An Introduction to Bayesian Inference in Spatial Econometrics

Donald J. Lacombe*

August 14, 2008

Abstract

This tutorial is designed to introduce readers to Bayesian variants of the standard SAR and SEM models that are the most widely used and applied models in spatial econometrics. Particular attention is paid to the mathematical derivations required to obtain the full conditional distributions required for Gibbs sampling. The models are derived using diffuse as well as natural conjugate priors for the parameters.

1 Introduction

Spatial econometric models have made substantial headway in recent years in various fields of study, such as economics, political science, and sociology. The standard models utilized, i.e. the SAR and SEM models, are usually estimated using maximum likelihood techniques. In recent years, several papers have demonstrated the estimation of these models via generalized method-of-moments (GMM) techniques. However, Bayesian variants of these models have been gaining ground, due to the fact that in certain situations, e.g. the spatial probit, standard techniques are either too cumbersome or impossible to utilize. The Bayesian paradigm is conceptually fairly easy to understand, however, the mathematical details surrounding estimation issues are difficult to comprehend unless one has some help. In fact, once

*Associate Professor of Economics, Ohio University, Department of Economics, 345 Bentley Annex, Athens, OH, 45701 USA email: lacombe@ohio.edu.

a few basic mathematical ideas are known and applied, the derivation of such models becomes much easier to understand and apply. In particular, this tutorial will focus on the mathematical details required to implement various Markov Chain Monte Carlo techniques, such as the popular Gibbs sampler, and give the reader complete and full mathematical details so that they may understand the underlying mathematical derivations.

2 The SAR and SEM models

The SAR and SEM models have their genesis in the general spatial model shown in (1).

$$\begin{aligned} y &= \rho W_1 y + X\beta + u \\ u &= \lambda W_2 u + \varepsilon \\ \varepsilon &\sim N\left(0, \sigma^2 I_n\right) \end{aligned} \tag{1}$$

where y contains an n x 1 vector of dependent variables and X represents an n x k matrix of explanatory variables. W_1 and W_2 are exogenous n x n spatial weight matrices, usually containing first-order contiguity relations or functions of distance. The spatial weight representation does not make any difference for the derivation of the Bayesian SAR and SEM models.

The SAR model can be obtained from (1) above by setting $W_2 = 0$. This models includes a spatially lagged dependent variable and the usual error term structure. The SEM model can be obtained by setting $W_1 = 0$ in equation (1) above. Thus we can obtain either the SAR or SEM models by simply altering the more general spatial model presented in (1). The SAR and SEM models are usually estimated via maximum likelihood methods or other methods, such as generalized method of moments (GMM). However, this introduction wishes to explain how the Bayesian variants of these two models are obtained.

3 A Brief Introduction to the Bayesian Paradigm

The Bayesian paradigm consists of three entities: the posterior distribution, the prior distribution, and the likelihood function. The likelihood function is probably the most familiar to those working in the frequenist domain and indeed it is the same entity in the Bayesian paradigm. In terms of the two different spatial econometric models, the likelihood function is as follows

$$L(\beta, \sigma, \rho, y, X) = \frac{1}{(2\pi)^{n/2}\sigma^n} |I_n - \rho W| \exp\left\{-\frac{1}{2\sigma^2}(\varepsilon'\varepsilon)\right\} \tag{2}$$

The $\varepsilon'\varepsilon$ term is equal to $(I_n - \rho W)y - X\beta$ in the SAR model and is equal to $(I_n - \rho W)(y - X\beta)$ in the SEM model[1].

The next entity to be discussed in the Bayesian paradigm is the role of the prior distribution. The prior distribution is designed to capture the prior beliefs of the researcher and to formalize those beliefs in a probability distribution. Although the designation of the prior distribution is of utmost importance, in this introduction, we are interested only in how the specification of certain priors changes the mathematical derivations of the full conditional distributions. As a starting point, so-called "non-informative","diffuse", or "ignorant" priors for the parameters in the model, namely β, σ, and ρ, will be specified. Mathematically, this can be accomplished as follows

$$\begin{aligned} p(\beta) &\propto constant \\ p(\sigma) &\propto \frac{1}{\sigma} \\ p(\rho) &\propto constant \end{aligned}$$

In other words, the β and ρ terms are constants (usually set to one), and the prior on the σ term is a standard diffuse prior for this parameter that is used frequently in the literature. In fact, the use of this diffuse parameter for σ allows one to easily recognize the form of the conditional distribution required for Gibbs sampling, which will be illustrated later.

The final entity, and perhaps the most important from a Bayesian perspective, is the specification of the posterior distribution. The posterior distribution summarizes all of the information about the parameters of the model and is the focus of all Bayesian inference. The posterior distribution is derived from Bayes Rule, which can be summarized as follows

$$p(\theta | y) = \frac{p(y | \theta) p(\theta)}{p(y)} \tag{3}$$

[1] In general, the spatial dependence parameter in the SAR model is designated as ρ and in the SEM model it is designated λ. I use the same symbol for both models and the interpretation will depend on the context.

Frequently, the relationship between the posterior, prior, and likelihood is summarized in the phrase "the posterior is proportional to the prior times the likelihood" and thus can be written

$$p(\theta \,|y) \propto p(y \,|\theta)\, p(\theta)$$

This relationship holds because the denominator in equation (3) does not involve the parameters, so the relationship is a proportional one[2].

Now that the basic elements of the Bayesian paradigm are in order, it is time to start assembling the constituent parts in order to derive the full conditional distributions required for Gibbs sampling.

4 Deriving Full Conditional Distributions

The first order of business in deriving the full conditional distributions required for Gibbs sampling is to mathematically derive the posterior distribution. This is accomplished by multiplying the likelihood and the priors together to form the posterior distribution. The priors in this example are assumed to be independent, i.e. $p(\beta, \sigma, \rho) = p(\beta)\, p(\sigma)\, p(\rho)$, so that they may be multiplied by the likelihood without any problem. The likelihood can be conveniently expressed mathematically as follows

$$L(\beta, \sigma, \rho, y, X) = (2\pi)^{-n/2}\, \sigma^{-n}\, |I_n - \rho W| \exp\left\{-\frac{1}{2\sigma^2}(\varepsilon'\varepsilon)\right\}$$

which is simply taking the leading terms in the likelihood function in equation (2) above and inverting. When we take this likelihood and multiply it by all of the priors, we obtain the following posterior distribution

$$p(\rho, \beta, \sigma \,|y, W) \propto |I_n - \rho W|\, \sigma^{-(n+1)} \exp\left\{-\frac{1}{2\sigma^2}(\varepsilon'\varepsilon)\right\}$$

The posteriors for each of the two competing models (i.e. SAR and SEM) will depend on the particular ε. The posterior for the SAR model is

$$p(\rho, \beta, \sigma \,|y, W) \propto |I_n - \rho W|\, \sigma^{-(n+1)} \exp\left\{-\frac{1}{2\sigma^2}(Ay - X\beta)'(Ay - X\beta)\right\}$$

[2] The denominator in equation (3) is referred to as the marginal likelihood. The marginal likelihood will not play a role in the derivation of the full conditional distributions required for Gibbs sampling, however it does play a role in model choice.

and the posterior for the SEM model is

$$p(\rho, \beta, \sigma \,|y, W) \propto |I_n - \rho W| \, \sigma^{-(n+1)} \exp\left\{ -\frac{1}{2\sigma^2} (Ay - AX\beta)' (Ay - AX\beta) \right\}$$

where $A \equiv (I_n - \rho W)$ in both models.

One may wonder what happened to the $(2\pi)^{-n/2}$ term. In Bayesian analysis, since we are working with a proportional relationship, we can ignore any constants that do not involve the variable of interest, which is why the above term "disappears". When this term is ignored, the likelihood looks as follows

$$L(\beta, \sigma, \rho, y, X) = |I_n - \rho W| \, \sigma^{-n} \exp\left\{ -\frac{1}{2\sigma^2} (\varepsilon' \varepsilon) \right\}$$

Now, to obtain the posterior, we must multiply this likelihood by the priors. The priors on β and ρ are diffuse and equal to one, so there is no change to the likelihood. When we multiply the likelihood by the prior for σ, we obtain the following posterior distribution

$$\begin{aligned} L(\beta, \sigma, \rho, y, X) &= |I_n - \rho W| \, \sigma^{-n} \exp\left\{ -\frac{1}{2\sigma^2} (\varepsilon' \varepsilon) \right\} \times \sigma^{-1} \\ p(\rho, \beta, \sigma \,|y, W) &\propto |I_n - \rho W| \, \sigma^{-(n+1)} \exp\left\{ -\frac{1}{2\sigma^2} (\varepsilon' \varepsilon) \right\} \end{aligned}$$

where again the ε term will vary depending on the model. The $\sigma^{-(n+1)}$ term is obtained as follows: $\sigma^{-n} \times \sigma^{-1} = \sigma^{-(n+1)}$. We will see later on that specifying the prior in this manner will enable us to recognize the form of the conditional distribution for σ.

The posterior is now fully defined and explained so we are ready to derive the full conditional distributions required for Gibbs sampling. But first, a brief introduction to the idea behind Gibbs sampling.

4.1 A Brief Introduction to Gibbs Sampling

The common objective in all Bayesian exercises is to derive the marginal distributions that summarize knowledge about the parameters, θ, conditional on the data, y. The posterior distribution, however, is "difficult to work with" meaning that some of the required posterior marginal distributions, namely $p(\sigma \,|y)$, $p(\beta \,|y)$, and $p(\rho \,|y)$, are not available in closed form. In other words, one cannot integrate out the parameter of interest and derive the marginal distribution using pencil and paper. In general settings, this integration is difficult or impossible, which is a major impediment to analysis. The solution is to Gibbs sample.

The necessary conditions for Gibbs sampling the SAR or SEM model, or any model, for that matter, are two. First, the fully conditional distributions comprising the joint posterior must be available in closed form. Second, these forms must be tractable in the sense that it is easy to draw samples from them.

A simplified example may help in explaining the idea behind the Gibbs sampler. Suppose one has a joint density of the form $f(x, y, z)$ and that the marginal distributions of interest, $f(x)$, $f(y)$, and $f(z)$ are not available in closed form. Normally, this would lead to an intractable situation. However, we can use Gibbs sampling to achieve the desired goal. Essentially, Gibbs sampling requires that we obtain random draws from each of the component full conditional distributions (i.e. $f(x \mid y, z)$, $f(y \mid x, z)$, and $f(z \mid x, y)$) derived from the joint posterior distribution. Under weak regularity conditions, this chain converges in distribution to the true marginal quantities (the marginal posterior pdfs) that we seek. Therefore, Gibbs sampling is a very convenient alternative that allows inference to take place.

The next order of business is deriving the full conditional distributions for the SAR and SEM models. When deriving the full conditional distributions, there are three mathematical ideas that one should keep in mind. First, all non-essential constants can be ignored because they are subsumed into the constant of proportionality. Second, only the terms involved in the conditional distribution need be collected when deriving the conditional distribution. Lastly, "completing the square" will play a role in deriving conditional distributions as well.

The first place to start is to examine the posterior distribution for all of the parameters, which is as follows

$$p(\rho, \beta, \sigma \mid y, W) \propto |I_n - \rho W| \, \sigma^{-(n+1)} \exp\left\{-\frac{1}{2\sigma^2}(\varepsilon'\varepsilon)\right\}$$

where ε depends on whether it is the SAR or SEM model. The conditional distributions that we seek are the following

$$\begin{gathered} p(\sigma \mid \beta, \rho) \\ p(\beta \mid \sigma, \rho) \end{gathered}$$

$$p(\rho\,|\beta,\sigma)$$

We will start by examining the full conditional distribution for σ. In order to define the full conditional distribution for σ, or any other parameter, we simply pick out the terms in the posterior distribution that contain σ. This leads to the following conditional distribution for σ

$$p(\sigma\,|\beta,\rho) \propto \sigma^{-(n+1)} \exp\left\{-\frac{1}{2\sigma^2}(\varepsilon'\varepsilon)\right\}$$

where ε depends on whether it is the SAR or SEM model. The next step is to identify the distributional form of the conditional distribution for σ so that we may take random draws from this distribution.

A standard reference for distributional forms is Zellner (1996, p. 371). The conditional distribution for σ is in the form of an inverse Gamma distribution, which is equation A.37b and has the following form

$$f^{IG}(\sigma\,|\upsilon,s) = \frac{2}{\Gamma(\upsilon/2)}\left(\frac{\upsilon s^2}{2}\right)^{\upsilon/2}\frac{1}{\sigma^{\upsilon+1}}\exp^{-\upsilon s^2/2\sigma^2}$$

We can now use the first mathematical idea in eliminating the constant in the inverse Gamma distribution. The inverse Gamma distribution is defined in terms of σ so we can eliminate any terms that do not involve σ, which leads to the following functional form

$$f^{IG}(\sigma\,|\upsilon,s) \propto \frac{1}{\sigma^{\upsilon+1}}\exp^{-\upsilon s^2/2\sigma^2}$$

One further simplification to the standard formula for the inverse Gamma distribution will bring this formula into a form that is similar to the form of the conditional distribution for σ. Simply invert the first term (i.e. the $\frac{1}{\sigma^{\upsilon+1}}$ term) and the functional form for the inverse Gamma now looks like the following

$$f^{IG}(\sigma\,|\upsilon,s) \propto \sigma^{-(\upsilon+1)} \exp\left(\frac{-\upsilon s^2}{2\sigma^2}\right)$$

We can now let $\upsilon = n$ and $\upsilon s^2 = (Ay - X\beta)'(Ay - X\beta)$ in the SAR model and $\upsilon s^2 = (Ay - AX\beta)'(Ay - AX\beta)$ in the SEM model. The functional form of the posterior conditional distribution for σ is identical to the inverted Gamma distribution and the first step in the Gibbs sampling algorithm is to take a random draw from the inverted Gamma distribution with the appropriate parameters as defined by either the SAR or SEM model.

The next conditional distribution that needs explanation is the conditional distribution for β, which is $p(\beta\,|\sigma,\rho)$. Only the β terms in the posterior distribution need to be used, and this results in the following conditional distribution for β

$$p(\beta\,|\sigma,\rho) \propto \exp\left\{-\frac{1}{2\sigma^2}(\varepsilon'\varepsilon)\right\}$$

where again, the ε term is defined individually for the SAR and SEM models. Looking first at the SAR model, we have the following conditional distribution when we replace ε

$$p(\beta\,|\sigma,\rho) \propto \exp\left\{-\frac{1}{2\sigma^2}(Ay - X\beta)'(Ay - X\beta)\right\}$$

We can now use the second mathematical idea, completing the square, to find the conditional distribution for β.

The process of completing the square usually occurs when one is working with the kernel of the multivariate normal density[3]. The multivariate normal density has the following form (Zellner 1996, p. 379)

$$f^{MVN}(x\,|\theta,\Sigma) = (2\pi)^{-m/2}\,|\Sigma|^{-1/2}\exp\left\{-\frac{1}{2}(x-\theta)'\,\Sigma^{-1}(x-\theta)\right\}$$

We can once again ignore the leading constants and this will result is the following normal kernel[4]

$$f^{MVN}(x\,|\theta,\Sigma) \propto \exp\left\{-\frac{1}{2}(x-\theta)'\,\Sigma^{-1}(x-\theta)\right\}$$

We begin completing the square by expanding out the following term form the normal kernel

$$(x-\theta)'\,\Sigma^{-1}(x-\theta)$$

[3]The kernel of any probability density is the part excluding any integrating constants.

[4]Since the multivariate normal density is defined in terms of the variable x, we can ignore all terms that do not involve x and hence ignore the leading constants.

This expansion is a straightforward application of matrix algebra rules and results in the following

$$x'\Sigma^{-1}x - x'\Sigma^{-1}\theta - \theta'\Sigma^{-1}x + \theta'\Sigma^{-1}\theta$$

If we are interested in completing the square in x, the following expression results

$$x'\Sigma^{-1}x - x'\Sigma^{-1}\theta - \theta'\Sigma^{-1}x + constant$$

where we can ignore the $\theta'\Sigma^{-1}\theta$ term because it does not involve x. Alternatively, if we are interested in completing the square in θ, we would have the following expression

$$\theta'\Sigma^{-1}\theta - \theta'\Sigma^{-1}x - x'\Sigma^{-1}\theta + constant$$

where now we can ignore the $x'\Sigma^{-1}x$ term as it does not involve θ. The important point is that anything that is multivariate normal must fall into one of these two forms.

Completing the square in β is a similar exercise. After expansion of the terms in the normal kernel, we obtain the following general form

$$\beta'\left(something_1\right)\beta - \beta'\left(something_2\right) - \left(something_2\right)'\beta$$

When we expand the normal kernel into these three terms, then β has a multivariate normal distribution with

$$\begin{aligned} covariance &= \left(something_1\right)^{-1} \\ mean &= \left(something_1\right)^{-1} \times \left(something_2\right) \end{aligned}$$

We can now examine how this is applied in the context of the two spatial econometric models. Starting with the SAR model, we have the following conditional distribution[5]

$$p\left(\beta\,|\sigma,\rho\right) \propto \exp\left\{-\frac{1}{2}\left(Ay - X\beta\right)' C_\beta^{-1}\left(Ay - X\beta\right)\right\}$$

[5]The σ^2 term is no longer part of the leading fraction because it is now part of the C_β^{-1} term. The expressions are mathematically equivalent but it is consistent to express the covariance in this fashion so that it mimics the completing the square exercise above.

The portion of the normal kernel that we are interested in is the $(Ay - X\beta)' \, C_\beta^{-1} \, (Ay - X\beta)$ term. We also need to define a covariance matrix,[6] which can be expressed as $C_\beta^{-1} \equiv (\sigma^2 I_n)^{-1}$. We will also define $A \equiv (I_n - \rho W)$ in accordance with the A term in the normal kernel.

The expression that we wish to expand to complete the square is the following

$$(Ay - X\beta)' \, C_\beta^{-1} \, (Ay - X\beta)$$

When we perform this expansion using the standard rules of matrix algebra, we obtain the following

$$y'A'C_\beta^{-1}Ay - y'A'C_\beta^{-1}X\beta - \beta'X'C_\beta^{-1}Ay + \beta'X'C_\beta^{-1}X\beta$$

$$\beta'X'C_\beta^{-1}X\beta - \beta'X'C_\beta^{-1}Ay - y'A'C_\beta^{-1}X\beta$$

From the second line in the above equation, we see that by completing the square in β we put the expanded term into a form that is recognizable as a multivariate normal density[7]. In this particular example, $something_1 = X'C_\beta^{-1}X$, and $something_2 = X'C_\beta^{-1}Ay$. We can now place the $something_1$ and $something_2$ terms in the above formula to define the covariance and mean for the conditional distribution for β

$$\begin{aligned} covariance &= \left(X'C_\beta^{-1}X\right)^{-1} \\ mean &= \left(X'C_\beta^{-1}X\right)^{-1} \left(X'C_\beta^{-1}Ay\right) \end{aligned}$$

Therefore, β is distributed multivariate normal with mean $\left(X'C_\beta^{-1}X\right)^{-1} \left(X'C_\beta^{-1}Ay\right)$ and covariance $\left(X'C_\beta^{-1}X\right)^{-1}$.

A similar exercise can be performed in regards to the SEM model. In the SEM model, the normal kernel has the following form

$$p\left(\beta \,|\sigma, \rho\right) \propto \exp\left\{-\frac{1}{2}\left(Ay - AX\beta\right)' C_\beta^{-1} \left(Ay - AX\beta\right)\right\}$$

[6] The multivariate normal distribution formula from above has covariance matrix equal to Σ^{-1} so we need to define a covariance matrix for the normal kernel in the SAR model. We use $C_\beta^{-1} \equiv \left(\sigma^2 I_n\right)^{-1}$ because we are assuming a homoskedastic error structure.

[7] The $y'A'C_\beta^{-1}Ay$ term can be ignored in the expansion because it does not involve β.

We can now expand the following expression and complete the square in β

$$(Ay - AX\beta)' C_\beta^{-1} (Ay - AX\beta)$$

where $A \equiv (I_n - \rho W)$ and $C_\beta^{-1} \equiv (\sigma^2 I_n)^{-1}$ as before. When we perform the expansion, we are left with the following expression in terms of β[8]

$$y'A'C_\beta^{-1}Ay - y'A'C_\beta^{-1}AX\beta - \beta'X'A'C_\beta^{-1}Ay + \beta'X'A'C_\beta^{-1}AX\beta$$

$$\beta'X'A'C_\beta^{-1}AX\beta - \beta'X'A'C_\beta^{-1}Ay - y'A'C_\beta^{-1}AX\beta$$

In the SEM model, $something_1 = X'A'C_\beta^{-1}AX$ and $something_2 = X'A'C_\beta^{-1}Ay$. When we place these two terms into the formula for the multivariate normal density, we obtain

$$\begin{aligned} covariance &= \left(X'A'C_\beta^{-1}AX\right)^{-1} \\ mean &= \left(X'A'C_\beta^{-1}AX\right)^{-1}\left(X'A'C_\beta^{-1}Ay\right) \end{aligned}$$

In the SEM model, β is distributed multivariate normal with $covariance = \left(X'A'C_\beta^{-1}AX\right)^{-1}$ and $mean = \left(X'A'C_\beta^{-1}AX\right)^{-1}\left(X'A'C_\beta^{-1}Ay\right)$.

The Gibbs sampling algorithm would now involve a random draw from a multivariate normal density with mean and covariance terms defined by the particular model we were using.

The final conditional distribution concerns the ρ parameter. The full conditional distribution for this parameter is

$$p(\rho\,|\beta, \sigma) \propto |I_n - \rho W| \exp\left\{-\frac{1}{2\sigma^2}(\varepsilon'\varepsilon)\right\}$$

where ε is defined separately for the SAR and SEM model. Note that the full conditional distribution for the ρ parameter does not fall into any recognizable distributional form.

It would seem at this point that the Gibbs sampling algorithm fails us in that there is no "easy to sample from" full conditional distribution for the ρ parameter. Technically, this is correct. However, we can utilize what is referred to as the Metropolis-Hastings algorithm to obtain random draws from

[8]Once again, the $y'A'C_\beta^{-1}Ay$ term can be ignored in the expansion because it does not involve β.

this seemingly intractable distribution. The Metropolis-Hastings algorithm is an accept-reject type algorithm in which a candidate value is proposed and and then one decides whether to set the next value of the chain equal to this proposed value or to remain at the current value. The Metropolis-Hastings algorithm mimics the Gibbs sampling algorithm but the difference is that the Metropolis-Hastings algorithm can be used for conditional distributions that do not have any recognizable distributional form. When the Metropolis-Hastings algorithm is used in combination with standard Gibbs sampling techniques, it is referred to as "Metropolis-within-Gibbs".

We now have all of the ingredients required to obtain draws for all of the parameters in either the SAR or SEM model. At the start of this tutorial, all of the prior distributions were diffuse. However, parameters may be defined in terms of well-known prior distributions. We now examine the case of a normal prior for the β term.

4.2 Full Conditional Distributions Under a Normal Prior for β

In this section, we will utilize a multivariate normal prior for the β term. Why might we use such a prior? A well-defined prior distribution for the regression parameters becomes useful in model comparisons between competing models or when the researcher has valid prior information for the regression parameters. Regardless of the motivation, the specification of a normal prior for β will change the derivation of the full conditional distribution for these parameters relative to the diffuse prior case.

The first fact to recognize is that if the priors for the σ and ρ parameters remain unchanged, their full conditional distributions will not change. The full conditional distributions for these parameters were derived before and are as follows

$$p\left(\sigma\,|\beta,\rho\right) \propto \sigma^{-(n+1)} \exp\left\{-\tfrac{1}{2\sigma^2}\left(\varepsilon'\varepsilon\right)\right\}$$

$$p\left(\rho\,|\beta,\sigma\right) \propto |I_n - \rho W| \exp\left\{-\tfrac{1}{2\sigma^2}\left(\varepsilon'\varepsilon\right)\right\}$$

where ε is defined as before for the SAR and SEM models.

Attention now turns for the full conditional distribution for β. Instead of using a diffuse prior, we now will use a prior in the form of a multivariate normal distribution

$$\beta \sim f^{MVN}\left(\beta \left| \hat{\beta}_0, C_{\hat{\beta}_0}\right.\right)$$

where $\hat{\beta}_0$ is the mean and $C_{\hat{\beta}_0}$ is the covariance matrix. The process of defining the new conditional distribution for β given the new prior for β involves completing the square two times. First, we need to define the new posterior distribution for all of the parameters given the diffuse priors for σ and ρ and the multivariate normal prior for β. The new posterior is as follows

$$\begin{aligned} p\left(\rho, \beta, \sigma \left| y, W\right.\right) &\propto \left|I_n - \rho W\right| \sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2}\left(\varepsilon'\varepsilon\right)\right\} \times \sigma^{-1} \\ &\times \exp\left\{-\frac{1}{2}\left(\beta - \hat{\beta}_0\right)' C_{\hat{\beta}_0}^{-1}\left(\beta - \hat{\beta}_0\right)\right\} \end{aligned}$$

which is the likelihood as before multiplied by the multivariate normal density prior for β and the diffuse priors for σ and ρ.

The new conditional distribution for β, as mentioned before, requires that we complete the square in the normal kernel and in the multivariate normal prior distribution. When we complete the square in the normal kernel for the SAR model, we obtain the following expression, repeated here for convenience

$$\left(Ay - X\beta\right)' C_\beta^{-1}\left(Ay - X\beta\right)$$

$$y'A'C_\beta^{-1}Ay - y'A'C_\beta^{-1}X\beta - \beta'X'C_\beta^{-1}Ay + \beta'X'C_\beta^{-1}X\beta$$

$$\beta'X'C_\beta^{-1}X\beta - \beta'X'C_\beta^{-1}Ay - y'A'C_\beta^{-1}X\beta$$

As before, we can place $something_1 = X'C_\beta^{-1}X$, and $something_2 = X'C_\beta^{-1}Ay$ into the formula for the multivariate normal distribution to obtain the mean and covariance as follows

$$\begin{aligned} covariance &= \left(X'C_\beta^{-1}X\right)^{-1} \\ mean &= \left(X'C_\beta^{-1}X\right)^{-1}\left(X'C_\beta^{-1}Ay\right) \end{aligned}$$

At this point we are now half done with the derivation. We now need to complete the square in the multivariate prior distribution by expanding the following term

$$\left(\beta - \hat{\beta}_0\right)' C_{\hat{\beta}_0}^{-1}\left(\beta - \hat{\beta}_0\right)$$

When this term is expanded using standard rules of matrix algebra as before, we are left with the following expression

$$\beta' C_{\hat{\beta}_0}^{-1} \beta - \beta' C_{\hat{\beta}_0}^{-1} \hat{\beta}_0 - \hat{\beta}_0' C_{\hat{\beta}_0}^{-1} \beta + \hat{\beta}_0' C_{\hat{\beta}_0}^{-1} \hat{\beta}_0$$

In this situation, we would like to complete the square in β which leads to the following form

$$\beta' C_{\hat{\beta}_0}^{-1} \beta - \beta' C_{\hat{\beta}_0}^{-1} \hat{\beta}_0 - \hat{\beta}_0' C_{\hat{\beta}_0}^{-1} \beta_0$$

We can now identify the $something_1$ and $something_2$ terms from the above expression in order to place it into the predetermined multivariate normal form. The two terms are as follows

$$something_1 = \left(C_{\hat{\beta}_0}^{-1} \right)$$

$$something_2 = \left(C_{\hat{\beta}_0}^{-1} \hat{\beta}_0 \right)$$

All of the elements are now in place so we can now identify the multivariate normal density covariance and mean terms. In general, when a normal prior is used for β the covariance and mean terms will have the following form

$$covariance = (something_{1a} + something_{1b})^{-1}$$

$$mean = (something_{1a} + something_{1b})^{-1} (something_{2a} + something_{2b})$$

where $something_{1a} = X' C_{\beta}^{-1} X$, $something_{1b} = C_{\hat{\beta}_0}^{-1}$, $something_{2a} = X' C_{\beta}^{-1} Ay$, and $something_{2b} = C_{\hat{\beta}_0}^{-1} \hat{\beta}_0$. The "*somethings*" from above can now be put into the standard form for a multivariate normal density given above. Note that the terms are additive in that we simply add together the terms when we complete the square two times. The covariance and mean for the new conditional distribution for β is as follows

$$covariance = \left(X' C_{\beta}^{-1} X + C_{\hat{\beta}_0}^{-1} \right)^{-1}$$

$$mean = \left(X' C_{\beta}^{-1} X + C_{\hat{\beta}_0}^{-1} \right)^{-1} \left(X' C_{\beta}^{-1} Ay + C_{\hat{\beta}_0}^{-1} \hat{\beta}_0 \right)$$

With the new multivariate normal distribution defined for when we have a normal prior for β, we can use this information in the Gibbs sampling algorithm to obtain random draws for β in the SAR model.

A similar exercise can be carried out to determine the full conditional distribution for β in the SEM model. All that is required is that the $something_{1a}$ and $something_{2a}$ terms be replaced with $something_{1a} = X'A'C_{\beta}^{-1}AX$ and $something_{2a} = X'A'C_{\beta}^{-1}Ay$ respectively.

4.3 An Inverse Gamma Prior for σ

The final modification that will be explored in this introduction is the specification of a proper prior for the σ parameter. The most commonly used proper prior for this parameter is the inverse gamma distribution

$$f^{IG}\left(\sigma\,|\upsilon, s\right) \propto \sigma^{-(\upsilon+1)} \exp\left(\frac{-\upsilon s^2}{2\sigma^2}\right)$$

where the leading integrating constant has been suppressed as before. Whenever a new prior for any parameter is proposed, we must alter the posterior before deriving the conditional distribution for that parameter[9]. In this case, we will have the following priors

$$\begin{aligned} p\left(\sigma\right) &\propto \sigma^{-(\upsilon_0+1)} \exp\left(\frac{-\upsilon_0 s_0^2}{2\sigma^2}\right) \\ p\left(\beta\right) &\sim f^{MVN}\left(\beta\left|\hat{\beta}_0, C_{\hat{\beta}_0}\right.\right) \\ p\left(\rho\right) &\propto constant \end{aligned}$$

which consists of an inverse gamma prior for σ, a multivariate normal prior for β, and a diffuse prior for the ρ parameter.

We now examine the modified posterior, which is the likelihood times the priors. The likelihood can be expressed as follows

$$L\left(\beta, \sigma, \rho, y, X\right) = \left|I_n - \rho W\right| \sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2}\left(\varepsilon'\varepsilon\right)\right\}$$

where ε depends on whether it is the SAR or SEM model. The next step is to multiply this likelihood by all of the prior distributions, which results in

[9] Since the posterior distribution is the prior multiplied by the likelihood, whenever a new prior is proposed, the posterior has to be modified to accommodate the new prior distribution(s).

the following

$$
\begin{aligned}
p(\rho, \beta, \sigma \,|y, W) \quad &\propto \quad |I_n - \rho W| \, \sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2}(\varepsilon'\varepsilon)\right\} \\
&\times \quad \exp\left\{-\frac{1}{2}\left(\beta - \hat{\beta}_0\right)' C_{\hat{\beta}_0}^{-1}\left(\beta - \hat{\beta}_0\right)\right\} \\
&\times \quad \sigma^{-(\upsilon_0+1)} \exp\left(\frac{-\upsilon_0 s_0^2}{2\sigma^2}\right)
\end{aligned}
$$

Since neither the prior for the β term nor the ρ term has changed, their conditional distributions do not change either and are derived above. The only conditional distribution that does change is the conditional distribution for σ. Mathematically, this distribution can be expressed as follows after collecting only the terms in the posterior that depend on σ

$$
p(\sigma \,|\beta, \rho) \propto \sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2}(\varepsilon'\varepsilon)\right\} \times \sigma^{-(\upsilon_0+1)} \exp\left(\frac{-\upsilon_0 s_0^2}{2\sigma^2}\right)
$$

where ε will depend on whether we are looking at the SAR or SEM model. We now can collect terms in these two expressions to obtain the full conditional distribution for σ. First, we will collect the σ terms in each section of the equation

$$
\sigma^{-n} + \sigma^{-(\upsilon_0+1)} = \sigma^{-(n+\upsilon_0+1)}
$$

which is simply each exponent associated with each σ term added together. Next, we can add together the normal kernel parts as follows

$$
\begin{aligned}
&\exp \quad \left\{-\frac{1}{2\sigma^2}(\varepsilon'\varepsilon) - \frac{1}{2\sigma^2}\left(\upsilon_0 s_0^2\right)\right\} \\
&\exp \quad \left\{-\frac{(\varepsilon'\varepsilon + \upsilon_0 s_0^2)}{2\sigma^2}\right\}
\end{aligned}
$$

This leaves us with the followin expression for the full conditional distribution for σ

$$
p(\rho, \beta, \sigma \,|y, W) \propto \sigma^{-(n+\upsilon_0+1)} \exp\left\{-\frac{(\varepsilon'\varepsilon + \upsilon_0 s_0^2)}{2\sigma^2}\right\}
$$

which is in the form of an inverse gamma distribution

$$
f^{IG}(\sigma \,|\upsilon, s) \propto \sigma^{-(\upsilon+1)} \exp\left(\frac{-\upsilon s^2}{2\sigma^2}\right)
$$

Now we can let $\upsilon = n + \upsilon_0$ and $\upsilon s^2 = (Ay - X\beta)'(Ay - X\beta) + \upsilon_0 s_0^2$ in the SAR model whereas $\upsilon s^2 = (Ay - AX\beta)'(Ay - AX\beta) + \upsilon_0 s_0^2$ in the SEM model. The Gibbs sampling algorithm would now consist of a random draw from the inverse gamma distribution given the above arguments which depend on the model being used.

5 Conclusion

Bayesian econometrics has seen an incredible increase in use in recent years. This is primarily due to the fact that the development of the Gibbs sampler has revolutionized Bayesian computation and spatial econometrics is no different in this regard. This tutorial introduced the Bayesian paradigm and mathematically derived all of the necessary conditions required for Gibbs sampling in two of the most popular spatial econometric models, the SAR and the SEM.

Readers that are interested in furthering their knowledge of Bayesian spatial econometric models are urged to see LeSage (1997) which develops the Bayesian SAR model with homoskedastic as well as heteroskedastic errors. LeSage (2000) develops the Bayesian spatial probit model, while Holloway, Shankar, and Rahman (2002) offer a tutorial on the use and development of the Bayesian spatial probit model. Finally, Jim LeSage's website (www.spatial-econometrics.com) has MATLAB code and realated reading materials devoted to all aspects of spatial econometrics in general and Bayesian variants of these models in particular.

6 References

Holloway, Garth J., Bhavani Shankar, and Sanzidur Rahman. "Bayesian Spatial Probit Estimation: A Primer and an Application to HYV Rice Adoption" *Agricultural Economics*, vol. 27, no. 3, November 2002, pp. 383-402.

LeSage, James P. "Bayesian Estimation of Spatial Autoregressive Models" *International Regional Science Review*, vol. 20, no. 1&2, 1997, pp.113-129.

LeSage, James P. "Bayesian Estimation of Limited Dependent Variable

Spatial Autoregressive Models" *Geographical Analysis*, vol. 32, no. 1, January 2000, pp. 19-35.

LeSage, James P. www.spatial-econometrics.com

Zellner, Arnold. An Introduction to Bayesian Inference in Econometrics. New York: John Wiley and Sons, Wiley Classics Library Edition, 1996.