Challenges in the Analysis of Data with Clustering or Correlation

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Lisa Henn

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
Doctor of Philosophy

August, 2015

Acknowledgments

I thank my co-adviser Jim Hodges for his guidance over all the topics of this dissertation. I would like also to thank my co-adviser John Hughes for introducing me to the world of Gaussian copula regression models. The topic grew to encompass two-thirds of this dissertation, so engaging was the work. It has been a very informative collaboration.

Numerous additional people provided financial support and research opportunities that also facilitated my growth by expanding my knowledge of statistical practice. These include Julian Wolfson, my first academic adviser, Haitao Chu, and all the faculty members at the Biostatistical Design and Analysis Center, in particular Kyle Rudser. Nidhi Kohli and Chun Wang, respectively of the Departments of Educational Psychology and Psychology, provided support during a period when I worked remotely and introduced me to a new area of application in statistics. Also during that time, Brad Carlin facilitated additional support from the Division for tuition for research credits, for which I am grateful. The Division previously had supported me as a Teaching Assistant as well, which allowed me to work with Wei Pan and Richard MacLehose. The financial support from everyone was crucial to my being able to complete my degree.

Sally Olander helped me time and again to navigate the administrative requirements of the university, the importance of which cannot be overstated. The moral support and camaraderie from my classmates buoyed me at various stages of my program. I miss them already.

Dedication

To John Kasab, who went to great lengths to facilitate my matriculation at Minnesota, and who encouraged me endlessly. This dream would never have been realized without him.

Abstract

We examine several challenges associated with the analysis of clustered or correlated data.

First, multiple maxima can occur in posterior distributions or likelihoods for mixed linear models, though they are sparsely noted in the literature. For those problems with covariance structures that are diagonalisable in a specific sense, we present the restricted likelihood as a generalised linear model with gamma errors, identity link and a prior distribution on the error variance. The generalised linear model portion of the restricted likelihood can be made to conflict with the portion of the restricted likelihood that functions like a prior distribution on the error variance, and in doing so we demonstrate that multiple maxima can occur in the restricted likelihood as well. Adding an explicit conjugate prior distribution to variance parameters permits a second local maximum in the marginal posterior distribution even if the likelihood contribution has a single maximum. Moreover, reparameterisation from variance to precision can change the posterior modality; the converse also is true. Modelers should beware of these potential pitfalls when selecting prior distributions or using peak-finding algorithms to estimate parameters.

Second, Gaussian copula regression models provide a flexible, intuitive framework in which to model dependent responses with a variety of marginal distributions. The time required to compute the likelihood directly grows exponentially with sample size. Instead, we conduct inference for Gaussian copula regression models of non-continuous outcomes using three distinct approaches in a Bayesian setting: the continuous extension, the distributional transform, and the composite likelihood. The latter two include curvature correction. Using frequentist methods, we evaluate the inference resulting from these three approaches for several types of non-continuous data. We consider

the computational performance as well. In most cases, the distributional transform with curvature correction has acceptable but considerably faster performance, making it attractive for evaluating models of mutually dependent non-continuous responses.

Finally, we extend the Gaussian copula regression model to ordinal outcomes. The setting is femoroacetabular impingement (FAI), a condition in which subtle deformities of the femoral head and acetabulum (hip socket) result in pathologic abutment during hip motion. We apply a proportional odds model to relate T2* mapping data (an investigational, objective MRI sequence) to Beck's scale of cartilage damage, with adjustment for spatial proximity of responses. Each hip in the study is assigned its own parameter to describe the degree of association among the spatially grouped responses. A Dirichlet process allows clustering of these spatial association parameters. Using the fitted model, a patient-specific map of the entire acetabular cartilage displays gradations of cartilage quality with six colors representing the six grades of cartilage quality on Beck's scale. Such a map will facilitate patient education and clinical decision-making.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In this work we examine several challenges associated with the analysis of clustered or correlated data.

Our first challenge arises in the mixed linear modeling approach to analyzing such data. Multiple local maxima can arise in both restricted likelihoods and posterior distributions of variance (or precision) parameters. A fairly large class of such models can be written in scalar terms, which we exploit to understand the mechanism by which multiple maxima can arise.

The second area of research explores techniques for Gaussian copula regression models applied to non-continuous data. Gaussian copula regression models provide the ability to define an association among observations while allowing flexibility in the choice of marginal distributions of the observations. This combination makes these models a powerful and versatile tool applicable to a wide range of settings and response types. Analysis of non-continuous data with copula models necessitates the use of alternate approaches for likelihood-based inference. We consider frequentist performance of Bayesian analyses of three such approaches in three settings for non-continuous data.

Our third challenge concerns prior distributions for association parameters in Gaussian copula regression models. When the data favor association parameter values near the bound of 1, Bayesian analyses of such models can be stymied. Few continuous distributions can be parametrized intuitively such that the majority of its mass is clustered in this range. A Dirichlet process prior distribution for these parameters allows a finite number of candidate parameter values in this range to arise from a continuous

distribution and additionally provides a mechanism for the data to reveal clustering. The latter attribute can be especially attractive to clinicians seeking to group patients into treatment regimes based on current disease state. We illustrate the application of this Dirichlet process prior distribution with a proportional odds model of disease progression in the cartilage lining the acetabulum of the hip joint.

The rest of this dissertation proceeds as follows: Chapter 2 shows how multiple maxima can occur with restricted likelihood and with posterior distributions. In Chapter 3, we examine the performance of three approximation approaches for the likelihood of Gaussian copula regression models applied to non-continuous data. Chapter 4 applies one of the approximation approaches for the likelihood of Gaussian copula regression models to clinical data and explores the use of a Dirichlet process prior distribution for the association parameters. The collection of this data as well as preliminary modeling is the work of several collaborators named there. Finally, Chapter 5 summarizes the findings.

Chapter 2

Multiple Local Maxima in Restricted Likelihoods and Posterior Distributions for Mixed Linear Models

In this chapter, Section 2.1 provides the background to the challenge of multiple local maxima in restricted likelihoods and posterior distributions in mixed linear models. Section 2.2 establishes the notation for this work. Section 2.3 details the class of models of interest and how the restricted likelihood can be viewed as the likelihood for a GLM with an informative prior distribution on the variances, with examples showing the scope of this class of models. Section 2.4 demonstrates multiple maxima in a restricted likelihood, while Section 2.5 considers multiple maxima in the posterior distribution and explores performance under reparameterisation. Section 2.6 provides some guidance on detecting multiple maxima. Section 2.7 concludes.

Material in this chapter also appears in Henn and Hodges (2014).

2.1 Introduction

Maxima of functions or probability distributions play a central role in inference. The analyst may have intuition about the location of a maximum, particularly with generalised linear mixed models, or there may be very little intuition, necessitating a broad search through the parameter space. Most instruction concerns simple exponential family models with canonical links, which have a single maximum. Because most training is with single maximum cases, analysts may not look for other local maxima.

Multiple maxima are relatively common in likelihoods, which can be viewed as proper posterior distributions resulting from flat prior distributions. Little is known, however, about multiple maxima in restricted likelihoods, which are marginal posterior distributions for certain prior distributions.

Many procedures, such as the expectation-maximisation algorithm, guarantee convergence only to a local maximum. Woodard *and others* (2009) identified conditions under which chains for multimodal distributions constructed by parallel or simulated tempering mix slowly. Common convergence diagnostics, *e.g.* the Gelman-Rubin statistic, fail to detect the slow mixing. Second maxima may not be detected under those circumstances. Approximations for the likelihood, restricted likelihood or posterior distribution emerge from a central limit theorem regarding a maximum. Again, intuition concerning such an approximation arises from familiarity with canonical exponential families, which may not apply to other cases such as mixed linear models, which this work discusses.

Even within mixed linear models, the phenomenon of multiple maxima is subtle, and it appears that what little is known is not widely known. The existing literature is sporadic in coverage and thus produces an incomplete picture. This work shows that second local maxima can occur in posterior distributions and in restricted likelihoods from mixed linear models, and it discusses how they occur. They happen commonly in posterior distributions, the second local maximum arising most readily from the prior distribution. Because most prior distributions are intended to be non-informative, the posterior distribution's susceptibility to multiple maxima suggests a prior distribution that does not conform to the analysts intuition.

This work proposes an idea that unifies a large class of mixed linear models to illuminate the multiple maxima problem. Within this class, a restricted likelihood can be interpreted as the likelihood for a generalised linear model (GLM) with the canonical parameter being a function of the unknown variances, along with an informative inverse gamma prior distribution for the error variance. This class includes all two-variance models, described subsequently, and many other models. Tools for GLMs, then, facilitate understanding of multiple maxima in the restricted likelihood and posterior distribution. We use this formulation to show how restricted likelihoods with multiple maxima can occur and to create examples in which modality changes with reparameterisation from variance to precision or vice versa, the first time this has been reported. We provide some suggestions for detecting multiple maxima in posterior distributions and restricted likelihoods as well, although this is by no means a comprehensive list.

2.2 Notation

The mixed linear model for an observation vector y of length n has the form

$$\begin{aligned} y &= X\beta + Zu + \epsilon, \text{ with } \epsilon \sim \mathcal{N}(0, R) \text{ the vector of error terms,} \quad (2.1)\\ X &\quad \text{the } n \times p \text{ design matrix for the } p \text{ fixed effects } \beta, \text{ and} \\ Z &\quad \text{the } n \times m \text{ design matrix for the } m \text{ random effects } u, \text{ or} \\ y \mid \beta, u, R &\sim \mathcal{N}\left(X\beta + Zu, R\right), \\ u \mid G &\sim \mathcal{N}\left(0, G\right). \end{aligned}$$

Unconditionally with respect to u, we have $y \mid \beta, V \sim \mathcal{N}\left(X\beta, V\right)$, $V = ZGZ^T + R$, which specifies a well-defined likelihood for β and unknowns in G and R. While a non-zero mean could be specified for the random effects u, it is more common to set it equal to zero. Without loss of generality, we do so.

The restricted likelihood often is defined from a frequentist perspective, transforming the observations using a contrast that renders the mean structure equal to zero. It also can be defined from a Bayesian perspective. As shown in Searle *and others* (2006), if a flat prior distribution is used for the fixed effects β and for the unknowns in the variance structure, the result is mathematically the same as the restricted likelihood arising from transforming the observation y.

2.3 Unifying Idea

The unifying idea is that for a certain class of models, the restricted likelihood can be interpreted as a generalized linear model, with the canonical parameter a function of the unknown variances, and an informative inverse gamma prior distribution for the error variance. Members of this class of models have covariance structures that are diagonalizable in a sense described below. We introduce this unifying idea and class of models using the two-variance model. Then we show this class includes more general models.

Reich and Hodges (2008) and subsequently Welham and Thompson (2009) decomposed the restricted likelihood by diagonalizing the variance-covariance matrices. As a result, the restricted likelihood can be presented in scalar rather than matrix terms.

Following Reich and Hodges (2008), the mixed linear model with two unknown variances can be formulated as follows, with the normal distributions written in terms of their precision matrices instead of covariance matrices.

$$\begin{aligned} y \mid \beta_1, \sigma^2 &\sim \mathcal{N}\left(X_1\beta_1, \frac{1}{\sigma^2} I_n\right) \text{ (precision formulation)}, \\ \beta_1 \mid \beta_2, \tau^2 &\sim \mathcal{N}\left(Z_1\beta_2, \frac{1}{\tau^2} P\right) \text{ (precision formulation), } P \text{ a precision matrix} \\ \beta_2 &\sim \text{flat.} \end{aligned} \tag{2.2}$$

Integrating β_2 out of this model results in the following prior distribution for β_1:

$$\beta_1 \mid \tau^2 \sim \mathcal{N}\left(0, \frac{1}{\tau^2}[P - PZ_1(Z_1^T P Z_1)^{-1} Z_1 P]\right) \text{ (precision formulation)} \tag{2.3}$$

Define

$$\begin{aligned} \Upsilon &= (X_1^T X_1)^{-1/2}[P - PZ_1(Z_1^T P Z_1)^{-1} Z_1 P](X_1^T X_1)^{-1/2} \\ &= \Gamma \mathcal{D} \Gamma^T, \ \mathcal{D} \text{ diagonal, } \Gamma \text{ orthogonal} \end{aligned}$$

As Υ is real-valued and symmetric, this model's restricted likelihood decomposes as desired. Further, define

$$\begin{aligned} X_2 &= X_1(X_1^T X_1)^{-1/2}\Gamma^T \\ \gamma &= \Gamma(X_1^T X_1)^{1/2}\beta_1 \end{aligned}$$

The model in Equation (2.2) then becomes

$$\begin{aligned} y \mid \gamma, \sigma^2 &\sim \mathcal{N}\left(X_2\gamma, \frac{1}{\sigma^2} I_n\right) \text{ (precision formulation)}, \\ \gamma \mid \tau^2 &\sim \mathcal{N}\left(0, \frac{1}{\tau^2}\mathcal{D}^{-1}\right) \text{ (precision formulation)}. \end{aligned} \tag{2.4}$$

Adding and subtracting the product of X_2 and the unshrunk estimate of γ, $\hat{\gamma} = \left(X_2^T X_2\right)^{-1} X_2^T y$ to the observations y and defining $\text{SSE} = (y - X_2\hat{\gamma})^T(y - \Gamma^T X_2\hat{\gamma})$, the restricted likelihood becomes

$$p(\sigma^2, \tau^2 \mid y) \propto (\sigma^2)^{-\frac{(n-m)}{2}} \exp\left(-\frac{1}{2\sigma^2}\text{SSE}\right) \prod_{i=1}^{(m-p)} \left[\left(\frac{d_i^*}{\sigma^2 d_i^* + \tau^2}\right)^{\frac{1}{2}} \exp\left(-\frac{1}{2}\hat{\gamma}_i^2 \frac{d_i^*}{\sigma^2 d_i^* + \tau^2}\right)\right], \tag{2.5}$$

where m is the number of columns of Z and d_i^* are the diagonal elements of $\mathcal{D}$.

The first two terms of Equation (2.5),

$$(\sigma^2)^{-\frac{(n-m)}{2}} \exp\left(-\frac{1}{2\sigma^2}\text{SSE}\right), \tag{2.6}$$

can be called "free" terms, as they are free of τ^2. They fit the form of an inverse gamma distribution. The remaining terms of Equation (2.5),

$$\prod_{i=1}^{(m-p)} \left[\left(\frac{d_i^*}{\sigma^2 d_i^* + \tau^2}\right)^{\frac{1}{2}} \exp\left(-\frac{1}{2}\hat{\gamma}_i^2 \frac{d_i^*}{\sigma^2 d_i^* + \tau^2}\right)\right], \tag{2.7}$$

can be called "mixed" terms because both σ^2 and τ^2 appear in each term. As Welham and Thompson (2009) noted, these mixed terms can be viewed as a generalized linear model with gamma errors and an identity link, as we now make explicit.

In a generalized linear model (McCullagh and Nelder, 1989), the responses y have an exponential family distribution of the form

$$f_Y(y; \theta, \phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right\},$$

with θ being the parameter of interest. Other components include y, the observations; the function $b(\theta)$, a function only of θ; $a(\phi)$, a scaling function (the dispersion function); and $c(y, \phi)$, the remaining components that are functions solely of y and ϕ. The mean

of y is related to the linear model $X\gamma$ through a link function η. The resulting log-likelihood is then

$$l(\theta; y, \phi) = \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi).$$

From Equation (2.7) above, we have for a constant κ,

$$\begin{aligned} p_{GLM}(\sigma^2, \tau^2 \mid \mathbf{w}) &\propto \prod_{i=1}^{m-p} \left(\frac{d_i}{\sigma^2 d_i + \tau^2} \right)^{\frac{1}{2}} \exp\left(-\frac{1}{2} \hat{\gamma}_i^2 \frac{d_i}{\sigma^2 d_i + \tau^2} \right) \equiv \prod_{i=1}^{m-p} L_i \\ \log(p_{GLM}(\sigma^2, \tau^2 \mid \mathbf{w})) &= \frac{1}{2} \sum_{i=1}^{m-p} \left(\log\left(\frac{d_i}{\sigma^2 d_i + \tau^2} \right) + \hat{\gamma}_i^2 \left(\frac{-d_i}{\sigma^2 d_i + \tau^2} \right) \right) + \kappa. \end{aligned}$$

In terms of the GLM form,

$$\begin{aligned} \text{the data } y_i &= \hat{\gamma}_i^2,\ i = 1, \ldots, m-p \\ \text{the canonical parameter } \theta_i &= \frac{-d_i}{\sigma^2 d_i + \tau^2}, \\ \mathbb{E}(y_i) = \eta(\sigma^2, \tau^2, d_i) &= \mu_i = \frac{-1}{\theta_i} = \sigma^2 + \frac{1}{d_i}\tau^2, \\ b(\theta_i) &= -\log(-\theta_i) = -\log\left(\frac{d_i}{\sigma^2 d_i + \tau^2} \right), \text{ and} \\ a(\phi) &= 2. \end{aligned}$$

The free terms have the same form as a prior distribution on the error variance σ^2, specifically, an Inverse Gamma distribution as noted above, and can be viewed as a prior distribution as a convenience for the purpose of evaluating the contribution of the free terms to the restricted likelihood.

Alternatively, the free terms also can be viewed as another observation in the generalized linear model with gamma errors and identity link. To see this, rewrite a single mixed term as follows:

$$\left[\left(\frac{d_i}{\sigma^2 d_i + \tau^2} \right)^{\frac{1}{2}} \exp\left(-\frac{1}{2} \hat{\gamma}_i^2 \frac{d_i}{\sigma^2 d_i + \tau^2} \right) \right] = \left[\left(\frac{1}{\sigma^2 + \tau^2 / d_i} \right)^{\frac{1}{2}} \exp\left(-\frac{1}{2} \hat{\gamma}_i^2 \frac{1}{\sigma^2 + \tau^2 / d_i} \right) \right]. \tag{2.8}$$

As $d_i \to \infty$, this mixed term becomes

$$\left[\left(\frac{1}{\sigma^2} \right)^{\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2} \hat{\gamma}_i^2 \right) \right], \tag{2.9}$$

which has the same form as Equation (2.6). Conversely, Equation (2.6) can be written in the GLM form by setting

$$\begin{aligned} \text{the data } y &= \frac{\text{SSE}}{(n-m)}, \\ \text{the canonical parameter } \theta &= \frac{-1}{\sigma^2}, \\ \mathbb{E}(y) = \eta(\sigma^2) &= \mu = \frac{-1}{\theta} = \sigma^2, \\ b(\theta) &= -\log(-\theta) = -\log\left(\frac{1}{\sigma^2}\right), \text{ and} \\ a(\phi) &= \frac{2}{(n-m)}. \end{aligned}$$

Sometimes it can be advantageous to consider the free terms as contributing another "observation" in the GLM with gamma errors and identity link, while at other times, it may be fruitful to think of these terms as functioning as if they were a prior distribution on the error variance σ^2. For example, considering the free terms of Equation (2.6) as if it were a prior distribution on the error variance σ^2, it is easy to see how free terms and mixed terms can be set in conflict with one another through manipulation of the data values comprising the SSE and mixed terms, producing multiple maxima in the restricted likelihood.

Wedderburn (1976) surveyed maximum likelihood estimates for generalized linear models. He identified conditions yielding unique or finite estimates in the parameter space's interior and cataloged these attributes according to error model and link function combinations. For gamma likelihoods, it suffices for uniqueness that each likelihood term L_i be a strictly concave function of the observation y_i, in this case $\hat{\gamma}_i^2$. For a gamma likelihood and identity link, the maximum likelihood estimate is finite and in the interior, but its uniqueness is not established by this sufficient condition. Considering the free terms instead as an observation in the GLM with gamma errors and identity link, in the event that we produce multiple maxima in the restricted likelihood, as we do in Section 2.4 below, this provides an example in which this particular GLM in fact has multiple maxima.

2.4 Multiple Local Maxima in Restricted Likelihood

Since the restricted likelihood can be interpreted as a generalized linear model with a gamma likelihood with a prior distribution on the error variance, two local maxima can be produced by making the observations $\hat{\gamma}_i^2$ conflict with this prior distribution. We will use this insight to generate an example where the restricted likelihood has two maxima. An intrinsic conditionally autoregressive (spatial) model (ICAR) illustrates. The model is as follows.

$$\begin{aligned} y_{ij} &= \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varsigma_i + \epsilon_i, \text{ where } \epsilon_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \sigma^2\right), \\ f(\varsigma \mid \sigma_s^2) &\propto \left(\sigma_s^2\right)^{(N-S)/2} \exp\left(-\frac{1}{2\sigma_s^2}\varsigma^T Q \varsigma\right) \text{ an ICAR model, with } Q \text{ given below.} \end{aligned}$$

An ICAR model for spatial clustering uses a proximity matrix Q of the form

$$\begin{aligned} q_{ii} &= \text{a region's number of neighbors} \\ q_{ij} &= \begin{cases} -1 \text{ if regions } i \text{ and } j \text{ are neighbors} \\ 0 \text{ otherwise.} \end{cases} \end{aligned} \tag{2.10}$$

For this example, the proximity matrix Q was randomly generated for 341 observations at 339 sites; only two sites have repeat observations. The spatial map is connected, forming precisely one "island". The matrix size and the fixed effects were chosen to use all the observations from the insurance premium data set described in Hodges (1998) though discarding that data set's actual clustering of plans within jurisdictions.

Hodges (1998) modeled health maintenance organization (HMO) insurance premium data using Equation (2.2) and a Bayesian analysis with a flat prior distribution on the fixed effects β and a Gamma(1.1, 0.1) prior distribution on the group precision $1/\tau^2$. The Gamma(a, b) prior distribution was parameterized with mean equal to a/b. As described in Hodges (1998), the HMO data set describes 341 health maintenance organizations located in 45 states or similar political jurisdictions. Each jurisdiction had between 1 and 31 plans with a median of 5 plans. The data set originally was analyzed to assess the cost of moving military retirees and dependents from a Department of Defense health plan to plans serving the US civil service. In Hodges (1998), the data set was used to demonstrate a number of proposed diagnostic techniques for hierarchical

models. Specifically, the model is

$$\begin{aligned} y_{ij} &= \alpha_i + \epsilon_{ij} \\ \alpha_i &= \varrho_0 + \varrho_1 x_{1i} + \varrho_2 x_{2i} + \varsigma_i, \end{aligned} \tag{2.11}$$

where the fixed effects in α_i include an intercept, jurisdiction-average hospital expenses per admission (x_{1i}), and an indicator for plans in New England states (x_{2i}). The restricted likelihood has the same form as for Equation (2.2).

The resulting model based on this insurance premium data set had a large number of observations n but had a small difference between number of observations and number of groups, $n-m$. The maximum number of neighbor pairs q_{ii} was 321, the minimum, 22. This ICAR model yields eigenvalues d_i that are considerably greater than 1, ranging from 21.9 to 322.1. The estimate for σ^2 arising from the restricted likelihood's implicit gamma GLM can be made to conflict with the restricted likelihood's implicit "prior" distribution on the error variance. The maximizing value from the latter is $\hat{\sigma}^2 = \text{SSE}/(n-m)$; to produce a conflict, SSE was made quite small relative to $\hat{\gamma}_i^2$.

Figures 2.1 and 2.2 depict the result. Figure 2.1 shows the mixed terms without the implicit "prior" distribution, maximized at approximately $\sigma^2 = 160$, $\tau^2 = 6800$ (marked by a dot). The implicit "prior" distribution is maximized at approximately $\sigma^2 = 2 \times 10^{-3}$. Figure 2.2 shows the full restricted likelihood, which has two peaks arising from the conflict between the mixed and free terms. This is directly analogous to the familiar conflict between the likelihood and the prior distribution.

While most instruction concerns exponential families with canonical links, the structure of this GLM, now with a non-canonical link, may permit two or more maxima within the GLM portion alone. Wedderburn (1976) shows that the likelihood for the gamma GLM with identity link is not necessarily a strictly concave function of the parameter and therefore the uniqueness of the maximum is not assured. For definiteness, consider the estimation of a grand mean; call it β_0. The second partial derivative for this model, derived from the chain rule, is

$$\frac{\partial^2 \log L}{\partial \beta_0^2} = \sum_{i=1}^{n} \left[y_i \theta_i^3 + \frac{1}{2} \theta_i^2 \right] \tag{2.12}$$

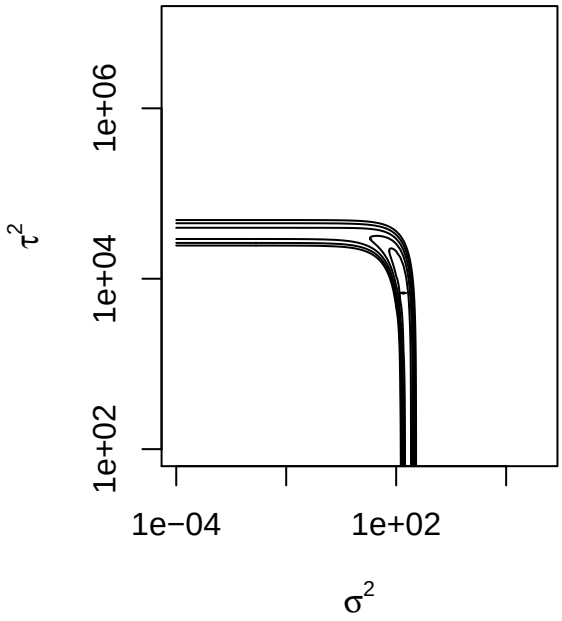

Figure 2.1: ICAR Model, 341 observations, 339 groups, gamma GLM terms only. Contours in 1 log increments. Axes have logarithmic scales.

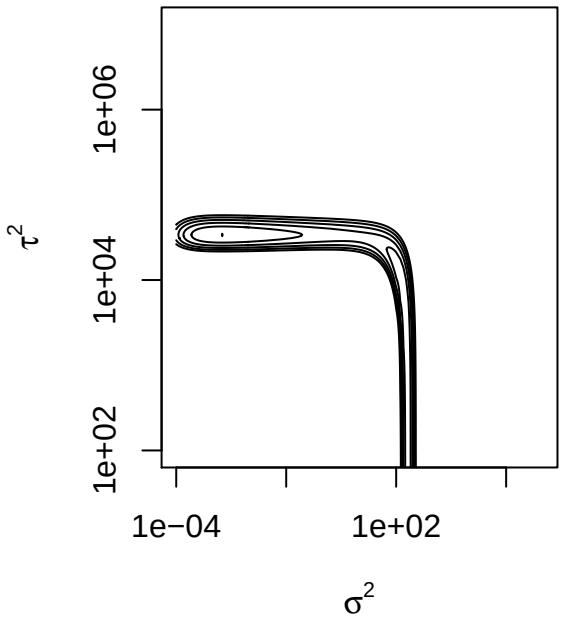

Figure 2.2: ICAR Model, 341 observations, 339 groups, gamma GLM terms and implicit prior distribution. Contours in 1 log increments. Axes have logarithmic scales.

with, in the case of the two-variance model,

$$\begin{aligned} \theta_i &= \frac{-d_i}{\sigma^2 d_i + \tau^2} \text{ and} \\ y_i &= \hat{\gamma}_i^2 \text{ in the nomenclature of Equation (2.5).} \end{aligned}$$

This second partial derivative is less than zero, and the gamma GLM likelihood is strictly concave, if $|\theta_i| > 1$ and $y_i > \frac{1}{2}$ for all i. This is sufficient but not necessary. Thus multiple maxima could occur in cases for which the second partial derivative is not less than zero.

As described in Section 2.3, the free terms of Equation (2.6) also can be considered as an element in the GLM with gamma errors and identity link. The present example thus demonstrates that such GLM's can themselves have multiple maxima. This fact suggests that it may be possible to produce two or more maxima among the mixed terms alone, so that considering the entire restricted likelihood with both mixed and free terms could result in three or more maxima in a two-variance model.

2.5 Modality Changes

Section 2.4 showed how to make restricted likelihoods with multiple local maxima. Doing so with posterior distributions is easier; for a flat prior distribution on the fixed effects, a posterior distribution is merely the restricted likelihood multiplied by explicit prior distributions for the variance components. In the following real example, bimodality occurred unexpectedly. Disconcertingly, the number of maxima in the posterior distribution is sensitive to whether it is a posterior distribution for variances or precisions. Here we consider only two-variance models and continue to assume a flat prior distribution on any fixed effects. While unnecessary, this yields some simplification.

If the restricted likelihood of Equation (2.5) is multiplied by an explicit Inverse Gamma(c, d) prior distribution of the form

$$p(\sigma^2) \propto \left(\sigma^2\right)^{-(c+1)} \exp\left(-\frac{1}{2}\frac{d}{\sigma^2}\right),$$

the first two terms of the resulting posterior distribution have the form

$$(\sigma^2)^{-\left[\frac{(n-m)}{2}+c+1\right]} \exp\left(-\frac{1}{2}\left[\frac{\text{SSE}}{\sigma^2}+d\right]\right).$$

Effectively, the restricted likelihood's implicit "prior" distribution absorbs the explicit prior distribution on σ^2. In most cases, $n-m$ dominates c and the SSE term dominates d, making the explicit prior distribution's effect modest.

By contrast, the restricted likelihood has no terms with the form of a prior distribution for the variance τ^2. An explicit prior distribution for τ^2 can thus have more influence on the shape of the posterior distribution.

These explicit prior distributions for σ^2 and τ^2 are independent of one another, so their product necessarily is unimodal. Thus, if the gamma GLM portion of the restricted likelihood has only one maximum, a posterior distribution of this form cannot have more than two local maxima. Other prior distributions for (σ^2, τ^2) may produce more maxima in the posterior distribution.

Hodges (1998) modeled health maintenance organization (HMO) insurance premium data using the form of Equation (2.2). A discussant of that paper, Jon Wakefield, analyzed the same data. Figure 2.4 shows the restricted likelihood for this problem, which has only one maximum. No change in contour shape indicative of a local maximum could be found in the flat $\sigma^2 \times \tau^2$ region of (500, 600) $\times$ (10^{-3}, 1), regardless of contour resolution. Figure 2.3, showing the Gibbs sampler draws, also indicates only one maximum. (For this and the other Gibbs sampler results in Figures 2.5, and 2.7, 13,000 samples total were acquired, and the first 3,000 were discarded as "burn-in".)

Wakefield used an Inverse Gamma(1.1, 0.1) prior distribution for the group variance τ^2 parameterized such that the mean is equal to $b/(a-1)$. Figures 2.5 and 2.6 reproduce his analysis. Two local maxima appear here. Comparing these figures to Figures 2.3 and 2.4, the prior distribution for the group variance clearly induces the second local maximum. Figures 2.7 and 2.8 reproduce Hodges' analysis, which parameterized the model in terms of precisions $1/\tau^2$ and $1/\sigma^2$ and used a Gamma(1.1, 0.1) prior distribution for the group precision $1/\tau^2$ and a flat prior distribution on $1/\sigma^2$. The Gamma(a, b) prior distribution is parameterized such that the mean is equal to a/b. This last analysis shows only one maximum, again confirmed by increasing the contour density in the flat portion of the plot, in this case the $1/\sigma^2 \times 1/\tau^2$ region of (0.0015, 0.0020) $\times$ (0.1, 10).

The second local maximum in the (σ^2, τ^2) parameterization (Figures 2.5 and 2.6), generated by the inverse gamma prior distribution for the group variance, has a higher peak than the primary local maximum, which arises from the restricted likelihood. Thus,

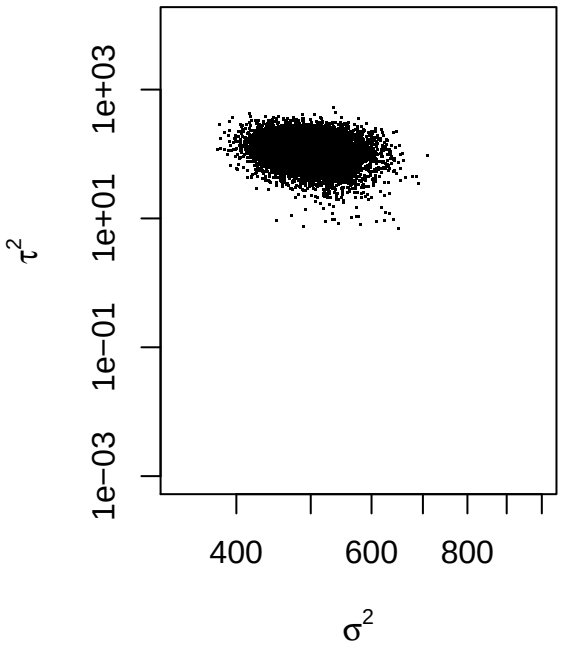

Figure 2.3: Gibbs sampler, variance parametrization, flat prior distribution. Axes have logarithmic scales.

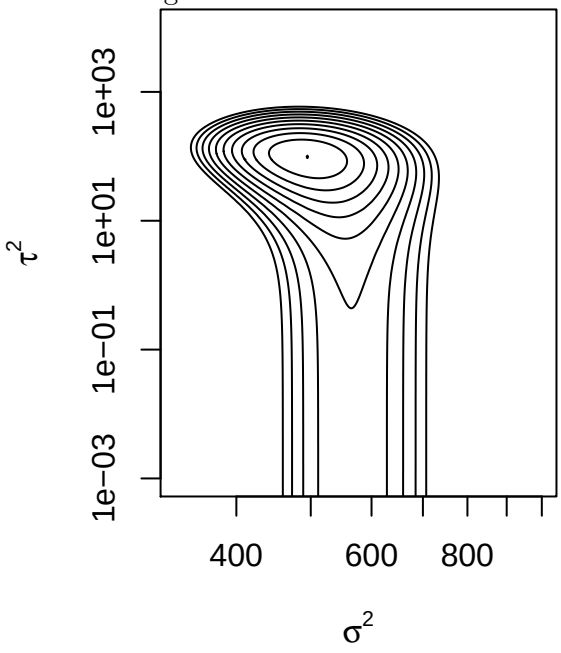

Figure 2.4: Contour plot, log restricted likelihood, variance parametrization, flat prior distribution. Contours have 1 log increments. Axes have logarithmic scales.

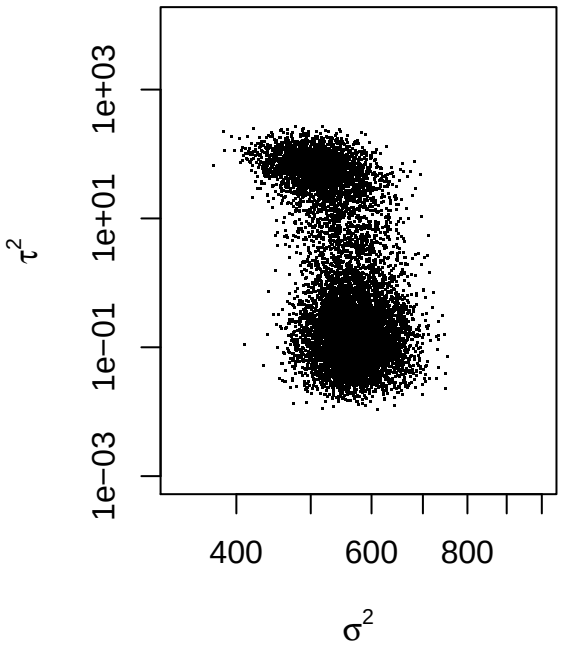

Figure 2.5: Gibbs sampler, variance parametrization, inverse gamma prior distribution. Axes have logarithmic scales.

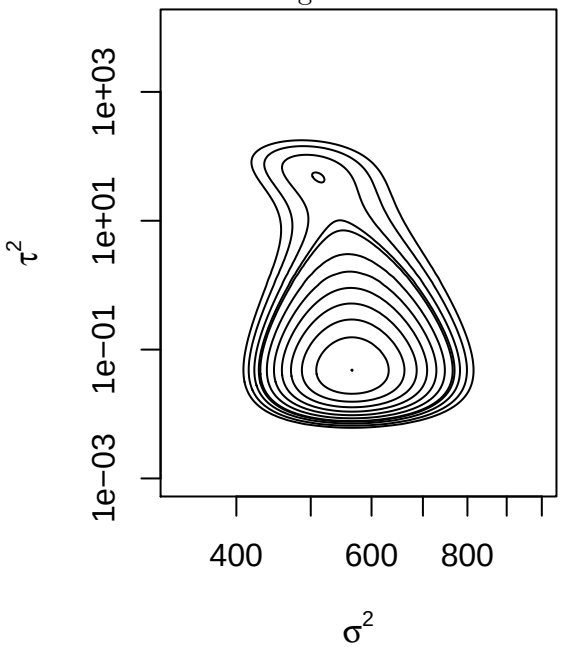

Figure 2.6: Contour plot, log posterior distribution, variance parametrization, inverse gamma prior distribution. Contours have 1 log increments. Axes have logarithmic scales.

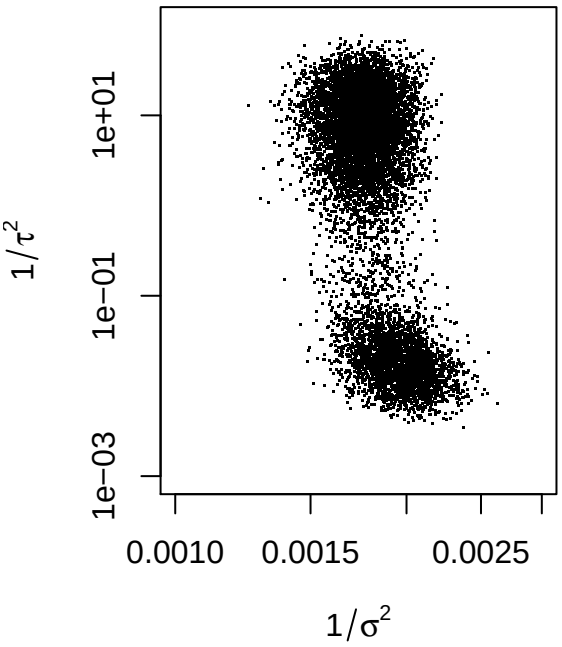

Figure 2.7: Gibbs sampler, posterior distribution, precision parametrization, gamma prior distribution. Axes have logarithmic scales.

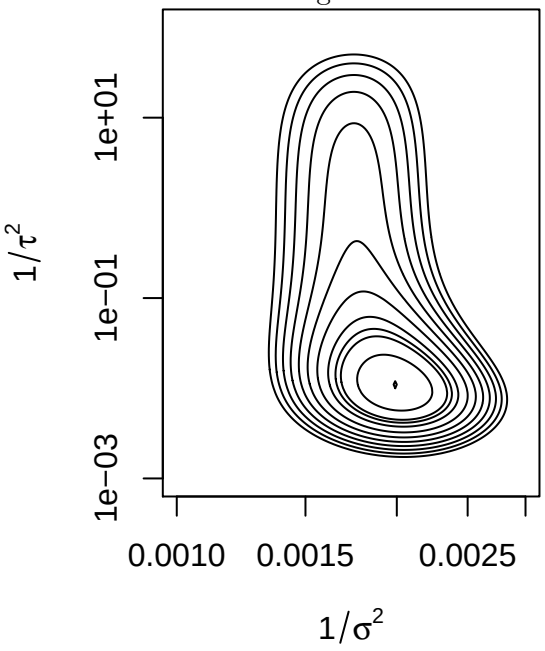

Figure 2.8: Contour plot, log posterior distribution, precision parametrization, gamma prior distribution. Contours have 1 log increments. Axes have logarithmic scales.

the maximum created by the prior distribution is the global maximum. This is true both for the Inverse Gamma(1.1, 0.1) distribution used by Wakefield (1998) and for the popular Inverse Gamma(0.001, 0.001) distribution (not shown). For both of these prior distributions, this second peak is taller but narrower, so it encompasses less volume than the peak arising from the restricted likelihood. A Gibbs sampler would mostly draw samples near the latter peak. This tendency, while agreeing with the data, might provide conditions under which the second maximum could easily be missed. Second maxima in posterior distributions therefore may be more common than reports in the literature would suggest.

Figures 2.5 through 2.8 show results from alternate parameterizations of the same problem using prior distributions Gamma(1.1, 0.1) and Inverse Gamma(1.1, 0.1) that are identical apart from the change of variables. The posterior distribution in Figures 2.5 and 2.6 have two local maxima, but in Figures 2.7 and 2.8, there is only one. When reparameterizing the density from variance to precision using the inverse gamma/gamma pairing, without changing anything else, the second local maximum, created by the prior distribution, spreads out over a wide range and become a shoulder rather than a local maximum.

The same phenomenon can appear in the reverse direction. Figure 2.9 depicts the posterior distribution for precisions. The gamma prior distribution on the error precision has parameters $a = 3$, $b = 1$; two local maxima are apparent. But if we simply change variables from precision to variance, there is only one local maximum, as shown in Figure 2.10.

2.6 Guidance for Diagnosing Multiple Local Maxima

The phenomenon of multiple maxima is subtle, occurring only under certain combinations of circumstances. No simple heuristic will identify all cases that display multiple maxima. Indeed, there is no known method for identifying multiple maxima. We offer some suggestions based on this paper's findings. Propitious strategies include the following.

1. In the case of a posterior distribution, use at least two sets of starting values for the Markov chain Monte Carlo algorithm. Potential candidates include the mode

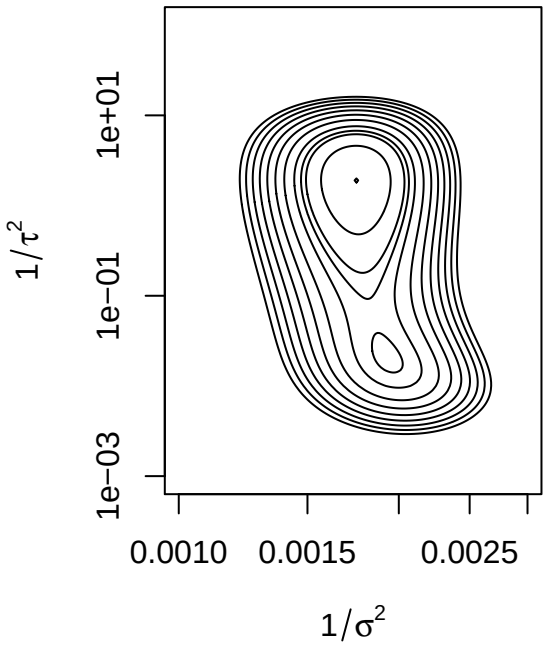

Figure 2.9: Precision parameterization, HMO data. Prior distribution for error precision has $a = 3$, $b = 1$. Contours have 1 log increments. Axes have logarithmic scales.

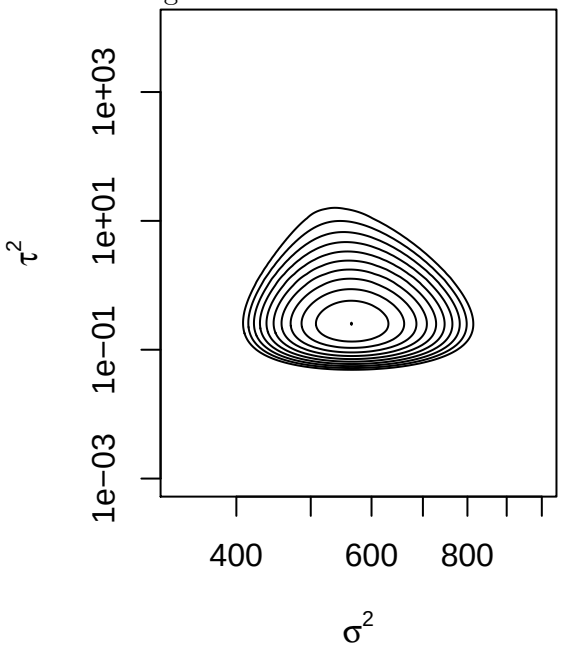

Figure 2.10: Variance parameterization, HMO data. Prior distribution for error variance has $a = 3$, $b = 1$. Contours have 1 log increments. Axes have logarithmic scales.

of the prior distribution and the mode of the function resulting when the actual prior distribution is replaced by a flat prior distribution.

2. In the case of the restricted likelihood, select somewhat extremely high and low values for each variance parameter and consider all the high-low combinations of the variance parameters as starting points. If this results in an excessive number of cases relative to the computational resources, use a fractional factorial design with a fraction small enough to reduce the number of starting values considered among all variance parameters to be less than 10. The idea is to start in very disparate locations in the parameter space.

3. Also in the case of the restricted likelihood, use starting values including the solution indicated by the free terms of the GLM formulation as well as one or more based on the mixed terms.

4. If feasible, run the analysis both for variance parameterization and precision parameterization.

2.7 Discussion

The literature has few results regarding multiple local maxima, particularly for mixed linear models. While many authors have reported multiple local maxima in posterior distributions, we have found only one precedent of multiple maxima in a restricted likelihood.

This work introduces the unifying idea of representing a restricted likelihood as a generalized linear model with a gamma distribution and identity link and an informative "prior" distribution for the error variance. The two-variance model and other more general models fit this form. The "data" in the generalized linear model can be made to conflict with the free terms, which have the form of a prior distribution on the error variance, or with an explicit prior distribution on the other variance component to produce two local maxima. We demonstrated this phenomenon with an ICAR model. The fact that the free terms for the error variance can also be viewed as part of the likelihood from a generalized linear model with gamma errors and identity link shows

that the GLM portion itself can have multiple maxima. That creates the possibility that restricted likelihoods in this class could have more than two maxima.

When a posterior distribution has multiple local maxima, the global maximum need not encompass more volume than a secondary maximum. In that case, a Gibbs sampler could draw more samples near the secondary local maximum. Moreover, when the density is reparameterized from variance to precision, the posterior distribution can change from bimodal to unimodal or vice versa. Simulated or parallel tempering algorithms also can mix slowly in the presence of multiple modes.

These findings underscore the need to choose data models and prior distributions carefully and to examine resulting posterior distributions critically. Otherwise, modelers may be misled by peak-finding algorithms.

Chapter 3

Alternate Likelihood-Based Inference for Gaussian Copula Regression Models of Non-continuous Data

This chapter is organized as follows. Section 3.1 describes motivation for this work. Section 3.2 presents Gaussian copula regression models and the forms of the likelihood for various types of outcome. Section 3.3 describes the three approaches to inference with Gaussian copula regression models for the non-continuous outcomes under consideration. Section 3.4 describes the simulation studies, the results of which appear in Section 3.5. Section 3.6 describes the application of these techniques to real data sets. Section 3.8 concludes.

3.1 Motivation for Alternate Likelihood-Based Inference

Mutually dependent data arise seemingly everywhere, from agriculture, economics, social science, medicine, finance, health services research, transport studies, geoscience, insurance, and dentistry. The copula regression model deftly handles them all with one straightforward structure. Copula-based models offer flexibility because they define

separately the specifications of the dependence structure and the marginal distribution of the responses. This malleability facilitates any combination of response types, continuous and non-continuous. Interpretation of the regression coefficients is natural as well—the same as it would be were the given marginal model fit treating the responses as uncorrelated. Use of a Gaussian copula in particular has well-known properties and provides a familiar and intuitive framework for the dependence structure. These features make the Gaussian copula regression model attractive compared to alternatives such as generalized estimating equations. The latter has difficulty in, for example, incorporating multiple outcomes of mixed type and moreover does not provide a fully parametric model (Song *and others*, 2009).

Song (2000) is the foundation for the Gaussian copula regression approach, with Song *and others* (2009) following in the same vein. Kazianka and Pilz (2010) use a Gaussian copula for geostatistical modeling. All these works generate the multivariate distribution for non-continuous data by considering the Radon–Nikodym derivative of the copula with respect to the counting measure. This approach results in a nested sum of a total of 2^n terms, where n is the number of observations. For all but the smallest data sets, the likelihood can be computationally infeasible.

To circumvent this limitation without resorting to a latent variable assumption, there is a need for alternate likelihood-based inference for use with larger data sets in Gaussian copula regression models with non-continuous responses. We apply three such approaches in a Bayesian framework to three settings using both real and simulated data. We use frequentist methods to evaluate the inference resulting from this work. The approaches are continuous extension, distributional transform with curvature correction, and composite marginal likelihood with curvature correction. The settings are a spatial model with Poisson marginal distributions, a model of incident disease counts over time using the negative binomial distribution, and a longitudinal model with Bernoulli marginal distributions. The marginal distributions were chosen to represent nearly all types of commonly used non-continuous data. The frequentist metrics are the proportion of highest posterior density credible sets that cover the true parameter and the Type II error rate, the proportion of such sets that cover zero. This effort is the first, to our knowledge, to consider these approaches to inference for parameters in Gaussian copula regression models for non-continuous data in a Bayesian framework.

3.2 Gaussian Copula Regression Models

A copula is a distribution function for the joint distribution of several random variables with standard uniform marginal distributions. Marginal distributions can then be incorporated into the model using the probability integral transform, *i.e.*, for a random variable Y with a continuous distribution function $\mathcal{F}$, the variable $U = \mathcal{F}(Y) \sim \mathcal{U}(0,1)$, the standard uniform distribution (see, for example, Casella and Berger (2002)). The copula relates several variables thusly transformed using the dependence structure defined by the copula.

A Gaussian copula specifically has the form

$$\mathcal{C}_V(u) = \Phi_V\left(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_n)\right), \tag{3.1}$$

where Φ denotes the univariate standard normal distribution function and Φ_V denotes the distribution function of a multivariate normal random variable with mean zero and covariance matrix V (Madsen and Fang, 2011). The variables u_i follow the uniform distribution. In most cases, V must be a correlation matrix to ensure identifiability.

For continuous marginal distributions, if Θ is a vector of unknown parameters appearing in the marginal distribution functions $\mathcal{F}_i$ for the observations $y_i, i = 1 \ldots n$, the likelihood can be expressed, per Song *and others* (2009), as

$$L(\Theta \mid y) \propto c\left(\mathcal{F}_1(y_1), \ldots, \mathcal{F}_n(y_n)\right) \prod_{i=1}^{n} f_i(y_i), \tag{3.2}$$

where $c(\cdot)$ is the density corresponding to the copula $\mathcal{C}(\cdot)$, *i.e.*,

$$c(u) = \frac{\partial^n \mathcal{C}(u)}{(\partial u_1 \ldots \partial u_n)},$$

and the f_i are densities corresponding to the $\mathcal{F}_i$.

If the marginal distributions are not continuous, the uniqueness of the copula is not assured, although a unique representation can be determined on the Cartesian product of the ranges of the marginal distributions, $\mathcal{R}(\mathcal{F}_1) \times \cdots \times \mathcal{R}(\mathcal{F}_n)$. The likelihood in this case must be expressed as follows:

$$L(\theta \mid y) = \sum_{j_1=0}^{1} \cdots \sum_{j_n=0}^{1} (-1)^l \mathcal{C}(u_{1j_1}, \ldots, u_{nj_n}), \text{ for } l = \sum_{i=1}^{n} j_i \tag{3.3}$$

(Song *and others*, 2009). In the above, $u_{i0} = \mathcal{F}_i(y_i)$ is the cumulative distribution function value for the observed datum, and $u_{i1} = \lim_{y \nearrow y_i} \mathcal{F}_i(y) = \mathcal{F}_i\left(y_i^-\right)$ is the limit of the cumulative distribution function as its argument approaches the datum from below.

Evaluating the summation in (3.3) can be infeasible for all but the smallest data sets because the number of summands grows exponentially with sample size. Also, the multivariate normal distribution function can be numerically unstable in high dimensions. In this work, we evaluate in detail three viable alternative approaches to likelihood-based inference for non-continuous observations.

3.3 Bayesian Inference

In this section we outline the three approaches for inference in a Bayesian framework for copula regressions with non-continuous outcomes. The continuous extension and distributional transform have the advantage of enabling the use of (3.2) to form the likelihood, thus averting the computational limitations of (3.3). This shared feature makes them both suitable for use in the copula regression models considered here. Each of these two approaches, along with the third approach, the composite likelihood, has attributes that confer an advantage over the other approaches for certain kinds of models.

3.3.1 Continuous Extension

Denuit and Lambert (2005) proposed the continuous extension. It associates a variable that takes integer values with a continuous variable. The non-continuous variable is "extended" by subtracting from it a continuous random variable with support on the unit interval, which Denuit and Lambert (2005) called a perturbation. By its nature, the associated continuous variable can make use of (3.2) for its likelihood.

For a random variable Y taking values on the set of non-negative integers, define

$$Y_i^* = Y_i - U_i, \tag{3.4}$$

where U_i is a continuous random variable on $(0, 1)$. Most commonly, $U_i \sim \mathcal{U}(0, 1)$ and

is independent of Y_i. Then Y_i^* is continuous with distribution function

$$\mathcal{F}_i^*(y) = \mathcal{F}_i(\lfloor y \rfloor) + (y - \lfloor y \rfloor) f_i(\lfloor y + 1 \rfloor). \tag{3.5}$$

The corresponding density is

$$f_i^*(y) = \mathbb{P}(Y_i = \lfloor y + 1 \rfloor). \tag{3.6}$$

In the above, $\lfloor y \rfloor$ is the largest integer not greater than y. Madsen (2009) further noted that, because of (3.6), parameters of the extended variables' marginal distributions are the same as those of the marginal distributions of the original non-continuous variables. This fact makes it a suitable alternative for inference in models for non-continuous data.

Because Y_i^* depends on the extending variable U_i, the likelihood formed with Y^* necessarily depends on U. The exact likelihood of Y_i is the expectation of the likelihood for Y_i^* with respect to U, *i.e.*,

$$L_{CE}(\Theta \mid y) \propto \mathbb{E}_U \left(c(\mathcal{F}_1^*(y_1^*), \ldots, \mathcal{F}_n^*(y_n^*)) \prod_{i=1}^{n} f_i(y_i) \right). \tag{3.7}$$

The fact that (3.7) is the exact likelihood makes this approach a natural standard of comparison for the others. To estimate (3.7), Madsen and Fang (2011) drew m values from the uniform distribution for each observation Y_i. The resulting U_i, termed a "jitter vector", is a vector of length m, and thus Y_i^* also is of length m. Averaging the m estimates of (3.7), one per jitter draw, approximates the expectation at Y_i and is more efficient than computing an integral. The jitter vector must be sufficiently large to approximate adequately the expectation. A jitter vector that is too long, however, taxes computational resources.

3.3.2 Distributional Transform

As Rüschendorf (2009) observed, the distributional transform allows us to treat general distributions with non-continuous components in largely the same way as distributions for continuous variables. In particular, we can use (3.2) to form the likelihood for a non-continuous variable whose distribution is thusly transformed. The expression for the distributional transform, shown below, can be seen as "extending" the distribution function of a non-continuous variable to one that covers the entire range of [0,1] in a

manner analogous to the way that the continuous extension "extends" a non-continuous variable to one that can assume non-integer values. Because the transformed distribution function is continuous over [0,1], it too can be used in (3.2).

Specifically, the marginal distribution of a non-continuous variable Y can be transformed as follows. If Y is non-continuous and $Y \sim \mathcal{F}$, let $U \sim \mathcal{U}(0,1)$ independently of Y. Then the distributional transform is

$$\mathcal{G}(U,Y) = (1-U)\,\mathcal{F}\left(Y^{-}\right) + U\,\mathcal{F}(Y). \tag{3.8}$$

The distributional transform has two useful properties. The first is that $\mathcal{G}(U,Y) \sim \mathcal{U}(0,1)$. This fact enables the use of the distributional transform with a copula. The second property is that $\mathcal{F}^{-1}(B = \mathcal{G}(U,Y)) = Y$ (Rüschendorf, 2009). The implication of this property is that parameters estimated with the distributional transform are consistent. As with the continuous extension, U is a nuisance parameter, removed by taking the expectation. For $i = 1, \ldots, n$,

$$u_i = \mathbb{E}\left(\mathcal{G}_i(U_i, Y_i) \mid Y_i = y_i\right) = \left(\mathcal{F}_i\left(y_i^{-}\right) + \mathcal{F}_i(y_i)\right)/2 = \left(u_{i1} + u_{i0}\right)/2. \tag{3.9}$$

In the above, $u_{i0} = \mathcal{F}_i(y_i)$ is the cumulative distribution function value for the observed datum, and $u_{i1} = \lim_{y \nearrow y_i} \mathcal{F}_i(y) = \mathcal{F}_i\left(y_i^{-}\right)$ is the cumulative distribution function value for the limit from below for the datum.

The resulting approximate likelihood using the distributional transform has a form analogous to that of (3.2):

$$\begin{aligned} L_{DT}(\Theta \mid y) &= c\left(\mathbb{E}\left(\mathcal{G}_1(U_1, Y_1) \mid Y_1 = y_1\right), \ldots, \mathbb{E}\left(\mathcal{G}_n(U_n, Y_n) \mid Y_n = y_n\right)\right) \prod_{i=1}^{n} f_i(y_i), \\ &= c(u_1, \ldots, u_n) \prod_{i=1}^{n} f_i(y_i). \end{aligned} \tag{3.10}$$

This transform requires evaluation of the distribution function for far fewer terms than either the continuous extension or the composite likelihood. It therefore runs considerably faster than the other two approaches, making it especially attractive for use in large problems.

3.3.3 Composite Likelihood

Lindsay (1988) described the composite likelihood as a "likelihood type object formed by

adding together individual component log likelihoods". The component log-likelihoods are formed for subsets of the full observation vector. Most commonly, pairs of observations are used. Because the composite likelihood uses subsets of the observation vector, the elements of the score function associated with the composite likelihood are valid score functions and, under regularity conditions listed in, *e.g.*, Molenberghs and Verbeke (2005), define unbiased estimating equations for the true parameter values. The resulting parameter estimate is thus consistent. Lindsay (1988) also provided a formula for the asymptotic variance of the estimator. The approach is analogous to that of generalized estimating equations.

That the composite likelihood uses subsets of the observation vector means it is misspecified in comparison to the exact, full likelihood. To compensate for this deficiency, Chandler and Bate (2007), in the context of independence estimating equations, proposed a curvature correction to composite likelihood, which we describe in Subsection 3.3.4. We apply composite likelihood with curvature correction to Bayesian inference, following Ribatet *and others* (2012). In this sense, composite likelihood with curvature correction can be considered an approximate likelihood.

For a set of marginal or conditional events $\{\mathcal{A}_e : e \in I\}$, $I \subset \mathbb{N}$, the general form of the composite likelihood is a weighted product:

$$L_{\mathrm{CL}}(\theta; y) = \prod_{e \in I} f(y \in \mathcal{A}_e; \theta)^{z_e} \tag{3.11}$$

for weights z_e and densities $f(y; \theta)$. As with most applications in the literature, we use pairwise Gaussian copulas to construct a composite likelihood. The (gh) elements of the overall Gaussian copula correlation matrix V form the correlation matrix for each pair (gh); we denote this matrix V^{gh}. The composite log-likelihood is then

$$\ell_{\mathrm{CL}}(\theta \mid y) = \sum_{\substack{g \in \{1,\ldots,n-1\} \\ h \in \{g+1,\ldots,n\}}} \log \left(\sum_{j_1=0}^{1} \sum_{j_2=0}^{1} (-1)^l \Phi_{V^{gh}}(u_{gj_1}, u_{hj_2}) \right), \tag{3.12}$$

for $l = j_1 + j_2$ (Song *and others*, 2009). In the above, $u_{g0} = \mathcal{F}_g(y_g)$, the cumulative distribution function value for the observed datum, and $u_{g1} = \lim_{y \nearrow y_g} \mathcal{F}_g(y) = \mathcal{F}_g(y_g^-)$ is the cumulative distribution function value for the limit from below for the datum.

For this approach to be valid, any posterior distribution resulting from this application of the composite likelihood with a chosen prior distribution must be integrable. As

we use proper prior distributions in this analysis, it is sufficient, following the argument set forth by Ribatet *and others* (2012), that for each term in the composite likelihood (3.11) there exists a finite b_e such that $\sup_\theta f\left(y \in \mathcal{A}_e; \theta\right) \leq b_e$. Then

$$\int L_{\mathrm{CL}}\left(\theta; y\right) \pi\left(\theta\right) d\theta = \int \prod_{e \in I} f\left(y \in \mathcal{A}_e; \theta\right)^{z_e} \pi\left(\theta\right) d\theta \leq \prod_{e \in I} b_e^{z_e} < \infty.$$

Inspection of (3.12) reveals that $b_e = 1$ is a suitable bound. The posterior distribution thus is well-defined and can be used in the Bayesian analysis described below.

Ribatet *and others* (2012) also addressed the question of convergence of the MCMC algorithm when the composite likelihood is used in this way. We return this point in Subsection 3.3.4.

Even for modest sample sizes, (3.12) has the largest number of terms to evaluate among the approaches considered here, making it the slowest. But in some cases, it may be the only viable option. For the continuous extension, Madsen and Fang (2011) referenced informal studies suggesting that a smaller jitter vector length m could be used when the sample size was large or when dependence was weak, implying the continuous extension would struggle with binary data in modest sample sizes and jitter vector lengths. But long jitter vectors lead to slow computational times, undermining the approach. Kazianka (2013) found that the distributional transform performed poorly with binary data unless a large sample of observations is available. These limitations are concordant with our preliminary work with the continuous extension in the longitudinal setting. For binary data of modest sample size, among the approaches we considered, only composite likelihood with curvature correction (described below in Subsection 3.3.4) produces suitable parameter estimates.

3.3.4 Curvature Correction

The pairwise composite likelihood is mis-specified in comparison to the exact likelihood. It considers the pairings that make up the terms of the composite likelihood as events independent of one another when, in general, they may not be. Correlation between variables implies similarity in observed values. Treating the pairings as independent fails to account for those similarities. Doing so tacitly regards the data as more informative than they actually are, resulting in estimates of variance that likely understate the

uncertainty of some parameters.

Kazianka and Pilz (2010) observed that the distributional transforms works well so long as the marginal variance is not too small, *i.e.*, when the distribution function is not too coarse. But if the marginal distribution is quite coarse, the distributional transform can be mis-specified as well.

To mitigate the possibility of overly optimistic inference, we apply a curvature correction both to the composite log-likelihood and the distributional transform. The correction, described by Chandler and Bate (2007), approximates the quadratic component of the exact log-likelihood by evaluating, in this study, either the composite log-likelihood or the distributional transform approximation, at a proxy value for the parameter vector; $\ell_{\text{curv}}(\theta; y) = \ell_{\text{CL}}(\theta_{\text{adj}}; y)$ or $\ell_{\text{curv}}(\theta; y) = \ell_{\text{DT}}(\theta_{\text{adj}}; y)$. The proxy for θ is determined through the relation $\theta_{\text{adj}} = \hat{\theta}_{\text{CL}} + \mathcal{B}\left(\theta - \hat{\theta}_{\text{CL}}\right)$ (Chandler and Bate, 2007), with a $\mathcal{B}$ matrix to be described presently and with $\hat{\theta}_{\text{CL}}$ being the estimate resulting from evaluating the composite log-likelihood. (In the case of the distributional transform, $\hat{\theta}_{\text{DT}}$ takes the place of $\hat{\theta}_{\text{CL}}$.)

The matrix $\mathcal{B}$ ensures the expected Hessian of the modified parameter vector equals the Godambe information of the original vector, that is to say,

$$\begin{aligned} \mathcal{B} &= \mathcal{M}^{-1}\mathcal{M}_A, \text{ where} \\ \mathcal{M}_A^T\mathcal{M}_A &= -\mathcal{H}(\theta_0)\,\mathcal{J}(\theta_0)^{-1}\,\mathcal{H}(\theta_0) \text{ and} \\ \mathcal{M}^T\mathcal{M} &= \mathcal{H}(\theta_0). \end{aligned} \tag{3.13}$$

In (3.13), $\mathcal{H}(\theta_0)$ is the Hessian of the original parameter vector for either $\ell_{\text{CL}}(\theta_0; y)$ or $\ell_{\text{DT}}(\theta_0; y)$ and $\mathcal{J}(\theta_0)$ is the variance of the score function for the original parameter vector, again evaluated with either $\ell_{\text{CL}}(\theta_0; y)$ or $\ell_{\text{DT}}(\theta_0; y)$. The technique thus effectively modifies the information of the original parameter vector, incorporating into the evaluation of $\ell_{\text{CL}}(\theta_0; y)$ or $\ell_{\text{DT}}(\theta_0; y)$ the robust variance estimate needed in the presence of clustering.

In practice, θ_0 is unknown. An estimate of the matrix $\mathcal{H}(\theta_0)$ is $\mathcal{H}\left(\hat{\theta}_{\text{CL}}\right)$ for composite log-likelihood, $\mathcal{H}\left(\hat{\theta}_{\text{DT}}\right)$ for distributional transform. The matrix $\mathcal{J}(\theta_0)$ can be

approximated by $\mathcal{J}\left(\hat{\theta}_{\mathrm{CL}}\right)$ or $\mathcal{J}\left(\hat{\theta}_{\mathrm{DT}}\right)$ as the case may be, which we estimate via parallel parametric bootstrap. As Chandler and Bate (2007) noted, the matrix square roots $\mathcal{M}_A$ and $\mathcal{M}$ are not unique. We use the singular value decomposition to calculate $\mathcal{M}_A$ and $\mathcal{M}$ to preserve any asymmetry that might be present.

Ribatet *and others* (2012) also noted an MCMC algorithm based on the curvature-corrected composite likelihood must be shown to converge to the correct target distribution. They argued that the conditions for Theorem 7.2 of Robert and Casella (2004) that establish detailed balance for the Metropolis-Hastings algorithm are valid for applying the algorithm to the composite likelihood with curvature correction. The proposal distribution fits the same criteria as (7.3) of Robert and Casella (2004). Theorem 7.2 of Robert and Casella (2004) builds upon Theorem 6.46 of that work and requires the target to be a distribution. As established in Subsection 3.3.3, while the composite likelihood is not the exact likelihood, it nevertheless is integrable and thus the posterior distribution is well-defined.

The same argument using Theorem 7.2 of Robert and Casella (2004) can be made for the Metropolis-Hastings algorithm applied to the distributional transform with curvature correction.

3.4 Settings for Simulation Studies

The continuous extension, distributional transform with curvature correction, and composite likelihood with curvature correction were evaluated in three settings. Those settings are a spatial copula model called "copCAR" (Hughes, 2015) with Poisson marginal distributions, a model of incident disease counts over time using the negative binomial distribution, and a longitudinal setting with Bernoulli marginal distributions. For the copCAR and negative binomial settings, the distributional transform with curvature correction was applied and results were compared to those from the continuous extension. For the longitudinal setting, as noted in subsection 3.3.3, neither of these methods performed satisfactorily. For example, point estimates for the correlation are biased toward zero. Instead, the composite likelihood with curvature correction was applied. Section 3.5 describes the results of these simulation studies. Additionally, Section 3.6 describes the results of applying these methods to the real data sets that motivated

these simulation settings. Descriptions of these settings appear here; details regarding the Markov Chain Monte Carlo setup appear in Appendix A.

3.4.1 The copCAR Spatial Model

The copCAR model (Hughes, 2015) combines a Gaussian copula with a Conditionally-Autoregressive (CAR) areal spatial model. Spatial models mathematically relate observations that are proximal to one another; doing so ensures smooth, continuous transition along the surface of interest. An areal model describes the association between data summarized over areas, with spatial association represented as similarity of adjacent areas. A CAR model defines the distribution of an observation as a function of its neighbors

The distribution of an observation Y_g, conditional on the observations of its neighbors, takes the form

$$Y_g \mid y_h, g \neq h \sim \mathrm{N}\left(\sum_h \mathcal{H}_{gh} y_h, \tau_g^2\right), g = 1, \ldots, n \tag{3.14}$$

(Banerjee *and others*, 2004). Algebraic manipulation and a technical lemma transform (3.14) into the corresponding joint distribution, having the form

$$p\left(y_1, \ldots, y_n\right) \propto \exp\left\{-\frac{1}{2\tau^2} \sum_{g \neq h} y^T \left(D - \xi W\right) y\right\}. \tag{3.15}$$

In a CAR model, the matrix $Q = D - \xi W$ describes the association between pairs of observations g and h. The diagonal matrix D has the gth diagonal term equal to the number of neighbors that observation g has, with ξ describing the strength of association between observations g and h. The parameter ξ ensures the precision matrix is invertible. Finally, W is a proximity matrix, with diagonal elements zero and ghth elements equal to 1 if observations g and h are neighbors. The form of Q ensures only neighboring observations are associated.

The parameter $\xi \in [0, 1)$, called the propriety parameter, is a range parameter and ensures Q is non-singular. Writing Q as $D - \xi W$, where D consists of the diagonal terms of Q and W includes the off-diagonal terms, $\xi \in \left(1/\lambda_{(\max)}, 1/\lambda_{(\min)}\right)$, where $\lambda_{(\min)}$

and $\lambda_{(\max)}$ are the minimum and maximum eigenvalues, respectively, of $D^{-1/2}WD^{-1/2}$ (Banerjee *and others*, 2004).

In the copCAR model, we have observations $Y = (Y_1, \ldots, Y_n)^T$ and marginal distribution functions $\mathcal{F}_1, \ldots, \mathcal{F}_n$, with $u_i = \mathcal{F}_i(Y_i)$. The Gaussian copula, (3.1), for the CAR then becomes:

$$\mathcal{C}_{Q^{-1}}(u) = \Phi_{Q^{-1}}\left(\Phi^{-1}_{\sigma_1^2}(u_1), \ldots, \Phi^{-1}_{\sigma_n^2}(u_n)\right), \tag{3.16}$$

with copula density

$$c_{Q^{-1}}(u) \propto |Q|^{1/2} |\Sigma|^{1/2} \exp\left(-\frac{1}{2} r^T \left(Q - \Sigma^{-1}\right) r\right), \tag{3.17}$$

where $r_i = \Phi^{-1}_{\sigma_i^2}(u_i) = \Phi^{-1}_{\sigma_i^2}(\mathcal{F}_i(Y_i))$ and $\Sigma = \operatorname{diag}\left(\sigma_1^2, \ldots, \sigma_n^2\right)$. The variances in (3.16), the diagonal terms of Σ, can be approximated efficiently as described in Hughes (2015), making it unnecessary to compute Q^{-1}.

The spatial association among the observables is thus described using the precision matrix Q. Their marginal distributions can be described flexibly through the distribution functions $\mathcal{F}_i$, and they are linked by the Gaussian copula. By eliminating the random effect by means of which CAR models are usually included in a model, copCAR eliminates any collinearity between the spatial random effect and fixed effects, which can create spatial confounding (*e.g.*, Reich *and others* (2006)).

3.4.2 Negative Binomial Count Time Series

The negative binomial model generalizes the Poisson model when the observations are counts exhibiting overdispersion. We parametrize the negative binomial distribution as follows:

$$\begin{aligned} \mathcal{P}(Y_i = y) &= \frac{\Gamma(y + \mu_i)}{\Gamma(\mu_i)\Gamma(y+1)} \left(\frac{\mu_i}{k + \mu_i}\right)^{\mu_i} \left(\frac{\mu_i}{k}\right)^{y}, \\ \text{with } \mathbb{E}(Y) = \mu &= \exp(X\beta), \text{ and } \operatorname{var}(Y) = \mu + \frac{\mu^2}{k}. \end{aligned} \tag{3.18}$$

The negative binomial model allows the variance to take on a value different from the mean and thus account for overdispersion.

The copula for this model is (3.1), with copula density

$$c_V(u) \propto |V|^{-1/2} \exp\left(-\frac{1}{2} r^T \left(V^{-1} - I_n\right) r\right), \tag{3.19}$$

where again $r_i = \Phi^{-1}_{\sigma_i^2}(u_i) = \Phi^{-1}_{\sigma_i^2}(\mathcal{F}_i(Y_i))$. The matrix V is block diagonal, with each block represented the association among repeated measurements.

For this class of models, we consider the first-order auto-regressive correlation structure in which the terms in each block have the form $v_{i,gg} = 1$, $v_{i,gh} = \xi^{\mathcal{D}(g,h)}$ for $\xi \in (0,1)$ and $\mathcal{D}(g,h) =$ distance between measurements g and h.

3.4.3 Longitudinal Model

Generally in longitudinal models, $y \mid \beta, V \sim \mathrm{N}(X\beta, V)$. As with the negative binomial model, V is block diagonal, each block representing the association among repeated measurements of the same individual. Often the longitudinal model is written as a mixed linear model, in which case $V = ZGZ^T + R$.

The copula describing the association among the observations is the same as for the negative binomial data, (3.19) above. The association matrix V is the same as well, first-order auto-regressive.

3.5 Simulation Study Results

3.5.1 copCAR Setting

For this study, the marginal distribution was Poisson. We simulated data on a 20×20 square lattice with Poisson rates $\ell_i = \exp(x_i + y_i)$ with x_i and y_i the coordinates of vertex i. We counted as neighbors those observations immediately adjacent to a given point along the x and y axes. Points on the interior thus have four neighbors, and the neighborhood forms the shape of a "plus" sign. Three values of ξ, $\{0.8, 0.975, 0.995\}$, were chosen to represent moderate, strong, and very strong dependence.

Table 3.1 shows the results. Both metrics concern the highest posterior density interval's coverage of the true parameter value or of zero, respectively. For the distributional transform with curvature correction, performance at all levels of dependence is almost universally good. The Type II error rate for the two β parameters for the $\xi = 0.995$ case can be understood to reflect the relatively low level of information available about these two parameters implied by ξ having such a high true value. The good coverage carries with it inflated Type II error rates. The distributional transform technique is by

far the faster technique, an important consideration.

For moderate and strong spatial dependence, the continuous extension has good to excellent performance. At $\xi = 0.975$, the performance for ξ is better than for the fixed effects β. For very strong spatial dependence, the performance falls short. Performance might have been improved with an even longer jitter vector, but computation time with such a vector is prohibitive. The Type II error rate for the two β parameters is elevated for $\xi = 0.995$. While the effect is less pronounced than with distributional transform with curvature correction, that might simply be a reflection of the lower coverage rate; were the coverage rate higher, the Type II error rate might be higher as well.

Table 3.1: Results for copCAR Setting

				Distributional Transform,	
	True	Continuous Extension		Curvature Correction	
Parameter	Value	Coverage	Type II	Coverage	Type II
β_1	1	93.00	0.00	94.00	0.00
β_2	1	93.75	0.00	94.50	0.00
ξ	0.8	95.25	0.00	95.75	0.00
β_1	1	89.50	2.25	95.25	5.00
β_2	1	92.00	2.00	96.25	6.50
ξ	0.975	94.75	0.00	95.25	0.00
β_1	1	89.75	7.75	93.25	16.50
β_2	1	87.75	11.75	94.00	21.25
ξ	0.995	86.00	0.00	93.5	0.00

We also evaluated the L1 norm for the difference in posterior distributions arising from the continuous extension and distributional transform with curvature correction (results not shown). In calculating this metric, we re-centered the posterior distributions resulting from the distributional transform with curvature correction to align them with those from the continuous extension method. Because of this re-centering, L1 norm values measure differences in tails and overall distributional shape.

The L1 norm values for comparisons of fixed effects posterior distributions in cases with moderate association ($\xi = 0.8$) were low. For the association parameter ξ, the L1 norms are elevated. The same is true for all parameters at true association values of 0.975 or 0.995. These findings demonstrate that in most cases, the distributional

transform with curvature correction method gives rise to different posterior distributions for the estimated parameters compared with those arising from the continuous extension method. As noted above, distributional transform with curvature correction is an approximation to the log-likelihood, whereas the estimate under continuous extension approaches the exact form with a sufficiently large jitter vector. This difference in posterior shape is one trade-off in using an approximation to yield a considerably shorter run time.

3.5.2 Negative Binomial Setting

The negative binomial simulation is based on incidence of malaria over a five-year period in two villages in Kenya reported in Noland *and others* (2012). Fixed effects include an intercept, β_0, the year, β_1, treated as a continuous regressor, and two seasonal effects, sine and cosine of the month, β_2 and β_3 respectively, with the period of each wave being one year.

We first fit simple generalized linear models using the R function `glm.nb()` to obtain starting values for the Markov Chain Monte Carlo (MCMC) scheme for continuous extension as well as the variance-covariance matrix used in β's proposal distribution. In the case of distributional transform with curvature correction, `glm.nb()` also supplies the initial parameter estimates to `optim()`, used to compute $\hat{\theta}_{\mathrm{DT}}$ in the curvature correction. We chose a simulated sample size of 240 observations and magnitudes of the true fixed effect sizes to ensure sufficient non-zero entries to stabilize `glm.nb()`.

Tables 3.2 and 3.3 show that the results for the negative binomial setting under the distributional transform with curvature correction are generally good, with coverage at, near, or even exceeding 95% for all parameters.

The Type II error rate for β_3 is very large for both levels of κ and all levels of ξ. As discussed in Subsection 3.4.2, we chose the true values to ensure a large number of non-zero observations to stabilize results from `glm.nb()`. The true value for β_3 is close to zero, and its posterior variance is large relative to the size of the effect. Most confidence intervals therefore span zero. This effect is the cosine term, a higher order effect than either the slope or intercept. Less information is available for it, leading to the higher variance. The variance for this effect is relatively large even in the fit for `glm.nb()` (not shown), which does not account for correlation.

The Type II error rate is high for ξ estimation when its true value is 0.2 for a similar reason—the true value of ξ is close to zero and its posterior variance is large relative to the size of the effect.

The Type II error rate for β_1, the slope parameter for year, also falls short of expectation for the two $\xi = 0.8$ cases, likely affected by the poor β_3 performance. A high association level means effectively less information in the observations on which to estimate parameters. Its Type II error rate increases as ξ increases both for $\kappa = 0.5$ and for $\kappa = 1$. The magnitude of the error is less for the $\kappa = 1$ case compared to the corresponding $\kappa = 0.5$ case. The $\kappa = 0.5$ level is the more adverse of the two. As can be seen in (3.18), the variance is larger for values of κ less than unity. The fact that the Type II error rate for β_1 at $\xi = 0.8$ is worse for $\kappa = 0.5$ than for $\kappa = 1$ reveals the sensitivity of the parameter's estimation to the effective amount of information in the data sets at these levels of κ and ξ.

In the main, performance of the distributional transform with curvature correction is quite similar to that of continuous extension. The distributional transform with curvature correction has better coverage for ξ when its true value is 0.8 and $\kappa = 0.5$. Note that for this case with the continuous extension, we used a sub-optimal length for the jitter vector, as explained in Subsection A.3. The association parameter appears in (3.7) only in the copula term, as does the jitter vector. So the estimation of ξ is especially sensitive to the jitter vector length. The data are not very informative about κ, either, so its estimate also is sensitive to imprecision in the log-likelihood estimation.

As with the copCAR setting, L1 norms mostly were low when comparing posterior distributions of either fixed effects or κ at the low association value of $\xi = 0.2$. Fixed effect L1 norm values were low at $\xi = 0.4$ for $\kappa = 1$ as well. For fixed effects at $\xi = 0.4$ for $\kappa = 0.5$ and at $\xi = 0.8$, the fraction of distribution pairs with low L1 norm values deteriorated. For κ and ξ posterior distributions, L1 norms almost always were elevated regardless of κ or ξ true values. The data convey little information about these parameters compared to the fixed effects and they are therefore harder to estimate precisely. By extension we can expect differences in estimates of posterior distributions for these parameters. The posterior distributions are sufficiently close to one another only for fixed effects at lowest association values, reflecting the relative abundance of information about those parameters at those parameter levels.

Table 3.2: Results for Negative Binomial, $\kappa = 0.5$

Parameter	True Value	Continuous Extension Coverage	Continuous Extension Type II	Distributional Transform, Curvature Correction Coverage	Distributional Transform, Curvature Correction Type II
β_0	3	94.50	0.00	96.25	0.00
β_1	-0.1	94.00	0.00	95.00	0.50
β_2	0	96.00	N/A	97.25	N/A
β_3	0.1	95.00	87.25	95.75	88.25
κ	0.5	93.25	0.00	93.75	0.00
ξ	0.2	95.00	17.00	96.50	20.00
β_0	3	95.00	0.00	96.75	0.00
β_1	-0.1	95.50	0.75	97.25	1.75
β_2	0	95.75	N/A	98.00	N/A
β_3	0.1	93.50	89.00	94.5	90.75
κ	0.5	95.00	0.00	97.75	0.00
ξ	0.4	94.75	0.00	98.00	0.00
β_0	3	89.50	0.00	97.75	0.25
β_1	-0.1	91.75	6.00	97.50	27.00
β_2	0	92.25	N/A	96.00	N/A
β_3	0.1	93.75	90.50	96.75	93.00
κ	0.5	74.50	0.00	93.25	0.00
ξ	0.8	52.00	0.00	95.75	0.00

Thus, with the exception of fixed effects comparisons at low association parameter values, results for the L1 norm of the difference in posterior distributions for parameters estimated with the two approaches show that the posterior distributions arising from the continuous extension method cannot be considered comparable to those arising from the distributional transform with curvature correction method.

3.5.3 Longitudinal Setting

The longitudinal simulation uses fixed effects from the so-called toenail data originally reported by de Backer *and others* (1996), and analyzed by Molenberghs and Verbeke (2005) as well as Madsen and Fang (2011), concerning a trial of two treatments of dermatophyte onychomycocis. Fixed effects include an intercept, treatment indicator,

Table 3.3: Results for Negative Binomial, $k = 1$

Parameter	True Value	Continuous Extension Coverage	Continuous Extension Type II	Distributional Transform, Curvature Correction Coverage	Distributional Transform, Curvature Correction Type II
β_0	3	94.25	0.00	94.5	0.00
β_1	-0.1	94.75	0.00	96.00	0.00
β_2	0	94.25	N/A	95.00	N/A
β_3	0.1	95.00	83.75	95.00	84.25
k	1	96.75	0.00	96.75	0.00
ξ	0.2	94.00	18.25	95.75	18.00
β_0	3	95.25	0.00	96.00	0.00
β_1	-0.1	94.50	0.00	96.50	0.00
β_2	0	93.25	N/A	94.75	N/A
β_3	0.1	95.75	85.25	96.25	86.25
k	1	92.00	0.00	95.00	0.00
ξ	0.4	94.00	0.25	94.00	0.25
β_0	3	90.50	0.00	96.00	0.00
β_1	-0.1	90.75	2.25	95.75	11.00
β_2	0	95.25	N/A	96.75	N/A
β_3	0.1	94.75	82.75	95.5	85.25
k	1	94.00	0.00	96.50	0.00
ξ	0.8	84.25	0.00	98.00	0.00

time in weeks, and time-treatment interaction.

The continuous extension method produces severely biased estimates of ξ. Table 3.4 shows results from maximizing the approximate likelihood from the continuous extension method for the four hundred simulated data sets and jitter vector lengths of 100. Verification studies of jitter vector length, described in detail in Subsection A.3, showed the jitter vector length of 100 to be wholly inadequate; adjusted standard errors were larger than unadjusted standard error by factors as high as 1.75. Yet it was impossible to run longer jitter lengths due to resource constraints. Binary data are the least informative of data types, and correlation parameters require numerous observations to be estimated well in any setting, so this poor performance is not surprising. Because the bias is so severe, the full Bayesian analysis was not attempted for this method in this

setting.

Table 3.4: Estimates for Continuous Extension in Longitudinal Setting, Based on 400 Data Sets

Parameter	True Value	Expected Point Estimate	95% Confidence Interval
β_0	-0.5	-0.5019	(-0.7553, -0.2485)
β_1	-1	-1.0117	(-1.4359, -0.5875)
β_2	0	-9E-04	(-0.0393, 0.0375)
β_3	1	1.0138	(0.8310, 1.1966)
ξ	0.2	0.1314	(0.0793, 0.1834)
β_0	-0.5	-0.4995	(-0.7720, -0.2271)
β_1	-1	-1.0159	(-1.5035, -0.5284)
β_2	0	-0.0017	(-0.041, 0.0377)
β_3	1	1.0075	(0.8063, 1.2087)
ξ	0.4	0.2077	(0.1611, 0.2543)
β_0	-0.5	-0.5174	(-0.8647, -0.1700)
β_1	-1	-0.9616	(-1.5077, -0.4155)
β_2	0	-3E-04	(-0.0408, 0.0402)
β_3	1	0.9895	(0.7964, 1.1827)
ξ	0.8	0.3923	(0.3408, 0.4439)

Table 3.5 shows results for the composite likelihood with curvature correction using the posterior mean as the point estimate. Coverage is excellent.

The Type II error rates for β_0 are elevated. Intercepts too can be difficult to estimate. In this case, the data are not centered, contributing to the difficulty. Moreover, the Type II error rates increase as the correlation increases. A higher correlation implies a smaller effective sample size, detracting from the ability of the method to estimate the parameter.

3.6 Application to Real Data

In this section, we apply the three methods studied in simulation to real data.

Table 3.5: Results for Composite Likelihood with Curvature Correction in the Longitudinal Setting

Parameter	True Value	Composite Likelihood, Curvature Correction Coverage	Type II
β_0	-0.5	94.75	1.25
β_1	1	96.25	0.75
β_2	0	94.25	N/A
β_3	-1	95.00	0.00
ξ	0.2	95.00	4.50
β_0	-0.5	93.50	6.00
β_1	-1	94.50	0.75
β_2	0	94.75	N/A
β_3	1	94.75	0.00
ξ	0.4	94.00	0.00
β_0	-0.5	95.25	14.75
β_1	-1	94.50	2.25
β_2	0	95.00	N/A
β_3	1	95.00	0.00
ξ	0.8	94.00	0.00

3.6.1 Slovenia Stomach Cancer

The first data set describes incidence of stomach cancer by municipality in Slovenia. At the time of the data collection, Slovenia had 192 municipalities, two of which were not contiguous. We considered the non-contiguous zones as separate areas, resulting in 194 spatial areas for analysis. We applied a Poisson model in the copCAR setting. In addition to an intercept, the model adjusts for centered and scaled socioeconomic status, a five-level indicator determined by Slovenia's Institute of Macroeconomic Analysis and Development. We applied all three techniques to these data. Details of the MCMC implementation are described in Appendix A.

All three techniques required 60,000 iterations for convergence. In each analysis, the MCMC standard error was more than two orders of magnitude smaller than the corresponding point estimate.

Results appear in Table 3.6. The continuous extension and distributional transformation with curvature correction techniques both estimated fixed effects that were smaller in absolute value than the corresponding estimate from composite likelihood with curvature correction. All were roughly similar to one another, however. Highest posterior density interval widths were slightly shorter for the first two techniques, but not notably so.

Table 3.6: Parameter Estimates (95% Highest Posterior Density Interval) for Slovenia Stomach Cancer Data by Technique

Technique	β_0	β_1	ξ
Continuous Extension	0.1524	-0.1277	0.2772
	(0.1095, 0.1916)	(-0.1702, -0.0858)	(0.0505, 0.5005)
Distributional Transform,	0.1542	-0.1316	0.2555
Curvature Correction	(0.1132, 0.1992)	(-0.1789, -0.0866)	(0.0500, 0.4330)
Composite Likelihood,	0.1696	-0.1524	0.2436
Curvature Correction	(0.1262, 0.2120)	(-0.2005, -0.1046)	(0.0474, 0.4153)

For estimates of the association parameter, the trend was reversed. Continuous extension and distributional transform with curvature correction produced estimates somewhat higher than that for composite likelihood, though not substantially so. The widths of their highest posterior density intervals were slightly longer than the one from composite likelihood with curvature correction.

Posterior densities for ξ using a random walk proposal exhibited an interesting attribute. The result from the distributional transform with curvature correction provides an example in Figure 3.1. The second peak near zero is an artifact of the proposal technique. It arises as follows. Our MCMC routine proposes values using a Gaussian random walk on the transformed scale, centered initially at the transformed value corresponding to the solution from `optim()`. Near-zero values have reasonable probabilities of being proposed. The data are somewhat ambiguous regarding the true peak, so there is a reasonable chance that the near-zero proposal is accepted. The random walk then recenters the proposal distribution around the transform of this near-zero value. Values of nearly zero for ξ correspond to nearly independent observations; the contribution to

the likelihood from the copula in any of (3.7), (3.10), or (3.12) is negligible. The log-likelihood value estimated for each proposal near zero is thus dominated by the terms from the marginal densities of the observations, and the estimated log-likelihood value in turn dominates the numerator or denominator of the Metropolis-Hastings ratio. As a result, nearly every proposal near zero is accepted.

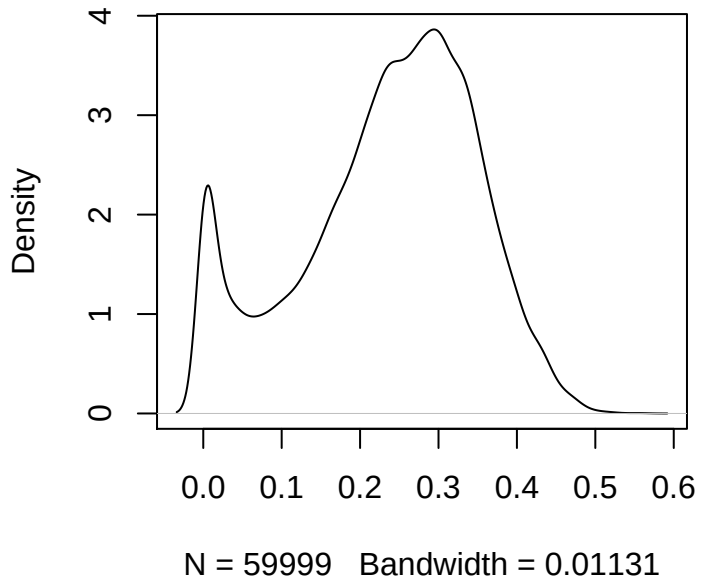

Figure 3.1: Density Estimate for Slovenia Data Set Analyzed with Distributional Transform with Curvature Correction.

When the random walk chain is replaced with an independence chain, a normal distribution centered about the point estimate from `optim()`, this peak near zero is nearly eliminated, confirming that its presence with the random walk proposal is an artifact of that approach. Results in Table 3.6 arose from analyses using this independence proposal distribution, though the estimates and intervals differ only slightly from those resulting from use of the original random walk proposal distribution.

While all three techniques produce similar estimates, distributional transform with

curvature correction has a decisive advantage in its run time, 22 minutes, faster than its nearest competitor, the continuous extension, by more than an order of magnitude. Complete run time comparisons appear in Section 3.7.

3.6.2 Toenail Ringworm Data

We analyzed the toenail data that provided the motivation for the longitudinal study. As noted above, it concerns a trial of two treatments of dermatophyte onychomycocis, a ringworm infection. We used composite likelihood with curvature correction. Fixed effects include an intercept, treatment indicator, time in weeks, and time-treatment interaction. Note that whereas the longitudinal simulation used an auto-regressive association structure, for these data, we use an exchangeable structure for comparison to results from Madsen and Fang (2011), who present results from generalized estimating equations.

Results appear in Table 3.7. Estimates and intervals were quite comparable to those from Madsen and Fang (2011). The highest posterior density intervals were somewhat shorter than the corresponding confidence intervals in the Madsen and Fang (2011) analysis.

Table 3.7: Parameter Estimates (95% Interval) for Toenail Data with results from Madsen and Fang (2011) for comparison. Madsen and Fang (2011) did not report an association parameter.

Parameter	Parameter Estimate (95% Interval)			
	Composite Likelihood, Curvature Correction		Madsen & Fang Analysis (GEE)	
β_0	-0.6302	(-0.9645, -0.3091)	-0.6472	(-1.055, -0.2393)
β_1	0.0010	(-0.4867, 0.4588)	0.0019	(-0.5994, 0.6032)
β_2	-0.2084	(-0.2631, -0.1497)	-0.2166	(-0.3003, -0.1329)
β_3	-0.0320	(-0.1097, 0.0498)	-0.0486	(-0.1834, 0.0862)
ξ	0.7726	(0.7066, 0.8375)	N/A	(N/A)

3.7 Run Times

Table 3.8 shows typical run times for each setting and method; these estimates do not include post-processing time. The distributional transform with curvature correction has far better run times than the other methods in all settings in which it could be used.

Table 3.8: Typical Run Times for Each Setting and Method

Setting	Approach	Typical Run Time
copCAR	Continuous Extension	12 – 74 hours
copCAR	Distributional Transform, Curvature Correction	1.1 hours
Longitudinal	Composite Likelihood, Curvature Correction	30 hours
Negative Binomial	Continuous Extension	20.3 hours
Negative Binomial	Distributional Transform, Curvature Correction	2.9 hours
Slovenia	Continuous Extension	5.5 hours
Slovenia	Distributional Transform, Curvature Correction	22 minutes
Slovenia	Composite Likelihood, Curvature Correction	10.8 hours
Toenail Data	Composite Likelihood, Curvature Correction	17.8 hours

3.8 Discussion

The distributional transform with curvature correction performs quite well on data other than binary. For binary data, composite likelihood is the only viable option because binary data is especially uninformative about the parameters. The substantially faster running times for the distributional transform with curvature correction as compared to other techniques make it quite attractive in suitable settings. This advantage is especially pronounced at high levels of association, in which the continuous extension method requires a long jitter vector. While the continuous extension makes use of

the exact form for the likelihood, resource limitations on implementation mean that in practice, its performance can fall short of the distributional transform with curvature correction. Even if sufficiently long jitter vectors could be used, the resulting posterior distribution may not have nominal coverage. The distributional transform with curvature correction is a viable alternative for parameter inference.

Chapter 4

Dirichlet Process Prior Distribution for Association Parameter Inference

This chapter describes the application of a Dirichlet process as a prior distribution for association parameter inference in Gaussian copula regression models. The setting is a proportional odds model relating T2* mapping data (an investigational, objective MRI sequence) to Beck's scale of cartilage damage, with adjustment for spatial proximity of responses. The clinical data motivating this study is femoroacetabular impingement (FAI), a condition in which subtle deformities of the femoral head and acetabulum (hip socket) result in pathologic abutment during hip motion. The data suggest a high degree of association among the observations. The challenge in this case is that few continuous distributions can be parametrized intuitively such that the majority of its mass is clustered near 1. The Dirichlet process can be parameterized more intuitively and also allows for clustering.

The first section of this chapter describes the clinical data providing the motivation for the analysis. It represents the work of several collaborators, namely John Hughes and Eleena Iisakka in the Division of Biostatistics, School of Public Health; Jutta Ellermann at the Center for Magnetic Resonance Research, Department of Radiology; Shabnam

Mortazavi and Mikko J. Nissi at the Center for Magnetic Resonance Research, Department of Radiology, Department of Orthopaedic Surgery; and Connor Ziegler and Patrick Morgan, Department of Orthopaedic Surgery; all of the University of Minnesota.

The second section describes the analysis of the femoroacetabular impingement (FAI). In Subsection 4.2.1 we present models and results of the statistical analyses. Note that the analysis of the random intercept model of this subsection originally appeared in Iisakka (2014). We repeated the random intercept model analysis, and the result of that replication appears here. The repeat of the analysis was performed in support of the spatial analysis work also described in Subsection 4.2.1. Preliminary spatial work also was performed and described by Iisakka (2014) as noted in the text. The application of the Dirichlet process prior distribution described in Subsection 4.2.1 is novel to this work. In Subsection 4.2.2 we apply the predictive models to unaggregated patient T2* data.

We conclude with a discussion in Section 4.3.

4.1 Clinical Motivation

4.1.1 Introduction

Femoroacetabular impingement (FAI), a common cause of hip pain, is a condition characterized by abnormal peri-articular morphology that results in pathological abutment between the head–neck junction of the femur and the acetabular rim (Ganz *and others*, 2003; Leunig *and others*, 2009). FAI has been shown to cause labral (edge, rim) and chondral (of or relating to cartilage) lesions and is a strong risk factor for osteoarthritis (Beck *and others*, 2005; Dudda *and others*, 2009; Allen *and others*, 2009; Ganz *and others*, 2008). For FAI patients with symptoms unresponsive to non-operative management, joint preservation surgery can be considered if cartilage damage is not significant (Philippon *and others*, 2007, 2009; Beck *and others*, 2004). But joint preservation procedures are contraindicated for patients with moderate to advanced cartilage changes because more severe cartilage abnormality is associated with early conversion to total hip replacement (Larson *and others*, 2011; Guanche *and others*, 2012). Unfortunately, moderate cartilage damage can be challenging to diagnose (Byrd and Jones, 2011). Radiographic evaluation with use of Tönnis grading (assignment of an ordinal score

in $\{0, 1, 2, 3, 4\}$ based on inspection) is the standard of care but has been shown to have poor inter-observer reliability (Clohisy *and others*, 2009; Carlisle *and others*, 2011; Tönnis *and others*, 1987). Magnetic resonance (MR) image evaluation seems a sensible alternative, but the accuracy of MR imaging and MR arthrography for detecting chondral damage in FAI is poor (Byrd and Jones, 2011; Anderson *and others*, 2009; Pfirrmann *and others*, 2008; Blankenbaker *and others*, 2011; Gold *and others*, 2012; Zlatkin *and others*, 2010).

The identification of cartilage damage in FAI may be difficult owing to the pattern of damage particular to the condition (Pfirrmann *and others*, 2006). In FAI, cartilage damage frequently is limited to the acetabulum and may occur initially deep within the tissue as a debonding of articular cartilage from acetabular bone (Beck *and others*, 2005). Since the superficial cartilage remains intact and traditional MR imaging is best suited for revealing damage at the articular surface, this pattern of damage hinders diagnosis. Thus investigators have turned to quantitative MR mapping techniques such as delayed gadolinium-enhanced MR imaging of cartilage (Kim *and others*, 2003; Mamisch *and others*, 2011; Zilkens *and others*, 2010; Lattanzi *and others*, 2012) and T2 mapping (Watanabe *and others*, 2007; Nishii *and others*, 2008, 2010; Mosher and Dardzinski, 2004). Delayed gadolinium-enhanced imaging is the most widely applied technique, but it can be time-consuming and logistically difficult to perform, and, due to limited resolution, currently does not allow segmentation of femoral and acetabular cartilage (Gold *and others*, 2009). Furthermore, use of gadolinium agents is contraindicated in patients with limited renal function. T2 relaxation time measurements (Watanabe *and others*, 2007; Nishii *and others*, 2008, 2010; Mosher and Dardzinski, 2004) and, more recently, T2* mapping have also been reported for the hip (Bittersohl *and others*, 2009, 012a,b; Apprich *and others*, 2012). T2* has the advantages that it (1) can be acquired quickly, (2) does not require contrast material (which must be injected in the joint or intravenously), and (3) has sufficient resolution to differentiate between femoral and acetabular cartilage.

The aim of our study was to determine whether quantitative T2* mapping can be used for routine cartilage assessment in FAI. To do this, we first developed an anatomically precise technique for extracting T2* data. Then we compared extracted and aggregated T2* data with a surgical gold standard. To link T2* measurements to the

cartilage's surgically revealed condition, we proposed two models. One model includes a simple hip-specific adjustment along with the predictors of interest. The other accounts for spatial proximity of the measurements in a given hip by assigning to each hip in the study an association parameter, the magnitude of which corresponds to the strength of influence of the adjacent observations. The tendency of association parameter estimates to cluster around a few values is modeled using a Dirichlet process. The resulting predictive models will allow clinicians to assess quickly the pattern, severity, and extent of damage.

In the remainder of this section, we describe the collection of our T2* data and corresponding arthroscopic data as well as provide some basic demographic information about the data set.

4.1.2 Collection of T2* Data and Corresponding Arthroscopic Data

Patients

We collected our data between February, 2010 and March, 2012. The study group included 26 patients (28 hips) who exhibited clinical and radiographic signs of FAI, showed no evidence of osteoarthritis, underwent the study imaging protocol, were diagnosed with a labral tear, and subsequently underwent hip arthroscopy after conservative treatment failed. The clinical diagnosis of FAI was established by the presence of moderate to severe persistent hip or groin pain that limited activity and worsened with flexion activity, and positive impingement sign (*i.e.*, sudden pain at 90° hip flexion with adduction and internal rotation, or with extension and external rotation). Radiographic confirmation of FAI was based on findings such as α angle (angle between femoral neck and line through center of the head and point where the head becomes aspherical) greater than 50°, pistol grip deformity (nonspherical femoral head), *coxa profunda* (a too-deep socket), and acetabular retroversion (the mouth of the acetabulum inclines posterolaterally). Exclusion criteria included osteoarthritis as evidenced by Tönnis grade > 1 (Tönnis *and others*, 1987), previous hip surgery, or diagnosis of other abnormalities to which the patient's hip pain could be attributed. All patients were examined by Dr. Morgan, who specializes in hip arthroscopy, and evaluated with standardized radiographs per published protocol (Clohisy *and others*, 2008).

There were nineteen female and seven male patients. There were twenty female hips and eight male hips. The mean age for all patients was 29.0 (range 12–53). The mean age for females was 28.9 (range 12–46). The mean age for males was 29.1 (range 16–53). All hips had a Tönnis grade of 0 or 1, indicating at most mild signs of osteoarthritis.

Imaging

A 3T clinical imaging protocol (Trio; Siemens Medical Solutions, Erlangen, Germany) was used. The protocol required approximately 45 minutes to complete, with the T2* data obtained during the final seven minutes to control for time dependence of T2* values after unloading (Apprich *and others*, 2012). T2* maps were generated using software provided by the scanner vendor (Mapit; Siemens Medical Solutions).

Region of interest (ROI) analyses were performed by Dr. Ziegler, a second-year orthopaedic resident, who was blinded to the patients' clinical information. Acetabular orientation was standardized on sagittal images by using a line passing through the center of the femoral head, perpendicular to the transverse acetabular ligament, defining the 12-o'clock position (Figure 4.1). Case regions of interest (ROI) were defined in the anterosuperior acetabulum because this area has the highest reported incidence of damage in patients with FAI (Beck *and others*, 2005; Anderson *and others*, 2009; Larson and Giveans, 2009). Using the image processing application OsiriX (Rosset *and others*, 2004), acetabular cartilage in this region was divided into five ROIs between the 12-o'clock position and the chondrolabral junction (Figure 4.1). This was done for three consecutive sagittal sections, yielding a total of fifteen case ROIs. Four control ROIs were defined in the posteromedial acetabulum, where articular cartilage damage is infrequent in FAI. Note that this landmark-based extraction resulted in ROIs comparable between the patients even though the volume (and number of voxels) varied from patient to patient.

Arthroscopy

To evaluate the utility of T2* for assessing cartilage damage, we needed a reference assessment. For our study, the reference assessment was obtained through arthroscopic surgery. All arthroscopic examinations were performed by Dr. Morgan, who was blinded to the T2* data. Dr. Morgan was presented with a patient-specific, flattened anatomical

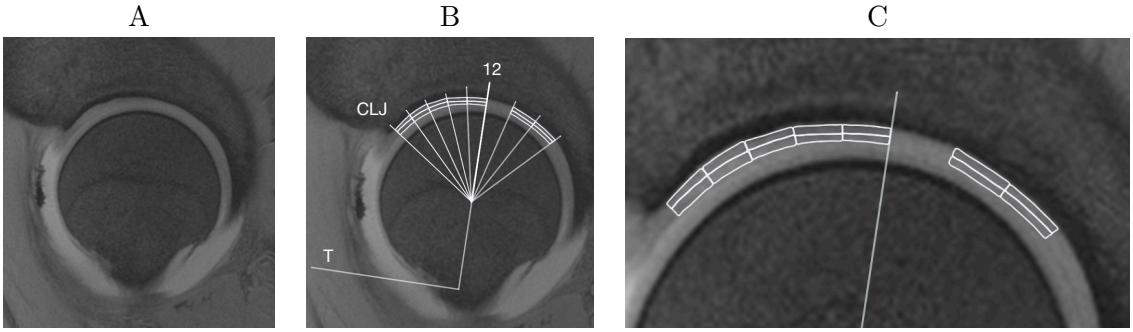

Figure 4.1: Selection of ROIs in acetabulum. (A) Gradient recalled echo image of hip, sagittal view. (B) Transverse ligament (T) used to find 12-o'clock position (12); angular guides for case ROIs in anterior superior labrum between 12-o'clock position and chondrolabral junction (CLJ); and angular guides for control ROIs (clockwise from 12-o'clock). (C) Magnified portion of panel B shows the dark line that delineates the boundary between acetabular and femoral cartilage.

map of the acetabulum (an example of which is shown in Figure 4.2), on which simple, obvious bony landmarks could be co-located during surgery. Once located, individual ROIs were measured relative to a flexible probe measuring 2 mm in diameter, which served as a ruler. Dr. Morgan recorded his surgical findings on the patient-specific acetabular projection. A modified Beck scale (Beck *and others*, 2005) (described in Table 4.1) was used to characterize the degree of articular cartilage damage. Both case and control ROIs were assessed (532 ROIs in all, 19 per hip).

Table 4.1: The modified Beck scale.

Score	Description	Criteria
1	normal	macroscopically sound cartilage
2	early changes	softening; fibrillation; cartilage remains adhered to bone
3	debonding	loss of fixation to bone; "carpet" phenomenon
4	cleavage	loss of fixation to bone; frayed edges; thinning of cartilage; flap
5	defect, fibrous base	full-thickness loss of cartilage; thin fibrous tissue-covered base
6	defect, eburnated base	full-thickness loss of cartilage; base of eburnated bone

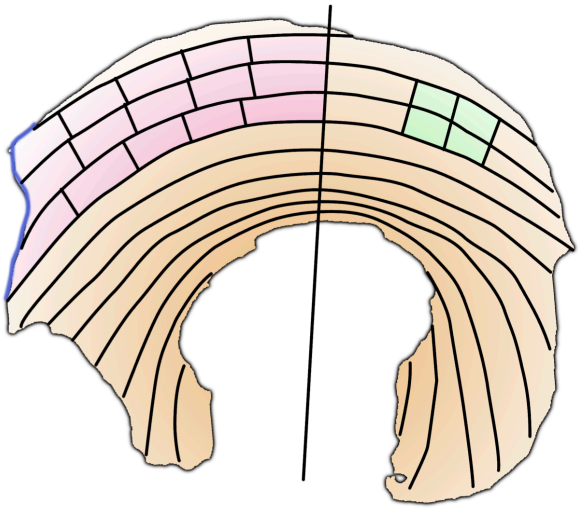

Figure 4.2: A flattened imaging slice and ROI map of acetabular cartilage.

The data are summarized in Table 4.2. Note that the T2* values for a given ROI were aggregated by taking the sample mean over all voxels in the ROI, and so the sample quantities given in the table are for samples of sample means. Box plots of T2* by Beck score are shown in Figure 4.3.

Table 4.2: Summaries, by Beck score, of the aggregated T2* values from our data set.

Beck Score	Frequency	Sample First Quartile	Sample Median	Sample Third Quartile
1	172	22.1	28.7	33.5
2	160	16.7	21.1	27.6
3	112	16.1	20.0	23.9
4	59	14.8	18.9	22.7
5	21	13.8	17.3	19.5
6	8	14.2	16.4	18.5

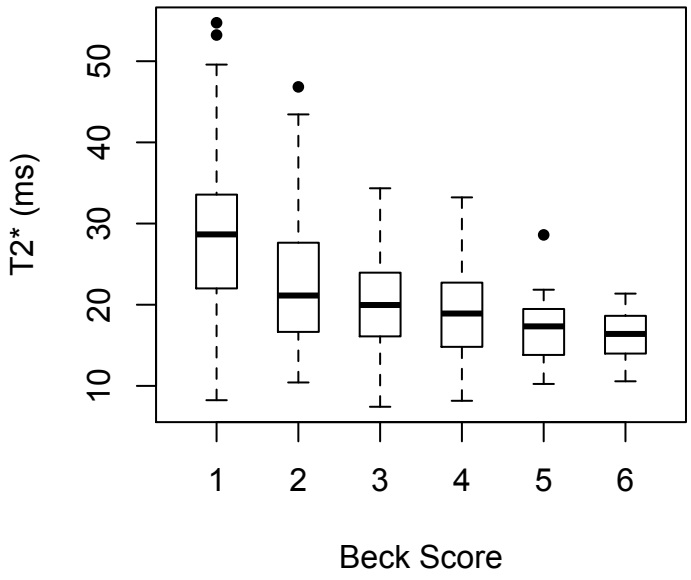

Figure 4.3: Box plots of aggregated T2* values by Beck score.

4.2 Dirichlet Process Prior Distribution

4.2.1 Development and Validation of the Predictive Models

We posit Beck scores as ordinal multinomial outcomes that satisfy the proportional odds assumption (Agresti, 2012). Specifically, we suppose that

$$\begin{aligned} \gamma_k(\boldsymbol{x}_{ij}) &= \mathbb{P}(Y_{ij} \leq k \mid \boldsymbol{x}_{ij}) = \frac{\exp(\alpha_k - \boldsymbol{\beta}'\boldsymbol{x}_{ij})}{1+\exp(\alpha_k - \boldsymbol{\beta}'\boldsymbol{x}_{ij})}, \quad i = 1,\ldots,28 \; ; \; j = 1,\ldots,19 \; ; \\ & \quad k = 1,\ldots,5, \end{aligned} \tag{4.1}$$

where Y_{ij} is the Beck score for the jth ROI in the ith hip, k is the Beck category, α_k is the intercept (or threshold) for category k, $\boldsymbol{x}_{ij}$ is a vector of predictors associated with ROI ij, and $\boldsymbol{\beta}$ are regression coefficients. The sequence of intercepts α_k is non-decreasing. Equivalently,

$$\text{logit } \gamma_k(\boldsymbol{x}_{ij}) = \alpha_k - \boldsymbol{\beta}'\boldsymbol{x}_{ij} \tag{4.2}$$

or

$$\pi_k(\boldsymbol{x}_{ij}) = \mathbb{P}(Y_{ij} = k \mid \boldsymbol{x}_{ij}) = \begin{cases} \gamma_1(\boldsymbol{x}_{ij}) & k = 1, \\ \gamma_k(\boldsymbol{x}_{ij}) - \gamma_{k-1}(\boldsymbol{x}_{ij}) & k \in \{2, 3, 4, 5\}, \text{ and} \\ 1 - \gamma_5(\boldsymbol{x}_{ij}) & k = 6. \end{cases} \tag{4.3}$$

Having no subscript on $\boldsymbol{\beta}$ implies that the predictors have the same effects for all categories. This assumption and the use of the logit link function together characterize the proportional odds model, the most common of the so-called cumulative link models for ordinal outcomes.

We also assume that Beck scores are dependent within a patient. We took two approaches to implementing this assumption. The first adds a simple random offset to the exponent in the model described in (4.1). The second provides an explicit model for the similarity between spatially proximate observations. It thus more closely mimics the physical conditions at the site of interest. This model additionally incorporates a Dirichlet process to describe similarities in the spatial association from hip to hip, which results in shrinkage of their estimates toward common cluster means. We elaborate on these two models in the next three subsections. All analyses were performed in R version 2.15.2 (R Core Team, 2014).

Random Intercept Model

First, as described in Iisakka (2014), we modeled the intra-patient dependence of Beck scores using a random intercept. That is, we assumed the outcomes for a given hip to be independent conditional on a shared random effect:

$$\text{logit } \gamma_k(\boldsymbol{x}_{ij}, \mathcal{W}_i') = \alpha_k - \boldsymbol{\beta}'\boldsymbol{x}_{ij} + \mathcal{W}_i', \tag{4.4}$$

where the $\mathcal{W}_i'$ follow a Gaussian distribution located at $\mathbf{0}$. This dependence model, although rudimentary, proved reasonably effective.

In addition to T2*, we included age, weight, and sex as predictors, as well as quadratic and cubic terms for the continuous predictors and two-way interactions among all predictors. Both weight and T2* were centered at their respective mean values. Corresponding quadratic and cubic terms were calculated from these centered values. We fit each candidate model using the function `clmm()` of the R (R Core Team, 2014)

package `ordinal` (Christensen, 2013). Since the random intercepts imply an analytically intractable likelihood (the likelihood contains an integral over $\mathbb{R}^{28}$), the likelihood must be approximated. Function `clmm()` approximates the integral using the Laplace approximation (Laplace *and others*, 1986) or adaptive Gaussian–Hermite quadrature (Abramowitz and Stegun, 2012). The number of quadrature points can be specified using the argument `nAGQ`. We found that ten quadrature points yielded sufficiently accurate maximum likelihood estimates. Two other arguments of interest are `link` and `threshold`. We initially chose the logit link function and flexible thresholds (*i.e.*, the only constraint on the thresholds was that they be strictly increasing).

The coefficient and threshold estimates for the most parsimonious model are given in Table 4.3. The best model among those considered, determined by backward elimination, is quadratic in both T2* and weight, has a significant effect for sex, and has a different effect of T2* for males and females. The inclusion of weight is justified because body mass index is known to be an independent risk factor for FAI (Kumar and Aggarwal, 2011; Clohisy *and others*, 2013). More specifically, a decrease (increase) in T2* (weight) is associated with larger Beck scores, and males have larger Beck scores than females on average. Males are considerably more likely to have cam-type FAI (Byrd, 2010), which is more damaging, and young, active males with FAI tend to have more severe damage (Clohisy and Nepple, 2013; Nawabi *and others*, 2014). The random effects are necessary, for the independence model lacks fit (LRT; p value < 0.001). The estimated standard deviation of the random effects was 1.640, a significant fraction of the effective range for parameter estimates on the logit scale. This implies very high within-hip correlation of observations. We found no evidence against the proportional odds assumption, nor did we find that a different link function or structure for the thresholds (symmetric or equidistant) would be more appropriate.

Figure 4.4 shows the estimated cumulative logits (as functions of centered T2* but without estimates of random effects) for females and males, with patient weights fixed at the sex-specific sample means (151.68 lbs and 180.75 lbs, respectively).

During a recently completed followup study we collected data for an additional 27 hips. We then applied the predictive model to the new data, and found that it performed well. The results are shown in Table 4.4.

We also combined the two data sets and repeated our analyses with the model

Table 4.3: Coefficient and threshold estimates for the most parsimonious mixed-effects proportional odds model.

Covariate	Estimate	P Value
T2*	-0.146	< 0.001
$(T2^*)^2$	-0.005	0.002
Weight	0.015	0.197
Weight^2	0.001	0.005
Sex (Male)	1.534	0.050
T2* $\times$ Sex	-0.089	0.003
Threshold	Estimate	95% CI
α_1	-0.517	(-1.538, 0.503)
α_2	1.912	(0.8774, 2.947)
α_3	4.037	(2.953, 5.122)
α_4	6.626	(5.349, 7.903)
α_5	8.636	(7.130, 10.141)

Table 4.4: Out-of-sample predictive performance. Prediction errors are given in Beck units.

Prediction Error	Percentage
0	29%
-1, 0, 1	76%
-2, -1, 0, 1, 2	95%

of Table 4.3. We found that the parameter estimates changed little, which suggests that our initial sample was probably large enough to yield stable estimates of model parameters.

Spatial Modeling of Beck Scores

With this data set, we may reasonably expect that the shorter the distance between measurement points, the more alike the measurements. Mathematically relating neighboring observations ensures smooth, continuous transition along the surface of interest. A spatial model describes the association between such observations, and we believe it is well-suited to this case.

In this data set, we have aggregated the T2* predictor over all the voxels in the ROI, and the evaluator-assessed Beck score applies to the ROI as a whole. An areal

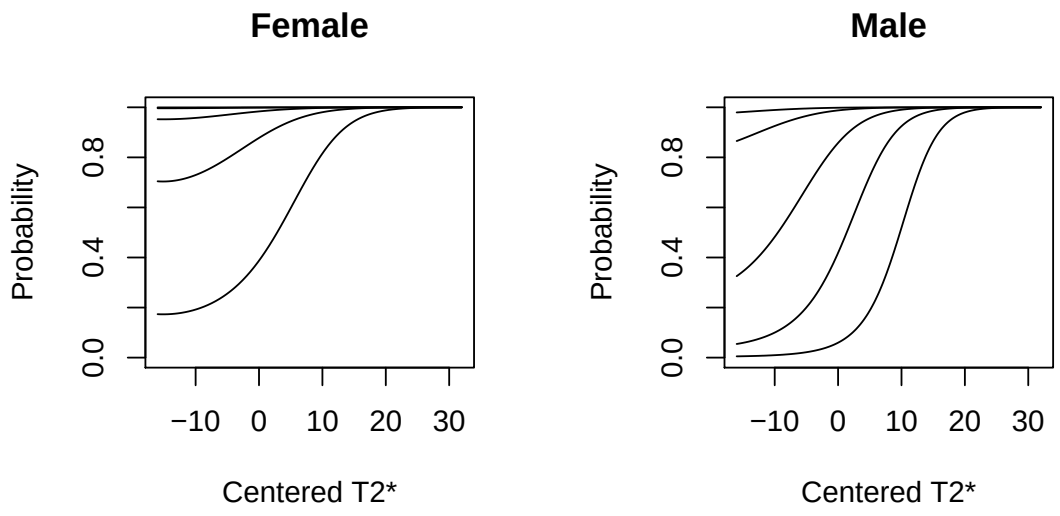

Figure 4.4: Estimated cumulative logits for females and males, as functions of centered T2* and with patient weights fixed at the sex-specific means. Points on the plots correspond to γ_k values from (4.1) for a given value of centered T2*. For example, for females at the mean T2* value (centered T2* = 0), γ_1 through γ_5 values are 0.387, 0.878, 0.984, 0.999, and 1.000 respectively.

model describes the association between data summarized over areas of this sort. For each hip, we use the copCAR model (Hughes, 2015) described in Subsection 3.4.1 with the proportional odds model of (4.1) for the marginal distributions.

Specifically, the precision matrix Q is block diagonal, one block for each hip of the form

$$\begin{aligned} q_{i,gg} &= \text{a region's number of neighbors} \\ q_{i,gh} &= \begin{cases} -\xi_i \text{ if regions } g \text{ and } h \text{ are neighbors} \\ 0 \text{ otherwise,} \end{cases} \\ \text{with } \xi_i &= \text{Association parameter for hip } i. \end{aligned}$$

We analyze it in a Bayesian framework using the composite likelihood with curvature correction described in Subsection 3.3.4.

In our analysis of Beck score data, we count as neighbors to an ROI those ROIs

immediately anterior and posterior to the ROI as well as those ROIs in the same position in adjacent sagittal sections, a total of four neighbors for observations in the interior of the measurement region. A trace on Figure 4.1 in Chapter 4 of a neighborhood so defined would take the shape of a "plus" sign. Clearly, other definitions of neighborhoods are possible as well, for example, one whose trace on Figure 4.1 would form the shape of a square, encompassing as many as eight neighbors.

The association parameters ξ_i take values in $[0, 1)$. The form of the curvature correction has the potential to place adjusted association parameters outside this range. To avoid this difficulty, in the bootstrap used to compute the curvature-correction matrix and in the Markov Chain Monte Carlo (MCMC) routine used for a Bayesian analysis, we generated values for the association parameter on a transformed scale, described below, instead of on the original scale.

We analyzed this model in a Bayesian framework using the Metropolis-Hastings algorithm. We re-parameterized the thresholds α_k for $k > 1$ to the form $\Delta_k = \alpha_k - \alpha_{k-1}$ to ensure, in conjunction with the prior and proposal distributions, that the corresponding α_k parameter proposals are non-decreasing. We applied a $\mathcal{N}\left(0, 10^6\right)$ prior distribution to α_1 and the β parameters, and we applied a Gamma(2,1) prior distribution to the Δ_k terms. The parameters of these gamma prior distributions place most of the mass in the range (0, 4), corresponding to results from earlier analyses. To propose α_1 and β values, we used random walks with normal distributions centered at the most recently accepted value. For the Δ_k parameters, we first proposed values using normal distributions centered at the log of the most recently accepted value. Then we exponentiated the values to transform them to Δ_k parameter proposals.

We used the same predictors as in the most parsimonious model listed in Table 4.3 from Section 4.2.1 in order to compare the two approaches. The prior distributions on these fixed effects were diffuse normal distributions centered at zero. We use random walks with normal distributions to propose parameter values.

For all random walk proposals, we used the variance-covariance sub-matrices from a proportional odds model fit without random intercepts. We scaled these sub-matrices to obtain acceptance rates in the range of 20%–50%.

Clustering of Spacial Association Parameters

In a preliminary spatial analysis of this data set, Iisakka (2014) found that a model with a single association parameter for all hips yielded implausible parameter estimates, whereas allowing a different association parameter for each hip resulted in a model for which the analysis was numerically unstable. A Dirichlet process prior distribution provides a compromise between these two extremes, allowing association parameter values to cluster. This is one mechanism by which the estimation process for a given hip's association parameter can "borrow strength" from the estimation for other hips.

We envision the association parameters, one for each hip, as arising from several members of a family of distributions. A Dirichlet process is a prior distribution for the mixing distribution of these family members. We therefore can think of it as a "distribution of distributions." Using it obviates the need to know *a priori* how many members of the family to use.

One characterization for the Dirichlet process is as a generalization of a beta distribution. A beta distribution can model proportions, the distribution of propensities to be in one of two categories. The Dirichlet distribution generalizes this idea to a mixture among t categories. The Dirichlet process generalizes further. It is the form resulting when $t \to \infty$. Each realization of the Dirichlet process is a Dirichlet distribution, discrete with probability one (Ferguson, 1973). As Escobar and West (1995) point out, there is a positive probability of coincident values in each realization of the Dirichlet process. These characteristics enable the Dirichlet process to model an arbitrary number of clusters. Each realization contains a finite number of elements, describing (non-zero) proportions in each of a number of categories not specified ahead of time. Sufficient sampling from a given realization inevitably results in repeats of values previously obtained. Such sampled values thus form clusters in particular categories.

The number of categories with non-zero proportions can change from realization to realization. Each category can be a parameter value. Alternately, each category can be an identifier for a member in a family of distributions, with the implication that each realization of the Dirichlet process describes a mixture of these distributions.

The Dirichlet process is specified by two attributes, a base distribution and a concentration parameter. As Neal (2000) writes, the Dirichlet process prior distribution

mathematically has the following form:

$$\begin{aligned} y_i \mid \theta_i &\sim F(\theta_i), \\ \theta_i \mid \mathcal{G} &\sim \mathcal{G}, \\ \mathcal{G} &\sim \mathcal{DP}(\mathcal{G}_0, \boldsymbol{a}), \end{aligned}$$

where $\mathcal{G}_0$ is the base distribution and $\boldsymbol{a}$ is the concentration parameter. The base distribution represents the prior expectation of the specific realization $\mathcal{G}$. The concentration parameter describes how closely the mass of a given realization concentrates around this expectation. For a measurable set $\mathcal{A}$, as $\boldsymbol{a} \to \infty$, $\mathcal{G}(\mathcal{A}) \to \mathcal{G}_0(\mathcal{A})$ (Teh, 2010). The smaller the concentration parameter, the fewer the clusters there likely are to be among the parameters of interest. (Ferguson, 1983)

Since we performed the MCMC analysis using a transformation of ξ_i, $v_{\xi_i} = \Phi^{-1}(\xi_i)$ instead of ξ_i, the specific choice of base distribution is arbitrary. In this analysis, we used a $\mathcal{N}(0.85, 0.0225)$ distribution as the base distribution. The concentration parameter's prior distribution was a Gamma(4,1), which applies most of the mass in the range (0, 10). Preliminary studies in which we fixed the concentration parameter to single-digit integer values showed, through the acceptance rates, that the data favor low values of $\boldsymbol{a}$. See Neal (2000) and Escobar and West (1995) for details on the Markov chain sampling methods.

Several metrics informed our assessment of the MCMC routine's convergence. They include stability of parameter estimate as a function of sample size, the ratio of the MCMC standard error to the estimated posterior mean, visual inspection of trace plots, qualitative assessment of the empirical distribution for spurious peaks, and review of acceptance rates.

For this analysis, we ran the chain for 120,000 iterations. With independent prior distributions, Markov chains for certain association parameters struggled to propose acceptable candidate values. With the Dirichlet Process prior distribution, all estimated parameters showed excellent acceptance rates.

Results appear in Table 4.5. Note that we constructed the α_2 to α_5 values from α_1 and Δ_k results at each iteration of the MCMC analysis. Means presented are the posterior means from the resulting vectors of values.

In this analysis, only centered T2* and the square of centered weight have intervals

that exclude zero, though the latter only barely so. The interpretation of the estimates is the same as with the random intercept model; a decrease in T2* is associated with higher Beck scores, as is an increase in the squared centered weight. The estimated α_k values were somewhat lower than their counterparts in the random intercept model, likely reflecting the smaller effect size for centered T2* estimated here. The interval widths are shorter or the same for this analysis as compared with the random intercept model. A side study showed that the estimates for α_1 values in particular changed if the number of neighbors for each observation was increased by changing the definition of neighbor pairs.

The posterior mean for the concentration parameter is approximately 1, indicating that the data strongly favor realizations featuring few distinct values of the association parameter. Indeed, the mean of the number of categories in each iteration was approximately 1.5, though the exact association parameter values in use changed from iteration to iteration.

All estimated association parameters were at least as large as 0.965. A review of Beck scores on a hip-by-hip basis showed very strong association among scores within a given hip. In some cases, all measurement locations showed the same Beck score. Large association parameter values clearly are appropriate for this data. The large standard deviation in random effects indicated this as well, as discussed in Section 4.2.1.

As will be seen in the next section, while the specific effect sizes differ somewhat between the two models, both models seem to fit the data reasonably well. The spatial model, since it accounts for the influence of neighbors of the observations, is the more intuitive model.

4.2.2 Application to Patient T2* Data

In this section we will present and interpret two maps of predicted Beck scores for one patient from the study.

Figure 4.5 shows the patient's predicted Beck map using the fits of the random intercept model and the spatial association model. We chose this color scheme for two reasons: (1) people with color vision deficiency can distinguish these colors, and (2) the colors for adjacent Beck values are sufficiently different that one can easily distinguish the various Beck values, whereas a gradient would make it difficult to distinguish the

Table 4.5: Coefficient and threshold estimates for areal model with Dirichlet Process prior distribution.

Covariate	Posterior Mean	95% Highest Posterior Density Interval
T2*	-0.0534	(-0.0819, -0.0249)
$(T2^*)^2$	-0.0014	(-0.0033, 5e-04)
Weight	0.0113	(-0.0037, 0.0255)
Weight^2	6.522E-4	(2.241E-4, 1.095E-3)
Sex (Male)	0.7403	(-0.2668, 1.759)
T2* $\times$ Sex	-0.0366	(-0.0841, 0.0086)
Threshold	Posterior Mean	95% Highest Posterior Density Interval
α_1	-0.1145	(-0.7865, 0.5347)
α_2	1.4357	(0.7489, 2.1273)
α_3	2.8555	(1.974, 3.7099)
α_4	4.6793	(3.4184, 6.0124)
α_5	6.0494	(4.4793, 7.6323)

values. We see from the map that this patient's cartilage damage is significant in several regions, with large bands of red and blue present. There are, however, two large regions of relatively healthy tissue, colored in black and white. We see mostly 1s and 2s in those regions, which suggests that the tissue there exhibits only minor damage. The spatial association model's prediction is slightly more conservative than that of the random intercept model in the sense that the predicted Beck scores are lower in some regions. Both models broadly indicate the same regions as damaged, providing confirmation as to the suitability of both models.

4.3 Discussion

Clinical MR imaging has significant limitations for providing the orthopaedic surgeon with the most relevant information on cartilage integrity. The limitations are threefold. First is the inadequate visualization of a thin, approximately spherical 3D structure by an imaging format comprising consecutive two-dimensional slices. Even when data are acquired with three-dimensional sampling, the resulting images occupy a 3D Cartesian grid composed of imaging planes. The obliquity of planes for viewing can be variable, but the visualization of the cartilage is possible only when such planes intersect with

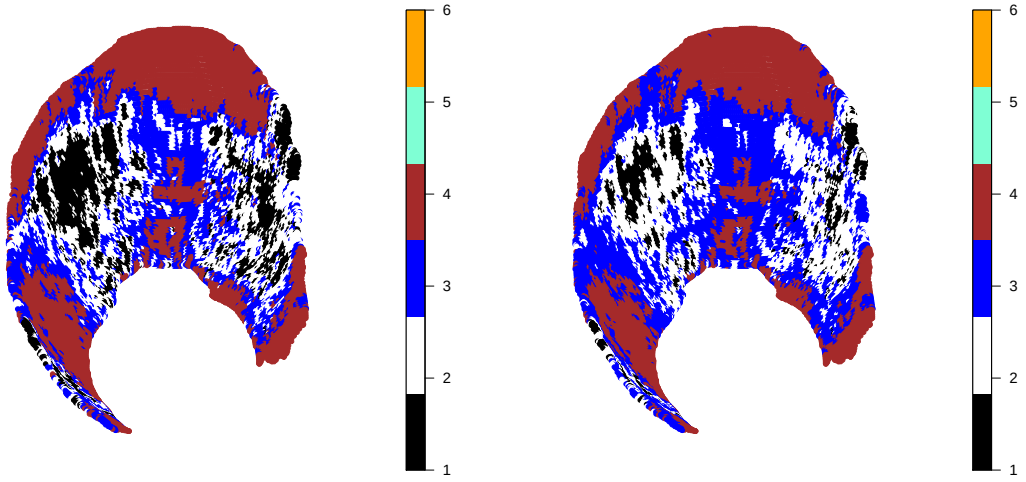

Figure 4.5: Predicted Beck maps of the patient's acetabulum (a) random intercept model (b) spatial association model

the approximately spherical shape. Second, the grayscale intensities in clinical images do not provide information for quantitative cartilage assessment. Finally, such clinical images do not predict probability of disease for therapy stratification.

In this work we developed two predictive models useful for assessing cartilage damage in femoroacetabular impingement. Given a patient's sex, weight, and T2* data, our models can be used to produce predictive maps of the acetabulum. The maps provide a non-invasive means of assessing the pattern, degree, and extent of articular cartilage damage, which can help clinicians to decide among non-operative management, joint preservation surgery, or joint replacement in FAI. The spatial model in particular facilitates representation of the physical configuration of the data and allows estimation of parameters to "borrow strength" across participants. It is more conservative than the random intercept model.

The Dirichlet process is an especially strong candidate for a prior distribution for inference on the association parameters in this data set. Few continuous distributions can be parameterized intuitively such that the majority of the mass is clustered in the

range just smaller than 1. Some hips had log-likelihood values that dropped off steeply as the association parameter value decreased, whereas for others, the decline was nearly flat. The implication of this shape difference is that it would be difficult to find a single proposal distribution that would yield good MCMC acceptance rates for all hips.

One potential way to redress this difficulty would be to define different proposal distributions for different groups of hips. But the small data set provides little information to suggest natural clustering. The Dirichlet process, by contrast, allows clusters to emerge from the data. Moreover, the mechanism by which values are sampled from realizations of the Dirichlet process facilitates good MCMC acceptance rates across all hips despite this difference, providing more flexibility in choice of base distribution.

We also can develop analogous models for dichotomized outcomes, *i.e.*, diseased or not. Such models necessarily discard information, namely, the degree of articular cartilage damage. For that reason, we recommend the full models presented here.

Several aspects of the study constrain the ability of the composite log-likelihood and the copCAR model to capture fully the variation in the acetabulum. Twenty-eight hips comprise this data set. Estimates via composite log-likelihood are consistent as the sample size becomes large. This data set may not have sufficient replicates to yield good estimates via composite likelihood.

Second, each hip has nineteen measurement locations. Having a relatively small number of spatial locations in the grid limits the achievable resolution for spatial effects. Moreover, each spatial measurement represents the aggregation of numerous raw T2* values. Such averaging also reduces variability. The ROIs were defined to match the measurement resolution of the flexible probe, which is 2 mm in diameter. This physical dimension imposes a coarseness on the resolution of the response variable.

Finally, as described in Section 4.1.2, data collection focused on a small section of the antero-superior acetabulum in which damage commonly is reported. This practice tended to yield relatively homogeneous Beck scores. With little variation in response, it is hard to estimate well the effect of predictors.

Chapter 5

Conclusion

Few reports exist in the literature concerning multiple maxima, especially for mixed linear models. We have shown how a restricted likelihood can be represented as a GLM with a gamma distribution and identity link, and an informative prior distribution for the error variance. The components of this model, the "data" and the free terms, which have the form of a prior distribution on the error variance, can create multiple maxima when they conflict with one another. The same phenomenon occurs when the "data" conflict with an explicit prior distribution on the second variance component.

In the posterior distribution, the global maximum need not encompass more volume than the secondary maximum. This means that the Gibbs sampler might draw from the secondary maximum more frequently. Additionally, the multiple maxima may not manifest under reparameterization of the variance components.

Modelers should choose data models and prior distributions carefully as well as scrutinize posterior distributions to minimize the chance of being misled by peak-finding algorithms.

For Bayesian inference with the Gaussian copula regression models of non-continuous outcomes discussed in Chapter 3, the distributional transform with curvature correction performs well in most circumstances, the exception being binary data. With such data, composite likelihood is the only viable alternative due to the fact that binary data is especially uninformative about the parameters. The substantially faster running times for the distributional transform with curvature correction as compared to other techniques make it particularly attractive in suitable settings. This is the case especially

at high levels of association in which a long jitter vector would be required for the continuous extension method. The continuous extension, while conceptually the better approach due to its use of an exact form for the likelihood, in practice often falls short due to computing resource limitations. Moreover, it may not produce intervals with good frequentist properties even with sufficiently long jitter vectors. While the shape of its posterior distribution generally is not the same as that for continuous extension, the distributional transform with curvature correction is a viable alternative for parameter inference.

When a data set favor very strong spatial association, such as the one in Chapter 4, the Dirichlet process is an especially good candidate for a prior distribution for the distribution of association parameter values. Few continuous distributions can be parameterized intuitively in this range. The fact that realizations of the Dirichlet process are discrete allows high association parameter values to emerge without the need to fine-tune a prior distribution for high values. The clustering property of the Dirichlet process facilitates high acceptance rates even when the data in different sub-groups favor different association parameter values. This attribute can be especially attractive to clinicians seeking to group patients into treatment regimes based on current disease state. This property additionally provides more flexibility in choice of base distribution.

References

Abramowitz, Milton and Stegun, Irene A. (2012). *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables*. Courier Dover Publications.

Adler, Daniel and Murdoch, Duncan. (2013). *rgl: 3D visualization device system (OpenGL)*. R package version 0.93.935.

Agresti, Alan. (2012). *Categorical Data Analysis*, Third edition. Wiley.

Allen, D, Beaulé, PE, Ramadan, O and Doucette, S. (2009). Prevalence of associated deformities and hip pain in patients with cam-type femoroacetabular impingement. *Journal of Bone & Joint Surgery, British Volume* **91**(5), 589–594.

Anderson, Lucas A, Peters, Christopher L, Park, Brandon B, Stoddard, Gregory J, Erickson, Jill A and Crim, Julia R. (2009). Acetabular cartilage delamination in femoroacetabular impingementrisk factors and magnetic resonance imaging diagnosis. *The Journal of Bone & Joint Surgery* **91**(2), 305–313.

Apprich, S, Mamisch, TC, Welsch, GH, Bonel, H, Siebenrock, KA, Kim, Y-J, Trattnig, S and Dudda, M. (2012). Evaluation of articular cartilage in patients with femoroacetabular impingement (FAI) using T2* mapping at different time points at 3.0 Tesla MRI: a feasibility study. *Skeletal Radiology* **41**(8), 987–995.

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

Bates, Douglas and Maechler, Martin. (2014). *Matrix: Sparse and Dense Matrix Classes and Methods*. R package version 1.1-4.

Beck, M, Kalhor, M, Leunig, M and Ganz, R. (2005). Hip morphology influences the pattern of damage to the acetabular cartilage femoroacetabular impingement as a cause of early osteoarthritis of the hip. *Journal of Bone & Joint Surgery, British Volume* **87**(7), 1012–1018.

Beck, Martin, Leunig, Michael, Parvizi, Javad, Boutier, Vincent, Wyss, Daniel and Ganz, Reinhold. (2004). Anterior femoroacetabular impingement: part ii. midterm results of surgical treatment. *Clinical Orthopaedics and Related Research* **418**, 67–73.

Bengtsson, Henrik. (2015). *matrixStats: Methods that Apply to Rows and Columns of Matrices (and to Vectors)*. R package version 0.14.2.

Bittersohl, Bernd, Hosalkar, Harish S, Hughes, Tim, Kim, Young-Jo, Werlen, Stefan, Siebenrock, Klaus A and Mamisch, Tallal C. (2009). Feasibility of T2* mapping for the evaluation of hip joint cartilage at 1.5T using a three-dimensional (3D), gradient-echo (GRE) sequence: A prospective study. *Magnetic Resonance in Medicine* **62**(4), 896–901.

Bittersohl, B, Miese, FR, Hosalkar, HS, Herten, M, Antoch, G, Krauspe, R and Zilkens, C. (2012a). T2* mapping of hip joint cartilage in various histological grades of degeneration. *Osteoarthritis and Cartilage* **20**(7), 653–660.

Bittersohl, Bernd, Miese, Falk R, Hosalkar, Harish S, Mamisch, Tallal C, Antoch, Gerald, Krauspe, Rüdiger and Zilkens, Christoph. (2012b). T2* mapping of acetabular and femoral hip joint cartilage at 3T: a prospective controlled study. *Investigative Radiology* **47**(7), 392–397.

Blankenbaker, Donna G, Ullrick, Steven R, Kijowski, Richard, Davis, Kirkland W, De Smet, Arthur A, Shinki, Kazuhiko, Muñoz del Rio, Alejandro and Keene, James S. (2011). MR arthrography of the hip: comparison of IDEAL-SPGR volume sequence to standard MR sequences in the detection and grading of cartilage lesions. *Radiology* **261**(3), 863–871.

Byrd, JW and Jones, Kay S. (2011). Arthroscopic management of femoroacetabular

impingement: minimum 2-year follow-up. *Arthroscopy: The Journal of Arthroscopic & Related Surgery* **27**(10), 1379–1388.

Byrd, JW Thomas. (2010). Cam-type femoroacetabular impingement. In: *AANA Advanced Arthroscopy: the Hip: Expert Consult: Online, Print and DVD*. Elsevier Health Sciences, p. 65.

Carlisle, John C, Zebala, Lukas P, Shia, Derek S, Hunt, Devyani, Morgan, Patrick M, Prather, Heidi, Wright, Rick W, Steger-May, Karen and Clohisy, John C. (2011). Reliability of various observers in determining common radiographic parameters of adult hip structural anatomy. *The Iowa Orthopaedic Journal* **31**, 52.

Casella, George and Berger, Roger L. (2002). *Statistical Inference*, Second edition. Duxbury Thomson Learning.

Chandler, Richard E and Bate, Steven. (2007). Inference for clustered data using the independence loglikelihood. *Biometrika* **94**(1), 167–183.

Christensen, R. H. B. (2013). ordinal—regression models for ordinal data. R package version 2013.9-30 http://www.cran.r-project.org/package=ordinal/.

Clohisy, John C, Baca, Geneva, Beaulé, Paul E, Kim, Young-Jo, Larson, Christopher M, Millis, Michael B, Podeszwa, David A, Schoenecker, Perry L, Sierra, Rafael J, Sink, Ernest L *and others*. (2013). Descriptive epidemiology of femoroacetabular impingement a north american cohort of patients undergoing surgery. *The American Journal of Sports Medicine* **41**(6), 1348–1356.

Clohisy, John C, Carlisle, John C, Beaulé, Paul E, Kim, Young-Jo, Trousdale, Robert T, Sierra, Rafael J, Leunig, Michael, Schoenecker, Perry L and Millis, Michael B. (2008). A systematic approach to the plain radiographic evaluation of the young adult hip. *The Journal of Bone & Joint Surgery* **90**(Supplement_4), 47–66.

Clohisy, John C, Carlisle, John C, Trousdale, Robert, Kim, Young-Jo, Beaule, Paul E, Morgan, Patrick, Steger-May, Karen, Schoenecker,

Perry L and Millis, Michael. (2009). Radiographic evaluation of the hip has limited reliability. *Clinical Orthopaedics and Related Research* **467**(3), 666–675.

Clohisy, John C and Nepple, Jeffrey J. (2013). Symptomatic femoroacetabular impingement are there gender-specific disease characteristics? *Orthopaedic Journal of Sports Medicine* **1**(4 suppl), 2325967113S00054.

de Backer, M., de Keyser, P., de Vroey, C. and Lesaffre, E. (1996). A 12-week treatment for dermatophyte toe onychomycosis: Terbinafine 250 mg/day vs. itraconazole 200 mg/day—a double-blind comparitive trial. *British Journal of Dermatology* **134**, 16–17. Supplement 46.

Denuit, Michel and Lambert, Philippe. (2005). Constraints on concordance measures in bivariate discrete data. *Journal of Multivariate Analysis* **93**, 40–57.

Dudda, Marcel, Albers, Christoph, Mamisch, Tallal Charles, Werlen, Stefan and Beck, Martin. (2009). Do normal radiographs exclude asphericity of the femoral head-neck junction? *Clinical Orthopaedics and Related Research* **467**(3), 651–659.

Escobar, Michael D and West, Mike. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* **90**(430), 577–588.

Ferguson, Thomas S. (1973). A bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 209–230.

Ferguson, Thomas S. (1983). Bayesian density estimation by mixtures of normal distributions. *Recent Advances in Statistics* **24**(1983), 287–302.

Furrer, Reinhard and Sain, Stephan R. (2010). spam: A sparse matrix R package with emphasis on MCMC methods for Gaussian Markov random fields. *Journal of Statistical Software* **36**(10), 1–25.

Ganz, Reinhold, Leunig, Michael, Leunig-Ganz, Katharina and Harris, William H. (2008). The etiology of osteoarthritis of the hip. *Clinical Orthopaedics and Related Research* **466**(2), 264–272.

Ganz, Reinhold, Parvizi, Javad, Beck, Martin, Leunig, Michael, Nötzli, Hubert and Siebenrock, Klaus A. (2003). Femoroacetabular impingement: a cause for osteoarthritis of the hip. *Clinical Orthopaedics and Related Research* **417**, 112–120.

Genz, Alan, Bretz, Frank, Miwa, Tetsuhisa, Mi, Xuefei, Leisch, Friedrich, Scheipl, Fabian and Hothorn, Torsten. (2014). *mvtnorm: Multivariate Normal and t Distributions*. R package version 1.0-0.

Gilbert, Paul and Varadhan, Ravi. (2012). *numDeriv: Accurate Numerical Derivatives*. R package version 2012.9-1.

Gold, Garry E, Chen, Christina A, Koo, Seungbum, Hargreaves, Brian A and Bangerter, Neal K. (2009). Recent advances in MRI of articular cartilage. *American Journal of Roentgenology* **193**(3), 628.

Gold, Stephanie L, Burge, Alissa J and Potter, Hollis G. (2012). MRI of hip cartilage: joint morphology, structure, and composition. *Clinical Orthopaedics and Related Research* **470**(12), 3321–3331.

Guanche, Carlos A, Chan, Keith A, Conner, Chad A and Sikka, Robby S. (2012). Paper 4: Arthroscopic treatment cam-type hip impingement lesions with 32.8-month mean follow-up. *Arthroscopy: The Journal of Arthroscopic & Related Surgery* **28**(6), e45.

Haran, Murali and Hughes, John. (2012). *batchmeans: Consistent Batch Means Estimation of Monte Carlo Standard Errors*. Minneapolis, MN. R package version 1.0-1.

Henn, Lisa and Hodges, James S. (2014). Multiple local maxima in restricted likelihoods and posterior distributions for mixed linear models. *International Statistical Review* **82**(1), 90–105. Copyright ©2014, *International Statistical Review*, International Statistical Institute, and Blackwell Publishing.

Hodges, James S. (1998). Some algebra and geometry for hierarchical models applied to diagnostics. *Journal of the Royal Statistical Society, Series B* **60**(3), 497–536.

Hughes, John. (2015). copCAR: A flexible regression model for areal data. *Journal of Computational and Graphical Statistics* (just-accepted), 00–00. Accepted author version posted online 31 Jul 2014.

Iisakka, Eleena. (2014). Magnetic resonance image segmentation for cartilage assessment in femoroacetabular impingement. Master of Science Plan B Report, Division of Biostatistics, University of Minnesota, Minneapolis, MN.

Kazianka, Hannes. (2013). Approximate copula-based estimation and prediction of discrete spatial data. *Stochastic Environmenal Research and Risk Assessment* **27**, 2015–2026.

Kazianka, Hannes and Pilz, Jürgen. (2010). Copula-based geostatistical modeling of continuous and discrete data including covariates. *Stochastic Environmental Research and Risk Assessment* **24**, 661–673.

Kim, Young-Jo, Jaramillo, Diego, Millis, Michael B, Gray, Martha L and Burstein, Deborah. (2003). Assessment of early osteoarthritis in hip dysplasia with delayed gadolinium-enhanced magnetic resonance imaging of cartilage. *The Journal of Bone & Joint Surgery* **85**(10), 1987–1992.

Kumar, Rajesh and Aggarwal, Aditya. (2011). Femoroacetabular impingement and risk factors: a study of 50 cases. *Orthopaedic Surgery* **3**(4), 236–241.

Laplace, Pierre Simon *and others*. (1986). Memoir on the probability of the causes of events. *Statistical Science* **1**(3), 364–378.

Larson, Christopher M and Giveans, M Russell. (2009). Arthroscopic debridement versus refixation of the acetabular labrum associated with femoroacetabular impingement. *Arthroscopy: The Journal of Arthroscopic & Related Surgery* **25**(4), 369–376.

Larson, Christopher M, Giveans, M Russell and Taylor, Mehul. (2011). Does arthroscopic FAI correction improve function with radiographic arthritis? *Clinical Orthopaedics and Related Research* **469**(6), 1667–1676.

Lattanzi, R, Petchprapa, C, Glaser, C, Dunham, K, Mikheev, AV, Krigel, A, Mamisch, TC, Kim, Y-J, Rusinek, H and Recht, M. (2012). A new method to analyze dGEMRIC measurements in femoroacetabular impingement: preliminary validation against arthroscopic findings. *Osteoarthritis and Cartilage* **20**(10), 1127–1133.

Leunig, Michael, Beaulé, Paul E and Ganz, Reinhold. (2009). The concept of femoroacetabular impingement: current status and future perspectives. *Clinical Orthopaedics and Related Research* **467**(3), 616–622.

Lindsay, Bruce G. (1988). Contemporary mathematics volume 80, 1988. In: *Statistical Inference from Stochastic Processes: Proceedings of the Ams-IMS-Siam Joint Summer Research Conference Held August 9-15, 1987, with Support from the National Science Foundation and the Army Research Office*, Volume 80. American Mathematical Soc. pp. 221–239. Chapter title is "Composite Likelihood Methods".

Madsen, L. (2009). Maximum likelihood estimation of regression parameters with spatially dependent discrete data. *Journal of Agricultural, Biological, and Environmental Statistics* **14**(4), 375–391.

Madsen, L. and Fang, Y. (2011, September). Joint regression analysis for discrete longitudinal data. *Biometrics* **67**(3), 1171–1175.

Mamisch, Tallal Charles, Kain, Micheal Sean Hillega, Bittersohl, Bernd, Apprich, Sebastian, Werlen, Stefan, Beck, Martin and Siebenrock, Klaus Arno. (2011). Delayed gadolinium-enhanced magnetic resonance imaging of cartilage (dgemric) in femoacetabular impingement. *Journal of Orthopaedic Research* **29**(9), 1305–1311.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. Chapman & Hall/CRC.

Molenberghs, Geert and Verbeke, Geert. (2005). *Models for Discrete Longituidnal Data*. Springer Science+Business Media.

Mosher, Timothy J and Dardzinski, Bernard J. (2004). Cartilage MRI T2 relaxation time mapping: overview and applications. In: *Seminars in Musculoskeletal Radiology*, Volume 8. pp. 355–368.

Nawabi, Danyal H, Bedi, Asheesh, Tibor, Lisa M, Magennis, Erin and Kelly, Bryan T. (2014). The demographic characteristics of high-level and recreational athletes undergoing hip arthroscopy for femoroacetabular impingement: A sports-specific analysis. *Arthroscopy: The Journal of Arthroscopic & Related Surgery* **30**(3), 398–405.

Neal, Radford M. (2000). Markov chain sampling methods for dirichlet process mixture models. *Journal of Computational and Graphical Statistics* **9**(2), 249–265.

Nishii, Takashi, Shiomi, Toshiyuki, Tanaka, Hisashi, Yamazaki, Youichi, Murase, Kenya and Sugano, Nobuhiko. (2010). Loaded cartilage T2 mapping in patients with hip dysplasia 1. *Radiology* **256**(3), 955–965.

Nishii, T, Tanaka, H, Sugano, N, Sakai, T, Hananouchi, T and Yoshikawa, H. (2008). Evaluation of cartilage matrix disorders by T2 relaxation time in patients with hip dysplasia. *Osteoarthritis and Cartilage* **16**(2), 227–233.

Noland, Gregory S, Ayodo, George, Abuya, Jackson, Hodges, James S, Rolfes, Melissa AR and John, Chandy C. (2012). Decreased prevalence of anemia in highland areas of low malaria transmission after a 1-year interruption of transmission. *Clinical Infectious Diseases* **54**(2), 178–184.

Pfirrmann, Christian WA, Duc, Sylvain R, Zanetti, Marco, Dora, Claudio and Hodler, Juerg. (2008). MR arthrography of acetabular cartilage delamination in femoroacetabular cam impingement1. *Radiology* **249**(1), 236–241.

Pfirrmann, Christian WA, Mengiardi, Bernard, Dora, Claudio, Kalberer, Fabian, Zanetti, Marco and Hodler, Juerg. (2006). Cam and pincer femoroacetabular impingement: characteristic MR arthrographic findings in 50 patients. *Radiology* **240**(3), 778–785.

Philippon, MJ, Briggs, KK, Yen, Y-M and Kuppersmith, DA. (2009). Outcomes following hip arthroscopy for femoroacetabular impingement with associated

chondrolabral dysfunction minimum two-year follow-up. *Journal of Bone & Joint Surgery, British Volume* **91**(1), 16–23.

Philippon, Marc, Schenker, Mara, Briggs, Karen and Kuppersmith, David. (2007). Femoroacetabular impingement in 45 professional athletes: associated pathologies and return to sport following arthroscopic decompression. *Knee Surgery, Sports Traumatology, Arthroscopy* **15**(7), 908–914.

Poncet, P. (2012). *modeest: Mode Estimation*. R package version 2.1.

R Core Team. (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Reich, B., Hodges, J. and Zadnick, V. (2006). Effects of residual smoothing on the posterior of the fixed effects in disease mapping models. *Biometrics* **62**(4), 1197–1206.

Reich, Brian J. and Hodges, James S. (2008). Identification of the variance components in the general two-variance linear model. *Journal of Statistical Planning and Inference* **138**, 1592–1604.

Ribatet, M, Cooley, D and Davison, AC. (2012). Bayesian inference for composite likelihood models and an application to spatial extremes. *Statist. Sinica* **22**, 813–845. (doi: 10.5705/ss. 2009.248).

Robert, Christian P and Casella, George. (2004). *Monte Carlo Statistical Methods*, Volume 319. Citeseer.

Rosset, Antoine, Spadola, Luca and Ratib, Osman. (2004). OsiriX: an open-source software for navigating in multidimensional DICOM images. *Journal of Digital Imaging* **17**(3), 205–216.

Rüschendorf, Ludger. (2009). On the distributional transform, sklar's theorem, and the empirical copula process. *Journal of Statistical Planning and Inference* **139**(11), 3921–3927.

Searle, Shayle R., Casella, George and McCulloch, Charles E. (2006). *Variance Components*. John Wiley & Sons, Inc.

Sevcikova, Hana and Rossini, Tony. (2012). *rlecuyer: R interface to RNG with multiple streams*. R package version 0.3-3.

Snow, Greg. (2013). *TeachingDemos: Demonstrations for teaching and learning*. R package version 2.9.

Song, Peter Xue-Kun. (2000, June). Multivariate dispersion models generated from gaussian copula. *Scandinavian Journal of Statistics* **27**(2), 305–320.

Song, Peter X.-K., Li, Mingyao and Yuan, Ying. (2009, March). Joint regression analysis of correlated data using gaussian copulas. *Biometrics* **65**, 60–68.

Teh, Yee Whye. (2010). Dirichlet process. In: *Encyclopedia of Machine Learning*. Springer, pp. 280–287.

Tierney, Luke, Rossini, A. J., Li, Na and Sevcikova, H. (2013). *snow: Simple Network of Workstations*. R package version 0.3-13.

Tönnis, Dietrich, Legal, Helmut, Graf, Reinhard and Telger, Terry C. (1987). *Congenital Dysplasia and Dislocation of the Hip in Children and Adults*, Volume 101. Springer-Verlag Berlin:.

Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S*, Fourth edition. New York: Springer. ISBN 0-387-95457-0.

Wakefield, Jon. (1998). Some algebra and geometry for hierarchical models applied to diagnostics. *Journal of the Royal Statistical Society, Series B* **60**(3), 497–536. Comment on article by James S. Hodges.

Watanabe, Atsuya, Boesch, Chris, Siebenrock, Klaus, Obata, Takayuki and Anderson, Suzanne E. (2007). T2 mapping of hip articular cartilage in healthy volunteers at 3T: a study of topographic variation. *Journal of Magnetic Resonance Imaging* **26**(1), 165–171.

Wedderburn, R. W. M. (1976). On the existence and uniqueness of the maximum likelihood estimates for certain generalized linear models. *Biometrika* **63**(1), 27–32.

Welham, S. J. and Thompson, R. (2009). A note on bimodality in the log-likelihood function for penalized spline mixed models. *Computational Statistics and Data Analysis* **53**, 920–931.

Woodard, Dawn B., Schmidler, Scott C. and Huber, Mark. (2009). Sufficient conditions for torpid mixing of parallel and simulated tempering. *Electronic Journal of Probability* **14**(29), 780–804.

Yu, Hao. (2002). Rmpi: Parallel statistical computing in R. *R News* **2**(2), 10–14.

Zilkens, Christoph, Holstein, Arne, Bittersohl, Bernd, Jäger, Marcus, Haamberg, Tanja, Miese, Falk, Kim, Young-Jo, Mamisch, Tallal C and Krauspe, Rüdiger. (2010). Delayed gadolinium-enhanced magnetic resonance imaging of cartilage in the long-term follow-up after perthes disease. *Journal of Pediatric Orthopaedics* **30**(2), 147–153.

Zlatkin, Michael B, Pevsner, Deborah, Sanders, Timothy G, Hancock, Chris R, Ceballos, Cesar E and Herrera, Mauricio F. (2010). Acetabular labral tears and cartilage lesions of the hip: indirect MR arthrographic correlation with arthroscopy—a preliminary study. *American Journal of Roentgenology* **194**(3), 709–714.

Appendix A

Gaussian Copula Regression Models Markov Chain Monte Carlo Implementation Notes

A.1 Sampling Schemes

In all settings, we sampled ξ and κ (if applicable) individually, and the β vector as a block. In each Metropolis-Hastings scheme described below, $L(\cdot)$ is the likelihood, $q(\cdot)$ is the proposal distribution, and $\pi(\cdot)$ is the prior distribution.

For the copCAR and longitudinal settings, the parameter vector is $\theta = (\beta, \xi)$. For these settings, the acceptance probabilities are as follows:

$$\begin{aligned} \text{probability for } \beta &= \frac{L\left(\beta^*, \xi^{(t-1)} \mid y\right) \pi\left(\beta^*\right)}{L\left(\beta^{(t-1)}, \xi^{(t-1)} \mid y\right) \pi\left(\beta\right)} \times 100\% \\ \text{probability for } \xi &= \frac{L\left(\beta^{(t-1)}, \xi^* \mid y\right) \pi\left(\xi^*\right)}{L\left(\beta^{(t-1)}, \xi^{(t-1)} \mid y\right) \pi\left(\xi^{(t-1)}\right)} \times 100\% \end{aligned}$$

Note that for techniques using curvature correction, the log-likelihood uses the adjusted theta proposal θ^*_{adj} and previous adjusted theta θ_{adj}. All other terms in the acceptance probability expression use the unadjusted theta proposal θ^* or previous unadjusted theta θ.

For the negative binomial setting, the parameter vector is $\theta = \left(\beta, \xi, \kappa\right)$. The acceptance probabilities are as follows:

$$\begin{aligned}
\text{probability for } \beta &= \frac{L\left(\beta^*, \xi^{(t-1)}, \kappa^{(t-1)} \mid y\right)\pi\left(\beta^*\right)}{L\left(\beta^{(t-1)}, \xi^{(t-1)}, \kappa^{(t-1)} \mid y\right)\pi\left(\beta\right)} \times 100\% \\
\text{probability for } \xi &= \frac{L\left(\beta^{(t-1)}, \xi^*, \kappa^{(t-1)} \mid y\right)\pi\left(\xi^*\right)}{L\left(\beta^{(t-1)}, \xi^{(t-1)}, \kappa^{(t-1)} \mid y\right)\pi\left(\xi^{(t-1)}\right)} \times 100\% \\
\text{probability for } \kappa &= \frac{L\left(\beta^{(t-1)}, \xi^{(t-1)}, \kappa^* \mid y\right)\pi\left(\kappa^*\right)}{L\left(\beta^{(t-1)}, \xi^{(t-1)}, \kappa^{(t-1)} \mid y\right)\pi\left(\kappa^{(t-1)}\right)} \times 100\%
\end{aligned}$$

Again when using the curvature correction, the log-likelihood uses the adjusted theta proposal θ^*_{adj} and previous adjusted theta θ_{adj}. All other terms in that expression use the unadjusted theta proposal θ^* or previous unadjusted theta θ.

A.2 Distributional Forms

All combinations of likelihood approaches and copula choice use the prior distribution $\pi\left(\beta\right) = \mathcal{N}\left(0, 10^6 I_p\right)$.

The association parameter ξ takes values in $[0, 1)$ for spatial models—the copCAR simulations and the analysis of Slovenia data—and it takes values in $[-1, 1]$ in the negative binomial and longitudinal settings. The form of the curvature correction described in Subsection 3.3.4 has the potential to place adjusted association parameters outside these ranges. To avoid this difficulty, in the bootstrap used to compute the curvature-correction matrix and in the Markov Chain Monte Carlo (MCMC) routine used for a Bayesian analysis, we generated values for the association parameter on a transformed scale instead of the original scale.

For the spatial models, the transformation is $\upsilon_\xi = \Phi^{-1}\left(\xi\right)$. For negative binomial and longitudinal settings, the transformation is

$$\upsilon_\xi = \Phi^{-1}\left(\frac{\xi + 1}{2}\right).$$

The prior distribution we use for parameters proposed on transformed scales is simply the standard normal distribution. A variable following this distribution on the transformed scale would, by the probability integral transform, follow the uniform distribution on the natural scale of ξ. It is thus well-suited for use as a prior distribution

in this study.

We generated values for the overdispersion parameter κ, used in the negative binomial setting, on a transformed scale as well. The transformation is

$$v_\kappa = \log(\kappa). \tag{A.1}$$

We used the standard normal distribution as the prior distribution for the transformed κ parameter as well, the rationale for which we now present.

Figure A.1 shows a sample from the standard normal distribution transformed in this manner. This prior distribution places most of the mass in the range of $(0.3, 1)$ on the natural scale of κ. This not only agrees well with the simulation settings of $\{0.5, 1\}$, but represents common values of the overdispersion parameter in various applications. More specifically, 50% of its values are less than 1, 94.5% of its values are less than 5, the mean is 1.65, and the mode is approximately 0.33. It thus represents relative equipoise about values understood via prior knowledge to be likely values.

Chandler and Bate (2007) note that the curvature correction "lacks invariance to reparameterization." They advise avoiding excessively skewed log-likelihoods. Our pilot studies with a test case suggest the transformations we employ here perform satisfactorily.

We use Gaussian random walks for all proposals except in the analysis of the Slovenia data. While a random walk can help ensure the chain fully and efficiently explores the parameter space, we found that, as described in Subsection 3.6.1 regarding analysis of the Slovenia data set, it gave rise to a spurious peak near zero in the posterior distribution. Instead, in that case we use a normal distribution centered about the point estimate for ξ found by `optim()`.

For Gaussian random walk proposals, the proposing distribution for the β parameters use the inverse of the observed Fisher information matrix for the corresponding generalized linear model fit without the correlation structure:

$$\Xi = \left(-\nabla^2 \ell_n \left(\hat{\beta}_n\right)\right)^{-1}. \tag{A.2}$$

Different scaling factors were used with Ξ to yield acceptance rates in the range of 20%–50%. The variance in the proposal distributions for transformed κ and transformed ξ also were scaled to tune the acceptance rates.

Table A.1 shows copula particulars and marginal distributions for each model setting.

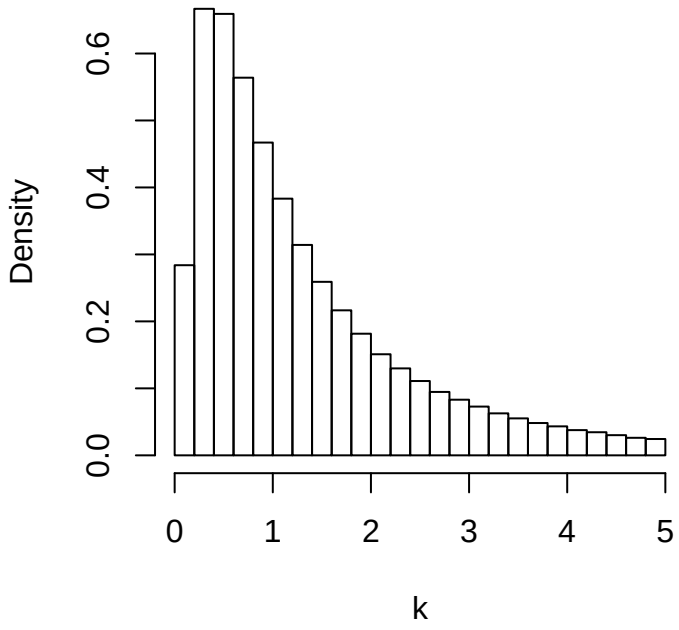

Figure A.1: Sample from the Standard Normal Distribution Transformed to the Natural Scale of k via (A.1)

Table A.1: Marginal Distributions and Copula Parametrization for Use with Sampling Schemes of Section A.1

Setting	σ_i^2	$\mathcal{F}_i$
copCAR	as per Hughes (2015)	$\mathcal{P}oi\left(\lambda_i\right)$
Longitudinal	1	$\mathcal{B}ern\left(\pi_{ij}\right)$
Negative Binomial	1	$\mathcal{N}egBin\left(\mu_i, \kappa\right)$

A.3 Simulation Setup

In each setting, we analyzed 400 simulated datasets for each level of the association parameter as well as, in the case of the negative binomial simulation, each value of the overdispersion parameter κ. To identify an appropriate length for the jitter vector U_i used with the continuous extension method, we compared performance of one analysis of one data set using one U-vector to the performance of 1000 analyses of the same data set using 1000 realizations of the U-vector. We made this comparison with a data set generated at each level of association parameter ξ and each level of overdispersion parameter κ from the negative binomial setting, analyzing it with the maximizer `optim()`. Adding the variance among the parameter estimates, the MCMC variance, to the mean of the variances calculated in each realization yields what Madsen and Fang (2011) calls an adjusted variance. They called the variance from a single realization the unadjusted variance. The ratio of the corresponding standard errors served as a metric to evaluate the length of the jitter vector. The adjusted standard error should not be much larger than the unadjusted standard error.

Table A.2 displays this ratio by true value of association parameter at both the starting jitter vector length of 100 and the final value used in simulation. Higher association parameter values required longer jitter vectors to achieve the same precision. This trend reflects the fact that observations with greater similarity to one another are less informative concerning variability.

Table A.3 shows the same metric in the negative binomial setting at the initial U-vector length of 100 for both $\kappa = 0.5$ and $\kappa = 1$. In most cases, 100 proved to

Table A.2: Ratios of Adjusted Standard Error to Unadjusted Standard Error for Initial and Final Lengths of Jitter Vector U, by True ξ Value for copCAR Setting

True ξ Value	Initial U			Final U		
	# U's	Adj. SE : Unadj. SE		# U's	Adj. SE : Unadj. SE	
0.8	100	β_1	1.004	100	β_1	1.004
		β_2	1.004		β_2	1.004
		ξ	1.001		ξ	1.001
0.975	100	β_1	1.098	800	β_1	1.050
		β_2	1.088		β_2	0.975
		ξ	1.050		ξ	0.978
0.995	100	β_1	1.123	1000	β_1	1.047
		β_2	1.179		β_2	0.963
		ξ	1.042		ξ	0.976

be a sufficient length for the U-vector. The exception is the $\kappa = 0.5$, $\xi = 0.8$ case. But increasing the length of the U-vector to 1000 did nothing to ameliorate this ratio. Longer U-vectors are unwieldy, as even with a jitter vector of length 100, a given case requires 20 hours to run. The $\kappa = 0.5$, $\xi = 0.8$ case was run with a jitter vector length of 100 so that some comparison could be made, even if the implementation this case is suboptimal.

Finally, a jitter vector length of 100 proved to be sufficient for the Slovenia data set. Ratios ranged from 0.997 to 1.002.

Other settings for each simulation are reported in Table A.4 below.

A.4 Convergence

Several metrics informed our assessment of the MCMC routine's convergence. They include stability of parameter estimate as a function of sample size, the ratio of the MCMC standard error to the estimated posterior mean, visual inspection of trace plots, qualitative assessment of the empirical distribution for spurious peaks, and review of acceptance rates.

Table A.3: Ratios of Adjusted Standard Error to Unadjusted Standard Error for Initial Jitter Vector U Length of 100, by True ξ Value for Negative Binomial Setting

True ξ Value	$\kappa = 0.5$		$\kappa = 1$	
	Adj. SE : Unadj. SE		Adj. SE : Unadj. SE	
0.2	β_0	1.000	β_0	1.001
	β_1	0.998	β_1	1.000
	β_2	1.005	β_2	1.000
	β_3	1.000	β_3	0.999
	κ	1.006	κ	1.000
	ξ	1.003	ξ	0.992
0.4	β_0	1.009	β_0	0.993
	β_1	1.007	β_1	0.990
	β_2	1.009	β_2	1.000
	β_3	1.006	β_3	0.997
	κ	1.037	κ	1.004
	ξ	1.062	ξ	1.014
0.8	β_0	1.031	β_0	0.929
	β_1	1.036	β_1	0.988
	β_2	0.996	β_2	0.997
	β_3	1.005	β_3	0.991
	κ	1.180	κ	1.059
	ξ	1.165	ξ	1.068

A.5 Hardware and Software

Three machine configurations supported this work. They are described in Table A.5. Software configurations for each machine appear in Table A.6.

Table A.4: Simulation Settings. Overdispersion parameter κ in Negative Binomial setting has true values of $(0.5, 1)$

Setting	Fixed Effects β	Dependence ξ	Chain Length
copCAR	$(1, 1)$	$(0.8, 0.975, 0.995)$	60,000
Longitudinal	$(-0.5, 1, 0, 1)$	$(0.2, 0.4, 0.8)$	120,000
Negative Binomial	$(3, -0.1, 0, 0.1)$	$(0.2, 0.4, 0.8)$	166,000 for CE, 360,000 for DT/CC
Slovenia Data	N/A	N/A	60,000
Toenail Data	N/A	N/A	60,000

Table A.5: Hardware Configurations

Machine #	Configuration	Purposes
1	64-bit Intel i5-3210M	Software development Some pilot runs Some checks of jitter vector length
2	64-bit Apple running Darwin	Some pilot runs Some checks of jitter vector length Analysis of real data sets
3	HP Linux cluster, Intel Xeon X5560 (Minnesota Supercomputing Institute)	Simulations Post-processing

Table A.6: Software Configurations

		Version		
R Package	Citation	Machine 1	Machine 2	Machine 3
base	R Core Team (2014)	2.15.2	3.1.2	3.1.1
rlecuyer	Sevcikova and Rossini (2012)	0.3-3	0.3-3	0.3-3
Rmpi	Yu (2002)	(N/A)	(N/A)	0.6-5
mvtnorm	Genz *and others* (2014)	0.9-9994	1.0-2	1.0-0
MASS	Venables and Ripley (2002)	7.3-22	7.3-40	7.3-33
snow	Tierney *and others* (2013)	0.3-10	0.3-13	0.3-13
Matrix	Bates and Maechler (2014)	1.0-9	1.1-4	1.1-4
spam	Furrer and Sain (2010)	0.29-2	1.0-1	1.0-1
numDeriv	Gilbert and Varadhan (2012)	2012.9-1	2012.9-1	2012.9-1
TeachingDemos	Snow (2013)	2.9	(N/A)	2.9
batchmeans	Haran and Hughes (2012)	1.0-1	(N/A)	1.0-2
modeest	Poncet (2012)	2.1	(N/A)	2.1
rgl	Adler and Murdoch (2013)	0.93.935	(N/A)	(N/A)
matrixStats	Bengtsson (2015)	(N/A)	(N/A)	0.14.2