A FLEXIBLE MULTIVARIATE CONDITIONAL AUTOREGRESSION WITH APPLICATION TO ROAD SAFETY PERFORMANCE INDICATORS

by

GRAHAM COOKSON

A dissertation presented for the examination

for the degree of Doctor of Philosophy

of Imperial College London

Centre for Environmental Policy

Faculty of Natural Sciences

Imperial College London

JULY 2009

Declaration of Own Work

I hereby declare that this submission is my own work and to the best of my knowledge it contains no material previously published or written by another person, nor material which has been accepted for the award of any other degree or diploma at Imperial College London or any other educational institution, except where due acknowledgement is made in the thesis.

Graham Cookson

To my mother, Jackie, who died at the start of this adventure.

The most important questions of life are indeed, for the most part, really only problems of probability.

Laplace (1812) *Théorie Analtique des Probabilitiés*

Table of contents

List of Figures

List of Tables

Acknowledgements

For many people, this will be the only page of this thesis that they read; I'll try and make this one comprehensible.

Starting from the top, the Economic Social and Research Council provided the funding for this research; I hope they got their money's worth. Like anything, none of this would have been possible without the love and support of my family over the years. Thanks especially to my (late) mum, dad, nan and pop.

The first three years of this thesis were completed in the Centre of Environmental Policy at Imperial College London where many people were influential and supportive. Professor Colin Thirtle, my Ph.D. supervisor, offered much needed support during some difficult times at the start, as well as his continual intellectual involvement throughout the thesis. My fellow cube dwellers at Imperial, Doctors Sian Morse-Jones, Lotte Ramsay and Ivan Scrase, kept me sane and hugely entertained.

The last 18 months have been spent as a teaching fellow and latterly lecturer in the Department of Management at King's College London where I began to think that I'd never finish this thesis. Everybody here has been wonderfully supportive offering sage words of advice and encouragement. I have been made to feel incredibly welcome at King's and for this, Ali Budjanovcanin, Professor Jenny Piesse, Dr. Martin Edwards and Dr. Mike Clinton deserve particular thanks.

I am very lucky to have a large group of friends that I've collected over the years who have provided much needed help and support throughout this, sometimes painful, process. Fond, yet sometimes painful memories, from the past four years include a trip to Oz, coming off my stead on a Welsh mountain, a coach journey in Wroclaw, cycling from Edinburgh to Newcastle, two new year's eves in Cornwall and Wilkin's farmhouse cider. Thank you all and I owe you all a drink (or two). There are too many people to mention individually but you know who you are: the King Edward boys, the Somerville posse, the Imperial boys, the Camberwell crew, the hobai and my housemates Dan Harewood and Nick Hill. A very special thanks must, however, go to Dan Taylor and Joanne Liddle for their Ph.D. completion incentive plan – they clearly need to read up on the Principal-Agent problem. And last, but certainly not least is Jayne Welburn – thanks for keeping me sane with plentiful supplies of tea and cake!

Thank you all. The onus has finally been put to bed.

Graham Cookson

London, July 2009

Abstract

There is a dearth of models for multivariate spatially correlated data recorded on a lattice. Existing models incorporate some combination of three correlation terms: (i) the correlation between the multiple variables within each site, (ii) the spatial autocorrelation for each variable across the lattice, and (iii) the correlation between each variable at one site and a different variable at a neighbouring site. These may be thought of as correlation, spatial autocorrelation and spatial cross-correlation parameters respectively.

This thesis develops a flexible multivariate conditional autoregression model where the spatial cross-correlation is asymmetric. A comparison of the performance of the FMCAR with existing MCARs is performed through a simulation exercise. The FMCAR compares well with the other models, in terms of model fit and shrinkage, when applied to a range of simulated data. However, the FMCAR out performs all of the existing MCAR models when applied to data with asymmetric spatial cross-correlations.

To demonstrate the model, the FMCAR model is applied to road safety performance indicators. Namely, casualty counts by mode and severity for vulnerable road users in London, taken from the STATS19 dataset for 2006. However, by exploiting correlation between multiple performance indicators within local authorities and spatial auto and cross-correlation for the variables across local

authorities, the FMCAR results in considerable shrinkage of the estimates of local authority performance. Whilst this does not enable local authorities to be differentiated based upon their road safety performance it produces a considerable reduction in the uncertainty surrounding their rankings. This is consistent with previous attempts to improve performance rankings. Further, although the findings of this thesis indicate that there is only mild evidence of asymmetry in the spatial cross-correlations for road casualty counts, the thesis provides a demonstration of the applicability of this model to real world social and economic problems.

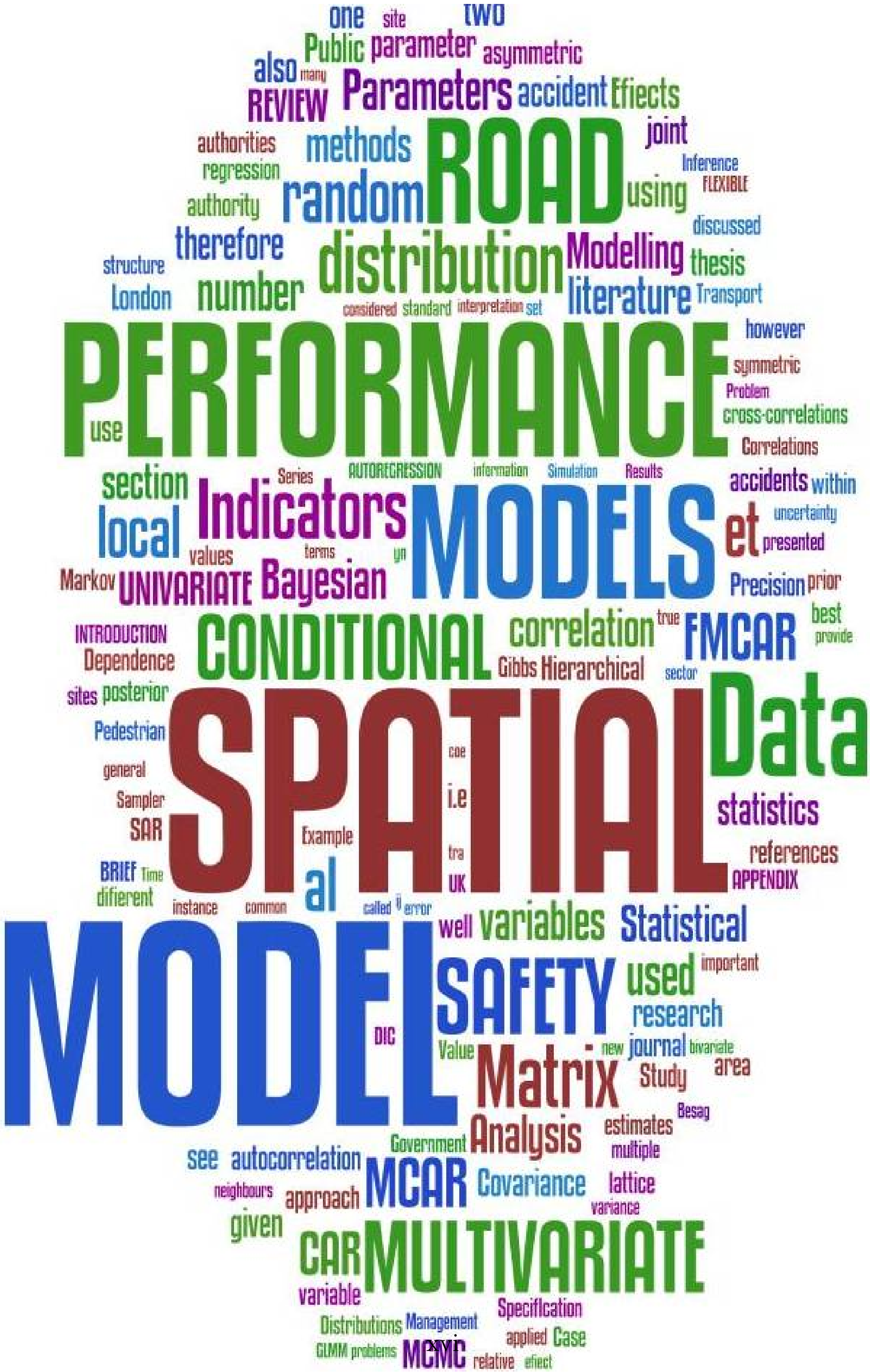

Figure 1: A Word Cloud representation of this thesis - created using http://www.wordle.net

Nomenclature & Notation

Introduction

Unlike the physical sciences there are no international standards for notation or symbols in economics and econometrics, only common conventions. As such, they are frequently broken. This section aims to set out the notation and nomenclature to be used consistently in this thesis.

Vectors and Matrices

Vectors are lowercase ($\boldsymbol{a}$) and matrices are uppercase ($\boldsymbol{A}$) symbols, but both are written in bold-italics. We write $\boldsymbol{a} = (a_{ij})$ to denote a typical element of matrix $\boldsymbol{A}$. The n columns of $\boldsymbol{A}$ are denoted by $\boldsymbol{a}_{\bullet 1}$, $\boldsymbol{a}_{\bullet 2}$, ..., $\boldsymbol{a}_{\bullet n}$, and the m rows by $\boldsymbol{a}'_{1\bullet}$, $\boldsymbol{a}'_{2\bullet}$, ..., $\boldsymbol{a}'_{m\bullet}$, where transpose is denoted by a prime. Two or more matrices (vectors) are *conformable* if their sum or product is defined.

Special vectors are:

$\mathbf{0}$, $\mathbf{0}_n$	null vector $(0, 0, \dots, 0)'$
$\boldsymbol{\imath}$, $\boldsymbol{\imath}_n$	sum vector $(1, 1, \dots, 1)'$

Special matrices are:

$\mathbf{O}$, $\mathbf{O}_{mn}$	null matrix of order $m \times n$
$\boldsymbol{I}$, $\boldsymbol{I}_n$	identity matrix of order $n \times n$.

Matrix Operations

The following matrix operations will be defined:

$\boldsymbol{A}'$	transpose
$\boldsymbol{A}^{-1}$	inverse
$\text{diag}(a_1, \ldots, a_n)$	diagonal matrix containing $a_1, \ldots, a_n$ on the diagonal
$\text{diag}(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_n)$	block-diagonal matrix with $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_n$ on the diagonal
$\boldsymbol{A}^2$	$\boldsymbol{AA}$
$\boldsymbol{A}^{1/2}$	(unique) square root of positive semidefinite matrix
$\boldsymbol{A}^p$	p-th power
$\boldsymbol{A}_k$	principal submatrix of order $k \times k$
$(\boldsymbol{A}, \boldsymbol{B})$, $(\boldsymbol{A} : \boldsymbol{B})$	partitioned matrix
$\text{rk}(\boldsymbol{A})$	rank
λ_i, $\lambda_i(\boldsymbol{A})$	i-th eigenvalue (of $\boldsymbol{A}$)
$\text{tr}\, \boldsymbol{A}$, $\text{tr}(\boldsymbol{A})$	trace
$\|\boldsymbol{A}\|$, $\det \boldsymbol{A}$, $\det(\boldsymbol{A})$	determinant
$\|\boldsymbol{A}\|$	norm of matrix ($\sqrt{(\text{tr}\, \boldsymbol{A}^* \boldsymbol{A})}$)
$\|\boldsymbol{a}\|$	norm of vector ($\sqrt{(\boldsymbol{a}^* \boldsymbol{a})}$)
$\boldsymbol{A} \geq \boldsymbol{B}$, $\boldsymbol{B} \leq \boldsymbol{A}$	$\boldsymbol{A} - \boldsymbol{B}$ positive semidefinite
$\boldsymbol{A} > \boldsymbol{B}$, $\boldsymbol{B} < \boldsymbol{A}$	$\boldsymbol{A} - \boldsymbol{B}$ positive definite ($>$, $<$)
$\boldsymbol{A} \otimes \boldsymbol{B}$	Kronecker product

If we have a symmetric matrix $\boldsymbol{A}$ of order $n \times n$, then the eigenvalues are real and can be ordered, such as

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n,$$

since there are many cases where it is desirable that λ_1 denotes the largest eigenvalue.

Mathematical symbols, functions and operators

Definitions, implications, convergence, and transformations are denoted by:

$\equiv$	identity, equivalence
$\Longrightarrow$	implies
$\Longleftrightarrow$	if and only if

We write $f(x) \approx g(x)$ if the two functions are approximately equal in some sense depending on the context. If $f(x)$ is proportional to $g(x)$ we write $f(x) \propto g(x)$).

The usual sets are denoted as follows:

$\mathbb{N}$	natural numbers $1, 2, \ldots$
$\mathbb{Z}$	integers $\ldots, -2, -1, 0, 1, 2, \ldots$
$\mathbb{Q}$	rational numbers
$\mathbb{R}$	real numbers
$\mathbb{C}$	complex numbers

Superscripts denote the dimension and subscripts the relevant subset. For example, $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ denotes the real plane, $\mathbb{R}^n$ the set of real $n \times 1$ vectors, and $\mathbb{R}^{m \times n}$ the set of real $m \times n$ matrices. The set $\mathbb{R}^n_+$ denotes the positive orthant of $\mathbb{R}^n$, while $\mathbb{Z}_+$ denotes the set of positive integers (hence, $\mathbb{Z}_+ = \mathbb{N}$) and $\mathbb{Z}_{0,+}$ denotes the non-negative integers. Finally, $\mathbb{C}^{n \times n}$ denotes the set of complex $n \times n$ matrices.

Other symbols used are:

$\in$	belongs to
$\notin$	does not belong to
$\{x : x \in S, x \text{ satisfies } P\}$	set of all elements of S with property P
$\subseteq$	is a subset of
$\subset$	is a proper subset of
$\cup$	union
$\cap$	intersection
$\emptyset$	empty set
A^c	complement of A
$B\backslash A$	$B \cap A^c$

We denote functions by:

$f : S \to T$	function defined on S with values in T
f, g, φ, ψ, ϑ	scalar-valued function
$\boldsymbol{f}$, $\boldsymbol{g}$	vector-valued function
$\boldsymbol{F}$, $\boldsymbol{G}$	matrix-valued function

Finally, various other symbols in common use are

e, exp	exponential
log	natural logarithm
$\log_a$	logarithm to the base a
!	factorial
$\lvert x\rvert$	absolute value (modulus) of scalar $x \in \mathbb{C}$
$1_{\mathcal{K}}$	indicator function (note the use of 1, not I):

equals 1 if condition $\mathcal{K}$ is satisfied, 0 otherwise

Statistical symbols, functions and operators

It is customary to use capital letters (e.g. X) for random variables and lowercase letters for their realisations, for example $\Pr(X = x)$. We cannot do this in a thesis on multivariate statistics as there will inevitably be the problem that $\boldsymbol{X}$ and $\boldsymbol{x}$ have been reserved for matrices and vectors respectively.

We follow the convention to denote the cumulative distribution function (c.d.f) by F and the probability density function (p.d.f) by f. In general, these will depend on a vector of m parameters, $\boldsymbol{\theta}$. An *estimator* of $\boldsymbol{\theta}$ is $\hat{\boldsymbol{\theta}}$ (and $\tilde{\boldsymbol{\theta}}$ if it's a second estimator) and its realisation is an estimate. We use the word *predictor* for an "estimator" of a random variable, employing the same hat and tilde conventions as for estimators. The realisation of an predictor is a prediction.

We denote the null hypothesis as H and alternatives as $\mathrm{H}_1, \mathrm{H}_2, ...$ The value of $\boldsymbol{\theta}$ under the hypothesis H_j is denoted as $\boldsymbol{\theta}^j$. If the hypothesis concerns only a subset of $\boldsymbol{\theta}_i$ then denote the value under the hypothesis H_j as $\boldsymbol{\theta}_i^j$.

The following symbols are commonly used:

$\sim$	is distributed as
$\overset{a}{\sim}$	is asymptotically distributed as
$\pi(\cdot)$	probability
$\mathrm{E}(\cdot)$	expectation
$\mathrm{E}(\cdot\|\cdot)$	conditional expectation
$\mathrm{var}(\cdot)$	variance (matrix)

$\text{cov}(\cdot, \cdot)$	covariance (matrix)
$\text{corr}(\cdot, \cdot)$	correlation (matrix)
$L(\boldsymbol{\theta})$	likelihood function
$\ell(\boldsymbol{\theta})$	log-likelihood function
t	t-statistic, t-value
$\to, \longrightarrow$	converges a.s.
$\xrightarrow{p}$	converges in probability
$\xrightarrow{d}$	converges in distribution
plim	probability limit

The main distributions in statistics are denoted as follows:

$\text{Bin}(n, p)$	binomial distribution
$\text{Poi}(\mu)$	Poisson distribution
$\text{U}(a, b)$	uniform distribution
$\text{N}_m(\boldsymbol{\mu}, \boldsymbol{\Omega})$	m-dimensional normal distribution
$\text{LN}(\mu, \sigma^2)$	lognormal distribution
$\phi(\cdot)$	standard-normal p.d.f.
$\Phi(\cdot)$	standard-normal c.d.f.
$\chi^2_n(\delta)$	chi-squared distribution with n d.f. and non-centrality parameter δ.
χ^2_n	central chi-squared ($\delta = 0$)
$\text{t}_n(\delta)$	Student distribution with n d.f. and noncentrality δ
t_n	central t ($\delta = 0$)
$\text{C}(a, b)$	Cauchy distribution

$\Gamma(\alpha, \lambda)$	gamma distribution
$\mathrm{B}(a, b)$	beta distribution
$W(\tau)$, $B(\tau)$	standard Wiener process (Brownian motion) on $\tau \in [0, 1]$

Quantiles are denoted as follows. If a random variable follows some distribution $\mathrm{D}(\boldsymbol{\theta})$, then the α^{th} quantile is $\mathrm{D}_\alpha(\boldsymbol{\theta})$. For example, $\mathrm{t}_{0.05}(n, \delta)$ denotes the 5 percent quantile of the non-central t-distribution.

We use the word 'expectation' to denote mathematical expectation of a random vector $\boldsymbol{x}$, written $\mathrm{E}(\boldsymbol{x})$. The word 'average' refers to taking the average of some numbers: $\bar{x} = (1/n)\sum_{i=1}^{n} x_i$. Like 'expectation', the words 'variance' (var), 'covariance' (cov), and 'correlation' (corr) indicate population parameters. The corresponding sample parameters are called 'sample variance', 'sample covariance' and 'sample correlation'.

Abbreviations and acronyms

2SLS	two-stage least squares
3SLS	three-stage least squares
AR(p)	autoregressive process of order p
CAR	conditional autoregression
c.d.f.	cumulative distribution function
d.f.	degrees of freedom
DW	Durbin-Watson
FMCAR	flexible multivariate conditional autoregression
GLM	generalized linear model
GLMM	generalized linear mixed model

GMM	generalized method of moments
GMRF	Gaussian Markov random field
i.i.d.	independent and identically distributed
IV	instrumental variable
LM	Lagrange multiplier
LR	likelihood ratio
LS[E]	least squares [estimator]; see also 2SLS, 3SLS,
MA(q)	moving-average process of order q
MCAR	multivariate conditional autoregression
MCMC	Markov chain monte carlo
ML[E]	maximum likelihood [estimator];
MRF	Markov random field
MSE	mean squared error
OLS	ordinary least squares
p.d.f.	probability density function
r.v.	random variable
SAR	simultaneous autoregression
s.e.	standard error
SEM	structural equation modelling
SFA	spatial factor analysis

CHAPTER 1

INTRODUCTION

Everything is related to everything else, but near things are more related than distant things. (Tobler 1970, p.236)

1.1 Space: the final frontier

Traditionally, economists have been more reluctant than geographers like Tobler to consider space as a relevant factor. In 1890, economist Alfred Marshall asserted the dominance of time by maintaining that the working of the market depends *"...chiefly on variation in the area of space, and the period of time over which the market in question extends; the influence of time being more fundamental than that of space"* (Marshall 1920, Bk. V chap. XV sec.1). Not until Isard (1956) do economists question this preoccupation with time when he famously commented that Hicks (1939) confines economic theory *"to a wonderland of no spatial dimensions"*. Isard coined the term "Anglo-Saxon bias" for the prevailing paradigm within general equilibrium analysis in the tradition of Walras, Pareto, and Hicks of failing to consider spatial dimensions explicitly. Thus, following the economic trends of the day econometric investigation has traditionally favoured the time rather than the spatial domain.

The regional science literature was the first to take the issue of 'space' seriously within economics, with Dutch-Belgian regional economist Jean Paelinck coining the term "spatial econometrics" and writing the first book on the subject (see Paelinck & Klaassen 1979). The first modern textbook treatment of spatial econometrics was Anselin (1988), another regional economist, with later texts including Anselin (2003*a*) and Anselin et al. (2004). At the same time, statisticians such as Besag (1974), Cliff & Ord (1981), Diggle (1983), Ripley (1981) and Cressie (1993) were developing appropriate methods from the statistician's perspective, much of which is of use in econometrics.

Once a marginal endeavour, modelling spatial interactions is now commonplace in applied econometrics. Economists are increasingly aware of the relevance of spatial interactions, spatial externalities and networking effects between agents in all fields of economic research (Florax & Nijkamp 2004) . This has prompted the development of the subdiscipline of spatial econometrics with methods to specify and estimate regression models that explicitly include and correct for spatial effects (Anselin et al. 2004). Spatial econometrics is now firmly established as a subdiscipline with its own professional association[1], several books[2] dedicated to the topic and a number of prestigous peer-reviewed journals publishing special issues[3] on spatial problems and methods.

Geography and the role of spatial interaction have gained prominence in the applied as well as the theoretical literature. This growing literature on empirical spatial econometrics includes many of the traditional fields of economics. A few

[1]Visit http://spatialeconometr.altervista.org/ for further details

[2]see for example Anselin (1988), Anselin & Florax (1995), Anselin et al. (2004), Arbia (2006), Bailey & Gatrell (1995), Banerjee et al. (2004), Cressie (1993), Haining (1991), Paelinck & Klaassen (1979) and Ripley (1981)

[3]see for example Anselin (1992), Anselin (2003*b*), Baltagi et al. (2007), Florax & van der Vlist (2003), Holloway (2007), LeSage et al. (2004), Nelson (2002), Pace et al. (1998) and Pace & LeSage (2004)

examples include studies in demand analysis (Case 1991), international economics (Aten 1996), labour economics (Topa 1996), public economics (Case et al. 1993), agricultural economics (Holloway & Lapar 2007), environmental economics (Nelson & Hellerstein 1997), microeconomic theory (Durlauf 1997), development economics (Nelson & Gray 1997), and financial economics (Ioannides 1997) among many others. This short list is far from exhaustive and it is not the aim of this thesis to survey the whole literature on either spatial econometric theory or applications. Interested readers are invited to consult one of the many journal special editions which provide an excellent overview of important research directions.

1.2 The importance of space

There are obvious differences between spatial and time-series data (Pinkse et al. 2007). The most commonly noted differences are that (i) time is unidirectional whereas space is (usually) multidirectional , (ii) time is one-dimensional whereas space is of higher dimensionality, (iii) time series observations are (normally) uniformly spaced on the time line whereas spatial data are rarely observed on a regular grid or lattice, (iv) time series observations are considered draws from a continuous stochastic process, where as in spatial data analysis it is common for the sample and the population to be the same (causing problems for inference and asymptotics)

Spatial effects include spatial heterogeneity and spatial dependence. Standard texts on spatial effects, specification strategies and an overview of inference for the standard spatial process models include Anselin (1988), Haining (1991) and Cressie (1993). Spatial heterogeneity refers to structural relations that vary over space, either in a discrete manner (for instance urban versus rural) or in a continous manner (such as a trend surface for ozone). Spatial dependence points to systematic spatial variation that results in observable clusters or systematic spatial pattern. The usual convention of using the terms spatial dependence, spatial autocorrelation and spatial clustering interchangeably is continued. Strictly speaking, spatial dependence is a characteristic of the joint probability density function. As such, it is only verifiable under simplifying conditions such as normality. Spatial autocorrelation is simply a moment of that joint distribution.

The presence of spatial heterogeneity does not necessarily have severe implications for the information that can be obtained from the spatial dataset. Spatial autocorrelation does, however, because an observation is partly predictable from neighbouring observations. A series of spatially dependent observations therefore contains less information. This is similar to the time series situation where a forecast can be partly inferred from the past

Although work on spatial autocorrelation can be traced back to the work of pioneering statisticians such as Moran (1950), Geary (1954) and Whittle (1954), the development of the literature is slow until Cliff & Ord (1981). Ignoring spatial autocorrelation when it is present has different consequences depending upon whether the correct model is a spatial lag or a spatial error specification[4] (see Anselin 2003*a*, for details). Ignoring a spatially lagged dependent variable is equivalent to an omitted variable error and will lead to Ordinary Least Squares (OLS) estimates that are biased and inconsistent. Monte Carlo studies have shown OLS estimates to biased by up to 35 percent when a spatially lagged dependent variable is incorrectly excluded from the specification (Darmofal 2006). Alternatively, ignoring the presence of spatially correlated errors will produce biased standard errors for the OLS estimates, but the OLS estimates themselves will remain unbiased; it is therefore more a problem of

[4]Unsurprisingly, the spatial lag model includes a spatially lagged dependent variable where as the spatial error model includes an autoregressive process for the error term.

efficiency. Yet this can cause serious Type I errors with the biased standard errors being as low as 50 percent of the true standard errors (Darmofal 2006). Moreover, to the extent that the spatially correlated errors mask a spatially varying omitted variable, the true consequences of ignoring this problem may be more serious than much of the literature acknowledges.

Research into the specification, estimation and application of spatial regression models is a legitimate and worthwhile enterprise with applications spanning the breadth of economics. Although much work has taken place in this arena there is still a long way to go until the range of spatial methods matches those available in the standard cross-sectional and time-series toolboxes.

1.3 Road Safety Performance Indicators

Performance management is a high profile activity throughout the public sector in the UK. Road safety is no exception and is typical of activities which are monitored by performance indicators (PIs) based on 'outcome measures' (Bailey & Hewson 2004). In particular, the UK Government has identified three traffic safety targets which are expected to be achieved by 2010[5]: a 40 percent reduction in the number of fatally or seriously injured casualties, a 10 percent reduction in the rate of slight casualties relative to the level of traffic, and a 50 percent reduction in the number of children who were fatally or seriously injured (DfT 2000). Related performance indicators, broken down by modal group, are monitored and published in the local authority league tables under the 'best value' requirements of the Local Government Act 1999 (Department for Transport, Local Government, and the Regions 1999).

[5] These were set relative to a baseline of the mean number of casualties that were reported between 1994 and 1998 inclusively

However, current UK road safety performance indicators continue to be expressed simply in the form of crude per capita numbers of reported collisions by type and modal group, with no allowance for geographically differing patterns in road infrastructure and usage, or spatially varying socioeconomic conditions. In fact, according to Bailey & Hewson (2004) there is no explicit consideration given to the extent to which differences in the raw rates reflect differential performance, rather than just inherent random variability in observed rates. In general, local government activity does not appear to have received anything like as much attention in the literature as that devoted to performance monitoring in other sectors. For instance, although local government (in the UK at least) plays a significant role in education, performance monitoring interest in that sector has largely focused on the school as the observational unit, rather than on the Local Educational Authority.

Traditional econometric methods for modelling performance and productivity of organisations such as Data Envelopment Analysis and Stochastic Frontier Analysis are problematic when applied to the public sector (Stone 2002*a*). This is partly due to the lack of prices for outputs and poor data on outputs, but there is also some concern that a single measure of (in)efficiency isn't appropriate for monitoring complex public service organisations (Smith & Street 2005). Given the extent of the public sector and the lack of appropriate tools available, developing more general methods for measuring and ranking public sector performance is clearly an important theme for social science research.

1.4 The contribution of this research

There are two motivations for this research. One is methodological — to extend the range of multivariate conditional autoregressive models available for spatially correlated data; and the second is applied — to reduce the uncertainty in multiple road safety performance indicators by exploiting the spatial correlation inherent in the data.

Many applied econometric problems are inherently multivariate in that more than one dependent variable is measured for each unit of observation. Multivariate spatial datasets are now prevelant in economics, particularly at areal level. Yet despite a growing spatial methodological literature there are limited empirical tools available to investigate multivariate spatial data. Chapter 3 provides an overview of the main models in the existing sparse literature. In general, existing multivariate approaches are shown to have severe constraints on the within site and across site correlations (see the discussion in Banerjee et al. 2004, chap. 7). The conditional means in the existing models are also directly dependent upon the number of neighbours, which in irregular lattices will not be constant across the lattice. Additionally these multivariate models are intended for the analysis of continuous dependent variables. Frequently, economically relevant variables are discrete and this will be the focus of this thesis.

This thesis will develop a new model for multivariate spatial data recorded on a lattice. The innovation in this model will be the incorporation of very general forms of intra and inter site correlations for the multiple variables i.e. allowing for the possibility of asymmetric spatial cross-correlations. To make this concrete, consider a lattice consisting of just three sites and two variables. In figure 1.1 there are three sites (1, 2 and 3) represented by the circles and for each site data on two variables (A, B) are recorded. There are five separate correlation parameters indicated in the figure. There is a *spatial* autocorrelation parameter α_1 relating observations of variable A across the sites on the lattice. Similarly, variable B has its own spatial autocorrelation parameter, α_2. Given that two variables are recorded at each site, there is the potential that these two variables are correlated, hence the correlation coefficient, ρ in figure 1.1. There is also what is termed a spatial cross-correlation or linking parameter in the literature, α_3, which links variable A at site i with variable B at site j. The existing literature consists of models that incorporate these four types of correlation. This thesis aims to develop a model that allows for the cross-correlation to be asymmetric i.e. that there is a fifth correlation parameter, α_4, linking variable A at site j with variable B at site i. The additional flexibility offered by incorporating asymmetric spatial cross-correlations will be the principal theoretical aim of this research.

The use of generalized linear mixed models (GLMMs) for performance measurement and ranking is well-established in the literature. A common finding is that these models result in an inability to differentiate performance. In an attempt to improve the estimation of local authority road safety performance, this new flexible spatial model will be incorporated into a GLMM for a dataset of multiple road traffic performance indicators recorded for the census output areas. Each performance indicator is nested within a census output area which is nested within a Local Authority (LA). The GLMM will provide estimates of a random effect which can be considered latent local authority performance. By incorporating a spatial model into this GLMM, additional structure will be imposed onto the random effects with the aim of reducing the variance or uncertainty associated with these performance measures. This is the empirical aim of this research.

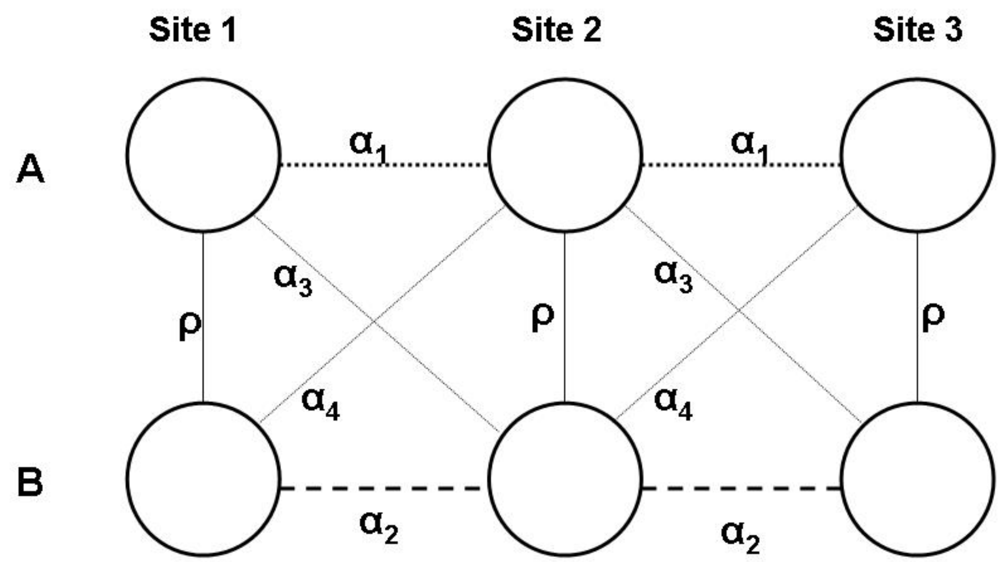

Figure 1.1: A representation of correlation coefficients on a two-dimensional bivariate lattice

1.5 Research Aims & Objectives

This research aims to extend the standard econometric toolbox to include a multivariate model with asymmetric spatial cross-correlations. After developing the model it will be demonstrated through an application to road safety performance indicators, a topic at the border between transport economics and public economics. The detailed objectives are therefore to:

1. Develop a flexible multivariate conditional autoregression that allows asymmetric inter site spatial correlations.

2. Demonstrate the performance of the model through a comparison with existing models using simulated data.

3. Demonstrate the applicability of this model through investigation of multiple traffic safety performance indicators in London.

4. Contribute to an improvement in public sector performance management by reducing the uncertainty of performance rankings through the inclusion of spatial correlation.

5. Provide a more general method for ranking public sector organisations than Data Envelopment Analysis and Stochastic Frontier Analysis.

6. Contribute to the road safety literature by identifying good and weak performing local authorities.

7. Provide the relevant computer code to perform parameter estimation, statistical inference and diagnostics within the Bayesian paradigm.

8. Provide a thorough introduction to multivariate conditional autoregression models.

1.6 Summary

The importance of space was first acknowledged in economics in the 1950s by Isard, yet it was not until the 1980s that there was any concerted effort to provide the necessary methodological framework to model spatial effects. There was a rapid expansion of the theoretical econometric literature in the 1990s and since then there have been a growing number of applied spatial econometric articles. Today, it is quite common to consider the possibility of spatial autocorrelation when performing econometric analysis. Ignoring it can lead, at best, to biased standard errors and more likely to biased and inconsistent parameter estimates.

Multivariate datasets are increasingly available in economics yet the tools to model them have been lacking from the literature. This thesis aims to partially fill this gap by developing a flexible multivariate spatial model that incorporates asymmetric cross-correlation parameters. This model will be demonstrated through an application to a multivariate dataset of road safety performance indicators for London indexed over census output areas. In doing so, it is hoped that the uncertainty in the individual rankings will be reduced and that this will enable differential performance to be identified.

This thesis is structured as follows. The next chapter provides a brief review of the necessary theory to model univariate spatial processes. Chapter 3 then looks at the existing literature on multivariate spatial models demonstrating a clear need for this current research into flexible multivariate models. In chapter 4 the use of performance indicators in public sector management as well as more specificaly within road safety performance management is discussed. The rather sparse literature in this area is reviewed. In particular, the main problems with the current system of crude headline indicators are presented, and a motivation for the application is provided. The main contribution to the theoretical literature is chapter 5, which presents an innovative flexible multivariate conditional autoregression model. The performance on the model in comparison with the existing approaches is considered through a simulation study in chapter 6. Chapter 7 is the principal empirical contribution of this thesis and is an application of the flexible multivariate conditional autoregression model to a set of multiple road safety performance indicators. Chapter 8 discusses the outcomes of the thesis and chapter 9 concludes.

CHAPTER 2

A BRIEF REVIEW OF UNIVARIATE SPATIAL MODELS

2.1 Introduction

This thesis considers multivariate spatial autoregression models. As a point of departure, this chapter reviews two well established univariate spatial autoregressive models. There are two forms of effects considered: spatial dependence and spatial heterogeneity. Spatial dependence is a particular case of cross-sectional dependence, in the sense that the structure of the correlation or covariance between observations at different locations is derived from the specific ordering, which in turn is determined by the relative position of the observations in geographic space (or, in more general terms, network space)[1]. While similar in concept to correlation in the time domain, spatial dependence requires specialised methods that are not straightforward extensions of the time-series techniques to the spatial domain. Therefore, this chapter reviews the essential univariate theory for spatial autoregressions. Spatial heterogeneity is a special case of observed (or unobserved) cross-sectional heterogeneity which is a well studied problem in standard econometrics. Unlike spatial dependence, tackling spatial heterogeneity does not require specialized tools and as such will not be explicitly

[1] See for instance Lee (2007) or Lin (2005), for examples of spatial autoregression models applied to the social interaction literature and Manski (1993) infamous *reflection problem*

considered in this thesis.

With geographically referenced data over a regular or irregular lattice it is common to incorporate the spatial dependence into the covariance structure either explicitly or implicitly via an autoregressive model. When the geographic location of the observations is known it is common to assume that observations at sites near each other may have a similar value on the omitted variables in the regression causing the error terms to be serially autocorrelated. Once a neighbourhood structure is determined (usually by the econometrician with reference to the actual lattice), models *resembling* autoregressive models from time-series econometrics are formed. Two popular models in the spatial literature are the conditionally autoregressive model (CAR) favoured by statisictians and the simultaneously autoregressive model (SAR) which dominates the econometric and regional science literature.

These models were originally developed for analysis on a regular (doubly infinite) lattice beginning with Whittle (1954) for the SAR model and Besag (1974) for the CAR model. As discussed in Cressie (1993) when used for modelling a doubly infinite regular lattice, these models are analgous to the well understood stationary autoregressive time series model defined on the integers. That is, the CAR is analgous in its Markov property, and the SAR specification in its functional form. Spatial autoregressive models were first deployed in economics to analyze data on regular lattices, which may arise for instance, in agricultural field trial experiments. In practice, particularly in economics where data are generated in non-experimental settings, these models are usually applied to irregular lattices and the effect of the neighbourhood structure and the spatial correlation parameter have on the implied covariance is not well understood. When applied to irregular lattices several authors have pointed out that the models exhibit some undesirable and often unexpected properties (for instance Besag & Kooperberg (1995) and Wall (2004)).

This chapter focuses on the two principal autoregressive models for a vector of observations $\boldsymbol{y}$ on a univariate random variable, Y, recorded for a set of locations, $\boldsymbol{s}$. It leans on the excellent textbook presentations of Anselin (1988), Haining (1991), Cressie (1993) and Rue & Held (2005) summarising and simplifying the methodological aspects required to understand the multivariate approach developed in this thesis. As such it can be omitted by readers familiar with these texts or confident in the specification, estimation and testing of these particular univariate spatial models. To begin consider how the econometrician codifies the geographical relationships in the data by specifying neighbourhoods.

2.2 Neighbourhoods

The concept of a neighbourhood is central to the study of spatial dependence. Consider a spatial location $\boldsymbol{s}$ and a random variable Y associated with each location. This location may be an actual geographical location, but it may also refer to a time of occurrence in panel/longitudinal data, or a grouping mechanism in a subsampling or repeated-measures study. For example, in a spatial problem, there may be $(s_i) = (u_i, v_i)$, where u_i is longitude or northing and v_i is latitude or easting; in a multivariate time-series application there may be $(s_i) = (k, t_k(j))$, where k indexes the variable and $t_k(j)$ is the time at which the j^{th} observation of the k^{th} variable is obtained. Letting $\boldsymbol{s}$ vary over the index set $\mathrm{D} \subset \mathbb{R}^{\mathrm{d}}$ generates a random field $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in \mathrm{D}\}$. For lattice data, D is commonly assumed to be given a finite (or countably infinite) collection of points. Lattices may be regular (like a grid) or more commonly in applied situations, irregular, such as census areas (ouput areas, wards, counties), regions, fields, etc. This thesis will consider irregular lattices.

An integral feature of spatial autoregressive models is the specification of neighbourhoods. For each site ($\boldsymbol{s}_i$) on the lattice, D, a neighbourhood is a collection of sites that are in spatial proximity. Many schemes exist for defining sites as neighbours (see Anselin (1988) for a discussion) but in this thesis two sites that share a common border (adjacency) will be considered neighbours.

The primary concept here is a *proximity matrix*, $\boldsymbol{W}$. The entries w_{ij} in $\boldsymbol{W}$ spatially connect locations i and j in $\boldsymbol{s}$. Typically $w_{ii} = 0$ and $w_{ij} = 1$ if and only if i and j share a common boundary. As a result, $\boldsymbol{W}$ is usually symmetric. However, it is common to standardize the w_{ij}'s by $\Sigma_j w_{ij} = w_{i+}$. If $\tilde{\boldsymbol{W}}$ has entries $\tilde{w}_{ij} = w_{ij}/w_{i+}$, then evidently $\tilde{\boldsymbol{W}}$ is row stochastic, i.e. $\tilde{\boldsymbol{W}}\mathbf{1} = \mathbf{1}$, but now $\tilde{\boldsymbol{W}}$ need not be symmetric. The entries in $\boldsymbol{W}$ can be viewed as weights. The specification of neighbourhoods via a spatial proximity matrix is how spatial dependence is considered. A key concept embodied by the structure of the spatial proximity matrix is conditional independence which is considered in the next section.

2.3 Time Series Autoregressions

2.3.1 Conditional Independence

Much of the methodological material presented in this thesis depends upon conditional independence as implied in the concept of a Markov chain, or more generally in the spatial case, a Markov random field. Conditional independence is a powerful concept and its discussion is motivated by reconsidering the time-series autoregressive models familiar to applied econometricians. Let $\boldsymbol{y} = (y_1, y_2, y_3)'$ be a random vector, then y_1 and y_2 are conditionally independent given y_3, if for a known value of y_3, discovering the value of y_2 provides no information about the distribution of y_1. Under conditional independence, the joint density $\pi(\boldsymbol{y})$ must be:

$$\pi(\boldsymbol{y}) = \pi(y_1 \mid y_3)\pi(y_2 \mid y_3)\pi(y_3) \tag{2.3.1}$$

which is a simplication of the more general representation

$$\pi(\boldsymbol{y}) = \pi(y_1 \mid y_2, y_3)\pi(y_2 \mid y_3)\pi(y_3). \tag{2.3.2}$$

The conditional independence assumption implies that $\pi(y_1 \mid y_2, y_3)$ can be simplified to $\pi(y_1 \mid y_3)$ because it provides no additional information.

2.3.2 AR(1) Processes

Consider a simple autoregressive time series of order one with white noise errors and with the autoregressive parameter $|\rho| < 1$ so as to ensure covariance stationarity. Any standard econometric textbook such as Greene (2003) or Davidson & MacKinnon (2004) will cover this model in detail. This 'textbook' autoregressive model is usually represented

$$y_t = \rho y_{t-1} + \boldsymbol{\epsilon}_t \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}_{iid}(0, 1), \qquad |\rho| < 1 \tag{2.3.3}$$

where the subscript, t, indexes time. Assumptions about conditional independence are not explicit in (2.3.3) but show up more clearly if expressed (2.3.3) in conditional form

$$y_t \mid y_1, ..., y_{t-1} \sim \mathcal{N}(\rho y_{t-1}, 1) \tag{2.3.4}$$

for $t = 2, ..., n$. In this model y_s and y_t with $1 \leq s \leq t \leq n$ are conditionally independent given $\{y_{s+1}, ..., y_{t-1}\}$ if $t - s > 1$.

In addition to (2.3.4), assume that the marginal distribution of y_1 is normal with mean zero and variance $1/(1-\rho^2)$, which is simply the stationary distribution of this AR(1) process. The joint density of $\boldsymbol{y}$ is then

$$\begin{aligned} \pi(\boldsymbol{y}) &= \pi(y_1)\pi(y_2 \mid y_1)...\pi(y_n \mid y_{n-1}) && (2.3.5) \\ &= \frac{1}{(2\pi)^{n/2}}\det(\boldsymbol{\Sigma}^{-1})^{1/2}\exp(-\frac{1}{2}\boldsymbol{y}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}). && (2.3.6) \end{aligned}$$

where $\boldsymbol{\Sigma}^{-1}$ is the *precision* matrix[2] or the inverse variance-covariance matrix which is the tridiagonal matrix

$$\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} 1 & -\rho & & & & \\ -\rho & 1+\rho^2 & -\rho & & & \\ & \ddots & \ddots & \ddots & & \\ & & & -\rho & 1+\rho^2 & -\rho \\ & & & & -\rho & 1 \end{pmatrix} \qquad (2.3.7)$$

with zero entries outside the diagonal and the first off-diagonals. The conditional independence assumptions impose certain restrictions on the precision matrix (Dempster 1972). The tridiagonal form is due to the fact that y_i and y_j are conditionally independent for $|y-j|>1$, given the rest. If $\boldsymbol{\Sigma}_{ij}^{-1}=0$ for $i\neq j$, then y_i and y_j are conditionally independent given the other observations $\{y_k : k\neq i \text{ and } k\neq j\}$ and vice versa. The simple relationship between conditional independence and the sparse nature of the precision matrix is not apparent from investigating the variance-

[2]The precision matrix is used in many Bayesian settings because it offers computational advantages as the conditional independence assumptions represented in it are readily exploited by Markov chain Monte Carlo methods. Therefore MRF methods can be employed without the need for computationally expensive matrix inversion.

covariance matrix, $\boldsymbol{\Sigma}$ which is completely dense with entries

$$\boldsymbol{\sigma}_{ij} = \frac{1}{1-\rho^2}\rho^{|i-j|}. \tag{2.3.8}$$

It is difficult to derive the conditional assumption from such a dense matrix. For example, for a sample of $n = 5$ the variance-covariance matrix would be:

$$\boldsymbol{\Sigma} = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \rho^4 \\ \rho & 1 & \rho & \rho^2 & \rho^3 \\ \rho^2 & \rho & 1 & \rho & \rho^2 \\ \rho^3 & \rho^2 & \rho & 1 & \rho \\ \rho^4 & \rho^3 & \rho^2 & \rho & 1 \end{pmatrix} \tag{2.3.9}$$

The enteries in $\boldsymbol{\Sigma}$, above, only give information directly about the marginal dependence — in this AR(1) model the observations y_s and y_t are marginally dependent whenever $\rho \neq 0$.

Simplied forms are obtained not only through the *directed* conditionals seen in (2.3.4) but also through the full or *undirected* conditionals $\{\pi(y_t \mid \boldsymbol{y}_{-t})\}$, where $\boldsymbol{y}_{-t}$ denotes all observations in the vector $\boldsymbol{y}$ excluding y_t. Returning to the AR(1) example,

$$y_t \mid \boldsymbol{y}_{-t} \sim \begin{cases} \mathcal{N}(\rho y_{t+1}, 1) & t = 1, \\ \mathcal{N}\left(\frac{\rho}{1+\rho^2}(y_{t-1} + y_{t+1}), \frac{\rho}{1+\rho^2}\right) & 1 < t < n, \\ \mathcal{N}(\rho y_{n-1}, 1) & t = n, \end{cases} \tag{2.3.10}$$

so, in general, y_t depends both on the previous (y_{t-1}) and the future (y_{t+1}) observations. Equation (B.1.1) illustrates an important alternative model specification through the full conditional distributions $\pi(y_t \mid \boldsymbol{y}_{-t})$ for $t = 1, ..., n$. As will be seen in more detail in section 2.4.1 beginning with the full conditional distributions allows an alternative, yet equivalent, specification for the joint density of $\boldsymbol{y}$ to be derived. This is not as obvious as using the directed conditional densities in 2.3.4 to form the joint densities as a product of these densities times the marginal density of y_1, which necessitates the level of detail provided in section 2.4.1.

2.4 Markov random fields

A Markov random field is the key to moving from the full conditional distributions presented in Section 2.3.2 to a joint distribution for $\boldsymbol{y}$. A *Markov random field* (MRF) is a name given to a natural generalisation of the well known concept of a Markov chain. It arrises by considering the chain itself as a simple graph and ignoring the directionality implied by 'time'. A Markov chain can then be seen as a chain graph of stochastic variables, each of which has the property of being independent of all the others (the future and the past) given its two neighbours. Using this interpretation of a Markov chain, a Markov random field is the same thing but instead of the chain graph we allow *any* graph to determine the relationship between the variables. Rozanov (1982) presents a very general treatment of MRFs and Rue & Held (2005) provide a thorough treatment of Gaussian Markov Random Field models with applications that include state-space models and time-series analysis.

A MRF is therefore a stochastic process Y indexed over some countable subset of $\mathbb{R}^k$. To any such MRF corresponds an acyclic algebraic graph with undirected edges[3](Whittaker 1990). This section is concerned with the construction of a joint distribution for $\boldsymbol{y}$, given a complete set of full (univariate) conditional distributions.

[3]An undirected graph $\mathcal{G}$ is an ordered pair $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ that is subject to the following conditions: (i) $\mathcal{V}$ is a set, whose elements are called vertices or nodes, and (ii) $\mathcal{E}$ is a multiset of unordered pairs of vertices (not necessarily distinct), called edges or lines. Bying acyclic there is no single cycle through all the nodes of the graph. See, for example, Lauritzen (1996) for further details.

As Besag (1974) discussed, these conditional distributions are the building blocks of a MRF.

2.4.1 Brook's Lemma and the Hammersley-Clifford Theorem

A useful technical result for constructing the joint distribution of the $\boldsymbol{y}$ is Brook's Lemma (Brook 1964) as described in the seminal paper by Besag (1974) on Conditionally Autoregressive (CAR) models. A technical presentation of this theorem is reserved to appendix B.1 and it is discussed only intuitively here. It is clear that given the joint distribution $\pi(y_1, ..., y_n)$, the so-called *full conditional* distributions, $\pi(y_i \mid \boldsymbol{y}_{-i})$, are uniquely determined, as demonstrated in section 2.3.2. Brook (1964) demonstrates how to retrieve the unique joint distribution given these full conditionals. It should be obvious that an arbitrary set of full conditional distributions can not simply be written down and it asserted that they uniquely determine the joint distribution. Textbooks such as Banerjee et al. (2004) and Cressie (1993) are replete with examples of contradicting and incompatible full conditionals. This thesis does not propose to examine these conditions for compatibility in much detail, although there has been considerable work in this area (see for example Arnold & Straus (1991) and references therein). Typically these conditions reduce to requiring that the precision matrix,$\boldsymbol{\Sigma}^{-1}$, is a symmetric and positive-definite matrix. As noted in Rue & Held (2005) an all too common approach to ensuring the positive-definite condition is met is to force the precision matrix to be *diagonal dominant.*[4] This is a sufficient but not necessary condition for positive-definiteness, which will be discussed in section 2.5.1.

[4]Meaning that for each row (or column) of $\boldsymbol{\Sigma}^{-1}$ the diagonal entry is larger than the sum of the absolute off-diagonal entries.

Although Brook (1964) illustrates how to create the joint density from the full conditionals up to a constant of proportionality, it is often cumbersome for a large number of geographical areas. Instead, it may be preferable to model the n full conditionals. In the context of a spatial model, it is expected that the full conditional distribution for y_i should depend only upon the neighbours of site i. Using the definition of a neighbourhood presented in section 2.2 let ∂_i represent the set of site $i's$ neighbours. Then a set of full conditionals of the following form are obtained

$$\pi(y_i \mid \boldsymbol{y}_{-i}) = \pi(y_i \mid \{y_j : j \in \partial_i\}) \tag{2.4.1}$$

All that is required is to be assured that (2.4.1) specifies a joint distribution i.e. if a Gibbs sampler (Geman & Geman 1984) is implemented to simulate realisations from the joint distribution that there actually exists a unique stationary joint distribution for this sampler. This concept of using a *local* specification such as (2.4.1) to determine a joint or *global* distribution should be a familiar notion to Bayesians and is called a *Markov random field*. The literature on this topic is voluminous and there is no attempt to cover it here, although a good starting place is Geman & Geman (1984). Section 5.8.2 provides a very brief overview of the Gibbs sampler in terms of implementing the model developed in chapter 5. Additionally, Gelfand & Smith (1990) provide a good introduction to the topic and Kaiser & Cressie (2000) provide a current perspective with numerous references.

A few important definitions are required, and a starting point is to determine a *clique*: a set of sites (or indices) such that all elements in the set are neighbours of all the other elements. Adopting the terminology from physics a *potential function* of order k is a function of k arguments that is exchangeable in those arguments. The arguments of the potential would be the values of the variables associated with the sites for a clique of size k. For continuous Y_i, a typical potential when $k = 2$ is $Y_i Y_j$, if i and j are a clique of size 2. If there were only cliques of size 1 then data are independent. Another important definition is that of a *Gibbs distribution*: $\pi(y_i, ..., y_n)$ is a Gibbs distribution if it is a function of the Y_i only through potentials on cliques. For example,

$$\pi(y_1, ..., y_n) \propto \exp\{\gamma \sum_k \sum_{\boldsymbol{\alpha} \in \mathcal{M}_k} \phi^{(k)}(y_{\alpha_1}, y_{\alpha_2}, ..., y_{\alpha_k})\}. \tag{2.4.2}$$

where $\phi^{(k)}$ is a potential of order k, $\mathcal{M}_k$ is the collection of all subsets of size k from $\{1, 2, ..., n\}$, $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_k)'$ indexes this set and $\gamma > 0$ is a scale parameter.

Informally then, the unpublished *Hammersley-Clifford Theorem* (see for example Clifford 1990) states that if there exists an MRF as per equation (2.4.1) then the unique joint distribution defined by it is a Gibbs distribution (see Besag 1974, for an exposition). Again, a formal treatment of this theorem is provided in appendix B.2 for interested readers. The result of the Hammersley-Clifford Theorem means that the MRF model can be written in the form of equation (2.4.2) with all of the ‘action’ coming in the form of potentials on cliques. The inverse was proven by Geman & Geman (1984); beginning with a Gibbs distribution such as (2.4.2) a MRF is determined. The significance of this was that simply sampling from the related Gibbs distribution provides samples from the related MRF.[5] MRFs that are Gaussian form a class of model introduced by Besag (1974) and labelled Conditional Autoregression (CAR) models in the literature. The CAR model, presented in the next section, forms the basis of the multivariate approach developed in this thesis.

[5] Hence Geman & Geman (1984) coined the term “Gibbs Sampler” to describe their method.

2.5 Conditional Autoregressions

2.5.1 Theory

Extensive literature on the origins, derivation and properties of this model can be found in Cressie (1993). Therefore only a brief summary is presented here. Assuming Y_i) is univariate, Besag (1974) shows that, under given consistency conditions (e.g. positivity), the conditional distributions

$$\pi(y_i \mid \boldsymbol{y}_{-i}) = \pi(y_i \mid \{y_j : j \in \partial_i\}) \tag{2.5.1}$$

can be used to determine the joint distribution

$$\pi(y_1, ..., y_n) \tag{2.5.2}$$

which is called a MRF. MRFs that are Gaussian define a class of models that were described in earlier sections as conditional autoregressive (CAR) models. Assuming the conditional distributions in (2.5.1) are Gaussian, the i^{th} distribution ($i = 1, ..., n$) is specified through

$$\mathrm{E}[y_i|\{y_j : j \in \partial_i\}] = \mu + \sum_{j \in \partial_i} b_{ij}(y_j - \mu_j), \tag{2.5.3}$$

$$\mathrm{Var}[y_i|\{y_j : j \in \partial_i\}] = \tau_i^2. \tag{2.5.4}$$

These full conditionals are compatible so via Brook's theorem (see Section 2.4.1) the joint distribution is

$$\pi(y_1, ..., y_n) \propto \exp\{-\frac{1}{2}(\boldsymbol{y} - \mu)'\boldsymbol{\Gamma}^{-1}(\boldsymbol{I} - \boldsymbol{B})(\boldsymbol{y} - \mu)\}, \tag{2.5.5}$$

with $\boldsymbol{B} = \{b_{ij}\}$ capturing the spatial dependence and $\boldsymbol{\Gamma} = \text{diag}(\tau_i^2)$. Equation (2.5.5) yields a multivariate joint distribution for $\boldsymbol{y}$ as

$$\boldsymbol{y} \sim \mathcal{N}_n(\boldsymbol{\mu}, (\boldsymbol{I} - \boldsymbol{B})^{-1}\boldsymbol{\Gamma}). \tag{2.5.6}$$

In the presence of covariates, $\boldsymbol{\mu}$ is reparameterized as $\boldsymbol{\mu} = \boldsymbol{X}'\boldsymbol{\beta}$ where $\boldsymbol{X}$ is an n x q matrix of known covariates and $\boldsymbol{\beta}$ is a q x 1 vector of regression parameters. Of course, the joint distribution in (2.5.6) must be well defined: the elements of $\boldsymbol{B}$ must be chosen so that $(\boldsymbol{I} - \boldsymbol{B})^{-1}\boldsymbol{\Gamma}$ is a symmetric, positive-definite matrix. To ensure symmetry the following is required

$$\frac{b_{ij}}{\tau_i^2} = \frac{b_{ji}}{\tau_j^2} \qquad \forall i, j. \tag{2.5.7}$$

From (2.5.7), $\boldsymbol{B}$ is not symmetric. However, assuming a symmetric proximity matrix, $\boldsymbol{W}$ (see Section 2.2) set $b_{ij} = w_{ij}/w_{i+}$ and $\tau_i^2 = \tau^2/w_{i+}$. Now (2.5.7 is symmetric and (2.5.3) becomes $\pi(y_i \mid y_j, j \neq i) \sim \mathcal{N}(\sum_j w_{ij}y_j/w_{i+}, \tau^2/w_{i+})$. Also (2.5.5) now becomes

$$\pi(y_1, ..., y_n) \propto \exp\{-\frac{1}{2\tau^2}(\boldsymbol{y} - \mu)'(\mathbf{D}_w - \boldsymbol{W})(\boldsymbol{y} - \mu)\}, \tag{2.5.8}$$

with $\mathbf{D}_w = \text{diag}(w_{i+})$. Unfortunately now $(\mathbf{D}_w - \boldsymbol{W})\mathbf{1} = \mathbf{0}$, i.e. the precision matrix is now singular so that covariance does not exist, hence distribution (2.5.8) is improper.[6] Expression (2.5.8) can be expressed in pairwise form as

$$\pi(y_1, ..., y_n) \propto \exp\{-\frac{1}{2\tau^2}\sum_{i \neq j} w_{ij}(y_i - y_j)^2\} \tag{2.5.9}$$

[6]When the precision matrix is singular there is a non-integrable density function i.e. there are too many variables and a contraint is required. Where as when the variance-covariance matrix is singular there is no density function but a proper distribution residing on a lower dimensional space, i.e. there are too few variables.

which more clearly shows the impropriety of the joint distribution (undefined mean and infinite variance) since any constant can be added to the observations y_i without affecting expression (2.5.9). Introduced by Besag et al. (1991) this specification of the CAR model is usually termed the *intrinsic autoregressive* model (IAR).

As a constraint cannot be imposed on (2.5.9) in order to restore propriety, model (2.5.9) cannot be used as a model for data. Nevertheless, a proper posterior distribution – in the sense of a Bayesian hierarchical model – will result when used (2.5.9) is used as a prior distribution and subjected to the (usual) requirement of a proper hyperprior distribution on the variance components. Thus, as intended by Besag et al. (1991) this can be used as a prior distribution for the spatial effects. That is, it could be employed as a prior distribution for spatial random effects in the second (or third) stage of a Bayesian hierarchical model. Section 2.5.2 provides further details.[7]

There is an obvious alternative solution to the problem of a singular precision matrix in expression (2.5.8). Redefine $\mathbf{\Sigma}^{-1} = (\mathbf{D}_w - \rho \boldsymbol{W})$ choosing ρ to ensure that $\mathbf{\Sigma}^{-1}$ is positive definite. This requires that $\rho \in (1/\lambda_{(1)}, 1/\lambda_{(n)})$, where $\lambda_{(1)} < \lambda_{(2)} < ... < \lambda_{(n)}$ are the ordered eigenvalues of $\mathbf{D}_w^{-1/2}\boldsymbol{W}\mathbf{D}_w^{-1/2}$ (refer to Cressie 1993, p. 471)[8]. A yet simpler constraint on the precision matrix is if the proximity matrix, $\boldsymbol{W}$ is reparameterized as the row-stochastic or row-standardized counterpart, $\tilde{\boldsymbol{W}} \equiv \text{diag}(1/w_{i+})\boldsymbol{W}$, introduced in section 2.2. $\tilde{\boldsymbol{W}}$ is no longer symmetric. Rewritting $\mathbf{\Sigma}^{-1} = \boldsymbol{M}^{-1}(\boldsymbol{I} - \alpha\tilde{\boldsymbol{W}})$ where $\boldsymbol{M}$ is diagonal. Then $(\boldsymbol{I} - \alpha\tilde{\boldsymbol{W}})$ is non-singular as long as $|\alpha| < 1$. Banerjee et al. (2004) show that $\mathbf{\Sigma}^{-1}$ is diagonal dominant and symmetric

[7]Gelman et al. (2003) provides a gentle yet complete introduction to the Bayesian approach for the unfamiliar. A very general introduction to Bayesian econometrics is provided in Poirier & Tobias (2006) and standard texts include Koop (2003) or Lancaster (2004).

[8]The appendices provide a brief review of the linear algebra necessary to prove these results for the unfamiliar reader

which makes it a positive-definite matrix. This forces $\alpha \in (-1, 1)$ irrespective of the neighbourhood structure chosen for $\boldsymbol{W}$ when specification (2.5.8) is used. The case of $\alpha = 0$ has immediate interpretation as conditional independence. There is also an intuitive interpretation of the conditional mean $\mathrm{E}(y_i \mid \boldsymbol{y}_{-i})$ as a weighted average of the random effects for all neighbours, but it does cause the conditional variance to be non-constant across areas (i.e. it induces heteroscedasticity). Further, interpretation of α is not straightforward – see Sun et al. (2000) for an interpretation of α as a spatial shrinkage factor – and a value close to the maximum $(1/\lambda_{(max)})$ is needed to reflect even moderate spatial dependence (Besag & Kooperberg 1995). This is essentially what Besag et al. (1991) do in the intrinsic autoregression presented earlier; in effect they force α to its upper limit of 1.

The CAR model is reconsidered in the next chapter when the current literature on multivariate spatial models is reviewed. In this section the specification of a global or joint distribution based solely upon a local specification through the n full conditionals was considered. Given its prominence in the Bayesian literature on MCMC sampling this should not be controversial material. The main consideration when constructing a CAR is ensuring that the precision matrix is non-singular. A number of ways to impose this constraint were presented. Typically, the CAR model will not be used as a model for data, but instead as a second stage prior specification for the spatial random effects. How to implement the CAR model as a prior in a hierarchical model is considered in the next section.

2.5.2 Hierarchical Modelling for Non-Gaussian Data

When using the CAR specification to model the data directly, the dependent variable will often not be normally distributed. Common extensions to the Gaussian

CAR specification of Besag (1974) include models for binary data, often called the *autologistic* CAR (see Banerjee et al. 2004, chap. 3). As Section 2.5.1 discussed, the CAR model is often used as a prior for spatial random effects in a hierarchical (Bayesian) setting rather than as a model for the data. In this situation, the Instrinsic Autoregression (IAR) model presented in Equation (2.5.9) is suitable as it leads to a proper posterior distribution despite being improper itself. This hierarchical framework is particularly useful when the data are counts because the Generalized Linear Modelling (GLM) framework of Nelder & Wedderburn (1972) can be employed.[9] Including a spatial random effect term changes the GLM into a GLMM or Generalized Linear Mixed Model — see, for example, Breslow & Clayton (1993), Clayton & Kaldor (1987) and Besag et al. (1991).

GLMMs are appropriate for accommodating the overdispersion (extra-Poisson variation) in count data, for modelling the dependence between dependent variables in multilevel and multivariate analyses and for producing shrinkage estimators in multi-parameter problems. It is not surprising therefore, that the most common application of a CAR prior in the GLMM regression framework is within epidemiological disease mapping. There, the spatial random effects model is modelling the underlying risk surface which is assumed to come from some common distribution. See Elliott et al. (2000) for a review.

Consider the following GLM set-up for a discrete random variable Y_i such as the number of children killed or seriously injured in each area, i:

$$Y_i \mid \theta_i \sim \text{Poi}(E_i e^{\theta_i}), \tag{2.5.10}$$

$$\theta_i = \alpha + \boldsymbol{X\beta} + \psi_i + \phi_i + \epsilon. \tag{2.5.11}$$

[9] A standard reference on GLMs is McCullagh & Nelder (1989).

Equation (2.5.10) represents the first two stages of a hierarchical Poisson GLM with a matrix of explanatory variables, $\boldsymbol{X}$, and coefficients $\boldsymbol{\beta}$. The explanatory variables are area level and although interpretation must be done carefully to avoid ecological bias it is hoped that they will explain some of the spatial variation in the observed counts, Y_i. A common procedure when modelling count data is to standardize the data by the expected number of counts in each area, E_i. For instance, the number of accidents in a particular area will be dependent upon the local population and thus perhaps a more interesting model investigates any difference from the expected number i.e. increased road traffic risk. Letting Y_i be the observed number of accidents in site $\boldsymbol{s}_i$ and E_i to be the expected number. There are two methods of standardization employed, internal standardization and external standardization, and these will be discussed in more detail in chapter 7. The ψ_i in (2.5.10) capture the global or region-wide heterogeneity (i.e. they capture any global extra-Poisson variation) and is usually modelled via an exchangeable normal prior

$$\psi_i \sim \mathcal{N}(0, \tau_h^2). \qquad (2.5.12)$$

The ϕ_i in (2.5.10) are the spatial random effects and capture the extra-Poisson variation that occurs locally (i.e. through neighbourhoods). Although alternatives exist, the most common prior for this spatial random effect term (ϕ_i) is the Gaussian CAR presented in (2.5.3) or the IAR presented in (2.5.9). Thus $\phi_i \sim CAR(\tau_c^2)$ where τ_c^{-1} is the precision of the CAR model. The hierarchical framework is completed by specifying priors for the precision hyperparameters in the priors for ψ_i and ϕ_i. Alternatives to the CAR and IAR priors include a jointly specified multivariate model ($\phi_i \sim \mathcal{N}_n(\mu, \boldsymbol{\Sigma})$) applied to environmental data in Diggle et al. (1998) as well as semiparametric specifications such as the mixture model of Green & Richardson (2002) and the partition model of Knorr-Held & Raßer (2000).

Due to the MRF formulation of the CAR model (see Section 2.4.1) hierarchical models employing the CAR or IAR prior are computationally convenient. The Gibbs sampler (Geman & Geman 1984) for simulating from the posterior distribution of ψ_i and ϕ_i operates by successively sampling from the full conditional distribution of each parameter (i.e. the distribution of each parameter in the model given the data and the other parameters). Therefore besides not requiring any computationally expensive matrix inversion, there is no need for the joint distribution of ϕ_i at all. The full conditional of ϕ_i is

$$\pi(\phi_i \mid \boldsymbol{\phi}_{-i}, \boldsymbol{\psi}, \boldsymbol{\beta}, \boldsymbol{y}) \propto \text{Poi}(y_i \mid E_i e^{\boldsymbol{X}'\boldsymbol{\beta}+\psi_i+\phi_i}) \mathcal{N}(\phi_i \mid \bar{\phi}_i, \tau_c^2). \tag{2.5.13}$$

As noted in section 2.5.1, there are some difficulties in implementing the CAR specification. For instance, the impropriety of the IAR version of the CAR model in (2.5.9) as introduced by Besag et al. (1991) was discussed. The IAR is improper in that it does not integrate to one and hence it is not a legitimate probability distribution i.e. the precision matrix $\boldsymbol{\Sigma}^{-1} = (\mathbf{D}_w - \boldsymbol{W})$ is singular. Section 2.5.1 presented options for ensuring $\boldsymbol{\Sigma}^{-1}$ is non-singular including asserting diagonal dominance. When using the IAR model as a prior for the random spatial effects in (2.5.10) it does not matter that (2.5.9) is improper because the posterior in (2.5.10) will be proper. However, in order to identify the intercept (α) in (2.5.10) the following constraint $\sum_{i=1}^{N} \phi_i = 0$ is required. One way to implement this constraint is to recentre the vector ϕ_i around its own mean after each iteration of the Gibbs sampler.

The model is also sensitive to the choice of hyperpriors for τ_h^2 and τ_c^2. These precision terms control the amount of extra-Poisson variation in the global term, ψ_i, and in the clustering or spatial random effect term, ϕ_i. If they are set arbitrarily vague then the model becomes *unidentified* because only univariate Y_i is observed whilst the model attempts to estimate two random effects for each i. One choice of hyperprior would be the gamma distribution as it is conjugate. Such computational issues will be discussed in more detail when the models are fitted to real data in chapters 2.7 and 7, but note here that although CAR models are computationally convenient that care must be taken when specifying the hierarchical components to ensure identifiability and good convergence of the Gibbs sampler.

Implementing Gibbs sampling or alternative MCMC strategies requires repeated evaluation of the multiple undirected conditional densities. Given a spatial random effect, GLM hierarchical models such as the one presented in (2.5.10) necessitates repeated evaluation of the likelihood and/or conditional densities requiring calculation of the determinant and quadratic forms of the precision matrix. Even using Cholesky decomposition (see Rao & Rao 1998, p. 173) and taking advantage of the sparse nature of the precision matrices (see for instance Rue 2001), with large n computation of n x n matrices can become unstable and convergence can be difficult to achieve. When this problem is extended to a p dimensional multivariate dataset, evaluation of np x np matrices is required. Therefore, careful consideration will be given to efficient and stable sampling alogorithms when the multivariate CAR model is developed in Section 5.

Despite their wide application in statistics, biostatistics, environmetrics and epidemiology, conditional autoregressions have been overlooked within econometrics in favour of the Simultaneous Autoregression (SAR) models, which are presented in the next section. In fact, to date, the only application in the econometric literature that was found is Parent & LeSage (2008) which adopts a CAR model for knowledge spillovers. Section 2.7 attempts to explain this lack of interest in the CAR model among econometricians whilst presenting a case for the superiority of the CAR model in certain situations.

2.6 Simultaneous Autoregressions

When modelling areal data, the Simultaneous Autoregression (SAR) of Whittle (1954) is the main alternative to the CAR presented in Section 2.5.1. A brief introduction to Simultaneous Autoregession (SAR) models can be found in Anselin (2006) and Anselin (1988) provides a comprehensive treatment of them, while Wall (2004) and Cressie (1993) compare the SAR and CAR specifications. As this thesis concentrates solely on CAR models the SAR model is introduced only briefly as a comparison. Consider expression (**??**) again. Instead of letting $\boldsymbol{y}$ induce a distribution for $\boldsymbol{\epsilon}$ let $\boldsymbol{\epsilon}$ induce a distribution for $\boldsymbol{y}$. Adopting the usual time series assumption of independent innovations for $\boldsymbol{\epsilon}_i$ and assuming $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \tilde{\boldsymbol{\Gamma}})$ where $\tilde{\boldsymbol{\Gamma}} = \text{diag}(\sigma_i^2)$.[10] Analogous to (2.5.3) the SAR can be written as:

$$\text{E}[y(\boldsymbol{s}_i)] = \mu_i + \sum_j b_{ij}(y_i - \mu_j) + \boldsymbol{\epsilon}_i, \tag{2.6.1}$$

for $i = 1, ..., n$ and $\boldsymbol{\epsilon}_i \sim \mathcal{N}(0, \sigma_i^2)$. Providing that $(\boldsymbol{I} - \boldsymbol{B})$ is full rank the following joint distribution is obtained

$$\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu}(\boldsymbol{I} - \boldsymbol{B})^{-1}\tilde{\boldsymbol{\Gamma}}((\boldsymbol{I} - \boldsymbol{B})^{-1})') \tag{2.6.2}$$

This model is called *simultaneous* because, in general, $\boldsymbol{\epsilon}_i$ is correlated with $\{y_j : j \neq i\}$. To ensure that $(\boldsymbol{I} - \boldsymbol{B})$ is full rank it is possible to redefine $\boldsymbol{B} = \rho \boldsymbol{W}$ where

[10] Importantly, $\tilde{\boldsymbol{\Gamma}}$ is not the same matrix as $\boldsymbol{\Gamma}$ from Section 2.5.1 hence the use of the tilde yet it serves the same purpose in the model. The matrix $\boldsymbol{B}$ in this section may (or may not) be the same as $\boldsymbol{B}$ adopted in Section 2.5.1

$\boldsymbol{W}$ is the proximity matrix introduced in Section 2.2 incorporating elements $w_{ij} = 1$ if areas i and j are neighbours and being 0 otherwise. Again, setting $w_{ii} = 0$. As in the CAR example (see Section 2.5.1) the term $(\boldsymbol{I} - \rho\boldsymbol{W})$ will be non-singular as long as $\rho \in (1/\lambda_{(min)}, 1/\lambda_{(max)})$ where $\lambda_{(min)}$ an $\lambda_{(max)}$ are the smallest and largest of the ordered eigenvalues of $\boldsymbol{W}$. In this form, ρ is commonly called the *spatial autoregression parameter* and therefore $y_i = \rho \sum_j y_j \boldsymbol{I}(j \in \partial_i) + \boldsymbol{\epsilon}_i$. Similarly to Section 2.5.1 the row-stochastic neighbourhood matrix $\tilde{\boldsymbol{W}}$ could be adopted and then $\boldsymbol{B} = \alpha\tilde{\boldsymbol{W}}$. Now, $y_i = \alpha \sum_j y_i \boldsymbol{I}(j \in \partial_i)/w_{i+} + \boldsymbol{\epsilon}_i$ and α is called a *spatial autocorrelation parameter*. Analogous to the CAR situation with row-stochastic, $\tilde{\boldsymbol{W}}, (\boldsymbol{I} - \alpha\tilde{\boldsymbol{W}})$ will be non-singular if $\alpha \in (-1, 1)$ hence the name, autocorrelation parameter. It is perhaps the intuitive interpretation of α that has lead to so many econometric applications of the SAR model with row-stochastic proximity matrix, $\tilde{\boldsymbol{W}}$.

Typically, the SAR model is employed in a regression context in which the residuals $\boldsymbol{U} = \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}$ are assumed to follow a SAR model. Yet considering expression (**??**) again, gives

$$\boldsymbol{y} = \boldsymbol{B}\boldsymbol{y} + (\boldsymbol{I} - \boldsymbol{B})\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{2.6.3}$$

demonstrating that $\boldsymbol{y}$ is modelled through a combination of a traditional OLS regression and a spatial weighting of the neighbours. Expression (2.6.3) does not induce any spatial effects as the errors are independent. As a result, the SAR model cannot be used in conjunction with a Generalized Linear Model (GLM) and is therefore not convenient for modelling discrete data such as the count data considered in this thesis. In fact, no progress has been made in the discrete data field to date.

A popular specification for the SAR model is to incorporate the spatially lagged dependent variable as an explanatory variable. See, for instance Anselin (2006) for details or Kim et al. (2003) for an empirical example. This model is inevitably used with a row-stochastic neighbourhood proximity matrix, $\tilde{\boldsymbol{W}}$ and therefore a spatial autocorrelation parameter, α. The specification is:

$$\boldsymbol{y} = \alpha\tilde{\boldsymbol{W}}\boldsymbol{y} + \boldsymbol{X\beta} + \boldsymbol{\epsilon} \tag{2.6.4}$$

$$\begin{aligned} \boldsymbol{y} &= (\boldsymbol{I} - \alpha\tilde{\boldsymbol{W}})^{-1}\boldsymbol{X\beta} + (\boldsymbol{I} - \alpha\tilde{\boldsymbol{W}}^{-1}\boldsymbol{\epsilon} && (2.6.5) \\ (\boldsymbol{I} - \alpha\tilde{\boldsymbol{W}})^{-1} &= \boldsymbol{I} + \alpha\tilde{\boldsymbol{W}} + \alpha^2\tilde{\boldsymbol{W}}^2 + \alpha^3\tilde{\boldsymbol{W}}^3... && (2.6.6) \\ & && (2.6.7) \end{aligned}$$

where $\boldsymbol{X\beta}$ are a matrix $\boldsymbol{X}$ of explanatory variables with a vector of parameter coefficients, $\boldsymbol{\beta}$. One of the problems with specification (2.6.4) is that the partial derivatives of y_i with respect to the i^{th} observation of the r^{th} variable, x_{ir} involves also x_{jr}. Since the explanatory variables matrix $\boldsymbol{X}$ is transformed by the n x n matrix inverse $\boldsymbol{I} - \alpha\tilde{\boldsymbol{W}}^{-1}$ any change in one variable within one site will affect the dependent variable in other sites throughout the lattice and the conventional interpretation of the regression parameters no longer holds. In particular, $\partial y_i/\partial x_{jr} \neq 0$ and $\partial y_i/\partial x_{ir} \neq \beta_r$. As Kim et al. (2003) discuss, this is frequently ignored in applications of this model. Too often econometricians and applied economists are interpreting the regression coefficients, $\boldsymbol{\beta}$ as if they were from a standard OLS style linear regression model. Thus, even though the SAR model takes on a similar matrix form as the standard (non spatial) linear regresion model it is not true that it has the same ease of interpretation.

Textbooks by Anselin (1988), Anselin & Florax (1995), Florax & De Graaff (2004) and Arbia (2006) present a number of extensions to the standard SAR model presented here. There has been a growing literature adapting the SAR model to meet the demands of econometric applications, yet as Section 2.5.2 discussed there have been very few extensions of the CAR model to suit econometric problems. As a result there remains much research to be done.

2.7 Comparing the CAR and SAR

It is evident from the literature that in terms of spatial autoregressions, many statisticians prefer CAR over SAR. Unlike SAR, CAR can achieve minimum mean squared prediction error (see Cressie 1993, p. 408-410) and maximum entropy in some circumstances (Künsch 1981). Yet despite the reported benefits, the SAR model is still used extensively in many areas particularly economics and regional science (Anselin 2006).

This popularity may be due to the intuitive interpretation of the SAR model as a semiparametric estimator when the row-stochastic proximity matrix, $\tilde{\boldsymbol{W}}$, is used. According to Pace & LeSage (2003) using the CAR specification with a doubly stochastic proximity matrix (i.e. one in which row and column sum to 1) leads to the same intuitive interpretation. Although many authors appeal to the intuitive interpretation of the ρ spatial autoregression or the α spatial autocorrelation parameter in the SAR model, Martellosio (2006) argues using graph theory that this is not actually the case. Furthermore, Wall (2004) examines in detail the correlation structure implied by the CAR and SAR models. This study illuminates a number of alarming pecularities in the models, which were first highlighted in Besag & Kooperberg (1995). In particular, Wall (2004) argues that there is no intuitive interpretation of the implied spatial autocorrelations that result from fitting CAR and SAR models. With the use of a little graph theory Martellosio (2006) explains these pecularities in an appealing manner; the relevant quantity appears to be the length of *walks* between sites on the lattice.

An undirected graph $\mathcal{G}$ is an ordered pair $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ that is subject to the following conditions: (i) V is a set, whose elements are called vertices or nodes, and (ii) E is a multiset of unordered pairs of vertices (not necessarily distinct), called edges or lines. Two vertices are called neighbours if there is an edge between them. A walk from a vertex, u, to a vertex, v is a sequence of vertices $u = v_0, v_1, ..., v_r = v$ such that $(v_{-i}, v_i) \in \mathcal{E}(\mathcal{G})$ for $1 \leqslant i \leqslant r$ where r is the length of the walk. From Martellosio (2006) it is apparent that regardless of whether CAR or SAR is used the covariance between sites i and j are generating functions of the total weight of the walks of the same length between i and j in $\mathcal{G}$. Because there are fewer constraints on the form of the weights matrix, the correlation parameter will always be higher in the case of the SAR model (see p. 17 of Martellosio 2006, for details). It is fair to say (and unfortunate) that practitioners do not completely understand the properties of these spatial autoregressive models (Anselin (2003*a*), Wall (2004), and Martellosio (2006). Yet one thing is certain, the argument in favour of SAR models because of their intuitive interpretability is flawed.

CAR and SAR models are similar, both being spatial generalizations of time-series autoregressions (Brook 1964). As shown in Ripley (1981), the two models are only equivalent in the limiting case of their covariance matrices being equivalent (assuming that the mean has been successfully modelled). Using the CAR covariance matrix from (2.5.6) and the SAR covariance from (2.6.2) this implies:

$$(\boldsymbol{I} - \boldsymbol{B})^{-1}\boldsymbol{\Gamma} = (\boldsymbol{I} - \boldsymbol{B})^{-1}\tilde{\boldsymbol{\Gamma}}((\boldsymbol{I} - \boldsymbol{B})^{-1\prime}) \tag{2.7.1}$$

Cressie (1993) credits Brook (1964) with being the first person to compare the conditional and simultaneous approaches. As already stated in Section 2.5.1, the spatial proximity matrix, $\boldsymbol{W}$, must be symmetric in the CAR specification but not in the SAR alternative. Although this may appear an advantage, the spatial proximity matrix must be carefully specified to ensure identifiability of the spatial dependence parameters. Likelihood computations with both SAR and CAR models are expensive, and methods have to appeal to the sparse nature of the spatial weights matrix and by extension the precision matrix. Estimation of SAR parameters is inconsistent using OLS (Whittle (1954)) hence more sophisticated and often computationally burdensome estimation techniques must be employed. In fact, estimation of SAR model parameters is still an area of active research, and debate over the most appropriate method (e.g. MLE versus GMM) is rife. On the other hand, CAR models can be implemented with MCMC and has advantages when extending the model to multivariate data. Additionally, as muted in Section 2.5.2, the CAR model is available for use as a prior in a hierarchical GLM framework immediately opening up discrete variables to spatial analysis. The same is not the case for the SAR model.

Econometricians have focused almost exclusively on SAR models; research by R. Kelley Pace and James LeSage are the exception. However, despite their fondness for the simultaneous approach many econometricians interpret their results in a conditional expectations structure similar to standard linear regression. If a conditional expectations interpretation is more natural then perhaps the CAR approach should be adopted from the beginning.

2.8 Summary

This chapter provided a brief overview of the principal l univariate spatial models: the Conditional Autoregression of Besag (1974) and the Simultaneous Autoregression of Whittle (1954). The presence of spatial dependence can cause problems for econometric models, both in terms of efficiency and bias. When spatial autocorrelation is present the econometrician has the choice of a conditional or a simultaneous specification, as developed in sections 2.5.1 and 2.6 respectively. Both models require the econometrician to specify whether or not each area of the lattice are neighbours with the rest through the spatial proximity matrix, $\boldsymbol{W}$, discussed in section 2.2. The CAR and SAR models can be thought of as spatial analogues of the familiar time-series autoregressions in standard econometrics textbooks: the CAR is similar in its Markov property and the SAR in its functional form. Although the SAR model has dominated the econometrics literature to date, section 2.7 presents a case for the inclusion of the CAR model in the econometrician's toolbox. In particular, the CAR can achieve minimum mean squared prediction error and maximum entropy unlike the SAR model. Given the ease of implementation of the CAR model through MCMC methods and the relative computational burden of fitting SAR models, CARs can be considered a pragmatic alternative in the right circumstances. Additionally, the CAR model can be used as a spatial prior within a Bayesian hierarchical framework enabling Generalized Linear Modelling to be implemented. This extends the applicability of the CAR to discrete data such as counts.

Despite the rapid adoption of univariate spatial modelling in both statistics and econometrics, the modelling of multivariate data has received relatively little attention. In the next chapter recent innovations in the multivariate setting are discussed and the key weaknesses with the existing specifications are identified. In particular, the problem of incorporating differing spatial correlation parameters (α) within and between sites in multivariate models is considered as well as the difficulties in building MCMC sampling algorithms for multivariate CAR models.

CHAPTER 3

A REVIEW OF MULTIVARIATE SPATIAL MODELS

3.1 Introduction

The analysis of spatially correlated data is now an active area of research in both applied and theoretical econometrics. However, with the exception of Gamerman & Moreira (2004) and Kelejian & Prucha (2004)) this research has been limited to univariate data, yet many economic problems are inherently multivariate and there has been a long history of multivariate methods in econometrics. See for example, Harvey (1989) or West & Harrison (1997) for multivariate regression models in time series econometrics. The last chapter introduced the main theoretical contributions to univariate spatial analysis: the conditional autoregression (CAR) and the simultaneous autoregression (SAR). In comparison to econometrics, the statistics literature has seen significantly more progress with multivariate data, although this has been focused predominently on point-referenced or so called geostatistical data (i.e. data with a continuously varying spatial index) rather than on data distributed over a lattice (i.e. data with a discrete spatial index). In this chapter the relevant literature for multivariate spatially correlated data on a lattice is reviewed. The key theoretical developments in multivariate CAR models are presented and the weaknesses of the existing approaches are identified. These reduce to the problems of modelling inter and intra site correlations between the multiple variables and issues regarding sampling from the posterior distributions implied by these models. This paves the way for the development of the flexible multivariate CAR model in Chapter 5.

Consider a p-dimensional random variable $\boldsymbol{y}_i$ recorded at each site, i, which varies over an index set, $\mathrm{D} \subset \mathbb{R}^d$ generating a multivariate random field $\{\boldsymbol{y}_i \in \mathrm{D}\}$. For geostatistical data D is assumed to be a subset of $\mathbb{R}^d$ and i is assumed to vary continuously over D. For areal or lattice data, D is assumed to be a given finite or countably finite collection of points.

3.2 Multivariate MRF Models

Recall from section 2.4 that a Markov random field (MRF) is a generalization of a Markov process where, instead of the sequence or chain being indexed by time, any graph can be the index. Two popular specification of MRFs, the conditional autoregression of Besag (1974) and the simultaneous approach of Whittle (1954), were presented in section 2.7 as well as the limited situation in which these two forms are equivalent. In this section the focus is on presenting multivariate extensions of the CAR model, which will be the point of departure for this research. Before reviewing the literature on multivariate CAR models alternative multivariate MRF approaches are briefly discussed: (i) geostatistical approach for modelling data indexed on a continuous rather than discrete index, and (ii) multivariate extensions of the SAR model of section 2.6.

3.2.1 Geostatistical Data

Models for multivariate point-referenced or geostatistical data have been extensively explored in the literature. See, for example, papers by Royle & Berliner (1999), ver Hoef & Cressie (1994) or Gelfand et al. (2004) and the textbook by Wackernagel (1998). The essence of the analysis of geostatistical data is the construction of covariance functions between variables at different spatial locations as a function of their relative distance. Prediction at sites that were unobserved is then made using kriging[1] or co-kriging methods, such as in ver Hoef et al. (2004). These methods have been extended to lattice data by assigning the lattice measurements to one particular point in the lattice (e.g. the centroid of each site). Although on regular lattices this approach may be valid, it has limited use in irregular lattices such as those considered in this thesis and which constitute most economic datasets. This results largely from the arbitrary assignment of measurements to point sources (see chapter 7 of Cressie 1993, for a discussion). Lattice data are typically already aggregated over the site and collapsing this aggregate value to one point within the site adds to the problem of ecological bias as well as raising issues about the lack of a continuous underlying field and induced heteroskedasticity. As a result geostatistical methods are not optimal for lattice data applications.

3.2.2 Multivariate SAR models

Obviously, it is possible to extend either the SAR or CAR model for lattice data to the multivariate setting. For example, LeSage (1989) and LeSage & Reed (1989) consider

[1]Kriging is a group of techniques to interpolate the value of a random field at an unobserved location from observations of its value at nearby locations. Co-kriging is also a interpolation tool which expoits correlations between two, or more, variables to improve the estimation of variables at unobserved locations.

spatial Vector Autoregressions (VAR). More recently, Kelejian & Prucha (2004) introduce a very general class of multivariate SAR models that incorporate spatial lags in both endogenous and exogenous variables on top of the spatial dependence in the residuals. Frequentist two and three stage least squares (2SLS and 3SLS) estimators are derived and their properties discussed, generalizing those of Kelejian & Prucha (1998) from the univariate setting. This model is a multivariate extension of the traditional SAR model of Whittle (1954) as presented in Cliff & Ord (1981) and reviewed in section 2.6. However, the multivariate SAR has received very little attention in the applied literature to date and the open question about interpretation of regression parameters in the SAR model, univariate or multivariate, remains (see the discussion in section 2.6). This model will not be considered further.

3.2.3 Multivariate CAR models

In Chapter 5, the point of departure is the work of Mardia (1988) and recent extensions by Gelfand & Vounatsou (2003) and Jin et al. (2005) among others. This review begins by presenting the original multivariate CAR model of Mardia (1988). Mardia (1988) presents a summary of the very early literature on multivariate spatial approaches that will not be discussed further.

For a vector of univariate variables $\boldsymbol{y} = (y_1, y_2, ..., y_n)$, zero mean CAR models were developed in Besag (1974). Recall from section 2.5.1 on page 23, that under the Markov assumption, the n full unconditional distributions are specified as

$$\pi(y_i \mid y_j : j \in \partial_i) \sim \mathcal{N}(\alpha \sum_{j \in \partial_i} b_{ij} y_j, \tau_i^2), \quad i, j = 1, ..., n, \tag{3.2.1}$$

where $j \in \partial_i$ denotes that j is a neighbour of i which is captured in the spatial proximity matrix $\boldsymbol{W}$ with elements $w_{ii} = 0$ and $w_{ij} = 1$ if and only if sites i and j share a common boundary and $j \neq i$ – see page 15 in section 2.2 for details. Consider the spatial autocorrelation parameter version of the IAR model of Besag et al. (1991) (see page 24 in section 2.5.1) where the joint distribution, $\pi(\boldsymbol{y})$ is given by

$$\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu}, (\boldsymbol{I} - \alpha\boldsymbol{B})^{-1}\boldsymbol{\Gamma}), \tag{3.2.2}$$

and the spatial dependence is captured through $\boldsymbol{B} = \tilde{\boldsymbol{W}}$, which is the row-stochastic proximity matrix[2] and $\boldsymbol{\Gamma} = \text{diag}(\tau_i^2)$ and α is the spatial autocorrelation parameter from section 2.5.1. Recall that when $\alpha = 0$ there is spatial independence and when $\alpha = 1$ the improper IAR specification of Besag et al. (1991) is obtained. Cressie (1993) showed that a range of $\alpha \in (\lambda_{min}^{-1}, \lambda_{max}^{-1})$, where λ_{min} and λ_{max} are the minimum and maximum eigenvalues of $\boldsymbol{W}$, leads to non-singular covariance matrix, $(\boldsymbol{I} - \alpha\boldsymbol{B})$ and therefore a proper joint density. Carlin & Banerjee (2003) prove that $\alpha < |1|$ ensures the model's propriety.

For a multivariate CAR model (MCAR) let $\boldsymbol{y}' = (\boldsymbol{y}_1', \boldsymbol{y}_2', ..., \boldsymbol{y}_n')$ where each $\boldsymbol{y}_i'$ is a p-dimensional vector. Following Mardia (1988) the zero mean MCAR is

$$\boldsymbol{y}_i \mid \boldsymbol{y}_{-i} \sim \mathcal{N}(\sum_j \boldsymbol{B}_{ij}\boldsymbol{y}, \boldsymbol{\Sigma}_{\mathbf{i}}), \qquad i = 1, ..., n \tag{3.2.3}$$

where each $\boldsymbol{B}_{ij}$ and $\boldsymbol{\Sigma}_{\mathbf{i}}$ is p x p conditional covariance matrix. Analogous to the univariate case (equation (2.5.5)), Brook's lemma (Brook 1964) in Section 2.4.1

[2]gained by $\boldsymbol{D}^{-1}\boldsymbol{W}$ where $\boldsymbol{D} = \text{diag}(m_i)$ where m_i are the number of neighbours to site i and $\boldsymbol{W}$ is the spatial proximity matrix.

provides the joint density for $\boldsymbol{y}$ of the form

$$\pi(\boldsymbol{y} \mid \boldsymbol{\Sigma}) \propto \exp\left\{-\frac{1}{2}\boldsymbol{y}'\boldsymbol{\Gamma}^{-1}(\boldsymbol{I}-\tilde{\boldsymbol{B}})\boldsymbol{y}\right\}, \tag{3.2.4}$$

where $\boldsymbol{\Gamma}$ is block diagonal with blocks $\boldsymbol{\Sigma_i}$ and $\tilde{\boldsymbol{B}}$ is an np x np matrix with $(i,j)^{th}$ block $\boldsymbol{B}_{ij}$. From the MCAR specification in (3.2.4), different $\boldsymbol{\Gamma}$ and $\tilde{\boldsymbol{B}}$ matrices can be specified to produce different MCAR model structures. But, as in the univariate case, it is necessary to ensure that $\boldsymbol{\Gamma}^{-1}(\boldsymbol{I}-\tilde{\boldsymbol{B}})$ is a symmetric, positive-definite matrix if a proper joint density is to exist. Unfortunately, guaranteeing these conditions can be troublesome and has exercised a great part of the literature on multivariate spatial models.

Considering the case of symmetry, setting $\boldsymbol{B}_{ij} = b_{ij}\boldsymbol{I}$ yields $b_{ij}\boldsymbol{\Sigma}_j = b_{ji}\boldsymbol{\Sigma}_i$ analogous to (2.5.7). If $b_{ij} = w_{ij}/w_{i+}$ and $\boldsymbol{\Sigma}_i = w_{i+}^{-1}\boldsymbol{\Sigma}$, similar to section 2.5.1, then the symmetry condition is satisfied.

Kronecker product notation simplies the multivariate form of $\boldsymbol{\Gamma}^{-1}(\boldsymbol{I}-\tilde{\boldsymbol{B}})$ by letting $\tilde{\boldsymbol{B}} = \boldsymbol{B} \otimes \boldsymbol{I}$ with $\boldsymbol{B}$ the same as equation (2.5.5) and $\boldsymbol{\Gamma} = \boldsymbol{D}^{-1} \otimes \boldsymbol{\Sigma}$. This simplifies $\boldsymbol{\Gamma}^{-1}(\boldsymbol{I}-\tilde{\boldsymbol{B}})$ to

$$\boldsymbol{\Gamma}^{-1}(\boldsymbol{I}-\tilde{\boldsymbol{B}}) = (\boldsymbol{B} \otimes \boldsymbol{I})(\boldsymbol{D}^{-1} \otimes \boldsymbol{\Sigma}) = (\boldsymbol{D}_W - \boldsymbol{W}) \otimes \boldsymbol{\Sigma}^{-1}. \tag{3.2.5}$$

As discussed in section 2.5.1 in the univariate case, $\boldsymbol{D}_W - \boldsymbol{W}$ is singular implying the singularity of $\boldsymbol{\Gamma}^{-1}(\boldsymbol{I}-\tilde{\boldsymbol{B}})$. This distribution is denoted by $MCAR(1, \boldsymbol{\Sigma})$. This improper MCAR was ignored initially due to computational difficulties, but work by Knorr-Held & Rue (2002) employs block updating to conduct inference in a Bayesian MCMC setting with an application to disease mapping.

The literature since has focused on correcting this impropriety. Again, following the approach adopted for the univariate case, rewrite (3.2.3) in the general form of

$$\mathrm{E}[\boldsymbol{y}_i \mid \boldsymbol{y}_{-i}] = \boldsymbol{R}_i \sum_j \boldsymbol{B}_{ij}\boldsymbol{y}_j. \tag{3.2.6}$$

Substitute $\boldsymbol{\Gamma}^{-1}(\boldsymbol{I}-\tilde{\boldsymbol{B}})$ with $\boldsymbol{\Gamma}^{-1}(\boldsymbol{I}-\tilde{\boldsymbol{B}}_R)$ where $\tilde{\boldsymbol{B}}_R$ has $(i,j)^{th}$ block $\boldsymbol{R}_i\boldsymbol{B}_{ij}$ Then from Mardia (1988) the symmetry condition becomes $(\boldsymbol{\Sigma}_i^{-1}\boldsymbol{R}_i\boldsymbol{B}_{ij})' = \boldsymbol{\Sigma}_j^{-1}\boldsymbol{R}_j\boldsymbol{B}_{ji}$, or $\boldsymbol{\Sigma}_j\boldsymbol{B}_{ij}'\boldsymbol{R}_i' = \boldsymbol{R}_j\boldsymbol{B}_{ji}\boldsymbol{\Sigma}_i$. Setting $\boldsymbol{B}_{ij} = b_{ij}\boldsymbol{I}$ and $b_{ij} = w_{ij}/w_{i+}$ simplifies this symmetry condition to

$$w_{j+}\boldsymbol{\Sigma}_j\boldsymbol{R}_i' = w_{i+}\boldsymbol{R}_j\boldsymbol{\Sigma}_i. \tag{3.2.7}$$

Additionally, if $\boldsymbol{\Sigma}_i = w_{i+}^{-1}\boldsymbol{\Sigma}$ then $\boldsymbol{\Sigma}\boldsymbol{R}_i' = \boldsymbol{R}_j\boldsymbol{\Sigma}$ is obtained which reveals that $\boldsymbol{R}_i = \boldsymbol{R}_j = \boldsymbol{R}$ and as a result,

$$\boldsymbol{\Sigma}\boldsymbol{R}' = \boldsymbol{R}\boldsymbol{\Sigma}. \tag{3.2.8}$$

See Banerjee et al. (2004) for details. For any arbitrary positive-definite $\boldsymbol{\Sigma}$, a generic solution to (3.2.8) according to Carlin & Banerjee (2003) is $\boldsymbol{R} = \alpha\boldsymbol{\Sigma}'$ resulting without loss of generality to $\boldsymbol{R} = \alpha\boldsymbol{I}$. This now results in

$$\boldsymbol{\Sigma}^{-1}(\boldsymbol{I}-\tilde{\boldsymbol{B}}_R) = (\boldsymbol{D}-\alpha\boldsymbol{W}) \otimes \boldsymbol{\Sigma}^{-1} \tag{3.2.9}$$

Adopting the same restrictions on the range of α as we did in the univariate case in section 2.5.1 results in a non-singular matrix. This was that $\alpha \in (\lambda_{min}^{-1}, \lambda_{max}^{-1})$ where λ_{min} and λ_{max} are the minimum and maximum ordered eigenvalues of the proximity matrix. Carlin & Banerjee (2003) avoid the calculation of eigenvalues by using a row-stochastic proximity matrix, $\tilde{\boldsymbol{W}}$ and proving that $|\alpha| < 1$ ensures a non-singular matrix[3]. Refer back to section 2.5.1 for a review. This proper MCAR model

[3] If $\boldsymbol{\Sigma}$ is appropriately constrained to be diagonal with elements σ_i^2, $\boldsymbol{R}$ can be diagonal with

is denoted $MCAR(\alpha, \boldsymbol{\Sigma})$

To simplify, assume $\boldsymbol{R}_i = \alpha \boldsymbol{I}$ for $i = 1, ..., n$ where α is the spatial autocorrelation coefficient from section 2.5.1 or as Sun et al. (2000) termed it, a spatial smoothing parameter. Additionally, set $\boldsymbol{\Gamma} = \boldsymbol{D} \otimes \boldsymbol{\Lambda}$ where $\boldsymbol{D} = \text{diag}(m_i)$ and m_i are the number of neighbours to site i. With these two assumptions (3.2.4) becomes

$$\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu}, [(\boldsymbol{D}(\boldsymbol{I} - \alpha \boldsymbol{B})) \otimes \boldsymbol{\Lambda}]^{-1}), \tag{3.2.10}$$

where $\boldsymbol{\Lambda}$ is a p x p dimensional symmetric, positive-definite matrix of non-spatial variances. This simplication allow us to state that the covariance matrix in (3.2.10) is positive-definite as long as $\boldsymbol{\Lambda}$ is positive-definite. When $\boldsymbol{B} = \tilde{\boldsymbol{W}}$ (i.e. the row-stochastic spatial weights matrix) and $\boldsymbol{D} = \text{diag}(m_i)$ with m_i being the number of neighbours of site i and restricting $\alpha \in (-1, 1)$ then equation (3.2.10) reduces to

$$\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu}, [(\boldsymbol{D} - \alpha \boldsymbol{W}) \otimes \boldsymbol{\Lambda}]^{-1}). \tag{3.2.11}$$

This proper MCAR model is denoted $MCAR(\alpha, \boldsymbol{\Lambda})$ by Carlin & Banerjee (2003) and Gelfand & Vounatsou (2003). All of the univariate structures from chapter 2 can now be applied to the matrices in (3.2.10) to obtain equivalent multivariate structures. For instance, by forcing $\alpha = 1$ in (3.2.10) the multivariate IAR model of Besag et al. (1991) results.

All of the above MCAR models are generalizations of the univariate CAR models under the assumption $\boldsymbol{R}_i = \alpha \boldsymbol{I}$ and is thus applicable to any dimension, p. This assumption of a common $\boldsymbol{R}_i$ for all i may be too strong and is rather inflexible. Relaxing this assumption whilst maintaining a positive-definite covariance matrix has

---

elements $\alpha_{\boldsymbol{i}}$ which would yield p independent CAR models.

been the focus of the literature since Mardia (1988). Consider the bivariate case (i.e. $p = 2$ for each site $i = 1, ..., n$) and define $\boldsymbol{y}_1' = (y_{11}, ..., y_{n1})$ and $\boldsymbol{y}_2' = (y_{12}, ..., y_{n2})$. Then the $MCAR(\alpha, \boldsymbol{\Lambda})$ model from Equation (3.2.11) becomes

$$\begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{pmatrix} \sim \left(\begin{pmatrix} \boldsymbol{\mu_1} \\ \boldsymbol{\mu_2} \end{pmatrix}, \begin{pmatrix} (\boldsymbol{D} - \alpha_1 \boldsymbol{W})\boldsymbol{\Lambda}_{11} & (\boldsymbol{D} - \alpha_3 \boldsymbol{W})\boldsymbol{\Lambda}_{12} \\ (\boldsymbol{D} - \alpha_3 \boldsymbol{W})\boldsymbol{\Lambda}_{12} & (\boldsymbol{D} - \alpha_2 \boldsymbol{W})\boldsymbol{\Lambda}_{22} \end{pmatrix}^{-1} \right), \qquad (3.2.12)$$

where $\boldsymbol{\Lambda}_{ij}$, $i = 1, 2$ and $j = 1, 2$ are the elements of $\boldsymbol{\Lambda}$. From equation (3.2.12) three α parameters are required, representing a spatial autocorrelation for each variable ($\boldsymbol{y}_1$ and $\boldsymbol{y}_2$) and one to control for the correlation *between* variables at different locations. The covariance matrix includes α_1 and α_2 which are the spatial autocorrelation parameters (or spatial smoothing parameters to be more precise) for the two variables $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ and α_3 is the "linking" parameter (Kim et al. 2003) associating the two variables $\boldsymbol{y}_{i1}$ and $\boldsymbol{y}_{j2}$ $(i \neq j)$ at different sites. As shown in Gelfand & Vounatsou (2003) and Jin et al. (2005) it is difficult to confirm the positive-definiteness of the covariance matrix in (3.2.12) due to the unknown $\boldsymbol{\Lambda}$ matrix. Consequently, this model is particularly difficult to fit via traditional MCMC methods.

Gelfand & Vounatsou (2003) and Carlin & Banerjee (2003) instead generalize the basic $MCAR(\alpha, \boldsymbol{\Lambda})$ model to the case of two different spatial autocorrelation parameters, α_1 and α_2, which they denote as $MCAR(\alpha_1, \alpha_2, \boldsymbol{\Lambda})$ They write the precision matrix, $\boldsymbol{\Sigma}^{-1}$, for this MCAR model as

$$\begin{pmatrix} \boldsymbol{R}_1'\boldsymbol{R}_1\boldsymbol{\Lambda}_{11} & \boldsymbol{R}_1'\boldsymbol{R}_2\boldsymbol{\Lambda}_{12} \\ \boldsymbol{R}_2'\boldsymbol{R}_1\boldsymbol{\Lambda}_{12} & \boldsymbol{R}_2'\boldsymbol{R}_2\boldsymbol{\Lambda}_{22} \end{pmatrix} = \begin{pmatrix} \boldsymbol{R}_1' & 0 \\ 0 & \boldsymbol{R}_2' \end{pmatrix} (\boldsymbol{\Lambda} \otimes \boldsymbol{I}) \begin{pmatrix} \boldsymbol{R}_1 & 0 \\ 0 & \boldsymbol{R}_2 \end{pmatrix}, \qquad (3.2.13)$$

where $\boldsymbol{R}_k'\boldsymbol{R}_k = \boldsymbol{D} - \alpha_k \boldsymbol{W}$, $k = 1, 2$. Carlin & Banerjee (2003) take the Cholesky decomposition of $\boldsymbol{D} - \alpha_k \boldsymbol{W}$ as R_k so that it is an upper-triangular matrix. Alternatively, Gelfand & Vounatsou (2003) adopt a spectral (eigenvalue) decomposition such that $R_k = \text{diag}(1-\alpha_k\lambda_i)^{1/2}\boldsymbol{P}'\boldsymbol{D}^{1/2}\boldsymbol{P}$ where λ_i are the eigenvalues of $\boldsymbol{D}^{-1/2}\boldsymbol{W}\boldsymbol{D}^{-1/2}$ and $\boldsymbol{P}$ is an orthogonal matrix with the corresponding eigenvectors as its columns. Whichever way $\boldsymbol{R}_k$ is found, these generalization of the basic MCAR model of Mardia (1988) allow for a different spatial autocorrelation parameter, α_k, for each of the k variables observed. The non-spatial correlation between the variables at any location is captured by the $\boldsymbol{\Lambda}$ matrix.

The usefulness of these models is that as long as the Cholesky or spectral decompositions exist and $\boldsymbol{\Lambda}$ is positive-definite, then the necessary conditions to ensure that the covariance matrix is positive-definite are easy to find. As Jin et al. (2005) show, this reduces to restricting the spatial autocorrelation parameters to be less than 1 in the bivariate case (i.e. $|\alpha_1| < 1$ and $|\alpha_2| < 1$). From a computational perspective, the spectral decomposition favoured by Gelfand & Vounatsou (2003) is preferred because it saves the overhead of calculating the Cholesky decomposition at each MCMC iteration. However, both versions of (3.2.13) are limited because they do not allow for spatial autocorrelation between different variables across sites as in the general MCAR presented in (3.2.12). This is obvious from the fact that there is no spatial autocorrelation parameter, α, on the off-diagonal in (3.2.13) because the off-diagonal is determined by the diagonal to force positive-definiteness. Moreover, as $\boldsymbol{D} - \alpha_k\boldsymbol{W}$ is not unique there could exist different MCAR models with the same covariance structure as (3.2.13).

Kim et al. (2003) proposed a bivariate CAR model which they called the "two-fold conditionally autoregressive model". They specify the moments of the full conditional distributions as

$$\mathrm{E}[y_{ik} \mid y_{il}, y_{jk}, y_{jl}] = \frac{1}{2m_i + 1}\left(\alpha_k \sum_{j\in\partial_i} y_{jk} + \alpha_3\sqrt{\frac{\tau_l}{\tau_k}}\sum_{j\in\partial_i} y_{jl} + \alpha_0\sqrt{\frac{\tau_l}{\tau_k}} y_{il}\right) \tag{3.2.14}$$

and

$$\mathrm{Var}[y_{ik} \mid y_{il}, y_{jk}, y_{jl}] = \frac{\tau_k^{-1}}{2m_i + 1}, \qquad i, j = 1, ..., n, \quad l, k = 1, 2, \quad l \neq k. \tag{3.2.15}$$

They derive the joint distribution given by these full conditionals as

$$\begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} (2\boldsymbol{D} + \boldsymbol{I} - \alpha_1 \boldsymbol{W})\tau_1 & -(\alpha_0 \boldsymbol{I} + \alpha_3 \boldsymbol{W})\sqrt{\tau_1\tau_2} \\ -(\alpha_0 \boldsymbol{I} + \alpha_3 \boldsymbol{W})\sqrt{\tau_1\tau_2} & (2\boldsymbol{D} + \boldsymbol{I} - \alpha_2 \boldsymbol{W})\tau_2 \end{pmatrix}^{-1}\right) \tag{3.2.16}$$

where $\boldsymbol{y}_1' = (y_{11}, ..., y_{n1})$, $\boldsymbol{y}_2' = (y_{12}, ..., y_{n2})$, $\boldsymbol{D} = \mathrm{diag}(m_i)$, and $\boldsymbol{W}$ is the spatial proximity matrix. Notice that this has the same number of coefficients as the general specification in equation (3.2.12) for the bivariate case and so they are related to each other. In the Kim et al. (2003) specification in (3.2.16) there is one spatial autocorrelation parameter per variable (α_1 and α_2) and two additional correlation coefficients (α_0 and α_3) associating y_{i1} with y_{i2} and y_{j2} respectively. The so-called "bridging" parameter relates y_{i1} with y_{i2} (i.e. is a correlation coefficient for variables within a site) where as the so-called "linking" parameter relates y_{i1} with y_{j2} (i.e. it is a spatial correlation coefficient for different variables at different sites). The Kim et al. (2003) model therefore incorporates a more flexible correlation pattern than the $MCAR(\alpha_1, \alpha_2, \boldsymbol{\Lambda})$ models of Gelfand & Vounatsou (2003) and Carlin & Banerjee (2003) presented in (3.2.13).

However as noted in Banerjee et al. (2004), model (3.2.16) is designed solely for the bivariate case and it apears difficult to generalize it to larger numbers of dependent variables. The absense of any research in this direction over the intervening years is testament to this view. Moreover, it is unattractive because of the computational burden of fitting this "two-fold" model: it requires numerous inverse calculations, matrix multiplications and determinant evaluations at each iteration of the sampler requiring significant computational power and time for even reasonably small sample sizes. Finally, guaranteeing the positive-definite condition of the covariance matrix is not easy. Kim et al. (2003) provide a set of sufficient but unnecessary conditions: $|\alpha_l| < 1$, $l = 0, 1, 2, 3$. These conditions restrict the possible correlations between y_{i1} with y_{i2} and y_{j2} more than is practical in applied situations.

More recently, Jin et al. (2005) develop a "generalized MCAR" model. Reminiscent of the approach of Royle & Berliner (1999) for geostatistical (or continuous) data they specify the joint distribution for the multivariate data through the specification of simple conditional and marginal distributions. All of the MCAR models considered so far specify the precision matrix rather than the covariance matrix directly. This greatly improves MCMC computation but can make interpretation difficult. As demonstrated, specifying a valid joint covariance matrix is a difficult task. Jin et al. (2005) specify the joint distribution directly. Assuming a zero-mean joint bivariate distribution for $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ is

$$\begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{12} & \boldsymbol{\Sigma}_{22} \end{pmatrix}\right) \tag{3.2.17}$$

where $\boldsymbol{\Sigma}_{kl}, k, l = 1, 2$ are n x n covariance matrices. Appealing to standard multivariate theory (see for instance Mardia et al. 1979) for normal distributions

$$\mathrm{E}[\boldsymbol{y}_1 \mid \boldsymbol{y}_2] = \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{y}_2$$
$$\mathrm{Var}[\boldsymbol{y}_1 \mid \boldsymbol{y}_2] = \boldsymbol{\Sigma}_{11\cdot 2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{12}'$$

Writing $\boldsymbol{A} = \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}$ the joint distribution in (3.2.17) can be rewritten as

$$\begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11\cdot 2} + \boldsymbol{A}\boldsymbol{\Sigma}_{22}\boldsymbol{A}' & \boldsymbol{A}\boldsymbol{\Sigma}_{22} \\ (\boldsymbol{A}\boldsymbol{\Sigma}_{22})' & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right). \tag{3.2.18}$$

The conditions to ensure the propreity of (3.2.18) are that $\boldsymbol{\Sigma}_{22}$ and $\boldsymbol{\Sigma}_{11\cdot 2}$ are both positive-definite matrices (see Harville 1997, Corollary 14.8.5). Because $\boldsymbol{y}_1 \mid \boldsymbol{y}_2 \sim \mathcal{N}(\boldsymbol{A}\boldsymbol{y}_2, \boldsymbol{\Sigma}_{11\cdot 2})$ and $\boldsymbol{y}_2 \sim \mathcal{N}(0, \boldsymbol{\Sigma}_{22})$, it is possible to write $\pi(\boldsymbol{y}) = \pi(\boldsymbol{y}_1 \mid \boldsymbol{y}_2)\pi(\boldsymbol{y}_2)$. This allows the joint distribution of $\boldsymbol{y}$ to be specified by the matrices $\boldsymbol{\Sigma}_{11\cdot 2}, \boldsymbol{\Sigma}_{22}$ and $\boldsymbol{A}$.

Jin et al. (2005) adopt the univariate CAR structure discussed in section 2.5.1 and specify the conditional distribution, $\boldsymbol{y}_1 \mid \boldsymbol{y}_2$ as $\boldsymbol{y}_1 \mid \boldsymbol{y}_2 \sim \mathcal{N}(\boldsymbol{A}\boldsymbol{y}_2, [(\boldsymbol{D} - \alpha_1\boldsymbol{W})\tau_1]^{-1})$, and the marginal distribution as $\boldsymbol{y}_2 \sim \mathcal{N}(\mathbf{0}, [(\boldsymbol{D} - \alpha_2\boldsymbol{W})\tau_2]^{-1})$ where α_1 and α_2 are the spatial autocorrelation coefficients for the conditional and marginal distributions of $\boldsymbol{y}_1 \mid \boldsymbol{y}_2$ and $\boldsymbol{y}_2$ respectively. Similarly, τ_1 and τ_2 are the relevant conditional and marginal precisions. The resulting joint distribution will be proper providing that these two, simpler CAR distributions are proper. Again, as above, $\boldsymbol{D} = \mathrm{diag}(m_i)$ and $\boldsymbol{W}$ is, again, the spatial proximity matrix. As a result, the necessary conditions to ensure a valid covariance matrix for the joint distribution in (3.2.18) is the same as for the univariate case: $|\alpha_1| < 1$ and $|\alpha_2| < 1$ (Jin et al. 2005, p. 953).

The so called bridging and linking parameters can be introduced through the elements of the matrix $\boldsymbol{A}$. Since $\mathrm{E}[\boldsymbol{y}_1 \mid \boldsymbol{y}_2] = \boldsymbol{A}\boldsymbol{y}_2$ the elements are of the form

$$a_{ij} = \begin{cases} \eta_0 & \text{if} \quad j = i, \\ \eta_1 & \text{if} \quad j \in \partial_i, \\ 0 & \text{otherwise} \end{cases} \tag{3.2.19}$$

Therefore, $\boldsymbol{A} = \eta_0\boldsymbol{I} + \eta_1\boldsymbol{W}$ and $\mathrm{E}[\boldsymbol{y}_1 \mid \boldsymbol{y}_2] = (\eta_0\boldsymbol{I} + \eta_1\boldsymbol{W})\boldsymbol{y}_2$. It is clear then, that η_0 and η_1 are the bridging and linking parameters relating $\boldsymbol{y}_{i1}$ with $\boldsymbol{y}_{i2}$ and $\boldsymbol{y}_{j2}$ $(j \neq i)$ which are analogous to the α_0 and α_3 parameters in the two-fold model of Kim et al. (2003) in (3.2.16). Given these assumptions on the form of $\boldsymbol{A}$, the joint distribution covariance matrix from (3.2.18) becomes

$$\begin{pmatrix} \boldsymbol{\Sigma}_{11\cdot 2} + \boldsymbol{A}\boldsymbol{\Sigma}_{22}\boldsymbol{A}' & \boldsymbol{A}\boldsymbol{\Sigma}_{22} \\ (\boldsymbol{A}\boldsymbol{\Sigma}_{22})' & \boldsymbol{\Sigma}_{22} \end{pmatrix} \tag{3.2.20}$$

where

$$\begin{aligned} \boldsymbol{\Sigma}_{11\cdot 2} + \boldsymbol{A}\boldsymbol{\Sigma}_{22}\boldsymbol{A}' &= [\tau_1(\boldsymbol{D} - \alpha_1\boldsymbol{W})]^{-1} + (\eta_0\boldsymbol{I} + \eta_1\boldsymbol{W})[\tau_2(\boldsymbol{D} - \alpha_2\boldsymbol{W})]^{-1}(\eta_0\boldsymbol{I} + \eta_1\boldsymbol{W}) \\ \boldsymbol{A}\boldsymbol{\Sigma}_{22} &= (\eta_0\boldsymbol{I} + \eta_1\boldsymbol{W})[\tau_2(\boldsymbol{D} - \alpha_3\boldsymbol{W})]^{-1} \\ (\boldsymbol{A}\boldsymbol{\Sigma}_{22})' &= [\tau_2(\boldsymbol{D} - \alpha_2\boldsymbol{W})]^{-1}(\eta_0\boldsymbol{I} + \eta_1\boldsymbol{W}) \\ \boldsymbol{\Sigma}_{22} &= [\tau_2(\boldsymbol{D} - \alpha_2\boldsymbol{W})]^{-1} \end{aligned}$$

Jin et al. (2005) denote this model the Generalized Multivariate Conditional Autoregression (GMCAR). To avoid confusion and to draw out the similarities (and differences) with the models considered previously, this model is labelled the $MCAR(\alpha_1, \alpha_2, \eta_0, \eta_1, \tau_1, \tau_2)$. Many of the MCAR models that have already been reviewed can be seen within the Jin et al. (2005) framework when various assumptions are made about the six parameters in (3.2.20). For instance, assuming that $\alpha_1 = \alpha_2 = \alpha$ and applying standard multivariate theory (namely Harville 1997, Corollary 8.5.12) the $MCAR(\alpha, \eta_0, \eta_1, \tau_0, \tau_1)$ is the same as the $MCAR(\alpha, \mathbf{\Lambda})$ model of Mardia (1988) in (3.2.12). The relationship between the parameters becomes $\tau_1 = \mathbf{\Lambda}_{11}, \tau_2 = \mathbf{\Lambda}_{22} - \mathbf{\Lambda}_{12}^2/\mathbf{\Lambda}_{11}$, and $\eta_0 = -\mathbf{\Lambda}_{12}/\mathbf{\Lambda}_{11}$. Given these assumptions and functional relationships, setting $\alpha = 1$ then this MCAR becomes $MCAR(1, \eta_0, \tau_1, \tau_2)$ and is equivalent to the multivariate IAR model denoted $MCAR(1, \mathbf{\Lambda})$. Assuming that each variable has a separate spatial autocorrelation parameter (i.e. $\alpha_1 \neq \alpha_2$) but ignoring dependence between the multivariate components (i.e. $\eta_0 = \eta_1 = 0$) then two separate univariate CAR models result. Finally, if $\alpha_1 = \alpha_2 = 0, \eta_o \neq 0$, and $\eta_1 = 0$) then the MCAR reduces to a normal bivariate model.

One of the problems of the Jin et al. (2005) MCAR model of (3.2.20) is that when fitting the hierarchical model the econometrician has to decide whether to model the conditional distribution $\pi(\boldsymbol{y}_1 \mid \boldsymbol{y}_2)$ and then the marginal distribution $\pi(\boldsymbol{y}_2)$ or whether to model the conditional distribution $\pi(\boldsymbol{y}_2 \mid \boldsymbol{y}_1)$ and then $\pi(\boldsymbol{y}_1)$. In some applications there will be a natural ordering. For instance Royle & Berliner (1999) model the concentration of ozone at particular points which is scientifically explained by the maximum temperature at that location, but not the other way around. Similarly, Gelfand et al. (2004) model property price data and model the selling price for a block of apartments as a function of the rental income of that block. When no natural ordering is present, Jin et al. (2005) suggest that is possible to incorporate this as a model selection problem using DIC[4] (Spiegelhalter et al. 2002) to assess the best fit. However, remains a problem that has not been properly

[4]The deviance information criterion (DIC) is a hierarchical modeling generalization of the AIC (Akaike information criterion) and BIC (Bayesian information criterion), also known as the Schwarz criterion. It is particularly useful in Bayesian model selection problems where the posterior distributions of the models have been obtained by MCMC. See Gelman et al. (2003), Spiegelhalter et al. (1996) or Spiegelhalter et al. (2002) for further information.

addressed in practice and applications of this model are extremely limited.

As can be seen with all of the models discussed in this chapter, the linking parameters or conditional cross-correlations between variable k at site i and variable l and site j is symmetric due to the symmetrical specification of the spatial correlation parameters α_k. Only an asymmetric specification of these parameters will allow the multiple variables to have varying cross-correlations. Asymmetric spatial dependencies can occur in a range of natural processes important to modern economics such as agriculture, pollution, social mobility and disease. Even if there is no explicit a priori reason to suspect asymmetric spatial dependence between the variables it should still be explored during the analysis. ver Hoef & Cressie (1994) note that not considering this possible asymmetric structure can lead to researchers missing important elements of the process under study. This will be particularly important in situations where the relationships between the variables is important for the analysis itself, such as the multiple road safety performance indicators that are considered in this study.

An additional weakness of the MCAR models reviewed in this chapter is their reliance on the number of neighbours in calculating the precision of the measurements. For instance Carlin & Banerjee (2003) and Gelfand & Vounatsou (2003) both relate the precision at site i with the number of neighbours m_i in the neighbourhood of site i i.e. the number of $j \in \partial_i$. For instance, we defined $\boldsymbol{\Gamma} = \boldsymbol{D} \otimes \boldsymbol{\Lambda}$ and the covariance matrix as

$$\boldsymbol{\Sigma} = (\boldsymbol{D} - \alpha \boldsymbol{W}) \otimes \boldsymbol{\Gamma} \tag{3.2.21}$$

where $\boldsymbol{D} = \text{diag}(m_i)$ with m_i being the number of neighbours of site i. Hence both of these matrices depend explicitly on the number of neighbours which can vary from site to site. Using standard multivariate theory (see Appendix B.3.1 for details) the conditional covariance of two neighbouring locations (i and j) is

$$\begin{aligned}
\mathbf{\Sigma_{ij|-ij}} &= \begin{pmatrix} m_i & -\alpha \\ -\alpha & m_j \end{pmatrix}^{-1} \otimes \mathbf{\Gamma} \\
&= \mathbf{\Gamma}*_{ij} \begin{pmatrix} \boldsymbol{I} & -\frac{\alpha}{\sqrt{m_i m_j}}\boldsymbol{I} \\ -\frac{\alpha}{\sqrt{m_i m_j}}\boldsymbol{I} & \boldsymbol{I} \end{pmatrix}^{-1} \mathbf{\Gamma}*_{ij} \qquad (3.2.22) \\
&= \mathbf{\Gamma}*_{ij} \begin{pmatrix} \frac{1}{1-\alpha\sqrt{m_i m_j}}\boldsymbol{I} & \dfrac{\alpha/\sqrt{m_i m_j}}{1-\alpha\sqrt{m_i m_j}}\boldsymbol{I} \\ -\frac{\alpha}{\sqrt{m_i m_j}}\boldsymbol{I} & \frac{\alpha/\sqrt{m_i m_j}}{1-\alpha\sqrt{m_i m_j}}\boldsymbol{I} \end{pmatrix} \mathbf{\Gamma}*_{ij}
\end{aligned}$$

where

$$\mathbf{\Gamma}*_{ij} = \begin{pmatrix} \mathbf{\Gamma}^{1/2}/\sqrt{m_i} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Gamma}^{1/2}/\sqrt{m_j} \end{pmatrix}.$$

The conditional correlation matrix is then given by

$$\boldsymbol{R}_{ij|-ij} = \begin{pmatrix} 1 & \alpha/\sqrt{m_i m_j} \\ \alpha/\sqrt{m_i m_j} & 1 \end{pmatrix} \otimes \boldsymbol{R} \qquad (3.2.23)$$

where $\boldsymbol{R} = \boldsymbol{D}^{-1/2}\mathbf{\Gamma}\boldsymbol{D}^{-1/2}$.

It is clear that the correlations in (6.1) are functions of the number of neighbours as well, and this can vary with each pair of neighbours over irregular lattices such as the one considered in this thesis. Therefore, interpretation of the spatial dependence parameters, α will be very difficult, if not impossible, under such a parameterization of the MCAR model. In chapter 5 a novel solution is proposed based upon earlier work by Chan & Cressie (1989) that instead incorporates individual precision measures into the $\boldsymbol{m}$ rather than the number of neighbours.

Multivariate CAR models can also provide parameters in a multiple regression setting that are dependent and spatially varying at the site level. For instance, Gamerman et al. (2002) investigate a multivariate extension of the IAR model of Besag & Kooperberg (1995), in which they resolve the impropriety of the density by centering the spatial random effect, ϕ_i from equation (2.5.10) around some mean location. Additionally they remove the spatial random effect from the intercept term (i.e. they actually drop ϕ_i) and replace it with spatially varying regression coefficients (i.e. they replace β_1 from the vector $\boldsymbol{\beta}$ in equation (2.5.10) by β_{1i}) which follow a CAR distribution. This was termed a *space-varying coefficient model* by Assunção et al. (2002) who applied the model to estimating fertility schedules. A good survey of the methods in this particular approach can be found in Assunção (2003). The downside of this approach is the rather complicated MCMC blocking strategies proposed. Also noteworthy is the work of Langford et al. (1999) and Leyland et al. (2000) who create spatial random effects as proximity-based weighted averages of independent normal variables and use a hierarchical setting to improve estimates of each variable by shrinkage across the variables as well as across the levels of the hierarchical model.

3.3 Summary

Building from the presentation of univariate spatial models in Chapter 2, this chapter has reviewed the literature on multivariate spatial regression models. Multivariate extensions of the simultaneous autoregression model of Whittle (1954) were briefly introduced and the limitations of geostatistical approaches (i.e. those designed for point-referenced data over a continuous surface) were discussed. This chapter focused on the multivariate CAR (MCAR) model first proposed by Mardia (1988). Spatial dependence is captured through the covariance matrix, or rather its inverse, as in the univariate case. Previous research efforts have used simple forms for the covariance matrix that, although computationally convenient, unduly constrain the range and/or type of correlation modelled. For instance, both Gelfand & Vounatsou (2003) and Carlin & Banerjee (2003) restrict the degree of spatial dependence to be symmetric at different locations. Although Royle & Berliner (1999) develop a model that incorporates asymmetric spatial dependencies it is only suitable to geostatistical data. Jin et al. (2005) build a similar model for lattice data but do not explicitly model the cross-dependencies.

Given the growing number of multivariate spatial datasets available and the large number of problems in disciplines as diverse as econometrics, environmetrics, biostatistics, epidemiology and statistics that are inherently multivariate, work in this domain is a worthy avenue of research. Developing very general covariance structures that allow for flexible spatial correlation to be modelled whilst maintaining the positive-definite nature of the covariance matrix are important. This thesis will make a contibution to the literature on multivariate spatial regression models by introducing a novel conditional approach that allows for varying degrees of spatial dependence for different variables as well a asymmetric covariances between different variables at different locations. The model is presented in chapter 5, and efficient MCMC samplers for fitting the model are provided. In the next chapter the importance of modelling multiple performance indicators in the public sector is discussed.

CHAPTER 4

ROAD SAFETY PERFORMANCE INDICATORS

There can be economy only where there is efficiency. (Disraeli, 1868)

4.1 Introduction

The use of performance indicators is widespread in the public sector. Their use reaches across all levels of the public sector (from organisations to individuals) and across all branches. The limitations of performance indicators in general and league tables in particular are well reported and they are introduced only briefly here. Road accidents kill 3,000 people every day around the world; they are the 10th largest cause of death (Commission for Global Road Safety 2005). As such the UK government has set road safety a priority for local authorities and requires local authorities to publish a number of headline performance indicators of road causalities by mode as part of the statutory reporting of Best Value Performance Indicators (BVPI) produced annually. These crude measures do not reflect the true output under consideration, road safety, and are subject to high variability. In this chapter the literature on performance management in the public sector is reviewed and the problems with existing approaches highlighted. The importance of research into road safety monitoring and performance indicators is stressed. A number of the problems raised in this review will be addressed in chapter 7 through the application of the model developed in chapter 5 to data.

4.2 Performance Management in the Public Services

Performance monitoring in government has received significant attention since the 1980s, fostered by the "re-inventing government" movement (Osbourne & Gaebler 1992), and Smith (1990) provides an account of these early attempts at performance management in the UK public sector. The 1990s witnessed an explosion in the the UK's "government by measurement" (House of Commons Public Administration Select Committee 2003). Although good performance management is productive for all concerned, done badly, it can be costly, ineffective, harmful and destructive (Bird 2004). Performance management (PM) was introduced across government in an attempt to measure the process and outcomes of the public sector, and as an incentive to drive increases in efficiency and effectiveness. It also provided greater accountability to Parliament and the public for the government's "stewardship" of the public services (Bird 2005). PM takes places at all scales. For instance at the level of the programme (e.g. the impact of the NHS on health), at an organizational level (e.g. a local authority) or in extreme cases at the individual level (e.g. a surgeon). Bird (2005) identify three aims for public sector PM: to establish "what works" and therefore promoting best practice; to identify the functional competence of organisations or individuals; and accountability by Ministers to Parliament and the general public. These have been termed, respectively, research, managerial and democratic roles of PM. Increasingly, however, the government is using PM as a way of paying for performance (Burgess et al. 2002). An excellent example of this is the Public Service Agreements used by Her Majesty's Treasury.

4.2.1 The rationale for Performance Management

Performance in the public sector attracts significant attention. Stevens (2005) suggests there are two reasons for this: its size and the services that it provides. Government expenditure was £585 billion, £10,100 per capita or 41.7 percent of GDP in the fiscal year 2007-08 (H.M. Treasury 2008). Any inefficiency will therefore have a large impact on the country's welfare. Education, policing and healthcare affect everybody, the vulnerable in particular, and the private sector cannot be relied upon to provide these services due to a host of market failures, for instance:

- The service may be a public good (or quasi-public good) such as policing or national defence involving the problems of non-excludability and non-rivalry.
- The service may be a natural monopoly with enormous infrastructure costs such as the road network.
- A lack of (or asymmetric) information or time inconsistent preferences may result in lower than socially optimal consumption under a free market.
- There may be other market failures such as the presence of negative externalities.

The economic rationale for performance managament is clear and can be considered in a Principal-Agent framework (see Stiglitz 1987, for a brief presentation). In the first instance the principals are the electorate and taxpayers, whilst the agents are elected politicians in local and central government. They are held accountable largely through the ballotbox. In turn, however, the politicians are also the principals and the agents are the civil servants and managers of central government, local authorities, agencies and devolved organisations. Lastly, there is a complex hierarchy between central and local government (e.g. the Department for Transport and local highway authorities) and between central government departments and agencies (e.g. the Department of Health and the National Health Service).

In the private sector, incentives and targets on simple constructs such as profit are available to overcome the principal-agent problem. However, with a lack of prices for the output of the public services and no common yardstick, measuring performance is frought with theoretical, empirical and statistical problems.

4.2.2 How to measure performance

The problems in measuring the output of the public sector frequently occur for the same reasons that these goods and services have to be delivered by the public sector in the first place. There are three main problems with measuring public service performance: identifying outputs, the lack of prices and the problem of attribution. For instance, it may be considered that one of the outputs of local government is road safety. However, the local authority cannot produce "road safety" but undertakes a series of activities (e.g. road safety education or traffic calming and infrastructure improvements) which lead (hopefully) to the production of the output. There is also the problem of attribution. For instance, the level of the output "road safety" will also be affected by external factors such as central government policy e.g. the Department for Transport may introduce new seatbelt or speed legislation. Even when these outputs can satisfactorily be identified and measured they are difficult to aggregate because of the lack of prices resulting from the absence of markets. Prices are critical in economics because of the information they signal (see Deaton & Muellbauer 1980, for a full discussion). Consumers indicate the value of the good or service through their willingness to pay for it, and producers indicate the cost by the price they're willing to accept. Products and services have a range of attributes that are valued by consumers and signalled through prices. For instance, for waste collection provided by a local authority the frequency of collection, where the waste is collected from (roadside or home) and the availability of recycling schemes may all carry different values for each consumer. This lack of prices and the problem of aggregation makes it very difficult to employ traditional economic assessment of efficiency such as parametric Stochastic Frontier Analysis (SFA) (see Kumbhaker & Lovell 2003) or non-parametric Data Envelopment Analysis (DEA) (see Thanassoulis 2001).[1]

Despite these limitations, productivity analysis is still used extensively in the literature on public service performance management. See Martin & Smith (2005) and Hauck & Street (2006) for two recent examples from the health literature. SFA and DEA methods vary in their assumptions and estimation techniques but they both aim to identify the frontier of feasible performance and then estimate a single measure of the (in)efficiency of each organisation against this efficient frontier. The notion of efficiency used is that developed by Koopermans (1951) and Debreu (1951) and introduced into econometrics by Farrell (1957). These models have been employed countless times in the literature and specific software is now available (Coelli (1996*a*) and Coelli (1996*b*)). Outside of the public sector, they are used in industries as diverse as agriculture (Thirtle et al. 2003), banking (Khatri et al. 2002), fisheries (Holloway & Tomberlin 2006), healthcare (Koop et al. 1997) and viticulture (Conradie et al. 2006).

However, a growing number of academics challenge the relevance of productivity analysis to public sector data (Stone (2002*a*), Stone (2002*b*) and Smith & Street

[1]For a general introduction to both methods a good starting point is Coelli et al. (1998).

(2005)). The relevance of a single measure of (in)efficiency is questionable from a managerial perspective. There are also reasonable concerns about the relevance of a production function approach to the analysis of public sector organizations where the production process is not well understood, and probably cannot be modelled well by traditional production analysis. Moreover, the results from SFA and DEA are sensitive to the model assumptions and in particular, the level of inefficiency is dictated by the signal to noise ratio in the data. Given the number of omitted factors in public sector analyses this can be problematic for the correct interpretation of these analyses.

Given the lack of public information available on inputs and the total absence of output prices, the use of production analysis approaches for performance management in road safety is questionable. As a result, this research will focus on the multiple performance indicator approach where by the output "road safety" can be considered as a latent output measured (imperfectly) by a number of observable measures such as the number of road deaths per mode per capita. These measures will be discussed in more detail in section 4.4 on page 73.

4.2.3 Problems with Performance Measurement

In November 1995, Goldstein & Spiegelhalter (1996) read a seminal paper before the Royal Statistical Society (RSS) on the limitations of league tables which was widely discussed and is now a key reference in the literature (Bird 2004). Performance monitoring in the public sector has continued to increase since the Labour government came to power in 1997 (Propper & Wilson 2003). The collection, publication and linking of performance targets to financial resources is now widespread in British Government, and discussion of the target setting "culture" was an important topic of debate during the 2005 general election campaign. Although the ultimate aims of PM are honourable and important (to increase efficiency and garner transparency), the statistics community generally recognises that league tables of PIs are misleading due to the level of uncertainty present (see for example Goldstein & Myers 1996). Perhaps the pinnacle of this obsession with "government by numbers" is that the assessment (i.e. the ranking in the league table) is more important than the (latent) performance.

Any published ranking of organizations (public or private) identifies winners and losers, irrespective of whether these rankings were based on crude, uncontextualised outcome measures or so called "value-added" scores. All rankings are flawed. Research by Goldstein & Spiegelhalter (1996) suggests that rankings adopting robust procedures that incorporate uncertainty in a statistically valid manner frequently make it impossible to separate the organisations under study. There are several reasons for this.

First, no fancy statistical technique can rescue inappropriate, manipulated or otherwise bad data. Secondly, the statistical procedures that contextualise the numbers will produce estimates with a margin of error. This uncertainty creates uncertainty in the rankings and appropriate confidence (or credible) intervals for the rankings typically overlap. Thirdly, the PIs are based on the past and not the current state of the institution. For instance, school exam results are based on students who joined the school many years previously and local authority outcomes may be the result of old policies or staff. And, finally, there will always be omitted factors that could distort comparisons. The identification and measurement of these factors may be very difficult.

In January 2003, the RSS held a discussion meeting on performance management, reported in Bird (2004), which highlighted the extent to which performance indicators and performance management in general escapes the safeguards of national statistics. Consequently, the RSS launched a *Working Party on Peformance Management in the Public Services* which dealt with the statistical aspects of performance management which were identified by theHouse of Commons Public Administration Select Committee (2003). Their report, *Performance Indicators: good, bad and ugly* (Bird 2005) highlights the need for: independent scrutiny of PM schemes; formal and transparent PM protocols; and education for policy makers, politicians, the media, and the wider public on the difficult issues surrounding PM. It also highlighted the role statisticians, and academics in general, should play in both research, education and scrutiny of PM and PI. This research is aimed to meet this need for improved performance management tools.

4.3 Road Safety

4.3.1 The Problem

Worldwide, injuries and death resulting from road traffic accidents are of epidemic proportions. At present over 1 million people die every year and over 10 million people sustain permanent disabilities from road accidents (Bunn et al. 2003). Globally, road accidents are the 10th most common cause of death (Commission for Global Road Safety 2005). The World Health Organisation (WHO) predict road accidents will be the 6th leading cause of death worldwide, and the second leading cause of Disability-Adjusted Life Years (DALYs) lost in developing countries by 2020 (WHO 2004). The scale of the problem is vast and in developing countries is growing exponentially.

The cost of road accidents are generally thought to be high and an important variable in the analysis of transport projects and the formation of transport policy (Evans (1994) and Peirson et al. (1998)). For the UK, estimates of annual external costs of road accidents vary from £3 billion to £26 billion (see Maddison et al. (1996); Pearce (1993); Fowkes et al. (1990); Hansson & Marckham (1992) and Newbery (1988)). Besides the pure economic costs, accidents cause emotional and psychological suffering for family and friends, and pain and suffering for the individuals involved.

Road safety is a vast and active area of research with strong policy support throughout the world. Leading global organisations such as the World Bank, United Nations and World Health Organisation are at the forefront of research and policy design in this field. The FIA Foundation recently launched the Commission for Global Road Safety, supported by the G8, United Nations and World Bank, which is calling for a $300 million ten-year global road safety programme (Commission for Global Road Safety 2005). The need for evidence-based policy in this area has never been greater. This research will therefore contribute to the road safety literature by improving the performance management of local authorities in respect to road safety, the aim being that 'best practice' could be determined by identifying areas of strong performance. In this respect, this thesis will consider a dataset of casualty count for vulnerable road users in London. The justification for this focus is presented next.

4.3.2 Vulnerable Road Users

This thesis considers vulnerable road users which are considred to be pedestrians, cyclists and motorcyclists. Given the recent focus on integrated transport and shifting away from cars to alternative modes of transport, vulnerable road users numbers will increase. This promotion of sustainable transport is supported by iniatives from the Department of Health to promote lifestyle change to tackle obesity and coronary heart disease (Department of Health 2008). If these strategies are successful there will be increasing numbers of vulnerable road users and monitoring and ensuring their road safety will become increasingly important. Moreover, this thesis focuses on vulnerable road users because the underlying causal mechanism for road accidents involving them is similar. Other road users such as car drivers are excluded because the underlying causes, for example excessive speed or drink-driving, ar different. The safety programmes that should be targetted towards reducing vulnerable road accidents would be therefore be different. Performance measurement should be conducted separately for these two broad classes of road users.

Figure 4.1 clearly shows the extent of pedestrian road accidents. Every pedestrian road accident in 2006[2] within London is plotted on a plain, white background. The pattern mapped is easily recognizable as London. Several famous London features are clear including Hyde Park and the River Thames, as well as the main arterial roads and roundabouts.

All pedestrians are at risk in traffic but child pedestrians are particularly vulnerable because they are small and fragile, and their road sense and crossing skills are still developing. Children should be able to walk and cycle in safety, for their social development and to improve their health and fitness. Yet road traffic injury is the leading cause of accidental injury among children and young people (RSAP (2000) and Towner et al. (1993)). Two teenagers are killed or injured crossing London's roads every day (TfL 2006*b*). Across the UK, over 13,000 child pedestrian accidents were reported to the police in 2004, including more than 2,500 serious injuries (such as multiple fractures and extended hospital admission) and almost 100 deaths (DFT 2005).

[2] 2006 was seleted because it was the latest year of data available at the time of analysis.

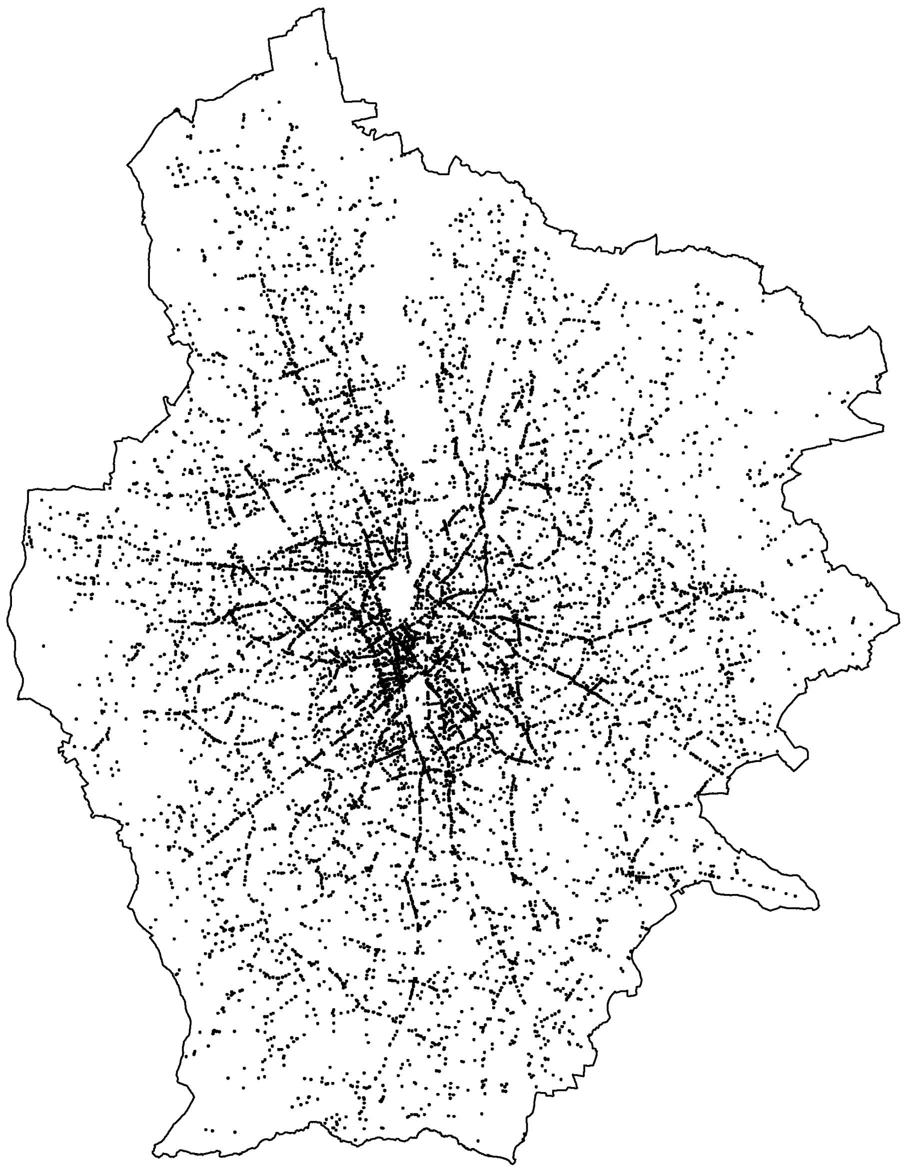

Figure 4.1: A plot of every road traffic accident in London in 2006

Cycling is a healthy, environmentally friendly way to travel, providing quick and cheap transport. Over 350,000 cycle journeys are made in the Greater London area every day, and in the last five years the number of recorded cycle trips has doubled, while the number of cyclists killed or injured has fallen by over 40 per cent (TfL 2008). Given the promotion of healthy and environmentally sustainable transport modes it is not surprising that there is renewed concern for the safety of pedestrians and (pedal) cyclists. However, given the vulnerable nature of these modes of transport, the number of accidents involving pedestrians and cyclists is still high, and in particular the number of killed and seriously injured victims is high.

Powered Two-Wheelers (PTW) is the generic term used in road safety to cover any motorised two-wheeled vehicle, from moped to motorcycle, although this thesis will use the more common term motorcycle or motorbike interchangeably for the catch-all term "powered two-wheelers". Motorcyclists have always been a vulnerable group for a number of reasons, not least of which is the fact that the affects of a collision are more severe for motorcycle riders than car drivers. In 2006 there were 839 riders or passengers killed or seriously injured in London (TfL 2006*a*). This is 9 percent over the target as laid out under the Best Value Performance Indicator, and the only road safety target which Transport for London did not meet. In addition there were 4,297 riders or passengers slightly injured on London's roads.

Motorcyclists are a particular concern for road safety professionals and for Transport for London in particular. There appears to be a significant difference in the type of motorcyclist in London when compared to the rest of the UK. London motorcyclists are more likely to be involved in more accidents, to be younger, higher earners, riding smaller capacity bikes and to be using those bikes for commuting (TfL 2004). Londoners buy motorcycles to avoid congestion and because they are relatively cheap to run and insure. Non-Londoners are more likely to use their motorcycles for freedom or leisure reasons. As such, Londoners are more likely to be classified as "new riders", which may partially explain why causalty statistics have increased in proportion with the number of new bikes registered in London. This is in stark contrast the rest of the UK and in contrast to the general trend of accident numbers which has steadily been in decline for the past decade.

4.3.3 The Solution

Recognizing the importance of road safety, the UK Government set out a strategy to their 2000 road safety white paper, *Tomorrow's Roads: safer for everyone* (DfT 2000). This commitment is resonant of the European Commission's commitment to improving pan-European road safety through the 2003 Road Safety Action Programme (European Transport Safety Council 2006). The focus of the UK strategy is on road safety education with some mention of 20mph zones, Home Zones[3] , increased speed enforcement and other engineering approaches. This is in stark contrast to the European Commission's focus on the harmonisation of European road safety legislation, programmes to improve (and test) drivers' abilities, and vehicle technologies (such as speed limiters, pedestrian recognition systems and adaptive control systems).

The identification of effective strategies for the prevention of road casualties is of major social, economic, political and health importance. But in order to establish what policies and practices work, more needs to be done to provide clear and robust information on the most effective local authorities. This includes research into the

[3] A home zone is a street or group of streets where pedestrians, cyclists and vehicles share the space on equal terms, with cars travelling at little more than walking pace.

causes of accidents, the effects of policy interventions and on producing more robust performance management tools. High quality evidence is required to make informed policy decisions as well as for the effective management of public sector agencies tasked with reducing road casualties.

Without good quality statistical research poor decisions may be made. For instance, superficial presentation of the raw numbers of child pedestrian casualties is taken to support the view that road safety education in schools has been successful (Hewson 2002). Data presented in figure 4.2 demonstrate that casualty rates have fallen consistently in recent decades despite rapidly growing traffic levels. However, there is no empirical evidence as to the cause of the remaining accidents and to the cause of this downwards trend. What is clear is that better information is required. For instance, once we account for the exposure to traffic in figure 4.3 this performance looks less impressive. Thus, the exposure-adjusted risk has been largely stable.

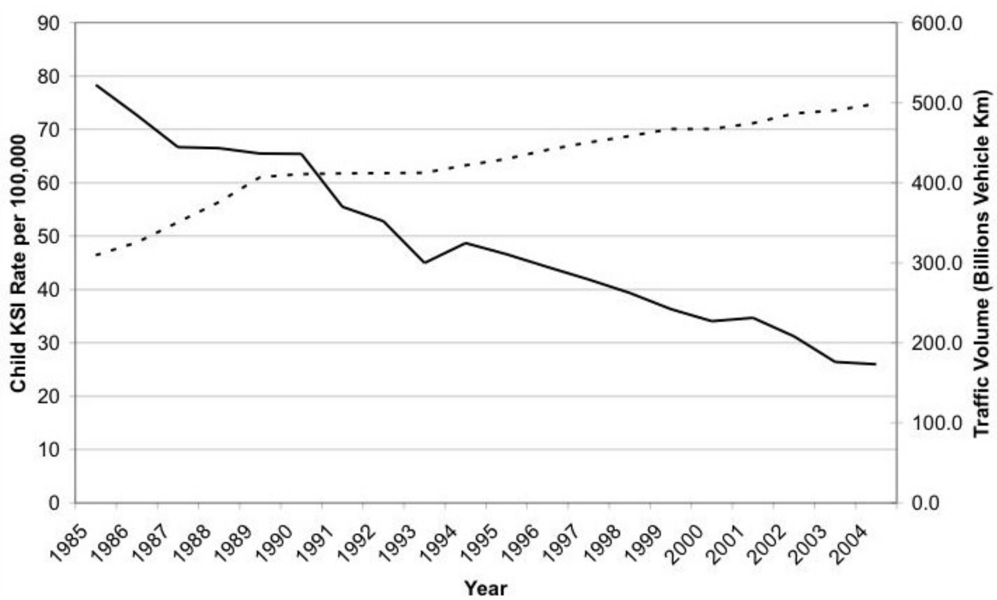

Figure 4.2: Downward Trend in Child Pedestrian Casualties

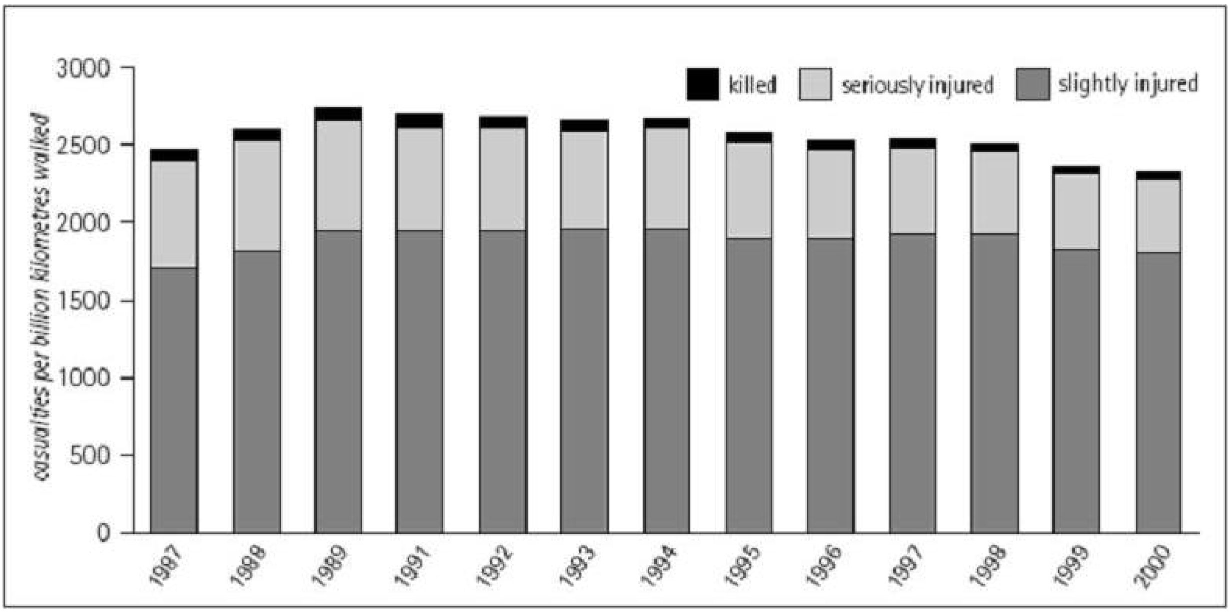

Figure 4.3: Exposure Adjusted Child Pedestrian Casualty Rates

One view, promulgated by a number of road safety researchers including Hewson (2002), Davis (1992), Plowden et al. (1984), and Grayling et al. (2002) (among others) is that increasing awareness of the risk of traffic through road safety education programmes, combined with busier roads and a cultural shift towards sedentary pastimes has led to a reduction in the amount children walk and play. Among the plethora of causes has been an inexorable shift into cars (Hewson (2002), Hillman et al. (1990), Roberts (1993) and Grayling et al. (2002)). The key point here is that reporting of single headline numbers is frequently misleading.

4.3.4 Explaining spatial variation

Very little attention has been paid to the issue of spatial structure in the road safety literature, although there have been many studies that have investigated the spatial variation in the incidence of pedestrian road casualties (including children). The largest body of work in this area seeks to explain the spatial variation through the inclusion variation of spatially varying covariates such as deprivation. For instance, Grayling et al. (2002), Graham et al. (2005) and Graham & Glaister (2005) have all attempted to explain the spatial variation in accident rates through area level deprivation covariates. Others have focused, at varying degrees of aggregation, on the influence of other factors on the spatial variation of accident rates, including Dee (1998) Noland (2001), Noland & Quddus (2002), Noland & Quddus (2004), and McCarthy (1999). In addition, Graham & Glaister (2003) looked at the spatial variation in pedestrian road casualties by examining the role of the urban environment. All of these studies have ignored the fact there may well be spatial dependence and spatial heterogeneity within the data, and as a result continue to apply methods of statistical inference that are not robust to such problems.

Although in themselves excellent studies, none of them have sought to understand and use the intrinsic spatial characteristics of the underlying data generation process. Recently work applied to data from Canada (MacNab 2003) and (MacNab 2004) and to Devon, UK (Hewson 2004) and (Hewson 2005) have sought to model the spatial dependence within the data. However, these studies are limited and the methods proposed have not been adopted by other researchers.

4.4 Road Safety Performance Indicators

As discussed in section 4.2, performance management is widespread in the public sector and is a high profile activity receiving significant attention from politicians, the media and the wider public. Road safety is no exception and is one of the many local government activities that is monitored through outcome based performance indicators. In particular, the UK Government has identified three traffic safety targets which are expected to be achieved by 2010[4]: a 40 percent reduction in the number of fatally or seriously injured casualties, a 10 percent reduction in the rate of slight casualties relative to the level of traffic, and a 50 percent reduction in the numbers of children who were fatally or seriously injured (DfT 2000). Most local authorities incorporate these targets into their Local Transport Plans (LTP) but it is related performance indicators, broken down by modal group, which are monitored and published in the local authority league tables under the 'best value' requirements of the Local Government Act 1999 (Department for Transport, Local Government, and the Regions 1999). It is these indicators called BVPI99 (Best Value Performance Indicators 99) that are used to judge the relative performance of each local authority with respect to road safety and they shall be the focus of this research. According to PACTS (2003*a*)[5], DfT guidance intimates that performance will be used in future to determine financial allocations to local authorities. This is supported by Department for Transport, Local Government and the Regions (2001) which suggests that good and improving performance will attract additional funding and increased autonomy under Public Service Agreements via pump priming. The Comprehensive Spending Review (H.M. Treasury 2007) has also indicated that budget allocation will increase from 5 percent allocated fr relative performance over the coming years.

There is also widespread support from within local authorities for a target-led approach (PACTS 2003*a*). However, there was concern expressed in PACTS (2003*a*) that the headline indicators should be aggregated together because of the variability in individual indicators. Obtaining statistically reliable results on performance is

[4]These were set relative to a baseline of the mean number of casualties that were reported between 1994 and 1998 inclusively

[5]PACTS is the Parliamentary Advisory Council for Transport Safety (PACTS) is a registered charity and an associate Parliamentary Group.

therefore a major concern of authorities themselves who do not possess the expertise and experience to produce robust methods internally.

These current UK traffic safety performance indicators are expressed simply in the form of crude per capita numbers of reported collisions by type and modal group, with no allowance for geographically differing patterns in road infrastructure and usage or spatially varying socioeconomic conditions, so called contextualisation in the literature (Goldstein & Spiegelhalter 1996). In fact, according to Bailey & Hewson (2004) there is no explicit consideration given to the extent to which differences in the raw rates reflect differential performance, rather than just inherent random variability in observed rates. In general, local government activity does not appear to have received anything like as much attention in the literature as that devoted to performance monitoring in other sectors. For instance, although local government (in the UK at least) plays a significant role in education, performance monitoring interest in that sector has largely focused on the school as the observational unit, rather than on the Local Educational Authority.

There have been significant developments in the modelling of performance and the uncertainty inherent in performance rankings (or league tables) in the education and health literatures. For example, see Laird & Louis (1989), Goldstein & Spiegelhalter (1996), Marshall & Spiegelhalter (1998), Lockwood et al. (2002), Kuhan et al. (2002), Draper & Gittoes (2004), Bratti et al. (2004) and references contained within. Unfortunately, this work has not been transferred to the area of local government performance management and performance indicators remain crude, uncontextualised numbers.

Although detecting a (statistically significant) departure in road safety performance between organisational units is only part of the larger picture of performance management, it is still important to ascertain if there is differential performance. This can, for example, indicate the need for further research e.g. detailed auditing to identify best practice and/or intervene in poor performaning authorities. It is also important to make the best possible statistical inference about the road safety performance of local authorities when their performance is tied to financial resources and Public Sector Agreements with Her Majesty's Treasury. Limited work has been done in this area. For instance, MacNab (2003) employed generalized additive modelling to smooth out year-in-year variance in area level accident rates in Canada. He emphasized the importance of separating signal from noise when investigating such "noisy" data. More generally, the Generalized Linear Mixed Model (GLMM) presented in section 2.5.10 is now well established in the wider literature on PM to model to uncertainty associated with performance indicators. See, for example, applications in the healthcare literature by Marshall & Spiegelhalter (1998) and Morris & Christiansen (1996).

The GLMM is discussed in more detail in chapters 5 and 7 but interested readers are directed to the appendix of Marshall & Spiegelhalter (1998) for a very concise and readable introduction to GLMM in the context of performance management. Briefly, the GLMM is characterized by the inclusion of a random effect as well as the traditional fixed effects of the GLM. In section 2.5.10 this random effect was used to capture the local spatial variation in the data. In the performance management context, the random effect captures the local authority specific performance that has not been directly measured: i.e. it can be considered a parameter for the latent (unobservable) local authority performance. It is this underlying construct of road safety that is important for performance management. This is in contrast to a fixed parameter (intercept) that would simply reproduce the observed performance indicator that was observed for each local authority. The random effect is a zero-mean process with unknown variance that is estimated when the model is fitted i.e. it explictly incorporates unknown uncertainty into the model for the performance indicators. The unknown variance component induces smoothing or shrinkage toward the global (zero mean) in the authority specific random effect depending upon the strength of evidence in the local authority's performance – it is therefore analogous to a partial pooling of the data to improve the efficiency of the estimation (see Gelman & Hill 2006, for a presentation).

Specifically within the road safety performance management literature, Papageorgiou & Loukas (1988) consider a bivariate binomial model for road safety in East Virginia based on the theory that fatalities should be correlated with injuries. More broadly, it is sensible to consider that there would be correlation between multiple road safety indicators because inherently they can all be considered as measures of the latent construct "road safety". In light of this Bailey & Hewson (2004) model a multivariate GLMM for nine road safety performance indicators, because the local authority specific random effects can borrow strength across the multiple variables to produce a better estimate (i.e. with smaller variance) of the local authority specific performance. If the variance is smaller than so will any credible interval used in ranking the organizations and therefore it may be possible to separate the performance between local authorities. Bailey & Hewson (2004) do find considerable shrinkage of the credible intervals and therefore an improvement in the rankings of local authorities. However, they report that the resulting league table still remains quite "fuzzy" with a great deal of overlap between authorities still remaining.

One potential reason for their findings remaining inconclusive is that they omitted to include any potentially important explanatory or contextualising variables.

Additionally, they failed to consider the spatial dependence in the data. Moreover, they applied their model to the entire UK dataset with full exchangeability inherent in the model. This is a questionable assumption. For instance, it is reasonable to expect a significant difference between rural and urban authorities or between metropolitan and county council areas. This thesis will model only the 32 boroughs of Greater London including a multivariate spatial model to improve inference in the model.

4.5 Summary

This chapter motivated the empirical objective of the thesis through a review of performance measurement. In particular, it argued that a common problem with performance indicators in the public services is that they usually make no allowance for the inherent uncertainty in both the underlying performance being measured or any rankings of this performance. Moreover, when attempts are made to incorporate uncertainty into performance measurement – for example through the use of generalized linear mixed models (GLMMs) – the resulting credible intervals relating to the performance rank are typically large and overlapping. The obvious result is that it becomes impossible to differentiate the relative performance of organisations. Given the increasing reliance on performance management in the public sector, and the trend towards ‘payment for performance’, this chapter argued that improving performance measurement was a worthy endeavour. Bird (2004) argues that the statistician plays an important role in safeguarding those that are monitored from misconceived reactions to uncertainty and to design effective performance monitoring tools. The important role that academics can play in deriving sensible PIs is also discussed by Stone (2002*a*). This thesis aims to address both of these roles by investigating the Best Value Performance Indicators for road safety. The concerns of PACTS (2003*a*) on the large variation in the individual measurements is addressed by adopting a multivariate approach that will allow for shrinkage in the estimators of performance and an improvement in the precision of these estimates. Additionally, it is hoped that more robust rankings of institutions will be produced by correcting for spatial dependence in the data and any omitted spatially varying variables as highlighted in Bailey & Hewson (2004). This will be achieved by the use of the multivariate spatial model to be developed in chapter 5. Robust rankings that properly account for uncertainty in the positions will be produced by adopting a Bayesian perspective (Lilford & Braunholtz 1996) and using Markov chain Monte Carlo techniques which will be presented in chapter 7. The next chapter therefore introduces a new, flexible multivariate CAR model into the literature which will later be used to model multiple road safety performance indicators.

CHAPTER 5

A FLEXIBLE MCAR MODEL

5.1 Introduction

As reviewed in chapter 2, Gaussian conditional autoregressions (CAR) have been used extensively to model the association between univariate random variables at sites on both regular and irregular lattices. Yet, as chapter 3 demonstrated, there is a rather limited body of work on conditional autoregressions for multivariate lattice data, so called MCAR models. In comparison, there is an extensive literature on geostatistical approaches to modelling multivariate data. The principal challenge when modelling multivariate lattice data is to develop conditional models that guarantee valid covariance matrices in the joint probability model whilst allowing for correlation both between variables *within* sites and between variables *across* sites. An additional complexity would be to allow these correlations to be asymmetric, something which ver Hoef & Cressie (1994) consider for the continuous (i.e. geostatistical) approach.

Given two variables, X and Y recorded at two locations i and j then symmetric spatial cross correlation would mean that $\rho_1(X_i, Y_j) = \rho_2(X_j, Y_i)$ where ρ captures the spatial cross correlation. Allowing these two cross correlations to be different i.e. asymmetric will be the key methodological contribution of this chapter. More concretely, consider two of the road safety performance indicators introduced in Chapter 4: severe motorcycle casualties and severe cyclist casualties and two neighbouring locations from London, Waterloo and London Bridge. Asymmetric spatial cross correlations allows the relationship between motorcycle accidents at Waterloo and cycling casualties at London Bridge to be different from the relationship between motorcycle accidents at London Bridge and cycling accidents at Waterloo. Differential relationships like the one posited here could occur for a range of unobserved or unrecorded reasons such as infrastructure differences, differing road user prioritisation and differing road safety policies adopted by Southwark and Lambeth councils (in which London Bridge and Waterloo are located).

Mardia (1988) provided the theoretical groundwork for multivariate Gaussian CAR models extending the seminal univariate work of Besag (1974). The problem with Mardia's original multivariate specification was that it required separable models that necessitated identical spatial parameters for each variable. The "two-fold CAR" model of Kim et al. (2003), which was described in chapter 3, provides a more flexible correlation structure incorporating both *bridging*[1] and *linking*[2] spatial parameters. However, this model is only suitable for the bivariate case and extension to higher dimensions seems problematic. The MCAR models of Gelfand & Vounatsou (2003) and Carlin & Banerjee (2003) are almost identical in their approach, although Carlin & Banerjee (2003) extend their model to spatio-temporal data. These MCAR models are suitable for non-separable models but do not allow for flexible between area correlations. In response to this rather limited cross-correlation structure, Jin et al. (2005) propose the Generalized MCAR model that specifies the joint distribution for a Markov random field in terms of a combination of simpler conditional and marginal distributions. In such, they are adapting the multivariate geostatistical model of Royle & Berliner (1999) to the lattice case. As discussed in chapter 3, one concern

[1]Bridging parameters refer to correlation coefficients between different variables at the same site.

[2]Linking parameters refer to spatial correlation coefficients that link different variables at different sites.

with this model is the rather arbitrary order in which the conditional and marginal variables are considered i.e. should $\pi(\boldsymbol{y}_1|\boldsymbol{y}_2)$ be modelled and then $\pi(\boldsymbol{y}_2)$ or the other way around. Jin et al. (2005) propose to use model comparison techniques such as the Deviance Information Criterion (DIC) to decide on the modelling order, yet this seems infeasible with more than a few variables.

Many problems in econometrics are multivariate and increasingly spatial datasets that record multiple observations for each location are becoming available. This thesis will consider the spatial dependence in a set of road safety performance indicators for Greater London local authorities which were briefly introduced in chapter 4. Chapter 7 will attempt to improve the measurement of the underlying or latent road safety performance by applying the model developed in this chapter to the data on road safety Best Value Performance Indicators. This will then be used to rank and select the top performing authorities from these results. Chapter 3 reviewed the current MCAR models for lattice data. The main problem with the existing approaches is the difficulty in relaxing the conditions on the cross-correlations whilst maintaining the propriety of the covariance matrix. In response, a very flexible model is introduced for multivariate spatial data recorded on a lattice whilst providing conditions to ensure a non-singular covariance matrix and hence a proper joint distribution. This is an important step beyond what is currently available to researchers.

5.2 Univariate Review

Recall from section 2.5.1 that when the variable y is univariate that, given some mild consistency conditions as given by Besag (1974), the full or undirected conditional distributions

$$\pi(y_i \mid \boldsymbol{y}_j : j \in \partial_i), \qquad i = 1, ..., n \tag{5.2.1}$$

determine a valid joint distribution

$$\pi(y_1, ..., y_n) \tag{5.2.2}$$

which is called a Markov Random Field (MRF). Gaussian MRFs are called Conditional Autoregression (CAR) models and these were introduced in chapter 2. Assuming that these n full conditionals are all Gaussian, the i^{th} distribution can be given as

$$\mathrm{E}[y_i \mid \boldsymbol{y}_j : j \in \partial_i] = \mu_i + \sum_{j \in \partial_i} b_{ij}(y_j - \mu_j), \tag{5.2.3}$$

$$\mathrm{Var}[y_i \mid \boldsymbol{y}_j : j \in \partial_i] = {\tau_i}^2. \tag{5.2.4}$$

Together these n full conditional distributions yield the joint distribution for $\boldsymbol{y}$

$$\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu}, (\boldsymbol{I} - \boldsymbol{B})^{-1}\boldsymbol{\Gamma}), \tag{5.2.5}$$

where $\boldsymbol{y} \equiv (y_1, ..., y_n)'$, $\boldsymbol{\mu} \equiv (\mu_1, ..., \mu_n), \boldsymbol{\Gamma} \equiv \mathrm{diag}(\tau_1^2, ..., \tau_n^2)$, and $\boldsymbol{B} \equiv (b_{ij})$. In the presence of explanatory variables, $\boldsymbol{\mu}$ can be reparameterized as ($\boldsymbol{\mu} = \boldsymbol{X\beta}$). (See section 2.5.1 for full details.) As discussed previously, for the joint distribution to be well defined $(\boldsymbol{I} - \boldsymbol{B})^{-1}\boldsymbol{\Gamma}$ must be a symmetric, positive-definite matrix. A common reparameterization of the model to ensure this condition holds is the so called intrinsic autoregression (IAR) of Besag et al. (1991) which requires $(\boldsymbol{D} - \alpha\tilde{\boldsymbol{W}})$ where $\tilde{\boldsymbol{W}}$ is a row stochastic proximity matrix, $\boldsymbol{D} = \mathrm{diag}(w_{i+})$ and w_{i+} are the number of neighbours to site i. As long as $|\alpha| \in (-1, 1)$ then Banerjee et al. (2004) show that the model leads to a valid covariance matrix and hence a valid joint distribution.

5.3 The FMCAR

inputfmar.tex

5.4 Ensuring the Existence of the Covariance

inputcov.tex

5.5 Interpreting the Spatial Parameters

inputparameter.tex

5.5.1 Conditional Correlations

5.5.2 A Brief Example

5.6 Unconditional Correlations

inputcorr.tex

5.7 Precision Measures

inputtau.tex

5.8 Implementation

This section deals with the implementation of the FMCAR model i.e. to the estimation of model parameters. A hierarchical framework[3] is used for pragmatic

[3]Also frequently called multilevel models, and sometimes by the more specific terms generalized linear mixed models, nested models, mixed models, random coefficient, random-effects models and

and theoretical reasons. The theoretical reasons for hierarchical modelling is best explained in Gelman & Hill (2006) but include shrinkage of estimators and increased efficiency of estimation, achieved largely through the (partial) pooling of data across levels of the model. There are also pragmatic reasons for employing a hierarchical framework, which centre on the desire to use Bayesian methods of inference for the estimation of model parameters. Bayesian methods are best modelled in a hierarchical framework because they easilylend to the inclusion of prior information in the model. Lastly, hierarcichal models provide a simple way to deal with the generalized linear model which is a well established method of modelling count data.

5.8.1 Hierarchical GLM

In most applications, CAR and MCAR models are not typically used as models for data but as priors for a spatial random effects parameter in a hierarchical model. Following the rest of the literature, this section demonstrates how to implement the FMCAR as a prior in a Bayesian hierarchical framework for count data. Without loss of generality, consider the bivariate case ($p = 2$) and let $\boldsymbol{Y}(\boldsymbol{s}_i) = (\boldsymbol{y}_1(\boldsymbol{s}_i), \boldsymbol{y}_2(\boldsymbol{s}_i))' \equiv (\boldsymbol{y}_{i1}, \boldsymbol{y}_{i2})'$ denote the observed accident casualty counts for two modal types (e.g. pedestrians and cyclists) respectively for each site $i = 1, ..., n$ (e.g. London boroughs or Lower Super Output Areas). The two variables could be generalized to any counts and the model can be generalized to any number of variables although, in this implementation, $p = 2$.

As proposed in chapter 2, a standard method for modelling such count data would be via Generalized Linear Modelling (see McCullagh & Nelder (1989) for details),

---

random parameter models. Multilevel or hierarchical model is the preferred label for this approach due to the confusion among disciplines as to exactly what is meant by random coefficients, effects and parameters.

which is used extensively in the literature on road casuality modelling – see, for instance, Bailey & Hewson (2004), Graham et al. (2005) or MacNab (2003). The first level of the hierarchical GLM model (the data model) will follow the convention from the road safety literature which relates the observed accident counts to some reference population level. This is a standard approach taken from epidemiology where the expected counts are calculated from the underlying population at risk in any particular geographic area. At the second stage of the model, the process model, departures from the expected level are explained by regression effects (e.g. exposure to traffic) or stochastic spatial-dependence modelled through a spatial random effects parameter.

Formally then, the data model can be written as:

$$\boldsymbol{Y}_{ik} \mid \theta_{\mathbf{ik}} \sim \mathrm{Poi}(\mathrm{E}_{ik}\mathrm{e}^{\theta_{ik}}), \qquad i = 1, ..., n; \quad k = 1, ..., p \tag{5.8.1}$$

where the E_{ik} are expected counts for variable k in site i derived from the standardized population. Chapter 7 provides further details on the standardization methods used in this application but the literature is extensive – see Mantel & Stark (1968). Departures from the expected counts are modelled by the parameter, $\theta_{\mathbf{ik}}$.

The process model, or second stage of the hierarchical GLM, models these departures from the expected counts through a combination of regression effects and stochastic spatial dependence parameters. The common notation of an n x p matrix, $\boldsymbol{X}$, of regression parameters with a q x p matrix, $\boldsymbol{\beta}$ of regression coefficients is adopted for these regression effects. The spatial dependence will be modelled using the FMCAR model. By using the vector operator to stack columns of the matrix so that $\boldsymbol{\theta}^v \equiv \mathrm{vec}(\boldsymbol{\theta}')$, the second stage of the GLM can be written as:

$$\boldsymbol{\theta}^v \mid \boldsymbol{\beta}, \boldsymbol{V}, \boldsymbol{C} \sim \mathcal{N}_{np}(\boldsymbol{\mu}^v, \boldsymbol{\Sigma}), \tag{5.8.2}$$

where $\boldsymbol{\mu}^v = \text{vec}(\boldsymbol{\mu}')$, $\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta}$, $\boldsymbol{V} = \boldsymbol{\Gamma}^{-1}$, $\boldsymbol{M}_i \equiv \text{diag}(\text{E}_{\text{i1}}, ..., \text{E}_{\text{ip}})$, $i = 1, ..., n$ and $\boldsymbol{\Sigma}$ is as defined in equation (**??**).

To complete the hierarchical GLM specification, the third level or priors must be specified for the matrices, $\boldsymbol{\beta}, \boldsymbol{C}$, and $\boldsymbol{V}$. Assume that the regression coefficients have a normal prior distribution, so that each column vector ($\beta_{\mathbf{k}}$) of the matrix $\boldsymbol{\beta}$ for each of the p dependent variables ($k = 1, ..., p$) can be written as:

$$\beta_{\mathbf{k}} \sim \mathcal{N}_p(\mathbf{0}, \sigma^2 \boldsymbol{I}). \tag{5.8.3}$$

The usual choice of prior for precision (inverse covariance) matrices is the Wishart distribution (Mardia et al. 1979), which is adopted here,

$$\boldsymbol{V} \sim \text{Wishart}(\rho, (\rho \boldsymbol{A})^{-1}), \tag{5.8.4}$$

where $\rho > p$ and $\boldsymbol{A}$ is a predetermined symmetric positive-definite matrix. An obvious candidate would be $\boldsymbol{A} = \boldsymbol{I}$. The choice of the final prior, for $\boldsymbol{C}$, is not straightforward however. It is selected to be proportional to

$$\exp\{-(\boldsymbol{C}^v)' \boldsymbol{C}^v / \xi^2\} \tag{5.8.5}$$

where $\boldsymbol{C}^v \equiv \text{vec}(\boldsymbol{C})$. The prior distribution must be truncated to ensure that only values of $\boldsymbol{C}$ that provide for a positive-definite $\boldsymbol{G}$ (see section 5.4) are permitted. A hyperprior distribution for ξ is not used but values of ξ are predetermined instead. If a hyperprior for ξ was used the computational burden on the model would be significant. Smaller values for ξ are preferred as it results in a peaked prior distribution for $\boldsymbol{C}$

centered around zero. Therefore, posterior values of $\boldsymbol{C}$ far from zero are strong statistical evidence for spatial dependence in the data. Note that use of this prior for $\boldsymbol{C}$ does not favour either symmetric or asymmetric spatial dependence.

The posterior distribution of this hierarchical model can be obtained by the simple application of Bayes' theorem, which yields:

$$\pi(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{V}, \boldsymbol{C} \mid \boldsymbol{Y}) \propto \pi(\boldsymbol{Y} \mid \boldsymbol{\theta})\pi(\theta \mid \boldsymbol{\beta}, \boldsymbol{V}, \boldsymbol{C})\pi(\boldsymbol{\beta})\pi(\boldsymbol{V})\pi(\boldsymbol{C}). \tag{5.8.6}$$

Substituting in the model specification at each of the three levels of the hierarchical model gives

$$\pi(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{V}, \boldsymbol{C} \mid \boldsymbol{Y}) \propto \tag{5.8.7}$$

$$\prod_{i=1}^{n}\prod_{k=l}^{p} \exp(-\mathrm{E_{ik}}\mathrm{e}^{\theta_{ik}})(\mathrm{E_{ik}}\mathrm{e}^{\theta_{ik}})^{\boldsymbol{Y}_{ik}} \tag{5.8.8}$$

$$\mathrm{x} \mid \boldsymbol{V} \mid^{n/2} \mid \boldsymbol{G} \mid^{1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{\theta}^v - \boldsymbol{\mu}^v)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}^v - \boldsymbol{\mu}^v)\right\} \tag{5.8.9}$$

$$\mathrm{x}\ \exp\left\{-\frac{1}{2\sigma^2}\boldsymbol{\beta}_1'\boldsymbol{\beta}_1\right\} \mathrm{x} \cdots \mathrm{x} \exp\left\{-\frac{1}{2\sigma^2}\boldsymbol{\beta}_p'\boldsymbol{\beta}_p\right\} \tag{5.8.10}$$

$$\mathrm{x} \mid \boldsymbol{V} \mid^{(\rho-p-1)/2} \exp\left\{-\frac{\rho}{2}\mathrm{tr}(\boldsymbol{A}^{-1}\boldsymbol{V})\right\} \tag{5.8.11}$$

$$\mathrm{x}\ \exp\left\{-(\boldsymbol{C}^v)'\boldsymbol{C}^v/\xi^2\right\}. \tag{5.8.12}$$

5.8.2 Statistical Inference

Statistical inference is carried out in the Bayesian paradigm where the objective is to obtain the posterior distribution given by Bayes' theorem. Frequently these distributions are high-dimensional and numerical integration or analytic solutions are not feasible. One of the major limitations to the widespread implementation of Bayesian methods is that frequently the posterior distribution is not available analytically and therefore high dimensional numerical integration is required to conduct statistical inference. Computationally, this can be very difficult. Several alternative approaches that do not rely on direct integration of these high dimensional functions have been proposed in the literature; reviews can be found in Tanner (1996), Smith (1991) and Evans & Swartz (1995). The dominant approach is currently *Markov chain Monte Carlo* (MCMC). MCMC methods are so-called because they use previous sample values to randomly generate the next sample value thus generating a Markov chain. Excellent introductions to MCMC methods are Robert & Casella (2004), Gamerman (1997) and the edited volume by Gilks et al. (1996). A briefer and more recent reference is Besag (2001) which covers recent innovations such as Langevin-Hastings diffusions. MCMC is approximately 50 years old and originated in physics, yet it came to the attention of a wider audience when Geman & Geman (1984) used the Gibbs sampler to sample from the joint probability distribution in a spatial imaging problem. The signifcance of this method was reviewed in section 2.4.1 when the Hammersley-Clifford theorem was introduced.

The introduction of the Gibbs sampler by Geman & Geman (1984), which is simple yet as Gelfand & Smith (1990) demonstrated applicable to a broad class of problems, generated significant interest in MCMC methods for statistical inference. Although MCMC has been widely adopted and adapted by Bayesian statisticians it is important to separate the Bayesian paradigm from MCMC methods. MCMC is simply a method for conducting approximate high dimensional numerical integration and although it is used heavily in Bayesian statistics it is also useful in Frequentist statistics. Besag (2001) goes to significant effort to make this distinction clear by demonstrating the usefulness of MCMC methods in Frequentist analysis, for example the exact Monte Carlo p-values of Barnard (1963) or maximum likelihood estimation.

Bayesian statistics in general, and MCMC techniques in particular, are well established and accepted, and this thesis will not discuss the underlying 'philosophical' or statistical debates surrounding these methods. Recommended textbooks on Bayesian methods include Gelman et al. (2003) and Lancaster (2004); the later adopts an econometric perspective. For a more rigorous treatment of the underlying mathematical statistics, readers should consult Bernardo & Smith (1994), O' Hagan (1994) or Berger (1985). A comprehensive treatment of MCMC methods is given in Gilks et al. (1996) or Gamerman (1997), and more succinct coverage can be found in either Besag (2001) or Green (2001). For discussion of the frequentist-Bayesian 'debate' readers are directed to Bernardo (2003), Berger (1985) and de Finetti (1970), and references therein. For information and advice on implementing MCMC methods for Bayesian inference in Matlab consult LeSage (1999), in R see Gelman et al. (2003) and in Python or C refer to Wilkinson (2008). These authors also provide freely available code for implementing basic MCMC samplers all of which can be ammended to implement the FMCAR model.

It is possible to implement MCMC methods in any computer language or matrix computing environment and each option has trade-offs in terms of financial cost, opportunity cost and computational cost. Two of the most popular ways to conduct Bayesian analysis is either through the extensive MATLAB econometrics toolbox compiled by James P. LeSage[4] or by using one of the derivatives of the BUGS software (Bayesian inference Using Gibbs Sampling), which is introduced below. Many econometricians and statisticians continue to write their own MCMC samplers. Throughtout this thesis the Python language is used because it offers the power of an

[4] Available freely along with a 350 page e-book from http://www.spatial-econometrics.com

object-orientated language without the usual overheads of development and training time. Python is free as well as incredibly easy to learn and deploy. Compared to other languages it is very succinct and runs across all operating and hardware configurations. Additionally, it works very well with the R statistical environment as well as with leading Geographical Information Systems (GIS) such as the commercial ArcGIS suite (from ESRI) or the open source GRASS[5] implementation. ArcGIS is used for the geographic elements of this thesis and basic data analysis and figure generation is done in R. The appendix provides further details.

For researchers who do not want to programme their own MCMC alogrithms or are not comfortable with amending the vast amount of computer and application code already available there is BUGS (Bayesian inference Using Gibbs Sampling), a self-contained piece of computer software for performing Bayesian analysis using MCMC methods. There are many different versions of the BUGS software[6], which was originally written for DOS systems. These include WinBUGS[7], OpenBUGS[8], JAGS[9], and BRugs[10]. There are also tools in the statistical languages R and S useful for convergence diagnostics[11] which are also built into BUGS and its derivatives.

Sampling from the posterior in (5.8.7) requires the use of an MCMC sampler.

[5]Geographical Resources Analysis Support System.

[6]All of the versions introduced here can be downloaded, for free, from http://www.mrc-bsu.cam.ac.uk/bugs/. There is also a link to JAGS from this website.

[7]Probably the most widely known Bayesian computer "application" this version runs on the Microsoft Windows platform and was developed by staff at Imperial College London.

[8]This is an open source version of BUGS which runs natively on Linux systems and Intel-based Mac computers and was developed by the University of Helsinki

[9]Just Another Gibbs Sampler is a re-written version of BUGS for UNIX users that uses the same syntax and model description tools as BUGS. It is written by Martyn Plummer.

[10]an R interface to OpenBugs

[11]CODA or Convergence Diagnostics and Output Analysis for R and S is available from http://www-fis.iarc.fr/coda/.

Gibbs sampling is applicable when the joint distribution is not known explicitly, but the conditional distribution of each variable is known. The Gibbs sampling algorithm generates a sample or draw from the distribution of each variable in turn, conditional on the current values of the other variables. It has been shown (see, for example, Gamerman (1997) for a thorough explanation) that the sequence of samples produced by the sampler constitutes a Markov chain, and the stationary distribution of that Markov chain is just the required joint distribution.

The Gibbs sampling algorithm is actually a special case of the Metropolis-Hastings (Metropolis et al. (1953) and Hastings (1970)) algorithm, although it is usually faster and easier to implement. However, the Gibbs sampler is actually less useful in practice. Unlike the Gibbs sampler which relies on sampling from simple, univariate probability distributions, the Metropolis-Hastings algorithm can draw samples from any probability distribution $\pi(x)$, requiring only that a function dominating the density can be calculated at x. Similarly to the Gibbs sampler, the Metropolis-Hastings algorithm generates a series of autocorrelated samples using a proposal density $Q(x', x^t)$, which depends on the current state x^t, to generate a new proposed sample x'. This proposal is 'accepted' as the next value ($x^t + 1 = x'$) if α drawn from $\mathcal{U}(0, 1)$ satisfies

$$\alpha < \frac{\pi(x')Q(x^t|x')}{\pi(x^t)Q(x'|x^t)}.$$

If the proposal is not accepted, then the current value of x is retained: $x^t + 1 = x^t$. The Gibbs sampler simply has an $\alpha = 1$ i.e. the sample is always accepted. Chib & Greenberg (1995) provide an intuitive introduction to the Metropolis-Hastings method and Gamerman (1997) provides a comprehensive treatment. Many advances have been made in the field of MCMC techniques but these two alogrithms are adequate to sample efficiently from the posterior (5.8.7) of the FMCAR.

To implement the FMCAR thousands of samples are produced from the posterior (5.8.7) using Metropolis-Hastings steps within a Gibbs sampler. After discarding some initial samples called 'burn-in'[12] because they will be autocorrelated, quantities of interest (e.g. the expectation) can be calculated from the posterior. This will be discussed in more detail in chapter 7. For (5.8.7), one iteration of the Gibbs sampler requires sampling from:

$$1.\quad \pi(\boldsymbol{\beta}_k \mid \boldsymbol{\beta}_{-k}, \boldsymbol{V}, \boldsymbol{C}, \boldsymbol{\theta}), \quad k = 1, ..., p \tag{5.8.13}$$

$$2.\quad \pi(\boldsymbol{V} \mid \boldsymbol{\beta}, \boldsymbol{C}, \boldsymbol{\theta}), \tag{5.8.14}$$

$$3.\quad \pi(C_{kl} \mid \boldsymbol{\beta}, \boldsymbol{V}, \boldsymbol{\theta}), \quad k, l = 1, ..., p \tag{5.8.15}$$

$$4.\quad \pi(\boldsymbol{\theta}_i \mid \boldsymbol{\theta}_{-i}, \boldsymbol{\beta}, \boldsymbol{V}, \boldsymbol{\Gamma}, \boldsymbol{Y}), \quad i = 1, ..., n \tag{5.8.16}$$

where $\boldsymbol{\beta}_{-k}$ denotes all of the columns of the regression coefficient matrix, $\boldsymbol{\beta}$ except for the kth and similarly for the parameter matrix, $\boldsymbol{\theta}_{-i}$. The term C_{kl} in step 3 represents the (k, l)th element of $\boldsymbol{C}$.

Deriving the conditional distributions for each step in the Gibbs sampler requires examining the terms in (5.8.7) that involve that parameter. This is done as follows.

Conditional distribution of $\boldsymbol{\beta}_k$

The conditional distribution is

$$\pi(\boldsymbol{\beta}_k \mid \boldsymbol{\beta}_{-k}, \boldsymbol{V}, \boldsymbol{C}, \boldsymbol{\theta}) \propto \exp\left\{-\frac{1}{2}(\boldsymbol{\theta}^v - \boldsymbol{\mu}^v)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}^v - \boldsymbol{\mu}^v)\right\} \exp\left\{-\frac{1}{2\sigma^2}\boldsymbol{\beta}_k'\boldsymbol{\beta}_k\right\} \tag{5.8.17}$$

The distribution in (5.8.17) can be simplified by reordering the elements in the quadratic term. Instead of ordering the stacked parameter vector by site ($i = 1, ..., n$)

[12] A good introduction to this and other key concepts of MCMC methods for the unfamiliar is Green (2001).

and then by variable ($k = 1, ..., p$) it could be ordered by variable and then by site. Hence, instead of writing

$$\boldsymbol{\theta}^v = (\theta_{11}, \theta_{21}, \cdots, \theta_{1p}, \cdots, \theta_{np})' \tag{5.8.18}$$

consider reordering the stacked vector as

$$\boldsymbol{\theta}^* = (\theta_{11}, \theta_{21}, \cdots, \theta_{n1}, \cdots, \theta_{np})'. \tag{5.8.19}$$

The result of this reordering is that the parameter matrix $\boldsymbol{\theta}^*$ has an np x np covariance matrix of the form Block($\boldsymbol{S}_{kl}$) where the elements ($\boldsymbol{S}_{kl}$) are n x n matrices which are derived below.

The conditional distribution in (5.8.17) becomes

$$\exp\left\{-\frac{1}{2}\left[(\boldsymbol{\theta}_k^* - \boldsymbol{\mu}_k^*)'\boldsymbol{S}_{kk}(\boldsymbol{\theta}_k^* - \boldsymbol{\mu}_k^*) + 2(\boldsymbol{\theta}_k^* - \boldsymbol{\mu}_k^*)'\sum_{l \neq k}(\boldsymbol{\theta}_l^* - \boldsymbol{\mu}_l^*)\right]\right\} \exp\left\{-\frac{1}{2\sigma^2}\boldsymbol{\beta}_k'\boldsymbol{\beta}_k\right\}, \tag{5.8.20}$$

where $\boldsymbol{\theta}^*$ is written as $((\boldsymbol{\theta}_1^*)', ..., (\boldsymbol{\theta}_p^*)')'$. Reparameterizing the mean as $\mu_k^* = \boldsymbol{X}\boldsymbol{\beta}_k$, (5.8.20) can be rewritten as

$$\exp\{-\frac{1}{2}\left[\boldsymbol{\beta}_k'\left(\boldsymbol{X}'\boldsymbol{S}_{kk}\boldsymbol{X} + \frac{1}{\sigma^2}\boldsymbol{I}\right)\boldsymbol{\beta}_k - 2\boldsymbol{\beta}_k'\left(\boldsymbol{X}'\boldsymbol{S}_{kk}\boldsymbol{\theta}_k^* + \boldsymbol{X}'\sum_{l \neq k}\boldsymbol{S}_{kl}(\boldsymbol{\theta}_l^* - \boldsymbol{X}\boldsymbol{\beta}_l\right)\right]\} \tag{5.8.21}$$

The conditional distribution in (5.8.21) is proportional to a multivariate normal distribution with covariance matrix

$$\boldsymbol{\Sigma}_{\boldsymbol{\beta_k}} = \left(\boldsymbol{X}'\boldsymbol{S}_{kk}\boldsymbol{X} + \frac{1}{\sigma^2}\boldsymbol{I}\right)^{-1}, \tag{5.8.22}$$

and with the mean given by

$$\boldsymbol{\mu}_{\boldsymbol{\beta_k}} \equiv \boldsymbol{\Sigma}_{\boldsymbol{\beta_k}} \left[\boldsymbol{X}'\boldsymbol{S}_{kk}\boldsymbol{\theta}_k^* + \boldsymbol{X}'\sum_{l\neq k}\boldsymbol{S}_{kl}(\boldsymbol{\theta}_l^* - \boldsymbol{X}\boldsymbol{\beta}_l) \right]. \tag{5.8.23}$$

As this is a multivariate normal distribution, sampling of this particular conditional distribution can be done directly in the relevant step of the Gibbs sampler.

It is still necessary to derive the matrix $\boldsymbol{S}_{kl}$, $(k, l = 1, ..., p)$. If $k = l$, then for each site $i = 1, ..., n$, the ith diagonal block element of $\boldsymbol{S}_{kk}$ is given by

$$\boldsymbol{m}_{ik}^{1/2} V_{kk} \boldsymbol{m}_{ik}^{1/2}$$

where $\boldsymbol{m}_{ik}$ is the kth diagonal element of the matrix $\boldsymbol{M}_i$ and V_{kk} is the kkth element of $\boldsymbol{V}$. Allow $\boldsymbol{U} \equiv (U_{kl} = \boldsymbol{V}^{1/2}\boldsymbol{C}\boldsymbol{V}^{1/2})$, then the off-diagonal elements are

$$\boldsymbol{m}_{ik}^{1/2} V_{kk} \boldsymbol{m}_{jk}^{1/2}$$

if $j \in \partial_i$ and zero otherwise. In the case that $k \neq l$ then the diagonal elements of $\boldsymbol{S}_{kl}$ are given by

$$\boldsymbol{m}_{ik}^{1/2} V_{kl} \boldsymbol{m}_{il}^{1/2}.$$

For $i < j$ the off-diagonal elements are

$$\boldsymbol{m}_{ik}^{1/2} U_{kl} \boldsymbol{m}_{jl}^{1/2},$$

and when $i > j$ they are given by

$$\boldsymbol{m}_{ik}^{1/2} U_{lk} \boldsymbol{m}_{jl}^{1/2}.$$

Conditional Distribution of $\boldsymbol{V}$

Recall that $\boldsymbol{V} \equiv \boldsymbol{\Gamma}^{-1}$. The conditional distribution is

$$\pi(\boldsymbol{V} \mid \boldsymbol{\beta}, \boldsymbol{C}, \boldsymbol{\theta}) \propto \tag{5.8.24}$$

$$\mid \boldsymbol{V} \mid^{n/2} \exp\left\{-\frac{1}{2}(\boldsymbol{\theta}^v - \boldsymbol{\mu}^v)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}^v - \boldsymbol{\mu}^v)\right\} \mid \boldsymbol{V} \mid^{(\rho-p-1)/2} \exp\left\{-\frac{\rho}{2}\mathrm{tr}(\boldsymbol{A}^{-1}\boldsymbol{V})\right\} \tag{5.8.25}$$

If there was no spatial dependence (i.e. $\boldsymbol{C} = \boldsymbol{0}$) then (5.8.24) would reduce to a Wishart distribution that could be sampled directly. However, in the presence of spatial dependence (i.e. $\boldsymbol{C} \neq \boldsymbol{0}$) then this is not possible and a Metropolis-Hastings alogrithm using a Wishart proposal density must be used to generate realisations from (5.8.24) in step 2 of the sampler. See Metropolis et al. (1953) and Hastings (1970) for details or the overview in either Gilks et al. (1996) or Green (2001). It is important to select a precision parameter for the Wishart proposal density that ensures a sufficiently high level of acceptance of the random draws and therefore reasonable mixing for $\boldsymbol{V}$.

Conditional distribution for $\boldsymbol{C}$

From the posterior distribution (5.8.7) the conditional distribution is

$$\mid \boldsymbol{G} \mid^{1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{\theta}^v - \boldsymbol{\mu}^v)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}^v - \boldsymbol{\mu}^v)\right\} \exp\{-(\boldsymbol{C}^v)'\boldsymbol{C}^v/\xi^2\}. \tag{5.8.26}$$

In the step 3 of the sampler where values for $\boldsymbol{C}$ are sampled, a Metropolis-Hastings algorithm with a uniform proposal density must be employed to generate random draws from (5.8.26) because the distribution is not directly available. The algorithm samples values for C_{kl} conditional upon the values of the other components. During this step, the order of the components generated is randomly selected. As in the proposal density for $\boldsymbol{V}$, the uniform density must be truncated with upper and lower bounds to reasonable mixing of $\boldsymbol{C}$ and a sufficiently high acceptance rate.

Conditional distribution for $\boldsymbol{\theta}_i$

It only remains to specify the conditional distribution for the fourth step in the sampler, the conditional distribution for $\boldsymbol{\theta}$, which is proportional to

$$\prod_{i=1}^{n}\prod_{k=l}^{p}\exp(-\mathrm{E_{ik}}\mathrm{e}^{\theta_{ik}})(\mathrm{E_{ik}}\mathrm{e}^{\theta_{ik}})^{\boldsymbol{Y}_{ik}}\exp\left\{-\frac{1}{2}(\boldsymbol{\theta}^v-\boldsymbol{\mu}^v)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}^v-\boldsymbol{\mu}^v)\right\}. \tag{5.8.27}$$

It is simpler to interpret, if the conditional distribution is considered for each $\boldsymbol{\theta}_i$, or rather as the process parameters for each site. From section 5.3, these conditional distributions are proportional to

$$\prod_{k=l}^{p}\exp(-\mathrm{e}^{\theta_{ik}}\mathrm{E_{ik}})(\mathrm{e}^{\theta_{ik}}\mathrm{E_{ik}})^{\boldsymbol{Y}_{ik}}\exp\left\{-\frac{1}{2}(\boldsymbol{\theta}_i-\mu^*{}_i)'\boldsymbol{M}_i^{1/2}\boldsymbol{V}\boldsymbol{M}_i^{1/2}(\boldsymbol{\theta}_i-\boldsymbol{\mu}*_i)\right\} \tag{5.8.28}$$

where

$$\mu^*{}_i \equiv \quad \boldsymbol{\mu}_i+\sum_{i<j}\boldsymbol{M}_i^{-1/2}\boldsymbol{V}^{-1/2}\boldsymbol{C}\boldsymbol{V}^{1/2}\boldsymbol{M}_j^{1/2}(\boldsymbol{\theta_j}-\boldsymbol{\mu_j})I(j\in\partial_i) \tag{5.8.29}$$

$$+\sum_{i>j}\boldsymbol{M}_i^{-1/2}\boldsymbol{V}^{-1/2}\boldsymbol{C}'\boldsymbol{V}^{1/2}\boldsymbol{M}_j^{1/2}(\boldsymbol{\theta_j}-\boldsymbol{\mu_j})I(j\in\partial_i) \tag{5.8.30}$$

Once again, the distribution is not immediately available and sampling is performed through a Metropolis-Hastings algorithm with a multivariate normal proposal density with the covariance matrix set to $\varepsilon^2\boldsymbol{I}_p$ with ε^2 chosen to ensure sufficiently high levels of acceptance for the candidate draws and therefore reasonable mixing of the $\boldsymbol{\theta}_i$.

Starting values for the sampler

Starting values for the sampler could be obtained by setting $\theta_{ik} = \log((y_{ik}+1)/\mathrm{E_{ik}})$

for $i = 1, ..., n$ and $k = 1, ..., p$ where y_{ik} and $\mathrm{E_{ik}}$ are the observed and expected counts respectively. With these estimates of $\boldsymbol{\theta}$, maximum likelihood estimates could be used to obtain starting values for the regression coefficients in $\boldsymbol{\beta}$ and the residuals could be used to select a suitable starting value for the precision matrix $\boldsymbol{V} = \boldsymbol{\Gamma}^{-1}$. A coarse grid search could be used to maximise (5.8.26) given the other parameters to find a starting value for $\boldsymbol{C}$.

5.9 Summary

In this chapter the Flexible Conditional Autoregressive (FMCAR) model was introduced. The innovation in this model is the inclusion of flexible (i.e. asymmetric) cross-correlations between different variables at different sites. Another novel feature is the removal of the dependence of the number of neighbours in estimating the conditional means, and instead the inclusion of a precision measure that allows for variation in the data to influence the correlation. The conditions necessary to ensure that the joint covariance matrix exists were introduced and interpretation of the spatial autocorrelation parameters was discussed. The conditional and posterior distributions for a hierarchical GLM model adopting the FMCAR model as a spatial prior distribution were derived and the sampling steps of an MCMC sampler were presented. This hierarchical formulation will be applied in later chapters and is the standard approach for incorporating spatial dependence into a model for count data. Chapter 7 uses this model in a hierarchical Generalized Linear Mixed Model to model multiple road safety performance indicators for 33 London Boroughs. The impact of allowing asymmetric correlations is investigated in the context of the multiple performance indicators and the relevance of the precision measures is explored. In the next chapter the relative performance of this model in comparison to existing

MCAR approaches is considered through a simulation exercise.

CHAPTER 6

A COMPARISON OF MCAR MODELS

6.1 Introduction

The last chapter developed the FMCAR, a flexible multivaritate conditional autoregression which incorporates an asymmetric linking or cross-correlation parameter. This additional flexibility comes at a cost; there are more parameters to estimate and the model is computationally more difficult to implement than the original MCAR model of Mardia (1988). It is therefore important to consider how well the FMCAR compares to the alternative models considered in chapter 3. When the spatial relationships between variables across sites on the lattice is asymmetric then the FMCAR model should be preferred. However, when the relations are symmetric how well does the FMCAR perform in comparison to the existing approaches? This question is answered in this chapter by comparing the performance of the main multivariate spatial models using both simulated and real data.

In most spatial modelling situations, the MCAR is used as a prior in a Bayesian hierarchical framework. The main use of these models to date has been in the disease mapping literature, a sub-speciality of epidemiology. Here, Bayesian methods are particularly useful because they allow for statistical inference over a fine geographic resolution where data are sparse by nature (rare diseases recorded over small areas results in a low expected count per grid on the lattice) and observational noise is commonplace. Hierarchical methods are easily incorporated into a Bayesian method of inference and are useful because they allow borrowing of strength in the estimation of small area point estimates across the whole lattice, yet also allow for variance reduction by the use of shrinkage estimators.

It is a well established property of Bayesian inference that Bayesian procedures offer a trade-off between bias and variance reduction of estimates (Carlin & Louis 2000). Gilks et al. (1996) provide a good discussion of this property as well as demonstrating the use of Bayesian methods for small area estimation. This is particularly prevelant in the disease mapping literature as discussed in Elliott et al. (1992) and Elliott et al. (2000), where Bayesian hierarchical spatial methods are demonstrated to produce point estimates with good properties in terms of Minimum Squared Error loss. Variance reduction in Bayesian methods is achieved through the borrowing of strength or information within the hierarchical structure. The result is point estimates that are shrunk towards a 'global average' from the distribution of all the units included in the hierarchy. The effect of this shrinkage is dependent upon the prior structure assumed and conditional upon this structure being close to the 'true' population model. Returning to the spatial setting, the different MCAR models will produce different levels of shrinkage when used as spatial priors in a hierarchical model. Therefore it will be useful to compare the variance and bias trade-offs implied by each model.

The empirical goal of this thesis is to model multiple road accident performance indicators in small areas with a view to producing overall road safety performance measures that have low variability and therefore improve performance ranking. Thus, the aim of this chapter is to compare the performance of the various MCAR formulations in terms of model complexity and fit to sets of simulated and 'real-world' data whilst also considering the impact upon variance and bias.

Due to their extensive use in the disease mapping and epidemiology literature, a small number of studies already exist that adopt a simulation framework to compare spatial models. For instance, Lawson et al. (2000) compared a range of univariate spatial models according to goodness of fit criteria and Richardson et al. (2004) compared the smoothing of disease risk performed by different univariate models and therefore their ability to detect hightened risk. More recently, Best et al. (2005) produced a thorough comparison of univariate models that extended their coverage beyond CAR models to semi-parametric and moving average models. Currently no similar simulation study exists for multivariate model, although both Kim et al. (2003) and Jin et al. (2005) demonstrate their models using simulated data. Therefore this chapter will be a useful aid to applied researchers in selecting the most appropriate MCAR specification for use in a Bayesian hierarchical framework.

For reasons of space and due to the undelying aim of comparing the FMCAR the focus of this simulation study will be on variants of the MCAR. After presenting the simulation to be used in this comparison along with the models, the comparative performance of the models is discussed. A comparison of the models applied to cancer data for West Yorkshire completes the chapter.

6.2 Comparing Spatial Priors

There are five different ‘correlation’ parameters that could possibly be specified in the multivariate (bivariate) models considered: (1) for variable one there is a spatial autocorrelation parameter (α_1), (2) for variable two there is also a spatial autocorrelation parameter (α_2), (3) for each site there is a non-spatial correlation parameter (α_0) which has also been called a bridging parameter in the literature, (4) there is a spatial cross-correlation parameter, also called a linking parameter in the literature, between variable one at site i and variable two at site$j$ which is labelled α_3, (5) and for asymmetric specifications there is a second linking or spatial cross-correlation parameter, α_4, which relates variable two at sitei with variable one at site j. For a symmetric specification $\alpha_3 = \alpha_4$. These multiple relationships are illustrated in figure 6.1 which depicts a lattice of four sites (labelled 1 to 4) arranged on a simple grid or square. There are two variables (labelled a and b) and therefore they can be viewed as two overlapping grids as in figure 6.1. The solid and dashed lines represent each of the five correlations described above.

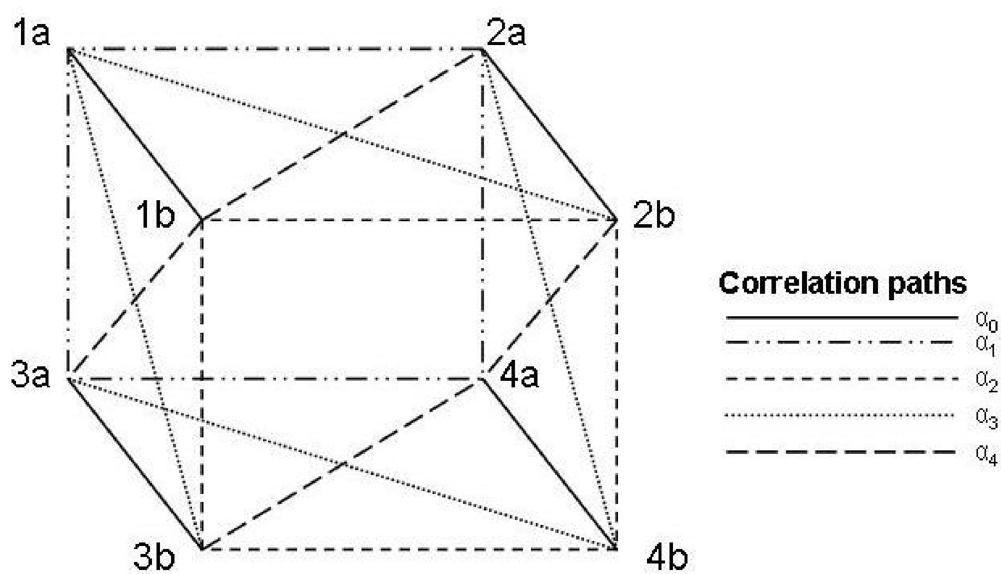

Figure 6.1: Illustration of the correlations in a bivariate dataset recorded on a four site lattice.

The FMCAR model incorporates all five types of correlation into the model and is therefore the most *flexible* of the models. Note that this is not the same as the FMCAR being more general because the other models (GMCAR, MCAR and 2fCAR) are not nested within the FMCAR. These parameters can be extracted from the matrix $\boldsymbol{C}$ in the FMCAR model – see section 5.5. However, it is also the most complex (i.e. has the most parameters) but is not necessarily the most computationally burdensome. The GMCAR of Jin et al. (2005) allows for a cross-correlation or linking parameter (α_3) but this cross-correlation is assumed to be symmetric (i.e. $\alpha_3 = \alpha_4$). As indicated in chapter 3, Jin et al. (2005) use the notation η_o for α_0 and η_1 for α_3, but this simulation study will use α_i for all of the correlation parameters to make interpretation of the findings simpler. The GMCAR is also difficult to implement in practice because there is very little guidance as to whether the right model is $\pi(\boldsymbol{y}_1 \mid \boldsymbol{y}_2)$ and $\pi(\boldsymbol{y}_2)$ or whether it is actually $\pi(\boldsymbol{y}_2 \mid \boldsymbol{y}_1)$ and $\pi(\boldsymbol{y}_1)$. Chapter 3 covered this issue in some detail. The two-fold CAR model (2fCAR) of Kim et al. (2003) also has four correlation parameters like the GMCAR model, but it can not be generalized beyond the bivariate case and is rather troublesome to implement, as was discussed in chapter 3. The last model that is included in this comparative study is the MCAR version specified in Gelfand & Vounatsou (2003) and discussed in chapter 3. Here there are three types of correlation that are included, the two spatial autocorrelation parameters and the non-spatial covariance which is captured in the covariance matrix, Λ rather than through a separate α_0 term. Table 6.1 summarizes the four models to be used as spatial priors in this chapter and the parameters they contained.

Table 6.1: Summary of model parameters

Model	Parameters
1. FMCAR	α_0, α_1, α_2, α_3, α_4
2. GMCAR	α_0, α_1, α_2, α_3
3. MCAR	α_0, α_1, α_2
4. 2fCAR	α_0, α_1, α_2, α_3

6.3 Simulation Study

A small simulation study is used to compare the performance of the MCAR prior specifications presented in the last section. Bivariate data are simulated based on the geographical layout of the 44 Health Authorities in Greater London because the adjacency map is readily available in WinBUGS for other researchers to compare results. For each health authority ($i = 1, .., 44$), assume that the data y_{ij} arise from the following bivariate Gaussian process:

$$y_{ij} \sim \mathcal{N}(\beta_j + \boldsymbol{\phi}_{ij}, \sigma^2), \qquad i = 1, ..., 44 \quad j = 1, 2 \tag{6.3.1}$$

where the β_j are fixed constants and the $\boldsymbol{\phi}_{ij}$ are random effects. Four simulation studies are performed where in each study one of the four models listed in table 6.1 are assumed to be the data generating or 'true' model. In the first study the $\boldsymbol{\phi}_{ij}$ are generated from the FMCAR model from chapter 5 and hence the data have asymmetric cross-correlations. In the second study the $\boldsymbol{\phi}_{ij}$ are produced using the GMCAR as the assumed model with symmetric cross-correlations. The MCAR model of Gelfand & Vounatsou (2003) is assumed to be the true data generating model in the third study and the spectral decomposition approach is used to produce the R_k matrices – see chapter 3 if this is unfamiliar. This model has a different spatial parameter for each variable (i.e. α_1 and α_2) but no spatial cross-correlations between variables at different sites ($\alpha_3 = \alpha_4 = 0$). Lastly, the two-fold CAR model of Kim et al. (2003) is used for the fourth study which includes a single symmetric spatial cross-correlation parameter (α_3) similarly to the GMCAR model. The true values of the parameters that were assumed for each of these models to simulate the data are shown in table 6.2.

Table 6.2: The true parameter values used in the simulation studies

Study	True Model	β_1	β_2	σ^2	τ_1	τ_2	α_0	α_1	α_2	α_3	α_4	Λ_{12}
1	FMCAR	-2.0	-5.0	0.01	10	10	0.90	0.20	0.90	0.50	0.30	—
2	GMCAR	-2.0	-5.0	0.01	10	10	0.90	0.20	0.90	0.50	—	—
3	MCAR	-2.0	-5.0	0.01	10	10	—	0.20	0.90	—	—	6.1
4	2fCAR	-2.0	-5.0	0.01	10	10	0.90	0.20	0.90	0.50	—	—

To compare the performance of the FMCAR with the existing multivariate approaches in the literature, 100 datasets were simulated for each study where each study uses one of the four models to generate the data. The four models were then fitted to each of the 100 datasets from each study using MCMC methods. To improve convergence of the MCMC chains the models were recentered (Gelfand et al. 1995) so that the hierarchical model becomes

$$y_{ij} \sim \mathcal{N}(\boldsymbol{Z}_{ij}, \sigma^2), \qquad i = 1, ..., n \quad j = 1, 2 \tag{6.3.2}$$

where

$$\boldsymbol{Z}_{ij} = \beta_j + \boldsymbol{\phi}_{ij}$$

and the mean of $\boldsymbol{Z}_{ij}$ becomes β_j rather than zero. This leads to trivial changes in the conditional distributions, $\pi(\boldsymbol{Z}_1 \mid \boldsymbol{Z}_2)$, for each of spatial priors adopted. The variance, σ^2 is given a non-informative inverse gamma distribution ($\sigma^2 \sim \mathcal{IG}(1, 0.1)$) and the intercept for each variable is given a vague normal prior ($\beta_j \sim \mathcal{N}(0, \infty)$). All that is required to complete the four model specifications are the hyperpriors i.e. the prior distributions for(α_i and τ_j. To keep the models as comparable as possible the same prior distributions are used for each of the MCAR models where possible, and non-informative priors are chosen. The precisions are given vague gamma distributions ($\tau_j \sim \mathcal{G}(1, 0.1)$, which is equivalent to specifying a Wishart distribution for the multivariate variance matrix, $\mathbf{\Gamma}$. Lastly, prior distributions for the correlation coefficients ($\alpha_0, \alpha_1, \alpha_2, \alpha_3$ and α_4) are required. The MCAR doesn't have an α_3 or α_4 (spatial cross-correlations), and the GMCAR and two-fold CAR priors don't have an asymmetric spatial cross-correlation parameter, α_4. In the bivariate case it is easier to specify each parameter individually as taking uniform priors ($\alpha_i \sim \mathcal{U}\rangle\{(0,1)$).

To monitor convergence of the MCMC chains five overdispersed (relative to the posterior) parallel chains[1] were run for each model and covergence checked visually using sample trace plots as well as numerically using summaries such as Gelman's $\sqrt{R}$-statistic (Gelman 1996) which are available in the CODA[2] package for S or R (see Best et al. (1996) and Best et al. (1997)). The chains had converged (i.e. $\sqrt{R}$-statistic $\approx$ 1.0) in all cases by iteration 5,000. These were then discarded as "burn-in" and an additional 25,000 samples were produced from which to summarize the posterior distributions of the models. Although all of these models could be implemented in the popular WinBUGS application[3] the MCMC program already written in Python for the FMCAR was adapted for the simulation study. Random number generation, posterior summarization, data visualization, and convergence diagnosis was performed using R[4] by using the RPy[5] Python interface.

[1]This requires that one chain is initially run and after signs of convergence the posterior is inspected. From the posterior four initial values for four additional chains that are overdispersed relative to the posterior are selected and are then run.

[2]Convergence Diagnosis and Output Analysis - see chapter 5 for further details.

[3]See chapter 5 or visit http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml

[4]www.r-project.org

[5]http://rpy.sourceforge.net/

6.3.1 Model Complexity and Fit

The complexity and fit of the four models in consideration were compared using the deviance information criterion (DIC), a simple and intuitive extension of the established Akaike information criterion (AIC)(Akaike 1974) for hierarchical models. The DIC is based on the posterior distribution of the deviance statistic, $D(\boldsymbol{\theta}) = -2\log f(\boldsymbol{y} \mid \boldsymbol{\theta}) + 2\log h(\boldsymbol{y})$, where $f(\boldsymbol{y} \mid \boldsymbol{\theta})$ is the likelihood function for the observed data ($\boldsymbol{y}$) given the vector of parameters ($\boldsymbol{\theta}$) and $h(\boldsymbol{y})$ is some standardizing function of the data. Analogously to the AIC, the DIC is defined in terms of the posterior expected deviance and an estimate of the 'effective' number of parameters i.e. DIC $= \bar{D} + p_D$. In classical nonhierarchical models this concept is well defined but in a Bayesian hierarchical setting, the shrinkage properties of the prior distribution essentially restrict the model parameters.

The effective number of parameters term, p_D, in the DIC was proposed by Spiegelhalter et al. (2002) to capture the amount of shrinkage performed by the prior. p_D was shown (for approximately normal likelihoods) in Spiegelhalter et al. (2002) to be equal to the ratio of the likelihood to the total information contained in the posterior distribution ($\propto$ likelihood x prior). Thus a p_D that is small relative to the number of observations highlights that the prior distribution is providing a lot of structural information about the parameters and that there is considerable 'borrowing of strength', while a p_D that is relatively large indicates that the prior is not providing much information. With models that provide very little prior information the effective number of parameters will be approximately equal to the actual number of parameters and the DIC will be almost equivalent to the AIC. Spiegelhalter et al. (2002) also demonstrate that the DIC can be interpreted as the expected posterior loss in prediction when adopting a particular model and therefore DIC can be considered to be selecting the model that produces the best prediction of the spatial random effect. The effective number of parameters, p_D, is the posterior mean of the deviance minus the deviance of the posterior means. In normal hierarchical models this is the trace of the "hat" matrix that maps the observed data to their fitted values. For linear models the trace of the hat matrix is equal to the rank of the matrix of independent variables i.e. it is the number of linearly independent parameters in the model. Gelman (2009) suggests using half the variance of the deviance as an estimate of p_D because it is invariant to parameterisation and is trivial to calculate.

The DIC, then, can be thought of as a combination of a goodness of fit measure ($\bar{D}$) and a complexity measure (p_D). As small deviance values indicate good fit and small number for p_D indicates a parsimonious model, small values of the DIC are preferred. Due to $\bar{D}$ being scale-free so is the DIC and hence there is no substantive interpretation to be placed on the absolute values of the DIC; only the rankings of the DIC between models is of interest. An important question is how large a difference in the DIC between models is noteworthy. According to Spiegelhalter et al. (2002) models with DIC values within 1 or 2 of the 'best' model (i.e. the one with the lowest DIC) are also strongly supported, those with DIC values between 3 and 7 of the 'best' are only weakly supported, and any other models (i.e. with a DIC greater than 7 away from the 'best') are substantially inferior.

In addition to computing the DIC, the average mean squared error (AMSE) is also calculated for the 100 datasets in each study. The mean squared error (MSE) of an estimator is one method of quantifing the difference from the true value of the quantity being estimated. While particular values of (A)MSE other than zero are meaningless (which indicates that the estimator completely accurately predicts) the MSE values may, once again, be used for comparative purposes. Once again, the model with the lowest (A)MSE is preferred. As the true $\boldsymbol{Z}_{ij}$ values are known in the simulation the $\hat{\text{AMSE}}$ can be estimated as

$$\frac{1}{Nnp}\sum_{t=1}^{N}\sum_{j=1}^{p}\sum_{i=1}^{n}\left(\hat{\boldsymbol{Z}}_{ij}-\boldsymbol{Z}_{ij}\right)^2$$

with associated Monte Carlo standard error estimate, $\hat{se}(\hat{\text{AMSE}})$ calculated as

$$\sqrt{\frac{1}{(Nnp)(Nnp-1)}\sum_{t=1}^{N}\sum_{j=1}^{p}\sum_{i=1}^{n}\left[\left(\hat{\boldsymbol{Z}}_{ij}-\boldsymbol{Z}_{ij}\right)^2-\hat{\text{AMSE}}\right]^2},$$

where for this simulation study $N = 100$, $p = 2$ and $n = 44$.

Tables 6.3 and 6.4 report the DIC and AMSE comparisons. Recall that study 1 used the FMCAR to generate data and is called model 1, study 2 used the GMCAR to generate data and is labelled model 2, model 3 is the MCAR and is used in study 3, and lastly model 4 is the two-fold CAR model and is used in study 4 to generate the data. Table 6.3 summarizes the relative performance of each of the models against that study's true model. Therefore the values reported are the amount by which each model is above (positive numbers) or below (negative numbers) the DIC of the true model. Hence a negative number would indicate that the model in question 'beat' the true model. The true model in each study is indicated by a dash (—). The median alongside the 2.5 and 97.5 percentiles are recorded for the DIC difference. Table 6.4 reports the estimated average mean square error and the related Monte Carlo estimates of the standard errors, again for each model in the simulation. The percentage change[6] (Δ) in AMSE for each model relative to the true model (indicated by a dash) is also reported. Once more, negative values for Δ indicate that the model

[6]Calculated as follows: $\Delta = (\hat{\text{AMSE}}_{\text{i}} - \hat{\text{AMSE}}_{\text{true}})/\hat{\text{AMSE}}_{\text{true}}$ x 100 for models $i = 1, ..., 4$.

is superior to the true model.

Table 6.3: Percentiles of estimated DIC difference between the true model and the other models

Study 1			Study 2			Study 3			Study 4		
2.5%	50%	97.5%	2.5%	50%	97.5%	2.5%	50%	97.5%	2.5%	50%	97.5%
—	—	—	-6.24	2.74	5.64	-15.3	-0.59	8.57	-12.1	-2.24	2.37
-4.80	19.6	56.7	—	—	—	-11.29	2.32	13.0	-14.3	-2.15	3.16
3.56	34.9	68.8	-3.62	1.48	8.47	—	—	—	-10.0	0.83	4.89
3.03	23.7	65.1	0.50	30.4	63.7	2.76	20.9	53.3	—	—	—

The symbol "—" indicates the model is the true model for this study

Table 6.4: Average mean squared error (x10^{-3}), associated MC standard errors (SE x10^{-5}), and percentage change in AMSE (Δ,%) relative to the true model.

Study 1			Study 2			Study 3			Study 4		
AMSE	(SE)	Δ	AMSE	(SE)	Δ	AMSE	(SE)	Δ	AMSE	(SE)	Δ
7.51	(8.26)	—	7.91	(8.57)	1.54	8.49	(9.20)	-3.08	5.46	(5.92)	-7.92
8.17	(8.91)	8.79	7.70	(8.44)	—	8.81	(9.67)	0.571	5.44	(5.87)	-8.26
9.22	(10.1)	22.8	7.82	(8.50)	0.385	8.76	(9.55)	—	5.85	(6.34)	-1.35
8.22	(8.80)	9.45	9.86	(10.9)	26.6	11.2	(12.3)	27.8	5.93	(6.44)	—

The symbol "—" indicates the model is the true model for this study

The results of the simulation study appear reasonably consistent; larger DIC differences in table 6.3 correspond with larger AMSE values in table 6.4. Starting with table 6.3, when asymmetric cross-correlations are present in the data (i.e. when the FMCAR was used to generate the data) the FMCAR is the best model with the DIC difference between it and the rest of the models being substantial when the indicator of a substantially inferior model is being more than 7 away from the best model. In the case of symmetric cross-correlations (i.e. in study 2 using thr GMCAR) both the FMCAR and the standard MCAR are close to the best model, the GMCAR. Surprisingly, the 2fCAR model of Kim et al. (2003) is over 30 units away from the GMCAR when it includes a cross-correlation parameter. In fact in no cases is the

2fCAR even weakly supported by the DIC using the criteria from Spiegelhalter et al. (2002). Even when it is used to generate the data (study 4) it is beaten by the two other models that also include cross-correlation parameters (the FMCAR and GMCAR) and the MCAR is effectively tied with the 2fCAR model being only 0.83 above it in terms of the DIC. Finally, in the case of no cross-correlations (study 3) the FMCAR and GMCAR both perform well being strongly supported by the DIC criteria.

These general findings are supported by the AMSE reported in table 6.4. All of the models have a better AMSE than the data generating 2fCAR model in study 4. And, with the exception of study 1, the 2fCAR model performs exceedingly badly in terms of AMSE being more than 25 percent worse than the true model. Again, the FMCAR is the best model under an asymmetric cross-correlation situation with the alternative models performing relatively poorly. In study 2, under a symmetric cross-correlation structure, both the FMCAR (+1.5%) and the MCAR (+0.4%) are close in terms of AMSE to the true model.

The additional complexity of the FMCAR appears to offer benefits in terms of shrinkage of the spatial random effects in comparison to the other MCAR priors tested. Figure 6.2 plots a histogram of the random effect variance ($\sigma_1^2 = 1/\tau$) for variable 1 for each of the four models fitted to the simulated data where the FMCAR is the true model. This, in effect, measures the precision. There is greater shrinkage occuring by the use of additional information contained in the spatial correlation structure which is allowing for the estimates of the random effect to be "shrunk" towards the global mean of zero. This indicates that the FMCAR will be an appropriate model to use in the ranking of road safety performance indicators in chapter 7 as the aim will be to generate narrow credible intervals for the random

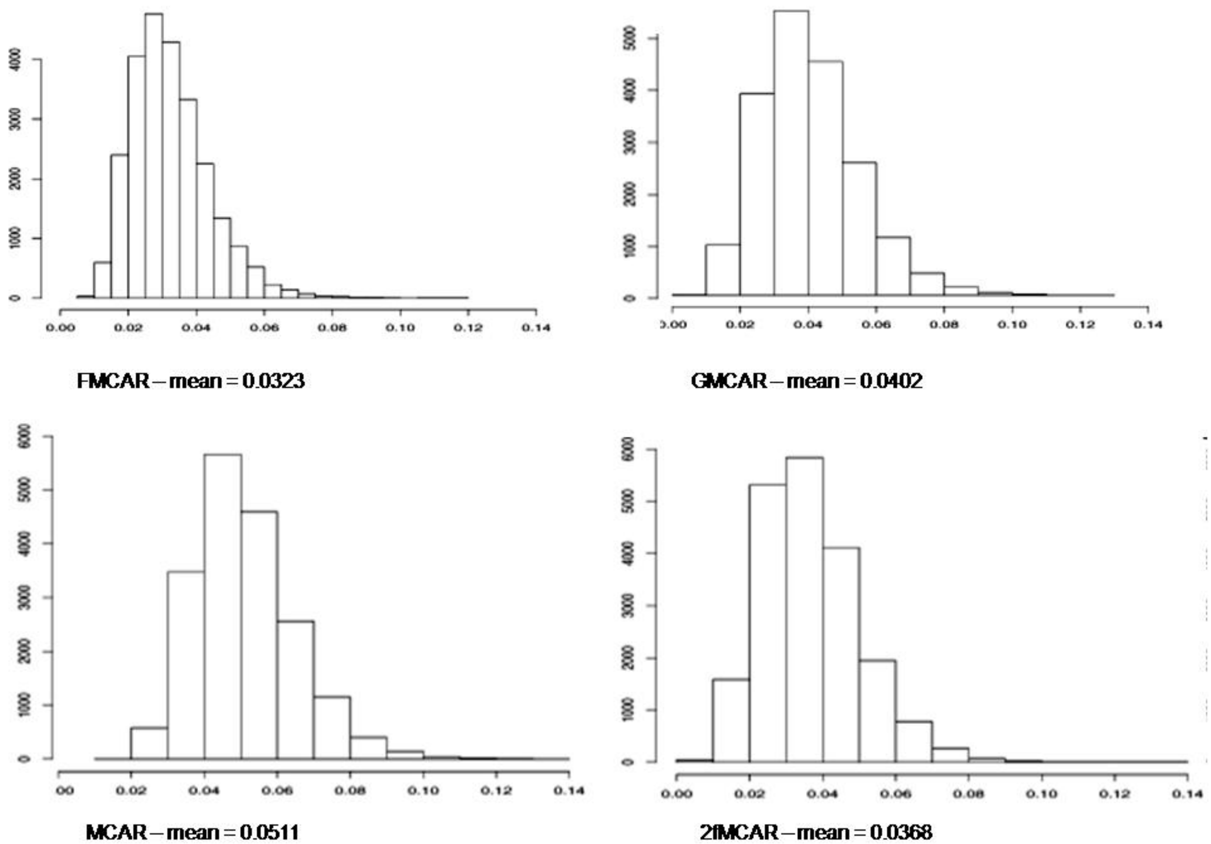

Figure 6.2: Histograms of posterior conditional variance for variable 1 from each model

effects.

In terms of predictive ability, the general conclusion that can be drawn from this simulation study is that the FMCAR is by far the superior model when asymmetric cross-correlations (study 1) are present in the data. However, even when the cross-correlations are symmetric (study 2) or non-existent (study 3) the FMCAR is strongly supported by the DIC and is within a few points of the 'best' model in each case. Given this and the ease of implementation and interpretation of the FMCAR which was discussed in chapter 5 there does not seem to be a downside to using the FMCAR as the default model for modelling spatial random effects. Additionally, the FMCAR model does offer attractive properties in terms of shrinkage of the estimated spatial random effects variance, which will be of considerable use in modelling performance indicators. However, the FMCAR offers little or no benefit over existing spatial priors in the absence of asymmetric cross-correlations. The conclusion drawn was that, given the negligible penalty of adopting the FMCAR by default, the attractiveness of this model to other researchers with depend upon the relative overhead of implementing this model.

6.4 Data Example

Although the results of the last section suggest that the FMCAR performs well regardless of the true data generation process, this section uses real data to compare the models. Further, this example illustrates the use of the various MCAR prior distributions for the joint modelling of non-Gaussian data. This is particularly salient given the empirical objective of this thesis is to model accident counts. In this comparison, the data represent observed and age and sex standardised expected counts of incidenct cases of oral cavity and lung cancer in each of 126 electoral wards in the West Yorkshire region of England between 1986 and 1991.[7]

Since both cancers are rare, the mortality counts y_{ij} for cancer j, (j =1, 2) in area i (i =1, ..., 126) are assumed to follow independent Poisson distributions, conditional on an unknown mean θ_{ij}

$$Y_{ik} \sim \text{Poi}(E_{ij} e^{\theta_{ij}}) \tag{6.4.1}$$

[7]This dataset is used because it is available along with the West Yorkshire adjacency file in the WinBUGS library allowing these results to be freely replicated. This dataset was previously used by Best et al. (2005) to compare univariate CAR models.

$$\log\theta_{ij} = \log E_{ij} + \beta_j + \phi_{ij} \tag{6.4.2}$$

where E_{ij} is the age and sex standardised expected count (offset) for cancer j in area i, β_j is an intercept term representing the baseline (log) relative risk of cancer j across West Yorkshire, and ϕ_{ij} is the area- and cancer-specific log relative risk. The ϕ_{ij} are therefore spatial random effects and are modelled using the MCAR models from the simulation exercise as spatial priors.

By inspecting figures 6.3 and 6.4 it is apparent that the log relative risks for oral cavity and lung cancer are spatially correlated across West Yorkshire. Comparing the two figures there also appears to be within area correlation. This is confirmed by the correlation between risk of oral cavity and lung cancers being 0.84 suggesting strong shared geographical pattern of risk between the two diseases. This could be the result of some underlying common cause such as smoking prevelance.

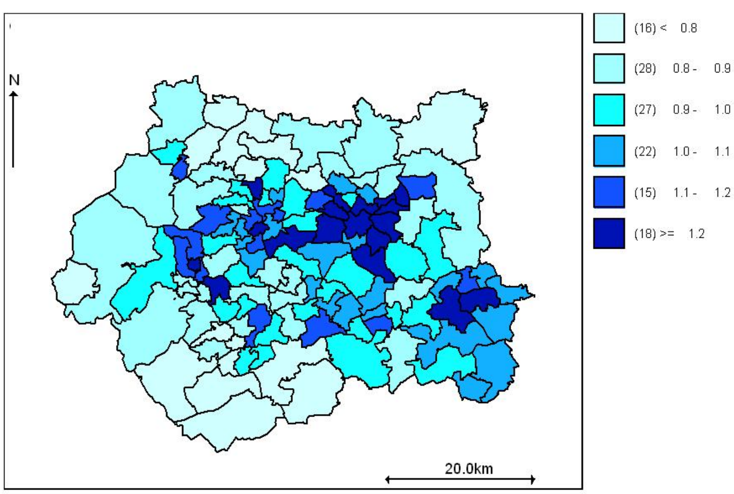

Figure 6.3: Map of relative risk of oral cavity cancer for West Yorkshire.

All four models are fitted to the West Yorkshire cancer data using the same

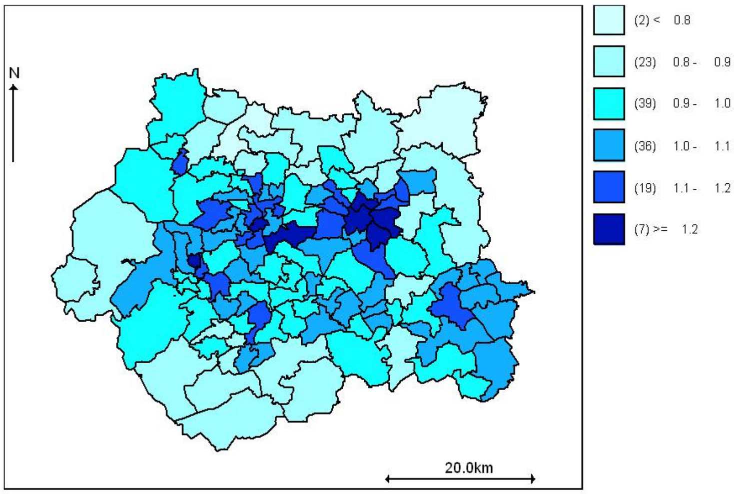

Figure 6.4: Map of relative risk of lung cancer for West Yorkshire.

priors and hyperpriors as section 6.3. In the first study the $\boldsymbol{\phi}_{ij}$ are estimated with the FMCAR prior and therefore there are five model parameters. In the second model the $\boldsymbol{\phi}_{ij}$ are fitted using the GMCAR as the prior and there are four parameters i.e. the cross-correlations are assumed to be symmetric. The MCAR model of Gelfand & Vounatsou (2003) is the prior in the third model and the spectral decomposition approach is once again used to produce the R_k matrices. This model has a different spatial parameter for each variable (i.e. α_1 and α_2) but no spatial cross-correlations between variables at different sites ($\alpha_3 = \alpha_4 = 0$). Lastly, the two-fold CAR model of Kim et al. (2003) is used for the fourth model which includes a single symmetric spatial cross-correlation parameter (α_3) similarly to the GMCAR model. The same hyperpriors from section 6.3 are used along with the same MCMC methods (Metropolis-Hasting step within a Gibbs sampler). For each model, five overdisperesed chains were run to check for convergence which occured around the

5,000th iteration. These were then dropped as 'burn-in' and another 10,000 iterations were run giving 50,000 samples for posterior summarization.

As the main concern of this chapter is comparing these competing spatial priors in terms of model fit, table 6.5 reports the separate contributions of fit $\bar{D}$ and complexity, p_D, to the the DIC scores. A discussion of the fixed and random effects is omitted as understanding the spatial distribution of two cancers in West Yorkshire is not the focus of this chapter.

The first thing to note from table 6.5 is that all four models allow considerable degrees of shrinkage or borrowing of strength with between 50 and 70 effective parameters to fit 126 data points. The FMCAR has the smallest p_D indicating that there is a lot of structural information in the prior leading to considerable borrowing of strength. However, despite this the GMCAR has a marginally better model fit $\bar{D}$. Note that the effective number of parameters, p_D, may actually be smaller for more complex models precisely because it allows for more complex forms of shrinkage. This has been reported by Best et al. (2005) and could help to explain the potentially contradictory findings from this exercise. Putting these two measures together, the DIC shows that the GMCAR model is the best supported model for *this* dataset (i.e. has the smallest DIC) although the FMCAR is also strongly supported by the data being within 2 of the DIC score achieved by the 'best' model. The MCAR model is weakly supported, lying less than 7 away from the best model, but the two-fold CAR model is substantially inferior.

The results aren't suprising. Whilst it seems sensible that there could be a cross-correlation to these two cancers (i.e. high lung cancer rates at site 1 may be related to high oral cavity cancer rates at site 2) there is no a priori reason to believe that this relationship would be asymmetric. Therefore the two models (GMCAR and FMCAR) that incorporate these linking or cross-correlation parameters would both fit the data well. However, the additional complexity and structure imposed by the asymmetric FMCAR is not warranted in this case. Yet, model fit is robust to changes between these two spatial priors and the effects of inappropriately fitting the FMCAR to these data seem limited (i.e. the change in DIC is minimal).

Table 6.5: Model comparison using DIC for West Yorkshire cancer data

Model	$\bar{D}$	P_D	DIC
1. FMCAR	485.4	58.2	543.6
2. GMCAR	477.6	63.8	541.9
3. MCAR	483.6	64.3	548.9
4. 2fCAR	492.6	65.1	557.7

6.5 Summary

The aim of this chapter was to compare the three principal MCAR models from the extant literature with the FMCAR model developed in chapter 5. This was achieved through a combination of a simulation study and an application to a small real-world dataset on cancer mortality in West Yorkshire. Four studies were used where each of the four competing models was assumed to be the true data generating process. In each of these studies 100 datasets were simulated and the four models were fitted to the datasets. This allows a comparison of the models across a range of possible spatial configurations. In general, the performance between the FMCAR, GMCAR and MCAR was comparable for all situations except the presence of asymmetric linking parameters (cross-correlations) i.e for all but the case where the FMCAR generated the data. This is both good and bad news. On the positive side, fitting the FMCAR when asymmetric spatial cross-correlations are not present does not have deleterious effects. It is therefore safe to fit the FMCAR without a priori reasons to suspect asymmetric cross-correlations as there is little impact in terms of model fit. However, the FMCAR offers little or no benefit over existing spatial priors in the absence of asymmetric cross-correlations. Therefore persuading researchers to adopt the FMCAR will depend upon the relative costs of implementing this model over the potential benefits of avoiding model mis-specification. The ease of implementing and interpreting the model will be discussed in the chapter 7. The FMCAR model does offer attractive properties in terms of shrinkage of the estimated spatial random effects variance, which will be of considerable use in modelling multiple road safety indicators as achieving narrow credible intervals for the random effects is the principal aim.

CHAPTER 7

MODELLING ROAD SAFETY PERFORMANCE INDICATORS

7.1 Introduction

This chapter applies the FMCAR model developed in chapter 5 to multiple road safety performance indicators. Currently, local authority road safety performance is measured through a series of crude per capita accident rates by modal type and accident severity. Given the rare nature of road accidents (particularly fatalities) the data are sparse and subject to great variability. This makes robust inference about local authority performance difficult. The aim of this chapter is two-fold. Firstly, this chapter seeks to demonstrate the applicability of the FMCAR to a policy-relevant problem. Additionally, inference and interpretation of the model output will be illustrated. Therefore, this chapter aims to persuade other researchers to adopt the FMCAR for their research. Secondly, by adopting a multivariate spatial modelling framework it is hoped that the correlation structure in the data (both within local authorities and across local authorities) can be exploited to reduce the uncertainty in the estimates of road safety performance. This is the major empirical contribution of this thesis: to reduce the uncertainty involved in the estimation of local authority specific performance and to improve performance ranking of local authorities.

7.2 Background

7.2.1 Why manage performance?

Worldwide, injuries and death resulting from road traffic accidents are of epidemic proportions: over 1 million people die every year and over 10 million people sustain permanent disabilities from road accidents (Bunn et al. 2003). Globally, road accidents are the 10th most common cause of death (Commission for Global Road Safety 2005). By 2020, the World Health Organisation (WHO) predict road accidents will be the 6th leading cause of death worldwide, and the second leading cause of Disability-Adjusted Life Years (DALYs) lost in developing countries (WHO 2004).

Recognizing the extent of the problem, the UK Government identified three road safety targets which are expected to be achieved by 2010 (these were set relative to a baseline of the mean number of casualties that were reported between 1994 and 1998 inclusively): a 40 percent reduction in the number of fatally or seriously injured casualties, a 10 percent reduction in the rate of slight casualties relative to the level of traffic, and a 50 percent reduction in the numbers of children who were fatally or seriously injured (DfT 2000). Most local authorities incorporate these targets into their Local Transport Plans (LTP) but it is related performance indicators, broken down by modal group, which are monitored and published in the local authority league tables under the 'best value' requirements of the Local Government Act 1999 (Department for Transport, Local Government, and the Regions 1999). It is these indicators called BVPI99 (Best Value Performance Indicators 99) that are used to judge the relative performance of each local authority with respect to road safety.[1]

According to PACTS (2003*a*) DfT guidance intimates that performance will be

[1]Although arguably they are measures of 'unsafety'.

used to determine financial allocations to local authorities in future. This is supported by Department for Transport, Local Government and the Regions (2001) which suggests that good and improving performance will attract additional funding and increased autonomy under Public Service Agreements. The Comprehensive Spending Review (H.M. Treasury 2007) has also indicated that in future increasing attention will be paid to performance when allocating resources. There is also widespread support from within local authorities for a target-led approach (PACTS 2003*a*). However, there was concern expressed in PACTS (2003*a*) that the headline indicators should be aggregated together because of the variability in individual indicators. Obtaining statistically reliable results on performance is therefore a major concern of authorities themselves who do not possess the expertise and experience to produce robust methods internally. Futher, if central government is moving towards a culture of 'payment for performance' then producing improved measures of performance is a worthwhile endeavour.

7.2.2 Current Practice

Currently, UK traffic safety performance indicators are expressed simply in the form of crude per capita numbers of reported collisions by type and modal group, with no allowance for geographically differing patterns in road infrastructure and usage or spatially varying socioeconomic conditions, so called contextualisation in the literature (Goldstein & Spiegelhalter 1996). In fact, according to Bailey & Hewson (2004) there is no explicit consideration given to the extent to which differences in the raw rates reflect differential performance, rather than just inherent random variability in observed rates. As chapter 4 discussed, it is bad practice to use point estimates of indicators to rank performance across observational units (e.g. local authorities) without accommodating the uncertainty inherent in these estimates. Research by Goldstein & Spiegelhalter (1996) suggests that rankings adopting robust procedures that incorporate uncertainty in a statistically valid manner frequently make it impossible to separate the organisations under study.

There have been significant developments in the modelling of performance and the uncertainty inherent in performance rankings (or league tables) in the education and health literatures. For example, see Laird & Louis (1989), Goldstein & Spiegelhalter (1996), Marshall & Spiegelhalter (1998), Lockwood et al. (2002), Kuhan et al. (2002), Draper & Gittoes (2004), Bratti et al. (2004) and references contained within. This literature was discussed in chapter 4. Unfortunately, this work has not been transferred to the area of local government performance management and performance indicators remain crude, uncontextualised numbers.

Although detecting a (statistically significant) departure in road safety performance between organisational units is only part of the larger picture of performance management, it is still important to ascertain if there is differential performance. This can, for example, indicate the need for further research e.g. detailed auditing to identify best practice and/or intervene in poor performaning authorities. It is also important to make the best possible statistical inference about the road safety performance of local authorities when their performance is tied to financial resources. Limited work has been done in this area. For instance, MacNab (2003) employed generalized additive modelling to smooth out year-in-year variance in area level accident rates in Canada. He emphasized the importance of separating signal from noise when investigating such ‘noisy’ data. More generally, the Generalized Linear Mixed Model (GLMM) is now well established in the wider literature on performance management to model the uncertainty associated with performance indicators.

GLMMs are characterised by the inclusion of 'random effects' in addition to the fixed parameters found in conventional generalized linear models[2]. The use of fixed and random effects is now commonplace in a number of disciplines, and modelling using random effects is the norm in hierarchical or multilevel modelling. In terms of modelling road accident performance, the traditional GLM would be extended to include a random effect term to capture the latent or unobserved local authority performance. This term reflects the difference in performance between local authorities that has not been observed and must be estimated from the data. This is assumed to be a latent (i.e. unobserved) zero mean process with unknown variance. In contrast, a local authority fixed effect would simply reproduce the observed performance indicator for that local authority. The strength of this GLMM approach is the recognition of a source of uncertainty (or variance) that is related to local authority specific performance that is not captured by the indicators (accident counts) included in the model. This uncertainty could be the result of omitted variables, mismeasured exposure or noisy data. The inclusion of this random effect or variance component induces smoothing of these local authority specific effects. In effect, the estimate of these local authority random effects 'borrow strength' from each other and are therefore 'shrunk' to the global mean (of zero). How much shrinkage occurs depends on the relative strength of the evidence i.e. the relative size of the variance of the data relating to that local authority to the variance of the random effect as a whole.

If performance ranking is of interest – for instance for producing league tables – then the local authority specific random effect can be used to produce estimates (along with relevant credible intervals) of each local authority's performance which

[2]See, for example, Gelman & Hill (2006) for an excellent introduction.

has taken into account uncertainty in the estimate of this effect. Given the sparse nature of the data, univariate GLMMs for each road user type lead to large and overlapping confidence intervals: very little can be inferred from the rankings (Bailey & Hewson 2004).

An alternative approach would be to model multiple performance indicators simultaneously. For example, Papageorgiou & Loukas (1988) used a bivariate negative binomial model for vehicle accidents in East Virginia. The underlying motivation is that for each transport mode there should be significant correlation between fatalities and serious and slight injuries. More broadly, it seems sensible that there should be some correlation between all road safety indicators. The various activities that local authorities may perform to improve road safety may affect multiple transport modes. For example, driver awareness campaigns[3] that aim to increase car drivers' awareness of motorbikes may simultaneously increase their awareness of cyclists and pedestrians. Accordingly, it may be possible to reduce the uncertainty associated with local authority specific random effects by borrowing strength from multiple, related, variables. In essence, the random effects are correlated between the road accident variables within each local authority and their estimates are therefore shrunk across the variables to increase precision. This is the approach adopted by Bailey & Hewson (2004). They hoped that the shrinkage would reduce the credible intervals on the performance rankings for individual indicators enough to allow for judgements to be made of relative performance. The result of their analysis was a reduction in the credible intervals, however there remained significant overlap and it remained impossible to discriminate (statistically) between local authorities' performance.

[3]For example, the "Think Bike!' campaign that is running in London in 2008.

7.2.3 The impact of space

There is an additional source of structure within the data that could provide additional sources of shrinkage and therefore improve the precision of the performance estimates. As well as correlation between the multiple road accident indicators there is also correlation across local authorities. More specifically, modelling spatial dependence in each of the variables at a lower level of aggregation (the LSOAs) allows the large variability in accident rates to be smoothed across the areas. Moreover, there is also the potential to model correlation between different variables across sites – the so called linking or cross-correlation parameter. There may be unobserved characteristics of the area (e.g. the road network) that may result in a high correlation between motorcycle accidents in one area and cycling accidents in a neighbouring area. Further, it is feasible that this correlation structure is asymmetric i.e. that there is a different correlation between motorcycle accidents at site 1 and cycling accidents at site 2 to the correlation between cycling accidents at site 1 and motorcycle accidents at site 2.

Very little attention has been paid to the issue of spatial structure in the road safety literature, although there have been many studies that have investigated the spatial variation in the incidence of pedestrian road casualties. Most of this research seeks to explain the spatial variation through the variation in area deprivation. Grayling et al. (2002), Graham et al. (2005) and Graham & Glaister (2005) have all attempted to explain the spatial variation in accident rates through area level deprivation covariates. Others have focused, at varying degrees of aggregation, on the influence of other factors on the spatial variation of accident rates, including Dee (1998) Noland (2001), Noland & Quddus (2002), Noland & Quddus (2004), and McCarthy (1999). In addition, Graham & Glaister (2003) looked at the spatial variation in pedestrian road casualties by examining the role of the urban environment. All of these studies have ignored the fact there may well be spatial dependence and spatial heterogeneity within the data, and as a result continue to apply methods of statistical inference that are not robust to such problems. Recently work applied to data from Canada (MacNab 2003) and (MacNab 2004) and to Devon, UK (Hewson 2004) and (Hewson 2005) have sought to formally test the assumption of spatial independene in the data and firmly reject this hypothesis. They both adopt univariate CAR models in their analyses but the impact of this research has so far been limited.

There are two principal reasons for the application of spatial models to multiple road accident data. The first reason concerns the need to reduce the variability in the estimates for individual areas and consequently improve the confidence placed in these estimates. Just as modelling random effects through GLMMs can produce shrinkage across observational units and improve the stability of the estimates, so can the introduction of spatial structure. This is the principal empirical concern of this chapter. This smoothing effect is particularly important for count data where the is high variability in the data as a result of excessive zero counts. The second reason concerns the need to address potentially omitted spatially varying covariates when any model of road accident data is extended to a regression context. With spatial data it is common to incorporate the spatial dependence into the covariance structure either explicitly or implicitly via an autoregressive model because it is assumed that observations at sites near each other may have a similar value on the omitted variables in the regression causing the error terms to be serially autocorrelated. It is hoped by demonstrating the presence of spatial dependence in the road accident data and providing a model for incorporating these spatial effects that other researchers will adopt the FMCAR for their research.

In the analysis that follows, the FMCAR model developed in chapter 5 is applied to a subset of the Best Value Performance Indicators for road safety for 2006. This subset includes the multiple performance indicators reported for vulnerable road users[4] for the 33 London boroughs recorded at the Lower Super Output Area (LSOA). The FMCAR is employed as a prior for the random effects in a hierarchical generalized linear mixed model (GLMM). Given the recent focus on integrated transport and shifting away from cars to alternative modes of transport, vulnerable road users numbers will increase. This promotion of sustainable transport is supported by iniatives from the Department of Health to promote lifestyle change to tackle obesity and coronary heart disease (Department of Health 2008). If these strategies are successful there will be increasing numbers of vulnerable road users and monitoring and ensuring their road safety will become increasingly important. Moreover, this thesis focuses on vulnerable road users because the underlying causal mechanism for road accidents involving them is similar. Other road users such as car drivers are excluded because the underlying causes, for example excessive speed or drink-driving, ar different. The safety programmes that should be targetted towards reducing vulnerable road accidents would be therefore be different. Performance measurement should be conducted separately for these two broad classes of road users.

[4]Vulnerable road users are defined as pedestrians, cyclists and motorcylists

7.3 Data

7.3.1 Best Value Performance Indicators

Road casualties per capita by mode and severity are performance indicators that are statutarily reported by local authorities to the Audit Commission as part of their reporting on 'best value'. Collectively they are known as BVPI99 and collecting and reporting these data was introduced in the Local Government Act (1999), which requires local authorities to publish details of these indicators in their 'best value performance plans'. Additionally, these indicators are collated by central government and used to produce league tables of local authority performance.

This thesis analyses a subset of BVPI99 relating to vulnerable road users. Bailey & Hewson (2004) argue that vulnerable road user casualties are problematic as indicators of road safety performance due to the small number (and therefore high variability) involved. This problem is indicated in table 7.1 which provides summary statistics for the nine performance indicators considered in this study. Three road user or transport modes are considered: pedestrians, (pedal) cyclists and motorcylists (sometimes referred to as powered two-wheelers). In addition, the accident data are broken down by the severity of the injuries sustained: fatal, serious and slight. The severity is assessed by the police officer completing the report and a serious injury is usually defined as one that requires hospitalisation (most commonly multiple fractures or cranial and spinal injuries). Therefore, the accident dataset consists of nine accident counts.

One commonly used solution to the problem of small numbers is to aggregate the data. Frequently in the literature this is done over multiple years (e.g. Graham & Glaister (2005) or Edwards et al. (2006)) but this is not possible for local authorities

Table 7.1: Median (and upper and lower Quartiles) for Vulnerable Road User Casualties by London Borough in 2006

Mode	Fatal	Serious	Slight
Pedestrians	3 (1,4)	33 (27,41)	122 (87,162)
Motorcylists	1 (0,3)	24 (16,30)	96 (78,143)
Cyclists	0 (0,1)	9 (5,16)	61 (43,95)

reporting annual performance indicators. More commonly during publication of best value indicators, fatal and serious injuries are aggregated or less frequently, as suggested by PACTS (2003*b*), road user categories are merged – for example, pedestrians and cyclists or even all vulnerable road users. Although this superficially can smooth out random fluctuations it also hides a significant amount of information. This is problematic when the government sets road safety targets that include specific modal groups (e.g. pedestrians) such as those laid down in *Tomorrow's Roads* (DfT 2000). In the future there is likely to be growing interest in vulnerable road users, especially pedestrians and cyclists, as central government continues to promote integrated transport and modal shift away from cars.

This promotion of sustainable transport is supported by iniatives from the Department of Health to promote lifestyle change to tackle obesity and coronary heart disease (Department of Health 2008)[5]. Almost 25 percent of adults in England are obese with this figure set to reach 90 percent by 2050 Department of Health (2009). The cost of obesity to the NHS is estimated to be £4.2 billion and is forecasted to more than double by 2050 (*ibid*). If these strategies are successful there will be increasing numbers of vulnerable road users and monitoring and ensuring their road safety will become increasingly important. Therefore disaggregated analysis of these performance indicators must be preferred.

[5] These include Let's Get Moving, Local Exercise Action Plans, and Change4Life.

7.3.2 STATS19 data

The administrative data on road casualties reported by local authorities originates from the police via a recording system referred to as "STATS19".[6] Bull & Roberts (1973) summarize a number of problems with these data. For the purposes of this thesis, the most significant is the under-reporting of accidents involving vulnerable road users. More recent evidence of this under-reporting is found in an article on 25 case studies by James (1991). It also appears that up to 60 percent of slight accidents go unreported (Cryer et al. 2001). However, hospital episode data are equally problematic: only those casualties that require hospital treatment (serious and fatal accidents) will be included in the dataset and no accident location will be recorded. The STATS19 are the definitive data source in terms of policy and practice. It is these data that local authorities must report and which are used for determining road safety targets locally and nationally, and it is these data that are used almost exclusively in the literature on road safety modelling. Therefore, despite their limitations, the STATS19 data will be used for modelling local authority road safety performance in this thesis.

The individual accidents reported in the STATS19 dataset were aggregated by casualty type and accident severity at census Lower Super Output Area level (LSOAs) to form nine area-level accident counts. The LSOAs are geographic areas containing an average of 1,500 people and are defined by the Office for National Statistics (ONS) using measures of population size, mutual proximity and social homogeneity to provide robust small-area statistics for use in comparative analyses. In London there are 4,765 LSOAs contained within 33 boroughs. The STATS19 file contained

[6] STATS19 is the colloqial name for the dataset Road Accident Statistics GB collated by the Department for Transport. The name refers to the title of the form used to collect the data.

data on 13,184 vulnerable road user casualties within London in 2006[7], all of which could be linked to a LSOA based on the location of the collision.

The STATS19 data were downloaded from the UK Data Archive[8] at Essex University and matched to a LSOA using a six figure grid reference (easting and northing) for the accident location, which is recorded by the police at the scene of the accident. Digital boundary datasets for the LSOA were downloaded from the UKBorders repository at the EDINA[9] data archive at the University of Edinburgh. Population 'forecast' data at LSOA level for 2004 were made available by the Small Area Population Estimates team at the Office of National Statistics[10], which is based upon the 2001 census. All of the data were matched using ArcGIS.[11]

For a given road user type and accident severity, an expected casualty count for each LSOA can be calculated to create a model offset using the data sources outlined above. This is based upon the London-wide accident rate per capita (by modal type and severity) and the local population in each LSOA. A simple ratio of the actual (i.e. observed) accident count to the expected accident count (also known as relative risk) can be used as a performance indicator for each transport mode and level of severity. This is a standard approach used in epidemiology and statistics more broadly. However, the reference populations used may not be entirely satisfactory. For example, the number of people 'at risk' of a pedestrian accident in any particular area may be more (or theoretically less) than the population resident in that area.

[7]It is worth noting that strictly speaking these data are accidents that occured in 2004 and were reported for the financial year 2005-6.

[8]www.data-archive.ac.uk

[9]www.edina.ac.uk

[10]http://www.statistics.gov.uk/sape

[11]See www.esri.com/software/arcgis/ for further details.

This would be particularly true in town centres or parts of central London where people may congregate for work, shopping or leisure for instance. Similarly, as noted by Woodward (1983) a better estimate of the motorcyling population may be the number of motorcyle owners in a particular area. The argument can be extended to cyclists as well. However, as table 7.2[12] illustrates, accidents tend to happen close to home for both pedestrians and cyclists but this doesn't tend to hold for motorcyclists; this is intuitive as motorcyclists would tend to make longer journies. Given the focus of government targets to date is on per capita accident rates, the data seem adequate for this purpose but the data could be improved if data on 'exposure' were available by area.

Table 7.2: Distance (Km) between Home and Collision Location for Vulnerable Road Users in 2006

Mode	Median	5 percentile	95 percentile
Pedestrians	1.06	0.06	12.26
Motorcylists	4.22	0.38	16.75
Cyclists	2.14	0.16	10.10

7.3.3 Exploratory Data Analysis

The use of multiple road safety performance indicators relies on a reasonable degree of correlation between the various accident counts. Papageorgiou & Loukas (1988) reported high correlations between fatal road accidents and injuries for data from East Virginia in a bivariate negative binomial model and Bailey & Hewson (2004) report strong correlations between 9 different accident variables recorded for highway authorities in the UK. There are reasonable precedents for assuming that the variables

[12] The home postcode was available for 50% of the accidents in 2006. The distance was calculated as the straight line distance from the postcode centre to the six-figure grid reference for the accident location using ArcGIS.

will be correlated which is confirmed by figure 7.1 – a matrix scatterplot of the nine vulnerable user variables on a log scale. The figure also reports the Pearson correlation coefficients with the size of the font also representing the correlation strength.

The data appear to be reasonably correlated as a set with cycling fatalities being the only departure from this general trend. This is likely to be the result of the very small number (19) of cycling fatalities in 2006. This isn't peculiar to this dataset as Bailey & Hewson (2004) also report very small correlation coefficients for cycling fatalities and the other variables. The highest observed correlation was between cycling slight and motorcyling slight ($\rho = 0.91$). All of the correlations between serious and slight injuries within the same modal type were 0.88 or 0.89.

Similarly, figures 7.2 to 7.4 gives an indication that there is a fair degree of spatial correlation present in the data. Each of the transport modes exhibits spatial clustering with high levels of accidents (darker regions) in the centre of London. A comparison across figures 7.2 to 7.4 indicates that there is a fair amount of shared spatial correlation which may indicate that there is spatial cross-correlations present in the data.

This intuition is supported by formal tests of spatial autocorrelation. Table 7.3 presents the Geary C statistic (Geary 1954)[13]. The statistic ranges from 0 to 2 with a value of 1 indicating that the data were spatially independent, and a number lower (higher) indicating positive (negative) spatial dependence (Cliff & Ord 1981). The p-values are Monte Carlo p-values generated from producing 1,000 replicates. The data have been aggregated by mode to match figures 7.2, 7.3 and 7.4 but this does not affect the result that positive spatial autocorrelation is present in the data.

[13]Geary's C is calculated as $C = \frac{(N-1)\sum_i \sum_j w_{ij}(\boldsymbol{X}_i - X_j)^2}{2W\sum_i(\boldsymbol{X}_i - \bar{X})^2}$ where N is the number of spatial units indexed by i and j; $\boldsymbol{X}$ is the variable of interest; $\bar{X}$ is the mean of $\boldsymbol{X}$; w_{ij} is a matrix of spatial weights; and W is the sum of all w_{ij}

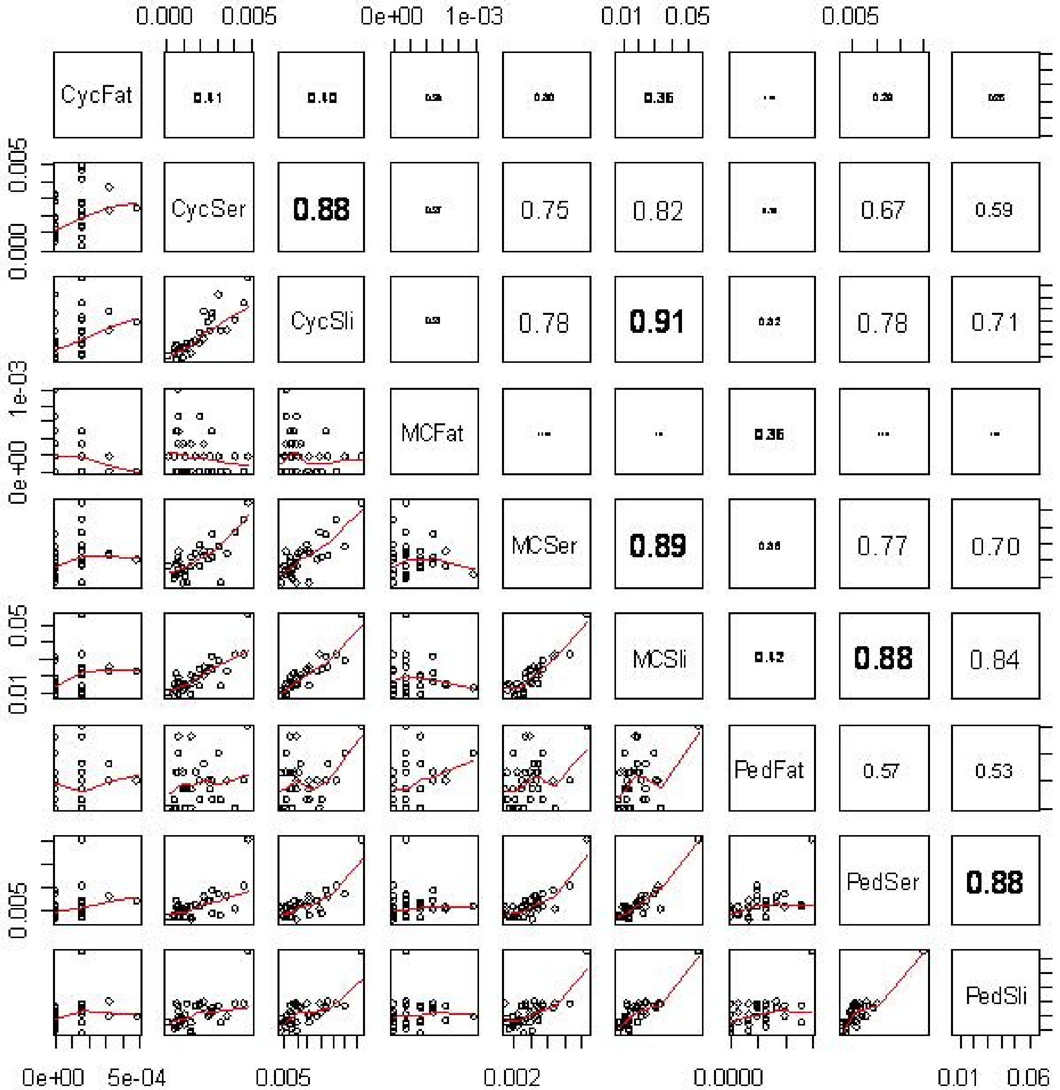

Figure 7.1: Bivariate log-log scatterplots and correlations of accident casualty rates

Table 7.3: Monte Carlo simulation of Geary C statistics

	Statistic	p-value
Pedestrians All	0.654	0.035
Motorcyclists All	0.894	0.046
Cyclists All	0.821	0.044

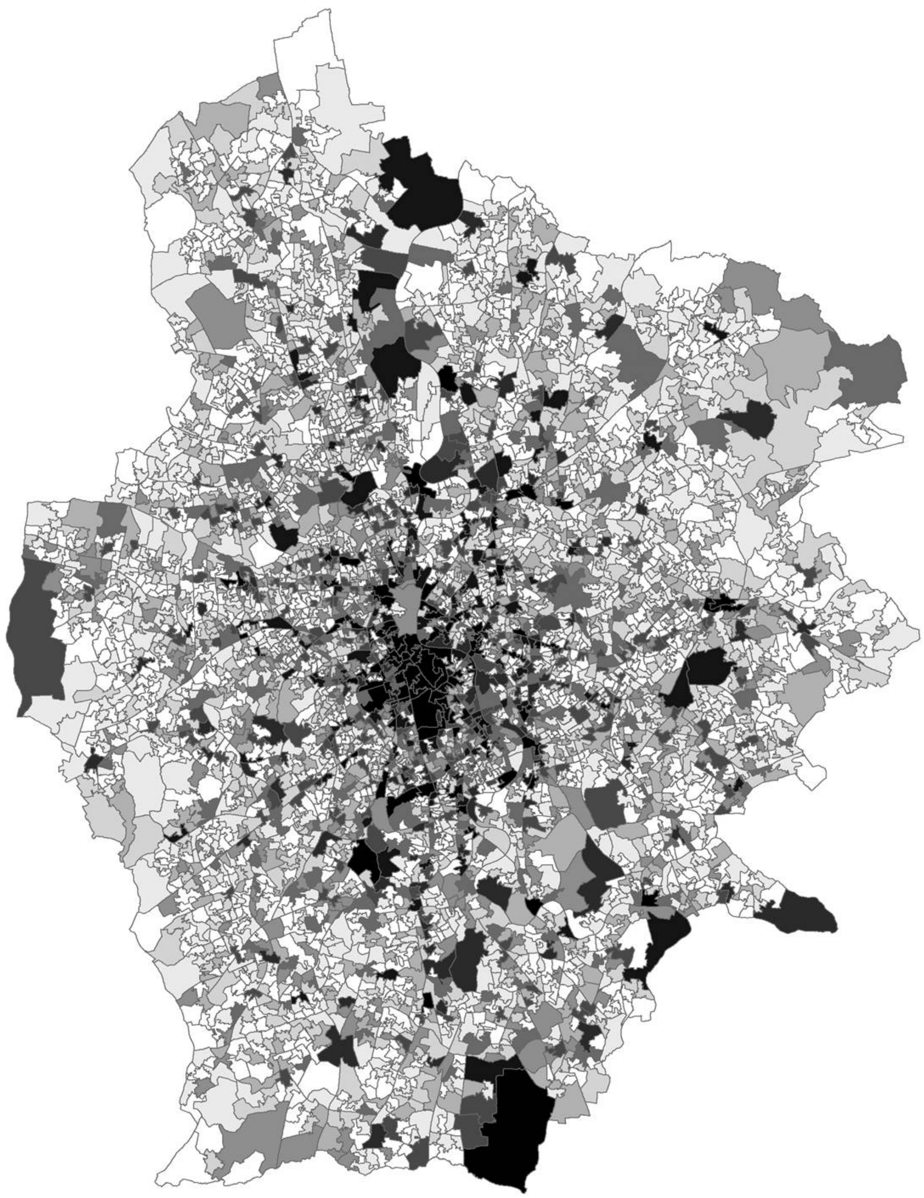

Figure 7.2: A map indicating deciles of London pedestrian casualties in 2006

Figure 7.3: A map indicating deciles of London motorcycle casualties in 2006

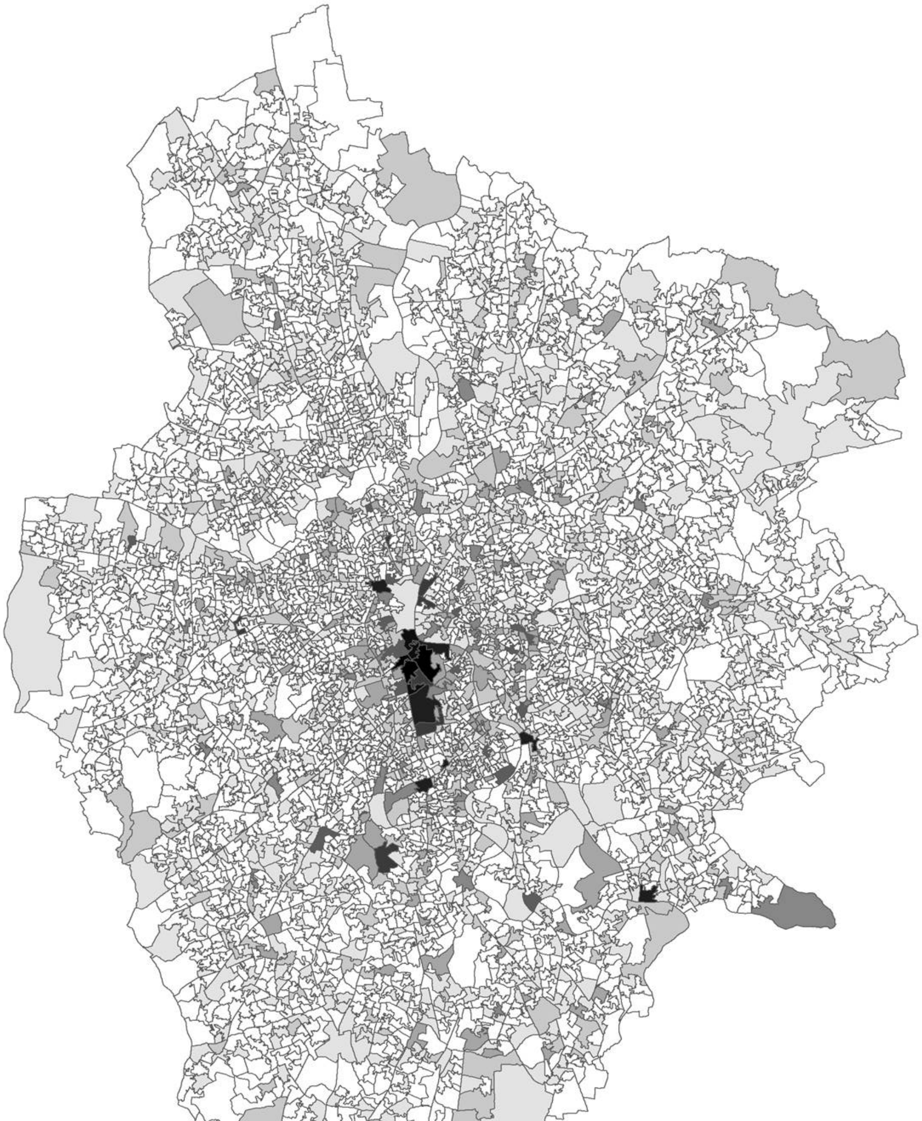

Figure 7.4: A map indicating deciles of London bicycle casualties in 2006

7.4 The Spatial GLMM

7.4.1 Modelling approach

The GLM is well established in the literature for modelling count data (Cameron & Trivedi 1998). Given observed accident counts as described in the last section, the typical model for each univariate count, Y_i in each area i ($i = 1, ..., n$) is the Poisson model: $Y_{ij} \sim \text{Poi}(E_i e^{\theta_i})$, where $E_i = rN_i$ is the offset or expected accident count in area i (where r is the lattice wide accident rate and N_i is the population in area i), and e^{θ_i} is the relative risk or performance of area i. The second stage of the GLM is typically to model the log-linear function: $log(\theta_i) = \beta + \phi_i$ where ϕ_i are area specific random effects. The inclusion of the random effect, which is usually specified in the third stage of the hierarchical model as a zero mean process (e.g. $\phi_i \sim \mathcal{N}(0, \sigma^2)$) results in the GLMM described earlier on page 99.

The principal motivation for modelling accident counts using a GLMM is to interpret these effects as a latent measure of local authority performance. An alternative argument for the inclusion of a random effect term is to capture the lattice-wide heterogeneity or over-dispersion in the data. Frequently count data exhibits over-dispersion (i.e. variability beyond that imposed by the Poisson model's equality of mean and variance). Bailey & Hewson (2004) used a multivariate version of this GLMM to model vulnerable road user casualties recorded for highway authorities. This is a straightforward extension of the univariate case, where the collection of accident counts y_{ij} are a set of $j = 1, ..., p$ variables recorded for each site i ($i = 1, ..., n$) on the lattice. The GLMM now becomes: $y_{ij} \sim \text{Poi}(E_{ij} e^{\theta_{ij}})$, where the offset E_{ij} is now calculated as $r_j N_i$ where r_j is the lattice wide accident count for mode/severity variable j and N_i is the site population. Again a log-linear model is used at the second stage with $\log\theta_{ij} = \beta_j + \phi_{ij}$ with the vector of area specific random effects for any authority $\phi_{i1}, ..., \phi_{ip}$, modelled as a multivariate normal density wih zero mean and unknown p x p variance-covariance matrix i.e. $\boldsymbol{\phi} \sim (0, \boldsymbol{\Sigma})$.

The innovation in this chapter is two-fold. Firstly, by modelling the random effects using the FMCAR prior spatial structure is incorporated into the model. The borrowing of strength across areas as well as across variables will hopefully reduce the uncertainty in the estimates of the local authority random efects. Secondly, an additional level of hierarchy is created as accident counts ($j = 1, ..., 9$) are nested within LSOAs ($i = 1, ..., 4766$) which are nested within local authorities ($k = 1, ..., 33$). This allows for a fine degree of spatial smoothing within the model yet still provides estimates of local authority specific performance.

The data model becomes:

$$y_{ijk} \mid \theta_{\mathbf{ik}} \sim \text{Poi}(\text{E}_{ij}\text{e}^{\theta_{ij}}), \qquad i = 1, ..., n; \;\; j = 1, ..., p; \;\; k = 1, ..., r. \tag{7.4.1}$$

where the E_{ij} are expected counts for variable j in site i derived from the standardized population. Departures from the expected counts are modelled by the parameter, θ_{ij}.

As per chapter 5, using the vector operator to stack columns of the matrix so that $\boldsymbol{\theta}^v \equiv \text{vec}(\boldsymbol{\theta}')$, the second stage of the GLM can be written as:

$$\boldsymbol{\theta}^v \mid \boldsymbol{\beta}, \boldsymbol{V}, \boldsymbol{C} \sim \mathcal{N}_{np}(\boldsymbol{\mu}^v, \boldsymbol{\Sigma}), \tag{7.4.2}$$

where $\boldsymbol{\mu}^v = \text{vec}(\boldsymbol{\mu}')$, $\boldsymbol{\mu} = \boldsymbol{\beta}$, $\boldsymbol{V} = \boldsymbol{\Gamma}^{-1}$, $\boldsymbol{m}_i \equiv \text{diag}(\text{E}_{\text{i1}}, ..., \text{E}_{\text{ip}})$, $i = 1, ..., n$ and $\boldsymbol{\Sigma}$ is as defined in equation (**??**).

To complete the hierarchical GLM specification, the third level or priors must be specified for the matrices, $\boldsymbol{\beta}, \boldsymbol{C}$, and $\boldsymbol{V}$. Vague priors are adopted for the intercept and precision matrix as discussed in chapter 5: $\beta_{\mathbf{k}} \sim \mathcal{N}_p(\mathbf{0}, \sigma^2 \boldsymbol{I})$ and $\boldsymbol{V} \sim \text{Wishart}(\rho, (\rho \boldsymbol{A})^{-1})$. Lastly, to ensure a positive-definite covariance matrix it is important to specify the prior for $\boldsymbol{C}$ carefully. Again, adopting the same priors suggested in chapter 5, it is proportional to $\exp\{-(\boldsymbol{C}^v)'\boldsymbol{C}^v/\xi^2\}$ where $\boldsymbol{C}^v = \text{vec}(\boldsymbol{C})$. The prior distribution must be truncated to ensure that only values of $\boldsymbol{C}$ that provide for a positive-definite $\boldsymbol{G}$ (see section 5.4). A hyperprior distribution for ξ is not used but values of ξ are predetermined. If a hyperprior was used the computational burden on the model would be significant. Smaller values for ξ are preferred as it results in a peaked prior distribution for $\boldsymbol{C}$ centered around zero. Therefore, posterior values of $\boldsymbol{C}$ far from zero is strong statistical evidence against no spatial dependence in the data. Note that use of this prior for $\boldsymbol{C}$ does not favour either symmetric or asymmetric spatial dependence.

7.4.2 Fitting models

To implement the FMCAR thousands of samples are produced from the posterior (5.8.7) using Metropolis-Hastings steps within a Gibbs sampler. After discarding some initial samples called 'burn-in'[14] because they will be autocorrelated, quantities of interest (e.g. the expectation) can be calculated from the posterior. This will be discussed in more detail in chapter 7. For (5.8.7), one iteration of the sampler requires

[14] A good introduction to this and other key concepts of MCMC methods for the unfamiliar is Green (2001).

sampling from:

$$1.\quad \pi(\boldsymbol{\beta}_k \mid \boldsymbol{\beta}_{-k}, \boldsymbol{V}, \boldsymbol{C}, \boldsymbol{\theta}), \quad k = 1, ..., p \tag{7.4.3}$$

$$2.\quad \pi(\boldsymbol{V} \mid \boldsymbol{\beta}, \boldsymbol{C}, \boldsymbol{\theta}), \tag{7.4.4}$$

$$3.\quad \pi(C_{kl} \mid \boldsymbol{\beta}, \boldsymbol{V}, \boldsymbol{\theta}), \quad k, l = 1, ..., p \tag{7.4.5}$$

$$4.\quad \pi(\boldsymbol{\theta}_i \mid \boldsymbol{\theta}_{-i}, \boldsymbol{\beta}, \boldsymbol{V}, \boldsymbol{\Gamma}, \boldsymbol{Y}), \quad i = 1, ..., n \tag{7.4.6}$$

where $\boldsymbol{\beta}_{-k}$ denotes all of the columns of the regression coefficient matrix, $\boldsymbol{\beta}$ except for the kth and similarly for the parameter matrix, $\boldsymbol{\theta}_{-i}$. The term C_{kl} in step 3 represents the (k, l)th element of $\boldsymbol{C}$. The posterior distribution was given in equation (5.8.7) along with the conditional distributions used in each step in the sampler were presented in chapter 5.

Starting values for the sampler could be obtained by setting $\theta_{ij} = \log((y_{ik}+1)/\mathrm{E_{ij}})$ for $i = 1, ..., n$ and $j = 1, ..., p$ where y_{ij} and $\mathrm{E_{ij}}$ are the observed and expected counts respectively. This is a common approach in the epidemiology literature. Given these estimates of $\boldsymbol{\theta}$, a non-spatial GLMM was estimated to obtain the regression intercepts for each accident count ($\boldsymbol{\beta}$) and the residuals were extracted as starting values for the precision matrix $\boldsymbol{V} = \boldsymbol{\Gamma}^{-1}$. To generate initial values for the uniform proposal density for $\boldsymbol{C}$ a coarse grid search was used to maximise (5.8.26).

7.5 Results & Discussion

7.5.1 Model Checking

Five over-dispersed chains were run using an MCMC sampler written in Python for a burn-in period of 5,000 iterations. Convergence of the model parameters was assessed visually using autocorrelation trace plots as well as numerical summaries (e.g. the $\sqrt{R}$ statistic of Gelman (1996)) available in CODA. All of the parameters had $\sqrt{R}$ statistic of approximately 1 and below the 1.2 value suggested in Gelman (1996). The convergence of the precision, $\boldsymbol{V}$, which is important due to the focus on modelling random effects is shown as elipses for each chain at various iterations in figure 7.5. Model fit was assessed using the DIC introduced in the previous chapter. The spatial GLMM was compared to the non-spatial GLMM used by Bailey & Hewson (2004) (although they used data from 2000 for England and Wales indexed at the highway authority level) as a comparison. The FMCAR model produced a DIC of 2859 whereas the standard GLMM had a DIC of 4696 suggesting that the spatial GLMM is significantly preferred.

7.5.2 Random Effects

The principal empirical aim of this chapter is to produce improved estimates of local authority road safety performance. This rests on the spatial multivariate modelling approach generating significant shrinkage of the estimates of the random effects between variables and across sites. This warrants close inspection of the posterior variance-covariance structure. Posterior mean estimates of the correlation between the random effects for variables within sites are presented in table 7.4. These are the bridging parameters or α_0 from the bivariate models formulated in chapters 3, 5 and 6.

Given the range of correlation values (from high positive correlation to low negative correlation) it is apparent that the hyperiors from the FMCAR didn't dominate the data. This isn't surprising given the sample size and the non-informative nature of the hyperpriors chosen. A pattern emerging from the intra site correlations

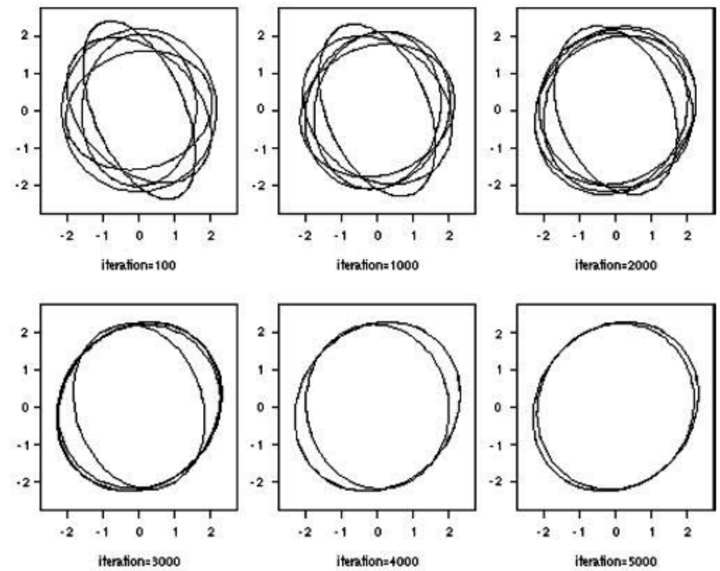

Figure 7.5: Ellipses of the precision matrix for the 5 chains as the iteration number is increased: 100, 1000, 2000, 3000, 4000 and 5000.

Table 7.4: Posterior mean estimates of within site correlations of random effects

Accident Count	1	2	3	4	5	6	7	8	9
1 Pedestrian Fatal	1	0.91	0.82	-0.04	0.44	0.32	-0.28	-0.07	0.19
2 Pedestrian Serious	0.91	1	0.78	-0.09	0.49	0.26	-0.32	-0.11	0.02
3 Pedestrian Slight	0.82	0.78	1	-0.37	0.08	0.34	-0.69	-0.52	0.06
4 Cyclist Fatal	-0.04	-0.09	-0.37	1	0.66	0.76	0.29	0.64	0.64
5 Cyclist Serious	0.44	0.49	0.08	0.66	1	0.64	-0.02	0.53	0.22
6 Cyclist Slight	0.32	0.26	0.34	0.76	0.64	1	-0.49	0.01	0.44
7 Motorcyclist Fatal	-0.28	-0.32	-0.69	0.29	-0.02	-0.49	1	0.67	0.13
8 Motorcyclist Serious	-0.07	-0.11	-0.52	0.64	0.53	0.01	0.67	1	0.43
9 Motorcyclist Slight	0.19	0.02	0.06	0.64	0.22	0.44	0.13	0.43	1

is that there appears to be stronger correlation between accident counts for the same transport mode (e.g. fatal pedestrian accidents and serious pedestrian accidents). There is also some evidence of negative correlations implying that performing well on one area of road safety may occur alongside poor performance on another element of road safety. These negative correlations are most noticeable for motorcycle casualties with pedestrians and cyclists. Bailey & Hewson (2004) reported similar results and posited that this may be the result of separate component structures for "non-motorized transport and for two-wheeled transport" (Bailey & Hewson 2004, p. 510). This seems sensible; pedestrians and cyclists may well share similar characteristics or face similar road safety 'dangers'. Equally, pedestrians and cyclists are fewer in numbers in areas with faster roads and a greater danger for motorcylists. This idea is supported in part by these findings and an additional spatial factor analysis using the same data, which is not reported here. Whatever, one takes from the mixture of positive and negative correlation it must be apparent that aggregating casualty data to overcome the problem of high variability due to the sparse data is not appropriate, and could certainly mask some important information.

Examining the impact of the spatial dependence parameters in the matrix $\boldsymbol{C}$ contained in the variance-covariance matrix will also indicate the extent that the FMCAR prior produces shrinkage beyond the non-spatial GLMM approach considered by Bailey & Hewson (2004). Table 7.5 provides the posterior mean and 95 percent credible intervals for the spatial autocorrelation parameters for each of the nine casualty variables. Recall from the discussion in chapter 5 that although these 'correlations' range from 0 to 1, that they are not correlation coefficients in the usual (Pearson) sense. On the whole, the posterior mean estimates are fairly moderate with wide credible intervals indicative of relatively moderate spatial autocorrelation.

This reflects the moderate patterns seen in the maps in figure **??** and reflect the underlying sparsity of the accident data. Nevertheless none of the credible intervals include zero and there is spatial autocorrelation present, thus modelling it will improve the precision of the random effects.

Table 7.5: Posterior mean estimates of spatial autocorrelation coefficients (and 95 percent credible intervals)

Casualty Variable	α_j	2.5%	97.5%
1 Pedestrian Fatal	0.757	0.173	0.841
2 Pedestrian Serious	0.638	0.088	0.703
3 Pedestrian Slight	0.815	0.215	0.997
4 Cyclist Fatal	0.602	0.027	0.979
5 Cyclist Serious	0.699	0.080	0.970
6 Cyclist Slight	0.589	0.056	0.893
7 Motorcyclist Fatal	0.643	0.142	0.953
8 Motorcyclist Serious	0.713	0.199	0.978
9 Motorcyclist Slight	0.514	0.031	0.694

Turning to the linking parameters (or cross-correlation parameters) there are again signs of moderate correlation although many of the marginal posterior distributions are quite wide. The important advantage of the FMCAR over existing MCAR models is the incorporation of asymmetric linking parameters. The easiest way to compare the two cross-correlation parameters (α_3 and α_4 from the bivariate models in chapter 3, 5 and 6) is to plot the kernel density of their respective posteriors in figures 7.6 to 7.8. Only a representative selection of the plots are presented as there are 72 cross-correlation parameters. Recall that the objective is to reduce the uncertainty in the random effect measure of performance rather than to make substantive interpretation of the spatial autocorrelation and cross-correlation parameters. There are some interesting findings that can be extracted from inspection of figures 7.6 to 7.8. Firstly, the majority of the densities fall to the right of zero indicating positive spatial cross-correlations. This is supported by inspection of the unmodelled accident data in figures 7.2 to 7.4. Generally speaking, central London and hotspots around Heathrow Airport and Croydon have high numbers of all types of accidents a severities. The cross-correlations (the equivalent of α_3 andα_4) often exhibit similar shaped densities which usually overlap although a small number are noteworthy for exhibiting signs of asymmetry. These are presented in figures 7.6 to 7.8

Figures 7.6, 7.7, and 7.8 present a sample of the posterior densities for the cross-correlation or linking parameters for the FMCAR. Of particular interest is figure 7.8 which shows the spatial cross-correlation parameters for motorcylist severe at

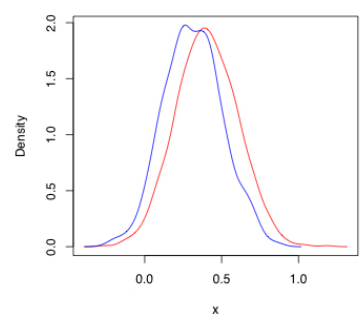

Figure 7.6: Posterior densities of Motorcyle Fatal and Motorcycle Severe spatial cross-correlation parameters

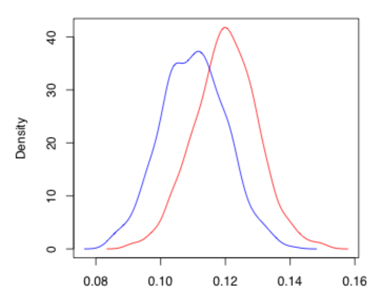

Figure 7.7: Posterior densities of Cyclist Severe and Pedestrian Severe spatial cross-correlation parameters

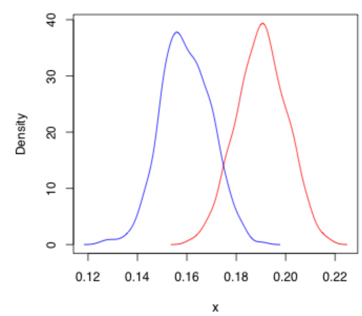

Figure 7.8: Posterior densities of Motorcycle Severe and Cyclist Severe spatial cross-correlation parameters

site i with cyclist severe at site j (right) and motorcyclist severe at site j with cylist severe at site i (left). A comparison of these two densities shows considerable difference in the expectation. The relationship between motorcylist severe at site i with cyclist severe at site j is stronger with less slightly variance than the relationship between motorcyclist severe at site j with cylist severe at site i. Other instances of posterior cross-correlations that are indicative of potentially asymmetric relations are also shown in figures 7.6, 7.7. Figure 7.6 plots the posterior cross-correlations for motorcycle fatal and motorcycle severe and figure 7.7 plots cyclist severe and pedestrian severe. The remaining cross-correlations are not reported as they were, broadly speaking, symmetric.[15] There is considerable overlap in the distributions, however, these cross-correlations deserve further study. Why these particular variables exhibit asymmetric spatial cross-correlation is not clear and there is no guidance from the road safety literature on this topic. Hence this is a completely new finding that deserves further attention.

[15]This is to remain focused on the chapter's objective of modelling road safety ranks.

7.5.3 Road safety performance

The key aim of this chapter was to improve the precision of the local authority random effects and therefore the ranks of local authority relative performance. A major interest is in whether increased precision allows for the separation of local authorities in terms of differential performance or whether the credible intervals of the performance rankings remain overlapping. It is of considerable interest therefore to compare the ranks of the standard (i.e. non-spatial) GLMM model with the ranks of the random effects produced by the FMCAR model.

Figures 7.9 to 7.11 are lattices of plots in which the ranks of each local authority's performance is plotted. Each figure shows the "before and after" ranks with credible intervals for one level of severity. Within each figure the left hand side column are the posterior summaries of the ranks of the of a non-spatial GLMM random effects compared alongside the posterior summaries of the ranks of where the use of a spatial GLMM (FMCAR) has been used to model the random effects. Each row represents a different mode or casualty type: the top row is pedestrians, the middle row is cyclists and the bottom row is motorcyclists. Figure 7.9 shows the fatal accident counts, figure 7.10 the severe accident counts and finally, figure 7.11 plots the slight accident counts. As the local authority specific random effect is estimated at each iteration of the MCMC sampler the rank of that random effect (the latent performance) will vary from iteration to iteration i.e. it will have a posterior distribution which allows the median rank and a 95 percent credible interval for this rank to be produced. This summarizes the uncertainty inherent in the estimate of the performance ranking.

The use of the FMCAR prior is most advantageous for modelling local authority performance for the fatalities indicator. This can be seen clearly in Figure 7.9 where the 95% credible intervals produced by the non-spatial GLMM (left column) are

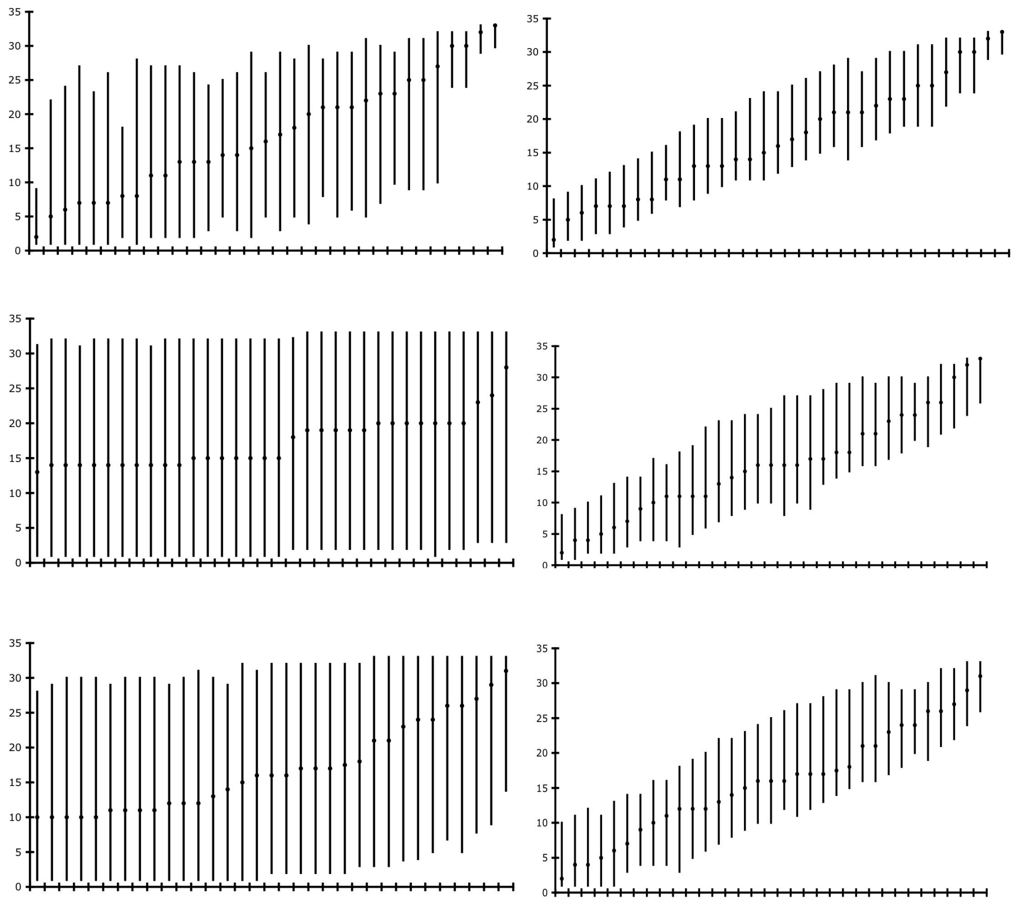

Figure 7.9: Plots of the posterior rank of the local authority performance for fatal accidents. The left column is a standard GLMM and the right column reports the FMCAR. The rows represent pedestrian, cyclist and motorcylist accidents starting from the top.

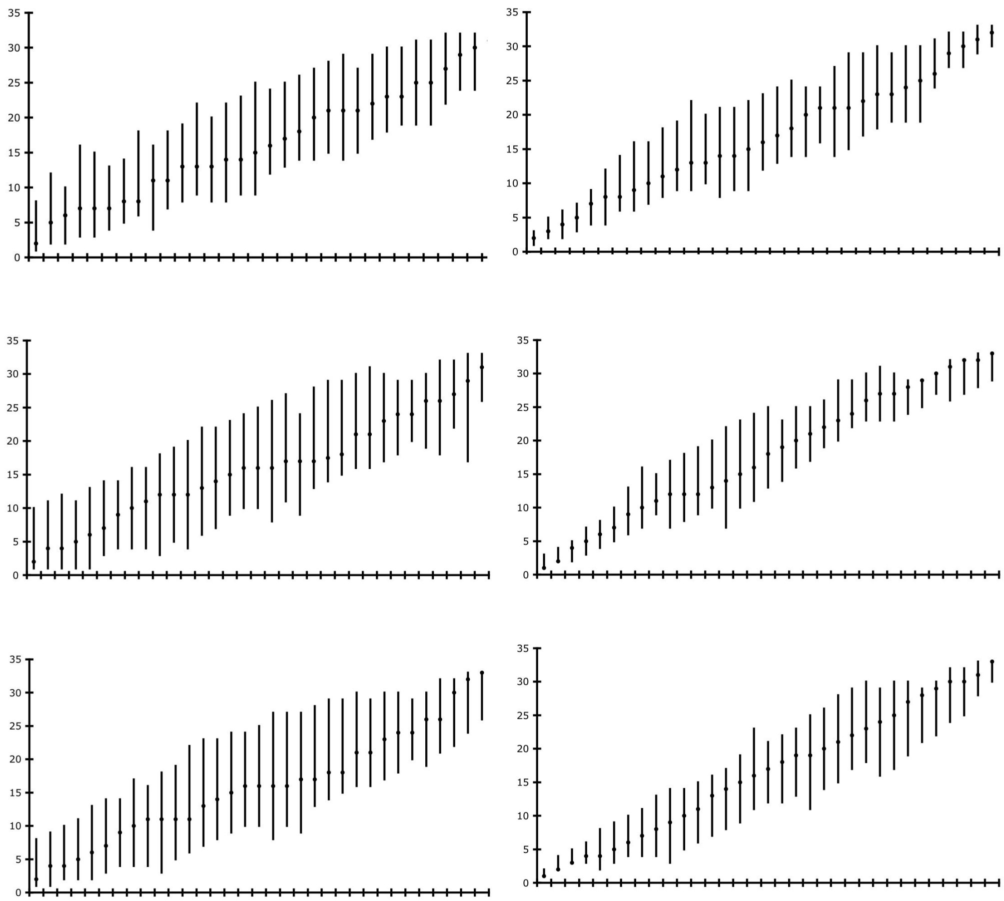

Figure 7.10: Plots of the posterior rank of the local authority performance for severe accidents. The left column is a standard GLMM and the right column reports the FMCAR. The rows represent pedestrian, cyclist and motorcylist accidents starting from the top.

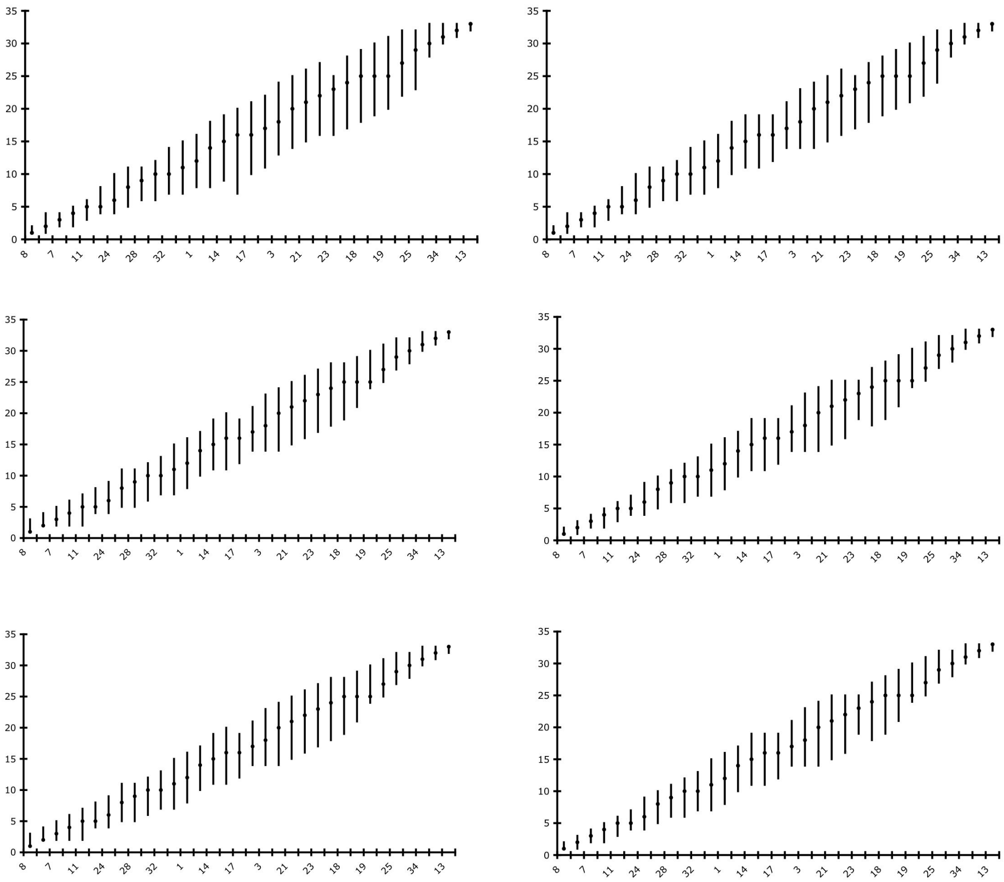

Figure 7.11: Plots of the posterior rank of the local authority performance for slight accidents. The left column is a standard GLMM and the right column reports the FMCAR. The rows represent pedestrian, cyclist and motorcylist accidents starting from the top.

considerably wider than for the FMCAR produced credible intervals. Given the sparse data for fatalities, i.e. they are thankfully rare, there is a great deal of uncertainty in the estimates of local authority ranking. By enabling spatial smoothing across these low count variables the variance in the local authority random effect is shrunk considerably generating much narrower credible intervals. Thus it becomes possible when using the FMCAR prior to separate the top quartile of local authorities from the bottom quartile in a statistically meaningful manner. A similar effect occurs when considering the severe casualties (Figure 7.10) and the slight casualties (Figure 7.11) however the narrowing of the credible intervals is much reduced for these accident types. This is due to the larger number of accident counts and the lower variability in accident counts across areas. As shown in Figures 7.10 amd 7.11, less spatial smoothing is occurring as a result of the FMCAR spatial prior for severe and slight accidents and more information is being provided by the data. Thus the effect of the spatial prior on the posterior local authority ranks is diminished.

It is obvious from even a casual inspection of figures 7.9 to 7.11 that the spatial GLMM using the FMCAR has considerably narrowed the credible interval for the performance ranking. That is, the modelling approach applied in this chapter is substantially superior to the non-spatial GLMM model in terms of the precision of the random effects: there is less uncertainty in the rankings. This effect is most apparent for indicators that have very sparse data (i.e. the fatalities) where the extra structure provided by the spatial correlations has allowed significant shrinkage in the variance estimates. This is the effect of spatial smoothing over neighbouring areas and accident types within areas. This demonstrates the usefulness of the FMCAR model (and spatial smoothing in general) to small area estimates where high variability (excessive zero counts) are encountered. In comparison, relatively little shrinkage has occurred for the slight accident counts due to the relatively large amount of data. Here the data outweighs the structure suggested by the FMCAR prior. Yet despite, the considerable narrowing of credible intervals from adopting the spatial prior, there remains significant overlap and it remains impossible to statistically differentiate local authority road safety performance. The narrowing of the credible intervals is enough to separate the best performing local authorities from the worst performing ones (i.e. there is no overlap in their credible intervals). This would allow for some very broad measure of comparative performance such as ‘above average’ and ‘below average’. For example, Lambeth and Bexley are consistently above average whereas Barking and Dagenham is consistently below average. This may be warrant further research to discover why there is a statistically valid difference between these two extremes. However, despite the reduction in the credible intervals, a number of local authorities would not be able to be statistically categorised into one of these groups. Overall therefore, the picture remains unclear.

7.6 Summary

The results of this chapter have been mixed. The road safety performance indicators have given ample opportunity to demonstrate the implemention and interpretation of the FMCAR parameters in the GLMM setting. Although not conclusive, there is some evidence of asymmetric spatial cross-correlations in the data which indicate a complex relationship between the underlying variables that warrants further exploration. The interaction of people, vehicles and the physical and natural environment creates complex interactions that are difficult to model. The empirical aim of this chapter however, was not to model or understand the casual mechanisms of road accidents but to improve estimation of local authorities’ road safety performance where these accidents are indicators of their (poor) performance. Therefore the inclusion of this spatial structure in the GLMM was to generate additional sources of shrinkage for the estimation of the random effects and their variances. Consideration of the catterpillar plots in figures 7.9 to 7.11 illustrates that this has been achieved yet despite a significant reduction in the uncertainty relating to the performance ranks, differentiating between local authorities remains problematic.

This chapter presented an application of the FMCAR to multiple road safety performance indicators. In doing so it demonstrated the implementation and interpretation of the model developed in chapter 5 with the aim of persuading other researchers to adopt this method. This chapter demonstrated that the FMCAR provides an easy to implement and interpret method for incorporating a very general set of correlations for multivariate data. Additionally, the chapter hoped to improve the estimation and ranking of local authority road safety performance. By applying the FMCAR model that includes additional correlation parameters (for spatial autocorrelation and cross-correlations), the aim was to reduce the uncertainty involved in the estimation of local authority random effects (performance) to enable performance between authorities to be differentiated. Although the use of the FMCAR did improve the precision of these estimates and reduced the credible intervals for the random effects, there was still a significant degree of overlap. Therefore, although the FMCAR achieved the considerable reduction in uncertainty it aimed for, it failed to make performance management significantly clearer. Mirroring the findings of Bailey & Hewson (2004), despite an improvement in the methods the resulting performance rankings remain 'fuzzy'. Thus, there seems to have been little progress since Goldstein & Spiegelhalter (1996) stated that rankings that incorporate the uncertainty in the rankings in a statistically valid manner frequently make it impossible to separate the organisations under study.

CHAPTER 8

GENERAL DISCUSSION

8.1 Introduction

This thesis has made several contributions to the theoretical and applied literatures. From a theoretical perspective a new, more flexible, multivariate conditional autoregression has been developed and its performance against existing approaches tested. From an empirical perspective, the application of this FMCAR model to multiple road safety performance indicators has led to a significant reduction in the uncertainty of local authority performance rankings. The aim of this chapter then is to discuss the findings of this research in light of the original research aims and objectives, as well as placing the findings in the context of the extant literature.

8.2 Aims and Objectives

Recall from chapter 1 that there are two motivations for this research. One is methodological — to extend the range of multivariate models available for spatially correlated data; and the second is applied — to improve the road safety performance ranking of local authorities. In particular, the specific objectives of this research are:

1. Develop a flexible multivariate conditional autoregression that allows asymmetric inter site spatial correlations.

2. Demonstrate the performance of the model through a comparison with existing models using simulated data.

3. Demonstrate the applicability of this model through investigation of multiple traffic safety performance indicators in London.

4. Contribute to an improvement in public sector performance management by reducing the uncertainty of performance rankings through the inclusion of spatial correlation.

5. Provide a more general method for ranking public sector organisations than Data Envelopment Analysis and Stochastic Frontier Analysis.

6. Contribute to the road safety literature by identifying good and weak performing local authorities.

7. Provide the relevant computer code to perform parameter estimation, statistical inference and diagnostics within the Bayesian paradigm.

8. Provide a thorough introduction to multivariate conditional autoregression models.

Each of these objectives will be discussed in turn, in terms of the findings of the research.

8.2.1 Develop a flexible multivariate conditional autoregression

This thesis has focused on multivariate CAR (MCAR) models. As in the univariate models presented in chapter 2, spatial dependence is captured through the covariance matrix, or rather its inverse. Previous research efforts have used simple forms for the covariance matrix that, although computationally convenient, unduly constrain the range and/or type of correlation modelled. Mardia (1988) provided the theoretical groundwork for multivariate Gaussian CAR models. The problem with Mardia's original multivariate specification was that it required separable models that necessitated identical spatial parameters for each variable. The MCAR models of Gelfand & Vounatsou (2003) and Carlin & Banerjee (2003) are almost identical in their approach, although Carlin & Banerjee (2003) extend their model to spatio-temporal data. These MCAR models are suitable for non-separable models but do not allow for between area correlations. In comparison, the "two-fold CAR" model of Kim et al. (2003) provides a more flexible correlation structure incorporating both *bridging* and *linking* spatial parameters i.e. within area correlations and between area cross-correlations. Despite being very flexible, this model is only suitable for the bivariate case and extension to higher dimensions using the same approach has been impossible. Jin et al. (2005) propose an alternative framework for including cross-correlations into the traditional MCAR model which they term the Generalized MCAR model. This model specifies the joint distribution for a Markov random field in terms of a combination of simpler conditional and marginal distributions. In such, they are adapting the multivariate geostatistical model of Royle & Berliner (1999) to the lattice case. However, as discussed in chapter 3, the order in which the conditional and marginal variables are considered (i.e. should $\pi(\boldsymbol{y}_1|\boldsymbol{y}_2)$ be modelled and then $\pi(\boldsymbol{y}_2)$ or the other way around) is completely arbitrary. To combat this complaint, Jin et al. (2005) propose to use model comparison techniques such as the Deviance Information Criterion (DIC) to decide on the modelling order, yet this seems infeasible with more than a few variables.

In summary, the existing literature provides for only a handful of MCAR models, none of which are entirely satisfactory. The first objective of this thesis therefore was to develop a flexible multivariate conditional autoregression that allows asymmetric inter site spatial correlations. Chapter 5 clearly achieved this with a model that is both flexible and easy to implement. The conditional distributions and MCMC sampling scheme were discussed in great depth and this model should be fairly easy to implement for researchers confident in statistical programming. A novel feature is the removal of the dependence on the number of neighbours in estimating the conditional means, and instead the inclusion of a precision measure that allows for variation in the data to influence the correlation. The conditions necessary to ensure that the joint covariance matrix exists were introduced and interpretation of the spatial autocorrelation parameters was discussed. The FMCAR, therefore, fills an important gap in the literature.

8.2.2 Demonstrate the model's performance through a comparison

The additional flexibility of the FMCAR comes at a cost; there are more parameters to estimate and the model is computationally more difficult to implement than the original MCAR model of Mardia (1988). It is therefore important to consider how well the FMCAR compares to the alternative models considered in chapter 3. This was the motivation for this research objective. When the spatial relationships between variables across sites on the lattice is asymmetric then the FMCAR model should be preferred. However, when the relations are symmetric how well does the FMCAR perform in comparison to the existing approaches? This question was answered by comparing the performance of the main multivariate spatial models using both simulated and real data.

In addition, very little guidance is available for applied researchers on which MCAR model is most appropriate in which situations. In comparison there a number of comparative studies of univariate models that can help guide researchers in their model choice. For instance, Lawson et al. (2000) compared a range of univariate spatial models according to goodness of fit criteria and Richardson et al. (2004) compared the smoothing of disease risk performed by different univariate models and therefore their ability to detect hightened risk. More recently, Best et al. (2005) produced a thorough comparison of univariate models that extended their coverage beyond CAR models to semi-parametric and moving average models. Currently no similar simulation study exists for multivariate models. Chapter 6 therefore plugs an important gap.

MCAR models are typically deployed in a Bayesian hierarchical framework. It is a well established property of Bayesian inference that Bayesian procedures offer a trade-off between bias and variance reduction of estimates (Carlin & Louis 2000). Bayesian hierarchical spatial methods are known to produce point estimates with good properties in terms of Minimum Squared Error loss. Variance reduction in Bayesian methods is achieved through the borrowing of strength or information within the hierarchical structure. The result is point estimates that are shrunk towards a 'global average' from the distribution of all the units included in the hierarchy. The effect of this shrinkage is dependent upon the prior structure assumed and conditional upon this structure being close to the 'true' population model. Returning to the spatial setting, the different MCAR models will produce different levels of shrinkage when used as spatial priors in a hierarchical model. Therefore it will be useful to compare the variance and bias trade-offs implied by each model.

A comparison of the model fit, variance and bias trade-off was achieved through a combination of a simulation study and an application to a small real-world dataset on cancer mortality in West Yorkshire. For the simulation exercise, four studies were used where each of the four competing MCAR models was assumed to be the true data generating process. In each of these studies 100 datasets were simulated and the four models were fitted to the datasets. This allows a comparison of the models across a range of possible spatial configurations. In general, the performance between the FMCAR, GMCAR and MCAR was comparable for all situations except the presence of asymmetric linking parameters (cross-correlations) i.e for all but the case where the FMCAR generated the data. This is both good and bad news. On the positive side, fitting the FMCAR when asymmetric spatial cross-correlations are not present does not have deleterious effects. It is therefore safe to fit the FMCAR without a priori reasons to suspect asymmetric cross-correlations as there is little impact in terms of model fit. However, the FMCAR offers little or no benefit over existing spatial priors in the absence of asymmetric cross-correlations. Therefore persuading researchers to adopt the FMCAR will depend upon the relative costs of implementing this model over the potential benefits of avoiding model mis-specification. However, this is to be expected. A parsimonious model should be preferred wherever possible and fitting the more complex FMCAR when it isn't required should be avoided.

8.2.3 Demonstrate the applicability of this model

The application of the FMCAR model to multiple road safety performance indicators provided a good opportunity to demonstrate the implementation and interpretation of the model to real-world data. Chapter 5 went into considerable depth regarding the theoretical implementation of the model using MCMC within a Bayesian hierarchical framework. This was accompanied by detailed discussion about interpretation of model parameters. However, one aim of chapter 7 was to demonstrate the model applied to data. Examining the spatial dependence parameters in the matrix $\boldsymbol{C}$

contained in the variance-covariance matrix indicates the extent that the FMCAR prior produces shrinkage beyond a non-spatial GLMM. On the whole, the posterior mean estimates are fairly moderate with wide credible intervals indicative of relatively moderate spatial autocorrelation. Nevertheless none of the credible intervals include zero and there is spatial autocorrelation present, thus modelling it will improve the precision of the random effects.

Furthermore, the majority of the posterior kernel densities for the spatial cross-correlations fall to the right of zero indicating positive spatial cross-correlations. This is evidence that a model that allows for cross-correlations (e.g. GMCAR or FMCAR) should be preferred over traditional MCAR varients. However, the innovation in the FMCAR is the ability to model *asymmetric* cross-correlations in spatial datasets. From the findings in chapter 7, it is apparent that the road accident data used in this thesis exhibit only mild asymmetric cross-correlations at best. There were only a few noteworthy departures from symmetric cross-correlations. For example, the relationship between motorcylist severe at site i with cyclist severe at site j is stronger with less variance than the relationship between motorcyclist severe at site j with cylist severe at site i. Although it wasn't possible to know a priori that there wouldn't be asymmetric cross-correlations, it would have obviously been better to apply the FMCAR to a dataset where it was truly beneficial and where meaningful interpretation of the asymmetric cross-correlations could take place.

One important thing that applying the FMCAR has demonstrated is that there is spatial correlation and cross-correlation present in the STATS19 data. These findings are consistent with the most recent research on this topic such as the two studies by Hewson (Hewson (2004) and Hewson (2005)) using STATS19 for the county of Devon, UK. Both papers explictly test and model for spatial autocorrelation using a univariate CAR model and find some degree of positive spatial autocorrelation. This has severe implications for the existing road safety research briefly reviewed in chapter 4. In particular, a series of papers (Grayling et al. (2002), Graham et al. (2005) and Graham & Glaister (2005)) have all attempted to explain accident rates through area level deprivation covariates. It seems that both of these variables could be spatially correlated and that the findings of these papers – that there is a high correlation between area deprivation and accident rates – may require reviewing.

8.2.4 Reducing the uncertainty of performance rankings through the inclusion of spatial correlation.

As chapter 4 discussed, it is bad practice to use point estimates of indicators to rank performance across observational units (e.g. local authorities) without accommodating the uncertainty inherent in these estimates. A well-established method for accounting for uncertainty in the estimates of performance is the use of Generalized Linear Mixed Models where the performance measure is assumed to be a random effect with unknown variance. Research by Goldstein & Spiegelhalter (1996) suggests that rankings adopting robust procedures that incorporate uncertainty in a statistically valid manner frequently make it impossible to separate the organisations under study. (Bailey & Hewson 2004) attempts to reduce the uncertainty in the random effects (and therefore the performance ranking) via modelling multiple performance indicators simultaneously. Assuming that these multiple measures are correlated allows for 'borrowing of strength' across the multiple measures of performance and a reduction in the uncertainty associated with each organisation's performance. Additionally, as these data are count data exhibiting significant variability (due to the preponderance of zeros) spatial smoothing induced by spatial autoregressive models should improve the interpretability of th results. Despite, considerable improvement in the resulting league table Bailey & Hewson (2004) are still unable to separate organisations based upon differential performance. This was the motivation behind this research objective. By adding additional forms of correlation through the modelling of spatial correlations it was hoped that this uncertainty in the performance ranking could be shrunk further. This is especially important when the performance indicators are road accidents as the data are sparse and subject to great variability. This makes robust inference about local authority performance difficult.

Chapter 7 was remarkably successful in achieving this objective. It is obvious from even a casual inspection of ranks reported in chapter 7 that the spatial GLMM using the FMCAR has considerably narrowed the credible interval for the performance ranking. That is, the modelling approach applied in this chapter is substantially superior to the non-spatial GLMM model in terms of the precision of the random effects: there is less uncertainty in the rankings. This effect is most apparent for indicators that have very sparse data (i.e. the fatalities) where the extra structure provided by the spatial correlations has allowed significant shrinkage in the variance estimates. In comparison, relatively little shrinkage has occured for the slight accident counts due to the relatively large amount of data.

8.2.5 Provide a more general method for ranking public sector organisations

The problems in measuring the output of the public sector frequently occur for the same reasons that these goods and services have to be delivered by the public sector in the first place. There are three main problems with measuring public service performance: identifying outputs, the lack of prices and the problem of attribution. This lack of prices and the problem of aggregation makes it very difficult to employ traditional economic assessment of efficiency such as parametric Stochastic Frontier Analysis or non-parametric Data Envelopment Analysis. However, a growing number of academics challenge the relevance of productivity analysis to public sector data (Stone (2002*a*), Stone (2002*b*) and Smith & Street (2005)). The relevance of a single measure of (in)efficiency is questionable from a managerial perspective. There are also reasonable concerns about the relevance of a production function approach to the analysis of public sector organizations where the production process is not well understood, and probably cannot be modelled well by traditional production analysis. Moreover, the results from SFA and DEA are sensitive to the model assumptions and in particular, the level of inefficiency is dictated by the signal to noise ratio in the data. Given the number of omitted factors in public sector analyses this can be problematic for the correct interpretation of these analyses. More importantly for this research, traditional methods of performace measurement do not take into account the inherent uncertainty in the estimates themselves.

Chapter 7 envisaged road safety performance as a latent output that was measured by several related and correlated measures: accident counts by modal type and severity. By exploiting the correlation across observational units (space) and across variables, this thesis has produced quite a simple yet statistically robust method of estimating organisation's performance.

8.2.6 Identifying good and weak performing local authorities

A more specific objective than the one to reduce the uncertainty in the performance ranking was to reduce the uncertainty by enough to enable differential performance to be identified i.e. to be able to statistically separate organisations based upon their performance. Ultimately, this is the goal of league tables and performance rankings. Yet despite, the considerable narrowing of credible intervals from adopting the FMCAR as a spatial prior, there remains significant overlap and it remains impossible to statistically differentiate local authority road safety performance. The narrowing of the credible intervals is enough to separate the best performing local authorities from the worst performing ones (i.e. there is no overlap in their credible intervals). This would allow for some very broad measure of comparative performance such as ‘above median’ and ‘below median’, although for the majority of local authorities it would not be possible for them to statistically categorised into one of these groups. Overall therefore, the picture remains unclear. This result mirrors the findings of Bailey & Hewson (2004), in which despite an improvement in the methods the resulting performance rankings remain ‘fuzzy’. Thus, there seems to have been little progress since Goldstein & Spiegelhalter (1996) stated that rankings that incorporate the uncertainty in the rankings in a statistically valid manner frequently make it impossible to separate the organisations under study.

8.2.7 Provide the relevant computer code

This objective has been met through the inclusion of Python code for the FMCAR in the appendix. In addition, and more usefully, the full conditional distributions and an MCMC sampling scheme are discussed in chapter 5. Specifics of selecting priors, starting values and issues relating to the monitoring of convergence are discussed in some depth in chapter 7.

8.2.8 Provide a thorough introduction to MCAR models

Taken as a whole this thesis should provide a solid introduction to MCAR models. A great many textbooks exist that cover univariate SAR and CAR models or geostatistical in depth (e.g. Cressie (1993)) but this thesis allows the reader to progress from time-series autoregressions to MCAR models in a succinct yet thorough fashion. The comparison of MCAR models provided in chapter 6 and the demonstration of the implementation and interpretation of the FMCAR in chapter 7 complete the coverage.

8.3 Summary

This chapter has discussed the findings of this research in light of the original research objectives and the literature. This research has made two main contributions. Firstly, a new flexible model has been presented for multivariate spatial data recorded on lattice. The principal innovation in this model is the incorporation of asymmetric spatial cross-correlation parameters. The second contribution is empirical. By applying the FMCAR to data on multiple road safety performance indicators considerable shrinkage in the estimation of the organisations' performance has been achieved. This extends the range of models available for performance measurement and is particularly useful in the public sector where traditional econometric approaches to measuring efficiency are inappropriate. The next chapter concludes by summarising the research and by discussing the limitations, future directions and policy relevance of the thesis.

CHAPTER 9

CONCLUSION

9.1 Introduction

In the last 30 years since Jean Paelinck introduced 'spatial econometerics' into the economic lexicon, there has been an exponential increase in the methods for, and applications of, spatial econometrics. The presence of spatial dependence can cause problems for econometric models, both in terms of efficiency and bias. When the geographic location of the observations is known it is common to assume that observations at sites near each other may have a similar value on the omitted variables in the regression causing the error terms to be serially autocorrelated. This autocorrelation is typically modelled through an autoregressive model, two of which dominate the literature: the Simultaneous Autoregression (SAR) introduced by Whittle (1954) and the Conditional Autoregression (CAR) promoted by Besag (1974). The analysis of spatially correlated data is now an active area of research in both applied and theoretical econometrics. With the exception of Gamerman & Moreira (2004) and Kelejian & Prucha (2004)) this research has been limited to univariate data, yet many economic problems are inherently multivariate and there has been a long history of multivariate methods in econometrics. In comparison, there have been a number of significant contributions to multivariate methods within statistics, including recent contributions from Gelfand & Vounatsou (2003), Carlin & Banerjee (2003), Kim et al. (2003), and Jin et al. (2005) – which were reviewed in chapter

3. This thesis therefore focused on extending the range of multivariate models and in doing so draws on both the statistics and econometrics literatures. This chapter reviews the principal findings of the thesis in light of the original research aims and objectives. It considers both the limitations and contributions made by this thesis and suggests future directions for research.

9.2 Findings and Contributions

Recall from chapter 1 that there were two motivations for this research. One was methodological — to extend the range of multivariate models available for spatially correlated data; and the second was applied — to improve the road safety performance ranking of local authorities. Chapter 2 bridged the gap between time series autoregressions and univariate spatial autoregressions, introducing some of the key concepts in the analysis of spatially correlated data such as neighbourhoods. The key contribution of this chapter was the justification of the conditional modelling approach adopted in this thesis. This rested on four main arguments: (i) that the CAR achieves minimum mean squared prediction error and maximum entropy (ii) the CAR model is naturally interpreted in the conditional expectations sense familiar to econometricians that isn't true of the SAR model, (iii) the ease with which CAR models can be implemented through MCMC methods due to their conditional specification, and (iv) the ease with which the CAR can be used to model discrete data through GLMMs.

Building from chapter 2's presentation of univariate models, chapter 3 reviewed the literature on multivariate spatial regression models. It identified the principal challenge when modelling multivariate lattice data: guaranteeing valid covariance matrices in the joint probability model whilst allowing for correlation both between variables *within* sites and between variables *across* sites. Three existing multivariate conditional autoregression (MCAR) models were discussed in detail along with their limitations. The MCAR models of Gelfand & Vounatsou (2003) and Carlin & Banerjee (2003), which originated in the work of Mardia (1988), are the least general. They incorporate a spatial autocorrelation parameter for each variable and a non-spatial correlation term for the variables. The "two-fold CAR" model of Kim et al. (2003) and the Generalized MCAR (GMCAR) model of Jin et al. (2005) both adopt an additional linking parameter (or cross-correlation) which allows spatial correlation between variable 1 at site i and variable 2 at site j. As such they are more general than the simple MCAR model, but chapter 3 argued that they also have their weaknesses. In particular, the "two-fold CAR" is only suitable for bivariate data and the GMCAR suffers from problems with implementation and interpretation. More generally, neither of the models allow for this spatial cross-correlation to be asymmetric, and this resulted in the theoretical motivation for this thesis.

In chapter 4, the empirical objective of the thesis was motivated through a review of performance measurement. This chapter suggested that traditional econometric methods for measuring and ranking performance (e.g. Stochastic Frontiers) were inappropriate for multi-output public sector organisations where prices are missing and input data are scarse. Looking specifically at output-based performance indicators, it was argued that a common problem with performance indicators in the public services is that they usually make no allowance for the inherent uncertainty in both the underlying performance being measured, or any rankings of this performance. Moreover, when attempts are made to incorporate uncertainty into performance measurement – for example through the use of generalized linear mixed models (GLMMs) – the resulting credible intervals relating to the performance rank are typically large and overlapping. The obvious result is that it becomes impossible to differentiate the relative performance of organisations. Given the increasing reliance on performance management in the public sector, and the trend towards 'payment for performance', chapter 4 argued that improving performance measurement was a worthy endeavour. Additionally, given the large and growing number of road traffic accidents – with the resulting impact on the economy and society – any methods that could identify road safety excellence or weakness is worthwhile. Chapter 4 suggested that one method of reducing the uncertainty in performance ranking would be to add further structure to the GLMM random effects. One source of additional structure for accident data could be spatial autocorrelation and spatial cross-correlations.

Having identified a gap in the theoretical literature in chapter 3, the innovation in this thesis was the development of a *flexible* MCAR model i.e. one that allows asymmetric cross-correlations between different variables at different sites. Chapter 5 presented this innovation. Another novel feature of this model is the removal of the dependence on the number of neighbours in estimating the conditional means for each site, and instead the inclusion of a precision measure that allows for variation in the data to influence the correlation. As identified in chapter 3, the major challenge in the literature has been ensuring that a valid joint covariance matrix exists. The conditions that ensure this were also presented in chapter 5 along with an MCMC sampling scheme for fitting the model to data.

The additional flexibility that is the hallmark of the FMCAR comes at a cost; there are more parameters to estimate and the model is computationally more difficult to implement than existing approaches. It was therefore important to consider how well the FMCAR compared to the alternative models and this was the purpose of chapter 6. This comparison was achieved through the use of simulated and real world data and included both continuous and discrete variables. In general, the performance between the FMCAR, GMCAR and MCAR was comparable except in the presence of asymmetric cross-correlations in the simulated data. The major finding of chapter 6 was that this was both good and bad news. On the positive side, fitting the FMCAR when asymmetric spatial cross-correlations are not present will not have deleterious effects. Therefore it is safe to fit the FMCAR without a priori reasons to suspect asymmetric cross-correlations. Additionally, the FMCAR model does offer attractive properties in terms of shrinkage of the estimated spatial random effects variance, which was expected to be of considerable use in modelling performance indicators. However, the FMCAR offers little or no benefit over existing spatial priors in the absence of asymmetric cross-correlations. The conclusion drawn was that, given the negligible penalty of adopting the FMCAR by default, the attractiveness of this model to other researchers will depend upon the relative overhead of implementing this model.

Chapter 7 had two empirical aims. Firstly, originating from the discussion in chapter 4, to reduce the uncertainty in the estimates of (road safety) performance measurement through the addition of the spatial structure inherent in the FMCAR model. Secondly, given the results of the model comparison in chapter 6, to demonstrate the implementation and interpretation of the FMCAR model to real data. The results of this chapter were mixed. The road safety performance indicators provided a good opportunity to demonstrate the implemention and interpretation of the FMCAR parameters in the GLMM setting. Although not conclusive, there was some evidence of asymmetric spatial cross-correlations in the casualty variables which warrants further exploration. The use of the FMCAR improved the precision of the local authority performance estimates and therefore reduced the credible intervals for the rankings. Mirroring the findings of Bailey & Hewson (2004), there remained a significant degree of overlap of these credible intervals making it impossible to statistically differentiate performance. The tentative conclusion of this chapter was therefore, that there seems to have been little progress since Goldstein & Spiegelhalter (1996) stated that rankings that incorporate uncertainty in a statistically valid manner frequently make it impossible to separate the organisations under study. Despite this negative conclusion the FMCAR did induce a considerable degree of shrinkage, in part achieving the empirical goal of this chapter. Moreover, the spatial smoothing induced by the model aids intreptation of a dataset with a large proportion of zero counts.

In chapter 1, the specific objectives of this research were specified as:

1. Develop a flexible multivariate conditional autoregression that allows asymmetric inter site spatial correlations.

2. Demonstrate the performance of the model through a comparison with existing models using simulated data.

3. Demonstrate the applicability of this model through investigation of multiple traffic safety performance indicators in London.

4. Contribute to an improvement in public sector performance management by reducing the uncertainty of performance rankings through the inclusion of spatial correlation.

5. Provide a more general method for ranking public sector organisations than Data Envelopment Analysis and Stochastic Frontier Analysis.

6. Contribute to the road safety literature by identifying good and weak performing local authorities.

7. Provide the relevant computer code to perform parameter estimation, statistical inference and diagnostics within the Bayesian paradigm.

8. Provide a thorough introduction to multivariate conditional autoregression models.

This thesis has met all of these objectives. The development of a flexible conditional autoregression that allows asymmetric inter site spatial correlations was reported in chapter 5 and a demonstration of the performance of the model through a comparison with existing models using simulated data was then reported in chapter 6. Chapter 7 met the next four objectives including: demonstrating the applicability of the the FMCAR to performance management data; a reduction in the uncertainty of rankings of local authority road safety performance; demonstrating that the Generalized Linear Mixed Model is a more general method for ranking public sector performance than traditional econometric techniques; the identification of good and weak performing local authorities. As already reported, this last aim was only partially fulfilled by the FMCAR model. Chapter 5 provided the full conditional distributions and a sampling algorithm to fit the FMCAR model and chapter 7 demonstrated how to interpret and validate the model output. The actual Python computer code used in this research is reserved to the appendix, which satisfies objective 7. Lastly chapters 2 and 3 provide a thorough introduction to multivariate conditional autoregression models, especially when read in conjunction with the detailed appendices and chapter 6 which compares multivariate conditional autoregressions used both simulated and real data.

Besides the production of this thesis, the achievement of these research objectives are demonstrated through the following specific outputs:

1. Chapter 2 (the univariate review) with additional material has been accepted,

via peer review, for presentation at The Academy of Management annual meeting in August 2009 in the research methods division under the title "*Space the final frontier: spatial regression models in organizational research*". It has also been invited to be submitted to Organizational Research Methods which will be done after the conference.

2. Chapter 5 introduced the Flexible Multivariate Conditional Autoregression model, which incorporates asymmetric cross-correlations (or linking parameters). This has been accepted, via peer review, for the European meeting of the Econometric Society in August 2009 under the title "*A Flexible Multivariate Model for Areal Data*". It will be revised with a new application and ultimately submitted to an econometrics journal.

3. Chapter 6 used both simulated and real data to compare the FMCAR to existing MCAR specifications. This chapter with chapter 3 (the multivariate literature review) has been submitted to Statistics Surveys under the title "*A Comparison of Multivariate Conditional Autoregressions*".

4. Parts of chapters 4 and 7 have been presented at the King's College London Social Science conference under the title: *"Road traffic accidents: can we assume spatial independence?"* and in revised form have been accepted, via peer review, for presentation at The Royal Statistical Society annual conference in September 2009 under the title "*Improving performance ranking through a spatial GLMM."* A revised version of this chapter will eventually be submitted to a social statistics or public policy journal.

9.3 Challenges and Opportunities

9.3.1 Limitations

As with all pieces of research, there are inevitably some limitations to this thesis. In terms of the FMCAR model itself there are a number of minor issues that could be addressed in further work. Firstly, in section 5.4 the conditions necessary to ensure the existence of a valid covariance matrix were presented and discussed. Recall from the discussion in chapter 3 that this has proved the greatest challenge to researchers developing MCAR models in the literature. This difficulty explains why there is such a limited body of work on MCARs. Early versions of the FMCAR relied on the rather blunt and restrictive condition of ensuring diagonal dominance of the covariance matrix through restricting the MCMC sampler. This was replaced by a neater solution which imposed a restriction on the singular values of the spatial correlation matrix. This condition may be seen as a limitation and there is room to investigate the relaxation of this condition in the future.

Another potential limitation of the model is the complexity of fitting the model to data using MCMC methods. It is not possible to use a Gibbs sampler on its own; instead two Metropolis-Hastings steps must be used to simulate draws for the precision and the spatial correlation parameters. This combined with a number of large matrix inversions makes the MCMC computationally challenging, requiring liberal use of sparse matrix methods. There are a range of tools available in most programming and statistical languages to handle sparse matrix manipulation. Further, the MCMC methods required are not at the cutting-edge of stochastic simulation. Nevertheless, a limitation of this model is the need to have a fairly good understanding of computational statistics and stochastic simulation in order to be able to adapt the

FMCAR to problems beyond the hierarchical GLMM presented in this thesis. This will no doubt limit the appeal of the model to other researchers and an important area of future research would be to produce 'routines' or scripts to automate this. In fact, this need and the skills learnt during the course of this research have prompted the development of a MCMC sampling 'application' developed in Python. Extending this to a broad class of 'everyday' models is still in the early stages.

Given the large number of parameters that the FMAR is estimating, large datsets are required to ensure efficient estimation of model parameters. In this thesis there were 9 variables recorded for each of 4,766 observations (sites) which provided enough data that this wasn't a concern for this particular application. However, this would certainly be a concern when considering applying the FMCAR to other datasets. Unfortunately, there is no solution to this problem.

Turning to the limitations with the empirical elements of this research, the most obvious weakness is the usefulness of the data in demonstrating the full potential of the FMCAR. The innovation in the FMCAR is the ability to model asymmetric cross-correlations in spatial datasets. From the findings in chapter 7, it is apparent that the road accident data used in this thesis exhibit only mild asymmetric cross-correlations at best. Although it wasn't possible to know a priori that there wouldn't be asymmetric cross-correlations, it would have obviously been better to apply the FMCAR to a dataset where it was truly beneficial and where meaningful interpretation of the asymmetric cross-correlations could take place. Given the original aim to model road safety performance indicators there was little that could be done to address this problem without significant divergence from the stated research aims. Hopefully, it will be possible to address this limitation in the future.

There remains one other empirical weakness with this research relating to one of the key objectives as stated on page 9: to "produce a more robust ranking of London local authorities' road safety performance through the application of this model." Whilst it is true that the addition of spatial structure to the GLMM through the adoption of the FMCAR as a spatial prior did create a significant improvement in the precision of the local authority performance estimates, it was not possible to differentiate between individual local authority ranks. There was, however, enough narrowing of the credible intervals to identify groups of local authorities that were statistically above or below average. Thus, this aim was only partially achieved which can be considered a weakness. One of the conclusions that may be drawn from this result is that it is likely that *almost* all statistically robust methods of performance ranking will leave it impossible to compare individual organisational units.

On reflection, the limitations presented above largely reflect the inadequacy of the accident data (and the related problem of performance management) in demonstrating the properties and usefulness of the FMCAR. This leads neatly to areas of future work.

9.3.2 Future Directions

The most obvious future direction for this research is to find alternative applications using data that exhibit stronger asymmetric cross-correlations, demonstrating the full potential of this model. Areas where this model *may* be more appropriate are the natural environment and biostatistics. For instance, the complex relationship between environmental factors like rainfall and temperature. This requires more thought. Beyond applying the model to different data, the model could be extended to other domains e.g. panel data.

There are also other theoretical innovations of the FMCAR that could be investigated. The most exciting opportunity, but one which is still very much in its infancy, is the attempt to link MCAR models to dynamic Vector Autoregressions (VARs) and to offer new approaches to the specification and estimation of well-established econometric models. The approach would rest of revisiting what is meant by 'space' and instead viewing the macroeconomic panel data on a lattice similar to spatial models. There is much room for research as this is a completely uninvestigated topic and there is, so far, very little of substance.

Other, more obvious, extensions of the FMCAR model would be to incorporate it within a common factor model to develop a spatial common factor model. One use of this would be to represent in one (or more) latent variables road accident performance rather than in separate random effects for each indicator. This offers intuitive simplicity but may prove to obfuscate the mechanisms through which local authority activity affects accidents. Further, given the presence of some negative correlations between variables in the analysis conducted in chapter 7, this could be problematic.

From the performance management side, work has started on developing multivariate estimates of police performance with the Metropolitan Police Service. Similarly to the road casualty data, crimes have a geographic component that may make spatial methods useful. It is hoped that this research will plug into existing research streams that attempt to map crime. Currently these maps are statistically naive and report graphically uncontextualised per capita crime rates; they are therefore a visualisation rather than a statistical tool. The adoption of the FMCAR model could lead to the mapping of smoothed individual crime rates (which counts for the inherent variability in the crime counts) as well as maps of police performance.

9.3.3 Policy Implications

There are two broad policy implications of this research. Firstly, the results presented in chapter 7 indicate that there is spatial autocorrelation and spatial cross-correlation in the accident counts considered in this thesis. This has implications for the efficiency and accuracy of existing research using these data. More importantly, it raises the question as to whether there are one (or more) underlying spatially varying common factors that may explain the spatial distribution of these casualty variables. Obvious candidates for further research are environmental variables (e.g. varying patterns of weather), physical variables (e.g. the built environment), transport variables (e.g. the transport network) and socioeconomic factors (e.g. area deprivation). Whether or not all or any of these variables can be accurately measured and incorporated is questionable, which implies that adopting a spatial autoregressive model would be important as it would allow for spatially correlated errors. Given the importance of road safety, this finding has immediate policy relevance.

The second policy implication relates to the measurement of public sector performance. GLMMs are already promoted by statisticians as more reliable ways of estimating 'performance' than the use of crude, uncontextualised indicators. Recent innovations in the statistical literature have seen the use of multiple indicators simultaneously within a GLMM framework to improve the precision of the performance estimates by 'borrowing strength' across the multiple indicators. Extending this idea, this research provided extra levels of structure through the incorporation of spatial correlation parameters. There are a number of applications where the use of spatial information may also help to improve estimates of performance – for example in performance management for police services. However, what is also clear from this research is that even the most sophisticated models cannot reliably differentiate the relative performance of organisational units. This has implications for the promotion of policies that allocate resources based upon performance measurement.

9.4 Final remarks

This thesis considers the theory and application of multivariate conditional autoregressions. Given an understanding of time series autoregressions, it provides a self-contained introduction to the theory and methods of both univariate and multivariate conditional autoregressions. Beyond this it makes two substantial contributions to the literature. Firstly, it makes a significant methodological contribution by introducing a flexible multivariate conditional autoregression which allows asymmetric spatial cross-correlation to be modelled. A complete derivation, implementation and interpretation of the model is presented and through a simulation exercise the model is compared to the current MCAR models available to researchers. The second major contribution is empirical. The FMCAR model is applied to a selection of road safety performance indicators for London. By adopting the FMCAR as a spatial prior in the GLMM, considerable shrinkage of the estimates of local authority performance is produced. Whilst this does not enable local authorities to be differentiated based upon their road safety performance, it produces a considerable reduction in the uncertainty surrounding their rankings. It also provides further evidence to support the conjecture in Goldstein & Spiegelhalter (1996) that statistically robust methods of performance ranking make it impossible to separate observational units. Thus, whilst it is highly unlikely that the findings of this thesis will change policy relating to the use of performance management in the public sector, it rises to the call in Bird (2005) that researchers should work to improve the methods of performance measurement available to the public sector. Further, although the findings of this thesis indicate that there is only mild evidence of asymmetry in the spatial cross-correlations for road casualty counts, the thesis provides a demonstration of the applicability of this model to real world social and economic problems. Thus while the model may not be immediately applicable it remains insightful and advances the literature on multivariate spatial methods.

References

Akaike, H. (1974), 'A new look at the statistical model identification.', *IEEE Transactions on Automatic Control* **19**(6), 716723.

Anselin, L. (1988), *Spatial econometrics: methods and models*, Studies in Operational Regional Science, Dordrecht, The Netherlands: Kluwer Academic Publishers.

Anselin, L. (1992), 'Space and applied econometrics: Introduction', *Regional Science and Urban Economics* **22**(3), 307–316.

Anselin, L. (2003*a*), *A Companion to Theoretical Econometrics*, Oxford: Blackwell, chapter Spatial Econometrics, pp. 310–330.

Anselin, L. (2003*b*), 'Spatial externalities', *International Regional Science Review* **26**(2), 147–152.

Anselin, L. (2006), *Palgrave Handbook of Econometrics Volume 1: Econometric Theory*, Basingstoke: Palgrave Macmillan, chapter Spatial Econometrics, pp. 901–969.

Anselin, L. & Florax, R. J. G. M. (1995), *New directions in spatial econometrics*, Advances in Spatial Science, Berlin: Springer-Verlag.

Anselin, L., Florax, R. J. G. M. & Rey, S. J. (2004), *Advances in spatial econometrics: methodology, tools and applications*, Advances in Spatial Science, Berlin: Springer-Verlag.

Arbia, G. (2006), *Spatial Econometrics: Statistical Foundations and Applications to Regional Convergence*, Berlin: Springer-Verlag.

Arnold, B. C. & Straus, D. J. (1991), 'Bivariate distributions with conditionals in prescribed exponential families.', *Journal of the Royal Statistical Society Series B (Methodology)* **53**(2), 365–375.

Assunção, R. (2003), 'Space-varying coefficient models for small area data.', *Environmentrics* **14**(5), 453–473.

Assunção, R., Potter, J. & Cavenaghi, S. (2002), 'A bayesian space varying parameter model applied to estimating fertility schedules', *Statistics in Medicine* **21**(14), 2057–2075.

Aten, B. (1996), 'Evidence of spatial autocorrelation in international prices', *Review of Income and Wealth* **42**(2), 149–63.

Bailey, T. C. & Hewson, P. J. (2004), 'Simultaneous modelling of multiple traffic safety performance indicators by using a multivariate generalized linear mixed model', *Journal of the Royal Statistical Society Series A (Statistics in Society)* **167**(3), 501–517.

Bailey, T. & Gatrell, A. (1995), *Interactive spatial data analysis*, Longman Scientific & Technical.

Baltagi, B. H., Kelejian, H. H. & Prucha, I. R. (2007), 'Analysis of spatially dependent data', *Journal of Econometrics* **140**(1), 1–4.

Banerjee, S., Carlin, B. P. & Gelfand, A. E. (2004), *Hierarchical Modeling and Analysis for Spatial Data*, Boca Raton, US: Chapman & Hall.

Barnard, G. A. (1963), 'Some logical aspects of the fiducial argument', *The Journal of Royal Statistical Society Series B: Methodology* **25**(1), 111–4.

Berger, J. O. (1985), *Statistical Decision Theory and Bayesian Analysis*, Springer, Berlin, Germany.

Bernardo, J. M. (2003), *Encyclopedia of Life Support Systems (EOLLS)*, Vol. Probability and Statistics, Oxford: UNESCO, chapter Bayesian Statistics.

Bernardo, J. M. & Smith, A. F. M. (1994), *Bayesian Theory*, John Wiley & Sons, Chichester, UK.

Besag, J. (1974), 'Spatial interaction and the statistical analysis of lattice systems (with discussion)', *Journal of the Royal Statistical Society Series B (Methodology)* **35**(2), 192–236.

Besag, J. (2001), 'Markov chain monte carlo for statistical inference', *Center for Statistics and Social Sciences Working Paper, University of Washington* (9).

Besag, J. & Kooperberg, C. (1995), 'On conditional and intrinsic autoregressions', *Biometrika* **82**(4), 733–746.

Besag, J., York, J. & Mollie, A. (1991), 'Bayesian image restoration, with two applications in spatial statistics', *Annals of the Institute of Statistical Mathematics* **43**(1), 1–20.

Best, N. G., Cowles, M. K. & Vines, K. (1996), *CODA – Convergence Diagnosis and Output Analysis, Version 3.0 Manual*, MRC Biostatistics Unit, Cambridge.

Best, N. G., Cowles, M. K. & Vines, K. (1997), *CODA – Convergence Diagnosis and Output Analysis, Version 4.0 Manual*, MRC Biostatistics Unit, Cambridge.

Best, N., Richardson, S. & Thomson, A. (2005), 'A comparison of bayesian spatial models for disease mapping', *Statistical Methods in Medical Research* **14**(1), 35–59.

Bierens, H. J. (2004), *Introduction to the Mathematical and Statistical Foundations of Econometrics*, Cambridge: Cambridge University Press.

Bird, S. M. (2004), 'Editorial: Performance monitoring in the public services', *Journal of the Royal Statistical Society, Series A: Statistics in Society* **167**(3), 381–383.

Bird, S. M. (2005), 'Performance indicators: good, bad, and ugly', *Journal of the Royal Statistical Society, Series A: Statistics in Society* **168**(1), 1–27.

Bratti, M., McKnight, A., Naylor, R. & Smith, J. (2004), 'Higher education outcomes, graduate employment and university performance indicators', *Journal of the Royal Statistical Society, Series A: Statistics in Society* **167**(3), 475–496.

Breslow, N. E. & Clayton, D. G. (1993), 'Approximate inference in generalized linear mixed models', *Journal of the American Statistical Association* **88**(421), 9–25.

Brook, D. (1964), 'On the distinction between the conditional probability and the joint probability approaches in the specification of nearest-neighbour systems.', *Biometrika* **51**(3/4), 481–483.

Bull, J. P. & Roberts, B. J. (1973), 'Road accident statistics - a comparison of police and hospital information', *Accident Analysis and Prevention* **5**(5), 45–53.

Bunn, F., Collier, T., Frost, C., Ker, K., Roberts, I. & Wentz, R. (2003), 'Traffic calming for the prevention of road traffic injuries: systematic review and meta-analysis', *Injury Prevention* **9**(3), 200–204.

Burgess, S., Propper, C. & Wilson, D. (2002), 'Does performance monitoring work? a review of evidence from the uk public sector, excluding health care', *CMPO, University of Bristol Working Paper* **02**(49).

Cameron, C. & Trivedi, P. K. (1998), *Regression analysis of count data*, Econometric Society monographs ; 30, Cambridge University Press.

Carlin, B. & Banerjee, S. (2003), *Bayesian Statistics 7: Proceedings of the Seventh Valencia International Meeting on Bayesian Statistics*, Oxford: Oxford University Press, chapter Hierarchical multivariate CAR models for spatio-temporally correlated survival data (with discussion), pp. 45–63.

Carlin, B. P. & Louis, T. A., eds (2000), *Bayes and Empirical Bayes Methods for Data Analysis*, Boca Raton, US: Chapman & Hall, New York.

Case, A. (1991), 'Spatial patterns in household demand', *Econometrica* **59**(4), 953–965.

Case, A., Rosen, H. & Hines, J. R. (1993), 'Budget spillovers and fiscal policy interdependence', *Journal of Public Economics* **52**(3), 285–307.

Chan, N. H. & Cressie, N. (1989), 'Spatial modelling of regional variables', *Journal of the American Statistical Association* **84**(406), 393–401.

Chib, S. & Greenberg, E. (1995), 'Understanding the Metropolis-Hastings algorithm', *American Statistician* **49**(4), 327–335.

Clayton, D. & Kaldor, J. (1987), 'Empirical bayes estimates of age-standardised relative risks for use in disease mapping', *Biometrics* **43**(3), 671–81.

Cliff, A. D. & Ord, J. K. (1981), *Spatial processes: models & applications*, London: Pion.

Clifford, P. (1990), *Disorder in Physical Systems. A Volume in Honour of John M. Hammersley*, Oxford: Oxford University Press, chapter Markov random fields in statistics, pp. 19–32.

Coelli, T. (1996*a*), A guide to DEAP version 2.1: a data envelopment analysis (computer) program, Technical report, Centre for Efficiency and Productivity Analysis Working Paper 96/08, Department of Econometrics, University of New England.

Coelli, T. (1996*b*), A guide to Frontier version 4.1: A computer program for stochastic production and cost function estimation, Technical report, Centre for Efficiency and Productivity Analysis Working Paper 96/07, Department of Econometrics, University of New England.

Coelli, T., Rao, D. & Battese, G. (1998), *An introduction to efficiency and productivity analysis*, Dordrecht, The Netherlands: Kluwer Academic Publishers.

Commission for Global Road Safety (2005), Make roads safe: A new priority for sustainable development, Technical report.

Conradie, B., Cookson, G. & Thirtle, C. (2006), 'Efficiency and farm size in western cape grape production: Pooling small datasets', *South African Journal of Economics* **74**(2), 334 – 343.

Cressie, N. A. C. (1993), *Statistics for spatial data*, Wiley series in probability and mathematical statistics. Applied probability and statistics, rev. edn, New York: Wiley.

Cryer, P., Westrup, S., Cook, A. C., Ashwell, V., Bridger, P. & Clarke, C. (2001), 'Investigation of bias after data linkage of hospital admissions data to police road traffic crash reports', *Injury Prevention* **7**(3), 234–241.

Darmofal, D. (2006), 'Spatial econometrics and political science', Paper presented at the Southern Political Science Association. Retrieved from http://polmeth.wustl.edu/retrieve.php?id=575 on 5th January, 2007.

Davidson, R. & MacKinnon, J. G. (2004), *Econometric Theory and Methods*, Oxford: Oxford University Press.

Davis, R. (1992), *Death on the streets: cars and the mythology of road safety*, Leeds, UK: Leading Edge Press.

de Finetti, B. (1970), *Teoria delle Probabilit*, Einaudi, Turin. English translation as *Theory of Probability* in 1975, Chichester: Wiley.

Deaton, A. & Muellbauer, J. (1980), *Economics and Consumer Behaviour*, Cambridge: Cambridge University Press.

Debreu, G. (1951), 'The coefficient of resource utlization', *Econometrica* **19**(3), 273–292.

Dee, T. (1998), 'Reconsidering the effects of seat-belt laws and their enforcement status', *Accident Analysis and Prevention* **30**(1), 1–10.

Dempster, A. P. (1972), 'Covariance selection', *Biometrics* **28**(1), 157–175.

Department for Transport, Local Government, and the Regions (1999), Local Government Act 1999, Technical report, London: HMSO.

Department for Transport, Local Government and the Regions (2001), Strong local leadership quality public services (the local government white paper), Technical report, London: HMSO.

Department of Health (2008), *Healthy Weight, Healthy lives: A Cross-Government strategy for England*, London: HMSO.

Department of Health (2009), 'Website. accessed 1st july, 2009'.
URL: *http://www.dh.gov.uk/en/Publichealth/Healthimprovement/Obesity/index.htm*

DfT (2000), Tomorrow's roads - safer for everyone, Technical report, London: HMSO.

DFT (2005), *Great Britain Road Accident Statistics*, STATS19, London: HMSO.

Diggle, P. J. (1983), *Statistical analysis of spatial point patterns*, New York: Academic Press.

Diggle, P. J., Tawn, J. A. & Moyeed, R. A. (1998), 'Model-based geostatistics (with discussion)', *Journal of the Royal Statistical Society Series C (Applied Statistics)* **47**(3), 299–350.

Draper, D. & Gittoes, M. (2004), 'Statistical analysis of performance indicators in uk higher education', *Journal of the Royal Statistical Society Series A (Statistics in Society)* **167**(3), 449–474.

Durlauf, S. N. (1997), *The economy as an evolving complex system II.*, Reading, US: Addison-Wesley, chapter Statistical mechanics approaches to socioeconomic behaviour., pp. 81–104.

Edwards, P., Green, J., Roberts, I., Grundy, C. & Lachowycz, K. (2006), *Deprivation and Road Safety in London: A report to the London Road Safety Unit.*, London School of Hygiene and Tropical Medicine.

Elliott, P., Cuzick, J., English, D. & Stern, R., eds (1992), *Geographical and Environmental Epidemiology: Methods for Small-Area Studies*, Oxford: Oxford University Press, Oxford.

Elliott, P., Wakefield, J. C., Best, N. G. & Briggs, D. J. (2000), *Spatial epidemiology: methods and applications*, Oxford medical publications, Oxford: Oxford University Press.

European Transport Safety Council (2006), 'Road safety: we must do more', Press Release.

Evans, A. (1994), 'Transport safety: Editorial', *Journal of Transport Economics and Policy* **28**(Special Issue), 3–6.

Evans, M. & Swartz, T. (1995), 'Methods for approximating integrals in statistics with special emphasis on bayesian integration problems', *Statistical Science* **10**(3), 254–272.

Farrell, M. J. (1957), 'The measurement of productive efficiency (with discussion)', *Journal of the Royal Statistical Society, Series A: Statistics in Society* **120**(3), 253–290.

Florax, R. J. G. M. & De Graaff, T. (2004), *Advances in Spatial Econometrics: Methodology, Tools and Applications*, Berlin: Springer-Verlag, chapter The performance of diagnostic tests for spatial dependence in linear regression models: A meta-analysis of simulation studies.

Florax, R. J. G. M. & Nijkamp, P. (2004), *Encyclopedia of Social Measurement*, New York: Academic Press, chapter Misspecification in Linear Spatial Regression Models.

Florax, R. J. G. M. & van der Vlist, A. (2003), 'Spatial econometric data analysis: Moving beyond traditional models.', *International Regional Science Review* **26**(3), 223–243.

Fowkes, A., Nash, C. & Tweddle, G. (1990), 'The track and external costs of road transport', *ITS Working Paper 312, University of Leeds* .

Gamerman, D. (1997), *Markov chain Monte Carlo : stochastic simulation for Bayesian inference*, Boca Raton, US: Chapman & Hall.

Gamerman, D. & Moreira, A. R. B. (2004), 'Multivariate spatial regression models', *Journal of Multivariate Analysis* **91**(22), 262–281.

Gamerman, D., Moreira, A. R. B. & Rue, H. (2002), 'Space-varying regression models: Specifications and simulation', *Computational Statistics and Data Analysis* **42**(3), 513–533.

Geary, C. (1954), 'The contiguity ratio and statistical mapping', *The Incorporated Statistician (renamed as Journal of the Royal Statistical Society Series D)* **5**(3), 115–145.

Gelfand, A. E., Sahu, S. K. & Carlin, B. P. (1995), 'Efficient parameterizations for normal linear mixed models.', *Biometrika* **82**(3), 479–488.

Gelfand, A. E., Schmidt, A. M., Banerjee, S. & Sirmans, C. F. (2004), 'Nonstationary multivariate process modeling through spatially vary coregionalization', *Sociedad de Estadística e Investigación Opertiva Test* **13**(2), 263–312.

Gelfand, A. E. & Smith, A. F. M. (1990), 'Sampling-based approaches to calculating marginal densities', *Journal of the American Statistical Association* **85**(410), 398–409.

Gelfand, A. E. & Vounatsou, P. (2003), 'Proper multivariate conditional autoregressive models for spatial data analysis', *Biostatistics* **4**(1), 11–25.

Gelman, A. (1996), *Markov Chain Monte Carlo in Practice*, London: Chapman & Hall, London: UK, chapter Inference and monitoring convergence, pp. 131–144.

Gelman, A., Carlin, J. S., Stern, H. S. & Rubin, D. B. (2003), *Bayesian Data Analysis*, Texts in statistical science, 2nd edn, Chapman & Hall.

Gelman, A. & Hill, J. (2006), *Data Analysis Using Regression and Hierarchical/Multilevel Models*, Cambridge: Cambridge University Press.

Geman, S. & Geman, D. (1984), 'Stochastic relaxation, gibbs distributions, and the bayesian restoration of images', *IEEE transactions on pattern analysis and machine intelligence* **6**(6), 721–741.

Gilks, W. R., Richardson, S. & Spiegelhalter, D. J., eds (1996), *Markov chain Monte Carlo in practice*, Boca Raton, US: Chapman & Hall.

Goldstein, H. & Myers, K. (1996), 'Freedom of information: towards a code of ethics for performance indicators', *Research Intelligence* **57**, 12–16.

Goldstein, H. & Spiegelhalter, D. (1996), 'League tables and their limitations: Statistical issues in comparisons of institutional performance', *Journal of the Royal Statistical Society, Series A: Statistics in Society* **159**(3), 385–443.

Graham, D. & Glaister, S. (2005), 'Decomposing the effects of deprivation on child pedestrian casualties', *Working Paper, Centre for Transport Studies, Imperial College London* .

Graham, D., Glaister, S. & Anderson, R. (2005), 'The effects of area deprivation on the incidence of child and adult pedestrian casualties in england', *Accident Analysis and Prevention* **37**(1), 125–135.

Graham, D. J. & Glaister, S. (2003), 'Spatial variation in road pedestrian casualties: the role of urban scale, density and land-use mix', *Urban Studies* **40**(8), 1591–1607.

Grayling, T., Hallam, K., Graham, D., Anderson, R. & Glaister, S. (2002), Streets ahead: safe and liveable streets for children, Technical report, Institute of Public Policy Research, London, UK.

Green, P. J. (2001), *Complex Stochastic Systems*, London: Chapman & Hall, chapter A Primer on Markov Chain Monte Carlo.

Green, P. & Richardson, S. (2002), 'Hidden markov models and disease mapping.', *Journal of the American Statistical Association* **97**(460), 1055–70.

Greene, W. H. (2003), *Econometric analysis (5th Edition)*, New York: Prentice-Hall International.

Haining, R. (1991), *Spatial data analysis in the social and environmental sciences*, Cambridge: Cambridge University Press.

Hansson, L. & Marckham, J. (1992), Internalisation of external effects in transportation, *in* 'Paris UIC'.

Harvey, A. C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge: Cambridge University Press.

Harville, D. D. (1997), *Matrix Algebra from a Statistician's Perspective*, Berlin: Springer-Verlag.

Hastings, W. K. (1970), 'Monte carlo sampling methods using markov chains and their applications', *Biometrika* **57**(1), 97–109.

Hauck, K. & Street, A. (2006), 'Performance assessment in the context of multiple objectives: A multivariate multilevel analysis', *Journal of Health Economics* **25**(6), 1029–1048.

Hewson, P. J. (2002), 'Child pedestrian accidents in the uk: Monitoring the uk government's road safety strategy', *Radical Statistics* **79/80**(Summer/Winter 2002).

Hewson, P. J. (2004), 'Deprived children or deprived neighbourhoods? a public health approach to the investigation of links between deprivation and injury risk with specific reference to child road safety in devon county, UK', *BMC Public Health* **4**(15). [Online].

Hewson, P. J. (2005), 'Epidemiology of child pedestrian casualty rates: Can we assume spatial independence?', *Accident Analysis and Prevention* **37**(4), 651–659.

Hicks, J. R. (1939), *Value and Capital*, Oxford: Oxford University Press.

Hillman, M., Adams, J. & Whitelegg, J. (1990), *One false move : a study of children's independent mobility*, London: Policy Study Institute.

H.M. Treasury (2007), Comprehensive spending review 2007, Technical report, HMSOLondon: HMSO.

H.M. Treasury (2008), Public finances databank, Technical report, London: HMSO.

Holloway, G. J. (2007), 'Special feature: Spatial issues in agricultural economics', *Journal of Agricultural Economics* **58**(3), 484–486.

Holloway, G. J. & Lapar, M. L. (2007), 'How big is your neighbourhood? spatial implications of market participation among filipino smallholders.', *Journal of Agricultural Economics* **58**(3), 37–60.

Holloway, G. J. & Tomberlin, D. (2006), 'Bayesian rank and selection of fishing boat efficiencies', *Marine Resource Economics* **21**(4), 415–432.

House of Commons Public Administration Select Committee (2003), On target?: Government by measurement, Technical report, London: HMSO.

Ioannides, Y. M. (1997), *The economy as an evolving complex system II.*, New York: Addison-Wesley, chapter Evolution of Trading Strategies, pp. 129–167.

Isard, W. (1956), *Location and Space-economy; a General Theory Relating to Industrial Location, Market Areas, Land Use, Trade, and Urban Structure.*, Cambridge, US: Technology Press of Massachusetts Institute of Technology and Wiley.

James, H. F. (1991), 'Under-reporting of road traffic accidents', *Traffic Engineering and Control* **32**(12), 574–583.

Jin, X., Carlin, B. & Banerjee, S. (2005), 'Generalized hierarchical multivariate car models for areal data', *Biometrics* **61**(4), 950–961.

Kaiser, M. S. & Cressie, N. (2000), 'The construction of multivariate distributions from markov random fields', *Journal of Multivariate Analysis* **73**(2), 199–220.

Kelejian, H. H. & Prucha, I. R. (1998), 'A generalized two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances', *Journal of Real Estate Finance and Economics* **17**(1), 99–121.

Kelejian, H. H. & Prucha, I. R. (2004), 'Estimation of simultaneous systems of spatially interrelated cross sectional equations', *Journal of Econometrics* **118**, 27–50.

Khatri, Y., Leruth, L. & Piesse, J. (2002), Corporate performance and governance in Malaysia, Technical report, IMF Working Paper.

Kim, C. W., Phipps, T. T. & Anselin, L. (2003), 'Measuring the benefits of air quality improvement: a spatial hedonic approach', *Journal of Environmental Economics and Management* **45**(1), 24–39.

Knorr-Held, L. & Raßer, G. (2000), 'Bayesian detection of clusters and discontinuities in disease maps.', *Biometrics* **56**(1), 13–21.

Knorr-Held, L. & Rue, H. (2002), 'On block updating in markov random field models for disease mapping', *Scandanavian Journal of Statistics* **29**(4), 597–614.

Koop, G. (2003), *Bayesian econometrics*, New York: Wiley.

Koop, G., Osiewalski, J. & Steel, M. F. J. (1997), 'Bayesian efficiency analysis through individual effects: Hospital cost frontiers', *Journal of Econometrics* **76**(1-2), 77–105.

Koopermans, T. C. (1951), *Activity Analysis of Production and Allocation*, New York: Wiley.

Kuhan, G., Marshall, E. C., Abidia, A., Chetter, I. C. & McCollum, P. T. (2002), 'A bayesian hierarchical approach to comparative audit for cartoid surgery', *European Journal of Vascular and Endovascular Surgery* **24**(6), 505–510.

Kumbhaker, S. & Lovell, C. (2003), *Stochastic Frontier Analysis*, Cambridge: Cambridge University Press.

Künsch, H. (1981), 'Thermodynamics and statistical analysis of gaussian random fields', *Probability Theory and Related Fields* **58**(3), 407–21.

Laird, N. M. & Louis, T. A. (1989), 'Empirical bayes ranking methods', *Journal of Educational Statistics* **14**, 29–46.

Lancaster, T. (2004), *An Introduction to Modern Bayesian Econometrics*, Oxford: Blackwell.

Langford, I. H., Leyland, A. H., Rasbash, J. & Goldstein, H. (1999), 'Multilevel model of the geographical distributions of diseases', *Journal of the Royal Statistical Society Series C (Applied Statistics),* **48**(2), 253–268.

Lauritzen, S. L. (1996), *Graphical Models*, Oxford: Clarenden Press.

Lawson, A. an Biggeri, A., Böhnring, D., Lesaffre, E., Clark, A., Schlattmann, P. & Divino, F. (2000), 'Disease mapping models: An empirical evaluation', *Statistics in Medicine* **19**(17-18), 2217–41.

Lee, L. (2007), 'Identification and estimation of econometric model with group interactions, contextual factors and fixed effects', *Journal of Econometrics* **140**(2), 333–374.

LeSage, J. P. (1989), 'Incorporating regional wage relations in local forecasting models with a bayesian prior', *International Journal of Forecasting* **5**(1), 37–47.

LeSage, J. P. (1999), Applied econometrics using matlab.
URL: *http://www.spatial-econometrics.com/*

LeSage, J. P., Pace, R. K. & Tiefelsdorf, M. (2004), 'Methodological developments in spatial econometrics and statistics', *Geographical Analysis* **36**(2), 87–89.

LeSage, J. P. & Reed, J. D. (1989), 'The dynamic relationship between export, local, and total area employment', *Regional Science and Urban Economics* **19**(4), 615–636.

Leyland, A. H., Langford, I. H., Rasbash, J. & Goldstein, H. (2000), 'Multivariate spatial models for event data', *Statistics in Medicine* **19**(17-18), 2469–2478.

Lilford, R. J. & Braunholtz, D. (1996), 'The statistical basis of public policy: a paradigm shift is overdue', *British Medical Journal* **313**(7057), 603–7.

Lin, X. (2005), Peer effects and student academic achievement: An application of spatial autoregressive model with group unobservables. Job Market Paper, Department of Economics, The Ohio State University.

Lockwood, J., Louis, T. A. & McCaffrey, D. F. (2002), 'Uncertainty in rank estimation: Implications for value-added modeling accountability systems', *Journal of Educational and Behvioral Statistics* **27**(3), 255–270.

MacNab, Y. (2003), 'A bayesian hierarchical model for accident and injury surveillance', *Accident Analysis and Prevention* **35**(1), 91–102.

MacNab, Y. C. (2004), 'Bayesian spatial and ecological models for small-area accident and injury analysis', *Accident Analysis and Prevention* **36**(6), 1019–1028.

Maddison, D., Pearce, D., Johansson, O., Calthrop, E., Litman, T. & Verhoef, E. (1996), *Blueprint 5: The True Costs of Road Transport*, Earthscan, London, UK.

Manski, C. F. (1993), 'Identification of endogenous social effects: the reflection problem', *Review of Economic Studies* **60**(3), 531–542.

Mantel, N. & Stark, C. (1968), 'Computation of indirect adjustment rates in the presence of confounding', *Biometrics* **24**(4), 997–1005.

Mardia, K. V. (1988), 'Multi-dimensional multivariate gaussian markov random fields with application to image processing', *Journal of Multivariate Analysis* **24**(2), 265–284.

Mardia, K. V., Kent, J. T. & Bibby, J. (1979), *Multivariate Analysis*, New York: Academic Press.

Marshall, A. (1920), *Principles of Economics: Revised Edition*, Macmillan: reprinted by Prometheus Books.

Marshall, E. C. & Spiegelhalter, D. (1998), 'Reliability of league tables of in vitro fertilisation clinics: retrospective analysis of live birth rates', *British Medical Journal* **316**(7146), 1701–1705.

Martellosio, F. (2006), 'The correlation structure of spatial autoregressions on graphs', *Munich Personal RePEc Archive* **1695**.

Martin, S. & Smith, P. (2005), 'Multiple public service performance indicators: Toward an integrated statistical approach', *Journal of Public Administration Research and Theory* **15**(4), 599–613.

McCarthy, P. S. (1999), 'Public policy and highway safety: a city-wide perspective', *Regional Science and Urban Economics* **29**(2), 231–244.

McCullagh, P. & Nelder, J. (1989), *Generalized Linear Modelling*, London: Chapman & Hall.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953), 'Equations of state calculations for fast computing machines', *Journal of Chemical Physics* **21**(6), 1087–1091.

Moran, P. A. P. (1950), 'Notes on continuous stochastic phenomena', *Biometrika* **37**(1-2), 17–23.

Morris, C. & Christiansen, C. (1996), *Bayesian Statistics 5: Proceedings of the Fifth Valencia International Meeting, June 5-9, 1994*, Oxford: Oxford University Press, chapter Hierarchical models for ranking and for identifying extremes, with applications, pp. 277–296.

Nelder, J. A. & Wedderburn, R. W. M. (1972), 'Generalized linear models', *Journal of the Royal Statistical Society Series A (Statistics in Society)* **135**(3), 370–384.

Nelson, G. C. (2002), 'Introduction to the special issue on spatial analysis', *Agricultural Economics* **27**(3), 197–200.

Nelson, G. C. & Gray, D. A. (1997), 'The use of spatial analysis at the world bank', *The World Bank Environmental Economics Series* **56**(Feb).

Nelson, G. C. & Hellerstein, D. (1997), 'Do roads cause deforestation? using satellite images in econometric analysis of land use.', *American Journal of Agricultural Economics* **79**, 80–88.

Newbery, D. (1988), 'Road user charges in Britain', *Economic Journal* **Conference Issue**, 161–176.

Noland, R. B. (2001), Traffic fatalities and injuries: are reductions the result of improvements in highway design standards?, *in* '80th Annual Meeting of the Transport Research Board'.

Noland, R. & Quddus, M. (2002), Improvements in medical care and technology and reductions in traffic related fatalities in Great Britain, *in* 'European Regional Science Association Conference Papers', Dortmund.

Noland, R. & Quddus, M. (2004), 'A spatially disaggregate analysis of road casualties in england', *Accident Analysis and Prevention* **36**, 973–984.

O' Hagan, A. (1994), *Bayesian Inference*, Vol. 2 of *Kendall's Advanced Theory of Statistics*, London: Edward Arnold.

Osbourne, D. & Gaebler, T. (1992), *Re-inventing Government*, New York: Addison-Wesley.

Pace, R. K., Barry, R. & Sirmans, C. (1998), 'Spatial statistics and real estate', *Journal of Real Estate Finance and Economics* **17**(1), 5–13.

Pace, R. K. & LeSage, J. P. (2003), Conditional autoregressions with doubly stochastic weight matrices. http://www.spatial-statistics.com/spatial statistical manuscripts/doubly stochastic/ doublystochastic1.pdf.

Pace, R. K. & LeSage, J. P. (2004), 'Spatial statistics and real estate', *Journal of Real Estate Finance and Economics* **29**(2), 147–148.

PACTS (2003*a*), Best value, local transport plans and road safety: Listening to and learning from the profession, Technical report, London: Parliamentary Advisory Council for Transport Safety.

PACTS (2003*b*), 'Speed cameras: 10 criticisms and why they are flawed', London: Parliamentary Advisory Council for Transport Safety and the Slower Speeds Initiative.

Paelinck, J. H. P. & Klaassen, L. H. (1979), *Spatial Econometrics*, Farnborough, UK: Saxon House.

Papageorgiou, H. & Loukas, S. (1988), 'On estimating the parameters of a bivariate model applicable to traffic accidents', *Biometrics* **44**(2), 495–504.

Parent, O. & LeSage, J. P. (2008), 'Using the variance structure of the conditional autoregressive spatial specification to model knowledge spillovers', *Journal of Applied Econometrics* **23**(2), 235–256.

Pearce, D. (1993), *Blueprint 3: Measuring Sustainable Development*, Earthscan, London, UK.

Peirson, J., Skinner, I. & Vickerman, R. (1998), 'The microeconomic analysis of the external costs of road accidents', *Economica* **65**(259), 429–440.

Pinkse, J., Shen, L. & Slade, M. (2007), 'A central limit theorem for endogenous locations and complex spatial interactions', *Journal of Econometrics* **140**(1), 215–225.

Plowden, S., Hillman, M. & Policy Studies, I. (1984), *Danger on the road: the needless scourge. A study of obstacles to progress in road safety*, London: Policy Studies Institute.

Poirier, D. J. & Tobias, J. L. (2006), *Palgrave Handbook of Econometrics Volume 1: Econometric Theory*, Basingstoke, UK: Palgrave Macmillan, chapter Bayesian Econometrics, pp. 841–870.

Propper, C. & Wilson, D. (2003), 'The use and usefulness of performance measures in the public sector', *Oxford Review of Economic Policy* **19**(2), 250–267.

Rao, C. R. & Rao, M. B. (1998), *Matrix Algebra and Its Applications to Statistics and Econometrics*, World Scientific Publishing Co.

Richardson, S., Thomson, A., Best, N. G. & Elliott, P. (2004), 'Interpreting posterior relative risks estimates in disease mapping studies', *Environmental Health Perspectives* **112**(9), 1016–25.

Ripley, B. D. (1981), *Spatial Statistics*, London: Wiley.

Robert, C. P. & Casella, G. (2004), *Monte Carlo Statistical Methods*, 2nd edn, Berlin: Springer-Verlag.

Roberts, I. (1993), 'Why have child pedestrian death rates fallen?', *British Medical Journal* **307**(6896), 116–116.

Royle, J. A. & Berliner, L. M. (1999), 'A hierarchical approach to multivariate spatial modeling and prediction', *Journal of Agricultural, Biological and Environmental Statistics* **4**(1), 29–56.

Rozanov, Y. A. (1982), *Markov Random Fields*, Berlin: Springer-Verlag.

RSAP (2000), Improving child road safety - improving options for next steps, Technical report, London: Department for Transport, Local Government and the Regions, Road Safety Advisory Panel Papers.

Rue, H. (2001), 'Fast sampling of Gaussian Markov random fields', *Journal of the Royal Statistical Society, Series B: Methodology* **63**(2), 325–338.

Rue, H. & Held, L. (2005), *Gaussian Markov Random Fields: Theory and Applications*, Chapman & Hall.

Smith, A. F. M. (1991), 'Bayesian computational methods', *Philosophical Transactions of the Royal Society A: Physical, Mathematical and Engineering Sciences* **337**(1647), 369–386.

Smith, P. (1990), 'The use of performance indicators in the public sector', *Journal of the Royal Statistical Society Series A (Statistics in Society)* **153**(1), 53–72.

Smith, P. C. & Street, A. (2005), 'Measuring the efficiency of public services: the limits of analysis', *Journal of the Royal Statistical Societ, Series A: Statistics in Society* **168**(2), 401–417.

Spiegelhalter, D., Best, N. G., Carlin, B. & van der Linde, A. (2002), 'Bayesian measures of model complexity and fit (with discussion)', *Journal of the Royal Statistical Society Series B: Statistical Methodology* **64**(4).

Spiegelhalter, D., Thomas, A., Best, N. G. & Gilks, W. R. (1996), *BUGS: Bayesian inference using Gibbs sampling*, Cambridge: Medical Research Council.

Stevens, P. A. (2005), 'Assessing the performance of local government', *National Institute of Economic Review* **193**, 90–101.

Stiglitz, J. (1987), *The New Palgrave - A Dictionary of Economics: Volume 3*, Basingstoke, UK: Palgrave-MacMillan, chapter Principal and Agent, pp. 966–71.

Stone, M. (2002*a*), 'Can public service efficiency measurement be a useful tool of government? the lesson of the spottiswoode report', *Public Money and Management* **July-September**, 33–39.

Stone, M. (2002*b*), 'How not to measure the efficiency of the public services (and how one might)', *Journal of the Royal Statistical Society, Series A: Statistics in Society* **165**(3), 405–434.

Sun, D., Tsutakawa, R. K., Kim, H. & He, Z. (2000), 'Spatio-temporal interaction with disease mapping', *Statistics in Medicine* **19**(15), 2015–2035.

Tanner, M. A. (1996), *Tools for Statistical Inference*, New York: Springer-Verlag.

TfL (2004), Differences between london motorcyclists and those from the rest of the uk, Technical report, London Road Safety Unit: Research Summary No 4, London: UK.

TfL (2006*a*), *Best Value Performance Indicators Report 2006*, London: Transport for London.

TfL (2006*b*), 'Teen street safety campaign press release'.

TfL (2008), *London Travel Report 2007*, London: Transport for London.

Thanassoulis, E. (2001), *Introduction to the theory and application of data envelopment analysis*, Dortmondt, The Netherlands: Kluwer Academic Publishers.

Thirtle, C., Beyers, L., Ismael, Y. & Piesse, J. (2003), 'Can GM-technologies help the poor? The Impact of Bt cotton in Makhathini Flats, KwaZulu-Nata', *World Development* **31**(4), 717–732.

Tobler, W. R. (1970), 'A computer movie simulating urban growth in the Detroit region', *Economic Geography* **46**(Supplement: Proceedings. International Geographical Union), 234–240.

Topa, G. (1996), Social interactions, local spillovers and unemployment., PhD thesis, Department of Economics, University of Chicago.

Towner, E., Dowswell, T. & Jarvis, S. (1993), Reducing childhood accidents. the effectiveness of health promotion interventions: a literature review., Technical report, Health Education Authority, London, UK.

ver Hoef, J. M. & Cressie, N. (1994), 'Multivariable spatial prediction', *Mathematical Geology* **26**(2), 219–240.

ver Hoef, J. M., Cressie, N. & Pace, R. K. (2004), 'Flexible spatial models for kriging and cokriging using moving averages and the Fast Fourier Transform (FFT)', *Journal of Computational and Graphical Statistics* **13**(2), 265–282.

Wackernagel, H. (1998), *Multivariate Geostatistics*, Springer.

Wall, M. M. (2004), 'A close look at the spatial structure implied by the CAR and SAR models', *Journal of Statistical Planning and Inference* **121**(2), 311–324.

West, M. & Harrison, P. J. (1997), *Bayesian Forecasting and Dynamic Models*, New York: Springer.

Whittaker, J. (1990), *Graphical Models in Applied Multivariate Statistics*, New York: Wiley.

Whittle, P. (1954), 'On stationary processes in the plane', *Biometrika* **41**(3-4), 434–449.

WHO (2004), World report on road traffic injury prevention, Technical report, World Health Organization, Geneva, Switzerland.

Wilkinson, D. (2008), 'Implementation of stochastic simulation algorithms using c and python', Online.
URL: *http://www.mas.ncl.ac.uk/ ndjw1/teaching/sim/*

Woodward, A. (1983), 'Time trends in motorcycle accidents in britain', *Journal of Epidemiology and Community Health* **37**(1), 66–69.

APPENDIX A

COMPUTATIONAL APPENDIX

A.1 Introduction to Python

Python is a clear and powerful object-oriented programming language, comparable to Perl, Ruby, Scheme, or Java. Some of Python's notable features:

- Uses an elegant syntax, making the programmes easier to read.
- Is an easy-to-use language that makes it simple to get programmes working.
- Comes with a large standard library that supports many common programming tasks such as connecting to web servers, searching text with regular expressions, reading and modifying files.
- Python's interactive mode makes it easy to test short snippets of code.
- Is easily extended by adding new modules implemented in a compiled language such as C or C++.
- Can also be embedded into an application to provide a programmable interface.
- Runs on many different computers and operating systems: Windows, MacOS, many brands of Unix, OS/2, etc
- Is free software in two senses. It doesn't cost anything to download or use Python, or to include it in your application. Python can also be freely modified

and re-distributed, because while the language is copyrighted it's available under an open source license.

The ActivePython Python distribution is a distribution focusing on an easy install and use. It provides versions for a number of platforms: Linux, Windows, Mac OS X, Solaris, HP-UX and AIX. On Windows, ActivePython also includes Win32All (a.k.a. pywin32). You can download this distribution from:
http://www.activestate.com/Products/ActivePython/

There are numerous free manuals and tutorials available online for Python. A good starting point is the official Python programming language website: http://www.python.org/

Python is incredibly easy to learn and deploy. Compared to other languages it is very succinct and runs across all systems. Additionally, it works very well with the R statistical environment as well as with leading Geographical Information Systems (GIS) such as the commercial ArcGIS suite (from ESRI) or the open source GRASS implementation.

A.2 Introduction to R

R is a free software environment for statistical computing and graphics. Among other things it has:

- an effective data handling and storage facility,
- a suite of operators for calculations on arrays, in particular matrices,
- a large, coherent, integrated collection of intermediate tools for data analysis,
- graphical facilities for data analysis and display either directly at the computer or on hardcopy, and

- a well developed, simple and effective programming language (called S) which includes conditionals, loops, user defined recursive functions and input and output facilities.

The term "environment" is used by the designers to characterize it as a fully planned and coherent system, rather than an incremental accretion of very specific and inflexible tools, as is frequently the case with other data analysis software. R is very much a vehicle for newly developing methods of interactive data analysis. It has developed rapidly, and has been extended by a large collection of packages. However, most programs written in R are essentially ephemeral, written for a single piece of data analysis.

R is freely available for download from the Internet: http://www.r-project.org/ A number of free manuals and tutorials are also available for download from the same website.

The sampler for the FMCAR is made up of a Metropolis-Hastings step within a Gibbs Sampler and the code is quite modular and reuseable. The purpose of this section is to allow people familiar with MCMC methods and basic programming to be able to understand simple Python MCMC samplers before the Python code for the FMCAR is presented.

A.3 A Python Gibbs Sampler

Consider the simplest possible Gibbs sampler for a bivariate normal distribution.

```
from sys import argv

from math import *

from whrandom import random
```

```
def genexp(lamb):
  return (-1.0/lamb)*log(random())


def gennor(mu,sigma):
  theta=random()*2*pi
  rsq=genexp(0.5)
  z=sqrt(rsq)*cos(theta)
  return mu+z*sigma


n=eval(argv[1])
rho=eval(argv[2])
x=0
y=0
sig=sqrt(1-rho*rho)
for i in range(n):
  x=gennor(rho*y,sig)
  y=gennor(rho*x,sig)
  print x,y
```

A.4 A Python Metropolis-Hastings Sampler

Consider a very simple independence sampler for a Gamma distribution which uses a normal distribution as a proposal distribution.

```
from sys import argv
from math import *
from whrandom import random
```

```
def genexp(lamb):
  return (-1.0/lamb)*log(random())


def gennor(mu,sigma):
  theta=random()*2*pi
  rsq=genexp(0.5)
  z=sqrt(rsq)*cos(theta)
  return mu+z*sigma


def sdnorm(x,mu,sigma):
  return exp(-0.5*pow((x-mu)/sigma,2))


def sdgamma(x,a,b):
  if (x>0):
      return pow(x,a-1)*exp(-b*x)
  else:
      return 0


n=eval(argv[1])
a=eval(argv[2])
b=eval(argv[3])
x=(a+0.0)/b
mu=(a+0.0)/b
sig=sqrt((a+0.0)/(b*b))
```

```
for i in range(n):
  can=gennor(mu,sig)
  aprob=min(1,(sdgamma(can,a,b)/sdgamma(x,a,b))       \
    /(sdnorm(can,mu,sig)/sdnorm(x,mu,sig)))
  u=random()
  if (u<aprob):
    x=can
  print x
```

A.5 A Python Implementation of the FMCAR

includefmcarcode.tex

APPENDIX B

MATHEMATICAL APPENDIX

B.1 Specification through full conditionals

Chapter 2 introduced the Conditional Autogression of Besag (1974). Besag pioneered the specification of Gaussian Markov random field models via their full conditionals rather than the mean and precision. Here we provide a technical presentation of this approach and the principal theorems required by it. Following the notation adopted in chapter 2, suppose we specify the full conditional distributions as normals with

$$\mathrm{E}[y_i \mid \boldsymbol{y}_{-i}] = \mu_i + \sum_{j \in \partial_I} b_{ij}(y_j - \mu_j) \tag{B.1.1}$$

and

$$\mathrm{Var}[y_i \mid \boldsymbol{y}_{-i}] = \tau_i^{-1} \tag{B.1.2}$$

for $i = 1, ..., n$ and $\tau > 0$ for some neighbourhood matrix $\boldsymbol{B}$ with elements $b_{ij} = 1 \Longleftrightarrow j \in \partial_i$.

Theorem B.1.1. *Given the n normal full conditional distributions with conditional mean and variance as in (B.1.1) and (B.1.2), the $\boldsymbol{y}$ is a Gaussian Markov random field with respect to a labelled graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with mean $\boldsymbol{\mu}$ and precision matrix*

$\mathbf{\Sigma}^{-1} = \Sigma_{ij}^{-1}$, *where*

$$\Sigma_{ij}^{-1} = \begin{cases} \tau_i b_{ij} & i \neq j \\ \tau_i & i = j \end{cases}$$

provided $\tau_i b_{ij} = \tau_j b_{ji}$, $i \neq j$, *and* $\mathbf{\Sigma}^{-1} > 0$.

To prove this result we need Brook's lemma, which is due to Brook (1964) and demonstrates how the joint and full conditional distributions are related. We discussed the usefulness of this lemma in Section 2.4.1 and it has been used throughout the thesis.

Lemma B.1.2 (Brook's Lemma)**.** *Let* $\pi(\boldsymbol{y})$ *be the density for* $\boldsymbol{y} \in \mathbb{R}^n$ *and define* $\Omega = \{\boldsymbol{y} \in \mathbb{R}^n : \pi(\boldsymbol{y}) > 0\}$. *Let* $\boldsymbol{y}, \boldsymbol{y}' \in \Omega$, *then*

$$\frac{\pi(\boldsymbol{y})}{\pi(\boldsymbol{y}')} = \prod_{i=1}^{n} \frac{\pi(y_i \mid y_1, ..., y_{i-1}, y'_{i+1}, ..., y'_n)}{\pi(y'_i \mid y_1, ..., y_{i-1}, y'_{i+1}, ..., y'_n)} \tag{B.1.3}$$

$$= \prod_{i=1}^{n} \frac{\pi(y_i \mid y'_1, ..., y'_{i-1}, y_{i+1}, ..., y_n)}{\pi(y'_i \mid y'_1, ..., y'_{i-1}, y_{i+1}, ..., y_n)} \tag{B.1.4}$$

If we fix $\boldsymbol{y}'$ then (B.1.3) (and (B.1.4)) represents $\pi(\boldsymbol{y})$ up to a constant of proportionality, employing the n full conditional distributions $\pi(y_i \mid \boldsymbol{y}_{-i})$. As $\pi(\boldsymbol{y})$ integrates to 1 we can find the constant.

Proof (Brook's lemma). Start with the identity

$$\frac{\pi(y_n \mid y_1, ..., y_{n-1})}{\pi(y'_n \mid y_1, ..., y_{n-1})} = \frac{\pi(y_1, ..., y_{n-1}, y_n)}{\pi(y_1, ..., y_{n-1}, y'_n)}$$

from which it follows that

$$\pi(y_1, ..., y_n) = \frac{\pi(y_n \mid y_1, ..., y_{n-1})}{\pi(y'_n \mid y_1, ..., y_{n-1})} \pi(y_1, ..., y_{n-1}, y'_n)$$

Express the last term on the right hand side similarly to obtain

$$\begin{aligned}\pi(y_1, ..., y_n) &= \frac{\pi(y_n \mid y_1, ..., y_{n-1})}{\pi(y'_n \mid y_1, ..., y_{n-1})} \\ &\text{x} \quad \frac{\pi(y_{n-1} \mid y_1, ..., y_{n-2}, y'_n)}{\pi(y'_n \mid y_1, ..., y_{n-2}, y'_n)} \\ &\text{x} \quad \pi(y_1, ..., y_{n-2}, y'_{n-1}, y'_n)\end{aligned}$$

Repeat this process until Expression (B.1.3) results. To prove the alternative (B.1.4) start with

$$\frac{\pi(y_1 \mid y_2, ..., y_n)}{\pi(y'_1 \mid y_2, ..., y_n)} = \frac{\pi(y_1, ..., y_{n-1}, y_n)}{\pi(y'_1, y_2, ..., y_n)}$$

and proceed forward as we did with (B.1.3). □

Proof (Theorem B.1.1). Assume $\boldsymbol{\mu} = \mathbf{0}$ and fix $\boldsymbol{y}' = \mathbf{0}$. Then (B.1.3) simplifies to

$$\log \frac{\pi(\boldsymbol{y})}{\pi(\mathbf{0})} = -\frac{1}{2} \sum_{1=1}^{n} \tau_i y_1^2 - \sum_{1=2}^{n} \sum_{j=1}^{i-1} \tau_i b_{ij} y_i y_j. \tag{B.1.5}$$

Using (B.1.4) this becomes

$$\log \frac{\pi(\boldsymbol{y})}{\pi(\mathbf{0})} = -\frac{1}{2} \sum_{1=2}^{n-1} \tau_i y_1^2 - \sum_{1=2}^{n} \sum_{j=i+1}^{n} \tau_i b_{ij} y_i y_j. \tag{B.1.6}$$

As (B.1.5) and (B.1.6) are clearly identical it follows that $\tau_i b_{ij} = \tau_j b_{ji}$, $i \neq j$. The joint distribution of $\boldsymbol{y}$ can therefore be expressed as

$$\log \pi(\boldsymbol{y}) = \text{const} - \frac{1}{2} \sum_{1=2}^{n-1} \tau_i y_1^2 - \frac{1}{2} \sum_{i \neq j} \tau_i b_{ij} y_i y_j.$$

hence $\boldsymbol{y}$ is zero mean multivariate normal distribution provided that $\boldsymbol{\Sigma}^{-1} > 0$. The precision marix, $\boldsymbol{\Sigma}^{-1}$, has elements $\Sigma_{ij}^{-1} = \tau_i b_{ij}$ for $i \neq j$ and $\Sigma_{ii}^{-1} = \tau_i$. □

B.2 Hammersley-Clifford Theorem

As discussed in Section 2.5.1 it can be difficult to specify a full set of n full conditional distributions that provide for a valid joint distribution. This would make Markov random field (MRF) models impossible to work with. Fortunately, as presented informally in Section 2.5.1, the Hammersely-Clifford Theorem says that a random field with the Markov property is equivalent to this random field having a Gibbs distribution. Therefore, instead of concerning ourselves with the full conditional distributions we can concentrate instead on the Gibbs distribution. If a MRF has a Gibbs distribution then it has a valid joint distribution. Clifford (1990) provides a thorough proof of the theorem but we aim to produce a simpler one here. We begin with some important definitions from graph theory which we initially presented in Section 2.5.1.

Definition A set of nodes C is **complete** if all distinct nodes in C are neighbours of each other. A **clique** is a maximal complete set of all nodes. C is a clique if it is complete and no other complete set of nodes D that strictly contain C exists.

Definition Let $\mathcal{G}$ be a finite graph. A **Gibbs distribution** with respect to $\mathcal{G}$ is a probability mass function that can be expressed in the form

$$\pi(y) = \prod_{C \quad \text{complete}} V_C(y)$$

where each V_C is a function that depends only on the values $y_C = (y_s : s \in C)$ of y at the nodes in the clique C. By combining functions V_C that are subsets of the same clique this product can be further reduced to

$$\pi(y) = \prod_{C \text{ clique}} V_C(y).$$

Theorem B.2.1 (Hammersely-Clifford)**.** *Suppose that a random variable* $Y = (y_1, ..., y_n)$ *has a positive*[1] *joint probability mass function.* Y *is a Markov random field on* $\mathcal{G}$ *if and only if* Y *has a Gibbs distribution with respect to* $\mathcal{G}$

Before we provide a proof of B.2.1 note the following notational conventions that are adopted for convenience. Each random variable y_s takes its values in some finite set $\mathcal{S}_s$ and the entire MRF $Y = (y_1, ..., y_n)$ takes values in the state space $\mathcal{S} = \mathcal{S}_1 \text{x} ... \text{x} \mathcal{S}_n$. The names of the elements of the sets $\mathcal{S}_s$ are irrelevant so we adopt the labelling $\mathcal{S}_s = \{0, 1, ..., m_s\}$ for convenience in the following proof. One element has been arbirtrarily labelled "0".

Proof (Hammersley-Clifford). One direction of the proof is easy. Suppose that Y has a Gibbs distribution. It is sufficient to demonstrate that the ratio

$$\frac{\pi(Y_s = y_s \mid Y_{\neq s} = y_{\neq s})}{\pi(Y_s = 0 \mid Y_{\neq s} = y_{\neq s})} \triangleq \frac{\pi(y_s \mid y_{\neq s})}{\pi(0_s \mid y_{\neq s})} = \frac{\pi(y_s, y_{\neq s})}{\pi(0_s, y_{\neq s})}$$

depends only on $y_{\mathcal{N}(s)}$. This ratio is

$$\frac{\pi(y_s, y_{\neq s})}{\pi(0_s, y_{\neq s})} = \frac{\prod_{s \in C} V_C(y_s, y_{\neq s})}{\prod_{s \in C} V_C(0_s, y_{\neq s})} \left[\frac{\prod_{s \notin C} V_C(y_s, y_{\neq s})}{\prod_{s \notin C} V_C(0_s, y_{\neq s})} \right]$$

[1]The reason this theorem was never published despite its clear importance was that Hammersley and Clifford were convinced that it should be possible to relax the restrictive positivity condition. The theorem was eventually published in **?**

But the last fraction on the right hand side is 1 because changing y_s to 0_s obviously does not change V_C if node s is not in C. Therefore the ratio only involves functions V_C where node s is in the complete set C and every other member of C is a neighbour of node s. These functions only depend on $y_{\mathcal{N}(s)}$.

For the converse, suppose Y is a Markov random field. We need to prove that we can write the probability distribution of Y in the following form:

$$\pi(y) = \prod_{A} V_A(y) \tag{B.2.1}$$

where $V_A \equiv 1$ whenever A is not complete. Define V_A such that

$$\pi(y_D, 0_{D^c}) = \prod_{A \subseteq D} V_A(y) \tag{B.2.2}$$

holds for all $D \subseteq \{1, ..., n\}$ with $V_A \equiv 1$ when A is not complete, and (B.2.1) will follow from (B.2.2) by taking $D = \{1, ..., n\}$. Functions V_D that satisfy (B.2.2) are found recursively beginning with $D = \emptyset$, then singleton sets D, and so forth. For $D = \emptyset$, (B.2.2) says that $\pi(0) = V_\emptyset(y)$ which is a constant function taking on value $\pi(0)$. For singleton set $D = \{s\}$, (B.2.2) says

$$\pi(y_s, 0_{\neq s}) = V_\emptyset(y)V_{\{s\}}(y) = \pi(0)V_{\{s\}}(y),$$

so that

$$V_{\{s\}}(y) = \frac{\pi(y_s, 0_{\neq s})}{\pi(0)}$$

Continue this pattern recursively, with the general recursion

$$V_D(y) = \frac{\pi(y_D, 0_{D^c})}{\prod_{A \subset D} V_A(y)} \tag{B.2.3}$$

These definitions guarantee that (B.2.2) holds for all D. Completing the proof requires that we show that if D is not complete then $V_D(y) = 1$ for all y. This will be proved by induction on the number of nodes in D, to be called $\#(D)$. It is vacuously true for $\#(D) \leq 1$ since all elements must be neighbours and all sets D are complete. Start by assuming that the condition holds for $\#(D) \leq k$. We will prove that it is also true for $\#(D) = k + 1$ by induction. Suppose that when $\#(D) = k + 1$ then D is not complete and that it contains two nodes s and t that are not neighbours: $D = \{s, t\} \cup B$, where $\#(B) = k - 1$. By (B.2.3) we aim to prove that

$$\pi(y_D, 0_{D^c}) = \prod_{A \subset D} V_A(y). \tag{B.2.4}$$

Start with

$$\pi(y_D, 0_{D^c}) = \pi(y_s, y_t, y_B, 0_{D^c}) = \left[\frac{\pi(y_s, y_t, y_B, 0_{D^c})}{\pi(0_s, y_t, y_B, 0_{D^c})}\right] \pi(0_s, y_t, y_B, 0_{D^c}).$$

Since s and t are not neighbours then, by the Markov property, we have

$$\frac{\pi(y_s, y_t, y_B, 0_{D^c})}{\pi(0_s, y_t, y_B, 0_{D^c})} = \frac{\pi(y_t \mid y_s, y_B, 0D^c)}{\pi(0_s \mid y_t, y_B, 0_{D^c})} = \frac{\pi(y_t \mid 0_s, y_B, 0D^c)}{\pi(0_s \mid 0_t, y_B, 0_{D^c})} = \frac{\pi(y_t, 0_s, y_B, 0D^c)}{\pi(0_s, 0_t, y_B, 0_{D^c})}.$$

Thus

$$\begin{aligned}\pi(y_D, 0_{D^c}) &= \left[\frac{\pi(y_s, y_t, y_B, 0_{D^c})}{\pi(0_s, y_t, y_B, 0_{D^c})}\right] \pi(0_s, y_t, y_B, 0_{D^c}) \\ &= \frac{\left(\prod_{A\subseteq B\cup\{s\}} V_A(y)\right)\left(\prod_{A\subseteq B\cup\{t\}} V_A(y)\right)}{\prod_{A\subseteq B} V_B(y)} = \prod_{A\subset D\{s,t\}\nsubseteq A} V_A(y).\end{aligned}$$

However, by the induction process $V_A \equiv 1$ if $\{s,t\} \subseteq A \subset D$. Therefore,

$$\pi(y_D, 0_{D^c}) = \prod_{A\subset D\{s,t\}\nsubseteq A} V_A(y) = \left(\prod_{A\subset D\{s,t\}\nsubseteq A} V_A(y)\right)\left(\prod_{A\subset D\{s,t\}\subseteq A} V_A(y)\right) = \prod_{A\subset D} V_A(y),$$

which proves (B.2.4).

□

B.3 Conditional Distributions

Two important results from multivariate statistics concerning conditional distributions are reported for convenience. For a more detailed treatment, see for example Bierens (2004). Partition a vector of observations, $\boldsymbol{y}$ as

$$\boldsymbol{y} = \left(\frac{\boldsymbol{y}_1}{\boldsymbol{y}_2}\right),$$

and let

$$\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

with

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu_1} \\ \boldsymbol{\mu_2} \end{pmatrix} \qquad \text{and} \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma_{11}} & \boldsymbol{\Sigma_{12}} \\ \boldsymbol{\Sigma_{21}} & \boldsymbol{\Sigma_{22}} \end{pmatrix}.$$

The conditional distribution of $\boldsymbol{y}_1$ given $\boldsymbol{y}_2$ is then also multivariate normal with

$$\boldsymbol{\mu_1} \mid \boldsymbol{\mu_2} = \boldsymbol{\mu_1} + \boldsymbol{\Sigma_{12}}\boldsymbol{\Sigma_{22}}^{-1}(\boldsymbol{y}_2 - \boldsymbol{\mu_2}) \qquad \text{and} \qquad \boldsymbol{\Sigma_{1|2}} = \boldsymbol{\Sigma_{11}} - \boldsymbol{\Sigma_{12}}\boldsymbol{\Sigma_{22}}^{-1}\boldsymbol{\Sigma_{21}}, \quad \text{(B.3.1)}$$

where $\boldsymbol{\Sigma_{12}}\boldsymbol{\Sigma_{22}}^{-1}$ are regression coefficients and in which $\boldsymbol{\Sigma_{1|2}}$ is the Schlur complement of $\boldsymbol{\Sigma_{22}}$ in $\boldsymbol{\Sigma}$. Note that knowing $\boldsymbol{y}_2$ shifts both the mean and variance. Now let $\boldsymbol{\Sigma}$ be

$$\begin{pmatrix} \boldsymbol{\Sigma_{11}} & \boldsymbol{\Sigma_{12}} \\ \boldsymbol{\Sigma_{21}} & \boldsymbol{\Sigma_{22}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma^{11}} & \boldsymbol{\Sigma^{12}} \\ \boldsymbol{\Sigma^{21}} & \boldsymbol{\Sigma^{22}} \end{pmatrix}^{-1}.$$

Then we have

$$\begin{aligned} \boldsymbol{\Sigma_{11}} &= [\boldsymbol{\Sigma^{11}} - \boldsymbol{\Sigma^{12}}(\boldsymbol{\Sigma^{22}})^{-1}\boldsymbol{\Sigma^{21}}]^{-1} \\ \boldsymbol{\Sigma_{22}} &= [\boldsymbol{\Sigma^{22}} - \boldsymbol{\Sigma^{21}}(\boldsymbol{\Sigma^{11}})^{-1}\boldsymbol{\Sigma^{12}}]^{-1} \end{aligned}$$

and

$$\begin{aligned} \boldsymbol{\Sigma_{12}} &= -(\boldsymbol{\Sigma^{11}})^{-1}\boldsymbol{\Sigma^{12}}\boldsymbol{\Sigma_{22}} &= -\boldsymbol{\Sigma_{11}}\boldsymbol{\Sigma^{12}}(\boldsymbol{\Sigma^{22}})^{-1} \\ \boldsymbol{\Sigma_{21}} &= -(\boldsymbol{\Sigma^{22}})^{-1}\boldsymbol{\Sigma^{21}}\boldsymbol{\Sigma_{11}} &= -\boldsymbol{\Sigma_{22}}\boldsymbol{\Sigma^{21}}(\boldsymbol{\Sigma^{11}})^{-1} \end{aligned}.$$

Combining these two results for the inverses of partitioned matrices we obtain

$$\begin{aligned} \boldsymbol{\Sigma_{1|2}} &= \boldsymbol{\Sigma_{11}} - (\boldsymbol{\Sigma^{22}})^{-1}\boldsymbol{\Sigma^{21}}\boldsymbol{\Sigma_{11}}(\boldsymbol{\Sigma^{22}})^{-1}\boldsymbol{\Sigma^{21}}\boldsymbol{\Sigma_{11}} \\ &= (\boldsymbol{\Sigma^{11}})^{-1}(\boldsymbol{\Sigma^{11}} - \boldsymbol{\Sigma^{12}}(\boldsymbol{\Sigma^{22}})^{-1}\boldsymbol{\Sigma^{21}})\boldsymbol{\Sigma_{11}} \\ &= (\boldsymbol{\Sigma^{11}})^{-1} \end{aligned}$$

B.4 Conditional & Covariance Structure of the Multi-normal Distribution

The presentation of the MCAR in Chapter 3 and the point of departure for the FMCAR developed in Chapter 5 is based upon a key result from Mardia (1988) for the multivariate normal distribution. Consider a p-dimensional set of random variables $\boldsymbol{y}_1, ..., \boldsymbol{y}_n$ where

$$\mathrm{E}(\boldsymbol{y}_i|\boldsymbol{y}_{-i}) = \boldsymbol{\mu_i} + \sum_{j \in \partial_i} \mathbf{B}_{ij}(\boldsymbol{y}_j - \boldsymbol{\mu_j}), \qquad i = 1, ..., n \tag{B.4.1}$$

and

$$\mathrm{Var}[\boldsymbol{y}_i|\boldsymbol{y}_{-i}] = \mathbf{\Gamma_i}, \qquad i = 1, ..., n \tag{B.4.2}$$

where the $\boldsymbol{y}_{-i} \equiv (\boldsymbol{y}_j : j \in \partial_i)$ refers to the "rest" of the variables ($\boldsymbol{y}$) at the other sites in the neighbourhood of i. Mardia (1988) proves that given a multivariate vector $\boldsymbol{y} = (\boldsymbol{y}_1', ..., \boldsymbol{y}_n')$ of length np that the joint distribution for $\boldsymbol{y}$ is $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

Theorem B.4.1 (Mardia)**.** *Given the n conditional multi-normal distributions, $\boldsymbol{y}$ is $\mathcal{N}_{np}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where*

$$\boldsymbol{\mu} \equiv (\boldsymbol{\mu_1'}, ..., \boldsymbol{\mu_n'})' \quad \text{and} \quad \boldsymbol{\Sigma} \equiv [\mathrm{Block}(-\mathbf{\Gamma_i^{-1}B_{ij}})]^{-1}. \tag{B.4.3}$$

provided

$$\mathbf{B_{ij}\Gamma_j} = \mathbf{\Gamma_i B_{ji}'}, \qquad i, j = 1, ..., n \tag{B.4.4}$$

and

$$\text{Block}(-\mathbf{\Gamma_i^{-1}B_{ij}}) \qquad \text{or} \qquad \text{Block}(\mathbf{B_{ij}}) \tag{B.4.5}$$

are positive-definite. Further, the p.d.f. of $\boldsymbol{y}$ *is*

$$(2\pi)^{-np/2} \prod_{i=1}^{n} |\Gamma_i|^{-1/2} |\text{Block}(-\mathbf{B_{ij}})|$$

$$\text{x} \quad \exp \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} (\boldsymbol{y}_i - \boldsymbol{\mu}_i)' \mathbf{\Gamma}_i^{-1} \mathbf{B}_{ij} (\boldsymbol{y} - \boldsymbol{\mu}_j)\ . \tag{B.4.6}$$

Proof. The Brook expansion of B.1 can be extended to the multivariate case i.e. if $\boldsymbol{x} = (\boldsymbol{x}_1', ..., \boldsymbol{x}_n')'$ and $\boldsymbol{y} = (\boldsymbol{y}_1', ..., \boldsymbol{y}_n')'$ are identically distributed with the joint p.d.f $f()$, then

$$f(\boldsymbol{x})f(\boldsymbol{y}) = \prod_{i=1}^{n} f(\boldsymbol{x}_i|\boldsymbol{x}_1, ..., \boldsymbol{x}_{i-1}, \boldsymbol{y}_{i+1}, ..., \boldsymbol{y}_n) f(\boldsymbol{y}_i|\boldsymbol{x}_1, ..., \boldsymbol{x}_{i-1}, \boldsymbol{y}_{i+1}, ..., \boldsymbol{y}_n). \tag{B.4.7}$$

Assuming a zero mean process and that $\boldsymbol{y} = \mathbf{0}$ then from (B.4.1) and (B.4.2), equation (B.4.7) simplifies to

$$-2\log f(\mathbf{0})f(\boldsymbol{y}) = \sum_{i=1}^{n} \boldsymbol{y}_i' \mathbf{\Gamma}_i^{-1} \boldsymbol{y}_i - 2 \sum_{i=2}^{n} \sum_{j=1}^{i-1} \boldsymbol{y}_i' \mathbf{\Gamma}_i^{-1} \mathbf{B_{ij}} \boldsymbol{y}_j. \tag{B.4.8}$$

The forward version[2] of (B.4.7) can be written as

$$f(\boldsymbol{x})f(\boldsymbol{y}) = \prod_{i=1}^{n} f(\boldsymbol{x}_i|\boldsymbol{y}_1, ..., \boldsymbol{y}_{i-1}, \boldsymbol{x}_{i+1}, ..., \boldsymbol{x}_n) f(\boldsymbol{y}_i|\boldsymbol{y}_1, ..., \boldsymbol{y}_{i-1}, \boldsymbol{x}_{i+1}, ..., \boldsymbol{x}_n). \tag{B.4.9}$$

[2]See Mardia (1988) p. 267 for an explanation.

which leads in the same manner as (B.4.8) to

$$-2\log f(\mathbf{0})f(\boldsymbol{y}) = \sum_{i=1}^{n} \boldsymbol{y}_i'\boldsymbol{\Gamma}_i^{-1}\boldsymbol{y}_i - 2\sum_{i=1}^{n-1}\sum_{j=i}^{n} \boldsymbol{y}_i'\boldsymbol{\Gamma}_i^{-1}\mathbf{B_{ij}}\boldsymbol{y}_j. \tag{B.4.10}$$

Since (B.4.8) and (B.4.10) must be identical, the coefficients $\mathbf{B_{ij}}$ must satisfy (B.4.6). Further, from (B.4.10) we have

$$-2\log f(\boldsymbol{y}) = \text{Const.} + \sum_{i=1}^{n} \boldsymbol{y}_i'\boldsymbol{\Gamma}_i^{-1}\boldsymbol{y}_i - \sum_{i\neq j} \boldsymbol{y}_i'\boldsymbol{\Gamma}_i^{-1}\mathbf{B_{ij}}\boldsymbol{y}_j. \tag{B.4.11}$$

Thus $\boldsymbol{y}$ is $\mathcal{N}_{np}(\mathbf{0}\boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is as defined in (B.4.3), provided that $\boldsymbol{\Sigma}$ is positive definite i.e. if (B.4.4) holds. □