8

Bayesian Models for Spatial Data II: Hierarchical Models with Spatial Dependence

8.1 Introduction

One of the fundamental properties of spatial data is the property of spatial dependence – data values close together in geographical space tend to be more alike than data values that are further apart.[1] For the purpose of what is to follow in this chapter, we can express this property in a slightly different way: an observation on a variable at location i carries some information about what is observed for the same variable in areas that are close to i.[2] Now, if this property is evident in spatial *data* as identified, for example, through an exploratory spatial autocorrelation analysis (see Chapter 6), is it not reasonable to assume that it will be a property of *parameters* too – so that, for example, average household income levels (i.e., the θ_is as in Chapter 7) will be more alike in MSOAs that are close together than in MSOAs that are far apart? We argue that this assumption can be invoked to further strengthen and improve small area estimates. The exchangeable, global, hierarchical model in Chapter 7 (Strategy 3) borrowed information from all other areas in the study region. Now, we propose to be more discriminating, but how can this spatial approach to borrowing strength be implemented? We turn to that in this chapter.

As discussed in Chapter 7, the options we consider for information sharing are based on prior assumptions that we place on the spatial dependence structure of the area-specific parameters. To express such spatial dependency so that we are able to implement these processes of information sharing, we need models. The discussion of such models forms a significant part of this chapter. We draw on the Newcastle income data to provide the illustrative example, and the reader is referred back to Section 7.2 for details of the dataset and Strategy 4 (see Figure 7.2 in Chapter 7).

We shall discuss various spatial models for localised information sharing, all of them involving some form of the conditional autoregressive (CAR) modelling structure – a modelling structure, like the simultaneous autoregressive (SAR) structure to be encountered in Chapter 10, commonly used to capture spatial dependence when analysing data associated with irregular polygons (e.g. outcomes attached to census tracts) or fixed points (such as a set of retail outlets).

The CAR structure imposes a neighbourhood structure on the spatial units using a spatial weights matrix (Chapter 4), and it is through that neighbourhood structure that local information sharing is achieved. We discuss the intrinsic conditional autoregressive

[1] As expressed informally in Tobler's First Law of Geography (see Chapter 3).

[2] This way of looking at spatial dependence underlies geostatistics and in particular methods for spatial interpolation including kriging (see Chapter 2).

(ICAR) model in Section 8.2, a version of the CAR modelling structure that is widely used in practice, and the proper CAR (pCAR) model in Section 8.3. We shall consider the key issues that arise when specifying binary and general weights matrices in the ICAR model (Sections 8.2.1 and 8.2.2). In Section 8.4, we look at adaptive ICAR models, which allow the elements in the spatial weights matrix to be estimated using data. In Section 8.5, we introduce the Besag-York-Mollie (BYM) model, which combines both an exchangeable model and the ICAR model, so that borrowing information is carried out both globally *and* locally. In section 8.6, we summarise all the modelling results and provide some insights into the application of these different modelling options.

8.2 The Intrinsic Conditional Autoregressive (ICAR) Model

The presence of positive spatial autocorrelation implies that the attribute values of areas that are geographically close together are more alike than those that are geographically widely separated. For the purpose of estimation, the presence of spatial dependence implies that we can borrow strength "locally" so that the parameter of area i will be similar to those from the *neighbours* of this area. The Intrinsic Conditional AutoRegressive (ICAR) model (Besag et al., 1991) is widely used to give a spatial structure to a set of area-specific parameters, $S_1, \ldots, S_N$, over N areas within a study region. Here, we denote the area-specific parameters as $S_1, \ldots, S_N$, as opposed to $\theta_1, \ldots, \theta_N$, to emphasise that these parameters are spatially structured. $\boldsymbol{S} = (S_1, \ldots, S_N)$ denotes a collection of these area-specific parameters. The ICAR model is used as a *prior model* to impose a spatial dependence structure on $\boldsymbol{S}$.

However, this does not entirely resolve the issue, because the ICAR model depends on the selected spatial weights matrix, $\boldsymbol{W}$ (see Chapter 4). The choice of $\boldsymbol{W}$ will have a significant impact on how information is borrowed amongst areas, and as we saw in Chapter 4, there are many ways to define a spatial weights matrix. In Section 8.2.1, we discuss the ICAR model that defines a neighbouring structure using a spatial weights matrix with binary (0 or 1) entries. In Section 8.2.2, we discuss the ICAR model using a general form of the spatial weights matrix, the entries of which are continuous-valued.

8.2.1 The ICAR Model Using a Spatial Weights Matrix with Binary Entries

In many statistical applications, the spatial structure of a map is often (but not always) defined through spatial contiguity. For example, recall from Section 4.2, a spatial weights matrix, $\boldsymbol{W}$, defined based on rook's move contiguity consists of 0/1 entries, where $w_{ij} = 1$ if areas i and j share a common border (or edge) and $w_{ij} = 0$ otherwise. Also, $w_{ii} = 0$ for all $i = 1, \ldots, N$. Using the ICAR model with a rook's move $\boldsymbol{W}$ matrix induces a particular spatial dependence structure on the area-specific parameters. The choice of model (ICAR) and the choice of spatial weights matrix $\boldsymbol{W}$, together, will have a strong influence on the geography of the spatial smoothing. Choosing a different spatial weights matrix, for example using the queen's move definition of contiguity, will result in a (slightly) different spatial smoothing.[3] Other neighbourhood definitions based on geographical distance, for example, $w_{ij} = 1$ if the distance between the centroids of areas i and j is less than a predefined

[3] In the case of irregular spatial units, there may be very few instances of spatial units joined at a vertex, so in practice these two definitions of the weights matrix typically lead to very similar if not identical smoothings.

threshold and $w_{ij}=0$ otherwise, will again yield a binary weights matrix but will produce yet another pattern of local smoothing.[4]

The ICAR model defines a *conditional* normal distribution for the parameter of each area i so that the specification of this distribution depends on the parameters of the neighbours of i. Based on a binary weights matrix, the conditional distribution for each S_i ($i=1, \ldots, N$) is given by

$$S_i \mid \boldsymbol{S}_{\{-i\}}, v, \boldsymbol{W} \sim N\left(\frac{\sum_{j\in\Delta i} S_j}{m_i}, \frac{v}{m_i}\right) \tag{8.1}$$

where

- $\boldsymbol{S}_{\{-i\}}$ denotes the set of area-specific parameters excluding S_i, namely, $\boldsymbol{S}_{\{-i\}} = (S_1, \ldots, S_{i-1}, S_{i+1}, \ldots, S_N)$.
- v is an unknown variance parameter whose role in this model will be discussed later
- Δi denotes the set of neighbours of area i as defined in $\boldsymbol{W}$.
- m_i is the number of neighbours that area i has. It equals the sum of the ith row of the $\boldsymbol{W}$ matrix and is also denoted as w_{i+}.

The conditional specification given in Eq. 8.1 implies a local dependence structure that leads to a local or neighbourhood smoothing of the area-specific parameters. Given the parameters of all other areas, the conditional mean of S_i, namely, $\frac{\sum_{j\in\Delta i} S_j}{m_i}$, only depends on the parameters of the neighbouring areas, implying a local dependence structure rather than a global dependence structure (in the latter case, the conditional mean of S_i would depend on all other S_j with $j \neq i$). In addition, the conditional mean is taken as the mean of the parameters of the neighbouring areas so that the parameters are similar locally. When applied to the income modelling, the ICAR model defined in Eq. 8.1 assumes that MSOAs that are close in space have similar average income levels. In effect, the ICAR model assumes positive spatial autocorrelation on the area-specific parameters, an assumption that should be tested at the exploratory stage of data analysis (see Chapter 6).

Positive spatial dependence induces local smoothing in the process of estimating the area-specific parameters. This local smoothing is evident in the conditional mean of S_i, which is set to be the mean of the parameters in the neighbouring areas. Instead of pulling towards the global mean as in the global smoothing model defined in Eq. 7.6 and Eq. 7.7 in Chapter 7, the ICAR model pulls the estimate of S_i towards the mean of the parameters in the neighbouring areas, thus smoothing the parameter estimates locally. For example, if we know (or are informed) that the average income levels of the three neighbours of area i are, say, 610, 570 and 620, then, prior to obtaining data for area i, we would expect the average income of area i to be around 600, the mean of its neighbours. Once we have obtained some data for this area, the posterior estimate for this area will lie between its sample mean and the average of its three neighbours. The question is, how close would the estimated income level be to the average of its neighbours? In other words, what are the factors that can affect the amount of local shrinkage under the ICAR model?

As in the case of global smoothing, the larger the sample size, n_i, the less local shrinkage a parameter will receive. Expressing this from a Bayesian standpoint, the ICAR model acts as a prior distribution for the area-specific parameters. If the dataset is sufficiently large,

[4] The ICAR model also requires the $\boldsymbol{W}$ matrix to be symmetric. Therefore $\boldsymbol{W}$ cannot be derived based on the k nearest neighbours or spatial interactions discussed in Chapter 4. This requirement will be discussed in Section 8.2.2.

the likelihood dominates the posterior distribution, and hence the prior distribution plays a minimal role. Another factor that can affect the amount of local shrinkage is m_i, the number of spatial neighbours of area i. Eq. 8.1 suggests that the more neighbours an area has, the more local smoothing it will receive. This is because, compared to an area that has only one or two neighbours, an area with a large number of neighbours (a large m_i), the conditional variance, $\frac{v}{m_i}$, as in Eq. 8.1, is small. Then its parameter, S_i, is more likely to be close to the average of the area's neighbours. In the extreme (although unrealistic) case where m_i tends to infinity (i.e. area i has an infinite number of neighbours), the conditional variance tends to 0. This is equivalent to placing a strong prior that the estimate of S_i simply takes the value of the local neighbourhood average, even before obtaining data for this area. When strong positive spatial autocorrelation is present everywhere on the map, it is reasonable to assume that more neighbours lead to more local smoothing. This is because when the neighbours are shown to be similar, more neighbours means that more data can be appropriately borrowed. However, when there is weak or no spatial autocorrelation, or when the (strong) spatial autocorrelation is only present in some part of the study region but not elsewhere, such a "global" effect of m_i on local shrinkage may not be appropriate. The latter situations lead to the discussion of adaptive smoothing, a topic that we shall return to in Section 8.4.

Finally, the variance parameter v controls the amount of local smoothing for the entire study region. Specifically, when v is large, the conditional variances for all the parameters in $\boldsymbol{S}$ are also large. The amount of local smoothing will, as a result, reduce for all areas, giving a less smooth-looking map of the posterior means of the parameters in $\boldsymbol{S}$. This reduction of smoothing applies to every area regardless of its neighbourhood structure. This variance parameter v is the only unknown parameter in the ICAR model, and we will provide an example of its prior distribution in the analysis of the income data.

It is important to point out that local information sharing is reciprocated. This means that when the estimation of S_i borrows information from area j, a neighbouring area of i, then at the same time, the estimation of S_j borrows information from area i. This reciprocity is important and explains why the neighbourhood smoothing process is *not* limited to just those neighbours defined by the spatial weights matrix $\boldsymbol{W}$. When estimating S_i, information is borrowed not only from the neighbours of area i as defined in $\boldsymbol{W}$ but also from its neighbours' neighbours, and so on. However, it is also the case that the *amount* of information borrowed from areas further away is less than from areas that are nearby (see also Section 4.9.1).

8.2.1.1 The WinBUGS Implementation of the ICAR Model

In WinBUGS, the ICAR model is implemented via the `car.normal` function, and the syntax is given below:

```
S[1:N] ~ car.normal(adj[], weights[], num[], precision)
```

where `N` denotes the number of areas in the study region. The first three arguments of the `car.normal` function, `adj[]`, `weights[]` and `num[]`, together define the chosen spatial weights matrix $\boldsymbol{W}$. Specifically, `num[]` is an array of size N with the ith entry showing the number of neighbours of area i. The IDs of the neighbours and their associated weights are stored in `adj[]` and `weights[]`, respectively. See the Appendix 4.13.2 in Chapter 4 for the steps to create these three data arrays in R. We will provide an example of the three arrays derived from the Newcastle MSOA map later in Figure 8.1.

The last term in the `car.normal` function, `precision`, is the precision parameter, the inverse of the variance parameter v in Eq. 8.1 (i.e. *precision* $= 1/v$). This is a hyperparameter

```
#  The WinBUGS code for specifying the model with the ICAR prior
model {
  #  a for-loop to go through all the household level income data
  for (j in 1:nhhs) {
    #  defining the normal likelihood for each household
    y[j] ~ dnorm(theta[msoa[j]],prec.y)
  }
  for (i in 1:nmsoas) {
    #  each theta is the sum of the Newcastle average and the
    #  corresponding S[i]
    theta[i] <- alpha + S[i]
  }
  #  modelling all the parameters in S using the ICAR model
  S[1:nmsoas] ~ car.normal(adj[],weights[],num[],prec.S)

  #  the improper uniform prior on the whole real line for the
  #  intercept
  alpha ~ dflat()

  #  a vague prior for the hyperparameter in the ICAR model
  sigma.S ~ dunif(0.0001,1000)

  #  a vague prior for the sampling standard deviation
  sigma.y ~ dunif(0.0001,1000)

  #  calculate the two precisions
  prec.S <- pow(sigma.S,-2)
  prec.y <- pow(sigma.y,-2)

  #  calculate the household-level variance for the VPC calculation
  var.y <- pow(sigma.y,2)
  #  calculate the MSOA-level variance (the unconditional variance
  #  of the S parameters) for the VPC calculation
  var.S.un <- sd(S[1:nmsoas]) * sd(S[1:nmsoas])
  #  calculate VPC defined in Eq. 7.8 in Chapter 7
  vpc <- var.S.un / (var.S.un + var.y) * 100
}

#  household-level income data from survey with a spatial
#  neighbourhood structure
list(nhhs=760,
nmsoas=109,
y=c(501, 616, 472, 816, 637, 500, 506, 560, 542, 447, 644, 522,
   ,487, 275,...),
msoa=c(1,1,1,1,1,2,2,4,4,4,4,5,5,...),
#  elements of the W matrix (defined via rook's move spatial
#  contiguity)
num=c(1,6,4,...),
adj=c(2,
      1,3,4,11,23,24,
      2,4,20,22,
      ...),
weights=c(1,1,1,1,1,1,1,1,1,1,...)
)

#  initial values for chain 1
list(alpha=200,sigma.y=50,sigma.S=20,S=c(1,-1,0,0,0,...))

#  initial values for chain 2
list(alpha=700,sigma.y=80,sigma.S=50,S=c(-1,1,0,0,0,...))
```

FIGURE 8.1
The WinBUGS code, the data list and the two sets of initial values for fitting the local smoothing model defined in Eq. 8.2 with the ICAR model using rook's move contiguity. The data list and the sets of initial values are shown only in part for the purpose of illustration.

for which a hyperprior is required. The `car.normal` function specifies a joint distribution for the entire set of area-specific parameters, which is denoted as `S[1:N]` (on the left-hand side of the expression).

The `car.normal` function in WinBUGS imposes the so-called sum-to-zero constraint on $\boldsymbol{S}$. That is, $\sum_{i=1}^{N} S_i = 0$. If we define the following line in WinBUGS and monitor the node `sum.S`, the value of this node is always 0.

```
sum.S <- sum(S[1:N])
```

This sum-to-zero constraint is required because the joint distribution for all the parameters in $\boldsymbol{S}$ (which can be derived from the conditional distributions given in Eq. 8.1) is an improper probability distribution, meaning that the joint distribution does not integrate to 1. We will defer discussion of the impropriety of the joint distribution to Section 8.2.2.2. Because of the sum-to-zero constraint, we always need to include a separate intercept term, say α. The prior distribution for the intercept must be the improper uniform prior defined on the whole real line, i.e. $\alpha \sim Uniform(-\infty, +\infty)$ (see Appendix 1 – under the Intrinsic CAR model section – of the GeoBUGS manual and also Section 8.2.2.2). This improper uniform distribution is coded as the `dflat()` function in WinBUGS.

We now turn to the Newcastle income example to illustrate the implementation of a Bayesian hierarchical model with the ICAR model using a binary $\boldsymbol{W}$ matrix derived from spatial contiguity.

8.2.1.2 Applying the ICAR Model Using Spatial Contiguity to the Newcastle Income Data

Recall y_{ij} is the income value for household j in MSOA i ($i = 1, \ldots, 109$). In Eq. 8.2, the (unknown) average income in MSOA i, θ_i, is modelled as the sum of α, the Newcastle average, and S_i. Because of the sum-to-zero constraint, each S_i is either above or below (or equal to) 0, and thus θ_i will be either above or below (or the same as) the Newcastle average accordingly. In other words, the term S_i measures the deviation of the income level in area i from the global average – the set $\boldsymbol{S} = (S_1, \ldots, S_{109})$ is also called a set of random effects because they vary from one area to another and they follow a common probability distribution with unknown hyperparameters. We can in fact rewrite the exchangeable model in Eq. 7.6 (Chapter 7) in the same format so that the MSOA average income is a sum of the global average and a local deviation. We will explore that formulation further when we discuss the BYM model in Section 8.5.

$$
\begin{aligned}
y_{ij} &\sim N(\theta_i, \sigma_y^2) \\
\theta_i &= \alpha + S_i \\
\boldsymbol{S}_{1:109} &\sim ICAR\left(\boldsymbol{W}, \sigma_S^2\right) \\
\alpha &\sim Uniform(-\infty, +\infty) \\
\sigma_y &\sim Uniform(0.0001, 1000) \\
\sigma_S &\sim Uniform\left(0.0001, 1000\right)
\end{aligned}
\tag{8.2}
$$

In this model, the prior distribution for $\boldsymbol{S} = \left(S_1, \ldots, S_{109}\right)$ is the ICAR model with the spatial structure defined by the spatial weights matrix $\boldsymbol{W}$. In this example, $\boldsymbol{W}$ is defined using

rook's move contiguity and the elements of the $\boldsymbol{W}$ matrix are entered into WinBUGS as data. σ_S^2 is the variance parameter v in Eq. 8.1. A vague uniform prior, $\text{Uniform}(0.0001, 1000)$, is assigned independently to the (conditional) standard deviation of the random effects, σ_S, and the household-level standard deviation, σ_y. The intercept α has the improper uniform prior, $\text{Uniform}(-\infty, +\infty)$.

Figure 8.1 summarises the WinBUGS code for the model, the data list and the initial values. For the model specification, Line 14 uses the `car.normal` function to assign the ICAR model to the vector of parameters $\boldsymbol{S} = (S_1, \ldots, S_{109})$. Line 36 calculates the VPC to compare the partition of the variance in the data between the household-level and the MSOA-level. Note that the MSOA-level variance, measuring the variability of average income across MSOAs, is calculated using the unconditional variance of the parameters $\boldsymbol{S}$ (see Line 34; the function `sd` calculates the standard deviation of a vector of values). We cannot use σ_S^2 for the VPC calculation because it measures the variability of the *conditional distribution* for $\boldsymbol{S}$ (see Eq. 8.1), rather than the *unconditional* variability of $\boldsymbol{S}$.

We now turn our attention to the three arrays, `adj[]`, `weights[]` and `num[]`, in the data list. The three arrays are derived from the off-diagonal elements in the $\boldsymbol{W}$ matrix. Using rook's move contiguity, $w_{ij} = 1$ if MSOAs i and j share a common border, and otherwise $w_{ij} = 0$ (for $i \neq j$). WinBUGS automatically assigns 0 to all the diagonal elements in $\boldsymbol{W}$. Using the Newcastle map in Figure 7.1 (Chapter 7), we can see that MSOA 1, the polygon at the southern tip of Newcastle, has only one neighbour, which is MSOA 2. Hence, the first entry of `num[]` is 1, meaning one neighbour for MSOA 1, and the first entry of `adj[]` is 2, the label (or ID) for that single neighbour. MSOA 2 has six neighbours, namely, MSOAs 3, 4, 24, 11, 23 and 1. Thus the second entry in `num[]` is 6 and the second to the seventh entries of `adj[]` are the labels of its neighbours. The above procedure is then used to derive the elements in `num[]` and `adj[]` for all the remaining MSOAs.

The elements in the array `weights[]` define the weights assigned to the neighbours of each MSOA. With a binary weights matrix, the weights of all neighbours are set to 1. This assumes all the neighbours of MSOA i contribute equally towards the conditional mean, and hence the estimation of S_i. For example, the conditional mean for MSOA 1 is $\frac{S_2}{1}$, and the conditional mean for MSOA 2 is $\frac{S_1 + S_3 + S_4 + S_{11} + S_{23} + S_{24}}{6}$. When using a binary $\boldsymbol{W}$ matrix, the weights of the neighbours are fixed to 1. However, in some situations, we might want to assign different weights to the neighbours based on, for example, the length of the common border or some socio-economic factors so that the extent of information sharing depends on not only the neighbourhood structure but also the similarity between neighbours based on some chosen set of local characteristics that are considered to be relevant to the analysis. In Section 8.2.2, we will explore a more general specification of the ICAR model using a more general form of weighting.

The initial values for the set of parameters $\boldsymbol{S}$ also need to be subjected to the sum-to-zero constraint. A simple way to achieve this, as shown in Figure 8.1, is to set the first element of $\boldsymbol{S}$ to 1, the second element to –1 and then set the rest to 0 (Line 57). For a second set of initial values, the same trick is used, but this time the positions of the values 1 and –1 are swapped (i.e., –1 and 1 for the first and the second elements, respectively) so that the initial values of $\boldsymbol{S}$ in the second set are different from those in the first set.[5]

[5] For a set of parameters modelled using the ICAR model, WinBUGS does not require the user to supply a set of initial values that sum to 0. However, OpenBUGS does. Here, we recommend specifying the initial values to sum to 0 so that the same set of codes can also be fitted in OpenBUGS.

8.2.1.3 Results

Figure 8.2 compares the posterior means from the Bayesian hierarchical model with the ICAR prior (Eq. 8.2) to those from the non-hierarchical model with independent parameters (Eq. 7.4 and Eq. 7.5 in Chapter 7). The first thing to remark is that under the ICAR model, each point estimate is smoothed towards the mean of its neighbouring areas, instead of towards the global mean. Initially this can be spotted by noticing that the lines, going from left to right, do not necessarily converge towards the Newcastle average, which is indicated by the horizontal dashed line. For example, for the lines lying above the Newcastle mean, a close inspection of Figure 8.2 shows that some lines are tilted upward while some are tilted down. Using a selection of MSOAs, Figure 8.3 further illustrates the local smoothing nature of the ICAR model. For each MSOA, the posterior mean of the average income (the triangle at the bottom of the plot) always lies between the sample mean (the cross) and the mean of its neighbours (the inverted triangle). In addition to the point values, in Figure 8.3, we have also shown the posterior distributions of the average income estimated for MSOA i and the mean income of that MSOA's neighbours to emphasise that each of these two quantities is associated with a probability distribution.

Figure 8.4 demonstrates the effect of the number of neighbours on the amount of local shrinkage. The latter, as defined along the vertical axis of Figure 8.4, is measured by $r_i = d_{i1}/d_{i2}$, where d_{i1} is the absolute difference between the posterior mean of θ_i and the sample mean and d_{i2} is the absolute difference between the posterior mean of θ_i and the local mean, which is taken as the average of the posterior means of the contiguous neighbours of i. The calculation of d_{i1} and d_{i2} can be seen in each panel of Figure 8.3 as follows: d_{i1} measures how far the triangle is from the cross and d_{i2} measures how far the triangle is

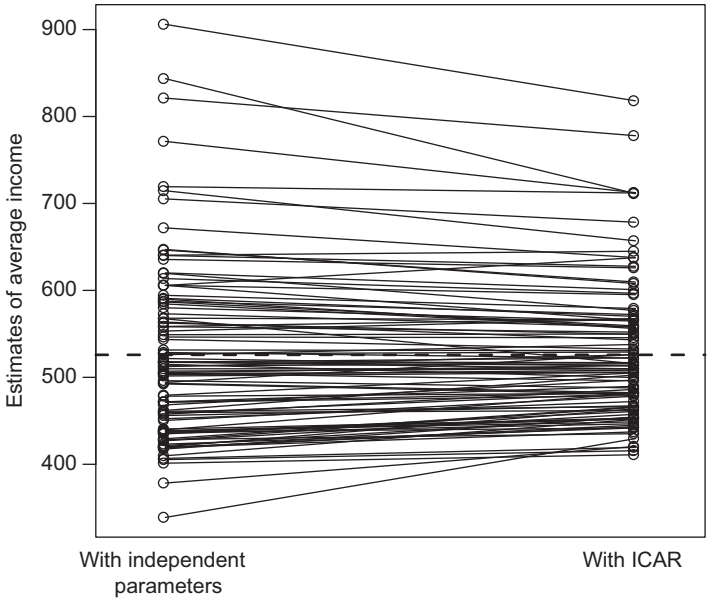

FIGURE 8.2
A shrinkage plot comparing the point estimates from the independent-parameters model (see Section 7.3.2) with those from the local smoothing model using the ICAR model and with the $\boldsymbol{W}$ matrix based on areas sharing a common border. This plot only shows the 99 MSOAs with survey data.

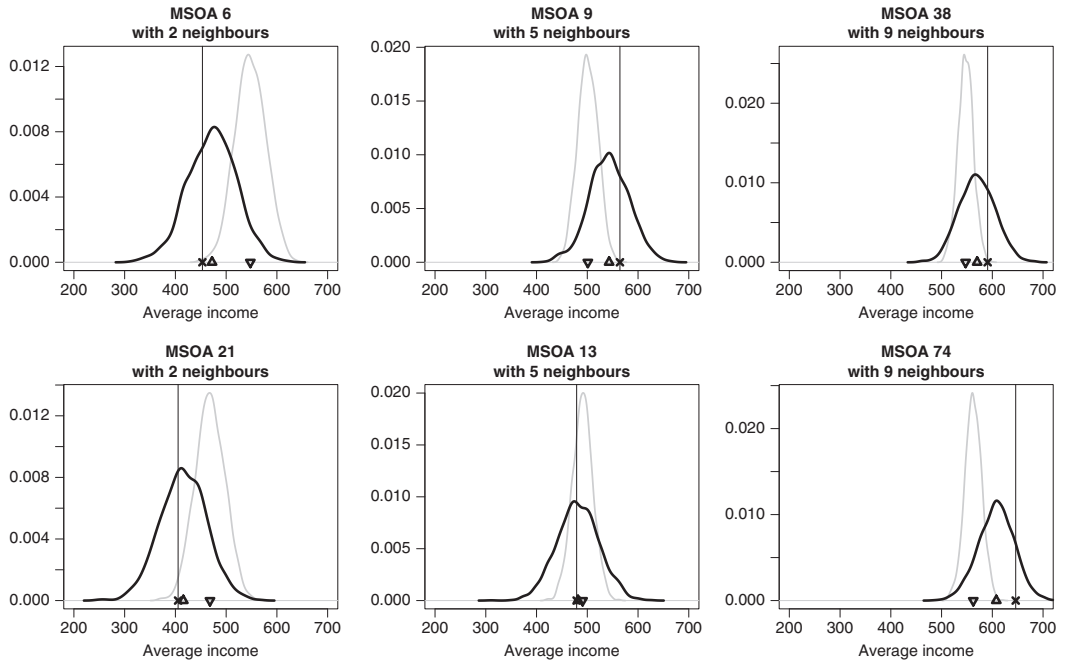

FIGURE 8.3
An illustration of the local smoothing nature under the ICAR model with six selected MSOAs. In each panel, the black curve denotes the posterior distribution of the average income θ_i in MSOA i, and it is constructed using the MCMC iterations for θ_i from WinBUGS. The grey curve is the corresponding conditional distribution of θ_i as defined in Eq. 8.1. That conditional distribution is calculated based on the posterior distributions of the θ_js from the neighbouring MSOAs ($j \in \Delta i$). That is, for each MCMC iteration, the mean of the sampled values of the θ_js is calculated, and the grey distribution is the distribution of those mean values. The vertical line indicates the sample mean from the survey data for MSOA i. At the bottom of each plot, the sample mean, the posterior mean and the conditional mean (given the neighbours) are denoted by a cross, a triangle and an inverted triangle, respectively. Because of local smoothing, the posterior mean (the triangle) always sits between the sample mean (the cross) and the mean of the neighbours (the inverted triangle).

from the inverted triangle. An r_i that is less than 1 means the posterior mean of θ_i is closer to the sample mean, and hence there is less local shrinkage, whereas r_i greater than 1 suggests more local shrinkage. It is evident from Figure 8.4 that the point estimate of an MSOA with more neighbours tends to be closer to the mean of its neighbours, thus receiving more local smoothing. On the other hand, an MSOA with fewer neighbours tends to receive less local shrinkage, so its posterior mean is closer to its sample mean.

Figure 8.4 also shows a considerable amount of variability in r_i even with the number of neighbours fixed. For example, with six neighbours, the values of r_i range from 0.28 to 2.30. As illustrated in Figure 8.5, such variability is primarily due to the varying sample size. With the number of neighbours fixed at 4 and 6 respectively in Figure 8.5(a) and Figure 8.5(b), the more data that an MSOA has, the smaller the value of r_i tends to be and the less information is borrowed from the neighbouring MSOAs. However, Figure 8.5 shows that there is still some variability left in the amount of local shrinkage even after accounting for the difference in sample size and in the number of neighbouring areas. For example, in Figure 8.5(a), MSOA 40 has a larger r_i than MSOA 48, although both have the

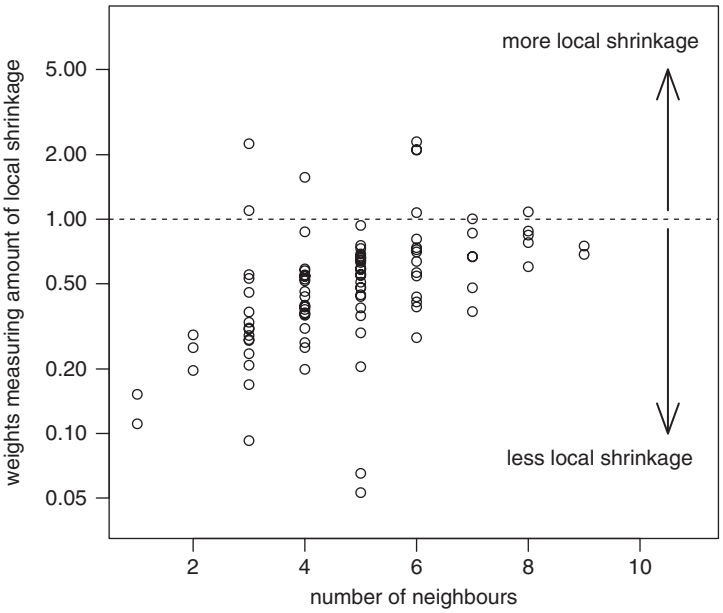

FIGURE 8.4
The effect of the number of neighbours on the amount of local shrinkage. The more neighbours that an area has, the more local shrinkage its parameter will receive.

same number of data values and the same number of neighbours. A closer inspection of their respective neighbours reveals that the income levels of the four neighbours of MSOA 40 are less variable than the income levels of the neighbours of MSOA 48 – for MSOA 40, the standard deviation (SD) of the posterior means of its neighbours is 5, while the SD for the neighbours of MSOA 48 is 26. This observation suggests that the consistency of information from the neighbouring areas also plays a role in the amount of local smoothing. For example, if the information provided from the neighbours is quite variable (e.g. some neighbours have high income levels while others have low), then the ICAR model borrows less from its neighbours. The two MSOAs highlighted in Figure 8.5(b) provide another example where MSOA 7 has more smoothing because its six neighbours show somewhat similar levels of income (the SD of the posterior means of the neighbours is 33). There is less local smoothing for MSOA 79 due to the more variable income levels of its neighbours (the SD of the neighbours is 53).

Figure 8.6 shows the summary of the posterior distributions for the average income across the 109 MSOAs. Because smoothing is local, the 10 MSOAs without data now have different posterior distributions, each of which depends on the mean of the parameters in their neighbouring areas. The map of the posterior means of the average income is given in Figure 8.7(a) while, for comparison, the map of the posterior means from the exchangeable model (Eq. 7.6 in Chapter 7) is given in Figure 8.7(b). Compared to the global smoothing model, the income estimates from the local smoothing model tend to be more similar in MSOAs that are spatially close. As a result, the spatial pattern in Figure 8.7(a) becomes more apparent. The MSOAs in the northwest and the northeast of Newcastle tend to have higher income, while those along the River Tyne show relatively lower income levels.

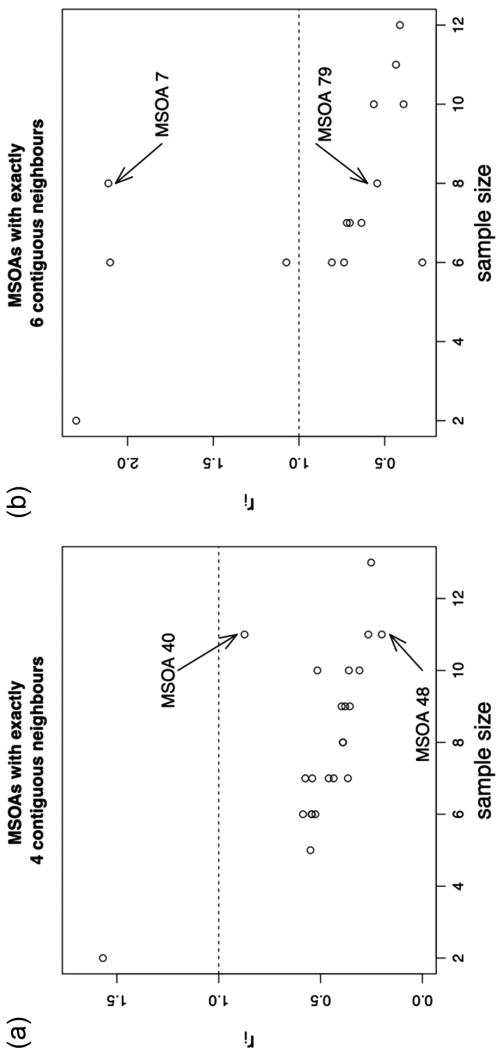

FIGURE 8.5
With the number of neighbours fixed at four in Panel (a) and at six in Panel (b), both plots show that the amount of local shrinkage tends to reduce with increasing sample size. The MSOAs highlighted in each panel demonstrate the additional variability in their weights (see the discussion in the main text).

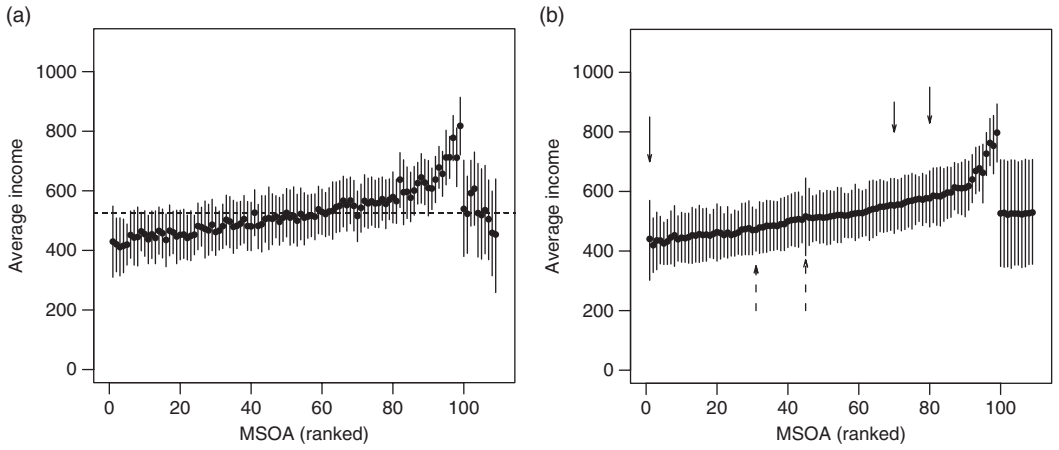

FIGURE 8.6
(a) A summary of the posterior distributions of the estimated average income for the 109 MSOAs obtained from Eq. 8.2, and for comparison, (b) posterior estimates for the MSOA-level average income based on the model with global smoothing (Eq. 7.6; an exact copy of Fig. 7.8(a) in Chapter 7 and see there for the explanation of the arrows). The solid dots are the posterior means and the vertical bars denote the 95% credible intervals.

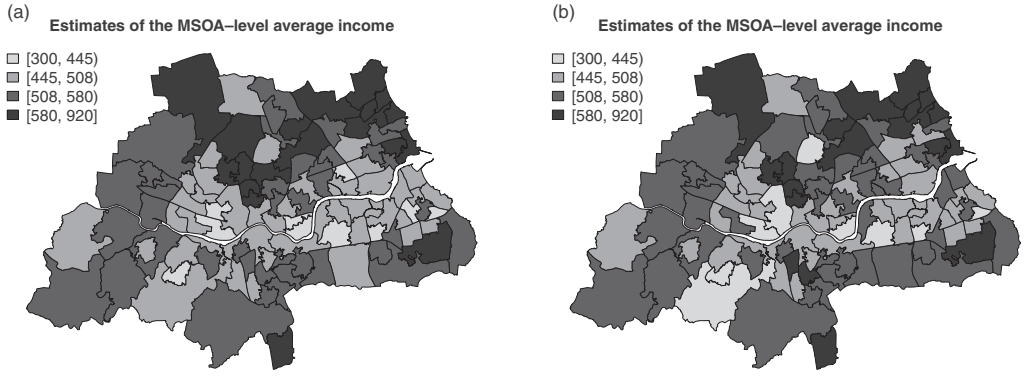

FIGURE 8.7
The maps of the posterior means of the average income (θ_is) from: (a) the local smoothing model using the ICAR model (Eq. 8.2) and (b) the global smoothing model based on exchangeability (Eq. 7.6 in Chapter 7).

8.2.1.4 A Summary of the Properties of the ICAR Model Using a Binary Spatial Weights Matrix

- The ICAR model assumes positive spatial autocorrelation and imposes a spatial neighbourhood structure on the area-specific parameters via the chosen spatial weights matrix.
- Through positive spatial autocorrelation, the ICAR model shares information locally (i.e. across a neighbourhood of nearby areas) and, as a result, produces estimates that are locally smoothed.
- The amount of local smoothing depends on the sample size in an area and the variability or consistency of the information from its neighbours.

- When using a spatial weights matrix with binary 0/1 entries, the number of neighbours of an area also affects the extent of local smoothing – the more neighbours an area has, the more local shrinkage its parameter will receive.

8.2.2 The ICAR Model with a General Weights Matrix

In Section 8.2.1, we introduced a version of the ICAR model where the spatial dependence structure is imposed via a binary spatial weights matrix with the weight associated with each neighbour fixed at 1. Using the equal weights of 1 has a direct implication on how information is shared locally: all neighbours of an area contribute the same amount of information towards the estimation of the parameter of this area. However, in addition to the spatial configuration of the areas, in some situations, other factors may also influence how similar an area is to each of its neighbours. For example, in estimating the income level, perhaps more information should be shared between two neighbouring areas if they are also similar in terms of their demographic structure. To do this, Eq. 8.3 presents the ICAR model that allows for a more general weights matrix:

$$S_i \mid \boldsymbol{S}_{\{-i\}}, v, \boldsymbol{W} \sim N\left(\frac{\sum_{j=1}^{N} w_{ij} S_j}{w_{i+}}, \frac{v}{w_{i+}}\right) \tag{8.3}$$

In Eq. 8.3, w_{ij} is the element of the ith row and the jth column in a spatial weights matrix $\boldsymbol{W}$, where the diagonal elements of $\boldsymbol{W}$ are all 0. $w_{i+} = \sum_{j=1}^{N} w_{ij}$ is the sum of the elements in the ith row. Eq. 8.3 reduces to Eq. 8.1 when the weights of the neighbours are set to 1 and the weights of the non-neighbouring areas are set to 0. This specification of the ICAR model can accommodate a more general form of $\boldsymbol{W}$ while possessing all the features of the ICAR model that we have discussed in the previous section. However, as we shall now explain, the ICAR model can only permit a *symmetric* spatial weights matrix. The weights matrix derived from contiguity (as used in the previous section), as well as other definitions introduced in Chapter 4, are symmetric, but those defined via the k-nearest neighbours in Section 4.3 and through interactions/flows in Section 4.6 are not. We will also explain the reason behind the sum-to-zero constraint imposed on $\boldsymbol{S}$ that we alluded to in the previous section. The exposition of these two topics requires defining the ICAR model via its joint distribution.

8.2.2.1 Expressing the ICAR Model as a Joint Distribution and the Implied Restriction on $\boldsymbol{W}$

Both Eq. 8.1 and Eq. 8.3 define the ICAR model via a set of N full *conditional* distributions (often referred to as the full conditionals), and each of these N full conditionals defines a probability distribution for each parameter given all the other parameters. Using the full conditionals in Eq. 8.3, one can derive *the joint probability distribution* for all the area-specific parameters. That is,

$$\begin{aligned} \Pr(S_1, \ldots, S_N) &= \Pr(\boldsymbol{S}) \\ &\propto \exp\left\{-\frac{1}{2v}\boldsymbol{S}'\left(\boldsymbol{D}_w - \boldsymbol{W}\right)\boldsymbol{S}\right\} \end{aligned} \tag{8.4}$$

In Eq. 8.4, the notation $\propto$ is the proportionality symbol, meaning that the expression on the second line of Eq. 8.4 ignores all the multiplicative constants (i.e. the terms that do not involve any elements of $\boldsymbol{S}$) in the joint distribution. $\boldsymbol{D}_w$ is an $N \times N$ diagonal matrix with the diagonal elements equal to the row sums of the $\boldsymbol{W}$ matrix. That is, $(\boldsymbol{D}_w)_{ii}$, the ith diagonal element of $\boldsymbol{D}_w$, is equal to $w_{i+} = \sum_{j=1}^{N} w_{ij}$ for all $i = 1, \ldots, N$. The derivation from Eq. 8.3 to Eq. 8.4 is beyond the scope of this book, but readers are referred to Cressie (1991, p.410–419) for more details. It is, however, easier to go from Eq. 8.4 to Eq. 8.3 (see Exercise 8.1), which can be used to verify the derivation of Eq. 8.4.

Eq. 8.4 resembles the density of a multivariate normal distribution with a mean vector of $\mathbf{0}$ and a covariance matrix $\boldsymbol{\Sigma}$ specified on a set of N random variables $\boldsymbol{X} = (X_1, \ldots, X_N)$, as given in Eq. 8.5:

$$\Pr(\boldsymbol{X}) \propto \exp\left\{-\frac{1}{2}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right\} \tag{8.5}$$

For the multivariate normal distribution, the covariance matrix $\boldsymbol{\Sigma}$ (or, equivalently, its inverse $\boldsymbol{\Sigma}^{-1}$) is required to be symmetric. This is because the covariance between two random variables is symmetric, namely, $cov(X_i, X_j) = cov(X_j, X_i)$. For a joint probability distribution to exist under the ICAR model, this same requirement is applied to the matrix $(\boldsymbol{D}_w - \boldsymbol{W})$ in Eq. 8.4. Since $\boldsymbol{D}_w$ is a diagonal matrix, the spatial weights matrix $\boldsymbol{W}$ must be symmetric, i.e. $w_{ij} = w_{ji}$ for all i and j, with $i \neq j$.

This symmetric restriction on $\boldsymbol{W}$ implies that the ICAR model assumes a spatial model in which the influence of area i on its neighbour j is the same as the influence of area j on i. However, if such an assumption is not appropriate (e.g. i has a larger influence on j than j has on i), then the ICAR model is not appropriate. For example, when modelling small area spatial variation in an economic variable, it is often the case that the economic relationships are asymmetric (see Chapter 2). In such cases, the simultaneous autoregressive model (see Chapter 10) or a spatial-temporal model (see Cressie, 1991, p.410) is a more suitable choice.

8.2.2.2 The Sum-to-Zero Constraint

As discussed in Section 8.2.1.1, the ICAR model places a sum-to-zero constraint on the area-specific parameters $\boldsymbol{S}$, namely, $\sum_{i=1}^{N} S_i = 0$. This constraint is required because the joint distribution in Eq. 8.4 is improper, meaning that this distribution does not integrate to 1. Here, we will first explain why the joint distribution is improper, then how the sum-to-zero constraint resolves the impropriety problem and the implication of the constraint for the interpretation of the area-specific parameters.

To see why the joint distribution is improper, rewrite the joint distribution in Eq. 8.4 as follows (see Exercise 8.2):

$$\Pr(S_1, \ldots, S_N) \propto \exp\left\{-\frac{1}{2v}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} w_{ij}(S_i - S_j)^2\right\} \tag{8.6}$$

We can now see that $S_1, \ldots, S_N$ enter the joint distribution as pairwise differences, and the joint distribution is not affected by the addition (or subtraction) of a constant to the vector $\boldsymbol{S}$. For any constant c, $\Pr(S_1, \ldots, S_N)$ is exactly the same as $\Pr(S_1 + c, \ldots, S_N + c)$. Put simply, under the ICAR model, the mean of $\boldsymbol{S}$ is not defined. This is why the ICAR model is only

used as a prior model for a set of area-specific parameters but not used as a likelihood function to directly model the data. It is unrealistic to assume that the data we observe would have come from a process in which the overall mean is undefined. Besag et al. (1991, p.8) interpret Eq. 8.6 as a stochastic version of linear interpolation. Eq. 8.6, and hence the ICAR model belongs to "the so-called intrinsic models of geostatistics (Matheron, 1973) where all possible spatial differences are modelled simultaneously" (Künsch, 1987, p.517).

To solve the problem of an undefined overall mean, the sum-to-zero constraint fixes the mean of $\boldsymbol{S}$ to 0. Due to this constraint, a separate intercept is then required in the regression to represent the overall mean (of the outcome). The area-specific parameters are effectively measuring the deviations of the local estimates from the global mean. For example, in the income example, the parameters in $\boldsymbol{S}$ measure the deviations of the MSOA-level income from the Newcastle average (see the model in Eq. 8.2). The prior for the intercept must be the improper uniform distribution defined on the whole real line so that the constrained parameters with a separate intercept are equivalent to the unconstrained parameters (see Appendix 1 – under the Intrinsic CAR model section – of the GeoBUGS manual).

8.2.2.3 Applying the ICAR Model Using General Weights to the Newcastle Income Data

We now outline a Bayesian hierarchical model with the ICAR prior for the income example. We derive a general spatial weights matrix based on some demographic data. We explain the underlying assumption and the interpretation of results but leave the model fitting to Exercise 8.3.

Eq. 8.7 defines a local smoothing model similar to Eq. 8.2. In both models, the ICAR model with rook's contiguity is used to carry out local smoothing. However, in Eq. 8.7, the non-zero entries in the spatial weights matrix are derived based on h_i, an MSOA-level variable from the 2011 UK census measuring the number of residents aged 16 to 74 working full time for at least 31 hours per week (hereafter, for simplicity, we call this variable "the number of full-time working residents"). We use the notation $\boldsymbol{W}_{hour}$ to emphasise this specification. The use of $\boldsymbol{W}_{hour}$ in the ICAR model assumes that the income levels of two areas are more alike if (a) they share a common border *and* (b) they have similar numbers of full-time working residents. The latter seems reasonable since, as shown in Figure 8.8, an MSOA with a higher h_i tends to have a higher average income, although there are three MSOAs that appear to have high average incomes with relatively small full-time working populations. The global Moran's I based on the number of full-time working residents is 0.33, with a p-value of 0.001 derived from 999 random permutations, suggesting strong positive spatial autocorrelation.

$$
\begin{aligned}
y_{ij} &\sim N(\theta_i, \sigma_y^2) \\
\theta_i &= \alpha + S_i \\
\boldsymbol{S}_{1:109} &\sim ICAR\left(\boldsymbol{W}_{hour}, \sigma_S^2\right) \\
\alpha &\sim Uniform\left(-\infty, +\infty\right) \\
\sigma_y &\sim Uniform\left(0.0001, 1000\right) \\
\sigma_S &\sim Uniform\left(0.0001, 1000\right)
\end{aligned}
\tag{8.7}
$$

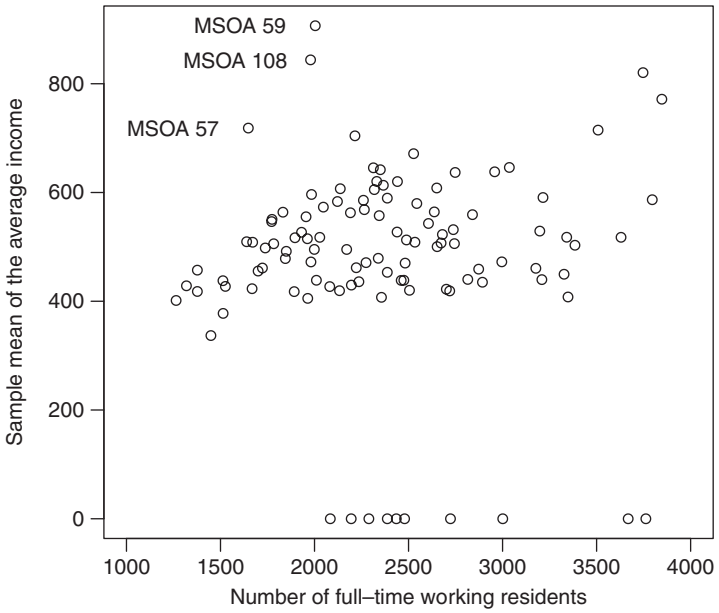

FIGURE 8.8
A scatter plot between the numbers of residents who worked full time with at least 31 hours and the sample means calculated from the survey data. The 10 data points at the bottom (with sample means equal to 0) are those MSOAs without income survey data. Also highlighted are the three MSOAs that appear to deviate from the overall positive relationship.

The diagonal elements in $\boldsymbol{W}_{hour}$ are 0, and the off-diagonal elements are set to be $w_{ij} = \dfrac{1}{|h_i - h_j|}$ if MSOAs i and j share a common border and $w_{ij} = 0$ otherwise.

8.2.2.4 Results

Figure 8.9 shows that for most of the MSOAs, the posterior means from the two specifications of the ICAR model are different but reasonably close, scattering around the diagonal line. The difference between the two sets of estimates is due to the different weights being assigned to the contiguous neighbours. Most of the estimates are similar between the two specifications because of (a) the strong positive spatial autocorrelation present in the sizes of the working populations and (b) the strong relationship between the working population characteristic and income. Point (a) implies that the weights of the neighbours tend to be similar while, even when large weights are assigned to some neighbours, point (b) suggests the resulting (weighted) mean of the neighbours tends not to be too different from the case where the weights of the neighbours are fixed at 1. If there is no apparent correlation between income and the size of the working population, the two ways of specifying weights can give quite different local conditional means.

There are, however, some exceptions. The two points above the diagonal line, MSOAs 28 and 33, in Figure 8.9 are two MSOAs with no income survey data. Compared to the estimates with weights fixed at 1, the posterior means of both MSOAs are higher when $\boldsymbol{W}_{hour}$ is used. This is because, according to $\boldsymbol{W}_{hour}$, these two MSOAs are more similar to some of their respective neighbours that have high average income. For example, looking at the row standardised $\boldsymbol{W}_{hour}$, two out of eight of the neighbours of MSOA 28 have weights of

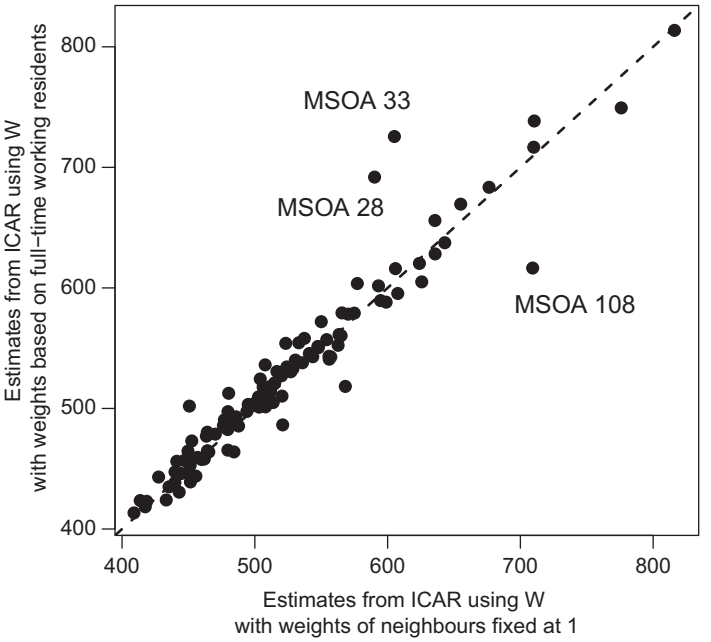

FIGURE 8.9
A scatter plot comparing the posterior means from the ICAR model with contiguity but where the weights between neighbours are either fixed at 1 (x-axis) or computed using data on the size of the working population (y-axis). The dashed line is the diagonal line.

0.31 and 0.26, and the posterior means of the average income of these two neighbours are quite large: 747 and 721. For MSOA 33, one neighbour (out of seven) has a weight of 0.79, and the posterior mean of this neighbour is 747. The non-equal weights in $\boldsymbol{W}_{hour}$ push the estimates of these MSOAs higher.

MSOA 108 is an MSOA that is located below and quite far away from the diagonal line in Figure 8.9. With a sample size of six, this MSOA has a particularly high sample mean of 844. It has five contiguous neighbours with relatively low incomes, and four of them have sample means less than 530. When using the fixed weight of 1 for the neighbours, the estimate of MSOA 108 is unsurprisingly shrunk towards the neighbours. When adding in the similarity measure based on the working population, the estimate of this MSOA is further smoothed towards the neighbours because the numbers of full-time working residents amongst MSOA 108 and its neighbours are similar and low.

8.3 The Proper CAR (pCAR) Model

As discussed in Section 8.2.2.2, the joint distribution of the ICAR model is an improper distribution because the overall mean is not defined. The ICAR model can be made proper by introducing a multiplicative parameter ρ to the spatial weights. This gives rise to the proper CAR model, which is defined via a set of N full conditionals as in Eq. 8.8 or, equivalently, via the joint multivariate distribution given in Eq. 8.9.

$$S_i \mid \boldsymbol{S}_{\{-i\}}, v, \boldsymbol{W} \sim N\left(\rho \frac{\sum_{j=1}^{N} w_{ij} S_j}{w_{i+}}, \frac{v}{w_{i+}}\right) \tag{8.8}$$

$$\Pr(\boldsymbol{S}) \propto \exp\left\{-\frac{1}{2v}\boldsymbol{S}'(\boldsymbol{D}_w - \rho \boldsymbol{W})\boldsymbol{S}\right\} \tag{8.9}$$

The notations in the above two distributions are the same as those in Eq. 8.3 and Eq. 8.4, the full conditionals and the joint distribution respectively under the ICAR model. Like the ICAR model, the spatial weights matrix $\boldsymbol{W}$ needs to be symmetric (see Section 8.2.2.1). Furthermore, for the joint distribution (Eq. 8.9) to be proper (i.e. integrating to 1), the precision matrix $\boldsymbol{Q} = (\boldsymbol{D}_w - \rho \boldsymbol{W})$ needs to be invertible so that the covariance matrix, $\boldsymbol{\Sigma} = \boldsymbol{Q}^{-1}$, exists. This requirement places a constraint on the multiplicative parameter ρ, under which ρ can only lie in the open interval $(1/\lambda_{min}, 1/\lambda_{max})$, where λ_{min} and λ_{max} respectively are the smallest and the largest eigenvalues of the row-standardised spatial weights matrix $\boldsymbol{W}^*$ (see Section 3.3.1 in Banerjee et al., 2004; Haining, 2003, p.300 and Appendix 1 – under the Conditional specification section – of the GeoBUGS manual).[6] Given ρ within the permissible range, the pCAR model specifies a multivariate normal distribution for $\boldsymbol{S}$, namely,

$$\boldsymbol{S} \sim MVN\left(\boldsymbol{0}, v(\boldsymbol{D}_w - \rho \boldsymbol{W})^{-1}\right) \tag{8.10}$$

The parameter ρ measures the strength of spatial autocorrelation. A value of ρ that is close to 0 signals very weak or no spatial autocorrelation. $\rho < 0$ implies negative spatial autocorrelation (i.e. things nearby in space tend to be dissimilar), while $\rho > 0$ implies positive spatial autocorrelation (i.e. things nearby in space tend to be similar).[7] Moreover, the largest eigenvalue of the row-standardised weights matrix is 1, while all other eigenvalues are between –1 and 1 (Exercise 8.4). Therefore, the bounds for the open interval within which ρ lies are typically simplified to –1 and 1.

The WinBUGS function to implement the proper CAR model is `car.proper` and the syntax is

```
S[1:N] ~ car.proper(a[],C[],adj[], num[], M[], tau, rho)
```

From the GeoBUGS manual, the definitions of the arguments of the `car.proper` function are given as follows:

---

[6] To derive the constraint on ρ, rewrite the precision matrix $\boldsymbol{Q}$ as $\boldsymbol{D}_w(\boldsymbol{I} - \rho \boldsymbol{W}^*)$, where $\boldsymbol{W}^*$ denotes the row-standardised weights matrix and $\boldsymbol{I}$ is the identity matrix of size $N \times N$. A valid precision matrix is positive-definite, so its determinant, $\det(\boldsymbol{Q})$, is positive. So, $\det(\boldsymbol{Q}) = \det(\boldsymbol{D}_w(\boldsymbol{I} - \rho \boldsymbol{W}^*)) = \det(\boldsymbol{D}_w)\det(\boldsymbol{I} - \rho \boldsymbol{W}^*)$. Given that the row sums of $\boldsymbol{W}$ are all positive, $\det(\boldsymbol{D}_W) > 0$. Thus we require $\det(\boldsymbol{I} - \rho \boldsymbol{W}^*) > 0$. From matrix algebra, $\det(\boldsymbol{I} - \rho \boldsymbol{W}^*) = \prod_{k=1}^{N}(1 - \rho\lambda_k)$, where $\lambda_1, \ldots, \lambda_N$ are the eigenvalues of $\boldsymbol{W}^*$ (Abadir and Magnus, 2005, Exercise 7.26). It is sufficient for $\det(\boldsymbol{I} - \rho \boldsymbol{W}^*) > 0$ if $\rho \in (1/\lambda_{min}, 1/\lambda_{max})$ where λ_{min} and λ_{max} are the smallest and largest eigenvalues of $\boldsymbol{W}^*$. Ord (1975, Appendix C, p.125) shows that $\boldsymbol{W}^*$ and $\boldsymbol{D}_w^{-1/2}\boldsymbol{W}\boldsymbol{D}_w^{-1/2}$ have identical eigenvalues – the latter is used in WinBUGS to derive the lower and the upper bounds for ρ. Haining (2003) uses the unstandardised spatial weights matrix, the largest eigenvalue of which is generally not equal to 1, but $\lambda_{min} < 0$ and $\lambda_{max} > 0$ are still true.

[7] It is worth noting that $\lambda_{min} < 0$ and $\lambda_{max} > 0$ with λ_{min} and λ_{max}, the smallest and the largest eigenvalues of $\boldsymbol{D}_w^{-1/2}\boldsymbol{W}\boldsymbol{D}_w^{-1/2}$ or, equivalently, of $\boldsymbol{W}^*$. This is because $tr(\boldsymbol{W}^*) = 0$, i.e. the trace of $\boldsymbol{W}^*$ (the sum of the diagonal elements) is equal to 0 and, from Abadir and Magnus (2005, Exercise 7.27), $tr(\boldsymbol{W}^*) = \sum_{k=1}^{N}\lambda_k$. Thus, $\sum_{k=1}^{N}\lambda_k = 0$, implying that some of the eigenvalues are negative and some are positive, so $\lambda_{min} < 0$ and $\lambda_{max} > 0$ (see also Section 3.3.1 in Banerjee et al., 2004).

- `a[]`: A vector of length N giving the mean for each area – each element in this vector can be (a) specified as a fixed value (thus, the mean is not estimated but a fixed constant), (b) assigned with a prior distribution then each mean is estimated or (c) specified deterministically within the model code, e.g. as a function of some covariates (see Section 8.3.2 below).
- `C[]`: A vector the same length as `adj[]` (defined below), storing the non-zero weights from the row-standardised weights matrix $\boldsymbol{W}^*$.
- `adj[]`: A vector storing the IDs of the neighbouring areas (the same as in the `car.normal` function).
- `num[]`: A vector of length N storing the number of neighbours for each area (the same as in the `car.normal` function).
- `M[]`: A vector of length N specifying the inverse of the elements in the diagonal matrix $\boldsymbol{D}_w$ (i.e. the ith element in this vector is equal to $\frac{1}{w_{i+}}$, as defined in Eq. 8.4).
- `tau`: The unknown precision parameter controlling the overall local smoothing.
- `rho`: The unknown spatial autocorrelation parameter representing the overall degree of spatial dependence. See below for its prior specification.

8.3.1 Prior Choice for ρ

To ensure the constraint, a typical prior for ρ is a uniform distribution defined on the open interval $\left(1/\lambda_{min}, 1/\lambda_{max}\right)$, where λ_{min} and λ_{max} are the smallest and the largest eigenvalues of the row standardised weights matrix $\boldsymbol{W}^*$ or, equivalently, of the matrix $\mathbf{D}_w^{-1/2}\boldsymbol{W}\mathbf{D}_w^{-1/2}$. The latter is the matrix used in WinBUGS to compute the two bounds via the functions `min.bound` (for $1/\lambda_{min}$) and `max.bound` (for $1/\lambda_{max}$). The syntax of the two functions is given as follows:

```
lower.bound <- min.bound(C[], adj[], num[], M[])
upper.bound <- max.bound(C[], adj[], num[], M[])
```

where the four arguments for both functions are as defined in the `car.proper` function. Then the uniform prior is

```
rho ~ dunif(lower.bound, upper.bound)
```

8.3.2 ICAR or pCAR?

Compared to the ICAR model, the pCAR model has three advantages. First, the joint distribution is a proper probability distribution, so the area-specific parameters under the pCAR model are not constrained to sum to zero. This allows the flexibility to model the mean of each S_i via the vector `a[]` in the `car.proper` function. Second, while the ICAR model assumes positive spatial autocorrelation, the pCAR model can accommodate both positive (i.e. when $\rho > 0$) and negative (i.e. when $\rho < 0$) spatial autocorrelation. Third, the pCAR model encompasses, to some extent, the global smoothing model with exchangeability introduced in Section 7.4. When ρ is estimated to be close to 0, the joint distribution in Eq. 8.10 effectively reduces to a multivariate normal distribution with a diagonal covariance matrix, i.e. $\boldsymbol{S} \sim MVN\left(\mathbf{0}, v\boldsymbol{D}_w^{-1}\right)$, which can be rewritten as $S_i \sim N\left(0, v \cdot w_{i+}^{-1}\right)$ for all i. Thus, conditioning on the variance v, the mean 0 and w_{i+}^{-1}, all areas are considered to be

independent of each other. However, there is still a degree of "local smoothing" left in the pCAR model because each individual variance, $v \cdot w_{i+}^{-1}$, still depends on the spatial neighbourhood structure via the sum of the ith row in $\boldsymbol{W}$.

There are, however, a few limitations with the pCAR model. First is the interpretation of the spatial autocorrelation parameter ρ. Despite its permissible range lying between –1 and 1, ρ is not calibrated to have the same interpretation as a correlation coefficient. To illustrate, we follow the simulation outlined in Section 3.3.1 in Banerjee et al. (2004). For a range of values for ρ, we simulate 100 sets of values using the joint distribution in Eq. 8.10, where $\boldsymbol{W}$ defines the neighbourhood structure of the MSOAs in Newcastle based on rook's contiguity (Exercise 8.5). For each set, we compute the global Moran's I to measure the strength of spatial autocorrelation. Figure 8.10 shows the results of the simulation. The strength of spatial autocorrelation, as measured by Moran's I, increases as ρ increases, and a negative (or positive) ρ tends to result in a negative (or positive) value for Moran's I. Moran's I tends to be close to 0 when ρ is close to 0. However, even when ρ is 0.9, Moran's I merely ranges between 0.12 and 0.56 (see the right-hand plot in Figure 8.10). Only when $\rho = 0.9999$ does the average of Moran's I over the 100 sets of simulated data lie above 0.9. These simulations suggest that the interpretation of ρ remains qualitative rather than quantitative, at best suggesting how likely it is that positive spatial autocorrelation is present in the data rather than providing a measure of the strength of spatial autocorrelation. In a Bayesian analysis, this is equivalent to calculating the posterior probability $\Pr(\rho > 0 \mid \text{data})$. We will illustrate the interpretation of ρ in the income example.

The second limitation is the interpretation of the conditional mean in Eq. 8.8. Under the ICAR model, given the neighbours, S_i is similar to the average of the neighbours (Eq. 8.1 and Eq. 8.3), whereas in the pCAR model, S_i is similar to some proportion of the local average (Eq. 8.8), which, as pointed out by Banerjee et al. (2004, p.81), may not have any sensible interpretation.

A third limitation of the pCAR model is the need to estimate the parameter ρ, which, particularly in complex models, may introduce computational issues (e.g. slower convergence; maybe even non-convergence). We will return to this in Section 8.5.1 when the BYM model is discussed.

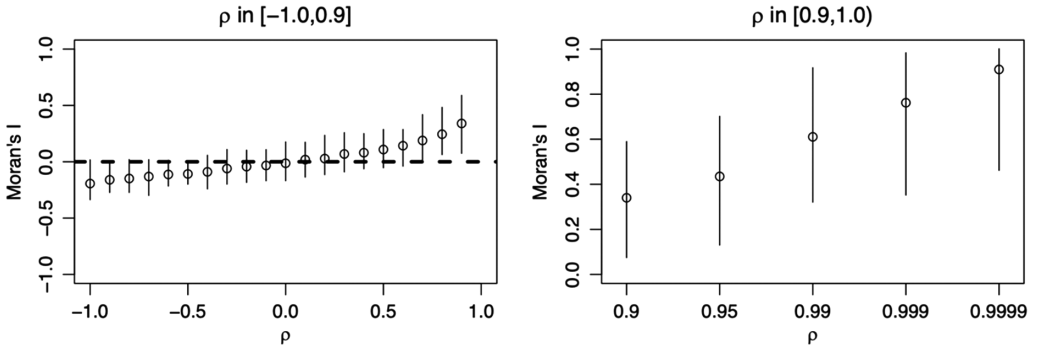

FIGURE 8.10
The strength of spatial autocorrelation of the values simulated from the proper CAR model using different values of ρ ranging from –1 to 1. With each ρ value, 100 sets of values are simulated based on the Newcastle MSOA map and the global Moran's I statistic is calculated for each set to measure the strength of spatial autocorrelation. Each open circle represents the mean value of Moran's I, and the vertical bar shows the 95% sampling variability (i.e. the values of the 2.5 and 97.5 percentiles). The left-hand panel shows $\rho \in [-1.0, 0.9]$, and the right-hand panel shows $\rho \in [0.9, 1.0)$.

8.3.3 Applying the pCAR Model to the Newcastle Income Data

We apply the pCAR model where the spatial neighbourhood structure is defined via rook's spatial contiguity and the weights of the neighbours are based on similarity in terms of the size of the full-time working population. That is, the weight of two neighbours i and j is set to be $w_{ij} = 1/|h_i - h_j|$, with h_i and h_j being the numbers of full-time working residents in the two areas. The model is given as follows:

$$\begin{aligned}
y_{ij} &\sim N\left(\theta_i, \sigma_y^2\right) \\
\theta_i &= \alpha + S_i \\
\boldsymbol{S}_{1:109} &\sim pCAR\left(\boldsymbol{W}_{hour}, \sigma_S^2, \rho\right) \\
\alpha &\sim N\left(0, 100000000\right) \\
\sigma_y &\sim Uniform\left(0.0001, 1000\right) \\
\sigma_S &\sim Uniform\left(0.0001, 1000\right) \\
\rho &\sim Uniform\left(a, b\right)
\end{aligned} \tag{8.11}$$

The lower and the upper bounds (denoted as a and b) of the uniform prior for ρ are computed using the built-in functions, as explained in Section 8.3.1. The WinBUGS implementation of this model is in Figure 8.11.

Lines 49 to 52 in Figure 8.11 illustrate the structure of the array `C[]`, which contains the row standardised weights of the neighbours, and Line 53 shows the first few elements in the array `M[]`. To understand how the two arrays are constructed, take the example of the first MSOA. MSOA 2 is the only neighbour of MSOA 1, and their numbers of full-time working residents are $h_1 = 2650$ and $h_2 = 3385$, respectively, which gives the weight $w_{12} = 1/|2650 - 3385| = 0.00136$, and the corresponding row sum is $w_{1+} = 0.00136$. Therefore, the row-standardised weight becomes 1 (the first element in the `C[]` array) and the first element in `M[]` is $\frac{1}{w_{1+}} = \frac{1}{0.0014} = 735$. The same procedure applies to the calculation for the remaining MSOAs. Readers are encouraged to verify the values in the arrays `C[]` and `M[]` for MSOA 2 (3385), which has six neighbours – MSOAs 1 (2650), 3 (2289), 4 (2606), 11 (1723), 23 (2133) and 24 (2196); the numbers in brackets are their numbers of full-time working residents.

8.3.4 Results

Figure 8.12 compares the posterior means (Panel a) and the lengths of the 95% credible intervals (Panel b) of the MSOA-level average income from the pCAR model and from the ICAR model; the latter model was discussed in Section 8.2.2.3. The two models yield very similar posterior point and interval estimates. This is perhaps not surprising given that ρ is estimated to be close to 1. The posterior mean of ρ is 0.95 with a 95% credible interval of (0.85–1.00) and its posterior distribution in given in Figure 8.13.

At first sight it might seem that fitting the pCAR model where the spatial parameter is estimated should offer more flexibility in comparison to the ICAR model. However, as

```
#  The WinBUGS code for specifying the model with the pCAR prior
model {
  #  a for-loop to go through all the household level income data
  for (j in 1:nhhs) {
    #  defining the normal likelihood for each household
    y[j] ~ dnorm(theta[msoa[j]],prec.y)
  }
  for (i in 1:nmsoas) {
    theta[i] <- alpha + S[i]
    a[i] <- 0  #  set the mean of each MSOA to 0 in the pCAR model
  }
  #  modelling all the elements in S using the pCAR model
  S[1:nmsoas] ~ car.proper(a[],C[],adj[], num[], M[], prec.S, rho)

  #  a vague prior for the Newcastle average
  alpha ~ dnorm(0,0.00000001)

  #  a vague prior for the conditional SD in the pCAR model
  sigma.S ~ dunif(0.0001,1000)

  #  a vague prior for the sampling standard deviation
  sigma.y ~ dunif(0.0001,1000)

  #  calculate the two precisions
  prec.S <- pow(sigma.S,-2)
  prec.y <- pow(sigma.y,-2)

  #  a uniform prior on rho and the calculation of the lower and
  #  upper bounds
  rho ~ dunif(lower.bound, upper.bound)
  lower.bound <- min.bound(C[], adj[], num[], M[])
  upper.bound <- max.bound(C[], adj[], num[], M[])
}

#  household-level income data from survey with a spatial
#  neighbourhood structure
list(nhhs=760
    ,nmsoas=109
    ,y=c(501,616,472,816,637,500,506,560,542,447,644,522
    ,487,275,...)
    ,msoa=c(1,1,1,1,1,2,2,4,4,4,4,4,5,5,...)
#  elements of the W matrix for pCAR (defined via rook's spatial
#  contiguity)
    ,num=c(1,6,4,...)
    ,adj=c(2
          ,1,3,4,11,23,24
          ,2,4,20,22
          ,...)
    ,C=c(1
        ,0.235, 0.157, 0.221, 0.104, 0.138, 0.145
        ,0.023, 0.078, 0.852, 0.048
        ,...)
    ,M=c(735.00,172.47,24.70,...)
)

#  initial values for chain 1
list(alpha=200,sigma.y=50,sigma.S=20,S=c(1,-1,0,0,0,...),rho=0.1)

#  initial values for chain 2
list(alpha=700,sigma.y=80,sigma.S=50,S=c(-1,1,0,0,0,...),rho=0.05)
```

FIGURE 8.11
The WinBUGS code for fitting the pCAR model (Eq. 8.11) to the Newcastle income data.

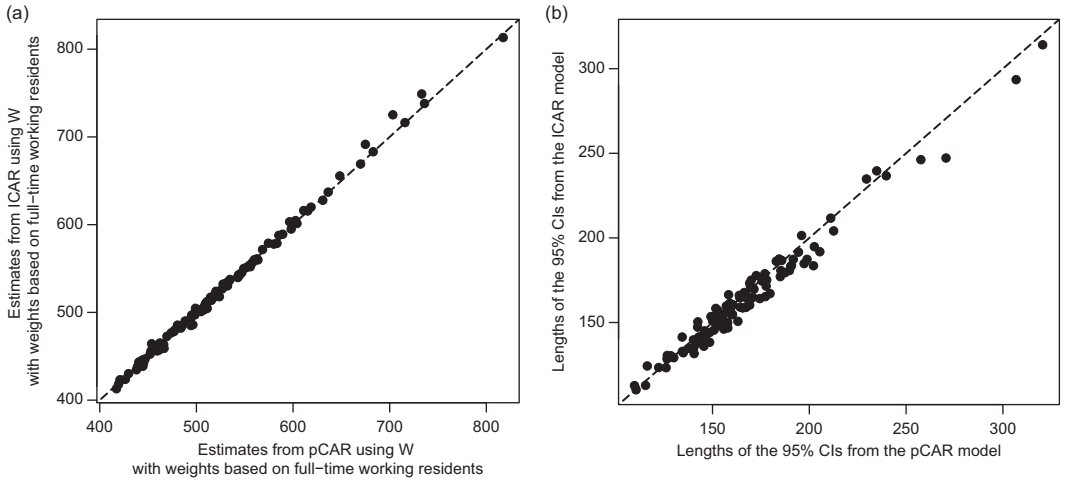

FIGURE 8.12
(a) Comparing the posterior means of the MSOA average income from the ICAR model and from the pCAR model. For both models, the spatial neighbourhood structure is defined via rook's contiguity and the weights of the neighbours are derived based on the similarity in the size of the full-time working population. Panel (b) compares the widths of the 95% credible intervals (the upper bound – the lower bound) from the two models. In both plots, the dashed line indicates the diagonal line.

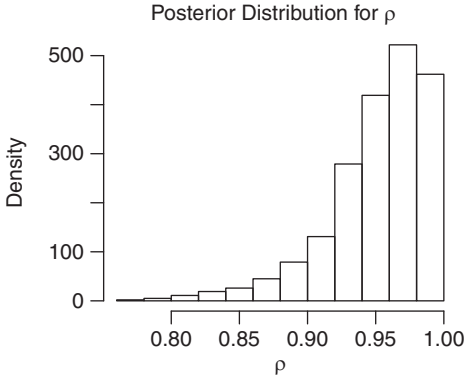

FIGURE 8.13
The posterior distribution of the parameter ρ from the pCAR model.

remarked in Gelfand and Vounatsou (2003, p.15), setting "$\rho = 1$ is analogous to the non-stationary or random walk case in familiar autoregressive time series models and can be advantageous in accommodating *more irregular spatial behaviour*". Making a similar point, Banerjee et al. (2004, p.81) also remark that in the case of the pCAR model, "the breadth of spatial pattern may be too limited. In the case where a CAR model is applied to random effects, an improper choice may actually enable wider scope for posterior spatial pattern".

When positive spatial autocorrelation is found to be present, ρ is often estimated to be very close to 1 (as in, for example, the Newcastle income data). In many practical applications, this is another reason for adopting the ICAR model (with ρ fixed to 1) for imposing spatial structure on a set of area-specific parameters (Mollié, 1996).

However, if a localised pattern of spatial autocorrelation, a form of spatial heterogeneity, is suspected or its detection is of interest, one option is to employ locally adaptive models, a topic that we will discuss next.

8.4 Locally Adaptive Models

In some cases, the spatial data that we observe may exhibit more complex spatial dependence structures or even discontinuities. The evidence for the former may come from the use of local statistics such as the local Moran's I (Chapter 6). Such complexity is often seen in urban data when mapped at the scale of small spatial units such as census tracts. For example, some sections of an urban area may display quite a smooth-looking pattern of average household income values, while abrupt changes (or spatial discontinuities) may exist in other parts. In some cities, high income, affluent neighbourhoods are located geographically close, even adjacent, to poorer neighbourhoods. Such abrupt changes may be evident by simply inspecting a map of the data. Wombling is a statistical method for determining the boundaries where spatial discontinuity or zones of rapid change occur (Womble, 1951; see also Section 4.10). In Section 8.4.2, we consider a method of wombling and we also refer readers to Gelfand and Banerjee (2015) and the references therein. For these types of situations, the CAR models that we have discussed so far may not be appropriate because their strategy for borrowing information from the neighbours of an area is applied *uniformly* across the whole study region.

To demonstrate the issue of spatial discontinuity, consider again the Newcastle income data. Figure 8.14 shows an overall spatially-smooth map, where high-income (or low-income) MSOAs tend to be clustered together. There are, however, some sub-regions in Newcastle where spatial discontinuity is evident (see the two sets of cross-hatched areas). For example, for the pocket of MSOAs in the middle of Newcastle, there appears to be two clusters, where MSOAs 31, 32, 39 and 43 tend to have high-income levels, while the income levels of MSOAs 48, 50 and 55 can be seen to be low. While it is reasonable to share information amongst the MSOAs within each cluster, is it reasonable for MSOAs 32 and 50 to borrow information from each other or for 43 and 50 to do likewise? The same point applies to the subset of MSOAs in the northeast corner.

To address this problem, a number of locally adaptive spatial models have been proposed. The rationale behind these models is to recognise where it is appropriate to borrow information from one's neighbours and where it is not. Because the spatial neighbourhood structure is encoded in the $\boldsymbol{W}$ matrix, for most of these models adaptive spatial smoothing is achieved by modelling the non-zero elements in $\boldsymbol{W}$. The basic modelling idea is to start with the $\boldsymbol{W}$ matrix defined by spatial contiguity with the weights of neighbours set to 1. However, instead of treating $\boldsymbol{W}$ as fixed, each $w_{ij}=1$ in the matrix is potentially subject to modification so that two spatially-contiguous areas can be considered as neighbours (by confirming the weight as 1) or not neighbours (by re-setting the weight to 0). To operationalise this idea, there are three main approaches: (1) choosing an optimal $\boldsymbol{W}$ matrix from all possible specifications; (2) modelling the elements in the $\boldsymbol{W}$ matrix; (3) grouping areas via mixture models. We will explore the first two approaches in this section and illustrate the mixture modelling approach through an application in Chapter 9.

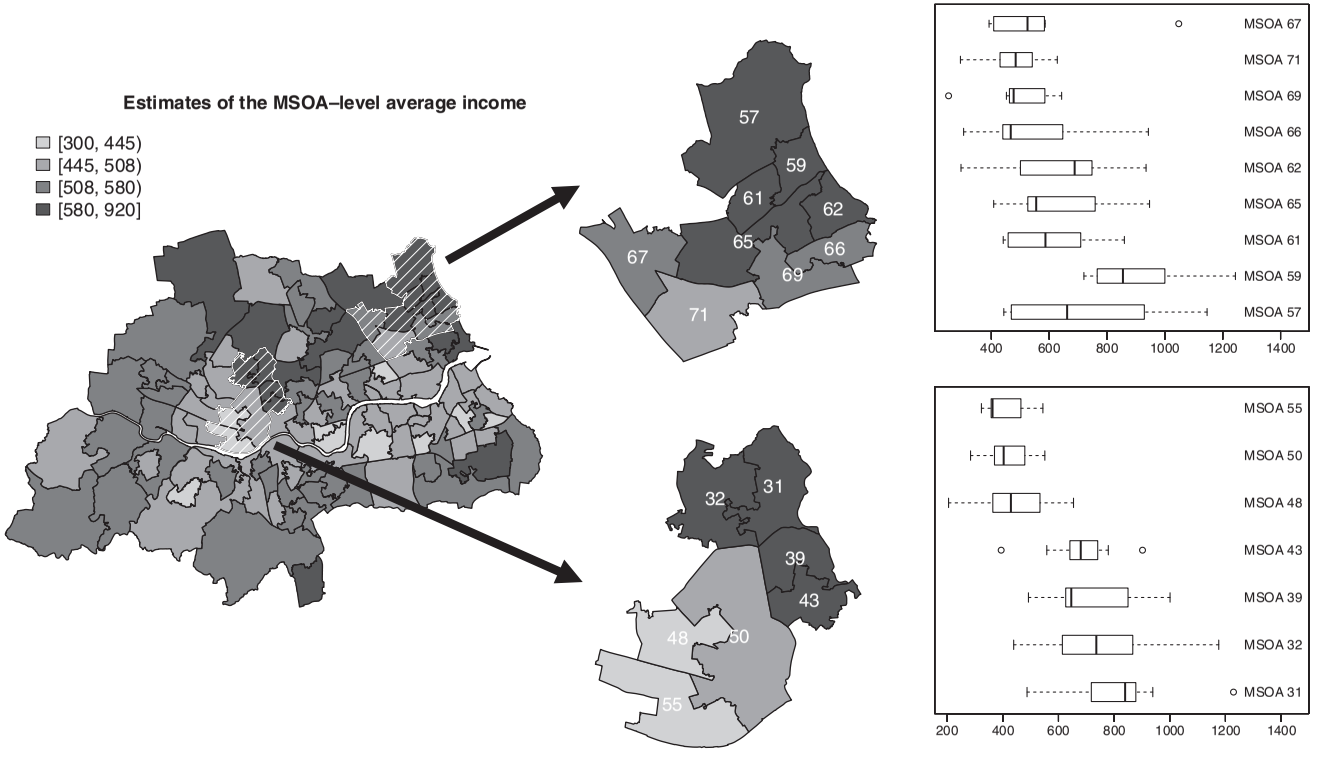

FIGURE 8.14
Examples of spatial discontinuity in the Newcastle income data at the MSOA scale. The left-hand map shows the posterior means of the MSOA-level income from the ICAR model in Eq. 8.2. The two cross-hatched sub regions are examples where spatial discontinuity is evident. The two boxplots on the right-hand side show the survey data for each MSOA. The labelling of the MSOAs is the same as that used in Figure 7.1 in Chapter 7.

8.4.1 Choosing an Optimal *W* Matrix from All Possible Specifications

The first approach is to consider all possible modifications to the $\boldsymbol{W}$ matrix derived from spatial contiguity. Different versions of the $\boldsymbol{W}$ matrix give rise to different models. Model choice is then based on, for example, the Bayesian information criterion (Li et al., 2011) or an automatic model selection algorithm using the reversible jump MCMC coupled with simulated annealing (Seya et al., 2013). However, an immediate challenge to this approach is the large number of $\boldsymbol{W}$ matrices, each requiring the fitting of a spatial model to the data. Since the weight of each pair of spatially-contiguous areas can take either 1 or 0, the number of possible $\boldsymbol{W}$ matrices and hence models is $2^{\mathbf{1W1}'/2}$, where $\boldsymbol{W}$ is derived from spatial contiguity with binary 0/1 weights, $\mathbf{1}$ denotes a vector of 1s of size $1 \times N$ and $\mathbf{1}'$ is its transpose. For the MSOA Newcastle map, $\mathbf{1W1}'/2$ gives 263 unique pairs of spatially-contiguous MSOAs, and hence the possible modifications to $\boldsymbol{W}$ is $2^{263} = 1.48 \times 10^{79}$(!).

To circumvent this challenge, Anderson et al. (2014, 2016) proposed to reduce the set of $\boldsymbol{W}$ matrices by eliciting possible configurations of $\boldsymbol{W}$ using data observed from q time periods, say five years, immediately before the current study period. Specifically, a hierarchical agglomerative clustering algorithm is used to group areas together based on spatial contiguity and how similar they are in terms of the outcome values over the past q time periods. Over N areas, because of the hierarchical agglomerative nature of the clustering method, N cluster configurations are formed by, effectively, scanning from the bottom of the resulting dendrogram (which gives rise to a cluster configuration where each area is assigned to its own singleton cluster) to the top (which forms another cluster configuration where all areas are grouped into a single cluster). When modelling the data at the current period, the $\boldsymbol{W}$ matrices to consider are reduced vastly to the N cluster configurations derived from the past data. While reducing the computational burden, this method relies on the availability of past data. Deriving the potential $\boldsymbol{W}$ matrices from past data for the current time period also assumes that the clustering of areas is temporally stable. This assumption is perhaps justifiable for modelling chronic diseases whose risk factors are stable over time "but would be unsuitable for epidemic diseases such as influenza, where the spatial pattern in disease risk in the years prior to an outbreak would be vastly different to the pattern during an outbreak" (Anderson et al., 2016, p.12). This method is implemented in R; see Anderson et al. (2016) for more detail.

8.4.2 Modelling the Elements in the *W* Matrix

The second approach is to model directly the elements of the spatially-contiguous pairs in $\boldsymbol{W}$. Recall that using rook's contiguity, if areas i and j share a common border, then $w_{ij} = 1$, they are considered to be neighbours and information is shared between them. However, if the two areas are shown to be sufficiently different based on some criteria, for example, one is a high average income area while the other is a low average income area, then there is little evidence to support the sharing of information between them. In this case, we might wish to set w_{ij} to 0, despite the fact that they are spatially contiguous. In other words, we are introducing a "hard boundary" or spatial discontinuity between the two areas. The methods discussed below aim to construct such "hard boundaries" and are referred to as boundary detection methods.[8]

Based on the above idea, Lu et al. (2007) model the weight of a pair of spatially-contiguous areas using a logistic regression with a single covariate, while the weight of two areas

[8] The introduction of impenetrable boundaries into models of geographical diffusion processes has a history (see for example Hagerstand, 1967).

that are not spatially-contiguous remains as 0. Under their model, w_{ij}, the weight of two *spatially-contiguous* areas, i and j, is modelled as

$$w_{ij} \sim \text{Bernoulli}\left(p_{ij}\right)$$

$$logit\left(p_{ij}\right) = \gamma_0 + \gamma_1 d_{ij}, \tag{8.12}$$

where γ_0 and γ_1 are regression coefficients to be estimated using the data and d_{ij} is a similarity measure defined as the difference between the two areas with respect to the chosen covariate. A small value of d_{ij} means that the two areas have similar values of the chosen covariate (e.g. average income). When Lu et al. (2007) applied this method to their Minnesota breast cancer late detection data (where the outcome values are the numbers of late detection cases across the 87 counties), several specifications of d_{ij} were considered. Using their labelling of the models (see Section 4 in Lu et al., 2007), these specifications are:

- Model 1: d_{ij} is the distance between the centroids of areas i and j scaled (divided) by the maximum distance on the map.
- Model 2: d_{ij} is the absolute difference in cancer mortality rate between the two areas, i.e. $d_{ij} = \left|\text{mortality}_i - \text{mortality}_j\right|$. The county-level mortality rates were available to the study as a set of covariate values.

Similar to Model 2, Models 3 and 4 were two additional specifications where the mortality rates in Model 2 were replaced by mammogram, a county-level covariate derived from a separate survey measuring the percent of patients aged 40 and over who reported having had a mammogram within the last two years, or by incidence, the county-level cancer incidence rate.[9]

In addition to d_{ij}, another important aspect of this modelling is the choice of priors for γ_0 and γ_1 in Eq. 8.12. In Lu et al. (2007), two moderately informative priors were chosen: $\gamma_0 \sim N\left(2, 0.5\right)$ and $\gamma_1 \sim N(-2, 0.5)$. Why are these two priors chosen? And given the priors on γ_0 and γ_1 and given a permissible value of d_{ij}, what is the equivalent prior on p_{ij} which determines the probability that two spatially-contiguous areas will be treated as neighbours ahead of seeing the data? To answer these two questions, we need to understand the roles of the two coefficients. First, γ_0 is the logit transform of p_{ij} when $d_{ij} = 0$ so it determines the probability that i and j are neighbours ($w_{ij} = 1$; hereafter referred to as the "neighbouring probability") when the two areas have the same covariate values (e.g. having the same mortality rates). Before seeing the data, when $d_{ij} = 0$, this neighbouring probability is expected to be high, which is reflected in the chosen prior for γ_0: the prior neighbouring probability has a mean of 0.86 and there is a 95% chance that this prior neighbouring

[9] In Lu et al. (2007), spatial contiguity was also considered for defining d_{ij}, where $d_{ij} = 0$ if areas i and j share a common boundary and $d_{ij} = 1$ otherwise. However, Lu et al. (2007, p. 445–446) considered the specification (their Model 0) where $d_{ij} = 1$ if i and j share a common boundary and 0 otherwise. When combined with their prior for $\gamma_1 \sim N(-2, 0.5)$, this latter specification of d_{ij} suggests that two spatially-contiguous areas are *less likely* to be considered as neighbours *a priori*. Such a (prior) assumption is not reasonable. See Figure 8.15 and the discussion of how the priors on γ_0 and γ_1 are translated to the prior on w_{ij}. Defining d_{ij} based on contiguity introduces some "general randomness or uncertainty" in deciding whether two contiguous areas will be allowed to remain as neighbours, albeit a strong prior belief (through the choice of prior by the authors) towards the two areas staying as neighbours. Such randomness/uncertainty may be difficult to quantify in the absence of covariate(s). This option would not be recommended as a candidate model in practice.

probability varies between 0.65 and 0.97 (see Figure 8.15 for the calculation where the results from Lines 18 and 20 give the values above). However, γ_1 quantifies the effect of d_{ij} on the neighbouring probability. Now, the smaller d_{ij} is, the more alike two areas are in terms of their attributes, in the case of Models 2–4, and thus the more likely it is that the two areas will be treated as neighbours, *a priori*. In the case of Model 1, the closer d_{ij} is to 0 the more likely it is the two areas will be treated as neighbours. To reflect these expectations, the prior on γ_1 is centred at a negative value so that a smaller d_{ij} gives a larger p_{ij}, and hence w_{ij} is more likely to take the value 1 (see Eq. 8.12). In the case of Model 1 where $d_{ij}=1$ (i.e. for two areas the maximum distance apart), the prior mean of the neighbouring probability is around 0.50 and there is a 95% chance that the neighbouring probability varies between 0.12 and 0.88. The calculation of these values is in Figure 8.15 and the results from Lines 26 and 28 give the values above. The same calculation can be used to assess the prior information on p_{ij} with a given value of the absolute mortality difference (simply change the value multiplied with `gamma1` in Line 24, Figure 8.15).

The exposition above gives an example of how we can investigate the behaviour of the model based on the chosen priors. When dealing with complex models such as the one above, this type of investigation is particularly important and should be carried out before

```
1   #  define the function expit, the inverse function of logit
2   #  (where logit(x) = log(x/(1-x)))
3   expit <- function(a) exp(a)/(1+exp(a))
4
5   #  define number of draws from the prior distributions
6   ndraws <- 100000
7   #  obtain a set of random draws (of size ndraws) from the prior
8   #  for gamma0
9   gamma0 <- rnorm(ndraws, 2, sqrt(0.5))
10  #  obtain a set of random draws (of size ndraws) from the prior
11  #  for gamma1
12  gamma1 <- rnorm(ndraws, -2, sqrt(0.5))
13
14  #  calculate the neighbouring probability when the similarity
15  #  measure is equal to 0
16  p0 <- expit(gamma0)
17  #  the prior mean of the neighbouring probability
18  mean(p0)
19  #  the 95% prior interval of the neighbouring probability
20  quantile(p0, c(0.025,0.975))
21
22  #  calculate the neighbouring probability when the similarity
23  #  measure is equal to 1
24  p1 <- expit(gamma0 + gamma1*1)
25  #  the prior mean of the neighbouring probability
26  mean(p1)
27  #  the 95% prior interval of the neighbouring probability
28  quantile(p1, c(0.025,0.975))
```

FIGURE 8.15
The R code to calculate the prior probability of treating two spatially-contiguous areas as neighbours given the value of the similarity measure, d_{ij}, and the moderately informative priors on γ_0 and γ_1. The basic idea is to carry out a Monte Carlo simulation, in which values are randomly drawn for γ_0 and for γ_1 from their respective prior distributions. For a given value of d_{ij}, we can then calculate p_{ij}, the (prior) neighbouring probability (of i and j being neighbours), and investigate its properties. Modify Lines 9 and 12 if different prior distributions are used. For example, change Line 9 to `gamma0 <- runif(ndraws, -10, 10)` if $\gamma_0 \sim \text{Uniform}(-10,10)$.

fitting the model to the data to ensure that the model with the chosen priors behaves as expected. Readers are encouraged to explore the impact on the prior neighbouring probability if (much) less informative priors, say Uniform(–10,10) or $N(0,10)$, are chosen for both γ_0 and γ_1 (Exercise 8.7). However, when vaguely/weakly informative priors are used, the data may not have sufficient information to estimate the two parameters.

Lee and Mitchell (2012) extended the model by Lu et al. (2007) to incorporate several covariates in the logistic regression for the weights. Instead of treating w_{ij} as a Bernoulli random variable, Lee and Mitchell (2012) modelled it deterministically:

$$w_{ij} = \begin{cases} 1 & \text{if } \exp\left(-\sum_{k=1}^{K} d_{ijk}\gamma_k\right) \geq 0.5 \text{ and } i \text{ and } j \text{ share a common boundary} \\ 0 & \text{otherwise} \end{cases} \tag{8.13}$$

The above model measures the similarity between two spatially-contiguous areas using K covariates, as represented by d_{ijk}, with $k = 1,\ldots,K$. As in Lu et al. (2007), $d_{ijk} = 0$ if two areas take the same value for covariate k. The regression coefficients $\gamma_1,\ldots,\gamma_K$ quantify the importance of each covariate on the neighbouring weights. In their application of modelling the lung cancer risk across the 271 administrative units in Glasgow, three covariates were included – smoking prevalence; ethnicity, which is measured by the percentage of school children from ethnic minorities; and the natural log of the median house price (Lee and Mitchell, 2012). These three covariates were also used in the Poisson regression to explain the risks of lung cancer.

The prior for the regression coefficient γ_k ($k = 1,\ldots,K$) takes the form $\gamma_k \sim \text{Uniform}(0, M_k)$. This prior choice restricts the regression coefficients to be positive so that a large value of d_{ijk} corresponds to a small probability of w_{ij} being 1. In other words, before seeing the data, it is expected that if two spatially-contiguous areas are very different in terms of the chosen covariates, they have a small chance of being treated as neighbours. The upper bound of the uniform prior is chosen "so that at most 50% of borders in the study region can be classified as boundaries" (Lee and Mitchell, 2012, p.419). This corresponds to a restriction that at most 50% of all pairs of spatially-contiguous areas are considered to be non-neighbours. The fitting of their model is done in R.

Both the models by Lu et al. (2007) and by Lee and Mitchell (2012) rely on covariates to inform where the "hard boundaries" are, or equivalently, where spatial discontinuity occurs. However, such covariates may not be available in some cases due to either the availability of the covariate data or the available covariates being uninformative in determining the boundaries. In the absence of such covariate information, Lee and Mitchell (2013) developed another locally adaptive spatial model which sets the weights of the spatially contiguous areas based on the posterior estimates of the area-specific parameters. Their model was constructed in the context of disease mapping where the outcome values are the numbers of disease cases (Y_i with $i = 1,\ldots,N$) across N spatial units:

$$Y_i \sim \text{Poisson}(\mu_i)$$

$$\log(\mu_i) = \log(E_i) + \alpha + \sum_{k=1}^{K} \beta_k x_{ik} + S_i \tag{8.14}$$

where E_i is the expected number of cases and x_{ik} is the kth covariate value for area i with β_k being the associated coefficient. S_i is the area-specific parameter that represents the residual risk that is not explained by the covariates included in the model. The set of $\boldsymbol{S} = (S_1,\ldots,S_N)$

is then modelled hierarchically using a spatial model in which the $\boldsymbol{W}$ matrix induces the spatial dependence structure. To allow adaptive smoothing, the modelling idea of Lee and Mitchell (2013) is to set two spatially-contiguous areas i and j as neighbours if the estimated S_i and S_j are similar (or, in other words, they have similar residual risks). If that is the case, it is then reasonable to borrow information between the two areas in estimating S_i and S_j. If, however, the estimated S_i and S_j are not similar, there is little support for borrowing information between them. The similarity between two estimated parameters is measured by whether their 95% credible intervals overlap or not. The weight of the two spatially-contiguous areas i and j is set as

- $w_{ij} = 1$ if the 95% credible intervals for S_i and S_j overlap.
- $w_{ij} = 0$ if the 95% credible intervals for S_i and S_j do not overlap.

This adaptive model is fitted iteratively. To start, the model in Eq. 8.14 with an exchangeable model for $\boldsymbol{S}$ (e.g. Eq. 7.6 in Chapter 7) is fitted to the data. The $\boldsymbol{W}$ matrix is then modified according to the rules given above and the model in Eq. 8.14 is refitted using the resulting $\boldsymbol{W}$ matrix via a chosen spatial model (e.g. the ICAR model or the Leroux model[10], and the latter model is used in Lee and Mitchell (2013)). This iterative process is carried out until either there is no change to the $\boldsymbol{W}$ matrix between two consecutive iterations or the current $\boldsymbol{W}$ matrix has already appeared in a previous iteration. To achieve computational efficiency, Lee and Mitchell (2013) used INLA for the iterative fitting of the spatial model. In the example below, we will apply this method to a subset of the Newcastle income data where WinBUGS is used.

8.4.3 Applying Some of the Locally Adaptive Spatial Models to a Subset of the Newcastle Income Data

To illustrate some of the locally adaptive smoothing models, we will use the income data for the seven MSOAs located in the middle of Newcastle (see Figure 8.14). For ease of discussion, the MSOAs are relabelled as MSOAs 1–7 (see Figure 8.16(a)) and the $\boldsymbol{W}$ matrix

(b)

$$\boldsymbol{W}_{cont} = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

FIGURE 8.16
(a) Map of the selected seven MSOAs. For ease of discussion, the MSOAs are relabelled as MSOA 1 to MSOA 7; and (b) the $\boldsymbol{W}$ matrix is derived using rook's spatial contiguity. Note that in (a), due to imprecision in the drawing, MSOAs 2, 3, 4 and 5 may look as if they all join at a single point, but they do not (MSOAs 2 and 4 share a common border but MSOAs 3 and 5 do not).

[10] The Leroux model was proposed by Leroux et al. (2000) and is a prior model that combines both global and local smoothing for the area-specific parameters. The model structure is similar to the BYM model, which will be introduced in Section 8.5.

Some possible modifications	The implied spatial configuration	Notes
A global smoothing model: Eq. 8.15 but with $S_i \sim N(0, \sigma_S^2)$ for $i = 1, \ldots, 7$ implying the following neighbourhood structure $\boldsymbol{W}_a = \begin{pmatrix} 0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0 \end{pmatrix}$		None of the spatially-contiguous MSOAs are treated as neighbours to each other but information is still shared globally through the unknown overall mean and variance.
A local smoothing model (Eq. 8.15) with $\boldsymbol{W}_b = \begin{pmatrix} 0&1&0&0&1&0&0 \\ 1&0&0&0&1&0&0 \\ 0&0&0&1&0&0&0 \\ 0&0&1&0&0&0&0 \\ 1&1&0&0&0&1&1 \\ 0&0&0&0&1&0&1 \\ 0&0&0&0&1&1&0 \end{pmatrix}$		MSOAs 3 and 4 are neighbours but neither is a neighbour to other spatially-contiguous MSOAs (e.g. MSOAs 2 and 3 are not treated as neighbours).
A local smoothing model (Eq. 8.15) with $\boldsymbol{W}_c = \begin{pmatrix} 0&1&0&0&0&0&0 \\ 1&0&1&1&0&0&0 \\ 0&1&0&1&0&0&0 \\ 0&1&1&0&0&0&0 \\ 0&0&0&0&0&1&1 \\ 0&0&0&0&1&0&1 \\ 0&0&0&0&1&1&0 \end{pmatrix}$		To reflect their differences in income level (see the bottom-panel boxplot in Figure 8.14), MSOA 5 is not considered as a neighbour to MSOAs 1, 2 and 4 but it is still a neighbour to MSOAs 6 and 7.

FIGURE 8.17
Three examples of possible modification to $\boldsymbol{W}_{cont}$. The maps in the middle column illustrate the spatial configurations implied by the corresponding $\boldsymbol{W}$ matrices. Thicker black lines denote the boundaries where two spatially-contiguous MSOAs are not treated as neighbours.

derived from rook's contiguity (denoted as $\boldsymbol{W}_{cont}$) is given in Figure 8.16(b). With 10 unique pairs of spatially-contiguous MSOAs, Figure 8.17 presents a selection of the 1024 possible modifications to $\boldsymbol{W}_{cont}$ that, together with $\boldsymbol{W}_{cont}$, we will investigate. Readers may construct other versions and compare the fits with those presented here (Exercise 8.8). Eq. 8.15 is the Bayesian hierarchical model with the ICAR prior fitted to the data – this model is the same as that defined in Eq. 8.2 but only for the seven MSOAs.

$$y_{ij} \sim N\left(\theta_i, \sigma_y^2\right)$$

$$\theta_i = \alpha + S_i$$

$$\boldsymbol{S}_{1:7} \sim ICAR\left(\boldsymbol{W}, \sigma_S^2\right)$$

$$\alpha \sim Uniform\left(-\infty, +\infty\right)$$

$$\sigma_y \sim Uniform\left(0.0001, 1000\right)$$

$$\sigma_S \sim Uniform\left(0.0001, 1000\right) \tag{8.15}$$

To find an optimal spatial structure, we fit the model in Eq. 8.15 using each version of the $\boldsymbol{W}$ matrix in Figure 8.17, apart from $\boldsymbol{W}_a$, for which the ICAR prior on $\boldsymbol{S}$ is replaced by $S_i \sim N(0, \sigma_S^2)$ for $i = 1, \ldots, 7$. Table 8.1 compares these models using the Deviance Information Criterion (DIC). Amongst the four models, the ICAR model with $\boldsymbol{W}_c$ gives the smallest DIC value of 974, suggesting it is the most parsimonious model for this set of data. With a difference in DIC value of only 1, the global smoothing (exchangeable) model performs equally well. However, the ICAR model with either $\boldsymbol{W}_{cont}$ or $\boldsymbol{W}_b$ yields a much poorer fit. The results therefore suggest that this dataset is modelled well by imposing either a global smoothing structure on the MSOA average income or a two-clusters neighbouring structure, as implied by $\boldsymbol{W}_c$ (see the spatial configuration in Figure 8.17). Of course, there are other configurations that should be considered. Instead of fitting them all here, readers are encouraged to fit other versions of the $\boldsymbol{W}$ matrix.

As an example of the second approach to locally adaptive smoothing, we take the method proposed by Lee and Mitchell (2013). The starting model is an exchangeable model for S_i with $S_i \sim N(0, \sigma_S^2)$ for $i = 1, \ldots, 7$, namely, the global smoothing model as presented in Figure 8.17.

For each of the 10 pairs of spatially-contiguous MSOAs, Figure 8.18(a) compares the 95% credible intervals (CIs) of the estimated S_is from the exchangeable model. For each pair, the corresponding element in $\boldsymbol{W}$ is set to 1 if the two CIs overlap and is set to 0 otherwise. The weight for a pair of MSOAs that do not share a common boundary is always fixed at 0. After Iteration 0, the weights of all contiguous pairs are set to 1, thus resulting in $\boldsymbol{W}_{cont}$, as presented in Figure 8.16(b). Using $\boldsymbol{W}_{cont}$, the next iteration fits the local smoothing model in Eq. 8.15 to the data. The 10 pairs of 95% CIs are shown in Figure 8.18(b) and the modification gives rise to $\boldsymbol{W}_c$, as in Figure 8.17. Iteration 2 uses $\boldsymbol{W}_c$ in Eq. 8.15, and Figure 8.18(c) compares the 95% CIs, which results in the same weights matrix, namely $\boldsymbol{W}_c$, as in Iteration 1. Thus, the iterative algorithm stops and $\boldsymbol{W}_c$ is the optimal spatial structure for these seven MSOAs, the same spatial configuration as determined from the first approach.

TABLE 8.1

Comparison of the Four Bayesian Hierarchical Models Fitted to a Subset of the Newcastle Income Data with Seven MSOAs

$\boldsymbol{W}$ used	Smoothing Type	$\bar{D}$	pD	DIC
$\boldsymbol{W}_{cont}$	Local	972	13	985
$\boldsymbol{W}_a$	Global	967	8	975
$\boldsymbol{W}_b$	Local	972	13	985
$\boldsymbol{W}_c$	Local	967	7	974

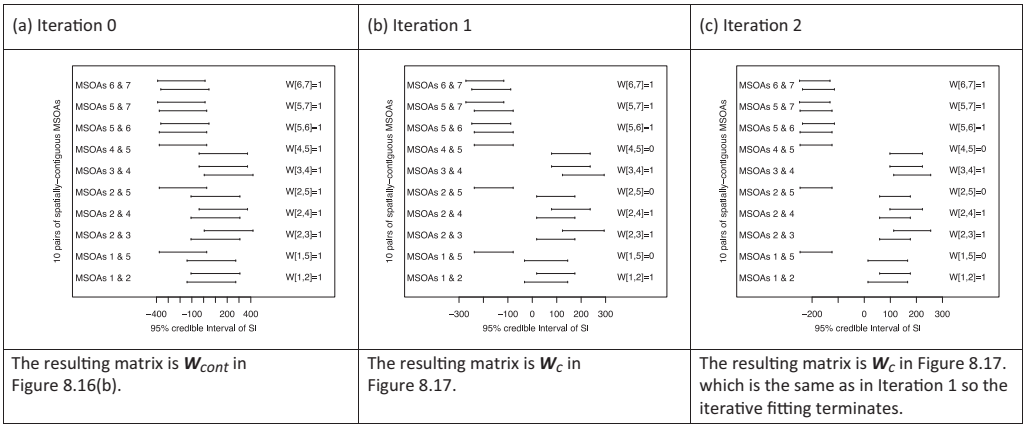

FIGURE 8.18
Comparison of the 95% credible intervals (CIs) of the area-specific parameters S_is from two spatially-contiguous MSOAs. If the two 95% CIs overlap, the weight is set to 1 and is set to 0 otherwise. Note that $w_{ij} = w_{ji}$.

Finally, Figure 8.19 compares the posterior estimates of the MSOA-level average income across three models, the exchangeable model and the two versions of the ICAR model where the spatial structure is imposed either by $\boldsymbol{W}_{cont}$ or by $\boldsymbol{W}_c$. Between the exchangeable model and the ICAR model with $\boldsymbol{W}_{cont}$, there is little difference in the posterior means for MSOAs 1–4, 6 and 7 (Figure 8.19(a)), but the posterior mean for MSOA 5 under the ICAR model with $\boldsymbol{W}_{cont}$ is pulled upwards due to the influence from the three high-income neighbours, MSOAs 1, 2 and 4. However, when $\boldsymbol{W}_c$ is used, those three high-income MSOAs are not treated as neighbours with MSOA 5 and, as a result, the posterior mean of MSOA 5 is less affected by those high-income areas and as a result remains low. The posterior means

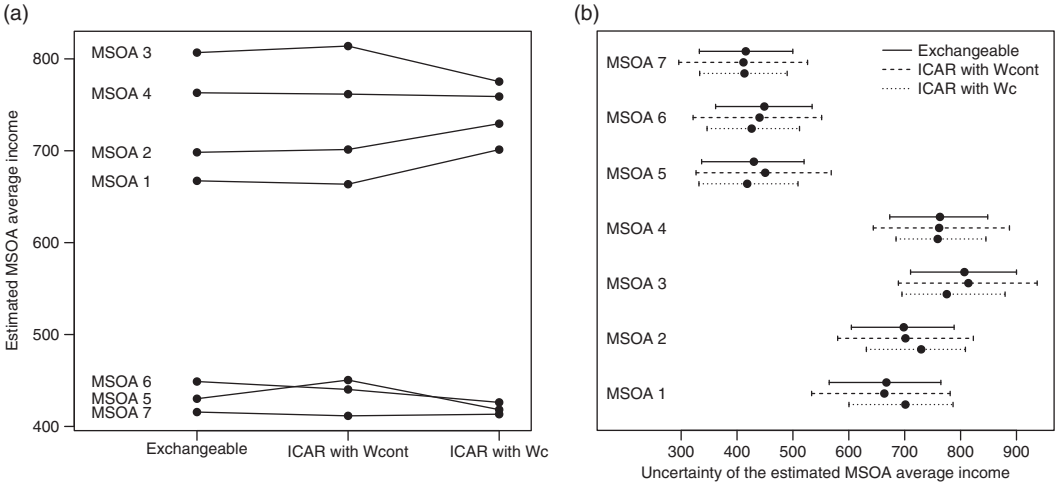

FIGURE 8.19
Comparison of (a) the posterior means and (b) the 95% credible intervals (with the dots denoting the posterior means) of the average income for the seven MSOAs from three models: the exchangeable model, the ICAR model with $\boldsymbol{W}_{cont}$ and the ICAR model with $\boldsymbol{W}_c$. See Figure 8.16 and Figure 8.17 for the specifications of $\boldsymbol{W}_{cont}$ and $\boldsymbol{W}_c$.

from the ICAR model with $\boldsymbol{W}_c$ (Figure 8.19(a)) better reflects the clustering feature, as evident in the exploratory plot in Figure 8.14 – the average income levels of MSOAs 1–4 tend to be similarly high, and those of MSOAs 5–7 tend to be similarly low, thus providing a better fit to the data.

As shown in Figure 8.19(b), the uncertainty intervals from the ICAR model with $\boldsymbol{W}_{cont}$ are wider compared to those from the other two models. The larger uncertainty may be due to the fact that this weights matrix imposes a spatial structure that is somewhat incompatible with the data. For the other two models, the uncertainty intervals are more comparable, although those from the ICAR model with $\boldsymbol{W}_c$ are slightly narrower, which may be due to the more appropriate sharing of local information through $\boldsymbol{W}_c$.

Some practical implications follow from this example. Albeit simple with only seven areas, the above example highlights the fact that the spatial neighbourhood structure that we impose on the area-specific parameters is an *assumption* so that different definitions of the neighbourhood structure can result in different estimates of the parameters. From a Bayesian viewpoint, the specification of $\boldsymbol{W}$ can be thought of as comparable to a prior – a prior which is assigned to the neighbourhood structure of the area-specific parameter. In Section 5.6.2, we called this type of information *spatial knowledge*. As with any Bayesian analysis therefore, assessing the sensitivity of modelling results to different prior specifications is an important, indeed integral, part of any Bayesian spatial analysis. The locally adaptive spatial models discussed in this section allow us to systematically (and automatically) explore different specifications of the spatial structure in order to assess their potential impacts on the estimates as well as on the conclusions we draw from the analysis.

Although used in the demonstration here, WinBUGS may not be suitable for fitting these adaptive models for two reasons. First, the weights derived from the $\boldsymbol{W}$ matrix can only enter the `car.normal` function (via the `weights[]` argument) or the `car.proper` function (via the `C[]` and the `M[]` arguments) as data. The elements in these arguments cannot depend on unknown parameters. As a result, WinBUGS cannot fit an adaptive model where the weights are modelled as a function of covariates (e.g. those proposed by Lu et al. (2007) and Lee and Mitchell (2012)). Second, when dealing with a study region with a large number of spatial units, WinBUGS may need a long time in order to fit a spatial model. This makes WinBUGS inefficient for exploring different specifications of the $\boldsymbol{W}$ matrix. Other programmes for fitting Bayesian models such as INLA (e.g. in Lee and Mitchell (2013)) can be used or the fitting algorithm can be specially written in R (e.g. in Lee et al. (2014) and Anderson et al. (2016)) or in C (e.g. in Li et al. (2011)).

8.5 The Besag, York and Mollié (BYM) Model

When displayed on a map, a set of outcome values often exhibits positive, global (or whole-map) spatial autocorrelation. The spatial distribution of these values is said to be *spatially-structured*. For example, high-income (or low-income) areas tend to cluster in space because of the spatial structure of the urban housing market. However, there may also exist some pattern-elements in the data that are not spatially-structured and which are said to be *spatially-unstructured*. One source for such a pattern element might be associated with individual areas whose income levels differ from their neighbours due, for example, to localised patterns of gentrification or housing in-fill. Another source is where there is some element of randomness in the degree of similarity between otherwise quite similar

contiguous areas. For example, some degree of income divergence might be expected in well-established housing areas as some streets or neighbourhoods improve and others decline, due to the effects of tenure history.

Such spatially-structured and spatially-unstructured patterns can be accounted for if the relevant covariates are available. Unfortunately, this may not be the case in practice, yet if our model is to provide a good basis for inference, it still needs to accommodate both the spatially-structured and the spatially-unstructured variation that may be present in the outcome data. In this section we introduce the BYM model, which combines two model components, one capturing spatially-structured variability and the other spatially-unstructured variability. This model is named after its authors, Besag, York and Mollié, who developed the model in their 1991 paper.

We formulate the BYM model in the context of the Newcastle income example. However, as we will illustrate subsequently in the book, this model can be applied to both continuous-valued and discrete-valued outcome data. As before y_{ij}, the income level from the survey for household j in MSOA i is modelled by $y_{ij} \sim N(\theta_i, \sigma_y^2)$. The aim of the analysis is to estimate the θ_is, the average income levels for the 109 MSOAs. Under the BYM model, the area-specific parameter θ_i is modelled as follows:

$$\theta_i = \alpha + S_i + U_i, \tag{8.16}$$

where S_i is a parameter from the spatially-structured component $\boldsymbol{S} = (S_1, \ldots, S_N)$ and U_i is a parameter from the spatially-unstructured component $\boldsymbol{U} = (U_1, \ldots, U_N)$. In the absence of all the necessary covariates in the model, $\boldsymbol{S}$ and $\boldsymbol{U}$ accommodate, respectively, the spatially-structured and the spatially-unstructured patterns present in the data. The parameters in each component are area-specific, and they are modelled hierarchically. For $\boldsymbol{S}$, the ICAR model is used so that the parameters in $\boldsymbol{S}$ are structured spatially according to the chosen spatial weights matrix. The parameters in $\boldsymbol{U}$ are modelled through an exchangeable model, for which a common choice is a normal distribution centred at 0 with an unknown variance, i.e. $U_i \sim N(0, \sigma_U^2)$, independently for all $i = 1, \ldots, N$. The parameters in $\boldsymbol{U}$ are therefore spatially-unstructured.

Because of the sum-to-zero constraint on $\boldsymbol{S}$ and the prior mean of 0 for the elements in $\boldsymbol{U}$, α, the intercept, is included to represent the overall level. As discussed in Section 8.2.1.2, $(S_i + U_i)$ can be interpreted as the deviation of the average income level in area i from the overall level α. Such local deviations can result from a combination of area-level characteristics that tend to be more alike in neighbouring areas and area-level characteristics that are specific to certain areas. In the absence of such covariate information, Besag et al. (1991, p.7) interpret the two components "as surrogates for unknown or unobserved covariates; the u_is represent variables that, if observed, would display substantial spatial structure in that the values for a pair of contiguous zones would be generally much more alike than for two arbitrary zones, whereas the v_is represent unstructured variables" – u_i and v_i in the quoted text correspond to S_i and U_i, respectively, in our notation.

When area-level covariates are available, Eq. 8.16 can be extended to

$$\theta_i = \alpha + \sum_{k=1}^{K} \beta_k x_{ik} + S_i + U_i \tag{8.17}$$

Then the two components measure the spatially-structured and the spatially-unstructured patterns in the *residuals* after the K covariate effects have been accounted for.

While the two components are included to represent the unmeasured/unobserved covariates, in terms of the estimation of the area-specific parameters, the hierarchical structures assigned to the two components imply that information is shared both locally and globally. Specifically, information is shared locally through the ICAR model on $\boldsymbol{S}$ while the exchangeable model on $\boldsymbol{U}$ allows information to be shared globally. In practice, for a given dataset, the relative importance of the two components is not known but can be quantified using the spatial fraction, a version of the variance partition coefficient that compares the variability in $\boldsymbol{S}$ to the variability of $\boldsymbol{S}$ and $\boldsymbol{U}$ combined. This relative importance has an implication on the estimates of the θ_is. As Besag et al. (1991) explain in the context of estimating small area disease risks, "If u [dominates], then the estimated risks will display spatial structure; if v, then the effect will be to shrink the estimated risks towards the overall mean." We will illustrate such a measure of importance in the example below.

8.5.1 Two Remarks on Applying the BYM Model in Practice

The first remark concerns the implementation of the BYM model. There are two equivalent formulations of the BYM model, the hierarchically-centred version and the non-hierarchically-centred version. The non-hierarchically-centred version is given in Eq. 8.16 (or Eq. 8.17 with covariates included), while the hierarchically-centred version with covariates is given below

$$\theta_i \sim N\left(\mu_i, \sigma_U^2\right)$$

$$\mu_i = \alpha + \sum_{k=1}^{K} \beta_k x_{ik} + S_i \tag{8.18}$$

The hierarchically-centred version is a re-parameterisation of the non-hierarchically-centred version so that the parameter estimates from both versions are virtually identical. Note that when fitted in WinBUGS, the posterior estimates from both versions will be very close but not identical because of the simulation nature of the MCMC method (see MC error in Section 5.3.5.3). From the hierarchically-centred version, we can recover the spatially-unstructured parameter U_i by calculating $\theta_i - \mu_i$ (see the income example below). When fitting a BYM model in WinBUGS, the hierarchically-centred version often yields faster convergence and better mixing.[11]

The second note is on the potential issue of non-identifiability. The BYM model partitions a single map of outcome values into two components. This partitioning causes a non-identifiability issue because each area has only one data value available to estimate two unknown parameters, S_i and U_i.[12] To overcome this problem, the two components are assumed to have different spatial structures *a priori*. The strong spatial positive autocorrelation from an ICAR model on $\boldsymbol{S}$ contrasts with the non-spatial structure from the exchangeable model on $\boldsymbol{U}$ so that the two components can be estimated separately. Although other

[11] First proposed in Gelfand et al. (1995), hierarchical centring is a method that re-parameterises a hierarchical model in order to achieve faster convergence when fitting the model using MCMC methods. Browne (2004) and Browne et al. (2009) further show the efficiency of the technique when fitting hierarchical models.

[12] In the context of count data, the single data value observed in each area refers to the number of crime or disease cases that occurred in the area (within a given period). In the case of the income survey example, although, for most of the areas, we have multiple data points, these data points are at the household-level. At the area-level, however, we only have one "observed value", which is the mean of the sample data in each area.

spatial models can be specified on $\boldsymbol{S}$ (e.g. a finite mixture model; see Chapter 9), the spatial dependence structure needs to be markedly different from the non-spatial structure of the exchangeable model. The proper CAR (pCAR) model for $\boldsymbol{S}$, on the other hand, is generally not recommended to be used in the BYM model. This is to avoid the potential non-identifiability between the two components when spatial autocorrelation in the data is weak – when ρ is close to 0, both the pCAR model and an exchangeable model with a common normal distribution imply similar structures on the area-specific parameters (see Section 8.3.2). Non-identifiability can lead to non-convergence of the MCMC chains for the three hyperparameters: σ_U^2, the variance of the non-spatial component; ρ, the spatial autocorrelation parameter; σ_S^2, the conditional variance in the pCAR model; and/or poor mixing of the MCMC chains for the above parameters (i.e. the iterations within each MCMC chain are highly autocorrelated).[13]

8.5.2 Applying the BYM Model to the Newcastle Income Data

Eq. 8.19 specifies a Bayesian hierarchical model with the BYM structure on the average income levels across the MSOAs. The spatial neighbourhood structure is defined using rook's contiguity, and the neighbouring weights are set to 1. The last line of Eq. 8.19 defines the spatial fraction, the ratio of $\tilde{\sigma}_S^2$, the unconditional variance of the spatially-structured random effect terms (which is not equal to the conditional variance σ_S^2), to the sum of the two variances, $\tilde{\sigma}_S^2$ and $\tilde{\sigma}_U^2$, where the latter is the variance of all the spatially-unstructured random effect terms. The WinBUGS implementation is given in Figure 8.20, and Lines 35–38 calculate the spatial fraction.

$$
\begin{aligned}
y_{ij} &\sim N\left(\theta_i, \sigma_y^2\right) \\
\theta_i &\sim N\left(\mu_i, \sigma_U^2\right) \\
\mu_i &= \alpha + S_i \\
\boldsymbol{S}_{1:109} &\sim ICAR\left(\boldsymbol{W}, \sigma_S^2\right) \\
\alpha &\sim Uniform\left(-\infty, +\infty\right) \\
\sigma_y &\sim Uniform\left(0.0001, 1000\right) \\
\sigma_S &\sim Uniform\left(0.0001, 1000\right) \\
\sigma_U &\sim Uniform\left(0.0001, 1000\right) \\
spatial\ \ fraction &= \frac{\tilde{\sigma}_S^2}{\tilde{\sigma}_U^2 + \tilde{\sigma}_S^2}
\end{aligned}
\tag{8.19}
$$

[13] Other modelling forms have been proposed to estimate the strength of spatial autocorrelation while simultaneously incorporating the spatially-unstructured component (see, for example, the Cressie model by Cressie (1992) and Stern and Cressie (2000) and the Leroux model first proposed by Leroux et al. (2000) then further explored in MacNab (2003)). Readers are referred to the aforementioned papers regarding the specifications of these models. MacNab (2003) and Lee (2011) both provide detailed comparisons of these models (including the BYM model) in simulation settings as well as in analysing real datasets.

```
#  The WinBUGS code for specifying the model with the BYM model
model {
  #  a for-loop to go through all the household level income data
  for (j in 1:nhhs) {
    #  defining the normal likelihood for each household
    y[j] ~ dnorm(theta[msoa[j]],prec.y)
  }
  #  the BYM model with hierarchical centring
  for (i in 1:nmsoas) {
    mu.theta[i] <- alpha + S[i]
    theta[i] ~ dnorm(mu.theta[i],prec.U)
    #  recovering the spatially-unstructured terms
    U[i] <- theta[i] - mu.theta[i]
  }
  #  modelling all the elements in S using the ICAR model
  S[1:nmsoas] ~ car.normal(adj[], weights[],num[], prec.S)

  #  the improper uniform prior for the Newcastle average
  alpha ~ dflat()

  #  a vague prior for the sampling standard deviation
  sigma.y ~ dunif(0.0001,1000)
  #  a vague prior for the conditional SD in the ICAR model
  sigma.S ~ dunif(0.0001,1000)
  #  a vague prior for the SD in the exchangeable model
  sigma.U ~ dunif(0.0001,1000)

  #  convert the SD to the precision
  prec.y <- pow(sigma.y,-2)
  prec.S <- pow(sigma.S,-2)
  prec.U <- pow(sigma.U,-2)

  #  calculate the spatial fraction to measure the relative
  #  importance of S and U
  var.S.un <- sd(S[])*sd(S[]) #  the unconditional variance of S
  var.U <- sd(U[])*sd(U[])    #  the variance of the spatially-
                                 #  unstructured parameters
  spatial.frac <- var.S.un / (var.S.un + var.U)

  #  calculate the posterior probability that the average income
  #  of MSOA i is above the Newcastle average
  for (i in 1:nmsoas) {
    diff[i] <- theta[i] - alpha
    prob.gt.average[i] <- step(diff[i])
  }
}

#  the data list is the same as that in Figure 8.1 for fitting
#  the model with the ICAR prior

#  initial values for chain 1
list(alpha=200,sigma.y=50,sigma.S=20,sigma.U=10
    ,S=c(1,-1,0,0,0,…),theta=c(300,300,300,...))

#  initial values for chain 2
list(alpha=700,sigma.y=80,sigma.S=50,sigma.U=20
    ,S=c(-1,1,0,0,0,…),theta=c(600,600,600,...))
```

FIGURE 8.20
The WinBUGS code for fitting the BYM model (Eq. 8.19) to the income data. The hierarchically centred version of the BYM model is used for the implementation.

We also calculate the posterior probability that the average income level of each MSOA exceeds the Newcastle average, i.e. $pp_i = \Pr(\theta_i > \alpha \mid \text{data})$. This posterior probability is calculated in Lines 42–45 using the `step` function. At each MCMC iteration, the `step` function returns 1 if its argument is non-negative (i.e. when θ_i is greater than α) and returns 0 otherwise. Thus, for each MSOA, the posterior mean of `prob.gt.average[i]` gives the proportion of iterations exceeding the Newcastle average and therefore the required posterior probability.

Figure 8.21 illustrates the structure of the BYM model. All maps in Figure 8.21 show the departures of the MSOA-level average income from the Newcastle average. For all

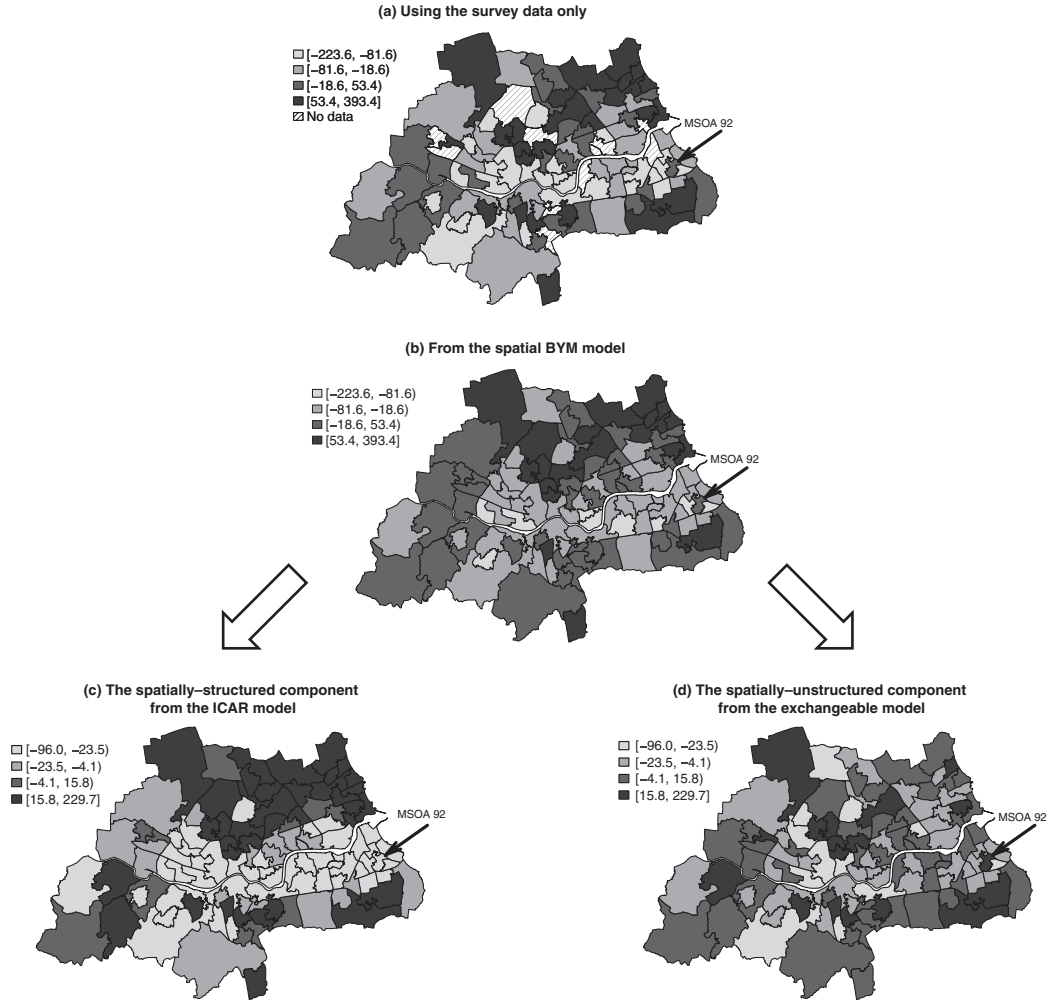

FIGURE 8.21

The maps of the local departures in average income from the Newcastle average (a) using only the survey data (no values assigned to the 10 MSOAs with no survey data) and (b) estimated based on the BYM model. The values in map (a) are calculated by subtracting the grand mean of the survey data, which is very close to the posterior mean of α from the BYM model, from the sample mean of each MSOA. The values in map (b) are the posterior means of $(S_i + U_i)$, for $i = 1,\ldots,109$. The posterior means of the spatially-structured and the spatially-unstructured components of the BYM model are shown in (c) and (d) respectively.

maps, an MSOA with a darker (or a lighter) colour is estimated to have an average income level above (or below) the Newcastle average. Based on the survey data, the BYM model estimates the average income levels across the MSOAs by combining the two model components, $\boldsymbol{S}$ and $\boldsymbol{U}$. Therefore, adding the maps in panels (c) and (d) of Figure 8.21 gives the map in panel (b), which, together with the posterior estimate of the overall intercept α, gives the estimates of the average income levels across MSOAs. Compared to the raw data map in panel (a), the posterior means from the BYM model in map (b) are smoother spatially in the sense that the income levels amongst the MSOAs that are geographically close are more alike compared to those from MSOAs that are far apart. The locally smooth feature of map (b) is a direct result of the spatially-structured parameters, $\boldsymbol{S}$, whose point estimates, as mapped in panel (c), show a spatial pattern that is broadly comparable to that in map (b). But there are clearly some differences between the two maps due to the spatially-unstructured component, $\boldsymbol{U}$, whose job is to capture the local variability that does not conform to the spatially-smooth pattern from the ICAR model.

To provide an example, MSOA 92 is highlighted in each map. With a sample size of six, this MSOA has a sample mean of 563.79, which is slightly above the Newcastle average of 525.59 (the grand mean of the survey data) and is higher than those of its contiguous neighbours, which tend to have a light grey colour in map (a) of Figure 8.21. Because of the local smoothing nature, the posterior mean of this MSOA under the spatially-structured component is –23.9, meaning that its average income is lower than the Newcastle average. It has clearly been overly smoothed. But this over-smoothed estimate is counterbalanced by the spatially-unstructured component, under which the posterior mean is 22.4. Therefore, the sum of the two point estimates is –1.5, suggesting that this MSOA may have an average income level that is close to the Newcastle average, in line with the observed data. This illustrates the joint effort of the two components in the BYM model in describing the patterns in the observed data. Exercise 8.9 asks the reader to investigate the BYM model further in order to provide some general insight into the data properties that determine the global and local smoothing of this model.

Note also that the above calculation is based on posterior means, only one of the many ways to summarise a posterior distribution. An advantage of Bayesian inference is that, using the posterior distribution, we can compute the posterior probability of how likely it is that the average income level of an MSOA is above (or below) a certain threshold (see Exercise 7.4 in Chapter 7 as well as the calculation of the posterior probability, $pp_i = \Pr(\theta_i > \alpha \mid \text{data})$, as defined above).

As has been noted, the spatial fraction is a way to gauge the relative importance of the two spatial components. For this dataset, the posterior median of the spatial fraction is 76.8% with a 95% CI of (23.3%, 99.9%), suggesting that compared to the spatially-unstructured component, the spatially-structured component accounts for a large proportion (about 77%) of the between-MSOA variability in average income. In other words, $\boldsymbol{S}$ dominates $\boldsymbol{U}$. This explains why the combined map (b) in Figure 8.21 is closer in pattern to the map of the posterior means of $\boldsymbol{S}$ than to that of $\boldsymbol{U}$. Since the spatial fraction is estimated to lie between 0 and 1, modelling the spatial distribution of the MSOA-level income data requires both the spatially-structured component and the spatially-unstructured component, although the former is more important. Note also that we use the posterior median, as opposed to the posterior mean, as a point estimate for the spatial fraction due to its skewed posterior distribution (see Figure 8.22).

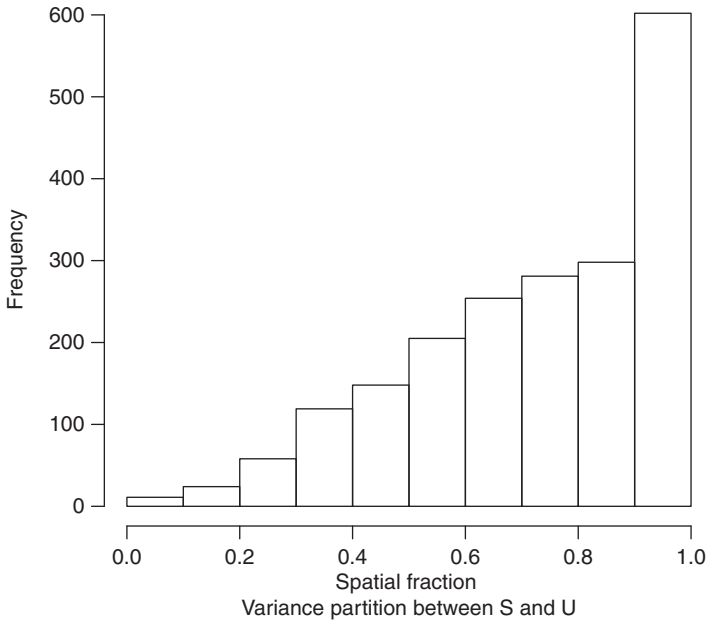

FIGURE 8.22
The posterior distribution of the spatial fraction.

Finally, Figure 8.23 shows the history plots for $\tilde{\sigma}_U^2$ and $\tilde{\sigma}_S^2$, the variance of the spatially-unstructured component and the unconditional variance of the spatially-structured component, respectively, and the spatial fraction. While the two chains have converged for all three quantities, the mixing is somewhat poor and the iterations within each chain are highly autocorrelated. Furthermore, a close inspection shows a negative correlation between the values for the two variances. For example, higher values of $\tilde{\sigma}_S^2$ tend to be associated with lower values of $\tilde{\sigma}_U^2$ (see, for example, the right-hand end of the black chains for $\tilde{\sigma}_S^2$ and $\tilde{\sigma}_U^2$). This is due to the identifiability issue between the two components discussed earlier. When the parameters in $\boldsymbol{S}$ vary more, then they explain more of the variability in the observed data, and hence the parameters in $\boldsymbol{U}$ are forced to vary less and vice versa. The two variances, $\tilde{\sigma}_S^2$ and $\tilde{\sigma}_U^2$, are thus intrinsically negatively correlated, and their correlation also results in the poor mixing of the MCMC chains. This poor mixing issue becomes less pronounced when the two components contribute to the combined map equally (or, equivalently, the spatial fraction is estimated to be well away from either 0 or 1). Figure 8.23 serves as an example of a typical mixing when one component dominates the other. While the autocorrelation of the MCMC iterations can be reduced via thinning whereby only every, say, 10th iteration is retained, we can alleviate the issue of poor mixing by simply running the MCMC chains longer. We can proceed with the posterior summary as soon as the requirements on convergence and efficiency are met (see Sections 5.3.4.2 and 5.3.4.3).

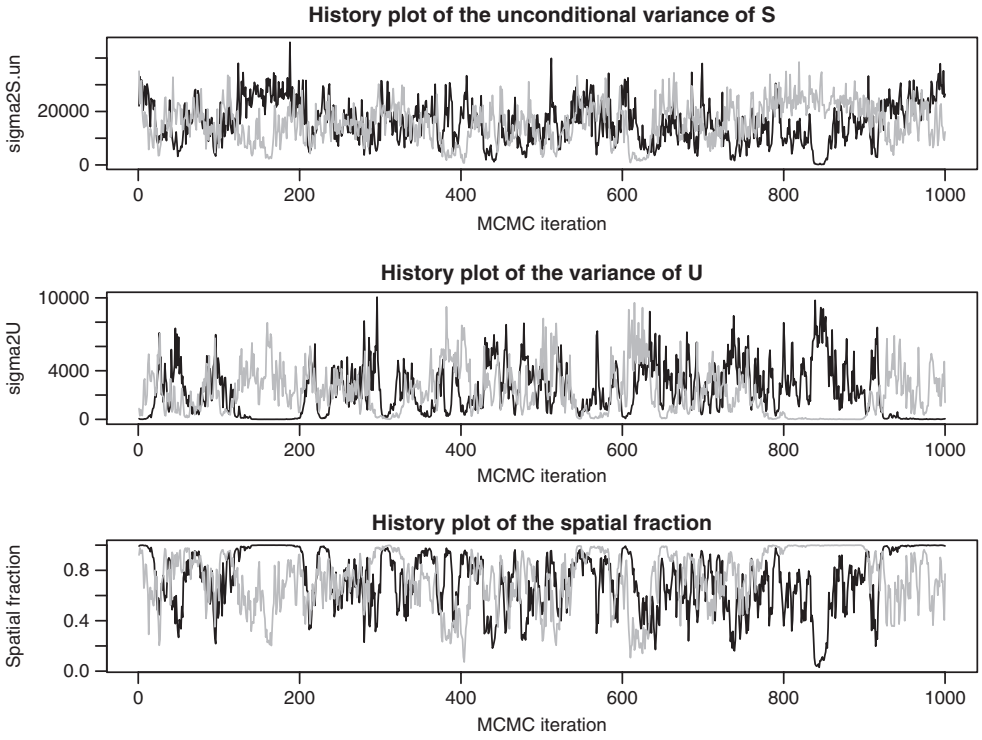

FIGURE 8.23
History plots of the unconditional variance of the spatially-structured component (top panel), the variance of the spatially-unstructured component (middle panel) and the spatial fraction (bottom panel). Each history plot is produced based on two MCMC chains run over 20000 iterations, with the first 10000 iterations discarded as burn-in. For the second half of each chain, every 10th iteration was stored, and each plot shows the resulting 1000 iterations from each chain. The two MCMC chains for each parameter have converged but show poor mixing with a high degree of autocorrelation within each chain.

8.6 Comparing the Fits of Different Bayesian Spatial Models

As we saw in Section 5.5, the Deviance Information Criterion (DIC), together with various other accuracy checks, can be used to compare Bayesian spatial models. We apply this approach now to the results we have obtained from fitting various models to the Newcastle income data.

8.6.1 DIC Comparison

Table 8.2 tabulates the values of $\bar{D}$, *pD* and *DIC* for five Bayesian models. Compared to the non-hierarchical model with identical parameters, the four Bayesian hierarchical models all have much lower DIC values and thus are more in line with the observed data. Amongst the four hierarchical models, the three models that include a form of local smoothing perform equally well and are all better than the model with exchangeable parameters, indicating the benefit of borrowing information locally in this application. In addition to the

evidence from the raw data (see for example Figure 8.21(a)), this message has been consistently revealed by various models, for example, through the estimated ρ in the pCAR model (Figure 8.13) and the spatial fraction from the BYM model (Figure 8.22).

Inspecting the values more closely, the BYM model has a slightly lower $\bar{D}$ value (by 2) than the ICAR model, suggesting the addition of the spatially-unstructured component leads to a slightly better fit to the data. The BYM model, however, is more complex than the ICAR model, so the two models are not different in terms of DIC. The similarity between these two models is also evident from the fit of the BYM model, where the spatially-structured component dominates the spatially-unstructured component – including $\boldsymbol{U}$ matters, but it is less important than $\boldsymbol{S}$. For the pCAR model, the spatial autocorrelation parameter ρ is estimated to be close to 1, thus the pCAR model and the ICAR model (where ρ is fixed at 1) have similar $\bar{D}$ and *pD* and the same DIC value.

With larger numbers of effective parameters (*pD*), the four hierarchical models are more complex than the non-hierarchical model, which has only two parameters, the overall average and the sampling variance – for a non-hierarchical model, *pD* is close to the actual number of parameters. However, for a hierarchical model, because of information sharing, *pD* is different from the actual number of parameters in the model. For example, for the exchangeable parameters model, *pD* is estimated to be 75, while there are 112 actual parameters (α, σ_θ^2, σ_y^2 and $\theta_1,\ldots,\theta_{109}$; see Eq. 7.6). See also the discussion in Section 5.5.

The DIC value of the non-hierarchical Bayesian model with independent parameters is 9569 (*pD* and $\bar{D}$ are 99 and 9469 respectively), much lower than those in Table 8.2. This is because the independent parameters model has a distinct feature of allowing the sampling variance to vary across areas (see Section 7.3.2 in Chapter 7, where the variance in the normal likelihood, σ_i^2, is indexed by *i*, the area indicator), while all the models in Table 8.2 assume a constant sampling variance (thus, there is only one variance parameter, σ_y^2, in, for example, Eq. 7.6 (Chapter 7) of the exchangeable model and in Eq. 8.2 of the ICAR model). The independent parameters model is a heteroscedastic model (also referred to as a heterogeneous variance model) and those in Table 8.2 are homoscedastic models. The lower DIC value from the independent parameters model suggests the benefit of allowing the within-area variance to be area-specific. Intuitively, the "varying-variance" assumption is more realistic since it seems very likely that the within-area variation in household income will differ from one area to another. The models in Table 8.2 can be extended to allow for area-specific within-area variances. One can plug in the sample variances as formulated in the independent parameters model, but this may give rise to problems

TABLE 8.2

Summary of the Posterior Mean Deviance ($\bar{D}$), the Effective Number of Parameters (*pD*) and the DIC Value for Each of the Five Bayesian Models Applied to the Newcastle Household-Level Income Data

Bayesian Models	Hierarchical?	$\bar{D}$	*pD*	*DIC*
With identical parameters (Eq. 7.1)	No	9910	2	9912
With exchangeable parameters (Eq. 7.6)	Yes	9647	75	9722
With an ICAR model (Eq. 8.2)	Yes	9647	69	9717
With a pCAR model (Eq. 8.11)	Yes	9646	71	9717
With a BYM model (Eq. 8.19)	Yes	9645	72	9717

The ICAR, pCAR and the BYM models all use the same spatial weights matrix, which is defined by rook's spatial contiguity with the weights of the neighbours set to 1.

when the data are sparse within each area. An alternative model-based approach is to model the (log) area-specific variances hierarchically, an idea that is similar to modelling the unknown area-level means. While such heterogeneous-variance models can be implemented in WinBUGS, they are beyond the scope of this book. Readers are referred to, for example, Leckie et al. (2014) for an exposition of such models. In addition, the reader is encouraged to study the paper by Best et al. (2005) where the authors compare a number of Bayesian spatial models in the context of disease mapping.

8.6.2 Model Comparison Based on the Quality of the MSOA-Level Average Income Estimates

Recall that the household-level income data were simulated using the known MSOA-level average (see the Appendix to Chapter 7). For each model, we can compare the posterior estimates of the MSOA-level average to the true values. The quality of the estimates is measured by three metrics, two of which are: (a) the average bias (avg. bias), the mean of the differences between the posterior means and the true values, and (b) the root mean square error (RMSE), the square root of the sum of the squared differences. Metrics (a) and (b) assess the quality of the point estimates so that a model with a smaller average bias or a smaller RMSE overall gives better estimates. Metric (c), the 95% coverage, measures how well a model represents parameter uncertainty. To calculate the 95% coverage, we assign the value 1 to an area if its 95% credible interval for the average income contains the true value and 0 if it does not. The mean of the assigned values of 0/1 across the areas gives the 95% coverage. If a model represents the uncertainty well, its 95% coverage should cover 95% of the true values. A 95% coverage that is higher (or lower) than 95% would suggest the estimates of this model are overly uncertain (or overly confident). These three metrics were calculated separately for the MSOAs with and without survey data, and the results are tabulated in Table 8.3.

For the 99 MSOAs with survey data, apart from the identical parameters model, all the other five models have similar summaries in terms of average bias, RMSE and 95% coverage. Due to the unrealistic assumption that all MSOAs have the same average income level, the identical parameters model performs poorly, with a high RMSE value and a very low 95% coverage. All the other five models allow the average income levels to vary across MSOAs, a feature that gives rise to posterior estimates of reasonably high quality. Across these five models, the point estimates are relatively accurate with small biases, and the

TABLE 8.3

Comparing the Quality of the Estimates for the MSOA-Level Average Income from Six Different Bayesian Models

	For the 99 MSOAs with Survey Data			For the 10 MSOAs with No Survey Data		
	Avg. bias	RMSE	95% Coverage	Avg. Bias	RMSE	95% Coverage
Independent	5.6	45.5	92.9%	–	–	–
Identical	5.8	95.4	9.1%	30.5	117.1	0.0%
Exchangeable	4.9	43.4	92.9%	29.9	117.4	80.0%
ICAR	5.4	41.6	92.9%	30.3	89.2	90.0%
pCAR	5.2	41.3	91.9%	29.2	95.6	80.0%
BYM	4.8	41.0	93.9%	30.1	92.8	90.0%

The comparison is carried out for the MSOAs with and without survey data separately.

95% CIs represent the uncertainty well since the 95% coverages are reasonably close to the nominal 95% level.

For the 10 MSOAs without survey data, there are no estimates from the independent parameters model. The identical parameters model has a poor 95% coverage of 0%. Although not yet hitting the 95% nominal level, the coverages from the four hierarchical models are much better, indicating a better representation of the uncertainty from these Bayesian hierarchical models. While these four hierarchical models do not differ much in terms of average bias, the three models that incorporate a form of local smoothing, i.e., the ICAR, the pCAR and the BYM models, yield similar values of RMSE and all are lower than the RMSE value from the exchangeable model. This highlights the benefit of borrowing information locally in this application, the same conclusion that we drew from the DIC comparison.

In a full simulation study, such model comparisons should be performed over a number of simulated datasets to obtain a more representative view on model performance (Exercise 8.10). The interested reader is referred to Gómez-Rubio et al. (2008a and 2008b) for a full investigation of using different Bayesian hierarchical models for small area estimation.

8.7 Concluding Remarks

At this point let us stand back from the detail of this chapter, and the previous one, to make clear where we have got to. Throughout these two chapters, our aim has been to provide estimates of a set of parameters, $\theta_1,\ldots,\theta_N$, for N small geographical areas using "information borrowing". In our running example, these N parameters represent the average household income levels across the MSOAs in Newcastle, England. The data available to us for estimating these parameters consist of income levels for sample households in the N areas. However, sample sizes vary between areas and in some cases are small whilst some areas have no sample data at all – the challenge of data sparsity. As an important part of this endeavour, we have shown, for each spatial prior model and sample size, *the mechanism* under which information is borrowed across spatial units and the *consequences* in terms of their effect on parameter estimates.

We adopted the Bayesian hierarchical modelling approach to the estimation of these parameters. To this end we specified a *data model* for the sample data, $y_{ij} \sim N(\theta_i, \sigma_y^2)$, where y_{ij} is the income value for household j in area i,and θ_i, the parameter of interest, is the average household income for area i. These data values are assumed to be independent given the data generating process. So, in addition to the data model, we also specified a *process model* for the set of parameters, $\theta_1,\ldots,\theta_N$. Typically, a process model includes an intercept term, a set of observable covariates that are thought to be associated with the outcome of interest and a set of area-specific *random effects*. It is the specification of the last of these three components that Chapters 7 and 8 have focused on. In Chapter 7, Strategy 3 specifies a *hierarchical* structure for these random effects in which the random effects are considered to be exchangeable. This exchangeable hierarchical structure allows *partial information borrowing* – a compromise between two non-hierarchical modelling structures, namely, complete information pooling (Strategy 1) and no information pooling (Strategy 2) – thereby addressing two of the challenges associated with modelling spatial data: spatial heterogeneity and data sparsity.

In this chapter, we have gone a stage further by modelling these N random effects according to the spatial dependence structure (amongst other properties) that we have

reason to believe is present in the parameters. Of course, that spatial dependence is, by definition, unknown so that *how* it is specified is a modelling assumption. A natural starting point for modelling these random effects is to invoke the idea that parameters in areas that are near in geographical space tend to be more alike compared to those for areas that are far apart. We may arrive at such an assumption drawing on relevant research studies that have suggested the presence of this property in the system that we wish to study. In the absence of such relevant studies, such an assumption may come from general assertions, such as Tobler's First Law of Geography or the First Law of Geostatistics (Chapter 3), or it might come from a quite basic intuition we hold to, that in a world of continuous space the above property is more likely to arise than not and hence may be considered as a reasonable assumption to make. In a *Bayesian* hierarchical model, we represent this spatial assumption by assigning *a spatial prior probability distribution* (or, equivalently, a *spatial prior model*) to the random effects. In this chapter we have looked at various prior models to formalise such spatial assumptions: the ICAR, the pCAR and the BYM models. We have also considered some local adaptive models where spatial discontinuity is suspected.

The various spatial priors (ICAR, pCAR, BYM) on the random effects may be considered to be "somewhat" *informative* because they represent our prior assumptions (or beliefs) on the spatial structure of these parameters. However, depending on what we invoke as the source of our prior knowledge, the quality of this prior information is likely to be variable. Very often, when specifying these spatial priors, we are not drawing on the accumulated evidence of many previous experiments; often we are simply invoking some general ideas we have about how "parameters vary spatially" (spatial knowledge as we termed in Section 5.6.2). For this reason, as part of our analysis we should consider the sensitivity of our results (parameter estimates) to our modelling assumptions – both the type of spatial model and the choice of $\boldsymbol{W}$. In doing so we should also recognise that these particular spatial models capture only a limited range of possibilities as to how attributes can vary spatially. We will pick up on this point in Chapter 17.

All the models considered in Chapters 7 and 8 are referred to as unit-level models because they deal with household-level outcome data (see Chapter 4 in Rao and Molina, 2015). However, our interest is in obtaining parameter estimates at the area level. Complications arise if we want to model outcome at the unit level (e.g. the household level, or in some cases the individual level) using covariates *and* derive area level estimates too. This is not the subject of this book, although we will discuss this in Chapter 17.

In the next chapter, Chapter 9, through a series of case studies we discuss several applications of Bayesian hierarchical modelling which will engage with other challenges in addition to the challenges and opportunities presented by spatial dependence. The applications we visit all involve the inclusion of covariates in the process model. Covariates are added to the process model in order to account for the variation in the observed outcome values so that we can make inferences about the underlying association between each covariate and the outcome of interest (the fixed effects). Such inference allows us to test theories or make predictions. Including spatial random effects in these circumstances means that information borrowing takes place across areas *after controlling for the effects of the included covariates*. In addition, including random effects enables us to acknowledge the presence of omitted covariates, which can display spatially-smooth and/or spatially-non-smooth patterns. Whilst our hope is that by including spatial random effects we shall obtain better estimates of the fixed effects, which usually are our principal interest, there is no guarantee that this will be so. As a warning, there remains the possibility of spatial confounding. See the discussion of this topic in Section 4.9.3.

8.8 Exercises

Exercise 8.1. Taking the joint probability distribution in Eq. 8.4 as the starting point, derive the conditional distributions under the ICAR model (Eq. 8.3) using the $\boldsymbol{W}$ matrix given in Figure 8.16(b).

Exercise 8.2. Show that the joint probability distribution for the ICAR model (Eq. 8.4) can be written in the form of Eq. 8.6, thereby illustrating the feature that only the pairwise differences of the random variables enter the ICAR model.

Exercise 8.3. Explore the relationship between income and the number of full-time working residents at the MSOA-level. Then fit the model presented in Eq. 8.7 to the Newcastle income data in WinBUGS, paying particular attention to the derivation of the spatial weights matrix $\boldsymbol{W}_{hour}$, whose non-zero elements are defined using the number of full-time working residents in each MSOA. (Hint: First derive $\boldsymbol{W}$ using rook's contiguity (see Appendix 4.13.2 in Chapter 4) then modify the non-zero weights accordingly; also set the difference to 10 if two MSOAs have the same number of full-time working residents.)

Exercise 8.4. Show that the largest eigenvalue of the row standardised weights matrix is 1 while all other eigenvalues are between –1 and 1.

Exercise 8.5. Carry out the simulation as described in Section 8.3.2 to investigate using ρ, the spatial autocorrelation parameter in the pCAR model, as a way to measure the strength of spatial autocorrelation. In the simulation, set v in Eq. 8.10 to 1.

Exercise 8.6. Fit the model in Eq. 8.11, using the pCAR prior for the spatially-structured random effects, to the Newcastle income data.

Exercise 8.7. The R code given in Figure 8.15 carries out a Monte Carlo simulation to investigate the impact of the choice of prior on the neighbouring probability in the locally adaptive spatial smoothing model (8.4.2) of Lu et al. (2007). Modify the R code in Figure 8.15 to explore the impact of the choice of prior on the neighbouring probability if the following two vague priors are used for both γ_0 and γ_1: (a) Uniform(–10,10) or (b) $N(0,10)$ assigned to both parameters.

Exercise 8.8. Follow the description in Section 8.4.3 to carry out the fitting of the model in Eq. 8.15 with several plausible specifications of the $\boldsymbol{W}$ matrix other than those given in Figure 8.17. Compare your model fits to those considered in Section 8.4.3 via DIC as well as through the posterior estimates of the MSOA-level average income (i.e. in the form of Figure 8.19).

Exercise 8.9. At various points in Chapters 7 and 8 we have provided insight into what is pulling estimates in different directions. For example, in the case of the exchangeable model (Section 7.4) we have commented on what determines the extent to which properties of the observed data pull the local estimates away from the sample mean and towards the global mean. In Section 8.2.1.4, we have indicated what influences the pull towards the local neighbours. In the case of the BYM model, provide some general insight into the data properties that determine the relative pull towards the sample mean, the global mean and the mean of the local neighbours. (Hint: A careful inspection of the results from fitting the model in Eq. 8.19 to the Newcastle income data may shed some light.)

Exercise 8.10. Carry out the model evaluation procedure as discussed in Section 8.6.2 on multiple (say 50) datasets, each simulated following the description given in the Appendix of Chapter 7.

Exercise 8.11. The zip file, `PHIA_with_covariates.zip`, on the book's website contains the shapefile of the Census Output Areas (COAs) in Sheffield with the binary 0/1 outcome values that indicate whether a COA was considered as a high intensity crime area (HIA) by the police (see Section 5.2.2). Four COA-level covariates are also included in the shapefile (see Exercise 5.11 in Chapter 5). Fit a logistic regression model to the binary outcome data with the four covariates and a set of area-specific random effects. Use the exchangeable model, the ICAR model, the pCAR model and the BYM model for the random effects. Compare the covariate effects estimated from the different models. Give reasons if you encounter non-convergence and/or poor mixing when fitting some of these models. Calculate residuals and test the correlation structure of the residuals of each model.

Exercise 8.12. Repeat the same analysis outlined in Exercise 8.11 for the data in the zip file, `EHIA_with_covariates.zip`, which contains the binary outcome data for the empirically-defined HIAs (see Exercise 5.12 in Chapter 5). Are the conclusions on the covariate effects different from those for the PHIAs? We will return to the analysis of the PHIA and EHIA data in Chapter 9.