A Gibbs Sampler for Multilevel Variance Components Models with Separable Spatial Effects

Levi John Wolf

August 6, 2016

Abstract

When stated, a multilevel variance components model with spatial effects can appear to look somewhat difficult to sample. However, by noting two inherent properties of the problem structure, one can derive a sampler structured around a separable univariate scale parameter and univariate spatial correlation parameter for spatial covariance structures where correlation is separable from a univariate scale parameter *at arbitrary levels*. This note proceeds to define the following Gibbs sampler for variance components models of the form:

$$Y = X\beta + \Delta\alpha + \epsilon$$

where the covariance of α, ϵ will be shown to be defined by the separable spatial correlation and scale parameters, $\Psi(\lambda), \sigma_\alpha^2$ and $\Psi(\rho), \sigma_e^2$, respectively. Using this form, a Gibbs sampler for multilevel variance components models with spatial effects at any level is possible. Having an exact solution for this problem provides exact samplers for the following combinations of Spatial Moving Average (SMA) and Simultaneous Autoregressive Spatial Error models (SE) for the variance components at different levels:

- SMA(1), SMA(2), SMA(1,2)
- SE(1), SE(2), SE(1,2)
- SE(1)-SMA(2), SMA(1)-SE(2)

Extending this result to broader classes of models would require defining the correct conditional posteriors for models with endogenous SAR-lag components at any level or combinations thereof.

1 Defining the problem

In Bayesian multilevel modeling literature, the following styles of models are typically called "variance components" models [3]:

$$Y = X\beta + \Delta\alpha + \epsilon \tag{1}$$

Where $X\beta$ is the typical linear predictor which may or may not contain an upper-level predictor component, $\Delta Z\gamma$, with Δ being the dummy variable matrix relating the lower and upper levels. Then, $\Delta\alpha + \epsilon$ is intended to capture the relation between a group-wise variance and an individual-level variance. These *variance components* models are used to partition variance in multilevel setups, and are common in school efficacy literature.

As noted in the review paper, these two error terms, $\Delta\alpha$ and ϵ can be unified without problem and joint covariance matrices derived:

$$\begin{aligned} Y &= X\beta + v \\ v &= \Delta\alpha + \epsilon \\ v &\sim \mathcal{N}(0, \Sigma_v) \end{aligned} \tag{2}$$

And, the review walks through various forms this dependence can take, conditional on where and what kind of spatial dependence is introduced. Focusing only on cases where the error is in the SMA or SAR-error forms, we can state the following fully:

$$\begin{aligned} Y &= X\beta + v \\ v &= \Delta\alpha + \epsilon \\ \epsilon &\sim \mathcal{N}(0, \Psi(\rho, \sigma_e^2)) \\ \alpha &\sim \mathcal{N}(0, \Psi(\lambda, \sigma_\alpha^2)) \end{aligned} \tag{3}$$

Where we parameterize the the variance in spatial structure is, in part an intrinsic variance term, $\sigma_.^2$ and due to a function of spatial relationships, $\Psi(.)$. Note that, at this point, the scale and correlation are *not separable*. This means that a full statement of a model for Y is:

$$Y = X\beta + \Delta\mathcal{N}(0, \Psi(\lambda, \sigma_\alpha^2)) + \mathcal{N}(0, \Psi(\rho, \sigma_e^2))$$

Which is rather complex. The full variance-covariance discussion of Y is provided above.

1.1 Defining a conditional likelihood

To derive a sampler, let us first condition the likelihood on more information. We consider $\mathcal{L}(Y|\beta, \alpha, \sigma_e^2)$ as the data model. That is, consider the problem of estimating Y as if the α were known and state:

$$Y|\beta, \alpha, \sigma_e^2, \rho \sim \mathcal{N}(X\beta + \Delta\alpha, \Psi(\rho, \sigma_e^2)) \tag{4}$$

This is a feasible method to sample the problem because we can state the following conditional as well

$$\alpha|\lambda, \sigma_\alpha^2 \sim \mathcal{N}(0, \Psi(\lambda, \sigma_\alpha^2)) \tag{5}$$

So, if samples of α can be constructed, samples from the conditional likelihood $Y|.$ can be made that include this term. This strategy makes the problem structure amenable to Gibbs sampling. Thus, for a choice of conjugate priors:

$$\begin{aligned} \sigma_e^2 &\sim IG(a_e, b_e) \\ \sigma_\alpha^2 &\sim IG(a_\alpha, b_\alpha) \\ \beta &\sim \mathcal{N}(\mu_\beta, \Sigma_{\beta 0}) \end{aligned} \tag{6}$$

We can proceed in defining conditional posteriors for the various parameters in accordance with Gibbs sampling. Note that the priors for the spatial parameters are left unspecified at this point. But, let us stipulate only that they are:

- independent of each other
- functions of the parameter and their hyperparameters alone
- proper

To proceed in defining the conditional posteriors for the problem, it is first helpful to state how this is done. Briefly, consider the approach outlined in Gelman et al. [7]. For a generic model with parameters $\theta_j \in \theta$, note that the likelihood of Y conditions on a set of random variables. This constructs a kind of graph relating $L(Y|\cdots)$ and each of the parameter distributions $P(\theta_j|\cdots)$. Indeed, there may be parameters that θ_j depends on that depend on *yet more* parameters. Thus, by forming a graph of conditional dependency, one can define the conditional posterior as the set of nodes in the dependency graph that reach down to a leaf node of the dependency tree. Formally, this recalls the familiar statement of Bayes's Theorem:

$$P(\theta_j|\neg\theta_j) \propto L(y|\theta_j) \times P(\theta_{\neg j}|\theta_j) \times P(\theta_j)$$

Thus, the conditional posterior is proportional to the product of the likelihood of Y and all conditional distributions $\theta_{k \neq j}$ until the required leaf of a conditional dependency graph is reached. Fortunately, most of the distributions encountered in this process are constant with respect to the parameter of interest. This means the distributions do not mix with the rest of the probability graph, and can be discarded.

For the specific problem of variance components models with potential spatial structure, this will mean that some statements that involve the likelihood $L(Y|\alpha, \beta, \cdots)$ will *also* require $P(\alpha|\Psi(\lambda), \sigma^2_\alpha)$. This is the "cost" of conditioning on the random effect α, that some statements about the likelihood of Y will include the conditional dependence on α, which is random potentially-correlated effect with zero expectation. From the Bayesian perspective, where parameters are viewed as realizations from random distributions anyway, this is not a significant concern.

In addition, the problem structure can be further exploited. First, let us refer to generic spatial correlation parameters and scale parameters, $\phi, \sigma^2_\cdot$, modeling scale and correlation at any level of a variance components model. When the matrix-valued function $\Psi(\phi)$ reflects an SMA, SMA-like, or SAR-error covariance the prior for the scale and the spatial correlation parameters is necessarily separable:

$$P(\alpha|\Psi(\lambda, \sigma^2_\alpha) \propto N(0, \Psi(\lambda, \sigma^2_\alpha)) \times P(\lambda) \times P(\sigma^2_\alpha)$$

So, for explicitness, let us note the forms of $\Psi(\phi)$, for any level of dependence:

$$\begin{aligned} \Psi(\phi)_{SMA} &= (I + \phi W)(I + \phi W)' \\ \Psi(\phi)_{SMA'} &= WW' \\ \Psi(\phi)_{SAR-E} &= ((I - \phi W)'(I - \phi))^{-1} \end{aligned} \tag{7}$$

Thus, since $\Psi(\phi, \sigma^2_\cdot) = \Psi(\phi)\sigma^2_e$, we can separate out the $\sigma^2_\cdot$ terms from the covariance determinant and inversion in the likelihood kernel, *only since* we have conditioned the likelihood on α. This allows us to define conjugate relationships on priors for the univariate scale parameters $\sigma_\cdot$ and spatial parameters ϕ, separately.

Other choices of spatial covariances may not exhibit this separability. But, for the purposes of SMA, SMA-like, and SAR-error forms, these are indeed fully separable priors, meaning that we can define univariate priors for spatial correlation and scale parameters at any level, avoiding the need of specifying a matrix-valued covariance prior and all the issues that are incumbent on such a choice [8, 1, 9, 13].[1]

1.2 Conditional Posterior for β

Defining this conditional posterior is simplest. To do so, note the following:

$$P(\beta|\alpha, \sigma^2_\alpha, \sigma^2_e, Y) \propto L(Y|.) \times P(\beta) \times P(\alpha|\lambda, \sigma^2_\alpha), \times P(\lambda) \times P(\sigma^2_\alpha) \times P(\sigma^2_e) \times P(\rho)$$

[1] While this solves the problem at hand of specifying a variance components model for a certain class of exogenous W and $\Psi(\phi)$, a Bayesian approach to endogenous W would probably focus on this exact concern, as is done in the hierarchical coregionalization problems Gelfand et al. [6].

Fortunately, however, all of the terms behind $P(\beta)$ are not functions of β and thus do not mix with the conditional posterior. They can be dropped since they are constants with respect to β, and the likelihood updated like a typical Bayesian regression update. Thus, letting $\Sigma_y = \Psi(\rho)\sigma_e^2$ for convenience, let us proceed:

$$\begin{aligned}
&\propto exp\left[-\frac{1}{2}(Y - X\beta - \Delta\alpha)'(\Sigma_y^{-1})(Y - X\beta - \Delta\alpha)\right] \times exp\left[-\frac{1}{2}(\beta - \mu_\beta)'(\Sigma_{\beta 0})^{-1}(\beta - \mu_\beta)\right] \\
&\propto exp\left[-\frac{1}{2}\left(-\beta' X' \Sigma_y^{-1}(Y - \Delta\alpha) - (Y - \Delta\alpha)'\Sigma_y^{-1} X\beta + \beta' X' \Sigma_y^{-1} X\beta - \beta'\Sigma_{\beta 0}^{-1}\mu_\beta - \mu_\beta'\Sigma_{\beta 0}^{-1}\beta + \beta'\Sigma_{\beta 0}^{-1}\beta\right)\right] \\
&\propto exp\left[-\frac{1}{2}\left(-\beta'(X'\Sigma_y^{-1}(Y - \Delta\alpha) + \Sigma_{\beta 0}^{-1}\mu_\beta) - ((Y - \Delta\alpha)'\Sigma_y^{-1} X + \mu_\beta'\Sigma_y^{-1})\beta + \beta'(X'\Sigma_Y^{-1} X + \Sigma_{\beta 0}^{-1})\beta\right)\right]
\end{aligned}$$

following Lindley and Smith [12], let us complete the square using $b'\Sigma_\beta b$, where:

$$\Sigma_\beta = \left(X'\Sigma_y^{-1} X + \Sigma_\beta^{-1}\right)^{-1} \qquad\qquad b = \Sigma_\beta\left[X'\Sigma_y^{-1}(Y - \Delta\alpha) + \Sigma_{\beta 0}^{-1}\mu_\beta\right]$$

$$\propto \exp\left[-\frac{1}{2}\left(\beta' b - b'\beta + \beta'\Sigma_\beta^{-1}\beta + b'\Sigma_\beta b - b'\Sigma_\beta b\right)\right]$$

Which, since $b'Bb$ is constant with respect to β, means that we can drop the extra trailing $-b'\Sigma_\beta b$ and factor into a p-variate normal distribution of the form:

$$\begin{aligned}
&\propto \exp\left[-\frac{1}{2}(\beta - \Sigma_\beta b)'\Sigma_\beta^{-1}(\beta - \Sigma_\beta b)\right] \\
P(\beta|\cdots) &\propto \mathcal{N}_p(\Sigma_\beta b, \Sigma_\beta)
\end{aligned} \tag{8}$$

Note that this posterior only involves the spatial parameters in one level, since $\Sigma_Y = Phi(\rho)\sigma_e^2$. Therefore, any of the covariance forms mentioned in Eq. 7 can be used in this distribution.

1.3 Conditional Posterior for Upper-Level Random Effect, α

This analysis proceeds like that for $P(\beta|\cdots)$, but is somewhat simpler. First, note:

$$P(\alpha|Y, \beta, \lambda, \rho, \sigma_\alpha^2, \sigma_e^2) \propto L(Y|\cdots) \times P(\alpha|\Psi(\lambda), \sigma_\alpha^2) \times P(\lambda) \times P(\sigma_\alpha^2)$$

Since the last two priors are constant with respect to the parameter, they do not mix in the conditional posterior. This leaves us reducing the following:

$$\propto exp\left[-\frac{1}{2}\left((y - X\beta - \Delta\alpha)'\Sigma_Y^{-1}(Y - X\beta - \Delta\alpha)\right)\right] \times exp\left[-\frac{1}{2}\left(\alpha'\left(\Psi(\lambda)\sigma_\alpha^2\right)^{-1}\alpha\right)\right]$$

Since the mean α is zero, this can be reduced more simply than $P(\beta|\cdots)$.[2] This leaves us with the following conditional posterior, $P(\alpha|\cdots)$, using similar conditional posterior terms, $\Sigma_{\alpha'}$, a:

$$a = \Delta\Sigma_Y^{-1}(y - X\beta) \qquad\qquad \Sigma_{\alpha'} = \left(\Delta'\Sigma_Y^{-1}\Delta + \frac{\Phi(\lambda)^{-1}}{\sigma_\alpha^2}\right)^{-1}$$

$$P(\alpha|\cdots) \propto \mathcal{N}_J(\Sigma_{\alpha'}a, \Sigma_{\alpha'}) \tag{9}$$

1.4 Conditional Posterior for σ_e^2

When deriving the conditional posteriors for the scale parameters, we must exploit the separability of $\Psi(\phi, \sigma_\cdot^2)$ noted above for the SAR-error and SMA-style models. Thus, let us proceed to define the conditional posterior, $P(\sigma_e^2)$, the lower-level scale parameter by noting its conditional dependency graph:

$$P(\sigma_e^2|\cdots) \propto L(y|\cdots) \times P(\sigma_e^2)$$

The fact that $\Psi(\rho)$, the lower-level spatial correlation matrix, is not involved is due to the separability of scale and correlation in this model, but this will be made more clear as the conjugate prior is derived. Proceeding:

$$\propto |\Psi(\rho)\sigma_e^2|^{-\frac{1}{2}} exp\left[-\frac{1}{2}\left((y - X\beta - \Delta\alpha)'(\Psi(\rho)\sigma_e^2)^{-1}(y - X\beta - \Delta\alpha)\right)\right] \times \left(\sigma_e^2\right)^{-a_{e0}-1} exp\left[\frac{-b_{e0}}{\sigma_e^2}\right]$$

denoting the residual vector $\eta = y - X\beta - \Delta\alpha$, note that, if we *were not* able to separate $\Psi(\rho, \sigma_e^2)$ multiplicatively, the determinant and inverse expressions in the likelihood would require one to specify a fully-multivariate covariance prior. However, this can be avoided, since:

$$|\Psi(\rho)\sigma_e^2|^{-\frac{1}{2}} = |\Psi(\rho)|^{-\frac{1}{2}}\left(\sigma_e^2\right)^{-\frac{N}{2}} \qquad\qquad (\Psi(\rho)\sigma_e^2)^{-1} = (Psi(\rho)^{-1})\sigma_e^{-2}$$

Therefore, we can separate the scale from the correlation $\Psi(\rho)$ and proceed conjugating *just* the scale parameter:

$$\propto |\Psi(\rho)|^{-\frac{1}{2}}\left(\sigma_e^{-2}\right)^{-\frac{N}{2}} exp\left[-\frac{1}{2\sigma_e^2}\left(\eta'\Psi(\rho)^{-1}\eta\right)\right] \times \left(\sigma_e^2\right)^{-a_{e0}-1} exp\left[\frac{-b_{e0}}{\sigma_e^2}\right]$$
$$\propto |\Psi(\rho)|^{-\frac{1}{2}}\left(\sigma_e^{-2}\right)^{-\left(\frac{N}{2}+a_{e0}\right)-1} exp\left[-\frac{1}{\sigma_e^2}\left(\frac{\eta'\Psi(\rho)^{-1}\eta}{2} + b_{e0}\right)\right]$$

[2] If it were of practical interest, $Z\gamma$ could be included in the model for α rather than pulled out as we did earlier. But, this provides no real gains, since its inclusion in $X\beta$ simplifies all of the mathematics. That is, of course, until one begins stating endogenous Lag forms on a joint hierarchical-multilevel model, such as done in Lacombe and McIntyre [10], which require separate treatment from the SMA/SAR-error forms due to potential conjugacy issues.

Thus, we note that $|\Psi(\rho)|^{-\frac{1}{2}}$ is constant with respect to σ_e^2 and state this as proportional to the following Inverse Gamma distribution:

$$P(\sigma_e^2|\cdots) \propto IG\left(\frac{N}{2} + a_{e0}, \frac{\eta'\Psi(\rho)^{-1}\eta}{2} + b_{e0}\right)$$

Note that $\Psi(\rho)^{-1}$ *does* show up in the scaling parameter of the conditional posterior, even though it's conditionally constant.

1.4.1 Conditional Posterior for Upper-Level Scale Parameter σ_α^2

This argument proceeds similarly to that above, but pertains to the mixture of the $P(\alpha|\lambda, \sigma_\alpha^2) \times P(\sigma_\alpha^2)$, rather than to $L(Y|\cdots) \times P(\sigma_e^2)$:

$$P(\sigma_\alpha^2|\cdots) \propto L(Y|\cdots) \times P(\alpha|\sigma_\alpha^2, \lambda) \times P(\sigma_\alpha^2) \times P(\lambda)$$

Noting that the likelihood *conditional on* σ_u^2 is thus constant with respect to it, the only two distributions that mix in this case are:

$$P(\sigma_\alpha^2|\cdots) \propto P(\alpha|\sigma_\alpha^2, \lambda) \times P(\sigma_\alpha^2)$$

Which leads to the same Normal, Inverse Gamma behavior shown above:

$$\begin{aligned}
&\propto |\Psi(\lambda)|^{-\frac{1}{2}} \left(\sigma_\alpha^2\right)^{-\frac{J}{2}} exp\left[-\frac{\alpha'\Psi(\lambda)^{-1}\alpha}{2\sigma_\alpha^2}\right] \times \left(\sigma_\alpha^2\right)^{-a_{\alpha0}-1} exp\left[\frac{-b_{\alpha0}}{\sigma_\alpha^2}\right] \\
&\propto \left(\sigma_\alpha^2\right)^{-\left(\frac{J}{2}+a_{\alpha0}\right)-1} exp\left[-\frac{1}{\sigma_\alpha^2}\left(\frac{\alpha'\Psi(\lambda)^{-1}\alpha}{2} + b_{\alpha0}\right)\right] \\
&\propto IG\left(\frac{J}{2} + a_{\alpha0}, \frac{\alpha'\Psi(\lambda)^{-1}\alpha}{2} + b_{\alpha0}\right)
\end{aligned}$$

1.4.2 Conditional posterior for Lower-Level Correlation Parameter ρ

For the spatial parameters, a similar process applies, but it is necessary to note a few stipulations. Above, recall that we have specified the prior for ρ over a restricted support, namely the eigenrange of the relevant W matrix over which the correlation occurs. This means that the *conditional posterior* observes these restrictions as well. First, however, let us denote the bounds of ρ as $(\rho_-, \rho_+) = (1/e_{W_N min}, 1/e_{W_N max})$

$$P(\rho|\cdots) \propto L(y|\cdots) \times P(\rho)\Big|_{\rho_-}^{\rho_+}$$

This is relatively straightforward, since the likelihood only depends on ρ through its covariance and none of the other priors involve $P(\rho)$, due to the separability mentioned above. However, it is unlikely that a conjugate prior will exist for this choice, and the following distribution is likely not analytically tractable for any non-trivial choice of $P(\rho)$:

$$\propto |\Psi(\rho)|^{-\frac{1}{2}} exp\left[-\frac{1}{2\sigma_e^2}\left(\eta'\Psi(\rho)^{-1}\eta\right)\right] \times P(\rho)\Big|_{\rho_-}^{\rho_+}$$

While this might share the same form as a multivariate normal distribution in terms of y, the function is decidedly *not* in terms of y. But, this provides a simple parameterization to sample from using Metropolis-Hastings or Slice sampling methods. While the log determinant may be expensive to sample in large problems, the analytical complexity of the log probability distribution is rather simple:

$$log(P(\rho|\cdots)) \propto -\frac{1}{2}log(|\Psi(\rho)|) + \frac{\eta'\Psi(\rho)^{-1}\eta}{2 * \sigma_e^2} + P(\rho)\Big|_{\rho_-}^{\rho_+} \tag{10}$$

For the choice of a uniform prior, one can simply drop the last constant term, since it does not change with respect to ρ, either. However, if one chooses the objective or reference priors [4] or the informative priors chosen in LeSage and Pace [11], the prior term must be kept.

Since $P(\rho)$ did not mix in any other previous conditional posterior, this means that one could specify any informative prior on ρ, such as a scaled-shifted $\mathcal{B}(a_\rho, b_\rho)$ or a truncated $\mathcal{N}(\mu_\rho, \sigma_\rho^2)\Big|_{\rho_-}^{\rho_+}$ and have a similar (yet likely still analytically intractable) conditional posterior, as long as it is examined for propriety. In most reviewed cases of Bayesian spatial econometric models, authors in this literature do not thoroughly examine the choice of uniform priors for posterior propriety [5, 10], while the objective and reference prior discussions in broader statistical contexts are quite concerned with posterior propriety, since the information Jeffrey's Prior can lead to improper posteriors [4, 2, 14].

1.4.3 Conditional Posterior distribution for Upper-Level Correlation Parameter λ

This proceeds identically to the discussion of ρ, except that the conjugation occurs over $P(\alpha|\lambda, \sigma_u^2)$, since the likelihood conditioned on the group-wise random effects does not mix with the upper-level correlation parameter.

$$\begin{aligned} P(\lambda|\cdots) &\propto L(Y|\cdots) \times P(\alpha|\lambda, \sigma_\alpha^2) \times P(\lambda) \\ &\propto P(\alpha|\lambda\sigma_\alpha^2) \times P(\lambda) \\ &\propto |\Psi(\lambda)|^{-\frac{1}{2}} exp\left[-\frac{1}{2\sigma_\alpha^2}\left(\alpha'\Psi(\lambda)^{-1}\alpha\right)\right] \times P(\lambda) \end{aligned}$$

This yields a posterior of the same form as above, trained on the upper-level random effects rather than the lower-level linear predictor η:

$$log(P(\lambda|\cdots)) \propto -\frac{1}{2}log(|\Psi(\lambda)|) + \frac{\alpha'\Psi(\lambda)^{-1}\alpha}{2 * \sigma_e^2} + P(\lambda)\Big|_{\rho_-}^{\rho_+} \tag{11}$$

And, this can be sampled from using usual sampling methods for unknown parameters.

References

[1] Ignacio Alvarez. "Bayesian inference for a covariance matrix." In: *arXiv preprint arXiv:1408.4050* (2014).

[2] James O Berger, Victor De Oliveira, and Bruno Sansó. "Objective Bayesian analysis of spatially correlated data." In: *Journal of the American Statistical Association* 96.456 (2001), pp. 1361–1374.

[3] William J Browne, David Draper, et al. "A comparison of Bayesian and likelihood-based methods for fitting multilevel models." In: *Bayesian analysis* 1.3 (2006), pp. 473–514.

[4] Victor De Oliveira and Joon Jin Song. "Bayesian analysis of simultaneous autoregressive models." In: *The Indian Journal of Statistics, Series B (2008-)* (2008), pp. 323–350.

[5] Guanpeng Dong and Richard Harris. "Spatial autoregressive models for geographically hierarchical data structures." In: *Geographical Analysis* 47.2 (2015), pp. 173–191.

[6] Alan E Gelfand et al. "Nonstationary multivariate process modeling through spatially varying coregionalization." In: *Test* 13.2 (2004), pp. 263–312.

[7] Andrew Gelman et al. *Bayesian data analysis*. Vol. 2. Taylor & Francis, 2014.

[8] Alan Huang, Matthew P Wand, et al. "Simple marginally noninformative prior distributions for covariance matrices." In: *Bayesian Analysis* 8.2 (2013), pp. 439–452.

[9] Robert E Kass, Ranjini Natarajan, et al. "A default conjugate prior for variance components in generalized linear mixed models (comment on article by Browne and Draper)." In: *Bayesian Analysis* 1.3 (2006), pp. 535–542.

[10] Donald J Lacombe and Stuart G McIntyre. "Local and global spatial effects in hierarchical models." In: *Applied Economics Letters* (2016), pp. 1–5.

[11] J. LeSage and R. K. Pace. *Introduction to Spatial Econometrics*. CRC Press, 2009.

[12] Dennis V Lindley and Adrian FM Smith. "Bayes estimates for the linear model." In: *Journal of the Royal Statistical Society. Series B (Methodological)* (1972), pp. 1–41.

[13] Ranjini Natarajan and Robert E Kass. “Reference Bayesian methods for generalized linear mixed models.” In: *Journal of the American Statistical Association* 95.449 (2000), pp. 227–237.

[14] Cuirong Ren. “Objective Bayesian analysis of SAR models.” In: *Spatial Statistics* 4 (2013), pp. 68–78.