol{\theta}^{(r)} \mid \mathbf{y}_{-k}\right)}{p^*\left(\boldsymbol{\theta}^{(r)} \mid \mathbf{y}\right)} \tag{18}$$

3.2 Main result: Bayesian case influence analysis for SAR model

$$\underset{(n-1\times n)}{\mathbf{D}_k} = \begin{bmatrix} 1 & & & 0 & & & \\ & \ddots & & \vdots & & & \\ & & 1 & 0 & & & \\ & & & 0 & 1 & & \\ & & & \vdots & & \ddots & \\ & & & 0 & & & 1 \end{bmatrix} \tag{19}$$

$$\underset{(n\times 1)}{\mathbf{s}_k} = (0, \ldots, 0, 1, 0, \ldots, 0)^T \tag{20}$$

$$\mathbf{y}_{-k} = \mathbf{D}_k \mathbf{y} \tag{21}$$

$$\begin{aligned} \mathbf{y}_{-k} =& \rho \mathbf{D}_k \mathbf{W} \mathbf{D}_k^T \mathbf{y}_{-k} + \rho^2 \mathbf{D}_k \mathbf{W} \mathbf{s}_k \mathbf{s}_k^T \mathbf{W} \mathbf{D}_k^T \mathbf{y}_{-k} \\ &+ \mathbf{D}_k \mathbf{X} \boldsymbol{\beta} + \rho \mathbf{D}_k \mathbf{W} \mathbf{s}_k \mathbf{s}_k^T \mathbf{X} \boldsymbol{\beta} \\ &+ \mathbf{D}_k \boldsymbol{\epsilon} + \rho \mathbf{D}_k \mathbf{W} \mathbf{s}_k \mathbf{s}_k^T \boldsymbol{\epsilon} \end{aligned} \tag{22}$$

$$\boldsymbol{\epsilon}_{-k} = \mathbf{D}_k \boldsymbol{\epsilon} + \rho \mathbf{D}_k \mathbf{W} \mathbf{s}_k \mathbf{s}_k^T \boldsymbol{\epsilon} \tag{23}$$

$$\boldsymbol{\epsilon}_{-k} \sim MVN\left(\mathbf{0}, \mathbf{V}_{-k} = \sigma^2 \left(\mathbf{D}_k \mathbf{D}_k^T + \rho^2 \mathbf{D}_k \mathbf{W} \mathbf{s}_k \mathbf{s}_k^T \mathbf{s}_k \mathbf{s}_k^T \mathbf{W}^T \mathbf{D}_k^T\right)\right) \tag{24}$$

$$p\left(\mathbf{y}_{-k} \mid \rho, \boldsymbol{\beta}, \sigma^2\right) = \frac{\det\left(\mathbf{A}_{-k}\right)}{(2\pi)^{(n-1)/2} \det\left(\mathbf{V}_{-k}\right)^{1/2}} \exp\left(-\frac{1}{2} \boldsymbol{\epsilon}_{-k}^T \mathbf{V}_{-k}^{-1} \boldsymbol{\epsilon}_{-k}\right) \tag{25}$$

$$\boldsymbol{\epsilon}_{-k} = \mathbf{A}_{-k} \mathbf{y}_{-k} - \mathbf{B}_{-k} \mathbf{X} \boldsymbol{\beta} \tag{26}$$

$$\mathbf{A}_{-k} = \mathbf{I} - \rho \mathbf{D}_k \mathbf{W} \mathbf{D}_k^T - \rho^2 \mathbf{D}_k \mathbf{W} \mathbf{s}_k \mathbf{s}_k^T \mathbf{W} \mathbf{D}_k^T \tag{27}$$

$$\mathbf{B}_{-k} = \mathbf{D}_k + \rho \mathbf{D}_k \mathbf{W} \mathbf{s}_k \mathbf{s}_k^T \tag{28}$$

$$w_{-k}\left(\rho^{(r)}, \boldsymbol{\beta}^{(r)}, \sigma^{2(r)}\right) = \frac{p\left(\mathbf{y}_{-k} \mid \rho^{(r)}, \boldsymbol{\beta}^{(r)}, \sigma^{2(r)}\right)}{p\left(\mathbf{y} \mid \rho^{(r)}, \boldsymbol{\beta}^{(r)}, \sigma^{2(r)}\right)} \tag{29}$$

4 Examples on artificial data

4.1 Rationale

Our objective is to identify outliers given a fitted simultaneous autoregressive (SAR) model. Consider n spatial units and p parameters, we derived the analytic expression to case influence analysis under the Gaussian assumption.

We then run through the Bayesian estimation for p parameters. It can be shown that $\boldsymbol{\beta}$ and σ^2 follow standard conjugacy results, but ρ needs extra care. In particular, the conditional posterior of the spatial lag coefficient ρ does not take any known form. As a result, we draw $\boldsymbol{\beta}$ and σ^2 via Gibbs and ρ via rejection algorithm.

Coupling the analytic expression with R retained Markov chain Monte Carlo (MCMC) draws, we end up getting an $R \times n$ matrix. Each element calculates the case influence statistic given a set of MCMC parameters draw with single spatial observation removed.

To complete the analysis, we perform a principal component analysis on the original $R \times n$ matrix of case-deletion importance sampling weights. The consequence of removing single cases is mapped to a reduced space (usually two or three dimensions given by the leading principal components). When higher leverage data points are removed, the associated columns in the $R \times n$ matrix will have relatively high loadings (given by the corresponding eigenvectors or singular vectors). We employ standard biplot for visualisation.

4.2 Simulation design

We now perform a simulation exercise to examine the efficacy in identifying influential SAR observations. Thomas et al. (2018) includes artificial and empirical examples for models exhibiting conditional independence and Markovian dependence. The similar procedure is applied here, but under a SAR dependence structure. Results are compared and contrasted.

Again, our goal is to detect abnormal data points which do not agreed with the assumed model. Therefore, we consider the following:

1. Assume conditional independence
   (a) No outlier
   (b) Single outlier with leverage varies
2. Assume SAR dependence and single outlier with leverage varies
   (a) Examine the effect of changing SAR parameter
   (b) Examine the effect of changing neighbourhood structure

Note that 1a and 1b are straightforward iterations of Thomas et al. (2018), which are nested in our proposed SAR framework. Since the degree of spatial dependence is governed by ρ (which is the regression coefficient on the heterogenous spatial lag $\mathbf{Wy}$) and the spatial weight matrix $\mathbf{W}$ (given by the binary network $\mathbf{B}$), changing one of the SAR parameter and neighbourhood structure examines the performance of our methodology.

4.3 Cases 1 and 2

In previous section, we see that the case influence statistic is proportional to the ratio of two likelihood functions, conditional on case-deleted and full data sets respectively. Being a function of the parameters, we can evaluate the likelihood function at different MCMC draws $\boldsymbol{\theta}^{(r)}$. In order to draw samples from the targeted posterior, we need n spatial units and prior assumption on parameters.

First, we simulate $n = 50$ observations from the SAR model (for some ρ and $\mathbf{B}$ values), $\mathbf{y} = (y_1, y_2, \ldots, y_{50})^T$. Next, a single data point is randomly selected to be the artificial *additive* outlier. Here, y_{23} is being chosen, where δ times the corresponding standard deviation σ_{23} is added to y_{23} :

$$y_{23} = y_{23} + \delta\sigma_{23}$$

Note, δ controls the degree of leverage.

The foregoing discussions mainly focus on cases 1a and 1b, by defining $\rho = 0$. Now, let us consider cases 2a and 2b. To simplify the discussion, we set $\mathbf{B}$ fixed when investigating ρ and vice versa. For a given spatial connection network, $\mathbf{B} = \mathbf{B}_1$:

$$\mathbf{B}_1 = \begin{bmatrix} & 1 & & \\ 1 & & \ddots & \\ & \ddots & & 1 \\ & & 1 & \end{bmatrix}$$

we simulate different $\mathbf{y}$s by varying ρ (see Table 1 for details). On the other hand, we study the effect of changing neighbourhood structure by defining $\delta = 8$ and $\rho = 0.9$. That is, we assume strong positive spatial autocorrelation and the existence of a high leverage data point. Now, we modify the spatial *inflow* and *outflow* structures around the outlier y_{23} (Karlsson et al. 2015). For example, blue ones represent spatial inflow to y_{23} from $\mathbf{y}_{-23}$, and purple ones represent

[1] Exercise 2b considers different spatial weight matrices based on this δ/ρ combination.

Table 1: Larger δ and ρ produce higher leverage data and stronger SAR dependence, respectively. Black stars denote cases 1a and 1b. White stars denote exercise 2a.

$\rho \backslash \delta$	0	2	4	6	8
0	★		★		★
0.1	☆		☆		☆
0.3					
0.5	☆		☆		☆
0.7					
0.9	☆		☆		☆[1]

spatial outflow from y_{23} to $\mathbf{y}_{-23}$:

$$
\mathbf{B}_i = \begin{bmatrix}
 & & & \vdots & & & \\
\ddots & \ddots & \ddots & 1 & & & \\
 & 1 & 0 & 1 & & & \\
\cdots & 1 & 1 & 0_{23} & 1 & 1 & \cdots \\
 & & & 1 & 0 & 1 & \\
 & & & 1 & \ddots & \ddots & \ddots \\
 & & & \vdots & & &
\end{bmatrix} \quad i = 2, 3, 4
$$

Say if there exist spatial inflow to y_i from y_j and outflow from y_i to y_j simultaneously, we treat y_i and y_j as neighbours, provided $i \neq j$. In essence, we are manipulating the number of neighbours (around y_{23}) in exercise 2b. Note that, we row standardised $\mathbf{B}_i$ to get the respective spatial weight matrices $\mathbf{W}_i$.

4.4 Output

See Figure 1 and Figure 2.

5 Empirical study

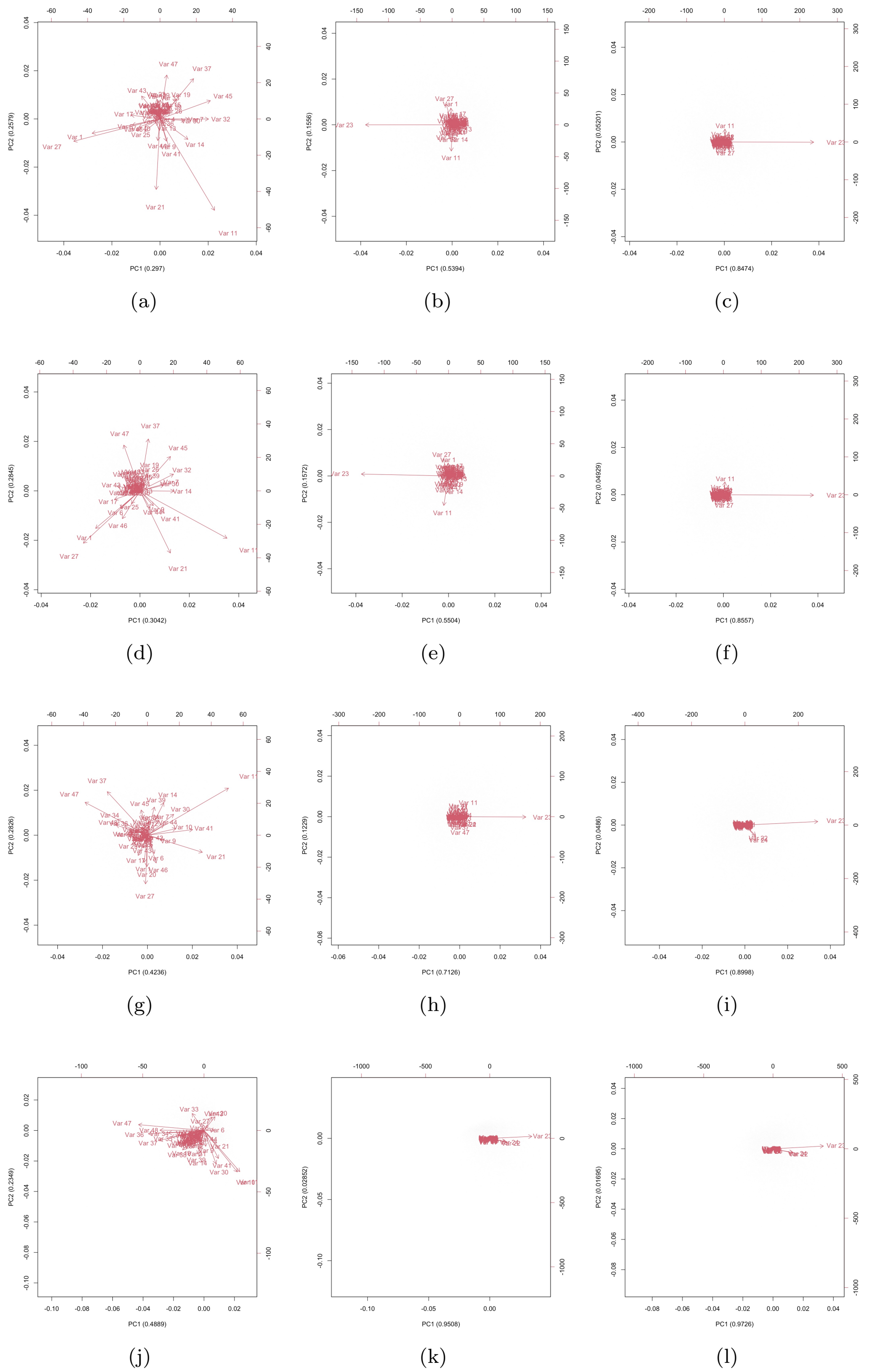
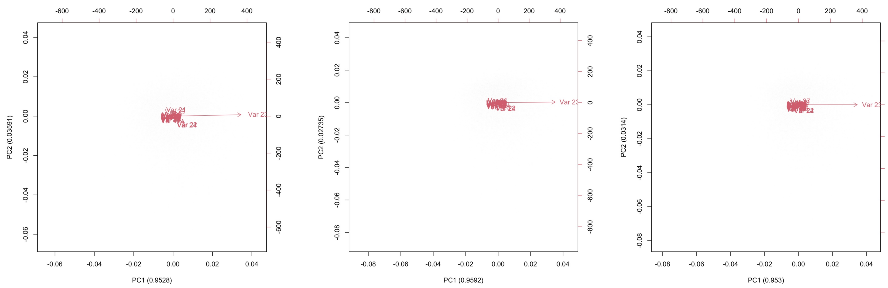


Figure 1: The order of this figure matrix matches to the δ/ρ combinations as in Table 1. Each figure showcased the corresponding biplot of principal component analysis on the log case-deletion importance sampling weights matrix.



(a) (b) (c)

Figure 2: Biplots showcasing an additional of two, four and six neighbours (for the outlier y_{23}), respectively.

References

Karlsson, C., Andersson, M. & Norman, T. (2015), *Handbook of Research Methods and Applications in Economic Geography*, Edward Elgar Publishing.

Thomas, Z. M., MacEachern, S. N. & Peruggia, M. (2018), 'Reconciling curvature and importance sampling based procedures for summarizing case influence in bayesian models', *Journal of the American Statistical Association* **113**(524), 1669–1683.

A Math background

A.1 PCA

Data

$$\mathbf{X} = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,p} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,p} \end{bmatrix}$$

Covariance

$$s_{j,k} = \frac{1}{n} \sum_{i=1}^{n} \left(x_{i,j} - \bar{x}_j\right)\left(x_{i,k} - \bar{x}_k\right) \quad j,k = 1,2,\ldots,p$$

Correlation

$$r_{j,k} = \left. \frac{\sum_{i=1}^{n} \left(x_{i,j} - \bar{x}_j\right)\left(x_{i,k} - \bar{x}_k\right)}{n} \middle/ s_j s_k \right.$$

Eigendecomposition

$$\mathbf{A} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{-1}$$

Eigenvalue

$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_p \end{bmatrix}$$

Eigenvector (`rotation` in `prcomp` and `loadings` in `princomp`)

$$\mathbf{V} = \begin{bmatrix} v_{1,1} & v_{1,2} & \cdots & v_{1,p} \\ v_{2,1} & v_{2,2} & \cdots & v_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ v_{p,1} & v_{p,2} & \cdots & v_{p,p} \end{bmatrix}$$

Reduced data (`x` in `prcomp` and `scores` in `princomp`)

$$\widetilde{\mathbf{X}} = \mathbf{X}\mathbf{V}$$

A.2 sPCA

Spatial data

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Spatial weight

$$\mathbf{W} = \begin{bmatrix} w_{1,1} & w_{1,2} & \cdots & w_{1,n} \\ w_{2,1} & w_{2,2} & \cdots & w_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n,1} & w_{n,2} & \cdots & w_{n,n} \end{bmatrix}$$

Spatial correlation (I and $LISA$)

$$\begin{aligned} I &= \left. \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j} \left(x_i - \bar{x}\right) \left(x_j - \bar{x}\right)}{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j}} \middle/ s^2 \right. \\ &= \sum_{i=1}^{n} \left. \frac{I_i}{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j}} \middle/ s^2 \right. \end{aligned}$$

A.3 Spatial filtering