Lecture 5: Spatial probit models

James P. LeSage
University of Toledo
Department of Economics
Toledo, OH 43606
jlesage@spatial-econometrics.com

March 2004
1 A Bayesian spatial probit model with individual effects

Probit models with spatial dependencies were first studied by McMillen (1992), where an EM algorithm was developed to produce consistent (maximum likelihood) estimates for these models. As noted by McMillen, such estimation procedures tend to rely on asymptotic properties, and hence require large sample sizes for validity. An alternative hierarchical Bayesian approach to non-spatial probit models was introduced by Albert and Chib (1993) which is more computationally demanding, but provides a flexible framework for modeling with small sample sizes. LeSage (2000) first proposed extending Albert and Chib’s approach to models involving spatial dependencies, and Smith and LeSage (2001) extend the class of models that can be analyzed in this framework. They introduce an error structure that involves an additive error specification first introduced by Besag, et al. (1991) and subsequently employed by many authors (as for example in Gelman, et al. 1998). Smith and LeSage (2001) show that this approach allows both spatial dependencies and general spatial heteroscedasticity to be treated simultaneously.

1.1 Choices involving spatial agents

For a binary 0, 1 choice, made by individuals $k$ in region $i$ with alternatives labeled $a = 0, 1$:

\[
\begin{align*}
U_{ik0} & = \gamma'\omega_{ik0} + \alpha_0's_{ik} + \theta_{i0} + \varepsilon_{ik0} \\
U_{ik1} & = \gamma'\omega_{ik1} + \alpha_1's_{ik} + \theta_{i1} + \varepsilon_{ik1}
\end{align*}
\]

(1)

Where:

- $\omega$ represent **observed** attributes of the $a = 0, 1$ alternative
- $s$ represent **observed** attributes of individuals $k$
- $\theta_{ika} + \varepsilon_{ika}$ represent **unobserved** properties of individuals $k$, regions $i$ or alternatives $a$

We decompose the **unobserved** effects on utility into:

- a regional effect $\theta_{ia}$, assuming homogeneity across individuals $k$ in region $i$.
- an individualistic effect $\varepsilon_{ika}$

The individualistic effects, $\varepsilon_{ika}$ are assumed conditionally independent given $\theta_{ia}$, so unobserved dependencies between individual utilities for alternative $a$ within region $i$ are captured by $\theta_{ia}$.

Following Amemiya (1995, section 9.2) one can use utility differences between individuals $k$ along with the utility maximization hypothesis to arrive at a probit regression relationship.

\[
\begin{align*}
z_{ik} & = U_{ik1} - U_{ik0} \\
& = x_{ik}'\beta + \theta_i + \varepsilon_{ik}
\end{align*}
\]

(2)
1.2 Spatial autoregressive unobserved interaction effects

Smith and LeSage (2001) model the unobserved dependencies between utility differences of individuals in separate regions (the regional effects $\theta_i : i = 1, \ldots, m$) as following a spatial autoregressive structure:

$$\theta_i = \rho \sum_{j=1}^{m} w_{ij} \theta_j + u_i, \quad i = 1, \ldots, m$$

$$u \sim N(0, \sigma^2 I_m)$$

Intuition here is that unobserved utility-difference aspects that are common to individuals in a given region may be similar to those for individuals in neighboring or nearby regions.

It is convenient to solve for $\theta$ in terms of $u$ which we will rely on in the sequel. Let

$$B_\rho = I_m - \rho W$$

and assume that $B_\rho$ is nonsingular, then from (3):

$$\theta = B_\rho^{-1} u \Rightarrow \theta | (\rho, \sigma^2) \sim N[0, \sigma^2 (B_\rho B_\rho^{-1})]$$

1.3 Heteroscedastic individual effects

Turning next to the individualistic components, $\varepsilon_{ik}$, observe that without further evidence about specific individuals in a given region $i$, it is reasonable to treat these components as exchangeable and hence to model the $\varepsilon_{ik}$ as conditionally iid normal variates with zero means and common variance $v_i$, given $\theta_i$. In particular, regional differences in the $v_i$’s allow for possible heteroscedasticity effects in the model. Hence, if we now denote the vector of individualistic effects of region $i$ by $\varepsilon_i = (\varepsilon_{ik} : k = 1, \ldots, n_i)'$, then our assumptions imply that $\varepsilon_i | \theta_i \sim N(0, v_i I_{n_i})$.

We can express the full individualistic effects vector $\varepsilon = (\varepsilon_i' : i = 1, \ldots, m)'$ as

$$\varepsilon | \theta \sim N(0, V)$$

where the full covariance matrix $V$ is shown in (7).

$$V = \begin{pmatrix} v_1 I_{n_1} & & \\ \vdots & \ddots & \\ & & v_m I_{n_m} \end{pmatrix}$$

We emphasize here that as motivated earlier, all components of $\varepsilon$ are assumed to be conditionally independent given $\theta$.

Expression (2) can also be written in vector form by setting $z_i = (z_{ik} : k = 1, \ldots, n_i)'$ and $X_i = (x_{ik} : k = 1, \ldots, n_i)'$, so the utility differences for each region $i$ take the form:
\[ z_i = X_i \beta + \theta_i \mathbf{1}_i + \varepsilon_i, \quad i = 1, \ldots, m \quad (8) \]

where \( \mathbf{1}_i = (1, \ldots, 1)' \) denotes the \( n_i \)-dimensional unit vector. Then by setting \( n = \sum_i n_i \) and defining the \( n \)-vectors \( z = (z'_i : i = 1, \ldots, m)' \) and \( X = (X'_i : i = 1, \ldots, m)' \), we can reduce (8) to the single vector equation,

\[ z = X \beta + \Delta \theta + \varepsilon \quad (9) \]

where

\[
\Delta = \begin{pmatrix}
1 \\
\vdots \\
1_m
\end{pmatrix}
\quad (10)
\]

If the vector of regional variances is denoted by \( v = (v_i : i = 1, \ldots, m) \), then the covariance matrix \( V \) in (6) can be written using this notation as

\[ V = \text{diag}(\Delta v) \quad (11) \]

2 Albert and Chib (1993) latent treatment of \( z \)

\[
\Pr(Y_{ik} = 1|z_{ik}) = \delta(z_{ik} > 0) \\
\Pr(Y_{ik} = 0|z_{ik}) = \delta(z_{ik} \leq 0) \quad (12)
\]

Where: \( \delta(A) \) is an indicator function \( \delta(A) = 1 \) for all outcomes in which \( A \) occurs and \( \delta(A) = 0 \) otherwise.

If the outcome value \( Y = (Y_{ik} \in 0, 1) \), then [following Albert and Chib (1993)] these relations may be combined as follows:

\[
\Pr(Y_{ik} = y_{ik}) = \delta(y_{ik} = 1) \delta(z_{ik} > 0) + \delta(y_{ik} = 0) \delta(z_{ik} \leq 0) \quad (13)
\]

Which produces a conditional posterior for \( z_{ik} \) that is a \textit{truncated} normal distribution, which can be expressed as follows:

\[
\left. z_{ik} \right| \ast \sim \begin{cases}
N(x'_i \beta + \theta_i, v_i) & \text{left-truncated at 0, if } y_i = 1 \\
N(x'_i \beta + \theta_i, v_i) & \text{right-truncated at 0, if } y_i = 0
\end{cases} \quad (14)
\]

\(^1\text{Note again that by assumption } X \text{ always contains } m \text{ columns corresponding to the indicator functions, } \delta(\cdot), i = 1, \ldots, m.\)
2.1 Hierarchical Bayesian Priors

The following prior distributions are standard [see LeSage, (1999)]:

\[
\begin{align*}
\beta & \sim N(c, T) \\
r/v_i & \sim ID\chi^2(r) \\
1/\sigma^2 & \sim \Gamma(\alpha, \nu) \\
\rho & \sim U[(\lambda^{-1}_{\min}, \lambda^{-1}_{\max}] 
\end{align*}
\]

These induce the following priors:

\[
\begin{align*}
\pi(\theta | \rho, \sigma^2) & \sim (\sigma^2)^{-m/2}|B_{\rho}|\exp \left( -\frac{1}{2\sigma^2}\theta' B_{\rho}' B_{\rho} \theta \right) \\
B_{\rho} & = I_m - \rho W \\
\pi(\varepsilon | V) & \sim |V|^{-1/2}\exp \left( -\frac{1}{2}\varepsilon' V^{-1} \varepsilon \right) \\
\pi(z | \beta, \theta, V) & \propto |V|^{-1/2}\exp \left\{ -\frac{1}{2} e' V^{-1} e \right\} \\
e & = z - X\beta - \Delta\theta
\end{align*}
\]

3 Estimation via MCMC

Estimation will be achieved via Markov Chain Monte Carlo methods that sample sequentially from the complete set of conditional distributions for the parameters. The complete conditional distributions for all parameters in the model are derived in Smith and LeSage (2001)

A few comments on innovative aspects:

3.1 The conditional distribution of \( \theta \):

\[
p(\theta | \beta, \rho, \sigma^2, V, z, y) \sim N(A_0^{-1} b, A_0^{-1})
\]

where the mean vector is \( A_0^{-1} b \) and the covariance matrix is \( A_0^{-1} \), which involves the inverse of the \( m \times m \) matrix \( A_0 \) which depends on \( \rho \). This implies that this matrix inverse must be computed on each MCMC draw during the estimation procedure. Typically a few thousand draws will be needed to produce a posterior estimate of the parameter distribution for \( \theta \), suggesting that this approach to sampling from the conditional distribution of \( \theta \) may be costly in terms of time if \( m \) is large.

In our illustration we rely on a sample of 3,110 US counties and the 48 contiguous states, so that \( m = 48 \). In this case, computing the inverse was relatively fast allowing us
to produce 2,500 draws in 37 seconds using a compiled c-language program on an Anthalon 1200 MHz. processor.

In the Appendix an alternative approach that involves only univariate normal distributions for each element $\theta_i$ conditional on all other elements of $\theta$ excluding the $i$th element.

This approach is amenable to computation for much larger sizes for $m$. It was used to solve a problem involving 59,025 US census tracts with the $m = 3,110$ counties as the regions. The time required was 357 minutes for 4500 draws.

3.2 The conditional distribution of $\rho$:

$$p(\rho|*) \propto |B_\rho| \exp \left( -\frac{1}{2\sigma^2} \theta'(I_m - \rho W)'(I_m - \rho W)\theta \right)$$

(23)

where $\rho \in [\lambda^{-1}_{\min}, \lambda^{-1}_{\max}]$. As noted in LeSage (2000) this is not reducible to a standard distribution, so we might adopt a M-H step during the MCMC sampling procedures. LeSage (1999) suggests a normal or $t-$ distribution be used as a transition kernel in the M-H step. Additionally, the restriction of $\rho$ to the interval $[\lambda^{-1}_{\min}, \lambda^{-1}_{\max}]$ can be implemented using a rejection-sampling step during the MCMC sampling.

Another approach that is feasible for this model is to rely on univariate numerical integration to obtain the conditional posterior density of $\rho$. The size of $(I_m - \rho W)$ will be based on the number of regions, which is typically much smaller than the number of observations, making it computationally simple to carry out univariate numerical integration on each pass through the MCMC sampler.

An advantage of this approach over the M-H method is that each pass through the sampler produces a draw for $\rho$, whereas acceptance rates in the M-H method are usually around 50 percent requiring twice as many passes through the sampler to produce the same number of draws for $\rho$.

3.3 Special cases of the model

• The homoscedastic case, where we let: individual variances are assumed equal across all regions, so the regional variance vector, $v$ reduces to a scalar

• The individual spatial-dependency case: where individuals are treated as ‘regions’ denoted by the index $i$. In this case we are essentially setting $m = n$ and $n_i = 1$ for all $i = 1, \ldots, m$.

• Note that although one could in principle consider heteroscedastic effects among individuals, the existence of a single observation per individual renders estimation of such variances problematic at best.

4 Applied Examples

4.1 Generated data examples

This experiment used $n = 3,110$ US counties to generate a set of data. The $m = 48$ contiguous states were used as regions. A continuous dependent variable was generated using the following procedure. First, the spatial interaction effects were generated using:
where $\rho$ was set equal to 0.7 in one experiment and 0.6 in another. In (24), $W$ represents the 48x48 standardized spatial weight matrix based on the centroids of the states.

Six explanatory variables which we label $X$ were created using county-level census information on: the percentage of population in each county that held high school, college, or graduate degrees, the percentage of non-white population, the median household income (divided by 10,000) and the percent of population living in urban areas. These are the same explanatory variables we use in our application to the 1996 presidential election, which should provide some insight into how the model operates in a generated data setting.

4.2 Presidential election application

To illustrate the model in an applied setting we used data on the 1996 presidential voting decisions in each of 3,110 US counties in the 48 contiguous states. The dependent variable was set to 1 for counties where Clinton won the majority of votes and 0 for those where Dole won the majority.\(^2\) To illustrate individual versus regional spatial interaction effects

\[^{2}\text{The third party candidacy of Perot was ignored and only votes for Clinton and Dole were used to make this classification of 0,1 values.}\]
Figure 2: Individual effects estimates from homoscedastic and heteroscedastic spatial probit models

Table 1: Generated data results, averaged over 100 samples

<table>
<thead>
<tr>
<th>Estimates</th>
<th>ols</th>
<th>probit</th>
<th>sprobit</th>
<th>sregress</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1 = 3$</td>
<td>0.2153</td>
<td>1.5370</td>
<td>2.9766</td>
<td>2.9952</td>
</tr>
<tr>
<td>$\beta_2 = -1.5$</td>
<td>-0.1291</td>
<td>-0.8172</td>
<td>-1.5028</td>
<td>-1.5052</td>
</tr>
<tr>
<td>$\beta_3 = -3$</td>
<td>-0.0501</td>
<td>-1.5476</td>
<td>-2.9924</td>
<td>-2.9976</td>
</tr>
<tr>
<td>$\beta_4 = 2$</td>
<td>0.1466</td>
<td>1.0321</td>
<td>2.0019</td>
<td>2.0013</td>
</tr>
<tr>
<td>$\beta_5 = -1$</td>
<td>-0.0611</td>
<td>-0.5233</td>
<td>-0.9842</td>
<td>-1.0013</td>
</tr>
<tr>
<td>$\beta_6 = 1$</td>
<td>0.0329</td>
<td>0.5231</td>
<td>0.9890</td>
<td>1.0006</td>
</tr>
<tr>
<td>$\rho = 0.7$</td>
<td>0.6585</td>
<td>0.6622</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2 = 2$</td>
<td>2.1074</td>
<td>2.0990</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Standard deviations</th>
<th>ols</th>
<th>probit</th>
<th>sprobit</th>
<th>sregress</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_{\beta_1}$</td>
<td>0.0286</td>
<td>0.2745</td>
<td>0.1619</td>
<td>0.0313</td>
</tr>
<tr>
<td>$\sigma_{\beta_2}$</td>
<td>0.0434</td>
<td>0.3425</td>
<td>0.1463</td>
<td>0.0393</td>
</tr>
<tr>
<td>$\sigma_{\beta_3}$</td>
<td>0.0346</td>
<td>0.4550</td>
<td>0.2153</td>
<td>0.0390</td>
</tr>
<tr>
<td>$\sigma_{\beta_4}$</td>
<td>0.0256</td>
<td>0.2250</td>
<td>0.1359</td>
<td>0.0252</td>
</tr>
<tr>
<td>$\sigma_{\beta_5}$</td>
<td>0.0176</td>
<td>0.1630</td>
<td>0.1001</td>
<td>0.0293</td>
</tr>
<tr>
<td>$\sigma_{\beta_6}$</td>
<td>0.0109</td>
<td>0.1349</td>
<td>0.0819</td>
<td>0.0244</td>
</tr>
<tr>
<td>$\sigma_\rho$</td>
<td>0.1299</td>
<td>0.1278</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_\sigma$</td>
<td>0.5224</td>
<td>0.3971</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
we treat the counties as individuals and the states as regions where the spatial interaction effects occur.

As explanatory variables we used: the proportion of county population with high school degrees, college degrees, and graduate or professional degrees, the percent of the county population that was non-white, the median county income (divided by 10,000) and the percentage of the population living in urban areas. These were the same variables used in the generated data experiments, and we applied the same studentize transformation here as well. Of course, our application is illustrative rather than substantive.

Diffuse or conjugate priors were employed for all of the parameters $\beta, \sigma^2$ and $\rho$ in the Bayesian spatial probit models.

A hyperparameter value of $r = 4$ was used for the heteroscedastic spatial probit model, and a value of $r = 40$ was employed for the homoscedastic prior.

The heteroscedastic value of $r = 4$ implies a prior mean for $r$ equal to $r/(r - 2) = 2$ and a prior standard deviation equal to $\sqrt{2/r} = 0.707$. A two standard deviation interval around this prior mean would range from 0.58 to 3.41, suggesting that posterior estimates for individual states larger than 3.4 would indicate evidence in the sample data against homoscedasticity.

The posterior mean for the $v_i$ estimates was greater than this upper level in 13 of the 48 states, with a mean over all states equal to 2.86 and a standard deviation equal to 2.36. The frequency distribution of the 48 $v_i$ estimates suggests the mean is not representative for this skewed distribution. We conclude there is evidence in favor of mild heteroscedasticity.
Table 2: 1996 Presidential Election results

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. deviation</th>
<th>P-level†</th>
</tr>
</thead>
<tbody>
<tr>
<td>high school</td>
<td>0.0976</td>
<td>0.0419</td>
<td>0.0094</td>
</tr>
<tr>
<td>college</td>
<td>-0.0393</td>
<td>0.0609</td>
<td>0.2604</td>
</tr>
<tr>
<td>grad/professional</td>
<td>0.1023</td>
<td>0.0551</td>
<td>0.0292</td>
</tr>
<tr>
<td>non-white</td>
<td>0.2659</td>
<td>0.0375</td>
<td>0.0000</td>
</tr>
<tr>
<td>median income</td>
<td>-0.0832</td>
<td>0.0420</td>
<td>0.0242</td>
</tr>
<tr>
<td>urban population</td>
<td>-0.0261</td>
<td>0.0326</td>
<td>0.2142</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.5820</td>
<td>0.0670</td>
<td>0.0000</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.6396</td>
<td>0.1765</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. deviation</th>
<th>P-level†</th>
</tr>
</thead>
<tbody>
<tr>
<td>high school</td>
<td>0.0898</td>
<td>0.0446</td>
<td>0.0208</td>
</tr>
<tr>
<td>college</td>
<td>-0.1354</td>
<td>0.0738</td>
<td>0.0330</td>
</tr>
<tr>
<td>grad/professional</td>
<td>0.1787</td>
<td>0.0669</td>
<td>0.0010</td>
</tr>
<tr>
<td>non-white</td>
<td>0.3366</td>
<td>0.0511</td>
<td>0.0000</td>
</tr>
<tr>
<td>median income</td>
<td>-0.1684</td>
<td>0.0513</td>
<td>0.0002</td>
</tr>
<tr>
<td>urban population</td>
<td>-0.0101</td>
<td>0.0362</td>
<td>0.3974</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.6176</td>
<td>0.0804</td>
<td>0.0000</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.9742</td>
<td>0.3121</td>
<td></td>
</tr>
</tbody>
</table>

† see Gelman, Carlin, Stern and Rubin (1995) regarding p-levels

References


Appendix

This appendix derives a sequence of univariate conditional posterior distributions for each element of \( \theta \) that allows the MCMC sampling scheme proposed here to be applied in larger models. For models with less than \( m = 100 \) regions it is probably faster to simply compute the inverse of the \( mxm \) matrix \( A_0 \) and use the multinormal distribution presented in (22). For larger models this can be computationally burdensome as it requires large amounts of memory.

First, note that we can write:

\[
p(\theta|\star) \propto \pi(z|\beta, \theta, V) \cdot \pi(\theta|\rho, \sigma^2) \\
\propto \exp \left\{ -\frac{1}{2} \left( \Delta \theta - (z - X\beta) \right)'V^{-1} \left( \Delta \theta - (z - X\beta) \right) \right\} \cdot \\
\exp \left\{ -\frac{1}{2} \theta' B'_\rho B_\rho \theta \right\} \\
= \exp \left\{ -\frac{1}{2} \left[ \theta' \Delta'V^{-1}\Delta \theta - 2(z - X\beta)'V^{-1}\Delta \theta + \theta' (\sigma^{-2} B'_\rho B_\rho) \theta \right] \right\} \\
= \exp \left\{ -\frac{1}{2} \left[ \theta' (\sigma^2 B'_\rho B_\rho + \Delta'V^{-1}\Delta) \theta - 2(z - X\beta)'V^{-1}\Delta \theta \right] \right\} \tag{25}
\]

The univariate conditional distributions are based on the observation that the joint density in (25) involves no inversion of \( A_0 \), and hence is easily computable. Since the univariate conditional posteriors of each component, \( \theta_i \) of \( \theta \) must be proportional to this density, it follows that each is univariate normal with a mean and variance that are readily computable.

To formalize these observations, observe first that if for each realized value of \( \theta \) and each \( i = 1, \ldots, m \) we let \( \theta_{-i} = (\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_m) \), then:

\[
p(\theta_i|\star) = \frac{p(\theta, \beta, \rho, \sigma^2, V, z, y)}{p(\theta_{-i}, \beta, \rho, \sigma^2, V, z|y)} \propto p(\theta, \beta, \rho, \sigma^2, V, z|y) \\
\propto \pi(z|\beta, \theta, V) \cdot \pi(\theta|\rho, \sigma^2) \\
\propto \exp \left\{ -\frac{1}{2} \left[ \theta' (\sigma^{-2} B'_\rho B_\rho + \Delta'V^{-1}\Delta) \theta - 2(z - X\beta)'V^{-1}\Delta \theta \right] \right\} \tag{26}
\]

This expression can be reduced to terms involving only \( \theta_i \) as follows. If we let \( \phi = (\phi_i : i = 1, \ldots, m)' = ((z - X\beta)'V^{-1}\Delta)' \), then the bracketed expression in (26) can be written as,

\[
\theta' (\sigma^{-2} B'_\rho B_\rho + \Delta'V^{-1}\Delta) \Delta \theta - 2(z - X\beta)'V^{-1}\Delta \theta \\
= \frac{1}{\sigma^2} \theta' (I - \rho W')(I - \rho W) \theta + \theta' \Delta'V^{-1}\Delta \theta - 2\phi' \theta \\
= \frac{1}{\sigma^2} [\theta' \theta - 2\rho \theta' W \theta + \rho^2 \theta' W' W \theta] + \theta' \Delta'V^{-1}\Delta \theta - 2\phi' \theta \tag{27}
\]
But by permuting indices so that $\theta' = (\theta_i, \theta'_{-i})$, it follows that

$$
\theta'W \theta = \theta' \begin{pmatrix} w_i & W_{-i} \end{pmatrix} \begin{pmatrix} \theta_i \\ \theta_{-i} \end{pmatrix}
= \theta' (\theta_i w_i + W_{-i} \theta_{-i})
= \theta_i (\theta' w_i) + \theta' W_{-i} \theta_{-i}
$$

(28)

where $w_i$ is the $i$th column of $W$ and $W_{-i}$ is the $m \times (m-1)$ matrix of all other columns of $W$. But since $w_{ii} = 0$ by construction, it then follows that

$$
\theta'W \theta = \theta' \left( \sum_{j \neq i} \theta_j w_{ji} \right) + \theta_i \begin{pmatrix} \theta' w_i \\ W_{-i} \theta_{-i} \end{pmatrix}
= \theta_i \sum_{j \neq i} \theta_j (w_{ji} + w_{ij}) + C
$$

(29)

where $C$ denotes a constant not involving parameters of interest. Similarly, we see from (28) that

$$
\theta'W'W \theta = (\theta_i w_i + W_{-i} \theta_{-i})' (\theta_i w_i + W_{-i} \theta_{-i})
= \theta_i^2 w_i' w_i + 2\theta_i (w_i' W_{-i} \theta_{-i}) + C
$$

(30)

Hence, by observing that

$$
\theta' \theta = \theta_i^2 + C
$$

(31)

$$
\theta' \Delta' V^{-1} \Delta \theta = n_i \theta_i^2 / v_i + C
$$

(32)

$$
-2\phi' V^{-1} \theta = -2\phi_i \theta_i + C
$$

(33)

where the definition of $\phi = (\phi_i : i = 1, \ldots, m)'$ implies that each $\phi_i$ has the form

$$
\phi_i = \frac{1_i'(z_i - X_i \beta)}{v_i}, \quad i = 1, \ldots, m
$$

(34)

Finally, by substituting these results into (27), we may rewrite the conditional posterior density of $\theta_i$ as

$$
p(\theta_i | \star) \propto \exp \left\{ -\frac{1}{2} \left[ -2\rho \theta_i \sum_{j \neq i} \theta_j (w_{ji} + w_{ij}) \theta_i 
+ \rho^2 \theta_i^2 w_i' w_i + 2\rho^2 \theta_i (w_i' W_{-i} \theta_{-i})) \frac{1}{\sigma^2} + n_i \theta_i^2 / v_i - 2\phi_i \theta_i \right] \right\}
= \exp \left\{ -\frac{1}{2} \left( a_i \theta_i^2 - 2b_i \theta_i \right) \right\}
\propto \exp \left\{ -\frac{1}{2} \left( a_i \theta_i^2 - 2b_i \theta_i + b_i^2 / a_i \right) \right\}
= \exp \left\{ -\frac{1}{2(1/a_i)} \left( \theta_i - \frac{b_i}{a_i} \right)^2 \right\}
$$

(35)
and $a_i$ and $b_i$ are given respectively by

$$a_i = \frac{1}{\sigma^2} + \frac{\rho^2}{\sigma^2} w_i w_i + \frac{n_i}{v_i} \quad (36)$$

$$b_i = \phi_i + \frac{\rho}{\sigma^2} \sum_{j \neq i} \theta_j (w_{ji} + w_{ij}) \theta_j - \frac{\rho^2}{\sigma^2} w_i W - \theta_\cdot \quad (37)$$

Thus the density in (35) is seen to be proportional to a univariate normal density with mean, $b_i/a_i$, and variance, $1/a_i$, so that for each $i = 1, \ldots, m$ the conditional posterior distribution of $\theta_i$ given $\theta_\cdot$ must be of the form

$$\theta_i | (\theta_\cdot, \beta, \rho, \sigma^2, V, z, y) \sim N\left(\frac{b_i}{a_i}, \frac{1}{a_i}\right) \quad (38)$$